

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Instituto de Ciências Exatas**  
**Programa de Pós-Graduação em Ciência da Computação**

Felipe Glicério Gomes Marcelino

**Detecção automática de glaucoma de ângulo aberto usando algoritmos de  
aprendizado profundo**

Belo Horizonte  
2023

Felipe Glicério Gomes Marcelino

**Detecção automática de glaucoma de ângulo aberto usando algoritmos de  
aprendizado profundo**

**Versão Final**

Dissertação apresentada ao Programa de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Minas  
Gerais, como requisito parcial à obtenção do título de Mestre  
em Ciência da Computação.

Orientador: Adriano Alonso Veloso

Belo Horizonte  
2023

## [Ficha Catalográfica em formato PDF]

A ficha catalográfica será fornecida pela biblioteca. Ela deve estar em formato PDF e deve ser passada como argumento do comando `ppgccufmg` no arquivo principal `.tex`, conforme o exemplo abaixo:

```
\ppgccufmg{  
    ...  
    fichacatalografica={ficha.pdf}  
}
```

# [Folha de Aprovação em formato PDF]

A folha de aprovação deve estar em formato PDF e deve ser passada como argumento do comando `ppgccufmg` no arquivo principal `.tex`, conforme o exemplo abaixo:

```
\ppgccufmg{  
    ...  
    folhadeaprovacao={folha.pdf}  
}
```



*Dedico este trabalho aos meus pais falecidos, por terem me dado a base e o suporte necessários para que eu me tornasse a pessoa que sou hoje.*

# Agradecimentos

Agradeço, primeiramente, a minha esposa Isabel, por sempre ter me apoiado, oferecido suporte e companheirismo durante a minha jornada acadêmica. Agradeço às minhas tias pelo cuidado que tiveram comigo durante a minha formação e por valorizarem o meu esforço nos estudos. Agradeço ao meu orientador, Adriano, por todo o auxílio e por ter me proporcionado grandes oportunidades durante o meu trajeto acadêmico. Por fim, agradeço à Universidade Federal de Minas Gerais pela sua excelência no ensino e pesquisa.



# Resumo

O glaucoma é uma doença ocular que se desenvolve em decorrência de uma alteração no nervo óptico, que o torna mais frágil. Como resultado, as fibras nervosas são danificadas de forma irreversível, o que faz com que o paciente perca o seu campo visual gradativamente e, nos casos mais avançados, fique cego.

É recomendável uma visita anual ao médico oftalmologista a partir dos 40 com o intuito de descartar a presença da doença, ou, em caso positivo, iniciar o tratamento em suas fases iniciais e os devidos cuidados sejam tomados para que danos maiores ao globo ocular sejam evitados. Quanto mais idoso, maior probabilidade de incidência da doença. Os principais fatores de riscos são: pressão, idade e fatores genéticos.

A Fundoscopia e a Tomografia de Coerência Óptica são exames laboratoriais comumente utilizados para a detecção do glaucoma. Considerando diversos estudos que concluíram que ocorrerá, nos próximos anos, um grande aumento de casos positivos da doença, novos métodos para a detecção e tratamento do glaucoma vêm sendo desenvolvidos, com o objetivo de dar suporte aos especialistas.

Este trabalho estuda a utilização de algoritmos de aprendizado de máquina, através do uso de dados clínicos e de imagens da retina, para detectar o glaucoma de ângulo aberto, com diagnósticos rápidos, eficientes e automatizados. Além disso, o trabalho também apresenta diagnósticos e decisões do modelo que suportam explicabilidade, dando ainda mais recursos para o profissional ter uma decisão mais assertiva junto ao paciente.

Ao final, os experimentos mostraram que diferentes arquiteturas de modelos tiveram assertividade alta quanto ao diagnóstico, inclusive modelos que utilizam recurso computacional de forma mais eficiente, dando flexibilidade de sua utilização em diferentes tipos de infraestrutura computacional disponível, sendo este um smartphone ou conjunto de computadores centralizados.

**Palavras-chave:** Aprendizado de Máquina. Computação Visual. Modelagem. Glaucoma. Diagnóstico. Explicabilidade.

# Abstract

Glaucoma is an eye disease that develops due to a change in the optic nerve, making it more fragile. As a result, the nerve fibers are irreversibly damaged, causing the patient to lose their visual field gradually and, in more advanced cases, to go blind.

An annual visit to the ophthalmologist from the age of 40 is recommended to rule out the presence of the disease or, if positive, to start treatment in its early stages and the proper care taken so that more significant damage to the eyeball is avoided. The older you are, the more likely the disease is to occur. The main risk factors are pressure, age, and genetic factors.

Fundoscopy and Optical Coherence Tomography are laboratory tests commonly used to detect glaucoma. Considering several studies that have concluded that there will be a significant increase in positive cases of the disease in the coming years, new methods for detecting and treating glaucoma are being developed to withstand specialists.

This paper studies the use of machine learning algorithms using clinical and retinal imaging data to detect open-angle glaucoma with fast, efficient, and automated diagnosis. In addition, the work also presents diagnoses and model decisions that withstand explainability, giving even more resources for the professional to have a more assertive decision with the patient.

In the end, the experiments indicated that different model architectures had high assertiveness regarding diagnosis, including models that use computational resources more efficiently, giving flexibility of use in various types of computational infrastructure available, whether a smartphone or a set of centralized computers.

**Keywords:** Machine Learning. Computer Vision. Modeling. Glaucoma. Diagnostic. Explainability

# Lista de Figuras

3.1	Estrutura interna do olho . . . . .	26
3.2	Glaucoma progredindo e danificando o nervo óptico . . . . .	28
3.3	Disco óptico e escavação do disco óptico de um olho normal e outro com glaucoma. . . . .	28
3.4	Disco óptico e escavação do disco óptico de um olho normal e outro com glaucoma e o valor da CDR e a regra ISNT. . . . .	29
3.5	Exemplos de imagem de fundo de olho colorida e fluorescente . . . . .	30
3.6	Exemplos de imagem do fundo de olho com presença e ausência de glaucoma. . . . .	31
3.7	Exemplo de ISNT capturado por uma Fundoscopia. (a) Olho normal, (b) Olho com glaucoma . . . . .	31
3.8	<i>Smartphone</i> com câmera de imagem de fundo de olho. . . . .	32
3.9	Exemplo do exame OCT, sendo o olho direito (OD) com a presença de glaucoma e o esquerdo (OS) em estado saudável. . . . .	33
3.10	Uma rede neural MLP, com 3 camadas. Uma de entrada com 3 neurônios, uma camada intermediária com 4 neurônios e uma de saída com dois neurônios. . . . .	35
3.11	Uma imagem representando o problema de sobreajuste e subajuste. . . . .	39
3.12	Exemplo de uma convolução com filtro $5 \times 5$ . Cada quadrado com este tamanho produz um número utilizando operações matemáticas. Em seguida, o quadrado é deslocado para o lado ou para baixo e um novo filtro é formado. . . . .	40
3.13	Corpo de um modelo RegNet contendo 4 (Fixo) estágios. Cada estágio contém número $X$ de blocos. . . . .	43
3.14	Exemplo de um bloco $X$ baseado em uma rede residual. . . . .	43
3.15	A imagem da esquerda mostra os <i>Design Spaces</i> A, B, e C, sendo o C mais restritivo entre eles. A imagem da direita mostra a distribuição de erro de cada desses espaços, sendo o C aquele que possui o melhor desempenho entre eles. . . . .	44
3.16	Arquitetura de modelos. O <i>Backbone</i> pode ser um dos modelos citados anteriormente (RegNet, ResNet, Inception e “etc”), e o <i>Classifier</i> é uma rede neural de múltiplas camadas (MLP), assim como o modelo <i>MLP</i> que recebe os dados tabelados e também o que faz a operação de concatenar ( <i>concat</i> ) juntando todas as saídas e passando para o classificador. . . . .	45
3.17	AUC e curva ROC. . . . .	48
3.18	Imagem ilustrando a validação cruzada. Cada parte assume o papel de teste uma vez e treino $K-1$ vezes. . . . .	49

3.19	<i>Learning Rate Decay</i> aplicado a modelos visuais. . . . .	51
3.20	Gráfico sumarizado para o conjunto de dados Boston Housing ordenado pelas características mais importantes do modelo. . . . .	53
3.21	Explicação de uma imagem utilizando SHAP. Os pontos vermelhos e azuis mostram onde o modelo julgou como parte importante para a imagem. Seguindo a mesma lógica do 3.20, pontos vermelhos aumentam a probabilidade na saída do modelo, enquanto pontos azuis diminuem. . . . .	54
4.1	Duas imagens do fundo de olho com distâncias diferentes . . . . .	58
4.2	Contagem de casos de glaucoma por gênero e por olhos . . . . .	60
4.3	Distribuição do glaucoma por variável . . . . .	61
4.4	Boxplot do glaucoma por variável. . . . .	62
4.5	Correlação absoluta das características entre 0 e 1, sendo 1 correlação perfeita. . . . .	63
4.6	Correlação das características criadas pela operação razão. . . . .	65
5.1	Avaliação geral dos <i>Backbones</i> divididos por arquiteturas. O gráfico violino mostra a distribuição dos valores máximos atingidos nas 4 arquiteturas de modelos testados. . . . .	68
5.2	Comparação das estatísticas da métrica AUC. O eixo y é o melhor resultado do <i>Backbone</i> durante todas as épocas, enquanto o eixo x é a média do modelo durante o treinamento. . . . .	69
5.3	Distribuição dos desempenhos separados por arquiteturas de modelos. . . . .	71
5.4	Comportamento da função de custo durante as épocas de treinamento. O gráfico mostra um valor médio e um intervalo de confiança de 0.95 . . . . .	72
5.5	Comparação do uso do <i>Learning Rate Decay</i> e <i>Early Stopping</i> nos modelos selecionados na primeira fase. Os melhores modelos da etapa anterior foram re-executados, sem a utilização das técnicas citadas anteriormente, utilizando validação cruzada para uma comparação justa. . . . .	73
5.6	Imagem com explicabilidade do modelo. A coluna da esquerda se refere a arquitetura do modelo utilizando somente uma imagem e, a da direita, utiliza duas imagens. Foto contendo a borda preta . . . . .	78
5.7	Imagem com explicabilidade do modelo. A coluna da esquerda mostra o modelo utilizando somente uma imagem; a da direita, duas imagens. Fonte recortada sem a borda preta . . . . .	78
5.8	Importância das características no melhor modelo em desempenho relativo a AUC. O modelo atingiu AUC de 0.99 usando dados do OCT . . . . .	80
5.9	Mapa de calor importância (importância normalizada entre 0 e 1) das características para cada <i>Backbones</i> presentes na etapa de validação cruzada. As características estão ordenadas de acordo com sua importância média geral. . . . .	81
5.10	Gráfico de cascata para casos com presença de glaucoma. . . . .	82

---

5.11 Gráfico de cascata para casos com ausência de glaucoma. . . . .	83
--	----



# Lista de Tabelas

3.1	Tabela de restrições dos modelos até chegar no modelo RegNet [61]. O índice $i$ é o estágio, fixado em 4. O $b_i$ representa razão do gargalo, $g_i$ representa o total de canais da convolução, $d_i$ representa o total de blocos e o $w_i$ representa o tamanho do bloco. . . . .	43
3.2	Matriz de confusão identificando . . . . .	47
4.1	Tabela contendo sumarização dos dados utilizando as estatísticas: máximo, mínimo, média, desvio e contagem. . . . .	63
4.2	Versões dos modelos visuais utilizados como Backbones e o total de parâmetros de cada um. O modelo ViT é o que possui maior número de parâmetros, aproximadamente 86 milhões, enquanto o modelo ShuffleNet possui somente 2.5 milhões. . . . .	66
5.1	Tabela contendo os resultados quanto à métrica AUC separado por Backbone. . . . .	70
5.2	Tabela contendo os resultados separados por grupos: OCT e Dual Image. A tabela mostra que utilizar duas imagens não foi benéfico, enquanto o uso do exame OCT agrega desempenho aos resultados . . . . .	70
5.3	Tabela com os resultados comparando a utilização da técnica LRD e ES na validação cruzada. Foram calculados a média e o desvio padrão entre os <i>Backbones</i> dos melhores modelos. Os <i>Backbones</i> com desvio padrão igual a 0 significa que só houve uma arquitetura de modelo entre os modelos selecionados na fase anterior. . . . .	74
5.4	Tabela contendo informações relacionadas ao tempo de inferência e o uso de memória. A memória é medida em megabytes (mb) e o tempo de inferência em milissegundos (ms). . . . .	75
5.5	Experimentos utilizando diferentes parâmetros de <i>RandAugment</i> (coluna OP). O número na coluna representa o número de operações seguidas aplicada na imagem. . . . .	76

# Lista de Algoritmos

5.1 Algoritmo utilizado para calcular o tempo de inferência dos modelos. Código retirado de DECI . . . . . 75

,

# Sumário

<b>1</b>	<b>Introdução</b>	<b>16</b>
1.1	Aprendizado de máquina aplicado a diagnósticos com imagens . . . . .	17
1.2	Motivação . . . . .	18
1.2.1	Explicabilidade . . . . .	18
1.2.2	Flexibilidade . . . . .	19
1.3	Declaração de Dissertação . . . . .	19
1.4	Contribuição . . . . .	19
1.5	Organização . . . . .	20
<b>2</b>	<b>Referências Teóricas</b>	<b>21</b>
2.1	Aplicações de Aprendizado de Máquina Geral . . . . .	21
2.2	Aplicações de Aprendizado de Máquina para glaucoma . . . . .	23
<b>3</b>	<b>Fundamentação Teórica</b>	<b>26</b>
3.1	Estrutura Ocular . . . . .	26
3.2	Glaucoma . . . . .	27
3.3	Fundoscopia . . . . .	30
3.4	Tomografia de Coerência Óptica . . . . .	32
3.5	Conceitos de Aprendizado de Máquina . . . . .	34
3.5.1	Camadas . . . . .	34
3.5.2	Pesos ou Parâmetros . . . . .	34
3.5.3	Taxa de aprendizagem . . . . .	35
3.5.4	Treino, Teste e Inferência . . . . .	36
3.5.5	Batch e Época . . . . .	36
3.5.6	Função de custo . . . . .	37
3.5.7	Otimizadores . . . . .	37
3.5.8	Hiperparâmetros . . . . .	38
3.5.9	Sobreajuste . . . . .	38
3.5.10	Convolução . . . . .	39
3.6	Modelos . . . . .	40
3.7	RegNet: O modelo mais flexível para visão computacional . . . . .	41
3.7.1	Hiperparâmetros . . . . .	42
3.7.2	Estrutura . . . . .	42

3.8	Arquiteturas de Modelos . . . . .	44
3.9	Métricas de Avaliação . . . . .	46
3.10	Treinamento e Avaliação dos Modelos . . . . .	49
3.11	<i>Fine-Tuning</i> . . . . .	50
3.12	<i>Learning Rate Decay</i> . . . . .	51
3.13	RandAugment . . . . .	51
3.14	Explicabilidade: SHAP . . . . .	52
3.14.1	Explicabilidade para o glaucoma . . . . .	55
<b>4</b>	<b>Metodologia</b>	<b>56</b>
4.1	Setup . . . . .	56
4.2	Aquisição dos Dados . . . . .	56
4.2.1	CrITÉRIOS de Inclusão . . . . .	57
4.2.2	CrITÉRIO de Exclusão . . . . .	57
4.3	Sumário dos dados . . . . .	57
4.4	Limpeza . . . . .	58
4.5	Exploração . . . . .	59
4.6	Engenharia de características e Normalização . . . . .	63
4.7	Experimentação . . . . .	65
<b>5</b>	<b>Resultados</b>	<b>67</b>
5.1	Desempenho Geral . . . . .	67
5.2	Validação Cruzada . . . . .	72
5.2.1	<i>RandAugment</i> . . . . .	76
5.3	Explicabilidade . . . . .	77
5.3.1	Explicabilidade Visual . . . . .	77
5.3.2	Importância das Características . . . . .	79
5.3.3	Análise de casos individuais . . . . .	82
<b>6</b>	<b>Conclusão</b>	<b>84</b>
	<b>Referências Bibliográficas</b>	<b>85</b>
	<b>Apêndice A Validação Cruzada Resultados</b>	<b>94</b>

# Capítulo 1

## Introdução

O glaucoma é uma doença que acomete os olhos caracterizada pela degeneração das células ganglionares e do nervo óptico (ONH, do inglês *optic nerve head*), o que pode levar a perda progressiva de visão. É uma doença grave que, em seu estágio mais avançado, pode levar a cegueira, e, por isso, faz-se imprescindível que o paciente receba o diagnóstico nas fases iniciais da doença, para os sintomas não evoluam para a perda total e irreversível da visão. Sabe-se, no entanto, que o diagnóstico precoce da doença se torna prejudicado devido às fases assintomáticas da doença. Por isso, o número de pessoas que não recebem um diagnóstico adequado é alto [82], o que contribui para o agravamento da doença e para o paciente ter a perda completa de sua visão.

O glaucoma do tipo ângulo aberto é um dos principais tipos de glaucoma e está relacionado pelo aumento e/ou descontrole da pressão intraocular (PIO). Sendo este tipo de glaucoma o mais comum encontrado em pacientes com mais de 49 anos. É possível utilizar intervenções médicas para diminuir a pressão e/ou controlar a PIO por colírios, laser ou cirurgia, e evitar que a doença avance para estágios mais severos e a perda da visão [9]. Em seguida, está o glaucoma secundário, causado por fatores externos, o glaucoma congênito e o glaucoma de ângulo fechado, sendo este último o mais agressivo entre eles [76].

O diagnóstico do glaucoma é efetuado por especialistas, através da realização de exames clínicos, como a Fundoscopia e a Tomografia de Coerência Óptica. A Fundoscopia é realizada via aparelhos de retinografia. Esse exame é bastante utilizado por não ser invasivo e por ser de fácil acesso ao profissional de saúde [52], sendo possível ser feita por um smartphone e uma lente especial[52, 39]. Este exame de imagem consegue capturar características importantes nas imagens do fundo de olho, como, por exemplo, os vasos sanguíneos e o disco óptico (DO), fornecendo dados úteis para a detecção do glaucoma [52].

A Tomografia de Coerência Óptica (OCT, do inglês *Optical Coherence Tomography*) também é utilizada para avaliar o paciente quanto a presença do glaucoma. Tal procedimento, assim como a Fundoscopia, não é invasivo e gera imagens de cortes seccionais da retina com alta resolução e detalhamento [10]. As imagens seccionais fornecem a espessura da camada de fibras nervosa da retina, possibilitando a verificação da de-

gradação da camada de fibras nervosas retinais do paciente com diagnóstico positivo de glaucoma.

Estudos comprovam que o número de pessoas com glaucoma aumentará em 111,8 milhões até 2040 [54], portanto, métodos que utilizam inteligência artificial vem sendo desenvolvidos para automatizar o processo de diagnóstico do glaucoma [7], dentre eles técnicas baseadas em aprendizado de máquina (ML, do inglês, *Machine Learning*). ML vem se destacando no campo de ciência e tecnologia, tornando-se referência, inclusive na área médica. Uma das principais aplicações e técnicas de aprendizado de máquina utilizado no campo médico são os modelos de classificação. Isto porque o uso de modelos classificadores possibilita o diagnóstico de doenças utilizando dados demográficos, clínicos, imagens e até mesmo áudio. Estes modelos podem ser baseados em técnicas de aprendizado estatístico ou aprendizado profundo [7]. Sistemas inteligentes de diagnósticos conseguem prover interpretações clínicas para auxiliar o especialista em suas decisões quanto aos tratamentos do paciente, aumentando as chances de sucesso [32, 67].

## 1.1 Aprendizado de máquina aplicado a diagnósticos com imagens

O desenvolvimento de novas tecnologias envolvendo inteligência artificial e aprendizado de máquina vem possibilitando grandes avanços em várias áreas do conhecimento, reforçando a importância da interdisciplinaridade para a criação de aplicações como solução eficiente de problemas complexos. Dentre as soluções, técnicas de reconhecimento de voz, reconhecimento de objetos, predição de demanda e veículos autônomos têm se mostrado eficientes. Na área médica, além do diagnóstico, é possível utilizar IA para descobrir novas drogas, melhorar a comunicação entre médico e paciente, transcrever documentos (como receita de medicamentos) e realizar cirurgias e tratamentos a longa distância.

O aprendizado de máquina é uma área de IA que utiliza dados como pilar central em suas aplicações. Em alguns casos, o desempenho de modelo de ML consegue ser melhor do que a do humano, como, por exemplo, na detecção de objetos [65], jogos eletrônicos [49], imitação de estilo e de arte [24] e em previsão comportamental [25].

O presente trabalho analisa o uso de diversos modelos de ML para diagnóstico de glaucoma através do estudo de artigos publicados entre o ano 2014 e 2019 em que existe um problema de classificação (presença ou ausência do glaucoma) [6]. Os modelos analisados se dividem em métodos que extraem as características da imagem e fazem redução de dimensionalidade do espaço amostral e modelos de aprendizado profundo utilizando redes

convolucionais (CNN, do inglês, *Convolution Neural Network*)[29] para análise de imagens. A maior influência da escolha do método se resume ao número de dados disponíveis e o recurso computacional disponível.

## 1.2 Motivação

Este trabalho apresenta uma análise detalhada do uso de diferentes arquiteturas de modelos de aprendizado profundo que consigam auxiliar no diagnóstico do glaucoma de ângulo aberto eficientemente. Foram exploradas diferentes arquiteturas de modelos, com diferentes técnicas de sofisticação e uso de recurso computacional. Podendo este recurso ser supercomputadores ou combinações de computadores (*cluster*<sup>1</sup>) ou, até mesmo, um smartphone com acesso a uma câmera. Além do diagnóstico, o uso dos modelos também traz o benefício da explicabilidade, fornecendo racionalidade na decisão.

### 1.2.1 Explicabilidade

A explicabilidade dos modelos será explorada tanto nas arquiteturas mais simples, utilizando somente uma imagem do fundo de olho, quanto nas arquiteturas mais complexas, que utilizam múltiplas imagens do fundo de olho e as mensurações extraídas pelo exame OCT do paciente. O especialista irá entender quais foram os fatores ou características mais importantes para a decisão do modelo e o quão influente, em magnitude, cada uma delas foram.

Para explicar os modelos, foi utilizada uma biblioteca conhecida como SHAP [68]. Esta biblioteca será abordada na seção 3.14 com maiores detalhes, contudo, a ideia do SHAP é calcular o impacto médio de cada uma das características da imagem ou dos dados clínicos quanto ao diagnóstico do modelo.

---

<sup>1</sup>Um *cluster* (do inglês *cluster*: 'grupo, aglomerado') consiste em computadores fortemente ligados trabalhando em conjunto, de modo que, em muitos aspectos, podem ser considerados um único sistema. Diferentemente dos computadores em grade, computadores em *cluster* têm cada conjunto de nós, para executar a mesma tarefa, controlado e programado por software. Fonte: Wikipedia

### 1.2.2 Flexibilidade

O Brasil, por ser um país de escala continental, carece de acesso à saúde em locais remotos ou que não se adéquem a realidade social daqueles que lá vivem [17]. Soluções que sejam acessíveis em questão de custo e mobilidade são as melhores opções a serem utilizadas. Com isso, o trabalho, além de trazer o aspecto explicativo para o diagnóstico da doença, também irá trazer diferentes arquiteturas que se encaixam em diferentes usos e acessos a recursos computacionais.

Regiões remotas do Brasil, onde a infra-estrutura de saúde possui poucos recursos acessíveis, modelos que consigam ser executados em dispositivos móveis como um celular são recomendados. Outra opção é adotar um servidor central que ficará responsável por executar todo o processo de diagnóstico e tratamento dos dados e retornar os resultados para o especialista responsável. Cada uma das opções anteriores possui seus benefícios, mas como requerimento, independentemente da opção, modelos eficientes e com baixo custo são necessários.

## 1.3 Declaração de Dissertação

As principais perguntas a serem respondidas neste projeto são: **(Q1)** Modelos computacionalmente eficientes e flexíveis (em termos de acesso à infra-estrutura) desempenham a tarefa, de forma eficaz, igual, ou melhor, do que modelos computacionalmente mais caros e menos flexíveis? **(Q2)** De acordo com a explicabilidade do modelo, quais são os fatores ou características mais pertinentes quanto a sua decisão?

## 1.4 Contribuição

Este projeto contribui com diferentes modelagens de aprendizado profundo e análise dos resultados para o diagnóstico do glaucoma de ângulo aberto. Os modelos aqui apresentados são focados em tarefas que envolvem computação visual, ou seja, quando se utiliza de imagens para resolver uma tarefa.

- Automatização do processo de detecção do glaucoma utilizando modelos de apren-



dizado profundo

- Flexibilidade e explicabilidade
  - Diferentes arquiteturas de modelos implementadas, dando ênfase nas arquiteturas mais eficientes e eficazes
  - Utilização do SHAP para trazer explicabilidade para os modelos
- Aplicação e experimentação de técnicas de aprendizado de máquina para criação de modelos mais robustos e performáticos (*RandAugment/Early Stopping*)
- Análise de resultados quanto as métricas escolhidas e também quanto ao uso de recurso computacional e a eficiência do modelo

**Observação:** A partir deste momento o glaucoma de ângulo aberto será identificado somente como glaucoma como forma de simplificar a escrita e a leitura do texto.

## 1.5 Organização

O segundo capítulo 2 contém referências teóricas quanto a utilização de modelos de aprendizado de máquina para diagnóstico e detecção de doenças. O terceiro capítulo 3 explicita conceitos de aprendizado de máquina e do glaucoma, das métricas utilizadas, das arquiteturas testadas e das técnicas aplicadas. O quarto capítulo 4 contém todos os passos realizados antes da experimentação dos modelos, que envolve processamento e aquisição dos dados, configurações de hiperparâmetros e softwares utilizados. O quinto capítulo 5 contém os resultados dos experimentos realizados e suas avaliações. Também está incluído a explicabilidade do modelo. O último capítulo 6 conclui esta dissertação.

# Capítulo 2

## Referências Teóricas

O uso de inteligência artificial vem revolucionando o mercado mundial, base para o desenvolvimento de diversas aplicações inovadoras. O aprendizado de máquina, área promissora da inteligência artificial, tem sido cada vez mais estudado por pesquisadores de todo o mundo, devido às contribuições trazidas a múltiplas áreas do conhecimento. Na medicina, vários tipos de modelos de aprendizado de máquina auxiliam os profissionais de saúde na detecção de doenças e outras aplicações, beneficiando não só os especialistas da área, mas os pacientes e a sociedade como um todo.

Na seção 2.1 serão avaliados alguns trabalhos voltados para o diagnóstico de doenças de forma geral, enquanto na seção 2.2 serão apresentados trabalhos focados no diagnóstico do glaucoma.

### 2.1 Aplicações de Aprendizado de Máquina Geral

Modelos de redes neurais convolucionais possuem múltiplas aplicações, inclusive na detecção de doenças de pele [84]. Os autores exploram diferentes arquiteturas de redes neurais convolucionais, entre elas estão RegNet [83] e EfficientNet [77], que aparecem como modelos mais eficientes, e também modelos tradicionais como ResNet [35] e VGG [45]. Os autores também trouxeram uma versão modificada do *RandAugment*, aplicado neste projeto e explicitado na seção 3.13, para melhorar o desempenho do modelo utilizando técnicas de *Data Augmentation*. Utilizando um conjunto de dados extremamente desbalanceado, os autores implementaram uma função de custo (Seção 3.5.6) que considera o desbalanceamento dos dados. Ao final, os resultados dos diferentes modelos serão apresentados. Os modelos computacionalmente mais eficientes, como o RegNet, demonstraram um desempenho próximo, ou melhor, do que os modelos menos eficientes computacionalmente (ResNet, VGG) com uma acurácia balanceada de 0.858.

É possível identificar doenças nas folhas da maçã utilizando modelos profundos, conforme citado no artigo [42]. Na referida pesquisa, os autores utilizaram um conjunto de

dados que possui um total de 2200 imagens, aproximadamente, e 5 diferentes doenças para serem detectadas. Os modelos testados foram ShuffleNet, EfficientNet-B0, MobileNetV3, RegNet e ViT [89, 77, 83, 37, 92]. Dentre os modelos citados, o que atingiu a melhor acurácia de 0.99 foi o modelo RegNet. Todos os modelos foram treinados com 4 diferentes otimizadores, que serão explicados na seção 3.5.7, e 5 diferentes taxas de aprendizagem (Seção 3.5.3).

A partir da combinação de imagens e dados demográficos/clínicos, modelos de aprendizado profundo foram utilizados para detectar o câncer de mama [75]. Uma rede neural convolucional (Seção 3.5.10) processa múltiplas imagens com diferentes ângulos de uma tomografia do tórax. Cada imagem é processada por um modelo independente, ou seja, eles não compartilham os parâmetros entre si (Seção 3.5.2). Os dados clínicos, contendo idade, histórico familiar em relação ao câncer e presença de diabetes, são processados por outra rede neural multicamadas (Seção 3.6). Ao final, os dados processados pela rede neural convolucional e da rede neural multicamadas são concatenados por outra rede neural multicamadas e, em seguida, o modelo efetua o diagnóstico quanto a presença de câncer. Os autores concluem que a adição de um modelo convolucional com múltiplas entradas (*multi-input*) e dados clínicos conseguem trazer resultados melhores do que o modelo utilizando somente uma imagem (*single-input*) frontal do tórax, com uma AUC de 0.99 e acurácia de 0.97.

Utilizando modelos de aprendizado profundo, os autores conseguiram detectar doenças cardiovasculares [3]. Os modelos foram divididos em dois grupos: *single-input* e *multi-input*. As arquiteturas de modelo *single-input* utilizaram o a rede neural Inception-V3 [74] quando os dados são imagens, e redes neurais multicamadas com modelos de árvore de decisão [8] quando os dados são clínicos. A arquitetura *multi-input* utiliza o Inception-V3 para os dados de imagem e o de redes neurais multicamadas para processar os dados clínicos e, também, como classificador. A arquitetura de modelo *multi-input* atingiu uma acurácia 0.79, enquanto o *single-input* atingiu uma acurácia de 0.75.

Com o uso de *Visual Transformers* (ViT) apresentados na seção 3.6, foi possível a detecção de pneumonia nos pulmões, inclusive quando advindo da Covid-19 [13]. As imagens foram coletadas de uma tomografia computadorizada e divididas em vários pedaços menores. Cada um destes pedaços foram codificados no modelo ViT e, em seguida, repassados para uma rede neural multicamadas para detectar a presença de Pneumonia (ocasionada pelo vírus da COVID-10 ou outro vírus). O modelo dos autores apresentou uma acurácia de 0.97 na detecção da doença.

O trabalho apresentado a seguir traz uma forma diferente de trabalhar com redes neurais convolucionais utilizando o modelo DCGAN [70]. Os modelos DCGAN são baseados em modelos neurais generativos [29, 30]. O intuito destes modelos é utilizar dados reais para criar dados que sejam sintéticos e similares aos dados reais. Para isso, o modelo se divide em duas partes: codificador e decodificador. O modelo codificador

aprende uma representação vetorial dos dados de entrada, no caso as imagens, enquanto o decodificador processa essa representação vetorial e tenta recriar uma imagem parecida com a original. Neste artigo [70], os autores utilizam o codificador para criar a representação vetorial mais próxima possível as imagens do fundo de olho. Em seguida, os autores utilizaram essa representação vetorial em um modelo classificador baseado em regressão logística para detectar a presença de Alzheimer nos pacientes. Os dados utilizados pelo modelo DCGAN são imagens neurais multimodais.

## 2.2 Aplicações de Aprendizado de Máquina para glaucoma

Conforme os autores, que utilizaram redes neurais convolucionais, mais especificamente uma ResNet34, o modelo conseguiu prever a espessura das camadas de fibras nervosas do disco óptico [48]. Para evitar que haja qualquer tipo de vazamento de dados entre os dados de treino e de teste (Seção 3.10), a divisão dos dados foi feita por paciente, desta forma, nenhum dado de paciente está presente no treino e no teste em simultâneo. As imagens foram padronizadas com altura e largura de  $256 \times 256$  píxeis, e os valores de píxeis foram normalizados entre 0 e 1. Foram utilizadas algumas técnicas de *Data Augmentation* para aumentar a variabilidade das amostras do conjunto de dados, diminuindo a chance de sobreajuste do modelo (Seção 3.5.9). Um segundo modelo classificador binário foi treinado para detectar anormalidades (ou normalidades) na espessura da camada de fibras utilizando fotografias do disco óptico. As métricas utilizadas foram MAE (*Mean Absolute Error*) para predição da espessura da camada de fibras nervosas e AUC (*Area under the ROC curve*) para a classificação binária. Foram utilizados aproximadamente 1200 pacientes como amostras e um total de 2312 olhos. Os modelos apresentaram resultados de 0.944 para curva ROC, uma acurácia de 0.84 e erro de predição da espessura da camada de fibras nervosas de aproximadamente 7,39 micromilímetros.

Neste trabalho, a imagem do fundo de olho foi utilizada para extrair uma característica utilizando redes CNN, que o autor deu o nome de *CNN Degree* [53]. Além da característica extraída pelas redes CNNs, foram também utilizados dados clínicos para um segundo modelo, com o papel de diagnosticar o glaucoma, como, por exemplo: espessura da camada de fibras nervosas (temporal, nasal, inferior, superior), espessura da córnea, tamanho da escavação do disco óptico (CDO, do inglês, *Physiologic Cup* ou *Optic Cup*) e do disco óptico, idade, sexo do paciente, etc. Para o modelo diagnosticador, foram utilizados modelos com estruturas diferentes. Dentre eles, estão: Support Vector Machine

(SVM), Random Forest, C5.0 e XGBoost. As métricas utilizadas foram AUC, ROC, sensibilidade e especificidade. Os autores constataram que o melhor modelo foi o XGBoost, apresentando AUC igual a 0.945, sensibilidade de 0.950 e especificidade de 0.945. Além disso, os autores apresentaram explicabilidade para o modelo, sendo a característica mais importante a espessura das camadas das fibras nervosas da retina superior (*RFNLSuperior*). Pacientes com glaucoma apresentam valores menores para essa característica e pessoas saudáveis apresentam valores altos. As características apontadas pelo modelo como a segunda e terceira mais importantes são *RNFLInferior* (espessura da camada de fibras nervosas inferior) e a pressão intraocular (PIO).

Outro trabalho que vale a pena ser analisado para a detecção do glaucoma é a aplicação do modelo Inception-V3 [74] para detecção do glaucoma [43]. Os autores trazem técnicas de *Data Augmentation* (Seção 3.13) para aumentar heterogeneidade dos dados. As fotos foram padronizadas para o tamanho  $299 \times 299$  e os píxeis foram normalizados entre 0 e 1. Quanto ao *Data Augmentation*, foram aplicadas as seguintes operações: rotações randomizadas de 90, 180 e 270 graus e também um deslocamento entre 0 e 3 píxeis na horizontal. Este último é aplicado para aumentar o detalhe de captura das imagens sem a necessidade de aumentar o número de píxeis das imagens. Quanto aos hiperparâmetros do treino, o autor utilizou o otimizador Adam [40] com uma taxa de aprendizagem de 0.002 e um, tamanho de *batch* de 32 imagens. Estes hiperparâmetros serão explicados na seção 3.5. Com um total de 48 mil imagens e em um intervalo de confiança de 0.95, o autor atinge um desempenho de AUC entre 0.984 e 0.988. Além disso, o autor aponta quais foram as características que mais geraram falso negativo:

- O glaucoma normalmente aparece com outras doenças, como, por exemplo, miopia, retinopatia diabética e degeneração macular devido à idade.
- Defeito nas camadas das fibras nervosas da retina ou disco óptico causado por uma hemorragia.

Para um uso mais extensivo ao público, foi implementado uma framework de diagnóstico do glaucoma utilizando a combinação de dados públicos e privados com acesso ao modelo de aprendizado de máquina profundo [23]. Esta framework<sup>1</sup> é acessível para o público geral, que envia as fotos do fundo de olho e, em seguida, recebe o resultado decisivo do modelo. Quando negativo, o resultado é enviado diretamente para o paciente. Quando positivo, este é enviado para um oftalmologista para uma análise mais profunda e o paciente encaminhado. A framework é dividida em duas partes: a parte pública, onde

---

<sup>1</sup>Um framework em desenvolvimento de software, é uma abstração que une códigos comuns entre vários projetos de software provendo uma funcionalidade genérica. Um framework pode atingir uma funcionalidade específica, por configuração, durante a programação de uma aplicação. Ao contrário das bibliotecas, o framework é quem dita o fluxo de controle da aplicação, chamado de Inversão de Controle. Fonte: Wikipedia

as pessoas podem carregar as imagens do fundo de olho e a parte privada, onde é realizado todo o processamento e a geração do relatório. O acesso público pode ser pela web ou por um dispositivo MobileNet. O resultado é baseado na proporção entre o disco e escavação do disco óptico: *cup-to-disc ratio* (CDR). Se o valor da CDR é maior que 0.5 o modelo retorna a possibilidade de glaucoma. Para segmentar e classificar a imagem, os modelos utilizados foram: EfficientNet [77] (Codificador) e UNet++ [92] (Decodificador). O codificador cria um mapa de características que será decodificado, pelo decodificador, de forma que cada segmento da imagem seja separado e o valor da CDR seja calculado. O acesso ao banco de dados também é manuseado por um servidor público, de forma que todas as pessoas tenham acesso. Enquanto o processamento da imagem é feito por um servidor privado.

O seguinte trabalho faz uma comparação da acurácia do modelo de aprendizado de máquina com fotos coletadas nos centro ópticos e fotos coletadas utilizando um smartphone e um aparelho acoplado D-Eye Lens [50]. O autor descreve que a utilização do smartphone possibilita que mais pessoas sejam avaliadas devido ao fácil acesso e locomoção do dispositivo. Para captar as fotos pelo smartphone, um vídeo de 1 minuto é gravado, e do vídeo. A partir do vídeo, uma imagem é extraída, sendo esta imagem a que melhor captura o disco óptico. Isto possibilita que pacientes com demência também sejam diagnosticados. Os autores utilizaram uma ResNet para classificar a presença do glaucoma. Quanto aos dados, foram utilizadas técnicas de *Data Augmentation*, abordado na seção 3.13, para diversificar o dado e por consequência melhorar o desempenho do modelo quanto a generalização. Dentre as operações realizadas nas imagens, estão: rotação da imagem (10 graus), translação vertical e horizontal, mudança de contraste e saturação de forma randômica e variação do tamanho das bordas da imagem. O modelo apresentou um desempenho AUC de 0.98 para a câmera padrão e 0.84 para o uso do smartphone. Os autores discutem que a diferença dos desempenhos, que apesar de serem suficientemente bons para uma análise, se dá pela capacidade de foco da câmera, gerando imagens que nem sempre possuem qualidade adequada para uma avaliação mais assertiva.

## Capítulo 3

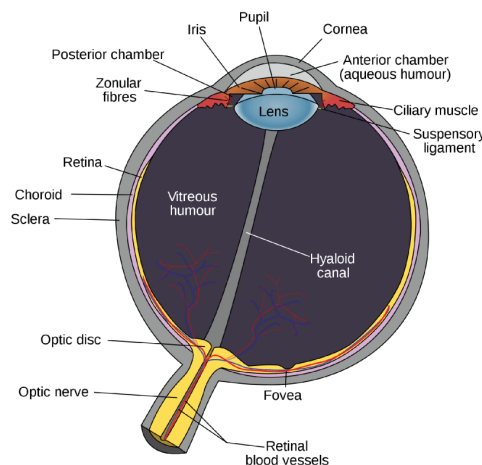
# Fundamentação Teórica

### 3.1 Estrutura Ocular

O conhecimento sobre a anatomia do olho humano é importante para um melhor entendimento dos resultados que serão apresentados no capítulo 5. O olho é um órgão sensorial complexo, capaz de capturar relações espaciais e temporais de algum objeto, captando a energia luminosa refletida e transformada em sinais elétricos. A energia luminosa passa pela retina e através do nervo óptico, onde se encontra o disco óptico (DO), a energia transformada em impulsos elétricos é enviada para o cérebro [55].

A figura a seguir 3.1 mostra o olho humano com uma geometria similar a um globo. O globo ocular possui formato esférico, revestido por três camadas: externa, intermediária e interna.

**Figura 3.1: Estrutura interna do olho**



Fonte: Imagem retirada de [72]

A retina (do inglês, *retina*) é a membrana mais interna do olho e se estende por toda a porção posterior da parede, composta por células fotossensíveis (cones e bastonetes) que recebem o estímulo, luminoso, transformando-o em estímulo elétrico. Desta forma, a retina é responsável pela sensação da imagem visual projetada pelas estruturas frontais

do olho [55, 28, 56].

Na retina existe uma pequena região conhecida como fóvea (do inglês, *fovea*) onde a imagem se torna mais detalhada e nítida. A fóvea possui a capacidade de capturar as cores dos objetos e se encontra na região central da mácula (do inglês, *macula*) [36].

A luz que penetra no olho percorre o meio óptico composto pela córnea (do inglês, *cornea*), íris (do inglês, *iris*), cristalino (do inglês, *lens*), humor vítreo (do inglês, *vitreous humour*) e humor aquoso (do inglês, *aqueous humour*) antes de chegar na retina. A íris é responsável por controlar a quantidade de luz que penetra no olho através da pupila (do inglês, *pupil*) [56, 55].

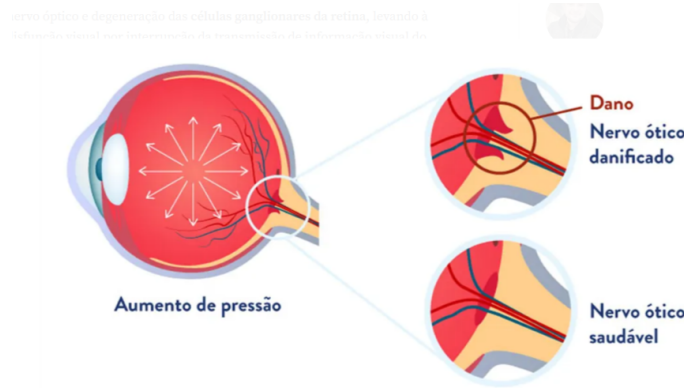
## 3.2 Glaucoma

O glaucoma é uma neuropatia que degenera as células ganglionares da retina (RGC, do inglês, *retinal ganglion cells*) e seus axônios, resultando em uma aparência diferenciada para a região central do nervo óptico, conhecida como escavação papilar ou escavação do disco óptico [81].

Dentre os principais fatores relacionados ao aparecimento do glaucoma estão: idade acima de 49 anos, portadores de miopia, pessoas de origem africana, pessoas com histórico familiar da doença, pressão intraocular alta, pessoas com diabetes, pessoas que utilizam ou já utilizaram esteroides/cortisona, pessoas lesionadas visualmente, com hipertensão arterial alta e desregulação vascular [51, 27].

A incidência do glaucoma está diretamente relacionada ao humor aquoso, um fluido que circula na parte interna do olho, com o objetivo de nutrir a região da íris e da córnea e exercer pressão para manter o formato do olho. Esse líquido é secretado constantemente e, quando não é drenado adequadamente, aumenta a pressão intraocular. Consequentemente, o nervo óptico é danificado, como mostra a figura abaixo.



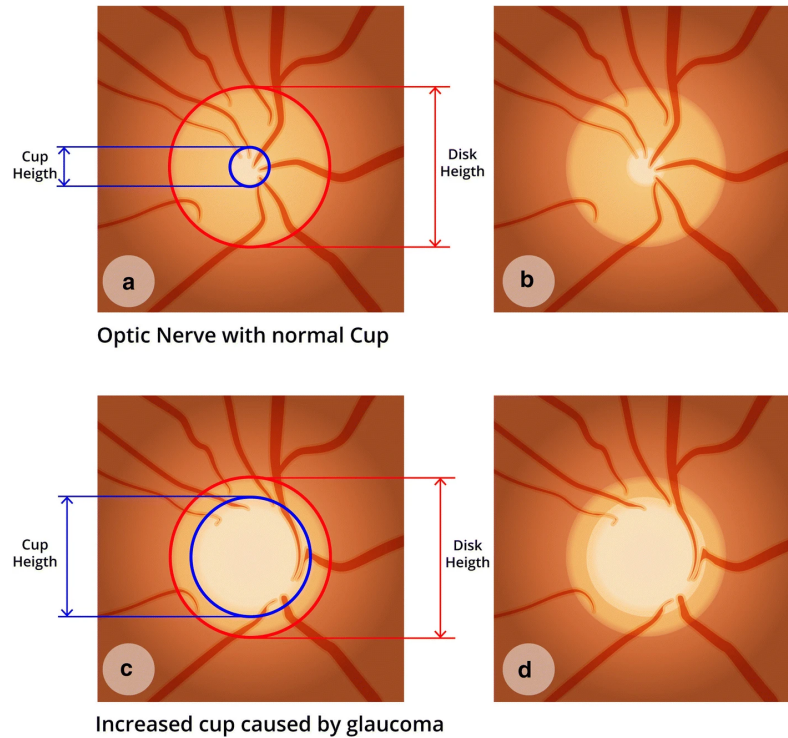
**Figura 3.2: Glaucoma progredindo e danificando o nervo óptico**

Fonte: Imagem retirada de [Medium](#)

Uma das formas de detecção do glaucoma é o cálculo da razão conhecida como *cup-to-disc ratio* (CDR) [2].

$$CDR = \frac{CDO}{DO} \quad (3.1)$$

A equação 3.1 mostra como é feito o cálculo do valor da CDR que nada mais é do que a razão do diâmetro vertical da escavação do disco óptico e do diâmetro vertical do disco óptico.

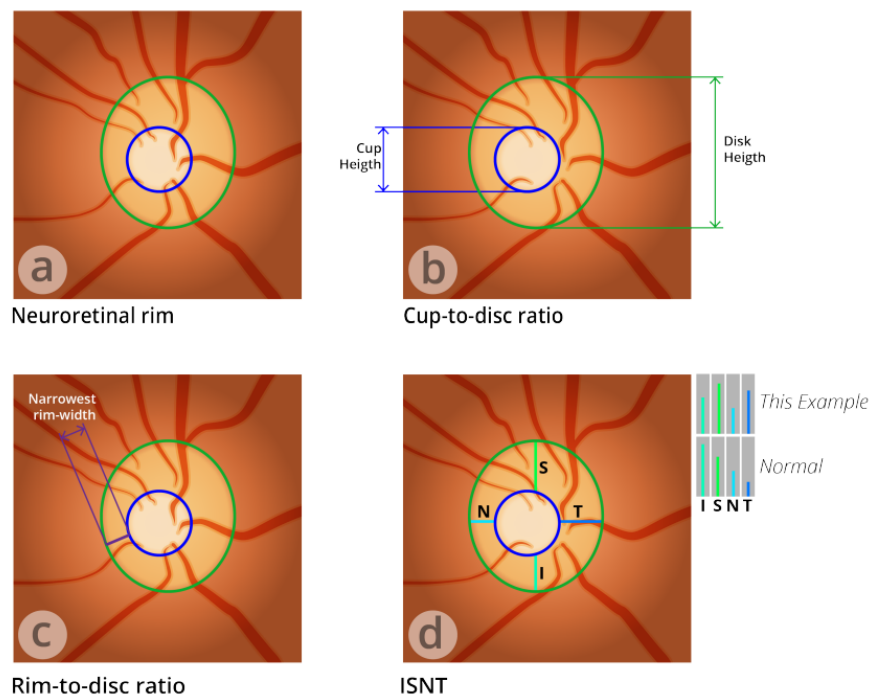
**Figura 3.3: Disco óptico e escavação do disco óptico de um olho normal e outro com glaucoma.**

Fonte: Imagem retirada de [6]

A figura acima 3.3 possui 4 imagens de 2 retinas diferentes. As duas imagens na linha inferior possui a escavação do disco óptico modificada pela presença do glaucoma. Nestas duas imagens, a altura do diâmetro da escavação do disco óptico (*cup height*) está próxima da altura do diâmetro do disco, e conseqüentemente, a equação 3.1 assume valores maiores. Casos positivos do glaucoma podem ser considerados quando a razão for maior 0.5 [2] ou 0.7 [90]. De qualquer forma, este método é sensível à diferença de iluminação e o especialista pode demorar de 1 até 3 minutos para fazer uma análise subjetiva [18].

Outra forma de fazer o diagnóstico do glaucoma é utilizando a regra ISNT (Inferior, Superior, Nasal e Temporal) [34]. Para cada quadrante da região do disco óptico é calculado uma energia *wavelet* através de diferentes sub-bandas, usando diferentes famílias *wavelet*. Em seguida, é coletado a espessura da camada de fibras nervosas de cada uma das regiões. Um olho sem a presença do glaucoma deve apresentar a seguinte ordem de espessura da camada de fibras: inferior  $\geq$  superior  $\geq$  nasal  $\geq$  temporal.

**Figura 3.4: Disco óptico e escavação do disco óptico de um olho normal e outro com glaucoma e o valor da CDR e a regra ISNT.**



Fonte: Image retirada de [6]

A figura 3.4 mostra o valor da CDR e a ordem ISNT do olho glaucoma. O olho com glaucoma não segue a ordem estabelecida para um olho saudável quanto a ordenação da espessura da camada de fibras nervosas nos 4 setores ISNT.

Para se obter as duas informações citadas anteriormente, ISNT e CDR, os exames de Fundoscopia e Tomografia de Coerência Óptica são utilizados para diagnosticar os pacientes com suspeita de glaucoma. Estes dois exames serão detalhados nas duas próximas seções.

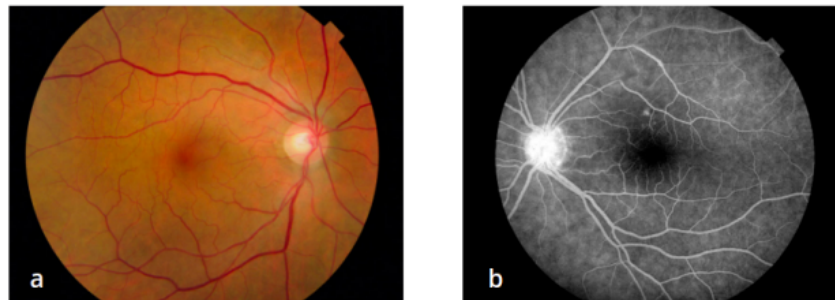
### 3.3 Fundoscopia

Atualmente, um dos recursos mais utilizados para detectar a doença é a imagem da retina ou imagem do fundo de olho (Fundoscopia). A Fundoscopia é um exame que avalia a parte posterior do olho e, através da imagem, é possível verificar o formato e o tamanho do disco óptico e da escavação do disco óptico, já que este último sofre alterações na presença da doença [88, 71, 5, 93]. Além disso, a Fundoscopia detecta os vasos sanguíneos, a mácula e a fóvea, dando recursos para detectar outras doenças e lesões como, por exemplo, microaneurismas, sangramentos, exsudatos e manchas de algodão [79].

Existem dois tipos de imagens geradas pela Fundoscopia apresentada na figura 3.5 [16]:

- **Colorida:** Documenta possíveis alterações retinianas por intermédio das imagens, realizado em toda a faixa do espectro visível, permitindo que o DO seja capturado.
- **Fluorescente ou Angiofluoresceinografia:** Usa uma iluminação especial e filtros, podendo ser realizado com ou sem contraste, ajudando a detectar melhor os vasos sanguíneos.

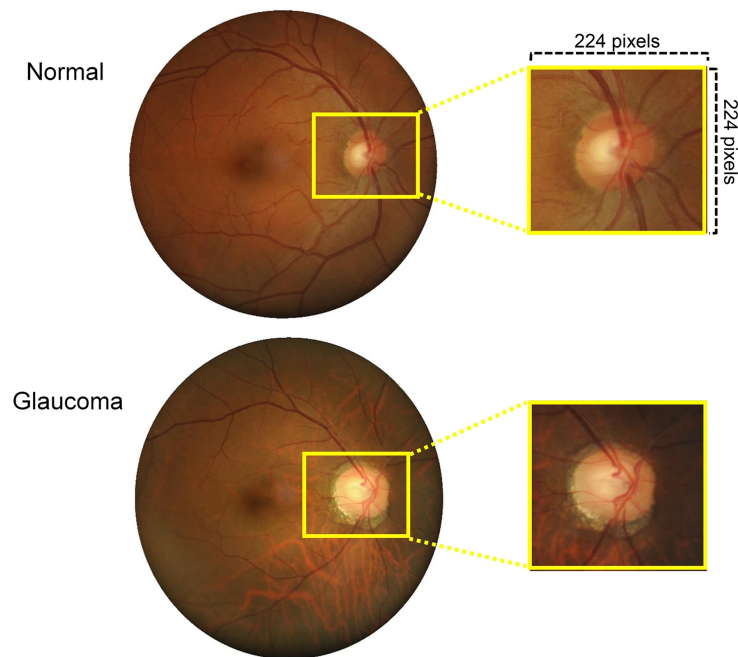
**Figura 3.5: Exemplos de imagem de fundo de olho colorida e fluorescente**



Fonte: Imagem retirada de [16]

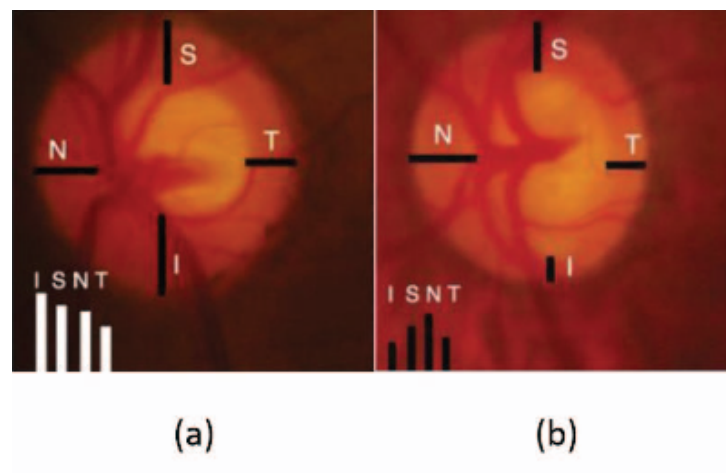
A figura 3.6 traz duas imagens de fundo de olho, uma delas com a presença do glaucoma. É possível conferir na imagem inferior do fundo de olho, em que há o diagnóstico de glaucoma, o aumento do valor da CDR. Já a figura 3.7 traz a diferença de espessura do ISNT da escavação do disco.

**Figura 3.6:** Exemplos de imagem do fundo de olho com presença e ausência de glaucoma.



Fonte: [Nature, 2021](#)

**Figura 3.7:** Exemplo de ISNT capturado por uma Fundoscopia. (a) Olho normal, (b) Olho com glaucoma

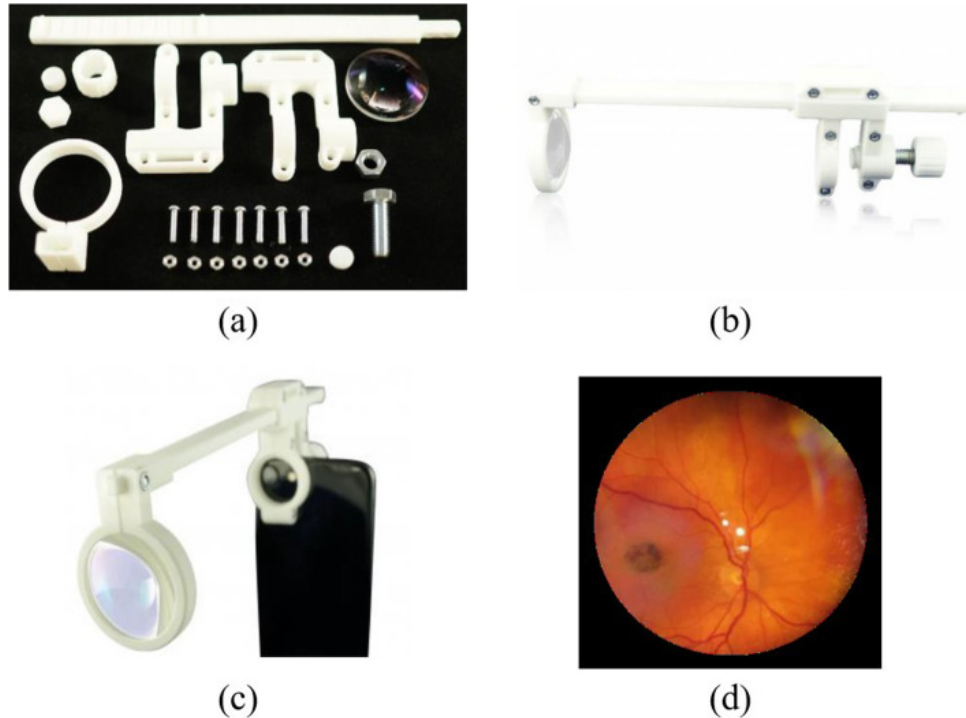


Fonte: Imagem retirada de [\[64\]](#)

Por mérito do avanço tecnológico de dispositivos móveis e o uso de redes neurais eficientes, é possível que o glaucoma seja detectado utilizando *Smartphones* [\[11\]](#). Utilizando uma lente e um smartphone (Android ou IOS) os autores desenvolveram um aplicativo que extraí a imagem do fundo de olho e, em seguida, um modelo de aprendizado de máquina profundo faz a segmentação do disco e da escavação do disco óptico. A segmentação fornecia os valores da CDR, e por fim, o diagnóstico do glaucoma era dado conforme o valor da razão, atingindo uma acurácia de 0.78. A figura a seguir mostra os dispositi-

vos (imagem a e b) utilizados pelos autores para detectar o glaucoma com a aplicação desenvolvida:

**Figura 3.8: *Smartphone* com câmera de imagem de fundo de olho.**



Fonte: Imagem retirada de [31]

## 3.4 Tomografia de Coerência Óptica

O exame OCT, Tomografia de Coerência Óptica, mensura a espessura da camada de fibras nervosas da retina peri-papilares assim como o tamanho da escavação do disco óptico. Estas medições fornecem aos profissionais inúmeras informações, dando mais recurso para uma melhor avaliação do glaucoma.

A obtenção de imagens pelo exame OCT baseia-se na técnica de interferometria<sup>1</sup>. O interferômetro óptico usa o princípio de comparação entre feixes de luz para alta resolução de estruturas, ou seja, mede o tempo de atraso em escala de femtossegundos. Um feixe é refletido e o outro é transmitido ao olho, refletido por estruturas oculares a partir de distâncias axiais variadas. O feixe de luz que reflete do olho é composto por ecos que informam a distância e espessura das estruturas interiores dos olhos. O segundo feixe é refletido a partir de um espelho de referência situado a uma determinada distância. A

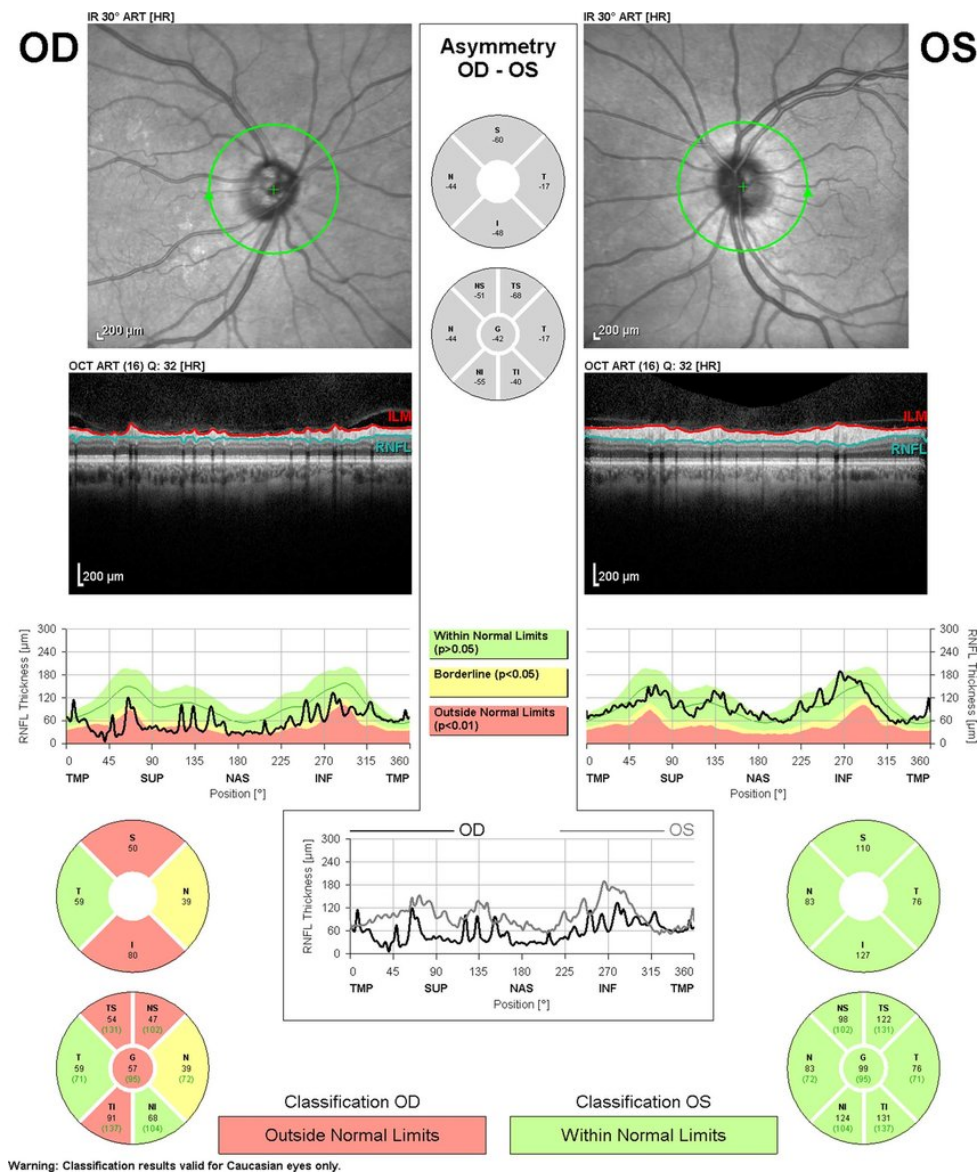
<sup>1</sup>Interferometria é uma técnica que utiliza a interferência de ondas sobrepostas para extrair informações. Fonte: Wikipedia



chave do processo consiste no fato de que o interferômetro pode medir o tempo de atraso de ecos ópticos pela comparação do feixe de luz refletido com o feixe de luz padrão de referência [69].

A figura 3.9 mostra um exame OCT, com suas respectivas medidas de espessura da camada de fibras nervosas da retina, e o valor G, sendo a média dos setores. Quanto menor o valor G, maior a probabilidade de se ter glaucoma. A camada de fibras nervosas é dividida em 6 setores: nasal (N), nasal superior (NS), nasal inferior (NI), temporal (T), temporal superior (TS) e temporal inferior (TI). A figura nos mostra que o olho direito está com as camadas de fibras nervosas danificadas e pode ter glaucoma, enquanto o olho esquerdo está saudável.

**Figura 3.9: Exemplo do exame OCT, sendo o olho direito (OD) com a presença de glaucoma e o esquerdo (OS) em estado saudável.**



Fonte: [Glaucoma Today, 2016](#)

## 3.5 Conceitos de Aprendizado de Máquina

### 3.5.1 Camadas

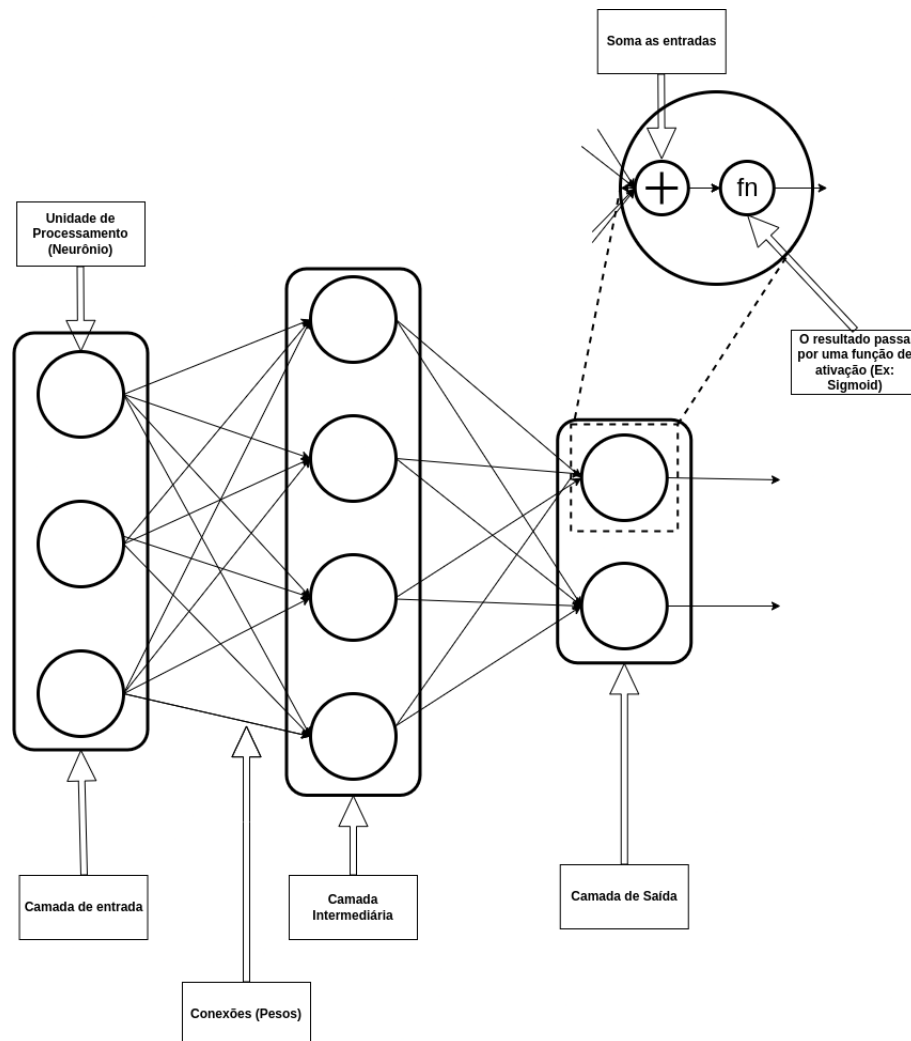
As camadas são onde se encontram os neurônios ou unidades de processamento. Uma rede neural normalmente possui múltiplas camadas, ou, pelo menos, duas camadas de entrada e saída. Quanto mais camadas na rede neural, mais profunda ela é, ou seja, mais neurônios são utilizados. No entanto, é necessário um cuidado, pois quanto mais neurônios mais operações matemáticas são realizadas e, por consequência, mais dados são necessários para utilizar o modelo, demandando mais poder computacional.

### 3.5.2 Pesos ou Parâmetros

As redes neurais são compostas por várias unidades de processamento, cujo funcionamento é o mais simples possível. As unidades de processamento das redes neurais, conhecidas também como neurônios, são interligadas entre si, e cada uma dessas unidades possui um parâmetro associado, também conhecido como peso. As unidades de processamento fazem operações matemáticas com os dados recebidos de outras unidades de processamento ou das tabelas de dados se a unidade estiver na primeira camada do modelo, também conhecida como camada de entrada (do inglês, *input layer*).

A figura 3.10 representa uma rede neural multicamada (MLP, do inglês, *Multi Layer Perceptron*) contendo três camadas. Uma camada de entrada, uma camada intermediária e uma camada de saída. Cada uma dessas camadas possui um número  $N$  de neurônios. Cada neurônio, ou unidade de processamento, possui conexões com outras unidades, sendo que essa conexão seria o peso ou parâmetro. O parâmetro então é multiplicado pela saída de um neurônio que entra em um próximo neurônio. No neurônio, todas as entradas são somadas. Em seguida, o resultado da soma passa em uma função de ativação, como, por exemplo, uma sigmoid [29], para transformar ou limitar os valores. Todos os neurônios de uma camada possuem ligações com todos os neurônios da próxima camada.

**Figura 3.10:** Uma rede neural MLP, com 3 camadas. Uma de entrada com 3 neurônios, uma camada intermediária com 4 neurônios e uma de saída com dois neurônios.



Fonte: Imagem elaborada pelo autor.

### 3.5.3 Taxa de aprendizagem

A taxa de aprendizagem (LR, do inglês, *Learning Rate*), é o valor de atualização dos parâmetros. Toda vez que uma rede neural erra sua previsão ou resposta, é necessário que estes parâmetros se ajustem, para diminuir o valor do erro em uma próxima iteração. A taxa de aprendizagem é a magnitude desta atualização. Quanto mais alta a taxa de aprendizagem, maior será o valor adicionado ou subtraído do parâmetro. É necessário balancear o valor desta taxa, para as atualizações acontecerem de acordo e em direção da diminuição do erro do modelo [29].



### 3.5.4 Treino, Teste e Inferência

O processo para se obter e usar um modelo de aprendizado de máquina se divide em três fases: treino, teste e inferência. O treino é a fase onde se ajusta os parâmetros do modelo para que ele complete uma tarefa da melhor forma possível, que neste projeto seria diagnosticar o glaucoma. O teste é a fase onde irá ser avaliado o desempenho do modelo, utilizando as métricas previamente estabelecidas. E a fase de inferência é quando um modelo já treinado e avaliado segundo as métricas estabelecidas já está sendo utilizado pelo público alvo. Normalmente, os modelos na fase de inferência são colocados na nuvem<sup>2</sup> ou em alguma infraestrutura computacional local para poderem ser acessados por diferentes aplicações e profissionais que queiram agregar valor ao negócio [29].

Cada uma dessas fases trabalha com um conjunto de dados único que não se intersectam. É necessário ressaltar que se deve ter cuidado em dividir o conjunto de dados do treinamento e do teste para não haver vazamento de dados entre os dois conjuntos e o modelo atinja um desempenho que não representa a realidade. O conjunto de dados da inferência é coletado quando o modelo já está em produção, ou seja, todos aqueles dados que não estavam presentes no conjunto de treino e teste e ainda não possuem um *ground truth*<sup>3</sup>. Após um certo tempo, normalmente os dados coletados na inferência são adicionados aos dados de treino e teste e o modelo é retreinado já que o modelo pode sofrer de *Distribution Shift*<sup>4</sup>.

### 3.5.5 Batch e Época

O treinamento se divide em épocas. Cada época é um processamento completo dos dados de treino e de teste pelo modelo. Logo, consegue-se atualizar seus parâmetros e avaliar o modelo. No entanto, não é possível passar todos os dados para o modelo de uma única vez devido ao custo computacional e ao tamanho limitado da memória. Portanto, os dados são separados em *batches* que nada mais são que um subgrupo de um grupo de dados. O tamanho deste subgrupo é ajustado conforme a tarefa e recursos computacionais disponíveis. Há algumas restrições quanto a este tamanho. O ideal é que

---

<sup>2</sup>Infraestrutura de computação necessários para rodar uma aplicação que seja acessada através da internet. Os recursos virtuais refletem uma infraestrutura física local, com servidores, memórias, clusters e armazenamentos. Tudo isso voltado para uma aplicação que seja escalável e de fácil uso.

<sup>3</sup>*Ground truth* é uma informação estabelecida como real utilizando métricas ou através da observação.

<sup>4</sup>*Distribution Shift* é quando a distribuição de dados muda durante um tempo devido às sazonalidades dos dados ou mudança de comportamento do grupo de amostra.

o tamanho dos *batches* seja menor que a totalidade dos dados do conjunto e maior do que 1. Neste último caso, o modelo precisará fazer por cada amostra processada e usará recurso computacional excessivamente, além de que, a generalização poderá ser afetada para pior, já que 1 amostra dificilmente irá representar ou aproximar minimamente a distribuição do conjunto de dados [29].

### 3.5.6 Função de custo

A função de custo mede o quão longe o modelo está da resposta correta presente nos dados. Também conhecida como função de perda, esta função irá indicar a qualidade do modelo e quais são os valores de atualização necessários para os parâmetros utilizando um número real como indicador. O que esta função faz é comparar a *ground truth* com a decisão do modelo e será analisado a proximidade ou igualdade entre dois elementos. Neste projeto, a *ground truth* é a classe dos dados, indicando presença ou ausência do glaucoma.

### 3.5.7 Otimizadores

Os otimizadores são algoritmos ou funções responsáveis por atualizar os parâmetros dos modelos usando o número gerado pela função de custo. O otimizador calcula os melhores valores para cada um dos parâmetros visando diminuir o erro do modelo em uma devida tarefa. Para realizar este cálculo são necessários os parâmetros do modelo, a taxa de aprendizado e o erro calculado na função de custo. O otimizador conseguinte escolhe a melhor atualização para os parâmetros. Todavia, é essencial escolher o melhor otimizador para cada uma das tarefas ou modelos a serem utilizados, não existindo um otimizador que seja global e eficiente em todos os contextos.

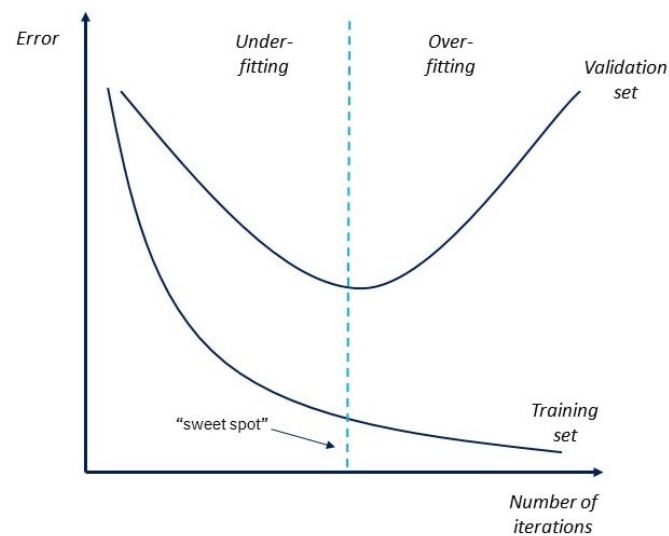
### 3.5.8 Hiperparâmetros

Os hiperparâmetros dos modelos são configurações ligadas ao processo de aprendizado. Por exemplo, como citado anteriormente, a taxa de aprendizagem (Seção 3.5.3) é um hiperparâmetro a ser escolhido e sinaliza a magnitude de atualização dos parâmetros. Outro hiperparâmetro é o tamanho do *batch* (Seção 3.5.5) e também qual otimizador (Seção 3.5.7) será utilizado [29]. Normalmente, para encontrar os melhores hiperparâmetros para o modelo é necessário repetir o processo de treino e teste múltiplas vezes, tornando este processo extremamente caro.

### 3.5.9 Sobreajuste

O sobreajuste (do inglês, *overfitting*) é um cenário onde o modelo se ajusta de forma quase que perfeita para um conjunto de dados  $X$  e, por consequência, acaba tendo uma perda de desempenho no conjunto de dados  $Y$  que possui uma distribuição diferente de  $X$ . O ideal é que o modelo consiga generalizar bem para todos os conjuntos de dados disponíveis, no entanto, é praticamente impossível, já que qualquer comportamento diferente nos dados pode causar distribuições totalmente diferentes entre eles. Logo, se os conjuntos de dados  $X$  e  $Y$  possuem distribuições próximas, mas não iguais, é possível utilizar de um para conseguir inferir o outro. Portanto, se o modelo sobreajusta para  $X$  e tenta inferir  $Y$ , este modelo irá ter um erro grande quanto a sua inferência, já que apesar de as distribuições serem próximas, estas diferem entre si [29].

Para se evitar o sobreajuste do modelo, este projeto aplicou uma técnica conhecida como *Early Stopping* (ES) (Seção 3.10). A técnica tem como intuito parar o processo de treinamento quando o modelo começa a aumentar seu desempenho no conjunto de dados de treino ( $X$ ) e sofre uma piora significativa no conjunto de dados de teste ( $Y$ ) utilizado para avaliação. A figura 3.11 a seguir mostra visualmente como acontece o sobreajuste.

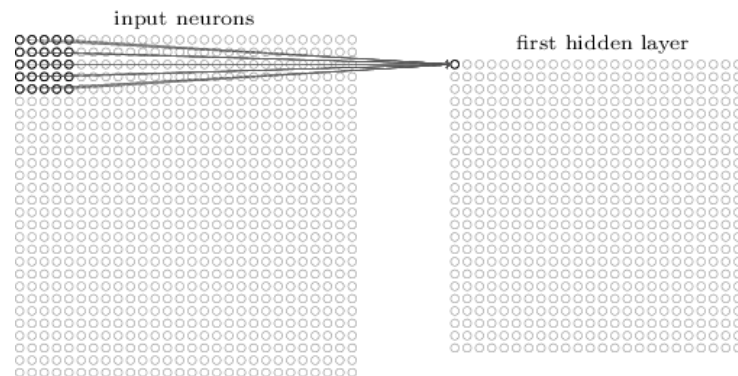
**Figura 3.11:** Uma imagem representando o problema de sobreajuste e subajuste.

Fonte: IBM

### 3.5.10 Convolução

As convoluções funcionam como filtros que enxergam pequenos quadrados e vão “escorregando” por toda a imagem, captando os traços mais marcantes. Explicando melhor, com uma imagem  $32 \times 32 \times 3$  e um filtro que cobre uma área de  $5 \times 5$  da imagem com movimento de 2 saltos (chamado de *stride*), o filtro passará pela imagem inteira, por cada um dos canais, formando no final um *feature map* ou *activation map* de  $28 \times 28 \times 1$ . A imagem 3.12 mostra como uma convolução funciona.

**Figura 3.12: Exemplo de uma convolução com filtro  $5 \times 5$ . Cada quadrado com este tamanho produz um número utilizando operações matemáticas. Em seguida, o quadrado é deslocado para o lado ou para baixo e um novo filtro é formado.**



Fonte: [Yale](#)

A profundidade da saída de uma convolução é igual à quantidade de filtros aplicados (total de canais). Quanto mais profundas são as camadas das convoluções, mais detalhados são os traços identificados com o *activation map*. No entanto, quanto maior o número de canais, maior será o número de operações matemáticas e, conseqüentemente, uma demanda maior de hardware.

## 3.6 Modelos

Os modelos mencionados a seguir são modelos focados em visão computacional. Todos os modelos são baseados em redes neurais convolucionais, com exceção do modelo Vision Transformers (ViT). Com intuito de trazer mais flexibilidade ao diagnóstico de glaucoma, foram testados modelos com número total de parâmetros (Seção 3.5.2) variável e por consequência, o uso computacional destes modelos. Alguns dos modelos conseguem ser executados em dispositivos menores, como, por exemplo, um smartphone. Outros modelos trazem rápida inferência quando é necessário que a resposta do modelo seja eficiente. Toda a experimentação é com objetivo de trazer diferentes arquiteturas de redes neurais que se encaixam em diferentes contextos e infraestrutura disponível. Logo, os modelos testados foram: RegNet, MobileNet, ShuffleNet, EfficientNet, ResNet, Inception-V3, ViT [83, 37, 89, 35, 74, 21, 77]

Cada um dos modelos citados anteriormente possui suas vantagens. A MobileNet, como o próprio nome já diz, é otimizada para rodar em celulares ou dispositivos menores do que um computador [37]. As redes MobileNet, são menores dos que as redes convolucionais tradicionais, utilizando menos energia para atender recursos limitados e também possuem baixa latência para ter respostas rápidas. A EfficientNet é bastante eficiente

em tarefas que envolvem o reconhecimento visual, dando respostas rápidas. Estas redes podem ser escaladas facilmente para serem utilizadas em larga escala ou melhorar seu desempenho sem trazer custos adicionais exagerados e de forma simples [77]. A ShuffleNet traz uma rede neural que seria o extremo de redução de custo tentando manter a acurácia do modelo. A ShuffleNet pode ter uma eficiência 13 vezes maior que uma ResNet sem perder muito do seu desempenho quanto a acurácia do modelo. As ShuffleNet possuem o número de operações por segundo reduzidas (MFLOps<sup>5</sup>) e também o tempo de resposta [89].

Quanto aos modelos mais custosos, foram testados ViT, Inception-V3 e ResNet. A ResNet traz um modelo robusto com o foco em assertividade alta devido ao número de parâmetros [35]. As ViT são redes neurais que conseguem trazer capacidade de aprendizado mesmo com um número pequeno de dados, mas mantendo o poder computacional das redes neurais convolucionais mais profundas, como a ResNet, citada anteriormente. Como toda rede neural, quanto mais dados disponíveis, melhor será o poder de generalização da rede, no entanto, redes neurais ViT apresentam resultados melhores utilizando técnicas de *Self-Attention* [41, 21]. A Inception-V3 utiliza um classificador auxiliar que evita o problema de *Vanish Gradient*, ou, quando as camadas mais próximas da entrada não conseguem modificar seus parâmetros, com objetivo de diminuir o erro, devido à profundidade da rede.

E, além dos modelos convolucionais citados anteriormente, foram utilizados também modelos de redes neurais de múltiplas camadas (MLP, do inglês, *Multi Layer Perceptron*). As redes MLP são redes neurais mais simples, no entanto, com alto poder de generalização. Neste projeto, as MLP serão utilizadas na parte de classificação e concatenação dos resultados das arquiteturas que serão citadas na próxima seção 3.8.

## 3.7 RegNet: O modelo mais flexível para visão computacional

O modelo RegNet é um tipo de rede neural criada via um *Network Design Space* (NDS). Diferente dos modelos de arquitetura padrão, em que uma arquitetura de rede neural foi descoberta por NAS<sup>6</sup> [94], o RegNet é um espaço contendo todas as redes neurais possíveis com as configurações de hiperparâmetros disponíveis. Ou seja, NDS fornece um

<sup>5</sup>Medida de operação de ponto flutuante por segundo. Quanto menor este número para uma rede neural, menos energia é gasta para serem calculados os resultados.

<sup>6</sup>*Neural Architecture Search* é uma técnica de automatizar o processo de busca de arquitetura de redes neurais, desempenhando melhor do que redes neurais montadas manualmente.

mecanismo de testagem de arquiteturas mais sofisticado, sem a necessidade de testar cada uma das arquiteturas, mas sim uma amostragem robusta do espaço, reduzindo custo computacional e, simultaneamente, fornece modelos simples e eficazes.

O NDS é um conjunto de possíveis arquiteturas em um espaço de busca. Em cada iteração, esse espaço de amostragem é restringido pelos hiperparâmetros, com o intuito de querer obter redes neurais com desempenho maior ou pelo menos igual, no entanto, utilizando uma arquitetura muito mais simples. Desta forma, com a amostragem, verifica-se que utilizando uma certa combinação de grupo de hiperparâmetros, é possível testar se uma população amostral de arquitetura de redes neurais consegue ter um bom desempenho em tarefa ao invés de procurar uma arquitetura única que traga o melhor resultado possível, conforme utilizado na metodologia NAS [94].

A composição dos hiperparâmetros e estruturas de uma RegNet serão abordados nos seguintes tópicos.

### 3.7.1 Hiperparâmetros

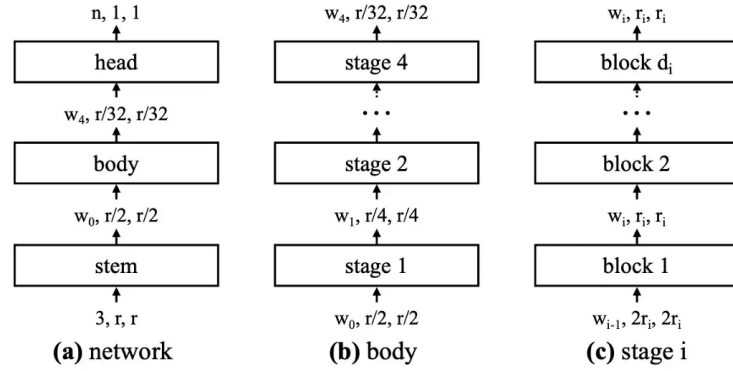
- Total de blocos: Cada bloco é composto por um grupo de convoluções (Seção 3.5.10).
- Largura do bloco: Total de canais (Seção 3.5.10) de cada um das convoluções.
- Razão do gargalo: Utilizado para diminuir o número de canais entre uma convolução e outra, diminuindo o total de operações a serem executadas.
- Largura do grupo: Total de operações de convoluções para cada bloco.

### 3.7.2 Estrutura

- Tronco (*Stem*): Contém as camadas (Seção 3.5.1) de entrada da rede neural.
- Corpo (*Body*): A parte principal da rede neural, sendo esta a estrutura influenciada pelos parâmetros citados acima.
- Cabeça (*Head*): Camada de saída do modelo, contendo o classificador.

A figura a seguir 3.13, apresenta um modelo RegNet.

**Figura 3.13: Corpo de um modelo RegNet contendo 4 (Fixo) estágios. Cada estágio contém número  $X$  de blocos.**



Fonte: Imagem retirada do artigo [61].

Um ponto de atenção citado pelos autores é a decisão de manter todos os blocos iguais e com a mesma estrutura de uma ResNet [35]. A imagem 3.14 a seguir exemplifica:

**Figura 3.14: Exemplo de um bloco  $X$  baseado em uma rede residual.**

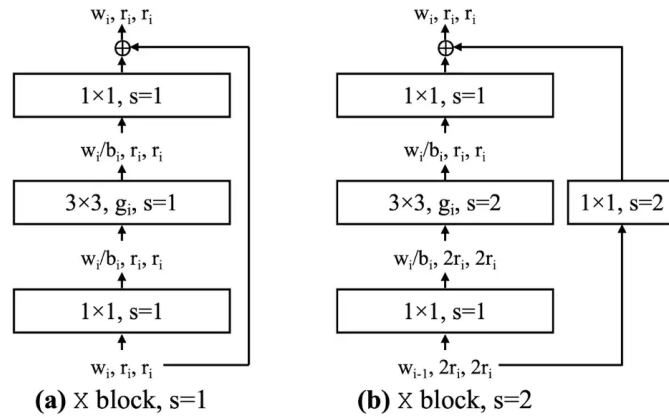


Imagem retirada do artigo [61].

**Tabela 3.1: Tabela de restrições dos modelos até chegar no modelo RegNet [61]. O índice  $i$  é o estágio, fixado em 4. O  $b_i$  representa razão do gargalo,  $g_i$  representa o total de canais da convolução,  $d_i$  representa o total de blocos e o  $w_i$  representa o tamanho do bloco.**

Versão	Restrições	Dim	Combinações	Total
AnyNet <sub>A</sub>	none	16	$((16 \cdot 128 \cdot 3 \cdot 6)^4)$	$\sim 1.8 \cdot 10^{18}$
AnyNet <sub>B</sub>	$(+b_i + 1 = b_i)$	13	$((16 \cdot 128 \cdot 6)^4 \cdot 3)$	$\sim 6.8 \cdot 10^{16}$
AnyNet <sub>C</sub>	$(+g_i + 1 = g_i)$	10	$((16 \cdot 128)^4 \cdot 3 \cdot 6)$	$\sim 3.2 \cdot 10^{14}$
AnyNet <sub>D</sub>	$(+w_i + 1 \geq w_i)$	10	$((16 \cdot 128)^4 \cdot 3 \cdot 6 / (4!))$	$\sim 1.3 \cdot 10^{13}$
AnyNet	$(+d_i + 1 \geq d_i)$	10	$((16 \cdot 128)^4 \cdot 3 \cdot 6 / (4!)^2)$	$\sim 5.5 \cdot 10^{11}$
RegNet	quantized linear	6	$(\sim 64^4 \cdot 6 \cdot 3)$	$\sim 3.0 \cdot 10^8$

Fonte: Designing Design Network Spaces

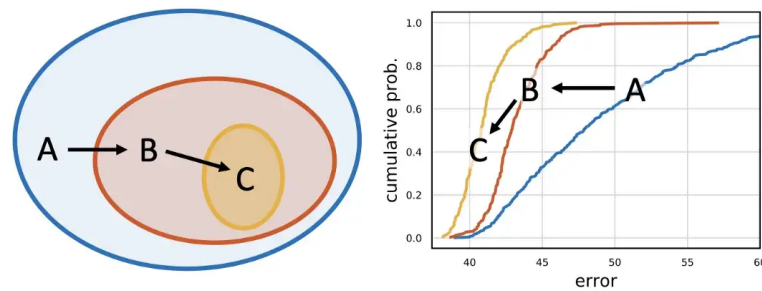
A tabela 3.1 explica um pouco como o modelo RegNet é construído. Como citado anteriormente, são infinitas as combinações de arquitetura de redes neurais, a depender



dos hiperparâmetros a serem testados. A primeira arquitetura de rede neural, AnyNetA, não possui nenhuma restrição, totalizando, aproximadamente,  $1.8 \cdot 10^{18}$  combinações possíveis. A cada nova versão de AnyNet mais restrições são adicionadas ao NDS. No último modelo, após adicionar todas as restrições e avaliar todas as AnyNets, o modelo enfim chega no RegNet utilizando uma metodologia conhecida como *quantized linear parameterization*, para que todos os blocos do mesmo estágio tenham o mesmo tamanho.

O desempenho dos modelos amostrados do NDS é calculado por um procedimento que se resume em treinar o modelo no conjunto de dados ImageNet [20] por 10 épocas (Seção 3.5.5) e calcular o erro de inferência nesse mesmo conjunto de dados. É amostrado em torno de 100 modelos para cada NDS com seu conjunto de parâmetros e restrições.

**Figura 3.15:** A imagem da esquerda mostra os *Design Spaces* A, B, e C, sendo o C mais restritivo entre eles. A imagem da direita mostra a distribuição de erro de cada desses espaços, sendo o C aquele que possui o melhor desempenho entre eles.



Fonte: Imagem retirada do artigo [61].

Ao final, os autores chegam a conclusão que os modelos amostrados do espaço RegNet possuem desempenho melhor do que os modelos amostrados do AnyNet para qualquer valor dos parâmetros escolhidos. Além disso, os autores apresentam o modelo RegNetY que é uma variação do RegNetX contendo uma técnica conhecida como *squeeze-and-excitation* [38] para que cada bloco adapte os pesos para capturar as partes mais importantes da imagem com um adicional de custo computacional baixo.

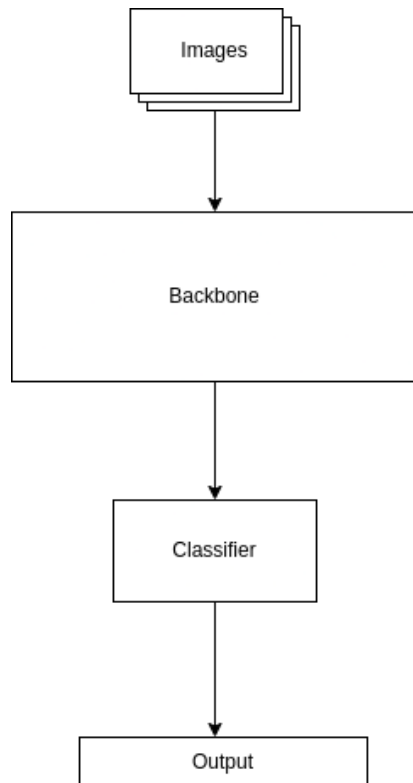
## 3.8 Arquiteturas de Modelos

Neste projeto, 4 arquiteturas de modelos foram testadas. Diferentes das arquiteturas de redes neurais, uma arquitetura de modelo combina uma ou mais redes neurais. Cada uma das arquiteturas de modelos apresentados a seguir será testada com diferentes combinações de hiperparâmetros. O intuito de testar diferentes arquiteturas é que cada uma delas pode se sair melhor a depender da tarefa a ser realizada e dos dados disponíveis.

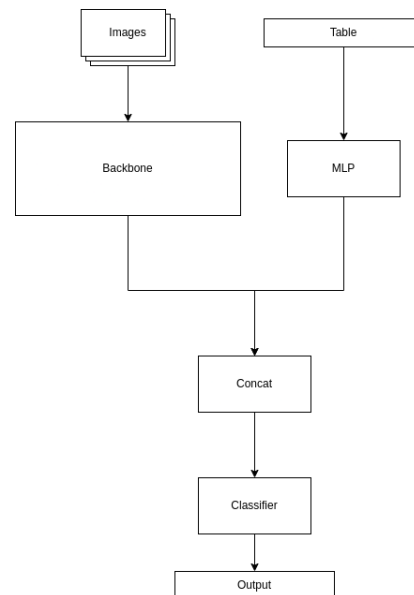
Foram testadas as seguintes arquiteturas mostradas na figura 3.16.

**Figura 3.16: Arquitetura de modelos.** O *Backbone* pode ser um dos modelos citados anteriormente (RegNet, ResNet, Inception e “etc”), e o *Classifier* é uma rede neural de múltiplas camadas (MLP), assim como o modelo *MLP* que recebe os dados tabelados e também o que faz a operação de concatenar (*concat*) juntando todas as saídas e passando para o classificador.

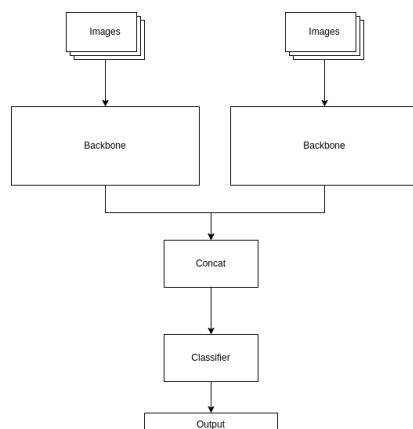
(a) Arquitetura utilizando uma imagem de olho



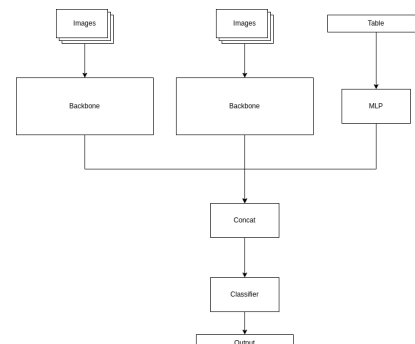
(b) Arquitetura utilizando uma imagem do olho e os dados tabulados



(c) Arquitetura utilizando duas imagens do mesmo olho



(d) Arquitetura utilizando duas imagens do mesmo olho e os dados tabulados



Fonte: Imagem elaborada pelo autor.

Todas as arquiteturas de modelos citados na figura 3.16, representados por um

retângulo e com exceção do *Backbone*, são modelos MLP. Os modelos *Backbone*, com tradução literária de *espinha dorsal*, são os modelos principais utilizados para extrair os padrões mais importantes dos dados de imagens de fundo de olho.

A primeira arquitetura 3.16a usou somente uma imagem do fundo de olho. Entre as arquiteturas citadas na figura 3.16 esta é a arquitetura mais simples que utiliza a menor quantidade de recurso computacional. A segunda arquitetura, além de uma imagem do fundo de olho, também utilizou os dados tabelados do exame OCT 3.16b. Esta segunda arquitetura junta os dados de imagens e os dados clínicos para trazer mais informações e ajudar na classificação/diagnóstico da presença/ausência do glaucoma.

A terceira arquitetura 3.16c utilizou duas imagens do fundo de olho. Para que isso fosse possível, foi feita uma limpeza dos dados e uma avaliação manual de cada uma das fotos para cada paciente da base de dados. Foram escolhidas dois tipos de imagens: uma mais distante, contendo as bordas pretas; outra mais focada e próxima do disco óptico. A quarta e última arquitetura é a que utiliza duas imagens do fundo de olho e dados clínicos. Esta é a arquitetura que utiliza mais recurso computacional, já que o modelo de imagem é duplicado (*Backbone*), utilizando o dobro de memória e uma rede MLP para processar os dados do exame OCT e demográficos.

Um **ponto de atenção** é que a arquitetura de modelo, que utiliza somente uma imagem, seleciona aleatoriamente um dos dois tipos de imagens disponíveis para processar. Desta forma, este modelo aprende a trabalhar com imagens com bordas e sem bordas. Para o modelo que utiliza duas imagens, cada tipo de imagem passa por um *Backbone* diferente. Logo, o mesmo padrão de imagem deve passar pelos seus respectivos *Backbones* em todas as etapas e iterações, inclusive no momento de inferência.

## 3.9 Métricas de Avaliação

Para a modelagem de diagnóstico na área médica, as métricas certas devem ser utilizadas para que se possa avaliar a qualidade do modelo com robustez. Para isso, faz-se necessário o entendimento dos seguintes conceitos: Verdadeiros Positivos (TP, do inglês, *True Positive*), Falsos Negativos (FN, do inglês, *False Negative*), Falsos Positivos (FP, do inglês, *False Positive*) e Verdadeiros Negativos (TN, do inglês, *True Negative*) [19], conforme explicitado a seguir:

- Verdadeiros Positivos (TP): Casos positivos de glaucoma, identificados pelo modelo como positivo.

- Falsos Negativos (FN): Casos positivos de glaucoma, identificados pelo modelo como negativo.
- Falso Positivos (FP): Casos negativos de glaucoma, identificados pelo modelo como positivo.
- Verdadeiros Negativos (TN): Casos negativos de glaucoma, identificados pelo modelo como negativo.

**Tabela 3.2: Matriz de confusão identificando**

		Verdadeiro	
		Glaucoma	Normal
Predição	Glaucoma	TP	FP
	Normal	FN	TN

Fonte: Tabela retirada de [19]

A tabela acima sumariza os conceitos de TP, TN, FP e FN utilizando uma matriz de confusão 3.2. Abaixo, verificam-se as métricas utilizadas para a avaliação das arquiteturas de modelos sugeridas:

- Especificidade (SP): É a capacidade do modelo em detectar casos que são negativos.

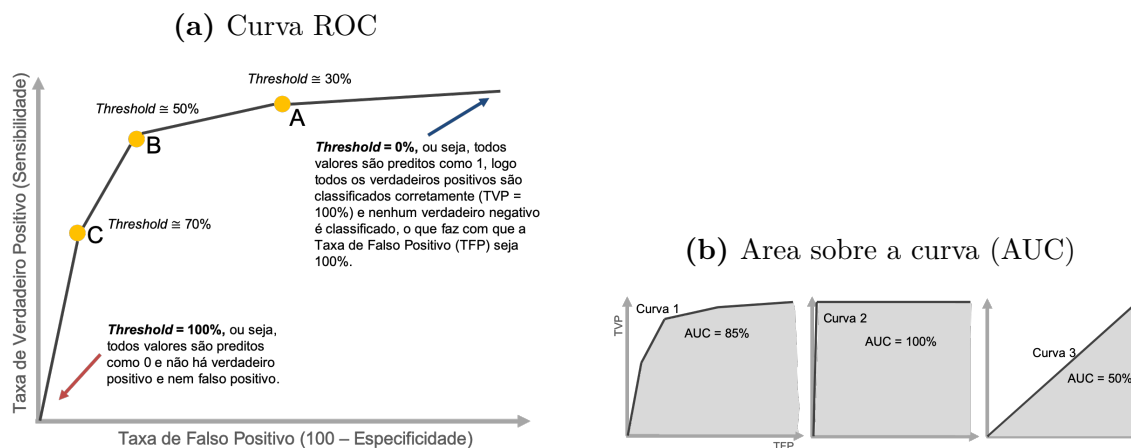
$$\frac{TN}{TN + FP} \quad (3.2)$$

- Sensibilidade (SN): É a capacidade do modelo em detectar todos os casos verdadeiros de glaucoma.

$$\frac{TP}{TP + FN} \quad (3.3)$$

- Característica de Operação do Receptor (ROC): Esta curva ajuda os profissionais e desenvolvedores do modelo a escolher um limiar de corte entre sensibilidade e especificidade que concorde com o problema de negócio. Esta métrica é a mais importante e utilizada em diagnósticos, pois como citado anteriormente, é possível se ter uma flexibilidade consoante a regra de negócio, dando mais importância para sensibilidade ou para especificidade [33].
- Área sob a curva ROC (AUC): Permite verificar o quão bom o modelo classificador binário desempenha no problema. O cálculo é feito utilizando a área sob a curva ROC. Quanto maior este valor, mais flexível o modelo consegue ser quanto a especificidade e sensibilidade. Esta métrica varia de 0 a 1, sendo o melhor modelo aquele que tem valor igual a 1.

Figura 3.17: AUC e curva ROC.



Fonte: [Ciências e Negócios, 2020](#)

Para selecionar o melhor modelo entre todos os experimentados, foi utilizada a AUC como métrica principal, já que esta métrica consegue sumarizar, utilizando somente um número, o quão bom a curva ROC se encontra em cada um dos modelos. A figura 3.17 explica, de forma ilustrativa, como a curva ROC e a AUC funcionam. Em aplicações médicas, é mais válido trazer o maior número de verdadeiros positivos sem aumentar significativamente a taxa dos falsos positivos. Normalmente, é mais prejudicial deixar de diagnosticar uma doença do que dar um diagnóstico positivo falso. Neste caso, o interesse maior é na sensibilidade. No entanto, se abaixarmos o *threshold* demais, o número de falsos positivos aumenta significativamente, o que faz com que muitos casos analisados sejam enviados para a verificação de um especialista, gerando um maior custo, mão de obra e de tempo 3.17a. As duas métricas, sensibilidade e especificidade, interagem entre si. Quando se aumenta uma, movimentando threshold, é possível que se perca desempenho na outra. Contudo, o valor da métrica AUC carrega consigo a informação de variação desse ganho e dessa perda, ou seja, quanto melhor o AUC, mais fácil movimentar o threshold ganhando de um lado e abdicando menos do outro. O modelo perfeito, com 100% de AUC, na imagem da direita 3.17b, mostra que se movimentarmos o threshold não se perderia nada, e seria possível capturar todos os casos verdadeiros sem nenhum caso de falso positivo. No entanto, é impossível construir um modelo perfeito, e normalmente os modelos ficam entre 50% e 100%.

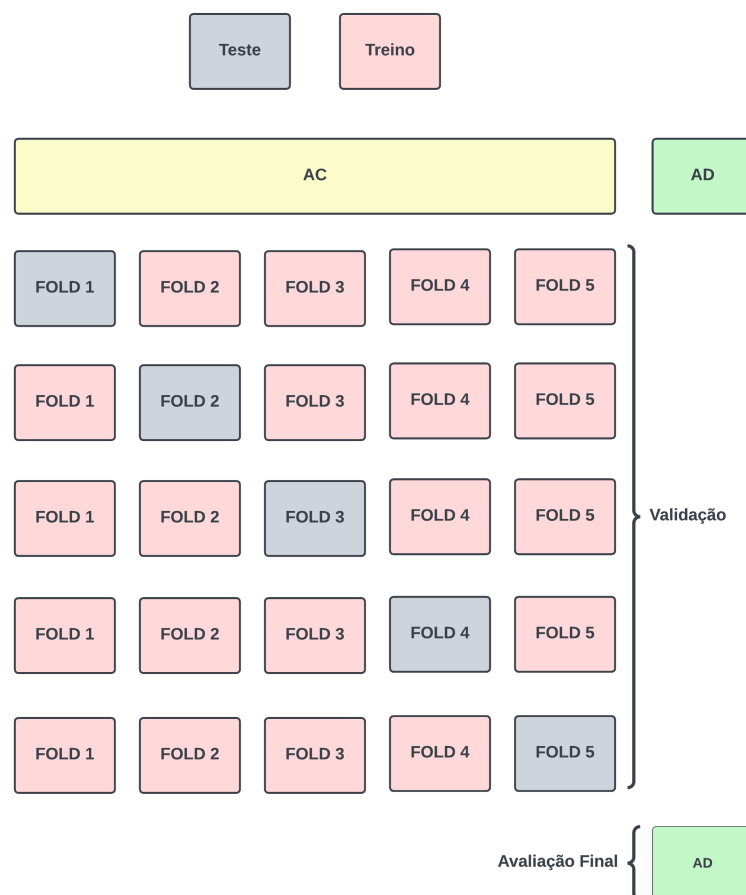
Devido ao aspecto do problema, a métrica de especificidade também irá ser importante, pois esta métrica mede a capacidade do modelo de capturar todos os casos positivos, ou seja, os verdadeiros positivos. É mais prejudicial para o paciente, com glaucoma, ser diagnosticado como falso negativo e não receber o tratamento adequado, podendo evoluir sintomas mais graves da doença, como a cegueira.

## 3.10 Treinamento e Avaliação dos Modelos

Os modelos foram treinados e avaliados utilizando uma técnica conhecida como validação cruzada (do inglês, *cross-validation*). A validação cruzada avalia a capacidade de generalização de um modelo a partir de diferentes combinações do conjunto de dados. Busca-se, então, estimar o quão preciso é este modelo quando a avaliação é feita em um conjunto de dados nunca visto antes pelo modelo. Foi utilizada a metodologia K-Fold [62], sendo este método mais utilizado para se obter modelos confiáveis quanto ao aspecto de desempenho do diagnóstico [6].

O método K-Fold consiste em separar os dados em K partes, sendo que K-1 partes serão utilizadas para o treino do modelo e a parte restante será utilizada para a validação do modelo. Em uma próxima iteração, um novo grupo é selecionado para teste e os outros K-1 para treino, até que todos os grupos passem pela fase de teste. Com isto, é possível que o modelo seja avaliado com mais robustez e com variabilidade controlada.

**Figura 3.18: Imagem ilustrando a validação cruzada. Cada parte assume o papel de teste uma vez e treino K-1 vezes.**



Fonte: Elaborado pelo autor.

A imagem 3.18 mostra como os dados são divididos. Existem dois conjuntos de dados que são AC (Amostras Conhecidas) e o AD (Amostras Desconhecidas). Utilizando o conjunto AC, aplica-se a metodologia de validação cruzada, avaliando os melhores modelos. Em seguida, o melhor modelo é avaliado no conjunto AD para mensurar seu desempenho real. Este conjunto AD não entra nas iterações da validação cruzada, portanto é possível avaliar a capacidade de generalização do modelo em dados que simulam os dados na etapa de inferência.

### 3.11 *Fine-Tuning*

Os modelos utilizados como *Backbones* são modelos que já foram ajustados anteriormente com algum conjunto de dados de imagens (não necessariamente similares aos utilizados neste projeto). Existem três formas mais utilizadas para fazer o treinamento para cada modelo. A primeira, conhecida como *Scratch*, é um treinamento onde os modelos são inicializados com parâmetros de forma randômica e, em seguida, os dados são processados para o modelo conseguir aprender padrões do zero. O segundo método é conhecido como *Feature Extraction*, onde somente a última camada do modelo é, normalmente, a camada classificadora, com seus parâmetros ajustados para o conjunto de dados que deseja. Neste caso, todo o treinamento será conduzido apenas nessa última camada adicionada, enquanto todas as outras camadas terão os seus parâmetros congelados. Existe, também, um terceiro modo, conhecido como *Fine-Tuning*, onde todos os pesos sofrem um ajuste fino para os dados da tarefa a ser realizada.

Dos métodos citados acima, as literaturas apontam que o *Fine-Tuning* consegue desempenhar melhor na grande maioria das tarefas. No treinamento em que se utiliza o método *Scratch*, é necessário um grande volume de dados para ajuste do modelo. Desta forma, quanto mais parâmetros, mais dados são necessários para o modelo desempenhar da melhor forma. No segundo método, utilizando *Feature Extraction*, o modelo não consegue se especializar da melhor forma já que se têm algumas poucas adaptações dos pesos, limitado em pequenas mudanças nesta última camada classificadora. Já no terceiro método, o *Fine-Tuning*, consegue aproveitar a complexidade do modelo treinado em um conjunto de dados mais volumoso e heterogêneo e, em seguida, se adaptar para o conjunto de dados desejado [44, 26, 4, 1, 86].

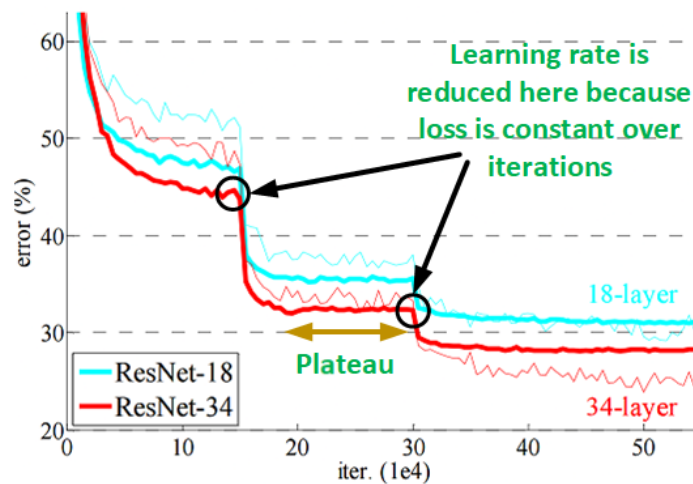
Logo, para este projeto, foi utilizado somente o *Fine-Tuning* como método de treinamento dos modelos.

## 3.12 *Learning Rate Decay*

*Learning Rate Decay* (LRD) é uma técnica de treinamento de rede neural. O modelo começa com uma taxa de aprendizagem e vai decaindo essa taxa múltiplas vezes durante o treinamento. Dessa maneira, o modelo começa atualizando os pesos mais rapidamente até que se chega em um certo ponto, em que é necessário trabalhar com magnitudes de atualizações diferentes. A taxa de aprendizagem do modelo modifica, podendo aumentar ou reduzir a magnitude dos valores de atualização dos parâmetros atingir valores melhores quanto ao desempenho.

Os seguintes trabalhos trazem um pouco do benefício do *Learning Rate Decay* [87, 35]. A imagem na figura 3.19 mostra que em um certo momento os modelos ResNet-18 e ResNet-34 chegam em um platô onde o erro do modelo não decrementa. A partir deste momento, a taxa de aprendizagem é subtraída por um valor e, após esta mudança de valor, o modelo volta a melhorar seu desempenho. Com isso, o modelo utilizando taxas menores de aprendizado consegue fazer ajustes mais finos, ou, em outras palavras, dar passos menores para um mínimo local da função de erro.

**Figura 3.19:** *Learning Rate Decay* aplicado a modelos visuais.



Fonte: [Towards Data Science](#)

## 3.13 RandAugment

*Data Augmentation* [58] é uma técnica de aumento da heterogeneidade dos dados de forma sintética. Neste caso, operações são aplicadas na imagem de forma que novas



imagens sejam criadas a partir das originais, aumentando a variabilidade dos dados para modelagem. Com os dados mais heterogêneos, é possível atingir desempenhos melhores para o modelo sem aumentar de forma exponencial o custo computacional (mais poder computacional) que será refletido em custo real.

Para realizar o *Data Augmentation* foi utilizado neste projeto o *RandAugment* [15]. Conforme os autores, o *RandAugment* traz benefícios como: diminuir drasticamente o uso computacional, pois não se utiliza políticas para escolher as operações a serem aplicadas nas imagens, mas sim uma probabilidade  $X$  de acontecer para cada uma delas. É possível manter a diversidade das imagens e ter um custo baixo para isso. As transformações possíveis são: transformação, identidade, alto-contraste, equalizador, rotação, solarização, cor tremida, mudança de contraste, mudança de brilho, posterização, mudança de formato, translações e cortes.

## 3.14 Explicabilidade: SHAP

O Shapley Additive Explanations (SHAP) é uma das ferramentas mais utilizadas quando o assunto é explicabilidade. O objetivo por trás do SHAP é trazer um entendimento quanto a decisão de modelos complexos, modelos antes conhecidos como caixa preta devido sua falta de interpretabilidade. Logo, modelos estatísticos como XGBoost, LightGBM e modelos profundos como redes neurais convolucionais (CNN) agora são possíveis de serem entendidos ou interpretados por pessoas [46].

A interpretação dos modelos trouxe uma forma inovadora de aplicar modelos de aprendizado de máquina, principalmente na área da saúde. O modelo interpretável traz informações extras para que o profissional consiga tomar melhores decisões quanto ao problema que enfrenta.

Diferentes modelos foram propostos para diagnosticar e verificar a progressão da evolução da doença de Alzheimer [22]. Os autores aplicaram o modelo SHAP para pegar as características que são mais críticas para o modelo e entender como essa doença evoluiu durante o tempo. Além disso, o modelo também forneceu explicações individuais dos pacientes. Logo após, foi utilizado um sistema baseado em lógica *fuzzy* [80] e linguagem natural para poder criar um formulário que os médicos e os pacientes utilizassem e o especialista conseguisse explicar de forma mais completa e concisa os fatores que causaram a doença e sua evolução.

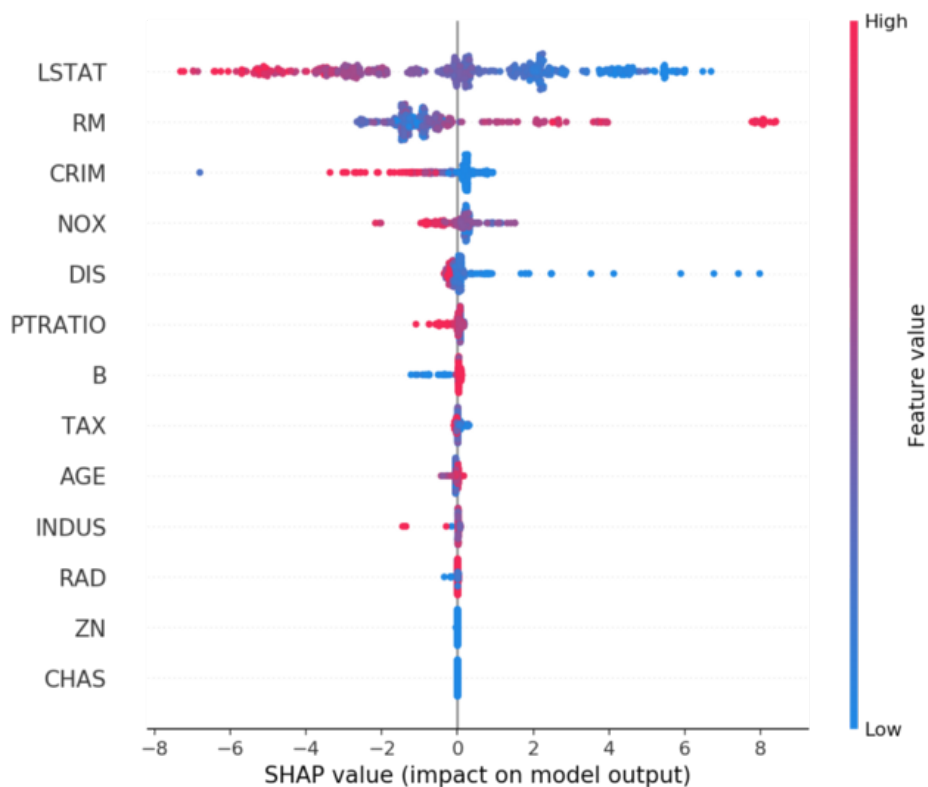
Utilizando modelos mais simples, como Árvores de Decisão [60], Regressão Logística [14] e kNN [78] e modelos mais complexos, como SVM [66], XGBoost [12], e RF [8], os autores criaram um sistema de prognóstico da doença de hepatite [57]. Utilizando o

modelo SHAP, fornecem explicação para cada um dos pacientes e também geraram gráficos de dependência entre as variáveis (características) de entrada e o rótulo dos dados (presença ou ausência de hepatite).

Foram utilizados modelos XGBoost [12] para selecionar qual cirurgia refratária deve ser feita em um paciente [85]. A escolha certa de uma cirurgia aumenta a satisfação do paciente no final do processo. Antes da escolha da cirurgia, o paciente consulta um especialista da área e, em seguida, 80 características da tomografia da córnea são extraídas, 40 características demográficas são coletadas e mais de 22 características de exames oftalmológicos, como, por exemplo, diâmetro da pupila. Ao final, os autores fazem uma validação cruzada de 10 grupos e atingem uma acurácia média de 72% quanto a cirurgia a ser aplicada. Os autores também apresentam as características gerais mais importantes para cada um dos 3 tipos de cirurgias possíveis: LASEK, LASIK, SMILE e contraindicação.

O SHAP utiliza os *Shapley Values* ou valores de *Shapley*. Os valores de *Shapley* são uma média ponderada da contribuição marginal de cada uma das características [68]. Sumarizando, ele mede o impacto que cada uma destas características tem no valor decisivo do modelo [73].

**Figura 3.20:** Gráfico resumido para o conjunto de dados Boston Housing ordenado pelas características mais importantes do modelo.



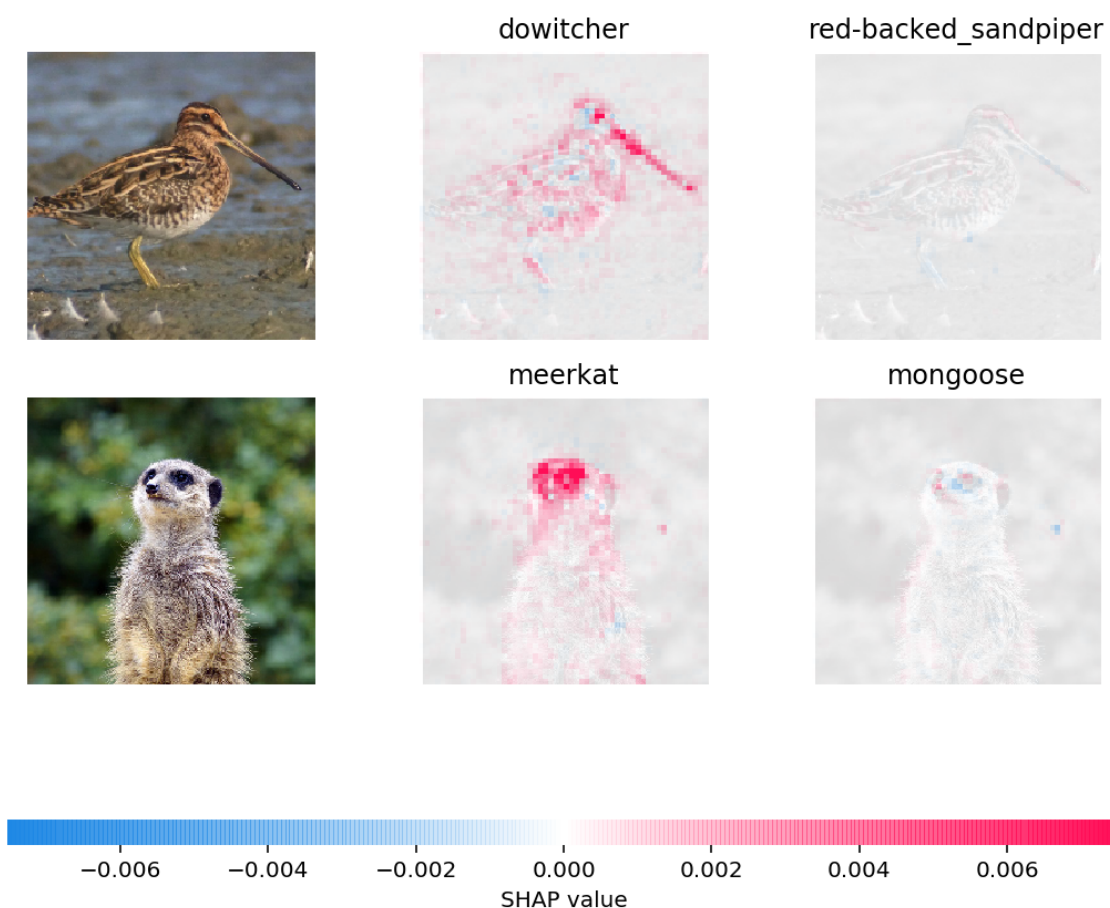
Fonte: SHAP

Github

A figura 3.20 representa um gráfico SHAP para previsão dos preços de moradias

em Boston-MA. Cada um dos pontos é uma amostra com o valor de suas características. Se o ponto é azul, essa amostra apresenta um valor baixo para esta característica, se o ponto é vermelho ela apresenta um valor alto, se o ponto está do lado direito quer dizer que aquele valor da variável contribui para a resposta ser positiva ou maior; se do lado esquerdo, negativa ou menor. Por exemplo, a variável mais importante é a LSTAT, que corresponde a proporção de adultos sem ensino superior e a proporção de trabalhadores braçais do sexo masculinos em uma área. Significa que quanto maior essa proporção, menor será o preço das casas na vizinhança. Enquanto a variável RM é a quantidade de quartos na casa. Quanto mais quartos na casa, maior será o valor atribuído para aquela residência.

**Figura 3.21: Explicação de uma imagem utilizando SHAP. Os pontos vermelhos e azuis mostram onde o modelo julgou como parte importante para a imagem. Seguindo a mesma lógica do 3.20, pontos vermelhos aumentam a probabilidade na saída do modelo, enquanto pontos azuis diminuem.**



Fonte: [SHAP Github](#)

Na figura 3.21 a explicabilidade do modelo forneceu quais são as áreas das imagens mais importantes para a identificação de um animal. Por exemplo, para o pássaro, observa-se que existem vários pontos vermelhos fortes na mesma área do bico, indicando que essa parte foi uma das mais importantes para a identificação do pássaro.

### 3.14.1 Explicabilidade para o glaucoma

A explicabilidade no glaucoma se dará de duas formas:

- O modelo de imagem irá apontar as áreas mais importantes para diagnosticar o glaucoma. Será possível verificar, em imagens, se o modelo consegue capturar a escavação do disco óptico e o disco óptico em si. Esta visualização se dará com acúmulo de pontos nestas regiões, demonstrando que o modelo se adapta a diferente posições da retina, como também diferentes recortes de imagens do fundo de olho.
- As medições mensuradas pelo exame OCT irá fornecer espessuras das camadas em diferentes regiões. O especialista irá utilizar da explicabilidade para verificar suas teorias quanto a degeneração das fibras em diferentes setores. Além disso, o especialista irá poder de dialogar com o paciente utilizando recursos visuais que facilitam o entendimento do diagnóstico da doença.

O objetivo é que o modelo explicativo dê suporte ao especialista quanto as decisões do tratamento. A explicabilidade, além de fortalecer argumentos, consegue ainda fornecer informações para que novos métodos sejam estudados e criados e que métodos antigos sejam validados ou confrontados baseados em resultados interpretáveis do modelo.

# Capítulo 4

## Metodologia

### 4.1 Setup

Todos os experimentos foram executados em uma máquina com processador i7-3770, 16 gigabytes de RAM e uma placa de vídeo Nvidia RTX-3060 com 12 gigabytes de memória.

Dentre os *softwares* utilizados estão as seguintes bibliotecas de Python<sup>1</sup>: *Pytorch*<sup>2</sup>, *Seaborn*<sup>3</sup>, *SHAP*<sup>4</sup> e *Pandas*<sup>5</sup>. Para execução das análises foi utilizada o *Jupyter Notebook*<sup>6</sup>.

### 4.2 Aquisição dos Dados

Os dados foram coletados de pacientes por especialistas do Hospital São Geraldo, especializado em atendimento médico nas áreas de Oftalmologia e Otorrinolaringologia. Os exames foram armazenados em um banco de dados contendo a data do exame e a identificação do paciente. Todos os dados, que possuem casos confirmados de glaucoma, são do tipo glaucoma primário de ângulo aberto.

Os exames OCT foram realizados por um dispositivo da Heidelberg (Spectralis SD-OCT, Software Version 5.4.7.0) e os exames de Fundoscopia foram realizados por um dispositivo da Cannon (CR2, Canon USA).

---

<sup>1</sup><https://www.python.org/>

<sup>2</sup><https://pytorch.org/>

<sup>3</sup><https://seaborn.pydata.org/>

<sup>4</sup><https://github.com/slundberg/shap>

<sup>5</sup><https://pandas.pydata.org/>

<sup>6</sup><https://jupyter.org/>

### 4.2.1 Critérios de Inclusão

- Pacientes com glaucoma de ângulo aberto
- Pacientes com idade entre 3 e 96 anos
- Pacientes com incidência de glaucoma em somente um dos olhos ou nos dois olhos

### 4.2.2 Critério de Exclusão

- Todos os pacientes que possuem glaucoma de ângulo fechado, secundário e congênito
- Pacientes que realizaram exames de retinografia utilizando dispositivos móveis
- Pacientes que não realizaram os dois exames: OCT e Fundoscopia

## 4.3 Sumário dos dados

Os dados recebidos possuem múltiplas imagens de fundo de olho. As imagens estão separadas por paciente, conforme o olho em que foi realizado o exame, direito ou esquerdo. Há dados de 442 pacientes e 887 imagens de olhos. Entre os dados, consta uma tabela com as seguintes informações: informações demográficas do paciente e do exame OCT. Na tabela recebida contém os seguintes dados:

- dados demográficos: nome, idade, gênero;
- presença ou ausência do glaucoma;
- indicação do olho examinado: esquerdo ou direito. Quando não há indicação, ambos os olhos foram examinados;
- medições de espessura da camada de fibras nervosas coletadas do exame OCT: região nasal (N), nasal inferior (NI), nasal superior (NI), temporal (T), temporal inferior (TI), temporal superior (TS) e a média de todos os quadrantes (G);

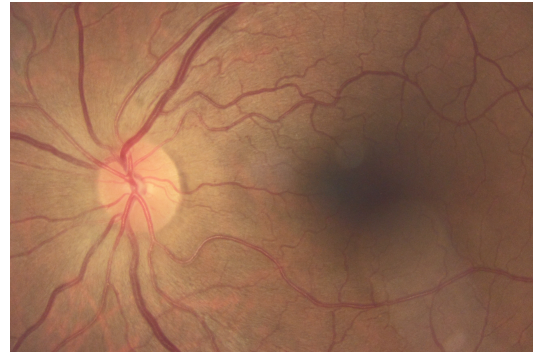
Um exemplo na figura 4.1 contendo duas imagens do fundo de olho:

**Figura 4.1: Duas imagens do fundo de olho com distâncias diferentes**

(a) Imagem do fundo de olho distante do disco e copo



(b) Imagem do fundo de olho próxima do disco e copo



Fonte: Imagem extraída do conjunto de dados.

A primeira imagem 4.1a é uma imagem de fundo de olho contendo bordas pretas e engloba toda retina. A segunda imagem é focada no centro da retina e possui o disco e escavação do disco óptico mais próximo e também a mácula e a fóvea 4.1b. Cada paciente testado neste trabalho deve possuir pelo menos duas fotos similares a essa para cada olho onde o exame OCT foi realizado.

Portanto, faz-se necessária a realização da limpeza dos dados, observando-se os requisitos dos experimentos executados neste projeto. Na próxima seção, será descrito todo o processo de limpeza dos dados realizado como requisitos dos experimentos executados neste projeto.

## 4.4 Limpeza

Conforme explicitado acima, foi realizada a padronização dos dados utilizados, para que todo o processo fique mais simples, automatizado e consistente. Assim, algumas manipulações manuais foram aplicadas aos dados, conforme se observa a seguir:

- verificação de duas vias: foi verificado se o paciente está presente nos dados tabulares do exame OCT e se existe uma pasta contendo as fotos do fundo de olho (Esquerdo e/ou Direito);
- contabilização do número de fotos: foi contabilizado o número de fotos para cada um dos pacientes, considerando o requisito de, no mínimo, duas fotos de cada olho. Todos que possuíam somente uma foto ou nenhuma foram eliminados;

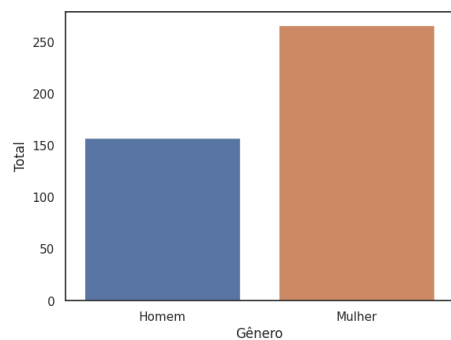
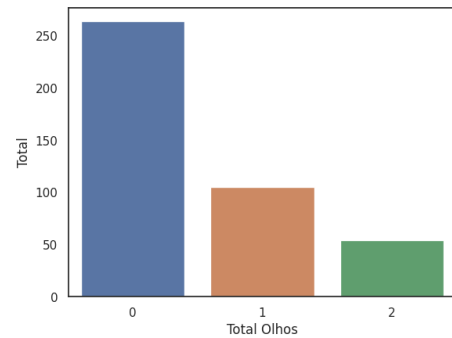
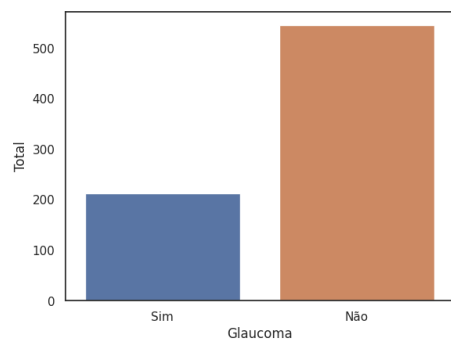
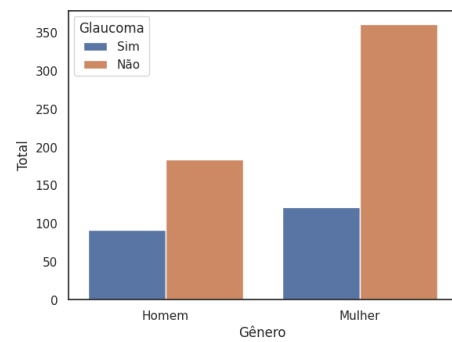
- proximidade da imagem: foram selecionadas, manualmente, duas fotos para cada olho de cada paciente. Uma foto mais distante, contendo a parte marginal preta, e uma segunda foto mais próxima, sem a parte marginal preta. As fotos foram renomeadas para indicar qual das duas imagens é a mais próxima ou distante, de forma que a arquitetura do modelo, que utiliza duas imagens, sempre receba o mesmo tipo de foto;
- cor das imagens: foram selecionadas somente fotos coloridas. Fotos preto e branco foram eliminadas, já que a maioria dos pacientes não possuía este tipo de foto;
- data dos exames: alguns pacientes possuem diferentes datas dos exames OCT e Fundoscopia. Foram consideradas as fotos de fundoscopia tiradas em datas próximas à realização do exame OCT.

No total de 442 pacientes e 885 olhos, após a limpeza dos dados, restaram 426 pacientes e 758 amostras de olhos utilizados na modelagem.

## 4.5 Exploração

Para um melhor entendimento do problema, faz-se necessário que os dados sejam explorados e que a sua distribuição seja compreendida. Nesta seção, os dados serão avaliados, observando-se suas características, como, por exemplo, a correlação entre as espessuras da camada de fibras nervosas e a presença do glaucoma, a idade e o gênero dos pacientes, dentre outras características importantes que deram base para a obtenção dos resultados apresentados na seção 5.3.



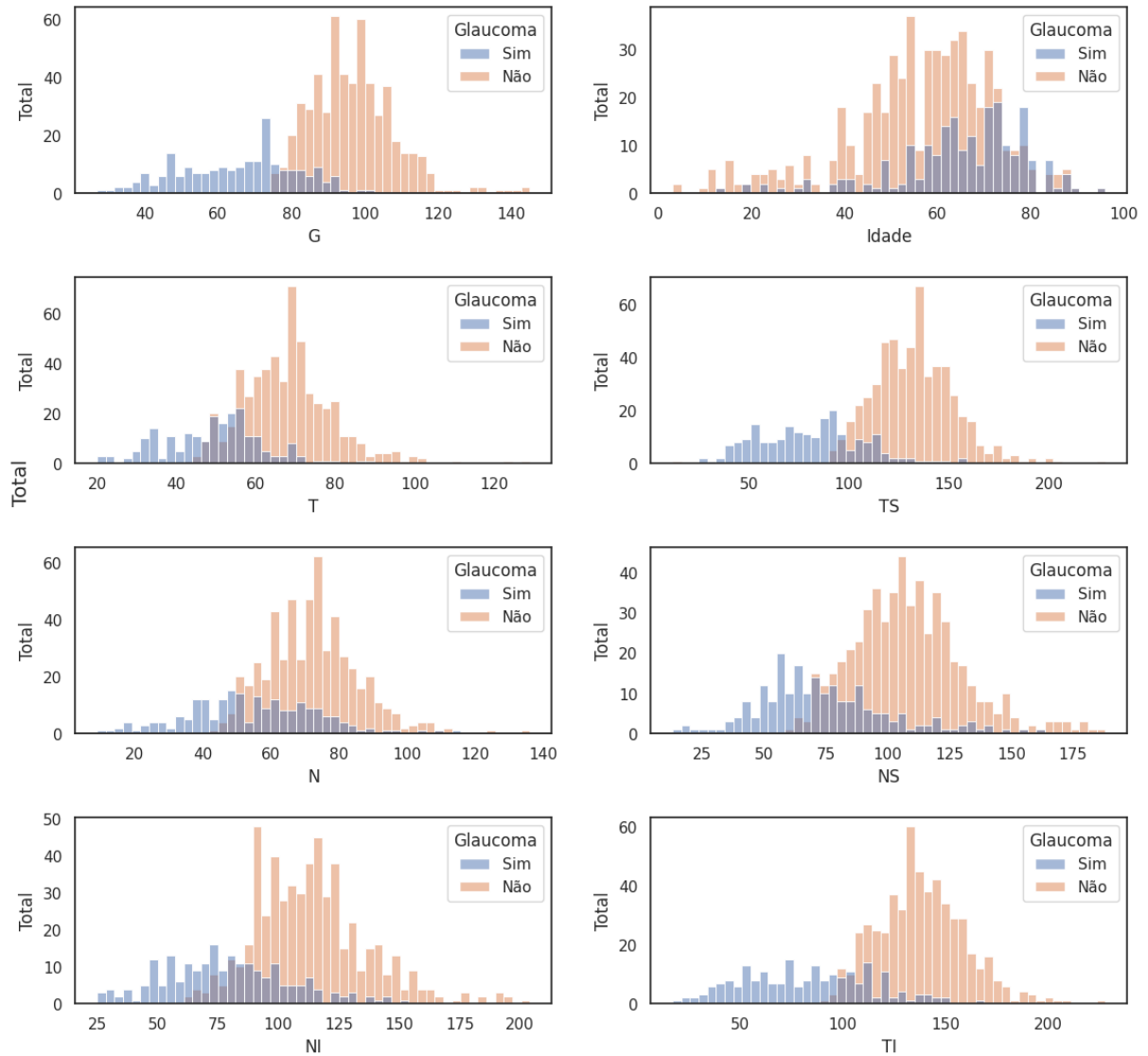
**Figura 4.2: Contagem de casos de glaucoma por gênero e por olhos****(a)** Total de pacientes por gênero**(b)** Total de pacientes com presença de glaucoma em 0(ausência), 1, 2**(c)** Total de olhos com e sem glaucoma**(d)** Total de olhos com ou sem glaucoma por gênero

Fonte: Elaborado pelo autor.

Os histogramas da figura 4.2 demonstram que os dados possuem mais casos de pessoas contendo glaucoma em somente um dos olhos do que nos dois olhos. Há mais pacientes do gênero feminino do que do masculino. No entanto, proporcionalmente, a presença de glaucoma é maior em homens do que em mulheres. Possivelmente, tal fato se deve a fatores culturais observados em pessoas do sexo feminino que, conforme a Pesquisa Nacional de Saúde [59], são as mulheres que têm mais hábito de ir ao médico manter-se saudável. Assim, apesar de as mulheres serem a maioria dos pacientes que realizaram os exames, são as que menos registram a presença do glaucoma. Na seção de resultados, será avaliado se o modelo coloca a variável gênero como uma das principais características para a predição do glaucoma.

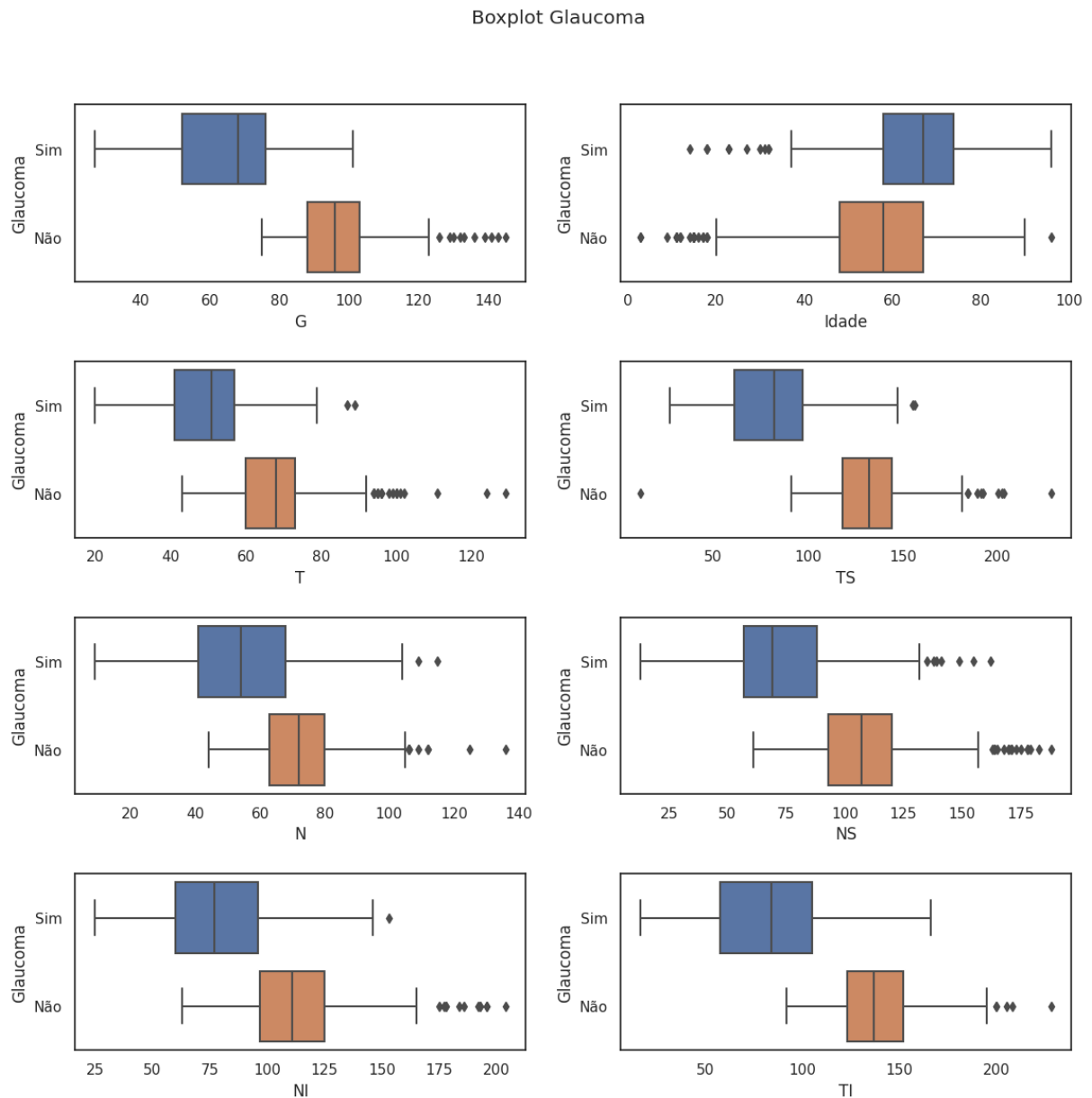
**Figura 4.3: Distribuição do glaucoma por variável**

Distribuição das dos dados numéricos baseado nas classes



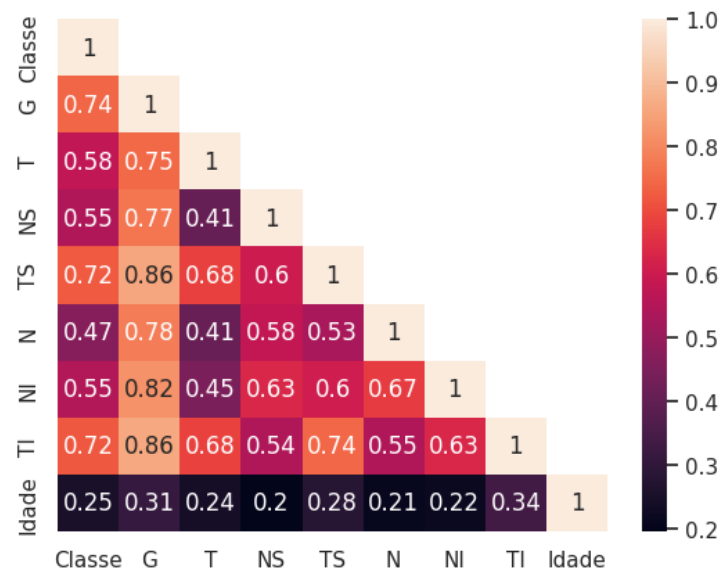
Fonte: Elaborado pelo autor.

Os gráficos da figura 4.3 mostram a distribuição das variáveis do exame OCT quanto a presença de glaucoma. Existe uma separação visível das distribuições das características G, TS, N e TI quando há ou não glaucoma.

**Figura 4.4: Boxplot do glaucoma por variável.**

Fonte: Elaborado pelo autor.

Observou-se a presença de valores atípicos mostrados na figura 4.4. As características idade, NI, G e NS são as que mais possuem valores fora do comportamento padrão da distribuição. A idade é, como se vê na figura, um bom divisor, já que existe uma tendência a esquerda (menor idade) quanto a ausência do glaucoma. As variáveis TI, G e TS são as que mais distanciam os casos quanto a presença ou não do glaucoma.

**Figura 4.5: Correlação absoluta das características entre 0 e 1, sendo 1 correlação perfeita.**

Fonte: Elaborado pelo autor.

Existe uma forte correlação entre a variável TS e G 4.5. Sendo G uma média global, existe a possibilidade de que a variável TS tenha uma influência forte na média. As variáveis que mais se correlacionam com a classe são: TS, TI e G. A sumarização dos dados após a limpeza está descrita abaixo:

**Tabela 4.1: Tabela contendo sumarização dos dados utilizando as estatísticas: máximo, mínimo, média, desvio e contagem.**

	Classe	G	T	NS	TS	N	NI	TI	Idade
Contagem	758.00	758.00	758.00	758.00	758.00	758.00	758.00	758.00	758.00
Média	0.28	87.73	62.65	98.15	117.77	67.15	103.76	122.49	58.57
Desvio Padrão	0.45	18.82	14.26	28.11	31.97	16.94	28.63	34.75	15.90
Mínimo	0.00	27.00	20.00	13.00	12.00	9.00	25.00	17.00	3.00
25%	0.00	79.00	54.00	80.00	101.00	57.00	87.00	106.00	50.00
50%	0.00	91.00	63.00	100.00	122.00	68.00	104.00	129.00	61.00
75%	1.00	100.00	71.00	117.00	139.00	78.00	121.00	145.75	70.00
Máximo	1.00	145.00	129.00	188.00	228.00	136.00	204.00	228.00	96.00

Fonte: Elaborado pelo autor.

## 4.6 Engenharia de características e Normalização

Quanto a engenharia de características, apenas uma operação matemática foi aplicada. Sabendo que uma das características importantes para a detecção do glaucoma é o valor da CDR (Equação 3.1) presente na imagem de fundo de olho, as características criadas são a razão entre as medições numéricas do exame do OCT [91]. Cada caracte-

ística numérica do exame OCT irá criar uma característica de razão entre pares com as outras. No total, 15 novas características foram criadas para serem utilizadas nos modelos. Além de serem características facilmente criadas, são fáceis de serem interpretadas pelos profissionais responsáveis ao analisar os resultados e a explicabilidade do modelo.

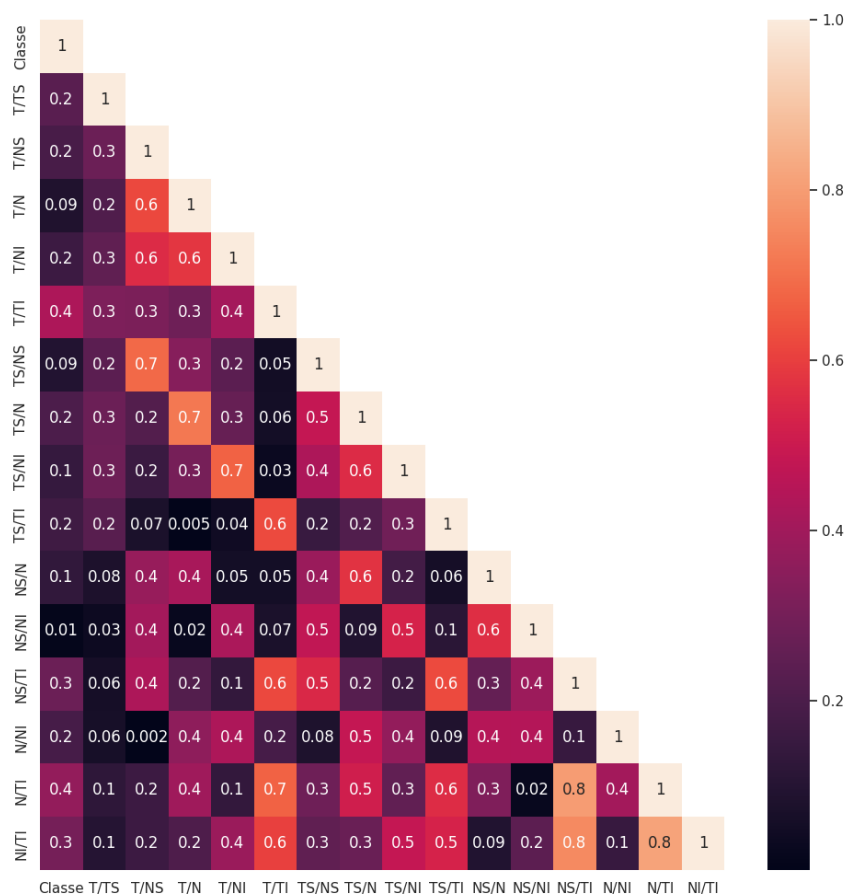
A normalização é um processo de redimensionar os valores das características numéricas para limites entre 0 e 1. A importância da normalização nas redes neurais está diretamente ligada ao fator de atualização dos parâmetros e otimização do modelo. Os algoritmos de otimização utilizam a descida gradiente<sup>7</sup> para encontrar os melhores valores para diminuir o erro. Na fórmula de descida gradiente, os valores de cada uma das características são utilizados para atualizar os parâmetros e diminuir o erro do modelo. Contudo, se os valores das características não forem normalizados e os dados possuírem características com valores de magnitude diferentes, as características com magnitudes maiores terão mais importância, o que não pode se assumir como verdade conforme os dados. Portanto, para evitar que tal fato ocorra e prejudique o desempenho do modelo, o ideal é que todas as características tenham a mesma magnitude e o modelo escolha a importância de cada uma delas durante o processo de treinamento.

Uma comparação de aplicação da técnica de normalização foi realizada entre os modelos de redes neurais, modelos de árvore de decisão como Tree Decision, Regressão Logística e modelos matemáticos como SVM [47]. Para os modelos que não são redes neurais, 22 características foram extraídas do exame OCT. Uma destas características é a razão entre o copo e o disco óptico. Todas as características foram normalizadas antes de serem processadas pelos modelos. Os resultados foram benéficos para os modelos de redes neurais profundas e invariante para os outros modelos. Uma **observação** é que todas as estatísticas para aplicar a normalização devem ser calculadas utilizando o conjunto de dados de treino. Por exemplo, na normalização entre 0 e 1, os valores mínimo e máximos são calculados usando os dados de treino e aplicados nos dados de teste, evitando-se que haja vazamento de dados entre os dois conjuntos.

Todas as características numéricas, como a idade, as mensurações do exame OCT e as características sintéticas derivadas do exame OCT mencionadas foram anteriormente normalizadas entre 0 e 1 para serem processadas pelo modelo. Quanto ao gênero do paciente, foi utilizado 1 para o gênero masculino e 0 para o gênero feminino. <https://www.facebook.com/messages/t/100002734850472/>

---

<sup>7</sup>O método do gradiente (ou método do máximo declive) é um método numérico usado em otimização. Para encontrar um mínimo (local) de uma função usa-se um esquema iterativo, onde em cada passo se toma a direção (negativa) do gradiente, que corresponde à direção de declive máximo. Pode ser encarado como o método seguido por um curso da água, na sua descida pela força da gravidade. Fonte: Wikipédia

**Figura 4.6: Correlação das características criadas pela operação razão.**

Fonte: Elaborado pelo autor.

A figura 4.6 mostra a correlação das características criadas sinteticamente com a operação de razão. Duas características que mais apresentam correlação com a classe (presença ou ausência de glaucoma) são: N/TI e T/TI. No capítulo de resultados 5, mais especificamente na seção de explicabilidade 5.3, será averiguado se alguma destas características criadas foram importantes para o modelo.

## 4.7 Experimentação

Os experimentos foram faseados, da seguinte forma:

- Primeiro, utilizando diferentes configurações de hiperparâmetros, foram testados os seguintes modelos: Mobile, RegNet (X e Y), ShuffleNet, EfficientNet, Inception-V3, ResNet, e ViT. Cada um dos modelos citados foram testados com 4 diferente otimizadores (Adam, RAdam, SGD, Ranger) de parâmetros e 4 valores de taxa de

aprendizagem (0,01, 0,001, 0,0001, 0,0005). Ao final desta etapa, foram selecionados os melhores modelos conforme a métrica AUC para a próxima etapa. Esta etapa também forneceu o melhor momento para aplicar o *Learning Rate Decay* e o *Early Stopping* para ser utilizado nas próximas etapas.

- Em seguida, os melhores K modelos selecionados foram re-executados com validação cruzada e divididos em dois grupos. Um grupo utilizando as técnicas *Learning Rate Decay* e *Early Stopping* e outro grupo sem incluir estas técnicas.
- E, por último, os melhores M modelos dos K foram selecionados e aplicado o *RandAugment* (Seção 3.13) para trazer mais diversidade para os dados de treino e verificar o impacto, em questão de desempenho e variabilidade, do *Data Augmentation*.

A partir do segundo passo, todos os testes foram feitos utilizando a metodologia de validação cruzada com o parâmetro de divisão de grupos igual a 5. A decisão de aplicar somente a partir do segundo passo é o número exponencial de modelos a serem treinados caso fosse aplicado desde o primeiro passo. Em todas as etapas, o número de épocas padrão é 100 e a função de custo é uma entropia cruzada binária[63].

Uma **observação** importante é que a separação dos dados de treino e teste aconteceu na granularidade de paciente e não de olho, para que, assim, um paciente não esteja presente no conjunto de treino e teste em simultâneo, evitando vazamento de dados e o modelo demonstre um desempenho irreal.

**Tabela 4.2: Versões dos modelos visuais utilizados como Backbones e o total de parâmetros de cada um. O modelo ViT é o que possui maior número de parâmetros, aproximadamente 86 milhões, enquanto o modelo ShuffleNet possui somente 2.5 milhões.**

Versão	Backbone	Parâmetros
RegNet800MFX	RegNetX	6.586.329
RegNet800MFY	RegNetY	5.648.297
RegNet1.6GFX	RegNet16X	8.279.048
RegNet1.6GFY	RegNet16Y	10.314.319
RegNet3.2GFX	RegNet32X	14.288.561
RegNet3.2GFY	RegNet32Y	17.824.851
MobileV3Large	MobileNet	4.203.313
ResNet50	ResNet	23.510.081
EfficientNetb0	EfficientNet	4.008.829
ShuffleNetV1X1.5	ShuffleNet	2.479.649
VitBase16_224	ViT	85.799.425
InceptionV3	Inception	24.346.082

Elaborado pelo autor.

A tabela 4.2 possui a versão e o total de parâmetros dos modelos citados na seção 3.6 implementados na biblioteca Pytorch<sup>8</sup>.

<sup>8</sup>Biblioteca Python de aprendizado de máquina

# Capítulo 5

## Resultados

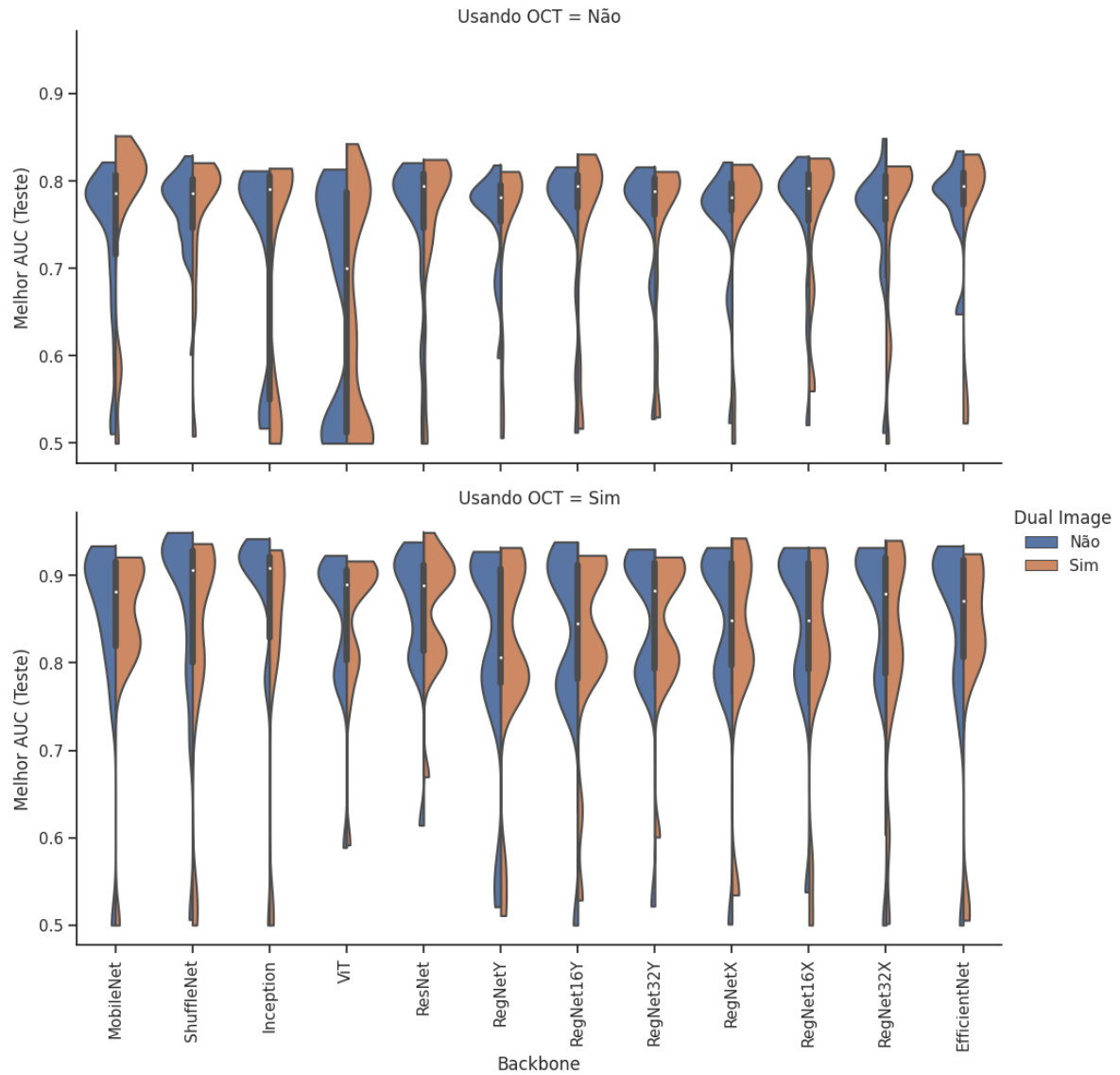
Neste capítulo os resultados serão apresentados. A primeira seção abordará uma visão geral dos modelos quanto ao desempenho relacionado a métrica principal, AUC (Seção 3.9). Além disso, pelo gráfico da função de custo (*loss*), um ponto ideal foi definido para ser aplicado as técnicas de *Learning Rate Decay* (Seção 3.12) e *Early Stopping*. Na seção seguinte 5.2 será avaliado o desempenho do modelo utilizando validação cruzada e as técnicas mencionadas anteriormente. Nesta mesma seção será também avaliado o RandArgument. E por fim, na última seção 5.3 será avaliado a explicabilidade do modelo. Arquiteturas de modelos utilizando duas imagens de fundo de olho são apontados como *Dual Image*.

### 5.1 Desempenho Geral

A figura 5.1 mostra como foi o desempenho geral dos *Backbones*. Os *Backbones* utilizando dados do exame OCT atingiram melhores resultados nas arquiteturas de uma e duas imagens. A distribuição de desempenho dos modelos que não utilizam o exame OCT é menos esparsa, com exceção do ViT e do Inception.



**Figura 5.1:** Avaliação geral dos *Backbones* divididos por arquiteturas. O gráfico violino mostra a distribuição dos valores máximos atingidos nas 4 arquiteturas de modelos testados.



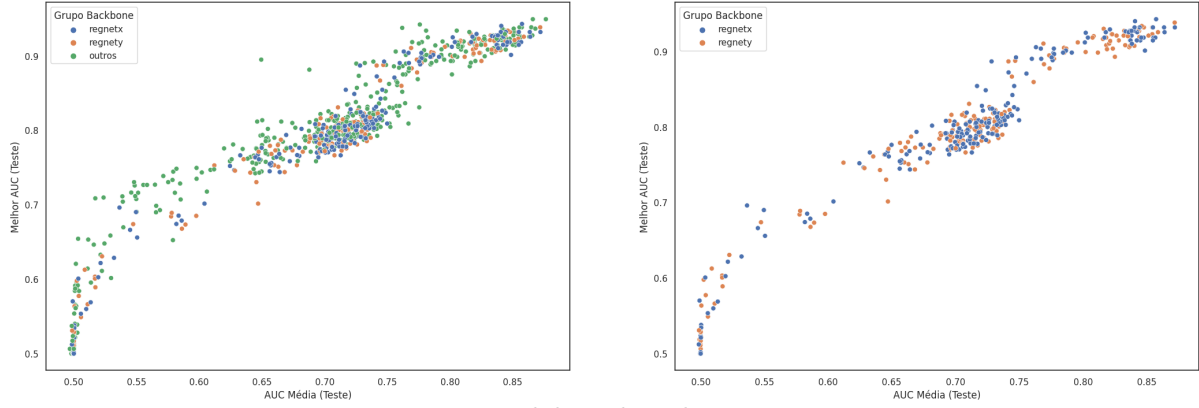
Fonte: Elaborado pelo autor.

O gráfico 5.2 faz uma relação entre os valores máximo e médio que o *Backbone* atinge quanto a métrica AUC para o conjunto de teste. Quando se compara os grupos, percebe-se que os modelos estão distribuídos uniformemente no gráfico. Existe uma relação direta entre a média de AUC (Teste) e o valor máximo que essa métrica atinge durante todas as iterações, com raros casos de discrepância. Portanto, é possível perceber que as arquiteturas de modelos não variam seu desempenho drasticamente durante o processo de treinamento e avaliação.

**Figura 5.2: Comparação das estatísticas da métrica AUC. O eixo y é o melhor resultado do *Backbone* durante todas as épocas, enquanto o eixo x é a média do modelo durante o treinamento.**

(a) Comparação entre grupo de modelos RegNet e todos os outros.

(b) Comparação entre os modelos RegNet por si.



Fonte: Elaborado pelo autor.

As tabelas 5.1 e 5.2 resumizam todas as execuções dos *Backbone* em todas as arquiteturas discutidas na seção 3.16. O primeiro cabeçalho da tabela contém as colunas Melhor AUC e AUC Média, indica o valor máximo atingido pela métrica AUC e valor médio durante todo o processo de treinamento avaliado no conjunto de teste, respectivamente; No segundo cabeçalho estão as estatísticas calculadas das duas colunas citadas anteriormente, **std**, **min**, **max**, **mean**, significando, respectivamente: desvio padrão, valor mínimo, valor máximo, a média do valor. Ou seja, na coluna Melhor AUC/min igual 0.5 significa que houve arquitetura de modelos que não conseguiram aprender utilizando certas combinações de hiperparâmetros para aquele *Backbone* específico e AUC Média/mean significa que na média as arquiteturas de modelos para aquele *Backbone* específico atingiu 0.71 na métrica AUC durante o processo de treinamento avaliado no conjunto de dados de validação/teste. Os valores máximos atingidos pelos modelos ficaram entre 0.92 e 0.95. A AUC Média mostra como o processo de generalização do modelo é um processo que exige experimentação, para que assim o modelo consiga alcançar resultados que ultrapassem o desempenho médio. Quanto as arquiteturas testadas, a presença do exame OCT é benéfico para o modelo, aumentando até em 12% o desempenho dos modelos 5.2. No entanto, não houve melhora dos modelos quando se utiliza duas imagens na arquitetura. Dito isso, todas as arquiteturas avaliadas daqui para frente utilizará somente uma imagem do fundo de olho como entrada.

**Tabela 5.1:** Tabela contendo os resultados quanto à métrica AUC separado por Backbone.

Backbone	Melhor AUC				AUC Média			
	min	max	mean	std	min	max	mean	std
EfficientNet	0.5	0.93	0.8	0.11	0.5	0.85	0.72	0.1
Inception	0.5	0.94	0.79	0.14	0.5	0.86	0.7	0.12
MobileNet	0.5	0.93	0.79	0.11	0.5	0.84	0.69	0.1
RegNet16X	0.5	0.93	0.79	0.1	0.5	0.86	0.71	0.09
RegNet16Y	0.5	0.94	0.79	0.11	0.5	0.87	0.71	0.09
RegNet32X	0.5	0.94	0.79	0.11	0.5	0.87	0.71	0.1
RegNet32Y	0.52	0.93	0.8	0.1	0.5	0.86	0.71	0.09
RegNetX	0.5	0.94	0.79	0.11	0.5	0.86	0.71	0.09
RegNetY	0.51	0.93	0.78	0.11	0.5	0.85	0.7	0.09
ResNet	0.5	0.95	0.79	0.1	0.5	0.87	0.7	0.1
ShuffleNet	0.5	0.95	0.8	0.09	0.5	0.88	0.7	0.1
ViT	0.5	0.92	0.75	0.15	0.5	0.85	0.67	0.12

Fonte: Elaborado pelo autor.

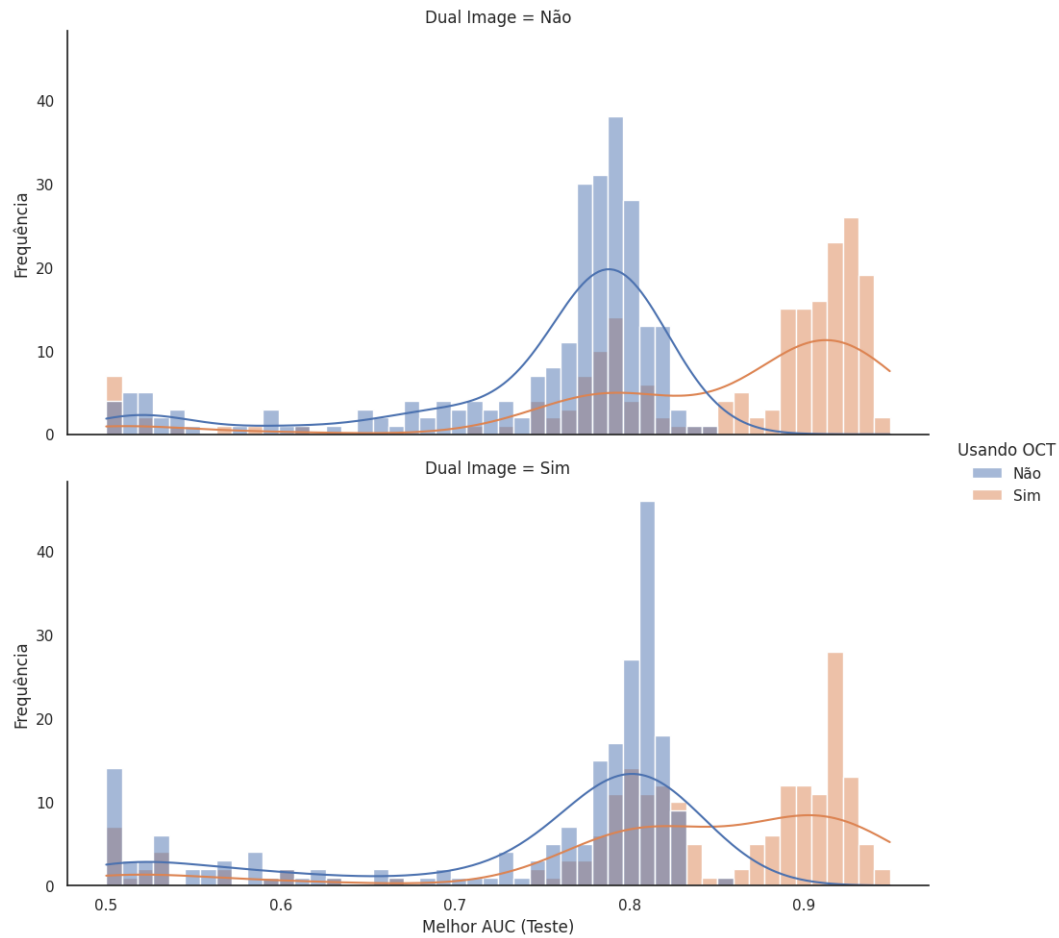
**Tabela 5.2:** Tabela contendo os resultados separados por grupos: OCT e Dual Image. A tabela mostra que utilizar duas imagens não foi benéfico, enquanto o uso do exame OCT agrega desempenho aos resultados

Dual Image	Usando OCT	Melhor AUC				AUC Média			
		min	max	mean	std	min	max	mean	std
Não	Não	0.5	0.85	0.75	0.08	0.5	0.74	0.66	0.07
Não	Sim	0.5	0.95	0.85	0.1	0.5	0.88	0.75	0.1
Sim	Não	0.5	0.85	0.74	0.11	0.5	0.77	0.66	0.09
Sim	Sim	0.5	0.95	0.83	0.11	0.5	0.87	0.74	0.1

Fonte: Elaborado pelo autor.

A distribuição dos desempenhos apresentados na figura 5.3 mostra uma separação clara entre a utilização e não utilização do exame OCT. No entanto, como citado anteriormente, não há diferença do modelo utilizando uma ou duas imagens. Foram selecionados os melhores 5% modelos baseados na AUC das arquiteturas que utilizam uma imagem com a inclusão e não inclusão do exame OCT. Para a seleção, foi utilizado o percentil<sup>1</sup> 95 e todos os modelos abaixo das 95 partes dos dados foram filtrados. Ao final, restaram um total de 28 arquiteturas de modelos e configurações de hiperparâmetros dos modelos.

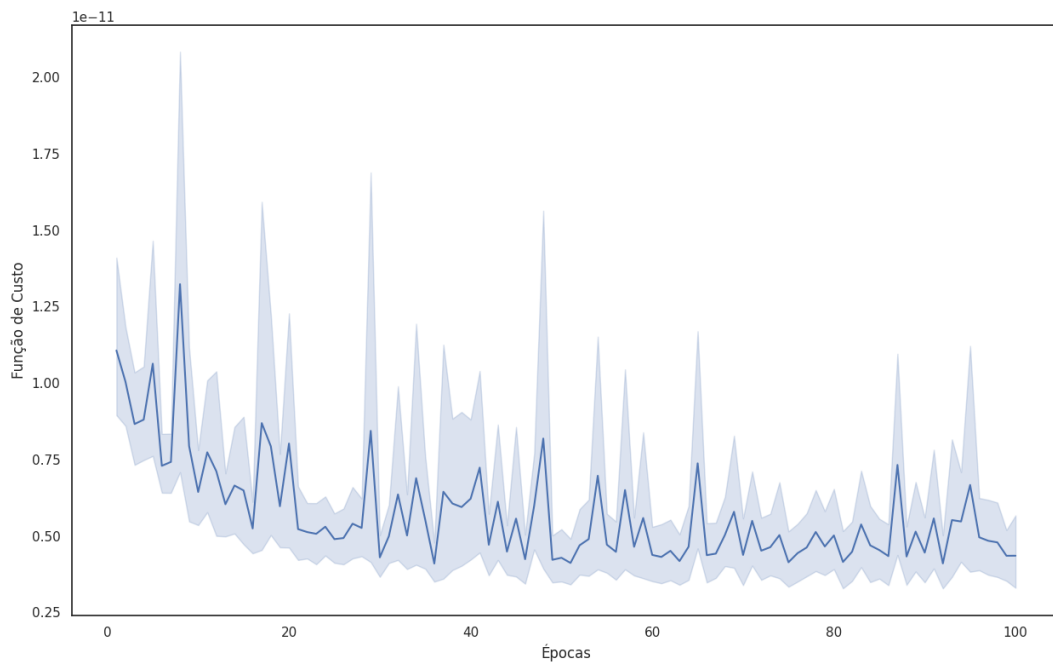
<sup>1</sup>Em estatística descritiva, os percentis são medidas que dividem a amostra (por ordem crescente dos dados) em 100 partes, cada uma com uma percentagem de dados aproximadamente igual. O  $k$ -ésimo percentil  $P_k$  é o valor  $x$  ( $x_k$ ) que corresponde à frequência cumulativa de  $N.k/100$ , onde  $N$  é o tamanho amostral.

**Figura 5.3: Distribuição dos desempenhos separados por arquiteturas de modelos.**

Fonte: Elaborado pelo autor.

Por último, tem-se o gráfico 5.4 da função de custo durante o treinamento. O gráfico apresenta a média global de todos os modelos treinados com intervalo de confiança de 0.95. O gráfico mostra, ainda, que a partir de 30 iterações (épocas), os modelos não conseguem ter uma melhora significativa no seu desempenho. Portanto, as técnicas *Early Stopping* e o *Learning Rate Decay* foram aplicadas, a partir deste momento, na próxima fase de experimentação, a validação cruzada.

**Figura 5.4: Comportamento da função de custo durante as épocas de treinamento. O gráfico mostra um valor médio e um intervalo de confiança de 0.95**



Fonte:

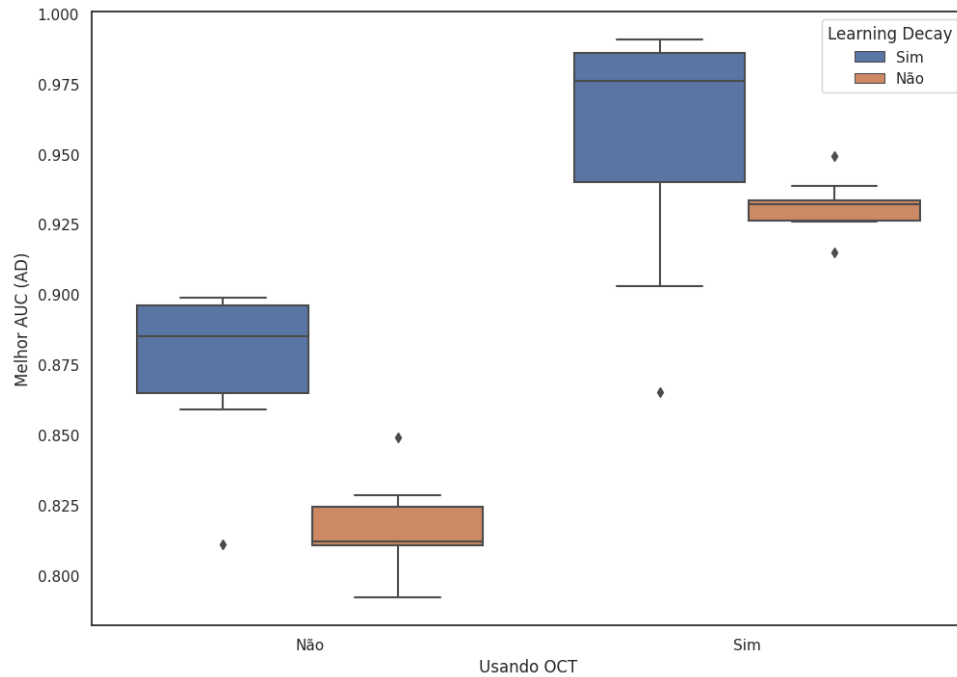
Elaborado pelo autor.

## 5.2 Validação Cruzada

Nesta seção, os melhores modelos selecionados na etapa anterior serão avaliados para a verificação do seu desempenho utilizando a validação cruzada [62]. O valor de grupos para a validação cruzada, fixo para todo o experimento, é igual a 5, ou seja, cada configuração de hiperparâmetros e arquiteturas geraram 5 modelos diferentes, totalizando 140 modelos.

A figura 5.5 mostra a distribuição do desempenho dos modelos testados. O *Learning Rate Decay* e o *Early Stopping*, indicado nas colunas somente como LRD, trouxeram uma melhora significativa para os resultados. Apesar de terem aumentado a variabilidade da AUC, as técnicas trouxeram benefícios positivos quanto ao máximo atingido pela métrica. O salto maior foi para a arquitetura de modelos utilizando somente imagem.

**Figura 5.5: Comparação do uso do *Learning Rate Decay* e *Early Stopping* nos modelos selecionados na primeira fase. Os melhores modelos da etapa anterior foram re-executados, sem a utilização das técnicas citadas anteriormente, utilizando validação cruzada para uma comparação justa.**



Fonte: Elaborado pelo autor

A tabela a seguir [5.3](#) sumariza as melhorias nas métricas quanto a aplicação da técnica LRDs e ES. O modelo RegNet16X teve o maior ganho na métrica AUC, no entanto, possui uma variabilidade alta nos resultados. Esta variabilidade é considerável quando se analisa as médias de todas as métricas, por outro lado, os valores máximos de desempenho atingidos pelos modelos trouxeram menos variabilidade. Para a visualização completa dos resultados, uma tabela que contém os hiperparâmetros utilizados e os resultados detalhados será anexada ao apêndice dessa dissertação [A](#).

**Tabela 5.3:** Tabela com os resultados comparando a utilização da técnica LRD e ES na validação cruzada. Foram calculados a média e o desvio padrão entre os *Backbones* dos melhores modelos. Os *Backbones* com desvio padrão igual a 0 significa que só houve uma arquitetura de modelo entre os modelos selecionados na fase anterior.

Backbone	AUC (%)		SN (%)		SP (%)	
	Melhor	Média	Melhor	Média	Melhor	Média
Efficient-Net	5.57 ± (0.0)	-2.10 ± (0.0)	9.65 ± (0.0)	0.77 ± (0.0)	2.86 ± (0.0)	3.56 ± (0.0)
Inception	5.50 ± (4.44)	5.54 ± (4.06)	17.33 ± (8.1)	15.8 ± (8.74)	2.41 ± (2.1)	1.45 ± (4.19)
MobileNet	0.26 ± (0.0)	7.79 ± (0.0)	8.24 ± (0.0)	34.54 ± (0.0)	1.83 ± (0.0)	-0.99 ± (0.0)
Reg-Net16X	9.01 ± (3.92)	9.32 ± (2.13)	14.58 ± (7.09)	12.54 ± (12.77)	8.88 ± (9.09)	9.59 ± (10.42)
Reg-Net16Y	3.41 ± (3.01)	4.34 ± (5.8)	4.66 ± (4.16)	19.1 ± (9.8)	3.43 ± (1.48)	3.32 ± (2.36)
Reg-Net32X	2.33 ± (5.26)	6.06 ± (7.63)	9.51 ± (8.63)	17.62 ± (8.33)	0.53 ± (2.63)	-0.53 ± (3.82)
Reg-Net32Y	1.38 ± (0.23)	4.28 ± (4.73)	5.62 ± (5.87)	9.47 ± (16.26)	0.48 ± (1.92)	1.92 ± (2.57)
RegNetX	6.61 ± (1.03)	7.03 ± (3.3)	14.18 ± (3.27)	32.93 ± (13.03)	0.95 ± (1.83)	2.21 ± (0.7)
ResNet	1.56 ± (1.51)	-10.57 ± (18.94)	2.73 ± (4.06)	-28.48 ± (30.21)	2.31 ± (2.02)	5.26 ± (4.48)
ShuffleNet	1.36 ± (4.13)	2.14 ± (7.55)	5.48 ± (4.56)	8.53 ± (20.61)	0.88 ± (1.4)	1.23 ± (1.89)

Fonte: Elaborado pelo autor.

Quanto ao uso de memória e tempo de inferência, a tabela 5.4 mostra que os modelos RegNets possuem tempo de inferência próximo a dos modelos mais eficientes, como ShuffleNet e MobileNet, e menor do que os modelos mais complexos, como, por exemplo, o Inception. Isto proporciona flexibilidade para o modelo RegNet possa ser utilizado em diferentes contextos em que uma resposta rápida é necessária, tal como o processamento de quadros de um vídeo ou algum sistema nuvem que recebe milhares de requisições simultaneamente. Quanto ao uso de memória, os modelos RegNets apresentam versões com diferentes usos deste recurso. Variando de 1600 até 4182 megabytes, o modelo consegue apresentar desempenho melhor ou próximo quando comparado aos modelos que utilizam mais memória (ou seja, usam mais parâmetros e fazem mais operações matemáticas), e também aqueles que são mais eficientes computacionalmente. Conclui-se, que é possível atingir desempenho alto com modelos que utilizam recursos computacionais limitados, dando flexibilidade para sua utilização, tanto para supercomputadores ou para dispositivos móveis, e desta forma, se responde à questão **Q1** (Seção 1.3).

**Tabela 5.4:** Tabela contendo informações relacionadas ao tempo de inferência e o uso de memória. A memória é medida em megabytes (mb) e o tempo de inferência em milissegundos (ms).

<i>Backbone</i>	Memória (mb)	Inferência (ms)	<i>Batch</i>
EfficientNet	3002	0.033 ± (0.002)	16
Inception	4747	0.041 ± (0.003)	16
MobileNet	1662	0.019 ± (0.000)	16
RegNet16X	2405	0.022 ± (0.001)	16
RegNet16Y	2812	0.024 ± (0.002)	16
RegNet32X	3600	0.024 ± (0.002)	16
RegNet32Y	4182	0.027 ± (0.003)	16
RegNetX	1679	0.020 ± (0.001)	16
ResNet	3762	0.044 ± (0.002)	16
ShuffleNet	1057	0.018 ± (0.001)	16

Fonte: Elaborado pelo autor.

```

1     model = init_pretrained_model(model_name)
2     device = torch.device("cuda")
3     model.to(device)
4     dummy_input = torch.randn(1, 3,224,224, dtype=torch.float).to(
device)
5     starter, ender = torch.cuda.Event(enable_timing=True), torch.
cuda.Event(enable_timing=True)
6     repetitions = 300
7     timings=np.zeros((repetitions,1))
8     for _ in range(10):
9         _ = model(dummy_input)
10    with torch.no_grad():
11        for rep in range(repetitions):
12            starter.record()
13            _ = model(dummy_input)
14            ender.record()
15            torch.cuda.synchronize()
16            curr_time = starter.elapsed_time(ender)
17            timings[rep] = curr_time
18    mean_syn = np.sum(timings) / repetitions
19    std_syn = np.std(timings)
20

```

**Algoritmo 5.1:** Algoritmo utilizado para calcular o tempo de inferência dos modelos. Código retirado de DECI



### 5.2.1 *RandAugment*

Tendo em vista os resultados da seção anterior, 3 modelos foram selecionados para aplicar *data augmentation*. Um dos hiperparâmetros do *RandAugment* (coluna ROP) é o número de operações de transformações sequenciais aplicado em uma foto. Por esse motivo, foram testados diferentes valores para este hiperparâmetro: 1, 2, 3 e 4. Nesta etapa, a validação cruzada também foi utilizada para avaliar o modelo. A tabela 5.5 a seguir mostra os resultados dos experimentos e os respectivos hiperparâmetros associados a este experimento.

**Tabela 5.5: Experimentos utilizando diferentes parâmetros de *RandAugment* (coluna OP). O número na coluna representa o número de operações seguidas aplicada na imagem.**

<i>Backbone</i>	<i>Op-tim</i>	LR	ROP	OCT	Melhor AUC	AUC Média	Melhor SP	Melhor SN	SP Médio	Mé-dio	SN Mé-dio
Inception	ranger	0.001	0	Não	0.899	0.795 ( $\pm 0.066$ )	0.935	0.882	0.923		0.666
Inception	ranger	0.001	1	Não	0.89	0.802 ( $\pm 0.057$ )	0.971	0.882	0.914		0.69
Inception	ranger	0.001	2	Não	0.847	0.787 ( $\pm 0.045$ )	0.971	0.765	0.884		0.69
Inception	ranger	0.001	3	Não	0.871	0.787 ( $\pm 0.055$ )	0.943	0.853	0.897		0.677
Inception	ranger	0.001	4	Não	0.855	0.785 ( $\pm 0.056$ )	0.935	0.794	0.892		0.679
Reg-Net16X	radam	0.001	0	Não	0.899	0.818 ( $\pm 0.052$ )	0.929	0.882	0.913		0.722
Reg-Net16X	radam	0.001	1	Não	0.872	0.811 ( $\pm 0.038$ )	0.962	0.794	0.911		0.711
Reg-Net16X	radam	0.001	2	Não	0.865	0.808 ( $\pm 0.049$ )	0.941	0.872	0.88		0.737
Reg-Net16X	radam	0.001	3	Não	0.871	0.805 ( $\pm 0.047$ )	0.962	0.853	0.913		0.697
Reg-Net16X	radam	0.001	4	Não	0.855	0.8 ( $\pm 0.046$ )	0.971	0.794	0.93		0.669
Reg-Net16X	radam	0.01	0	Sim	0.991	0.927 ( $\pm 0.04$ )	0.99	1.0	0.963		0.891
Reg-Net16X	radam	0.01	1	Sim	0.973	0.922 ( $\pm 0.031$ )	0.991	0.955	0.945		0.9
Reg-Net16X	radam	0.01	2	Sim	0.984	0.919 ( $\pm 0.039$ )	0.991	0.977	0.971		0.868
Reg-Net16X	radam	0.01	3	Sim	0.984	0.93 ( $\pm 0.028$ )	0.991	0.977	0.969		0.892
Reg-Net16X	radam	0.01	4	Sim	0.989	0.928 ( $\pm 0.038$ )	1.0	0.977	0.954		0.902

O *RandAugment* trouxe uma variância menor nas métricas, principalmente na métrica principal AUC. Os resultados demonstram que *Data Augmentation* é uma ferramenta robusta para lidar com o problema de sobreajuste (Seção 3.5.9) dos modelos e variabilidade do desempenho.

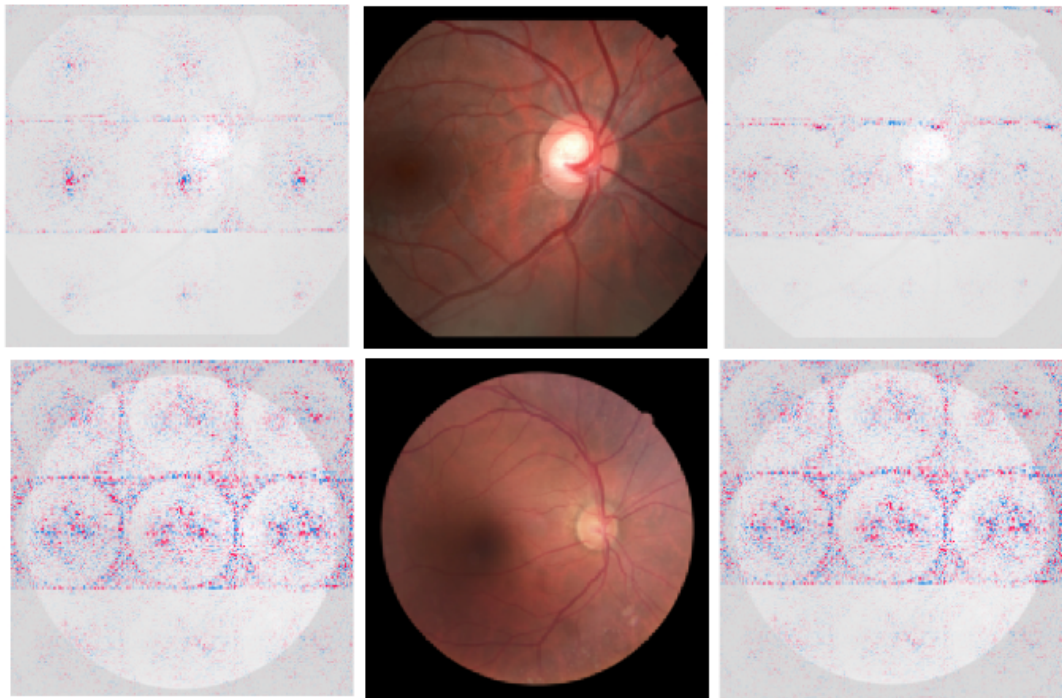
## 5.3 Explicabilidade

Serão analisadas, nesta seção, a explicabilidade e a interpretação dos melhores modelos. O objetivo é verificar a interação dos modelos com os dados, tanto os de imagem quanto os tabulares, para ser possível entender o porquê da decisão do modelo. Foram avaliadas tanto as imagens do fundo de olho, quanto os valores tabelados contendo as medições do exame OCT e as características criadas e descritas na seção 4.6.

### 5.3.1 Explicabilidade Visual

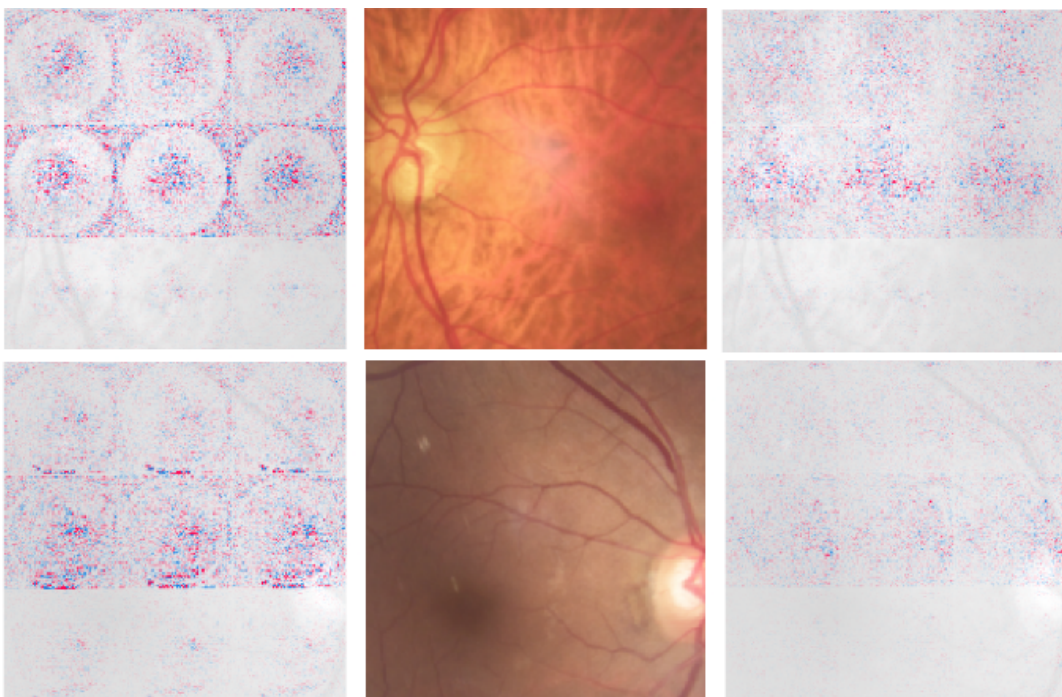
A figura abaixo 5.6 compara a explicabilidade de duas arquiteturas de modelos. A explicabilidade da coluna da esquerda é de uma arquitetura que utiliza somente uma imagem e a da direita utiliza duas imagens como entrada. Na imagem superior esquerda existe uma concentração de pontos no meio dos círculos e também na parte inferior da imagem. Isto se dá ao fato de que a arquitetura que utiliza somente uma entrada deve aprender padrões dos dois tipos de imagem citadas anteriormente 4.1. Para o modelo que utiliza duas imagens, imagens da direita, existe um *Backbone* para cada tipo de imagem. Na imagem da linha inferior, todas as duas arquiteturas de modelos deram a mesma explicabilidade para a imagem do fundo de olho. O modelo que utiliza somente uma imagem consegue se adaptar a diferentes tipos e recortes de imagens geradas pela Fundoscopia.

**Figura 5.6: Imagem com explicabilidade do modelo. A coluna da esquerda se refere a arquitetura do modelo utilizando somente uma imagem e, a da direita, utiliza duas imagens. Foto contendo a borda preta**



Fonte: Elaborado pelo autor.

**Figura 5.7: Imagem com explicabilidade do modelo. A coluna da esquerda mostra o modelo utilizando somente uma imagem; a da direita, duas imagens. Fonte recortada sem a borda preta**



Fonte: Elaborado pelo autor.

A segunda figura 5.7 traz a explicabilidade para as imagens sem a margem preta

presente no primeiro tipo de imagem de fundo de olho 5.6. A primeira arquitetura de modelo, na coluna da esquerda, adaptou a explicabilidade para extrair características desse tipo de imagem, formando círculos e uma pequena concentração de pontos vermelhos e azuis arredondada no meio. Da mesma forma, como na figura anterior 5.6, a primeira arquitetura de modelo possui uma concentração de pontos na parte inferior da imagem. Estas concentrações inferiores se devem ao fato de que o disco e o copo óptico podem estar em qualquer posição da imagem, e, para isto, o modelo precisa analisar todos os cantos possíveis. É importante ressaltar que a arquitetura utilizando duas imagens como entrada, na coluna da direita, praticamente não extraiu nenhum benefício da segunda foto. Isto corrobora com os resultados obtidos nesta tabela 5.2, onde a utilização de duas imagens nada acrescenta para o desempenho da tarefa.

A explicabilidade se concentra nas mesmas regiões para identificar tanto a presença como a ausência do glaucoma. Isto ocorre pela mistura de pontos azuis (diagnóstico negativo) e pontos vermelhos (diagnóstico positivo) na área da imagem em geral, e como elemento principal, os pontos se concentram onde fica o disco e escavação do disco óptico.

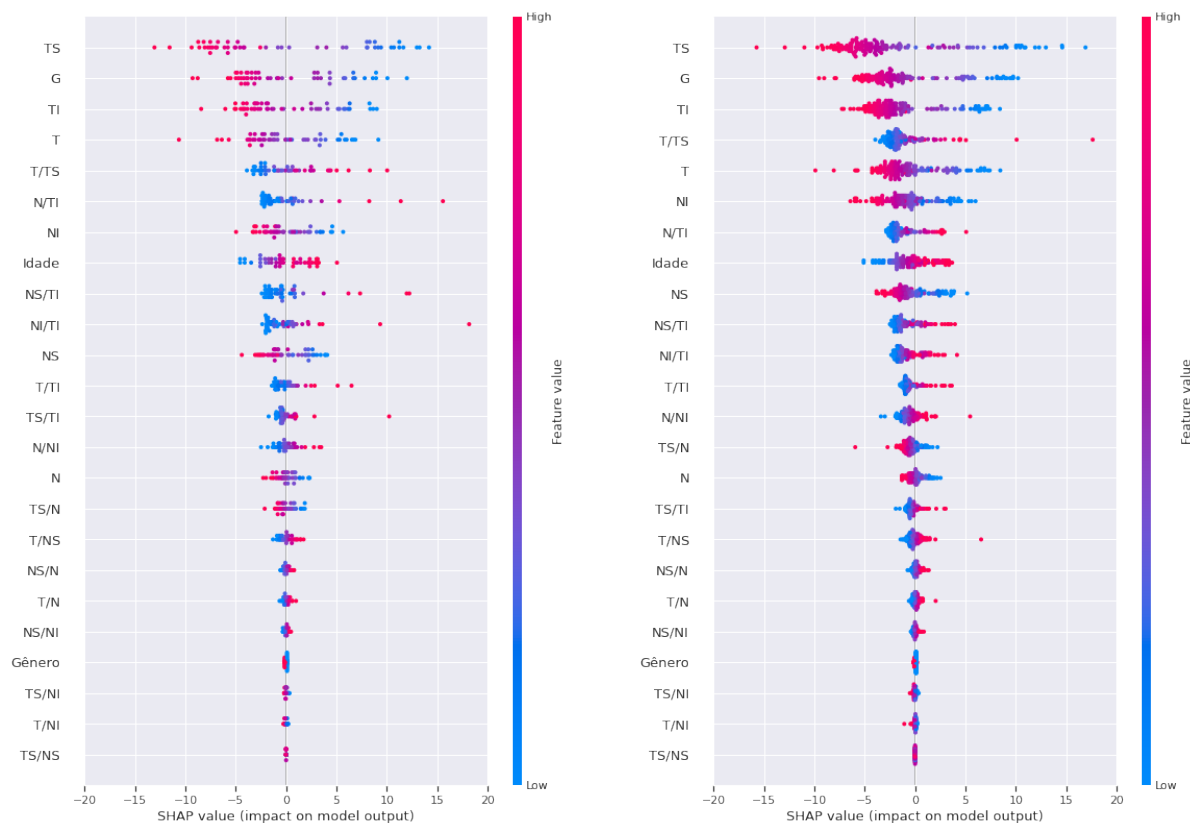
### 5.3.2 Importância das Características

Nesta seção será respondido à questão **Q2** (Seção 1.3) explanando a importância de cada uma das características do exame OCT, bem como as demográficas, utilizadas pelo modelo, serão analisadas, inclusive aquelas criadas por engenharia de características. As análises a seguir mostram quais foram as características mais importantes para o modelo quanto ao diagnóstico.

**Figura 5.8: Importância das características no melhor modelo em desempenho relativo a AUC. O modelo atingiu AUC de 0.99 usando dados do OCT**

(a) Importância das características nos dados de treino do conjunto AC.

(b) Importância das características nos dados de teste AD.



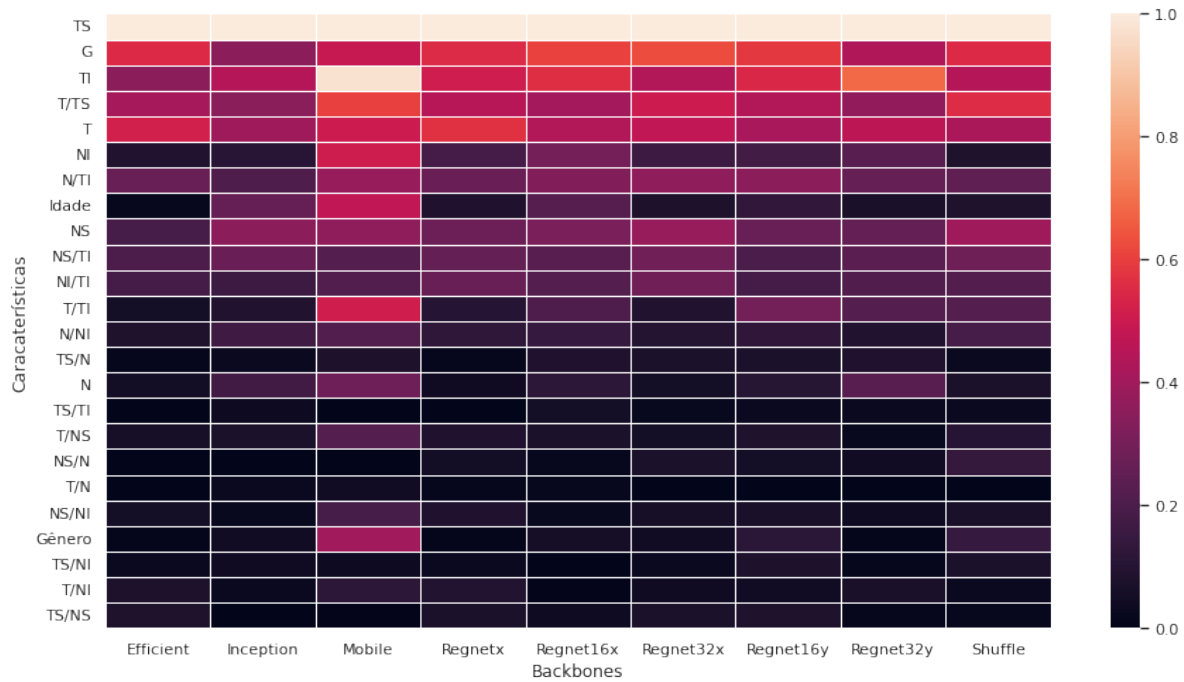
Fonte: Elaborado pelo autor.

A figura 5.8 mostra, de cima para baixo, as características mais importantes para os modelos. O gráfico da esquerda é do conjunto AC (Amostra Conhecida) e o da direita do AD (Amostra Desconhecida). As 3 primeiras características, valor médio global (G), temporal superior (TS) e temporal inferior (TI) estão na mesma posição de importância para os dois conjuntos de dados. Isto mostra que, nos dois conjuntos de dados, o impacto médio dessas características não sofre alterações bruscas. Com relação as outras características, suas posições de importância mudam entre os dois conjuntos de dados, no entanto, nenhuma mudança é significativa.

Ainda analisando a figura 5.8, os valores altos das características TS/G/TI representando camada de fibras nervosas mais espessas, está associado a ausência do glaucoma. Outro ponto importante a ser considerado é que os dados mostram que quanto mais avançada a idade do paciente, maiores são as chances do glaucoma ser diagnosticado. Com relação ao gênero do paciente, o modelo não trouxe nenhum tipo de relação importante. Para as características criadas sinteticamente, quando a razão T/TS aumenta, o modelo mostra que a espessura de TS degenera mais rápido. Para N/TI pode ser o oposto, ou seja, significa que a camada de fibras nervosas se degradam mais rápido na área temporal

comparado a Nasal (N).

**Figura 5.9:** Mapa de calor importância (importância normalizada entre 0 e 1) das características para cada *Backbones* presentes na etapa de validação cruzada. As características estão ordenadas de acordo com sua importância média geral.



Fonte: Elaborado pelo autor.

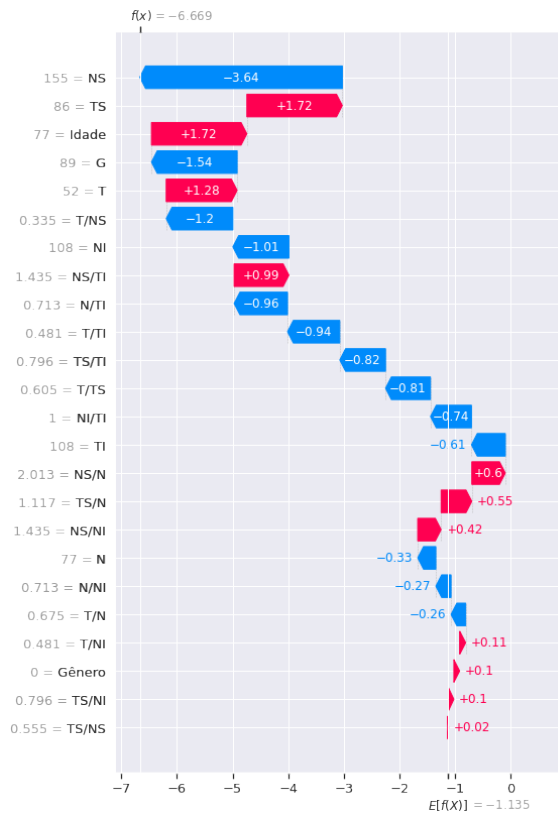
O mapa de calor 5.9 mostra que TS assume um papel importante em quase todos os *Backbones*. Pode-se dizer, então, que esta característica é de fato importante para o diagnóstico do glaucoma pelo exame OCT e, por isso, deve ser analisada com cuidado pelos profissionais. Em suma, todas as características relacionadas a área temporal, com exceção das criadas sinteticamente, ficaram acima das características nasais, dando ainda mais relevância para esta região. A área TI assumiu um papel importante quando se utilizou o *Backbone* Mobile. Outra análise relevante é que este mesmo *Backbone* trouxe importância para as caraterísticas que envolvem gênero e idade, diferentemente do que ocorreu nos outros *Backbones*. Algumas outras características parecem ser irrelevantes para qualquer modelo, como por exemplo: TS/NI, T/NI, TS/NS, NS/NI, T/N, NS/N e TS/TI. Para características não-sintéticas e regionais, a menos relevante para a modelagem foi a área nasal geral (N).

### 5.3.3 Análise de casos individuais

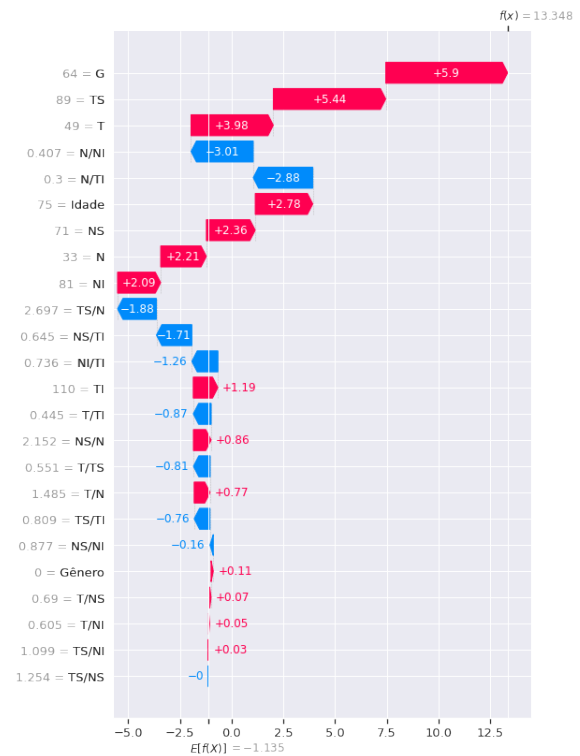
Nesta seção, casos individuais de pacientes serão avaliados, tendo em vista a decisão tomada pelo modelo. Na primeira parte, serão analisados dois casos confirmados de glaucoma. Observou-se que, em um, o modelo tomou a decisão correta; em outro, o modelo errou em sua decisão. Para isto, um gráfico de cascata (*waterfall*) do SHAP foi utilizado. Este gráfico contém as informações das características do paciente e como cada uma delas contribuiu, em magnitude, para a decisão do modelo, de acordo com os valores atribuídos as características.

**Figura 5.10: Gráfico de cascata para casos com presença de glaucoma.**

(a) Decisão errada do modelo indicando ausência do glaucoma.



(b) Decisão correta do modelo indicando presença do glaucoma.



Fonte: Elaborado pelo autor.

A figura 5.10 possui dois gráficos de cascata, sendo o da direita, o modelo que tomou a decisão correta e, o da esquerda, a decisão errada. O valor na parte inferior da figura mostra o valor esperado  $E[f(x)]$ , ou seja, a média de decisão do modelo. Para valores acima dessa média, o modelo decide o diagnóstico como positivo, caso contrário, negativo. A saída do modelo é vista na parte superior do gráfico  $f(x)$ . Na imagem da direita, a média global G foi a característica que mais contribuiu para a decisão do modelo. Pela distribuição da figura 4.3, é possível verificar que os valores de G entre 40 e 70 são



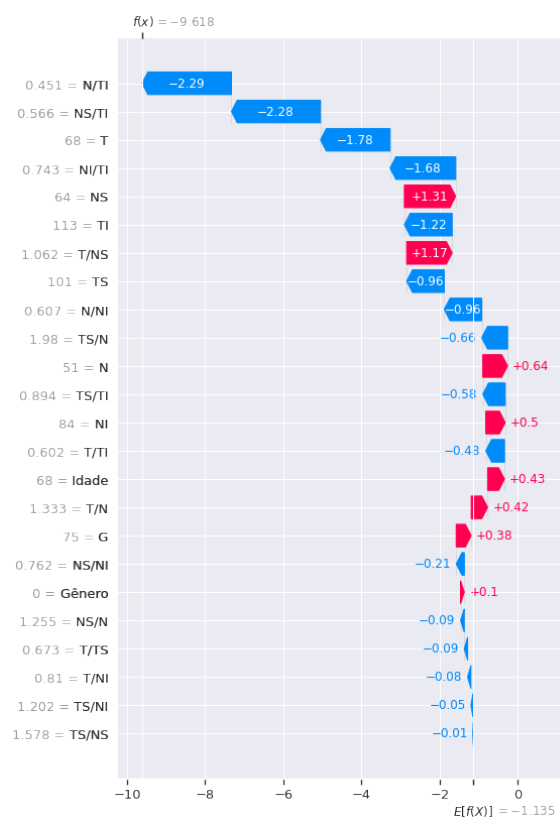
um grupo composto majoritariamente por pessoas diagnosticadas com o glaucoma. O mesmo vale para TS e para T, apesar de a última estar no limiar entre a população com glaucoma e sem glaucoma. A idade de 75 anos foi um fator também que contribuiu para o modelo diagnosticar como positivo. No gráfico de cascata da esquerda, a característica mais importante foi NS com um valor de 155. Este valor, na figura 4.3, está na cauda direita da população, onde existe uma concentração maior de casos negativados, levando o modelo a decidir erradamente.

**Figura 5.11: Gráfico de cascata para casos com ausência de glaucoma.**

(a) Decisão errada do modelo indicando presença do glaucoma.



(b) Decisão correta do modelo indicando ausência do glaucoma.



Fonte: Elaborado pelo autor.

Na figura 5.11a o valor de TS e razão entre T/TS na imagem da esquerda foram os causadores da distorção na decisão do modelo. Com a medição de TS igual a 12, a razão ficou 6 vezes maior entre T e TS. O valor coletado de TS pode ser um possível *outlier*, tendo em vista a figura 4.3, um problema de medição ou apenas um erro esporádico na anotação dos dados. O modelo foi impactado por este valor e, como consequência, foi dado um diagnóstico positivo errado. Quanto ao diagnóstico da imagem da direita 5.11b, TI tem mais que o dobro de espessura que a região N. A partir dessa informação, o modelo diagnosticou corretamente a ausência do glaucoma.



## Capítulo 6

### Conclusão

Dentre os 28 melhores modelos em relação ao desempenho da métrica AUC, o modelo RegNet aparece 16 usando diferentes arquiteturas de modelos e configurações de hiperparâmetros. O modelo se mostrou uma escolha robusta quanto a detecção do glaucoma utilizando imagens de fundo de olho e o exame OCT. Seu melhor desempenho apresentou uma AUC de 0.99, sendo o melhor modelo de todos utilizando a versão RegNet16X.

O *RandAugment* mostrou benefícios quanto a variabilidade dos resultados, sendo uma ferramenta essencial para evitar o sobreajuste. A técnica de *Learning Rate Decay* e *Early Stopping* trouxeram um incremento no desempenho dos modelos, aumentando os valores máximos das métricas AUC, SP e SN.

A arquitetura que utiliza somente uma imagem como entrada consegue extrair características de diferentes fotos de fundo de olho, sem a necessidade de se ter um segundo modelo para processamento de uma segunda imagem. As características G, T e TS foram as mais importantes para os modelos utilizando dados tabulares do exame OCT.

# Referências Bibliográficas

- [1] Pulkit Agrawal, Ross B. Girshick, and Jitendra Malik. Analyzing the performance of multilayer neural networks for object recognition. *ArXiv*, abs/1407.1610, 2014.
- [2] Hafsah Ahmad, Abubakar Yamin, Aqsa Shakeel, Syed Ormer Gillani, and Umer Ansari. Detection of glaucoma using retinal fundus images. In *2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE)*, pages 321–324, 2014.
- [3] Ioannis D. Apostolopoulos, Dimitris I. Apostolopoulos, Trifon I. Spyridonidis, Nikolaos D. Papathanasiou, and George S. Panayiotakis. Multi-input deep learning approach for cardiovascular disease diagnosis using myocardial perfusion imaging and clinical data. *Physica Medica*, 84:168–177, 2021.
- [4] Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. Factors of transferability for a generic convnet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38:1790–1802, 2016.
- [5] baidaa al bander, waleed al nuaimy, bryan m. williams, and yalin zheng. multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. *biomedical signal processing and control*, 40:91–101, 2018.
- [6] Daniele Barros, Julio Moura, Cefas Freire, Alexandre Taleb, Ricardo Valentim, and Philippi Morais. Machine learning applied to retinal image processing for glaucoma detection: Review and perspective. *BioMedical Engineering OnLine*, 19, 04 2020.
- [7] Daniele M. S. Barros, Julio C. C. Moura, Cefas R. Freire, Alexandre C. Taleb, Ricardo A. M. Valentim, and Philippi S. G. Morais. Machine learning applied to retinal image processing for glaucoma detection: review and perspective. *BioMedical Engineering OnLine*, 19(1):20, Apr 2020.
- [8] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.
- [9] Claude Burgoyne, J. Downs, Anthony Bellezza, Jun-Kyo Suh, and Richard Hart. The optic nerve head as a biomechanical structure: A new paradigm for understanding the role of iop-related stress and strain in the pathophysiology of glaucomatous optic nerve head damage. *Progress in retinal and eye research*, 24:39–73, 02 2005.

- [10] Juan Carrillo, Lola Bautista, Jorge Villamizar, Juan Rueda, Mary Sanchez, and Daniela Rueda. Glaucoma detection using fundus images of the eye. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–4, 2019.
- [11] Qin A Chang, Abdulraheem Qasem, Sheshadri Madhusudhana, and Alexey Glazyrin. The  $t(14;18)(q32;q21)$  with extra MYC signal - is it a gray zone lymphoma? *Int. J. Clin. Exp. Pathol.*, 8(8):9602–9608, August 2015.
- [12] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.
- [13] Gabriel Costa, Anselmo Paiva, Geraldo Braz Júnior, and Marco Ferreira. Covid-19 automatic diagnosis with ct images using the novel transformer architecture. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 293–301, Porto Alegre, RS, Brasil, 2021. SBC.
- [14] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232, 1958.
- [15] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical data augmentation with no separate search. *CoRR*, abs/1909.13719, 2019.
- [16] Daniele Montenegro da Silva Barros. *Uma arquitetura digital baseado em aprendizado de máquina para diagnóstico de doenças oculares: um estudo aplicado ao glaucoma*. PhD thesis, Universidade Federal do Rio Grande do Norte, 2020.
- [17] Danielle Monteiro. Comunidades rurais remotas carecem de políticas públicas adequadas às realidades locais. Disponível em: <https://informe.ensp.fiocruz.br/noticias/53419>, Acesso em 26 de janeiro de 2023, 2022.
- [18] Rosana Aurélio de Jesus. *Investigação da importância dos canais de cores em imagens de fundoscopia para auxílio na identificação do glaucoma via abordagens de aprendizado de máquina*. PhD thesis, Universidade Federal do Espírito Santo, 2021.
- [19] Alan Carlos de Moura Lima. *Aprendizagem Profunda Aplicado Ao Diagnóstico do Glaucoma*. PhD thesis, Universidade Federal do Maranhão, 2019.
- [20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

- [21] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020.
- [22] Shaker El-Sappagh, Jose M. Alonso, S. M. Riazul Islam, Ahmad M. Sultan, and Kyung Sup Kwak. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for alzheimer’s disease. *Scientific Reports*, 11(1):2660, Jan 2021.
- [23] Hitendra Garg, Neeraj Gupta, Rohit Agrawal, Shivendra Shivani, and Bhisham Sharma. A real time cloud-based framework for glaucoma screening using efficientnet. *Multimedia Tools and Applications*, 81(24):34737–34758, Oct 2022.
- [24] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.
- [25] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states. *Proceedings of the National Academy of Sciences*, 114(50):13108–13113, 2017.
- [26] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014.
- [27] Glaucoma Foundation. Who’s at risk. Disponível em: <https://glaucomafoundation.org/aboutglaucoma/whos-at-risk/>. Acesso em 07 de fev de 2023, 2023.
- [28] R.C. Gonzalez and R.E. Woods. *Processamento Digital De Imagens*. ADDISON WESLEY BRA, 2009.
- [29] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. <http://www.deeplearningbook.org>.
- [30] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- [31] F. Guo, W. Li, X. Zhao, J. Qiu, and Y. Mai. A mobile app for Glaucoma diagnosis and its possible clinical applications. *BMC Med Inform Decis Mak*, 20(Suppl 3):128, Jul 2020.

- [32] Yuki Hagiwara, Joel En Wei Koh, Jen Hong Tan, Sulatha V. Bhandary, Augustinus Laude, Edward J. Ciaccio, Louis Tong, and U. Rajendra Acharya. Computer-aided diagnosis of glaucoma using fundus images: A review. *Computer Methods and Programs in Biomedicine*, 165:1–12, 2018.
- [33] Karimollah Hajian-Tilaki. Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, 4:627–635, 09 2013.
- [34] M. S. Haleem, L. Han, J. van Hemert, and A. Fleming. Glaucoma classification using Regional Wavelet Features of the ONH and its surroundings. *Annu Int Conf IEEE Eng Med Biol Soc*, 2015:4318–4321, Aug 2015.
- [35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [36] Otaviano Helene and André Frazão Helene. Alguns aspectos da óptica do olho humano. *Revista Brasileira de Ensino de Física*, 33(Rev. Bras. Ensino Fís., 2011 33(3)), Jul 2011.
- [37] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861, 2017.
- [38] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. *CoRR*, abs/1709.01507, 2017.
- [39] Usama Iqbal. Smartphone fundus photography: a narrative review. *International Journal of Retina and Vitreous*, 7(1):44, Jun 2021.
- [40] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.
- [41] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. Vision transformer for small-size datasets. *CoRR*, abs/2112.13492, 2021.
- [42] Lili Li, Shujuan Zhang, and Bin Wang. Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. *Sensors (Basel)*, 22(1):173, December 2021.
- [43] Zhixi Li, Yifan He, Stuart Keel, Wei Meng, Robert T. Chang, and Mingguang He. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*, 125(8):1199–1206, 2018.

- [44] Zhizhong Li and Derek Hoiem. Learning without forgetting. *CoRR*, abs/1606.09282, 2016.
- [45] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734, 2015.
- [46] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.
- [47] Stefan Maetschke, Bhavna Antony, Hiroshi Ishikawa, Gadi Wollstein, Joel Schuman, and Rahil Garnavi. A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS One*, 14(7):e0219126, July 2019.
- [48] Felipe A Medeiros, Alessandro A Jammal, and Atalie C Thompson. From machine to machine: An OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology*, 126(4):513–521, April 2019.
- [49] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dhharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb 2015.
- [50] Kenichi Nakahara, Ryo Asaoka, Masaki Tanito, Naoto Shibata, Keita Mitsuhashi, Yuri Fujino, Masato Matsuura, Tatsuya Inoue, Keiko Azuma, Ryo Obata, and Hiroshi Murata. Deep learning-assisted (automatic) diagnosis of glaucoma using a smartphone. *British Journal of Ophthalmology*, 106(4):587–592, 2022.
- [51] Janika Nättinen, Antti Jylhä, Ulla Aapola, Roger W. Beuerman, Matti Nykter, Juha Kesseli, and Hannu M. T. Uusitalo. Effects of preserved and preservative free glaucoma drugs on proteomic expression levels in corneal and conjunctival epithelial cells in vitro. *Investigative Ophthalmology & Visual Science*, 56(7):1978–1978, Jun 2015.
- [52] Jagadish Nayak, Rajendra Acharya U, P Subbanna Bhat, Nakul Shetty, and Teik-Cheng Lim. Automated diagnosis of glaucoma using digital fundus images. *J. Med. Syst.*, 33(5):337–346, October 2009.

- [53] Sejong Oh, Yuli Park, Kyong Jin Cho, and Seong Jae Kim. Explainable machine learning model for glaucoma diagnosis and its interpretation. *Diagnostics*, 11(3), 2021.
- [54] José A. A. Ottaiano, Marcos P. Ávila, Cristiano C. Umbrelino, and Alexandre C. Taleb. As condições de saúde ocular no brasil. *CBO*, 1, 2019.
- [55] Erinaldo Costa Passos, A. V. Andrade-Neto, and Thierry Lemaire. Comportamento otico do olho humano e suas ´ ametropias, 2008.
- [56] H. Pedrini and W.R. Schwartz. *Análise de imagens digitais: princípios, algoritmos e aplicações*. Thomson Learning, 2008.
- [57] Junfeng Peng, Kaiqiang Zou, Mi Zhou, Yi Teng, Xiongyong Zhu, Feifei Zhang, and Jun Xu. An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. *Journal of medical systems*, 45(5):61–61, 2021.
- [58] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning, 2017.
- [59] PNS. Pesquisa nacional de saúde. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=sobre>. Acesso em 08 de outubro de 2022, 2019.
- [60] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.
- [61] Ilija Radosavovic, Raj Prateek Kosaraju, Ross B. Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. *CoRR*, abs/2003.13678, 2020.
- [62] Sebastian Raschka. Model evaluation, model selection, and algorithm selection in machine learning. *CoRR*, abs/1811.12808, 2018.
- [63] Usha Ruby and Vamsidhar Yendapalli. Binary cross entropy with deep learning technique for image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, 9, 10 2020.
- [64] W. Ruengkitpinyo, P. Vejjanugraha, W. Kongprawechnon, T. Kondo, P. Bunnun, and H. Kaneko. An automatic glaucoma screening algorithm using cup-to-disc ratio and isnt rule with support vector machine. In *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society*, pages 000517–000521, 2015.
- [65] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014.

- [66] Sebastian Schlag, Matthias Schmitt, and Christian Schulz. Faster support vector machines. *CoRR*, abs/1808.06394, 2018.
- [67] Ursula Schmidt-Erfurth, Amir Sadeghipour, Bianca S. Gerendas, Sebastian M. Waldstein, and Hrvoje Bogunović. Artificial intelligence in retina. *Progress in Retinal and Eye Research*, 67:1–29, 2018.
- [68] L.S. Shapley and A.E. Roth. *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press, 1988.
- [69] Fabrício Reis Da Silva. *Diagnóstico de glaucoma baseado em classificadores de parendizagem de máquina utilizando dados do Spectral Domain-OCT e Perimetria Automatizada Acromática*. PhD thesis, Universidade Estadual de Campinas, 2013.
- [70] Mahmoud Smaida, Serhii Yaroshchak, and Youness El Barg. Dcgan for enhancing eye diseases classification. In *International Workshop on Computer Modeling and Intelligent Systems*, 2021.
- [71] Jaemin Son, Joo Young Shin, Hoon Dong Kim, Kyu-Hwan Jung, Kyu Hyung Park, and Sang Jun Park. Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. *Ophthalmology*, 127(1):85–94, 2020.
- [72] M. Caroline Viola Stella Mary, Elijah Blessing Rajasingh, and Ganesh R. Naik. Retinal fundus image analysis for diagnosis of glaucoma: A comprehensive survey. *IEEE Access*, 4:4327–4354, 2016.
- [73] Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3):647–665, Dec 2014.
- [74] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.
- [75] Raquel Sánchez-Cauce, Jorge Pérez-Martín, and Manuel Luque. Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Computer Methods and Programs in Biomedicine*, 204:106045, 2021.
- [76] Karim Talaat, Obada T Fathi, Saeed M Alamoudi, Muhanad G Alzahrani, Rayan M Mukhtar, and Muhammad A Khan. Types of glaucoma and associated comorbidities among patients at king abdulaziz medical city, jeddah. *Cureus*, 13(6):e15574, June 2021.



- [77] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946, 2019.
- [78] Kashvi Taunk, Sanjukta De, Srishti Verma, and Aleena Swetapadma. A brief review of nearest neighbor algorithm for learning and classification. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 1255–1260, 2019.
- [79] Kenneth W Tobin, Edward Chaum, V Priya Govindasamy, and Thomas P Karnowski. Detection of anatomic structures in human retinal imagery. *IEEE Trans. Med. Imaging*, 26(12):1729–1739, December 2007.
- [80] Michael Voskoglou. A study on fuzzy systems. *American Journal of Computational and Applied Mathematics*, 2, 12 2012.
- [81] R. N. Weinreb and P. T. Khaw. Primary open-angle glaucoma. *Lancet*, 363(9422):1711–1720, May 2004.
- [82] Robert N. Weinreb, Tin Aung, and Felipe A. Medeiros. The pathophysiology and treatment of glaucoma: A review. *JAMA*, 311(18):1901–1911, 2014.
- [83] Jing Xu, Yu Pan, Xinglin Pan, Steven Hoi, Zhang Yi, and Zenglin Xu. Regnet: Self-regulated network for image classification, 2021.
- [84] Peng Yao, Shuwei Shen, Mengjuan Xu, Peng Liu, Fan Zhang, Jinyu Xing, Pengfei Shao, Benjamin Kaffenberger, and Ronald X. Xu. Single model deep learning on imbalanced small datasets for skin lesion classification. *CoRR*, abs/2102.01284, 2021.
- [85] Tae Keun Yoo, Ik Hee Ryu, Hannuy Choi, Jin Kuk Kim, In Sik Lee, Jung Sub Kim, Geunyoung Lee, and Tyler Hyungtaek Rim. Explainable Machine Learning Approach as a Tool to Understand Factors Used to Select the Refractive Surgery Technique on the Expert Level. *Translational Vision Science & Technology*, 9(2):8–8, 02 2020.
- [86] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *NIPS*, 2014.
- [87] Kaichao You, Mingsheng Long, Michael I. Jordan, and Jianmin Wang. Learning stages: Phenomenon, root cause, mechanism hypothesis, and implications. *CoRR*, abs/1908.01878, 2019.
- [88] Camila Zangalli, Shelly Gupta, and George Spaeth. The disc as the basis of treatment for glaucoma. *Saudi journal of ophthalmology : official journal of the Saudi Ophthalmological Society*, 25:381–7, 10 2011.

- [89] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. *CoRR*, abs/1707.01083, 2017.
- [90] Zhuo Zhang, Beng Hai Lee, Jiang Liu, Damon Wing Kee Wong, Ngan Meng Tan, Joo Hwee Lim, Fengshou Yin, Weimin Huang, Huiqi Li, and Tien Yin Wong. Optic disc region of interest localization in fundus image for glaucoma detection in argali. In *2010 5th IEEE Conference on Industrial Electronics and Applications*, pages 1686–1689, 2010.
- [91] R. Zhao, X. Chen, X. Liu, Z. Chen, F. Guo, and S. Li. Direct Cup-to-Disc Ratio Estimation for Glaucoma Screening via Semi-Supervised Learning. *IEEE J Biomed Health Inform*, 24(4):1104–1113, Apr 2020.
- [92] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018.
- [93] Chengzhang Zhu, Bei Zou, Rongchang Zhao, Jinkai Cui, Xuanchu Duan, Zailiang Chen, and Yixiong Liang. Retinal vessel segmentation in colour fundus images using extreme learning machine. *Computerized Medical Imaging and Graphics*, 55:68–77, 2017. Special Issue on Ophthalmic Medical Image Analysis.
- [94] Barret Zoph and Quoc V. Le. Neural architecture search with reinforcement learning. *CoRR*, abs/1611.01578, 2016.

## Apêndice A

### Validação Cruzada Resultados

A tabela a seguir mostra os resultados completos obtidos pela utilização do *Learning Rate Decay/Early Stopping* (LRD) comparado aos experimentos sem sua utilização (C).

<i>Back- bone</i>	Op- tim	LR	Usando OCT	AUC				SP				SN			
				Melhor		Média		Melhor		Média		Melhor		Média	
				C	LRD	C	LRD	C	LRD	C	LRD	C	LRD	C	LRD
Reg-Net16X	radam	0.001	Não	0.792	0.899	0.733	0.818	0.78	0.929	0.753	0.913	0.72	0.882	0.601	0.722
Reg-Net16X	radam	0.0	Não	0.828	0.888	0.746	0.801	0.874	0.924	0.857	0.906	0.783	0.853	0.711	0.695
Incep-tion	ranger	0.0	Não	0.81	0.859	0.724	0.781	0.946	0.951	0.911	0.891	0.674	0.794	0.526	0.67
Incep-tion	sgd	0.01	Não	0.812	0.811	0.76	0.756	0.928	0.962	0.896	0.88	0.696	0.818	0.554	0.633
Incep-tion	ranger	0.001	Não	0.812	0.899	0.745	0.795	0.928	0.935	0.894	0.923	0.696	0.882	0.627	0.666
Reg-Net32X	radam	0.001	Não	0.849	0.882	0.763	0.805	0.937	0.947	0.906	0.905	0.761	0.882	0.59	0.706
ResNet	adam	0.01	Sim	0.939	0.943	0.885	0.673	0.964	1.0	0.916	0.993	0.913	0.912	0.703	0.353
ResNet	ranger	0.01	Sim	0.937	0.961	0.813	0.836	0.982	0.991	0.949	0.969	0.891	0.941	0.756	0.702
Reg-Net32Y	adam	0.0	Sim	0.93	0.945	0.833	0.897	0.991	0.982	0.95	0.951	0.87	0.955	0.696	0.842
Reg-Net32Y	adam	0.01	Sim	0.926	0.937	0.837	0.845	0.982	1.0	0.953	0.989	0.87	0.882	0.787	0.771
mobile	radam	0.01	Sim	0.926	0.928	0.828	0.892	0.982	1.0	0.955	0.946	0.87	0.941	0.623	0.839
Shuffle-Net	ranger	0.01	Sim	0.949	0.981	0.858	0.935	0.964	0.975	0.945	0.963	0.935	1.0	0.748	0.907
Shuffle-Net	radam	0.001	Sim	0.934	0.926	0.867	0.826	0.955	0.953	0.915	0.924	0.913	0.971	0.747	0.728
Shuffle-Net	ranger	0.001	Sim	0.932	0.903	0.829	0.795	0.973	0.971	0.939	0.927	0.891	0.882	0.776	0.663

Fonte: Elaborado pelo autor.

<i>Back- bone</i>	Op- tim	LR	Usando OCT	AUC				SP				SN			
				Melhor		Média		Melhor		Média		Melhor		Média	
				C	LRD	C	LRD	C	LRD	C	LRD	C	LRD	C	LRD
Shuffle- Net	radam	0.01	Sim	0.932	0.989	0.865	0.937	0.973	1.0	0.945	0.976	0.891	0.977	0.691	0.897
RegNetX	adam	0.01	Sim	0.926	0.984	0.855	0.934	0.982	0.991	0.951	0.968	0.87	0.977	0.635	0.9
RegNetX	adam	0.001	Sim	0.915	0.986	0.885	0.914	0.982	0.973	0.949	0.967	0.848	1.0	0.618	0.86
RegNetX	radam	0.01	Sim	0.932	0.986	0.859	0.933	0.973	1.0	0.94	0.968	0.891	1.0	0.76	0.897
Reg- Net16Y	ranger	0.01	Sim	0.932	0.973	0.829	0.89	0.973	0.991	0.94	0.947	0.891	0.955	0.74	0.833
Reg- Net16Y	adam	0.01	Sim	0.939	0.939	0.884	0.863	0.964	1.0	0.941	0.991	0.913	0.912	0.712	0.815
Reg- Net16Y	radam	0.01	Sim	0.934	0.989	0.864	0.933	0.955	1.0	0.936	0.973	0.913	0.977	0.685	0.893
efficient	adam	0.01	Sim	0.928	0.979	0.865	0.847	0.964	0.992	0.927	0.96	0.891	0.977	0.728	0.734
Reg- Net32X	radam	0.01	Sim	0.926	0.986	0.829	0.94	0.982	0.991	0.953	0.958	0.87	1.0	0.716	0.922
Reg- Net32X	radam	0.001	Sim	0.926	0.865	0.852	0.801	0.982	0.943	0.958	0.892	0.87	0.824	0.673	0.711
Reg- Net32X	ranger	0.01	Sim	0.932	0.951	0.853	0.899	0.973	0.991	0.953	0.957	0.891	0.977	0.711	0.841
Reg- Net32X	adam	0.01	Sim	0.932	0.986	0.84	0.94	0.973	1.0	0.941	0.973	0.891	1.0	0.782	0.907
Reg- Net16X	radam	0.01	Sim	0.932	0.991	0.85	0.927	0.973	0.99	0.948	0.963	0.891	1.0	0.744	0.891
Incep- tion	ranger	0.01	Sim	0.929	0.979	0.84	0.909	0.946	0.99	0.909	0.969	0.913	0.977	0.738	0.85

Fonte: Elaborado pelo autor.