Journal of Health Informatics

J Health Inform. 2026;18:e1875
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1875

ARTIGO COMPLETO
CBIS 2026

Classificação de estruturas oculares em retinógrafos portáteis utilizando deep learning para triagem oftalmológica

Classification of ocular structures in portable retinal cameras using deep learning for ophthalmologic screening

Iaze Guilherme Soares Carneiro Santos[1] ✉, Amanda dos Santos Almeida[1], João Dallyson Sousa de Almeida[1], Aristófanes Corrêa Silva[1], Elaine de Paula Fiod Costa[1], Marcos Adriano Garcia Campos[2]

ORCID: 0009-0001-6071-5740 · 0009-0007-0946-3131 · 0000-0001-7013-9700 · 0000-0003-0423-2514 · 0000-0001-9357-2427 · 0000-0001-8924-1203

[1] Universidade Federal do Maranhão, São Luís (MA), Brasil.

[2] Duke University, Durham (NC), Estados Unidos.

FICHA DO ARTIGO

TIPO
Artigo completo · CBIS 2026

IDENTIFICADOR
e1875

DOI
10.59681/2175-4411.v18.iEspecial.2026.1875

RECEBIDO
17/05/2026

APROVADO
20/07/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Iaze Guilherme Soares Carneiro Santos
iaze@nca.ufma.br

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Introdução: A triagem oftalmológica enfrenta desafios em regiões com escassez de especialistas, cenário em que soluções baseadas em telemedicina tornam-se essenciais. Contudo, a captura de imagens por operadores não especializados frequentemente resulta em capturas da estrutura anatômica incorreta ou fora do escopo. **Objetivo:** Propor um modelo de classificação para validar a estrutura anatômica das capturas, categorizando-as em quatro classes: Retina colorida, Retina Red-free, Segmento Anterior e No-eye. **Método:** Foram avaliadas cinco arquiteturas de deep learning para classificação de imagens (MobileNetV3-Large, EfficientNet-B0, DenseNet121, ConvNeXt-Tiny e Vision Transformer (ViT)) utilizando um conjunto de 11.679 imagens. **Resultados:** As redes atingiram acurácia de 99%, porém a MobileNetV3-Large destacou-se pela eficiência, apresentando menor latência (12,5ms) e menor consumo de armazenamento (5,8MB). Ainda, este modelo alcançou 97,15% de acurácia e 97,17% de F1-score macro em datasets usados para validação externa. **Conclusão:** A MobileNetV3-Large apresentou a melhor performance para a aplicação proposta, considerando latência e espaço de memória, aspectos importantes em um cenário de escassez de recursos computacionais.

Descritores: Retinografia portátil; Redes Neurais de Computação; Triagem oftalmológica.

ABSTRACT

Introduction: Ophthalmological screening faces challenges in regions with a shortage of specialists, a scenario in which telemedicine-based solutions become essential. However, images captured by non-specialized operators often result in images of the incorrect anatomical structure or outside the scope. **Objective:** To propose a classification model to validate the anatomical structure of the images, categorizing them into four classes: Color Retina, Red-free Retina, Anterior Segment, and No-eye. **Method:** Five deep learning architectures for image classification (MobileNetV3-Large, EfficientNet-B0, DenseNet121, ConvNeXt-Tiny, and Vision Transformer (ViT)) were evaluated using a dataset of 11,679 images. **Results:** The networks achieved 99% accuracy; however, MobileNetV3-Large stood out for its efficiency, exhibiting lower latency (12.5 ms) and lower storage consumption (5.8 MB). Furthermore, this model achieved 97.15% accuracy and a 97.17% macro F1-score on datasets used for external validation. **Conclusion:** MobileNetV3-Large showed the best performance for the proposed application, considering latency and memory space, important aspects in a scenario of scarce computational resources.

Keywords: Portable fundus photography; Neural Networks; Computer; Ophthalmic screening.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Artigo apresentado no XXI Congresso Brasileiro de Informática em Saúde (CBIS 2026), realizado de 23 a 25 de setembro de 2026, em Brasília (DF), e publicado na seção CBIS 2026 do Journal of Health Informatics.

COMO CITAR ESTE ARTIGO: Santos IGSC, Almeida AS, de Almeida JDS, Silva AC, Costa EPF, Campos MAG. Classificação de estruturas oculares em retinógrafos portáteis utilizando deep learning para triagem oftalmológica. J Health Inform 2026;18:e1875. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1875

1 INTRODUÇÃO

A partir de exames oftalmológicos, é possível diagnosticar patologias silenciosas precocemente e, consequentemente, iniciar tratamentos eficazes. Se identificadas tardiamente, tais patologias podem evoluir para cegueira irreversível ou danos permanentes à estrutura ocular. A deficiência visual evitável permanece um problema global relevante, com projeção de crescimento expressivo até 2040, especialmente em casos tratáveis por diagnóstico precoce, como catarata e erros refrativos não corrigidos [1,2]. Ainda, no Brasil, populações quilombolas, indígenas, ribeirinhas e comunidades rurais enfrentam barreiras geográficas e socioeconômicas que dificultam o acesso ao atendimento oftalmológico especializado [3-5].

Nesse cenário, a oftalmoscopia baseada em *smartphones* surge como uma solução que viabiliza a triagem em larga escala, como demonstrado em estudo que avaliou modelos baseados em *Transformers* para a detecção de glaucoma em retinografias adquiridas por oftalmoscópio portátil [6], bem como em trabalho que propôs um método de *stitching* para retinografias de baixa resolução [7]. Ao empregar essa tecnologia, é possível obter imagens médicas com custo reduzido e portabilidade, permitindo integração com sistemas de telemedicina que permitem a análise remota das imagens por especialistas [8]. Diferentemente dos retinógrafos de mesa com suportes fixos e foco automático, a portabilidade deste dispositivo exige maior estabilidade do operador e do paciente para garantir a boa qualidade das imagens adquiridas [9]. Assim, sob operadores não especialistas, é comum a ocorrência de capturas inadequadas, como imagens fora de foco, registros acidentais do segmento anterior ou sem conteúdo relevante.

Diante desse desafio, a classificação e validação automática do tipo das imagens em tempo real tornam-se indispensáveis na triagem. Este trabalho propõe um estudo comparativo entre cinco arquiteturas de *deep learning*, incluindo quatro redes neurais convolucionais (CNNs) e um Vision Transformer (ViT) para classificar imagens de dispositivos móveis em quatro classes: Retina colorida, Retina Red-free, Segmento Anterior e imagens fora do escopo (No-eye). A identificação da estrutura é fundamental no processo de avaliação da qualidade para encaminhar as capturas a algoritmos específicos. Assim, a pesquisa avalia o equilíbrio entre acurácia e latência, visando determinar a arquitetura mais adequada para garantir *feedback* imediato ao profissional operador.

2 TRABALHOS RELACIONADOS

Trabalhos recentes abordam controle de qualidade em retinografias por regressão [10], avaliação automática do segmento anterior com EfficientNet [11] e sistemas específicos para classificação de qualidade ocular com arquiteturas leves como MobileEye [12]. O trabalho proposto por Vohra *et al.* [13] apresenta um sistema de baixo custo baseado em um dispositivo de hardware e uma arquitetura web armazenada na nuvem para diagnóstico automatizado, demonstrando, no diagnóstico de patologias, desempenho superior ao dos trabalhos anteriores. Devido à computação de borda, também é possível incorporar o processamento de imagens no próprio dispositivo, tornando viável o uso de modelos de inteligência artificial para classificação de qualidade da imagem [14]. Além disso, Chakour *et al.* [15] propõem um sistema de diagnóstico de retinopatia diabética que tem como diferencial o funcionamento *offline* em um smartphone conectado a uma câmera de fundo de olho portátil. Para a validação do estudo, foram comparados os desempenhos de modelos especializados em dispositivos móveis.

Apesar de os trabalhos citados apresentarem evoluções na avaliação da qualidade e diagnóstico automatizado, é notável a lacuna na etapa de identificação da estrutura capturada em dispositivos portáteis. Atualmente, as abordagens desenvolvem trabalhos relacionados à qualidade da retina [10] ou à qualidade do segmento anterior [12], assumindo que a captura já contém a anatomia correta. Ademais, soluções que usam arquiteturas na nuvem [13] enfrentam desafios de latência e conectividade em regiões remotas. Em segunda análise, propostas de arquiteturas *offline* [15] abordam apenas o diagnóstico final das patologias, uma tarefa computacionalmente custosa que poderia ser otimizada se precedida por uma triagem eficiente. Diferentemente disso, este trabalho propõe um modelo leve de classificação multiclasse, que atua como um mecanismo de roteamento, garantindo que apenas imagens com conteúdo anatômico válido sejam submetidas aos *pipelines* de avaliação de qualidade e diagnóstico subsequentes.

3 MÉTODOS

Esta seção apresenta a metodologia adotada neste estudo. A seguir, estão descritas as etapas que se iniciam na definição e processamento do *dataset*, em seguida a escolha e treinamento dos modelos, e por fim, a avaliação dos resultados. O processo está ilustrado na Figura 1.

Figura 1. Etapas da metodologia.



Fonte: Elaborado pelos autores (2026).

3.1 Conjunto de Dados

Para o desenvolvimento e avaliação dos modelos, foi construído um conjunto de dados composto por imagens médicas proprietárias, adquiridas com aprovação do comitê de ética CAAE: 82867424.6.0000.5087, e imagens de domínio geral [16], totalizando 11.679 amostras. Durante a aquisição, foram capturadas imagens de 309 pacientes do município de Itapecuru Mirim no interior do estado do Maranhão. A aquisição das imagens foi realizada utilizando retinógrafo portátil, sem a dilatação da pupila e por profissionais que não são da área médica oftalmológica, refletindo o cenário real de aquisições realizadas em áreas remotas e com escassez de recursos.

A base de dados é composta por quatro classes: Retina (Colorida), Retina Red-free, Segmento Anterior (Olho) e fora do escopo (No-eye). As imagens da classe Retina são capturas de retinografia padrão utilizadas para diagnóstico de retinopatia diabética, glaucoma, degeneração macular relacionada à idade e outras patologias. Já a retinografia Red-free é um exame oftalmológico de imagem, realizado em preto e branco e obtido automaticamente a partir da retinografia original. É utilizada para destacar os vasos sanguíneos, a mácula e as fibras nervosas da retina, sendo crucial para diagnosticar e monitorar retinopatia diabética, hipertensiva e glaucoma sem usar contraste. Exemplos de cada classe são apresentados na Figura 2.

Para compor a classe No-eye, foram utilizadas imagens do Imagenette [16], um subconjunto do ImageNet contendo 10 classes distintas de objetos e cenários. Foram selecionadas aleatoriamente 2.800 imagens desta fonte para garantir que o modelo rejeitasse capturas acidentais em que o dispositivo fosse apontado para objetos ou cenas sem conteúdo ocular, apesar de imagens de domínio geral não representarem borrões ou artefatos que mimetizam a retina. Cabe ressaltar a presença de imagens de baixa qualidade que não foram removidas, já que o objetivo do modelo é classificar a estrutura, assim, imagens degradadas continuam sendo consideradas válidas quando contêm a estrutura correta.

Todas as imagens foram redimensionadas para a resolução de 224 × 224 pixels. Para mitigar o desbalanceamento durante o treinamento, adotou-se a estratégia de *Class Weighting*. Os pesos foram calculados a partir da distribuição real de classes em cada *fold*, considerando tanto imagens clínicas quanto amostras No-eye, resultando em penalidades maiores para erros nas classes minoritárias e menores para classes predominantes. A Tabela 1 representa a distribuição geral de dados.

Figura 3. Curvas de perda do treinamento e da validação para a MobileNetV3-Large.



Fonte: Elaborado pelos autores (2026).

Figura 2. Seleção de amostras do conjunto de dados.



Fonte: Elaborado pelos autores (2026).

3.2 Pré-processamento e Modelos Utilizados

As CNNs fundamentam-se na aprendizagem hierárquica proposta por LeCun *et al.* [17]. Neste estudo, avaliaram-se cinco arquiteturas com distintos perfis computacionais: MobileNetV3-Large, otimizada para dispositivos móveis via convoluções *depthwise separable* e mecanismos Squeeze-and-Excitation [18,19]; EfficientNet-B0, que utiliza escalonamento composto para equilibrar profundidade, largura e resolução [20].

A DenseNet121 serviu como *baseline* devido à sua conectividade densa, que favorece o fluxo de gradientes e a reutilização de características [21]. A ConvNeXt-Tiny representou as convoluções modernas, integrando *designs* inspirados em Transformers, como Layer Normalization e blocos invertidos [22]. O Vision Transformer (ViT) foi incluído para comparar a atenção global com as convoluções tradicionais, processando imagens como *patches* para capturar relações espaciais amplas [23].

As CNNs foram inicializadas com pesos da ImageNet-1k, tendo suas camadas de classificação substituídas por um cabeçalho com Global Average Pooling, Dropout (0,3) e uma camada Softmax ajustada ao número de classes. O pré-processamento seguiu as especificações de cada arquitetura (Tabela 2). Já o ViT [23] foi inicializado com pesos pré-treinados, cabeça de classificação configurada para quatro classes e *fine-tuning* completo, mantendo o *backbone* descongelado durante o treino.

3.3 Treinamento dos Modelos

Para avaliação, utilizou-se validação cruzada estratificada por grupos em 5 *folds*, com agrupamento por paciente, de modo que imagens de um mesmo paciente não aparecem em *folds* distintos, evitando vazamento de dados. As imagens Red-free derivadas compartilham o identificador do paciente de origem, permanecendo no mesmo *fold*, o que também evita vazamento entre derivadas. 10% dos pacientes foram separados como conjunto de teste independente (1.316 imagens) e os 90% restantes constituíram a validação cruzada. Cada *fold* foi treinado de forma independente, reinicializando os pesos pré-treinados na ImageNet. O treinamento empregou o otimizador Adam (taxa de aprendizado inicial de $1{\times}10^{-3}$, reduzida por uma estratégia de redução em platô com fator 0,3 e paciência de 2 épocas), função de perda Sparse Categorical Crossentropy, *batch* de 32 e até 15 épocas, com Early Stopping com paciência de 5 épocas, restaurando os melhores pesos. Não foi aplicado aumento de dados e o pré-processamento restringiu-se ao redimensionamento para 224×224 e à normalização específica de cada arquitetura (Tabela 2). Os pesos do melhor *fold* foram preservados com base no F1-score de validação. Empregou-se também precisão mista (float16) e semente fixa (42). Todos os modelos foram treinados sob as mesmas configurações para garantir comparação justa e reprodutibilidade, em GPUs NVIDIA Tesla T4 (Kaggle).

Como etapa complementar de avaliação, foi realizado um experimento de *cross-dataset* utilizando bases independentes não presentes no treinamento. Cada classe foi obtida de uma fonte distinta: segmento anterior do *dataset* SLID (Slit-Lamp Image Dataset) [24], No-eye do Indoor Scenes CVPR 2019 e retina do Eye Disease Image Dataset, cujo conjunto foi dividido em dois subconjuntos disjuntos: um mantido como Retina Colorida e outro convertido sinteticamente para o padrão Red-free com extração do canal verde e ajuste de contraste. Dessa forma, não há sobreposição de imagens entre as duas classes, embora ambas derivem da mesma base, e o fato de cada classe provir de uma base distinta implica que parte do desempenho pode refletir a separação entre fontes. O conjunto final totalizou 2.527 imagens distribuídas entre as quatro classes avaliadas.

Tabela 1. Distribuição do conjunto de dados utilizado nos experimentos.

Classe	Fonte	Amostras
Retina Red-free	Proprietária	5.368
Retina Colorida	Proprietária	2.688
No-eye	Imagenette	2.800
Segmento Anterior	Proprietária	823
Total Geral	-	11.679

Fonte: Elaborado pelos autores (2026).

Tabela 2. Especificações das arquiteturas utilizadas e seus respectivos métodos de pré-processamento.

Modelo	Parâmetros (M)	Entrada (px)	Pré-processamento
MobileNetV3-Large	≈ 5,4	224 × 224	Normalização [-1, 1]
EfficientNet-B0	≈ 5,3	224 × 224	Rescaling [0, 255]
DenseNet121	≈ 8,0	224 × 224	Z-Score
ConvNeXt-Tiny	≈ 28,6	224 × 224	Z-Score
ViT-Base/16	≈ 86	224 × 224	Normalização [-1, 1]

Fonte: Elaborado pelos autores (2026).

Tabela 3. Comparação do desempenho médio na validação cruzada.

Modelo	Acurácia (%)	Especificidade (%)	Precisão (%)	Sensibilidade (%)	F1-Score (%)	AUC (%)
DenseNet121	99,98 ± 0,03	99,99 ± 0,01	99,98 ± 0,02	99,96 ± 0,06	99,97 ± 0,04	100,00 ± 0
ViT-Base/16	99,98 ± 0,04	99,99 ± 0,02	99,99 ± 0,02	99,93 ± 0,13	99,96 ± 0,08	100,00 ± 0
MobileNetV3-Large	99,95 ± 0,07	99,98 ± 0,03	99,92 ± 0,15	99,95 ± 0,08	99,93 ± 0,11	100,00 ± 0
ConvNeXt-Tiny	99,94 ± 0,09	99,98 ± 0,03	99,85 ± 0,27	99,93 ± 0,09	99,89 ± 0,18	99,99 ± 0,01
EfficientNet-B0	99,92 ± 0,04	99,97 ± 0,02	99,90 ± 0,09	99,88 ± 0,09	99,89 ± 0,07	100,00 ± 0

Fonte: Elaborado pelos autores (2026).

4 RESULTADOS

A seção a seguir apresenta os resultados do estudo estruturado, comparando as arquiteturas propostas e dando destaque para a relação entre acurácia do modelo e eficiência no cenário central.

4.1 Desempenho das Arquiteturas

Considerando a natureza do problema, todos os modelos obtiveram acurácia média próxima a 99,9%; contudo, esse resultado reflete a alta separabilidade das classes, pois diferem em textura, espaço de cor e domínio de origem das imagens, tornando a tarefa de discriminação favorável. Portanto, as métricas obtidas na avaliação interna devem ser interpretadas como um limite superior otimista do desempenho, e não como indicador de acurácia diagnóstica. O modelo DenseNet121 apresentou o melhor desempenho médio (99,98%), empatado com o ViT-Base/16, seguido pela MobileNetV3-Large, que, apesar de ser uma das arquiteturas mais leves, obteve 99,95% de acurácia média. Apesar da equivalência dos resultados, foram observadas diferenças relevantes na eficiência computacional discutidas nas seções subsequentes. A Tabela 3 apresenta um resumo geral da média das métricas mensuradas nos 5 *folds* da validação cruzada para cada modelo.

A Figura 3 apresenta as curvas de perda durante as etapas de treinamento e validação para o modelo MobileNetV3-Large. As curvas representam a média das métricas durante o treino, enquanto a área sombreada representa o desvio-padrão. Observa-se que o modelo apresenta estabilidade já com poucas épocas, o que demonstra a eficácia da estratégia e do ajuste de pesos para o problema em questão. Ainda, é possível notar que a arquitetura aprendeu características importantes que a tornam capaz de generalizar para dados ainda não vistos, não apresentando *overfitting* já que as curvas tendem à proximidade.

4.2 Análise da Eficiência Computacional

Embora a DenseNet121 e a MobileNetV3-Large tenham apresentado métricas comparáveis, a partir da Tabela 4 é possível observar que a rede MobileNetV3-Large apresenta uma latência menor que a DenseNet121, indicando maior eficiência no tempo de inferência. Para o cálculo do resumo da latência apresentado, os modelos foram transformados para o formato TensorFlow Lite com quantização float16 pós-treino em que os pesos são armazenados em 16 bits. Ainda, para garantir a padronização do experimento, o número de *threads* da CPU foi fixado em 1, evitando o paralelismo. A CPU utilizada foi Intel Xeon do ambiente Kaggle. Inicialmente, foram realizadas 20 execuções não contabilizadas, com o objetivo de estabilizar o carregamento de memória e reduzir o impacto no resultado. Em seguida, foram avaliadas 500 imagens com o pré-processamento necessário. O tempo de inferência foi medido individualmente para cada imagem considerando apenas o intervalo da chamada da inferência. Esse processo permitiu uma estimativa estável e representativa do desempenho das arquiteturas. A Tabela 4 apresenta a média, o desvio-padrão e o percentil 95 dos modelos.

Considerando que MobileNetV3-Large apresentou desempenho equivalente às demais arquiteturas em termos de acurácia, F1-score, precisão e *recall*, porém com latência expressivamente inferior, esta arquitetura foi selecionada como modelo final do estudo. Conforme apresentado na Tabela 4, a MobileNetV3-Large obteve uma latência média de 12,524 ms por imagem e percentil 95 de 17,774 ms, o que indica comportamento estável mesmo no pior caso. Em relação ao espaço de armazenamento, o modelo também apresentou o menor tamanho no formato TFLite, reduzindo o consumo de memória e favorecendo a implementação em dispositivos móveis. Cabe ressaltar que o desempenho em dispositivos móveis pode variar conforme o gerenciamento térmico, o uso de *delegates* de aceleração e outros detalhes. Portanto, a escolha é fundamentada pela relação entre acurácia e custo computacional, visto que a latência reduzida não levou a perda relevante nas métricas de classificação.

Tabela 4. Latência média, P95 e tamanho dos modelos TFLite avaliados

Modelo	Latência (ms/imagem)	P95 (ms)	Tamanho (.tflite) (MB)
ConvNeXt-Tiny	190,405 ± 62,233	278,782	54,4
DenseNet121	118,519 ± 40,388	172,992	13,7
EfficientNet-B0	29,241 ± 6,534	40,688	7,8
ViT-Base/16	13,053 ± 0,262	20,284	87,8
MobileNetV3-Large	12,524 ± 3,328	17,774	5,8

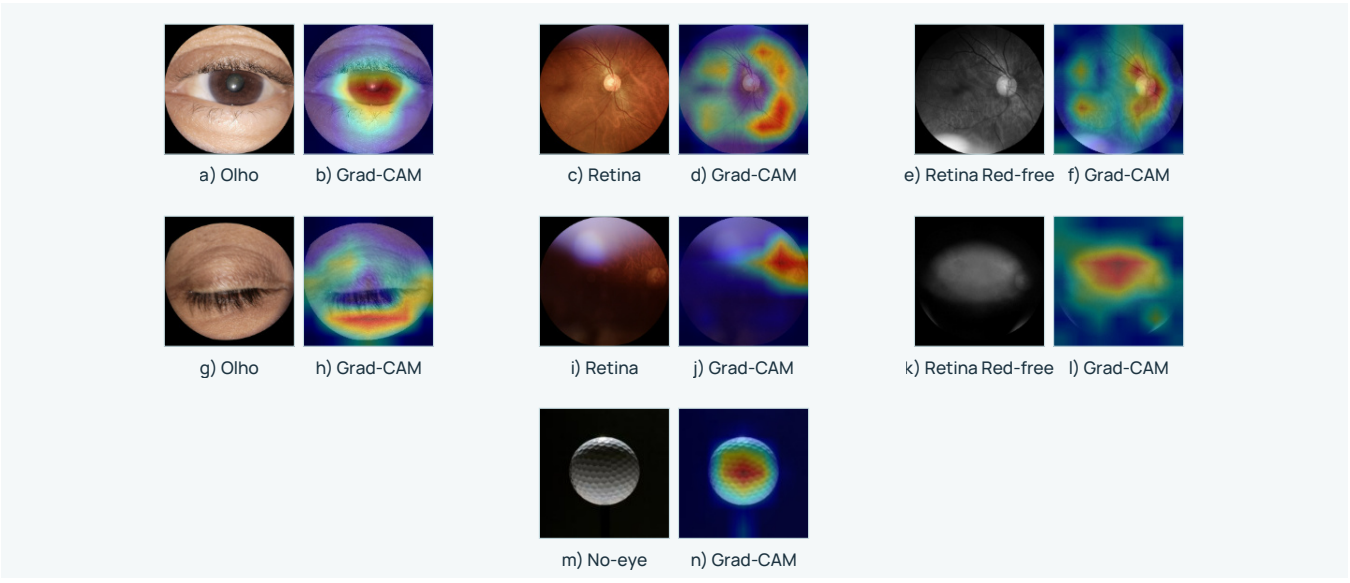
Fonte: Elaborado pelos autores (2026).

4.3 Análise do Modelo Selecionado

Para a análise do modelo selecionado, foram utilizadas duas abordagens: Grad-CAM (Gradient-weighted Class Activation Mapping) e a matriz de confusão. Os mapas de ativação foram gerados a partir da camada convolucional activation_19 do *backbone* da MobileNetV3-Large, com o objetivo de verificar se o modelo prioriza regiões anatômicas relevantes, como disco óptico e mácula nas retinografias, e pálpebras, cílios e pupila no segmento anterior. A Figura 4 apresenta exemplos das classes avaliadas, incluindo imagens com boa qualidade e casos mais desafiadores.

Observou-se que o modelo concentrou sua atenção em estruturas semanticamente relevantes para a classificação. No segmento anterior, destacaram-se pupila, pálpebras e cílios; nas retinografias coloridas e Red-free, região macular, disco óptico e vasos sanguíneos; e, na classe No-eye, a ausência de padrões anatômicos de exames oftalmológicos. A matriz de confusão (Figura 5), no conjunto de teste (10%), apresentou um erro. Ainda assim, o modelo atingiu acurácia de 99,92%, demonstrando potencial para apoio na validação e direcionamento das imagens. Por classe, o F1-score no teste interno foi de 100% para Olho e Retina, 99,91% para Red-free e 99,87% para No-eye. O único erro, uma imagem No-eye classificada como Red-free, reduziu a sensibilidade da classe No-eye para 99,75% e a precisão da Red-free para 99,82%.

Figura 4. Grad-CAM.



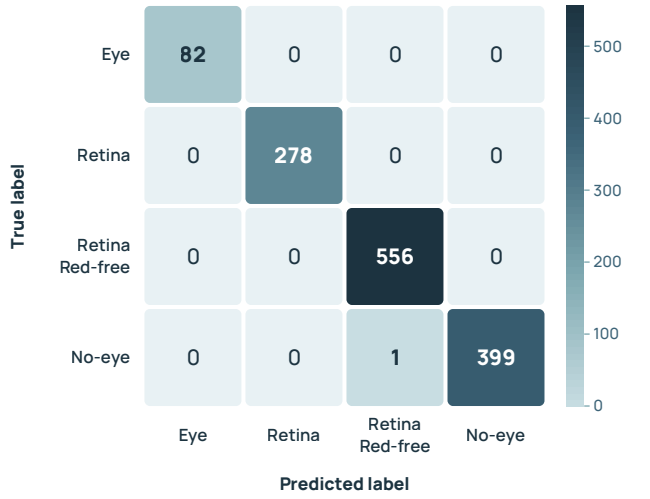
Fonte: Elaborado pelos autores (2026).

Tabela 5. Validação externa do modelo MobileNetV3-Large

Classe	Precisão (%)	*Recall* (%)	F1 (%)	N
Olho	99,81	88,17	93,63	600
Retina	90,84	99,86	95,13	695
Red-free	100	100	100	696
No-eye	99,81	100	99,91	536

Fonte: Elaborado pelos autores (2026).

Figura 5. Matriz de confusão para a arquitetura MobileNetV3-Large.



Fonte: Elaborado pelos autores (2026).

4.4 Validação Externa

Foi realizada validação externa com bases independentes do treinamento, incluindo imagens de segmento anterior do SLID [24] e ambientes internos do Indoor Scene Recognition para a classe No-eye. As imagens de retina foram obtidas de um único *dataset* (Eye Disease Dataset) e divididas em dois subconjuntos disjuntos: um mantido como Retina Colorida e outro convertido sinteticamente para o padrão *Red-free* (diferença de natureza de cor). Ao todo, foram avaliadas 2.527 imagens distribuídas entre as classes Olho (600), Retina (695), Retina Red-free (696) e No-eye (536). A MobileNetV3-Large atingiu acurácia de 97,15% e F1-score macro de 97,17%. A principal queda ocorreu na classe Olho (*recall* de 88,17%), devido ao foco exclusivo na pupila e íris e à ausência de estruturas como cílios e pálpebras nas imagens externas. A Tabela 5 resume as métricas.

5 CONCLUSÃO

Este estudo demonstrou a viabilidade de utilizar redes neurais convolucionais para a triagem automática de imagens oftalmológicas em dispositivos móveis. A partir do trabalho desenvolvido, a análise comparativa evidenciou que todas as arquiteturas avaliadas atingiram acurácia superior a 99%. Porém ressalta-se que esse desempenho decorre das diferenças de cor, textura e origem, assim, as métricas internas representam um limite superior otimista e não devem ser interpretadas como desempenho clínico. Sob essa ressalva, e considerando que a acurácia se manteve equivalente entre os modelos, a MobileNetV3-Large apresentou a melhor performance para a aplicação proposta, dentre os modelos comparados, considerando latência e espaço de memória. A escolha desta arquitetura é justificada tanto pela precisão quanto pela eficiência computacional, pois apresenta a menor latência e o menor consumo de armazenamento. Esses aspectos são importantes tendo em vista um cenário de escassez de recursos computacionais.

A visualização dos mapas de ativação indica que a classificação tende a se concentrar em estruturas oculares relevantes, como o disco óptico e a mácula, embora não constitua, por si só, prova de explicabilidade clínica. Além disso, a acurácia e F1-Score de 97% na validação externa com 2.527 imagens indicam boa capacidade de generalização sob variação de domínio. A implantação deste modelo como etapa preliminar direcionando cada imagem para o *pipeline* de análise mais adequado ao seu tipo pode contribuir para a democratização do acesso à saúde ocular ao proporcionar auxílio a profissionais que não são da área médica oftalmológica e otimizar campanhas de saúde ocular em populações vulneráveis. Como trabalhos futuros, planeja-se integrar o modelo a uma aplicação móvel, realizar testes de usabilidade com profissionais em campo e conduzir uma validação em diferentes populações, instituições e dispositivos, com imagens reais das quatro classes.

Contribuição dos autores. Os autores contribuíram igualmente em todas as etapas de construção do artigo: concepção e delineamento do estudo; coleta, análise e interpretação dos dados; redação e revisão crítica do manuscrito; e aprovação da versão final.

Agradecimentos. Os autores agradecem ao Núcleo de Computação Aplicada da Universidade Federal do Maranhão (NCA/UFMA) pelo suporte e pela infraestrutura disponibilizados, e aos pacientes do município de Itapecuru Mirim (MA) que participaram da coleta das imagens.

Financiamento. Este trabalho foi apoiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código de Financiamento 001, pela Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Processo nº APP-02680/25) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processos nº 406497/2024-9 e 305253/2025-5).

Disponibilidade de dados e materiais. As imagens clínicas são proprietárias e foram obtidas com aprovação do Comitê de Ética (CAAE 82867424.6.0000.5087), não podendo ser disponibilizadas publicamente; podem ser solicitadas ao autor correspondente, mediante justificativa e respeitadas as restrições éticas. O conjunto Imagenette é público.

Conflitos de interesse. Os autores declaram não haver conflitos de interesse.

Declaração de uso de IA. Os autores não utilizaram ferramentas de IA generativa na redação do manuscrito.

REFERÊNCIAS

1. GBD 2019 Blindness and Vision Impairment Collaborators, Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health. 2021;9(2):e144-e160. doi: https://doi.org/10.1016/S2214-109X(20)30489-7
2. Zhang Y, Chen J, Liao L, Ding X. Global Trends and Projections for Avoidable Visual Impairment among Working-Age Individuals: A Population-Based Study. Am J Ophthalmol. 2025;271:304-15. doi: https://doi.org/10.1016/j.ajo.2024.12.003
3. Allison K, Virk A, Alamri A, Patel D. Analysis of the Awareness and Access of Eye Healthcare in Underserved Populations. Vision (Basel). 2025;9(3):55. doi: https://doi.org/10.3390/vision9030055
4. Guimarães AF, Barbosa VLM, Silva MP, Portugal JKA, Reis MHS, Gama ASM. Acesso a serviços de saúde por ribeirinhos de um município no interior do estado do Amazonas, Brasil. Revista Pan-Amazônica de Saúde. 2020;11(0). doi: https://doi.org/10.5123/s2176-6223202000178
5. Martins MVP, Cunha SC, Godoi LF, Matos ITM, Sousa AFS, Alvares MCR, et al. Prevalência de cegueira em uma população idosa quilombola da região norte do Tocantins. Revista Brasileira de Oftalmologia. 2025;84:e0056. doi: https://doi.org/10.37039/1982.8551.20250056
6. Costa ROC, Pimentel PO, Pessoa ACP, Braz Júnior G, Almeida JDS. Diagnóstico de Glaucoma em Retinografias de Oftalmoscópio Portátil Utilizando Ensemble Baseado em Transformers. In: Anais do XXV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2025). [s.l.]: Sociedade Brasileira de Computação - SBC; 2025. p. 449-60. doi: https://doi.org/10.5753/sbcas.2025.7270
7. Nunes GGS, Almeida JDS, Quitanilha DBP, Cunha A. Image Stitching Baseado em Bifurcações Vasculares Aplicado a Retinografias de Baixa Resolução. In: Anais do XXV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2025). [s.l.]: Sociedade Brasileira de Computação - SBC; 2025. p. 485-96. doi: https://doi.org/10.5753/sbcas.2025.7312
8. Ahn SJ, Kim YH. Clinical Applications and Future Directions of Smartphone Fundus Imaging. Diagnostics (Basel). 2024;14(13):1395. doi: https://doi.org/10.3390/diagnostics14131395
9. Davila JR, Sengupta SS, Niziol LM, Sindal MD, Besirli CG, Upadhyaya S, et al. Predictors of Photographic Quality with a Handheld Nonmydriatic Fundus Camera Used for Screening of Vision-Threatening Diabetic Retinopathy. Ophthalmologica. 2017;238(1-2):89-99. doi: https://doi.org/10.1159/000475773
10. Abramovich O, Pizem H, Van Eijgen J, Oren I, Melamed J, Stalmans I, et al. FundusQ-Net: A regression quality assessment deep learning algorithm for fundus images quality grading. Comput Methods Programs Biomed. 2023;239:107522. doi: https://doi.org/10.1016/j.cmpb.2023.107522
11. Gu B, Christopher M, Lim SH, Baxter SL. Deep Learning-Based Analysis of Ocular Anterior Segment Diseases from Patient-Self-Captured Smartphone Images. Informatics. 2025;12(1):2. doi: https://doi.org/10.3390/informatics12010002
12. Xu C, Liu Y, Wang S, Wang Y, Lin X, Guan W, et al. Deep learning-based automatic image quality control system for anterior eye segment in slit lamp and smartphone. Biomedical Signal Processing and Control. 2025;106:107776. doi: https://doi.org/10.1016/j.bspc.2025.107776
13. Vohra H, Hasan MK, Norul Huda Sheikh Abdullah S, Islam S, Hadi Abd Rahman A, Bhojake SN, et al. A Low-Cost AI-Powered System for Early Detection of Diabetic Retinopathy and Ocular Diseases in Resource Limited Settings. IEEE Access. 2025;13:97322-36. doi: https://doi.org/10.1109/access.2025.3572471
14. Rêgo S, Monteiro-Soares M, Dutra-Medeiros M, Camila Dias C, Nunes F. Exploring the Feasibility of Opportunistic Diabetic Retinopathy Screening with Handheld Fundus Cameras in Primary Care: Insights from Doctors and Nurses. Diabetology. 2024;5(6):566-83. doi: https://doi.org/10.3390/diabetology5060041
15. Chakour EM, Sadok Z, Kachouri R, Mansouri A, Andaloussi IB, Akil M, et al. Mobile-based deep learning system for early detection of diabetic retinopathy. Intelligence-Based Medicine. 2025;12:100259. doi: https://doi.org/10.1016/j.ibmed.2025.100259
16. Howard J. Imagenette: a smaller subset of 10 easily classified classes from ImageNet [Internet]. 2019. Disponível em: https://github.com/fastai/imagenette.
17. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc. IEEE. 1998;86(11):2278-324. doi: https://doi.org/10.1109/5.726791
18. Howard A, Sandler M, Chen B, Wang W, Chen LC, Tan M, et al. Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). [s.l.]: IEEE; 2019. p. 1314-24. doi: https://doi.org/10.1109/iccv.2019.00140
19. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861. 2017.
20. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proc 36th Int Conf Mach Learn (ICML). PMLR; 2019. p. 6105–14.

21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [s.l.]: IEEE; 2017. p. 2261-9. doi: https://doi.org/10.1109/cvpr.2017.243

22. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [s.l.]: IEEE; 2022. p. 11966-76. doi: https://doi.org/10.1109/cvpr52688.2022.01167

23. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv: 2010.11929. 2020.

24. Xu M, Sun Y, Cheng H, Zhou Y, Maimaiti N, Chen P, et al. SLID: a slit-lamp image dataset for deep learning-based anterior eye anatomical segmentation and multi-lesion detection. Front Digit Health. 2025;7:1716501. doi: https://doi.org/10.3389/fdgth.2025.1716501

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde