

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE COMPUTAÇÃO

**Proposta de Projeto da Disciplina MC959/MO810 - Introdução ao
Aprendizado Auto-Supervisionado**

Campinas - 18 de Outubro de 2024

CLASSIFICAÇÃO DE IMAGENS DE TOMOGRAFIA DE COERÊNCIA ÓPTICA
USANDO APRENDIZADO AUTO-SUPERVISIONADO

Aluno: Arimã Batista (194347)

Aluno: José Lucas Amaro de Melo (155978)

Aluna: Juliana Midlej do Espírito Santo (200208)

Aluna: Maria Júlia de Lima Silva (184181)

Professor: Prof. Dr. Marcelo Reis

Sumário

1	Introdução	1
1.1	Objetivos	2
1.2	Hipóteses do Trabalho	2
2	Material e Métodos	3
2.1	Métodos Supervisionados e Auto-Supervisionados	3
2.1.1	Método Supervisionado	4
2.1.2	Método Auto-Supervisionado	4
2.2	Base de Dados	5
2.3	Métricas de Avaliação	5
2.4	Implementação e Recursos Computacionais	6
2.5	Resultados Esperados	6
3	Plano de Trabalho e Cronograma de Execução	7
	Bibliografia	8

Resumo

Este trabalho propõe comparar o desempenho de abordagens de aprendizado supervisionado e auto-supervisionado para a classificação de imagens de Tomografia de Coerência Óptica (OCT) em quatro categorias: neovascularização coroidal, edema macular diabético, drusas e condições normais. Imagens de OCT são amplamente utilizadas para o diagnóstico de distúrbios oculares, e métodos baseados em aprendizado de máquina têm sido eficazes na classificação dessas patologias. Tradicionalmente, técnicas supervisionadas, que requerem grandes volumes de imagens rotuladas, têm sido utilizadas. No entanto, o aprendizado auto-supervisionado, que aprende representações dos dados sem a necessidade de rótulos explícitos, pode oferecer uma alternativa eficiente, especialmente em cenários com dados parcialmente rotulados ou não rotulados.

Capítulo 1

Introdução

Imagens de Tomografia de Coerência Óptica (OCT) são utilizadas para diagnosticar distúrbios oculares, como neovascularização coroidal, edema macular diabético e drusas. Com o aumento da disponibilidade de imagens médicas, especialmente em estudos oftalmológicos, há um interesse crescente em aplicar técnicas avançadas de aprendizado de máquina para melhorar o diagnóstico dessas condições. Nesse contexto, Redes Neurais Convolucionais (CNNs) têm se mostrado eficazes na classificação dessas patologias.

Kermany et al. [3] propõem uma ferramenta de suporte à decisão clínica baseada em CNNs e aprendizado por transferência, utilizando pesos pré-treinados da ImageNet, para classificar doenças oculares. O modelo demonstrou desempenho comparável ao de especialistas humanos, alcançando uma acurácia de 96,6%. Além disso, o estudo destacou a interpretabilidade do modelo ao identificar as regiões de interesse usadas para o diagnóstico.

Khan et al. [4] combinou aprendizado por transferência e técnicas de otimização. A abordagem utiliza três modelos pré-treinados (DenseNet-201, InceptionV3 e ResNet-50) para extrair características das imagens, seguido pela seleção das melhores características por meio da Otimização por Colônia de Formigas (ACO). As características selecionadas foram então usadas para treinar algoritmos de k-vizinhos mais próximos (KNN) e máquinas de vetores de suporte (SVM). Esse método obteve acurácia de 99,1%.

Tradicionalmente, métodos supervisionados são amplamente usados para treinar modelos nessa tarefa, explorando grandes volumes de imagens rotuladas para alcançar bons resultados. No entanto, o aprendizado auto-supervisionado permite que os modelos aprendam representações úteis dos dados sem a necessidade de rótulos explícitos. Esse tipo de aprendizado pode, posteriormente, ser aproveitado em tarefas supervisionadas, aumentando a robustez e generalização do modelo.

O objetivo deste trabalho é comparar o desempenho entre abordagens de aprendizado supervisionado e auto-supervisionado na classificação de quatro categorias em imagens de OCT: neovascularização coroidal, edema macular diabético, drusas e condições normais. Embora haja uma quantidade significativa de imagens disponíveis, testar o aprendizado auto-supervisionado pode trazer *insights* sobre a eficiência do modelo e sua capacidade de extrair informações relevantes, mesmo com menos supervisão direta, sendo uma alternativa viável para o uso em cenários com dados parcialmente rotulados ou em maior escala.

1.1 Objetivos

O objetivo principal é comparar o desempenho entre abordagens de aprendizado supervisionado e auto-supervisionado na classificação de imagens de OCT em quatro categorias: neovascularização coroidal, edema macular diabético, drusas e condições normais.

Os objetivos específicos são:

- Desenvolver modelos supervisionados e auto-supervisionados para a classificação de imagens de OCT;
- Avaliar e comparar o desempenho das abordagens auto-supervisionada e supervisionada desenvolvidas, com base em métricas de classificação como acurácia, precisão, revocação e taxa F1;
- Avaliar e comparar o desempenho das abordagens propostas com outras disponíveis na literatura;
- Documentar os resultados obtidos.

1.2 Hipóteses do Trabalho

As seguintes hipóteses serão investigadas neste trabalho:

- O aprendizado auto-supervisionado será capaz de aprender representações eficazes das imagens de OCT, resultando em um desempenho comparável ao aprendizado supervisionado quando combinado com uma etapa supervisionada posterior;
- A abordagem auto-supervisionada apresentará melhor generalização em novos dados e maior robustez a variações na quantidade de rótulos disponíveis, em comparação ao aprendizado supervisionado;
- Para cenários com abundância de dados rotulados, o aprendizado supervisionado ainda oferecerá um desempenho superior, mas a diferença será reduzida à medida que a quantidade de rótulos disponíveis diminui.

Capítulo 2

Material e Métodos

Este capítulo descreve os métodos supervisionados e auto-supervisionados que serão utilizados, a base de dados, as métricas de avaliação e a implementação e recursos computacionais.

2.1 Métodos Supervisionados e Auto-Supervisionados

Neste trabalho, serão aplicados métodos de aprendizado supervisionado e auto-supervisionado para classificar imagens de tomografia de coerência óptica (OCT) em quatro categorias: neovascularização coroidal (CNV), edema macular diabético (DME), drusas e imagens normais.

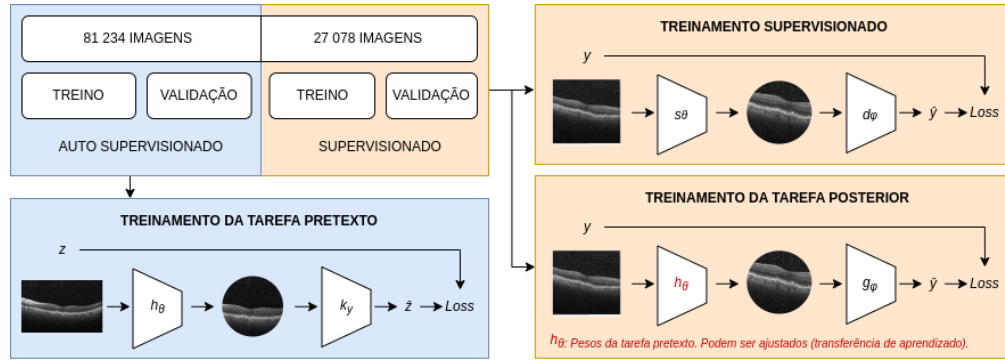


Figura 2.1: Esquema de treinamento dos modelos.

O esquema de treinamento dos modelos é ilustrado na Figura 2.1 e será detalhado nas próximas subseções. Para garantir uma comparação justa entre os modelos, a SqueezeNet [2] foi escolhida como a arquitetura principal devido ao seu pequeno tamanho e eficiência computacional, permitindo um treinamento mais rápido.

A SqueezeNet é uma rede neural convolucional projetada para ser leve, possuindo significativamente menos parâmetros do que outras arquiteturas mais complexas, mas mantendo um desempenho competitivo em tarefas de classificação de imagens.

2.1.1 Método Supervisionado

Um modelo supervisionado será treinado utilizando 25% dos dados de treinamento. Este conjunto de dados, composto por 27078 imagens rotuladas, permitirá que o modelo aprenda a identificar e classificar os padrões radiológicos associados a cada uma das classes. Essa abordagem possibilitará uma avaliação clara do desempenho do modelo supervisionado em um cenário com dados limitados.

- \mathbf{s}_ϕ : Extrator de características (SqueezeNet excluindo a última camada).
- \mathbf{d}_φ : Cabeça de classificação (última camada da SqueezeNet), onde as representações extraídas são mapeadas para as previsões \hat{y} .

2.1.2 Método Auto-Supervisionado

Para o método auto-supervisionado, a tarefa pretexto será realizada utilizando 75% dos dados disponíveis e a tarefa posterior será realizada com 25% dos dados disponíveis.

Treinamento da Tarefa Pretexto

O modelo será pré-treinado utilizando SimCLR [1] como tarefa pretexto. Durante essa fase, diversas transformações, como rotação, corte e mudanças de brilho, serão aplicadas nas imagens, gerando múltiplas visualizações. O modelo aprenderá a maximizar a similaridade entre representações de visualizações diferentes da mesma imagem e a minimizar a similaridade entre diferentes imagens. Essa tarefa permitirá que o modelo extraia características úteis e robustas das imagens de OCT.

O modelo consiste de duas partes:

- \mathbf{h}_θ : Representa a função codificadora que mapeia a imagem original para uma representação latente z (SqueezeNet excluindo a última camada).
- \mathbf{k}_ψ : Este bloco recebe a representação z e realiza a tarefa pretexto (como reconstrução de imagem ou predição de rotação). A perda (Loss) é calculada com base em \hat{z} , que é a saída da tarefa pretexto.

Treinamento da Tarefa Posterior

O modelo previamente treinado será ajustado para a tarefa de classificação, aproveitando as representações aprendidas durante o pré-treinamento e visando uma classificação mais precisa.

- \mathbf{h}_θ : Os pesos do codificador da tarefa pretexto são aplicados à nova tarefa supervisionada. Eles podem ou não ser ajustados durante o treinamento da tarefa posterior (SqueezeNet excluindo a última camada).
- \mathbf{g}_φ : Esta é a cabeça de classificação para a tarefa posterior, responsável por gerar a predição \hat{y} (última camada da SqueezeNet).

2.2 Base de Dados

O conjunto de dados utilizado é sobre imagens de tomografia de coerência óptica (OCT). Os dados de treino original são compostos por 108.312 imagens, sendo 37.206 com neovascularização coroidal, 11.349 com edema macular diabético, 8.617 com drusas e 51.140 normais, coletados de 4.686 pacientes. O conjunto de teste original é composto por 1.000 imagens (250 de cada categoria) de 633 pacientes [3].

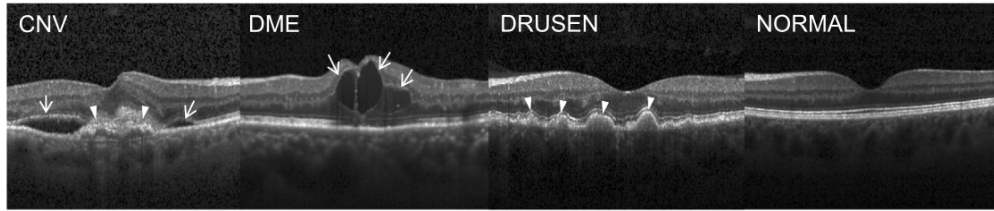


Figura 2.2: (Extrema esquerda) Neovascularização coroidal (CNV) com membrana neovascular (setas brancas) e fluido sub-retiniano associado (setas). (Meio esquerda) Edema macular diabético (DME) com espessamento retiniano associado a fluido intrarretiniano (setas). (Meio direita) Múltiplas drusas (setas) presentes em estágio inicial de DMRI (Degeneração Macular Relacionada à Idade). (Extrema direita) Retina normal com contorno foveal preservado e ausência de qualquer fluido/edema retiniano [3].

Como os dados estão desbalanceados entre as quatro classes de imagens, pretendemos colocar pesos maiores para as classes com menos imagens. Também é notável que o tamanho do conjunto de teste é muito pequeno, então vamos analisar a possibilidade de colocar algumas observações do conjunto de treino para o conjunto de teste.

2.3 Métricas de Avaliação

Para avaliar a classificação serão utilizadas as métricas acurácia, precisão, revocação e a taxa F1. Antes de definir a fórmula para calcular cada métrica, é importante definir alguns termos que serão usados nas equações:

- P: o número de casos positivos;
- N: o número de casos negativos;
- VP: o número de verdadeiros positivos;
- VN: o número de verdadeiros negativos;
- FP: o número de falsos positivos;
- FN: o número de falsos negativos.

A acurácia é uma métrica de avaliação muito usada em problemas de classificação. Seu valor é calculado como apresentado na Equação 2.1. A acurácia de uma classe é o número de casos preditos corretamente com relação a todos os casos existentes.

$$\text{Acurácia} = \frac{VP + VN}{P + N} \quad (2.1)$$

A precisão é calculada pela Equação 2.2. Essa métrica indica a capacidade de classificar como positivo um dado que realmente é positivo.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.2)$$

A revocação é representada pela Equação 2.3. A revocação verifica a capacidade de classificar corretamente dados positivos.

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (2.3)$$

A taxa F1 é dada pela Equação 2.4. Essa métrica corresponde à média harmônica da precisão e da revocação, de modo a obter um valor único que indique a qualidade geral do modelo.

$$F_1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (2.4)$$

2.4 Implementação e Recursos Computacionais

A implementação deste projeto será feita usando a linguagem de programação Python, dado o grande número de ferramentas desenvolvidas com suporte para essa linguagem. O projeto utilizará bibliotecas de aprendizado de máquina, aprendizado profundo, funções científicas e numéricas e visualização de dados. Algumas bibliotecas que podem ser destacadas são: NumPy¹, scikit-learn², PyTorch³ e Matplotlib⁴. Os experimentos deste projeto serão realizados em uma NVIDIA RTX 2060 SUPER E5-2699 v3.

2.5 Resultados Esperados

A meta deste experimento é demonstrar que um modelo auto-supervisionado pode alcançar um desempenho similar com o benefício de não depender fortemente de rótulos. A expectativa é que o modelo auto-supervisionado, após um ajuste adequado, seja tão eficaz quanto o supervisionado. Além disso, vamos comparar os resultados com outras abordagens da literatura, incluindo métodos que utilizam aprendizado por transferência e otimização de características.

¹<https://numpy.org>

²<https://scikit-learn.org>

³<https://pytorch.org>

⁴<https://matplotlib.org>

Capítulo 3

Plano de Trabalho e Cronograma de Execução

O plano de trabalho é composto pelas atividades descritas na Tabela 3.1.

Cronograma					
Tarefa	Semana				
	1	2	3	4	5
Seleção de uma base de dados	★	☆	☆	☆	☆
Definição da tarefa pretexto	★	☆	☆	☆	☆
Definição da arquitetura de CNN	★	☆	☆	☆	☆
Pré-processamento da base de dados	★	★	☆	☆	☆
Treinamento da tarefa pretexto	☆	★	★	★	☆
Treinamento da tarefa posterior	☆	☆	★	★	☆
Treinamento do modelo supervisionado	☆	☆	★	★	☆
Testes e análise dos resultados	☆	☆	☆	★	★
Documentação dos resultados	☆	☆	☆	☆	★
Apresentação	☆	☆	☆	☆	★

Figura 3.1: Cronograma e atividades.

Bibliografia

- [1] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020.
- [2] F. N. Iandola. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- [3] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131, 2018.
- [4] A. Khan, K. Pin, A. Aziz, J. W. Han, and Y. Nam. Optical coherence tomography image classification using hybrid deep learning and ant colony optimization. *Sensors*, 23(15):6706, 2023.