Classificação de Exames de Imagem Oftalmológicos por Redes Neurais Convolucionais com Transferência de Aprendizado

Luiz Henrique Pereira Niero[1]

[1]Universidade Estadual Paulista "Júlio de Mesquita Filho - UNESP)

luiz.niero@unesp.br

Abstract. *This article explores the use of Convolutional Neural Networks (CNNs), specifically the ResNet-18 architecture, to automatically classify digital ophthalmological images that lack textual content extractable via OCR. The goal is to validate a fully visual classification approach capable of identifying exam types (OCT, fundus photography, tonometry, among others) even in low-data settings. Leveraging transfer learning, stratified cross-validation, and supervised evaluation, the model achieved over 98% accuracy on the test set. The results suggest that text-independent visual classifiers can effectively replace OCR-based pipelines in privacy-sensitive clinical contexts.*

Resumo. *Este artigo investiga o uso de Redes Neurais Convolucionais (CNNs), com ênfase na arquitetura ResNet-18, para classificar automaticamente imagens oftalmológicas digitais que não possuem conteúdo textual extraível por OCR. O objetivo é validar uma abordagem puramente visual capaz de identificar o tipo de exame (OCT, retinografias, tonometria, entre outros) mesmo em contextos com dados limitados. Utilizando técnicas de transfer learning, validação cruzada e avaliação supervisionada, o modelo alcançou acurácia superior a 98% no conjunto de teste. Os resultados indicam que é possível substituir mecanismos baseados em texto por classificadores visuais robustos em contextos médicos sensíveis à privacidade e à estrutura dos dados.*

1

1. Introdução

O avanço da capacidade e popularidade de tecnologias de inteligência artificial [Singla et al. 2024] tem impulsionado um movimento crescente de integração entre sistemas legados e novas plataformas baseadas em IA, especialmente na área da saúde [Rajpurkar and Outros 2022]. Soluções como copilotos médicos, apoio ao diagnóstico, análise da adequação a tratamentos, prevenção a fraudes e automação da elaboração de contas médicas vêm se consolidando como estratégias relevantes para a redução do custo operacional [Sahoo et al. 2025, Esteva et al. 2019].

Porém, enquanto as ferramentas de IA tornam-se cada vez mais poderosas e acessíveis, um dos principais desafios enfrentados na área da saúde é a qualidade dos dados disponíveis. Embora existam padrões de interoperabilidade, como o HL7 (Health

[1]O código-fonte completo deste trabalho está disponível em: https://github.com/luiz-niero/CNN_Resnet18_exam_classification

Level Seven), que buscam organizar e uniformizar as informações médicas em escala global, a adoção desses padrões ainda é limitada. Relatórios mostram que muitos sistemas de informação hospitalares continuam utilizando fluxos não estruturados e obsoletos, como fax ou entrada manual, e apenas uma fração das instituições integra efetivamente seus dados via HL7 [Shortliffe and Sepúlveda 2018].

A esse cenário soma-se o fato de que aproximadamente 80% dos dados em saúde são não estruturados, incluindo textos livres, imagens médicas e anotações manuais [Esteva et al. 2019]. Em ambientes multi-institucionais, essa despadronização compromete significativamente a análise automática dos dados. Um mesmo exame, como uma tomografia de coerência óptica (OCT), pode ser identificado com nomes distintos em diferentes bancos de dados, tornando difícil a classificação automatizada baseada apenas em texto.

Mesmo em países com maior digitalização dos sistemas de saúde, a falta de dados rotulados para tarefas supervisionadas permanece como um obstáculo. Estimativas apontam que menos de 1% dos dados clínicos disponíveis são devidamente anotados para fins de treinamento de modelos de aprendizado de máquina [Johnson et al. 2016], o que limita o desempenho e a generalização dos modelos em ambientes reais.

No Brasil, existem iniciativas como a Rede Nacional de Dados em Saúde (RNDS), que visam padronizar e integrar o intercâmbio de dados clínicos em nível nacional. No entanto, essas iniciativas ainda enfrentam desafios técnicos e legais, além de baixa adesão entre instituições privadas e prestadores regionais.

Adicionalmente, utilizar grandes modelos de linguagem via APIs públicas para resolver esse problema esbarra em barreiras legais e operacionais. A Lei Geral de Proteção de Dados brasileira (LGPD) classifica os dados de saúde como sensíveis, exigindo consentimento explícito para usos que extrapolem o tratamento do paciente [Brasil 2018]. A Resolução CFM nº 2.314/2022 também reforça que a utilização de documentos clínicos para pesquisa só pode ocorrer com autorização formal do paciente [de Medicina 2022]. Além disso, o uso de APIs públicas impõe riscos de indisponibilidade e custos crescentes conforme a escala do sistema.

Nesse contexto, modelos de inteligência artificial executados em ambientes privados, locais e seguros surgem como uma alternativa viável. No entanto, treinar tais modelos requer um conjunto de dados representativo e bem anotado, o que é raro na área médica por dois fatores principais: i) a exigência legal de consentimento para reutilização de dados clínicos, e ii) o alto custo de profissionais qualificados para anotação dos dados. De acordo com estimativas de mercado, a remuneração média de médicos no Brasil está entre as mais altas do país, variando de R$13 mil a R$20 mil mensais.

Frente a essas dificuldades, técnicas baseadas em OCR (Reconhecimento Óptico de Caracteres) e expressões regulares têm se mostrado úteis. Quando palavras-chave como “tomografia”, “coerência” e “óptica” são reconhecidas em imagens, pode-se inferir que se trata de um exame OCT, sem a necessidade de classificação visual. Essa abordagem é especialmente vantajosa em imagens que contenham cabeçalhos ou anotações textuais.

Por outro lado, exames sem conteúdo textual exigem técnicas visuais, como extração de *features* via redes convolucionais [Goodfellow et al. 2016]. Entre as estratégias consideradas, destaca-se a utilização de redes convolucionais pré-treinadas,

como a ResNet-18, aplicando-se técnicas de *transfer learning* para adaptar o modelo ao domínio oftalmológico com um conjunto de dados limitado.

Este artigo propõe, portanto, o uso de arquiteturas de redes neurais convolucionais para a classificação de exames oftalmológicos digitais, visando superar os desafios apresentados anteriormente. O código-fonte completo, incluindo os scripts de pré-processamento, treinamento e avaliação, está disponível publicamente em: `https://github.com/luiz-niero/CNN_Resnet18_exam_classification`.

2. Metodologia

A metodologia adotada neste trabalho compreende as seguintes etapas: (i) coleta e preparação do conjunto de dados, (ii) pré-processamento das imagens, (iii) definição da arquitetura da rede convolucional, (iv) treinamento do modelo, e (v) avaliação do resultado. Todos os experimentos foram conduzidos em ambiente computacional controlado, garantindo reprodutibilidade.

2.1. Coleta e organização dos dados

O conjunto de dados utilizado neste estudo é composto por imagens de exames oftalmológicos, obtidas a partir da base de dados do sistema Vöiston – um copiloto para médicos oftalmologistas. Foram selecionadas as imagens com muito pouco ou nenhuma informação textual, cujos mecanismos de classificação baseados em OCR não foram capazes de classificá-las corretamente. Com o auxílio de um médico oftalmologista, as imagens foram divididas em 10 diferentes classes:

1. Angiografia fluoresceínica
2. Fixação de biometria tipo 1
3. Fixação de biometria tipo 2
4. OCT de segmento anterior
5. OCT macular
6. Relatório de retinografia
7. Retinografia em autofluorescência
8. Retinografia tipo fundo de olho
9. Tonometria

Para cada classe, foram coletadas 100 imagens, totalizando 900 imagens no experimento.

2.2. Pré-processamento

Todas as imagens passaram por uma etapa de transformação utilizando redimensionamento (*resize*) para que ficassem com o tamanho padrão de 224 x 224 pixels, valor compatível com a arquitetura ResNet-18 utilizada no experimento. Também foi aplicada uma normalização com média e desvio padrão correspondentes ao conjunto de dados ImageNet [Deng et al. 2009], que é utilizado como base de pré-treinamento para diversas arquiteturas convolucionais.

2.3. Divisão do Conjunto de Dados

O conjunto de dados foi estruturado inicialmente a partir de uma pasta com subdiretórios nomeados de acordo com as classes de interesse, sendo carregado com o auxílio da classe `ImageFolder` da biblioteca PyTorch [Paszke et al. 2019]. Para garantir uma avaliação justa e realista do modelo, o *dataset* foi dividido em três subconjuntos:

- **Conjunto de teste:** 20% das amostras foram separadas antes do treinamento para compor o conjunto de teste final, utilizando amostragem estratificada.
- **Conjunto de treino e validação:** os 80% restantes foram utilizados em validação cruzada K-Fold com $k = 5$ [Goodfellow et al. 2016], onde em cada iteração 80% dos dados eram usados para treinamento e 20% para validação.

Esse esquema de divisão permitiu avaliar o desempenho do modelo em múltiplas partições, reduzindo o viés de seleção e aumentando a generalização dos resultados. Além disso, o uso de validação estratificada garante que todas as classes estejam representadas proporcionalmente em todas as dobras.

2.4. Arquitetura do modelo

A arquitetura adotada é baseada na ResNet-18 [He et al. 2016], uma rede neural convolucional com conexões residuais, amplamente utilizada em tarefas de visão computacional devido à sua eficiência e profundidade moderada. A ResNet-18 foi carregada com pesos pré-treinados no *dataset* ImageNet [Deng et al. 2009] e teve seus parâmetros congelados, prática conhecida como *transfer learning*, para evitar sobreajuste nas primeiras camadas [Yosinski et al. 2014].

A camada final da rede foi substituída por uma nova cabeça de classificação composta por:

- uma camada totalmente conectada com 256 neurônios;
- uma função de ativação ReLU;
- uma camada de *Dropout* com taxa de 40%;
- e uma camada final com número de saídas igual ao número de classes.

Essa modificação permitiu adaptar a arquitetura ao domínio oftalmológico com custo computacional reduzido.

2.5. Treinamento e Validação

O treinamento foi conduzido com a função de perda `CrossEntropyLoss`, apropriada para tarefas de classificação multiclasse, e o otimizador Adam com taxa de aprendizado de 0,001. Em cada dobra do K-Fold, o modelo era treinado sobre 80% dos dados e validado nos 20% restantes. O modelo com melhor desempenho em cada iteração era salvo.

Todos os experimentos foram conduzidos na plataforma Google Colab, utilizando a linguagem Python e a biblioteca PyTorch [Paszke et al. 2019], em ambiente com GPU Nvidia T4 (15 GB VRAM) e 12 GB de memória RAM.

3. Resultados

A avaliação do desempenho do modelo foi realizada em duas etapas: (i) durante o treinamento, por meio de validação cruzada estratificada com $k = 5$ dobras; e (ii) posteriormente, em um conjunto de teste separado, composto por 20% dos dados iniciais e não utilizado durante o treinamento.

Durante a validação cruzada, o modelo apresentou desempenho elevado e consistente. A acurácia média global foi de 100%, com valores de *precision*, *recall* e *f1-score* superiores a 0,99 para todas as classes, conforme pode ser observado na Tabela 1.

Tabela 1. Métricas por classe na validação cruzada K-Fold

Classe	Precisão	Revocação	F1-Score	Suporte
Angiografia fluoresceínica	1.00	0.97	0.99	80
Fixação de biometria	0.99	1.00	0.99	80
Fixação de biometria 2	1.00	1.00	1.00	80
OCT de segmento anterior	1.00	1.00	1.00	80
OCT macular	1.00	1.00	1.00	80
Relatório de retinografia colorida	0.99	1.00	0.99	80
Retinografia em autofluorescência	0.99	0.99	0.99	80
Retinografia fundo de olho	1.00	1.00	1.00	80
Tonometria	1.00	1.00	1.00	80

A matriz de confusão agregada das dobras (Figura 1) mostra que os erros foram mínimos, com apenas dois casos de confusão observados: um entre “angiografia fluoresceínica” e “retinografia em autofluorescência”, e outro entre “angiografia fluoresceínica” e “retinografia fundo de olho”.

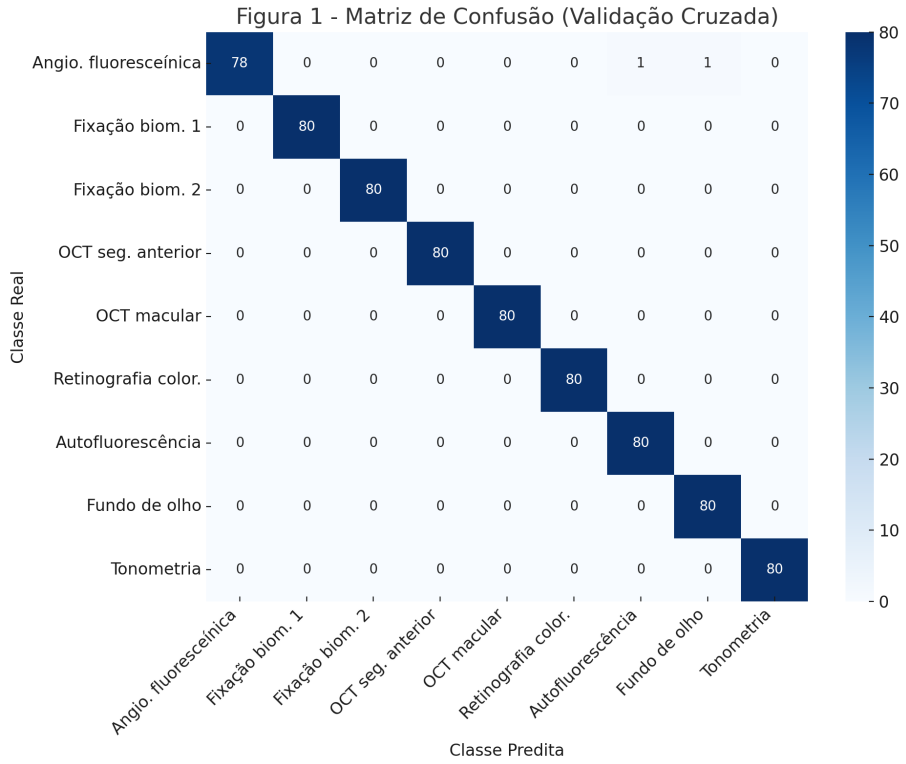

Figura 1. Matriz de confusão agregada na validação cruzada.

3.1. Avaliação no Conjunto de Teste

No conjunto de teste final, composto por 180 imagens (20 por classe), o modelo manteve excelente desempenho, com acurácia global de 98% e métricas macro médias de *precision* = 0,98, *recall* = 0,98 e *f1-score* = 0,98, conforme apresentado na Tabela 2.

Tabela 2. Métricas por classe no conjunto de teste

Classe	Precisão	Revocação	F1-Score	Suporte
Angiografia fluoresceínica	1.00	0.85	0.92	20
Fixação de biometria	1.00	1.00	1.00	20
Fixação de biometria 2	1.00	1.00	1.00	20
OCT de segmento anterior	1.00	1.00	1.00	20
OCT macular	1.00	1.00	1.00	20
Relatório de retinografia colorida	0.95	1.00	0.98	20
Retinografia em autofluorescência	1.00	1.00	1.00	20
Retinografia fundo de olho	0.91	1.00	0.95	20
Tonometria	1.00	1.00	1.00	20

A matriz de confusão do teste (Figura 2) indica que todos os erros de classificação ocorreram apenas na classe “angiografia fluoresceínica”, que foi confundida duas vezes com “retinografia fundo de olho” e uma vez com “relatório de retinografia colorida”.

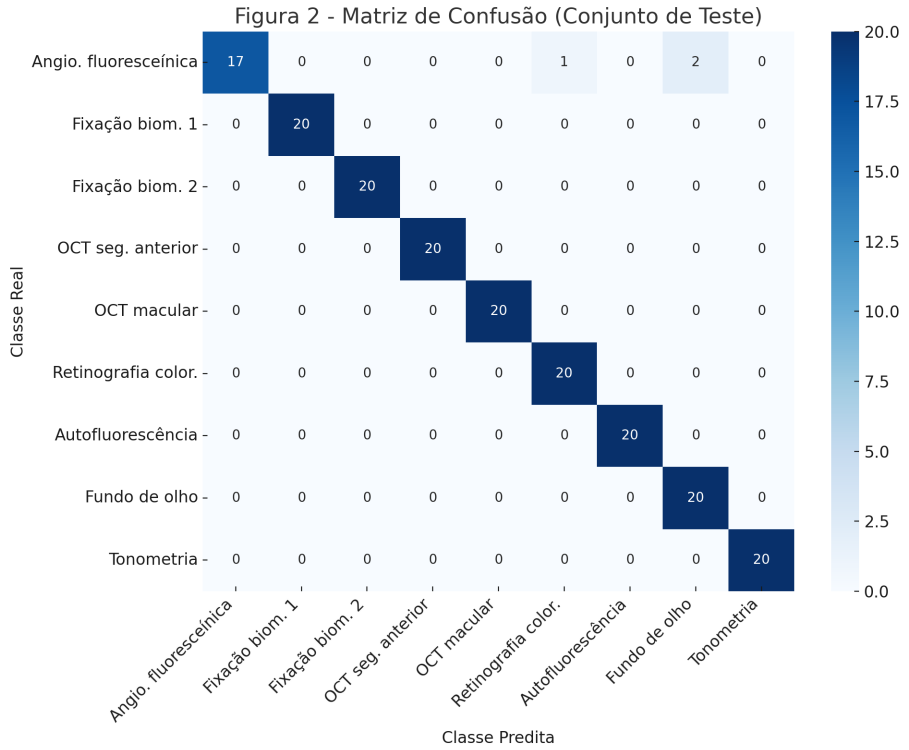

Figura 2. Matriz de confusão no conjunto de teste.

Esses resultados indicam que a arquitetura utilizada é altamente eficaz na distinção entre exames oftalmológicos com características visuais distintas, mesmo sem a presença de texto auxiliar extraído por OCR. O leve decréscimo na performance da classe "angiografia fluoresceínica" sugere que ela apresenta maior semelhança visual com outras classes, e pode ser foco de aprimoramentos futuros.

4. Conclusão

Este estudo demonstrou a viabilidade de utilizar redes neurais convolucionais, especificamente a arquitetura ResNet-18, para a classificação automática de exames oftalmológicos digitais em contextos onde os métodos baseados em OCR se mostram insuficientes. A abordagem proposta foi capaz de alcançar resultados robustos, com acurácia superior a 98% no conjunto de teste e desempenho quase perfeito durante a validação cruzada.

Os resultados confirmam que, mesmo com um conjunto de dados relativamente pequeno e limitado a imagens com baixa ou nenhuma informação textual, é possível obter classificadores visuais eficazes por meio de técnicas de *transfer learning* e *fine-tuning* supervisionado. A estratégia de congelamento das camadas convolucionais pré-treinadas da ResNet-18 — originalmente ajustadas ao ImageNet — e a substituição da camada final por uma cabeça densa adaptada ao domínio oftalmológico permitiram maximizar a generalização com custo computacional reduzido. Essa decisão técnica mostrou-se especialmente valiosa em ambientes com restrições de dados, recursos e privacidade.

No entanto, algumas limitações devem ser consideradas. A coleta de dados foi feita manualmente e com auxílio de um especialista, o que pode não ser escalável para bases maiores. Além disso, as confusões observadas na classe "angiografia fluoresceínica"

indicam que certos exames visuais compartilham características semelhantes e podem exigir arquiteturas mais profundas ou técnicas complementares, como atenção visual ou segmentação, para melhorar ainda mais a precisão.

Como trabalho futuro, propõe-se a expansão do conjunto de dados, incluindo variações de aparelhos e condições clínicas, bem como a comparação com outras arquiteturas mais recentes (como EfficientNet ou Vision Transformers). Além disso, planeja-se investigar abordagens híbridas que combinem análise textual e visual para cenários com imagens parcialmente anotadas.

Referências

[Brasil 2018] Brasil (2018). Lei nº 13.709, de 14 de agosto de 2018. lei geral de proteção de dados pessoais (lgpd). Acesso em: 20 abr. 2025.

[de Medicina 2022] de Medicina, C. F. (2022). Resolução cfm nº 2.314/2022. Dispõe sobre prontuário médico e guarda de dados. Disponível em: `https://portal.cfm.org.br`.

[Deng et al. 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE.

[Esteva et al. 2019] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, L., Cui, C., Corrado, G., Thrun, S., and Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25:24–29.

[Goodfellow et al. 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

[He et al. 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

[Johnson et al. 2016] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., and Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3(1):160035.

[Paszke et al. 2019] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32.

[Rajpurkar and Outros 2022] Rajpurkar, P. and Outros (2022). Título do artigo de 2022. *Nome do Periódico*.

[Sahoo et al. 2025] Sahoo, R. K., Sahoo, K. C., Negi, S., Baliarsingh, S. K., Panda, B., and Pati, S. (2025). Health professionals' perspectives on the use of artificial intelligence in healthcare: A systematic review. *Patient Education and Counseling*, 134:108680.

[Shortliffe and Sepúlveda 2018] Shortliffe, E. H. and Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA*, 320(21):2199–2200.

[Singla et al. 2024] Singla, A., Sukharevsky, A., Yee, L., Chui, M., and Hall, B. (2024). The state of ai in early 2024: Gen ai adoption spikes and starts to generate value. Acesso em: 20 abr. 2025.

[Yosinski et al. 2014] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems (NeurIPS)*, 27.