

# Classificação do glaucoma usando deep learning e imagens de fundo de olho obtidas com smartphone



Clerimar Paulo Bragança

Faculdade de Ciências e Tecnologia - Ciências da Informação - Sistemas,  
Tecnologias e Gestão da Informação

Universidade Fernando Pessoa

Tese submetida ao grau de

*Doutor*

2023

## Resumo

Nos últimos anos, os algoritmos de *deep learning* mostraram um potencial sem precedentes para auxiliar no diagnóstico da doença ocular glaucoma, considerada uma das principais causas de cegueira irreversível no mundo. Esses algoritmos são comumente usados em estudos para identificar essa doença a partir de imagens digitais do fundo de olho, amplamente empregadas na maioria das clínicas oftalmológicas e adquiridas por meio do exame tradicional de fundoscopia. Além dos algoritmos de *deep learning*, o desenvolvimento tecnológico permitiu a criação de dispositivos portáteis com capacidade de aplicar o exame fundoscópico por meio de smartphones, tornando o teste ainda mais acessível. Diante dessa situação, este trabalho coletou um novo conjunto de dados chamado Brasil Glaucoma (BrG), contendo 2.000 imagens da retina obtidas com o oftalmoscópio *panoptic Wech Allyn*, um dispositivo considerado simples e portátil, com conexão de um adaptador para smartphone. Em seguida, treinou-se um algoritmo desenvolvido por meio de um *ensemble* de redes neurais convolucionais, de modo que a sua precisão na classificação das imagens BrG em glaucomatosas e normais foi estimada. Assim, além de um teste em imagens globais, ou seja, apresentando toda a área fotografada, também foram realizados testes em três tipos de regiões de interesse: a região do disco óptico, obtida por meio de segmentação manual; a região específica do disco óptico; e o disco óptico e áreas adjacentes. Estes dois últimos procedimentos foram aplicados de forma automatizada com auxílio de uma rede neural *U-Net*. Considerando a classificação do glaucoma, o *ensemble* alcançou, no seu melhor desempenho, 95,4% de acurácia, 98,5% de sensibilidade e 98,2% de especificidade usando a região do disco óptico e adjacências como entrada. Deste modo, apesar de limitações como a falta de biomarcadores robustos e computáveis para o diagnóstico do glaucoma, concluiu-se que o algoritmo de *deep learning* desenvolvido possui capacidade substancial para a classificação desta neuropatia óptica a partir do reconhecimento de padrões em imagens fundoscópicas de baixa resolução adquiridas através de um smartphone. Portanto, essas tecnologias de aquisição de fotografias da retina combinadas com algoritmos de inteligência artificial podem ter um grande impacto em pesquisas sobre o glaucoma e no diagnóstico automatizado, contribuindo para a viabilidade de futuros programas de triagem populacional para essa doença ocular.

## Abstract

In recent years, deep learning algorithms have shown unprecedented potential to help diagnose glaucoma eye disease, considered one of the main causes of irreversible blindness in the world. These algorithms are commonly used in studies to identify this disease from digital images of the fundus, widely used in most ophthalmological clinics and acquired through the traditional fundoscopy examination. In addition to deep learning algorithms, technological development has allowed the creation of portable devices that can apply the fundoscopic examination using a smartphone, making the test even more accessible. Given this situation, this work established a new dataset called Brazil Glaucoma (BrG), containing retinal images 2.000 obtained with the Welch Allyn panoptic ophthalmoscope, a device considered simple and portable with a smartphone adapter connection. Then, an ensemble of convolutional neural networks was developed and its accuracy in classifying BrG images into glaucomatous and normal was estimated. Thus, besides a test on global images, that is, showing the entire photographed area, tests were also performed on three types of regions of interest: the optical disc region, obtained through manual segmentation; the specific optical disc region; and the optical disc and adjacent areas. The last two procedures were applied in an automated way with the aid of a U-Net neural network. Considering glaucoma classification, the it ensemble achieved its best performance, 95.4% accuracy, 98.5% sensitivity, and 98.2% specificity using the optic disc region and adjacencies as input. Thus, despite limitations such as the lack of robust and computable biomarkers for the diagnosis of glaucoma, it was concluded that the developed deep learning algorithm has substantial capacity for the classification of this optic neuropathy from the recognition of patterns in low-resolution images acquired through a smartphone. Therefore, these fundoscopic image acquisition technologies combined with artificial intelligence algorithms can have a great impact on glaucoma research and automated diagnosis, contributing to the feasibility of future population screening programs for this eye disease.

---

Dedico este trabalho aos meus irmãos (a) e esposa. Também dedico ao meu pai, Anísio José Bragança (in memoriam), que me ensinou a me reerguer diante das adversidades da vida, e à minha mãe, Maria Narcisa de Jesus Bragança, que sempre me apoiou nos meus sonhos e projetos.

## **Agradecimentos**

Em primeiro lugar, como cristão, agradeço a Deus por guiar-me, iluminar e permitir que eu caminhe em direção aos meus objetivos com tranquilidade e sem desanimar com as dificuldades.

Gostaria de agradecer aos meus orientadores. Prof. José Manuel Torres e Prof. Christophe Pinto de Almeida Soares, por toda a paciência e empenho com que sempre me orientaram neste trabalho e em todos aqueles realizados durante doutoramento. Muito obrigado por ter-me corrigido quando necessário sem nunca me desmotivar.

Por fim, agradeço aos meus amigos e familiares, por compreenderem a minha ausência em certos momentos e principalmente pelo constante apoio em minha caminhada.

# Índice

<b>Índice</b>	<b>v</b>
<b>Índice de Figuras</b>	<b>ix</b>
<b>Índice de Tabelas</b>	<b>xiv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Hipótese . . . . .	2
1.2 Objetivo geral . . . . .	3
1.3 Objetivos específicos . . . . .	3
1.4 Metodologia e resultados . . . . .	3
1.5 Estrutura da tese . . . . .	4
<b>2 Referencial teórico</b>	<b>5</b>
2.1 Inteligência Artificial (IA) . . . . .	5
2.1.1 Funções de ativação . . . . .	9
2.1.1.1 Função de ativação sigmoide . . . . .	9
2.1.1.2 Função de ativação (tanh) . . . . .	10
2.1.1.3 Função de ativação ReLU . . . . .	11
2.1.1.4 Função de ativação softmax . . . . .	12
2.2 Organização do processo de treino e classificação numa RNA . . . . .	12
2.2.1 Base de dados . . . . .	13
2.2.2 Pré-processamento de imagens . . . . .	14
2.2.3 Extração da região de interesse . . . . .	18
2.2.4 Extração de características . . . . .	20
2.2.5 Treinamento . . . . .	24
2.2.5.1 Função de perda . . . . .	26
2.2.5.2 Overfitting e underfitting . . . . .	27
2.3 Classificação . . . . .	28
2.4 Deep learning e redes neurais convolucionais CNN . . . . .	32
2.4.1 Redes neurais convolucionais CNN . . . . .	34

2.4.2	Extração de características com uso de redes CNNs . . . . .	37
2.4.3	Aprendizado por transferência . . . . .	39
2.5	Redes Adversárias Generativas (GANs) . . . . .	40
2.6	Métricas de avaliação . . . . .	40
<b>3</b>	<b>Estado da arte</b>	<b>44</b>
3.1	A degeneração do nervo óptico e o glaucoma . . . . .	44
3.2	Diagnóstico tradicional do glaucoma . . . . .	48
3.2.1	Danos estruturais glaucomatosos . . . . .	49
3.2.1.1	Relação escavação disco - CDR . . . . .	49
3.2.1.2	Regra ISNT . . . . .	51
3.2.1.3	Assimetria da escavação . . . . .	51
3.2.1.4	Exemplos de outros danos estruturais do disco óptico . . . . .	52
3.2.2	Dificuldades no diagnóstico precoce . . . . .	54
3.3	Epidemiologia . . . . .	55
3.4	Dificuldades de assistência oftalmológicas em países como o Brasil . . . . .	56
3.5	Avanços científicos e tecnológicos . . . . .	61
3.6	Trabalhos relacionados . . . . .	63
3.6.1	Principais bases de dados públicas . . . . .	63
3.6.2	Abordagens utilizando deep learning . . . . .	67
3.6.2.1	Classificação do glaucoma a partir da relação escavação disco . . . . .	69
3.6.2.2	Classificação do glaucoma a partir de vetores de características . . . . .	70
3.6.2.3	Classificação do glaucoma com uso de redes CNNs . . . . .	71
3.6.2.4	Classificação do glaucoma com uso de CGANs . . . . .	74
3.6.2.5	Classificação do glaucoma a partir de multitecnologias . . . . .	75
3.6.2.6	Síntese dos estudos relacionados . . . . .	76
3.6.3	Características gerais dos trabalhos relacionados . . . . .	77
<b>4</b>	<b>Construção da base de dados</b>	<b>81</b>
4.1	Escolha do aparelho de fotografias de fundo de olho . . . . .	83
4.1.1	D-eye . . . . .	84
4.1.2	Panoptic Welch Allyn . . . . .	85
4.1.3	Oftalmoscópio selecionado . . . . .	87
4.2	Locais de coleta das imagens . . . . .	89
4.2.1	Hospital de olhos HO . . . . .	89
4.2.2	Policlínica de Unai/MG . . . . .	93
4.3	Aquisição das imagens de fundo de olho . . . . .	93
4.3.1	Imagens glaucomatosas . . . . .	94

4.3.2	Imagens não glaucomatosas . . . . .	94
4.3.3	Dificuldades na aquisição das imagens . . . . .	95
4.3.4	Imagens com ruídos . . . . .	96
4.4	Formação da base de dados - BrG . . . . .	97
<b>5</b>	<b>Experimentos</b>	<b>101</b>
5.1	Plataforma de desenvolvimento . . . . .	103
5.2	Seleção de algoritmos individuais . . . . .	103
5.3	Configurações e treinamento dos classificadores . . . . .	105
5.3.1	Extração da região de interesse . . . . .	106
5.3.2	Divisão da base de dados . . . . .	107
5.3.3	Regularização dos algoritmos . . . . .	108
5.3.4	Ajustes de treinamento . . . . .	111
5.4	Testes e resultados individuais . . . . .	113
5.4.1	ResNet50v2 e ResNet101 . . . . .	113
5.4.2	Inception-v3 . . . . .	117
5.4.3	InceptionResNet . . . . .	119
5.4.4	MobileNet . . . . .	119
5.4.5	Xception . . . . .	120
5.4.6	DenseNet . . . . .	122
5.4.7	Resultados individuais . . . . .	125
5.5	Construção do ensemble . . . . .	127
5.6	Resultados do ensemble - corte manual . . . . .	130
5.7	Experimentos comparativos . . . . .	132
5.7.1	Algoritmo U-Net utilizado para segmentação . . . . .	134
5.7.1.1	Treinamento e testes da rede U-Net com imagens BrG . . . . .	135
5.7.1.2	Resultados do teste de segmentação . . . . .	137
5.7.2	Segmentação automática e específica do disco óptico . . . . .	138
5.7.3	Segmentação automática do disco óptico e adjacências . . . . .	139
5.7.4	Observações sobre as imagens segmentadas . . . . .	140
5.7.5	Segmentando imagens BrG com a rede U-Net . . . . .	142
5.8	Resultados definitivos da classificação . . . . .	142
<b>6</b>	<b>Análise de resultados e prática clínica</b>	<b>144</b>
6.1	Deep learning e triagem populacional do glaucoma . . . . .	144
6.2	Análise do algoritmo desenvolvido . . . . .	146
6.2.1	Cálculo sobre a população (A) . . . . .	148
6.2.2	Cálculo sobre a população (B) . . . . .	149

<b>7</b>	<b>Conclusões e trabalhos futuros</b>	<b>152</b>
7.1	Conclusões . . . . .	152
7.2	Recomendações para trabalhos futuros . . . . .	154
	<b>Referências Bibliográficas</b>	<b>155</b>

# Índice de Figuras

2.1	Exemplo de camadas de neurônios em uma rede totalmente conectada - adaptado de (Russell, 2010). . . . .	7
2.2	Exemplo modelo de um neurônio artificial - adaptado de (Russell, 2010). . . . .	8
2.3	Exemplo gráfico da função de ativação sigmoide. . . . .	10
2.4	Exemplo gráfico da função de ativação tanh. . . . .	11
2.5	Exemplo gráfico da função de ativação ReLU. . . . .	11
2.6	Exemplo gráfico das funções de ativação - sigmoide e <i>softmax</i> . . . . .	12
2.7	Exemplo de um modelo básico de algoritmo de <i>deep learning</i> . . . . .	13
2.8	Exemplos de dados, instâncias de dados e rótulos. . . . .	14
2.9	Coordenadas espaciais de uma imagem - adaptado de (E Woods and C Gonzalez, 2008). . . . .	15
2.10	Exemplo de vizinhança entre <i>pixels</i> - A) $N_4(px)$ e B) $N_8(px)$ . . . . .	16
2.11	Exemplo de uma imagem de fundo de olho com o sistema vascular segmentado - adaptado de (Staal et al., 2004). . . . .	19
2.12	Exemplo de uma imagem monocromática com matriz de <i>pixels</i> e um vetor de características. . . . .	20
2.13	Direções $\theta$ de aplicação GLCM. . . . .	21
2.14	Matriz de <i>pixels</i> seguida da matriz GLCM. . . . .	22
2.15	Exemplo de aplicação do algoritmo LBP. . . . .	23
2.16	Transforme dados brutos em novas coordenadas chamadas de componentes principais (PCs). . . . .	25
2.17	Exemplo de um modelo apropriado, seguido de um modelo com <i>underfitting</i> e <i>overfitting</i> . . . . .	28
2.18	Exemplo da arquitetura <i>random forest</i> com decisão conforme as decisões de cada uma das árvores. . . . .	29
2.19	Exemplos de hiperplanos (A, B, C) na separação de vetores de suporte (quadrante verde e triângulo vermelho). O hiperplano (C) representa a melhor classificação dos vetores. . . . .	30
2.20	Exemplo de meios de calcular probabilidades com uso do Naive Bayes. . . . .	31

2.21	Exemplo do algoritmo KNN apresentando a vizinhança referente a $k=3$ e $k=7$ . . . . .	32
2.22	Exemplo de aplicação de <i>kernel</i> por convolução. . . . .	35
2.23	Exemplo da aplicação <i>pooling</i> médio e máximo em uma matriz de <i>pixels</i> $4 \times 4$ (Russell, 2010). . . . .	36
2.24	Exemplo da aplicação <i>flattening</i> . . . . .	37
2.25	Exemplo de camadas de uma rede CNN. . . . .	38
2.26	Exemplo de aplicação da matriz de confusão. . . . .	41
2.27	Exemplo de curva roc. . . . .	42
2.28	Exemplo de aplicação da métrica IoU na segmentação da região do disco óptico. . . . .	43
3.1	Exemplo de glaucoma avançado com perda da visão periférica. . . . .	45
3.2	Exemplo do globo ocular apresentando a retina, o disco óptico e nervo óptico - adaptado de (AAO., 2023). . . . .	46
3.3	Exemplo das estruturas de fundo de olho - adaptado de (Budai et al., 2013). . . . .	47
3.4	Medidas consideradas no cálculo da CDR. . . . .	50
3.5	Exemplo de cálculo manual do valor de CDR com escavação de 0,6. . . . .	50
3.6	Exemplo dos diâmetros (VCD, VDD) e regra ISNT Inferior (I), Superior (S), Nasal (N) e Temporal (T). . . . .	51
3.7	Exemplo de assimetria de escavação. . . . .	52
3.8	Principais características glaucomatosas observadas nas imagens de fundo de olho - adaptado de (Remo Susanna Jr and Filho, 2019). . . . .	53
3.9	Gráfico de cobertura oftalmológica do Brasil e Estados brasileiros - adaptado de (Umbelino, 2021). . . . .	58
3.10	Relação médico por habitantes no Brasil - adaptado de (Scheffer et al., 2020). . . . .	59
3.11	Exemplo de imagens médicas oftalmológicas passíveis de aplicações com algoritmos de IA. . . . .	62
3.12	Exemplo de imagem com segmentação do disco e copo óptico. . . . .	68
4.1	Organograma das etapas de aquisição da base de dados BrG. . . . .	82
4.2	Imagens de aparelhos portáteis e úteis para realização de fotografias do fundo de olho com auxílio de smartphones . . . . .	83
4.3	Dispositivo <i>D-eye</i> e armação para acoplar-se em um smartphone. . . . .	84
4.4	Imagem do olho esquerdo tirada com smartphone e dispositivo <i>D-Eye</i> - Fonte: próprio autor. . . . .	85
4.5	Exemplo do oftalmoscópio <i>panoptic Welch Allyn</i> . . . . .	86
4.6	Exemplo de filtro de abertura de luz do <i>panoptic</i> . . . . .	87

4.7	Exemplo comparativo de imagens obtidas com o <i>D-eye</i> e <i>panoptic</i> - próprio autor. . . . .	88
4.8	Diferença entre uma imagem com ângulo de visão de 45° que registra toda a região da retina e uma imagem obtida com o <i>panoptic</i> com ângulo de visão de 25°. . . . .	89
4.9	Mapa do Estado de Minas Gerais com apresentação das cidades da região do Sul e a cidade de Unaí ao Noroeste do Estado. . . . .	90
4.10	Fluxograma do atendimento aos pacientes do programa de tratamento do glaucoma HO. . . . .	92
4.11	( A ) imagem com pequena abertura do olho, ( B ) disco óptico descentralizado, ( C ) presença de catarata, ( D ) ruído de foco de luz e pouca abertura do olho. . . . .	95
4.12	Principais tipos de ruídos presentes nas imagens do oftalmoscópio <i>panoptic</i> : ( A ) imagem ideal, ( B ) pouca iluminação, ( C ) ruído externo, ( D ) ruído de foco de luz. . . . .	96
4.13	Tamanho da imagem de 720 × 1280 <i>pixels</i> com corte central de 800 × 800 <i>pixels</i> na região que representa o disco óptico. . . . .	98
4.14	Exemplo de uma imagem original seguida por suas máscaras de segmentação padrão ouro para o glaucoma. . . . .	99
5.1	Etapas da metodologia aplicada neste trabalho. . . . .	102
5.2	Exemplo da arquitetura dos modelos individuais de classificação - adaptado de (Patil and Rane, 2021). . . . .	105
5.3	Extração manual da região de interesse - disco óptico. . . . .	106
5.4	Exemplo da divisão da base de dados BrG - adaptado de (Patil and Rane, 2021). . . . .	107
5.5	Exemplo da aplicação de <i>dropout</i> anulando alguns neurônios aleatoriamente. . . . .	108
5.6	Exemplo de aumento de dados com a biblioteca <i>ImageDataGenerator Keras</i> . . . . .	109
5.7	Exemplo <i>early stopping</i> com melhor ajuste na época de número 10. . . . .	110
5.8	Modelo base de pesos congelados, seguido de construção de camadas densas com aplicação de <i>dropout</i> . A saída foi obtida pela função de ativação <i>softmax</i> . . . . .	112
5.9	Exemplo de um bloco de construção residual ResNet (He et al., 2016a). . . . .	114
5.10	Curva de aprendizado, taxa de perda e curva ROC do modelo ResNet50v2 com 14 épocas de treinamento sobre a base de dados BrG. . . . .	115
5.11	Curva de aprendizado, taxa de perda e curva ROC do modelo ResNet101 com 12 épocas de treinamento sobre a base de dados BrG. . . . .	116
5.12	versão do módulo Inception v1 - adaptado de (Szegedy et al., 2016b). . . . .	117

5.13	Curva de aprendizado, taxa de perda e curva ROC do modelo Inception-v3 com 10 épocas de treinamento sobre a base de dados BrG. . . . .	118
5.14	Curva de aprendizado, taxa de perda e curva ROC do modelo Inception-ResNet com 17 épocas de treinamento sobre a base de dados BrG. . . . .	120
5.15	Exemplo de blocos convolucionais da segunda versão MobileNet (Sandler et al., 2018). . . . .	121
5.16	Curva de aprendizado, taxa de perda e curva ROC do modelo MobileNet com 25 épocas de treinamento sobre a base de dados BrG. . . . .	122
5.17	Curva de aprendizado, taxa de perda e curva ROC do modelo Xception com 20 épocas de treinamento sobre a base de dados BrG. . . . .	123
5.18	Curva de aprendizado, taxa de perda e curva ROC do modelo DenseNet com 25 épocas de treinamento sobre a base de dados BrG. . . . .	124
5.19	Deep DenseNet com três blocos densos. As camadas entre dois blocos adjacentes são chamadas de camadas de transição e alteram o tamanho do mapa de recursos por meio de convolução e pooling (Szegedy et al., 2016a).	125
5.20	Exemplo da utilização <i>k-fold</i> , adaptado de (Russell, 2010). . . . .	126
5.21	Matriz de confusão de cada um dos modelos de CNNs individuais. . . . .	127
5.22	Modelo de <i>ensemble</i> usando a melhor combinação dos classificadores individuais. . . . .	129
5.23	Exemplo da resposta final do classificador com base nas respostas individuais dos algoritmos selecionados para compor o <i>ensemble</i> . . . . .	131
5.24	Matriz de confusão com resultado do ensemble - corte manual. . . . .	131
5.25	Área sob a curva ROC para o modelo <i>ensemble</i> , cada linha ( corresponde a uma rodada de validação cruzada tripla. . . . .	132
5.26	Base de dados BrG - Imagem global à direita (A), ao centro exemplo de máscara da região do disco e adjacências (B), à esquerda exemplo de máscara específica do disco óptico (C). . . . .	133
5.27	Exemplos de imagens de câmeras com diferentes ângulos de visão - (drive, panoptic, d-eye). . . . .	133
5.28	Arquitetura de uma rede U-Net representando um bloco de entrada e mapa de segmentação como saída adaptado de (Ronneberger et al., 2015). . . . .	134
5.29	Exemplos de distorção aplicada a técnicas como aumento de dados. . . . .	136
5.30	Fluxograma para aplicação do algoritmo de segmentação <i>U-Net</i> . . . . .	137
5.31	Imagem global seguida pela máscara do disco óptico e região adjacente e a predição de segmentação na camada de saída. . . . .	137
5.32	Extração automática da região de interesse - disco óptico específico. . . . .	138
5.33	Extração automática da ROI - disco óptico e adjacências. . . . .	139
5.34	Imagem global, máscara do disco óptico, e região específica do disco óptico segmentada com toda estrutura preservada. . . . .	140

5.35	Imagem global, máscara do disco óptico, e região específica do disco óptico segmentada com destruição de parte do conteúdo. . . . .	140
5.36	Imagem global, máscara do disco óptico e adjacências, e região segmentada do disco óptico e adjacências com perfeita preservação de partes importantes do disco e <i>pixels</i> vizinhos. . . . .	141
5.37	Imagem global seguida da máscara manual e da região segmentada do disco óptico e adjacências com preservação do disco óptico e com danos aos <i>pixels</i> vizinhos. . . . .	141
5.38	Matriz de confusão das regiões: específica do disco óptico, disco óptico e adjacências, imagens globais. . . . .	143
6.1	Exemplo da distribuição de casos de glaucoma na base BrG e população real. . . . .	148
6.2	Matriz de confusão exemplo 1 - valores absolutos. . . . .	149
6.3	Matriz de confusão exemplo 2 - valores absolutos. . . . .	150

# Índice de Tabelas

3.1	Bases de dados e número de imagens glaucomatosas . . . . .	67
3.2	Trabalhos utilizando bases de dados públicas e privadas . . . . .	76
3.3	Trabalhos utilizando bases de dados públicas . . . . .	77
3.4	Trabalhos utilizando bases de dados particulares . . . . .	77
5.1	CNNs pré-treinadas com padrão de cores RGB utilizadas neste estudo . .	104
5.2	Resultados de classificadores individuais . . . . .	125
5.3	Combinações de CNNs avaliadas na construção <i>ensemble</i> . . . . .	130
5.4	Resultados obtidos pela combinação dos classificadores. . . . .	130
5.5	Precisão dos algoritmos usados para segmentar imagens BrG . . . . .	138
5.6	Resultados da classificação das imagens . . . . .	142

# Lista de Acrónimos

## Roman Symbols

ACC Acurácia

Adam Adaptive Moment Estimation

AHE Adaptive Histogram Equalization

AUC Area Under the Curve

BrG Brasil Glaucoma

CCV Color Coherence Vector

CDR Cup:Disc Ratio

CFNR Camada De Fibras Nervosas da Retina

cGAN Rede Adversarial Generativa Condicional

CLAHE Contrast Limited Adaptive Histogram Equalization

CNN Convolutional Neural Network

ES Especificidade

F1 F1-Score

FCN Fully Convolutional Networks

FFT Fast Fourier Transform

FN Falso-Negativo

FP Falso-Positivo

GANs Generative Adversarial Networks

GLCM Gray Level Co-occurrence Matrix

GPU	Graphics Processing Unit
HO	Hospital de Olhos
HOG	Histogram of Oriented Gradients
HSV	Hue Saturation Value
IA	Inteligência Artificial
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge
IoU	Interseção Sobre União
ISNT	Inferior, Superior, Nasal e Temporal
JPEG	Joint Photography Experts Group
KNN	K-Nearest Neighbors
LBP	Local Binary Pattern
MLP	Multilayer Perceptron
mmHg	Milímetros de Mercúrio
MSE	MeanSquareError
OCT	Optical Coherence Tomography
OMS	Organização Mundial da Saúde
P	Prevalência
PCA	Principal Component Analysis
PIO	Pressão Intraocular
PNG	Portable Network Graphic
PR	Precisão
ReLU	Rectified Linear Units
ResNet	Residual Network
RGB	Red, Green and Blue
RIGA	Retinal Fundus Images For Glaucoma Analysis

RNAs	Redes Neurais Artificiais
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SCES	Singapore Chinese Eye Study
SE	Sensibilidade
SUS	Sistema Único de Saúde
SVM	Support-Vector Machine
tanh	Tangente Hiperbólica
VCD	Diâmetro Vertical da Escavação
VDD	Diâmetro Vertical do Disco
VN	Verdadeiro-Negativo
VP	Verdadeiro Positivo
VPN	Valor Preditivo Negativo
VPP	Valor Preditivo Positivo

# Capítulo 1

## Introdução

O glaucoma é uma neuropatia de causa multifatorial que pode afetar o fundo de olho, causando perda gradual da visão e, em casos graves, cegueira. Tradicionalmente o diagnóstico do glaucoma é beneficiado com o auxílio de equipes clínicas prontamente disponíveis e equipamentos altamente especializados. A sensibilidade do diagnóstico é geralmente alta, pois, os exames aplicados nos consultórios de oftalmologia têm potencial clínico para auxiliar na identificação de praticamente todos os casos da doença. Porém, mesmo com o cenário sofisticado para diagnosticá-la, a sua evolução silenciosa e lenta, os valores dos exames e consultas, e a falta de acesso à rede pública de consultas oftalmológicas são problemas que fazem com que milhares de pessoas não procurem um oftalmologista nos estágios iniciais dessa neuropatia. Isso contribui para que cerca de 70% dos pacientes sejam autodiagnosticados, ou seja, alertados por sua própria deficiência visual e não por um diagnóstico precoce (Tan et al., 2020).

O glaucoma é considerado um problema global, e mesmo em países desenvolvidos, estima-se que pelo menos 50% dos pacientes glaucomatosos desconheça que têm a doença, sendo essa porcentagem ainda pior em países de baixa renda (Heijl et al., 2013; Quigley, 1996).

Embora os números estatísticos de subdiagnóstico na sociedade combinados com a necessidade de diagnóstico precoce para prevenir a cegueira possam sugerir que o glaucoma é um bom candidato para triagens populacionais, estudos mostraram que, pelo menos em países como o Reino Unido e Finlândia, a triagem do glaucoma de base populacional por métodos diagnósticos tradicionais não é viável devido ao alto custo de implantação / manutenção, e à prevalência relativamente baixa da doença na população em geral, sendo de aproximadamente 3,5% (Burr et al., 2014; Vaahtoranta-Lehtonen et al., 2007; Zaleska-Żmijewska et al., 2020). Da mesma forma, a força-tarefa de serviços preventivos dos Estados Unidos da América (EUA), com apoio da academia americana de médicos de família, não recomenda a triagem para o glaucoma no ambiente de atenção primária, citando evidências insuficientes para avaliar os seus benefícios ou malefícios (Mangione et al., 2022; Gedde et al., 2021).

---

Quanto aos custos de implantação e manutenção de uma possível triagem populacional do glaucoma, essa realidade vem mudando com o surgimento de novas tecnologias, como a telemedicina e o uso de Inteligência Artificial (IA) para auxiliar no diagnóstico tradicional da doença, considerada uma das maiores causas de cegueira irreversível do mundo (Tan et al., 2020; Health-Organization et al., 2019). Assim, diante de tais tecnologias, especialmente os algoritmos de *deep learning*, mostraram um potencial sem precedentes para auxiliar no diagnóstico dessa enfermidade, que pode comprometer a qualidade de vida dos pacientes e também impõe um enorme ônus econômico à sociedade (Girard and Schmetterer, 2020; Vaahtoranta-Lehtonen et al., 2007).

Além dos algoritmos de IA, a evolução tecnológica permitiu a criação de dispositivos de imagem de fundo de olho de pequeno porte, com baixo consumo de energia e acessíveis com o uso de smartphones. Esses dispositivos tais como o aparelho *panOptic* (Allyn, 2014), o *D-eye* (Russo et al., 2015) e *iNview* (Volk, 2022), juntamente com aplicativos para smartphones, são capazes de gerar imagens com processamento de dados integrado e permitir a comunicação dos resultados obtidos. Ademais, as imagens obtidas com esses aparelhos podem apresentar concordância substancial com os exames clínicos convencionais de fundo de olho.

Diante do contexto, o cenário mostra que os sistemas baseados em smartphones podem ser amplamente utilizados em diversas aplicações, inclusive em pesquisas de IA no campo da oftalmologia, o que se traduz naturalmente no campo do glaucoma a partir da identificação da doença em imagens digitais do fundo de olho. Assim, no contexto do glaucoma e dessas novas tecnologias de baixo custo capazes de adquirir o principal insumo para o diagnóstico dessa neuropatia óptica, bem como o avanço contínuo dos algoritmos de *deep learning* para reconhecimento de padrões em imagens digitais, as seguintes perguntas ainda precisam ser respondidas:

- Algoritmos de *deep learning* podem identificar efetivamente o glaucoma com base em imagens de fundo de olho, capturadas com smartphone?
- Uma combinação de tecnologias (*deep learning* e dispositivos portáteis de baixo custo para aquisição de imagens de fundo de olho) pode ser uma alternativa para viabilizar a triagem populacional do glaucoma?

## 1.1 Hipótese

Com base em trabalhos relacionados, foi investigada a hipótese de que o diagnóstico assistido por algoritmos de IA pode representar um excelente compromisso em: reduzir custos, alcançar uma precisão aceitável e viabilizar a triagem populacional do glaucoma, contribuindo para a redução do subdiagnóstico desta doença (Mirzania et al., 2021; Vilela et al., 2018; Hacısoftaoglu et al., 2020).

---

## 1.2 Objetivo geral

Embora em trabalhos relacionados os algoritmos de IA tenham alcançado resultados promissores para classificar o glaucoma em imagens fundoscópicas obtidas de dispositivos convencionais, resta saber se esses algoritmos são igualmente eficazes quando alimentados com imagens de baixa resolução obtidas de smartphones. Portanto, este trabalho apresenta um algoritmo *deep learning* treinado em imagens de fundo de olho obtidas de um dispositivo de baixo custo (smartphone conectado a um oftalmoscópio portátil). Além de ser uma forma de medir a capacidade de um algoritmo de IA em resolver o problema em questão, essa abordagem faz um levantamento estatístico da possibilidade de aplicações do algoritmo desenvolvido em programas de triagem populacional, como uma forma de esclarecer algumas dúvidas, lacunas ou dificuldades que possam ser encontradas pelos algoritmos de *deep learning* na prática clínica.

## 1.3 Objetivos específicos

Uma dificuldade encontrada ao usar algoritmos de IA para testar imagens de smartphones na detecção do glaucoma é a falta de conjuntos de dados disponíveis publicamente para esses estudos. Portanto, devido a essa deficiência, para atingir o objetivo geral, foram formuladas as seguintes etapas de trabalho, conhecidos como objetivos específicos:

1. Aquisição de fotografias de fundo de olho por meio de um smartphone acoplado em um oftalmoscópio portátil *panOptic*.
2. Seleção das fotografias adquiridas e formação de uma base de dados contendo imagens da retina, denominada como Brasil Glaucoma (BrG).
3. Construção de um *ensemble* de algoritmos de *deep learning*, mais especificamente, redes neurais convolucionais, para classificar as imagens BrG em classes "sim" ou "não" para glaucoma.
4. Identificar presumíveis fragilidades na aplicação prática do algoritmo desenvolvido no diagnóstico do glaucoma em possível triagem populacional.

## 1.4 Metodologia e resultados

Relativamente ao contexto metodológico, todo o trabalho de coleta de fotografias do fundo de olho foi elaborado em um cenário brasileiro no Estado de Minas Gerais e por um profissional não médico. As imagens fundoscópicas foram escolhidas por serem

---

adquiridas de forma não invasiva e amplamente utilizadas na maioria das clínicas oftalmológicas. Para a aquisição de tais imagens, acoplou-se um smartphone em um oftalmoscópio *panOptic* da *Weilch Allyn* (Allyn, 2014), escolhido por sua portabilidade, qualidade fotográfica, uso sem midríase, facilidade de uso e menor custo de aquisição em comparação com equipamentos fotográficos convencionais. Após a coleta de imagens fundoscópicas formou-se a base de dados BrG, contendo 2.000 fotografias de fundo de olho rotuladas para glaucoma.

Em seguida, a partir da concatenação de alguns algoritmos classificadores disponíveis pela biblioteca *Keras* (Heaton, 2020) e pré-treinados no conjunto de dados ImageNet (Russakovsky et al., 2015), formou-se um *ensemble*. O algoritmo desenvolvido apresentou acurácia máxima de 95,4% na classificação geral do glaucoma a partir de treinamento e testes nas imagens BrG.

Quanto à possibilidade de aplicação prática na triagem populacional da doença em evidência, foram levantadas estatísticas considerando como exemplo o cenário brasileiro, que apresenta extensões continentais e características de subdesenvolvimento. Deste modo, durante uma análise de resultados que se fez considerando os resultados alcançados com o algoritmo desenvolvido e a média da prevalência da doença em proporções mundiais, foi identificada que; embora ainda exista a necessidade de criar mecanismos que ajudem a reduzir ainda mais o número de resultados falso-positivos, a tecnologia desenvolvida tem capacidade substancial para o diagnóstico do glaucoma a partir do reconhecimento de padrões da doença em imagens de fundo de olho. Portanto, pode ser uma alternativa para auxiliar na identificação de novos casos de glaucoma, principalmente quando aplicada em grupos com maior risco de desenvolver a doença, e diante da situação de países como o Brasil, que apresenta um quadro desigual na distribuição de médicos em seu território e conseqüentemente apresenta desigualdade também na distribuição de serviços de saúde à população geral, com milhares de pessoas com dificuldade ao acesso à saúde, principalmente em algumas regiões mais afastadas dos grandes centros urbanos.

## 1.5 Estrutura da tese

A tese está estruturada em sete Capítulos apresentados a seguir, os quais buscam contemplar objetivos distintos de cada etapa da metodológica. Assim, o Capítulo 2 apresenta conceitos sobre as principais tecnologias utilizadas como base teórica para o desenvolvimento da tese. O Capítulo 3 apresenta o estado da arte aplicado em trabalhos relacionados. O Capítulo 4 descreve a construção da nova base de dados rotulada para o glaucoma e denominada BrG. O Capítulo 5 detalha a metodologia para desenvolver um modelo de classificação do glaucoma por meio de um *ensemble* de redes neurais convolucionais. Finalmente, o Capítulo 6 discute o estudo de forma geral com análise dos resultados alcançados e o Capítulo 7 apresenta a conclusão e perspectiva de trabalhos futuros.

# Capítulo 2

## Referencial teórico

Assim como em diversas outras áreas, a implementação de algoritmos de IA em oftalmologia tem sido alvo de grandes estudos, principalmente voltados para a automatização do processo de diagnóstico de doenças oculares como retinopatia diabética e glaucoma a partir de imagens digitais do fundo de olho. Diante do contexto, é importante entender quais são os tipos de tecnologias existentes para esta finalidade e como essas tecnologias são aplicadas, e assim, conhecer parcialmente os limites e avanços possíveis que podem acontecer com o uso de tais ferramentas. Diante disso, este Capítulo tem por objetivo apresentar o referencial teórico do estudo, a base conceitual e teórica da metodologia que sustenta algoritmos de IA, mais especificamente *machine learning*, algoritmos de redes neurais artificiais e *deep learning*, utilizados para reconhecimento de padrões em imagens digitais.

Deste modo, buscou-se no campo de *deep learning* as principais referências na definição desta tecnologia, dimensões e elementos, bem como as tipologias relacionadas aos algoritmos de redes neurais convolucionais, em inglês *Convolutional Neural Network* (CNN), considerados o padrão-ouro em tecnologias de visão computacional e reconhecimento de padrões em imagens digitais. Além disso, este Capítulo apresenta os conceitos sobre mecanismos de aprendizado por transferência e algoritmos de segmentação de imagens digitais, bem como uma descrição das principais métricas estatísticas utilizadas para a avaliação da precisão dos algoritmos de IA aplicados ao reconhecimento de padrões de imagens digitais.

### 2.1 Inteligência Artificial (IA)

A tecnologia IA data de 1956, foi desenvolvida por John McCarthy, professor de matemática do *Dartmouth College* no nordeste dos Estados Unidos. É uma tecnologia que abrange várias áreas do conhecimento e de modo geral refere-se ao desenvolvimento de sistemas computacionais capazes de realizar tarefas que imitam a inteligência humana por meio de

---

máquinas onde as suas derivações compõem hoje as diversas técnicas de *machine learning*, uma metodologia que ensina um sistema de computador a executar tarefas ou prever resultados sem que sejam explicitamente programados (McCarthy et al., 2006)

O termo *machine learning* engloba uma variedade de metodologias voltadas para processos de reconhecimento de padrões, regressão estatística e classificação de dados. Dentre os algoritmos de *machine learning*, destacam-se as técnicas de *deep learning*, que estão na vanguarda do desenvolvimento e dos avanços da computação e do big data nos últimos anos, principalmente a partir da introdução e desenvolvimento das redes CNNs (LeCun et al., 1998). Deste modo, as redes CNNs percorreram um longo caminho e ganharam cada vez mais credibilidade no desenvolvimento de algoritmos *machine learning*, especialmente nas áreas de visão computacional ou reconhecimento de padrões e classificação de imagens digitais (Ran et al., 2020).

Para melhor entendimento dos termos *machine learning* e *deep learning* e também a estrutura de tais algoritmos com a capacidade de aprender, primeiramente é importante conhecer sobre as redes multicamadas perceptron, em inglês *multilayer perceptron* (MLP). Assim, redes MLPs são algoritmos de redes neurais artificiais (RNAs) que evoluíram de seu antecessor perceptron simples. Com diversos campos de estudos e aplicações como classificação de padrões, controle e processamento de sinais, as redes MLPs consistem em combinações de múltiplas unidades de processamento chamadas de neurônios artificiais (inspirados no cérebro humano, são unidades de processamento de informações que podem operar de forma linear ou não linear).

Neurônios artificiais são agrupados e executam funções específicas. As suas características essenciais são a adaptação e representação do conhecimento baseado em conexões chamadas sinapses. Normalmente, as redes RNAs (independente de sua arquitetura) consistem em organizar esses neurônios em camadas. Assim, a estrutura básica de uma rede RNA (como uma rede MLP), consiste em uma camada de entrada e uma camada de saída, além de uma ou mais camadas intermediárias chamadas de camadas ocultas, como mostra a Figura 2.1 (Haykin, 2009).

As camadas ocultas de neurônios artificiais permitem que a rede aprenda tarefas complexas e apresente alto grau de conectividade, determinado pelas sinapses da rede, possibilitando a extração de estatísticas de alta ordem (Russell, 2010). As camadas são estruturas fundamentais de dados nas RNAs, e consistem em processamentos de dados que recebem um ou mais vetores de entrada e retornam um ou mais vetores de dados transformados para as camadas adjacentes. Assim, as camadas possuem pesos, que especificam as transformações a serem feitas nos dados, de modo a criar o modelo preditivo.

O comportamento inteligente de uma rede neural artificial advém das interações entre suas unidades de processamento, ou seja, entre os neurônios e suas camadas. Essas interações podem ser estruturadas de diversas formas, determinando diferentes tipos de redes, como a criação de redes *perceptron*, ou arquiteturas avançadas de *deep learning*,

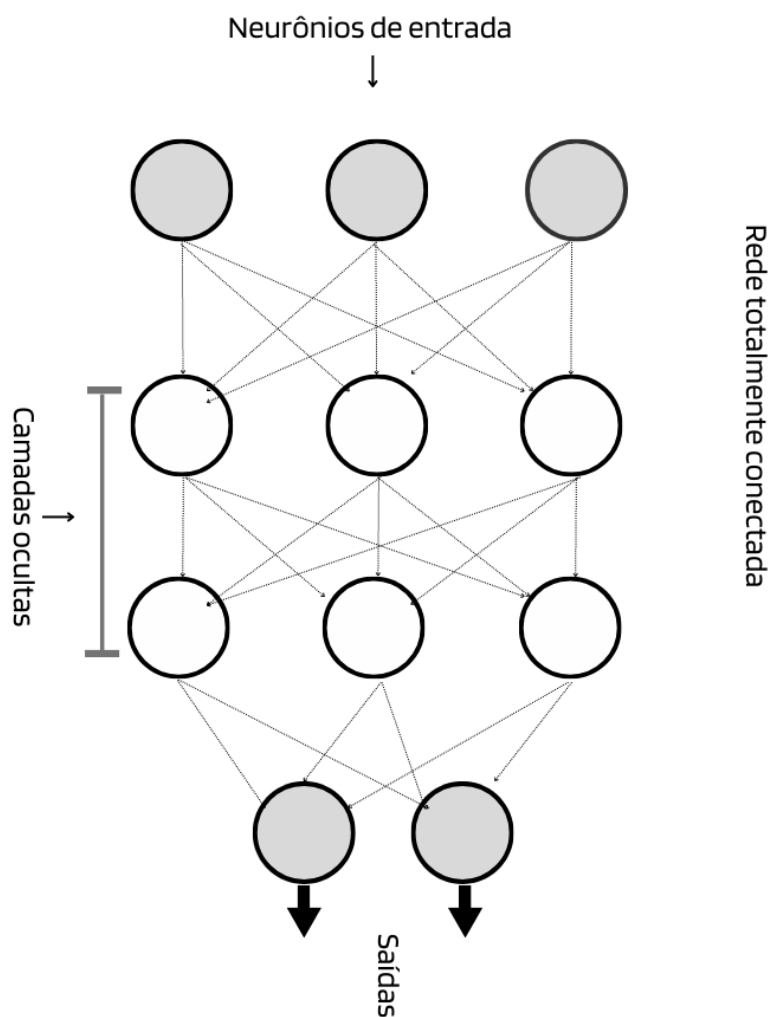


Figura 2.1: Exemplo de camadas de neurônios em uma rede totalmente conectada - adaptado de (Russell, 2010).

como as redes CNN.

Redes neurais podem ser do tipo propagação direta *feed-forward*, ou seja, a informação passa diretamente da entrada, passando por todas as camadas de neurônios existentes e segue para saída (Haykin, 2004). Assim, diz-se que uma rede está totalmente conectada quando cada nó numa camada está conectado a todos os nós na próxima camada adjacente, conforme se apresenta a Figura 2.1. Em contrapartida, quando alguns elos de comunicação são retirados, diz-se que a rede é parcialmente conectada (Russell, 2010).

Para melhor exemplificar o funcionamento de um neurônio artificial e as suas atribuições às RNAs, considere a Figura 2.2, que representa a arquitetura de um neurônio artificial, deste modo, é possível descrevê-lo da forma seguinte:

1. Primeiro, tem-se um conjunto de entrada denotado por  $x_1, \dots, x_n$ .

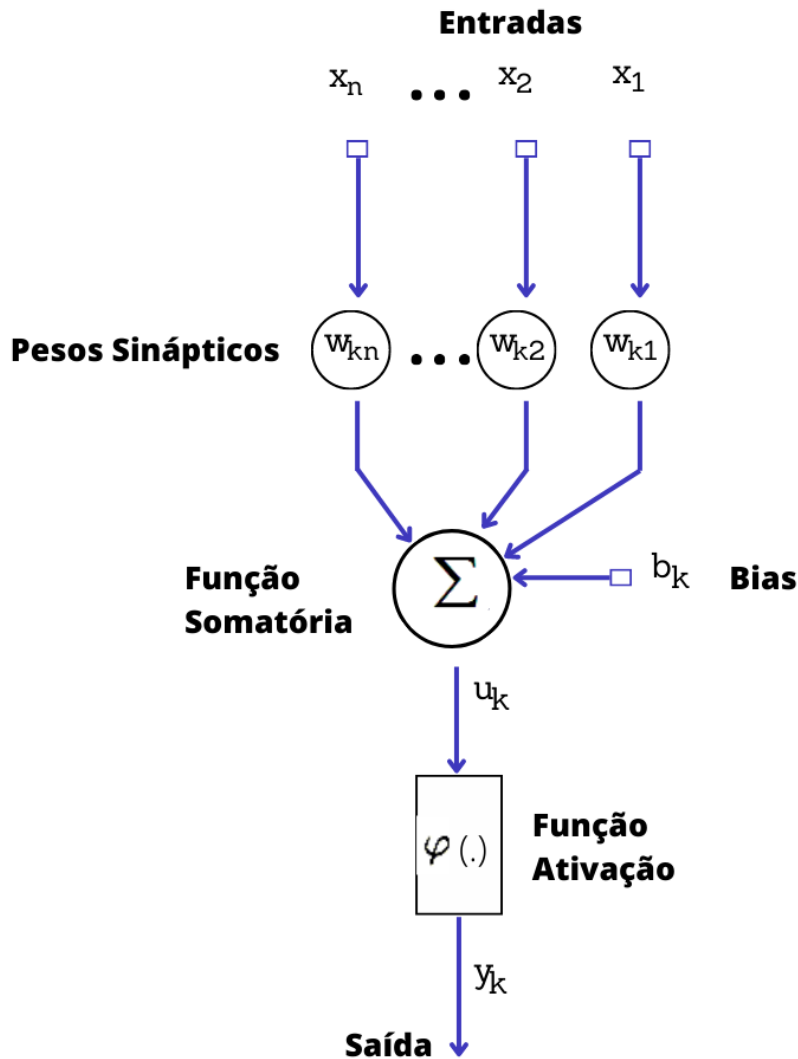


Figura 2.2: Exemplo modelo de um neurônio artificial - adaptado de (Russell, 2010).

2. Os pesos *sinápticos* são representados por  $w_1, \dots, w_n$ .
3. A função somatória realiza uma soma de todas as entradas multiplicadas por seus respectivos pesos.
4. O valor de *bias* denotado por  $b_k$  é ajustado da mesma forma que os pesos sinápticos, permitindo que um neurônio tenha saídas diferentes de nulo mesmo se todas as suas entradas forem nulas.
5. O valor de  $u_k$  corresponde à saída do combinador linear induzido pelo sinal de entrada.
6. A função de ativação denotada por  $\sigma(v)$  é usada para limitar a amplitude de saída do neurônio, ou seja, uma forma de restringir a faixa permitida da amplitude do sinal de saída a um valor limitado. É a diferença entre o valor produzido pelo combinador

---

linear e o limite de ativação. Normalmente, a faixa normalizada de amplitudes de saída do neurônio é escrita como a faixa de unidade fechada  $[0, 1]$  ou  $[-1, 1]$ . Desta forma, são consideradas transformações não lineares aquelas executadas ao longo do sinal de entrada.

7. A saída estimada pelo neurônio é representada por  $y_k$ , (Russell, 2010).

O seguinte par de Equações 2.1 e 2.2 pode ser entendido como um termo matemático para descrever um neurônio  $k$ :

$$u_k = \sum_{i=1}^n w_{ki}x_i \quad (2.1)$$

e

$$y_k = \sigma(u_k + b_k) \quad (2.2)$$

Como o *bias* é um aditivo externo ao neurônio, pode-se refazer a cálculo considerando as seguintes Equações 2.3 e 2.4:

$$v_k = u_k + b_k \quad (2.3)$$

e

$$y_k = \sigma(v_k) \quad (2.4)$$

## 2.1.1 Funções de ativação

As funções de ativação definidas por  $\sigma(v)$  nas equações seguintes, define matematicamente a saída de um neurônio em termos do campo local induzido  $v$ . Estas funções determinam o nível de ativação do neurônio artificial, limitando a amplitude do sinal de saída para uma faixa de valores finitos, ou seja, processa o sinal gerado pela combinação linear das entradas e dos pesos das sinapses, para gerar o sinal de saída do neurônio (Nwankpa et al., 2018; Russell, 2010).

As funções de ativação mais comumente usadas são: função logística ou sigmoide, tangente hiperbólica ( $\tanh$ ), função linear retificada, em inglês *Rectified Linear Units* (ReLU), e *softmax*, a saber:

### 2.1.1.1 Função de ativação sigmoide

A função sigmoide, exemplificada na Figura 2.3 e representada pela Equação 2.5, é uma função de ativação não linear que funciona bem em tarefas de classificação binária e regressão logística. Para entendimento da equação, a variável  $a$  determina a possibilidade

---

de diferentes inclinações da sigmoide. Essa função tende a 0 para valores muito pequenos e tende à 1 para valores muito grandes (Nwankpa et al., 2018).

$$\sigma(v) = \frac{1}{1 + \exp(-av)} \quad (2.5)$$

A função sigmoide tem o problema de dissipação de gradiente que é o meio de encontrar de forma iterativa, os valores dos parâmetros que minimizam determinada função de interesse, isso ocorre porque o gradiente propagado durante o processo de treinamento da rede é multiplicado por um valor menor que 1 em cada iteração e camada percorrida, assim, a rede encontra problemas para realizar uma técnica chamada de *backpropagation*, aplicada do gradiente da camada de saída para as primeiras camadas ocultas anteriores, pois tais replicações acabam chegando nas camadas iniciais da rede com valores ínfimos, necessitando de quantidade elevada de iterações para um ajuste minimamente relevante.

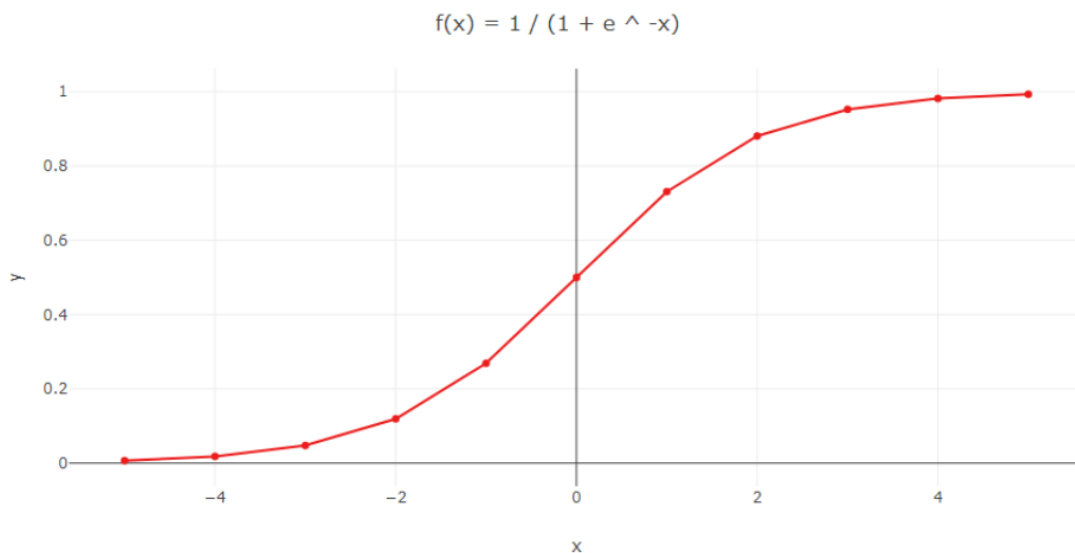


Figura 2.3: Exemplo gráfico da função de ativação sigmoide.

### 2.1.1.2 Função de ativação (tanh)

Esta é uma função que pode ser expressa na Equação 2.6 e representada pela Figura 2.4. Retém forma similar à função sigmoide, mas com a diferença que seus valores vão de -1 a 1 (Nwankpa et al., 2018). Ainda assim, essa função também sofre com dissipação de gradiente.

$$\sigma(v) = \tanh(v) \quad (2.6)$$

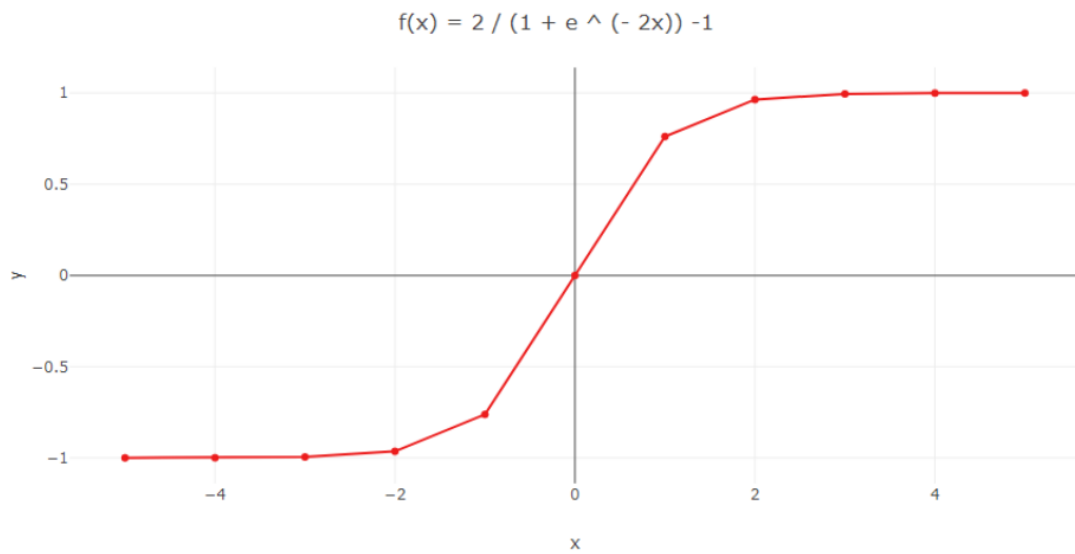


Figura 2.4: Exemplo gráfico da função de ativação tanh.

### 2.1.1.3 Função de ativação ReLU

A função de ativação ReLU, é a mais utilizada para facilitar o processo de treinamento dos algoritmos de *machine learning* (Patil and Rane, 2021). Essa função representada pela Equação 2.7, retorna 0 para todos os valores negativos, e o próprio valor para valores positivos, como mostra a Figura 2.5. Essa abordagem resulta em tempos menores na fase de treinamento para redes CNNs, comparado ao uso e aplicação de outros tipos de funções de ativação (Krizhevsky et al., 2017b).

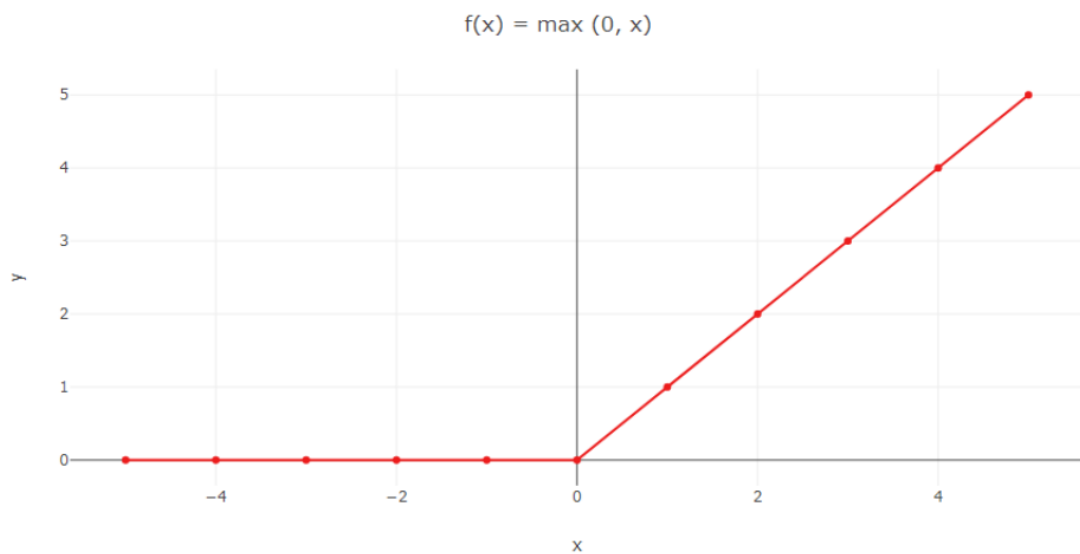


Figura 2.5: Exemplo gráfico da função de ativação ReLU.

$$\sigma(x) = \max(0, x) \quad (2.7)$$

### 2.1.1.4 Função de ativação softmax

Outra função de ativação amplamente utilizada é a função *softmax*, representada pela Equação 2.8 e Figura 2.6.

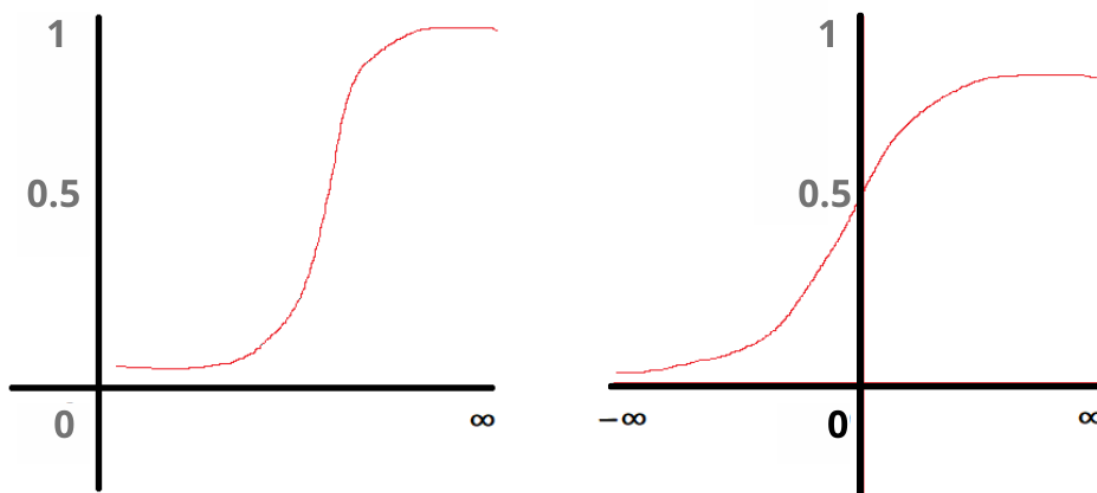


Figura 2.6: Exemplo gráfico das funções de ativação - sigmoide e *softmax*.

Esta função força a saída da rede neural a representar a probabilidade de cada classe para cada valor de entrada. Os valores na função *softmax* podem ser representados entre 0 e 1, onde a soma das probabilidades de todas as classes é igual a 1 (Agarap, 2018). Softmax é uma generalização da função sigmoide para casos não-binários e binários. Ela não costuma ser aplicada às camadas escondidas da rede neural, mas sim na camada de saída de problemas de classificação multiclasse e binários.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}, j = 1 \dots K \quad (2.8)$$

## 2.2 Organização do processo de treino e classificação numa RNA

Quanto aos algoritmos supervisionados e disponíveis para classificação de imagens digitais, embora existam diversos deles e, ainda que ofereçam tecnologias diferentes, de forma abrangente e geral, esses diversos classificadores são organizados e comumente treinados com as aquisições e procedimentos apresentados no organograma da Figura 2.7:

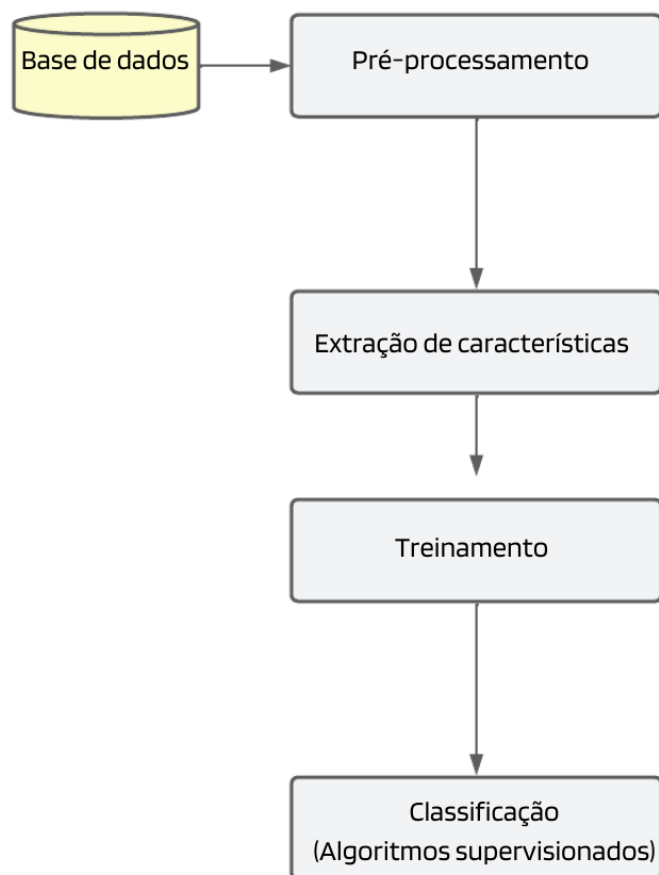


Figura 2.7: Exemplo de um modelo básico de algoritmo de *deep learning*.

### 2.2.1 Base de dados

Uma base de dados que também pode ser conhecida como conjunto de dados ou banco de dados pode ser entendido como conjunto de elementos inter-relacionados, organizados para permitir a recuperação da informação. Considerando como exemplo a Figura 2.8, tem-se uma base de dados contendo imagens digitais de cães e gatos, onde cada imagem na base de dados é chamada de instância de dados, e para cada instância de dados existe um rótulo que a classifica em uma categoria.

Algoritmos de RNAs, em sua grande maioria requerem grandes conjuntos de dados para serem treinados. O interessante para isso é que as bases de dados sejam representativas da população em estudo, para que assim, o algoritmo utilizado possa receber como entrada dados brutos e a partir do conhecimento adquirido durante o treinamento, tomar uma ação baseada na categoria do padrão identificado (Duda et al., 2006).

A deficiência quantitativa de dados, principalmente clínicos, pode estar relacionada às necessidades de rotulagem, ou seja, classificá-los em categorias definidas que possam ser utilizados por algoritmos de treinamento supervisionado, pois, os rótulos de dados

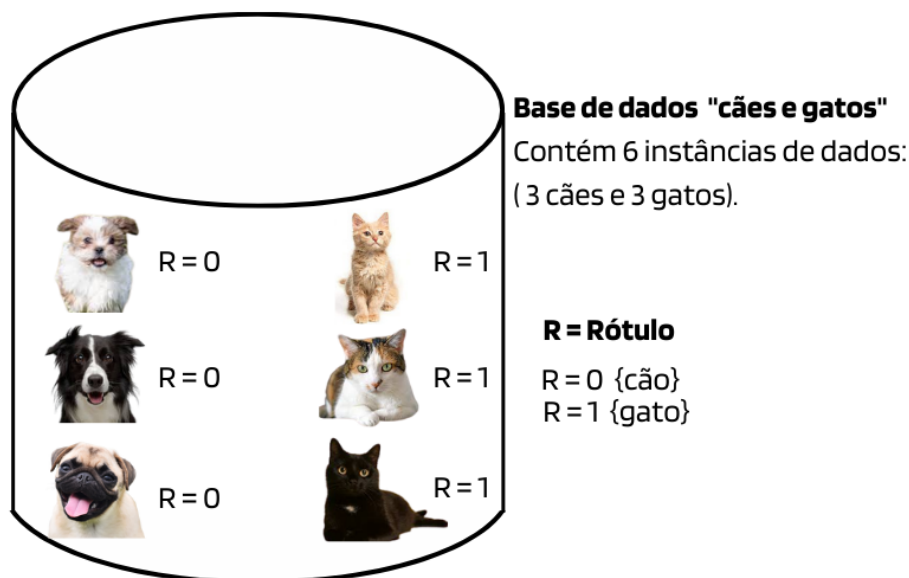


Figura 2.8: Exemplos de dados, instâncias de dados e rótulos.

servem como um guia para que o algoritmo possa entender, processar e aprender de forma abrangente com os padrões obtidos da entrada. A rotulagem de dados é primordial para o crescimento e evolução dos algoritmos de *deep learning*, mas, vem com vários desafios, como, falta de *hardware* necessário para lidar com uma força de trabalho massiva e volume excessivo de elementos, acesso restrito a ferramentas de etiquetagem ou rotulagem. Devido a essas dificuldades, muitas bases de dados existentes foram rotuladas pelo método conhecido como *crowdsourcing*, definido como um processo de aquisição de informações ou resultados necessários por solicitação de assistência de um grupo de muitas pessoas disponíveis por meio de comunidades online. Esta é uma forma de dividir e distribuir um grande projeto entre várias pessoas (Zhang et al., 2016).

Estima-se que mais de 80% das tarefas de engenharia em projetos de *machine learning* sejam dedicadas à preparação e rotulagem de dados e, apesar das técnicas de rotulagem como *crowdsourcing*, ainda não fornecem rótulos precisos e confiáveis para todos os casos de uso existentes. Portanto, rótulos de qualidade são altamente desejáveis, pois este é um pré-requisito para a execução de algoritmos supervisionados de *machine learning*, uma vez que o desempenho de um modelo em operação é diretamente afetado pela qualidade dos dados de treinamento.

## 2.2.2 Pré-processamento de imagens

As etapas de pré-processamento utilizadas em RNAs, geralmente preparam as imagens para serem usadas por técnicas como extração de características, segmentação de objetos ou de classificação. Algumas das técnicas de pré-processamento funcionam com imagens

coloridas, enquanto outras trabalham com imagens em tons de cinza ou binárias.

Para melhor compreender uma etapa de pré-processamento, é preciso entender como uma imagem é vista por um algoritmo computacional. Portanto, conforme mostrado na Figura 2.9, uma imagem digital em tons de cinza é definida como uma matriz numérica ou uma função bidimensional  $(x,y)$ , onde  $x$  e  $y$  são coordenadas espaciais cujo valor  $f$  para qualquer par de coordenadas  $(x,y)$  representa a intensidade ou nível de cinza da imagem (E Woods and C Gonzalez, 2008). Deste modo, em uma imagem monocromática, os *pixels* têm tons de cinza que variam em intensidade de 0 a 255, onde 0 é branco e 255 é preto.

Para melhor entendimento, uma imagem pode ter  $2^l$  níveis de cinza, sendo comum referir-se a ela como uma imagem  $l$  bits. Por exemplo, uma imagem com 256 valores discretos possíveis de níveis de cinza é chamada de uma imagem de 8 *bits*. Portanto, cada *pixel* pode estar associado a um valor da escala de cinza entre 0 e  $(2^l) - 1$ . Quanto maior o número de *pixels* numa imagem, melhor será sua resolução, permitindo uma melhor diferenciação entre as estruturas.

Imagens coloridas apresenta três vezes o tamanho da sua similar monocromática em *bits*. Deste modo, uma imagem de função  $f(x,y)$  será o conjunto de vetores tridimensionais associado ao espaço de cores. Assim, a quantidade de *bits* necessária para sua representação digital é dada por  $(3x \times y \times 2^l)$ . Diante do contexto, o processamento de uma imagem digital colorida torna-se diferenciado do processamento de imagens em tons de cinza, ou monocromático.

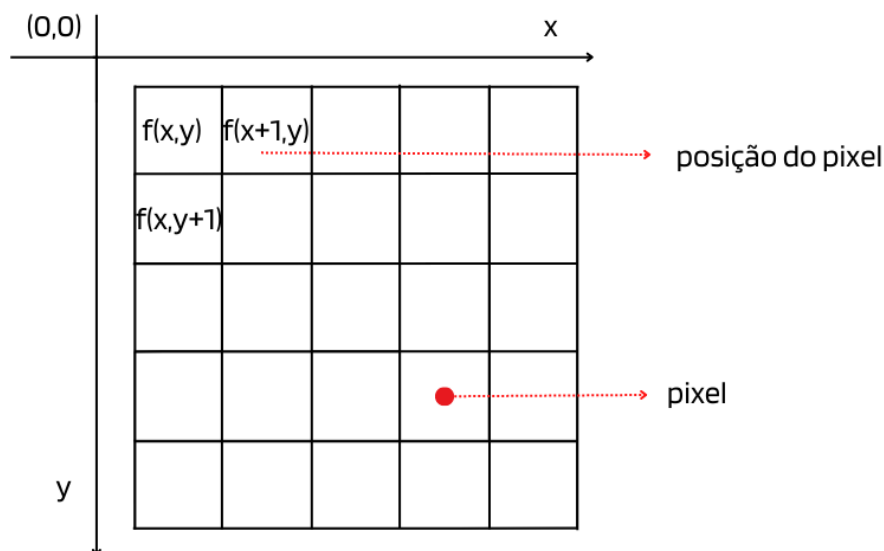


Figura 2.9: Coordenadas espaciais de uma imagem - adaptado de (E Woods and C Gonzalez, 2008).

As matrizes numéricas como apresentadas na Figura 2.9 são uma forma fundamental

de realizar cálculos matemáticos para o processamento de imagens ou para fazer medições quantitativas entre os *pixels*. Assim, para serem realizados os cálculos sobre as coordenadas das imagens, os algoritmos levam em consideração a localização dos *pixels*, conforme mostra a Figura 2.10. Desta forma, um *pixel*  $px$  na coordenada  $(x;y)$  pode ter quatro vizinhos distribuídos por localizações horizontais e verticais cujas coordenadas são:  $(x+1;y)$ ;  $(x-1;y)$ ;  $(x;y+1)$ ;  $(x;y-1)$ .

Esses vizinhos podem ser chamados de vizinhos 4 de  $px$ , expresso por  $N_4(px)$ , já outros quatro vizinhos diagonais de  $px$  têm coordenadas:  $(x+1;y+1)$ ;  $(x+1;y-1)$ ;  $(x-1;y+1)$ ;  $(x-1;y-1)$ , são os chamados vizinhos 8 de  $px$ , expresso por  $N_8(px)$ .

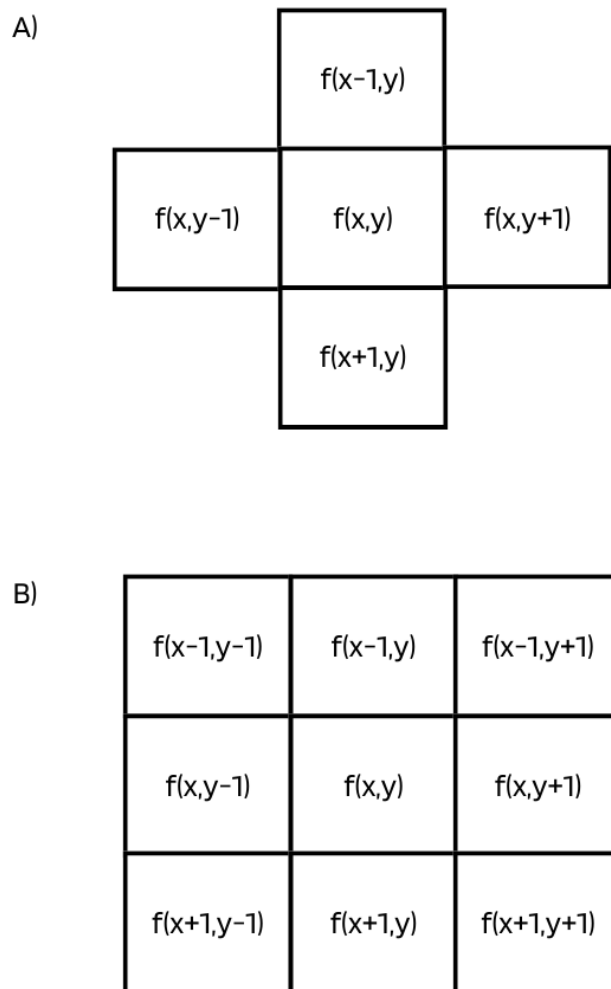


Figura 2.10: Exemplo de vizinhança entre *pixels* - A)  $N_4(px)$  e B)  $N_8(px)$ .

Existem muitos tipos de algoritmos destinados ao pré-processamento de imagens digitais. Esta vasta quantidade de algoritmos podem ser divididos em dois tipos, sendo, aqueles que operam o processamento no domínio espacial e aqueles que operam o processamento no domínio da frequência:

- **Domínio espacial:** Inclui operações aplicadas diretamente a matrizes de *pixels* de uma imagem. Algumas técnicas que se enquadram nessa categoria incluem: equalização de histograma, subtração de imagem, detecção de borda, alongamento de contraste, nitidez, suavização, etc. Inclui aplicação de filtros, tais como: (gaussiano, laplaciano, média, etc.)

Um exemplo de filtragem espacial de uma imagem de dimensão  $H \times L$  usando um filtro de dimensão  $h, l$ , pode ser representado pela Equação 2.9, onde  $x$  e  $y$  são aplicados de forma que cada *pixel* em  $z$  percorra todos os *pixels* em  $f$ :

$$g(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b z(s, t) f(x + s, y + t) \quad (2.9)$$

- **Domínio da frequência:** As transformações no domínio da frequência são outro tipo de processamento de imagem particularmente útil para reconhecimento de padrões, remover texturas muito finas e identificar estruturas periódicas repetidas. A transformação mais popular é a transformada de *Fourier*, que usa o algoritmo *Fast Fourier Transform* (FFT). Esses métodos permitem o isolamento de informações importantes no domínio das transformadas, principalmente aquelas que não podem ser separadas no domínio espacial. Para conhecimento, outras transformadas como (Hough, DCT, Walsh e Wavelet) também podem ser utilizadas com o mesmo propósito (Grande, 2012; E Woods and C Gonzalez, 2008).

Um exemplo de transformações lineares e bidimensionais podem ser representados pela Equação 2.10:

$$T(u, v) = \sum_{x=0}^{H-1} \sum_{y=0}^{L-1} f(x, y) r(x, y, u, v) \quad (2.10)$$

Neste caso,  $f(x, y)$  representa a imagem de entrada, onde  $x$  e  $y$  representam variáveis em coordenadas espaciais, e  $H$  e  $L$ , são as linhas e colunas de  $f$ . O *kernel* de entrada é denotado por  $r(x, y, u, v)$  e toda a equação é calculada para  $u = 0, 1, 2, \dots, H - 1$  e  $v = 0, 1, 2, \dots, L - 1$ . O resultado apresentado em  $T(u, v)$  é chamado de transformação direta de  $f(x, y)$ .

Para a transformação inversa, dado o valor de  $T(u, v)$ , pode-se aplicar a Equação 2.11, para  $x = 0, 1, 2, \dots, H - 1$  e  $y = 0, 1, 2, \dots, L - 1$ . Nesse caso,  $s(x, y, u, v)$  é chamado de *kernel* da transformação inversa.

$$f(x, y) = \sum_{u=0}^{H-1} \sum_{v=0}^{L-1} T(u, v) s(x, y, u, v) \quad (2.11)$$

---

Para imagens coloridas, é necessário considerar que diferentes espaços possuem diferentes expressões de cores, como *Red, Green and Blue (RGB)*, *Hue Saturation Value (HSV)*, etc. No entanto, independentemente desses espaços diferentes, as modalidades de extração de características, como descritores de histograma de cores, são frequentemente usados.

Um histograma de cores representa uma possível perspectiva de análise de imagem. Esta é uma forma de contar *pixels* semelhantes e combiná-los para descrevê-los em cada intervalo de cores ou compartimento de forma independente. Esta metodologia pode ser dividida em: histograma global de cores e histograma local de cores (Narwade and Kumar, 2016). A diferença entre as duas formas de aplicar o descritor é que, o global calcula o histograma de cores de toda a imagem e usa essa tabela de frequência como uma representação de baixa dimensão da imagem; O histograma local primeiro divide a imagem em blocos antes de aplicar a tabela de frequência.

Outra prática comumente utilizada é técnica de vetores de coerência de cores, em inglês, *Color Coherence Vector (CCV)*. Estes vetores são uma abordagem mais sofisticada do que os histogramas de cores. Funciona classificando cada *pixel* como coerente ou irrelevante. Um *pixel* coerente significa que faz parte de um grande componente conectado e um *pixel* incoerente significa que faz parte de um pequeno componente conectado. Uma etapa fundamental para fazer esse método funcionar é definir os critérios pelos quais os componentes conectados podem ou não ser grandes.

Este método, juntamente com a análise de histograma, visa construir uma representação de baixa dimensão da imagem. O recurso obtido com esta técnica é a possibilidade de criar vetores de características onde para cada cor pode ser calculado o número de *pixels* coerentes e o número de *pixels* incoerentes.

Por fim, o pré-processamento, seja no domínio espacial ou da frequência, seja uma imagem colorida ou não, trata-se de melhorar a qualidade da imagem para que as etapas subsequentes sejam bem-sucedidas. Assim, é possível nessa fase corrigir alguns tipos de ruídos, aplicar realce de bordas e realce de partes importantes das imagens para tarefas posteriores como segmentação de objetos ou classificação.

### **2.2.3 Extração da região de interesse**

A extração da região de interesse, em inglês, *Region of Interest (ROI)*, é um conceito que se refere a uma área de interesse em uma cena. Geralmente se refere a uma região importante e relevante na imagem. Por exemplo, regiões de interesse em imagens médicas são quase sempre áreas focais, como tumores, calcificações, hemorragias, etc. Considerando imagens fundoscópicas, a Figura 2.11 apresenta uma imagem do fundo de olho com o sistema vascular segmentado.

A tarefa de extrair regiões de interesse pode ser uma etapa importante em algorit-



Figura 2.11: Exemplo de uma imagem de fundo de olho com o sistema vascular segmentado - adaptado de (Staal et al., 2004).

mos de classificação de imagens digitais, pois permite, quando possível, treinar um classificador usando apenas as regiões mais relevantes da imagem, o que pode melhorar a eficiência geral do resultado pretendido. Este processo pode ser aplicado de forma manual, semiautomática ou totalmente automática. Nos casos semiautomáticos, o algoritmo utilizado pode auxiliar o usuário a estabelecer qual é a área desejável. O usuário especifica apenas uma parte da delimitação, que em seguida é continuada e completada por um algoritmo de avaliação e aquisição. Com este tipo de detecção, o usuário pode estabelecer aproximadamente uma área dentro da qual o algoritmo deverá operar. No caso do glaucoma, a ROI é a área próxima ao disco óptico, local onde geralmente a neuropatia afeta em maiores proporções.

No caso de extração automática da ROI, é necessário um algoritmo de segmentação. Esta metodologia se refere ao processo de decomposição de uma imagem digital nos múltiplos segmentos (regiões) que a formam. O resultado desejado do processo de segmentação é um número finito de regiões homogêneas que personalizam as diferentes regiões contidas na imagem. Portanto, uma imagem será formada por um conjunto de regiões conectadas por arestas e não se sobrepõem, de forma que cada *pixel* da imagem pertença apenas a uma determinada região (Acharya and Ray, 2005; Zaitoun and Aqel, 2015).

Técnicas de segmentação de imagens digitais podem ser aplicadas de diferentes formas, tais como: detecção de descontinuidade, detecção de ponto e linha, detecção de similaridade, crescimento de região, limiarização, agrupamento (*clustering*), segmentação semântica, etc. De forma geral, a segmentação de uma imagem tende a ser uma tarefa subjetiva e computacionalmente uma difícil tarefa para muitos tipos de problemas.

## 2.2.4 Extração de características

A extração de características se refere à tarefa de descrever as informações visuais de uma imagem em um vetor de características. Esse vetor é utilizado como entrada para os algoritmos de classificação de padrões. Por esse motivo, as características extraídas devem representar as diferenças visuais entre imagens. A Figura 2.12 apresenta um exemplo de vetor de características sobre uma imagem monocromática.

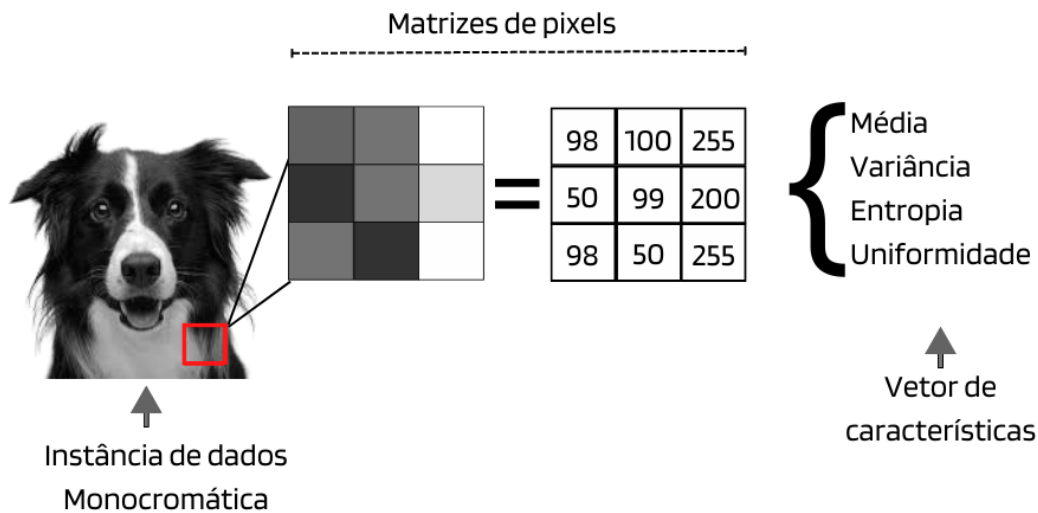


Figura 2.12: Exemplo de uma imagem monocromática com matriz de *pixels* e um vetor de características.

Os vetores de características podem ser formados a partir de informações geométricas, estatísticas, de textura, etc. Deve-se extrair o máximo de informações relevantes para cada tipo de imagem, tais como: contraste, entropia, suavidade, uniformidade, média, desvio padrão, energia, inclinação, perímetro, índice de irregularidade, diâmetro equivalente, área convexa e dureza, uniformidade, variância, etc (Mutlag et al., 2020).

A extração de características desempenha um papel decisivo para a tarefa de classificação, pois, é nesta etapa que se extrai as informações que irão de fato definir padrões com capacidade de distinguir uma classe de outra. Existem muitos algoritmos adequados para tais aplicações, então aqui estão alguns extratores de recursos amplamente utilizados em pesquisas científicas, como: Gray Level Co-occurrence Matrix (GLCM), *Local Binary Pattern* (LBP), Histogram of Oriented Gradients (HOG) e Contrast Limited Adaptive Histogram Equalization (CLAHE).

- **Gray Level Co-occurrence Matrix:** O modelo de algoritmo GLCM é um método estatístico para examinar texturas de imagens considerando a relação espacial de *pixels*. A função deste método caracteriza a textura de uma imagem contando a

frequência de ocorrência de pares de *pixels* com um valor específico e uma relação espacial especificada na imagem, criando uma matriz denominada GLCM. Em seguida, essa matriz, também conhecida como uma matriz de dependências espaciais em tons de cinza, servirá como ponto de partida para aplicações de extração de medidas estatísticas.

Deste modo, para extrair a matriz de dependência GLCM a partir de uma imagem digital, primeiramente, o algoritmo varre a matriz que representa a imagem e, para cada combinação dos níveis  $i$  e  $j$ , determina o número de coocorrências existentes entre eles, levando em consideração uma distância  $d$  e uma direção  $\theta$ , como apresentado na Figura 2.13.

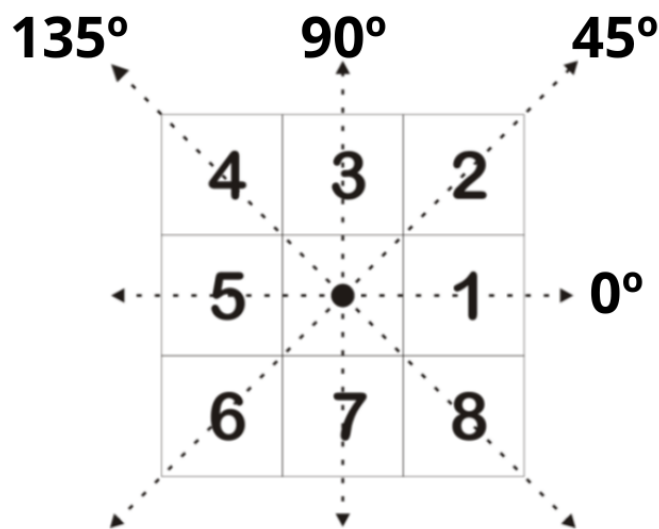


Figura 2.13: Direções  $\theta$  de aplicação GLCM.

Para melhor entendimento, o exemplo apresentado na Figura 2.14 (A), mostra uma matriz normalizada de *pixels* com tamanho de (4 x 6), seguida da matriz GLCM. Os *pixels* são representados contendo apenas quatro níveis de cinza como mostrado nos valores das células. Desta forma, a GLCM pode ser representada pela Figura 2.14 (B), com uma matriz de tamanho referente à quantidade de níveis de cinza da matriz de *pixels*.

A Figura 2.14 (B) representa a contagem do número de vezes que os pares de *pixels* coocorrem na matriz GLCM representada. Por exemplo, o valor sombreado em azul na matriz (A) representa um par de *pixels* com valor 33 seguido de outro *pixel* com valor 33, considerando a direção  $\theta = 0^\circ$  e a distância  $d = 1$ . As contagens de coocorrência são transferidas para o GLCM à direita. Neste exemplo: um *pixel* com valor 33 seguido por um *pixel* com valor 33 coocorre 2 vezes, conforme mostrado na geração da GLCM).

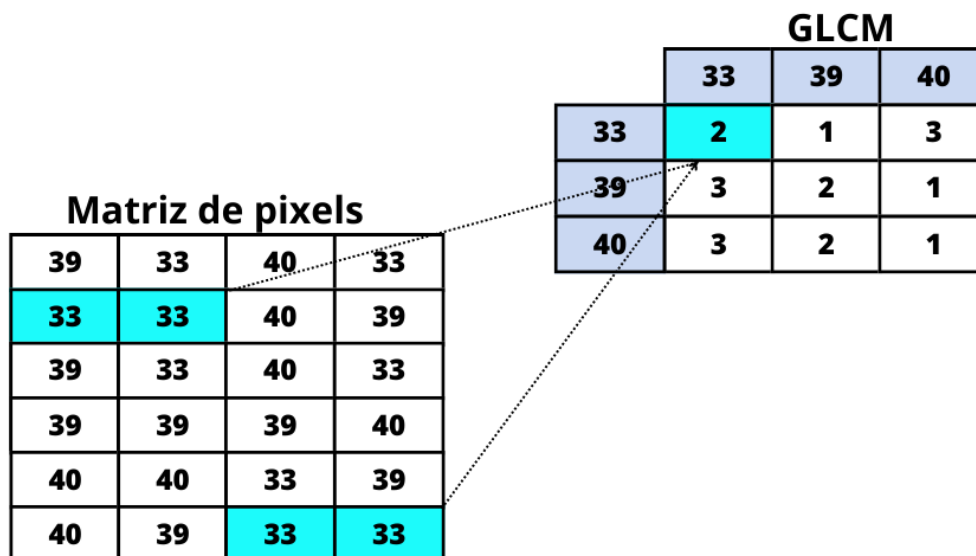


Figura 2.14: Matriz de *pixels* seguida da matriz GLCM.

Depois que a matriz GLCM é preenchida, os cálculos estatísticos são aplicados a ela. Assim, o autor (Haralick et al., 1973) estabeleceu 14 descritores para extração de textura da GLCM, incluindo: (correlação, contraste, uniformidade, entropia, homogeneidade, etc. ) Desta forma, de acordo com os cálculos estatísticos aplicados, são formados vetores de características, que podem ser utilizados em última instância no treinamento de redes neurais artificiais.

- **Local Binary Pattern (LBP):** O método de extração de características de padrão binário local, também se baseia nas características de vizinhança de *pixels* da imagem. Portanto, o LBP é definido como um conjunto de recursos de textura em tons de cinza, cujo valor é calculado atribuindo-se um valor binário a cada *pixel* da imagem, formando uma vizinhança linear de raio  $R$  em torno do *pixel* principal localizado na região central, ou seja, o operador analisa a vizinhança local e aplica o limiar consoante o *pixel* central, conforme exemplificado na Figura 2.15. O resultado da operação LBP é uma matriz binária, que então é usada como um descritor da imagem.

O exemplo dado na Figura 2.15 mostra uma definição LBP aplicada localmente a uma imagem digital, usando uma janela de tamanho  $(3 \times 3)$  *pixels*. O detector LBP usa uma janela de aplicação deslizante que se move ao lado da imagem. Em cada posição da janela do detector, calcula-se um descritor LBP. Para melhor entendimento, considere a janela definida para conter tons de cinza com a intensidade variando de  $(0 - 255)$  em cada *pixel*. Deste modo, o valor central do *kernel* será usado para definir os novos valores dos 8 *pixels* vizinhos da seguinte forma:

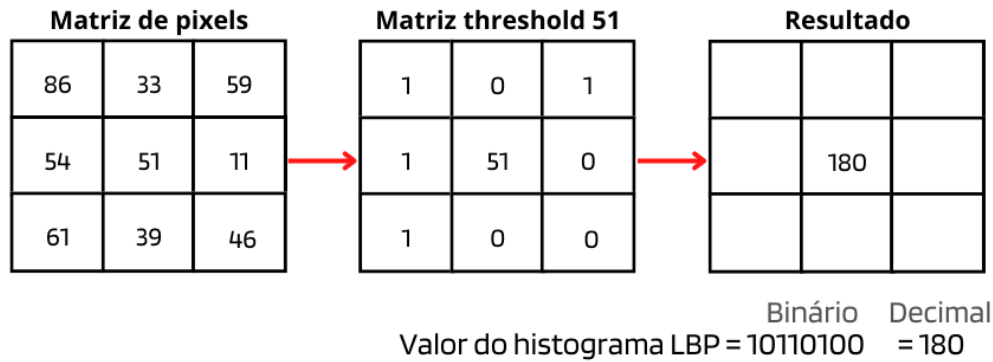


Figura 2.15: Exemplo de aplicação do algoritmo LBP.

1. Se o pixel central  $f(x,y)$  for maior que seu vizinho na vizinhança  $N_8(px)$  escreva '0', senão, escreva '1'.
  2. Após aplicação do limiar considerando todos os *pixels* vizinhos, será possível gerar um código com 8 números '0' ou '1', ou seja, um número binário de 8 *bits*.
  3. Em seguida, esse número binário deverá ser convertido em número decimal para obter um valor entre 0 e 255. No exemplo dado, o número binário calculado foi 10110100, que corresponde ao valor decimal 180.
  4. O descritor final será um histograma destes valores decimais atribuídos a cada *pixel*.
- **Histogram of Oriented Gradients (HOG):** Histograma de gradientes orientados é um descritor de recursos usado para fins de detecção de objetos em operações de reconhecimento de padrões e processamento de imagens. Essa técnica conta o número de ocorrências de direções de gradiente em partes locais de uma imagem. Para aplicações HOG, a imagem é dividida em pequenas regiões conectadas chamadas células, e para os *pixels* dentro de cada célula, um histograma de direções de gradiente é compilado. O descritor final é a concatenação desses histogramas, que se concentram na estrutura ou forma dos objetos na imagem, por esse motivo, é um dos recursos mais conhecidos para reconhecimento de objetos. Portanto, as características HoG codificam a estrutura dos objetos em uma imagem e podem ser aplicadas para todas as tarefas de detecção e reconhecimento de padrões, para as quais a informação da estrutura é considerada crucial.
  - **Contrast Limited Adaptive Histogram Equalization (CLAHE):** A equalização de histograma adaptativa de contraste limitado é aplicada por um algoritmo que separa a imagem em regiões contextuais e utiliza a equalização de histograma em cada uma delas.

---

O algoritmo CLAHE é uma evolução de uma técnica conhecida como “Adaptive Histogram Equalization (AHE)”, utilizada no pré-processamento de imagens para melhoramento de contraste. A técnica de AHE é adequada para melhorar o contraste local e aumentar as definições de bordas em cada região de uma imagem. Portanto, calcula vários histogramas, cada um correspondendo a uma seção distinta da imagem, e os utiliza para redistribuir os valores de luminância da imagem. Um dos problemas do AHE é que essa técnica tende a superamplificar o ruído em regiões relativamente homogêneas de uma imagem. Deste modo, buscando solucionar esse problema referente ao AHE, foi desenvolvido o algoritmo CLAHE que aplica o melhoramento de contraste evitando os efeitos de amplificação de ruídos.

Os algoritmos apresentados são alguns dos muitos disponíveis para extração de recursos. São exemplos clássicos ao lidar com a manipulação de *pixels*. Muitos desses métodos são práticos, mas podem ser exaustivos, dependendo do tamanho do conjunto de dados ou aplicabilidade. Foram apresentadas ferramentas de extração de características no domínio espacial, mas vale ressaltar que muitas características também podem ser obtidas no domínio da frequência por meio de transformadas.

## Principal Component Analysis (PCA)

A Análise de componentes principais é uma técnica estatística multivariada e não supervisionada que envolve a transformação de um conjunto original de variáveis em outro conjunto de variáveis com mesmas dimensões (chamadas de componentes principais), como apresentado na Figura 2.16. Cada componente principal (PC) é uma combinação linear de todas as variáveis originais, que são independentes umas das outras.

O objetivo da aplicabilidade é transformar os dados originais em um novo sistema de coordenadas, de modo que a maior variação de qualquer projeção dos dados esteja ao longo da primeira coordenada (o chamado primeiro componente), a segunda maior variação esteja ao longo da segunda coordenada e assim sucessivamente (Jolliffe, 2002).

A análise PCA auxilia no trabalho dos algoritmos de reconhecimento de padrões que lidam com grandes vetores de características, pois com o PCA é mais fácil enviar não todo o vetor de características, mas apenas os seus componentes principais para o algoritmo de treinamento e classificação, ou seja, os primeiros componentes que aparecem na lista de PCs, pois são eles os responsáveis por carregar a maior variação dos dados, portanto, são as características mais importantes em uma tarefa de divisão ou separação de classes.

### 2.2.5 Treinamento

A fase de treinamento é uma das principais tarefas de uma RNA, pois é nesta etapa que se mede a capacidade de fazer previsões bem-sucedidas, permitindo observar e quantificar

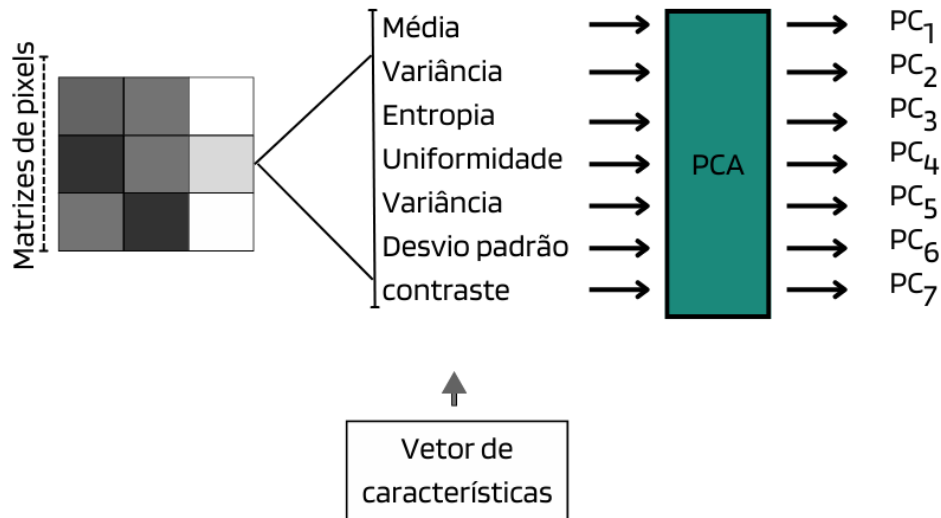


Figura 2.16: Transforme dados brutos em novas coordenadas chamadas de componentes principais (PCs).

qual é o poder de generalização do algoritmo em relação à aplicabilidade em novos dados.

Os algoritmos de *machine learning* podem ser classificados em pelo menos três tipos de aprendizado que se relacionam ao modo de treinamento, que são: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

Os algoritmos são chamados de supervisionados quando, pelo menos nos exemplos de treinamento, o algoritmo aprendiz pode avaliar se uma resposta provocada está correta, deste modo, o algoritmo usa essa supervisão para ajustar os pesos da rede, geralmente por meio da técnica de *backpropagation*, buscando diminuir o erro final de classificação durante o treinamento. Por fim, o algoritmo aprende a executar determinadas tarefas a partir de resultados conhecidos e obtidos das instâncias de dados de treinamento, para isso é preciso mapear toda a base de dados, identificando cada instancia que foi previamente identificada com o rótulo categórico.

Quanto ao uso de treinamento não supervisionado, os dados de treinamento consistem apenas em instâncias, sem nenhuma informação de rotulagem adicional sobre que tipo de saída é esperada para essas instâncias. O aprendizado não supervisionado identifica semelhanças nos dados e reage à presença ou ausência de tais semelhanças em cada novo dado de entrada. Geralmente está relacionado a métodos de *cluster*, que particiona objetos de dados em grupos. Como exemplo, têm-se os algoritmos de *K-means*, que avalia e aplica o agrupamento dos elementos de um conjunto de acordo com suas características e em *k*-grupos (Von Luxburg and Schölkopf, 2011; IBM, 2022c).

O treinamento por reforço é uma técnica que permite que as máquinas aprendam por meio de experimentação, ou seja, preocupa-se com o problema de um agente aprender,

---

por tentativa e erro, a atingir um objetivo interagindo com o seu ambiente sem que se tenha qualquer conhecimento prévio. Em vez de ser programado para realizar tarefas específicas, os algoritmos são incentivados a descobrir a melhor ação a ser tomada em cada situação por meio de recompensas e punições (Sutton and Barto, 1998).

Treinar uma RNA depende de um grande conjunto de dados, que durante o treinamento é geralmente dividido em três subconjuntos (treinamento, validação e teste). Dessa forma, o modelo é treinado no conjunto de treinamento e validado no conjunto de validação, após essas etapas, seu poder preditivo pode ser medido no conjunto de teste.

Segundo a literatura, não há uma regulamentação clara sobre a proporção da divisão do conjunto de dados durante o treinamento, mas, em geral, a divisão segue as seguintes medidas: 60% para treinamento, 10% para validação e 30% para teste.

### 2.2.5.1 Função de perda

Durante a etapa de treinamento de uma rede, seja de caráter supervisionado ou não, e seja a rede profunda ou superficial, faz-se necessário à utilização de uma função de perda. Esta função consiste em controlar a saída de uma rede neural, quantificando o quanto essa saída se desvia do valor esperado, ou seja, a função de custo ou perda como também é conhecida, calcula a pontuação da distância entre o resultado previsto e o resultado esperado de uma RNA durante o treinamento. Deste modo, para garantir maior poder de acurácia, a perda deve ser minimizada durante o treinamento, pois representa uma medida de erro do modelo.

As funções de perda mais utilizadas são:

- Entropia cruzada binária: para problemas de classificação de duas classes.
- Entropia cruzada multiclases: para problemas de classificação de muitas classes.
- Erro quadrático médio, em inglês, MeanSquareError (MSE): para problemas de regressão em variáveis com distribuição normal.

A entropia cruzada é uma entidade do campo da teoria da informação que mede a distância entre funções de probabilidade e, no caso de redes neurais para classificação, a distância entre a verdadeira distribuição de dados e as previsões do modelo.

Para entendimento desta importante função, toma-se como exemplo os algoritmos MLPs, que são frequentemente aplicados a problemas de aprendizagem supervisionada, assim, aplicando a função de custo com auxílio da técnica conhecida como *backpropagation*. Deste modo, o processo *backpropagation* aplica o aprendizado da rede por meio de um processo iterativo de ajuste de pesos, ou seja, procura atualizar os pesos sinápticos da rede, o que é feito por meio da minimização do erro quadrático médio *MSE-MeanSquareError* pelo método de Gradiente Descendente. Baseando-se neste método, o

fator de atualização do peso  $w_{ij}$  relativo à entrada  $i$  do neurônio  $j$  é dado pela Equação 2.12:

$$\nabla w_{ij} = -n \cdot \frac{\partial E}{\partial w_{ij}}, \quad (2.12)$$

onde  $\nabla w_{ij}$  é a variação do peso do neurônio  $j$  na conexão  $i$ ,  $n$  é a taxa de aprendizado, e  $E$  é o somatório do erro quadrático médio total.

Calcula-se  $E$  a partir da Equação 2.13, com a soma do erro quadrático médio de todos os padrões:

$$E = \frac{1}{2} \cdot \sum_p \sum_{i=1}^k (t_1^p - s_1^p)^2, \quad (2.13)$$

onde,  $p$  é o número de padrões,  $k$  é o número de neurônio na saída e  $t_1^p s_1^p$  são o valor desejado e a saída gerada pela rede para o neurônio  $i$  quando é apresentado o padrão  $p$ .

A minimização de cada padrão vai levar a redução do erro total, de modo que o erro passa a ser definido por:

$$E = \frac{1}{2} \cdot \sum_{i=1}^k (t_1 - s_1)^2 \quad (2.14)$$

### 2.2.5.2 Overfitting e underfitting

Quando um modelo de redes neurais não consegue obter um erro de treinamento suficientemente baixo, diz-se que o modelo apresenta um erro conhecido como *underfitting*, como exemplificado na Figura 2.17. Quando o erro obtido nos dados de treinamento é muito diferente do erro da classificação final, pode-se dizer que o modelo apresenta outro erro conhecido como *overfitting* (Theodoridis and Koutroumbas, 2008).

No exemplo da Figura 2.17, a tarefa da rede neural é aprender a diferenciar os círculos verdes (casos positivos) de quadrantes alaranjados (casos negativos). Analisando-se os cenários, o gráfico da rede (A) chegou próximo do esperado para um teste de classificação com poucos erros. A rede (B) não consegue estabelecer um limite que segue satisfatoriamente a distribuição de círculos e quadrantes na base de treinamento, apresentando resultados ruins durante o treinamento e possivelmente no teste. A rede (C) estabelece um limite perfeito que separa corretamente os círculos dos quadrantes. No entanto, esse limite sobreajusta os dados de treinamento e é improvável ter resultados semelhantes quando aplicado a outros dados de entrada.

O *overfitting* ocorre quando o modelo funciona bem nos dados de treinamento, mas não tão bem quando usados em dados de teste, ou elementos externos, sendo, portanto, incapaz de generalizar bem para novos dados, ou seja, ele não será capaz de realizar as tarefas de classificação ou previsão para as quais foi planejado.

O erro *underfitting*, geralmente ocorre quando o modelo não é treinado por tempo

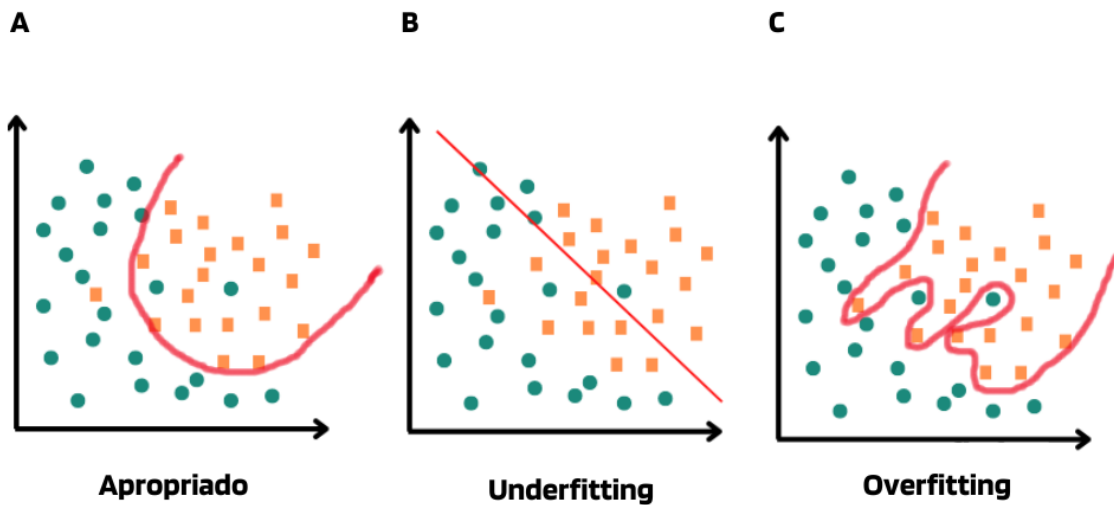


Figura 2.17: Exemplo de um modelo apropriado, seguido de um modelo com *underfitting* e *overfitting*.

suficiente ou não possui variáveis de entrada suficientes para determinar uma relação significativa entre as variáveis de entrada e saída. Para melhorias dos algoritmos, tanto o *underfitting* como o *overfitting* devem ser evitados ou ao menos tratados.

## 2.3 Classificação

Devido ao escopo do trabalho, esta seção trata em especial de algoritmos de classificação supervisionada e destinada à tarefa de classificação de imagens digitais. Deste modo, a tarefa de classificação se faz por meio de um procedimento de reconhecimento de padrões, que está contido nos algoritmos de *machine learning* como um campo científico que visa classificar objetos em um determinado número de classes ou categorias observando suas características. É considerada a etapa final de um algoritmo para classificar instâncias. Esta etapa envolve a aplicação de um método de decisão no qual novas entradas serão classificadas como pertencentes a uma determinada categoria preestabelecida.

Entre as técnicas utilizadas pelos algoritmos, o reconhecimento de padrões em imagens digitais é, sem dúvida, uma das áreas mais promissoras no campo da tecnologia de *deep learning*. O desempenho destes classificadores depende da relação entre sua complexidade e a qualidade do conjunto de dados de treinamento, ou seja, (quão bem ele representa a distribuição das classes) e o número de recursos usados.

Métodos de classificação e regressão de dados são exemplos de algoritmos supervisionados. Descreve-se a seguir alguns dos algoritmos amplamente utilizados para classificação de imagens, tais como: Random Forest, *Support-Vector Machine* (SVM),



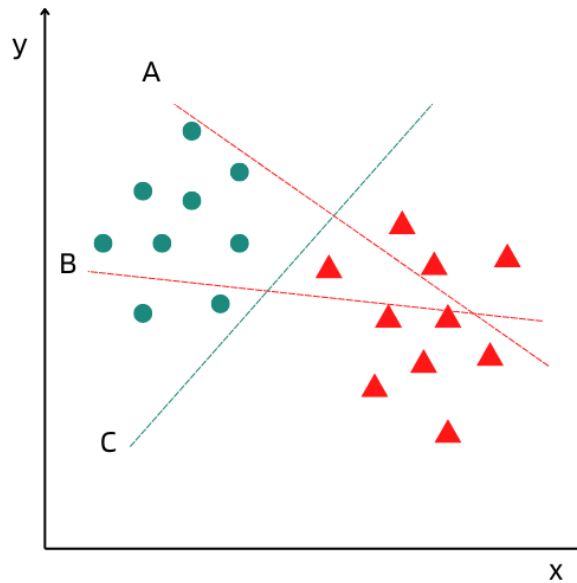


Figura 2.19: Exemplos de hiperplanos (A, B, C) na separação de vetores de suporte (quadrante verde e triângulo vermelho). O hiperplano (C) representa a melhor classificação dos vetores.

- **Classificador Naive Bayes:** O classificador Naive-Bayes se baseia na aplicação do teorema de Bayes, representado pela Equação 2.15. É um algoritmo que gera uma tabela de probabilidades a partir de uma técnica de classificação de instâncias, assumindo que a presença de uma característica particular em uma classe é independente da presença de qualquer outra característica. Deste modo, para calcular a classe provável de uma nova instância, as probabilidades de todas as classes possíveis são calculadas e, finalmente, a classe com maior probabilidade é selecionada como rótulo da nova instância (Rish et al., 2001).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (2.15)$$

$P(c|x)$  é a probabilidade posterior da classe ( $c$ ) dado o preditor ( $x$ ).  $P(c)$  é a probabilidade original da classe.  $P(x|c)$  é a probabilidade que representa a probabilidade do preditor dada a classe.  $P(x)$  é a probabilidade original do preditor.

Para entender melhor como o algoritmo funciona, é fornecido o seguinte exemplo; considere a Figura 2.20, de forma que, sejam 1000 alunos de uma escola, deste modo, faz-se uma contagem de quantos jogam futebol e quantos não jogam, ou ainda, quantos do sexo masculino ou não.

A partir dos dados apresentados na Figura 2.20, é possível calcular probabilidades como:



	 Futebol <span style="color: green;">✓</span>	 Não futebol <span style="color: red;">✗</span>	
Masculino	<b>P1</b> 610	<b>P2</b> 90	= 700
Não masculino	<b>P3</b> 100	<b>P4</b> 200	= 300
	<b>710</b>	<b>290</b>	<b>1000</b> Total

Figura 2.20: Exemplo de meios de calcular probabilidades com uso do Naive Bayes.

- P(Masculino) - Probabilidade de um integrante da escola ser masculino é:  
 $(P1 + P3)/100 = 7\%$
- P(Futebol) - (Probabilidade de um integrante da escola jogar futebol:  $(P1 + P3)/100 = 7,1\%$
- P(Masculino/Futebol) - Probabilidade de alguém do sexo masculino jogar futebol:  $(P1)/(P1 + P2) = 0,87\%$

Por fim, o algoritmo de Naive Bayes é um método similar ao apresentado para prever a probabilidade de classe diferente com base em vários atributos calculados sobre instâncias de dados.

- **Classificador KNN:** O objetivo do KNN é classificar cada instancia de um conjunto de dados avaliando sua distância em relação aos seus vizinhos mais próximos, como exemplificado na Figura 2.21.

A distância euclidiana é uma medida que o KNN geralmente usa para calcular o raio entre duas instâncias. Durante o processo, considera-se uma variável  $K$ , que representa a quantidade de vizinhos mais próximos que serão averiguados para analisar de qual classe a nova instância pertence (Kramer, 2013). Deste modo, se os vizinhos mais próximos forem à maioria de uma classe, a amostra em questão será classificada nesta categoria. Isso significa que novos pontos recebem um valor com base em sua similaridade com os pontos do conjunto de treinamento.

Considerando o exemplo dado na Figura 2.21 , por meio de votação uma nova instância com  $k = 5$  receberia 3 votos para ser classificada como classe (A) e 2 votos

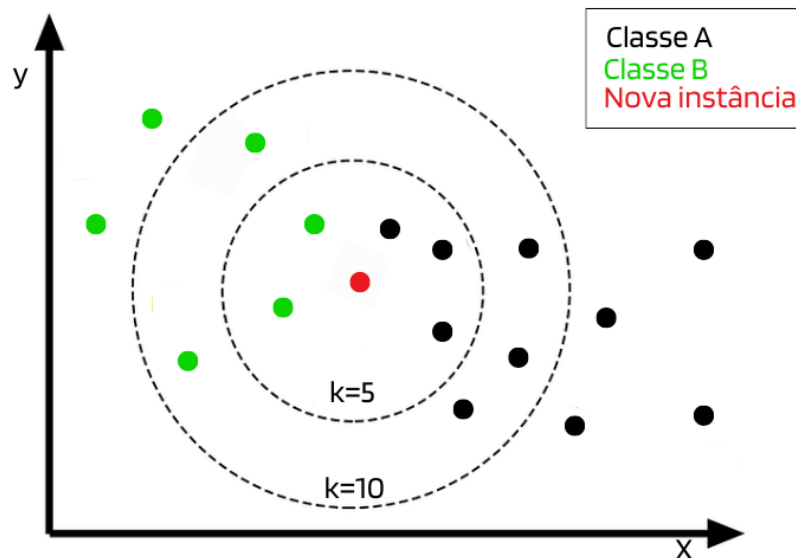


Figura 2.21: Exemplo do algoritmo KNN apresentando a vizinhança referente a  $k=3$  e  $k=7$ .

para ser classificada como classe (B), portanto a nova instância será classificada como pertencente à classe (A) por maioria de votos.

## 2.4 Deep learning e redes neurais convolucionais CNN

Algoritmos de *deep learning*, consistem em um método de aprendizado de representação com vários níveis de abstração para formular e processar dados de entrada e gerar saída, reconhecendo automaticamente estruturas complexas incorporadas em dados de alta dimensão. O termo “profundo” se refere ao número de camadas intermediárias na rede neural, quanto mais camadas a rede tem, mais profunda e complexa ela é, portanto, uma perspectiva sobre este tipo de metodologia é o conceito de profundidade, o que permite que o algoritmo aprenda determinada tarefa em várias etapas, ou seja, permite que modelos computacionais compostos por múltiplas camadas de processamentos aprendam representações de dados com múltiplos níveis de abstração (LeCun et al., 2015).

Quanto à forma de aprendizado, as redes *deep learning* geralmente se diferem das redes MLPs de metodologias superficiais na forma como essas redes aprendem a função  $f(\cdot)$ ; por exemplo, uma rede qualquer recebe como entrada uma imagem e pode classificá-la numa determinada classe. Dessa forma, pretende-se aprender a função,  $f(x) = y$ , em que  $x$  representa a imagem, e  $y$ , a classe provável para  $x$ .

Buscando apreender a função  $f(x)$ , redes superficiais ou não *deep* procuram diretamente uma única função que possa produzir o resultado desejado a partir de um conjunto de parâmetros. Redes *deep learning* têm uma forma de se organizar que considera uma

---

função  $f(x)$  por meio da composição de funções. Por exemplo, considere a composição de  $K$  funções da Equação 2.16:

$$f(x) = f_K(\dots f_2(f_1(x_1))\dots), \quad (2.16)$$

Cada uma dessas funções  $f_k(\cdot)$  recebe um vetor de dados  $x_k$  como entrada e gera o próximo vetor  $x_{k+1}$  como saída. O índice  $k$  se refere a uma camada. Assim, cada função usa parâmetros para realizar a transformação dos dados de entrada e, o conjunto desses parâmetros é representado por  $W_k$ , relacionando cada função  $f_k$ , de modo que, então, pode-se escrever conforme a Equação 2.17:

$$f_K(\dots f_2(f_1(x_1, W_1); W_2)\dots), W_K) \quad (2.17)$$

Na Equação 2.17,  $x_1$  representa os dados de entrada. Desta forma, cada função tem o seu próprio conjunto de parâmetros e a sua saída passada para a função adjacente seguinte, ou seja, trabalha em camadas de neurônios na forma *feed-forward* (Ponti and Da Costa, 2018).

Quanto ao uso de funções de ativação, em redes *deep learning* a função de ativação *ReLU*, tem sido uma das mais utilizada por facilitar o processo de treinamento. Isso se deve ao fato das funções *sigmoidais* saturarem a partir de um determinado ponto (Patil and Rane, 2021; Goodfellow et al., 2016). Quanto à aplicabilidade das redes *deep learning*, existem dois tipos de modelos amplamente utilizados por estes algoritmos (regressão e classificação), a saber:

- **Regressão:** aplicada de forma simples ou multivariada, é um dos métodos de previsão amplamente utilizados na comunidade estatística. O seu principal objetivo é verificar como a variável de interesse (variável preditora) afeta a variável resposta  $X$  (variável dependente), de modo a criar um modelo matemático capaz de prever o valor de  $X$  com base em novos valores da variável preditora  $Y$ , ou seja, é um método de estimar o valor de algo com base em uma variedade de outros dados históricos.
- **Classificação:** para a indução de modelos de classificação, os algoritmos são usados para introduzir modelos preditivos observando um conjunto de objetos rotulados, muitas vezes referidos como o conjunto de treinamento. Assim, o problema trabalha com dois tipos de espaços: o espaço de entrada  $X$  (também chamado de espaço de instâncias) e o espaço de saída  $Y$  (espaço de rótulos). Por exemplo, se a tarefa é classificar certas imagens de fundo de olho em um determinado conjunto finito de categorias "glaucoma" e "não glaucoma", então  $X$  consiste no espaço de todos os objetos possíveis (instâncias), em outras palavras, consiste em todos os possíveis tipos de exemplos de imagens com glaucoma e imagens sem glaucoma em uma determinada representação fixa, enquanto  $Y$  é o espaço de todas as cate-

---

gorias disponíveis, como mostrado na Equação 2.18. Para aprender, o algoritmo recebe alguns exemplos de treinamento

$$(X_1, Y_1) \dots (X_n, Y_n) \quad (2.18)$$

, ou seja, pares de objetos com o rótulo de categoria correspondente. O objetivo é então encontrar um mapeamento  $f : X \rightarrow Y$  que cometa o menor número de erros possível. O mapeamento  $f : X \rightarrow Y$  é chamado de classificador (Von Luxburg and Schölkopf, 2011).

Considerando metodologias voltadas para a classificação de imagens digitais, inúmeros algoritmos de *deep learning* tem mostrado desempenho muito alto nessa tarefa com uso de redes CNNs, portanto, descreve-se a seguir, este algoritmo, considerado o mais promissor no reconhecimento de padrões e classificação de imagens digitais.

### 2.4.1 Redes neurais convolucionais CNN

Redes CNNs são usadas em muitas áreas, como reconhecimento de padrões de imagens digitais, reconhecimento de fala, processamento de linguagem natural e análise de vídeo. Conforme exemplificado na Figura 2.25, a sua consistência está geralmente em uma ou mais camadas convolucionais, uma camada de subamostragem, em inglês *pooling*, seguida por uma ou mais camadas totalmente conectadas, estas últimas camadas são aplicadas da mesma forma aplicadas em uma rede neural padrão (Goodfellow et al., 2016).

#### Camada de convolução

O propósito de uma camada convolucional é aplicar  $n$  filtros (também conhecido como *kernel*), nos dados de entrada da rede para gerar  $n$  saídas chamadas mapas de ativação. Cada filtro é aplicado à profundidade da entrada. Por exemplo, para uma imagem contendo três canais de cores do sistema RGB, o filtro também deve conter três profundidades para gerar um mapa de ativação 2D que considere todos os canais de entrada, como altura, largura e componentes de cores RGB. Esses mapas são empilhados para formar um tensor com profundidade igual ao número de filtros. Desse modo, esse tensor é passado como entrada para a próxima camada da rede que consiste na camada de *pooling* (Krizhevsky et al., 2017a).

A camada convolucional de uma rede CNN atua como um extrator de recursos, em que os pesos do *kernel* dos filtros convolucionais são capazes de capturar com sucesso as dependências espaciais, permitindo mapear características por meio de ações de convolução (Farabet et al., 2009).

A Figura 2.22, mostra um exemplo de transformação dos valores de *pixels* de uma imagem de entrada usando um *kernel* de tamanho 3x3. A cada transformação aplicada,

busca-se por novas características representativas, deste modo, o filtro lê sucessivamente a matriz de pixels, da esquerda para a direita e de cima para baixo, todos os pixels da área de ação do *kernel*. Para obter o novo valor do *pixel*, multiplica-se o *kernel* pelo valor da imagem original em torno do *kernel* na posição original. Ao final, soma-se o produto e em seguida divide pelos valores de seus coeficientes, obtendo-se o valor do *pixel* na nova imagem. Os filtros vão sendo ajustados a cada iteração da rede de tal modo ao ponto ótimo, quando a entrada contiver uma determinada característica comum aos lotes de entrada, como por exemplos, arestas, cores, dentre outras características.

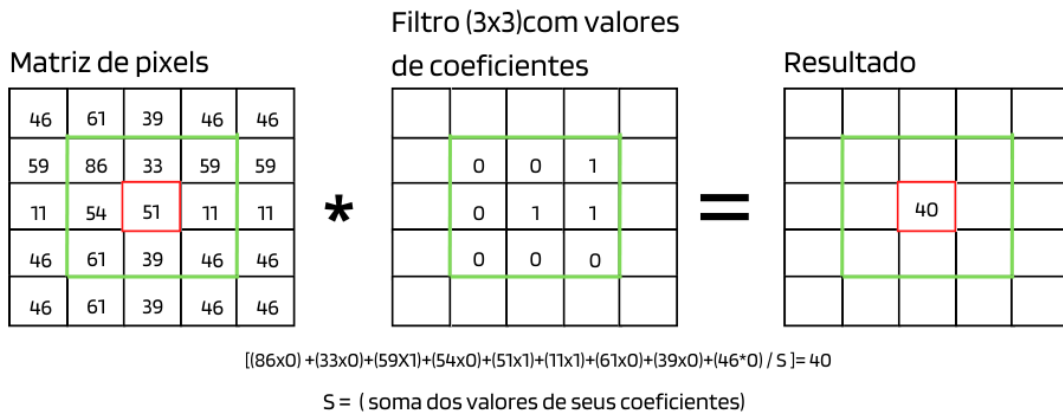


Figura 2.22: Exemplo de aplicação de *kernel* por convolução.

Na forma matemática, tem-se a convolução de um filtro  $w(x, y)$  de tamanho  $m \times n$  com uma imagem  $f(x, y)$ . Essa operação pode ser expressa como  $w(x, y) * f(x, y)$ . É definida como uma operação linear que computa a superposição de duas funções em função de um deslocamento, como mostra a Equação 2.19:

$$w(x, y) * f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x-s, y-t) \quad (2.19)$$

Além do *kernel* utilizado, a saída do processo de convolução é determinada pelas variáveis; *padding* e *stride*. O *padding* é caracterizado pela inserção de novos *pixels* simetricamente em torno da imagem de entrada. Isso é feito para que o tamanho da saída seja sempre controlado e as bordas da imagem original sejam devidamente consideradas durante a convolução. O passo (*stride*) é o número de pixels que o *kernel* move sobre a imagem de entrada para executar cada produto escalar durante o processo de convolução.

## Pooling

As camadas de *pooling*, conforme exemplificado na Figura 2.23, fazem que a rede CNN reduza o custo computacional e também possa se tornar invariante a pequenas transformações na imagem.

É uma camada que recebe como entrada a saída da camada de convolução e consiste

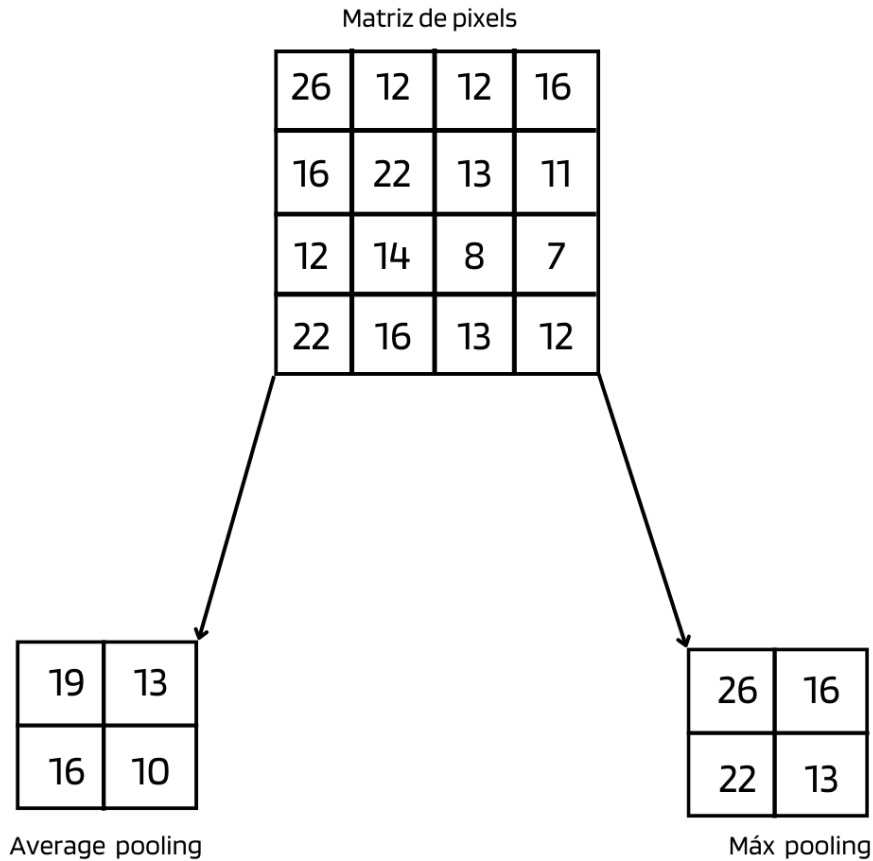


Figura 2.23: Exemplo da aplicação *pooling* médio e máximo em uma matriz de *pixels* 4x4 (Russell, 2010).

em realizar uma subamostragem de cada mapa de ativação recebido. Duas funções comumente usadas são *max pooling* e *average pooling*. A função *max pooling* encontrará o valor máximo no gráfico, e a função *average pooling* encontrará a média dos valores no gráfico (Goodfellow et al., 2016). A técnica de *pooling* também beneficia a rede neural, ao passo que reduz a quantidade de dados para a camada seguinte e melhora a regularização da rede, reduzindo o custo de memória e processamento.

Os dados produzidos pela operação de *pooling* não apresentam a configurações adequadas para serem utilizadas como sinal de entrada da rede neural conectada na etapa subsequente, pois o sinal deve ser disposto em formato vetorial. Por isso, é necessário que se use uma operação denominada *flattening*, sendo responsável por transformar a matriz da imagem obtida pela camada de *pooling*, alterando o seu formato para vetorial. Por exemplo, se a entrada para a camada plana for uma matriz (16 x 16 x 3) RGB, a saída será uma matriz com 768 posições, como mostra a Figura 2.24.

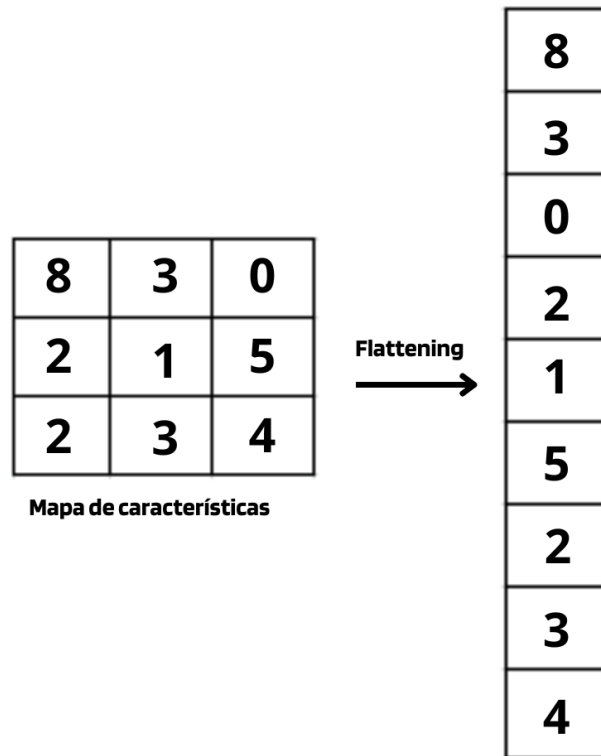


Figura 2.24: Exemplo da aplicação *flattening*.

### Camadas totalmente conectadas

Em redes CNNs usa-se uma camada conhecida como densa ou totalmente conectada no final da rede, após várias camadas convolucionais e *pooling*. Uma camada totalmente conectada é equivalente às camadas presentes nas redes neurais clássicas. Essa camada utilizada nas CNNs é responsável por dar a saída final com base na função de ativação utilizada. Assim, no contexto de reconhecimento de imagem, é comum projetar essa camada densa com uma função de ativação *ReLU* e, no final da rede, projetar outra camada densa com função de ativação *softmax*, com dimensões iguais ao número de classes a serem classificadas como saída.

### 2.4.2 Extração de características com uso de redes CNNs

Redes CNNs são muito utilizadas em áreas de reconhecimento de padrões em imagens digitais. Este campo envolve treinar máquinas para executar funções de reconhecimento de padrões consideravelmente mais rápido do que os humanos. Desta forma, se fornecer dados suficientes aos algoritmos, eles aprenderão a entender os dados e possibilitará a tarefa de distinguir, por exemplo, uma imagem de outra.

Um dos diferenciais entre a utilização de redes CNNs com os outros classificadores apresentados, é que a extração de características, considerada uma etapa necessária aos algoritmos de aprendizagem profunda, é inerente desta arquitetura, ou seja, as redes CNNs

tem a capacidade de utilizar a extração de características de forma automática, sem a necessidade de algoritmos paralelos ou interferências manuais. Por fim, conforme a entrada, as redes CNNs, selecionam os seus próprios *kernels* de extração de recursos e posteriormente utilizam as características extraídas das imagens para tarefas de treinamento e de classificação, segmentação ou regressão.

A Figura 2.25, exemplifica um modelo de rede CNNs e como ela se difere de abordagens clássicas como SVM, KNN, Naive Bayes, etc.

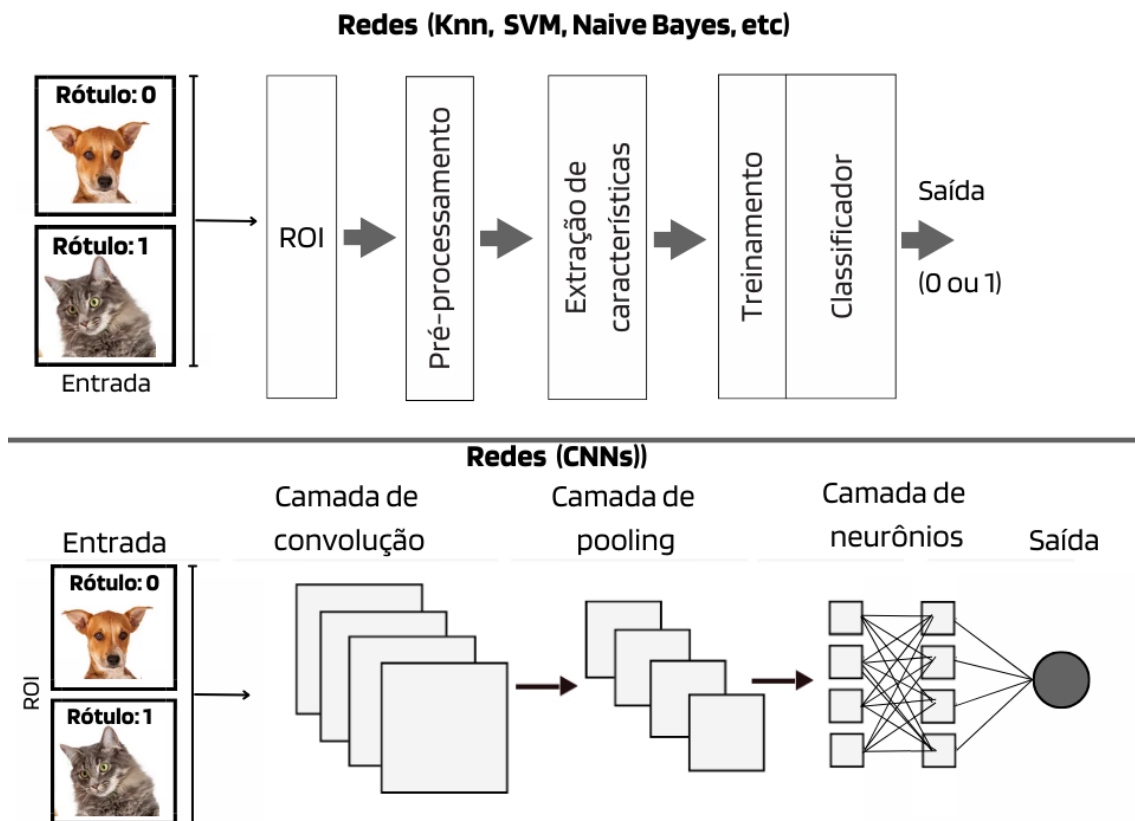


Figura 2.25: Exemplo de camadas de uma rede CNN.

A extração de recursos em aplicativos de redes CNNs é feita usando vários filtros ou *kernels* durante o estágio de convolução da rede. Os filtros são selecionados automaticamente com base em sua entrada, sendo possível manipular e determinar o valor de *stride*. Essa automação da seleção de filtros é importante, pois ajuda na seleção dos mapas de características mais sofisticados que possam melhor representar as imagens de entrada.

As redes CNNs, durante a extração de características das imagens digitais, em suas primeiras camadas de treinamento, identificam arestas sólidas e formas simples e, em seguida, realizam previsões iterativas para incorporar informações. Portanto, estes algoritmos requerem grandes quantidades de dados, que são analisados iterativamente até que se possam identificar características relevantes e, finalmente, compreender um determinado padrão.

A utilização das redes CNNs como métodos de extração de características, permite

---

aos algoritmos obterem resultados tão bons ou até melhores que os métodos clássicos. No entanto, isso não significa que a pesquisa de métodos baseados na extração de recursos seja ineficaz. Deve-se considerar que uma vez que os métodos que utilizam redes profundas possuem alto custo computacional e, muitas vezes, alto tempo de extração, além de exigirem uma abundância de dados para funcionarem adequadamente. Assim, em cenários em que não se tem muita informação ou poder de computação, os métodos de engenharia de características tiram vantagem por serem computacionalmente simples, mas muito robustos em múltiplas variáveis (Liu et al., 2017).

### 2.4.3 Aprendizado por transferência

Redes neurais são algoritmos que requerem muitos dados para treinamento, mas muitas vezes não há dados suficientes, principalmente considerando informações clínicas. Portanto, uma técnica amplamente utilizada que permite que redes neurais sejam aplicadas a pequenos conjuntos de dados é o processo de aprendizado por transferência.

O aprendizado por transferência é o método de transferir conhecimento adquirido durante o treinamento em um determinado domínio (uma base de dados), para ser aplicado em outro domínio, ou seja, outro problema semelhante. À vista disso, os algoritmos que oferecem essa tecnologia são chamados de pré-treinados. Uma das conveniências de se utilizar redes pré-treinadas é que elas já possuem pesos definidos, ou seja, os pesos são inicializados com valores obtidos de um treinamento já concluído.

Existem inúmeras opções de esse aprendizado ser aplicado, incluindo transferência total ou parcial do conhecimento, aplicado com uma técnica chamada de congelamento de pesos. A opção de utilização depende do problema a ser resolvido e do seu grau de semelhança com o problema utilizado para o pré-treinamento.

Ainda sobre o aprendizado por transferência, o *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) é uma competição anual realizada pela equipe *ImageNet* desde 2010, na qual equipes de pesquisa avaliam o desempenho de algoritmos de visão computacional e *machine learning* em várias tarefas de reconhecimento visual, como classificação e localização de objetos (Russakovsky et al., 2015). O *ImageNet* é um projeto visando fornecer grandes bibliotecas de imagens para uso em pré-treinamento de algoritmos a serem utilizados em diversas outras tarefas. Esta base de dados contém mais de 14 milhões de imagens divididas em mais de 20.000 categorias.

Juntamente com o aprendizado por transferência, a base de dados *ImageNet* tem sido fundamental para o avanço de pesquisas em visão computacional e *deep learning* auxiliando no melhor desempenho dos algoritmos de reconhecimento de padrões em imagens digitais; além disso, a utilização do aprendizado por transferência pode reduzir o custo computacional de treinar novos algoritmos, visto que não requer treinar todas as camadas da rede do zero (Russakovsky et al., 2015).

---

## 2.5 Redes Adversárias Generativas (GANs)

Redes Adversárias Generativas, em inglês, *Generative Adversarial Networks* (GANs), é uma arquitetura *machine learning* que consiste em duas redes que lutam uma contra a outra (dano origem ao nome "adversária"). O potencial das GANs é enorme porque elas podem aprender a imitar qualquer distribuição de dados da seguinte maneira: primeiro, uma rede neural chamada gerador gera novas instâncias de dados, enquanto outra rede neural chamada discriminador avalia sua autenticidade. Desta forma, o gerador produz imagens falsas na esperança de que mesmo as imagens falsas sejam consideradas reais pelo discriminador.

Com essa troca de informações, o gerador aprende a gerar dados plausíveis, enquanto o discriminador aprende a distinguir dados falsos do gerador. O discriminador penaliza o gerador para produzir resultados concretos e com isso o gerador vai se aperfeiçoando cada vez mais.

O treinamento é feito usando instâncias de dados reais como positivas e instâncias de dados falsas criadas pelo gerador como negativas durante. Após o treinamento, o classificador classifica os dados reais e falsos do gerador e propaga a perda do discriminador através da rede do discriminador para atualizar os pesos (Goodfellow et al., 2020).

## 2.6 Métricas de avaliação

Os modelos de *deep learning* utilizados para a classificação categórica são avaliados comparando a classe prevista pelo modelo com a classe verdadeira para instância testada. Assim, todas as métricas de classificação compartilham um objetivo comum de medir o quão próximo um modelo de *deep learning* está de uma classificação perfeita, mas atingem esse objetivo de maneiras diferentes. Uma forma que permite a visualização do desempenho de um algoritmo de classificação é por meio do cálculo de uma tabela chamada de matriz de confusão, conforme mostra a Figura 2.26. O exemplo considera uma classificação binária, onde uma instância de entrada pode terminar com um resultado positivo ou negativo.

Para análise da matriz de confusão deve se considerá-la como o resultado do classificador, deste modo, denota-se que: (VP) caracteriza os resultados verdadeiro-positivos e (VN) explica os verdadeiro-negativos. Para as classes incorretamente identificadas, (FP) refere-se a falso-positivos e (FN) refere-se a falso-negativos (Swift et al., 2020).

Com os dados obtidos pela matriz de confusão, é possível calcular as principais métricas para avaliar modelos de IA aplicados na classificação de dados, permitindo o cálculo da precisão, acurácia, sensibilidade, especificidade, *F1-score* e *Kappa*, a saber:

- **Precisão (PR):** essa métrica, representada pela Equação 2.20, fornece informações sobre o número de observações classificadas como VP e que foram realmente posi-

	SIM	NÃO
REALIDADE: SIM — Doente →	VERDADEIRO POSITIVO (VP)	FALSO NEGATIVO (FN)
REALIDADE: NÃO — Saudável →	FALSO POSITIVO (FP)	VERDADEIRO NEGATIVO (VN)

Figura 2.26: Exemplo de aplicação da matriz de confusão.

tivas, ou seja, dentre todas as classificações VP que o modelo previu, quantas estão corretas.

$$PR = \frac{VP}{VP + FP} \quad (2.20)$$

- **Acurácia (ACC):** representada pela Equação 2.21, é considerada uma das métricas simples e importantes. Ela avalia a taxa de acertos, que pode ser obtida pela razão entre o número de acertos e o número total de entradas, ou seja, é uma resposta de quantos indivíduos foram classificados corretamente considerando o total de indivíduos envolvidos.

$$ACC = \frac{VP + VN}{VP + FN + VN + FP} \quad (2.21)$$

- **Sensibilidade (SE):** essa métrica é representada pela Equação 2.22, e expressa a capacidade que um teste diagnóstico tem de discriminar, dentre os suspeitos de uma patologia, aqueles efetivamente doentes, ou seja, dentre todas as situações VP como valor esperado, quantas estão corretas.

$$SE = \frac{VP}{VP + FN} \quad (2.22)$$

- **Especificidade (ES):** descrito pela Equação: 2.23, esta métrica representa a capacidade de um teste ser negativo dada uma amostra de um indivíduo sabidamente não portador da doença, ou seja, de todos os saudáveis, quantos foram preditos corretamente.

$$ES = \frac{VN}{FP + VN} \quad (2.23)$$

- **F1-Score (F1) :** esta métrica é definida pela Equação 2.24 e indica o equilíbrio entre precisão e a sensibilidade. A pontuação F1 pode ser interpretada como a

média harmônica de precisão, em que o melhor valor da pontuação F1 é 1 e o pior valor é 0.

$$F1 = \frac{2 * PR * SE}{PR + SE} \quad (2.24)$$

- **Kappa - (k):** As contribuições relativas da métrica *Kappa*, (K), são analisadas da mesma forma que a F1 métrica. O coeficiente *Kappa*, é um método estatístico usado para avaliar o nível de concordância ou reprodutibilidade entre dois conjuntos de dados (Hoo et al., 2017).

Os resultados dos algoritmos de *deep learning* também podem ser avaliados graficamente por meio da característica de operação do receptor, em inglês, *Receiver Operating Characteristic* (ROC) e área sob a curva, em inglês (*Area Under the Curve* (AUC), ferramentas poderosas para medir e comparar o desempenho de modelos de classificação binária em aprendizado de máquina. A Figura 2.27 apresenta um exemplo de área sob a curva ROC.

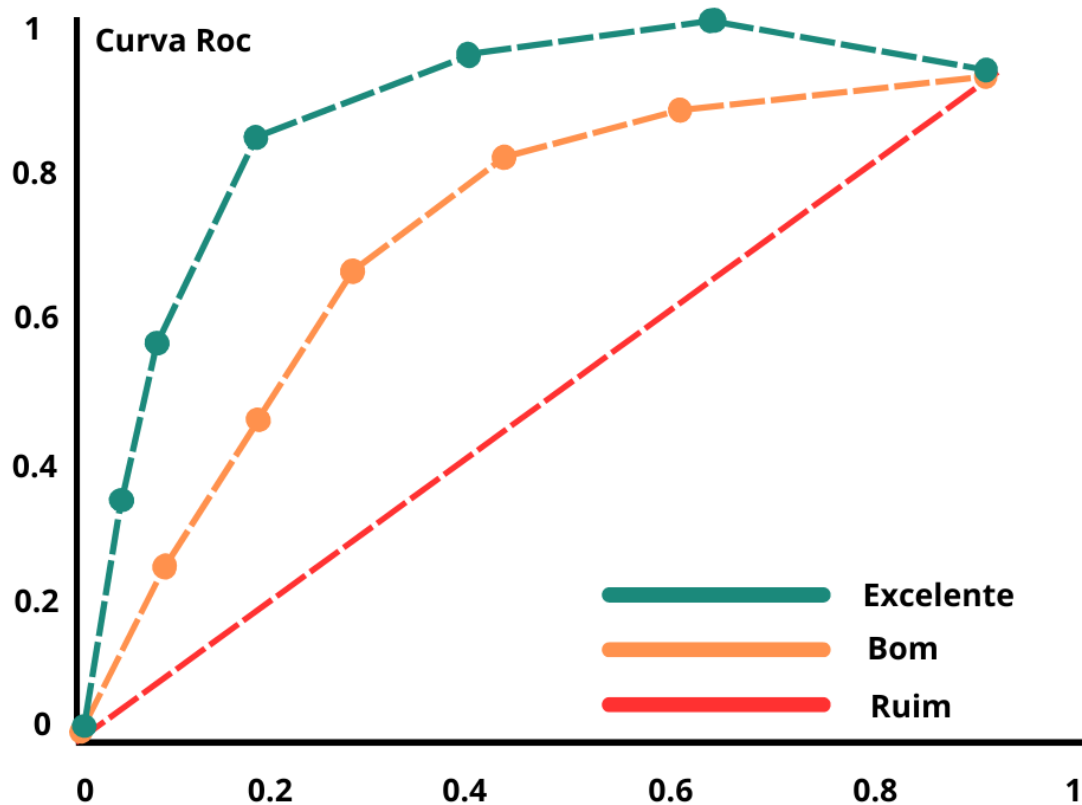


Figura 2.27: Exemplo de curva roc.

A curva ROC é um gráfico que mostra o desempenho de um modelo em todos os limites de classificação. Essa curva plota dois parâmetros: taxa de verdadeiro-positivos e taxa de falso-positivos. Os valores de AUC variam de [0,0] a [1,0], com um limite entre

---

classes de 0,5; portanto, um modelo que prediz 100% correto tem AUC igual a 1 (Hoo et al., 2017).

Para avaliar algoritmos de segmentação de imagens, a métrica Interseção Sobre União (IoU), exemplificada na Figura 2.28, mede a precisão de um detector de objetos aplicado a um conjunto de dados específico. Com essas métricas é possível medir a área comum entre as regiões prevista (P) e esperada (E), dividida pela área total das duas regiões, conforme apresentado na Equação 2.25:

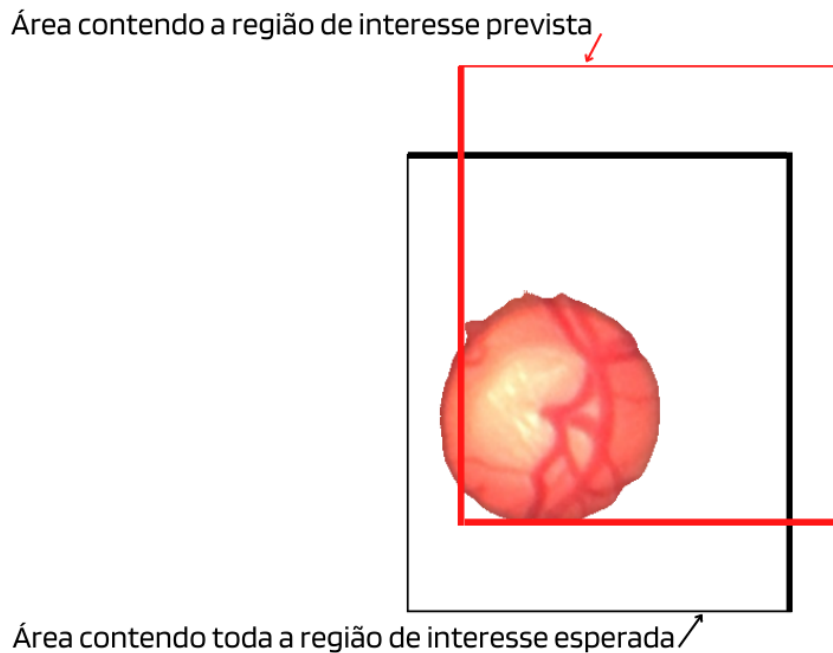


Figura 2.28: Exemplo de aplicação da métrica IoU na segmentação da região do disco óptico.

$$IoU = \frac{Area(P \cap E)}{Area(P \cup E)} \quad (2.25)$$

Outras métricas, como o coeficiente de *dice* e o índice de *Jaccard*, são amplamente utilizadas na literatura para descrever a qualidade de algoritmos de segmentação. Essas duas métricas podem representar a semelhanças entre duas imagens diferentes, por exemplo, uma imagem segmentada manualmente e outra segmentada automaticamente.

Ao desenvolver projetos de *machine learning* e ciência de dados, é crucial usar métricas apropriadas para cada problema. Seu valor reflete a qualidade do modelo, portanto, se for escolhido incorretamente, não será possível avaliar se o modelo atende efetivamente aos requisitos necessários para a generalização dos resultados. Em suma, este Capítulo apresentou algumas das principais tecnologias utilizadas por algoritmos de IA na classificação de imagens digitais. O próximo Capítulo apresenta como essas tecnologias foram dedicadas ao estudo de diagnóstico do glaucoma com uso de imagens de fundo de olho em trabalhos relacionados.

# Capítulo 3

## Estado da arte

Este Capítulo apresenta o progresso dos algoritmos de inteligência artificial no processamento de imagens digitais e nos estudos de diagnóstico automático da doença ocular glaucoma. Inicialmente apresenta-se uma introdução aos tipos de glaucoma existentes e às formas tradicionais de diagnóstico. Apresenta também a epidemiologia mundial da doença e a dificuldade de combater o subdiagnóstico desse distúrbio neurológico, principalmente em países subdesenvolvidos como o Brasil. Além disso, na Seção de trabalhos relacionados, são apresentadas as principais bases de dados públicas utilizadas na pesquisa de algoritmos de *deep learning* para classificação automática de glaucoma a partir de imagens digitais de fundo de olho, juntamente com as principais pesquisas relacionadas a esse tópico.

### 3.1 A degeneração do nervo óptico e o glaucoma

A neuropatia óptica glaucomatosa é progressiva, crônica e incurável, no entanto, geralmente, pode ser eficientemente controlada pelo diagnóstico e tratamento precoce. Conforme descrito em (Kanase and Yadav, 2019) e (Kanski and Bowling, 2011), o glaucoma pode ser classificado em ao menos 5 principais tipos:

- **Glaucoma de ângulo aberto:** este tipo de glaucoma está relacionado ao bloqueio lento e progressivo dos canais de drenagem do humor aquoso, um fluido que circula pela câmara anterior do olho. A função do humor aquoso é nutrir áreas ao redor da íris e da córnea. O fluido é produzido e drenado continuamente, mantendo um equilíbrio regulado por uma pressão conhecida como Pressão Intraocular (PIO). Quando esse fluido não escoar adequadamente, pode criar uma pressão muito alta e danificar o nervo óptico, alterando a sua estrutura e causando danos à visão. É o tipo mais comum do glaucoma, sendo importante obter um diagnóstico o mais rápido possível para evitar que a doença progrida para um estágio em que possa levar à cegueira total.

- **Glaucoma de ângulo fechado:** causado por um súbito bloqueio da saída do humor aquoso, resultando em sintomas imediatos, necessitando de atendimento emergencial. Caso não seja tratado de maneira rápida, o glaucoma de ângulo fechado também pode causar a perda definitiva e irreversível da visão.
- **Glaucoma de pressão normal:** é uma neuropatia óptica caracterizada por todos os efeitos estruturais e funcionais da doença, mas sem evidência de PIO elevada.
- **Glaucoma congênito:** o tipo de glaucoma congênito é o menos comum, mas pode ocorrer, pois, caso haja o aumento da PIO durante a formação fetal, a criança pode já nascer com a doença. O diagnóstico precoce é desafiador, porque a condição é não específica e varia com o grau de deformidade anatômica e com o início dos sintomas.
- **Glaucoma secundário:** os tipos de glaucoma geralmente têm fatores genéticos que possibilitam o desenvolvimento da doença, já as ocorrências secundárias são devidas a fatores externos, como, por exemplo, uma reação adversa ao uso de um medicamento. Desta forma, se a saída do humor aquoso for prejudicada por algum fator externo causando aumento da PIO, a doença se desenvolve.

Embora existam diferentes tipos de glaucoma, todos eles causam danos progressivos ao nervo óptico e, geralmente, ocorre lentamente, levando inicialmente à perda da visão média periférica. Em estágios avançados, como exemplifica a Figura 3.1, afeta a visão central, levando à cegueira irreversível. Ainda que não exista cura para o glaucoma, o tratamento pode retardar o seu desenvolvimento; portanto, é de grande importância o diagnóstico precoce (Giorgis et al., 2019; Kim et al., 2022).

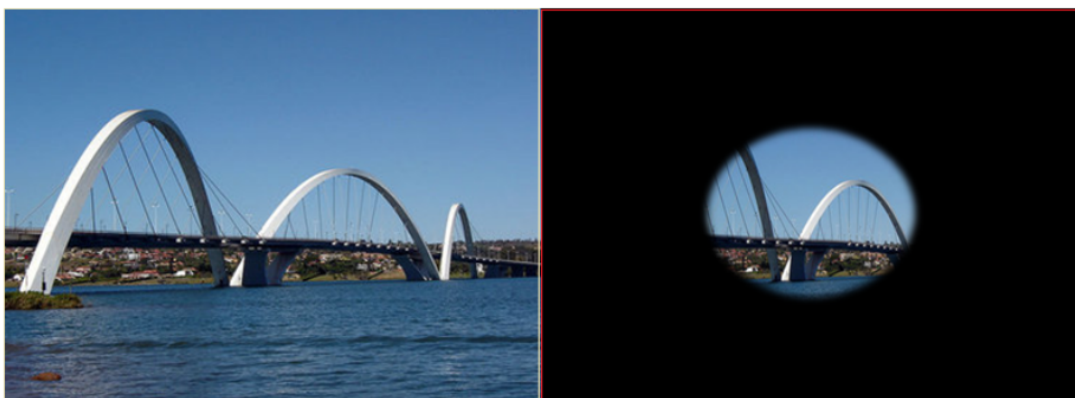


Figura 3.1: Exemplo de glaucoma avançado com perda da visão periférica.

Quanto aos métodos de examinar a região da retina, ou seja, do fundo de olho como exemplificado na Figura 3.2, o disco óptico corresponde a área onde os axônios de mais de um milhão de células ganglionares da retina convergem para formar o nervo óptico,

---

responsável por transmitir informações visuais ao cérebro (Smith and Czyz, 2022). Em se tratando dos danos ao disco óptico, eles podem ser vistos a partir de exames de fundo de olho, (também conhecidos como fundoscopia ou oftalmoscopia). O exame de fundo de olho é um exame utilizado por um clínico geral, optometrista ou oftalmologista para examinar a área da retina no diagnóstico e monitoramento de várias doenças, incluindo doenças oculares como degeneração macular, descolamento de retina, catarata, glaucoma, retinopatia diabética, etc.

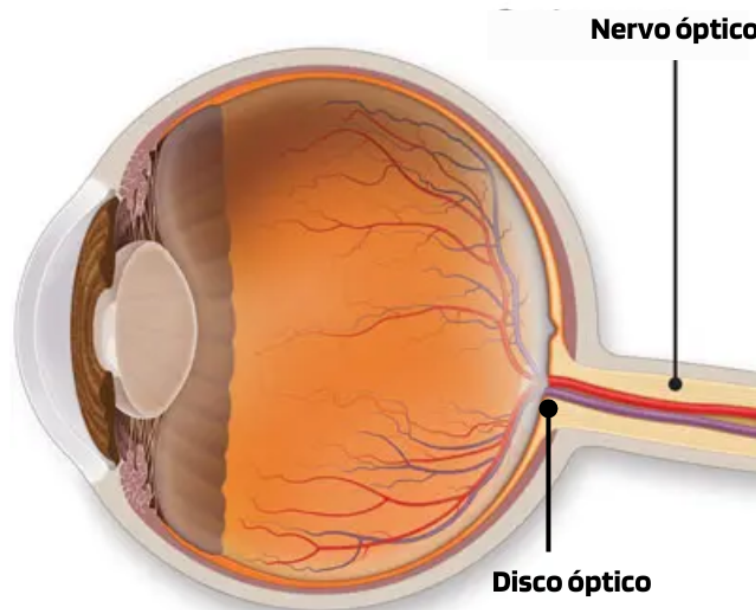


Figura 3.2: Exemplo do globo ocular apresentando a retina, o disco óptico e nervo óptico - adaptado de (AAO., 2023).

O exame fundoscópico pode ser realizado com uma variedade de equipamentos, como descrito a seguir:

- **Oftalmoscopia direta:** neste exame, um oftalmoscópio é usado para iluminar diretamente o olho do paciente para examinar o fundo. O dispositivo consiste em luzes com várias modificações ópticas, incluindo lentes que iluminam a retina através do canal normal da íris (pupila). Durante esse procedimento, são examinadas as estruturas mais internas do globo ocular (chamadas coletivamente de fundo), que corresponde à área da retina, vasos retinianos e disco óptico.
- **Oftalmoscopia indireta:** este exame utiliza um oftalmoscópio indireto especialmente projetado para oftalmologia. Este procedimento permite ver melhor toda a retina, incluindo a parte frontal, que é difícil de ver com outros métodos.
- **Biomicroscopia de lâmpada de fenda:** durante esse teste, o paciente é colocado em uma cadeira em frente a um instrumento chamado microscópio de lâmpada de

fenda, que combina vários equipamentos em um só. Consiste em um microscópio binocular montado sobre uma base que se move em arco com uma armação que sustenta a cabeça do paciente. O aparelho emite luz de alta intensidade ajustável para visualizar o fundo do olho. É semelhante a um oftalmoscópio indireto, mas a imagem pode ser mais ampliada.

- Retinografia: este exame representa uma fotografia digitalizada do fundo de olho adquirida usando um aparelho chamado retinógrafo de alta resolução e com ângulos de visão que geralmente variam de 30 a 50 °. O dispositivo consiste em uma câmera digital que capta imagens detalhadas do fundo de olho e as transfere para um computador ou similar. A técnica envolve um microscópio modificado acoplado a uma câmera que fornece uma visão vertical ampliada, como mostra a Figura 3.3. A retinografia pode ser usada para observar e documentar todos os tipos de doenças que envolvem o polo posterior do olho, proporcionando uma visão dos danos à camada de fibras da retina e região do disco óptico.

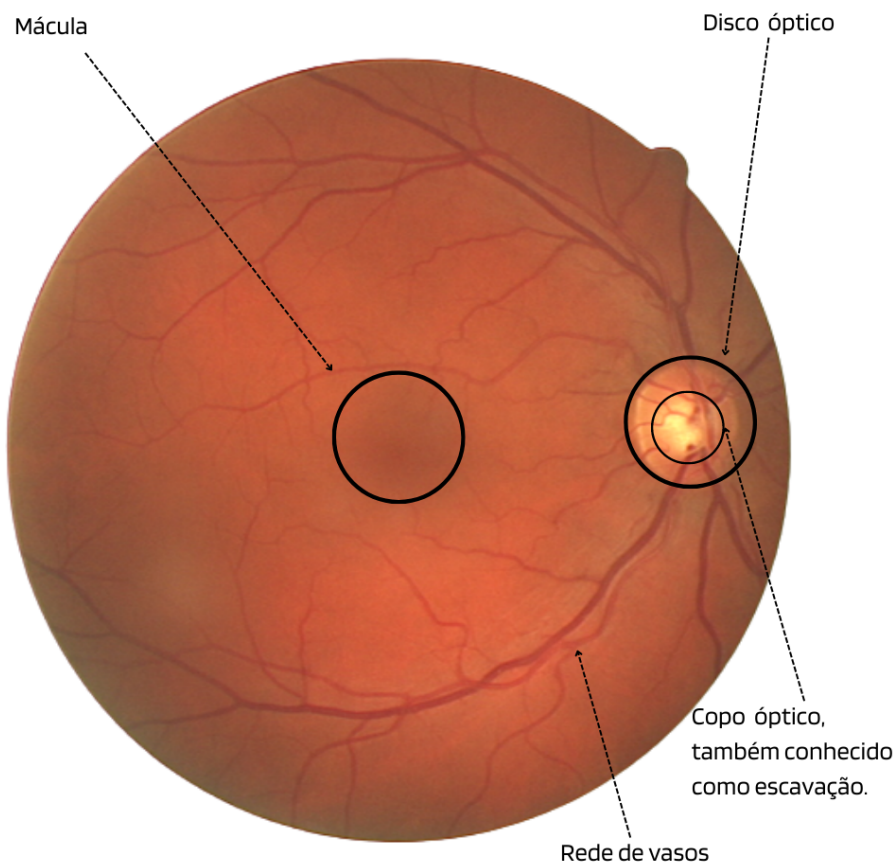


Figura 3.3: Exemplo das estruturas de fundo de olho - adaptado de (Budai et al., 2013).

Os aparelhos descritos são encontrados em quase todos os consultórios oftalmológicos e fornecem ao oftalmologista uma visão detalhada do globo ocular conforme mostrado

---

na Figura 3.3, onde a parte mais brilhante da retina representa o disco óptico, que, por sua vez, contém uma escavação também chamada de copo óptico, sendo a parte mais esbranquiçada do interior do disco. Assim, se o tamanho do copo óptico aumenta, ocorre o que é chamado de escavação óptica aumentada, considerado um dos principais indicadores de glaucoma (SBC, 2015).

## 3.2 Diagnóstico tradicional do glaucoma

Quanto aos métodos básicos e tradicionais de diagnóstico do glaucoma, além do exame de fundo de olho para examinar o disco óptico e a camada de fibras nervosas da retina (CFRN), os oftalmologistas geralmente usam tonometria, e testes de perimetria como auxiliares, a saber:

- Tonometria: é um exame para avaliar o grau de disfunção e medição da PIO em milímetros de mercúrio (mmHg). Pode ser aplicada pelo tonômetro de aplanção de *Goldmann*, considerado o método de medição mais utilizado atualmente, sendo uma referência para outros aparelhos similares. A faixa normal de pressão ocular é de 10 a 21 mmHg, que se baseia no nível médio de pressão ocular de uma pessoa normal. Embora o exame de tonometria seja muito importante no manejo e no tratamento do glaucoma, ele não pode ser considerado um diagnóstico devido à presença de casos de glaucoma de pressão normal. Outro fator que deve ser considerado na aplicação da tonometria é a medida da espessura da córnea por meio do teste de paquimetria, pois foi demonstrado que em córneas mais finas, a PIO pode ser subestimada (Oshika et al., 1997).
- Perimetria: por meio do exame de perimetria, campimetria ou campo visual, como também é conhecido, examina-se o grau de comprometimento funcional resultante da doença através dos resultados do mapa de campo visual obtido. Na prática clínica, o teste de campo visual é amplamente utilizado como padrão-ouro para avaliar se um paciente apresenta dano glaucomatoso funcional típico. O teste é feito com um aparelho de campo visual conhecido como perímetro ou campímetro. Existem vários tipos de perímetros, como *Humphrey Field Analyzers* ou *Oculus*. O exame consiste na quantificação do campo visual periférico e central, ou seja, é uma maneira de medir a visão de uma pessoa em cada olho e também o grau de perda de visão que pode ocorrer ao longo do tempo.

O teste de campo visual pode identificar pontos cegos chamados (escotomas) na visão de uma pessoa e a sua localização. É considerado um importante exame no cuidado oftalmológico regular para pessoas que correm o risco de cegueira. Sua principal indicação é o diagnóstico e o acompanhamento do tratamento do glaucoma.

---

Para aplicação, durante o teste são emitidos pontos de luz com intensidades variadas em posições fixas na esfera dentro do perímetro. Quando os pacientes veem os pontos, eles devem dar uma resposta pressionando um botão para que os clínicos recebam um relatório sobre a sensibilidade à luz em diferentes locais no campo visual. Finalmente, padrões de defeitos específicos, como degraus nasais e escotoma arqueado exibidos no campo visual, podem sugerir a presença de glaucoma (Li et al., 2018a).

Embora as características demográficas e clínicas associadas ao glaucoma sejam relativamente bem conhecidas, ainda não existe uma definição uniforme do diagnóstico desta doença pelos oftalmologistas. Por mais que muitos esforços internacionais tenham sido feitos para desenvolver tal definição, nenhum padrão de consenso real foi alcançado. Assim, o diagnóstico torna-se subjetivo e dependente da experiência médica, sendo geralmente incluídos como glaucomatosos, aqueles que apresentam PIO superior a 21 mmHg, acompanhado por dano característico no disco óptico ou por defeitos no campo visual compatível com glaucoma (Pachade et al., 2021; Garway-Heath, 2008).

### **3.2.1 Danos estruturais glaucomatosos**

Pela particularidade da abordagem diagnóstica do glaucoma, é importante documentar o aspecto do disco óptico como forma de avaliar possíveis danos estruturais causados pela doença, bem como acompanhar o paciente ao tratamento ou às consultas de rotina. Portanto, a partir de imagens de oftalmoscopia, os oftalmologistas podem avaliar pelo menos quatro características informativas importantes do glaucoma, a saber: relação escavação/disco; regra ISNT; assimetria de escavação e outros danos estruturas causados ao disco óptico.

#### **3.2.1.1 Relação escavação disco - CDR**

A escavação do disco é importante para auxiliar no diagnóstico do glaucoma, no entanto, muitas pessoas podem ter a escavação do nervo aumentada de tamanho e não necessariamente ter glaucoma. Isso é especialmente verdadeiro para pessoas míopes, que tendem a ter um nervo óptico maior, e conseqüentemente, uma escavação também maior. Deste modo, durante o diagnóstico do glaucoma, é importante que se avalie não apenas a escavação, mas sim a relação escavação/disco, em inglês *cup:disc ratio* (CDR).

Para melhor entendimento, a medida da CDR é calculada a partir da relação entre o *Vertical Cup Diameter* (VCD) e o *Vertical Disc Diameter* (VDD), como mostra a Figura 3.4.

Para calcular a relação CDR, deve-se considerar a divisão do disco óptico em 10 partes iguais, como na Figura 3.5, em seguida deve contabilizar o alcance da escavação

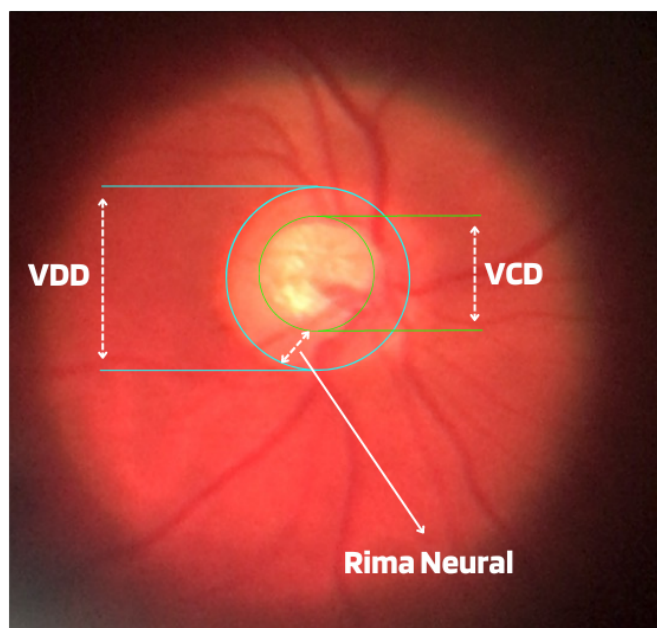


Figura 3.4: Medidas consideradas no cálculo da CDR.

em cada divisão efetuada. Deste modo, esta medida é aplicada como uma percentagem fracionária, geralmente feita horizontalmente e pode variar bastante entre indivíduos normais, no entanto, as escavações ópticas maiores que 0,65 indicam possíveis anormalidades, sendo sugestivas de uma investigação mais aprofundada (Murthi and Madheswaran, 2012; Schuster et al., 2020).

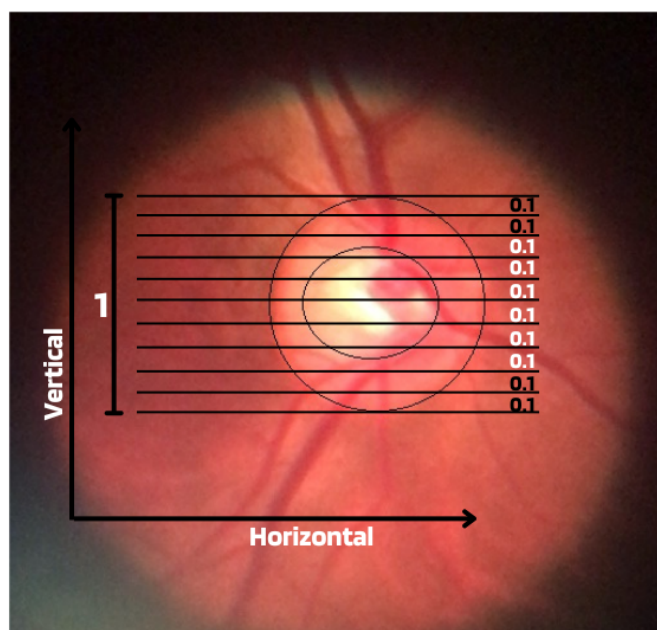


Figura 3.5: Exemplo de cálculo manual do valor de CDR com escavação de 0,6.

### 3.2.1.2 Regra ISNT

A borda formada entre o copo óptico e o disco óptico, chamada de anel neuroretiniano ou anel neural, também é considerada um indicativo de glaucoma, para o qual existe uma regra chamada ISNT, que faz alusão à orientação (inferior, superior, nasal e temporal) das bordas na imagem do fundo de olho, como apresentado na Figura 3.6 (Khalil et al., 2017; Budai et al., 2013).

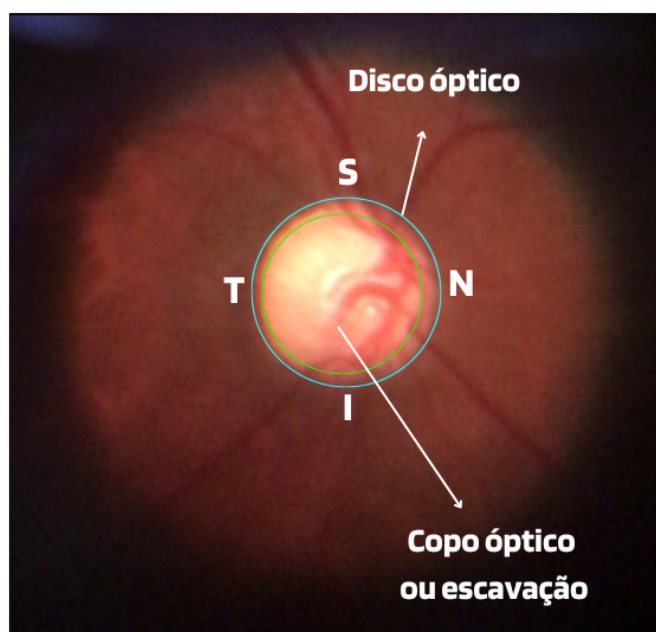


Figura 3.6: Exemplo dos diâmetros (VCD, VDD) e regra ISNT Inferior (I), Superior (S), Nasal (N) e Temporal (T).

Ao considerar a regra ISNT, em olhos não glaucomatosos, sugere-se que a espessura do anel neural deve ser maior no quadrante inferior, seguida pelo quadrante superior, nasal e temporal.

### 3.2.1.3 Assimetria da escavação

Avaliar a CDR é uma técnica que pode ser utilizada tanto para estimar o glaucoma diretamente quanto para examinar a assimetria de escavação entre os olhos de um mesmo paciente, considerada como outro indicativo do glaucoma. Esse cálculo é importante porque, se o glaucoma afetar apenas um olho, a escavação do nervo óptico pode não ser simétrica entre os dois olhos. Como resultado, o nervo óptico não afetado parecerá normal, enquanto o olho afetado parecerá com alterações em forma de concha. Para melhor entendimento, a CDR entre os olhos é simétrica na maioria das pessoas, e a assimetria é um sinal importante de suspeita de dano glaucomatoso. Isso se deve à observação de que 1% a 6% dos adultos normais podem apresentar uma discrepância de 0,2 na razão escavação/disco, enquanto 1% da população em geral pode apresentar uma assimetria de

0,3. Portanto, a assimetria da escavação ou copo óptico é um achado no exame oftalmológico que requer testes adicionais para descartar a presença de glaucoma ou outras possíveis complicações. Deste modo, durante o exame fundoscópico, a assimetria neural maior que 0,3 entre os olhos, também pode ser uma pista que pode ajudar a diagnosticar o glaucoma.

A Figura 3.7 demonstra um exemplo de assimetria entre os olhos, indicando a necessidade de uma avaliação mais aprofundada.

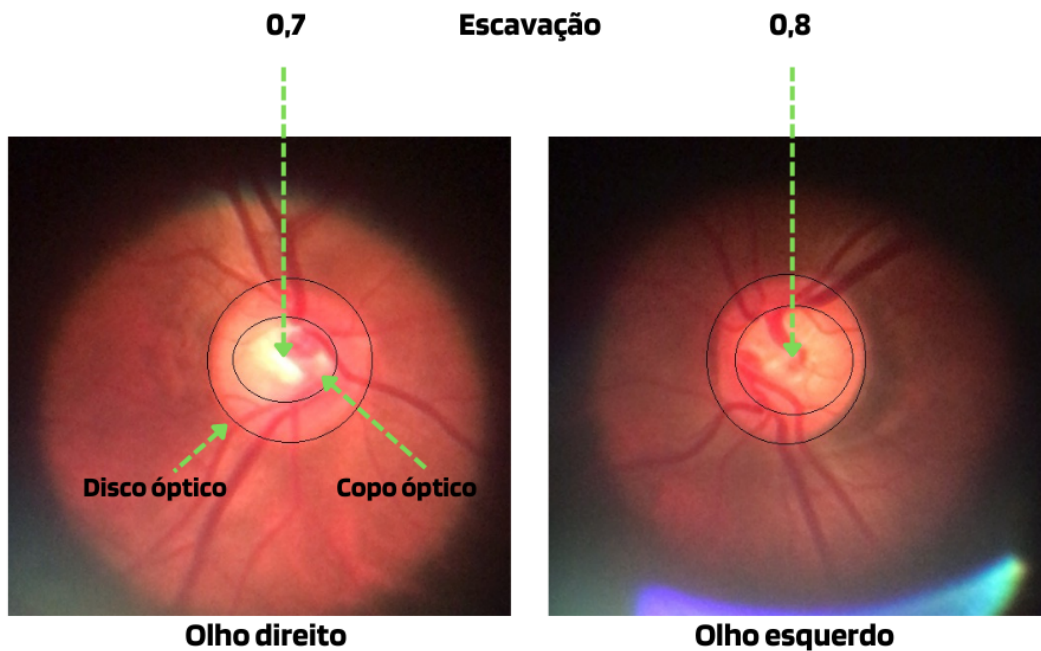


Figura 3.7: Exemplo de assimetria de escavação.

#### 3.2.1.4 Exemplos de outros danos estruturais do disco óptico

Conforme o exemplo mostrado na Figura 3.8, as principais descrições dos danos estruturais relacionados ao glaucoma são as seguintes:

- Alterações na CFNR: a presença de defeitos localizados na camada de fibras nervosas da retina é chamada de sinal de *hoyt* e caracteriza-se por uma zona escura, estendendo-se e alargando-se a partir do disco óptico, exibindo um formato arqueado. A Figura 3.8 (E) exemplifica essa característica glaucomatosa.
- Atrofia peripapilar: conforme aspecto oftalmológico, a atrofia peripapilar pode ser dividida em uma zona *alfa* periférica e uma zona *beta* central conforme mostrado na Figura 3.8(C). A zona *alfa* é caracterizada por hipopigmentação irregular e estreitamento das camadas de tecido coriorretiniano. Está lateralmente adjacente à retina e medialmente em contato com a área *beta* com esclera e grandes vasos coróides visíveis. Em olhos normais, as áreas *alfa* e *beta* estão geralmente localizadas na

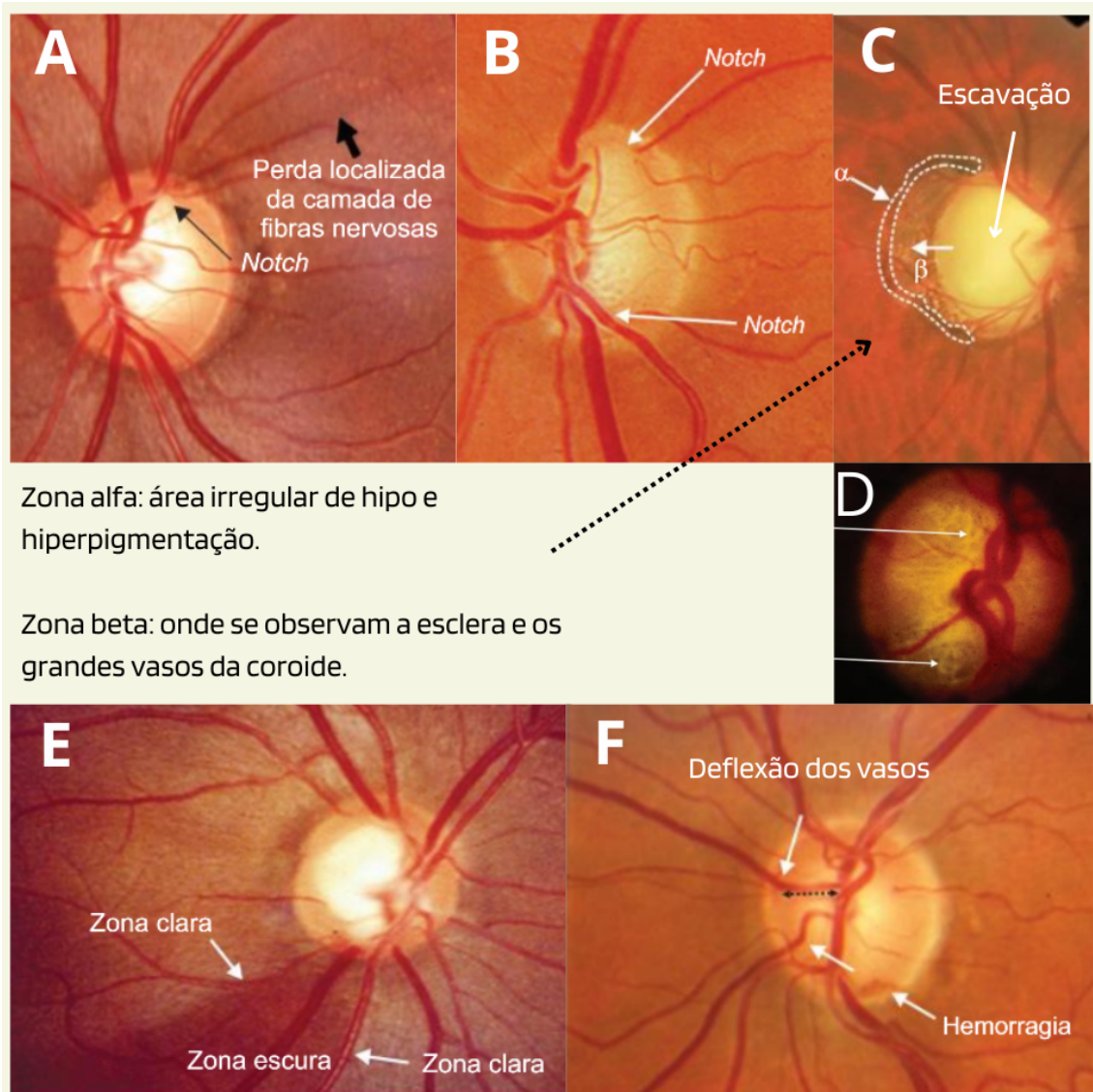


Figura 3.8: Principais características glaucomatosas observadas nas imagens de fundo de olho - adaptado de (Remo Susanna Jr and Filho, 2019).

área temporal, seguidas pelas áreas inferior e superior. Em olhos glaucomatosos, a área *beta* está mais presente na região temporal, e sua extensão está associada ao afinamento da CFNR.

- Escavação do disco óptico: além da escavação do disco representado na Figura 3.8(C), deve-se observar também o anel neuroretiniano ou rima neural, pois a escavação é influenciada pelo tamanho do disco óptico.
- Hemorragia de disco: a presença de hemorragias peripapilares é um sinal importante tanto no diagnóstico quanto no acompanhamento do glaucoma. Portanto, deve-se examinar a deflexão dos vasos e a escavação nasal. A Figura 3.8(F) apresenta um exemplo de hemorragia de disco.
- Desnudamento da lâmina, cribriforme: a presença de extinção visível da lâmina

---

cribriforme até a borda do disco óptico é chamada de *notch*, o qual representa a evolução de um defeito localizado na rima neural até que haja ausência completa de tecido na região, o que expõe a lâmina cribriforme e permitindo a visualização de seus poros, como apresentado nas imagens (A, B e D) da Figura 3.8. Embora seja muito sugestivo do glaucoma, esse sinal não é característico da doença (Remo Susanna Jr and Filho, 2019).

### 3.2.2 Dificuldades no diagnóstico precoce

Quanto às dificuldades associadas ao diagnóstico do glaucoma, considera-se que nos casos em estágios moderados ou avançados da doença, o diagnóstico costuma ser mais simplificado. No entanto, um grande desafio, é a melhor forma de detectar o glaucoma precoce, sendo primordial para um tratamento adequado, principalmente porque a qualidade de vida pode ser prejudicada mesmo com uma leve perda do campo visual (Tatham et al., 2014). Todavia, a tarefa de identificação precoce dessa doença, embora importante, pode ser desafiadora por vários motivos, entre eles, as características glaucomatosas que podem ser ambíguas na região do disco óptico, na CFNR ou nos resultados de campo visual no início da doença.

Devido a possíveis dificuldades no diagnóstico, o termo "suspeita de glaucoma" também é usado para denotar um indivíduo em quem o diagnóstico de glaucoma não pode ser confirmado, mas também não pode ser descartado no momento da consulta, exigindo avaliações futuras.

Ao longo dos anos, testes mais sensíveis foram desenvolvidos para identificar de forma mais confiável a perda precoce da função visual em pacientes com glaucoma, e dispositivos mais sofisticados de imagens foram criados para identificar os primeiros sinais de danos estruturais induzidos pela doença para auxiliar no diagnóstico precoce. Dentre esses aparelhos, destaca-se a tomografia de coerência óptica, em inglês, *Optical Coherence Tomography* (OCT), cujos resultados da sua avaliação são representados graficamente, o que também permite a comparação com dados normativos pareados por idade, auxiliando o oftalmologista a detectar doenças causadas por características diretamente observáveis no disco óptico.

A varredura OCT representa uma técnica de imagem axial não invasiva e de alta resolução que fornece aos médicos percepções únicas sobre a estrutura de retinas humanas normais e doentes. Além do exame de OCT propriamente dito, os aparelhos de OCT integram várias outras modalidades de imagem em um único instrumento. Os modos de operação incluem fundoscopia colorida, angiografia com fluoresceína, angiografia com indocianina verde, microperimetria e autofluorescência. Duas estratégias (domínio espacial e frequência de Fourier) podem ser usadas para detectar e analisar sinais ópticos. Recentemente, os métodos de OCT no domínio de Fourier demonstraram melhorias sig-

---

nificativas na velocidade sem perda de sensibilidade em comparação com as modalidades mais populares no domínio do tempo (Grulkowski et al., 2012).

Embora os exames de OCT tenham alta tecnologia e vantagens em relação ao uso de imagens de fundo de olho tradicionais, uma limitação considerável desses dispositivos de imagem é que eles geralmente estão disponíveis apenas em clínicas oftalmológicas especializadas ou hospitais, com pouca distribuição nas clínicas oftalmológicas mais gerais.

Outros importantes exames são a polarimetria de varredura a laser e a oftalmoscopia confocal de varredura a laser (Topouzis, 2007). Esses exames podem ser usados para fornecer medições quantitativas de parâmetros como espessura da CFNR ou medições da área marginal neuroretiniana, ambos importantes no diagnóstico e no acompanhamento de pacientes com glaucoma.

Ainda que os dispositivos tenham demonstrado uma boa capacidade de auxiliar os oftalmologistas a diagnosticar o glaucoma, poucos estudos examinaram especificamente o uso de tais tecnologias no início da doença, deixando permanecer o diagnóstico precoce do glaucoma como uma tarefa difícil para os especialistas, mesmo com auxílio de equipamentos sofisticados (Tatham et al., 2014).

Diante das dificuldades em diagnosticar precocemente o glaucoma, o que as clínicas oftalmológicas têm feito para tentar driblar essa dificuldade é a combinação de exames funcionais e estruturais. Posto que as alterações funcionais possam ser detectadas antes das alterações estruturais, em muitos casos a primeira manifestação detectável do glaucoma é uma anormalidade estrutural do disco óptico e CFNR, exigindo, portanto, que os testes sejam combinados para estabelecer os níveis de probabilidade da presença ou da ausência da doença (Tatham et al., 2014; Schuster et al., 2020; Pachade et al., 2021).

### **3.3 Epidemiologia**

Segundo a Organização Mundial da Saúde (OMS), pelo menos 2,2 bilhões de pessoas em todo o mundo, sofrem de algum tipo de deficiência visual. Em quase metade dos casos, esta deficiência poderia ter sido evitada ou ainda não foi tratada. Ao considerar esses dados, infere-se que hoje, milhões de pessoas vivem com deficiência visual ou cegueira que poderiam ter sido evitadas, mas, infelizmente, não foram. Embora o número exato seja desconhecido, estima-se que 11,9 milhões de pessoas em todo o mundo, tenham deficiência visual moderada ou grave, ou cegueira devido a doenças oculares como glaucoma, retinopatia diabética e tracoma.

A incapacidade visual e a cegueira podem ter um grande impacto na vida diária das pessoas afetadas por tais deficiências, pois a visão é o sentido dominante para os seres humanos em todas as fases da vida. No entanto, pesquisas estimam que até 2020, 76 milhões de pessoas em todo o mundo, terão glaucoma e, até 2030, esse número chegará a 95,4 milhões.

---

A deficiência visual, além de prejudicial à qualidade de vida dos pacientes, também apresenta um enorme encargo financeiro global, como demonstrado por pesquisas anteriores que estimaram os custos de perda de produtividade. Esses custos podem ser divididos em custos diretos e custos indiretos. Os custos diretos incluem medicamentos, cirurgias, consultas médicas, internações e exames complementares. Os custos médicos indiretos incluem principalmente os impactos econômicos provocados pela deficiência visual na produtividade do trabalho.

Embora o glaucoma geralmente progrida lentamente e tenha estimativas subdiagnósticas em todo o mundo, é a causa mais comum de cegueira irreversível globalmente, no entanto, evitável. A doença é considerada evitável porque, se detectada precocemente, há maneiras de controlá-la, mas as estatísticas mundiais mostram que há subdiagnóstico, resultando em inúmeras pessoas cegas. Este problema de subdiagnóstico pode ser ainda mais grave em países com baixa renda ou subdesenvolvidos. Por esse motivo, a Seção seguinte apresenta o cenário do glaucoma no Brasil, considerado pelo relatório de *World Inequality Lab* em 2018 (Vandekerckhove, 2018), como um dos países com maior desigualdade social e de renda do mundo, marcada por níveis extremos há muitos anos consecutivos.

O objetivo, portanto, não é usar o Brasil como um estudo de caso, mas como um exemplo cabal de como o glaucoma pode ser prejudicial quando combinado com deficiências de políticas públicas para combater essa neuropatia óptica. Antes de apresentar relativamente os problemas sobre o glaucoma, inicialmente descreve uma breve introdução ao modelo de saúde brasileiro.

### **3.4 Dificuldades de assistência oftalmológicas em países como o Brasil**

O Brasil pode ser um bom exemplo para mostrar o quanto pode ser difícil lidar com o problema de assistência oftalmológica e básica em países subdesenvolvidos. Além disso, é o único país com mais de 100 milhões de habitantes que conta com serviços públicos de saúde gratuitos para todos. Para essa garantia, o Brasil introduziu um sistema de seguridade social em 1988, embora a saúde também seja livre à iniciativa privada. Deste modo, o sistema público de saúde brasileiro se faz integralmente por meio de acesso ao programa criado pelo país e denominado Sistema Único de Saúde (SUS), sendo esse de acesso universal, financiado por meio de impostos e administrado pelos governos federal, estaduais e municipais. Os dados estatísticos mostram que 71,5% dos brasileiros usam exclusivamente o SUS para qualquer atendimento de saúde, ou seja, cerca de 161 milhões de brasileiros, enquanto 28,5% da população do país possui algum tipo de plano ou seguro de saúde privado (Scheffer et al., 2020).

O SUS abrange todos os 5.570 municípios do país, sendo que mais de um terço de

---

tais municípios não possui atendimento privado. Assim, segundo o cadastro nacional de estabelecimentos de saúde, 1.915 municípios, ou 34%, não oferecem serviços privados de saúde. Isso significa que aproximadamente 15,7 milhões de brasileiros nessas regiões são totalmente dependentes do SUS. Entre os municípios sem disponibilizar consultórios oftalmológicos particulares, a menor região tem aproximadamente 800 habitantes, enquanto a maior região tem 70.000 habitantes.

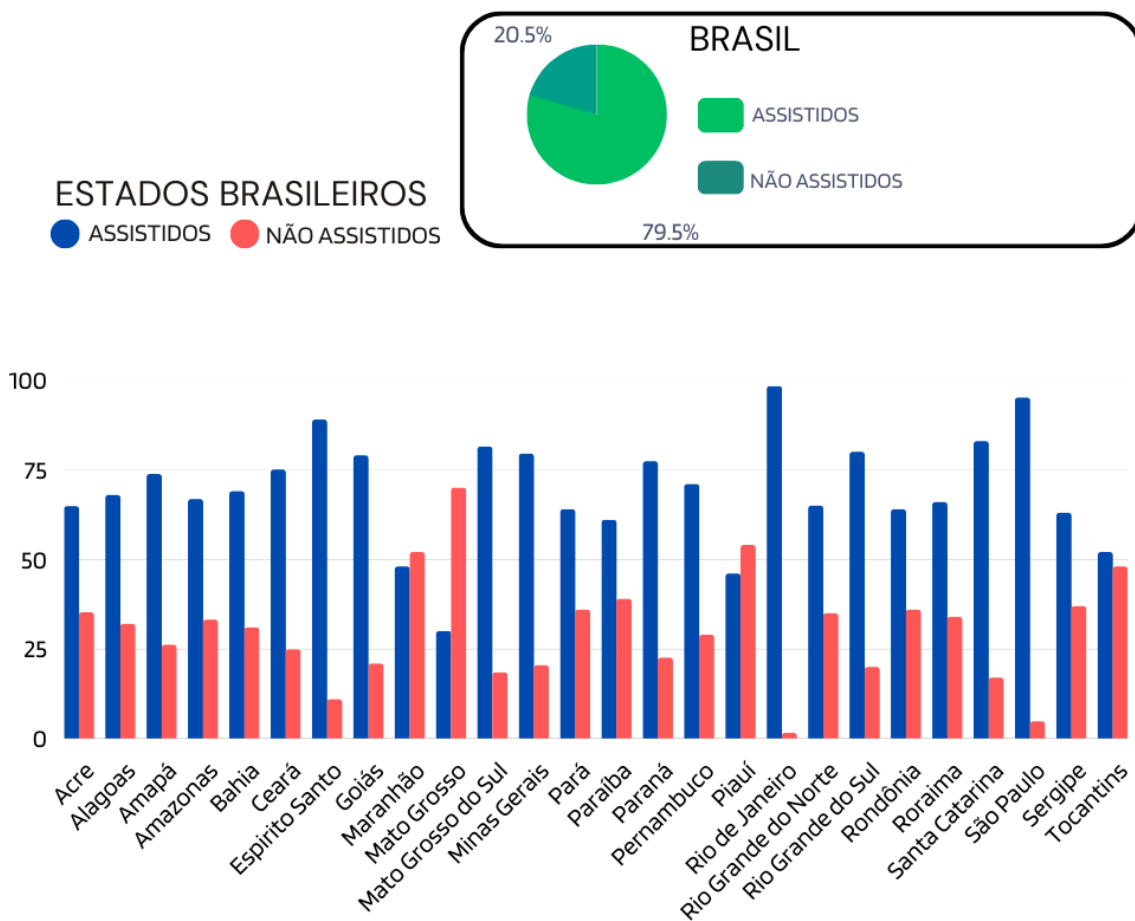
Considerando as informações estatísticas, ao desagregar os dados pelo número de residentes sem registros privados nesses municípios, o Nordeste foi à região mais dependente dos serviços do SUS: 8,9 milhões de pessoas viviam em municípios sem serviços médicos privados. No outro extremo está o Centro-Oeste, onde mais de 796 mil habitantes vivem em cidades totalmente dependentes do SUS para atendimento de saúde.

Embora uma parcela da população do país tenha plano de saúde, todos os brasileiros têm direito a assistência médica gratuita desde o nascimento, pois, no Brasil, a saúde é direito de todos e responsabilidade do Estado. No entanto, enquanto as pessoas têm direito à saúde fornecida pelo sistema público, faltam recursos e ações para oferecê-la com qualidade todos os brasileiros.

O SUS é considerado único por ter a mesma filosofia de atuação em todo o país e se organizar segundo a mesma lógica, mas, nem sempre essa igualdade é alcançada. Veja, por exemplo, o cenário da distribuição dos médicos oftalmologistas nos Estados brasileiros, mostrando que em algumas partes do Brasil a proporção de "oftalmologistas per capita" é mais que o dobro da recomendada pela OMS, para países desenvolvidos, correspondendo um oftalmologista para cada 17 mil pessoas (1:17.000). No entanto, algumas regiões mais distantes dos grandes e médios centros urbanos, existe uma escassez destes profissionais para atuarem no sistema público (Umbelino, 2021).

O cenário atual mostra que o número de municípios que conta com a presença de oftalmologistas representa menos de um terço do número total. No entanto, eles somam 168,4 milhões de habitantes assistidos, representando 79,5% da população do país. Os outros 43,3 milhões de habitantes não assistidos ou com necessidade de deslocamento regional para atendimento, estão distribuídos em aproximadamente 3.880 municípios, a maioria pequenos (Umbelino, 2021).

Para melhor exemplificar, a Figura 3.9 mostra uma comparação entre a população que recebe atendimento oftalmológico e a população que não recebe atendimento oftalmológico direto ou tem maiores dificuldades de acesso. Sobre a questão da distribuição desigual de serviços médicos no Brasil, a Figura 3.10 mostra a distribuição da proporção de especialistas por 100.000 habitantes, considerando oftalmologistas e outros profissionais médicos responsáveis por importantes áreas de prevenção de doenças, atendimento e acompanhamento de pacientes de todas as idades. Os dados representados na imagem mostram que, além do déficit oftalmológico em diversas regiões do Brasil, esse déficit também é frequente nas áreas mais básicas e importantes da saúde.



Distrito Federal - Brasília, apresenta uma densidade de um profissional para cada 4.168 residentes

Figura 3.9: Gráfico de cobertura oftalmológica do Brasil e Estados brasileiros - adaptado de (Umbelino, 2021).

Para números comparativos, a OMS não possui referências clínicas específicas, então o governo federal usa a proporção do Reino Unido (2,7 médicos por 1.000 pessoas) como referência. Em termos de números, em 2020, o Brasil passou a contar com mais de 500 mil médicos, contabilizando, portanto, 2,38 médicos para cada 1.000 habitantes, porém, a má distribuição desses profissionais nas regiões brasileiras consiste em um frequente problema na cobertura do SUS (Scheffer et al., 2020; Vandekerckhove, 2018).

Quanto à situação do glaucoma no Brasil, é impossível mensurar se a distribuição desigual de oftalmologistas está relacionada ao subdiagnóstico dessa doença no país, até porque, em geral, segundo estatísticas mundiais, metade ou mais dos pacientes com glaucoma permanecem sem diagnóstico, ou seja, as estatísticas vão além do território do Brasil.

Conforme o contexto observa-se que o glaucoma é uma doença subdiagnosticada na

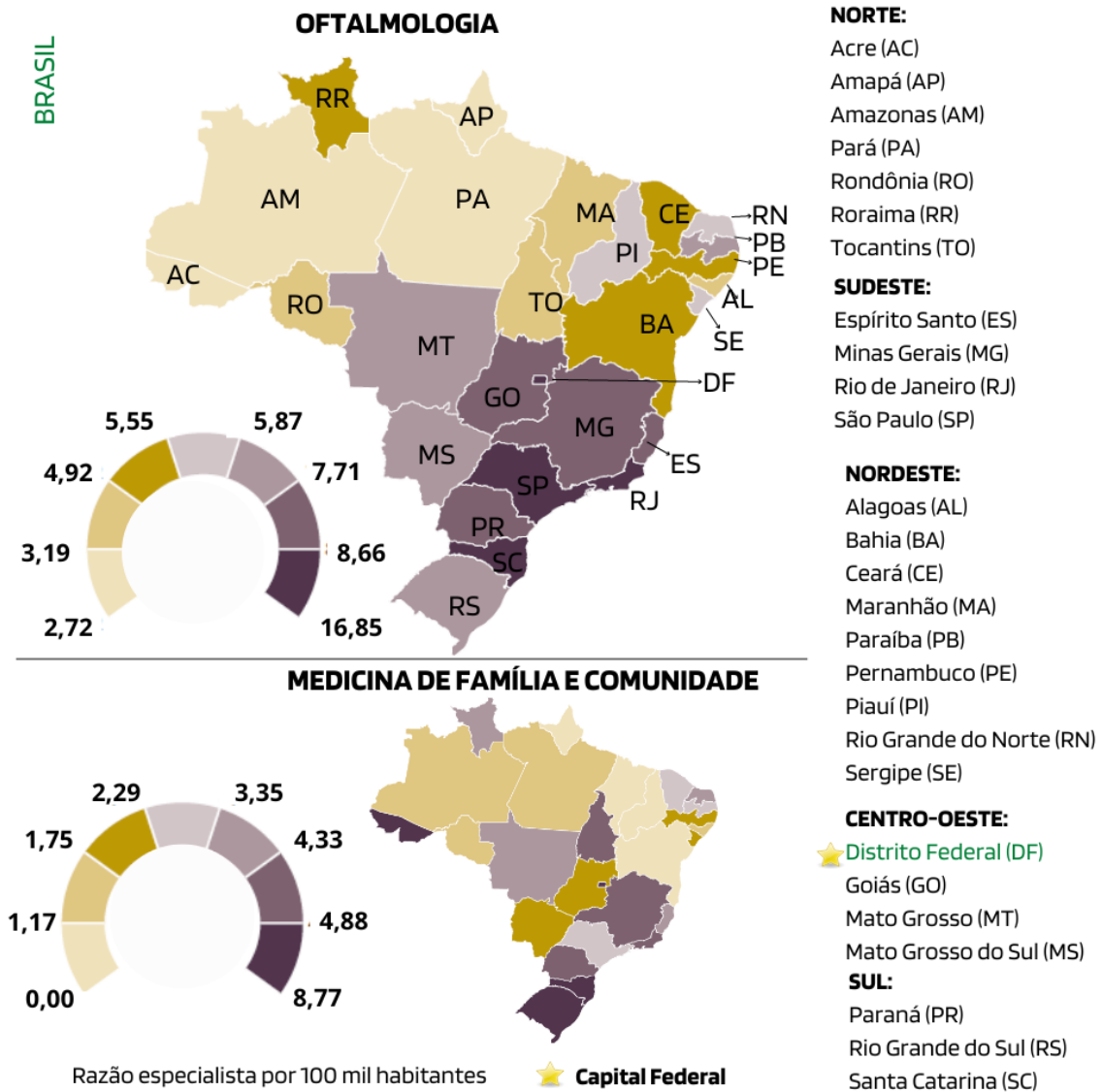


Figura 3.10: Relação médico por habitantes no Brasil - adaptado de (Scheffer et al., 2020).

maior parte do mundo, mas cada país é responsável por sua população e pode contribuir para a redução dos casos subdiagnosticados, no entanto, pensando no Brasil, a distribuição desigual dos profissionais oftalmologistas no mínimo pode contribuir em manter essa alta estatística negativa de subdiagnóstico do glaucoma no país.

Quanto à prevalência dessa neuropatia óptica no Brasil, algumas regiões apresentam dados populacionais insuficientes. Assim, a prevalência da deficiência visual no país não pode ser estimada com segurança, sobretudo, estima-se que a incidência de glaucoma seja de 1 a 2% na população geral, aumentando para 2-3% após os 40 anos, dos quais até aproximadamente 60% dos casos são referidos como glaucoma primário de ângulo aberto (Umbelino, 2021). A prevalência do glaucoma aumenta com a idade, no Brasil, estima-se que chegue a 6% ou 7% após os 70 anos. Geralmente, a doença é bilateral com característica genética que aumenta em 10 vezes a chance de um parente de primeiro grau ter a doença. Além disso, dados estatísticos apontam que a população idosa brasileira

---

(acima de 60 anos) vai dobrar até 2042, e isso pode contribuir para aumentar o número de pessoas glaucomatosas, exigindo que o país desenvolva políticas de saúde pública para abordar essa possibilidade (Ottaiano, José Augusto Alves Ávila and Umbelino, Cristiano Caixeta Taleb, 2019; Umbelino, 2021; Quigley and Broman, 2006).

Em resumo, é possível destacar de um contexto geral de assistência à saúde oftalmológica e básica, que não faltam médicos no Brasil, porém, uma fragilidade do SUS é que esses profissionais não estão distribuídos de forma igualitária e não conseguem cobrir todo o território brasileiro com qualidade. Destaca-se também que maioria dos cidadãos necessita do uso do SUS. Essa realidade demonstra claramente que o sistema de saúde brasileiro enfrenta problemas de acessibilidade, principalmente no que diz respeito a programas preventivos como consultas oftalmológicas de rotina. Contudo, a acessibilidade à saúde pode ser ainda mais crítico nos locais onde não se tem a opção de clínicas e hospitais particulares, ou seja, aproximadamente 34% da população geral.

Quanto a outras dificuldades no combate ao glaucoma, também é importante abordar a desigualdade social que impede que todos tenham acesso à saúde com qualidade. No Brasil existem centenas de centros privados de referência para pacientes glaucomatosos, no entanto, uma parcela da sociedade pode não ter condições de arcar com tais tratamentos particulares. Quanto aos serviços oferecidos pelo SUS, em algumas regiões do Brasil, também existem programas com excelência no combate ao glaucoma, como o hospital de olhos do Sul de Minas Gerais, apresentado no Capítulo 5 desta tese, onde os pacientes recebem consultas especializadas e exames complementares, além de colírios terapêuticos no âmbito do SUS. Estima-se que 140.000 pacientes foram registrados nesses programas até 2020, no entanto, tais programas de saúde não são descentralizados para atender igualmente todos os pacientes diagnosticados com glaucoma ou aqueles que requerem consultas e exames comprobatórios. Portanto, novamente a discrepância da distribuição de serviços de saúde em todos os municípios do Brasil, dificulta a aplicabilidade clínica. Dessa forma, esses serviços apresentam excelência em certas regiões do país, enquanto falta atendimento básico em outras regiões mais afastadas dos centros urbanos.

Diante do contexto, no processo de prevenção da deficiência visual induzida pelo glaucoma, o Brasil tem investido em campanhas para conscientizar a população sobre a gravidade da doença, com isso, foram criadas campanhas como o "abril marrom" dedicada ao combate, prevenção e reabilitação de todos os tipos de cegueira, chamando a atenção da população geral para uma das principais causas de cegueira no Brasil e no mundo.

Por fim, o sistema de saúde do Brasil foi apresentado como uma forma de demonstrar algumas das dificuldades que os governos podem enfrentar ao prover saúde pública. Este macro olhar sobre o Brasil, em especial ao tratamento e prevenção da doença ocular glaucoma, é uma forma de elucidar o quanto é difícil combater essa doença, pois isso pode envolver uma diversidade de problemas e prioridades de soluções. Portanto,

---

com base nos dados atuais apresentados, fica claro que é necessário e possível incorporar novas tecnologias para ajudar os profissionais da medicina oftalmológica a combater doenças graves como o glaucoma, principalmente em cenários como o apresentado. Evocando, tomou-se como exemplo o caso brasileiro, porém, condições semelhantes ou até mais graves podem existir em vários outros países, indicando a importância de avançar em pesquisas que busquem soluções para o diagnóstico precoce do glaucoma, para ser posteriormente tratado.

### 3.5 Avanços científicos e tecnológicos

Nos últimos anos, avanços científicos e tecnológicos abriram uma vasta gama de oportunidades clínicas e de pesquisas no campo dos cuidados oftalmológicos, o que pode auxiliar no combate ao glaucoma. Com esses avanços, tecnologias como *deep learning* provaram eficácia em áreas da medicina como radiologia, patologia, dermatologia, etc. Todos esses estudos são áreas afins que compartilham paralelos com a oftalmologia devido às suas raízes profundas no diagnóstico por imagem (Schmidt-Erfurth et al., 2018).

Considerando o campo geral da medicina e técnicas de processamento de imagens digitais para visão computacional e reconhecimento de padrões de diversas doenças. Aponta-se, a seguir, as principais modalidades de exames que podem fornecer dados de entradas para treinamento de algoritmos de *deep learning*:

- Ultrassom e raios-X; Tomografia computadorizada; Ressonância magnética;
- Fotografia da retina ou retinografia; Imagens de dermatoscopia, etc.

Todos esses tipos de imagens digitais são passíveis de aplicações de reconhecimento de padrões por meio de algoritmos de IA. No entanto, como apresentado no Capítulo anterior, uma das grandes deficiências nessa área é obtenção de grandes bases de dados, rotuladas e disponíveis publicamente para pesquisas. Entretanto, embora essa seja uma dificuldade, confia-se na evolução dos algoritmos de IA, uma vez que muitas novas tecnologias estarão disponíveis em breve, pois, o sucesso dos algoritmos de *machine learning* em tarefas de visão computacional nos últimos anos coincidiu com a crescente digitalização de registros médicos (Ker et al., 2017).

Quanto ao cenário oftalmológico do glaucoma, o uso de IA surge como uma ferramenta auxiliar no diagnóstico da doença através da detecção de alterações presentes, principalmente em resultados de OCT, resultados de exames de campo visual e imagens de fundo de olho. Estes tipos de imagens são exemplificados na Figura 3.11 (Han, 2022). Com isso, existem vários trabalhos direcionados a este tema, onde os estudos buscam o uso de algoritmos inteligentes na aplicação da automação do reconhecimento de padrões

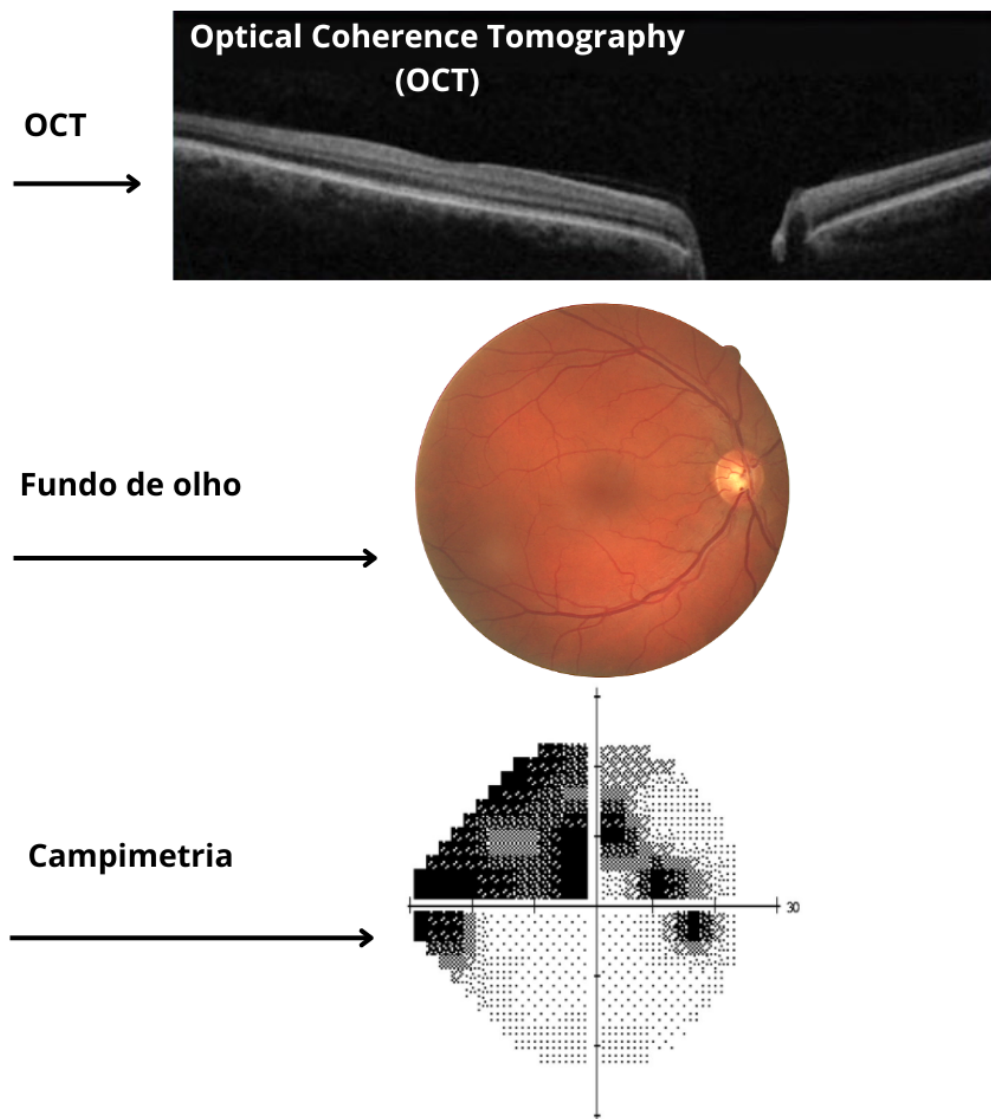


Figura 3.11: Exemplo de imagens médicas oftalmológicas passíveis de aplicações com algoritmos de IA.

do glaucoma, ou comumente referido como classificação do glaucoma a partir de imagens médicas oftalmológicas.

Apesar do potencial para aplicar a automação a diferentes tipos de imagens oftálmicas, as imagens de fundo (ou seja, imagens obtidas com equipamentos oftalmológicos convencionais) ganharam destaque em muitos trabalhos relacionados devido à disponibilidade, qualidade e custo-efetividade da aquisição. Portanto, a Seção seguinte apresenta o estado da arte em processamento de imagens digitais, ou seja, como os algoritmos de inteligência artificial propostos em trabalhos relacionados abordam o tópico de processamento de imagens e reconhecimento de padrões do glaucoma.

Quanto a aplicação de *deep learning* na oftalmologia, além dos estudos voltados para o diagnóstico automático do glaucoma, esta tecnologia geralmente se concentra em estudos sobre o diagnóstico de doenças, como catarata, degeneração macular relacionada à idade,

---

retinopatia diabética, a saber:

- **Degeneração macular relacionada com a idade:** causa lesão progressiva da mácula e também da área central mais vital da retina, tendo como consequência o aparecimento de manchas escuras, sombras ou distorção da visão. É uma doença que causa perda de visão, e o seu risco de desenvolvimento aumenta com o avanço da idade.
- **Catarata:** opacificação da lente ocular, causando visão cada vez mais turva. O risco de desenvolver catarata aumenta com a idade.
- **Retinopatia diabética:** este distúrbio na retina é devido a alterações nos vasos sanguíneos causadas pelo diabetes. À medida que a concentração de glicose no sangue continua a aumentar, as veias e artérias se estreitam e as paredes dos vasos sanguíneos se tornam mais fracas e mais propensas a dilatar e, posteriormente, romper. Vasos sanguíneos anormais também podem crescer a partir da retina, o que pode causar sangramento ou cicatrização, levando à cegueira.

## 3.6 Trabalhos relacionados

Para elaborar esta revisão, os manuscritos foram selecionados com base nos títulos e resumos dos métodos de IA utilizados para a classificação do glaucoma a partir de imagens digitais do fundo de olho. Deste modo, apresenta-se aqui, alguns dos relevantes trabalhos científicos publicados nos últimos anos. A busca pelos artigos foi aplicada por palavras-chave relacionadas ao tema nas principais plataformas de dados (*Scopus, Web of Science, Google Acadêmico, Scielo e Medline*). Devido ao escopo deste estudo, a busca limitou-se a algoritmos desenvolvidos para analisar imagens digitais do fundo de olho e principalmente com auxílio de algoritmos CNNs.

Antes de apresentar sobre métodos usando *deep learning*, descrevem-se as principais bases de dados públicas contendo imagens de fundo de olho utilizadas por muitos dos trabalhos relacionados descritos como instâncias para treinamentos e testes dos classificadores, que são na sua maioria supervisionados.

### 3.6.1 Principais bases de dados públicas

Quanto aos conjuntos de dados públicos e rotulados para glaucoma, estes foram compostos com imagens de vários pacientes de diferentes nacionalidades em diferentes locais com câmeras que podem variar em tamanho, concentração de foco e delineamento de regiões de interesse. Esses conjuntos de dados são construídos a partir de um dos exames mais populares em oftalmologia, o exame de fundo de olho. Portanto, tornou-se mais economicamente viável e acessível, à aplicação de pesquisas voltadas para a análise de

---

imagens digitais fundoscópicas, uma vez que esses dados são amplamente populares e de grande utilidade oftalmológica.

São descritas a seguir algumas bases de dados encontradas publicamente para trabalhos voltados para a classificação de glaucoma com uso de deep learning e imagens digitais do fundo de olho, obtidas por meio da retinografia com câmeras convencionais, as quais fornecem de 30 a 50° de visualizações da retina:

**ACRIMA:** criada pela *Spanish Ministerial Organization for Economy and Competition*. O conjunto de dados consiste em 396 imagens de glaucoma e 309 imagens normais, para um total de 705 imagens adquiridas com uma câmera retinal Topcon TRC configurada para um campo de visão de 35°. Dois especialistas em glaucoma rotularam todas as imagens deste conjunto de dados, a base de dados não inclui máscaras da região do disco óptico e copo óptico para algoritmos de segmentação (Diaz-Pinto et al., 2019a).

**DRIONs:** é uma base de dados pública para a avaliação comparativa da segmentação da cabeça do nervo óptico a partir de imagens retinianas digitais. As imagens foram adquiridas no Hospital Miguel Servet, no serviço de oftalmologia de Saragoça, Espanha. O conjunto de dados compreende 110 imagens de fundo de olho (55 saudáveis e 55 com glaucoma). Todas as imagens foram obtidas de indivíduos caucasianos usando uma câmera de fundo analógica colorida convencional centrada na região do disco óptico e armazenada em formato de *slides*. Posteriormente, as imagens foram digitalizadas em *scanner HP-PhotoSmart-S20* de alta resolução e salvas no tamanho de  $600 \times 400$  *pixels* (Carmona et al., 2008).

**DRISHTI-GS:** é um conjunto de dados de imagem da retina adquirido no Aravind Eye hospital em Madurai, Índia, para segmentação da cabeça do nervo óptico. Apresenta um conjunto de imagens da retina para avaliação e classificação do disco óptico em olhos normais e glaucomatosos, com máscara de segmentação manual realizada por especialistas, o que permite realizar medições da relação escavação / disco. Uma extensão do DRISHTI-GS chamado DRISHTI-GS1 consiste em 50 imagens de treinamento e 51 imagens de teste. As imagens foram capturadas usando uma câmera tradicional de alta resolução com 30° de visão do fundo de olho e um tamanho de  $2896 \times 1944$  *pixels* (Sivaswamy et al., 2014).

**DRIVE:** essas imagens de fundo de olho foram obtidas para a extração de vasos para um projeto de pesquisa de triagem oftalmológica na Holanda. O conjunto de dados inclui 40 imagens de fundo (34 saudáveis e 6 com glaucoma) anotadas por dois oftalmologistas. As imagens foram obtidas com uma câmera não midriática Canon CR5 3CCD com um campo de visão de 45° e dimensões de  $565 \times 584$  *pixels* (Staal et al., 2004).

**GLAUCOMADB:** nenhum registro de localização foi obtido para o conjunto de dados. Consiste em 120 imagens de fundo (85 imagens saudáveis e 35 imagens com glaucoma) de um conjunto maior de 462 imagens. Os rótulos de glaucoma foram aplicados por dois oftalmologistas. As imagens foram obtidas com uma câmera TopCon TRC 50EX

---

com resolução de  $11504 \times 1000$  pixels (Ramani and Shanthamalar, 2020).

HRF: as imagens foram coletadas em uma única clínica na República Tcheca. De um total de 45 imagens, 15 são glaucomatosas, 15 normais e 15 são rotuladas como retinopatia diabética. Todas as imagens do fundo foram adquiridas com uma câmera midriática Canon CR-1 e diferentes configurações de aquisição com um campo de visão de  $45^\circ$  e  $3504 \times 2336$  pixels. Não há registros de quantos oftalmologistas foram usados para rotular as imagens (Budai et al., 2013).

O banco de dados Sjchoi86-HRF (Abbas, 2017) possui 601 imagens de fundo de olho em 4 subconjuntos: normal (300 imagens); glaucoma (101 imagens); catarata (100 imagens); e doença da retina (100 imagens). As imagens foram capturadas usando 8 bits por plano de cores em  $2592 \times 1728$ ,  $2464 \times 1632$  ou  $1848 \times 1224$  pixels.

MESSIDOR: as imagens foram adquiridas por três departamentos oftalmológicos na França. O conjunto de dados contém um total de 1200 imagens de diferentes doenças, mas apenas 100 imagens são rotuladas para glaucoma (28 com glaucoma e 72 normais). Foi criado pelo projeto Messidor para avaliar o direcionamento automático de lesões retinianas e retinopatia diabética. As imagens foram adquiridas usando uma câmera não midriática Topcon TRC NW6 com um campo de visão de  $45^\circ$  e tamanhos de  $1440 \times 960$ ,  $2240 \times 1488$  ou  $2304 \times 1536$  pixels (Decencière et al., 2014).

ORIGA: este conjunto de dados tem um total de 650 imagens de fundo divididas em 168 imagens de glaucoma e 482 imagens normais. Foi construído usando dados de imagem da retina coletados do estudo do olho malaio de Cingapura em conjunto com o *Singapore Eye Institute*. Estatísticas relacionadas ao disco (como defeitos de CDR e RNFL compatíveis com ISNT) e resultados segmentados manualmente para discos ópticos e copos ópticos são fornecidos para cada imagem (Zhang et al., 2010).

PAPILA: essas imagens foram adquiridas por oftalmologistas ou técnicos da Unidade do Hospital Universitário Reina Sofía, Espanha, usando uma câmera retiniana Topcon TRC NW400 não midriática com resolução de  $2576 \times 1934$  pixels. O conjunto de dados que envolveu 244 pacientes, fornece um total de 488 imagens divididas em 333 saudáveis e 155 com glaucoma ou suspeita de glaucoma. A rotulagem para glaucoma foi baseada em dados clínicos (Kovalyk et al., 2022).

REFUGE: este conjunto de dados contém 1200 imagens divididas em 120 imagens de pacientes com glaucoma e 1080 de pacientes saudáveis. A aquisição das imagens foi realizada com duas câmeras retinianas (uma câmera de fundo Zeiss Visucam 500 e uma câmera Canon CR-2 com resolução de  $2124 \times 2056$  e  $1634 \times 1634$  pixels, respectivamente). Este conjunto de dados também fornece informações sobre o disco óptico e o copo óptico preparados por sete especialistas em glaucoma no *Sun Yat-Sen Eye Center* (localizado em Guangzhou, província de Guangdong, China) (Orlando et al., 2020).

RIM-ONE DL: o conjunto de dados RIM-ONE-DL (Batista et al., 2020) foi criado em 2020 combinando três versões abertas do conjunto de imagens do fundo do olho chamado

---

banco de dados de imagens da retina para avaliação do nervo óptico (RIM-ONE). A primeira versão aberta do conjunto de imagens de fundo de olho RIM-ONE foi publicada em 2011 por (Fumero et al., 2011). A segunda versão publicada em 2014 foi concebida como uma extensão da primeira; posteriormente, em 2015 (Fumero et al., 2015) publicou a terceira versão do conjunto de dados. Eles são referidos neste documento como RIM-ONE v1, v2 e v3, respectivamente. O conjunto de dados RIM-ONE-DL final consiste em 313 imagens de indivíduos saudáveis e 172 imagens de pacientes com glaucoma. Como o conjunto de dados foi criado recentemente, a maioria dos estudos acadêmicos anteriores são baseados em imagens RIM-ONE (v1, v2 ou v3). Portanto, eles são descritos a seguir.

1. RIM-ONE v1. O principal objetivo deste estudo em 2011 foi fornecer um banco de dados de retinografias de 118 indivíduos saudáveis e 51 pacientes classificados em vários estágios de glaucoma. As imagens do fundo foram adquiridas usando uma câmera não midriática Nidek AFC-210 com um corpo Canon EOS 5D Mark II com um campo de visão de 45°.
2. RIM-ONE v2. Contém 255 imagens de indivíduos saudáveis e 200 imagens de pacientes com glaucoma. É uma extensão da primeira versão e apresenta imagens segmentadas manualmente por um médico especialista. As imagens foram tiradas no HUC e HUMS usando a mesma câmera da versão 1.
3. RIM-ONE v3. Contém 85 imagens de indivíduos saudáveis e 74 imagens de pacientes com glaucoma. As imagens foram capturadas apenas no HUC com uma câmera de fundo de olho Kowa WX 3D não midriática com resolução total de 2144 × 1424 *pixels*.

G1020: as imagens do G1020 foram coletadas em uma clínica particular em Kaiserlautern, Alemanha, entre 2005 e 2017. As imagens foram adquiridas com um campo de visão de 45° usando midríase. O conjunto de dados contém 1.020 imagens de fundo de olho disponíveis publicamente (724 saudáveis e 296 com glaucoma). A rotulagem das imagens é fornecida, bem como a segmentação do disco óptico e do copo óptico. No conjunto de dados final, as imagens têm tamanhos entre 1944 × 2108 e 2426 × 3007 *pixels* (Bajwa et al., 2020).

BrG: as imagens BrG, foram coletadas na região sudeste do Brasil em dois locais distintos (hospital de olhos do sul de Minas gerais e centro de atenção básica policlínica da cidade de Unai, Minas Gerais. Este conjunto de dados conta com 1.000 fotografias de fundo de olho de pacientes com glaucoma e 1.000 fotografias de fundo de olho de voluntários não glaucomatosos. Todas as imagens deste conjunto de dados apresentam ângulo de visão de até 25° e foram coletadas com auxílio de um smartphone acoplado a um oftalmoscópio *panoptic Welch Allyn*. Para todas as imagens, os autores apresentam

máscaras de segmentação do disco óptico e do copo óptico, salvas no formato *portable network graphic* (PNG) (Bragança et al., 2022).

Outras bases de dados utilizadas na literatura incluem a base Singapore Chinese Eye Study (SCES) (Fu et al., 2018), composta por 1.676 imagens com 46 casos de glaucoma, e a Singapore Indian Eye Study (SINDI), composta por 5.783 imagens de 113 olhos com glaucoma.

A Tabela apresenta os conjuntos de dados publicamente disponíveis, levando em consideração somente as imagens rotuladas em "sim" ou "não" para o glaucoma, apresentando capacidade de serem utilizadas por algoritmos de deep learning para a classificação dessa doença.

Tabela 3.1: Bases de dados e número de imagens glaucomatosas

Dataset	Glaucoma	Normal	Total	Camera	Ângulo
Acrima	396	309	705	Topcon TRC	35
Drions	55	55	110	Câmera analógica	-
Drishti-gs	50	51	101	Câmera alta resolução	30
Drive	6	34	40	Canon CR5	45
GlaucomaDB	35	85	120	TopCon TRC 50EX	-
Hrf	15	15	30	Canon CR-1	45
Sjchoi86-hrf	101	300	401	Câmera alta resolução	-
Messidor	28	72	100	Topcon TRC NW6	45
Origa	168	482	650	Câmera alta resolução	-
Papila	155	333	488	Topcon TRC NW400	-
Refuge	120	1080	1200	Zeiss Visucam 500 e Canon CR-2	-
Rim-one- dl	172	313	485	Nidek AFC-210 e Kowa WX 3D	45
G1020	296	724	1020	Câmera alta resolução	45
BrG	1000	1000	2000	Smartphone / Panoptic	25
Sces	46	1641	1676	Câmera alta resolução	-
Sind	113	5670	5.783	Câmera alta resolução	-

Embora os conjuntos de dados públicos sejam apresentados aqui, as maiores coleções de imagens existentes são de propriedade privada, no entanto, quanto a esses dados, não foi possível catalogá-los devido à inacessibilidade de suas descrições, que se limita a hospitais ou universidades que mantêm tais bancos de dados. No entanto, alguns trabalhos científicos que os utilizam são apresentados a seguir, juntamente com outros trabalhos que utilizaram bases de dados públicas.

### 3.6.2 Abordagens utilizando deep learning

Devido às propriedades referentes aos tipos de glaucoma, a grande maioria dos trabalhos relacionados desenvolveram algoritmos de IA para classificação do glaucoma primário de

---

ângulo aberto, por esta ser considerada a forma mais comum da doença, além de ser mais propícia a aceitar modelos estudos como o proposto (King et al.,2013). Essa tipificação de glaucoma geralmente não apresenta sintomas porque o dano progride muito lentamente, apresentando, portanto, desafios para o diagnóstico da enfermidade, levando a atrasos consideráveis na sua detecção e início do tratamento.

Com base na análise da literatura que compõe os estudos relacionados, foi observado que os modelos de IA comumente utilizados nos estudos desta doença a partir de imagens digitais do fundo de olho são geralmente aplicados em duas formas específicas, cálculo da CDR ou identificação de padrões de glaucoma na região do disco óptico, a saber:

- **Cálculo do valor CDR:** uma das formas utilizadas por modelos de classificação do glaucoma tem sido por meio do cálculo da medida de CDR, Figura 3.12, geralmente obtida por meio da aplicação da segmentação das estruturas do disco e copo óptico. Em seguida, por meio da CDR calculada, os algoritmos estimam a presença ou a ausência do glaucoma.



Figura 3.12: Exemplo de imagem com segmentação do disco e copo óptico.

Embora muitos algoritmos tenham apresentado alta taxa de precisão na segmentação destas estruturas, este método pode ser considerado apenas como um indicativo do glaucoma e de necessidade de uma avaliação mais detalhada, pois, o diagnóstico dessa neuropatia se faz por meio do exame de toda a estrutura do disco óptico e não somente da escavação. Além disso, mesmo que a escavação aumentada seja sugestiva de glaucoma, nem toda escavação do nervo óptico está relacionada a essa doença, pois existem outras condições que podem causar o aumento da escavação do nervo óptico, como neurites, tumores, esclerose múltipla, etc. Devido à complexidade das imagens médicas, tais aplicações às imagens do fundo de olho são geralmente implementadas por métodos de segmentação semântica, também usada em larga escala como métodos de pré-classificação, onde primeiro é segmentada a região de interesse e em seguida aplica-se o classificador à imagem segmentada.

- 
- Reconhecimento de padrões glaucomatosos: enquanto os algoritmos de cálculo da CDR avaliam apenas a escavação do disco óptico, esta metodologia de reconhecimento de padrões busca avaliar toda a região do disco óptico em busca de características que possam levar ao reconhecimento do glaucoma. Conforme o contexto de *deep learning* e análise dos trabalhos relacionados descritos neste Capítulo, este tipo de aplicação pode ser operada ao menos por 4 metodologias distintas, tais como:

1. Extração de vetores de características e classificação: neste tipo de aplicação, várias técnicas de processamento de imagens e extração de características podem ser utilizadas sobre as imagens digitais, no entanto, um classificador será a parte do sistema responsável pela tarefa de categorização, ou seja, que irá aplicar o processo de decisão sobre a qual categoria uma determinada imagem pertence. Dentre os algoritmos que trabalham desta maneira, cita-se a SVM, KNN, Naive Bayes, etc. São algoritmos supervisionados que podem trabalhar com conjuntos de dados menores em comparação às redes CNNs.
2. Uso de redes CNNs: as redes CNN permitem que imagens globais ou fragmentos de imagens sejam usados como entrada para tarefas de classificação. Essa abordagem elimina a necessidade de extrair vetores de características, uma vez que as redes CNN podem extrair tais recursos por meio de mapas de características com suas camadas convolucionais. É uma rede supervisionada, e necessita de grandes conjuntos de dados para treinamento.
3. Uso de redes GANs: A modelagem generativa é uma tarefa de aprendizado não supervisionado em aprendizado de máquina, que basicamente envolve descobrir regularidades e padrões nos dados de entrada e aprendê-los automaticamente.
4. Uso de multitecnologias: este tipo de modelagem busca atingir o objetivo desejado utilizando uma combinação de técnicas, como KNN, SVM, CNN, etc.

Os benefícios buscados para essas possíveis aplicabilidades são variados, desde a potencial redução de custos associados ao diagnóstico tradicional do glaucoma, até o auxílio a aplicações de triagem populacional visando o diagnóstico precoce e a redução do índice de subdiagnóstico da doença.

### **3.6.2.1 Classificação do glaucoma a partir da relação escavação disco**

Em relação ao uso de algoritmos de *deep learning* para segmentação das estruturas do fundo de olho, em particular às estruturas correspondentes da relação escavação disco CDR, para estimar o dano glaucomatoso ou simplesmente segmentar a região do disco

---

óptico para outras tarefas relacionados, autores como (Sevastopolsky et al., 2018) utilizaram uma arquitetura U-Net para realizarem a segmentação da região do disco óptico nas bases de dados DRIONS-DB, RIM-One-v3 e DRISHTI-GS, com avaliação em um conjunto de dados privados.

Em (Gupta et al., 2021), os autores propõem uma estrutura para detecção de glaucoma usando redes CNN. Nesta estrutura, uma etapa de pré-processamento usa o algoritmo CLAHE para aumentar o contraste local das imagens. Além disso, os autores usaram dois modelos de segmentação conhecidos como (EfficientNet e U-Net) para segmentar o copo óptico, utilizado em seguida para calcular a relação CDR. O algoritmo foi testado para identificar o glaucoma com base na relação CDR, portanto, uma avaliação da capacidade de classificação foi aplicada em dois conjuntos de dados de referência (DRISHTI-GS1 e RIM-ONE). Os resultados experimentais mostraram que o algoritmo proposto superou métodos de última geração para detecção de glaucoma com uso de imagens da retina.

### **3.6.2.2 Classificação do glaucoma a partir de vetores de características**

(Abdel-Hamid, 2020) propuseram um algoritmo com uso da transformada wavelet para detecção de glaucoma. Esse método utiliza-se de aplicações no domínio do espaço e da frequência. Assim, uma combinação de características estatísticas e texturais baseadas em wavelets foram calculadas a partir da região do disco óptico para determinar se uma imagem da retina é saudável ou glaucomatosa. O estudo utilizou para a análise de desempenho do classificador os conjuntos públicos de imagens de fundo de olho (GLAUCOMADB e HRF). Uma precisão de 96,7% e AUC de 94,7% foram alcançadas para o conjunto de dados HRF usando o algoritmo KNN.

(Singh et al., 2021) usaram vários recursos estatísticos de imagens de fundo de olho e cálculos numéricos com base na regra ISNT e CDR para alimentar um modelo classificador. A saída do modelo proposto foi obtida por meio de um ensemble a partir da concatenação das saídas dos classificadores individuais (SVM, KNN, Naive Bayes e uma rede neural artificial). Os autores alcançaram 98,6% de precisão na classificação do glaucoma na base de dados DRIONS-DB.

Em (Shiny Christobel et al., 2022) as imagens retinianas são inicialmente pré-processadas usando o método CLAHE. Os vetores de características são extraídos usando PCA, em seguida, as características extraídas são fornecidas como entrada para o classificador SVM, Naive Bayes e K-NN para classificar as imagens normais e glaucomatosas. A partir dos resultados da classificação, a medida de desempenho de cada classificador é calculada e comparada com os demais. Os resultados dos testes mostraram que o classificador baseado em recursos PCA tem melhor desempenho. Desta forma, entre os classificadores utilizados, o Naive Bayes fornece alta precisão de 95%, sobre uma base de dados formada com imagens glaucomatosas obtidas de um conjunto de dados denominado STARE e imagens saudáveis obtidas do banco de dados MESSIDOR.

---

### 3.6.2.3 Classificação do glaucoma com uso de redes CNNs

Em outros estudos, como (Chen et al., 2015; Li et al., 2019a; Raghavendra et al., 2018; Aamir et al., 2020), os pesquisadores usaram CNNs para classificar imagens de fundo de olho devido às suas vantagens óbvias no processamento de imagens. Uma das dificuldades é que esses algoritmos precisam ser treinados em grandes conjuntos de dados, que nem sempre estão disponíveis publicamente. No entanto, autores como (Chen et al., 2015) desenvolveram uma estrutura contextualizada *deep learning* para diferenciar imagens oftalmológicas entre padrões glaucomatosos e não-glaucomatosos. Os experimentos foram realizados nos conjuntos de dados públicos ORIGA e SCES. Os resultados mostraram AUC de 0,838 e 0,898 para detecção de glaucoma nos dois bancos de dados, respectivamente.

Usando conjuntos privados de imagens, (Li et al., 2019a) propôs uma CNN baseada em atenção para detecção de glaucoma. Os mapas de atenção são utilizados para destacar as regiões mais importantes para detecção de glaucoma. Os autores usaram um conjunto de dados compreendendo 11.760 imagens de fundo de olho, com 4.878 marcadas como glaucomatosas e 6.882 normais. Em (Raghavendra et al., 2018), os autores realizaram o diagnóstico de glaucoma com uma rede CNN com dezoito camadas. No experimento foram extraídas características robustas de 1.426 imagens de fundo, entre as quais 589 eram normais e 837 apresentavam glaucoma. O método apresentou uma precisão de 98,13%, uma sensibilidade de 98% e uma especificidade de 98,3%.

Após comparar vários modelos pré-treinados, os autores (Li et al., 2019c) propuseram um modelo de aprendizado por transferência capaz de classificar o glaucoma com 94,71% de precisão. Além disso, os autores utilizaram explicações diagnósticas de modelos interpretáveis locais que introduzem explicações referentes à qual região da imagem foi mais importante para a classificação feita pelo algoritmo. Para isso, o artigo propõe uma rede CNN baseada em atenção. Assim, especificamente, primeiro foi estabelecido um conjunto de dados de glaucoma baseado em atenção, incluindo 5.824 imagens da retina marcadas com glaucoma, divididas em 2.392 casos positivos e 3.432 casos negativos.

Os autores (Aamir et al., 2020) coletaram um conjunto de dados de imagens da retina. As imagens são pré-processadas com um equalizador de histograma adaptável para reduzir o ruído. A arquitetura de rede neural convolucional profunda multinível proposta é usada para extração e classificação de recursos em dois estágios, o primeiro estágio é usado para detecção de glaucoma, chamado de rede de detecção, e a segunda rede de classificação é usada para classificar as imagens afetadas por glaucoma em três categorias diferentes: avançado, moderado e inicial. O modelo foi testado em 1338 imagens de glaucoma da retina e alcançou 98,2% de precisão.

Em (Chayan et al., 2022), propõe-se um modelo de aprendizado por transferência capaz de classificar o glaucoma com 94,71% de precisão em um conjunto de dados privado

---

contendo 4250 imagens para treinamento, 302 imagens para validação e 302 imagens para testes. Além disso, os autores introduzem interpretabilidade ao sistema usando *Local Interpretable Model-Agnostic Explanations*, mostrando quais são as regiões da imagem mais importantes para classificação. Esta melhoria, além da acurácia, dá aos profissionais médicos acesso a informações vitais e abrangentes para ajudá-los a tomar decisões.

Ainda com uso de conjuntos privados de imagens, (Ting et al., 2017) treinou um algoritmo baseado em deep learning com 125.189 imagens de fundo de olho para detecção do glaucoma. O desempenho do classificador foi avaliado em 71.896 imagens de teste, apresentando AUC de 0,942% na classificação de possível glaucoma. Este trabalho demonstra o poder diagnóstico e computacional dos algoritmos de *deep learning*, sobretudo utilizando grandes conjuntos de dados.

Continuando com o uso de conjuntos de dados privados, (Li et al., 2018b) desenvolveram um algoritmo de *deep learning* para detectar casos glaucomatosos com base em 48.116 fotografias de fundo de olho. Este estudo recrutou 21 oftalmologistas treinados para rotular as fotografias utilizadas na pesquisa. Na formação da base de dados os profissionais definiram o glaucoma encaminhava com relação escavação-disco vertical de 0,7 além de outras alterações típicas da neuropatia. A eficácia do algoritmo na classificação do glaucoma foi medida a partir de 8.000 imagens de validação, o que rendeu uma AUC de 0,986.

Os autores (Liu et al., 2018) usaram conjuntos de dados públicos e privados para classificar o glaucoma e concluíram que um algoritmo baseado em *deep learning* pode identificar o glaucoma com alta precisão. A base de dados foi formada com 4.394 imagens de fundo de olho, incluindo 3.768 imagens de estudos clínicos anteriores obtidas de 3 centros oftalmológicos terciários em Sydney, Austrália e 626 imagens de bancos de dados RIM-ONE e HRF.

No trabalho de (Fu et al., 2018), os autores apresentaram uma técnica de *deep learning* para obter informações adicionais relevantes para as imagens do fundo de olho. A técnica pode integrar quatro fluxos de profundidade em diferentes níveis de uma imagem fundoscópica. Os experimentos de classificação do glaucoma foram aplicados em dois conjuntos de dados de glaucoma (SCES e SINDI), obtendo resultados AUC de 91,83% no SCES e 81,73% no SINDI.

Em (Ahn et al., 2018), os autores usaram 1542 fotografias do fundo de olho com tamanhos diferentes e cortadas para o tamanho de 240 × 240 pixels. As imagens foram obtidas usando uma câmera não midiática Nidek AFC-330 e foram divididas em grupos de 754 para treinamento, 324 para validação e 464 para testes, entre os quais 786 eram normais, 467 apresentavam glaucoma avançado e 289 apresentaram glaucoma precoce. Os conjuntos de dados foram usados para classificação do glaucoma utilizando modelos pré-treinados. Ao final do experimento foram relatados 99,7% de precisão com dados de treinamento, 87,7% com dados de validação e 84,5% com dados de teste. Por fim, eles

---

também concluíram que tanto o glaucoma inicial quanto o avançado podem ser detectados corretamente pelo uso de *deep learning* com apenas fotos de fundo de olho.

(Diaz-Pinto et al., 2019c) usou cinco diferentes modelos pré-treinadas como classificadores de glaucoma (VGG16, VGG19, InceptionV3, ResNet50 e Xception). Todas essas arquiteturas foram alimentadas apenas com conjuntos de dados publicamente disponíveis (ACRIMA, HRF, DRISHTI-GS, RIM-ONE e Sjchoi86-HRF). O melhor modelo testado apresentou área média sob a curva ROC de 0,960 para classificação do glaucoma.

Os autores em (Gómez-Valverde et al., 2019) selecionaram os modelos pré-treinados VGG19, GoogleNet, ResNet50 e DENet para a classificação automática do glaucoma. No experimento comparou-se o desempenho do treinamento de modelos CNNs com o aprendizado por transferência e treinamento do zero. Os autores usaram 2.313 imagens da retina de três bases de dados diferentes: RIM-ONE, DRISHTI-GS e Esperanza (conjunto de dados privado). No conjunto de dados RIM-ONE, as imagens classificadas como suspeitas foram consideradas para o estudo como glaucomatosas. O melhor resultado foi obtido com o modelo VGG19 atingindo uma AUC de 0,94 usando aprendizado por transferência e validação cruzada de 10 vezes para confirmar o desempenho do algoritmo.

Em (Martins et al., 2020), as fotografias do fundo de olho foram analisadas com um pipeline de algoritmos com capacidade de diagnosticar o glaucoma com dispositivos móveis. Os autores usaram diferentes conjuntos de dados, incluindo ORIGA, DRISHTI-GS, iChallenge, RIM-ONE e *Retinal Fundus Images for Glaucoma Analysis* (RIGA). Por fim, com uso de modelos de CNNs pré-treinados, relatou-se 90% de precisão no reconhecimento do glaucoma.

Algoritmos com aprendizado por transferência também foram avaliados em vários outros estudos, como em (Li et al., 2019c), e também em aplicado em (Shibata et al., 2018; Norouzifard et al., 2018), em que conjuntos de dados privados foram usados e a precisão obtida excedeu 90%. Além disso, (Christopher et al., 2018) estudou três arquiteturas *deep learning* diferentes. Para cada arquitetura, duas versões diferentes foram avaliadas: aprendizado nativo (do zero) e aprendizado por transferência. Em todos os casos, os autores mostraram que o aprendizado por transferência pode melhorar o desempenho e reduzir o tempo de treinamento dos algoritmos.

Em (Serener and Serte, 2019), a classificação do glaucoma em estágios iniciais e avançados foi realizada em imagens de fundo de olho usando dois métodos de *deep learning*, ResNet50 e GoogLeNet. O desempenho dos dois modelos foi avaliado em termos de precisão, sensibilidade, especificidade e área sob a curva ROC; ao final, os resultados mostram que, para a detecção precoce, avançada e geral do glaucoma, o GoogLeNet supera o ResNet-50, considerando o experimento realizado.

Os autores em (Phene et al., 2019) desenvolveram um algoritmo para detecção do glaucoma usando um conjunto de dados retrospectivo de 86.618 imagens. O algoritmo *deep learning*, treinado apenas em imagens de fundo de olho, pôde detectar glaucoma

---

encaminhável com maior sensibilidade e especificidade comparável aos oftalmologistas. O algoritmo manteve um bom desempenho em um conjunto de dados independente com diagnósticos baseados em uma avaliação completa do glaucoma.

O trabalho de (Bragança et al., 2022) classificou o glaucoma com auxílio de um *ensemble* de redes CNNs pré-treinadas. Os autores alcançaram uma acurácia de 91% sobre a base de dados BrG, mostrando a capacidade dos algoritmos de *deep learning* na classificação do glaucoma em imagens de menor resolução e obtidas com auxílio de smartphones.

Em (Nandhini et al., 2022), os autores utilizaram a classificação do glaucoma em duas etapas; primeiro, utilizaram uma rede U-Net para segmentação das imagens e, posteriormente, as imagens segmentadas foram classificadas usando uma rede CNN. Os autores (Lee et al., 2019) usaram um método chamado de M-Net para segmentar o disco óptico e o copo óptico em imagens digitais do fundo de olho. Os testes foram validados com o conjunto de dados REFUGE para triagem de glaucoma. Os resultados relataram um coeficiente Dice de 94,26% para o disco óptico e 85,65% para o copo óptico, com AUC de 96,37% na tarefa de classificação do glaucoma.

(Shankaranarayana et al., 2019) propuseram uma arquitetura de rede totalmente convolucional para estimar a profundidade da escavação ocular por meio de segmentação e usando um mapa de profundidade como guia. Os autores validaram os experimentos em três conjuntos de dados ORIGA, RIMONE-r3 e DRISHTI-GS. (Edupuganti et al., 2018) implementaram um sistema de rede totalmente convolucional para realizar a segmentação do disco óptico e das regiões do copo óptico. Os autores validaram os experimentos com o conjunto de dados RIM-ONE utilizando a rede aplicada com uma e duas camadas. Os resultados relataram uma precisão de 95,6% e 96,9% considerando as formas de aplicação, respectivamente. Os autores também avaliaram os resultados conforme a métrica AUC, que correspondeu 98% com uma camada e 97,8% com duas camadas de rede.

Em (Al-Bander et al., 2018), os autores propuseram abordar a detecção da localização do disco óptico e da fóvea como desenvolvimento de ferramentas automáticas de diagnóstico de doenças da retina. Como metodologia, uma rede (CNN) foi projetada e treinada com os conjuntos de dados MESSIDOR e Kaggle disponíveis publicamente. Os centros da fóvea e o OD em cada imagem foram rotulados por avaliadores especializados como a verdade básica. O método proposto atinge uma precisão de 97%, 96,7% para a detecção do centro OD e 96,6%, 95,6% para a detecção do centro foveal dos conjuntos de teste MESSIDOR e Kaggle, respectivamente.

#### **3.6.2.4 Classificação do glaucoma com uso de CGANs**

Em (Singh et al., 2018), os autores usaram um método de segmentação de imagem da retina baseado em rede CGAN para segmentar o disco óptico. Os experimentos foram conduzidos em dois conjuntos de dados disponíveis publicamente; DRISHTI GS1 e RIM-

---

ONE. O modelo proposto atinge coeficientes de Jaccard e Dice de 0,96% e 0,98%, respectivamente, na segmentação do disco óptico.

O trabalho de (Chang et al., 2023) propõe um método para gerar imagens correspondentes a OCT a partir de imagens de fundo de olho usando uma rede GAN. O propósito é ajudar os médicos de família a julgar se um exame mais aprofundado é necessário com base nas imagens de OCT geradas. Os resultados experimentais mostram que a semelhança de cosseno entre a imagem OCT gerada e a imagem OCT real é de 97,8%. Combinado com o método de aprendizado de transferência proposto, a precisão de classificação do modelo chega a 83,17%. Combinado com o método incremental, a precisão de identificação do glaucoma é de cerca de 78,94%, o que é 8,77% maior do que a precisão de 70,17% do modelo inicial.

Em (Jain et al., 2022), uma abordagem de detecção de glaucoma é desenvolvida usando a rede adversarial Geral baseada em otimização de *forageamento rider manta-ray*. O processo de segmentação do disco óptico é realizado pelo agrupamento *Fuzzy Local Information C-Means (clustering)*. Além disso, o processo de centelhamento também é empregado neste método de detecção de glaucoma para o processo de detecção de vasos sanguíneos. Vários recursos significativos neste modelo de detecção, ou seja, média, desvio padrão, variância, curtose, assimetria, entropia e recursos CNN, são extraídos para detectar ainda mais as imagens glaucomatosas. O modelo GAN desenvolvido mostrou resultados aprimorados com a maior precisão de 0,96% na segmentação do disco óptico.

### **3.6.2.5 Classificação do glaucoma a partir de multitecnologias**

Embora os autores possam mesclar várias abordagens de IA, o objetivo final é obter uma aplicação de *deep learning* para a classificação do glaucoma a partir de imagens de fundo de olho obtidas por meio da retinografia. Assim, alguns pesquisadores, tais como (Shinde, 2021; Sreng et al., 2020; Abdel-Hamid, 2020), focaram na implementação de algoritmos para segmentar e calcular o valor de CDR ou aplicaram em algoritmos para a análise da textura de imagens de fundo da retina, por final, os autores atingiram o objetivo da classificação com o auxílio de diversas arquiteturas, como KNN, SVM, árvores de decisão e Naive Bayes, etc.

Em (Shinde, 2021) os autores desenvolveram um sistema de diagnóstico assistido por computador. Este aplicativo foi desenvolvido usando abordagens de processamento de imagem, *deep learning* e *machine learning*. A arquitetura Le-Net é usada para validação das imagens de entrada, enquanto a região de interesse é encontrada com auxílio do algoritmo *brightest spot*. Além disso, a segmentação do disco óptico e do copo óptico é realizada com a ajuda da arquitetura U-Net e a classificação é feita usando classificadores SVM, RNA e adaboost.

Também em dois estágios, os autores em (Sreng et al., 2020) introduziram um sistema automático de triagem de glaucoma. O sistema primeiro segmentou a região do

disco óptico e, em seguida, os autores usaram arquiteturas CNNs pré-treinadas para três propósitos: aprendizado por transferência (1), descritores de características usando uma SVM (2) e por meio de um ensemble contendo os dois métodos anteriores (1) e (2). Para a classificação do glaucoma, um conjunto de métodos teve melhor desempenho do que os métodos convencionais para conjuntos de dados RIM-ONE, ORIGA, DRISHTI-GS1 e ACRIMA com precisão de 97,37%, 90,00%, 86,84% e 99,53% e área sob a curva (AUC) de 100%, 92,06%, 91,67% e 99,98%, respectivamente.

Em (dos Santos Ferreira et al., 2018), primeiro, a aquisição de imagens é realizada a partir dos bancos de dados RIM-ONE, DRIONS-DB e DRISHTI-GS e, em seguida, uma rede neural convolucional para segmentação de disco óptico foi treinada. Após a segmentação, os vasos sanguíneos precisaram ser removidos, seguido de extração de recursos no canal RGB e imagens geradas em escala de cinza. Os atributos extraídos foram baseados apenas em características de textura usando um índice de diversidade filogenética. O método se mostrou promissor, alcançando 100% de acurácia, sensibilidade e especificidade na classificação do glaucoma usando métodos baseados em redes CNNs.

No trabalho disponível em (Claro et al., 2019), o método de extração de características inclui descritores como: LBP, GLCM, HOG, morfologia e sete arquiteturas CNN, resultando em 30.682 características. Em seguida, utilizou-se a taxa de ganho para ordenar os recursos por importância e selecionar o melhor conjunto para a classificação do glaucoma. Os testes foram realizados usando 1675 imagens dos bancos de dados DRISHTI, RIM-ONE, HRF, JSIEC e ACRIMA. O método alcançou uma precisão de 93,61% e finalmente concluiu-se que a combinação de GLCM e CNN pré-treinada tem potencial para ser usada em um sistema assistido por computador para detecção de glaucoma.

### 3.6.2.6 Síntese dos estudos relacionados

A Tabela 3.2 apresenta trabalhos relacionados que mesclaram bases de dados públicas e privadas. Na sequência, a Tabela 3.3 apresenta trabalhos relacionados que utilizaram apenas bases de dados públicas e a Tabela 3.4 apresenta trabalhos relacionados que utilizaram apenas bases de dados particulares.

Tabela 3.2: Trabalhos utilizando bases de dados públicas e privadas

AUTOR	BASE DE DADOS
(Sevastopolsky et al., 2018)	Drions-db, rim-one v.3, drishti-gs
(Liu et al., 2018)	4.394 imagens
(Gómez-Valverde et al., 2019)	Rim-one, drishti-gs e esperanza
(Martins et al., 2020)	Origa, drishti, iChallenge, rim-one e riga
(Nandhini et al., 2022)	Refuge
(Shinde, 2021)	Rim-one, drishti-gs, drions, Jsiec e drive
(Claro et al., 2019)	Drishti-ge, rim-one, hrf, jsiec, acrima

Tabela 3.3: Trabalhos utilizando bases de dados públicas

<b>AUTOR</b>	<b>BASE DE DADOS</b>
(Gupta et al., 2021)	Drishti-gs1 e rim-one
(Abdel-Hamid, 2020)	Glaucoma-db e hrf
(Singh et al., 2021)	Drions-db
(Shiny Christobel et al., 2022)	Messidor e stare
(Chen et al., 2015)	Origa e sces
(Fu et al., 2018)	Sindi e sces
(Diaz-Pinto et al., 2019c)	Acrima, hrf, drishti-gs1 e rim-one
(Bragança et al., 2022)	2.000 imagens
(Sengupta et al., 2020)	Refuge
(Lee et al., 2019)	Refuge
(Shankaranarayana et al., 2019)	Rim-one e drishti-gs
(Edupuganti et al., 2018)	Rim-one
(Al-Bander et al., 2018)	Drions, drishti-gs, rim-one e origa
(Singh et al., 2018)	Gans
(Sreng et al., 2020)	Drishti, rim-one, acrima, origa
(dos Santos Ferreira et al., 2018)	Drishti, rim-one, drions

Tabela 3.4: Trabalhos utilizando bases de dados particulares

<b>AUTOR</b>	<b>BASE DE DADOS</b>
(Li et al., 2019b)	11.760 imagens
(Raghavendra et al., 2018)	1426 imagens
(Li et al., 2019c)	5.824 imagens
(Aamir et al., 2020)	1338 imagens
(Chayan et al., 2022)	4250 imagens
(Ting et al., 2017)	125.189 imagens
(Li et al., 2018b)	48.116 imagens
(Ahn et al., 2018)	1542 imagens
(Shibata et al., 2018)	3132 imagens
(Norouzifard et al., 2018)	277 imagens
(Christopher et al., 2018)	14.822 imagens
(Phene et al., 2019)	86.618 imagens
(Chang et al., 2023)	Imagens OCT
(Jain et al., 2022)	Gans

### 3.6.3 Características gerais dos trabalhos relacionados

Em relação aos algoritmos de classificação do glaucoma, grandes conquistas foram alcançadas na distinção entre imagens digitais de glaucoma e não-glaucoma. Em alguns casos, como (Phene et al., 2019), os algoritmos são mais precisos do que a classificação feita por oftalmologistas experientes. No entanto, embora vários estudos concordem que os algoritmos de IA podem ser utilizados como uma tecnologia no auxílio ao diagnóstico do glaucoma,

---

ainda não há software para aplicações clínicas reais, sugerindo que é necessário mais envolvimento teórico e prático para melhorar a aplicabilidade de tais algoritmos.

Uma das dificuldades que compromete a metodologia de *machine learning* para testes mais representativos são as deficiências das bases de dados, que muitas vezes possuem um número pequeno de imagens, além disso, a maneira como essas imagens são rotuladas, também podem ser prejudicial aos algoritmos de classificação.

Em relação à rotulagem das bases de dados e imagens, em geral, os estudos apresentados nesta revisão solicitaram que avaliadores (oftalmologistas) fizessem tais anotações de rótulos apenas examinando imagens da retina para determinar a presença ou ausência de glaucoma. No entanto, em um estudo com 6 avaliadores especializados em glaucoma, sendo solicitados a diagnosticar a doença com base apenas nas fotografias do fundo ocular, a concordância entre eles foi de apenas 49% (Lee et al., 2021), mostrando que, por ser muito passível de erros, rotular base de dados apenas com observação das imagens do fundo de olho pode ser prejudicial à qualidade final dos resultados.

Uma forma de minimizar a inserção de erros em uma base de dados por meio da rotulagem das imagens é assegurar que essa prática seja aplicada sempre considerando a presença de oftalmologistas experientes e seguindo o padrão ouro de diagnóstico do glaucoma, ou seja, a união de exames funcionais e estruturais. Atingir esse patamar de qualidade sempre é considerado uma tarefa difícil; por esse motivo, muitos autores que trabalham com grandes conjuntos de dados privados, como (Ting et al., 2017; Li et al., 2018b; Phene et al., 2019), não rotularam suas bases de dados diante de um padrão ouro de diagnóstico, mas usaram um consenso de rotulagem avaliado por oftalmologistas experientes. No entanto, ainda tiveram as suas bases de dados rotuladas apenas com a informação visual de imagens de fundo de olho.

Quanto ao quantitativo de imagens rotuladas para glaucoma, as bases de dados disponíveis publicamente geralmente são pequenas e ainda racial ou clinicamente homogêneas, podendo não representar toda uma população que se deseja estudar. Deste modo, esses fatores podem dificultar a aplicabilidade algorítmica para uma ampla generalização. Estudos alternativos têm sido aplicados às redes GANs como métodos para geração artificial de imagens semelhantes às imagens originais, porém, mesmo que essas redes alcancem bons resultados, as imagens geradas por tais técnicas podem não suprir a homogeneidade dos conjuntos de dados com relação à raça, ou diferenças na possível ocorrência de dano retiniano glaucomatoso. Para tentar driblar os problemas envolvendo a composição das bases de dados, muitos autores usam a combinação de tais conjuntos de dados públicos para a classificação do glaucoma, como mostra a Tabela 3.2.

Ainda no que diz respeito à qualidade das bases de dados e à forma como são construídas, algumas das suas principais características observadas são:

- Todos os conjuntos de dados foram adquiridos com câmera retiniana de alta resolução, exceto o BrG, que foi adquirido com o auxílio de um smartphone conectado a um

---

oftalmoscópio portátil.

- Com exceção dos conjuntos de dados Refuge e Rim-one-dl formados com duas câmeras digitais de fundo de olho, todos os outros conjuntos de dados foram adquiridos com apenas uma câmera digital retiniana;
- A maioria das bases de dados foi rotulada com base na opinião oftalmológica sobre a observação apenas de imagens de fundo de olho. Poucas bases de dados são rotuladas com assistência oftalmológica e padrão-ouro de diagnóstico do glaucoma;
- Todas as bases de dados públicas são consideradas muito pequenas para treinar algoritmos de classificação do zero, ou seja, sem usar aprendizado por transferência.
- As bases de dados disponíveis publicamente são geralmente homogêneas em relação à etnia da população coletada.

Além das deficiências das bases de dados, outra dificuldade encontrada pelos algoritmos de *deep learning* na classificação do glaucoma é a falta de critérios diagnósticos objetivos e consistentes para qualificar a doença. Isso contribui para que os pesquisadores que investigam e testam o uso de IA neste contexto, elabore as suas próprias definições para classificar as instâncias em "sim" ou "não" para o glaucoma. Assim, enquanto alguns autores trabalharam com análise de textura, outros trabalharam na análise da razão CDR, ou de regras ISNT, etc (Lee et al., 2021). Não há problema com a divergência de métodos, mas muitas vezes esse contraste se deve à falta de biomarcadores específicos e computáveis para definir a doença. Portanto, muitos pesquisadores tentaram prever resultados semelhantes para o diagnóstico do glaucoma, mas usaram métodos diferentes, reduzindo a capacidade de comparar o desempenho de diferentes estudos. Tais biomarcadores também são importantes, pois, além das alegações médicas exigirem uma resposta "sim" ou "não" ao diagnóstico, também exigem uma razão para a referida resposta.

Devido a essa exigência médica, pelo menos até que os algoritmos sejam aceitos pela comunidade de saúde, muitos autores, como, por exemplo, (Christopher et al., 2018; Sreng et al., 2020; Martins et al., 2020), ao aplicar algoritmos de *deep learning* para classificação do glaucoma, também apresentaram as regiões mais importantes da imagem para as respostas dadas pelos algoritmos, como uma forma de justificar o porquê da resposta alcançada pela metodologia. Assim, enquanto alguns autores trabalham na análise de texturas, outros trabalham na análise de proporções CDR (Lee et al., 2021) ou regras ISNT, etc.

Outro fator importante na construção de banco de dados e estudos de classificação do glaucoma é a exclusão de indivíduos com múltiplas lesões oculares. Consequentemente, muitos autores relatam que excluíram especificamente dos conjuntos de dados de treinamento e teste indivíduos com doenças oculares diferentes do glaucoma e também

---

excluíram imagens comprometidas por outras doenças sistêmicas que poderiam afetar diretamente o nervo óptico ou o campo visual. Esse tipo de exclusão pode ser negativo, uma vez que é possível entendê-la como manipulação do mundo real em favor da precisão algorítmica. Quanto à generalização dos algoritmos para outras populações além das observadas pelo conjunto de dados, a homogeneidade racial dos conjuntos de dados contrasta com a verdadeira diversidade da população, dificultando tal generalização.

Diante do contexto, deve-se notar que o objetivo de todos os trabalhos relacionados não foi tornar o algoritmo pronto para o mercado, mas demonstrar os elementos essenciais para obter bons resultados ao usar imagens de fundo de olho para classificação do glaucoma em possíveis aplicações futuras. Assim, para maior evolução de tais pesquisas, é sugerido que, para melhor uso de algoritmos de *deep learning*, as bases de dados precisam ser rotuladas conforme o padrão ouro de diagnóstico, devendo apresentar uma clara separação entre os conjuntos de treinamento e teste, ter diversidade no conjunto de imagens, ou seja, ser capturadas por vários dispositivos, envolvendo também pacientes de diferentes etnias, além disso, devem apresentar imagens capturadas em diferentes condições de iluminação, contraste, ruído etc (Batista et al., 2020; Camara et al., 2022).

Quanto aos algoritmos relacionados, embora vários estudos científicos tenham mostrado alta precisão na distinção entre imagens glaucomatosas ou não, faltam ensaios clínicos e mais pesquisas para levantar e sanar possíveis fatores que venham a impedi-los de serem incorporados ao processo de tomada de decisão clínica. Quanto às bases de dados encontradas, nota-se que estas atendem parcialmente aos requisitos aqui relatados. Mesmo assim, são de grande importância para o treinamento de diversos algoritmos e para a evolução tecnológica.

Por fim, dada à importância da construção de bases de dados, o Capítulo seguinte, traz consigo um esforço do desenvolvimento de um novo conjunto de dados e um novo conceito de imagens digitais do fundo de olho obtidas a partir de oftalmoscópios portáteis acoplados em um simples smartphone.

## Capítulo 4

### Construção da base de dados

Este capítulo descreve as etapas para construção de um novo conjunto de dados denominado, Brasil glaucoma (BrG), que contém imagens de fundo de olho rotuladas para o glaucoma. Como forma de aquisição, todas as imagens foram obtidas com um smartphone acoplado a um oftalmoscópio portátil. A aquisição da base de dados que ocorreu no Estado de Minas Gerais - Brasil, segue o organograma da Figura 4.1 e faz parte da primeira fase para alcançar o objetivo geral desta tese.

A base de dados BrG contendo 2.000 imagens, está disponível publicamente e, além de subsidiar este trabalho, deve ser entendida como uma contribuição para o desenvolvimento de pesquisas futuras que busquem automatizar o diagnóstico do glaucoma, visto que não existiam bases de dados de imagens digitais de fundo de olho qualificadas e disponíveis para estudos como este, que buscam avaliar a classificação do glaucoma a partir de imagens fundoscópicas obtidas com auxílio de smartphones.

Diante disso, primeiramente, foi necessário examinar qual equipamento poderia ser utilizado para realizar tais fotografias, uma vez que, elas podem ser obtidas a partir de uma variedade de dispositivos, como câmeras de retina (retinógrafos), lâmpadas de fenda com adaptadores de câmera e uma variedade de oftalmoscópios portáteis que possuem as suas próprias câmeras ou podem ser conectados a um smartphone para a captura das fotografias.

Durante a escolha do equipamento para a construção da base de dados, foram realizadas buscas por aparelhos portáteis que tenham o menor custo de aquisição em relação aos aparelhos convencionais e que capturem fotografias de alta qualidade de fundo de olho.

Outra observação relevante na aquisição do aparelho para a construção da base de dados foi a não necessidade de dilatação ou midríase da pupila para examinar o fundo de olho, pois, esse procedimento ocorre farmacologicamente, sendo totalmente a critério de oftalmologistas. Portanto, atentou-se em usar um dispositivo que não requeira este fármaco, desse modo, facilitará a aplicação de uma possível triagem populacional com auxílio de profissionais não médicos.

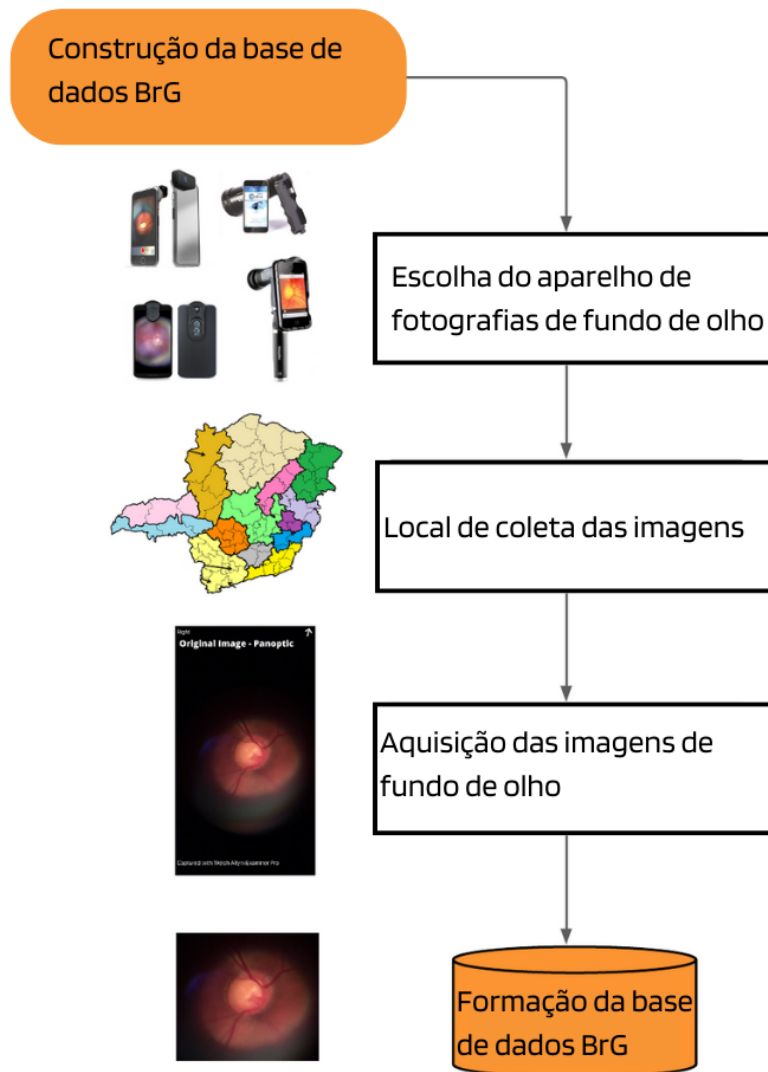


Figura 4.1: Organograma das etapas de aquisição da base de dados BrG.

Em suma e, conforme o escopo deste estudo, os requisitos básicos para a obtenção do equipamento utilizado para adquirir as imagens do fundo de olho e construir uma base de dados para este trabalho foram:

- Baixo custo de aquisição, para viabilizar aplicações de triagens populacionais;
- Portabilidade, facilidade de utilização por profissionais não médicos e não necessidade de midríase;
- Boa qualidade nas imagens adquiridas para melhor representar as regiões da retina e ser mais bem aceito pelos algoritmos de *deep learning*.

## 4.1 Escolha do aparelho de fotografias de fundo de olho

Com base nos pré-requisitos, dentre os aparelhos com menor custo de aquisição em comparação aos demais, estão os oftalmoscópios portáteis que não possuem uma câmera embutida, mas podem se conectar a um smartphone para a aquisição de fotografias. Portanto, conforme exemplificado na Figura 4.2, foram selecionados quatro modelos com tais características como candidatos para este estudo, que são:

- O *Peek Retina* (Maamari et al., 2014);
- Sistema Volk *iNview* (Volk, 2022);
- Oftalmoscópio direto *D-eye*, portátil de mão (Russo et al., 2015);
- Oftalmoscópio *panoptic* com *iExaminer* (WelchAllyn, 2020).

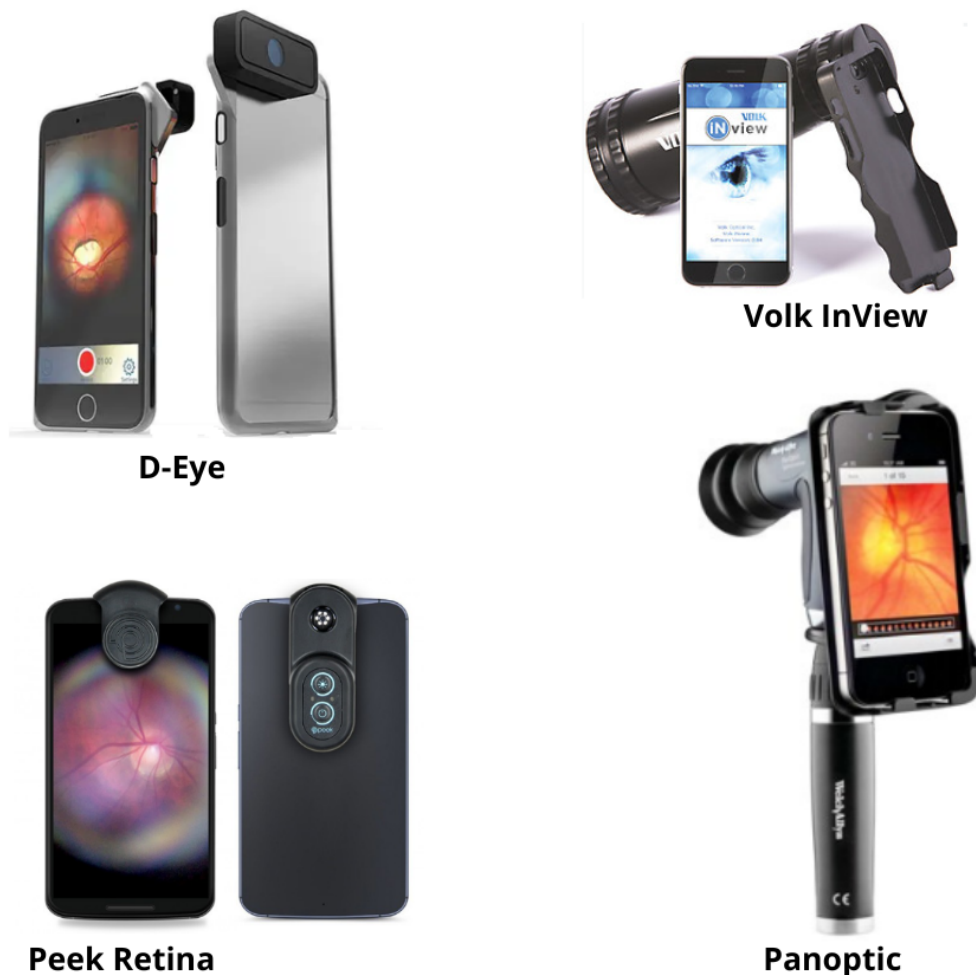


Figura 4.2: Imagens de aparelhos portáteis e úteis para realização de fotografias do fundo de olho com auxílio de smartphones

Cada um desses aparelhos listados tem características específicas. O *Peek Retina* inclui luz própria para iluminar a retina e a imagem é captada por um smartphone, porém,

---

ele foi descartado dessa pesquisa devido à necessidade de dilatação da pupila e as vendas descontinuadas do aparelho.

O sistema *Volk iNview* é um produto combinado que consiste em um aplicativo móvel e um acessório de oftalmoscópio indireto compatível com smartphone. É um aparelho que consegue captar um campo de visão retiniano de 50°, todavia, assim como o modelo anterior, foi descartado dessa pesquisa devido à necessidade mínima de dilatação pupilar.

Por fim, foram analisados os aparelhos *D-eye* e *panoptic* que, segundo os seus fabricantes, estes aparelhos não requerem dilatação da pupila. Por esse motivo, foram adquiridos previamente para este estudo e descritos separadamente a seguir.

#### 4.1.1 D-eye

O dispositivo *D-eye*, apresentado na Figura 4.3, é semelhante ao sistema *Peek Retina* e começou como um estudo comparativo prospectivo de base clínica para diabéticos. O sistema usa a câmera e a fonte de luz do smartphone, iluminando o interior do olho para o exame. Como o *D-eye* reflete a lanterna do smartphone para a retina usando seu design óptico com espelhos e lentes, ele não requer energia externa e fontes de luz adicionais.



Figura 4.3: Dispositivo *D-eye* e armação para acoplar-se em um smartphone.

As patologias comuns observadas são as doenças relacionadas ao nervo óptico, como glaucoma e retinopatia diabética. Durante o exame, o tipo de arquivo pode ser selecionado (vídeo ou foco automático do disco óptico). Depois de concluído, o exame pode ser armazenado no aplicativo *D-eye* para avaliações posteriores.

Em olhos sem dilatação da pupila, podem ser adquiridas imagens de até 8° de campo de visão da retina, conforme o exemplo mostrado na Figura 4.4, que consiste em uma fotografia capturada pelo próprio autor enquanto selecionava o equipamento a ser usado para construir o conjunto de dados. A imagem foi capturada em uma sala com pouca iluminação para maximizar a dilatação da pupila.

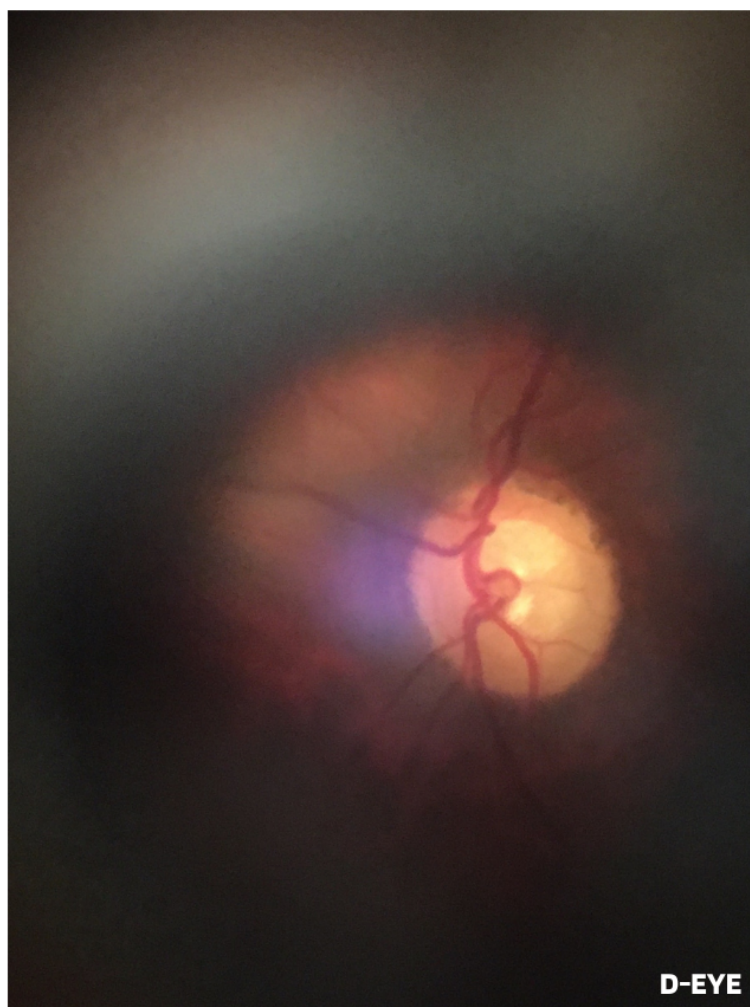


Figura 4.4: Imagem do olho esquerdo tirada com smartphone e dispositivo *D-Eye* - Fonte: próprio autor.

#### 4.1.2 Panoptic Welch Allyn

O oftalmoscópio *panoptic Welch Allyn* 11820 apresentado como o mesmo modelo mostrado na Figura 4.5, é um dispositivo que permite trabalhar com o aplicativo *IExaminer* para exames de fundo de olho.

O aplicativo *iExaminer* transforma o oftalmoscópio *panoptic* em um dispositivo móvel de imagem digital, que permite aos usuários visualizar e fotografar o fundo do olho por meio de um smartphone. O seu design óptico produz luz própria e facilita o acesso às pupilas pequenas com boa iluminação de fundo, permitindo fotografar sem dilatação pupilar. Para captar as fotografias, o oftalmoscópio pode ser alimentado por bateria (original de 3,5 volts), proporcionando um campo de visão de até 25° com ajuste de foco de -20 a +20 dioptrias (WelchAllyn, 2020).



Figura 4.5: Exemplo do oftalmoscópio *panoptic Welch Allyn*.

O oftalmoscópio também pode ser alimentado por baterias alcalinas, podendo ser utilizado em áreas remotas. Esse equipamento usa uma lâmpada halógena para iluminação, fornecendo um zoom de 5x para imagens adquiridas através de pupilas não dilatadas.

Para visualizar a retina, o oftalmoscópio geralmente possui três tamanhos de luz para usar dependendo do nível de dilatação da pupila, conforme mostrado na Figura 4.6:

- **Grande:** a luz grande é mais apropriada para ser utilizada após aplicação de colírios midriáticos para dilatar a pupila.
- **Médio:** esta luz é comumente utilizada para realização de fotografias em pupilas escuras e não dilatadas.
- **Pequeno:** a luz pequena é utilizada quando a pupila está muito contraída, ou seja, sala bem iluminada, sem uso de dilatadores de pupila.

Além das configurações da fonte de luz, os filtros apresentados pelo aparelho podem ser usados para destacar patologias específicas:

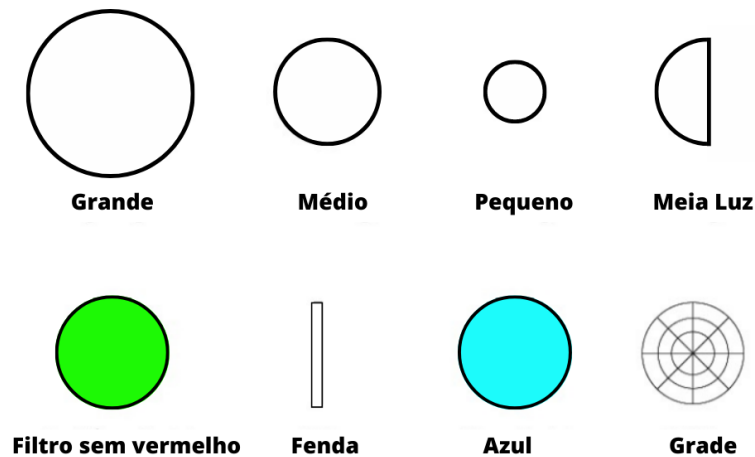


Figura 4.6: Exemplo de filtro de abertura de luz do *panoptic*.

- **Meia luz:** caso a pupila esteja parcialmente obstruída por uma lente com catarata, o semicírculo pode ser usado para passar a luz apenas pela parte clara da pupila.
- **Filtro sem vermelho:** também chamado de filtro verde, é usado para visualizar os vasos e hemorragias com mais detalhes, melhorando o contraste das imagens. Essencialmente, esse filtro bloqueia o "ruído" visual, permitindo visualização com maior contraste entre estruturas e tecidos. Isso é especialmente importante em exames de fundo de olho, onde a cor predominante do fundo da retina é um tom avermelhado. Essa configuração fará com que a retina pareça preta e branca.
- **Feixe de fenda:** usado para examinar anormalidades de contorno da córnea, cristalino e retina.
- **Filtro azul-cobalto:** alguns oftalmoscópios possuem esse recurso que pode ser utilizado para observar abrasões e úlceras da córnea após a coloração com fluoresceína.
- **Grade:** usado para fazer aproximações da distância relativa entre as lesões da retina.

### 4.1.3 Oftalmoscópio selecionado

Para selecionar o oftalmoscópio usado para este estudo, primeiramente, o aparelho *D-eye* foi comparado com o aparelho *panoptic*. Finalmente, o *panoptic* foi escolhido pelos seguintes motivos:

- O aparelho *panoptic* apresenta imagens com ângulo de visão mais ampliado em relação ao *D-eye*.
- As imagens do *panoptic* são mais nítidas visualmente.

- O aparelho é mais robusto e possui diversos recursos que permitem uma melhor visualização das áreas de interesse.

Como exemplo da qualidade das imagens dos dispositivos estudados, a Figura 4.7 mostra uma imagem comparativa entre o uso do aparelho *D-eye* e do *panoptic*.

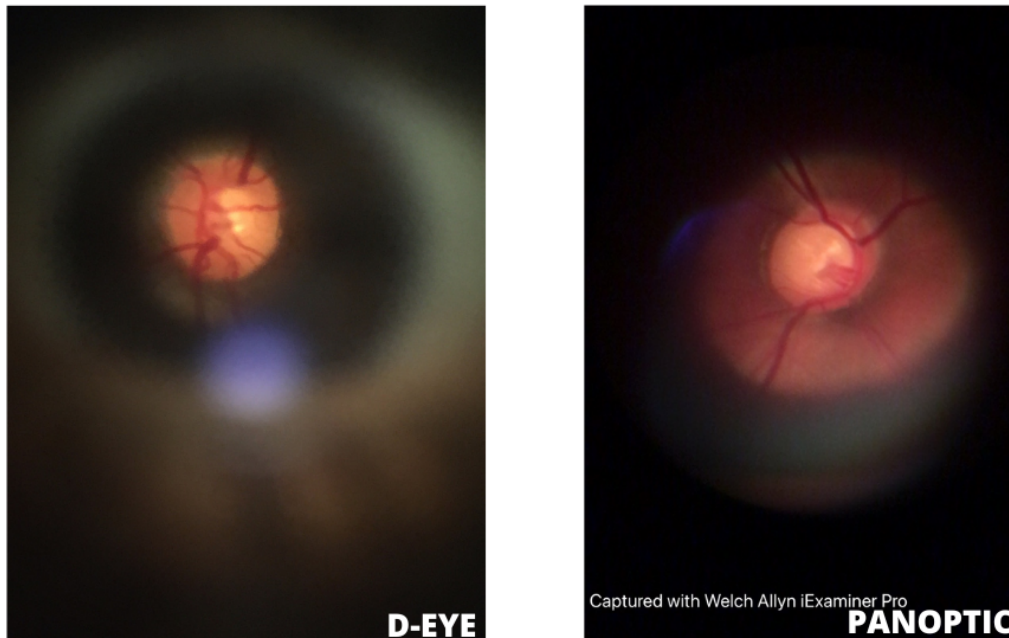


Figura 4.7: Exemplo comparativo de imagens obtidas com o *D-eye* e *panoptic* - próprio autor.

A oftalmoscopia é um exame do segmento posterior do olho, incluindo o vítreo, a retina, a coróide e o nervo óptico. Este tipo de exame é facilitado pela instilação de colírio dilatador para dilatar a pupila. Portanto, existem pelo menos dois fatores que afetam a capacidade de ver a retina através da oftalmoscopia, o primeiro tem a ver com o ângulo de visão que o aparelho consegue focar, e o segundo é a dilatação pupilar, pois esta tem a capacidade de afetar à abertura do ângulo de visão. Esses artefatos são notórios quando se comparam imagens do *panoptic* com outras imagens obtidas por retinógrafos convencionais, cuja alta resolução e amplo campo de visão de toda a área retiniana se destacam.

A Figura 4.8 mostra a diferença entre uma imagem do *panoptic* acoplado ao smartphone, com ângulo de visão de aproximadamente 25° e centradas na região do disco óptico e uma imagem global, ou seja, que mostra toda a região da retina obtida com ângulo de visão de 45°.

Após a escolha e a descrição do equipamento selecionado para a aquisição das imagens de fundo de olho, a seção seguinte descreve as regiões e locais onde foram realizadas todas as fotografias que compõe a base de dados BrG.

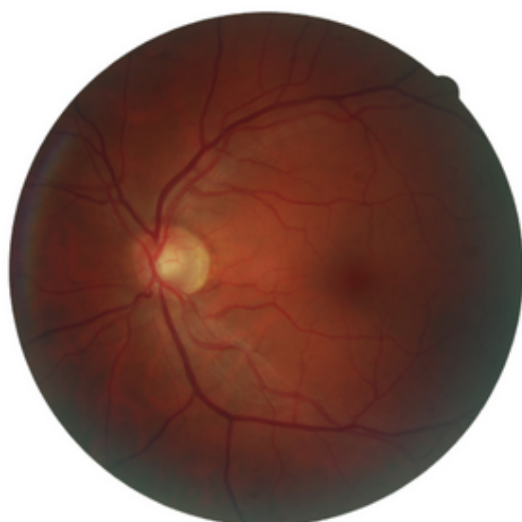


Imagem global - Base de dados HRF

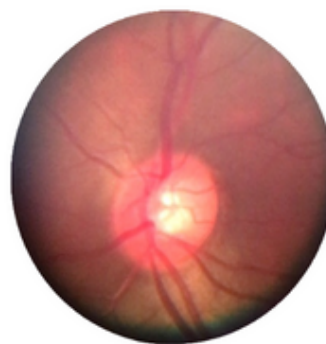


Imagem do fundo de olho Base de dados BrG

Figura 4.8: Diferença entre uma imagem com ângulo de visão de  $45^\circ$  que registra toda a região da retina e uma imagem obtida com o *panoptic* com ângulo de visão de  $25^\circ$ .

## 4.2 Locais de coleta das imagens

O estudo foi submetido e analisado pelo Comitê Nacional de Ética do Brasil, conforme CAAE: 29983120.0.0000.8078 – Número: 4056930. As imagens digitais de fundo de olho do conjunto de dados estabelecido neste estudo foram obtidas de dois locais diferentes, conforme mostrado na Figura 4.9: Hospital de Olhos (HO), do Sul de Minas Gerais (MG) e a policlínica da cidade de Unaí, ao noroeste de MG, a saber:

### 4.2.1 Hospital de olhos HO

Objetivando apresentar o local de coleta das imagens glaucomatosas de fundo de olho e também uma oportunidade de conhecer um modelo real de diagnóstico e tratamento do glaucoma nos moldes tradicionais, apresenta-se o HO (HO, 2022), um hospital privado com sede em Itanhandu (MG), Brasil. Esse hospital possui um programa de tratamento de glaucoma que abrange a região Sul do Estado de MG, com aproximadamente 2 milhões de habitantes, e mantém convênio com o SUS do governo brasileiro (da Saúde, 2022). As diretrizes do HO seguem as orientações do Ministério da Saúde, atendendo pacientes diagnosticados com glaucoma em outras unidades de saúde da região ou por meio de cotas de triagem disponibilizadas pelo próprio hospital e creditadas pelo SUS.

O atendimento oftalmológico realizado pelo HO em parceria com o SUS é descentralizado da unidade hospitalar da cidade de Itanhandu para outras cidades que compõem a região Sul de MG. As cidades de referência para as fases de aplicação e coleta de dados deste estudo são: Itajubá, Pouso Alegre e Três Corações.

Os procedimentos clínicos oferecidos pelo hospital são realizados trimestralmente em

## Divisão macro e microrregiões Estado de Minas Gerais - Brasil

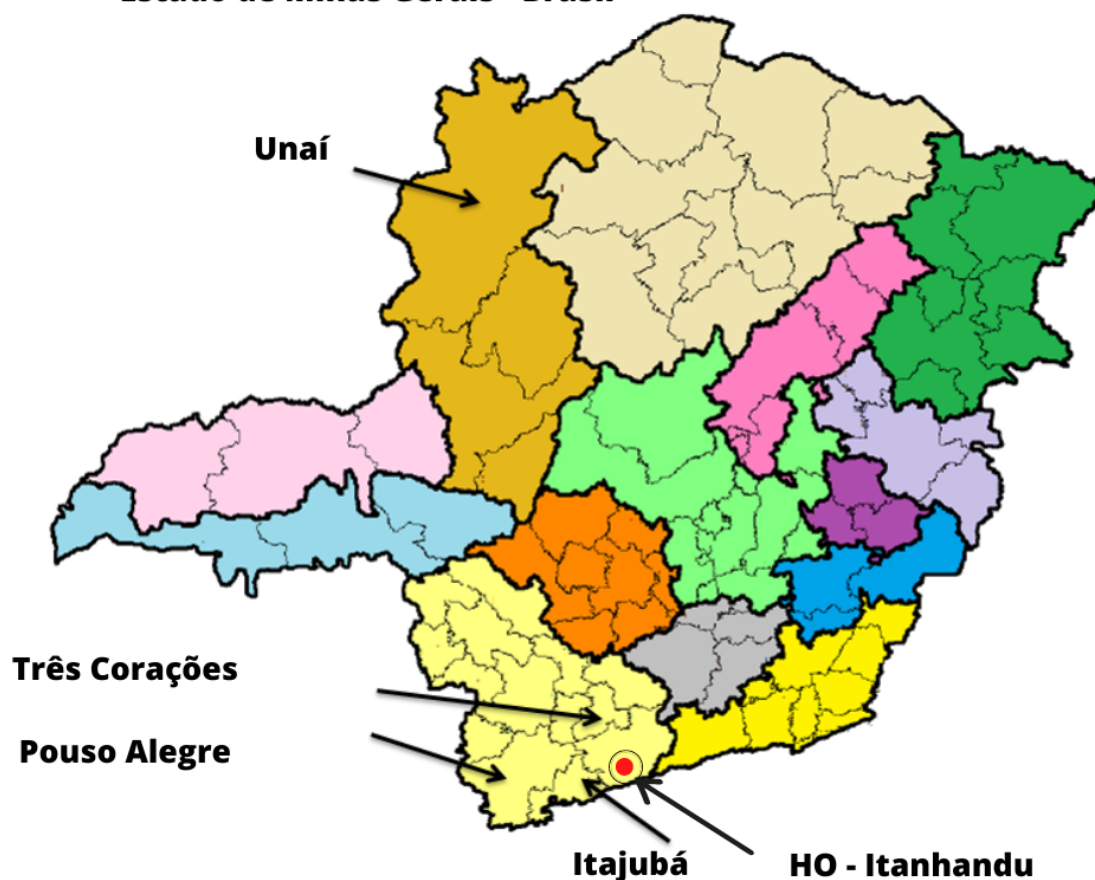


Figura 4.9: Mapa do Estado de Minas Gerais com apresentação das cidades da região do Sul e a cidade de Unaí ao Noroeste do Estado.

unidades hospitalares de cada município que recebe o programa de diagnóstico e tratamento do glaucoma. Assim, para atender toda a região Sul de MG, o HO conta com uma equipe móvel que se desloca para as cidades mais importantes da macrorregião. A equipe móvel é composta por 5 profissionais: 1 administrador e organizador de prontuários, 1 técnico em enfermagem para auxiliar nas consultas oftalmológicas, 1 operador para o exame de perimetria e pela distribuição de medicamentos, além de 2 oftalmologistas que trabalham em turnos rotativos.

De acordo com as obrigações legais, todos os pacientes atendidos pelo HO recebem em parceria com o SUS os seguintes exames a cada três meses: anamnese, medição da acuidade visual, medição da PIO, exame de paquimetria ultrassônica e avaliação do nervo óptico com auxílio de uma lâmpada de fenda. Além destes exames trimestrais, o hospital também oferece o exame de perimetria de caráter anual ou quando necessário para medir a qualidade visual dos pacientes. Desta forma, a aplicação do protocolo de atendimento ao paciente com glaucoma foi dividida em 5 etapas conforme se apresenta no fluxograma da Figura 4.10, a saber:

- 
- Etapa 1: O administrador do programa recebe o paciente, confere dados referentes à necessidade e legitimidade do atendimento e o transfere para o técnico em enfermagem.
  - Etapa 2: O técnico em enfermagem aplica dilatador pupilar, mede a PIO com auxílio de um tonômetro, realiza o exame de paquimetria e encaminha o paciente com os resultados dos exames ao oftalmologista.
  - Etapa 3: O oftalmologista realiza uma consulta clínica e examina o fundo de olho com o auxílio de uma lâmpada de fenda; após o exame, em casos de primeira consulta e não se confirmar o glaucoma, o oftalmologista encaminha o paciente para a Etapa 5 de encerramento. Em casos de confirmação do glaucoma na primeira consulta ou nos casos em que o paciente já esteja em tratamento da doença no HO, ele é encaminhado à Etapa 4.
  - Etapa 4: nessa etapa, caso seja recomendado clinicamente, o paciente fará um teste de campo visual e, em seguida, receberá colírios conforme as recomendações do oftalmologista, em quantidade necessária para utilização até a data da consulta subsequente de acompanhamento trimestral.
  - Etapa 5: nesta última etapa, agenda-se o retorno do paciente e encerra-se o atendimento.

Nos estágios iniciais do glaucoma, muitas vezes é difícil fazer um diagnóstico definitivo. Portanto, os pacientes que mesmo sem o diagnóstico conclusivo do glaucoma apresentem: PIO elevada ou/e disco óptico com aparência suspeita ou defeito inexplicável do campo visual, podem ser classificados como suspeitos de glaucoma (Perkins, 1994). Por conseguinte, nos casos suspeitos, o diagnóstico definitivo só pode ser feito com acompanhamento longitudinal do paciente e, muitas vezes, requer a repetição dos testes e exames adicionais. Esses casos são tratados de acordo com o entendimento médico até que um diagnóstico definitivo justifique a sua inclusão ou exclusão do plano de tratamento do HO.

O HO acolhe para tratamento e acompanhamento trimestral do glaucoma aqueles que apresentam pelo menos dois dos seguintes diagnósticos: PIO média não tratada acima de 21 mmHg, lesão típica do nervo óptico com perda da borda neurorretiniana identificada com auxílio de uma lâmpada de fenda e CDR igual ou superior a 0,5, ou campo visual compatível com danos glaucomatosos.

A forma como esse tratamento de glaucoma é oferecido é atípica no Brasil e, programas como esse estão disponíveis em pouquíssimas regiões do país. Conseqüente, o programa de tratamento oferecido por este hospital não representa os serviços médicos oftalmológicos prestados diretamente pelo SUS, mas sim os serviços médicos de um hospital privado que abre as suas portas para prestar serviços oftalmológicos gratuitos ao público

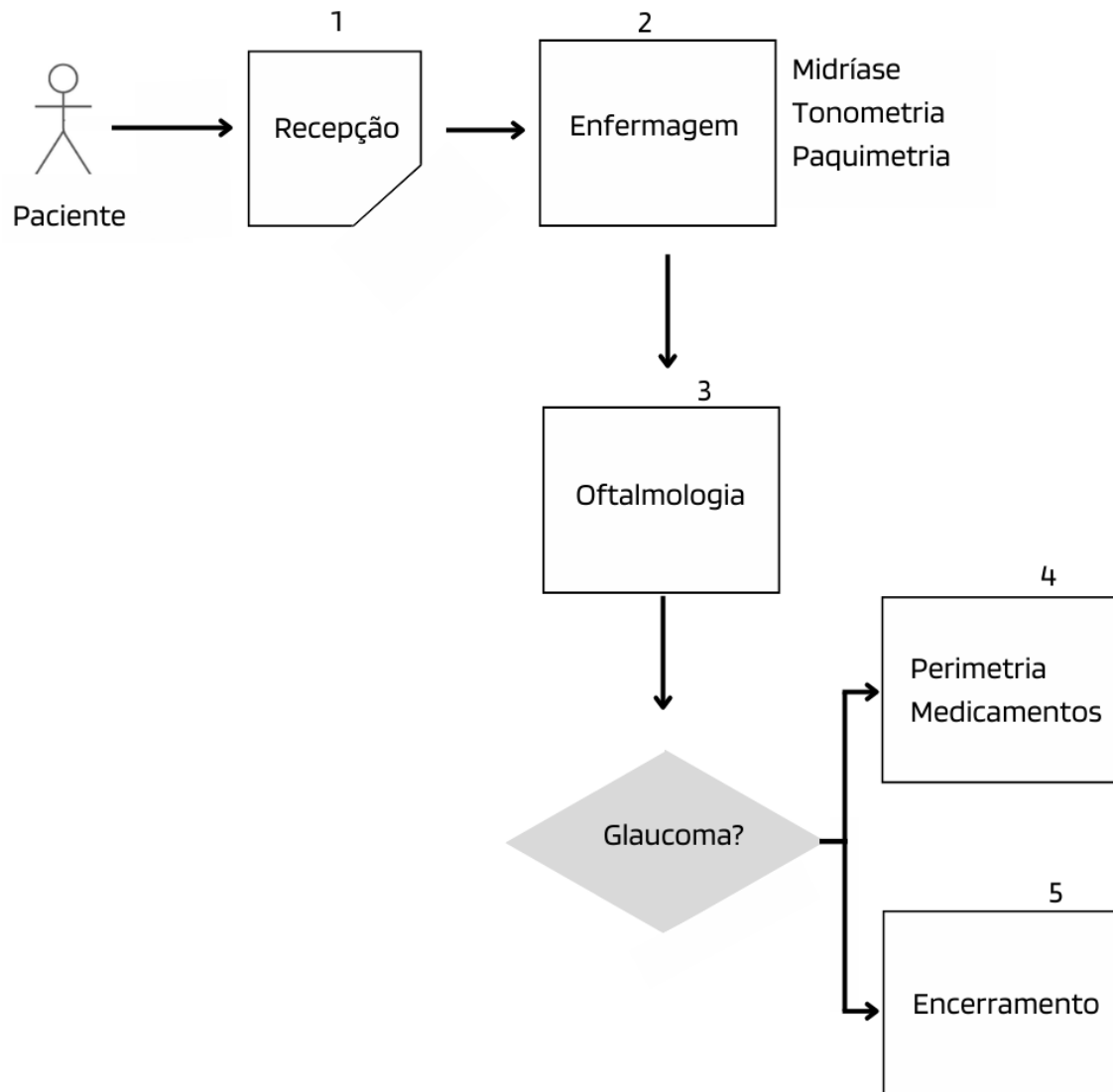


Figura 4.10: Fluxograma do atendimento aos pacientes do programa de tratamento do glaucoma HO.

com receitas pagas pelo SUS. Esse é um programa de grande ajuda para a população local, pois permite que os oftalmologistas disponíveis acompanhem a evolução e o tratamento do glaucoma nos pacientes regionais. Além disso, oferece distribuição gratuita de medicamentos para o tratamento da doença.

Uma das razões possíveis para a não oferta desses serviços à população em geral pode ser a falta de incentivo dos hospitais privados para oferecer serviços pagos pelo SUS que, além da burocracia de implementação, os repasses do SUS podem ser menos interessantes comercialmente do que os lucros com serviços pagos pelo setor privado.

Quanto à metodologia de triagem, embora o HO ofereça vagas para novas consultas, o programa não se enquadra como um modelo de triagem populacional, pois, não consegue oferecer atendimento a todos os casos necessários em uma determinada região ou população. Mesmo considerando apenas a parcela de maior risco de desenvolver a

---

doença, como a população de pessoas com mais de 40 anos de idade e portadores de outras doenças oculares que possam aumentar o risco de desenvolver o glaucoma, além disso, o hospital não estabeleceu regulamentos para triagem populacional.

#### **4.2.2 Policlínica de Unaí/MG**

Para esta tese, imagens de pacientes sem glaucoma foram coletadas durante consultas oftalmológicas eletivas no centro municipal de saúde policlínica na cidade de Unaí, MG, Brasil. A clínica é de caráter público em convênio com o SUS e oferece atendimento médico e oftalmológico à população em geral. A policlínica de Unaí está localizada na região noroeste do Estado de MG, cerca de 900 quilômetros de distância do Sul do Estado, onde se localiza o HO.

A diferença entre o programa de tratamento oferecido pelo HO e o programa de consultas oferecido pela policlínica de Unaí é o objetivo pretendido. No entanto, como o objetivo da policlínica de Unaí é oferecer consultas eletivas mais gerais, os exames incluíram apenas refração, medida da PIO, acuidade visual e exame de fundo de olho com auxílio de uma lâmpada de fenda.

O atendimento na policlínica é generalista no que diz respeito ao glaucoma, por este motivo, os pacientes considerados suspeitos nas consultas eletivas são encaminhados para outras unidades de saúde para realização de exames complementares como OCT, retinografia ou exame de campo visual.

Apresentado os locais de coleta de imagens para o objetivo desta pesquisa, a próxima seção descreve como foram adquiridas as imagens que compõem a base de dados utilizados neste trabalho.

### **4.3 Aquisição das imagens de fundo de olho**

Neste estudo, 1.000 voluntários tiveram o fundo de olho fotografado. Essas pessoas foram divididas em 500 pacientes com glaucoma (atendidos no programa nacional de glaucoma) oferecido pelo HO, e 500 pacientes sem glaucoma (que tiveram os seus olhos examinados na policlínica municipal de Unaí/MG). Todos os voluntários tiveram os dois olhos (esquerdo e direito) fotografados. Assim, um total de 2.000 imagens de fundo de olho foi obtido.

Os pacientes que consentiram voluntariamente em participar do estudo tiveram os seus olhos fotografados por um profissional não médico usando um smartphone acoplado ao oftalmoscópio *panoptic*.

O smartphone usado no estudo para realização de todas as fotografias foi um dispositivo da marca *Apple* modelo *iPhone 6s* com uma câmera de 12 *megapixels*. Assim, para tirar as fotos, um pequeno vídeoclipe foi gravado, em seguida, convertendo em fo-

---

tografias, as cinco melhores imagens foram pré-selecionadas manualmente com base no foco ideal e na visualização da vasculatura e, finalmente, a melhor entre as cinco imagens, foi selecionada.

Ambos voluntários, com ou sem glaucoma, tiveram o fundo de olho fotografado enquanto aguardavam atendimento oftalmológico no HO ou policlínica de Unaí. As imagens foram adquiridas usando a representação do sistema de cores RGB, e salvas no formato *Joint Photography Experts Group* (JPEG).

Todas as imagens foram tiradas com os olhos não dilatados, usando o oftalmoscópio centrado no disco óptico, com um campo de visão de aproximadamente 25°. As fotografias foram realizadas utilizando lâmpada halogena HPX 03800-U e bateria de 3,5 volts com autonomia de fotografar o fundo de olho de até 50 voluntários por dia - quantitativo máximo conseguido durante a realização da pesquisa.

### **4.3.1 Imagens glaucomatosas**

Imagens glaucomatosas foram coletadas de pacientes brasileiros atendidos pelo HO da zona sul de MG (HO, 2022). Foram incluídos pacientes que apresentaram as seguintes características:

- Faixa etária entre 18 e 80 anos de idade independentemente de gênero ou raça;
- Prova de glaucoma com diagnóstico completo fornecido pelo HO;
- Concordância com a participação na pesquisa científica.

Cada imagem de fundo de olho adquirida recebeu um rótulo de "sim" para o glaucoma com base nos achados clínicos do HO.

### **4.3.2 Imagens não glaucomatosas**

As imagens de pacientes sem glaucoma foram coletadas durante consultas eletivas de oftalmologia no centro de saúde policlínica na cidade de Unaí. Foram incluídos pacientes que apresentaram as seguintes características:

- Faixa etária entre 18 e 80 anos de idade independentemente de gênero ou raça;
- Teste negativo para glaucoma baseado em consulta oftalmológica auxiliada por tonometria e exame de fundo de olho com lâmpada de fenda;
- Concordância com a participação na pesquisa científica.

Para as imagens não glaucomatosas, a verdade absoluta de cada imagem rotulada como "não" para o glaucoma teve a confirmação de oftalmologistas responsáveis pelas consultas locais.

### 4.3.3 Dificuldades na aquisição das imagens

A Figura 4.11, representam imagens fundoscópicas obtidas com algum tipo de dificuldade.

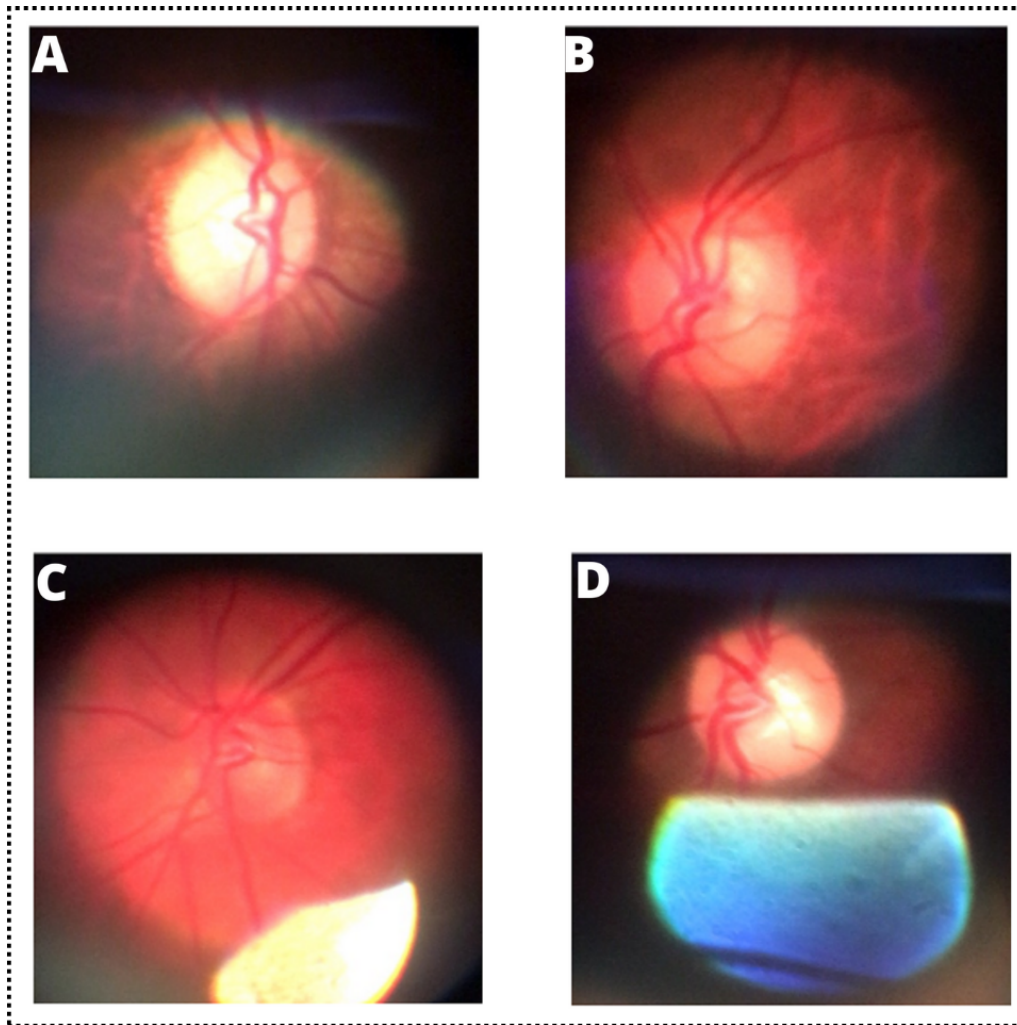


Figura 4.11: ( A ) imagem com pequena abertura do olho, ( B ) disco óptico descentralizado, ( C ) presença de catarata, ( D ) ruído de foco de luz e pouca abertura do olho.

Algumas dificuldades foram encontradas na obtenção das imagens do fundo de olho, algumas decorrentes de limitações inerentes ao equipamento utilizado, como autonomia das baterias, dificuldade em focar o disco óptico em determinados voluntários e dependência de locais pouco iluminados para melhor qualidade das imagens. Outras dificuldades foram relacionadas à falta de cooperação dos voluntários na realização do exame, como dificuldade para abrir os olhos ou focalizar um ponto fixo durante a realização das fotografias.

Controlar o ambiente de fotografia é considerado uma recomendação básica para obtenção de imagens com maior contraste, mas no programa de tratamento do glaucoma do HO ou na policlínica de Unai, gerenciar o local foi considerado uma tarefa difícil, pois a sala de aquisição das fotografias era compartilhada com outros profissionais.

A prática do operador do oftalmoscópio é essencial para obter imagens de alta qualidade, mas o paciente voluntário deve contribuir e manter os olhos o mais abertos possível e sempre fixar o olhar em um ponto determinado durante a aquisição das fotografias. Entretanto, muitos pacientes não conseguiram abrir totalmente os olhos ou eram muito sensíveis à luz do aparelho. Devido a essas objeções, cerca de 30 voluntários desistiram de participar do estudo.

#### 4.3.4 Imagens com ruídos

A Figura 4.12 mostra exemplos de imagens contendo alguns tipos de ruídos .

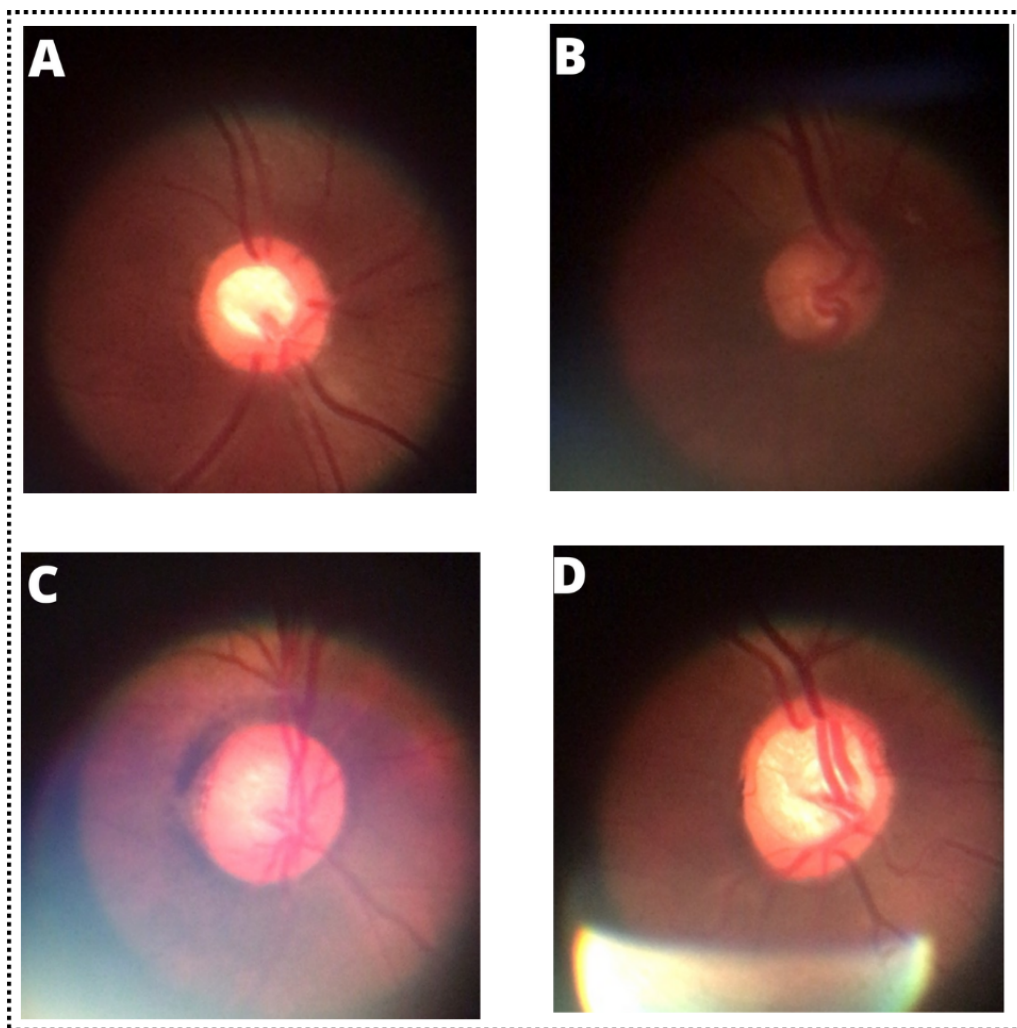


Figura 4.12: Principais tipos de ruídos presentes nas imagens do oftalmoscópio *panoptic*: ( A ) imagem ideal, ( B ) pouca iluminação, ( C ) ruído externo, ( D ) ruído de foco de luz.

Durante a aquisição das imagens fundoscópicas com o *panoptic*, descobriu-se que, além das dificuldades apresentadas, existem pelo menos três tipos potenciais de ruídos que podem interferir na qualidade das imagens e, conseqüentemente, na precisão geral dos algoritmos de *deep learning*:

- 
1. O primeiro tipo de ruído, ou apenas uma característica, está relacionado ao baixo contraste e à aparência de algumas imagens mais escuras do que outras. Esse efeito pode ocorrer devido à alimentação fornecida ao dispositivo, sendo feita por meio de uma bateria recarregável de 3,5 volts. Portanto, ao trabalhar continuamente, as primeiras imagens podem aparecer com iluminação mais alta, enquanto as imagens subsequentes podem aparecer com iluminação mais baixa. Embora os dispositivos panópticos tenham ajustes de iluminação, controlar esses efeitos é difícil.
  2. O segundo tipo de ruído surge da iluminação externa. Esse ruído ocorre quando a iluminação ambiente não pode ser controlada. Para reduzir esses efeitos, o oftalmoscópio possui um protetor ocular que bloqueia a luz externa e melhora o contraste da imagem. Porém, dependendo da posição do rosto ou da fisionomia de algumas pessoas, essa blindagem pode permitir a passagem de luz externa, o que pode causar efeitos indesejados.
  3. O terceiro e mais frequente tipo de ruído é obtido com a luz do próprio aparelho. Especificamente, quando apontado para um ângulo inadequado, o aparelho pode causar reflexos que podem prejudicar as imagens finais.

As Figuras 4.12 e 4.11 mostram os principais tipos de erros e ruídos que podem ocorrer na aquisição de imagens com o *panoptic*. Diante de tais possíveis erros e ruídos, para a construção da base de dados, o processo de considerar apenas as melhores imagens do equipamento para compor o conjunto de dados é uma atitude de escapismo do mundo real, porém, é permissível excluir ao menos aquelas imagens com ruídos oriundos da má prática do operador do oftalmoscópio. Desta forma, foram descartadas aproximadamente 30 imagens com mau posicionamento da região do disco óptico, devido a descuidos do operador do oftalmoscópio e imagens de baixo contraste por motivos de má alimentação do aparelho, ou seja, com ruídos causados por baterias em estágio de descarregamento de energia.

## 4.4 Formação da base de dados - BrG

Para construir o conjunto de dados BrG, foi recortada, primeiramente, a parte central de cada imagem original do fundo de olho adquirida neste estudo. Esta região foi extraída de forma automática, eliminando a área preta circundante, obtendo assim uma imagem de aproximadamente  $800 \times 800$  *pixels*, conforme mostrado na Figura 4.13.

Essa incisão teve uma margem de segurança para não danificar a região de maior interesse que é a área do disco óptico. Portanto, o corte exclui apenas as bordas horizontais da imagem, bem como os terços superior e inferior, correspondente a uma área retangular sobreposta à região do disco óptico. Após a aplicação da extração das bordas em todas as

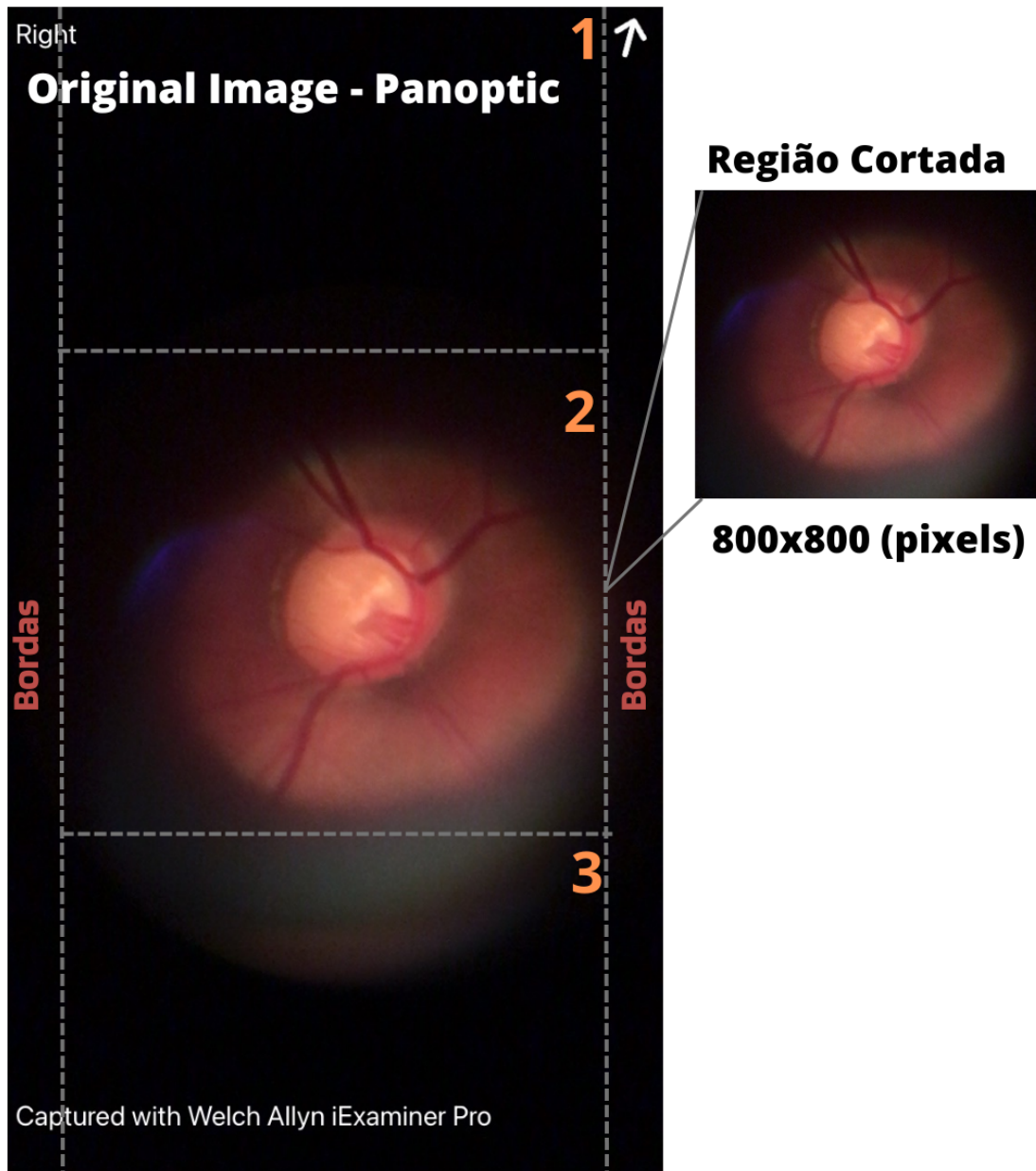


Figura 4.13: Tamanho da imagem de  $720 \times 1280$  pixels com corte central de  $800 \times 800$  pixels na região que representa o disco óptico.

imagens de fundo de olho, adquiridas neste trabalho, os recortes foram salvos no formato PNG, resultando em um novo conjunto de dados, que está disponível publicamente e denominado Brasil Glaucoma (BrG).

As imagens não sofreram nenhum tipo de processamento posterior à extração das bordas. Além disso, elas foram anonimizadas de dados pessoais e, para cada imagem no conjunto de dados, há uma máscara do disco óptico e uma máscara do copo óptico que podem ser usadas por algoritmos de segmentação, conforme mostrado na Figura 4.14. As máscaras foram obtidas por segmentação manual, em seguida foram revisadas por um oftalmologista experiente.



Figura 4.14: Exemplo de uma imagem original seguida por suas máscaras de segmentação padrão ouro para o glaucoma.

Uma característica relevante apresentada nas imagens do conjunto de dados BrG é que elas não foram divididas considerando os estágios do glaucoma. No entanto, é possível que haja um equilíbrio na base de dados considerando a gravidade dos danos causados por essa neuropatia óptica, que podem ser estimados de acordo com as seguintes categorias (Gedde et al., 2021):

- Leve: disco óptico definido, exibição anormal da CFNR, porém campo visual normal no teste de perimetria automatizado padrão;
- Moderado: disco óptico definido, exibição anormal da CFNR e campo visual apresentando anomalias no teste de perimetria automatizado padrão;
- Grave: anormalidades definitivas do disco óptico e CFNR, ou imagem macular consistentes com glaucoma e anormalidades do campo visual em ambos os hemisférios;
- Indeterminado: disco óptico definido, CFNR ou anormalidades de imagem macular consistentes com glaucoma e incapacidade do paciente de realizar testes de campo visual, dando assim resultados não confiáveis ou não interpretáveis.

A diversidade dos estágios do glaucoma inferida na base de dados BrG deve-se às campanhas populacionais propostas pelo HO para combater a doença. Assim, as pessoas são motivadas e instruídas a procurar mais vezes o oftalmologista, potencializando o diagnóstico precoce da doença. Portanto, a base de dados BrG é composta por imagens de pacientes que procuraram o HO por necessidade, ou seja, já apresentavam danos estruturais e funcionais que comprometiam a visão e, assim, procuraram atendimento oftalmológico. No entanto, outros pacientes também procuraram atendimento pela maior disponibilidade de consultas para a população regional e desta forma foram atraídos pelas campanhas de combate ao glaucoma, e tiveram o diagnóstico precocemente aplicado, ou seja, antes que o dano funcional comprometesse a sua qualidade de vida.

---

Considerando, portanto, o tempo de implementação do programa de consulta e tratamento de glaucoma pelo HO, o impacto e estabilidade da descoberta de novos casos da doença na região sul de MG, é possível inferir que o conjunto de dados BrG é constituído com uma distribuição mais uniforme entre os estágios ou gravidade da doença.

Das 2.000 fotos BrG, as estimativas de CDR são dadas para 160 pares de imagens (olho direito e olho esquerdo). Essa medida é importante porque na maioria das pessoas a relação escavação/disco entre os olhos é simétrica, e a assimetria é um sinal importante de suspeita de lesão glaucomatosa. Essas medidas foram aplicadas manualmente pelo oftalmologista do HO com o auxílio de um microscópio binocular lâmpada de fenda. Embora esses valores sejam imprecisos, eles podem ser importantes para estudos que buscam métodos para analisar a assimetria da depressão do disco óptico entre os olhos, como método alternativo para auxiliar no diagnóstico do glaucoma.

Por fim, as características mais relevantes da base de dados BrG podem ser resumidas da seguinte forma:

- A base de dados possui imagens com etnia representada pela população brasileira composta por negros, brancos e mestiços;
- Possui 2.000 imagens distribuídas entre os olhos direito e esquerdo. Apresenta 50% dos voluntários com glaucoma e 50% dos voluntários sem glaucoma;
- As imagens foram rotuladas com base em achados clínicos e com orientação de 2 oftalmologistas da policlínica da cidade de Unaí e 2 oftalmologistas do HO;
- Possui imagens obtidas com apenas uma câmera de retina (smartphone), acoplado em um oftalmoscópio portátil com ângulo de visão de até 25 ° centrado no disco óptico;
- Contém número aproximadamente igual de homens e mulheres entre 18 e 80 anos de idade;
- É uma base de dados disponível publicamente para pesquisas científicas, disponível em <https://www.globaleyeh.com>;

Diante da nova base de dados construída, no próximo capítulo, "Experimentos", serão conduzidos testes para medir a eficácia de algoritmos de *deep learning* na classificação das imagens BrG em "sim" ou "não" para o glaucoma.

# Capítulo 5

## Experimentos

Para o desenvolvimento deste experimento, primeiramente, observou-se que vários trabalhos relacionados foram propostos, demonstrando a capacidade dos algoritmos de IA em identificar padrões da doença ocular glaucoma em imagens digitais do fundo de olho. Embora esses algoritmos não tenham sido aplicados na prática clínica real, eles demonstram a poderosa capacidade dos métodos de *deep learning* na classificação de imagens glaucomatosas, ou seja, demonstraram forte capacidade de separar imagens glaucomatosas de imagens normais com alta precisão em conjuntos de dados utilizados como testes.

Em todos os trabalhos relacionados, as imagens da retina foram adquiridas com grandes câmeras oftalmológicas convencionais, conhecidas como retinógrafos. No entanto, com a evolução tecnológica, imagens fundoscópicas também passaram a ser obtidas com auxílio de smartphones e oftalmoscópios portáteis, reduzindo, consideravelmente, os custos de aquisição destas imagens médicas, além de disponibilizar novos tipos de imagens fundoscópicas que possam ser utilizadas em testes de classificação do glaucoma a partir de algoritmos de *deep learning*.

Dado este contexto, o objetivo principal desta tese foi desenvolver um algoritmo *deep learning* e analisar o comportamento da técnica algorítmica na identificação do glaucoma a partir de imagens digitais da retina adquiridas por meio de smartphones. O foco desta análise é entender como os algoritmos funcionam em imagens de resolução mais baixa adquiridas por esses novos meios. Além disso, essa análise é importante porque, uma vez que o glaucoma pode ser identificado em imagens com baixo custo de aquisição, softwares poderão ser propostos para o diagnóstico automático do glaucoma a partir dessas imagens oftálmicas, o que, aliás, também terá custos reduzidos. O propósito da redução de custo diagnóstico é a possibilidade de viabilizar um programa de triagem populacional e automatizada para o glaucoma.

Para realização deste experimento, primeiro foi construída a base de dados BrG, conforme descrita no capítulo anterior desta tese. Portanto, este estudo foi aplicado sobre a base de dados BrG, contendo 2.000 imagens rotuladas para glaucoma. Para melhor entendimento da ideia geral deste trabalho, os termos rastreamento ou triagem são descritos

como o ato de encontrar, detectar ou identificar casos de glaucoma em uma população. O termo classificação refere-se à tarefa de identificar se uma imagem digital do fundo de olho pertence à categoria glaucomatosa ou não glaucomatosa.

Dado o contexto, este capítulo descreve o uso de um algoritmo *deep learning* na classificação de imagens do conjunto de dados BrG por meio de um *ensemble*. O ensemble foi construído a partir da concatenação de redes CNNs, no entanto, antes de apresentá-lo, são descritos os algoritmos individuais que o compõem. No final do capítulo, são apresentados os resultados da classificação do glaucoma com base em diversos cortes da região de interesse sobre as imagens BrG. Deste modo, na ordem cronológica da pesquisa aplicada, o organograma da Figura 5.1 demonstra as etapas percorridas neste capítulo.

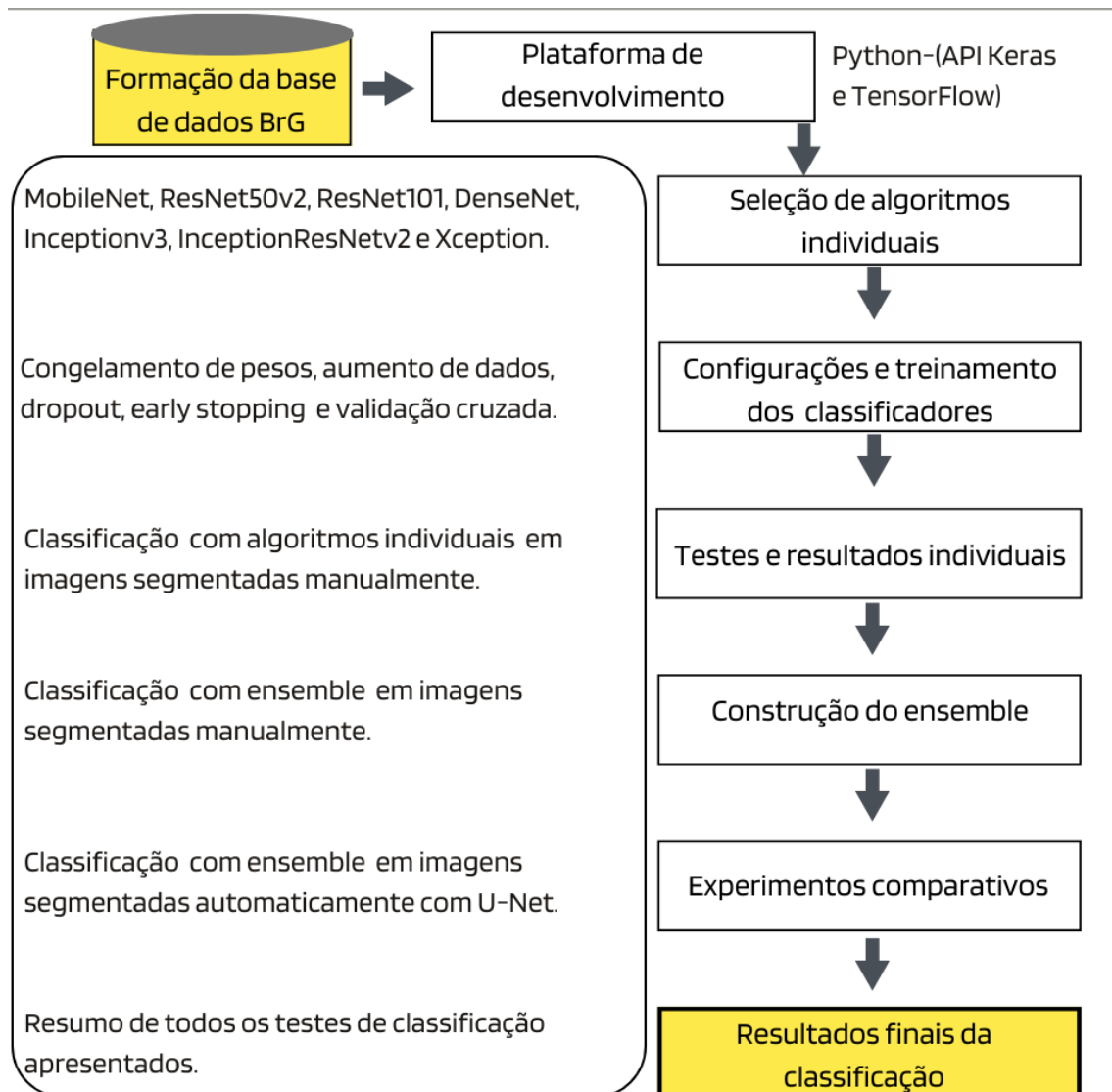


Figura 5.1: Etapas da metodologia aplicada neste trabalho.

---

## 5.1 Plataforma de desenvolvimento

Como forma de facilitar o desenvolvimento das redes CNNs, para implementar o código, são utilizadas as interfaces de programação *Keras* (Chollet et al., 2018) e *Tensorflow*. Para contextualizar, *TensorFlow* é uma biblioteca de código aberto para *machine learning* lançada em 2015. A *tf.keras* é uma *Application Programming Interface* de alto nível do *TensorFlow* para criar e treinar modelos de aprendizado profundo, é baseada na linguagem de programação *python* e fornece as funções e métodos para utilização de RNAs com poucas linhas de código. As ferramentas *Tensorflow*, *Keras* e *Python* foram utilizadas através da plataforma *Google Colaboratory*, um serviço na nuvem cujo objetivo é proporcionar estudos e aplicações em *machine learning* (Bisong and Bisong, 2019). Para o de A plataforma *Colaboratory* foi utilizada por meio de uma *Graphics Processing Unit* (GPU) e memória RAM em nuvem por um tempo limitado de uso e sem custos.

## 5.2 Seleção de algoritmos individuais

Para este trabalho foram selecionadas redes CNNs para formação do *ensemble*. A escolha dessa tecnologia se deu pelos seguintes motivos:

- Redes CNNs desde a sua criação, têm se mostrado muito eficazes na resolução de problemas de reconhecimento de padrões e classificação de imagens e são uma alternativa viável aos métodos tradicionais de resolução do problema proposto nesta tese.
- Quanto ao glaucoma, esta é uma doença sem métodos diagnósticos robustos que possam ser isolados, quantificados e usados como entrada para algoritmos convencionais de classificação como KNN, SVM, Naive Bayes, etc. Deste modo, abordagens com a opção de usar vetores de características como entrada para esses classificadores pode não ser tão boa quanto ao uso de mapas de características derivados automaticamente de redes CNNs. Portanto, redes CNNs podem ser a melhor escolha para esta pesquisa.

Quanto a exigência de grande conjuntos dados para treinamento de uma rede CNNs, esse problema pode ser atenuado com a possibilidade de utilizar uma base de dados em proporções do conjunto BrG juntamente com a metodologia de aprendizado por transferência. Esses algoritmos que fornecem essa transferência são conhecidos como pré-treinados, e além de ser uma técnica capaz de reduzir o *overfitting*, também servem como uma forma de evitar a trabalhosa etapa de treinar uma rede do zero. As etapas necessárias para aplicar o aprendizado de transferência em um algoritmo de classificação incluem selecionar uma rede que foi previamente treinada para uma tarefa semelhante à tarefa de destino e, em seguida, essa rede previamente treinada irá transferir algumas de suas camadas

para se adapta aos novos dados da tarefa alvo, permitindo fazer isso com pouco treinamento e provavelmente com maior poder de generalização de resultados em comparação com redes treinadas do zero, principalmente quando se trabalha com pequenos conjuntos de dados para tarefa alvo.

Dado o contexto, para os experimentos realizados neste trabalho, sete (7) arquiteturas de redes CNNs pré-treinadas foram selecionadas para construir um modelo classificador por meio de um *ensemble*, ou seja, por meio do agrupamento dos melhores algoritmos selecionados. Todos os modelos adquiridos são de código aberto e referidos neste trabalho como algoritmos individuais, sendo eles o DenseNet121 (Szegedy et al., 2016a), MobileNet (Howard et al., 2017), Inception-v3 (Szegedy et al., 2017b), InceptionResNet (He et al., 2016b), ResNet50v2 (Huang et al., 2017), ResNet101 (He et al., 2016a) e Xception (Chollet, 2017).

A Tabela 5.1 apresenta os sete modelos de CNNs selecionados para este estudo, o tamanho pré-definido da entrada e a profundidade topológica de cada modelo.

Tabela 5.1: CNNs pré-treinadas com padrão de cores RGB utilizadas neste estudo

CNN	Tamanho da Entrada	Camadas
DenseNet121	224 x 224 x 3 - (RGB)	242
MobileNet	224 x 224 x 3 - (RGB)	55
Inception-v3	299 x 299 x 3 - (RGB)	189
InceptionResNet	299 x 299 x 3 - (RGB)	449
ResNet50v2	224 x 224 x 3 - (RGB)	107
ResNet101	224 x 224 x 3 - (RGB)	209
Xception	299 x 299 x 3 - (RGB)	81

Os classificadores descritos foram pré-treinados no grande conjunto de dados *ImageNet*. As redes pré-treinadas dentro do *Keras* são capazes de reconhecer 1.000 categorias de objetos diferentes. Essas categorias representam objetos cotidianos como cães, espécies de gatos, vários utensílios domésticos, tipos de veículos etc. A arquitetura de tais modelos individuais utilizados pode ser exemplificada na Figura 5.2.

Deste modo, cada modelo é configurado com uma entrada, seguida por camadas convolucionais e de *pooling*, e por fim, uma camada totalmente conectada com uma saída configurável. Embora os algoritmos utilizados tenham capacidade de oferecer saída multicategórica, no trabalho aqui proposto a resposta de cada modelo foi ajustada para uma saída binária, portanto os modelos foram configurados para terem saídas por meio da função de ativação *softmax*.

As redes pré-treinadas incluídas na biblioteca principal do *Keras* e utilizadas neste trabalho são algumas das CNNs de maior desempenho no desafio ImageNet nos últimos anos. Além do bom desempenho no desafio mencionado acima, essas redes também demonstram uma poderosa capacidade de generalizar para treinamento com imagens fora do conjunto de dados ImageNet por meio do aprendizado de transferência, portanto, são

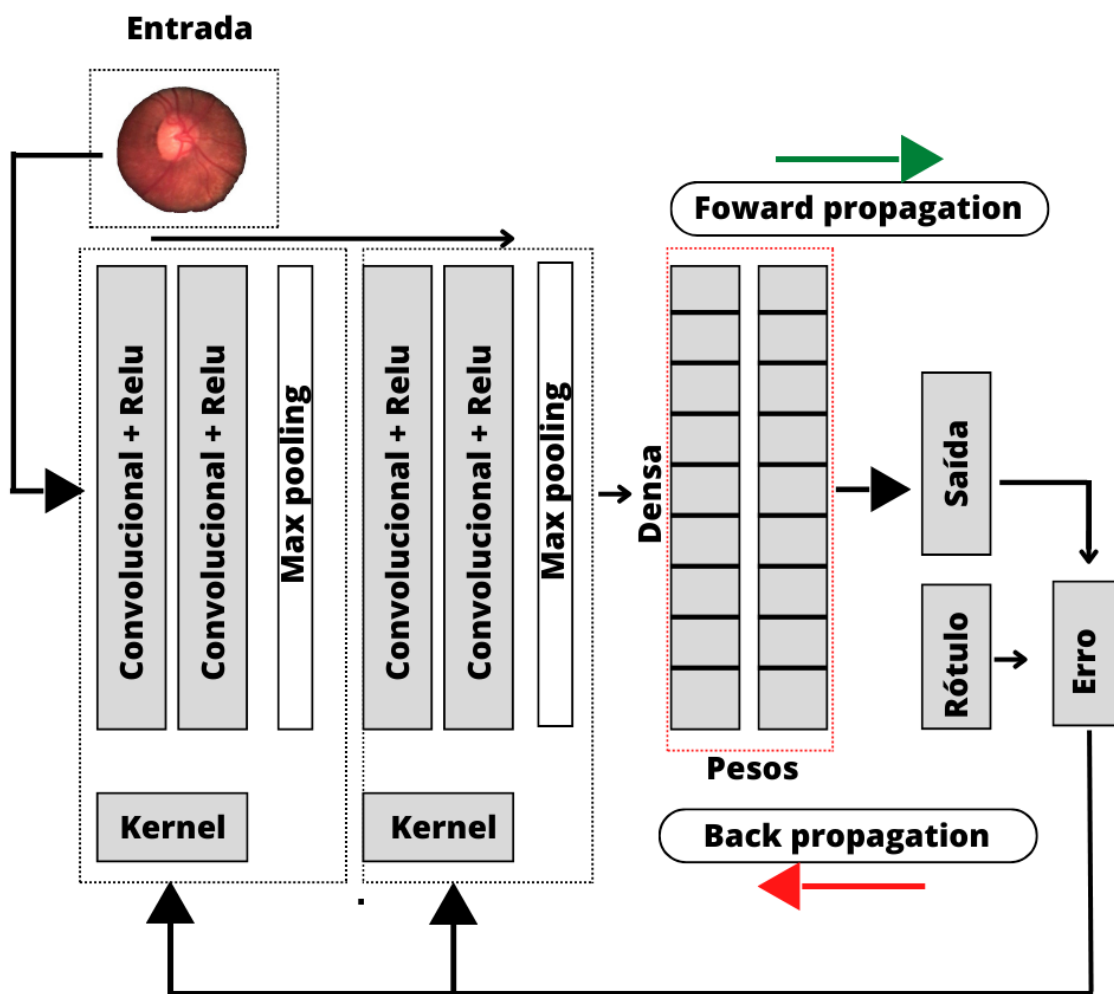


Figura 5.2: Exemplo da arquitetura dos modelos individuais de classificação - adaptado de (Patil and Rane, 2021).

modelos de reconhecimento de padrões amplamente utilizados para classificação de imagens digitais (Heaton, 2020) (Russakovsky et al., 2015).

### 5.3 Configurações e treinamento dos classificadores

A finalidade principal da metodologia aqui proposta é utilizar os algoritmos selecionados em forma de aprendizado por agrupamento, técnica mais comumente conhecida como *ensemble*. O conceito desta técnica baseia-se na ideia de combinar vários modelos preditivos mais simples, treinando-os na mesma tarefa, e produzindo a partir desse agrupamento, um modelo robusto, mais complexo e com redução de erros.

Dado o contexto, os modelos individuais apresentados na Tabela 5.1 são todos candidatos à concatenação e formação de um classificador mais robusto, porém, antes desta

---

formação, os sete modelos individuais foram treinados e testados separadamente, para que ao final, por meio de testes combinatórios, alguns deles fossem selecionados para formar o *ensemble* classificador. Para esse treinamento, algumas etapas foram aplicadas, tais como, extração da região de interesse, divisão da base de dados, regularização dos algoritmos e ajustes de treinamento, a saber:

### 5.3.1 Extração da região de interesse

Antes de aplicar o treinamento dos algoritmos individuais, a extração da região de interesse apresentada na Figura 5.3, foi o primeiro procedimento aplicado para que posteriormente os recortes pudessem ser utilizados pelos modelos pré-treinados. Esta etapa tem por objetivo descartar as regiões das imagens que apresentam menor importância na classificação do glaucoma.

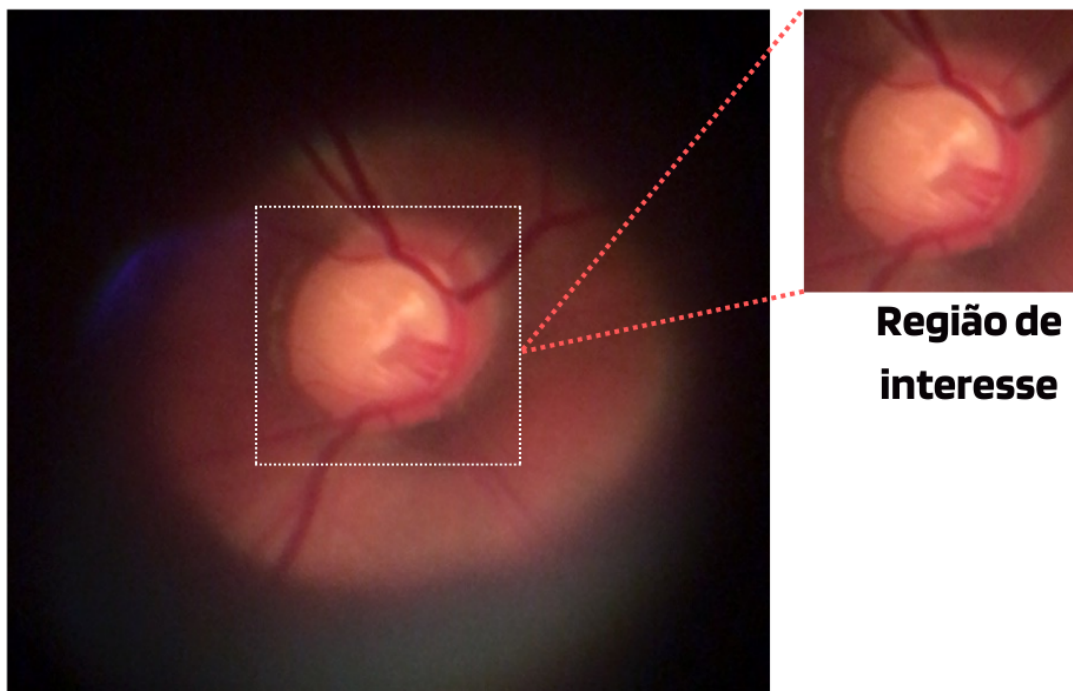


Figura 5.3: Extração manual da região de interesse - disco óptico.

Para realizar a extração da região de interesse foi necessário aplicar um processo de segmentação. Embora esse processo pudesse ter sido aplicado automaticamente por meio de algoritmos como a rede *U-Net*, em primeira análise, optou-se pela segmentação manual com auxílio de uma caixa delimitadora sobre a região do disco óptico. Essa opção é uma forma de não comprometer o processo de classificação das imagens aos possíveis erros que podem ocorrer com a segmentação automática, além de ser uma opção de remover ruídos das imagens conforme aqueles apresentados na Figura 4.12 (D).

Durante a extração da região de interesse, todas as imagens BrG foram recortadas ao centro, como mostra a Figura 5.3. Os recortes obtidos dessa operação foram salvos no

---

formato PNG e reservados para o treinamento e testes dos algoritmos de classificação.

### 5.3.2 Divisão da base de dados

Após a segmentação da região de interesse das imagens BrG, o conjunto formado foi dividido em 70% para treinamento e 30% para teste. A divisão foi realizada ao nível do paciente, o que significa que todas as imagens de um paciente foram incluídas na mesma parte do conjunto de dados (treinamento ou teste). Para usar a comparação de hiperparâmetros dos modelos *deep learning*, foram separadas 20% das imagens do conjunto de treinamento para criar um conjunto de validação, como mostra a Figura 5.4.

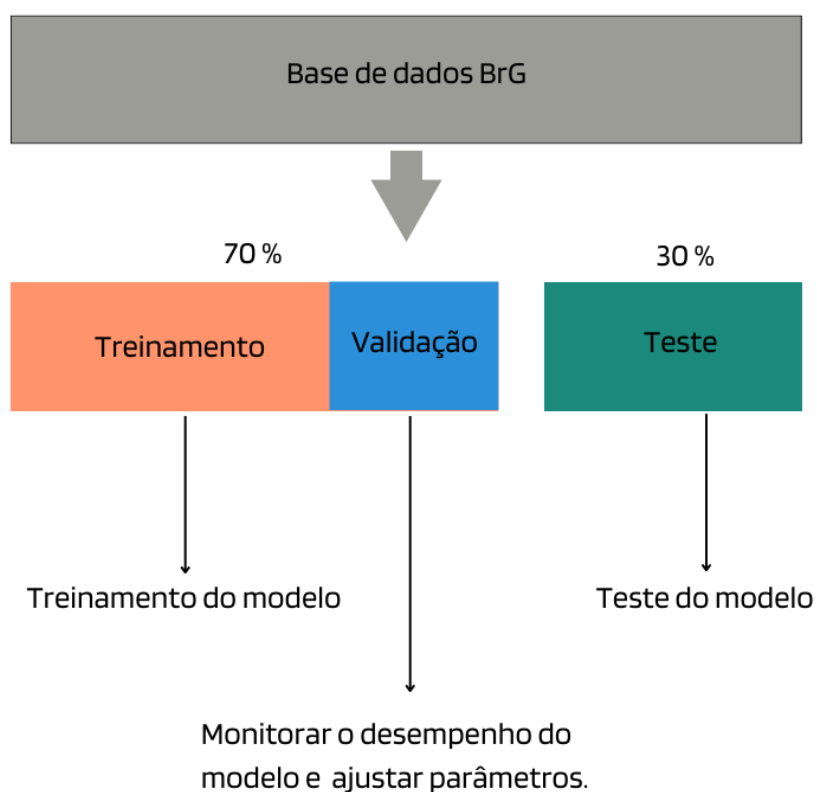


Figura 5.4: Exemplo da divisão da base de dados BrG - adaptado de (Patil and Rane, 2021).

Em síntese, o conjunto de treinamento foi usado para treinar cada modelo individual, onde os valores de perda são calculados via propagação direta e os parâmetros aprendíveis são atualizados via retropropagação.

O conjunto de validação também foi usado para monitorar o desempenho do modelo, além de ajustar os hiperparâmetros durante o treinamento, já o conjunto de teste é idealmente usado apenas uma vez, no final do treinamento, para avaliar o desempenho final do modelo treinado (Patil and Rane, 2021).

### 5.3.3 Regularização dos algoritmos

Para regularização das redes CNNs visando principalmente o tratamento de *overfitting*, algumas técnicas como *dropout*, aumento de dados e parada antecipada foram aplicadas neste trabalho durante as configurações dos algoritmos individuais, a saber:

- **Dropout:** essa técnica, como mostra a Figura 5.5, visa eliminar temporariamente alguns neurônios ocultos na rede de forma aleatória para ela poder se adaptar de diferentes maneiras durante o treinamento (Srivastava et al., 2014).

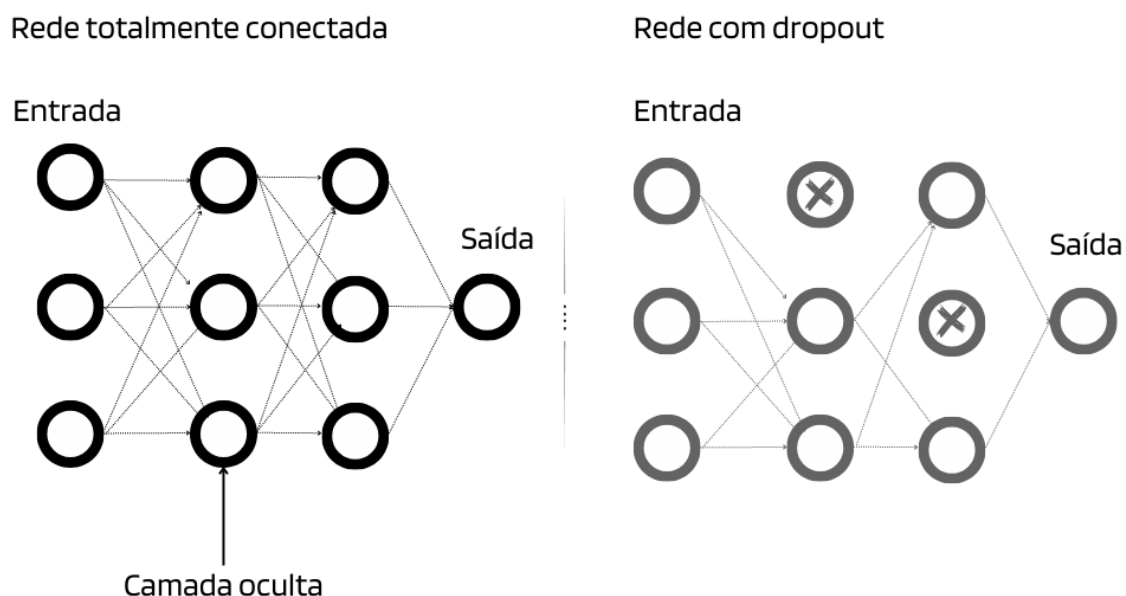


Figura 5.5: Exemplo da aplicação de *dropout* anulando alguns neurônios aleatoriamente.

Portanto, espera-se que o resultado do *dropout* seja reduzir o *overfitting* da rede aos dados de treinamento.

É considerada uma técnica que impede a coadaptação dos neurônios, forçando as unidades de neurônios a aprender extrair e representar características dos dados de entrada sem depender dos seus vizinhos, produzindo representações mais distribuídas.

- **Aumento de dados:** outra técnica também utilizada nas configurações dos algoritmos individuais foi o aumento de dados, em inglês *data augmentation* (Inoue, 2018). É uma técnica amplamente utilizada em muitas tarefas de *machine learning* para aumentar o tamanho do conjunto de dados de treinamento, como representado na Figura 5.6.

O aumento de dados é um método eficaz e de custo relativamente baixo para melhorar o desempenho e a precisão dos modelos de aprendizado de máquina no contexto de pequenos conjuntos de dados. Ao adicionar uma cópia da própria imagem e

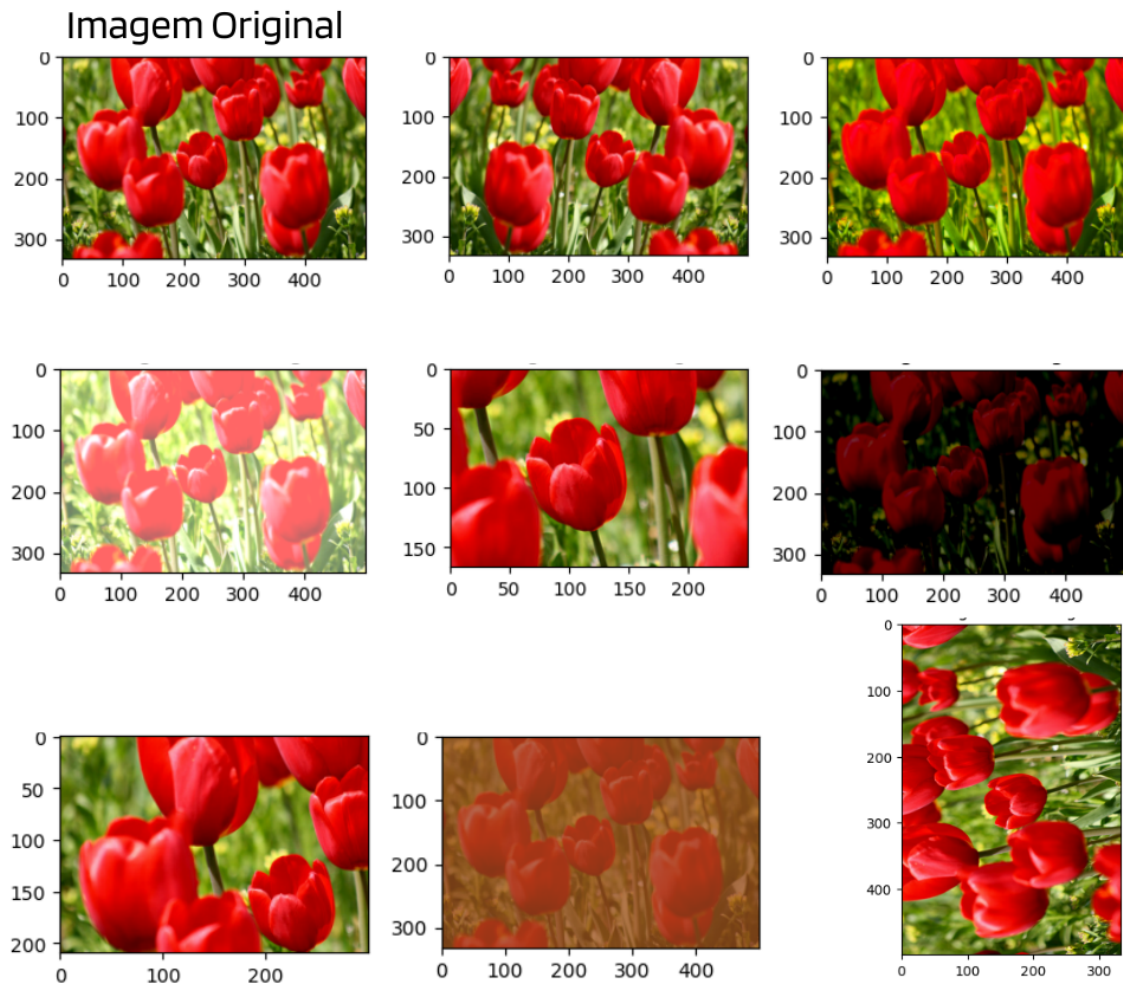


Figura 5.6: Exemplo de aumento de dados com a biblioteca *ImageDataGenerator* Keras.

modificá-la com rotação, corte, nitidez, dimensionamento e translação, as amostras usadas para treinar a máquina podem ser aumentadas.

A biblioteca *Keras* fornece a capacidade de aumentar dados artificialmente por meio da classe *ImageDataGenerator*. Desta forma, várias técnicas, bem como métodos de escalonamento de *pixels* são suportados. Ao utilizar esta ferramenta, as imagens do conjunto de dados não são salvas, ou seja, um novo lote de imagens é gerado a cada iteração, sendo alterado apenas para treinamento do modelo. Ao final do aumento de dados, o novo lote de imagens resultante é equivalente a 4 vezes o tamanho original do conjunto de treinamento. A seguir são apresentadas as técnicas utilizadas neste trabalho: deslocamento, inversão da imagem usando os argumentos horizontais e verticais, rotações, redução e aumento do brilho e feito de ampliação das imagens.

- **Parada antecipada:** conforme a configuração utilizada do algoritmo, a convergência do aprendizado ocorrerá após algumas iterações chamadas épocas, ponto em que o erro entre a saída desejada e a saída da rede é o menor possível (Haykin, 2004).

É um processo que determina se um algoritmo de aprendizado é bom ou ruim, analisando a diferença entre o erro de treinamento e o erro de classificação final (Goodfellow et al., 2016).

Determinar o número de épocas manualmente pode ser uma tarefa difícil, para a qual existe uma técnica chamada *early stopping* ou parada antecipada como demonstrado na Figura 5.7. Assim, para evitar que a rede perdesse generalidade por meio de *overfitting*, essa técnica também é aplicada ao algoritmo desenvolvido na tentativa de parar o treinamento no ponto ideal de aprendizado.

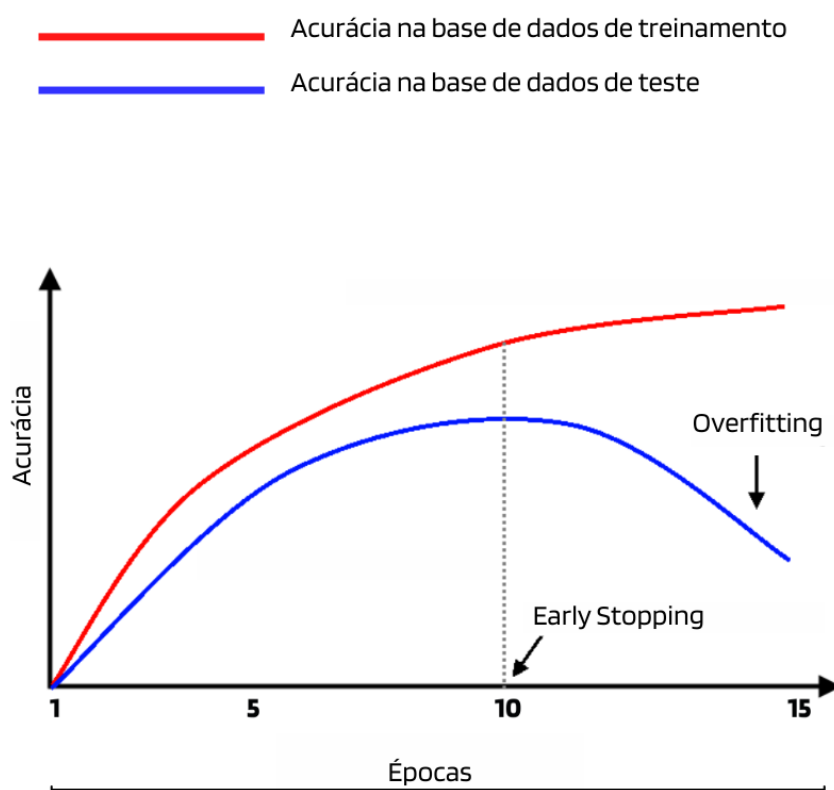


Figura 5.7: Exemplo *early stopping* com melhor ajuste na época de número 10.

Usar a técnica de *early stopping*, significa que, ao final de cada época de treinamento, a precisão da classificação é calculada sobre os dados de validação. Quando a precisão parar de melhorar, o algoritmo também deve parar de treinar. Isso facilita o processo de configuração do número necessário de épocas para treinamento de uma rede.

Neste trabalho a técnica de *early stopping* foi aplicada para observar 5 épocas de margem, ou seja, a partir do momento que a rede para de melhorar os resultados do

---

treinamento, ela ainda verifica 5 épocas adiante. Se não surgir nenhum resultado superior, a rede retorna a seu melhor resultado e finaliza o treinamento.

A regularização de dados se faz devido à obrigatoriedade de desenvolver um algoritmo que funcione bem não apenas em dados de treinamento, mas também em novas entradas. Esta generalização é considerada um problema central em técnicas de *machine learning*. Muitas estratégias usadas neste contexto são explicitamente projetadas para reduzir o erro nos dados de teste, possivelmente à custa da redução do erro nas etapas de treinamento da rede. No entanto, a prática de melhorar drasticamente um algoritmo para apresentar bons resultados nos dados de teste pode expô-lo ao risco de *overfitting*. Porém, devido a esses tipos de erros, a rede pode não conseguir generalizar seus resultados para novas instâncias de dados diferentes daquelas usadas nas fases de treinamentos e testes. Portanto, técnicas como a proposta são aplicadas, pois podem dar maior robustez e confiabilidade aos resultados obtidos pelos algoritmos.

### 5.3.4 Ajustes de treinamento

Os modelos pré-treinados são geralmente compartilhados na forma de milhões de parâmetros ou pesos obtidos durante o treinamento para um determinado problema. Existem potencialmente várias abordagens e configurações para aplicá-los a novos problemas, portanto, todas são úteis, desde que se adaptem as técnicas disponíveis ao uso desejado.

A biblioteca de *deep learning keras*, fornece interfaces de *download* para alguns modelos pré-treinados amplamente populares, entre eles os modelos apresentados na Tabela 5.1. Além disso, a biblioteca também fornece métodos de aplicar a transferência de aprendizado, dos quais este estudo escolheu a técnica de congelamento de pesos. Deste modo, neste trabalho, todos os modelos de classificadores são adaptados com base no conceito de congelamento de determinadas camadas da rede, considerando os meios mais comuns de aprendizado por transferência de conforme o seguinte fluxo de trabalho:

1. Congelou-se uma parte do modelo, mas conservando pesos adquiridos no treinamento original, ou seja, no treinamento obtido no conjunto de dados ImageNet. Para melhor entendimento, uma rede CNN típica revela três tipos de camadas, iniciais, intermediárias e posteriores. No aprendizado por transferência as camadas iniciais e intermediárias são (congeladas), ou seja, mantidas intactas, e apenas as últimas camadas são livres para novos treinamentos oriundas das tarefas de classificação em novos problemas.
2. Foram adicionadas duas camadas novas e treináveis no topo das camadas congeladas, conforme mostrado na Figura 5.8.
3. As novas camadas incluídas na rede foram treinadas usando as imagens de treinamento do conjunto de dados BrG com as seguintes configurações:

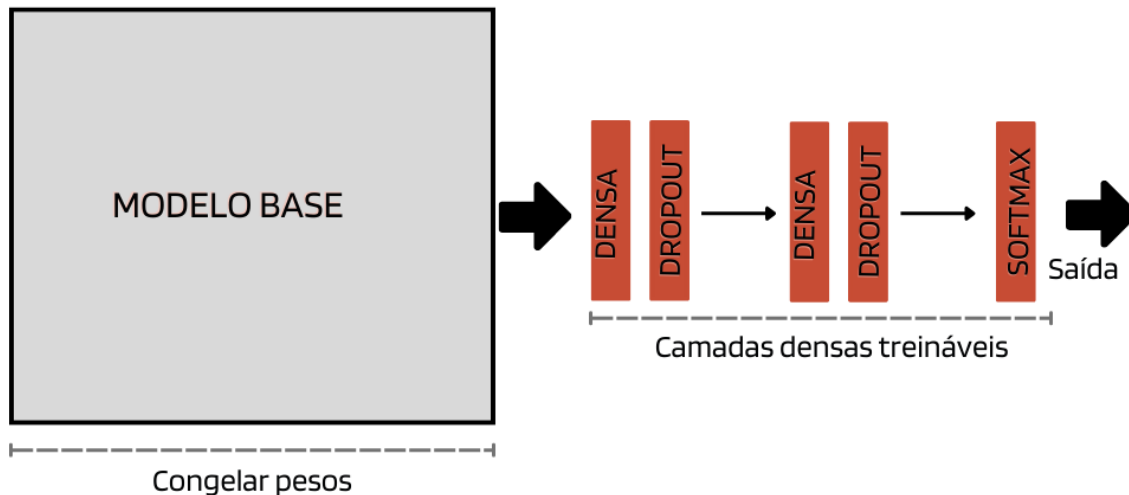


Figura 5.8: Modelo base de pesos congelados, seguido de construção de camadas densas com aplicação de *dropout*. A saída foi obtida pela função de ativação *softmax*.

- As imagens foram padronizadas para um tamanho fixo (224 x 224 x 3) ou (299 x 299 x 3), conforme exigência de cada modelo individual apresentado na Tabela 5.1. Esse tamanho representa a (altura x largura x profundidade 3, que representa o sistema de cores RGB).
- Na escolha do tamanho do lote *batch size*, foi utilizado o *mini-batch mode*, onde o tamanho do lote deve ser maior que 1 e menor que o tamanho total do conjunto de dados. Portanto, neste trabalho este parâmetro foi fixado em 32, número alcançado após testes de ajustes de lote. É um hiperparâmetro que define o número de amostras a serem trabalhadas em cada iteração da rede;
- Os valores de *pixels* das imagens foram normalizados entre [-1 e 1], usando funções de normalização dos próprios modelos. A normalização de imagens significa convertê-las em valores de forma que a média e o desvio padrão estejam dentro dos intervalos predefinidos;

4. Para cada classificador individual foram aplicadas as seguintes técnicas de regularização:

- O aumento de dados foi aplicado para gerar artificialmente novas amostras de dados de treinamento a fim de aumentar a generalidade do modelo. Nesta etapa, o número de imagens no conjunto de treinamento pode ser aumentado por um fator de 4.
- Uma taxa de *dropout* de 0,2 foi usada para camadas totalmente conectadas objetivando superar o *overfitting*;
- Aplicou-se a técnica de parada antecipada com 5 épocas de margem.

- 
- Para aplicações de retropropagação, o otimizador de estimativa de momento adaptativo, em inglês *Adaptive Moment Estimation* (Adam) foi usado como a função de perda nos classificadores (Kingma and Ba, 2014).

## 5.4 Testes e resultados individuais

Esta seção descreve brevemente cada um dos algoritmos utilizados neste trabalho. Descreve também os resultados alcançados com cada classificador individual por meio do gráfico de precisão e gráfico de perda. Assim, os dois gráficos apresentam curvas suaves que convergem para os valores esperados (1 e 0). Jundo aos gráficos também apresenta-se a curva ROC dos modelos individuais.

### 5.4.1 ResNet50v2 e ResNet101

Antes de descrever esses modelos de algoritmos, é importante entender um problema que as redes neurais enfrentam, conhecido como *vanishing gradient*, comum nas comunidades de *deep learning* e ciência de dados. Portanto, redes neurais artificiais que envolvem treinamento de retropropagação e aprendizado com base em métodos estocásticos de descida de gradiente para atualizar pesos, geralmente enfrentam esse problema quando aumentam a profundidade da rede.

A profundidade é um fator-chave na capacidade das RNAs, pois as redes neurais mais profundas possuem mais parâmetros, e as interações entre esses parâmetros permitem modelar funções complexas e alcançar resultados superiores nas tarefas executadas. Portanto, espera-se que arquiteturas mais profundas produzam resultados pelo menos tão bons ou até melhores do que aqueles obtidos por arquiteturas mais rasas. No entanto, nem sempre é esse o caso.

O problema *vanishing gradient* surge no momento que se aumenta o número de camadas da rede, principalmente quando se faz isso usando funções de ativação como sigmoide, de forma que com o passar das iterações o gradiente da função de perda se aproxime de zero, tornando o treinamento do algoritmo difícil ou disruptivo. Gradientes extremamente pequenos significam que os pesos e deslocamentos da camada inicial não são atualizados com eficiência a cada sessão de treinamento. Como essas camadas iniciais costumam ser críticas para identificar elementos-chave dos dados de entrada, isso pode levar a imprecisões gerais em toda a rede.

Uma das soluções para esse problema que impede trabalhar com redes amplamente profundas é a possibilidade de usar funções de ativação, como ReLU, que não levam a pequenas derivadas, ou usar uma camada de normalização em lote, ou mesmo uma aplicação de uma rede residual, como o modelo ResNet (do inglês Residual Network).

Para contextualizar, ResNet é uma rede neural convolucional que venceu a competição

ILSRVC de 2015 no conjunto de dados ImageNet. Sua arquitetura é caracterizada por conexões e blocos residuais, visando principalmente resolver o problema de *vanishing gradient*.

Um bloco residual consiste em duas ou mais camadas convolucionais, onde a saída do bloco é obtida adicionando a entrada do bloco com a saída da última camada do bloco. Isso permite que a rede aprenda uma representação residual, o que significa que ela aprende a diferença entre a entrada e a saída desejada, em vez de tentar aproximar diretamente a saída. A figura 5.9 exemplifica o bloco residual utilizado por esta rede.

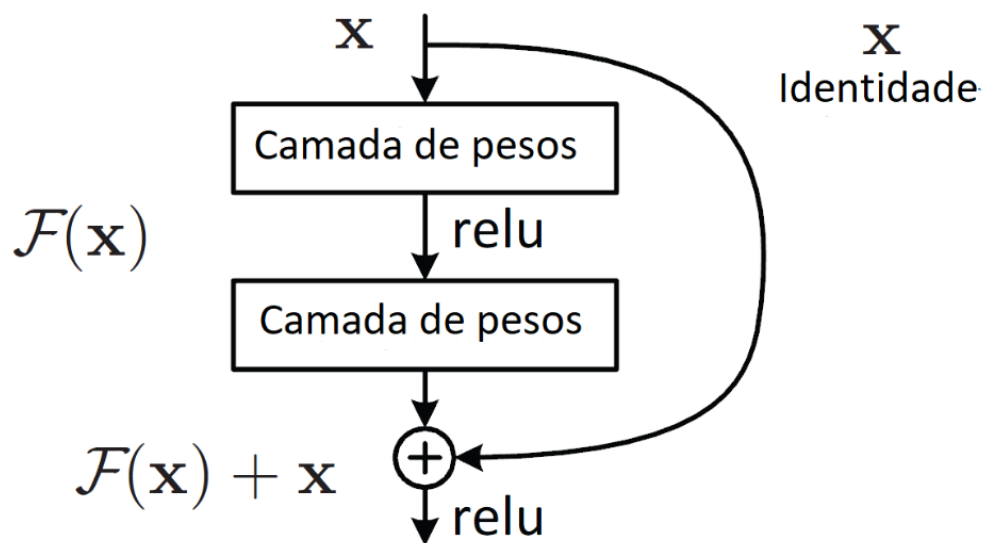


Figura 5.9: Exemplo de um bloco de construção residual ResNet (He et al., 2016a).

A representação matematicamente do bloco se faz pela Equação 5.1:

$$H(X) = F(X) + X \quad (5.1)$$

Para entendimento da equação,  $X$  é a entrada para o conjunto de camadas  $F(x)$  é a função residual e  $H(x)$  é a função de mapeamento da entrada para a saída. Portanto, os autores acreditam que otimizar o mapa residual é mais fácil do que otimizar o mapeamento original sem referência.

O avanço na arquitetura da ResNet permite treinar redes neurais extremamente profundas, com mais de 150 camadas. De modo geral, o número de camadas e parâmetros do modelo ResNet varia de versão para versão, incluindo camadas convolucionais, camadas de *pooling* intermediárias, camadas totalmente conectadas e funções finais *softmax*. Quanto a utilização dessa arquitetura neste trabalho, as Figuras 5.10 e 5.11 mostram o desempenho gráfico dos algoritmos ResNet-50v2 e ResNet101 sobre a classificação do glaucoma na base de dados BrG, passando os dados de treinamento e validação como entrada.

Existem muitas variantes da arquitetura ResNet, como:

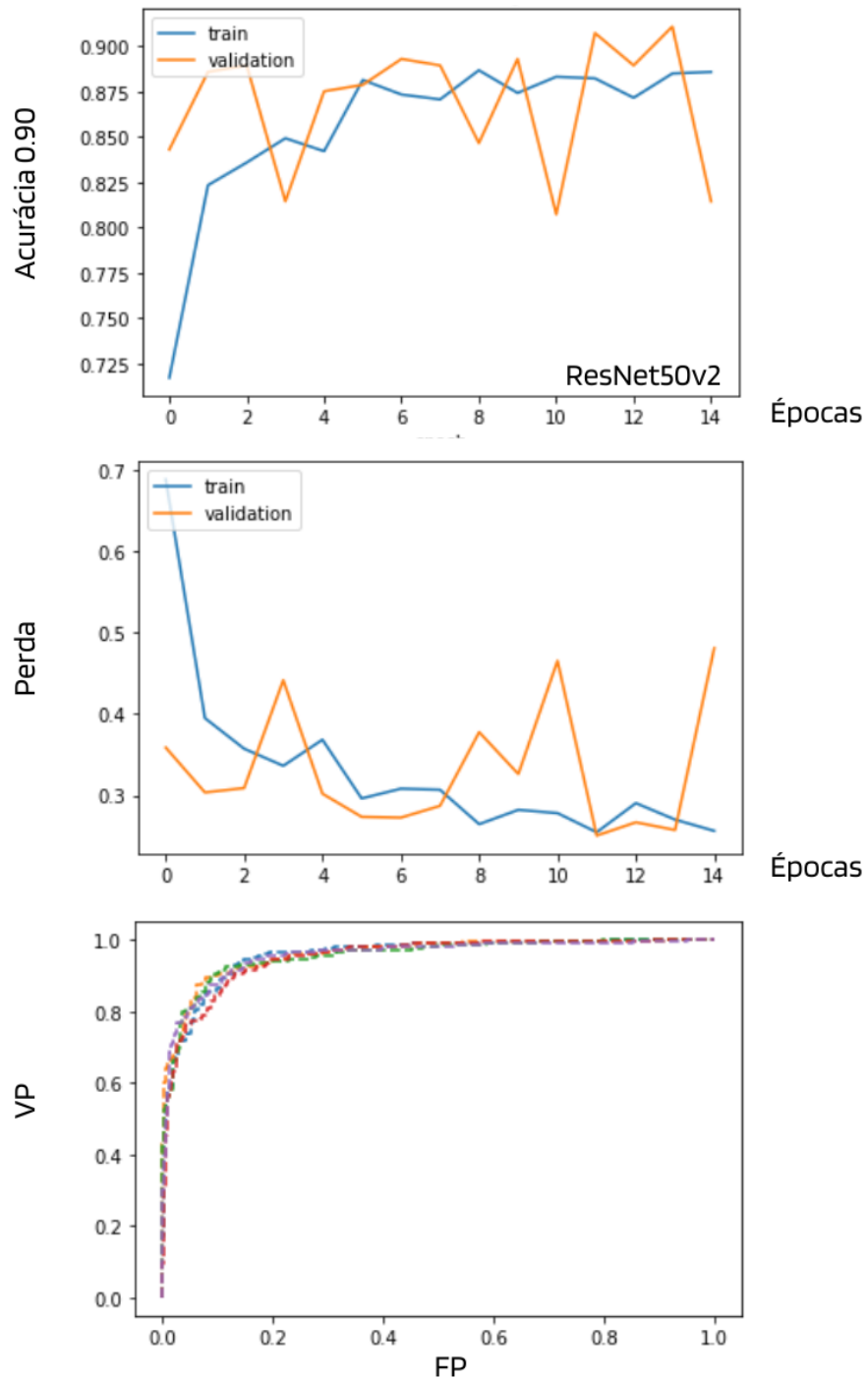


Figura 5.10: Curva de aprendizado, taxa de perda e curva ROC do modelo ResNet50v2 com 14 épocas de treinamento sobre a base de dados BrG.

- ResNet18 e ResNet34;
- ResNet50 e ResNet50v2;
- ResNetv101 e ResNetv152.

Para cada submodelo ResNet, o último número representa o número de camadas. A ResNet é uma rede convolucional clássica usada como base para muitas tarefas de visão

computacional. Embora haja diferença entre cada modelo ResNet, o princípio fundamental é muito semelhante entre os algoritmos (He et al., 2016b; Simonyan and Zisserman, 2014).

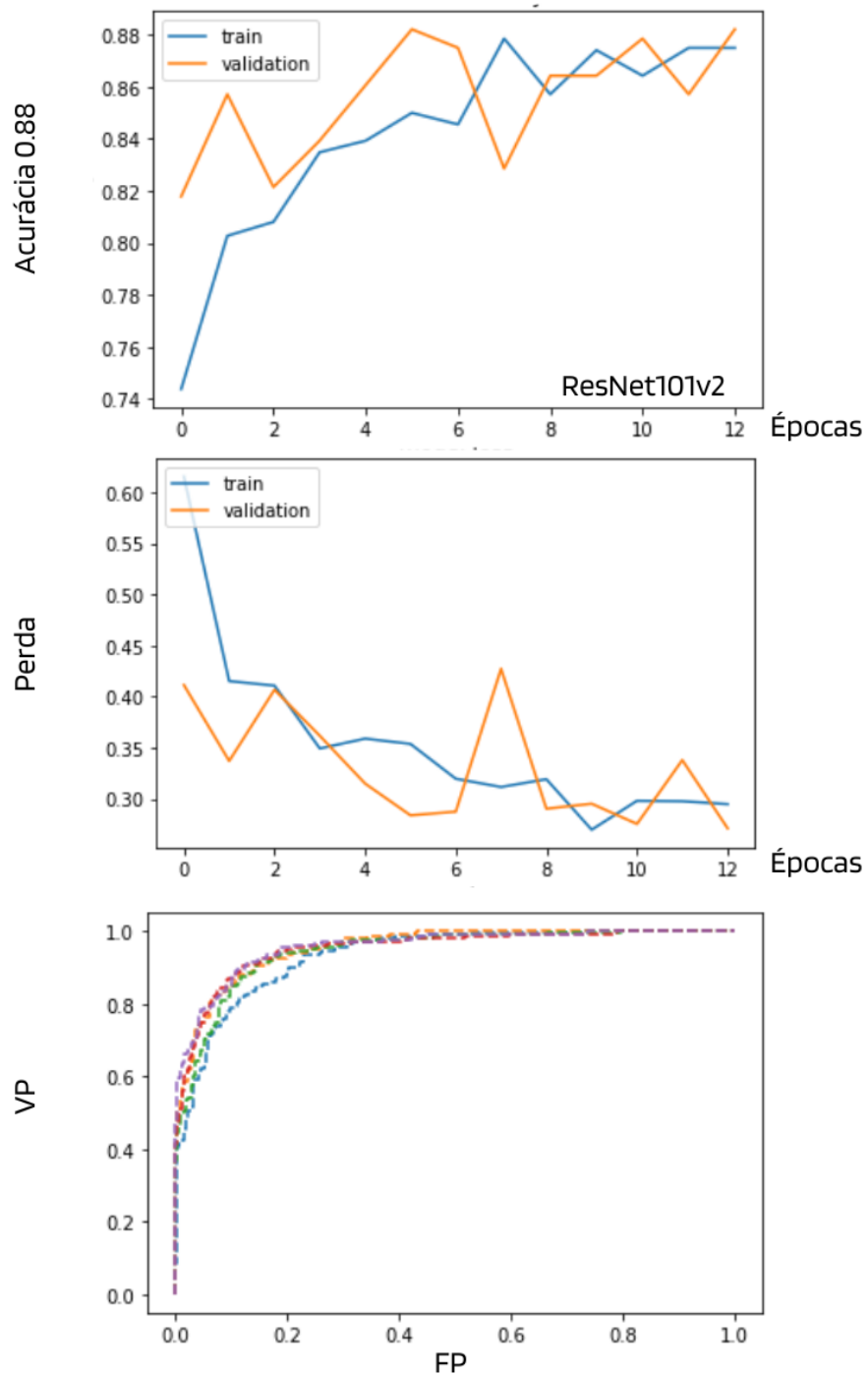


Figura 5.11: Curva de aprendizado, taxa de perda e curva ROC do modelo ResNet101 com 12 épocas de treinamento sobre a base de dados BrG.

### 5.4.2 Inception-v3

A rede Inception, também conhecida como GoogleNet, é uma rede desenvolvida pelos autores (Szegedy et al., 2016b) em 2014, ano em que foi vencedora do ILSVRC. Em sua versão v1, esse algoritmo foi um novo avanço no desenvolvimento de redes CNNs, pois, antes disso, a única forma de melhorar o desempenho da rede era aumentar o tamanho, profundidade ou largura das camadas. No entanto, com o surgimento no módulo Inception, em vez de usar várias camadas sequencialmente para tornar o modelo mais profundo, é usada uma estratégia de adicionar várias camadas no mesmo nível, ou seja, usar camadas em paralelo. Como tal, as arquiteturas mais amplas são preferidas às mais profundas.

O propósito do módulo de Inception é atuar como um extrator de recursos multinível, computando as convoluções 1x1, 3x3 e 5x5 dentro do módulo de rede, conforme mostrado na Figura 5.12. As saídas desses filtros são eventualmente empilhadas ao longo da dimensão do canal e então enviadas para as camadas subsequentes da rede. O módulo de iniciação é baseado em algumas pequenas convoluções para reduzir drasticamente o número de parâmetros e facilitar o treinamento.

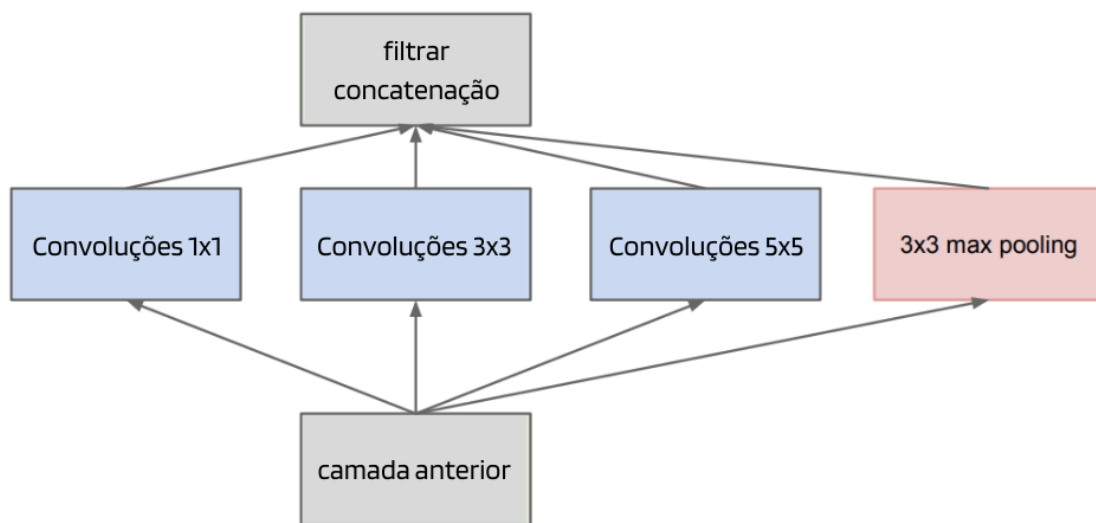


Figura 5.12: versão do módulo Inception v1 - adaptado de (Szegedy et al., 2016b).

Inception-v2 e v3 foram ambos introduzidos em (Szegedy et al., 2016b), na versão v2 foi introduzida a normalização de lote, na versão v3 foi adicionada convoluções 7x7 fatoradas. A quarta iteração, Inception-v4, foi uma melhoria adicional para v3 (Szegedy et al., 2017a). Com base no contexto, a estrutura do Inception V3 consiste em vários módulos, incluindo camadas convolucionais, camadas intermediárias, camadas de *max pooling*, *dropout*, camadas totalmente conectadas e camadas *softmax*, abrangendo 42 camadas e contendo quase 30 milhões de parâmetros.

Quanto ao uso do modelo Inception V3 neste trabalho, a Figura 5.13 mostra o desempenho gráfico do algoritmo Inception-V3 sobre a classificação do glaucoma na base de dados BrG, passando os dados de treinamento e validação como entrada:

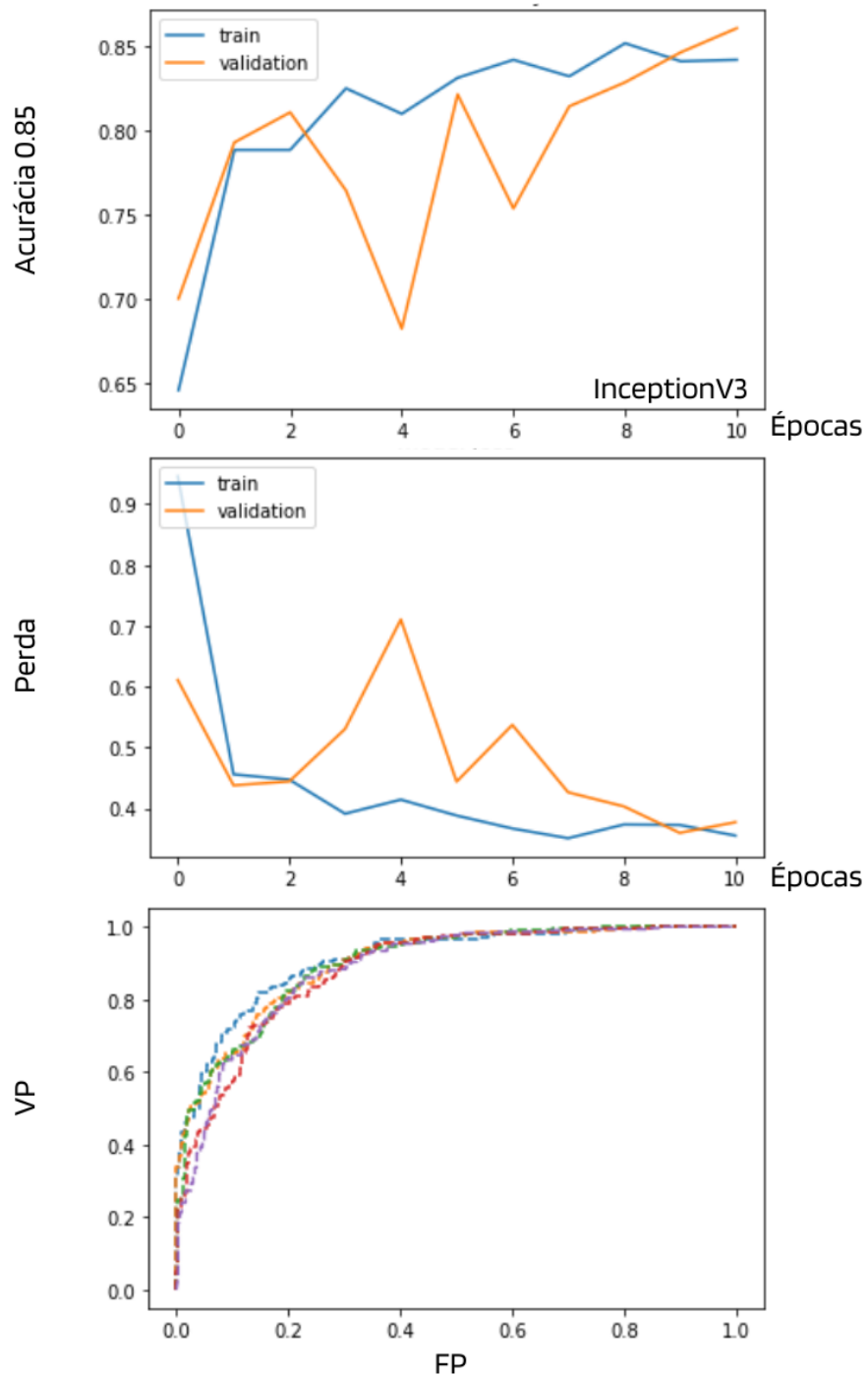


Figura 5.13: Curva de aprendizado, taxa de perda e curva ROC do modelo Inception-v3 com 10 épocas de treinamento sobre a base de dados BrG.

---

### 5.4.3 InceptionResNet

A ideia central do InceptionResNet é adicionar um módulo residual no final da convolução dos módulos Inceptions, ou seja, o InceptionResNet basicamente substitui a concatenação de filtros usados no Inception por conexões residuais, combinando assim as duas arquiteturas (ResNet e Inception).

Quando ao uso da rede Inception-ResNet neste trabalho, a Figura 5.14 mostra o desempenho gráfico do algoritmo InceptionResNet sobre a classificação do glaucoma na base de dados BrG, passando os dados de treinamento e validação como entrada:

Embora os autores da Inception v3 não parecem apoiar a visão de que as conexões residuais introduzidas por (He et al., 2016a) são inerentemente necessárias para treinar modelos convolucionais muito profundos, pesquisas mostraram que a conexão residual permite que o modelo proposto cresça significativamente, mantendo seu desempenho superior e não superajustado como outras redes neurais profundas (Ronald et al., 2021).

Para que haja a possibilidade de união das duas tecnologias (ResNet e Inception), as convoluções devem ter as mesmas dimensões de entrada e saída. Portanto, adiciona-se uma convolução pontual, que consiste em usar um *kernel* de tamanho  $1 \times 1$  no final da convolução (Szegedy et al., 2017b).

### 5.4.4 MobileNet

MobileNet é uma arquitetura CNN muito mais rápida, bem como, um modelo menor que faz uso de um novo tipo de camada convolucional, conhecida como convolução separável, como mostra a Figura 5.15. Devido ao pequeno tamanho do modelo, esses algoritmos são considerados muito úteis para serem implementados em dispositivos móveis e embarcados, portanto, é um modelo que se preocupa principalmente no uso de recursos computacionais.

Uma convolução separável em profundidade é feita a partir de duas operações (*Depthwise Convolution* e *Pointwise Convolution*). Em outras palavras, a convolução em profundidade, ou *Depthwise Convolution*, executa uma única convolução em cada canal de cor, em vez de combinar os três canais como nas redes tradicionais. Para extração de características, logo em seguida, vem a camada que aplica uma convolução  $1 \times 1$  chamada de *Pointwise Convolution*, para calcular uma combinação linear da *Depthwise Convolution*.

Essa técnica permite reduzir muito a quantidade de computação necessária ao realizar convoluções pela rede, reduzindo assim os custos computacionais. A Figura 5.15, exemplifica a estrutura da segunda versão do MobileNet, que possui duas camadas estruturadas da seguinte forma: A primeira camada é *Depthwise Convolution*, a segunda camada é *Pointwise Convolution*.

Quanto ao uso desta arquitetura neste trabalho, a Figura 5.16 apresenta o desempenho gráfico do algoritmo MobileNet sobre a classificação do glaucoma na base de dados BrG,

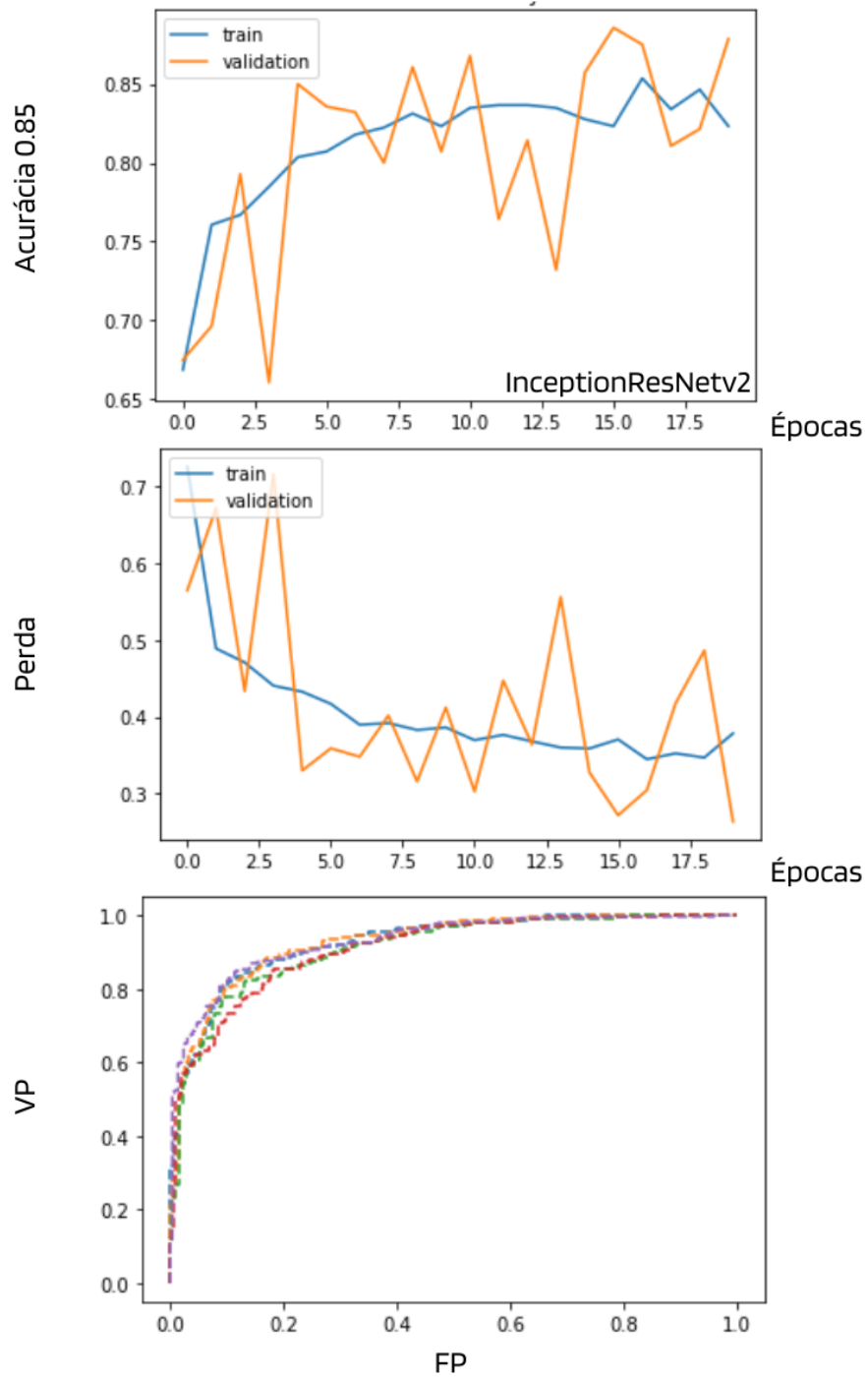


Figura 5.14: Curva de aprendizado, taxa de perda e curva ROC do modelo InceptionRes-Net com 17 épocas de treinamento sobre a base de dados BrG.

passando os dados de treinamento e validação como entrada:

### 5.4.5 Xception

Quanto a utilização desta rede neste trabalho, a Figura 5.17, mostra o desempenho gráfico do algoritmo Xception sobre a classificação do glaucoma na base de dados BrG, passando

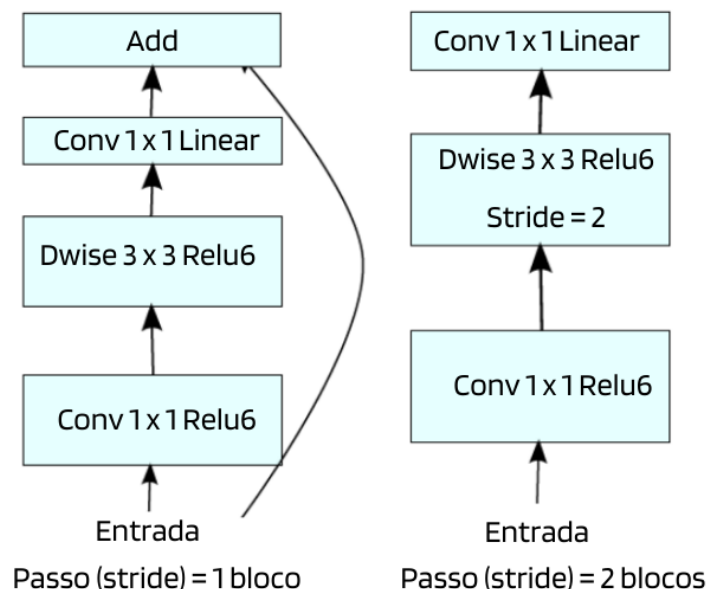


Figura 5.15: Exemplo de blocos convolucionais da segunda versão MobileNet (Sandler et al., 2018).

os dados de treinamento e validação como entrada:

O nome Xception vem de *Extreme Version of Inception*. Trata-se de uma rede CNN derivada de uma variação modular da arquitetura Inception. Na rede Xception, os blocos básicos da rede Inception são modificados para separar completamente o mapeamento de informações entre os canais de entrada, aplicando a técnica conhecida como a *depthwise convolution* seguida pela técnica *pointwise convolution*. Esta operação de convolução separada consiste em dividir a convolução tradicional em duas operações separadas, semelhante ao método utilizado pela MobileNet.

A arquitetura consiste em 36 camadas convolucionais organizadas em 14 módulos com conexões residuais lineares para acelerar a convergência e melhorar o desempenho da classificação (Chollet, 2017).

A diferença entre esse tipo de convolução e a convolução comum é que a convolução não precisa ser aplicada a todos os canais de uma vez, e cada convolução é aplicada ao canal individualmente, resultando em menos conexões de rede, resultando em um modelo mais leve. A operação de convolução 1x1 permite reduzir o número de canais ou filtros a serem passados adiante, reduzindo significativamente o número de computações nas camadas subsequentes. O efeito da convolução 1x1 é diferente da convolução com filtro de dimensão maior. Ao analisar apenas um pixel de cada canal, este filtro não extrai informações espaciais, mas pode atuar como um extrator de informações mais importante considerando a profundidade da imagem. (Chollet, 2017).

Outra diferença entre o Xception e o Inception é que os módulos Inception possuem um ativador ReLU não linear após a primeira operação, enquanto o Xception não possui

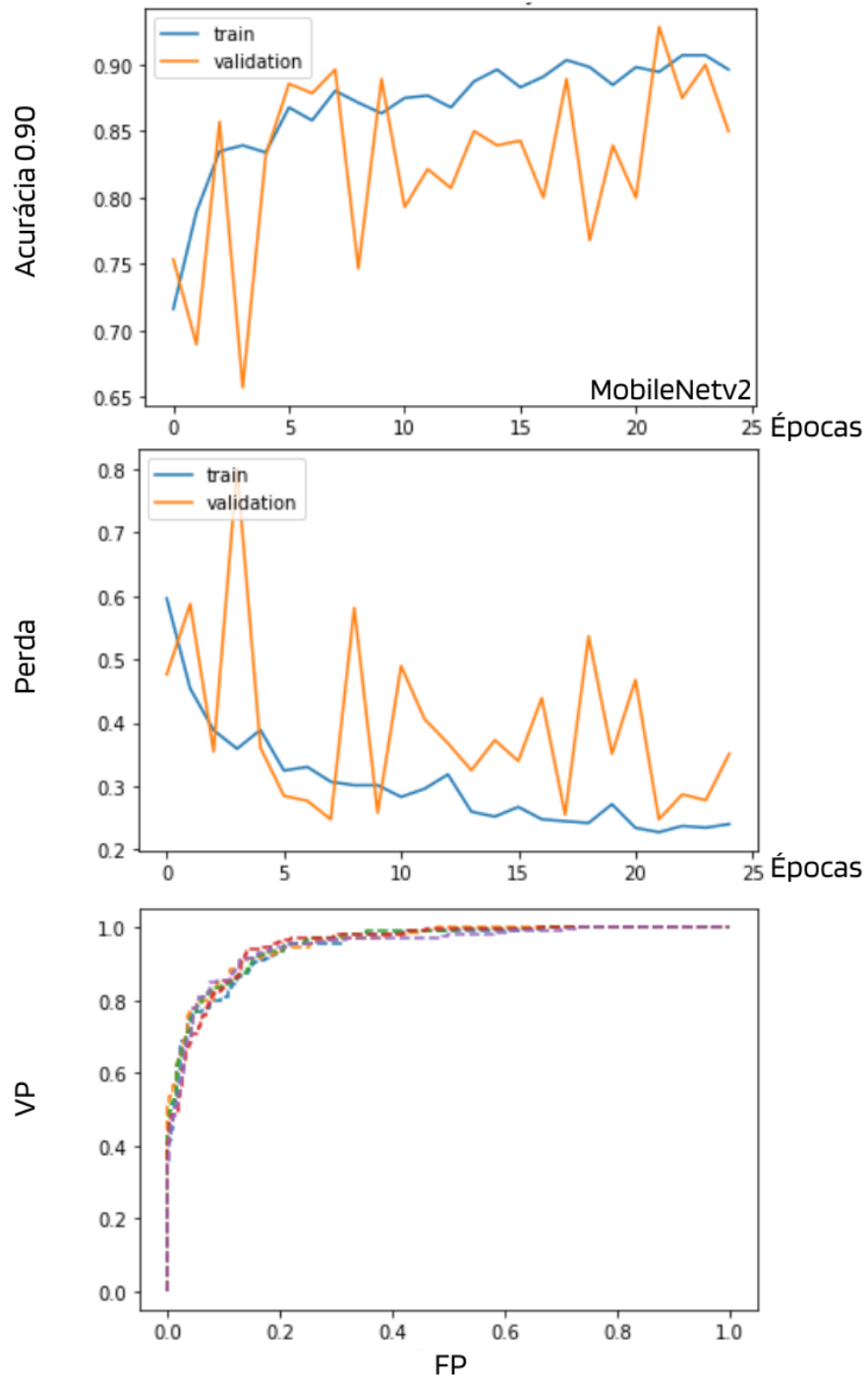


Figura 5.16: Curva de aprendizado, taxa de perda e curva ROC do modelo MobileNet com 25 épocas de treinamento sobre a base de dados BrG.

ativadores intermediários.

### 5.4.6 DenseNet

Quanto ao uso dessa rede neste trabalho, a Figura 5.18, mostra o desempenho gráfico do algoritmo DenseNet121 sobre a classificação do glaucoma na base de dados BrG,

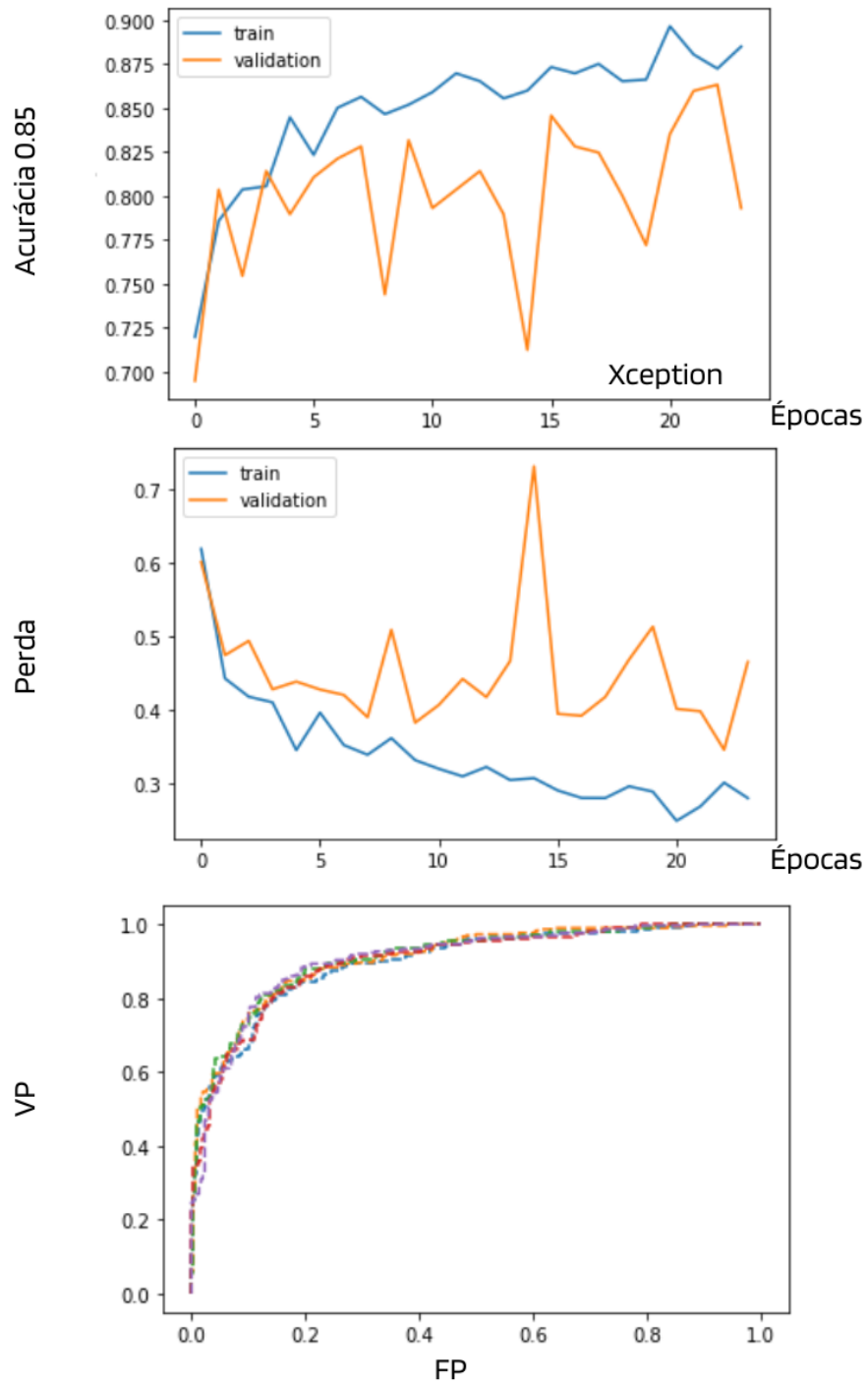


Figura 5.17: Curva de aprendizado, taxa de perda e curva ROC do modelo Xception com 20 épocas de treinamento sobre a base de dados BrG.

passando os dados de treinamento e validação como entrada:

Detalhes importantes e inovações nesta arquitetura são blocos densos. Esses blocos simplesmente são como o bloco residual com a seguinte alteração: em vez de somar os canais ao final do bloco, o resultado de  $F(x)$  é concatenado à entrada  $x$  aumentando o número de canais.

A fim de manter a natureza da rede *feed-forward*, cada camada tem uma entrada adi-

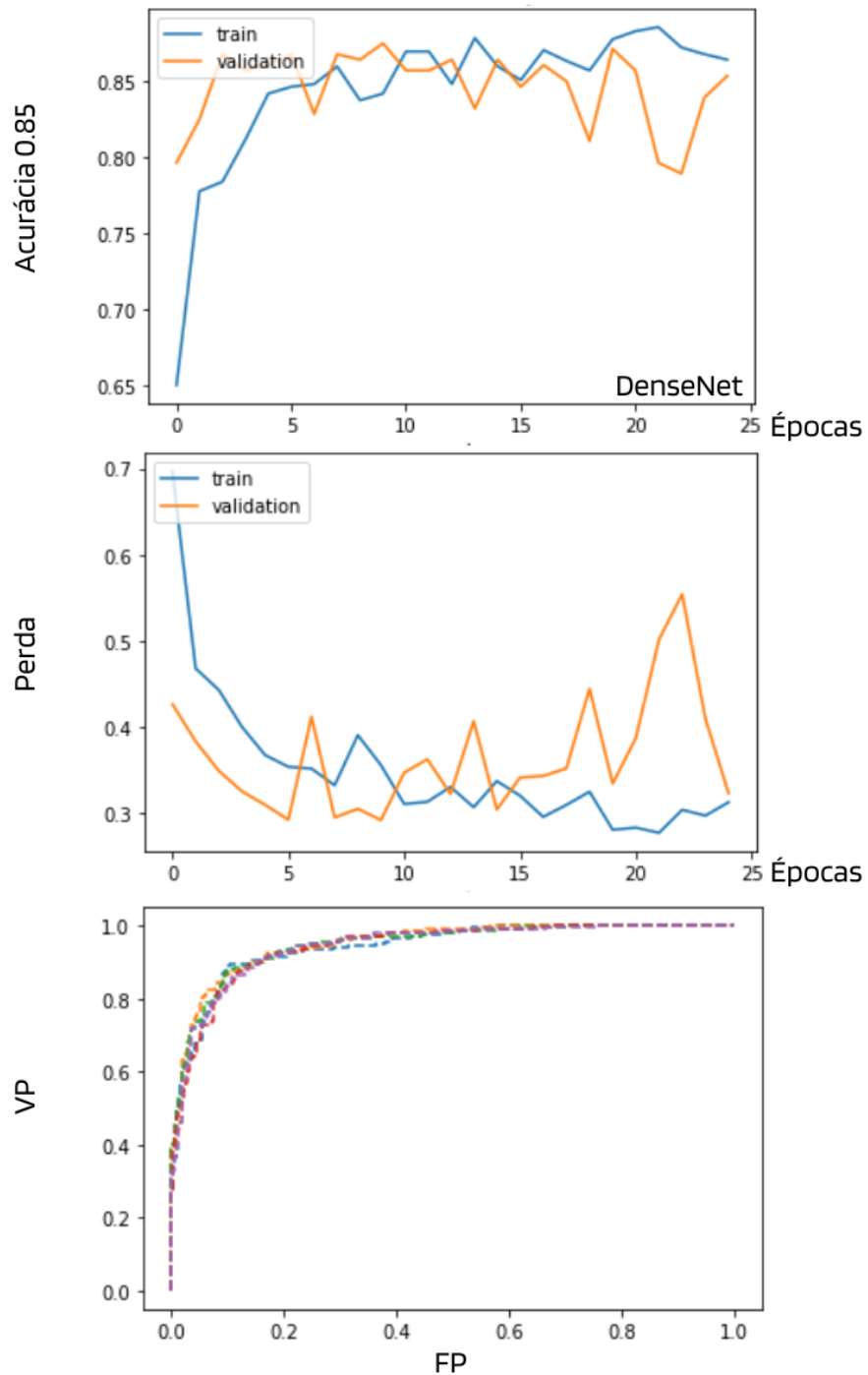


Figura 5.18: Curva de aprendizado, taxa de perda e curva ROC do modelo DenseNet com 25 épocas de treinamento sobre a base de dados BrG.

cional de todas as camadas anteriores e passam seus próprios mapas de recursos para todas as camadas subsequentes, ou seja, os mapas de características são passados como entrada não só para a camada subsequente, mas para todas as camadas até o final da rede. Devido ao seu padrão de conexão denso, esse método é chamado de Rede Convolutiva Densa (DenseNet) (Szegedy et al., 2016a). A Estrutura básica DenseNet consiste em blocos densos, camadas de transição que controlam a complexidade do modelo, camadas

convolucionais e camadas totalmente conectadas, como mostra a Figura 5.19.

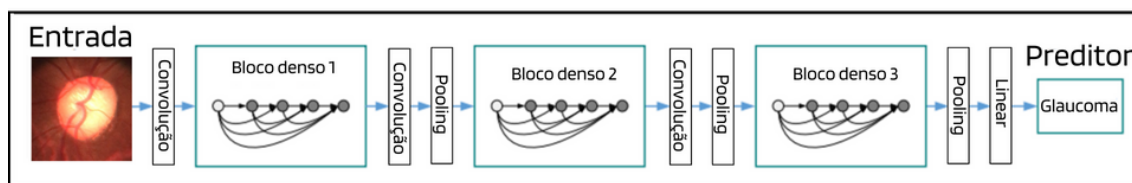


Figura 5.19: Deep DenseNet com três blocos densos. As camadas entre dois blocos adjacentes são chamadas de camadas de transição e alteram o tamanho do mapa de recursos por meio de convolução e pooling (Szegedy et al., 2016a).

Devido ao tamanho da rede, esta abordagem ajuda a preservar gradientes além de economizar parâmetros, pois a rede não precisa mais reaprender mapeamentos redundantes usando blocos residuais.

### 5.4.7 Resultados individuais

A saída dos modelos CNNs apresentados na Tabela 5.1 foi configurada com uma função de ativação (softmax), de forma que a rede aceite uma imagem digital como entrada e gere a probabilidade de a imagem de entrada representar um paciente com ou sem glaucoma como saída.

Tabela 5.2: Resultados de classificadores individuais

<b>CNN-Individuais</b>	<b>ACC</b>	<b>SE</b>	<b>ES</b>	<b>PR</b>	<b>F1</b>	<b>AUC</b>
DenseNet	0,870	0,933	0,807	0,828	0,877	0,954
MobileNet	0,836	0,957	0,718	0,771	0,854	0,947
Inception-v3	0,835	0,913	0,757	0,789	0,847	0,930
InceptionResNet	0,778	0,953	0,600	0,706	0,811	0,932
ResNet50v2	0,881	0,953	0,810	0,833	0,889	0,956
ResNet101	0,880	0,910	0,850	0,858	0,883	0,949
Xception	0,806	0,926	0,686	0,747	0,827	0,919

Na análise, os classificadores individuais classificaram as imagens do conjunto de dados BrG em glaucoma "positivo" ou "negativo", conforme mostrado na Tabela 5.2. Os testes foram aplicados sobre a técnica de validação cruzada como mostrada na Figura:5.20, uma metodologia de reamostragem adicional que pode ser utilizada para ajudar a encontrar modelos mais gerais e reduzir também o *overfitting*. O tempo médio de treinamento para cada um dos classificadores usando a GPU do Google Collaboratory foi de aproximadamente duas horas.

A técnica de reamostragem utilizada é a validação cruzada *k-fold*. Com esse método é possível treinar e testar o modelo com *k* "dobras" em diferentes subconjuntos de dados de treinamento e estimar o desempenho do modelo de *machine learning* com base em dados

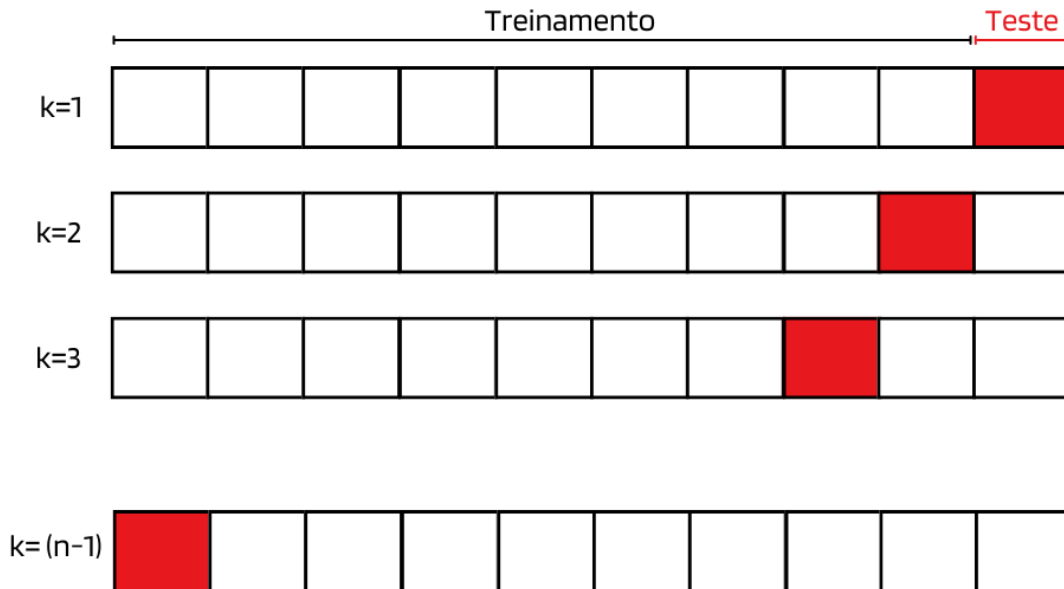


Figura 5.20: Exemplo da utilização  $k$ -fold, adaptado de (Russell, 2010).

não vistos. Assim, primeiramente, divide-se o conjunto de dados  $N$  em  $k$  subconjuntos de tamanhos iguais, também chamados de "dobras" sendo  $k > 1$ . Faz-se isso presumindo que  $k$  é divisível por  $N$ . Deste modo, uma das  $k$  dobras servirá como conjunto de teste, e as dobras restantes servirão para treinar o modelo. Esse processo se repete até que cada uma das dobras tenha agido como uma dobra de treinamento. Após cada avaliação, uma pontuação é mantida, e quando todas as iterações são concluídas, é calculada a média das pontuações para avaliar o desempenho geral do modelo (Russell, 2010), (IBM, 2022a).

Neste trabalho, por conveniência, foi utilizada a validação  $K$ -Fold, com  $k = 5$  e, portanto, a acurácia dos classificadores apresentados na Tabela 5.2, correspondeu à média dos resultados obtidos pela validação cruzada quádrupla.

Ainda buscando a avaliação dos modelos individuais, a Figura 5.21 apresenta uma matriz de confusão para cada um dos modelos individuais testados, onde as linhas representam os valores previstos do modelo e as colunas representam os valores reais.

Resumidamente, os resultados aqui apresentados foram obtidos com os classificadores individuais a partir de cortes manuais sobre as imagens BrG. Todos os resultados gráficos para taxas de aprendizado, taxas de perda e curvas ROC foram apresentados considerando a validação cruzada quádrupla. Considerando a curva ROC de cada um dos modelos CNNs individuais, cada linha apresentada no gráfico corresponde a uma rodada de validação cruzada. Os gráficos de acurácia e perda apresentada mostraram que a precisão dos testes segue a precisão do treinamento, indicando que os algoritmos sofrem de pouco *overfitting*. Além disso, os resultados das curvas também expressam a redução do erro das redes conforme se passam as épocas de treinamento.

Por fim, os resultados adquiridos até esta etapa serão utilizados como parâmetros para

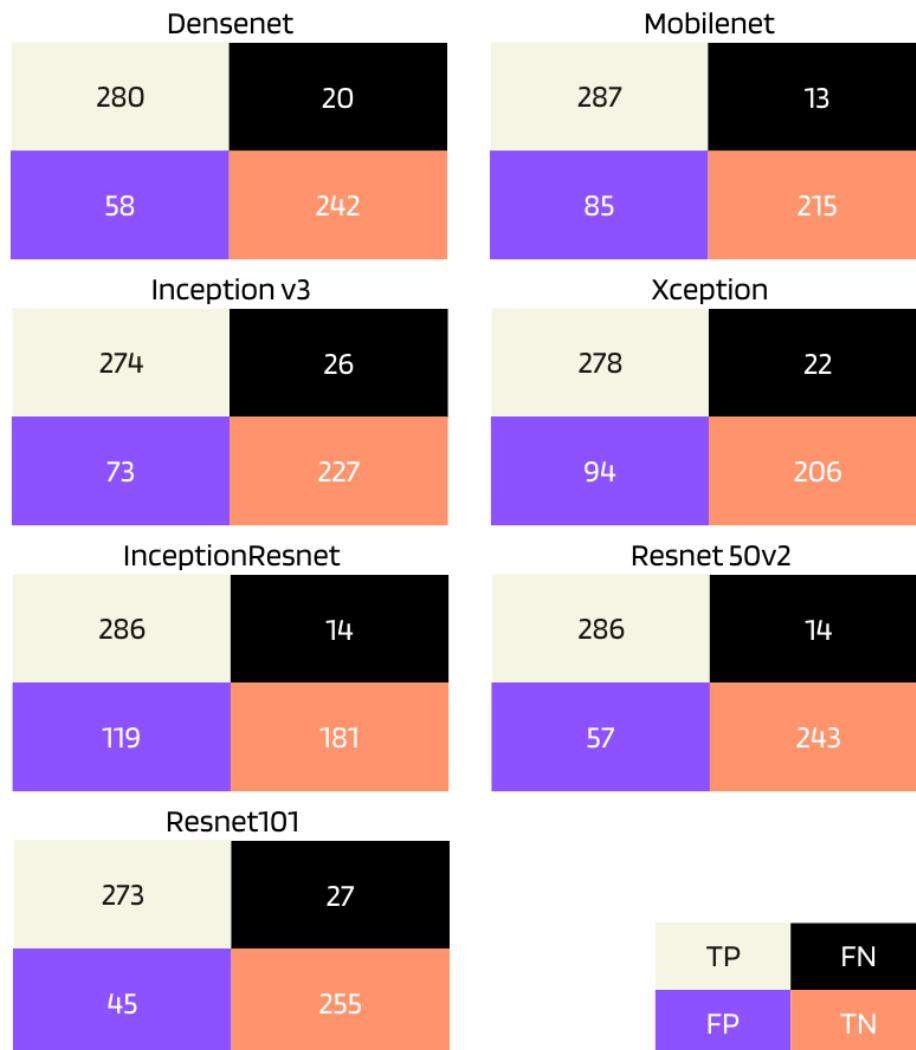


Figura 5.21: Matriz de confusão de cada um dos modelos de CNNs individuais.

comparação com resultados obtidos por um *ensemble* também treinado com as imagens BrG a partir da segmentação de outros cortes de regiões de interesse, bem como resultados de classificação sobre imagens globais.

## 5.5 Construção do ensemble

Existem vários métodos para programar tais combinações, como o método de combinação linear, de combinação não linear e baseados em estatística (Thomas, 1997). Desta forma, são citados três tipos principais de meta-algoritmos que visam combinar preditores individuais:

- *Bagging*: tem uma implementação simples e intuitiva; a diversidade é alcançada pela substituição com diferentes subconjuntos de dados criados aleatoriamente;

---

cada subconjunto é usado para treinar um classificador do mesmo tipo; as saídas dos classificadores são baseadas nas suas decisões por uma maioria de votos; para uma determinada instância, a categoria com mais votos torna-se a resposta (Bagui, 2005).

- *Boosting*: muitas vezes considera aprendizes fracos homogêneos, ensina-os sequencialmente de forma muito adaptativa (um modelo base depende dos anteriores) e os combina seguindo uma estratégia determinística, geralmente com saídas de modelos com os valores ponderados (Bagui, 2005).
- *Stacking*: geralmente considera preditores fracos heterogêneos, ensina-os em paralelo e os combina treinando um metamodelo para aprender o mapeamento entre as saídas e as classes corretas.

De maneira simplificada, pode-se dizer que *bagging* focará em obter um modelo *ensemble* com menos variância que os seus componentes, enquanto *boosting* e *stacking* tentarão produzir um modelo forte com menos viés que os seus componentes, além de também ser um modo de reduzir a variância dos resultados (Bagui, 2005).

Em resumo, esses meta-algoritmos se difunde em diversos meios de combinações, o mais comum das aplicações de *ensemble* é de forma linear por meio de uma votação simples entre os classificadores individuais, ou também por meio de uma votação com ponderação de confiança, onde os votos são ponderados com base no valor de confiança de cada predição individual (IBM, 2022b).

Quanto à construção do *ensemble* utilizado neste trabalho, primeiramente, observou-se que os algoritmos individuais são promissores na classificação do glaucoma, principalmente os modelos (ResNet50v2 e ResNet101), que apresentaram os melhores resultados, porém, para melhorar ainda mais a precisão da classificação geral das imagens de fundo de olho em estudo, os classificadores individuais na Tabela 5.1 são concatenados para formar um *ensemble*, conforme mostrado na Figura 5.22

Algo a se considerar é que não há nenhuma vantagem em um *ensemble* com classificadores idênticos ou muito parecidos, pois, votariam igualmente na classificação dos dados. Portanto, neste trabalho foram considerados os modelos de algoritmos que generalizam de maneiras diferentes e que provavelmente não cometerão erros coincidentes, ou seja, que apresente diferentes erros em diferentes instâncias dos dados. Assim, neste trabalho essa diversidade de classificadores foi encontrada utilizando os algoritmos individuais e pré-treinados.

Para chegar a combinação representada pela Figura 5.22, foram testadas combinações dos sete algoritmos listados em ordem de maior acurácia na Tabela 5.2, excluindo o algoritmo menos preciso em cada combinação testada: assim, a Combinação 1 foi realizada concatenando as saídas de todos os sete classificadores individuais, e a Combinação 2

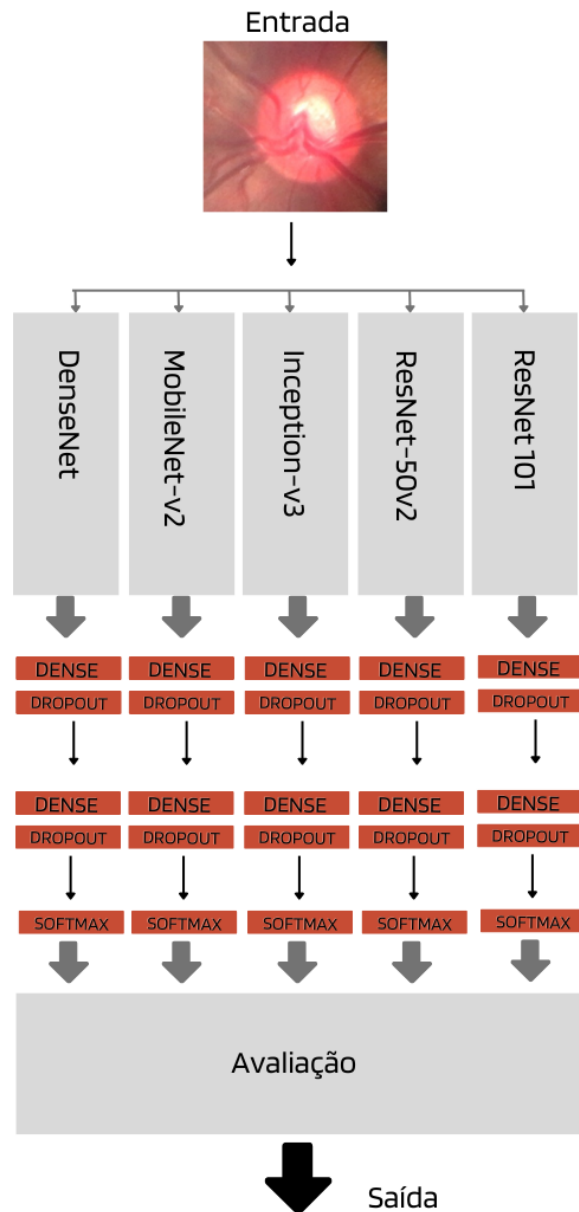


Figura 5.22: Modelo de *ensemble* usando a melhor combinação dos classificadores individuais.

foi conduzida concatenando as saídas dos seis modelos individuais mais precisos. Essas combinações foram realizadas até a Combinação 6, que possuía apenas os dois algoritmos individuais com maior acurácia, conforme Tabela 5.3.

Após a escolha dos algoritmos para a formação do *ensemble*, os resultados destes classificadores foram concatenadas. Dessa forma, o resultado do *ensemble* é obtido pela média das probabilidades dos classificadores individuais para obter a probabilidade única de que a imagem represente um paciente com ou sem glaucoma. Esta é uma metodologia que se aproxima da técnica conhecida como *bagging*. A proposta é que um *ensemble* de vários modelos CNNs possa melhorar significativamente a precisão das previsões, comparando com a precisão de qualquer modelo único incluído no *ensemble*. Além disso,

Tabela 5.3: Combinações de CNNs avaliadas na construção *ensemble*

CNN	COMBINAÇÕES					
	1	2	3	4	5	6
ResNet50v2	X	X	X	X	X	X
MobileNet	X	X	X	X	X	X
DenseNet	X	X	X	X	X	
Inception-v3	X	X	X	X		
ResNet101	X	X	X			
InceptionResNet	X	X				
Xception	X					

essa técnica também pode contribuir para redução de *overfitting*.

Para melhor compreensão da combinação dos classificadores individuais e formação do *ensemble*, a Figura 5.23 mostra um exemplo em que as imagens devem ser classificadas em duas categorias: normal ou glaucoma. Portanto, assumindo que a função *softmax* é utilizada na camada de saída de cada classificador que compõe o *ensemble* de CNNs, logo, a saída do teste é dada pela probabilidade de que a imagem de entrada pertença a uma das classes categorias preestabelecidas.

Considerando o exemplo dado, a resposta final do *ensemble* é derivada da média dessas probabilidades, gerando uma única probabilidade de uma imagem ser ou não glaucomatosa. Na ilustração dada como exemplo, o resultado mostra que a imagem tem 5,649% de probabilidade de não ser glaucoma, e 4,350% de probabilidade de ser glaucoma. Portanto, com base nesse exemplo, a imagem seria classificada como não glaucomatosa.

## 5.6 Resultados do ensemble - corte manual

Após a construção do *ensemble*, foram efetuados testes com a base de dados BrG. Os resultados para todas as combinações estabelecidas com os classificadores individuais são apresentados na Tabela 5.4.

Tabela 5.4: Resultados obtidos pela combinação dos classificadores.

Combinações	ACC	SE	ES	PR	F1	AUC
1	0,861	0,767	0,957	0,946	0,847	0,967
2	0,865	0,763	0,966	0,858	0,849	0,968
3	0,905	0,850	0,960	0,955	0,899	0,965
4	0,865	0,770	0,960	0,950	0,850	0,964
5	0,838	0,703	0,973	0,963	0,813	0,963
6	0,853	0,730	0,976	0,969	0,832	0,961

Por fim, após considerar o maior valor de acurácia, o melhor *ensemble* se deu pela Combinação 3, com a adição dos classificadores ResNet50v2, MobileNet, DenseNet,

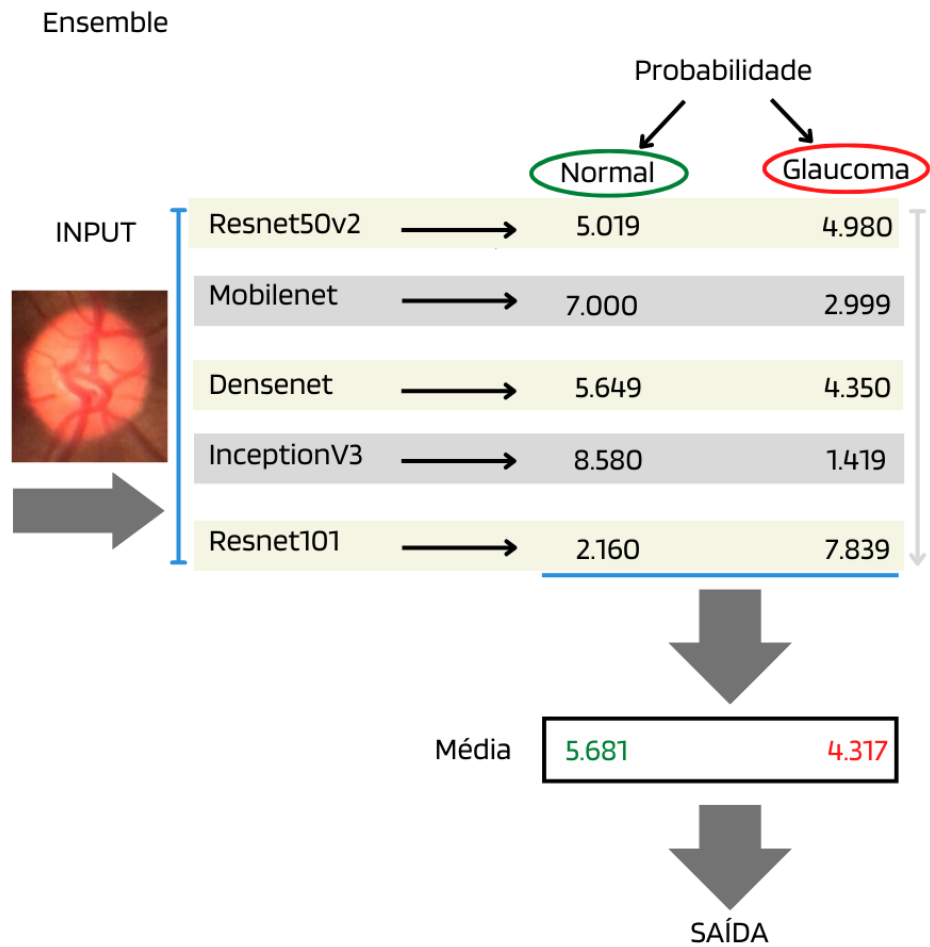


Figura 5.23: Exemplo da resposta final do classificador com base nas respostas individuais dos algoritmos selecionados para compor o *ensemble*.

Inception-v3 e ResNet101, com resultados apresentados na matrix de confusão da Figura 5.24

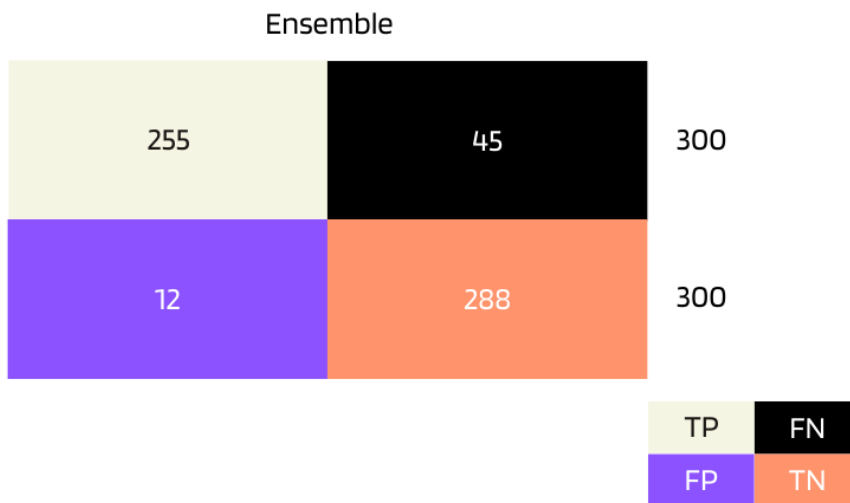


Figura 5.24: Matriz de confusão com resultado do ensemble - corte manual.

Consolidando o melhor desempenho na classificação das imagens BrG aplicado em cortes manuais da região de interesse, a Figura 5.25 mostra graficamente a curva ROC obtida pela combinação 3.

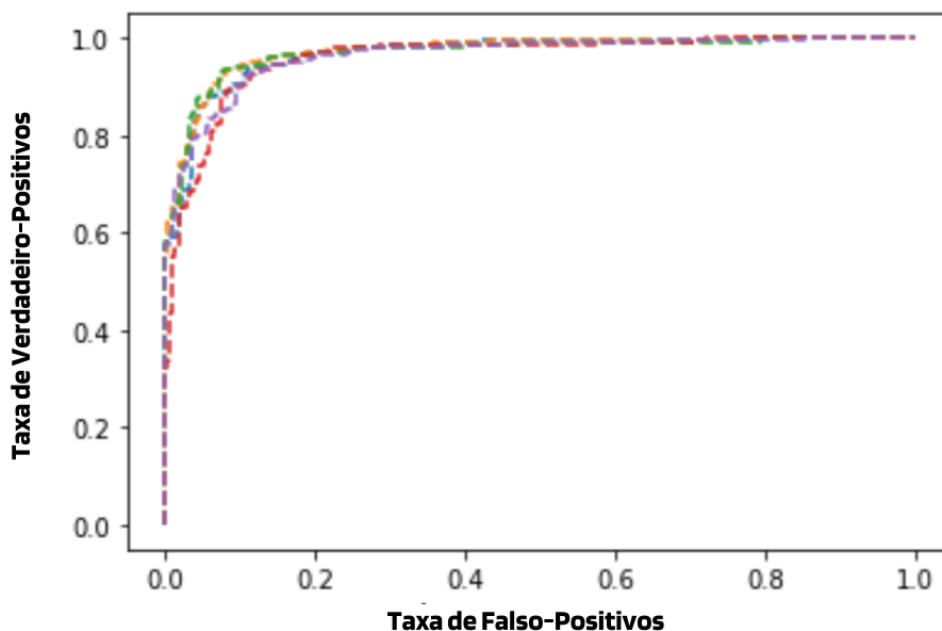


Figura 5.25: Área sob a curva ROC para o modelo *ensemble*, cada linha ( corresponde a uma rodada de validação cruzada tripla.

## 5.7 Experimentos comparativos

Para que a forma de coleta da região de interesse nas imagens BrG não seja um método questionável durante o processo de classificação, também foram realizados testes de classificação a partir de extrações adicionais da região de interesse. Para tanto, foram também realizados um teste de classificação utilizando imagens globais, ou seja, sem nenhum processo de segmentação, como mostra a Figura 5.26 (A). Aplicou-se também um teste de classificação com as imagens segmentadas automaticamente na região do disco óptico e áreas adjacentes, considerando a máscara apresentada na Figura 5.26 (B) e, por fim, um teste de classificação em imagens também automaticamente segmentadas nas regiões específicas do disco óptico, representada pela máscara da Figura 5.26 (C).

A escolha dessas regiões de interesse se fez devido ao ângulo de visão em que as imagens BrG foram fotografadas. O ângulo de visão pode determinar a quantidade de área do fundo de olho que aparecerá próximo ao disco óptico. Isso é relevante para os algoritmos de classificação, pois, no caso de ter como entrada uma imagem com amplo campo de visão da retina para classificação do glaucoma, pode ser necessário segmentar

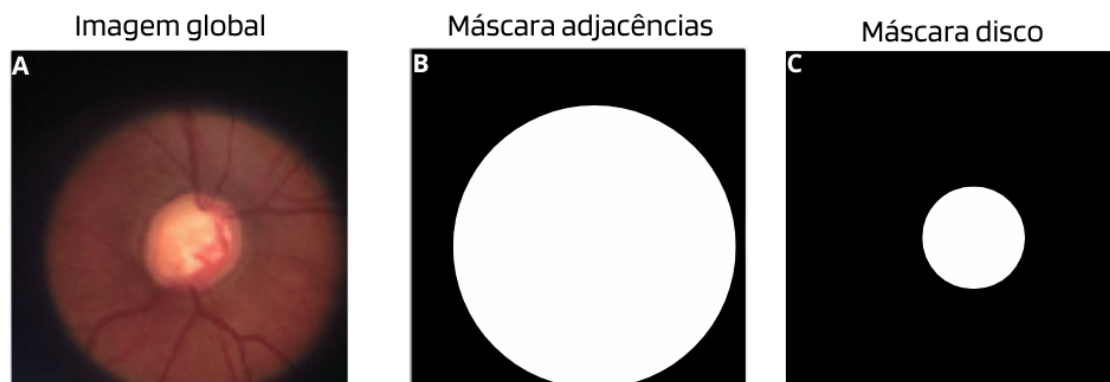


Figura 5.26: Base de dados BrG - Imagem global à direita (A), ao centro exemplo de máscara da região do disco e adjacências (B), à esquerda exemplo de máscara específica do disco óptico (C).

a região de interesse que é o disco óptico. Faz-se isso para que grandes áreas não sejam analisadas desnecessariamente pelo algoritmo.

Quando se trata de imagens com ângulos de visão menores, a necessidade de segmentar a região do disco óptico pode ser um pouco diferente, pois o que se vê ao redor do disco óptico é uma área muito pequena comparado às imagens com maiores ângulos de visão. Para melhor entendimento, diferentes aparelhos possuem diferentes ângulos de visão e consequentemente diferentes imagens. A Figura 5.27 mostra exemplos de imagens de fundo de olho vistas com diferentes dispositivos e de diferentes ângulos de visão.

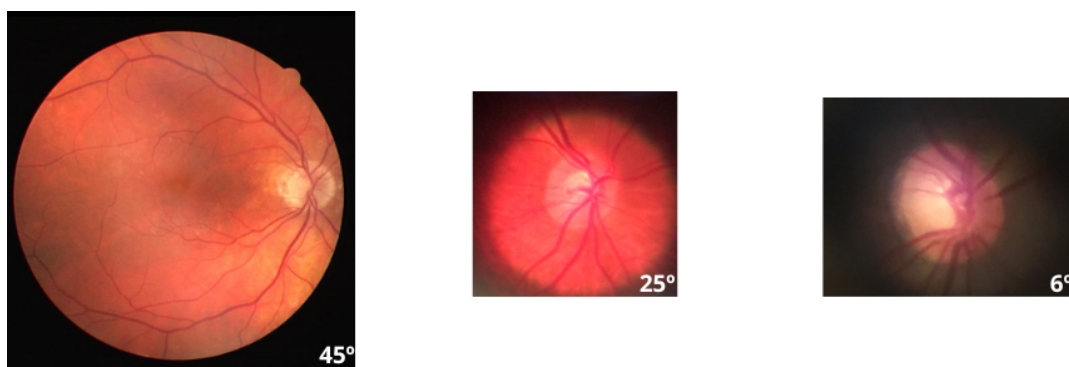


Figura 5.27: Exemplos de imagens de câmeras com diferentes ângulos de visão - (drive, panoptic, d-eye).

Deste modo, considerando imagens de 25° de visão, toda a área que representa o fundo de olho ou a retina poderá ser utilizada pelos algoritmos de classificação, sem comprometer muito os esforços necessários para análise das imagens por parte dos algoritmos, o que anularia ou no mínimo facilitaria os trabalhos de segmentação.

Facilitar ou eliminar a necessidade de segmentar regiões de interesse pode ser relevante no processamento de imagens de fundo de olho, pois segmentar tais estruturas é uma tarefa delicada e pode representar um risco para a precisão final dos algoritmos de

*deep learning* na classificação de doenças como o glaucoma, uma vez que o processo de segmentação pode levar a uma combinação de erros, ou seja, possíveis erros obtidos durante a tarefa segmentação e possíveis erros do algoritmo durante a tarefa de classificação. Portanto, nesta Seção, foi testado o uso de imagens globais e segmentadas em diferentes regiões e modalidades (automáticas ou manuais), a fim de saber qual ROI apresentaria melhores resultados no processo de classificação. Assim, para a tarefa de segmentação de imagens BrG, utilizou-se um algoritmo chamado U-Net, que é brevemente descrito a seguir.

### 5.7.1 Algoritmo U-Net utilizado para segmentação

A segmentação automática de imagens digitais (principalmente imagens médicas) é uma tarefa desafiadora devido às suas naturezas complexas, ainda assim, é uma técnica frequentemente utilizada na medicina para auxiliar em pesquisas anatômicas, como identificar regiões de interesse, medir tecidos volumétricos, etc.

Quanto aos algoritmos utilizados, o processo de segmentação também pode ser obtido por meio de redes CNNs, no entanto, quando essas redes são utilizadas para segmentação de imagens, elas não possuem camadas totalmente conectadas e são chamadas de redes totalmente convolucionais, em inglês *Fully Convolutional Networks* (FCN). Uma das principais redes com essa metodologia é a rede *U-Net*, exemplificada na Figura 5.28.

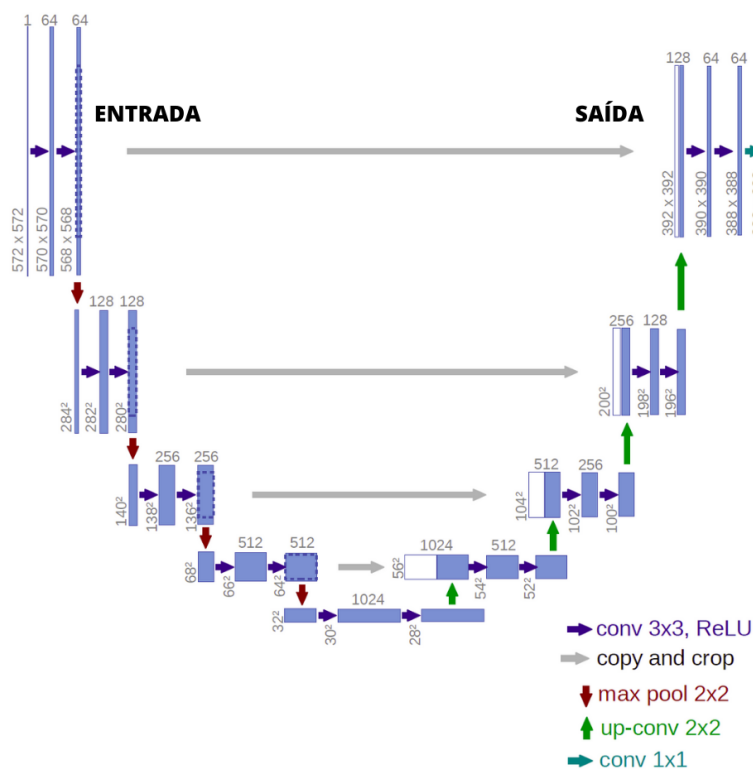


Figura 5.28: Arquitetura de uma rede U-Net representando um bloco de entrada e mapa de segmentação como saída adaptado de (Ronneberger et al., 2015).

---

A rede *U-Net*, proposta por (Ronneberger et al., 2015), é uma versão estendida das redes FCNs, que proporciona segmentações mais precisas com conjuntos de treinos pequenos. Essa arquitetura foi originalmente desenvolvida para segmentação de imagens biomédicas. A sua principal vantagem é o treinamento mais eficiente de redes *deep learning*, sem a necessidade de milhares de exemplos de treinamento rotulados. O objetivo dessa rede é fornecer uma forma de segmentação chamada semântica. Essa abordagem semântica consiste em atribuir a cada *pixel* um rótulo de classe e gerar a estrutura a ser segmentada.

O algoritmo *U-Net* consiste em dois caminhos, sendo divididos em caminho de contração à esquerda e caminho de expansão à direita. Devido à alta simetria entre os dois, esta rede possui o formato de *U*, dando origem ao seu nome. O caminho de contração segue a arquitetura tradicional das redes convolucionais e consiste em aplicações repetidas de duas convoluções  $3 \times 3$ , cada uma seguida por uma função de ativação *ReLU* e uma operação de *max pooling* com *stride*  $2 \times 2$  para redução da taxa de amostragem.

A reconstrução da entrada inicia-se por meio da expansão. Cada bloco de expansão é feito com camadas de *upsampling* ou convolução transposta  $2 \times 2$ . Ao final, uma camada de convolução  $1 \times 1$  é aplicada para reduzir um espaço multidimensional para uma imagem monocromática, tendo como saída a imagem já segmentada com a mesma resolução espacial do início, para ter a mesma dimensão (Weng and Zhu, 2021).

### 5.7.1.1 Treinamento e testes da rede U-Net com imagens BrG

Considerando o uso da rede *U-Net* para segmentar as estruturas das imagens BrG, a arquitetura do modelo utilizado consiste num codificador (para *downsampling*) e um decodificador (para *upsampling*), conforme exigência padrão da rede *U-Net*. Desta forma, as etapas de utilização da rede são apresentadas a seguir:

- Primeiramente as imagens e máscaras foram redimensionadas para  $128 \times 128$  *pixels*;
- A normalização das imagens foi aplicada para o intervalo de  $[-1, 1]$ ;
- O codificador (ou caminho de contração) foi criado e especificando o otimizador (Adam) e função de perda *crossentropy*, em seguida definiu-se o decodificador (ou caminho de expansão) da rede;
- Durante o treinamento, o modelo usou aumento de dados para aumentar a capacidade de aprendizado da rede, aplicando rotação, translação e distorção aos dados, resultando em um aumento de 4 vezes no número de imagens, como demonstrado na Figura 5.29.
- O número de épocas de treinamento é determinado pela técnica *early stopping*.

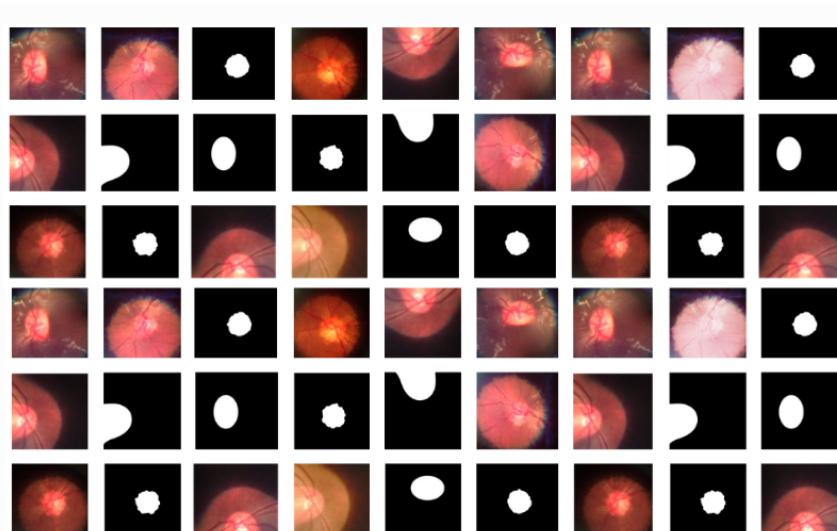


Figura 5.29: Exemplos de distorção aplicada a técnicas como aumento de dados.

O tipo de treinamento aplicado por esta rede é supervisionado, ou seja, para cada entrada de dados, uma saída desejada é fornecida. A rede usa essas informações para ajustar a função objetivo e, ao final do processo, as regiões de interesse sejam identificadas corretamente em parâmetros de precisão viáveis. Deste modo, para o treinamento da rede *U-Net* foi feito passando como entrada as imagens originais acompanhadas das suas máscaras da região a ser segmentada, e a saída é a previsão da região que espera ser segmentada.

Para entender melhor o processo de segmentação utilizada, o fluxograma da Figura 5.30 descreve desde a divisão da base de dados BrG até o processo de segmentação. Assim, as Etapas 1 e 2, do fluxograma, correspondem ao processo de treinamento da rede *U-Net*, enquanto as Etapas 3 e 4, correspondem a utilização da rede *U-Net* treinada para segmentação do restante das imagens BrG não utilizadas durante o treinamento da rede, na sequência foram utilizadas em testes de classificação.

Para o treinamento da rede *U-Net*, a Etapa 1 do fluxograma mostra que, primeiramente, foram separadas aleatoriamente 200 imagens da base de dados BrG, divididas em 100 imagens de glaucoma dos olhos esquerdo e direito, e 100 imagens sem glaucoma também dos olhos esquerdo e direito.

Na Etapa 2 do fluxograma, as 200 imagens selecionadas foram divididas em 70% para treinamento e 30% para teste da rede *U-Net*. O treinamento foi aplicado passando como entrada as imagens originais, seguidas das máscaras da região que pretende segmentar, ou seja, a máscara do disco óptico e região adjacente ou apenas a máscara específica do disco óptico, como mostra a Figura 5.31. Em seguida, para usar a comparação de hiperparâmetros dos modelos *deep learning*, separou-se 10% das imagens do conjunto de treinamento para criar um conjunto de validação.

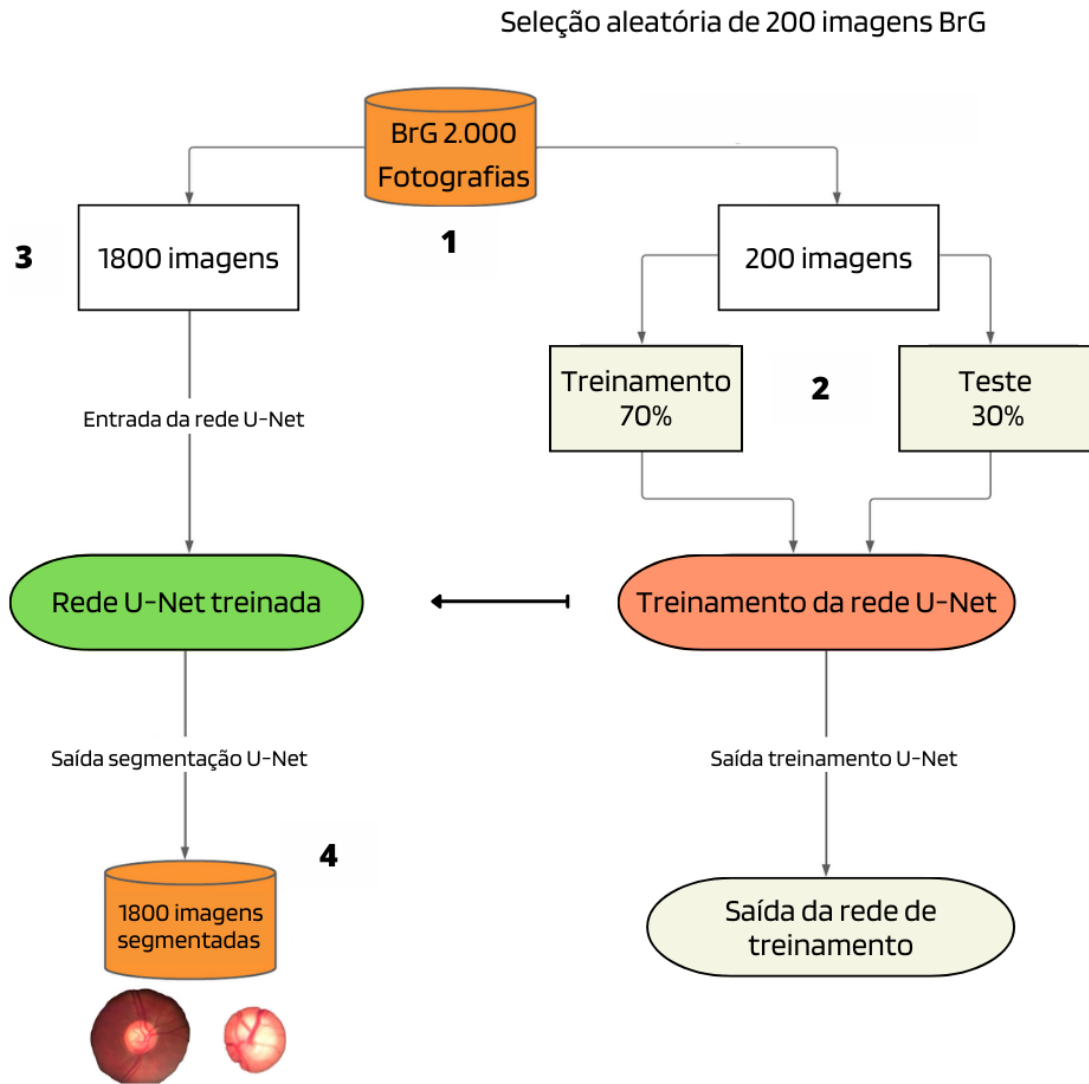


Figura 5.30: Fluxograma para aplicação do algoritmo de segmentação *U-Net* .

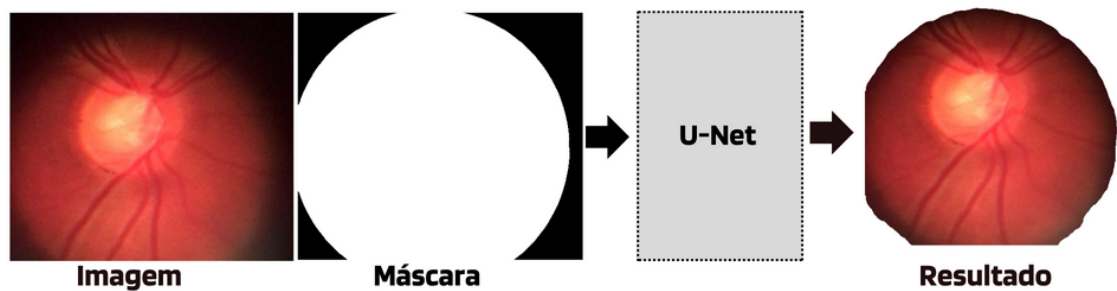


Figura 5.31: Imagem global seguida pela máscara do disco óptico e região adjacente e a predição de segmentação na camada de saída.

### 5.7.1.2 Resultados do teste de segmentação

Aplicada a divisão de dados, a rede *U-Net* foi treinada e testada usando 200 imagens, sendo: 140 para treinamento e validação, e 60 imagens para teste. O tempo médio

de treinamento foi de 50 minutos, levando em consideração as especificidades do disco óptico específico e áreas adjacentes. Em análise resultante, a rede *U-Net* apresentou IoU de 0,895 para segmentação do disco óptico e adjacências, e também IoU de 0,875 para segmentar regiões específicas do disco óptico. Os resultados desta segmentação são apresentados na Tabela 5.5.

Tabela 5.5: Precisão dos algoritmos usados para segmentar imagens BrG

ROI	Imagens Treinamento	Imagens Teste	IoU
Região específica do disco óptico	140	60	0,875
Disco óptico e adjacências	140	60	0,895

Após o teste de segmentação nas imagens BrG, o algoritmo U-Net treinado foi salvo para que o restante das imagens no banco de dados, ou seja, 1800 imagens, pudessem ser segmentadas de acordo com regiões específicas do disco e disco óptico e adjacências. O produto desse processo será levado em seguida para a tarefa de classificação do glaucoma com uso do ensemble de CNNs.

### 5.7.2 Segmentação automática e específica do disco óptico

Para realizar este procedimento, a rede *U-Net* (Ronneberger et al., 2015) foi utilizada e configurada para esta função, como mostra a Figura 5.32.

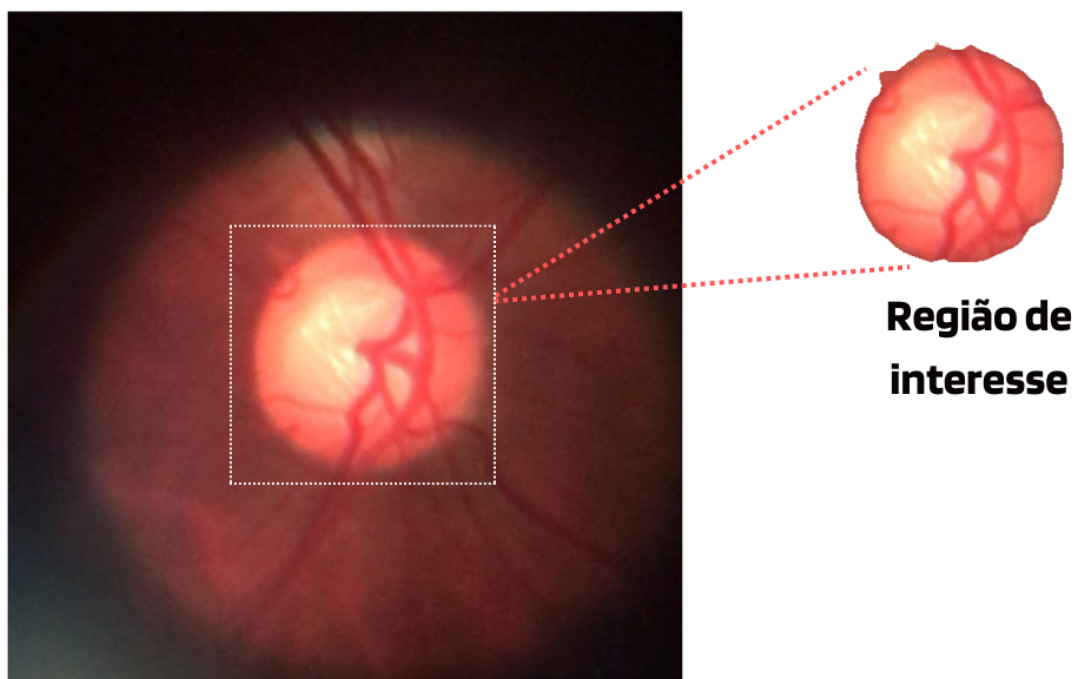


Figura 5.32: Extração automática da região de interesse - disco óptico específico.

Após a segmentação, os recortes gerados pela operação foram salvos no formato PNG. A tarefa de segmentação específica do disco óptico pode ser desafiadora para al-

goritmos gerais, principalmente considerando a base de dados BrG, que contém imagens com menor resolução e que geralmente apresentam maior quantidade de ruídos em comparação a imagens de equipamentos convencionais.

### 5.7.3 Segmentação automática do disco óptico e adjacências

Para aquisição deste tipo de região de interesse das imagens BrG, em vez de segmentar a região específica do disco óptico, como tem sido apresentado por abundante quantidade de estudos relacionados, neste trabalho, utilizou-se a rede *U-Net* básica e, aproveitando-se do ângulo de visão  $25^\circ$  em que as imagens foram capturadas, foi segmentada automaticamente a região do disco óptico e áreas adjacentes, como mostra a Figura 5.33. Em outras palavras, esta é uma forma de excluir regiões pretas das imagens, ou seja, aquelas em que não apresentam nenhuma imagem da retina. A diferença entre usar uma rede *U-Net* para tal segmentação e usar, por exemplo, algum algoritmo de limiarização para excluir áreas indesejadas é o resultado obtido, pois a segmentação semântica pode apresentar maior robustez em seus resultados.

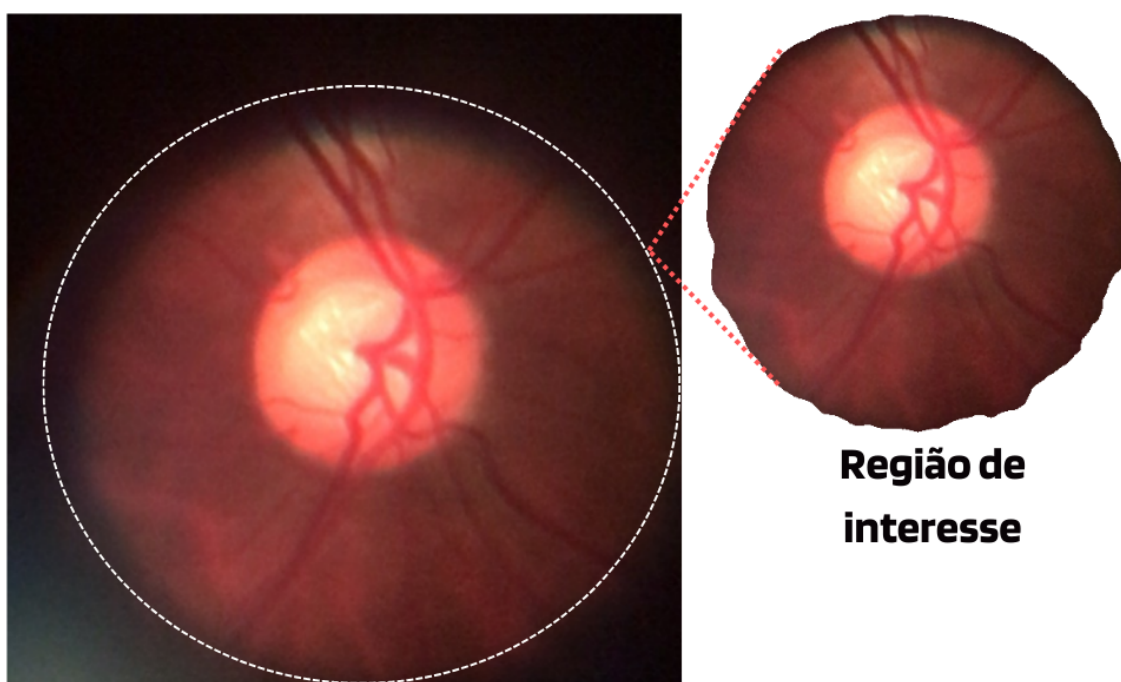


Figura 5.33: Extração automática da ROI - disco óptico e adjacências.

Uma limitação da segmentação do disco óptico e suas regiões adjacentes é que o resultado dessa segmentação pode não ser útil para estudos nos quais os valores de CDR são calculados. Essa limitação existe devido à área adjacente ao redor do disco óptico. No entanto, para algoritmos de classificação que usam a área do disco e não requerem cálculos de CDR, isso pode ser usado.

#### 5.7.4 Observações sobre as imagens segmentadas

Devido à diversidade estrutural da retina, percebe-se que para muitas imagens a segmentação produz resultados próximos aos esperados considerando as máscaras apresentadas no treinamento e testes. Isso pode ser verificado na Figura 5.34 que buscou segmentar a região específica do disco óptico.

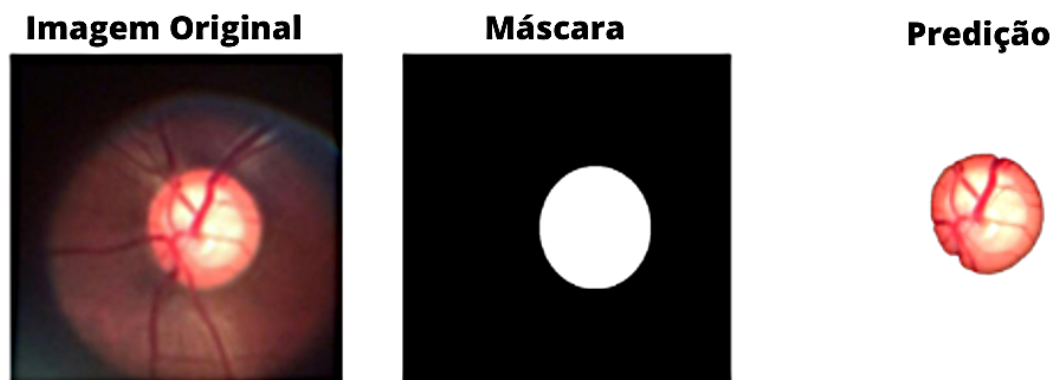


Figura 5.34: Imagem global, máscara do disco óptico, e região específica do disco óptico segmentada com toda estrutura preservada.

Ainda considerando a diversidade estrutural da retina, embora o algoritmo chegue próximo ao ideal para algumas imagens BrG na segmentação específica do disco óptico, para um percentual da base de dados, o algoritmo apresentou erros graves na busca pela extração da região de interesse nas imagens digitais, com isso foram destruídas partes importantes para o diagnóstico do glaucoma, como se vê na Figura 5.35.

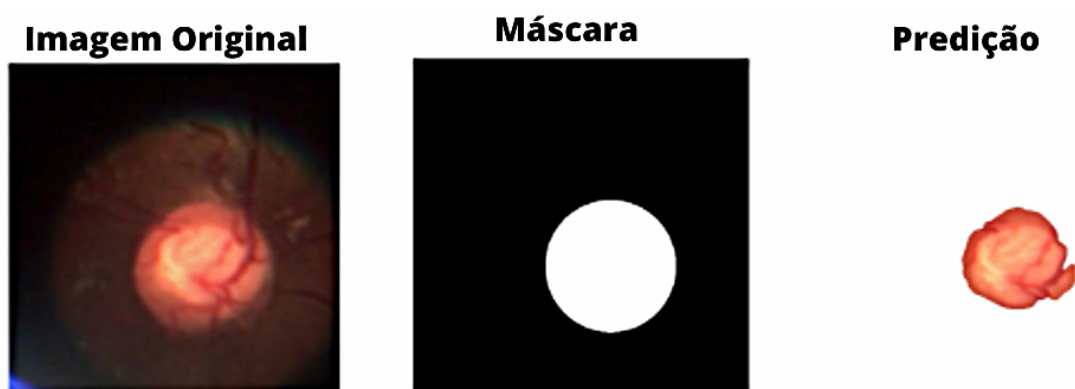


Figura 5.35: Imagem global, máscara do disco óptico, e região específica do disco óptico segmentada com destruição de parte do conteúdo.

Para a segmentação da região de interesse do disco óptico e adjacências, o algoritmo segue desempenho semelhante à segmentação específica do disco óptico. Portanto, o algoritmo também segmenta essa região de interesse em um limite muito próximo do ideal em um grande percentual das imagens BrG, como mostra a Figura 5.36.

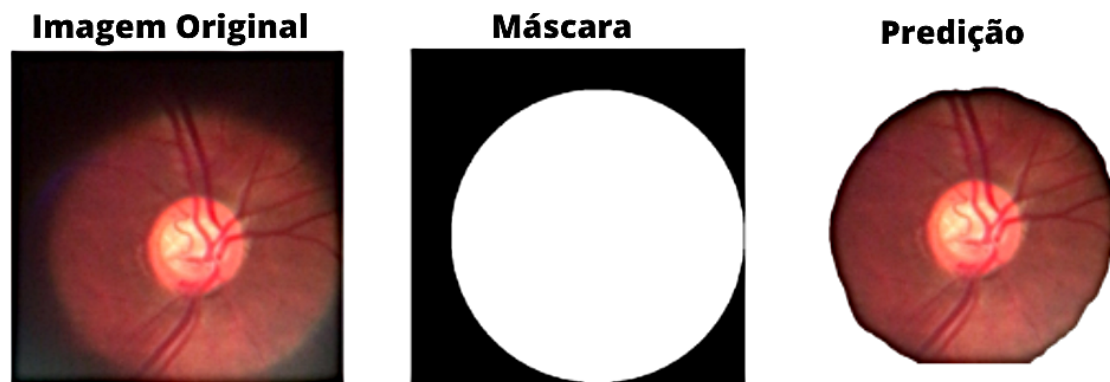


Figura 5.36: Imagem global, máscara do disco óptico e adjacências, e região segmentada do disco óptico e adjacências com perfeita preservação de partes importantes do disco e *pixels* vizinhos.

Por outro lado, a segmentação do disco óptico e adjacências também apresentou erros, no entanto, são erros que ao menos não comprometeu a região mais importante para o diagnóstico do glaucoma, ou seja, não comprometeu em nenhuma das imagens a região do disco óptico, como mostra a Figura 5.37.

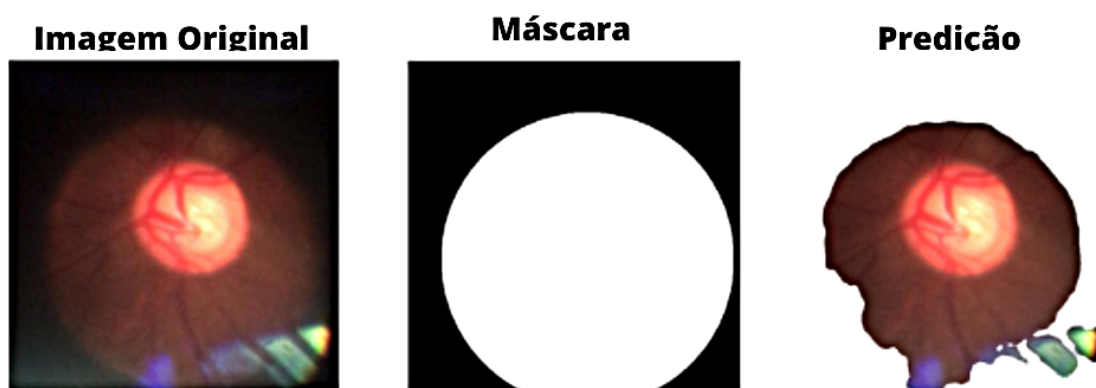


Figura 5.37: Imagem global seguida da máscara manual e da região segmentada do disco óptico e adjacências com preservação do disco óptico e com danos aos *pixels* vizinhos.

Em análise de resultados, há semelhanças entre os valores de IoU para os dois modos de segmentação (disco óptico e adjacências e específica do disco e disco. No entanto, as imagens com segmentação específica do disco óptico sofreram maiores problemas, pois quando o algoritmo erra no processo de segmentação, a região mais importante para o diagnóstico do glaucoma, o disco óptico, era afetado diretamente, retirando partes das imagens que poderiam ser primordiais para o diagnóstico automático do glaucoma. Em contrapartida, a segmentação da região do disco óptico e adjacências, quando afetada por erros de segmentação, perdiam partes da região das bordas, mas conservavam a região do disco óptico sem prejuízos.

---

### 5.7.5 Segmentando imagens BrG com a rede U-Net

Prosseguindo nas etapas do fluxograma da Figura 5.30, a Etapa 3 mostra a utilização da rede *U-Net* já treinada durante Etapa 2, possibilitando a segmentação automática da região de interesse das outras 1.800 imagens da base de dados BrG; preparando-as para a tarefa de classificação que ocorrerá com as imagens segmentadas apresentadas na saída da Etapa 4 do fluxograma.

Por fim, a tarefa de classificação utilizando o *ensemble* preparado para classificação das imagens com cortes manuais foi aplicado à classificação das imagens automaticamente segmentadas. Portanto, a seguir são apresentados os resultados dessas operações.

## 5.8 Resultados definitivos da classificação

Os resultados na tabela 5.6 mostram a precisão geral para todos os testes envolvendo a classificação das imagens BrG a partir da (segmentação manual do disco óptico, disco óptico e regiões adjacentes; regiões específicas do disco, bem como imagens globais).

Tabela 5.6: Resultados da classificação das imagens

Modelo / Região de interesse	ACC	SE	ES	PR	F1	AUC
Individual - ResNet50v2	0,881	0,953	0,810	0,833	0,889	0,956
Ensemble - Corte manual	0,905	0,850	0,960	0,955	0,899	0,965
Ensemble - Disco óptico específico	0,870	0,910	0,833	0,845	0,876	0,963
Ensemble - Disco óptico e adjacências	0,954	0,985	0,922	0,927	0,955	0,979
Ensemble - Imagens globais	0,928	0,918	0,937	0,936	0,927	0,976

Conforme os resultados apresentados pode-se inferir que o *ensemble* apresenta melhores resultados em comparação com o melhor dos classificadores individuais, mostrando que a metodologia de concatenação de algoritmos melhora a precisão geral da classificação do glaucoma, no estudo em questão.

A Figura 5.38 representa uma matriz de confusão, onde as linhas representam os valores reais do modelo e as colunas representam os valores previstos para todos os testes de classificação com referência as três últimas configurações de dados: região específica do disco, disco óptico e adjacências e imagens globais.

Em análise de resultados, a diferença no número de imagens representadas na matriz de confusão entre (imagens globais e segmentadas automaticamente) refere-se à proporção da partição do conjunto de dados BrG utilizada para treinar o algoritmo *U-Net*. Desta forma, os conjuntos de treinamento e testes com imagens segmentadas foram automaticamente formados a partir da divisão de 1.800 imagens, pois 200 imagens foram reservadas para o treinamento e testes do próprio algoritmo de segmentação, enquanto, os conjuntos de treinamento e testes com imagens globais ou segmentadas foram manualmente formados a partir de 2.000 imagens.

Este experimento é um exemplo claro do potencial dos algoritmos de *deep learning* para identificar padrões de glaucoma em imagens digitais de fundo de olho, mesmo quando se trata de imagens de menor resolução como é a base de dados BrG. Prova disso, o método proposto consegue identificar o glaucoma a partir das imagens do fundo de olho adquiridas por smartphones, e os resultados foram melhores quando as imagens foram segmentadas na área do disco óptico e regiões adjacentes, apresentando acurácia de 0,953% e AUC de 0,979%.

Os resultados obtidos até agora mostram capacidades algorítmicas para classificação de glaucoma ou possivelmente diagnóstico automatizado. No entanto, não foi exposta à possibilidade de aplicação prática de técnicas como o algoritmo desenvolvido, principalmente considerando o rastreamento populacional de glaucoma. Portanto, no próximo capítulo, são apresentadas algumas considerações estatísticas que devem ser levadas em consideração ao discutir o uso de algoritmos de *deep learning* para triagem populacional de glaucoma.

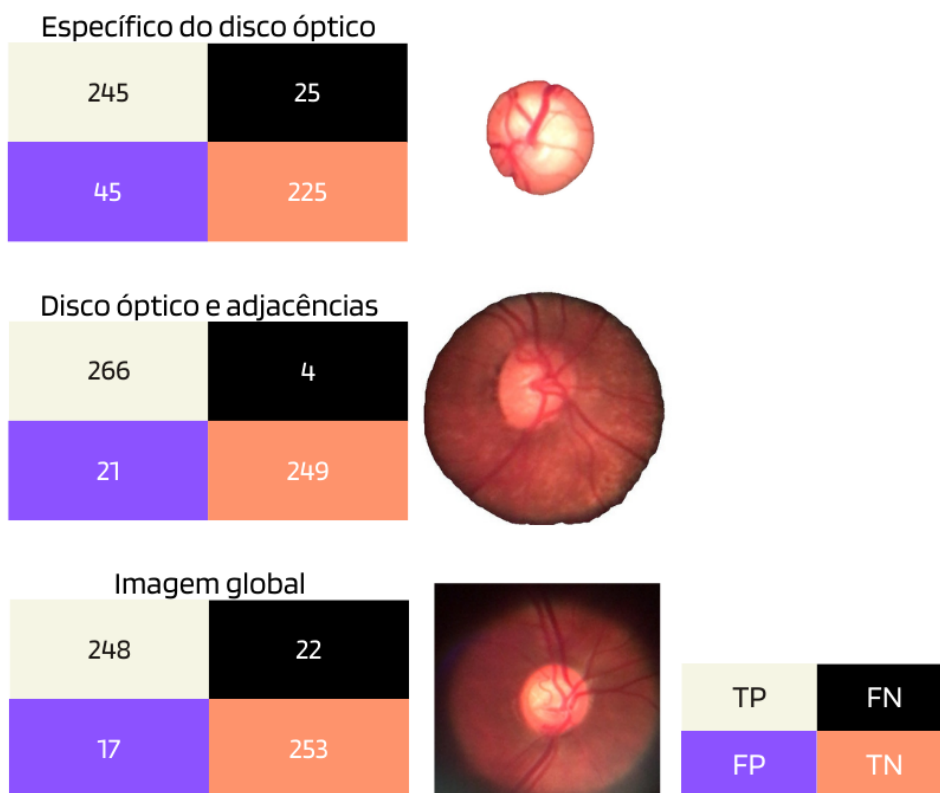


Figura 5.38: Matriz de confusão das regiões: específica do disco óptico, disco óptico e adjacências, imagens globais.

# Capítulo 6

## Análise de resultados e prática clínica

Este Capítulo apresenta uma análise dos resultados obtidos e levanta questões relacionadas à possibilidade de aplicações práticas de algoritmos de *deep learning* como ferramenta de triagem populacional do glaucoma. Portanto, são apresentadas as principais dificuldades encontradas na prática clínica dos algoritmos em geral, bem como do algoritmo desenvolvido neste trabalho.

### 6.1 Deep learning e triagem populacional do glaucoma

Considerando os algoritmos de IA, a sua aplicação oftálmica em casos de glaucoma pode ser entendida como um teste rápido e barato que pode indicar a necessidade de uma consulta mais detalhada com um oftalmologista, ou seja, uma tecnologia com capacidade de auxiliar no diagnóstico dessa neuropatia óptica. Assim, quando se considera a aplicabilidade de tais algoritmos, rapidamente se imagina a possibilidade de aplicá-los na forma de triagens populacionais. Especialmente porque é o método mais aceito, e pode realmente ajudar a reduzir o número de pessoas subdiagnosticadas com essa doença. Portanto, este tema foi objeto de muitos estudos relacionados, dos quais vários autores concluíram que algoritmos de *deep learning* podem ser usados como uma ferramenta para triagem de populacional do glaucoma.

Para analisar mais detalhadamente essa situação, é importante entender que o programa de triagem populacional aqui tratado, seria um processo de priorização do atendimento oftalmológico de pacientes com base na gravidade de sua condição de saúde ocular. É uma das estratégias que podem ser utilizadas na saúde pública em que: entre as pessoas consideradas "normais", certas pessoas sejam identificadas como de risco para a doença. Portanto, os indivíduos identificados se beneficiariam de uma investigação mais aprofundada para confirmar e/ou descartar tais riscos (WHO, 2020).

Um grande desafio com aplicações de triagem populacional em geral, é que esses testes muitas vezes não são 100% precisos. Assim, qualquer teste aplicado pode ter como

---

respostas casos de verdadeiro-positivos (VP) e verdadeiro-negativos (VN), além disso, sabe-se que uma pessoa pode ser informada de que tem uma doença quando ela não a tem, isso é conhecido como falso positivo – (FP) e pode levar algumas pessoas a fazer mais testes ou tratamento desnecessário. De modo análogo, os testes de triagem também podem falhar em relatar a presença da doença, isso é chamado de falso negativo - (FN) e podem levar as pessoas a ignorar os sintomas futuros (WHO, 2020).

Devido aos resultados FP e FN, procedimentos de triagem populacional devem ser cuidadosamente examinados antes de serem implantados, pois um programa de rastreamento contendo muitos casos de FP sobrecarregaria o sistema de saúde, uma vez que todos os casos VP e FP precisariam ser mais bem investigados. Ademais, quanto a um programa contendo elevados números de FN, além de não notificar a doença, ainda contribui para descrédito do programa.

Considerando o conhecimento sobre as características dos possíveis desfechos (FP e FN) apresentados por um programa de triagem populacional, analisou-se a possibilidade de implementação desse método diagnóstico no contexto de algoritmos de *deep learning* para auxiliar no combate ao glaucoma. Esta análise é importante e interessante principalmente para países como o Brasil, com a necessidade de oferecer serviços de saúde pública a uma gigante população distribuída em grandes extensões territoriais.

Embora os programas de triagens possam ser admirados como soluções para o diagnóstico precoce de uma diversidade de doenças, estes mecanismos não são fáceis de implementar, exigindo viabilidade econômica e análise rigorosa dos reais benefícios/malefícios que podem ser alcançados. Portanto, a OMS, propõe 10 princípios para avaliar se a triagem é um curso de ação apropriado para melhorar a saúde pública em independentemente da doença evidenciada:

1. A condição deve ser um problema de saúde importante;
2. Deve haver um tratamento aceito para pacientes com reconhecida doença;
3. Devem estar disponíveis instalações para diagnóstico e tratamento;
4. Deve haver uma fase sintomática latente ou precoce reconhecível;
5. Deve haver um teste ou exame adequado;
6. O teste deve ser aceito pela população;
7. A história natural da condição, incluindo o desenvolvimento de doença latente para doença declarada, deve ser adequadamente compreendida;
8. Deve haver uma política acordada sobre quem tratar como pacientes;

- 
9. O custo da detecção de casos (incluindo diagnóstico e tratamento de pacientes diagnosticados) deve ser economicamente equilibrado em relação a possíveis gastos com assistência médica na totalidade;
  10. A busca de casos deve ser um processo contínuo e não um projeto “de uma vez por todas” (WHO, 2020).

Dado o contexto, faz-se uma análise das estatísticas globais de subdiagnóstico do glaucoma e demonstra-se que: ainda que existam evidências de que os programas de triagem reduzirão as taxas de subdiagnóstico do glaucoma, isso por si só não é suficiente para introduzir tais serviços. Isso porque, quando se discute implantação de triagens populacionais, é necessário realizar uma análise abrangente dos prós e contras, considerando não apenas o método diagnóstico, mas também o valor dos exames de seguimento e tratamentos que podem ser indicados. Entretanto, requerer avaliações mais consistentes e realistas, como ensaios clínicos, que possam durante a testagem identificar e, se necessário, corrigir eventuais impasses relacionados à metodologia a ser empregada. Diante disso, a seção a seguir simula uma aplicação prática usando o algoritmo desenvolvido neste trabalho para melhor explicar essa discussão.

## 6.2 Análise do algoritmo desenvolvido

Para uma breve análise do algoritmo desenvolvido foram analisados alguns fatores que também podem influenciar a decisão de usar ou não tal algoritmo na prática clínica para triagem populacional dessa doença. Dentre esses fatores, destaca-se o valor preditivo alcançado pelo algoritmo e a prevalência da doença na população. Neste exemplo, considerou-se a acurácia do algoritmo desenvolvido neste trabalho, porém, esta é uma análise que pode ser amplamente aplicada a algoritmos em geral. Assim, diante da análise, os valores preditivos podem ser classificados em: valor Preditivo Positivo (VPP) e Valor Preditivo Negativo (VPN). São métricas de avaliação que depende essencialmente do resultado de três cálculos:

1. Cálculo de sensibilidade (SE) do algoritmo;
2. Cálculo de Especificidade (ES) do algoritmo;
3. Prevalência (P) da doença na população.

O valor de VPP pode ser definido pela Equação 6.1. Esta métrica representa a probabilidade de um indivíduo avaliado com resultado positivo ser realmente doente;

$$VPP = \frac{(SE \times P)}{(SE \times P) + (1 - ES) \times (1 - P)} \quad (6.1)$$

---

O valor de VPN pode ser definido pela Equação 6.2. Esta métrica representa a probabilidade de um indivíduo avaliado com resultado negativo ser realmente normal.

$$VPN = \frac{ES \times (1 - P)}{ES \times (1 - P) + (1 - SE) \times P} \quad (6.2)$$

Desta forma, quanto mais sensível for um teste, maior será o seu VPN e, (maior será a segurança em dizer que a pessoa com teste negativo não tem a doença). Quanto mais específico for um teste, maior será o seu VPP e, (maior será a segurança em dizer que a pessoa com teste positivo tem a doença).

Para demonstrar o quando os valores preditivos podem ser importantes avaliadores, a seguir será apresentado um exemplo de como essa métrica pode impactar positiva ou negativamente um programa de triagem populacional. Para realização deste exemplo será considerada a aplicabilidade do algoritmo desenvolvido neste trabalho em uma possível prática clínica, mostrando o quão é difícil analisar algoritmos de IA focando apenas nos seus dados de acurácia, sem considerar dados relevantes como prevalência da doença na população estudada.

Diante do exemplo, considere os resultados do algoritmo desenvolvido neste trabalho, onde a acurácia geral foi de 95,4%, a SE e a ES do teste de glaucoma foram de 98,5% e 92,2%, respectivamente. Em seguida, extrapola-se uma aplicabilidade do algoritmo desenvolvido em uma população real, assumindo que a prevalência P do glaucoma na população analisada seja de 3,5%, valor aceitável segundo estatísticas médias. Antes de iniciar os cálculos estatísticos, é interessante observar a distribuição das classes "glaucoma" e "normal" no conjunto de dados BrG utilizado neste trabalho. Essa observação permite perceber que a distribuição dessas classes na base de dados é uniforme, ou seja, 50% glaucoma e 50% normal. No entanto, essa uniformidade não existe na população real, onde a maioria das pessoas é normal e apenas 3,5% têm glaucoma, como representado pela Figura 6.1.

O exemplo de distribuição de classes equivale a dizer que a prevalência do glaucoma é de 50% no conjunto de dados BrG e 3,5% na população real. Desta forma, a partir desses dados foram analisados os efeitos da prevalência sobre uma possível triagem populacional do glaucoma, considerando os resultados do algoritmo desenvolvido. Para isso foram utilizadas duas populações hipotéticas:

- a população (A) com prevalência de 50% representando os exemplos de glaucoma base de dados BrG e,
- a população (B) com prevalência de 3,5%; representando uma população real, ambas contendo 1.000 pessoas para facilitar a visualização de resultados.

Por conveniência, algumas das variáveis importantes para essa extrapolação de resultados são descritas aqui, e são elas:

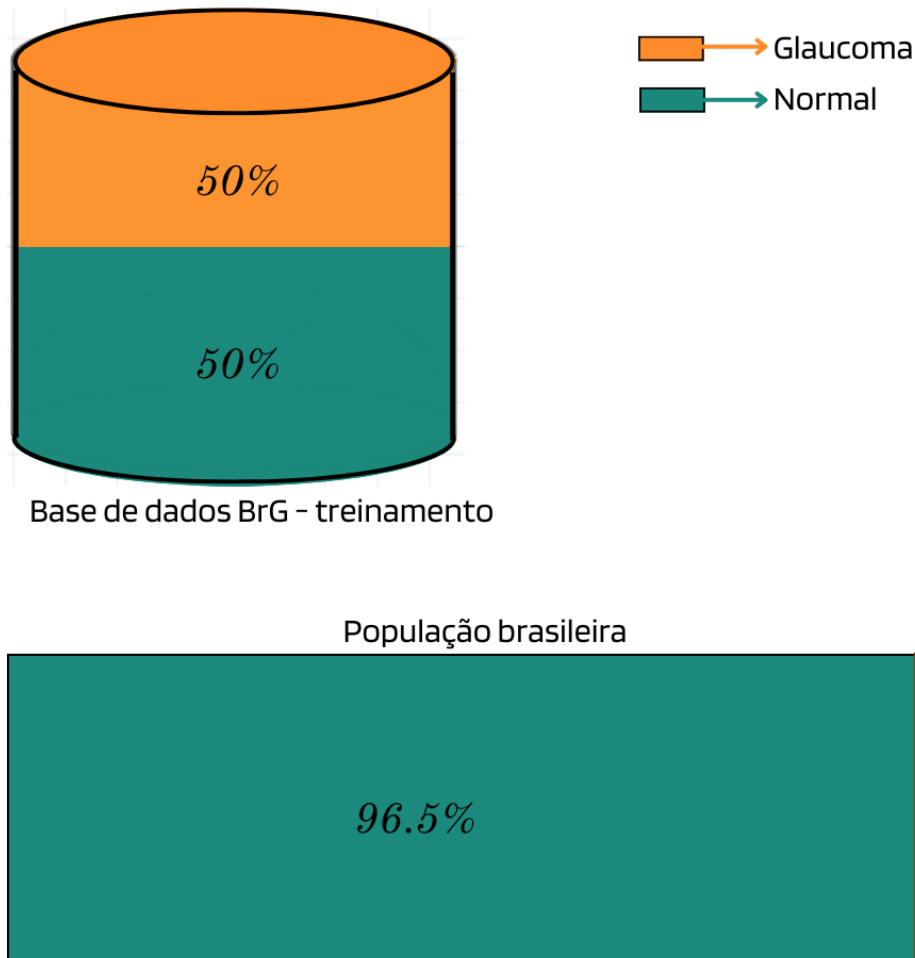


Figura 6.1: Exemplo da distribuição de casos de glaucoma na base BrG e população real.

- (*SE*) - Sensibilidade do teste = 98,5%;
- (*ES*) - Especificidade do teste = 92,2%;
- (*P*) - Prevalência da doença na população = 3,5%;
- (*PBrG*) - Prevalência da doença na base de dados BrG = 50%;
- (*TGA*) – Total de glaucomatosos na base de dados BrG = 500;
- (*TNA*) – Total de pessoas sem glaucoma (normais) na base de dados BrG = 500;

### 6.2.1 Cálculo sobre a população (A)

Considerando as métricas *SE* de 98,5% e a *ES* de 92,2%, ambas resultantes do algoritmo desenvolvido neste trabalho, calcula-se a quantidade de resultados *VP* foram gerados para essa população. Para isso, aplicando-se 98,5 de *SE* sobre a população (A), pode-se calcular pela Equação 6.3, que em um universo de 500 pessoas com glaucoma (número total de

glaucomatosos na população (A)), o algoritmo será capaz de identificar 493 casos como *VP* e 7 casos como *FN*.

$$VP = \frac{(TGA \times SE)}{100} \quad (6.3)$$

Logo, aplicando uma *ES* de 92,2% na mesma população (A), pode ser calculado que em um universo de 500 pessoas sem glaucoma (total de pacientes sem glaucoma na população (A)), o algoritmo será capaz de identificar pela Equação 6.4, 461 pessoas como *VN* e 39 pessoas como *FP*.

$$VN = \frac{(TNA \times ES)}{100} \quad (6.4)$$

Após os cálculos, todos os resultados sobre a população (A) são representados na matriz de confusão da Figura 6.2.

	GLaucoma	Normal	
Positivo	493 <small>VP</small>	7 <small>FN</small>	500
Negativo	39 <small>FP</small>	461 <small>VN</small>	500

Figura 6.2: Matriz de confusão exemplo 1 - valores absolutos.

### 6.2.2 Cálculo sobre a população (B)

Após aplicar o teste à população (A), foi aplicado o mesmo teste à população (B), mas antes, é importante aplicar a prevalência do glaucoma de 3,5% a esta população. Portanto, após considerar a prevalência *P* de 3,5%, foi calculado que 35 pessoas na população (B) têm glaucoma e 965 pessoas não.

Para melhor entendimento, algumas variáveis são criadas expostas aqui:

- (*TB*) = Número total de pessoas na população (B);
- (*TNB*) = Total de pessoas sem glaucoma (normais) na população (B);
- (*TGB*) = Total de pessoas com glaucoma na população (B), calculado pela Equação 6.5 ;

$$TGB = \frac{(TB \times P)}{100} \quad (6.5)$$

Após estimar o valor de  $TGB$  considerando a prevalência de 3,5%, calculou-se com o algoritmo desenvolvido o número de  $VP$  sobre a população (B). Deste modo, o algoritmo com  $SE$  de 98,5% foi capaz de identificar 34 pessoas dessa população como  $VP$  e 1 pessoa como  $FN$  de um total de 35 pessoas com glaucoma. Da mesma forma, com uma  $ES$  de 92,2 %, em uma população com 965 pessoas sem glaucoma, o algoritmo foi capaz de identificar 890 como  $VN$  e 75 pessoas como  $FP$ , conforme mostrado na Figura 6.3.

$$VP = \frac{(TGB \times SE)}{100} \quad (6.6)$$

	GLaucoma	Normal	
Positivo	34 <small>VP</small>	1 <small>FN</small>	35
Negativo	75 <small>FP</small>	890 <small>VN</small>	965

Figura 6.3: Matriz de confusão exemplo 2 - valores absolutos.

Olhando apenas para as duas matrizes de confusão nas Figuras 6.2 e 6.3, pode-se observar um aumento proporcional no número de falsos-positivos na população (B). Esse aumento se deve ao efeito da aplicação do teste na população (B) com baixa prevalência do glaucoma em comparação à prevalência da doença na populacional (A).

Esses valores podem ser mais bem observados calculando valores preditos, então, a partir dos dados obtidos na matriz de confusão da figura 6.3, é possível encontrar o valor de  $VPP$ , que dará a probabilidade da doença estar presente quando o teste for positivo para as pessoas da população (B), considerada uma população com características próximas da realidade. Deste modo, como já se considerou o valor preditivo da doença em estudo, o valor de  $VPP$  pode ser calculado pela equivalente Equação 6.7. Assim, foi possível encontrar o valor de 0,31%. Esse valor indica a probabilidade pós-teste, ou seja, o resultado quando positivo terá 31% de chance de ser realmente positivo, percentual que pode ser considerado baixo para uma aplicação em escala populacional.

$$VPP = \frac{VP}{VP + FP} \quad (6.7)$$

Da mesma forma, um valor de  $VPN$  de 0,99% é calculado a partir da equação 6.8,

---

mostrando a probabilidade de que o teste seja realmente negativo quando o resultado for VN.

$$VPN = \frac{VN}{VN + FN} \quad (6.8)$$

Esse exemplo apresentou uma extrapolação do resultado real obtido pelo algoritmo de *deep learning* desenvolvido para classificação utilizando-se da análise de imagens digitais do fundo de olho, pelo qual o valor de *VPP* obtido, faz o algoritmo aplicado na população hipotética apresentar números elevados de pacientes com diagnóstico *FP*, o que poderá sobrecarregar os sistemas de saúde. Todavia, esses dados estatísticos não devem ser considerados para descartar a possibilidade de uma triagem populacional com uso de IA, mas sim para reforçar a necessidade de ensaios clínicos para entender verdadeiramente os cenários de aplicação prática com uso de *deep learning*.

Dado o contexto, considerando a possível aplicação do algoritmo desenvolvido neste trabalho para triagem populacional, os resultados indicaram que é necessário aumentar o valor do *VPP* para uma população real. Assim, existem pelo menos duas formas de conseguir este aumento, designadamente: aplicando melhorias no código do algoritmo de classificação e direcionando a sua aplicabilidade para populações com prevalência maior do glaucoma.

Uma população com prevalência aumentada pode ser conseguida selecionando apenas os grupos de mais alto risco de desenvolver a doença. Para isso, há uma série de fatores, chamados de fatores de risco, sendo mais frequentemente associados ao glaucoma primário de ângulo aberto e, são usados como indicadores para um monitoramento mais efetivo do glaucoma (Lee et al., 2021), (Society, 2017), (Thakur, 2023). Diante disso, mesmo com uma possível triagem de glaucoma por meio de algoritmos de IA, é necessário começar pela seleção da população a ser rastreada. Essa opção deve ser considerada para aumentar a prevalência da doença no grupo filtrado para dar viabilidade de implantação de um possível processo de triagem populacional.

Conforme o conteúdo proposto, os dados apresentados juntamente com os cálculos dos valores preditivos demonstram que as aplicações de triagem populacional podem exigir um olhar muito além da simples interpretação de uma alta acurácia alcançada por algoritmos treinados e testados em conjuntos de dados controlados. No entanto, os algoritmos de IA, incluindo o algoritmo desenvolvido neste trabalho, tem potencial para distinguir imagens de fundo de olho glaucomatosas de normais com alta acurácia, mesmo usando imagens de baixa resolução obtidas de aparelhos de baixo custo. Isso mostra claramente que um dos caminhos para viabilizar a triagem populacional do glaucoma está praticamente concretizado e com potencial para auxiliar no combate ao subdiagnóstico da doença. Todavia, ainda são necessários ensaios clínicos com análises de outras variáveis para descobrir onde e como melhor aplicar a tecnologia dos algoritmos de *deep learning*.

# Capítulo 7

## Conclusões e trabalhos futuros

Neste capítulo são apresentadas as conclusões e as recomendações para a continuidade dos trabalhos nesta área de estudo.

### 7.1 Conclusões

Primeiro, nesta tese, uma nova base de dados chamada BrG foi construída no Brasil, contendo 2.000 fotografias de fundo de olho. Todas as imagens são rotuladas para glaucoma, e há uma distinção clara entre glaucomatosas de não glaucomatosas, o que é útil em estudos de classificação da referida doença ocular com uso de *deep learning*. Deste modo, inicialmente, a base de dados desenvolvida é comparada com outros conjuntos de dados similares e as suas principais características são analisadas. Portanto, observou-se que, enquanto as demais bases de dados obtiveram fotografias com auxílio de câmeras convencionais utilizadas na oftalmologia, o conjunto BrG utilizou apenas um smartphone acoplado ao dispositivo portátil (oftalmoscópio *panoptic*) para fotografar a região da retina.

Além do menor custo de aquisição das imagens BrG em comparação com imagens adquiridas por equipamentos oftalmológicos convencionais, elas também foram capturadas por profissionais não médicos, aumentando o seu acesso como alternativa em possíveis situações de triagem populacional do glaucoma. Outro ponto relevante é que as imagens BrG têm um campo de visão de no máximo  $25^\circ$ , focando apenas a área do disco óptico e proximidades. Em virtude do menor ângulo, menos esforço computacional é necessário para extrair regiões de interesse para treinar algoritmos de *deep learning* para classificar doenças degenerativas do nervo óptico.

O fato das imagens BrG não registrarem todo o fundo de olho, pode ser uma desvantagem em alguns casos onde este tipo de imagem é necessário; no entanto, considerando o glaucoma, essa particularidade não é necessariamente um problema, uma vez que, a área do disco óptico representou o conteúdo mais importante no diagnóstico da referida

---

patologia. À vista disso, todos os trabalhos aqui relacionados, utilizaram os recursos observáveis nesta região, durante o treinamento e testes de algoritmos para a classificação desta neuropatia óptica. Além disso, autores como (Fu et al., 2018) compararam a precisão do seu algoritmo sobre imagens globais e imagens segmentadas na região do disco óptico e, em todos os casos, a melhor acurácia foi obtida utilizando apenas a área do disco óptico, reforçando que as imagens BrG podem ser úteis para o diagnóstico de doenças lesivas a esta área da retina, como o glaucoma.

No que diz respeito à resolução das imagens BrG, esse fator não se apresentou prejudicial à tarefa de classificação, todavia, são necessários mais testes, principalmente voltados à segmentação das estruturas do disco óptico, por se tratar de um procedimento que responde melhor quando aplicado às imagens mais nítidas e menos ruidosas. Quanto à etnia da base de dados BrG, é difícil definir, pois, a composição racial e étnica da sociedade brasileira é resultado da fusão de raças de diferentes origens.

Considerando o número de imagens rotuladas para glaucoma, o novo conjunto BrG supera as bases de dados publicamente disponíveis. Por se tratar de uma base de dados pública, sua contribuição é significativa, proporcionando acesso a informações e auxiliando pesquisadores na condução de novos estudos relacionados à classificação do glaucoma e segmentação da estrutura do disco óptico. Viabilizando novas perspectivas sobre essa categoria de imagens fundoscópicas.

Ao analisar às limitações, percebe-se que o conjunto de dados BrG foi composto inteiramente por uma única câmera (smartphone), enquanto outros, por exemplo, o REFUGE e RIM-ONE são compostos por múltiplas câmeras, o que fornece maior generalidade para o treinamento de algoritmos *deep learning*. Outro ponto é que as imagens BrG não são rotuladas para outros tipos de doenças oculares, nem para o grau de dano glaucomatoso. Todavia, para novas pesquisas científicas, a base de dados BrG está disponível no site <https://www.globaleyeh.com/brgdataset>, de onde pode ser obtido todo o conteúdo deste trabalho.

Sobre a classificação do glaucoma, o *ensemble* de CNNs utilizado obteve uma precisão de 95,0% no conjunto de dados BrG, mostrando primeiramente ser superior a todos os classificadores individuais que o compõe, além disso, mostrou-se consistente com os resultados obtidos por outras pesquisas relacionadas. Assim, em análise comparativa, infere-se que os resultados de (Diaz-Pinto et al., 2019b), que também trabalharam com diversos classificadores, demonstraram semelhança entre a acurácia final que obtiveram com imagens de alta resolução e a acurácia alcançada neste trabalho. Deste modo, os resultados aqui apresentados são coerentes com as expectativas de classificação da referida doença, mesmo a partir de imagens obtidas com smartphone.

Quanto aos desafios na redução do subdiagnóstico do glaucoma, destaca-se a baixa frequência de visitas preventivas regulares às clínicas oftalmológicas pela população em geral. Portanto, os algoritmos de inteligência artificial podem se tornar uma ferramenta de

---

baixo custo e fácil de usar, projetada para rastrear o maior número possível de pessoas e alertá-las sobre a necessidade de visitar um consultório oftalmológico quando necessário.

Quanto aos obstáculos enfrentados pela implementação prática de algoritmos *deep learning* na classificação dessa neuropatia ocular, algumas dificuldades foram notadas, dentre as quais a falta de informações objetivas e de critérios consistentes para o diagnóstico dessa doença, uma vez que a sua investigação depende de uma avaliação holística das informações do paciente por parte do oftalmologista, incluindo histórico médico, exames estruturais e funcionais e, em muitos casos, longitudinais (Lee et al., 2021). Outro ponto é que, sem biomarcadores diagnósticos confiáveis, o glaucoma sofre de interpretação subjetiva, o que claramente dificulta uma solução diagnóstica calculável e computável. Por outro lado, a ausência de biomarcadores robustos e computáveis apenas requer aplicações mais complexas, mas não pode limitar a eficácia dos algoritmos *deep learning*, que, ao longo dos anos, vem se tornando presentes em diversas tarefas de reconhecimento de padrões. Apesar de que muitas vezes esses padrões reconhecidos sejam obscuros para os humanos, ao menos discorrendo o glaucoma, eles existem e devem ser considerados e aceitáveis, tendo a sua validação demonstrada em exaustivos trabalhos correlatos, constatando a sua eficácia na classificação de imagens glaucomatosas.

Com base no que foi apresentado e observando questões da pesquisa proposta, conclui-se que as imagens de fundo de olho, obtidas com um smartphone combinado com um oftalmoscópio portátil *panoptic*, podem ser utilizadas por algoritmos de *deep learning* para classificação do glaucoma, dada a capacidade de reconhecer padrões da referida doença em imagens digitais, sendo que mesmo com baixa resolução, o algoritmo proposto atingiu 95,4% de precisão, com sensibilidade e especificidade de 98,5% e 92,2%, respectivamente, evidenciando que o referido algoritmo pode, sim, possibilitar o diagnóstico automatizado do glaucoma. Quanto às aplicações práticas em triagens populacionais, justifica-se um estudo mais abrangente e cuidadoso, reduzindo ao máximo a taxa de falso-positivos e analisando também todas as variáveis possíveis para obter respostas sobre os benefícios ou malefícios da implementação de um programa de triagem.

## **7.2 Recomendações para trabalhos futuros**

Como trabalho futuro, deve ser possível aumentar o conjunto de dados BrG longitudinalmente para que algoritmos de *deep learning* possam ser usados para avaliar danos estruturais glaucomatosos ao longo do tempo, uma ferramenta para entender melhor a evolução do glaucoma a partir da imagem do fundo de olho.

# Referências Bibliográficas

- Muhammad Aamir, Muhammad Irfan, Tariq Ali, Ghulam Ali, Ahmad Shaf, Ali Al-Beshri, Tariq Alasbali, and Mater H Mahnashi. An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification. *Diagnostics*, 10(8):602, 2020. [71](#), [77](#)
- AAO. *Optic Nerve*. American Academy of Ophthalmology,, 7th ed. edition, 2023. URL <https://www.aao.org/eye-health/anatomy/optic-nerve-3>. accessed: june. 2021. [x](#), [46](#)
- Qaisar Abbas. Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning. *International Journal of Advanced Computer Science and Applications*, 8(6), 2017. [65](#)
- Lamiaa Abdel-Hamid. Glaucoma detection from retinal images using statistical and textual wavelet features. *Journal of digital imaging*, 33(1):151–158, 2020. [70](#), [75](#), [77](#)
- Tinku Acharya and Ajoy K Ray. *Image processing: principles and applications*. John Wiley & Sons, 2005. [19](#)
- Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018. [12](#)
- Jin Mo Ahn, Sangsoo Kim, Kwang-Sung Ahn, Sung-Hoon Cho, Kwan Bok Lee, and Ungsoo Samuel Kim. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PloS one*, 13(11):e0207982, 2018. [72](#), [77](#)
- 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. [74](#), [77](#)
- Welch Allyn. Panoptic™ ophthalmoscope, 2014. accessed: april. 2021. [2](#), [4](#)
- Subhash C Bagui. *Combining pattern classifiers: methods and algorithms*, 2005. [128](#)
- Muhammad Naseer Bajwa, Gur Amrit Pal Singh, Wolfgang Neumeier, Muhammad Imran Malik, Andreas Dengel, and Sheraz Ahmed. G1020: A benchmark retinal fundus

- image dataset for computer-aided glaucoma detection. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2020. [66](#)
- Francisco José Fumero Batista, Tinguaro Diaz-Aleman, Jose Sigut, Silvia Alayon, Rafael Arnay, and Denisse Angel-Pereira. Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. *Image Analysis & Stereology*, 39(3):161–167, 2020. [65](#), [80](#)
- Ekaba Bisong and Ekaba Bisong. Google colabatory. *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, pages 59–64, 2019. [103](#)
- Clerimar Paulo Bragança, José Manuel Torres, Christophe Pinto de Almeida Soares, and Luciano Oliveira Macedo. Detection of glaucoma on fundus images using deep learning on a new image set obtained with a smartphone and handheld ophthalmoscope. In *Healthcare*. Multidisciplinary Digital Publishing Institute, 2022. [67](#), [74](#), [77](#)
- Attila Budai, Rüdiger Bock, Andreas Maier, Joachim Hornegger, and Georg Michelson. Robust vessel segmentation in fundus images. *International journal of biomedical imaging*, 2013, 2013. [x](#), [47](#), [51](#), [65](#)
- Jennifer Burr, Rodolfo Hernández, Craig Ramsay, Maria Prior, Susan Campbell, Augusto Azuara-Blanco, Marion Campbell, Jill Francis, and Luke Vale. Is it worthwhile to conduct a randomized controlled trial of glaucoma screening in the united kingdom? *Journal of health services research & policy*, 19(1):42–51, 2014. [1](#)
- José Camara, Alexandre Neto, Ivan Miguel Pires, María Vanessa Villasana, Eftim Zdravevski, and António Cunha. Literature review on artificial intelligence methods for glaucoma screening, segmentation, and classification. *Journal of Imaging*, 8(2):19, 2022. [80](#)
- Enrique J Carmona, Mariano Rincón, Julián García-Feijoó, and José M Martínez-de-la Casa. Identification of the optic nerve head with genetic algorithms. *Artificial intelligence in medicine*, 43(3):243–259, 2008. [64](#)
- Chuan-Wang Chang, Chuan-Yu Chang, Yuan-Yi Lin, Wei-Wen Su, and Henry Shen-Lih Chen. A glaucoma detection system based on generative adversarial network and incremental learning. *Applied Sciences*, 13(4):2195, 2023. [75](#), [77](#)
- Touhidul Islam Chayan, Anita Islam, Eftykhar Rahman, Md Tanzim Reza, Tasnim Sakib Apon, and MD Golam Rabiul Alam. Explainable ai based glaucoma detection using transfer learning and lime. In *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6. IEEE, 2022. [71](#), [77](#)

- Xiangyu Chen, Yanwu Xu, Shuicheng Yan, Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu. Automatic feature learning for glaucoma detection based on deep learning. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pages 669–677. Springer, 2015. [71](#), [77](#)
- François Chollet. Xception: Deep learning with depthwise separable convolutions. arxiv. *arXiv preprint arXiv:1610.02357*, 2017. [104](#), [121](#)
- François Chollet et al. Keras: The python deep learning library. *Astrophysics source code library*, pages ascl–1806, 2018. URL <https://keras.io/guides/>. accessed: august. 2021. [103](#)
- Mark Christopher, Akram Belghith, Christopher Bowd, James A Proudfoot, Michael H Goldbaum, Robert N Weinreb, Christopher A Girkin, Jeffrey M Liebmann, and Linda M Zangwill. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Scientific reports*, 8(1):1–13, 2018. [73](#), [77](#), [79](#)
- Maila Claro, Rodrigo Veras, Andre Santana, Flavio Araujo, Romuere Silva, Joao Almeida, and Daniel Leite. An hybrid feature space from texture information and transfer learning for glaucoma classification. *Journal of Visual Communication and Image Representation*, 64:102597, 2019. [76](#)
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20: 273–297, 1995. [29](#)
- Ministério da Saúde. Sistema Único de saúde - sus, Junho 2022. URL <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/sus>. accessed: july. 2021. [89](#)
- Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, et al. Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, 2014. [65](#)
- Andres Diaz-Pinto, Sandra Morales, Valery Naranjo, Thomas Köhler, Jose M Mossi, and Amparo Navea. Cnns for automatic glaucoma assessment using fundus images: an extensive validation. *Biomedical engineering online*, 18:1–19, 2019a. [64](#)
- Andres Diaz-Pinto, Sandra Morales, Valery Naranjo, Thomas Köhler, Jose M Mossi, and Amparo Navea. Cnns for automatic glaucoma assessment using fundus images: an extensive validation. *Biomedical engineering online*, 18:1–19, 2019b. [153](#)

- Andres Diaz-Pinto, Sandra Morales, Valery Naranjo, Thomas Köhler, Jose M Mossi, and Amparo Navea. Cnns for automatic glaucoma assessment using fundus images: an extensive validation. *Biomedical engineering online*, 18:1–19, 2019c. [73](#), [77](#)
- Marcos Vinicius dos Santos Ferreira, Antonio Oseas de Carvalho Filho, Alcilene Dalilia de Sousa, Aristofanes Correa Silva, and Marcelo Gattass. Convolutional neural network and texture descriptor-based automatic detection and diagnosis of glaucoma. *Expert Systems with Applications*, 110:250–263, 2018. [76](#), [77](#)
- Richard O Duda, Peter E Hart, et al. *Pattern classification*. John Wiley & Sons, 2006. [13](#)
- Richard E Woods and Rafael C Gonzalez. *Digital image processing*, 2008. [ix](#), [15](#), [17](#)
- Venkata Gopal Edupuganti, Akshay Chawla, and Amit Kale. 25th ieee international conference on image processing (icip). *IEEE Journal of Biomedical and Health Informatics*, pages 2227–2231, 2018. [74](#), [77](#)
- Clément Farabet, Cyril Poulet, Jefferson Y Han, and Yann LeCun. Cnp: An fpga-based processor for convolutional networks. In *2009 International Conference on Field Programmable Logic and Applications*, pages 32–37. IEEE, 2009. [34](#)
- Huazhu Fu, Jun Cheng, Yanwu Xu, Changqing Zhang, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE transactions on medical imaging*, 37(11):2493–2501, 2018. [67](#), [72](#), [77](#), [153](#)
- Francisco Fumero, Silvia Alayón, José L Sanchez, Jose Sigut, and M Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In *2011 24th international symposium on computer-based medical systems (CBMS)*, pages 1–6. IEEE, 2011. [66](#)
- Francisco Fumero, Jose Sigut, Silvia Alayón, Marta González-Hernández, and M González De La Rosa. Interactive tool and database for optic disc and cup segmentation of stereo and monocular retinal fundus images. *23rd International Conference in Central Europe on Computer Graphics*, 2015. [66](#)
- David F Garway-Heath. Early diagnosis in glaucoma. *Progress in brain research*, 173: 47–57, 2008. [49](#)
- Steven J Gedde, Kateki Vinod, Martha M Wright, Kelly W Muir, John T Lind, Philip P Chen, Tianjing Li, Steven L Mansberger, et al. Primary open-angle glaucoma preferred practice pattern®. *Ophthalmology*, 128(1):P71–P150, 2021. [1](#), [99](#)

- Abeba T Giorgis, Abiye M Alemu, Sourabh Arora, Girum W Gessesse, Fikru Melka, Alemayehu Woldeyes, Samreen Amin, Faazil Kassam, Ayaz K Kurji, and Karim F Damji. Results from the first teleglaucoma pilot project in addis ababa, ethiopia. *Journal of Glaucoma*, 28(8):701–707, 2019. [45](#)
- Michael JA Girard and Leopold Schmetterer. Artificial intelligence and deep learning in glaucoma: current state and future prospects. *Progress in Brain Research*, 257:37–64, 2020. [2](#)
- Juan J Gómez-Valverde, Alfonso Antón, Gianluca Fatti, Bart Liefers, Alejandra Herranz, Andrés Santos, Clara I Sánchez, and María J Ledesma-Carbayo. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. *Biomedical optics express*, 10(2):892–913, 2019. [73](#), [76](#)
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. [33](#), [34](#), [36](#), [110](#)
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. [40](#)
- James C Grande. Principles of image analysis - metallography, microstructure, and analysis, 2012. URL <https://doi.org/10.1007/s13632-012-0037-5>. [17](#)
- Ireneusz Grulkowski, Jonathan J Liu, Benjamin Potsaid, Vijaysekhar Jayaraman, Chen D Lu, James Jiang, Alex E Cable, Jay S Duker, and James G Fujimoto. Retinal, anterior segment and full eye imaging using ultrahigh speed swept source oct with vertical-cavity surface emitting lasers. *Biomedical optics express*, 3(11):2733–2751, 2012. [55](#)
- Neeraj Gupta, Hitendra Garg, and Rohit Agarwal. A robust framework for glaucoma detection using clahe and efficientnet. *The Visual Computer*, pages 1–14, 2021. [70](#), [77](#)
- Recep E Hacisoftoglu, Mahmut Karakaya, and Ahmed B Sallam. Deep learning frameworks for diabetic retinopathy detection with smartphone-based retinal imaging systems. *Pattern recognition letters*, 135:409–417, 2020. [2](#)
- Jae-Ho Han. Artificial intelligence in eye disease: Recent developments, applications, and surveys. *Diagnostics*, 12(8):1927, 2022. [61](#)
- Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, pages 610–621, 1973. [22](#)
- Simon Haykin. *Kalman filtering and neural networks*. John Wiley & Sons, 2004. [7](#), [109](#)

- Simon Haykin. *Neural networks and learning machines, 3/E*. Pearson Education India, 2009. 6
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016a. xi, 104, 114, 119
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016b. 104, 116
- World Health-Organization et al. World report on vision. *World health Organisation*, 2019. accessed: June. 2022. 2
- Jeff Heaton. Applications of deep neural networks with keras. *arXiv e-prints*, pages arXiv–2009, 2020. 4, 105
- Anders Heijl, Boel Bengtsson, and Sigridur Erla Oskarsdottir. Prevalence and severity of undetected manifest glaucoma: results from the early manifest glaucoma trial screening. *Ophthalmology*, 120(8):1541–1545, 2013. 1
- HO. Hospital de olhos do sul de minas gerais, Junho 2022. URL <https://hosuldeminas.com.br/>. accessed: january. 2021. 89, 94
- Zhe Hui Hoo, Jane Candlish, and Dawn Teare. What is an roc curve?, 2017. 42, 43
- 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. *arXiv preprint arXiv:1704.04861*, 2017. 104
- Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 104
- IBM. What is overfitting? *IBM*, page <https://www.ibm.com/topics/overfitting>, 2022a. accessed: abr. 2022. 126
- IBM. What is bagging? *IBM*, pages <https://www.ibm.com/topics/deep-learning>, 2022b. accessed: abr. 2022. 128
- IBM. What is deep learning? *IBM*, pages <https://www.ibm.com/topics/deep-learning>, 2022c. accessed: abr. 2022. 25

- Hiroshi Inoue. Data augmentation by pairing samples for images classification. *arXiv preprint arXiv:1801.02929*, 2018. [108](#)
- Supiksha Jain, Sanjeev Indora, and Dinesh Kumar Atal. Rider manta ray foraging optimization-based generative adversarial network and cnn feature for detecting glaucoma. *Biomedical Signal Processing and Control*, 73:103425, 2022. [75](#), [77](#)
- Ian T Jolliffe. *Principal component analysis for special types of data*. Springer, 2002. [24](#)
- Shilpa Sameer Kanse and Dinkar Manik Yadav. Retinal fundus image for glaucoma detection: A review and study. *Journal of Intelligent Systems*, 28(1):43–56, 2019. [44](#)
- Jack J Kanski and Brad Bowling. *Clinical ophthalmology: a systematic approach*. Elsevier Health Sciences, 2011. [44](#)
- Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2017. [61](#)
- Tehmina Khalil, Muhammad Usman Akram, Samina Khalid, and Amina Jameel. Improved automated detection of glaucoma from fundus image using hybrid structural and textural features. *IET Image Processing*, 11(9):693–700, 2017. [51](#)
- Jongwoo Kim, Loc Tran, Tunde Peto, and Emily Y Chew. Identifying those at risk of glaucoma: A deep learning approach for optic disc and cup segmentation and their boundary analysis. *Diagnostics*, 12(5):1063, 2022. [45](#)
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [113](#)
- Oleksandr Kovalyk, Juan Morales-Sánchez, Rafael Verdú-Monedero, Inmaculada Sellés-Navarro, Ana Palazón-Cabanes, and José-Luis Sancho-Gómez. Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data*, 9(1):291, 2022. [65](#)
- Oliver Kramer. *Dimensionality reduction with unsupervised nearest neighbors*, volume 51. Springer, 2013. [31](#)
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017a. [34](#)
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017b. [11](#)

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

6

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015. 32

Eric Boya Lee, Sophia Ying Wang, and Robert T Chang. Interpreting deep learning studies in glaucoma: unresolved challenges. *The Asia-Pacific Journal of Ophthalmology*, 10(3):261–267, 2021. 78, 79, 151, 154

Joonseok Lee, J Lee, Hyoseob Song, and Chihoon Lee. Development of an end-to-end deep learning system for glaucoma screening using color fundus images. *JAMA Ophthalmol*, 137:1353–1360, 2019. 74, 77

Fei Li, Zhe Wang, Guoxiang Qu, Diping Song, Ye Yuan, Yang Xu, Kai Gao, Guangwei Luo, Zegu Xiao, Dennis SC Lam, et al. Automatic differentiation of glaucoma visual field from non-glaucoma visual field using deep convolutional neural network. *BMC medical imaging*, 18:1–7, 2018a. 49

Liu Li, Mai Xu, Hanruo Liu, Yang Li, Xiaofei Wang, Lai Jiang, Zulin Wang, Xiang Fan, and Ningli Wang. A large-scale database and a cnn model for attention-based glaucoma detection. *IEEE transactions on medical imaging*, 39(2):413–424, 2019a. 71

Liu Li, Mai Xu, Hanruo Liu, Yang Li, Xiaofei Wang, Lai Jiang, Zulin Wang, Xiang Fan, and Ningli Wang. A large-scale database and a cnn model for attention-based glaucoma detection. *IEEE transactions on medical imaging*, 39(2):413–424, 2019b. 77

Liu Li, Mai Xu, Xiaofei Wang, Lai Jiang, and Hanruo Liu. Attention based glaucoma detection: a large-scale database and cnn model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10571–10580, 2019c. 71, 73, 77

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, 2018b. 72, 77, 78

Li Liu, Paul Fieguth, Yulan Guo, Xiaogang Wang, and Matti Pietikäinen. Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, 62:135–160, 2017. 39

Sidong Liu, Stuart L Graham, Angela Schulz, Michael Kalloniatis, Barbara Zangerl, Weidong Cai, Yang Gao, Brian Chua, Hemamalini Arvind, John Grigg, et al. A deep

- learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs. *Ophthalmology Glaucoma*, 1(1):15–22, 2018. [72](#), [76](#)
- Robi N Maamari, Jeremy D Keenan, Daniel A Fletcher, and Todd P Margolis. A mobile phone-based retinal camera for portable wide field imaging. *British Journal of Ophthalmology*, 98(4):438–441, 2014. [83](#)
- Carol M Mangione, Michael J Barry, Wanda K Nicholson, Michael Cabana, David Chel-mow, Tumaini Rucker Coker, Esa M Davis, Katrina E Donahue, John W Epling, Carlos Roberto Jaén, et al. Screening for primary open-angle glaucoma: Us preventive services task force recommendation statement. *JAMA*, 327(20):1992–1997, 2022. [1](#)
- José Martins, Jaime S Cardoso, and Filipe Soares. Offline computer-aided diagnosis for glaucoma detection using fundus images targeted at mobile devices. *Computer Methods and Programs in Biomedicine*, 192:105341, 2020. [73](#), [76](#), [79](#)
- John McCarthy, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4):12–12, 2006. [6](#)
- Delaram Mirzania, Atalie C Thompson, and Kelly W Muir. Applications of deep learning in detection of glaucoma: a systematic review. *European Journal of Ophthalmology*, 31(4):1618–1642, 2021. [2](#)
- A Murthi and M Madheswaran. Enhancement of optic cup to disc ratio detection in glaucoma diagnosis. In *2012 International Conference on Computer Communication and Informatics*, pages 1–5. IEEE, 2012. [50](#)
- Wamidh K Mutlag, Shaker K Ali, Zahoor M Aydam, and Bahaa H Taher. Feature extraction methods: a review. In *Journal of Physics: Conference Series*, page 012028. IOP Publishing, 2020. [20](#)
- PS Nandhini, P Srinath, and P Veeramanikandan. Detection of glaucoma using convolutional neural network (cnn) with super resolution generative adversarial network (sr-gan). In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1034–1040. IEEE, 2022. [74](#), [76](#)
- Jyoti Narwade and Binod Kumar. Local and global color histogram feature for color content-based image retrieval system. In *Proceedings of the International Congress on Information and Communication Technology: ICICT 2015, Volume 1*, pages 293–300. Springer, 2016. [18](#)
- Mohammad Norouzifard, Ali Nemati, Hamid GholamHosseini, Reinhard Klette, Kouros Nouri-Mahdavi, and Siamak Yousefi. Automated glaucoma diagnosis using deep and

- transfer learning: Proposal of a system for clinical testing. *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 1–6, 2018. 77
- Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018. 9, 10
- José Ignacio Orlando, Huazhu Fu, João Barbosa Breda, Karel Van Keer, Deepti R Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Jeyoung Kim, JoonHo Lee, et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical image analysis*, 59:101570, 2020. 65
- Tetsuro Oshika, Fumiaki Yoshitomi, and Kohtaro Oki. The pachymeter guide: a new device to facilitate accurate corneal thickness measurement. *Japanese journal of ophthalmology*, 41(6):426–427, 1997. 48
- Marcos Pereira Ottaiano, José Augusto Alves Ávila and Alexandre Chater Umbelino, Cristiano Caixeta Taleb. As condições de saúde ocular no brasil. *Conselho Brasileiro de Oftalmologia Entidade*, 2019. URL [http://www.cbo.com.br/novo/publicacoes/condicoes\\_saude\\_ocular\\_brasil2019.pdf](http://www.cbo.com.br/novo/publicacoes/condicoes_saude_ocular_brasil2019.pdf). accessed: august. 2021. 60
- Samiksha Pachade, Prasanna Porwal, Dhanshree Thulkar, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Luca Giancardo, Gwenolé Quéllec, and Fabrice Mériaudeau. Retinal fundus multi-disease image dataset (rfmid): A dataset for multi-disease detection research. *Data*, 6(2):14, 2021. 49, 55
- Aseem Patil and Milind Rane. Convolutional neural networks: an overview and its applications in pattern recognition. *Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2020, Volume 1*, pages 21–30, 2021. xi, 11, 33, 105, 107
- Todd W Perkins. Clinical decisions in glaucoma. *Archives of Ophthalmology*, 112(12): 1518–1519, 1994. 91
- Sonia Phene, R Carter Dunn, Naama Hammel, Yun Liu, Jonathan Krause, Naho Kitade, Mike Schaekermann, Rory Sayres, Derek J Wu, Ashish Bora, et al. Deep learning and glaucoma specialists: the relative importance of optic disc features to predict glaucoma referral in fundus photographs. *Ophthalmology*, 126(12):1627–1639, 2019. 73, 77, 78
- Moacir Antonelli Ponti and Gabriel B Paranhos Da Costa. Como funciona o deep learning. *arXiv preprint arXiv:1806.07908*, 2018. 33

- Harry A Quigley. Number of people with glaucoma worldwide. *British journal of ophthalmology*, 80(5):389–393, 1996. [1](#)
- Harry A Quigley and Aimee T Broman. The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*, 90(3):262–267, 2006. [60](#)
- U Raghavendra, Hamido Fujita, Sulatha V Bhandary, Anjan Gudigar, Jen Hong Tan, and U Rajendra Acharya. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Information Sciences*, 441:41–49, 2018. [71](#), [77](#)
- R Geetha Ramani and J Jeslin Shanthamalar. Improved image processing techniques for optic disc segmentation in retinal fundus images. *Biomedical Signal Processing and Control*, 58:101832, 2020. [65](#)
- An Ran Ran, Clement C. Tham, Poemen C. Chan, Ching Yu Cheng, Yih Chung Tham, Tyler Hyungtaek Rim, and Carol Y. Cheung. Deep learning in glaucoma with optical coherence tomography: a review. *Eye (Basingstoke)*, 2020. URL <http://dx.doi.org/10.1038/s41433-020-01191-5>. [6](#)
- Tinku Remo Susanna Jr and Alberto Diniz Filho. *Simplificando o diagnóstico e tratamento do glaucoma*. Cultura Médica, 2019. x, [53](#), [54](#)
- Irina Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, pages 41–46, 2001. [30](#)
- Victor Francisco Rodriguez-Galiano, Bardan Ghimire, John Rogan, Mario Chica-Olmo, and Juan Pedro Rigol-Sanchez. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, 67:93–104, 2012. [29](#)
- Mutegeki Ronald, Alwin Poulouse, and Dong Seog Han. isplinception: An inception-resnet deep learning architecture for human activity recognition. *IEEE Access*, 9: 68985–69001, 2021. [119](#)
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. [xii](#), [134](#), [135](#), [138](#)
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115: 211–252, 2015. [4](#), [39](#), [105](#)

- Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010. ix, x, xii, 6, 7, 8, 9, 36, 126
- Andrea Russo, Francesco Morescalchi, Ciro Costagliola, Luisa Delcassi, and Francesco Semeraro. A novel device to exploit the smartphone camera for fundus photography. *Journal of ophthalmology*, 2015, 2015. 2, 83
- Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. xii, 121
- SBC. Manual De Exame Em Glaucoma. *Sociedade Brasileira de Glaucoma*, 2015. URL [https://www.sbglaucoma.org.br/wp-content/uploads/2015/10/Manual\\_de\\_Exames\\_Glaucoma\\_SBG.pdf](https://www.sbglaucoma.org.br/wp-content/uploads/2015/10/Manual_de_Exames_Glaucoma_SBG.pdf). accessed: may. 2020. 48
- Mário Scheffer et al. Demografia médica no brasil 2020, 2020. URL [https://www.fm.usp.br/fmusp/conteudo/estudo\\_demografico\\_FMUSP\\_AMB.pdf](https://www.fm.usp.br/fmusp/conteudo/estudo_demografico_FMUSP_AMB.pdf). accessed: may. 2021. x, 56, 58, 59
- 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. 61
- Alexander K Schuster, Carl Erb, Esther M Hoffmann, Thomas Dietlein, and Norbert Pfeiffer. The diagnosis and treatment of glaucoma. *Deutsches Ärzteblatt International*, 117(13):225, 2020. 50, 55
- Sourya Sengupta, Amitojdeep Singh, Henry A Leopold, Tanmay Gulati, and Vasudevan Lakshminarayanan. Ophthalmic diagnosis using deep learning with fundus images—a critical review. *Artificial Intelligence in Medicine*, 102:101758, 2020. 77
- Ali Serener and Sertan Serte. Transfer learning for early and advanced glaucoma detection with convolutional neural networks. *2019 Medical technologies congress (TIPTÉ-KNO)*, pages 1–4, 2019. 73
- Artem Sevastopolsky, Stepan Drapak, Konstantin Kiselev, Blake M Snyder, Jeremy D Keenan, and Anastasia Georgievskaya. Stack-u-net: refinement network for image segmentation on the example of optic disc and cup. *arXiv preprint arXiv:1804.11294*, 2018. 70, 76
- Sharath M Shankaranarayana, Keerthi Ram, Kaushik Mitra, and Mohanasankar Sivaprakasam. Fully convolutional networks for monocular retinal depth estimation

- and optic disc-cup segmentation. *IEEE journal of biomedical and health informatics*, 23(4):1417–1426, 2019. [74](#), [77](#)
- Naoto Shibata, Masaki Tanito, Keita Mitsuhashi, Yuri Fujino, Masato Matsuura, Hiroshi Murata, and Ryo Asaoka. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Scientific reports*, 8(1):14665, 2018. [77](#)
- Rutuja Shinde. Glaucoma detection in retinal fundus images using u-net and supervised machine learning algorithms. *Intelligence-Based Medicine*, 5:100038, 2021. [75](#), [76](#)
- J Shiny Christobel, D Vimala, J Joshan Athanesious, S Christopher Ezhil Singh, Sivaraj Murugan, et al. Effectiveness of feature extraction by pca-based detection and naive bayes classifier for glaucoma images. *International Journal of Digital Multimedia Broadcasting*, 2022, 2022. [70](#), [77](#)
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. [116](#)
- Law Kumar Singh, Hitendra Garg, Munish Khanna, and Robin Singh Bhadoria. An enhanced deep image model for glaucoma diagnosis using feature-based detection in retinal fundus. *Medical & Biological Engineering & Computing*, 59:333–353, 2021. [70](#), [77](#)
- Vivek Kumar Singh, Hatem A Rashwan, Farhan Akram, Nidhi Pandey, Md Mostafa Kamal Sarker, Adel Saleh, Saddam Abdulwahab, Najlaa Maarooof, Jordina Torrents-Barrena, Santiago Romani, et al. Retinal optic disc segmentation using conditional generative adversarial network. *Frontiers in Artificial Intelligence and Applications*, 308, 2018. [74](#), [77](#)
- Jayanthi Sivaswamy, SR Krishnadas, Gopal Datt Joshi, Madhulika Jain, and A Ujjwaft Syed Tabish. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, pages 53–56. IEEE, 2014. [64](#)
- Austen M Smith and Craig N Czyz. Neuroanatomy, cranial nerve 2 (optic). In *StatPearls [Internet]*. StatPearls Publishing, 2022. [46](#)
- European Glaucoma Society. European glaucoma society terminology and guidelines for glaucoma, -chapter 3: Treatment principles and options. supported by the egs foundation: Part 1: Foreword; introduction; glossary; chapter 3 treatment principles and options. *Br J Ophthalmol*, 101(6):130–195, 2017. [151](#)

- Syna Sreng, Noppadol Maneerat, Kazuhiko Hamamoto, and Khin Yadanar Win. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Applied Sciences*, 10(14):4916, 2020. [75](#), [77](#), [79](#)
- Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. [108](#)
- Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004. [ix](#), [19](#), [64](#)
- RS Sutton and AG Barto. Reinforcement learning: An introduction. massachusetts london. *England: The MIT Press Cambridge*, 1998. [26](#)
- Amelia Swift, Roberta Heale, and Alison Twycross. What are sensitivity and specificity? *Evidence-Based Nursing*, 23(1):2–4, 2020. [40](#)
- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016a. [xii](#), [104](#), [124](#), [125](#)
- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016b. [xi](#), [117](#)
- Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017a. [117](#)
- Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, 2017b. [104](#), [119](#)
- Nicholas YQ Tan, David S Friedman, Ingeborg Stalmans, Iqbal Ike K Ahmed, and Chelvin CA Sng. Glaucoma screening: where are we and where do we need to go? *Current opinion in ophthalmology*, 31(2):91–100, 2020. [1](#), [2](#)
- Andrew J Tatham, Robert N Weinreb, and Felipe A Medeiros. Strategies for improving early detection of glaucoma: the combined structure–function index. *Clinical ophthalmology*, pages 611–621, 2014. [54](#), [55](#)

- Sanjay KD Thakur. Comments on: A deep dive into the latest european glaucoma society and asia pacific glaucoma society guidelines and their relevance to india. *Indian Journal of Ophthalmology*, 71(1):314, 2023. 151
- S Theodoridis and K Koutroumbas. Pattern recognition—fourth edition, 2009, 2008. 27
- G Dietterich Thomas. Machine learning research: Four current directions. *Artificial Intelligence, Magazine*, 18(4):97–136, 1997. 127
- Daniel Shu Wei Ting, Carol Yim-Lui Cheung, Gilbert Lim, Gavin Siew Wei Tan, Nguyen D Quang, Alfred Gan, Haslina Hamzah, Renata Garcia-Franco, Ian Yew San Yeo, Shu Yen Lee, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *Jama*, 318(22):2211–2223, 2017. 72, 77, 78
- F Topouzis. Glaucoma—the importance of early detection and early treatment. *Journal-Glaucoma—The Importance of Early Detection and Early Treatment*, 2007. 55
- Neto José Umbelino, Cristiano Caixeta. Conselho brasileiro de oftalmologia - censo 2021. *Departamento de Oftalmologia da Associação Médica Brasileira – AMB*, 2021. URL [https://cbo.net.br/2020/admin/docs\\_upload/034327Censo2021.pdf](https://cbo.net.br/2020/admin/docs_upload/034327Censo2021.pdf). accessed: august. 2021. x, 57, 58, 59, 60
- Hanna Vaahtoranta-Lehtonen, Anja Tuulonen, Pasi Aronen, Harri Sintonen, Liisa Suoranta, Niina Kovanen, Miika Linna, Esa Läärä, and Antti Malmivaara. Cost effectiveness and cost utility of an organized screening programme for glaucoma. *Acta Ophthalmologica Scandinavica*, 85(5):508–518, 2007. 1, 2
- K. Vandekerckhove. Uitwerkingvan cardiomyopathie bij het jonge kind. *Tijdschrift voor Geneeskunde*, 70(9):474–477, 2018. 56, 58
- Manuel AP Vilela, Felipe M Valença, Pedro KM Barreto, Carlos EV Amaral, and Lúcia C Pellanda. Agreement between retinal images obtained via smartphones and images obtained with retinal cameras or fundoscopic exams—systematic review and meta-analysis. *Clinical Ophthalmology*, pages 2581–2589, 2018. 2
- Volk. Volk optical inview, February 2022. URL <https://www.volk.com/>. accessed: october. 2020. 2, 83
- Ulrike Von Luxburg and Bernhard Schölkopf. Statistical learning theory: Models, concepts, and results. In *Handbook of the History of Logic*, volume 10, pages 651–706. Elsevier, 2011. 25, 34

- WelchAllyn. Welch allyn iexaminer turns the panoptic ophthalmoscope, February 2020. URL <https://www.welchallyn.com/en/microsites/iexaminer.html/>. accessed: may. 2020. 83, 85
- Weihao Weng and Xin Zhu. Inet: convolutional networks for biomedical image segmentation. *Ieee Access*, 9:16591–16603, 2021. 135
- WHO. Screening programmes: a short guide. *WHO Press*, 1:1–70, 2020. URL <https://iris.who.int/bitstream/handle/10665/330829/9789289054782-eng.pdf?sequence=1>. accessed: june. 2021. 144, 145, 146
- Nida M Zaitoun and Musbah J Aqel. Survey on image segmentation techniques. *Procedia Computer Science*, 65:797–806, 2015. 19
- Anna Zaleska-Żmijewska, Jacek P Szaflik, Paweł Borowiecki, Katarzyna Pohnke, Urszula Romaniuk, Izabela Szopa, Jacek Pniewski, and Jerzy Szaflik. A new platform designed for glaucoma screening: identifying the risk of glaucomatous optic neuropathy using fundus photography with deep learning architecture together with intraocular pressure measurements. *Klinika Oczna/Acta Ophthalmologica Polonica*, 122(1):1–6, 2020. 1
- Jing Zhang, Xindong Wu, and Victor S Sheng. Learning from crowdsourced labeled data: a survey. *Artificial Intelligence Review*, 46:543–576, 2016. 14
- Zhuo Zhang, Feng Shou Yin, Jiang Liu, Wing Kee Wong, Ngan Meng Tan, Beng Hai Lee, Jun Cheng, and Tien Yin Wong. Origa-light: An online retinal fundus image database for glaucoma analysis and research. In *2010 Annual international conference of the IEEE engineering in medicine and biology*, pages 3065–3068. IEEE, 2010. 65