



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
IFCE *CAMPUS* FORTALEZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELECOMUNICAÇÕES
MESTRADO ACADÊMICO EM ENGENHARIA DE TELECOMUNICAÇÕES

ALAN SERGIO DOS SANTOS

DETECÇÃO DE GLAUCOMA ATRAVÉS DE PROCESSAMENTO DIGITAL DE
IMAGENS PARA AUXÍLIO AO DIAGNÓSTICO MÉDICO DE EXAMES DE FUNDO
DE OLHO

FORTALEZA

2025

ALAN SERGIO DOS SANTOS

DETECÇÃO DE GLAUCOMA ATRAVÉS DE PROCESSAMENTO DIGITAL DE IMAGENS
PARA AUXÍLIO AO DIAGNÓSTICO MÉDICO DE EXAMES DE FUNDO DE OLHO

Dissertação apresentada ao curso de Mestrado Acadêmico em Engenharia de Telecomunicações do Programa de Pós-Graduação em Engenharia de Telecomunicações do *Campus* Fortaleza do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Telecomunicações.

Orientador: Prof. Dr. Tarique da Silveira Cavalcante.

Coorientador: Prof. Dr. Auzuir Ripardo de Alexandria

FORTALEZA

2025

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

S237d Santos, Alan Sergio dos.

Detecção de glaucoma através de processamento digital de imagens para auxílio ao diagnóstico médico de exames de fundo de olho / Alan Sergio dos Santos. - 2025.
168 f. : il.

Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Engenharia de Telecomunicações, Campus Fortaleza, 2025.

Orientação: Prof. Dr. Tarique da Silveira Cavalcante.

Coorientação: Prof. Dr. Auzuir Ripardo de Alexandria.

1. Glaucoma. 2. Redes Neurais Convolucionais. 3. Ensemble de Classificadores. 4. Generalização Interdomínio. 5. Diagnóstico Automatizado. I. Título.

CDD 621.382

ALAN SERGIO DOS SANTOS

**DETECÇÃO DE GLAUCOMA ATRAVÉS DE PROCESSAMENTO DIGITAL DE
IMAGENS PARA AUXÍLIO AO DIAGNÓSTICO MÉDICO DE EXAMES DE FUNDO
DE OLHO**

Dissertação apresentada ao curso de Mestrado Acadêmico em Engenharia de Telecomunicações do Programa de Pós-Graduação em Engenharia de Telecomunicações do *Campus* Fortaleza do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Telecomunicações.

Aprovado em: ____/____/____.

BANCA EXAMINADORA

Prof. Dr. Tarique da Silveira Cavalcante (Orientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Prof. Dr. Auzuir Ripardo de Alexandria (Coorientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Prof. Dra. Jessyca Almeida Bessa
Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Prof. Dr. Alyson Bezerra Nogueira Ribeiro
Universidade Federal do Ceará - UFC

A Deus.

Aos meus pais.

Aos mestres.

AGRADECIMENTOS

Agradeço a Deus pela força e sabedoria ao longo desta jornada.

À minha família, pelo amor, apoio incondicional e incentivo constante.

Ao meu orientador, Tarique da Silveira Cavalcante, e ao meu coorientador, Auzuir Ripardo de Alexandria, pela orientação, dedicação e contribuições fundamentais para o desenvolvimento deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro por meio da bolsa de estudos.

Ao Programa de Pós-Graduação em Engenharia de Telecomunicações (PPGET) do Instituto Federal de Educação, Ciências e Tecnologia do Ceará, pela formação oferecida e pelo ambiente propício ao crescimento acadêmico.

Aos professores, colegas e demais colaboradores, que, de alguma forma, contribuíram para esta conquista, deixo meu sincero agradecimento.

"Não se apavore nem desanime, pois o SENHOR, o seu Deus, estará com você por onde você andar."

(JOSUÉ 1:9)

RESUMO

A detecção automatizada de glaucoma em imagens de fundo de olho é prejudicada pela baixa generalização inter-domínio dos modelos existentes, que apresentam quedas drásticas de desempenho ao serem aplicados em bases de dados com características de aquisição distintas. Para enfrentar este desafio, este trabalho propõe um sistema baseado em um *ensemble* híbrido e adaptativo. O método emprega três redes neurais convolucionais pré-treinadas (EfficientNet-B0, MobileNetV2 e ResNet-50), com pesos congelados, para extração de características, seguidas da consolidação e normalização L2 dos vetores resultantes. A classificação é realizada em paralelo por um modelo denso e por três algoritmos clássicos (SVM, Random Forest e Gradient Boosting), cujas previsões são combinadas por média ponderada. Os pesos de cada modelo no *ensemble* são definidos automaticamente com base em sua acurácia, obtida por meio de validação cruzada estratificada de 5 *folds*, conferindo adaptabilidade ao sistema. O *pipeline* é complementado por técnicas de *data augmentation* para mitigar o sobreajuste.

As principais contribuições científicas residem na concepção deste *ensemble* adaptativo, que elimina a necessidade de ajuste manual para novos contextos; em um *pipeline* totalmente livre da etapa de segmentação explícita do disco óptico, reduzindo complexidade; e em uma validação em múltiplos domínios, realizada em cinco bases públicas heterogêneas (ACRIMA, Drishti-GS1, ORIGA, REFUGE, RIM-ONE) e no conjunto TOTAL (2.341 imagens). Avaliado sob este protocolo, o modelo demonstrou alto desempenho e capacidade de generalização, atingindo 97,87% de acurácia e AUC de 0,995 no RIM-ONE. Os resultados consistentes em diferentes domínios confirmam a aplicabilidade da abordagem no contexto clínico de triagem.

Palavras-chave: Glaucoma; Redes Neurais Convolucionais; Ensemble de Classificadores; Generalização Inter-domínio; Diagnóstico Automatizado.

ABSTRACT

Automated glaucoma detection in fundus images is hindered by the limited cross-domain generalization of existing models, which often suffer significant performance drops when applied to datasets with distinct acquisition characteristics. To address this challenge, this work proposes an adaptive hybrid ensemble-based system. The method employs three pretrained convolutional neural networks (EfficientNet-B0, MobileNetV2, and ResNet-50), with frozen weights, for feature extraction, followed by consolidation and L2 normalization of the resulting vectors. Classification is performed in parallel by a dense model and three classical algorithms (SVM, Random Forest, and Gradient Boosting), whose predictions are combined through weighted averaging. The weights of each model in the ensemble are automatically assigned based on their accuracy, obtained via stratified 5-fold cross-validation, providing adaptability to the system. The pipeline is complemented by data augmentation techniques to mitigate overfitting. The main scientific contributions of this work are the conception of an adaptive ensemble that eliminates the need for manual adjustment in new contexts, the design of a pipeline entirely free from explicit optic disc segmentation, reducing complexity, and a comprehensive multi-domain validation conducted on five heterogeneous public datasets (ACRIMA, Drishti-GS1, ORIGA, REFUGE, and RIM-ONE) and on the combined TOTAL set (2,341 images). Under this protocol, the proposed model achieved 97.87% accuracy and an AUC of 0.995 on the RIM-ONE dataset. The consistent results across distinct domains confirm the applicability of the proposed approach in clinical screening scenarios.

Keywords: Glaucoma; Convolutional Neural Networks; Classifier Ensemble; Cross-Domain Generalization; Automated Diagnosis.

LISTA DE FIGURAS

Figura 1 – Distribuição das bases de dados mais utilizadas nos trabalhos revisados (2018-2024).	36
Figura 2 – Estrutura anatômica do olho humano.	44
Figura 3 – Comparação anatômica entre nervo óptico saudável e nervo óptico com dano glaucomatoso.	44
Figura 4 – Visualização do ângulo da câmara anterior e das estruturas de drenagem (canal de Schlemm) em um olho saudável.	45
Figura 5 – Imagem de biomicroscopia ultrassônica (UBM) demonstrando o mecanismo facomórfico de fechamento do ângulo por bloqueio pupilar e deslocamento anterior do complexo íris-cristalino.	47
Figura 6 – Retinografia de fundo destacando disco óptico, mácula, fóvea e vasculatura.	49
Figura 7 – Anatomia do disco óptico (DO) destacando a escavação central e a borda neurorretiniana circundante.	50
Figura 8 – Representação da regra ISNT para avaliação da borda neurorretiniana.	51
Figura 9 – Exemplos de câmeras de fundo de olho automatizadas utilizadas na aquisição de retinografias.	52
Figura 10 – Esquema óptico ilustrando o percurso da luz na iluminação da retina e na formação da imagem pela pupila.	52
Figura 11 – Representação esquemática do neurônio biológico, evidenciando corpo celular, dendritos, axônio e sinapses.	60
Figura 12 – Arquitetura típica de uma RNA.	61
Figura 13 – Operação de convolução na CNN.	66
Figura 14 – Representação do <i>max pooling</i>	67
Figura 15 – Camada de Classificação de uma CNN clássica.	68
Figura 16 – Representação conceitual do <i>Dropout</i>	70
Figura 17 – Visão esquemática da arquitetura ResNet-50 destacando os cinco estágios residuais.	75
Figura 18 – Arquitetura simplificada do EfficientNet-B0, composta por blocos invertidos <i>MBConv</i> de diferentes tamanhos de <i>kernel</i>	76
Figura 19 – Arquitetura da MobileNetV2 e bloco <i>bottleneck</i> invertido.	78

Figura 20 – Margem de separação em uma SVM, destacando vetores de suporte e o hiperplano.	85
Figura 21 – Representação do algoritmo Random Forest.	86
Figura 22 – Representação dos classificadores em <i>ensemble</i>	88
Figura 23 – Fluxograma da metodologia proposta, com as etapas de pré-processamento, extração de características, classificação e integração final por <i>ensemble</i> ponderado.	94
Figura 24 – Representação esquemática do processo de extração e consolidação de características.	97
Figura 25 – Arquitetura do modelo denso proposto, com redução progressiva da dimensionalidade e camadas de <i>dropout</i>	98
Figura 26 – Fluxograma da validação cruzada <i>K-Fold</i> (5 folds), ilustrando as etapas de particionamento dos dados, treinamento/avaliação dos modelos e cálculo dos pesos do <i>ensemble</i>	107
Figura 27 – Matrizes de confusão de ACRIMA, <i>Drishti-GS1</i> , ORIGA, REFUGE, RIM-ONE e TOTAL.	126
Figura 28 – Desempenho comparativo do modelo proposto nos diferentes bancos de dados.	131
Figura 29 – Curvas ROC do modelo proposto nos diferentes conjuntos de dados (a AUC quantifica a capacidade discriminativa).	132
Figura 30 – Exemplos representativos das variações de qualidade entre as imagens dos conjuntos ORIGA (esquerda) e <i>Drishti-GS1</i> (direita). Notam-se desafios como baixo contraste, iluminação irregular e presença de artefatos.	143

LISTA DE TABELAS

Tabela 1 – Análise crítica comparativa de métodos para detecção automática de glaucoma.	38
Tabela 2 – Classificação da gravidade do glaucoma com base na Razão Escavação/Disco (CDR).	50
Tabela 3 – Resumo da heterogeneidade técnica em bancos de dados públicos de imagens de retina.	54
Tabela 4 – Características das <i>Stem Layers</i> nas arquiteturas avaliadas	55
Tabela 5 – Parâmetros de compatibilidade arquitetural para o redimensionamento 160 × 160 pixels.	57
Tabela 6 – Fundamentação para exclusão de técnicas de pré-processamento	58
Tabela 7 – Funções de ativação mais citadas na literatura e principais características.	63
Tabela 8 – Análise comparativa de arquiteturas CNN para detecção de glaucoma	82
Tabela 10 – Resumo dos atributos principais do <i>boosting</i> e <i>bagging</i> .	84
Tabela 11 – Resumo dos bancos de dados utilizados neste trabalho.	91
Tabela 12 – Distribuição de imagens após o <i>data augmentation</i> .	92
Tabela 13 – Arquitetura detalhada do modelo denso proposto.	97
Tabela 14 – Pesos atribuídos a cada modelo no <i>ensemble</i> , proporcionais à acurácia média obtida no <i>K-Fold</i> .	102
Tabela 15 – Análise de ablação mostrando o impacto da remoção individual de cada modelo no desempenho do <i>ensemble</i> .	103
Tabela 16 – Hiperparâmetros ajustados para os classificadores tradicionais do <i>ensemble</i> .	111
Tabela 17 – Hiperparâmetros testados e valores finais adotados para o modelo denso.	111
Tabela 18 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto ACRIMA.	114
Tabela 19 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto <i>Drishti-GS1</i> .	115
Tabela 20 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto ORIGA.	116
Tabela 21 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto REFUGE.	117
Tabela 22 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto RIM-ONE.	118

Tabela 23 – Síntese comparativa dos resultados de hiperparametrização por base de dados.	120
Tabela 24 – Desempenho comparativo (acurácia média $\pm$ desvio padrão) entre os classificadores. Valores em percentagem.	121
Tabela 25 – Análise de ablação mostrando o impacto da remoção de cada modelo no desempenho do <i>ensemble</i> no conjunto TOTAL.	122
Tabela 26 – Desempenho discriminativo completo por classe: comparação entre casos glaucomatosos e normais em conjuntos públicos de imagens de retina (valores em %).	123
Tabela 27 – Desempenho do modelo proposto em conjuntos públicos de imagens de fundo de olho (valores em %, exceto AUC).	130
Tabela 28 – Comparação do método proposto para o conjunto de dados ACRIMA. . . .	139
Tabela 29 – Comparação do método proposto para o conjunto de dados <i>Drishti-GS1</i> . . .	139
Tabela 30 – Comparação do método proposto para o conjunto de dados ORIGA. . . .	139
Tabela 31 – Comparação do método proposto para o conjunto de dados REFUGE. . . .	140
Tabela 32 – Comparação do método proposto para o conjunto de dados RIM-ONE. . . .	140

LISTA DE SIGLAS

1D	Unidimensional
2D	Bidimensional
2D-T-EWT	do inglês <i>Two-Dimensional Tensor Empirical Wavelet Transform</i>
3D	Tridimensional
4D	Quadrilateral
AdaBoost	do inglês <i>Adaptive Boosting</i>
Adagrad	do inglês <i>Adaptive Gradient</i>
ADALINE	do inglês <i>Adaptive Linear Element</i>
Adam	do inglês <i>Adaptive Momentum</i>
ADASYN	do inglês <i>Adaptive Synthetic Sampling</i>
AG	do inglês <i>Attention Gate</i>
AMF	do inglês <i>Adaptive Median Filter</i>
AUC	do inglês <i>Area Under the Curve</i>
CAB	do inglês <i>Channel Attention Block</i>
CAD	do inglês <i>Computer Aided diagnosis</i>
Cb	Crominância azul
CDAM-Net	do inglês <i>Channel-shuffle dual attention-based multi-scale CNN</i>
CDR	do inglês <i>Cup-to-Disc Ratio</i>
CLAHE	do inglês <i>Contrast Limit Adaptive Histogram Equalization</i>
CNNs	Redes Neurais Convolucionais
CP	do inglês <i>Concatenation of Paths</i>
Cr	Crominância vermelha
CSAM	do inglês <i>Channel self-attention module</i>
CSDA	do inglês <i>Channel Shuffling Double Attention</i>
DO	disco óptico
DVD	Diâmetro Vertical do Disco
DVE	Diâmetro Vertical da Escavação
EGWO-SVM	do inglês <i>Enhanced Grey Wolf Optimized Support Vector Machine</i>
ELM	do inglês <i>Extreme learning machine</i>
ESAM	do inglês <i>Enhanced self attention module</i>
EWT	do inglês <i>Empirical Wavelet Transform</i>

FC	do inglês <i>Fully Connected</i>
FL	do inglês <i>Loss of Fusion</i>
FLOPs	do inglês <i>Floating Point Operations</i>
FN	Falsos Negativos
FP	Falsos Positivos
FPR	do inglês <i>False Positive Rate</i>
GAP	do inglês <i>Global Average Pooling</i>
GAP 2D	do inglês <i>Global Average Pooling 2D</i>
GBDT	do inglês <i>Gradient-Boosted Decision Trees</i>
GLCM	do inglês <i>Gray-Level Co-Occurrence Matrix</i>
GPAA	Glaucoma primário de ângulo aberto
GPAF	Glaucoma primário de ângulo fechado
GPUs	do inglês <i>Graphics Processing Units</i>
GWO	do inglês <i>Grey Wolf Optimization</i>
HOG	do inglês <i>Histogram of Oriented Gradients</i>
HRF	do inglês <i>High resolution fundus</i>
HTG	Glaucoma de Alta Tensão
IA	Inteligência artificial
ISNT	Inferior, Superior, Nasal e Temporal
ISO	do inglês <i>International Organization for Standardization</i>
ITU-T	do inglês <i>International Telecommunication Union - Telecommunication Sector</i>
JPEG	do inglês <i>Joint Photographic Experts Group</i>
LAG	do inglês <i>Large-scale attention-based glaucoma database</i>
LCR	Líquido Cefalorraquiano
LDA	do inglês <i>Linear discriminant analysis</i>
LSTM	do inglês <i>Long Short-Term Memory</i>
MC-LS-SVM	do inglês <i>Multiclass Least Squares Support Vector Machines</i>
MFR	do inglês <i>Multiscale Feature Representation</i>
MICCAI	do inglês <i>Medical Image Computing and Computer Assisted Intervention</i>
MLP	<i>Perceptron</i> multicamadas
MOD-PSO-ELM	do inglês <i>Modified particle swarm optimization</i>

NMRC	do inglês <i>National Medical Research Council</i>
NRR	do inglês <i>Neuro-retinal Rim</i>
NTG	do inglês <i>Normal Tension Glaucoma</i>
PCA	do inglês <i>Principal component analysis</i>
PIO	Pressão intraocular
PSO	do inglês <i>Particle swarm optimization</i>
REFUGE	do inglês <i>Retinal Fundus Glaucoma Challenge</i>
ReLU	Unidade Linear Retificada
RFI	do inglês <i>Retinal fundus image dataset</i>
RGB	do inglês <i>Red, Green, Blue</i>
RMSProp	Propagação da Raiz Quadrada Média
RNAs	Redes Neurais Artificiais
RNFL	Fibras nervosas da retina
RNN	Rede Neural Recorrente
ROC	do inglês <i>Receiver Operating Characteristic</i>
ROI	Região de interesse
SAB	do inglês <i>Spatial Attention Block</i>
SAP	Sinéquias Anteriores Periféricas
SBIs	do inglês <i>Sub-Band Images</i>
Self-ONNs	Redes neurais operacionais auto-organizadas compactas
SERI	do inglês <i>Singapore Eye Research Institute</i>
SiLU	do inglês <i>Sigmoid Linear Unit</i>
SiMES	do inglês <i>Singapore Malay Eye Study</i>
SSAM	do inglês <i>Spatial self-attention module</i>
SURF	do inglês <i>Speeded-up Robust Feature</i>
SVM	do inglês <i>Support Vector Machine</i>
Tanh	Tangente Hiperbólica
TPR	do inglês <i>True Positive Rate</i>
TRk-CNN	Rede Neural Convolucional de Classificação Transferível
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos

YCbCr do inglês *Luminance, Chroma Blue and Chroma Red*

YIQ do inglês *Luminance, In-phase and Quadrature*

SUMÁRIO

1	INTRODUÇÃO	20
1.1	Contribuições Científicas Esperadas	21
1.2	Objetivo Geral	22
1.3	Objetivos Específicos	22
1.4	Estrutura do trabalho	23
2	TRABALHOS CORRELATOS	25
2.1	Métodos Baseados em Segmentação	26
2.2	Métodos Leves para Dispositivos Móveis	27
2.3	Abordagens com Mecanismos de Atenção	28
2.4	Classificação Direta com CNNs	30
2.5	Métodos Metaheurísticos	31
2.6	Abordagens Híbridas e Baseadas em <i>Ensemble</i>	33
2.7	Abordagens para Conjuntos de Dados Pequenos	34
2.8	Síntese Crítica e Identificação de Lacunas	36
3	FUNDAMENTAÇÃO TEÓRICA	42
3.1	Estrutura Óptica do Olho Humano	42
3.2	Glaucoma	43
3.2.1	<i>Glaucoma primário de ângulo aberto (GPAA)</i>	45
3.2.2	<i>Glaucoma primário de ângulo fechado (GPAF)</i>	46
3.2.3	<i>Glaucoma Secundário</i>	47
3.2.4	<i>Epidemiologia</i>	47
3.3	Imagens de Fundo de Retina	48
3.3.1	<i>Aquisição de Imagens de Fundo de Retina</i>	51
3.4	Pré-processamento de Imagens Digitais	55
3.4.1	<i>Compatibilidade com as Arquiteturas Empregadas</i>	55
3.4.2	<i>Implementação do Pipeline Minimalista</i>	56
3.4.3	<i>Técnicas de Pré-processamento na Literatura e Razões para Exclusão</i>	57
3.4.4	<i>Justificativa Estratégica</i>	57
3.4.5	<i>Data Augmentation</i>	59
3.5	Fundamentos de Redes Neurais Artificiais	60
3.5.1	<i>Modelo do Neurônio Biológico</i>	60

3.5.2	<i>Perceptron e Formalização Matemática</i>	61
3.5.3	<i>Arquitetura de Redes Neurais</i>	61
3.5.3.1	<i>Redes Feedforward e Backpropagation</i>	62
3.5.3.2	<i>Redes Recorrentes</i>	62
3.5.4	<i>Funções de Ativação</i>	62
3.5.4.1	<i>Entropia Cruzada Binária</i>	63
3.5.5	<i>Limitações e Relação com CNNs para Detecção de Glaucoma</i>	64
3.6	<i>Redes Neurais Convolucionais (CNNs)</i>	65
3.6.1	<i>Arquitetura Geral</i>	65
3.6.1.1	<i>Camadas Convolucionais</i>	66
3.6.1.2	<i>Camadas de pooling</i>	66
3.6.1.3	<i>Camadas totalmente conectadas (FC)</i>	67
3.6.1.4	<i>Global Average Pooling 2D</i>	68
3.6.2	<i>Regularização em Redes Neurais</i>	69
3.6.2.1	<i>Dropout</i>	69
3.6.2.2	<i>Regularização L2</i>	70
3.6.3	<i>Treinamento e Otimização</i>	71
3.6.3.1	<i>Funções de perda</i>	71
3.6.3.2	<i>Otimizadores adaptativos</i>	72
3.6.3.3	<i>Taxa de aprendizado</i>	73
3.6.3.4	<i>Tamanho do lote</i>	73
3.6.4	<i>Avanços Arquiteturais e Transfer Learning</i>	74
3.6.4.1	<i>ResNet-50</i>	74
3.6.4.2	<i>EfficientNet-B0</i>	76
3.6.4.3	<i>MobileNetV2</i>	77
3.6.4.4	<i>Integração no Método Proposto</i>	77
3.6.5	<i>Mecanismos de Atenção</i>	79
3.6.6	<i>Análise Comparativa e Limitações</i>	80
3.6.7	<i>Tendências em Abordagens Híbridas</i>	81
3.7	<i>Classificadores de aprendizado de máquina</i>	82
3.7.1	<i>Bagging e Boosting</i>	83
3.7.2	<i>Máquina de Vetores de Suporte</i>	83
3.7.3	<i>Random Forest</i>	85

3.7.4	<i>Gradient Boosting</i>	86
3.8	Classificadores em <i>ensemble</i>	87
4	MATERIAIS E MÉTODOS	90
4.1	Descrição dos bancos de dados	90
4.1.1	<i>Data augmentation</i>	91
4.1.2	<i>Pré-processamento de Imagens</i>	92
4.2	Modelo Proposto para a Classificação de Imagens Glaucomatosas	93
4.2.1	<i>Extração de Características com CNNs Pré-treinadas</i>	95
4.2.2	<i>Consolidação do Vetor de Características</i>	96
4.3	Modelo Denso	96
4.3.1	<i>Otimizador Adam</i>	99
4.4	Classificadores em <i>ensemble</i>	100
4.4.1	<i>Seleção e Justificativa dos Algoritmos</i>	100
4.4.2	<i>Processo de Treinamento e Combinação</i>	101
4.4.3	<i>Análise de Diversidade do Ensemble</i>	102
4.4.4	<i>Análise de Contribuição Individual</i>	103
4.5	Protocolo Experimental	103
4.5.1	<i>Validação Cruzada K-Fold</i>	104
4.5.1.1	<i>Configuração Experimental e Controle de Vazamento de Dados</i>	105
4.5.1.2	<i>Aplicação da Validação Cruzada no Pipeline Proposto</i>	106
4.5.1.3	<i>Justificativa da Escolha</i>	106
4.5.2	<i>Ambiente de Execução</i>	108
4.5.3	<i>Métricas de Avaliação</i>	108
4.6	Treinamento	110
4.6.1	<i>Ajuste de Hiperparâmetros — Modelos Tradicionais</i>	110
4.6.2	<i>Ajuste de Hiperparâmetros — Modelo Denso</i>	110
4.7	Resumo da Metodologia Proposta	111
5	RESULTADOS	114
5.1	Análise de hiperparametrização	114
5.1.1	<i>ACRIMA</i>	114
5.1.2	<i>Drishti-GS1</i>	115
5.1.3	<i>ORIGA</i>	116
5.1.4	<i>REFUGE</i>	117

5.1.5	<i>RIM-ONE</i>	118
5.1.6	<i>Síntese e Análise Comparativa dos Resultados de Hiperparametrização</i> . .	119
5.2	Classificadores Individuais vs Ensemble	120
5.3	Análise Discriminativa e Padrões de Erro a partir das Matrizes de Confusão	123
5.3.1	<i>Métricas de Desempenho por Classe</i>	123
5.3.1.1	<i>Bases Equilibradas e de Alta Performance (ACRIMA e RIM-ONE)</i>	124
5.3.1.2	<i>Bases com Desempenho Assimétrico (Drishti-GSI, ORIGA e REFUGE)</i> . . .	124
5.3.2	<i>Análise Qualitativa de Erros a partir das Matrizes de Confusão</i>	125
5.3.2.1	<i>Falsos Negativos (FN) e Impacto Clínico</i>	127
5.3.2.2	<i>Falsos Positivos (FP) e Especificidade</i>	127
5.3.2.3	<i>Conjunto TOTAL: Um Cenário Realista</i>	128
5.3.3	<i>Síntese e Implicações</i>	129
5.4	Desempenho Geral do Modelo Proposto na Detecção de Glaucoma em Imagens de Fundo de Olho	130
5.4.1	<i>Análise Consolidada de Desempenho</i>	130
5.4.2	<i>Análise das Curvas ROC por Base de Dados</i>	132
5.4.2.1	<i>Análise Visual das Curvas ROC</i>	133
5.4.3	<i>Análise Crítica e Limitações</i>	134
5.4.4	<i>Implicações Práticas e Recomendações</i>	135
5.4.5	<i>Conclusão do Desempenho Geral</i>	136
5.5	Análise comparativa	137
5.6	Discussão dos Resultados	141
5.6.1	<i>Análise Crítica da Arquitetura: Leveza vs. Performance</i>	141
5.6.2	<i>Impacto do Desbalanceamento</i>	142
5.6.3	<i>Impacto da Qualidade das Imagens</i>	143
5.6.4	<i>Overfitting e Generalização: Uma Conquista Parcial</i>	144
5.6.5	<i>Limitações Computacionais do Ensemble</i>	145
6	CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS . . .	146
6.1	<i>Síntese das Conclusões e Contribuições</i>	146
6.2	<i>Limitações e Considerações sobre a Aplicabilidade</i>	147
6.3	<i>Trabalhos Futuros e Perspectivas</i>	148
6.4	<i>Considerações Finais</i>	149
	REFERÊNCIAS	150

1 INTRODUÇÃO

O glaucoma é um grupo de doenças oftalmológicas caracterizadas por danos progressivos e irreversíveis ao nervo óptico, resultando em perda gradual da visão, geralmente associada à elevação da pressão intraocular (PIO) e frequentemente assintomática nos estágios iniciais (MURTHY *et al.*, 2022). Trata-se de um problema de saúde pública relevante, sendo uma das principais causas de cegueira irreversível no mundo. Estima-se que atualmente mais de 76 milhões de pessoas sejam afetadas globalmente, número que pode alcançar 111,8 milhões até 2040 (ALLISON; PATEL; ALABI, 2020; THAM *et al.*, 2014), com maior prevalência na Ásia (cerca de 60% dos casos) e impacto significativo também nas Américas e Europa. Nos Estados Unidos, por exemplo, aproximadamente 1,2 milhão de casos de cegueira são atribuídos ao glaucoma anualmente (SINGH *et al.*, 2022). O diagnóstico precoce é fundamental, mas apresenta desafios clínicos devido à natureza crônica e silenciosa da doença, exigindo equipamentos de alto desempenho e avaliação especializada (SHEN *et al.*, 2020).

As imagens médicas desempenham papel central no diagnóstico e monitoramento do glaucoma, permitindo a documentação visual e o acompanhamento da evolução da doença (SHOBA; THERESE, 2020). Dentre essas modalidades, a retinografia — imagem de fundo de olho — é amplamente utilizada por ser não invasiva, de baixo custo e capaz de registrar detalhadamente estruturas como disco óptico (DO), escavação e rede vascular. Contudo, sua análise manual é demorada, depende de experiência especializada e está sujeita a variações interpretativas entre observadores (SHOUKAT *et al.*, 2023).

Apesar do potencial diagnóstico da retinografia, a análise automatizada dessas imagens impõe desafios técnicos críticos para modelos de visão computacional. A heterogeneidade técnica na aquisição — incluindo variações extremas de resolução (240×240 até 4752×3168 pixels), campos visuais díspares (10° a 60°), formatos de compressão heterogêneos (.JPG, .PNG ou não comprimidas) e alta incidência de artefatos como reflexos (23% no REFUGE) — compromete diretamente a extração de características e a capacidade de generalização dos modelos, como detalhado na Tabela 3.

Esses desafios são agravados por dois fatores críticos: um desbalanceamento acentuado entre classes — com proporções extremas como 1:9 no REFUGE (40 glaucoma vs 360 normais) e 1:2,8 no ORIGA (168 vs 482), resultando em desequilíbrio global de 1:1,77 (846 vs 1495) conforme a Tabela 11 — combinado a variações nos critérios de anotação entre bancos públicos devido a inconsistências nos protocolos de rotulagem. Essas condições, frequentes em cenários

clínicos reais, exigem abordagens consistentes capazes de manter desempenho estável mesmo em contextos heterogêneos (SHEN *et al.*, 2020; SHOBA; THERESE, 2020).

Nesse cenário, técnicas de Visão Computacional e Inteligência Artificial (IA) têm se mostrado promissoras para automatizar e padronizar a detecção do glaucoma. Sistemas de Diagnóstico Auxiliado por Computador (CAD), baseados inicialmente em aprendizado de máquina e, mais recentemente, em algoritmos de *deep learning*, já demonstraram desempenho competitivo na análise de fotografias de fundo de olho (SHOUKAT *et al.*, 2023; FAN *et al.*, 2023). Contudo, a literatura ainda apresenta lacunas importantes: a maioria dos métodos mantém desempenho elevado apenas em bases de dados específicas, sofrendo quedas significativas quando expostos a domínios clínicos diferentes (SHYAMALEE; MEEDENIYA, 2022; LATIF *et al.*, 2022). Essa limitação decorre, em grande parte, da dependência de segmentações precisas, do uso de conjuntos de treinamento restritos e da ausência de mecanismos de adaptação à variabilidade das imagens.

Para mitigar esses desafios, este trabalho propõe um sistema baseado em redes neurais convolucionais (CNNs) pré-treinadas como extratoras de características, combinadas com classificadores tradicionais em um *ensemble* ponderado, com pesos ajustados automaticamente por validação cruzada estratificada. Aliado a técnicas de aumento de dados, o método busca reduzir o risco de *overfitting* e manter alta acurácia estável em diferentes bases públicas, reforçando sua aplicabilidade em triagens clínicas.

Do ponto de vista científico, espera-se que o método desenvolvido nesta pesquisa:

- mantenha desempenho estável em múltiplas bases públicas heterogêneas;
- elimine a necessidade de segmentação manual ou pré-processamentos complexos; e
- apresente potencial de aplicação prática em triagens clínicas automatizadas, com validação quantitativa comparável ou superior ao estado da arte.

1.1 Contribuições Científicas Esperadas

Espera-se que este trabalho avance o estado da arte em detecção automatizada de glaucoma por meio das seguintes inovações técnicas e metodológicas:

1. *Ensemble* híbrido com três componentes complementares:

- Extração multi-CNN: utilização de EfficientNet-B0, MobileNetV2 e ResNet-50 com normalização L2, explorando complementaridade entre padrões lineares e não lineares;
 - Classificação paralela: combinação de um modelo denso e três classificadores tradicionais (SVM, Random Forest e Gradient Boosting) para melhorar estabilidade e precisão;
 - Ponderação automática: fusão das probabilidades por média ponderada, com pesos definidos proporcionalmente pela acurácia média obtida em validação cruzada estratificada de 5 *folds*, promovendo o controle da variabilidade e assegurando consistência dentro do mesmo conjunto de dados.
2. *Pipeline* livre de segmentação: extração direta de características de imagens de fundo de olho, eliminando a dependência de segmentação prévia e favorecendo tanto a simplicidade computacional quanto a viabilidade clínica.
3. Protocolo de validação multidomínio: avaliação quantitativa em cinco bases públicas heterogêneas (ACRIMA, *Drishti-GS1*, ORIGA, REFUGE, RIM-ONE DL), abrangendo variações de prevalência, qualidade e equipamentos de aquisição, possibilitando medir a estabilidade do desempenho frente a mudanças de domínio e reproduzir cenários clínicos reais.

1.2 Objetivo Geral

Desenvolver e avaliar um sistema automatizado para a classificação e detecção de glaucoma em imagens de fundo de olho, combinando redes neurais convolucionais pré-treinadas e classificadores tradicionais em um *ensemble* com ajuste automático de pesos, visando alto desempenho, estabilidade interdomínio e aplicabilidade clínica.

1.3 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Utilizar Redes Neurais Convolucionais (CNNs) pré-treinadas como extratoras de atributos, aplicando transferência de aprendizado para aproveitar representações visuais consistentes e reduzir a necessidade de grandes volumes de dados anotados;

- Aplicar técnicas de aumento de dados (*data augmentation*) para ampliar a variabilidade do conjunto de treinamento e melhorar a capacidade de generalização do modelo;
- Desenvolver um modelo denso (totalmente conectado) para a classificação das características extraídas, explorando sua capacidade de capturar relações não lineares entre atributos;
- Integrar múltiplos classificadores (SVM, Random Forest, Gradient Boosting e modelo denso) por meio de *ensemble* de probabilidades com pesos ajustados automaticamente via validação cruzada estratificada, a fim de aumentar a precisão e a estabilidade em diferentes contextos clínicos;
- Realizar experimentos utilizando múltiplos bancos de dados públicos heterogêneos (ORIGA, REFUGE, *Drishti-GSI*, RIM-ONE e ACRIMA), além de avaliar o desempenho com todos os bancos combinados;
- Comparar quantitativamente os resultados do sistema proposto com abordagens da literatura, analisando ganhos em desempenho, simplicidade operacional e aplicabilidade clínica.

1.4 Estrutura do trabalho

Este trabalho de dissertação é estruturado em 6 capítulos organizados da seguinte forma: Capítulo 1 (Introdução) está relacionado a introdução desse trabalho apresentando o contexto do problema e os objetivos do estudo.

O Capítulo 2 (Trabalhos Correlatos) revisa a literatura explorando as principais abordagens de aprendizado de máquina para a detecção do glaucoma, com ênfase em aprendizado profundo. São discutidos modelos *backbone* utilizados para classificação de imagens de fundo de olho, além da análise crítica de estudos anteriores, destacando lacunas e oportunidades de pesquisa.

O Capítulo 3 (Fundamentação Teórica) contextualiza o problema com uma visão geral das diferentes formas de glaucoma e sua relevância clínica. Em seguida, apresenta os fundamentos das redes neurais convolucionais e de outros classificadores de aprendizado de máquina, justificando as escolhas metodológicas adotadas no estudo.

O Capítulo 4 (Materiais e Métodos) descreve os bancos de dados utilizados, suas características e os critérios de seleção. Além disso, detalha a arquitetura do modelo de aprendizado de máquina desenvolvido para a classificação das imagens de fundo de olho.

O Capítulo 5 (Resultados) apresenta a seleção dos hiperparâmetros que impactam no desempenho do modelo e avalia sua eficácia em diferentes conjuntos de dados. Os resultados são comparados com estudos recentes para validar a abordagem proposta.

Por fim, o Capítulo 6 (Conclusões, Contribuições e Trabalhos Futuros) resume os principais achados do estudo, discute as limitações do modelo e sugere direções para pesquisas futuras na detecção automatizada do glaucoma.

2 TRABALHOS CORRELATOS

A detecção automatizada de glaucoma em imagens de fundo de olho consolida-se como um campo de pesquisa vital, dada a gravidade da doença e a urgência por métodos de triagem eficazes. A trajetória evolutiva das abordagens — que migrou de técnicas clássicas de segmentação e extração de características manuais para modelos de aprendizado profundo — espelha a busca dupla por maior precisão diagnóstica e, contemporaneamente, por generalização em cenários clínicos heterogêneos.

Este capítulo apresenta uma análise crítica e sistematizada do estado da arte, concentrando-se em contribuições publicadas entre 2018 e 2024. A seleção dos estudos considerou quatro critérios principais: relevância metodológica, com ênfase em propostas inovadoras ou paradigmáticas; impacto científico, mensurado pela adoção da comunidade e publicação em periódicos de referência; validação experimental em bases de dados públicas e reconhecidas; e alinhamento temático com os desafios centrais desta dissertação, notadamente a generalização interdomínio e a viabilidade de implantação.

Para organizar a diversidade de propostas, a análise foi estruturada em sete eixos metodológicos:

1. Métodos Baseados em Segmentação;
2. Classificação Direta com CNNs;
3. Abordagens com Mecanismos de Atenção;
4. Métodos Leves para Dispositivos Móveis;
5. Métodos Metaheurísticos;
6. Abordagens Híbridas e Baseadas em *Ensemble*; e
7. Estratégias para Conjuntos de Dados Pequenos.

Cada eixo será criticamente examinado sob múltiplas dimensões, que avalia méritos técnicos e inovação, desempenho quantitativo reportado, capacidade de generalização entre bases de dados diversas e, por fim, relevância clínica e operacional, um aspecto crucial para a tradução do algoritmo para o ambiente de saúde real.

Esta estrutura analítica é adotada não apenas para catalogar, mas para contrastar e sintetizar as tendências dominantes, os *trade-offs* inerentes e, sobretudo, as lacunas persistentes na

literatura. A conclusão do capítulo consolidará tais lacunas, como a instabilidade de desempenho em novos domínios e a dependência de ajustes manuais, fornecendo assim a base empírica que fundamenta e motiva as opções metodológicas originais adotadas nesta dissertação, as quais serão detalhadas a partir do Capítulo 3.

2.1 Métodos Baseados em Segmentação

As primeiras soluções automatizadas para detecção de glaucoma buscaram replicar o protocolo clínico tradicional, segmentando o disco óptico (DO) e a escavação para extrair atributos morfométricos quantificáveis, como a razão escavação/disco (CDR). Essas abordagens possuem alto valor explicativo, pois mimetizam o raciocínio especializado, mas introduzem vulnerabilidades significativas relacionadas à dependência da precisão da segmentação e à qualidade da imagem.

Diversos trabalhos ilustram essa abordagem e suas limitações inerentes. Guo *et al.* (2020) propuseram um *pipeline* baseado em uma UNet++ aprimorada para segmentação, seguida da extração de características de textura (GLCM, Gabor) em múltiplas escalas e classificação final com GBDT. Apesar do desempenho competitivo (acurácia de 84% em bases controladas), o método mostrou-se criticamente dependente da etapa de segmentação, com quedas superiores a 10 p.p. (pontos percentuais) em bases de dados heterogêneas. De forma semelhante, Civit-Masot *et al.* (2020) desenvolveram um sistema dual que combina a segmentação via U-Net com um classificador direto (MobileNetV2), fundindo suas saídas para aumentar a confiabilidade. Embora tenham alcançado 86% de acurácia, a validação restrita a conjuntos específicos compromete a confiança em sua capacidade de generalização.

Outros estudos, como os de Shinde (2021) e Arsyian e Maki (2022), reforçam esse paradigma e suas fragilidades. O primeiro emprega segmentação por U-Net e extração da região de interesse via método “*Brightest Spot*” para posterior classificação, dependendo fortemente de máscaras binárias detalhadas. O segundo apostou em um pré-processamento intensivo (CLAHE e morfologia matemática) e extração de momentos invariantes para descrição de formas, mas a qualidade variável das imagens e o desequilíbrio de classes comprometeram a estabilidade dos resultados. Ambos os métodos são sensíveis a artefatos e ruídos, dificultando sua aplicação em dados menos padronizados.

Em síntese, os métodos baseados em segmentação, embora clinicamente intuitivos, compartilham limitações fundamentais que restringem sua aplicação em cenários reais:

- **Dependência crítica de segmentações precisas:** A performance do sistema está intrinsecamente ligada ao sucesso de uma etapa anterior e vulnerável, que requer imagens rotuladas manualmente para treinamento e é sensível à variabilidade de aquisição.
- **Alta complexidade computacional e estrutural:** A necessidade de executar modelos complexos de segmentação antes da classificação torna o processo lento e computacionalmente custoso, inviabilizando a implantação em ambientes com infraestrutura limitada.
- **Sensibilidade a ruídos, artefatos e variações de imagem:** A propagação de erros é um problema central; variações no contraste, iluminação ou presença de patologias oclusivas afetam drasticamente a segmentação e, consequentemente, a classificação final.
- **Limitação de generalização:** Como demonstrado, a validação concentrada em bases homogêneas e controladas superestima o desempenho, que decai abruptamente frente à heterogeneidade de datasets clínicos reais.

Essas limitações motivaram o desenvolvimento de abordagens livres de segmentação explícita, cuja fundamentação é apresentada no Capítulo 3 (Fundamentação Teórica), a metodologia detalhada no Capítulo 4 (Materiais e Métodos) e a avaliação experimental no Capítulo 5 (Resultados).

2.2 Métodos Leves para Dispositivos Móveis

A popularização de dispositivos móveis e a necessidade de triagem acessível em ambientes com infraestrutura limitada motivaram o desenvolvimento de métodos computacionalmente eficientes para detecção de glaucoma. Essas abordagens priorizam baixo consumo computacional, velocidade de inferência e portabilidade, frequentemente enfrentando o *trade-off* entre eficiência e estabilidade de desempenho em cenários clínicos diversificados.

Guo *et al.* (2018) propuseram o sistema Yanbao, um dos primeiros aplicativos móveis para triagem de glaucoma, integrando segmentação automática do disco óptico, extração de parâmetros clínicos (CDR, NRR) e classificação por SVM ou Random Forest. Embora tenha obtido acurácia superior a 85% em ambientes controlados, a sensibilidade do método à qualidade da imagem e à precisão da segmentação resultou em perda significativa de desempenho na presença de artefatos ou entre diferentes dispositivos.

Em uma abordagem arquitetural diferente, Devecioglu *et al.* (2021) exploraram as Self-ONNs, demonstrando processamento em tempo real (inferência inferior a 0,1s por imagem) com

redução significativa de parâmetros computacionais. Apesar de superar CNNs convencionais em três bases públicas, a validação concentrou-se em contextos laboratoriais, não abordando adequadamente desafios de integração clínica ou adaptação a bases externas.

Complementando essas estratégias, Mrad *et al.* (2022) desenvolveram uma CNN compacta específica para *smartphones*, utilizando estratégias intensivas de aumento de dados (*data augmentation*) para compensar limitações amostrais. Apesar da execução rápida e requisitos mínimos de *hardware*, o desempenho mostrou-se instável entre diferentes bases de dados e dependente de ajustes manuais para adaptação, comprometendo sua escalabilidade clínica.

A revisão desses trabalhos revela desafios comuns aos métodos leves:

- **Instabilidade de desempenho:** Acurácia e sensibilidade variam significativamente (10-15 p.p.) entre diferentes dispositivos, bases de dados e condições de aquisição, refletindo sensibilidade a fatores não controlados em ambientes reais.
- **Dependência de pré-processamento:** A persistência de etapas como segmentação ou pré-processamento intensivo introduz pontos de falha e dependência de parâmetros ajustados manualmente, prejudicando a capacidade de generalização do sistema para diferentes cenários.
- **Validação insuficiente em ambientes reais:** A maioria dos estudos é validada em condições controladas ou poucas bases de dados, não demonstrando adequadamente performance em cenários clínicos diversificados com populações variadas e condições de aquisição heterogêneas.

Estas limitações evidenciam a necessidade de abordagens que conciliem eficiência computacional com generalização. Soluções que eliminam a dependência de etapas manuais de pré-processamento e oferecem validação abrangente em múltiplos contextos representam promessa para viabilizar aplicações móveis confiáveis, essas questões fundamentam a discussão apresentada no Capítulo 3 (Fundamentação Teórica), que aborda estratégias para equilibrar eficiência e generalização.

2.3 Abordagens com Mecanismos de Atenção

Os mecanismos de atenção representam uma evolução significativa em relação às abordagens baseadas em segmentação explícita. Em vez de depender de etapas manuais de pré-processamento, essas técnicas permitem que as redes neurais concentrem o processamento em

regiões morfológicamente relevantes das imagens de fundo de olho, integrando essa capacidade diretamente na arquitetura do modelo. Do ponto de vista clínico, isso oferece maior potencial de automatização, mas introduz novos desafios de eficiência computacional e generalização.

Diversas implementações ilustram essa tendência. Li *et al.* (2020) propuseram o AG-CNN, que integra mapas de atenção gerados por especialistas oftalmologistas a uma arquitetura ResNet. O modelo alcançou AUC de 0,975 na base LAG, destacando-se pela explicabilidade. No entanto, sua validação restrita a bases homogêneas e a dependência de anotações manuais de atenção limitam sua generalização para domínios com diferentes padrões de aquisição. De forma complementar, Shyamalee e Meedeniya (2022) empregaram *Attention U-Net* explorando diferentes arquiteturas de *backbone* (Inception-v3, VGG-19, ResNet-50) para segmentação precisa seguida de classificação, obtendo ganhos em sensibilidade e especificidade. A arquitetura, porém, mostrou-se computacionalmente custosa e propensa a *overfitting* em conjuntos pequenos.

Abordagens mais recentes buscaram maior sofisticação nos mecanismos de atenção. Das, Nayak e Pachori (2024) desenvolveram a AES-Net com módulos de auto-atenção adaptativos (ESAM) e adaptação espacial para classificação multiestágio, reportando até 99,3% de acurácia em bases específicas. De forma similar, Das *et al.* (2024) propuseram a CDAM-Net com atenção *cross-domain* e multi-escala, buscando manter estabilidade entre domínios diferentes. Ambos os métodos, contudo, exigem poder computacional substancial e *hardware* especializado, com validação ainda limitada a conjuntos específicos. Shyamalee *et al.* (2024) exploraram módulos de atenção autoajustáveis acoplados a CNNs profundas, mas relataram aumento na sensibilidade a ruídos em bases não homogêneas, elevando os riscos de *overfitting*.

A revisão dessas abordagens revela um padrão consistente de vantagens e limitações:

- **Vantagens na interpretabilidade e desempenho:** Os mecanismos de atenção proporcionam ganhos significativos em acurácia e oferecem mapas de atenção visuais sobre as regiões consideradas importantes pelo modelo, aumentando a confiança clínica potencial.
- **Complexidade computacional elevada:** A incorporação de módulos de atenção aumenta significativamente o número de parâmetros e operações, dificultando a implantação em ambientes com recursos limitados e aumentando o tempo de inferência.
- **Suscetibilidade a *overfitting*:** Especialmente em conjuntos pequenos ou pouco variados, as arquiteturas com atenção tendem a apresentar maior vulnerabilidade ao *overfitting*, necessitando de estratégias consistentes de regularização.

- **Validação limitada entre domínios:** A maioria dos estudos são validados em poucas bases de dados, frequentemente com características similares, não demonstrando adequadamente a capacidade de generalização para contextos clínicos diversificados.

Em síntese, embora promissores, os métodos baseados exclusivamente em atenção ainda enfrentam desafios de eficiência e generalização. Essas limitações reforçam a relevância de abordagens híbridas, como a proposta nesta dissertação, que buscam conciliar desempenho, estabilidade interdomínio e viabilidade clínica.

2.4 Classificação Direta com CNNs

A classificação direta utilizando Redes Neurais Convolucionais (CNNs) representa o paradigma dominante na detecção automatizada de glaucoma, superando as abordagens baseadas em segmentação ao extrair características relevantes diretamente das imagens de fundo de olho. O aprendizado por transferência (*transfer learning*) com arquiteturas pré-treinadas no ImageNet tornou-se prática padrão, permitindo alto desempenho mesmo com conjuntos de dados limitados. No entanto, a estabilidade e a capacidade de generalização desses modelos entre diferentes domínios clínicos permanecem desafios críticos.

Estudos seminalmente demonstraram tanto o potencial quanto as limitações dessa abordagem. Diaz-Pinto *et al.* (2019) investigaram extensivamente CNNs pré-treinadas (VGG, InceptionV3, ResNet-50, Xception) para detecção de glaucoma, reportando acurácias superiores a 93% em contextos específicos. Entretanto, ao avaliar o desempenho em diferentes domínios clínicos, observaram variações de até 12 p.p., evidenciando sensibilidade crítica à mudança de distribuição de dados. De forma complementar, Juneja *et al.* (2020) propuseram a GC-NET, uma arquitetura customizada que alcançou 97,51% de acurácia em bases específicas, porém exigindo ajuste manual de hiperparâmetros para cada novo contexto, o que compromete sua adaptabilidade.

Abordagens mais recentes buscaram explicitamente melhorar a generalização. Latif *et al.* (2022) desenvolveram a ODGNet, combinando localização do disco óptico, aumento de dados e *transfer learning* para mitigar desbalanceamento de classes. Apesar dos resultados promissores (até 95,75% de acurácia), a abordagem permaneceu dependente de configurações específicas por cenário, não demonstrando estabilidade consistente entre domínios diversos. Velpula e Sharma (2023) avaliaram sistematicamente arquiteturas profundas (ResNet-101, NasNetMobile, NasNetLarge) em múltiplos contextos, corroborando o achado fundamental: mesmo modelos

com excelente desempenho em domínios específicos apresentam variações superiores a 10% entre diferentes contextos clínicos.

A consolidação desses trabalhos revela o paradoxo central dos métodos baseados em CNNs:

- **Alta performance intra-domínio:** Quando treinadas e validadas em contextos similares, CNNs pré-treinadas alcançam desempenho excepcional, frequentemente superior a 95% de acurácia, estabelecendo um novo patamar de referência.
- **Sensibilidade interdomínio:** O desempenho decai significativamente (10–15 p.p.) quando os modelos são expostos a dados de diferentes distribuições, equipamentos, protocolos de aquisição ou populações de pacientes, revelando baixa adaptabilidade a mudanças de domínio.
- **Dependência de ajustes manuais:** A manutenção do desempenho normalmente exige re-treinamento ou ajuste fino conduzido por especialista para cada novo contexto, processo que demanda expertise técnica e recursos computacionais, inviabilizando implantação escalável.
- **Validação limitada:** A maioria dos estudos reporta resultados em poucos domínios (tipicamente 1–3 bases de dados), superestimando a capacidade de generalização real dos modelos.

Esse *gap* de generalização interdomínio constitui a principal limitação dos métodos atuais baseados em CNNs e representa a motivação central para o desenvolvimento de abordagens capazes de sustentar desempenho consistente em diferentes cenários clínicos, sem necessidade de ajustes manuais recorrentes. Essa lacuna fundamenta a discussão teórica apresentada no Capítulo 3 (Fundamentação Teórica) e orienta as escolhas metodológicas detalhadas no Capítulo 4 (Materiais e Métodos).

2.5 Métodos Metaheurísticos

Os métodos metaheurísticos representam uma abordagem distinta na detecção de glaucoma, focada na otimização de *pipelines* de aprendizado de máquina através de algoritmos inspirados em processos naturais. Estas técnicas são tipicamente aplicadas na seleção de características (*feature selection*) ou na otimização de hiperparâmetros, buscando automatizar a identificação das configurações ideais que maximizam o desempenho diagnóstico.

Latif *et al.* (2023) exemplificam esta abordagem ao propor um método híbrido que combina extração de características convencionais (SURF, HOG, estatísticas globais) com a metaheurística *Enhanced Grey Wolf Optimization* (EGWO) para seleção automática de atributos, seguida de classificação via SVM. O emprego da metaheurística resultou em acurácia e *F1-score* superiores a 90% e 91%, respectivamente, na base ORIGA, além de permitir redução significativa da dimensionalidade dos dados. Contudo, o desempenho mostrou-se fortemente dependente do ajuste manual dos parâmetros do algoritmo de otimização e restrito à estabilidade intra-base: quando aplicado a outros contextos ou populações, observam-se quedas superiores a 10 pontos percentuais.

De forma semelhante, Sharma *et al.* (2024) integraram técnicas evolutivas, como a otimização por enxame de partículas (MOD-PSO), à extração de características e classificação automática, combinando *deep learning* e metaheurísticas em *pipelines* híbridos. Embora tenham relatado ganhos de acurácia em bases específicas (até 98,46%), sua abordagem também se manteve dependente de parametrização manual e validação restrita a conjuntos homogêneos, o que limita a generalização e acentua a sobrecarga computacional inerente aos processos iterativos de otimização.

A revisão desses trabalhos revela desafios característicos dos métodos metaheurísticos:

- **Dependência de parametrização manual:** Os algoritmos metaheurísticos introduzem novos hiperparâmetros que requerem ajuste específico para cada cenário, muitas vezes através de processos manuais e não automatizados.
- **Limitações na generalização inter-contextual:** Modelos baseados em metaheurísticas apresentam quedas significativas de desempenho (superiores a 10%) quando avaliados fora do contexto de treino, evidenciando dificuldades de adaptação entre diferentes domínios clínicos.
- **Sobrecarga computacional:** Embora as metaheurísticas possam reduzir a dimensionalidade final, o processo iterativo de otimização impõe custo computacional adicional considerável, podendo inviabilizar aplicações que exigem resposta rápida ou operação em tempo real.
- **Validação limitada:** A maioria dos estudos é validada em poucas bases de dados, não demonstrando adequadamente a capacidade de generalização através de múltiplos contextos clínicos independentes.

Estas limitações indicam que, embora possam alcançar performance competitiva em domínios específicos, os métodos metaheurísticos compartilham desafios fundamentais de adaptabilidade e eficiência operacional. A forte dependência de parametrizações manuais e o elevado custo computacional reduzem sua viabilidade em cenários clínicos reais. Neste trabalho, optou-se por uma alternativa distinta: um *pipeline* adaptativo no qual os pesos do *ensemble* são definidos automaticamente via validação cruzada estratificada, eliminando a necessidade de calibração manual e garantindo maior consistência entre diferentes domínios. Essa escolha representa uma resposta direta às fragilidades identificadas na literatura, alinhando simplicidade operacional a requisitos práticos de aplicabilidade clínica.

2.6 Abordagens Híbridas e Baseadas em *Ensemble*

As abordagens híbridas e baseadas em *ensemble* representam uma evolução natural na busca por maior desempenho e estabilidade na detecção de glaucoma. Essas estratégias buscam explorar a complementaridade entre múltiplos modelos ou técnicas, combinando suas saídas para obter uma decisão final mais precisa. Contudo, muitas implementações ainda carecem de mecanismos de integração adaptativa, limitando sua capacidade de generalização entre domínios heterogêneos.

Gómez-Valverde *et al.* (2019) propuseram o DENet, um *ensemble* fixo de quatro arquiteturas CNN distintas, que obteve melhoria de acurácia em bases públicas como DRISHTI-GS, RIM-ONE e ESPERANZA. Apesar do desempenho competitivo, a arquitetura apresenta rigidez significativa: a combinação de modelos é estática, com pesos fixos e sem capacidade de adaptação a novos contextos ou incorporação de novos extratores de características. De forma similar, MK e Kalra (2024) desenvolveram uma CNN customizada que alcança sensibilidade e acurácia superiores em múltiplas bases públicas, porém permanece dependente de ajustes manuais específicos para cada domínio, não integrando mecanismos de adaptação automática.

Mrad *et al.* (2022), já discutido na Seção 2.2, também contribui para esta categoria ao combinar CNNs compactas em *ensemble* para triagem móvel. No entanto, assim como os demais trabalhos, enfrenta limitações de ajuste manual e validação restrita, que comprometem a aplicação clínica em ambientes diversos.

A revisão dessas abordagens revela desafios comuns nos métodos de *ensemble* atuais:

- **Arquiteturas estáticas:** A maioria dos *ensembles* opera com combinações fixas de modelos e pesos predeterminados, sem capacidade de adaptação automática a diferentes

contextos clínicos ou distribuições de dados.

- **Dependência de intervenção manual:** A otimização das combinações e pesos normalmente exige ajuste manual e re-treinamento para cada novo cenário, processo que demanda expertise técnica significativa e inviabiliza implantação escalável.
- **Validação limitada entre domínios:** A performance é frequentemente reportada em poucas bases de dados com características similares, não demonstrando adequadamente a capacidade de generalização através de contextos clínicos verdadeiramente diversos.
- **Ausência de mecanismos dinâmicos e automatizados de adaptação:** Falta de estratégias automáticas para priorização dinâmica dos modelos mais eficazes em cada contexto específico, resultando em subotimização em domínios não vistos durante o treinamento.

Essas limitações comprometem a viabilidade clínica em larga escala. Neste trabalho, optou-se por uma alternativa distinta: um *ensemble* adaptativo em que os pesos dos classificadores são definidos automaticamente por validação cruzada estratificada. Essa estratégia elimina a necessidade de ajustes manuais, amplia a consistência interdomínio e assegura maior previsibilidade dos resultados, representando uma resposta direta às fragilidades identificadas na literatura.

2.7 Abordagens para Conjuntos de Dados Pequenos

A escassez de imagens médicas anotadas é um desafio recorrente no desenvolvimento de modelos para detecção automatizada de glaucoma, especialmente em contextos clínicos com poucas amostras e alta variabilidade de aquisição. Estratégias específicas têm sido desenvolvidas para mitigar o risco de *overfitting* e manter consistência de desempenho nestes cenários restritos, embora ainda apresentem limitações significativas quanto à generalização e aplicabilidade clínica.

Parashar e Agrawal (2021) propuseram um *pipeline* baseado em decomposição por *wavelet* (2D-T-EWT) e classificação MC-LS-SVM, validando em RIM-ONE e Harvard Dataverse V1. O método obteve acurácias de 92,47% e 93,65%, respectivamente, demonstrando o potencial de técnicas otimizadas para cenários com dados limitados. Contudo, a performance depende criticamente de seleção manual de atributos e não inclui validação cruzada entre múltiplas bases, deixando em aberto sua capacidade de generalização para domínios mais diversos.

De forma complementar, Mahalakshmi *et al.* (2024) investigaram uma abordagem baseada em CNN aplicada às bases ACRIMA e HRF, evidenciando desempenho consistente (acurácia

> 91,87%) mesmo com amostras limitadas, graças a estratégias intensivas de regularização e pré-processamento. Apesar dos resultados promissores, o método mantém dependência da qualidade das imagens e de etapas manuais de validação, ressaltando desafios comuns em cenários de dados restritos.

Trabalhos previamente discutidos — como Latif *et al.* (2022) (ODGNet), Juneja *et al.* (2020) (GC-NET) e Mrad *et al.* (2022) — também reportaram acurácias superiores a 94% em bases pequenas como *Drishti-GS*, RIM-ONE e HRF, empregando aumento agressivo de dados e arquiteturas compactas. Entretanto, estes trabalhos compartilham três lacunas metodológicas relevantes:

- **Generalização limitada:** Mesmo com alta acurácia intra-domínio, a transferência entre bases distintas resulta em quedas de 10 a 15 p.p., evidenciando que soluções eficazes em contextos específicos não mantêm desempenho consistente em domínios diversos.
- **Dependência crítica de pré-processamento:** A performance ótima está condicionada a estratégias detalhadas de segmentação ou aumento de dados, que são sensíveis à qualidade das imagens e nem sempre escaláveis para aplicações clínicas reais.
- **Ausência de adaptação dinâmica:** Os métodos utilizam configurações estáticas ou requerem reparametrização manual para cada novo cenário, tornando a aplicação clínica menos ágil e pouco sustentável em fluxos de trabalho reais.

A revisão destas abordagens revela que o desafio fundamental vai além do tamanho dos conjuntos de dados:

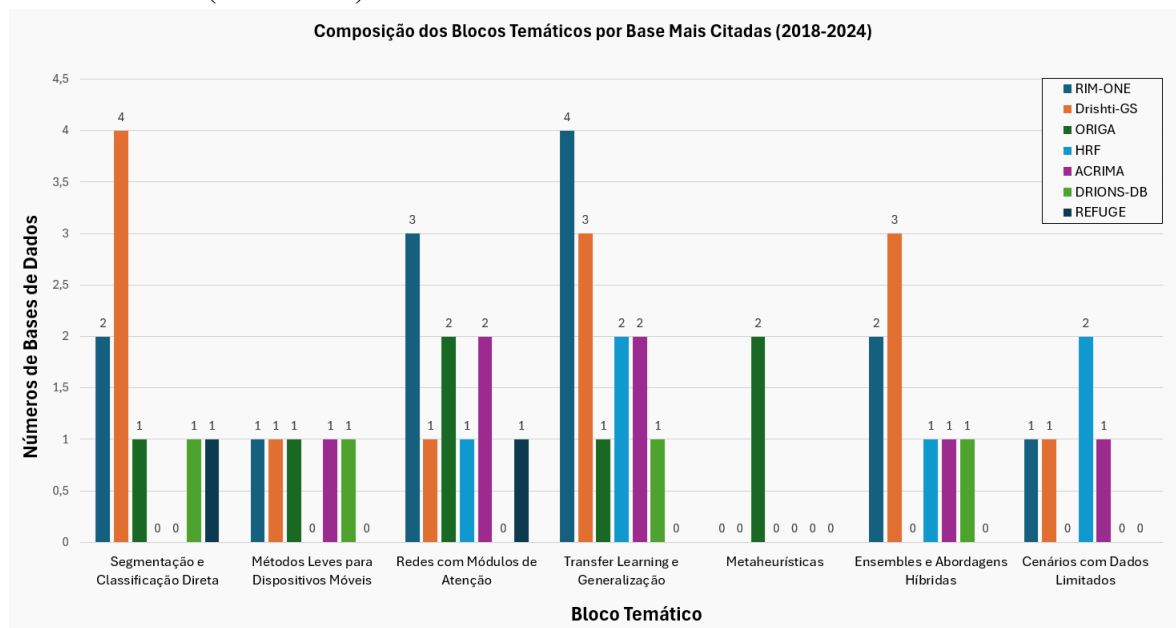
- **Foco excessivo em otimização intra-domínio:** As estratégias são majoritariamente voltadas para maximizar performance em bases específicas, negligenciando o desenvolvimento de mecanismos embutidos de generalização.
- **Soluções não transferíveis:** Técnicas que funcionam em um contexto específico (ex.: aumento de dados específico) frequentemente não se traduzem diretamente para outros domínios clínicos.
- **Validação inadequada:** A avaliação em poucas bases de dados (tipicamente 1–2) não permite inferir estabilidade real em cenários clínicos diversificados.

Estas limitações destacam a necessidade de abordagens que transcendam a otimização pontual para domínios específicos, priorizando arquiteturas adaptativas por construção, capazes de manter estabilidade. Soluções que incorporem mecanismos de generalização embutidos e validação multidomínio configuram uma direção necessária para avançar o estado da arte, tema aprofundado no próximo capítulo desta dissertação.

2.8 Síntese Crítica e Identificação de Lacunas

Os blocos anteriores discutiram em profundidade as principais abordagens para detecção automática de glaucoma em imagens de fundo de olho, destacando avanços e limitações das estratégias de segmentação, classificação direta, mecanismos de atenção e outras categorias metodológicas. Foram analisados 23 trabalhos (2018-2024), agrupados em sete categorias metodológicas: Métodos Baseados em Segmentação (4), Classificação Direta com CNNs (4), Abordagens com Mecanismos de Atenção (5), Métodos Leves para Dispositivos Móveis (3), Métodos Metaheurísticos (2), Abordagens Híbridas e *Ensemble* (2) e Abordagens para Conjuntos de Dados Pequenos (3). As bases de dados mais recorrentes foram RIM-ONE (13 ocorrências), *Drishti-GS* (12) e ORIGA (8), seguidas por ACRIMA (6), HRF (5), REFUGE (1) e DRIONS-DB (1), conforme ilustrado na Figura 1.

Figura 1 – Distribuição das bases de dados mais utilizadas nos trabalhos revisados (2018-2024).



Fonte: Autor (2024).

Cabe destacar que a soma total das ocorrências na Figura 1 é superior ao número de

artigos revisados (23), pois diversos trabalhos utilizaram múltiplas bases de dados em seus experimentos. Assim, cada contagem corresponde ao uso de uma base em um estudo, e não ao número total de artigos.

A análise consolidada dos 23 trabalhos permitiu identificar três limitações recorrentes e críticas no estado da arte:

- **Instabilidade interdomínio:** 20 de 23 estudos (87%) apresentaram variação de desempenho superior a 10 p.p. entre diferentes bases de dados, com quedas médias de 12,7 p.p. ($\pm 2,3$) em validação externa.
- **Dependência de ajustes manuais:** 18 de 23 estudos (78%) exigem reconfiguração específica e intervenção manual (e.g., *fine-tuning*) para cada novo contexto ou base de dados.
- **Validação restrita:** 18 de 23 estudos (78%) utilizaram três ou menos bases de dados públicas para validação, comprometendo a confiança na generalização dos resultados para cenários clínicos reais.

Estas limitações dão origem a três fenômenos observados:

1. **Paradoxo da Acurácia:** Valores de acurácia superiores a 95% concentram-se predominantemente em bases de dados pequenas e homogêneas (ex: RIM-ONE). Na presente revisão, não foram identificados métodos que mantivessem acurácia consistentemente superior a 90% quando avaliados de forma cruzada em cinco bases de dados heterogêneas.
2. **Efeito Dominó:** A validação restrita leva à superestimação do desempenho, que se traduz em quedas médias de 12,7 ($\pm 2,3$) p.p. em novos domínios, comprometendo a generalização. No conjunto analisado, a maior diferença de desempenho observada foi de 17,87 p.p. (ORIGA: 80,00% vs. RIM-ONE: 97,87%).
3. **Divisor Tecnológico:** Métodos não adaptativos (ex.: Shyamalee e Meedeniya (2022)) tendem a exigir reconfiguração manual por base de dados e apresentam quedas mais acentuadas de desempenho fora do domínio de treinamento (ex., *overfitting*), conforme documentado na coluna *Limitações* da Tabela 1.

Validação externa e avaliação multicêntrica são imperativas para a tradução clínica dessas tecnologias.

A acurácia reportada na literatura variou de 58,33% a 100%, com os valores mais elevados geralmente obtidos em conjuntos pequenos e controlados, seguidos por quedas expressivas quando os métodos são avaliados em domínios clínicos distintos.

Esta síntese crítica demonstra que, apesar dos avanços significativos em acurácia intra-domínio, o campo carece urgentemente de soluções consistentes que ofereçam:

1. desempenho estável entre domínios clínicos diversos;
2. autonomia operacional com mínima intervenção manual; e
3. validação abrangente em múltiplos contextos.

Estas lacunas justificam o desenvolvimento de uma abordagem metodológica concebida para superar tais limitações, conforme detalhado no Capítulo 4 (Materiais e Métodos). A arquitetura proposta será subsequentemente avaliada, no Capítulo 5 (Resultados), através de um protocolo rigoroso de validação multidomínio em cinco bases de dados públicas com perfis contrastantes, abrangendo as variações críticas de aquisição, prevalência e qualidade identificadas nesta revisão.

A análise crítica realizada anteriormente está consolidada na Tabela 1, que sintetiza as principais limitações de cada abordagem e posiciona esta dissertação frente ao estado da arte, evidenciando sua contribuição potencial.

Tabela 1 – Análise crítica comparativa de métodos para detecção automática de glaucoma.

Autor / Ano	Método	Base	Acc (%)	Limitações principais
Métodos Baseados em Segmentação				
Guo <i>et al.</i> (2020)	UNet++ + atributos + GBDT + ADASYN	ORIGA Drishti-GS1	83,70–84,30	• Segmentação crítica; • Variação > 10 p.p. em bases heterogêneas; • Pré-processamento intensivo.
Civit-Masot <i>et al.</i> (2020)	U-Net + MobileNetV2	RIM-ONE Drishti-GS1	86,00	• Segmentação crítica; • Validação limitada; • Generalização clínica limitada.
Shinde (2021)	U-Net + <i>Brightest Spot</i> + Classificadores	RIM-ONE Drishti-GS1 DRIONS-DB JSIEC DRIVE	≤99,10	• Segmentação crítica; • Dependência de qualidade.

Autor / Ano	Método	Base	Acc (%)	Limitações principais
Arsyan e Maki (2022)	Pré-proc (CLAHE + Morf. Mat.) + Momentos Inv. + RF/KNN	Drishti-GS1 REFUGE	58,33–90,00	• Dependência de qualidade; • Instável em heterogeneidade; • Vulnerável à desbalanceamento.
Métodos Leves para Dispositivos Móveis				
Guo <i>et al.</i> (2018)	Seg. DO/esc. + Parâm. clínicos (SVM/RF)	ORIGA	76,90	• Segmentação crítica; • Dependência de qualidade; • Instabilidade cross-dispositivo.
Devecioglu <i>et al.</i> (2021)	Self-ONNs	ACRIMA RIM-ONE ESOGU	$\leq 94,10$	• Validação limitada; • Falta integração clínica; • Generalização limitada.
Mrad <i>et al.</i> (2022)	CLAHE + black-hat + Otsu + MedianBlur + SVM	Drishti-GS1 DRIONS-DB RIAMP	$\leq 100,00$	• Instabilidade entre bases; • Ajustes manuais necessários; • Dependência de qualidade.
Abordagens com Mecanismos de Atenção				
Li <i>et al.</i> (2020)	AG-CNN (Attn. sup. + ResNet)	LAG RIM-ONE	85,20–96,20	• Validação limitada; • Queda em validação externa; • Alta complexidade.
Shyamalee e Meedeniya (2022)	Att. U-Net (InceptionV3 + VGG-19 + ResNet-50)	RIM-ONE ACRIMA	92,64–94,95	• Alta complexidade; • Segmentação crítica; • Generalização limitada.
Das, Nayak e Pachori (2024)	AES-Net (ESAM + adaptação espacial)	RFI, ORIGA HRF	$\leq 99,30$	• Alta complexidade; • Validação limitada; • Generalização limitada.
Das <i>et al.</i> (2024)	CDAM-Net (atenção cross-domain)	HDTV1, LMG RIM-ONE	$\leq 93,12$	• Alta complexidade; • Exige hardware avançado; • Inferência lenta.
Shyamalee <i>et al.</i> (2024)	Módulos de atenção + CNNs	RIM-ONE ACRIMA REFUGE Drishti-GS1 ORIGA	$\leq 98,97$	• Sensível a ruídos; • Hardware avançado; • Alta complexidade.
Classificação Direta com CNNs				

Autor / Ano	Método	Base	Acc (%)	Limitações principais
Diaz-Pinto <i>et al.</i> (2019)	CNNs pré-treinadas (VGG + InceptionV3 + ResNet-50 + Xception)	ACRIMA HRF Drishti-GS1 RIM-ONE sjchoi86-HRF	$\leq 80,00$	• Sensível à mudança de domínio; • Exige reajuste manual; • Desempenho inconsistente.
Juneja <i>et al.</i> (2020)	GC-NET (CNN customizada)	RIM-ONE Drishti-GS	97,51	• Tuning manual por base; • Generalização limitada; • Validação limitada.
Latif <i>et al.</i> (2022)	ODGNet (TL + localização)	ORIGA HRF DRIONS-DB DR-HAGIS RIM-ONE	$\leq 95,75$	• Config. específicas por cenário; • Estabilidade não validada; • Generalização limitada.
Velpula e Sharma (2023)	Resnet101 + NasNetMobile + NasNetLarge	ACRIMA RIM-ONE-V2 Drishti-GS1 FTVD HVD	$\leq 98,72$	• Variação > 10% entre contextos; • Desempenho inconsistente; • Generalização limitada.
Métodos Metaheurísticos				
Latif <i>et al.</i> (2023)	EGWO-SVM + SURF/HOG	ORIGA	94,00	• Parametrização manual; • Queda > 10 p.p. entre domínios; • Sobrecarga computacional.
Sharma <i>et al.</i> (2024)	CNN + MOD-PSO-ELM	G1020 ORIGA	97,80–98,46	• Parametrização manual; • Validação limitada; • Sobrecarga computacional.
Abordagens Híbridas e Baseadas em <i>Ensemble</i>				
Gómez-Valverde <i>et al.</i> (2019)	DENet	Drishti-GS RIM-ONE ESPERANZA	87,48	• Arquitetura rígida; • Sem adaptação automática; • Validação limitada.
MK e Kalra (2024)	CNN customizada	Drishti-GS1 sjchoi86-HRF RIMONE-DL ACRIMA	$\leq 98,60$	• Sem adaptação automática; • Validação limitada.
Abordagens para Conjuntos de Dados Pequenos				
Parashar e Agrawal (2021)	2D-T-EWT + MC-LS-SVM	Drishti-GS HRF LES-AV RIM-ONE	$\leq 92,47$	• Seleção manual de atributos; • Validação limitada; • Sem adaptação automática.

Autor / Ano	Método	Base	Acc (%)	Limitações principais
Mahalakshmi <i>et al.</i> (2024)	CNN	ACRIMA HRF	91,87–93,20	• Dependência da qualidade; • Validação limitada; • Generalização limitada.

Fonte: Autor (2024).

A literatura recente sobre detecção automatizada de glaucoma é rica em contribuições e variada em técnicas; ainda assim, persistem gargalos práticos:

1. dependência de dados rotulados e imagens de alta qualidade;
2. dificuldade de generalização para múltiplos domínios clínicos;
3. complexidade computacional elevada; e
4. sensibilidade a variações de aquisição.

Esses limites se manifestam nos três fenômenos discutidos nesta revisão — *Paradoxo da Acurácia, Efeito Dominó e Divisor Tecnológico* — reforçando a necessidade de soluções mais adaptativas.

Em resposta, este trabalho adota uma arquitetura híbrida adaptativa, que combina a extração de características por CNNs pré-treinadas com decisão baseada em *ensemble* ponderado, buscando reduzir a dependência de segmentação e aumentar a estabilidade interdomínios. A fim de contextualizar e sustentar essas escolhas, o próximo Capítulo 3 apresenta a Fundamentação Teórica, abordando desde conceitos clínicos do glaucoma até os princípios computacionais de processamento de imagens, aprendizado profundo e técnicas de *ensemble*, que compõem a base conceitual deste estudo.

3 FUNDAMENTAÇÃO TEÓRICA

O desenvolvimento de um sistema de detecção de glaucoma com capacidade de generalização frente a variações interdomínio exige uma base teórica que una a oftalmologia à ciência da computação. Este capítulo sintetiza essa base, guiando o leitor da manifestação clínica da doença na retina aos algoritmos concebidos para identificá-la. Inicialmente, são revisados os fundamentos clínicos — anatomia ocular, características do glaucoma e sua relevância epidemiológica. Em seguida, o foco se volta para as técnicas computacionais centrais ao método: os princípios de Redes Neurais Convolucionais (CNNs), as arquiteturas adotadas (ResNet-50, EfficientNet-B0 e MobileNetV2) e as estratégias de *Transfer Learning* e *Ensemble Learning*. A discussão destaca suas potencialidades e limitações especificamente no contexto de bases médicas heterogêneas e de tamanho reduzido. Por fim, este arcabouço conceitual serve de alicerce para a metodologia delineada no Capítulo 4 e para a análise dos resultados no Capítulo 5.

3.1 Estrutura Óptica do Olho Humano

Uma compreensão da óptica ocular é fundamental para projetistas de equipamentos de diagnóstico oftalmológico e sistemas ópticos visuais. Anatomicamente, o olho humano apresenta três camadas principais. A camada externa divide-se em duas partes: a córnea (anterior) e a esclera (posterior).

A córnea constitui a porção mais anterior do olho, situando-se anteriormente à íris e à pupila. Este tecido transparente e de formato aproximadamente esférico (raio de curvatura de aproximadamente $8mm$) representa a estrutura mais densamente innervada do organismo humano, com inervação sensorial majoritariamente proveniente do ramo oftálmico do nervo trigêmeo. Suas dimensões médias são aproximadamente $11,5mm$ no eixo horizontal e $10,5mm$ no vertical (WILLOUGHBY *et al.*, 2010).

A esclera consiste em um tecido fibroso, opaco e de coloração esbranquiçada, com função predominantemente protetora. Com formato aproximadamente esférico e diâmetro médio de $24mm$, seus centros de curvatura encontram-se deslocados em aproximadamente $5mm$ em relação aos centros de curvatura corneal (ATCHISON, 2018).

A camada intermediária, denominada trato uveal, compõe-se de três estruturas: íris (porção anterior), coroide (porção posterior) e corpo ciliar (porção intermediária). A íris desempenha funções ópticas cruciais no controle pupilar, enquanto o corpo ciliar atua nos processos de acomodação visual. Tanto o corpo ciliar quanto a coroide participam ativamente de processos

metabólicos de suporte (ATCHISON, 2018).

A camada interna corresponde à retina, extensão direta do sistema nervoso central que conecta-se ao encéfalo através do nervo óptico. A região interior do segmento posterior denomina-se fundo ocular, referindo-se à porção visível ao exame oftalmoscópico. Esta região inclui a retina propriamente dita, o disco óptico, e aspectos da coroide e esclera subjacentes (ATCHISON, 2018).

O espaço intraocular divide-se em três compartimentos principais:

1. Câmara anterior: entre córnea e íris, contendo humor aquoso;
2. Câmara posterior: entre íris, corpo ciliar e cristalino, também preenchida por humor aquoso; e
3. Câmara vítrea: entre cristalino e retina, preenchida pelo corpo vítreo (substância gelatinosa transparente).

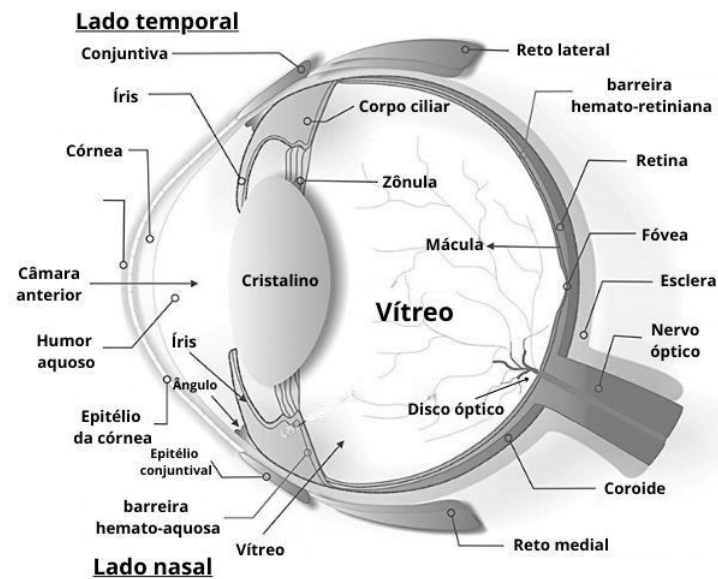
O cristalino posiciona-se imediatamente posterior à íris, conectando-se ao corpo ciliar através das fibras zonulares. A manutenção da pressão intraocular (PIO) em níveis superiores à pressão atmosférica é essencial para a manutenção da curvatura corneal e transparência dos meios oculares. Este equilíbrio pressórico é regulado pela produção de humor aquoso no corpo ciliar e sua drenagem através do ângulo da câmara anterior em direção ao canal de Schlemm e, subsequentemente, para o sistema venoso ocular (ATCHISON, 2018). A Figura 2 ilustra detalhadamente a anatomia ocular descrita.

3.2 Glaucoma

O glaucoma compreende um grupo de neuropatias ópticas progressivas crônicas, caracterizadas por alterações estruturais na cabeça do nervo óptico e na camada de fibras nervosas da retina (RNFL - *retinal nerve fiber layer*), conforme ilustrado na Figura 3. Essas alterações estão frequentemente associadas ao principal fator de risco modificável: a pressão intraocular (PIO) elevada.

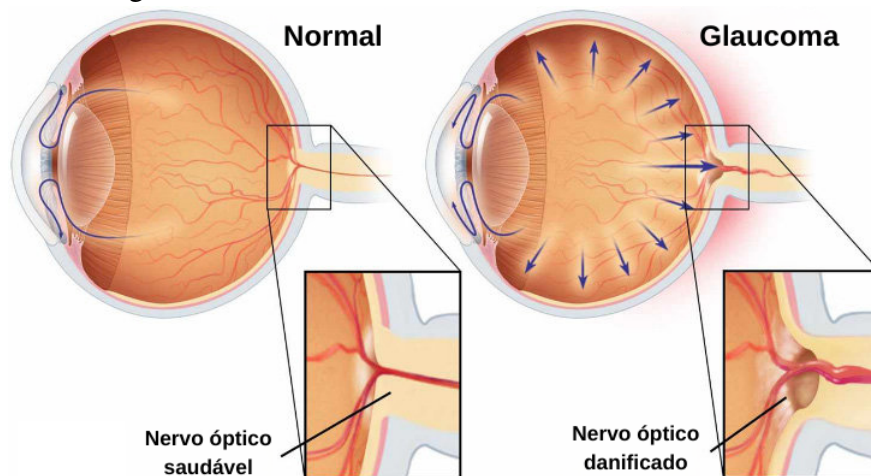
O termo *glaucoma* tem origem no período Homérico da Grécia Antiga (c. 1150 a.C.), derivando do adjetivo *glaukos* (verde ou azul brilhante) ou de *glaux* (coruja). Na literatura médica grega antiga, os termos *glaucoma* e *hipoquima* descreviam a mesma condição, atribuída a um suposto endurecimento (*pexis*) do um humor intraocular (LAIOS; MOSCHOS; ANDROUTSOS, 2016; LEFFLER *et al.*, 2015).

Figura 2 – Estrutura anatômica do olho humano.



Fonte: Adaptado de Willoughby *et al.* (2010).

Figura 3 – Comparação anatômica entre nervo óptico saudável e nervo óptico com dano glaucomatoso.



Fonte: Adaptado de Balakrishnan (2017).

Os principais fatores de risco para o desenvolvimento e progressão do glaucoma incluem:

- Pressão intraocular (PIO) elevada;
- Idade avançada;
- Histórico familiar e predisposição genética;
- Origem étnica (ex., maior prevalência em populações de origem africana e asiática);
- Hipotensão e hipertensão sistêmica;

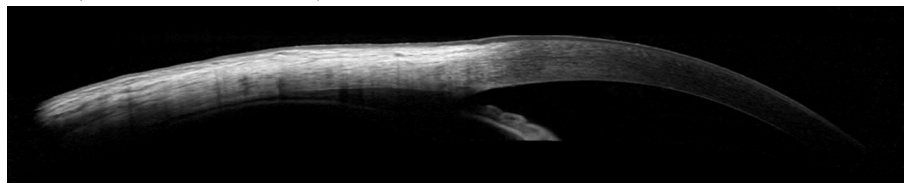
- Comorbidades como diabetes *mellitus*, enxaqueca e síndrome da apneia obstrutiva do sono.

A perda de fibras nervosas da RNFL resulta no afinamento da borda neural do disco óptico e, conseqüentemente, no aumento da razão escavação/disco (*Cup-to-Disc Ratio* - CDR) (Figura 3), culminando em defeitos característicos no campo visual (ALLISON; PATEL; ALABI, 2020; COAN *et al.*, 2023).

A visão é possibilitada pela transmissão de impulsos nervosos do nervo óptico ao cérebro. Danos significativos a essas fibras criam escotomas (pontos cegos) no campo visual, os quais, uma vez instalados, resultam em perda visual irreversível (KAUSU *et al.*, 2018). Pacientes nos estágios iniciais do glaucoma geralmente são assintomáticos, com a perda progressiva da visão periférica tornando-se evidente apenas em fases mais avançadas da doença (SHABBIR *et al.*, 2021). Um sinal observável em imagens da retina é a alteração no fluxo sanguíneo dos vasos retinianos (SHOBA; THERESE, 2020).

Morfologicamente, o glaucoma classifica-se em dois tipos principais: primário e secundário. Ambos se subdividem em glaucoma de ângulo aberto e de ângulo fechado, conforme a patogênese relacionada às estruturas de drenagem do humor aquoso no ângulo da câmara anterior, localizado entre a íris e a córnea (Figura 4) (ALLISON; PATEL; ALABI, 2020).

Figura 4 – Visualização do ângulo da câmara anterior e das estruturas de drenagem (canal de Schlemm) em um olho saudável.



Fonte: Asam *et al.* (2019).

3.2.1 Glaucoma primário de ângulo aberto (GPAA)

O GPAA é definido pela presença de um ângulo de câmara anterior aberto (avaliado por gonioscopia), associado a evidências de neuropatia óptica glaucomatosa e defeitos correspondentes no campo visual. Sua patogênese caracteriza-se pelo aumento da resistência ao escoamento do humor aquoso através do sistema trabecular no ângulo da câmara anterior.

O GPAA pode manifestar-se sob duas condições:

- **Glaucoma de Tensão Normal (*Normal-Tension Glaucoma* - NTG):** Ocorre com PIO

dentro da faixa estatística de normalidade (15–21 mmHg). Associa-se a perdas estruturais mais focais e defeitos funcionais centralizados, estando frequentemente ligado a fatores de risco vasculares como hipotensão sistêmica, enxaqueca e doença vascular periférica (JIANG *et al.*, 2023; VERNAZZA *et al.*, 2020; WOSTYN; KILLER, 2023).

- **Glaucoma de Alta Tensão (*High-Tension Glaucoma* - HTG):** Caracteriza-se por PIO consistentemente superior a 21 mmHg, apresentando geralmente um padrão de perda estrutural mais difusa e funcional periférica (THONGINNETRA *et al.*, 2010).

Wostyn e Killer (2023) propõem que a hipotensão sistêmica, um fator de risco conhecido para o NTG, poderia reduzir a produção de Líquido Cefalorraquidiano (LCR), impactando negativamente o transporte glinfático e a depuração de metabólitos no nervo óptico.

3.2.2 *Glaucoma primário de ângulo fechado (GPAF)*

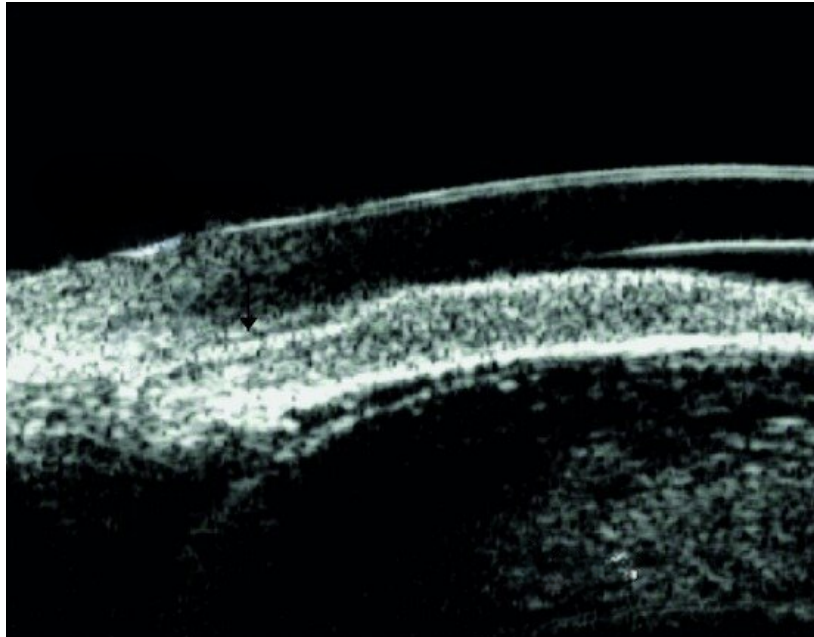
O GPAF distingue-se do GPAA pela sua patogênese central: o fechamento físico do ângulo da câmara anterior, que obstrui o acesso do humor aquoso à malha trabecular, resultando em elevação da PIO (SUN *et al.*, 2017). Este fechamento pode ser por aposição (contato reversível) ou sinequial (aderência permanente) da íris à córnea (COSTA *et al.*, 2020).

A classificação do GPAF baseia-se na cronicidade e sintomatologia do fechamento angular:

- **GPAF Agudo:** Apresenta início súbito, com contato iridotrabecular $\geq 180^\circ$ e PIO elevada, constituindo uma emergência oftalmológica.
- **GPAF Intermitente:** Episódios transitórios de fechamento angular com sintomas leves ou ausentes.
- **GPAF Crônico:** Fechamento angular progressivo ($\geq 180^\circ$) com PIO elevada, formação de Sinéquias Anteriores Periféricas (SAP) e dano glaucomatoso instalado (WRIGHT *et al.*, 2016).

Os mecanismos anatômicos predisponentes incluem câmara anterior rasa, cristalino de grande volume ou posicionado anteriormente, que empurram o diafragma íris-cristalino para frente, promovendo o fechamento do ângulo (Figura 5) (MARCHINI *et al.*, 2015).

Figura 5 – Imagem de biomicroscopia ultrassônica (UBM) demonstrando o mecanismo facomórfico de fechamento do ângulo por bloqueio pupilar e deslocamento anterior do complexo íris-cristalino.



Fonte: Dorairaj, Tsai e Grippo (2012).

3.2.3 *Glaucoma Secundário*

O glaucoma secundário distingue-se por ser consequência de uma condição ocular ou sistêmica subjacente, ou de intervenções farmacológicas. Embora menos prevalente que o GPAA, frequentemente apresenta maior agressividade e resistência ao tratamento convencional, conferindo um risco elevado de cegueira (NOLAN; YIP, 2015).

Pode manifestar-se tanto sob a forma de ângulo aberto (ex., síndromes pseudoesfoliativa ou de dispersão pigmentar, glaucoma neovascular, induzido por corticoides) quanto de ângulo fechado (ex., glaucoma facomórfico, facolítico, neovascular) (BERLIN *et al.*, 2018).

3.2.4 *Epidemiologia*

O glaucoma mantém-se como uma das principais causas de cegueira irreversível em escala global, rivalizando com a degeneração macular relacionada à idade (DMRI) (KELLY *et al.*, 2020).

Tham *et al.* (2014) estimaram a prevalência global de glaucoma em 3,54% na população entre 40 e 80 anos, projetando um aumento de 64,3 milhões de casos (2013) para 111,8 milhões em 2040. A distribuição é heterogênea:

- **África:** Apresenta a maior prevalência regional. Estudos como o de Kyari *et al.* (2015)

na Nigéria indicam que 1 em cada 20 adultos acima de 40 anos é portador da doença, majoritariamente glaucoma de ângulo aberto não diagnosticado.

- **Ásia:** Responde por aproximadamente 60% dos casos globais, concentrando 53,4% dos casos de GPAA e 76,7% dos casos de GPAF do mundo.

Em países ocidentais, como o Reino Unido, o GPAA afeta cerca de 2% da população acima de 40 anos, enquanto o GPAF é significativamente menos comum (0,17%) (IMRIE; TATHAM, 2016). O envelhecimento demográfico e as disparidades étnicas são os principais impulsionadores dessas projeções (KELLY *et al.*, 2020).

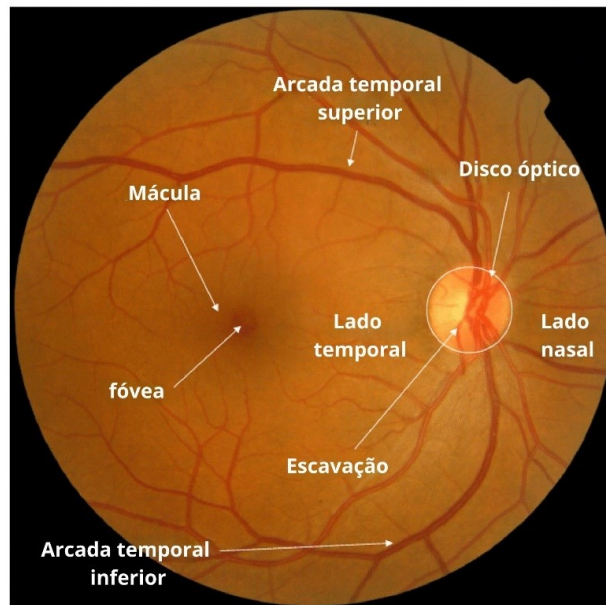
3.3 Imagens de Fundo de Retina

A degeneração do nervo óptico causada pelo glaucoma é irreversível. Uma vez que o nervo óptico é danificado, a perda de visão resultante não pode ser revertida. Por isso, a identificação da doença em seus estágios iniciais é crucial para impedir a progressão para perda visual permanente (SABA *et al.*, 2018). A avaliação do glaucoma depende da capacidade de detectar alterações patológicas no segmento posterior do olho, utilizando instrumentos ópticos em um procedimento denominado oftalmoscopia ou fundoscopia. O desenvolvimento de técnicas de registro imagético aprimorou esse processo, permitindo a documentação de características patológicas e a identificação de nuances morfológicas que podem escapar à observação clínica direta durante a biomicroscopia (KEANE; SADDA, 2014).

A retinografia (imagem do fundo) consiste na utilização da luz refletida para criar uma representação bidimensional (2D) dos tecidos retinianos, que são intrinsicamente semitransparentes e tridimensionais (3D). Esta representação projeta no plano da imagem pontos de referência anatômicos críticos, como o disco óptico (DO), a escavação, a mácula, a fóvea e, de forma crucial, o sistema vascular retinal (Figura 6) (ABRÀMOFF; KAY, 2013). A qualidade técnica da imagem retinal é frequentemente comprometida por artefatos como reflexos especulares da córnea e do cristalino. A degradação da imagem por opacidades do cristalino (catarata) constitui uma causa primária de iluminação não uniforme e desfocagem fotográfica (MITRA *et al.*, 2018).

Os sistemas convencionais de imagem de fundo oferecem um campo de visão que varia entre 30 e 60 graus do polo posterior da retina. Contudo, as modalidades avançadas contemporâneas suportam capacidades de campo ampliado (60 a 100 graus) e campo ultra-amplio (110 a 220 graus), permitindo uma avaliação mais abrangente da retina periférica (FOGEL-LEVIN *et al.*, 2022).

Figura 6 – Retinografia de fundo destacando disco óptico, mácula, fóvea e vasculatura.



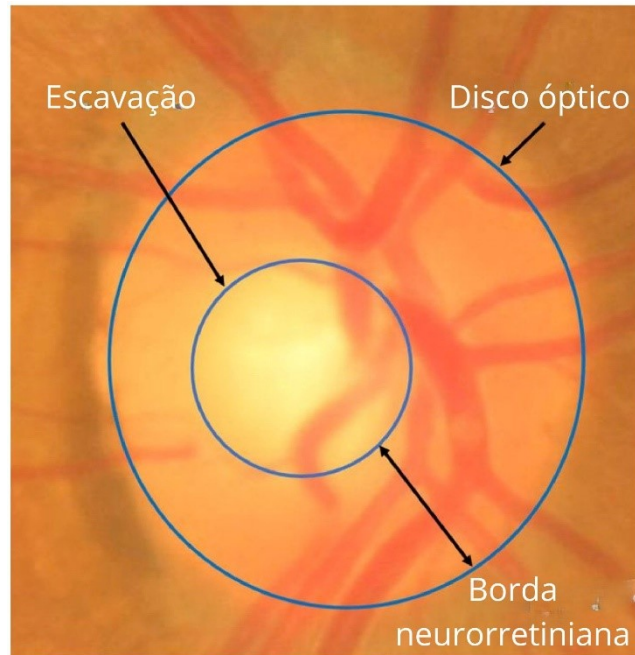
Fonte: Martins, Cardoso e Soares (2020).

A aquisição de imagens retinográficas de forma simplificada e de baixo custo apresenta um potencial significativo para a implementação de programas de rastreio em larga escala e para análises estatísticas populacionais. Diversos estudos clínicos demonstram que biomarcadores vasculares retinianos, como a tortuosidade e o calibre vascular, estão associados ao desenvolvimento de retinopatia diabética, glaucoma, hipertensão e outras patologias cardiovasculares (HUANG; DASHTBOZORG; ROMENY, 2018).

O disco óptico (DO), correspondente ao ponto cego da retina por carecer de fotorreceptores, é a região onde milhões de fibras nervosas da retina convergem para formar o nervo óptico. Morfologicamente, o DO é composto por três regiões principais: uma área central deprimida e de coloração esbranquiçada, denominada escavação; uma zona periférica em forma de anel, designada como borda neurorretiniana; e as fibras nervosas que constituem o nervo óptico (Figura 7) (KUMAR *et al.*, 2019).

O diagnóstico do glaucoma fundamenta-se predominantemente na fundoscopia, que permite a avaliação do DO e da RNFL. As alterações características do glaucoma incluem: perda de tecido na borda neurorretiniana, ampliação da escavação do nervo óptico, assimetria significativa das escavações entre os dois olhos, hemorragias splinter na margem do DO e defeitos localizados ou difusos da RNFL. A quantificação do tamanho e da morfologia do DO, da sua escavação e da borda neurorretiniana é, portanto, fundamental para a detecção precoce e monitoramento da progressão glaucomatosa (SCHUSTER *et al.*, 2020).

Figura 7 – Anatomia do disco óptico (DO) destacando a escavação central e a borda neurorretiniana circundante.



Fonte: Martins, Cardoso e Soares (2020).

A CDR é uma medida quantitativa fundamental na avaliação oftalmológica da saúde do nervo óptico, particularmente no contexto glaucomatoso. É calculada pela comparação entre o diâmetro vertical da escavação (DVE) e o diâmetro vertical do disco (DVD) (Equação 3.1).

$$CDR = \frac{DVE}{DVD} \quad (3.1)$$

Em indivíduos saudáveis, a CDR geralmente não excede 0,3 (30%). Valores acima deste limiar sugerem suspeita ou presença de glaucoma, com a gravidade da doença aumentando proporcionalmente ao valor da CDR (Tabela 2) (SABA *et al.*, 2018).

Tabela 2 – Classificação da gravidade do glaucoma com base na Razão Escavação/-Disco (CDR).

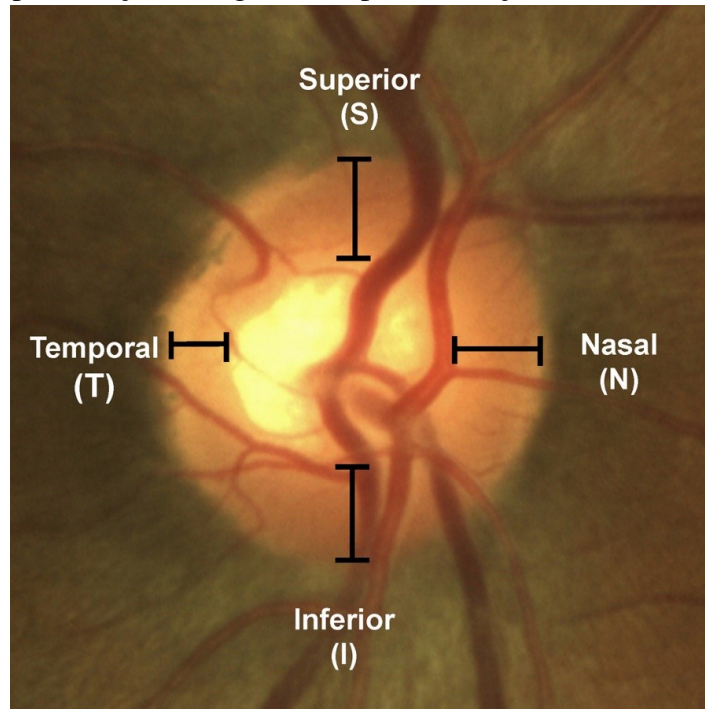
Estágio do glaucoma	Intervalo de CDR
Sem glaucoma	$< 0,3$
Leve	$0,3 \leq CDR < 0,5$
Moderado	$0,5 \leq CDR < 0,8$
Severo	$\geq 0,8$

Fonte: Adaptado de Hagiwara *et al.* (2018).

A regra ISNT, um acrônimo para Inferior (I) > Superior (S) > Nasal (N) > Temporal (T), estabelece um princípio clínico para avaliar a normalidade da espessura da borda neurorretiniana em cada quadrante do DO. Um padrão de espessura que se desvia significativamente desta regra

(ex., afinamento predominante nos quadrantes inferiores e superior) é altamente sugestivo de dano glaucomatoso (KUMAR *et al.*, 2019; MORGAN *et al.*, 2012; POON *et al.*, 2017). Dado que a perda da borda neurorretiniana é uma característica patognomônica do glaucoma, pacientes com desvios em relação à regra ISNT requerem acompanhamento clínico mais rigoroso. A Figura 8 ilustra a aplicação prática desta regra.

Figura 8 – Representação da regra ISNT para avaliação da borda neurorretiniana.



Fonte: Poon *et al.* (2017).

3.3.1 Aquisição de Imagens de Fundo de Retina

A aquisição de imagens de fundo de retina refere-se ao processo de captura de representações digitais bidimensionais (2D) da superfície tridimensional (3D) da retina. Este processo baseia-se em três componentes essenciais: um sistema óptico para concentrar a energia luminosa, um objeto de interesse (a retina) que a reflete, e um sensor que mede essa energia refletida para formar a imagem digital. A interação entre esses componentes possibilita diferentes métodos de aquisição, com resultados adequados ao uso clínico (MISHRA; KUMAR; SHUKLA, 2017; THAKUR; VENU; GURUSAMY, 2023).

Uma câmera de fundo de olho (Figura 9) funciona essencialmente como um microscópio de baixa potência, projetado para capturar as estruturas internas da retina — incluindo o disco óptico (DO), a mácula, a fóvea e a vasculatura — através do meio refrativo do olho, onde a

iluminação e a captação da luz refletida se dão pela pupila (MUKHERJEE; SENGUPTA, 2024; MISHRA; TRIPATHY, 2023).

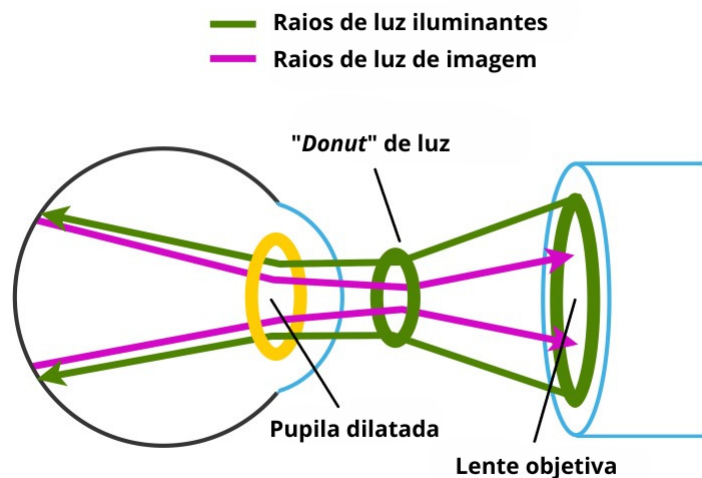
Figura 9 – Exemplos de câmeras de fundo de olho automatizadas utilizadas na aquisição de retinografias.



Fonte: Mukherjee e Sengupta (2024).

Na fotografia colorida convencional, filtros de contraste são frequentemente empregados para modificar o espectro luminoso, melhorando assim a visibilidade de estruturas específicas (BERNARDES; SERRANHO; LOBO, 2011). O sistema óptico tradicional, ilustrado na Figura 10, baseia-se no princípio da oftalmoscopia indireta monocular. Utiliza um conjunto de componentes ópticos (lentes, divisores de feixe, espelhos) que direcionam a luz através da pupila, coletam os reflexos da retina e projetam a imagem formada no sensor digital (TRAN *et al.*, 2012).

Figura 10 – Esquema óptico ilustrando o percurso da luz na iluminação da retina e na formação da imagem pela pupila.



Fonte: Mishra e Tripathy (2023).

Nesta configuração, uma fonte de iluminação contínua com um perfil em forma de anel (do tipo *donut*) é refletida por um espelho com uma abertura central, penetrando no olho pela

córnea.

A imagem forma-se quando os feixes de iluminação, dirigidos à borda pupilar, e os feixes de imagem, originados no centro pupilar, alinham-se opticamente. Esse alinhamento permite que a luz retroespalhada pela retina seja transmitida pela região central não iluminada do anel (MISHRA; TRIPATHY, 2023; PANWAR *et al.*, 2016).

Este processo de aquisição, contudo, varia significativamente entre diferentes fabricantes de equipamentos e protocolos clínicos, gerando uma heterogeneidade crítica que se reflete nos bancos de dados públicos de imagens. Como sintetizado na Tabela 3, destacam-se três dimensões primárias de variabilidade:

- **Perfil de Iluminação:** diferenças nos espectros e padrões de luz entre fabricantes;
- **Resolução Espacial:** varia de 768×584 (DRIVE) a 4752×3168 (DR-HAGIS);
- **Campo de Visão:** abrange desde 10° (HDV1) até 60° (HRF).

Essa diversidade técnica não deve ser vista como limitação, mas como reflexo fiel dos desafios encontrados em cenários clínicos reais, onde os sistemas de diagnóstico devem operar com:

- Equipamentos de diferentes gerações e fabricantes;
- Protocolos de aquisição não padronizados entre instituições de saúde;
- Variações intrínsecas na anatomia ocular entre pacientes.

Esta variabilidade inerente, associada à ausência de uma padronização global na aquisição de imagens de fundo de olho, impõe desafios significativos à generalização de modelos de aprendizagem profunda. Para enfrentar estes desafios, o *pipeline* metodológico proposto nesta dissertação incorpora estratégias específicas:

1. *Data augmentation* avançado para simular variações de iluminação e ângulos de aquisição;
2. Redimensionamento padronizado para dimensões uniformes (160×160 pixels);
3. Utilização de Redes Neurais Convolucionais (CNNs) pré-treinadas em grandes conjuntos de dados heterogêneos;
4. Técnicas de *transfer learning* para adaptação a domínios específicos; e

Tabela 3 – Resumo da heterogeneidade técnica em bancos de dados públicos de imagens de retina.

Referências	Bancos de Dados	Resolução	Campo de Visão	Aquisição
Ahn <i>et al.</i> (2018)	HDV1	240 × 240	10°	AFC 330
Bajwa <i>et al.</i> (2020)	G1020	1944 × 2108 2426 × 3007	45°	—
Carmona <i>et al.</i> (2008)	DRIONS-DB	600 × 400	—	Scanner HP-PhotoSmart-S20
Das, Nayak e Pachori (2023)	LMG	224 × 224	—	AFC 330
Devecioglu <i>et al.</i> (2021)	ESOGU	—	—	Kowa nonmyd alpha-DIII
Diaz-Pinto <i>et al.</i> (2019)	ACRIMA	—	35°	Topcon TRC
Fumero <i>et al.</i> (2011)	RIM-ONE	—	51°	Nidek AFC-210 EOS 5D Mark II
Gómez-Valverde <i>et al.</i> (2019)	ESPERANZA (Privado)	1024 × 680	45°	Nidek AFC-210
Holm <i>et al.</i> (2017)	DR-HAGIS	4752 × 3168 3456 × 2304 3216 × 2136 2896 × 1944 2816 × 1880	45°	Canon CR DGi Topcon TRC-NW6s Topcon TRC-NW8
Li <i>et al.</i> (2020)	LAG	582 × 597 3456 × 5184	—	Topcon Canon Carl Zeiss
Odstreilik <i>et al.</i> (2013)	HRF	3504 × 2336	60°	Canon CF-60 UVi Canon EOS-20D
Orlando <i>et al.</i> (2020)	REFUGE	2124 × 2056 1634 × 1634	—	Zeiss Visucam 500 Canon CR-2
Raghavendra <i>et al.</i> (2019)	RFI	1024 × 1024	—	Zeiss FF 450
Sivaswamy <i>et al.</i> (2014)	Drishti-GS	2996 × 1944	30°	—
Staal <i>et al.</i> (2004)	DRIVE	768 × 584	45°	Canon CR5
Zhang <i>et al.</i> (2010)	ORIGA	—	—	—

Fonte: Autor (2024).

5. Uma arquitetura *ensemble* que combina decisões de múltiplos classificadores.

Esta abordagem integrada permite ao sistema proposto manter precisão diagnóstica mesmo diante da heterogeneidade técnica documentada, evidenciando sua capacidade de adaptação a diferentes condições clínicas.

3.4 Pré-processamento de Imagens Digitais

O pré-processamento em imagens de fundo de olho pode ser formalizado como uma transformação $T : \mathcal{I} \rightarrow \mathcal{X}$, onde $\mathcal{I}$ representa o conjunto de imagens brutas e $\mathcal{X}$ o espaço de entrada das redes neurais convolucionais (CNNs). Para CNNs profundas aplicadas à detecção de glaucoma, o pré-processamento deve:

1. Preservar padrões estruturais críticos, tais como a conectividade vascular e a razão escavação/disco (CDR) (ELMOUFIDI *et al.*, 2022; GUPTA; GARG; AGARWAL, 2022).
2. Reduzir variações não informativas de escala, iluminação e contraste, frequentemente observadas em bases heterogêneas (JOSHI *et al.*, 2022; SANGHAVI; KURHEKAR, 2024).
3. Garantir compatibilidade dimensional (tamanho das imagens) e numérica (faixa de valores de intensidade) com os modelos pré-treinados utilizados (EfficientNet-B0, MobileNetV2 e ResNet-50) (ARCHANA; JEEVARAJ, 2024).

Esses requisitos fundamentam as escolhas adotadas neste trabalho, incluindo o redimensionamento para 160×160 *pixels*, a normalização dos valores de intensidade e a aplicação de estratégias específicas de *data augmentation*, detalhadas nas subseções seguintes.

3.4.1 Compatibilidade com as Arquiteturas Empregadas

As *stem layers* das CNNs modernas desempenham função análoga à de filtros clássicos de pré-processamento, realizando suavização adaptativa e padronização inicial da informação visual de forma aprendida (ELMOUFIDI *et al.*, 2022; ARCHANA; JEEVARAJ, 2024). Diferentes arquiteturas implementam essa camada inicial com estratégias distintas, conforme detalhado na Tabela 4, cada uma equilibrando custo computacional e capacidade de extração de características.

Tabela 4 – Características das *Stem Layers* nas arquiteturas avaliadas

Arquitetura	Operação Inicial	Recepção Inicial	Impacto no Pré-processamento
ResNet-50	Conv 7×7 + MaxPool	35×35 px	Alta suavização; reduz ruídos de alta frequência.
EfficientNet-B0	Conv 3×3	31×31 px	Foco em economia de parâmetros.
MobileNetV2	Conv 3×3 separável	27×27 px	Máxima eficiência computacional (FLOPs reduzidos).

Fonte: Autor (2024).

A configuração inicial da ResNet-50, apresentada nos dados da Tabela 4, consiste em uma convolução 7×7 seguida de *max pooling* que atua como filtro eficaz de suavização espacial, reduzindo ruídos de alta frequência (SANGHAVI; KURHEKAR, 2024). Em contraste, a EfficientNet-B0 prioriza a economia de parâmetros desde sua primeira convolução (3×3) (TAN; LE, 2019), e a MobileNetV2 emprega convoluções separáveis em profundidade, reduzindo significativamente os FLOPs sem perda relevante de detalhes anatômicos finos (??).

Essa diversidade de abordagens no nível arquitetural justifica a adoção de um pré-processamento minimalista. Investir em operações complexas de realce (ex., CLAHE) ou filtragem poderia, na prática, interferir negativamente ou redundar em operações já otimizadas realizadas pelas próprias *stem layers* das CNNs, que são calibradas para extrair informações maximamente discriminativas (TAN; LE, 2019). Dessa forma, o pipeline adotado prioriza a compatibilidade e a generalização, valendo-se da capacidade intrínseca dessas redes de aprender representações consistentes a partir de uma entrada minimamente processada.

3.4.2 Implementação do Pipeline Minimalista

Neste trabalho, foram aplicadas apenas duas operações, priorizando a eficiência computacional e a preservação de invariantes clínicos:

- **Redimensionamento** para 160×160 pixels. Esta dimensão foi selecionada por garantir compatibilidade com as *stem layers* de todas as CNNs utilizadas (Tabela 5) e, principalmente, por ser suficiente para preservar padrões morfológicos críticos, como a razão escavação/disco (CDR), especialmente para valores superiores a 0,3, limiar clinicamente relevante para o diagnóstico de glaucoma (ELMOUFIDI *et al.*, 2022). Este ajuste reduz consideravelmente o custo computacional, mantendo desempenho comparável ao de resoluções maiores em tarefas de classificação médica (MAYYA *et al.*, 2023).
- **Normalização** dos valores de pixel para a faixa $[0, 1]$, conforme definido pela Equação 3.2, onde I representa a imagem de entrada com valores na faixa $[0, 255]$. Esta operação assegura estabilidade numérica durante o treinamento e compatibilidade com as distribuições estatísticas do conjunto ImageNet, no qual as CNNs foram pré-treinadas (ARCHANA; JEEVARAJ, 2024).

$$I_{\text{norm}} = \frac{I}{255} \quad (3.2)$$

Tabela 5 – Parâmetros de compatibilidade arquitetural para o redimensionamento 160×160 pixels.

Backbone	Kernel da 1ª Camada	FLOPs (160×160)	Ganho vs. 224×224
ResNet-50	35×35	3,8G	49%
EfficientNet-B0	31×31	0,39G	52%
MobileNetV2	27×27	0,22G	61%

Fonte: Autor (2024).

O redimensionamento para 160×160 pixels, conforme mostrado na Tabela 5, proporciona ganhos substanciais de eficiência, com reduções de até 61% nas operações de ponto flutuante (FLOPs) em comparação com a resolução tradicional de 224×224 pixels, sem comprometer a capacidade de recepção inicial das redes.

3.4.3 Técnicas de Pré-processamento na Literatura e Razões para Exclusão

A seleção de um *pipeline* de pré-processamento é uma decisão estratégica que deve balancear potenciais ganhos de contraste em relação aos riscos de perda de informação, introdução de artefatos e redução da capacidade de generalização. Conforme resumido na Tabela 6, diversas técnicas frequentemente aplicadas na literatura para análise de imagens de fundo de olho foram deliberadamente excluídas neste trabalho. Essa decisão pautou-se no risco de comprometer a integridade de padrões morfológicos sutis, cruciais para o diagnóstico de glaucoma, e na priorização de uma abordagem estável e generalizável em diferentes bases heterogêneas.

Conforme evidenciado na Tabela 6, a análise crítica fundamenta a opção por um *pipeline* minimalista. Técnicas intensivas de realce ou filtragem, embora capazes de melhorar o contraste global, tendem a operar de forma indiscriminada, podendo suprimir sinais clinicamente relevantes ou exacerbar variações interbases relacionadas a equipamentos e protocolos de aquisição. A estratégia adotada, portanto, foi confiar na capacidade das CNNs modernas, pré-treinadas e calibradas, de aprender automaticamente representações estáveis e discriminativas a partir dos dados em seu estado original, reduzindo a necessidade de intervenções manuais que poderiam introduzir vieses no modelo.

3.4.4 Justificativa Estratégica

A decisão por um *pipeline* enxuto baseou-se em três pilares:

1. **Eficiência computacional:** redução de até 61% nos FLOPs em relação às imagens 224×224 , acelerando o treinamento e a inferência. (MAYYA *et al.*, 2023).

Tabela 6 – Fundamentação para exclusão de técnicas de pré-processamento

Técnica	Limitação Técnica	Impacto Potencial no Diagnóstico	Categoria do Impacto
CLAHE	Realce artificial em áreas de baixo contraste (GUPTA; GARG; AGARWAL, 2022).	Pode criar artefatos que confundem a CNN, gerando falsos positivos/negativos em bases heterogêneas (ELMOUFIDI <i>et al.</i> , 2022).	Artefato
Filtro Gaussiano	Suavização excessiva de bordas e detalhes finos (SANGHAVI; KURHEKAR, 2024).	Reduz a nitidez de microestruturas da escavação do DO e da camada de fibras nervosas (RNFL) (ARCHANA; JEEVARAJ, 2024).	Perda de Informação
Escala de Cinza	Perda irreversível de informação cromática (SANGHAVI; KURHEKAR, 2024).	Compromete a diferenciação de tecidos e a identificação de hemorragias ou outros sinais vasculares (ELMOUFIDI <i>et al.</i> , 2022).	Perda de Informação
Segmentação ROI	Depende criticamente da acurácia de um modelo de segmentação inicial (VALDEZ-RODRÍGUEZ; FELIPE-RIVERÓN; CALVO, 2021).	Erros de segmentação do DO são propagados e amplificados, degradando irreversivelmente a classificação final (ELMOUFIDI <i>et al.</i> , 2022).	Propagação de Erro

Fonte: Autor (2024).

- 2. Generalização entre bases:** minimiza a dependência de transformações que possam superajustar o modelo a características específicas de uma base (ELMOUFIDI *et al.*, 2022; JOSHI *et al.*, 2022).
- 3. Preservação diagnóstica:** mantém invariantes clínicos como a CDR e a morfologia vascular (SANGHAVI; KURHEKAR, 2024; ARCHANA; JEEVARAJ, 2024).

Assim, o pré-processamento adotado fornece a compatibilidade necessária com as arquiteturas escolhidas e mantém a integridade diagnóstica das imagens, alinhando-se às evidências recentes de que CNNs pré-treinadas podem extrair representações discriminativas mesmo a partir de imagens com transformações mínimas (TALEBI; MILANFAR, 2021; ELMOUFIDI *et al.*, 2022).

Os resultados experimentais do Capítulo 5 (Resultados) confirmam a eficácia do pré-processamento minimalista adotado. Mesmo sem etapas adicionais como equalização de histograma, filtragem ou segmentação, o modelo obteve acurácia de 97,87% (RIM-ONE) e AUC de 0,996 (ACRIMA). No conjunto TOTAL (junção das cinco bases), manteve desempenho consistente (AUC = 0,954) e métricas estáveis, indicando que o redimensionamento para 160×160

pixels e a normalização em $[0, 1]$ preservam invariantes clínicos como a CDR e a morfologia vascular, além de garantir compatibilidade com as *stem layers* das arquiteturas EfficientNet-B0, MobileNetV2 e ResNet-50. Esses achados validam, na prática, a opção por um *pipeline* enxuto, que alia eficiência computacional, preservação diagnóstica e alto desempenho.

3.4.5 Data Augmentation

Uma limitação recorrente no treinamento de modelos de aprendizado profundo, especialmente em bases médicas, é a escassez de amostras anotadas e o desequilíbrio entre classes (MIKOŁAJCZYK; GROCHOWSKI, 2018). No contexto da detecção de glaucoma, esses fatores são agravados pela heterogeneidade dos dispositivos de aquisição, variabilidade anatômica entre pacientes e diferenças na qualidade de imagem (ARCHANA; JEEVARAJ, 2024).

O *data augmentation* consiste em aplicar transformações controladas às imagens de treinamento, gerando amostras sintéticas que preservam o rótulo original, mas introduzem variações às quais o modelo deve ser invariante (MUMUNI; MUMUNI, 2022; SHORTEN; KHOSHGOFTAAR; FURHT, 2021). Estudos recentes demonstram que essa estratégia aumenta a capacidade de generalização sem necessidade de expandir fisicamente a base (ELMOUFIDI *et al.*, 2022; GUPTA; GARG; AGARWAL, 2022; JOSHI *et al.*, 2022; SANGHAVI; KURHEKAR, 2024).

As variações mais comuns incluem:

- **Geométricas:** rotações moderadas ($\pm 15^\circ$), espelhamento horizontal/vertical e translações, preservando as proporções e evitando distorções da morfologia do DO.
- **Escalação e recorte:** *zoom-in* controlado (até 10%) e recorte central (*cropping*), preservando a CDR e demais relações anatômicas.
- **Fotométricas:** variações de brilho, contraste e saturação para simular diferenças de iluminação e processamento entre equipamentos (ALOMAR; AYSEL; CAI, 2023).

Neste trabalho, adotou-se uma estratégia de *data augmentation* conservadora, evitando transformações que pudessem alterar características clínicas sutis, como micro-hemorragias ou contornos finos da escavação. Essa abordagem mantém a integridade diagnóstica e está alinhada ao pré-processamento minimalista descrito anteriormente, fornecendo diversidade suficiente para melhorar a generalização do modelo sem comprometer a fidelidade anatômica.

3.5 Fundamentos de Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados na organização e no funcionamento das redes neuronais biológicas, projetados para realizar tarefas de aprendizado e reconhecimento de padrões (SUZUKI, 2011).

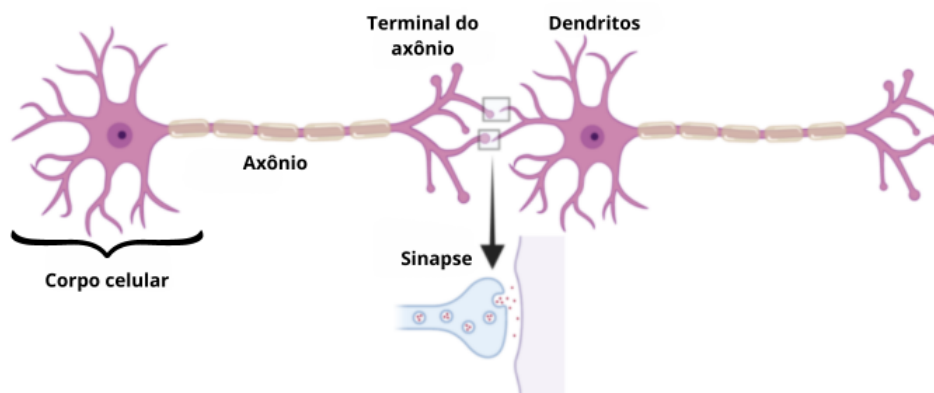
A pesquisa na área teve início em 1943 com o modelo matemático de McCulloch e Pitts. Em 1949, Hebb propôs sua regra de aprendizado, consolidando o princípio da plasticidade sináptica. Em 1958, Rosenblatt apresentou o *perceptron*, capaz de executar classificação linear supervisionada (KROGH, 2008; WALCZAK, 2019).

Na sequência, Widrow e Hoff desenvolveram o *ADALINE* (1959), que introduziu mecanismos de aprendizado adaptativo aplicados a problemas práticos (WU; FENG, 2018). Já em 1982, Hopfield propôs sua rede recorrente, incorporando propriedades dinâmicas não lineares e expandindo o escopo das RNAs (PAVLOVIC; SCHONFELD; FRIEDMAN, 2005).

3.5.1 Modelo do Neurônio Biológico

Conforme ilustrado na Figura 11, o neurônio biológico é constituído pelo corpo celular, dendritos (responsáveis pela recepção de sinais) e axônio (responsável pela emissão), transmitindo impulsos por meio de sinapses (ARIYO; AKTER, 2024). Essa organização serviu de inspiração para a formulação do neurônio artificial, no qual os sinais de entrada são ponderados, somados e posteriormente transformados por meio de uma função de ativação.

Figura 11 – Representação esquemática do neurônio biológico, evidenciando corpo celular, dendritos, axônio e sinapses.



Fonte: Das *et al.* (2023).

3.5.2 Perceptron e Formalização Matemática

O *perceptron*, proposto por Rosenblatt em 1958, é considerado a unidade fundamental das Redes Neurais Artificiais (RNAs) (PICTON, 1994; PETERS, 1995). Seu funcionamento pode ser descrito pela seguinte equação:

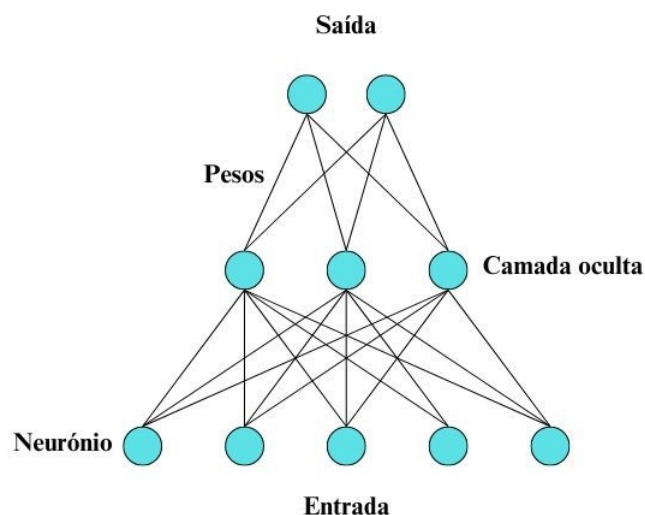
$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (3.3)$$

em que x_i representam as entradas, w_i os pesos sinápticos associados, b o termo de viés e $f(\cdot)$ a função de ativação. Essa formulação caracteriza um classificador linear, restrito a separar padrões que sejam linearmente separáveis em um espaço n -dimensional.

3.5.3 Arquitetura de Redes Neurais

Uma RNA típica é organizada em três tipos de camadas: entrada, ocultas e saída, conforme ilustrado na Figura 12. Cada camada aplica transformações — geralmente não lineares — que ampliam a capacidade de representação do modelo e permitem a modelagem de relações complexas entre as variáveis de entrada e saída (BENARDOS; VOSNIAKOS, 2007; WANG, 2003).

Figura 12 – Arquitetura típica de uma RNA.



Fonte: Adaptado de Wang (2003).

3.5.3.1 Redes Feedforward e Backpropagation

As RNAs do tipo *feedforward* transmitem sinais de forma unidirecional, ajustando os pesos por meio da retropropagação do erro (*backpropagation*), que emprega o método do gradiente descendente (SCHREUDER *et al.*, 2025; YADAV; YADAV; Anirbid Sircar, 2021):

$$w_{ij} \leftarrow w_{ij} - \eta \frac{\partial \mathcal{L}}{\partial w_{ij}} \quad (3.4)$$

em que $\eta > 0$ representa a taxa de aprendizado e $\mathcal{L}$ a função de perda, como a entropia cruzada amplamente utilizada em tarefas de classificação.

3.5.3.2 Redes Recorrentes

As Redes Neurais Recorrentes (RNNs) mantêm um estado interno, o que as torna adequadas para o processamento de dados sequenciais (GHANE; FAGERSTRÖM; MIRKHALAF, 2024). Embora não sejam empregadas neste trabalho, merecem menção como contraponto às CNNs, uma vez que o fluxo de informação recursivo permite modelar dependências temporais e contextuais.

3.5.4 Funções de Ativação

As funções de ativação introduzem não linearidade ao modelo, ampliando sua capacidade de representação e de aprendizado de relações complexas (LAUDANI *et al.*, 2015). Conforme resumido na Tabela 7, diferentes funções apresentam propriedades distintas, influenciando diretamente a estabilidade e o desempenho do treinamento.

Entre as funções mais utilizadas destacam-se:

- **Degrau:** aplicada nos primeiros perceptrons, atualmente obsoleta.
- **Sigmoide e Tanh:** suavizam a saída, mas apresentam saturação e gradiente desvanecente.
- **ReLU:** eficiente e amplamente utilizada em arquiteturas modernas.
- **SiLU e Softplus:** variantes suaves da ReLU, que reduzem saturação e podem melhorar a estabilidade do treinamento.

Tabela 7 – Funções de ativação mais citadas na literatura e principais características.

Referência	Nome	Expressão	Observações
Laudani <i>et al.</i> (2015)	Degrau	$f(y) = \begin{cases} a, & y \geq 0 \\ b, & y < 0 \end{cases}$	Usada nos primeiros Perceptrons; não diferenciável, dificulta <i>backpropagation</i> .
Song <i>et al.</i> (2023)	ReLU	$f(y) = \max\{0, y\}$	Simples e eficiente; risco de “neurônios mortos” para valores negativos persistentes.
Wang (2003)	Sigmoide	$\sigma(y) = \frac{1}{1+e^{-y}}$	Saturação para valores extremos causa gradiente desvanecente.
Elfwing, Uchibe e Doya (2018)	SiLU	$f(y) = y \times \sigma(y)$	Suaviza transição da ReLU; pode melhorar desempenho.
Wang <i>et al.</i> (2020)	Tanh	$\tanh(y) = \frac{1-e^{-2y}}{1+e^{-2y}}$	Saída centrada em zero; ainda sofre com gradiente desvanecente.
Zhang <i>et al.</i> (2022)	Softmax	$f(y_i) = \frac{e^{y_i}}{\sum_{j=1}^N e^{y_j}}$	Saída normalizada para classificação multiclasse.
Zhang <i>et al.</i> (2022)	Softsign	$f(y) = \frac{y}{1+ y }$	Mais suave que Tanh; menor saturação de gradiente.
Elsayed, Maida e Bayoumi (2018)	SoftPlus	$f(y) = \ln(1 + e^y)$	Aproxima ReLU com transição suave; maior custo computacional.

Fonte: Autor (2024).

3.5.4.1 Entropia Cruzada Binária

A função de perda constitui um componente essencial no treinamento de Redes Neurais Convolucionais (CNNs), pois mede a discrepância entre as previsões do modelo e os rótulos reais, orientando o ajuste dos pesos por meio do *backpropagation* (YU *et al.*, 2020; ZHOU *et al.*, 2019). No contexto da detecção de glaucoma, em que frequentemente há desbalanceamento entre as classes (com maior número de amostras não glaucomatosas), a escolha da função de perda é determinante para evitar que o modelo favoreça a classe majoritária e comprometa a identificação de casos positivos, que representam a prioridade clínica no rastreamento do glaucoma (XIONG *et al.*, 2021).

A entropia cruzada binária é amplamente utilizada em problemas de classificação binária, como no presente trabalho, por quantificar de forma eficiente a diferença entre a distribuição real (rótulos verdadeiros) e a distribuição prevista pelo modelo. Sua formulação é dada por:

$$\mathcal{L}_{BCE} = -\frac{1}{M} \sum_{m=1}^M [y_m \log(\hat{y}_m) + (1 - y_m) \log(1 - \hat{y}_m)] \quad (3.5)$$

em que:

- M é o número total de amostras no lote;
- $y_m \in \{0, 1\}$ representa o rótulo verdadeiro da m -ésima amostra (1 para glaucoma, 0 para não glaucoma);
- $\hat{y}_m = h_\theta(x_m)$ é a probabilidade prevista pelo modelo para a amostra x_m , parametrizada pelos pesos θ .

O termo $y_m \log(\hat{y}_m)$ atribui penalidade elevada quando a classe positiva (glaucoma) é classificada incorretamente, enquanto $(1 - y_m) \log(1 - \hat{y}_m)$ penaliza erros relacionados à classe negativa. Por exemplo, para um rótulo verdadeiro $y_m = 1$, uma previsão de $\hat{y}_m = 0,6$ gera perda maior do que $\hat{y}_m = 0,9$, incentivando o modelo a aumentar a confiança na classe correta.

Estudos recentes (HO; WOOKEY, 2020; YU *et al.*, 2020) demonstram que a entropia cruzada binária, quando associada a técnicas de *data augmentation* e balanceamento de classes, melhora a sensibilidade em cenários médicos, favorecendo a detecção de casos positivos mesmo em bases desbalanceadas, característica essencial para aplicações clínicas de rastreio e diagnóstico precoce do glaucoma.

3.5.5 Limitações e Relação com CNNs para Detecção de Glaucoma

Embora as RNAs genéricas sejam modelos poderosos, sua aplicação direta à detecção de glaucoma apresenta limitações relevantes:

- **Overfitting em bases reduzidas:** fenômeno recorrente em imagens médicas, devido ao número limitado de exemplos anotados.
- **Sensibilidade a desbalanceamento de classes:** tende a gerar vieses de predição, favorecendo a classe majoritária.
- **Dependência de atributos manuais:** modelos simples requerem *features* pré-definidas, restringindo a detecção de variações sutis na morfologia do DO.

Essas limitações justificam a adoção de Redes Neurais Convolucionais (CNNs), capazes de realizar extração automática de padrões espaciais diretamente das imagens, preservando estruturas críticas para o diagnóstico de glaucoma, como a CDR e a morfologia vascular.

3.6 Redes Neurais Convolucionais (CNNs)

As CNNs constituem uma classe de modelos de *deep learning* projetada para processar dados com estrutura de grade, como imagens bidimensionais, explorando propriedades de compartilhamento de pesos e invariância espacial (YU *et al.*, 2020; ZHOU *et al.*, 2019). Diferentemente das RNAs tradicionais, nas quais a extração de atributos relevantes geralmente depende de técnicas manuais de pré-processamento, as CNNs realizam essa etapa de forma automática, aprendendo representações hierárquicas diretamente a partir dos dados brutos.

Em tarefas médicas, e especialmente na detecção de glaucoma, essa capacidade é crucial: pequenas alterações morfológicas no DO, variações sutis na CDR e padrões de perda da RNFL demandam mecanismos de modelagem estáveis, capazes de capturar dependências espaciais em múltiplas escalas (DAS; NAYAK; PACHORI, 2024). Nesse contexto, as CNNs não apenas extraem características globais da imagem, mas também preservam detalhes locais decisivos para diferenciar entre um caso saudável e um com suspeita de glaucoma.

O sucesso das CNNs nesse domínio é amplificado pelo uso de estratégias modernas como *transfer learning*, que permite reaproveitar pesos de modelos treinados em bases extensas como o ImageNet, e pela incorporação de módulos de atenção (ex.: AG, CSAM, ESAM), que concentram o aprendizado em regiões anatomicamente relevantes. Essas abordagens têm se mostrado essenciais para lidar com conjuntos de dados médicos frequentemente limitados e heterogêneos, mantendo alto desempenho e capacidade de generalização.

3.6.1 Arquitetura Geral

As CNNs apresentam uma arquitetura composta por módulos especializados que, combinados, permitem extrair e processar características visuais em diferentes níveis de abstração. Essa capacidade é fundamental para tarefas como a detecção de glaucoma, em que alterações sutis na morfologia do DO, na escavação e no padrão vascular exigem representações espaciais precisas e hierárquicas (LI *et al.*, 2019; DAS; NAYAK; PACHORI, 2024; ALOM *et al.*, 2019). De forma geral, uma CNN típica é estruturada em três blocos funcionais principais: camadas convolucionais, camadas de *pooling* e camadas totalmente conectadas (FC).

3.6.1.1 Camadas Convolucionais

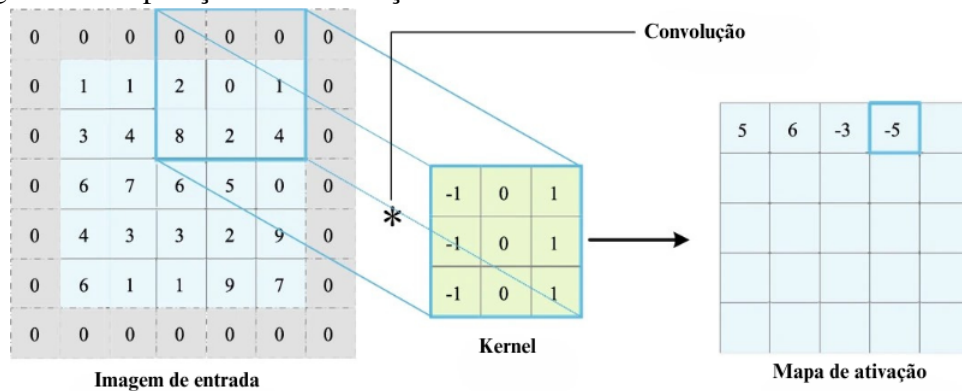
As camadas convolucionais realizam a operação de convolução aplicando filtros $K \in \mathbb{R}^{f \times f \times C}$ sobre a entrada X , produzindo mapas de ativação:

$$Y_{i,j,k} = \sigma \left(\sum_{u=1}^f \sum_{v=1}^f \sum_{c'=1}^C K_{u,v,c',k} \cdot X_{i+u,j+v,c'} + b_k \right) \quad (3.6)$$

em que f corresponde ao tamanho do filtro, C ao número de canais de entrada, b_k ao viés e $\sigma(\cdot)$ à função de ativação (ex., ReLU, SiLU). Nas primeiras camadas, os filtros tendem a capturar bordas do DO e padrões vasculares, enquanto nas camadas mais profundas emergem representações mais abstratas, como a CDR.

A operação de convolução pode levar à perda de detalhes nas bordas, pois o *kernel* não se sobrepõe totalmente aos elementos mais externos da entrada, reduzindo as dimensões do mapa de ativação. Em imagens de fundo de olho, isso pode implicar em perda de informações periféricas relevantes, como o contorno do DO. O *padding* com zeros mitiga esse problema ao adicionar linhas e colunas ao redor da imagem, conforme ilustrado na Figura 13, permitindo que o *kernel* processe também as regiões externas e preservando as dimensões originais da saída (modo *same padding*) (YAMASHITA *et al.*, 2018).

Figura 13 – Operação de convolução na CNN.



Fonte: Adaptado de Zhao *et al.* (2024).

3.6.1.2 Camadas de pooling

As camadas de *pooling* têm como finalidade reduzir a dimensionalidade espacial dos mapas de ativação, diminuindo o número de parâmetros e o custo computacional, além de aumentar a invariância a pequenas translações e deformações (NIRTHIKA *et al.*, 2022; ZHAO

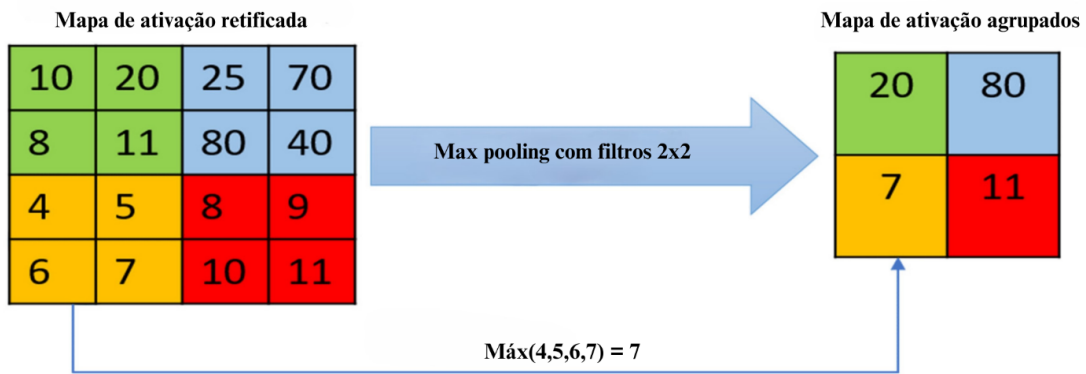
et al., 2024). A operação mais utilizada é o *max-pooling*, formalmente definida como:

$$Y_{i,j,k} = \max_{(u,v) \in \mathcal{R}} X_{si+u,sj+v,k} \quad (3.7)$$

em que $\mathcal{R}$ é a região de *pooling* e s o *stride*.

Em exames de fundo de olho, o *max-pooling*, ilustrado na Figura 14, contribui para a preservação de padrões salientes — como bordas de alta intensidade no contorno do DO e áreas de contraste na escavação — ao mesmo tempo em que atenua ruídos decorrentes de iluminação irregular ou artefatos de captura. Entretanto, por selecionar apenas o valor máximo em cada região, pode descartar detalhes finos importantes para a avaliação de estágios iniciais do glaucoma, o que demanda que camadas subsequentes compensem essa perda a fim de preservar a sensibilidade diagnóstica.

Figura 14 – Representação do *max pooling*.



Fonte: Adaptado de Zafar et al. (2022).

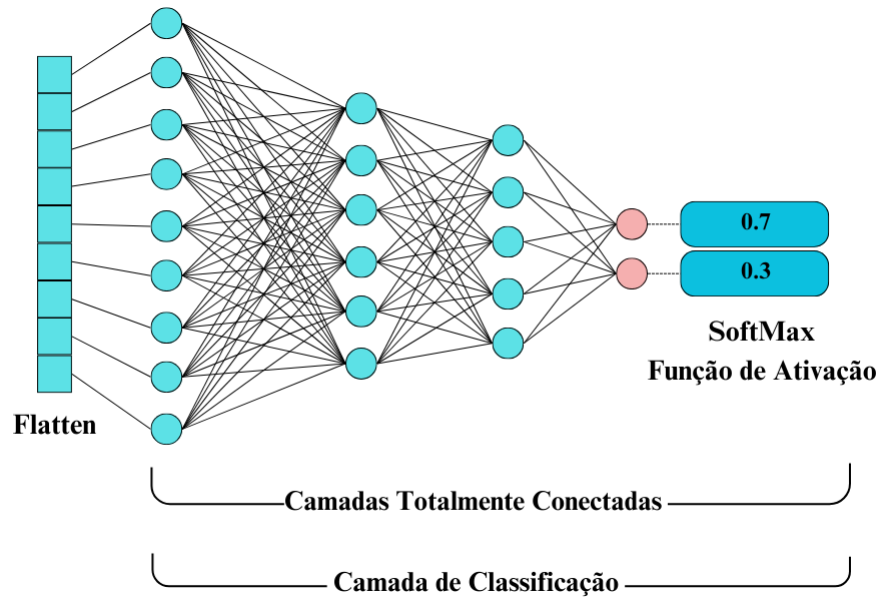
3.6.1.3 Camadas totalmente conectadas (FC)

As camadas totalmente conectadas (*Fully Connected* – FC) são responsáveis pela integração final das características extraídas, mapeando-as para as classes de saída. Cada neurônio conecta-se a todas as ativações da camada anterior, permitindo combinações globais de *features* e raciocínio de alto nível. Para classificação binária (glaucoma/não glaucoma), adota-se usualmente a função de ativação *sigmoid*, que comprime a saída no intervalo $[0, 1]$. Em problemas multiclasse, utiliza-se a *softmax*, que normaliza as probabilidades sobre todas as classes (YAMASHITA et al., 2018).

O número de camadas totalmente conectadas, a quantidade de neurônios e o volume de parâmetros treináveis influenciam diretamente o desempenho do modelo. Um número adequado

de unidades permite capturar funções discriminantes não lineares complexas no espaço de características, mas o excesso de parâmetros pode levar ao sobreajuste (*overfitting*), reduzindo a capacidade de generalização — fator crítico em bases médicas limitadas (BASHA *et al.*, 2020). Nesse contexto, a estrutura típica dessa camada de classificação é apresentada na Figura 15.

Figura 15 – Camada de Classificação de uma CNN clássica.



Fonte: Adaptado de Soudy *et al.* (2024).

O treinamento da rede é realizado por retropropagação do erro (*backpropagation*), ajustando os pesos para minimizar uma função de perda $\mathcal{L}$. Embora a entropia cruzada binária (*Binary Cross-Entropy* — BCE) seja a mais comum em cenários binários, sua descrição detalhada já foi apresentada na Seção 3.5.4.1 (Entropia Cruzada Binária). Aqui, destaca-se que, no caso da detecção de glaucoma, essa escolha é motivada por sua capacidade de penalizar fortemente erros na classe minoritária (glaucoma), o que é essencial para aplicações clínicas em bases desbalanceadas (HO; WOOKEY, 2020; YU *et al.*, 2020).

3.6.1.4 Global Average Pooling 2D

A camada *Global Average Pooling 2D* (GAP 2D) foi introduzida como alternativa às camadas FC em arquiteturas convolucionais modernas, reduzindo o risco de *overfitting* e o número de parâmetros treináveis (LIN; CHEN; YAN, 2014). Enquanto operações de *pooling* convencionais, como o *MaxPooling* ou *AveragePooling*, aplicam janelas deslizantes locais sobre o mapa de ativação, o GAP 2D realiza a média de todos os elementos de cada mapa de característica, condensando-o em um único valor escalar.

Formalmente, dado um mapa de ativação bidimensional $\mathbf{X} \in \mathbb{R}^{h \times w}$, o valor resultante da operação de GAP 2D é definido como:

$$y = \frac{1}{h \cdot w} \sum_{i=1}^h \sum_{j=1}^w x_{i,j}, \quad (3.8)$$

em que h e w representam, respectivamente, a altura e a largura do mapa de ativação, e $x_{i,j}$ corresponde ao valor do elemento localizado na posição (i, j) . Quando aplicado a múltiplos canais de saída de uma convolução, o GAP 2D gera um vetor $\mathbf{y} \in \mathbb{R}^c$, em que c é o número de canais (ou filtros) convolucionais.

A principal vantagem do GAP 2D é eliminar a necessidade de camadas densas volumosas, que tendem a introduzir milhões de parâmetros adicionais, sobretudo em arquiteturas profundas. Além disso, a camada estabelece uma correspondência direta entre cada canal convolucional e a classe de saída, facilitando a interpretabilidade dos modelos. Dessa forma, o GAP 2D é amplamente empregado em arquiteturas recentes, como as variantes do EfficientNet, MobileNet e ResNet, promovendo modelos mais compactos, eficientes e menos propensos ao sobreajuste. Como exemplo, Li *et al.* (2022), a substituição de camadas FC por GAP 2D em redes ResNeXt reduziu os parâmetros para apenas 2,5 milhões (vs. 44,5 milhões do ResNet101), mantendo acurácia superior a 97% no UCF101 (Conjunto de Dados de Reconhecimento de Ações) (SOOMRO; ZAMIR; SHAH, 2012).

3.6.2 Regularização em Redes Neurais

Devido à sua elevada capacidade de ajuste, modelos de aprendizado profundo são vulneráveis ao *overfitting*, principalmente em contextos biomédicos nos quais a quantidade de amostras é limitada, como na detecção de glaucoma. Nesses cenários, técnicas de regularização desempenham papel crucial ao restringir a complexidade do modelo e ampliar sua capacidade de generalização (SRIVASTAVA *et al.*, 2014; KROGH; HERTZ, 1991).

3.6.2.1 Dropout

O *Dropout* é uma técnica de regularização proposta por Srivastava *et al.* (2014), cujo objetivo é mitigar o *overfitting* em redes neurais profundas. Durante o treinamento, em cada iteração, um subconjunto de neurônios é desativado aleatoriamente com probabilidade p , impedindo que unidades específicas se tornem excessivamente dependentes umas das outras, conforme

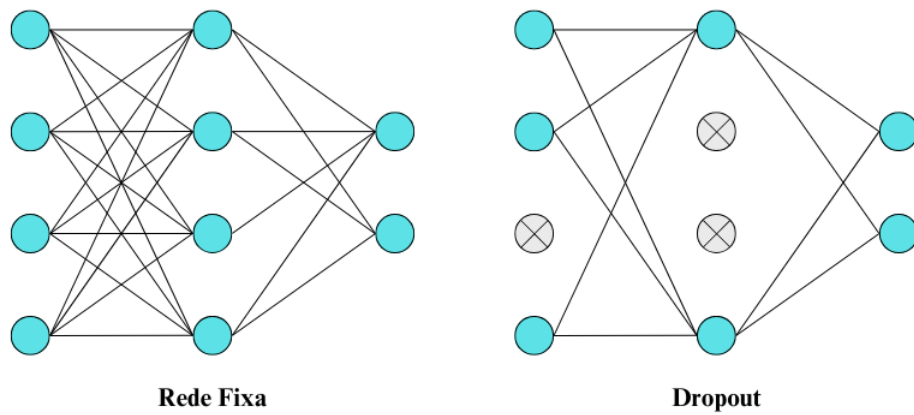
ilustrado na Figura 16. Esse mecanismo força a rede a aprender representações mais distribuídas e menos suscetíveis a ruídos ou correlações espúrias.

Formalmente, para uma camada com ativações $\mathbf{h} = [h_1, h_2, \dots, h_n]$, aplica-se uma máscara binária $\mathbf{r} \sim \text{Bernoulli}(p)$, resultando em:

$$\tilde{\mathbf{h}} = \mathbf{r} \odot \mathbf{h}, \quad (3.9)$$

onde $\odot$ representa a multiplicação elemento a elemento.

Figura 16 – Representação conceitual do *Dropout*.



Fonte: Adaptado de Garbin, Zhu e Marques (2020).

Na fase de inferência, todas as unidades são utilizadas, mas suas ativações são reescaladas por p , assegurando consistência entre treino e teste. Entretanto, estudos empíricos recentes como Garbin, Zhu e Marques (2020) alertam que o uso indiscriminado de *Dropout* em CNNs pode reduzir a acurácia, sugerindo sua aplicação criteriosa apenas em camadas densas ou com taxas adaptativas (ex.: $p \leq 0,25$).

3.6.2.2 Regularização L2

A regularização L2, também conhecida como *weight decay*, introduz um termo de penalização proporcional ao quadrado da norma dos pesos na função de perda. Assim, para uma função $L(\mathbf{w})$, a forma regularizada é:

$$L_{reg}(\mathbf{w}) = L(\mathbf{w}) + \lambda \sum_i w_i^2 \quad (3.10)$$

onde λ controla a intensidade da penalização. Esse mecanismo induz soluções de norma mínima, evitando a atribuição de pesos excessivos e, consequentemente, reduzindo a

variância do modelo. Em redes convolucionais aplicadas à análise de imagens de fundo de olho, a regularização L2 tem se mostrado eficaz para estabilizar o treinamento e melhorar a generalização, fator essencial em bases pequenas e desbalanceadas (KROGH; HERTZ, 1991).

3.6.3 Treinamento e Otimização

O processo de treinamento em redes neurais profundas consiste em ajustar iterativamente os pesos a fim de minimizar uma função de perda, que mede a discrepância entre as predições e os valores de referência. Esse ajuste é realizado por meio do algoritmo de retropropagação do erro (*backpropagation*), que calcula os gradientes em relação aos parâmetros e os utiliza em métodos de descida de gradiente para atualização dos pesos (RUMELHART; HINTON; WILLIAMS, 1986).

A escolha adequada da função de perda, da taxa de aprendizado e do otimizador influencia diretamente a convergência, a estabilidade e a capacidade de generalização do modelo. Entre os otimizadores modernos, destacam-se variantes adaptativas como o Adam, amplamente empregado em aplicações biomédicas por sua eficiência e convergência estável (ADAM *et al.*, 2014). Nesta subseção, discutem-se os principais elementos do processo de treinamento e otimização, incluindo funções de perda, estratégias de ajuste da taxa de aprendizado e métodos de otimização.

3.6.3.1 Funções de perda

As funções de perda (*loss functions*) orientam o processo de aprendizado ao quantificar a discrepância entre as predições da rede e os rótulos verdadeiros. O otimizador ajusta os pesos da rede minimizando essa medida por meio da retropropagação (RUMELHART; HINTON; WILLIAMS, 1986).

No presente trabalho, a BCE já apresentada na Seção 3.5.4.1 (Entropia Cruzada Binária), constitui a função base para tarefas de classificação binária como a detecção de glaucoma. Entretanto, dada a recorrência de bases desbalanceadas em contextos biomédicos, variantes têm sido propostas.

Uma das mais empregadas é a *Focal Loss*, que introduz um fator de ajuste para dar maior peso a exemplos difíceis de classificar (LIN *et al.*, 2017):

$$\mathcal{L}_{FL} = -\frac{1}{N} \sum_{i=1}^N \alpha (1 - \hat{y}_i)^{\gamma} y_i \log(\hat{y}_i) + (1 - \alpha) \hat{y}_i^{\gamma} (1 - y_i) \log(1 - \hat{y}_i) \quad (3.11)$$

em que α controla a importância entre classes e γ regula o foco nos exemplos de maior dificuldade.

Adicionalmente, funções de perda baseadas em métricas como AUC e F1 têm sido exploradas para alinhar diretamente o processo de otimização às métricas de avaliação mais relevantes em diagnósticos clínicos (YANG; YING, 2022). Essas alternativas são particularmente úteis quando o objetivo principal não é apenas a acurácia, mas a maximização da sensibilidade e da área sob a curva ROC.

3.6.3.2 Otimizadores adaptativos

O treinamento de redes neurais profundas tradicionalmente utilizava o *Stochastic Gradient Descent* (SGD) com taxa de aprendizado fixa. No entanto, esse método pode ser ineficiente em cenários de gradientes esparsos ou mal condicionados, comuns em imagens biomédicas. Para superar tais limitações, foram desenvolvidos otimizadores adaptativos, que ajustam dinamicamente a taxa de aprendizado para cada parâmetro ao longo do treinamento (CANONACO *et al.*, 2021; PARK; YI; JI, 2020).

O *Adaptive Moment Estimation* (Adam) tornou-se o mais difundido devido ao seu balanceamento entre eficiência e estabilidade. Ele combina o momento de primeira ordem (média dos gradientes) e de segunda ordem (média dos quadrados dos gradientes), normalizando-os por meio de correções de viés:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (3.12)$$

onde $\hat{m}_t$ e $\hat{v}_t$ representam estimativas corrigidas do gradiente e da variância acumulada, η é a taxa de aprendizado inicial e ϵ um termo de estabilidade numérica (ŞEN; ÖZKURT, 2020; YAQUB *et al.*, 2020). Sua simplicidade e baixo custo computacional tornam o Adam amplamente utilizado em CNNs para classificação de imagens médicas.

O *Root Mean Square Propagation* (RMSProp) segue ideia semelhante, mantendo uma média móvel exponencial dos quadrados dos gradientes. Isso estabiliza a descida em regiões de curvatura elevada e acelera a convergência, especialmente em problemas com ruído ou bases desbalanceadas (ADEM; KILIÇARSLAN, 2019; FAJAR *et al.*, 2023).

Já o *Adaptive Gradient* (Adagrad) ajusta a taxa de aprendizado de forma inversamente proporcional ao histórico acumulado dos gradientes. Isso favorece parâmetros raramente atualizados, mas pode levar à saturação da taxa de aprendizado ao longo do treino. Apesar dessa

limitação, continua relevante em tarefas com atributos esparsos (NUGROHO; YUNIARTI, 2022; PARK; YI; JI, 2020).

Em síntese, os otimizadores adaptativos oferecem maior estabilidade e rapidez de convergência em comparação ao SGD clássico, características fundamentais para treinar CNNs em conjuntos de dados biomédicos pequenos e desbalanceados, como os empregados na detecção de glaucoma.

3.6.3.3 Taxa de aprendizado

A taxa de aprendizado (η) é um dos hiperparâmetros mais críticos no treinamento de redes neurais, pois controla a magnitude das atualizações dos pesos a cada iteração (FAJAR *et al.*, 2023). Valores inadequados podem comprometer tanto a eficiência quanto a convergência do modelo.

Se η for muito baixa, a rede aprende lentamente, exigindo muitas épocas para atingir uma solução adequada, o que aumenta o custo de treinamento. Por outro lado, valores excessivamente altos provocam instabilidade, fazendo o algoritmo oscilar em torno de mínimos locais ou até divergir, impossibilitando a convergência (FAJAR *et al.*, 2023).

Para mitigar esses efeitos, utilizam-se estratégias de ajuste dinâmico, conhecidas como programação da taxa de aprendizado. Essas técnicas reduzem gradualmente o valor de η ao longo do treinamento, equilibrando velocidade inicial e estabilidade final da convergência. Entre os esquemas mais comuns destacam-se o *step decay*, o decaimento exponencial e o *cosine annealing* (recozimento cosseno), todos projetados para aprimorar a precisão e a capacidade de generalização do modelo (PARK; YI; JI, 2020).

3.6.3.4 Tamanho do lote

O tamanho do lote define o número de amostras processadas simultaneamente antes da atualização dos pesos da rede. Seu ajuste afeta diretamente a estabilidade do gradiente, o tempo de treinamento e a capacidade de generalização do modelo. Valores muito pequenos podem introduzir alta variância na estimativa do gradiente, resultando em treinamento instável; contudo, favorecem a exploração da função de perda e podem contribuir para melhor generalização. Por outro lado, valores muito grandes aproximam o treinamento de um gradiente determinístico, acelerando a convergência, mas com maior risco de sobreajuste e demanda de memória (KESKAR *et al.*, 2016; MASTERS; LUSCHI, 2018).

Na prática, o treinamento de CNNs para imagens biomédicas adota o regime de *mini-batch*, em que tamanhos intermediários equilibram custo computacional e capacidade de generalização. Estudos indicam que tamanho do lote moderados (por exemplo, 16 a 64) apresentam desempenho competitivo, ao permitir convergência eficiente sem comprometer a estabilidade do aprendizado (SHALLUE *et al.*, 2019). De forma consistente, nos experimentos deste trabalho também se verificou que valores intermediários de tamanho do lote impactaram a estabilidade e a velocidade de convergência. Esses efeitos foram analisados em maior profundidade na Seção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso).

Em síntese, o desempenho de redes convolucionais depende do equilíbrio entre função de perda, algoritmo de otimização, taxa de aprendizado e tamanho do lote. Esses elementos afetam tanto a estabilidade quanto a velocidade de convergência, além de influenciarem a capacidade de generalização em bases biomédicas reduzidas. Neste trabalho, tais escolhas foram ajustadas para conciliar eficiência computacional e acurácia diagnóstica, sustentando a metodologia proposta.

3.6.4 Avanços Arquiteturais e Transfer Learning

A aprendizagem por transferência (*transfer learning*) visa aprimorar o desempenho de um modelo no domínio de destino aproveitando conhecimentos adquiridos previamente em um domínio de origem relacionado. Essa estratégia reduz a necessidade de grandes volumes de dados anotados e o custo computacional de treinamento (ZHUANG *et al.*, 2021).

Em vez de iniciar o aprendizado de redes profundas do zero, processo intensivo em tempo e recursos, utiliza-se um modelo pré-treinado em grandes bases genéricas, como o ImageNet, que já extraiu características universais (bordas, texturas, formas) aplicáveis a diferentes tarefas. Essas redes podem ser adaptadas ao novo domínio via ajuste fino (*fine-tuning*) ou empregadas como extratoras de atributos, permitindo detectar padrões sutis no DO e na escavação em exames de fundo de olho (ZHUANG *et al.*, 2021; RAJARAMAN *et al.*, 2018).

3.6.4.1 ResNet-50

A ResNet-50 é composta por 50 camadas treináveis e introduziu *skip connections* para implementar o mapeamento residual (HE *et al.*, 2016; SHAFIQ; GU, 2022), atenuando o problema do gradiente desvanecido em redes profundas. O bloco residual é definido por:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x} \quad (3.13)$$

onde $\mathcal{F}(\cdot)$ representa convoluções, normalizações e ativações, e a soma direta com $\mathbf{x}$ preserva o fluxo de gradientes (GULHANE; KUMAR, 2023). Cada bloco segue a estrutura *bottleneck*, com:

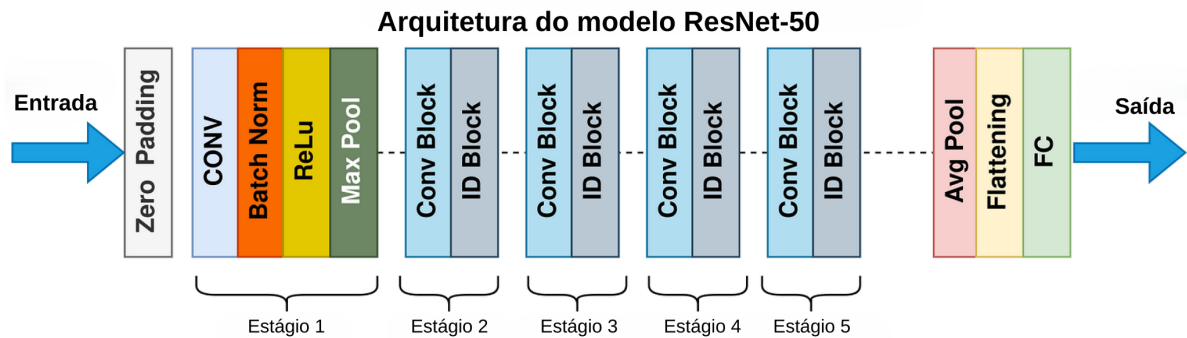
1. Convolução 1×1 para redução dimensional;
2. Convolução 3×3 para extração de atributos;
3. Convolução 1×1 para restauração dimensional.

Quando há mudança no número de canais ou resolução, aplica-se projeção 1×1 :

$$\mathcal{P}(\mathbf{x}) = W_s * \mathbf{x} \quad (3.14)$$

Os blocos residuais da arquitetura ResNet-50, compostos por convoluções (*conv_block* e *identity_block*) e normalização em lote, conforme ilustrado na Figura 17, utilizam conexões de atalho que preservam a informação original pela soma da entrada à saída do bloco. Esse mecanismo facilita o fluxo de gradientes e mitiga o problema de degradação em redes profundas (MANDAL; OKEUKWU; THEIS, 2021).

Figura 17 – Visão esquemática da arquitetura ResNet-50 destacando os cinco estágios residuais.



Fonte: Adaptado de Molčan *et al.* (2023).

Na análise oftálmica, a profundidade da ResNet-50 permite capturar desde padrões vasculares de baixo nível até representações mais abstratas associadas ao aumento da escavação. Embora exija cerca de 25,6 milhões de parâmetros, seu desempenho consistente e a estabilidade no treinamento justificam sua inclusão no *ensemble* proposto.

3.6.4.2 EfficientNet-B0

O EfficientNet-B0 otimiza simultaneamente acurácia e custo computacional por meio do *compound scaling* (TAN; LE, 2020):

$$d = \alpha^\phi, \quad w = \beta^\phi, \quad r = \gamma^\phi \quad (3.15)$$

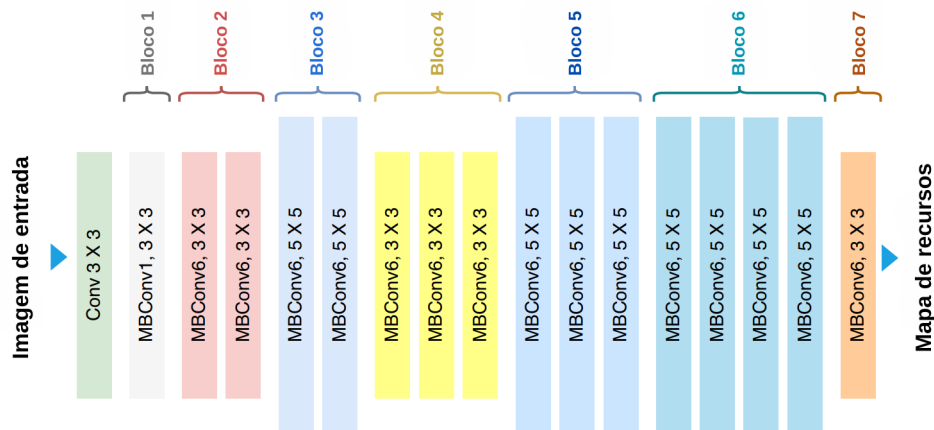
com $\alpha, \beta, \gamma > 0$ e $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$. Esse escalonamento conjunto evita degradação de desempenho ao balancear profundidade (d), largura (w) e resolução (r). Sua unidade básica é o bloco MBConv, derivado da MobileNetV2, que inclui:

1. Convolução 1×1 (expansão);
2. Convolução *depthwise* $k \times k$;
3. Módulo *Squeeze-and-Excitation* (SE);
4. Convolução 1×1 (projeção e conexão residual).

A ativação *Swish*, $\text{Swish}(x) = x \cdot \sigma(x)$, favorece a propagação de gradientes.

A arquitetura do EfficientNet-B0, conforme esquematizada na Figura 18, utiliza blocos *MBConv* como unidades fundamentais. Esses blocos realizam convoluções invertidas com diferentes tamanhos de *kernel* (3×3 e 5×5), combinando expansões e projeções lineares para minimizar a complexidade computacional, preservando a capacidade de extração de atributos (TAN; LE, 2019).

Figura 18 – Arquitetura simplificada do EfficientNet-B0, composta por blocos invertidos *MBConv* de diferentes tamanhos de *kernel*.



Fonte: Adaptado de Ahmed e Sabab (2022).

Com cerca de 5,3 milhões de parâmetros, o EfficientNet-B0 alia eficiência e boa capacidade de generalização, sendo especialmente adequado para dispositivos clínicos limitados. No contexto do *ensemble* proposto, ele contribui com representações multiescala e reforço de regiões discriminativas via módulos SE.

3.6.4.3 MobileNetV2

A MobileNetV2 (SANDLER *et al.*, 2018; ZAIDI *et al.*, 2022) foi projetada para alta eficiência computacional mantendo boa capacidade de extração de atributos. Seu bloco fundamental, o *inverted residual* com *bottleneck* linear, realiza:

$$\mathbf{X}_{\text{expand}} = \text{Conv}_{1 \times 1}(\mathbf{X}_{\text{in}}) \quad (3.16)$$

$$\mathbf{X}_{\text{dw}} = \text{ReLU6}(\text{DWConv}_{3 \times 3}(\text{ReLU6}(\mathbf{X}_{\text{expand}}))) \quad (3.17)$$

$$\mathbf{X}_{\text{proj}} = \text{Conv}_{1 \times 1}^{\text{linear}}(\mathbf{X}_{\text{dw}}) \quad (3.18)$$

com conexão residual quando dimensões coincidem. A *ReLU6* limita saídas a $[0, 6]$, beneficiando implementações em baixa precisão. O uso de convoluções *depthwise separable* reduz drasticamente as operações (*MAdds*), preservando detalhes finos como contornos vasculares.

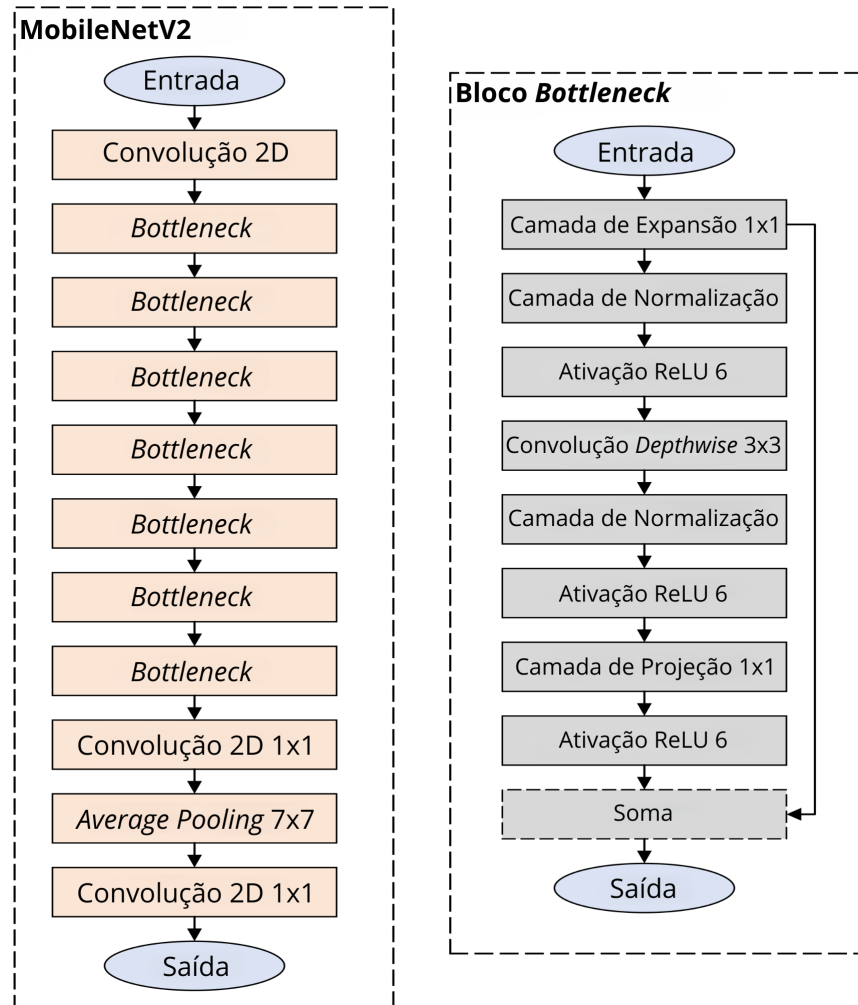
A arquitetura da MobileNetV2, conforme esquematizada na Figura 19, é organizada em blocos *bottleneck*, nos quais convoluções invertidas (*inverted residuals*) combinam operações *depthwise* e *pointwise*. Esse mecanismo interno possibilita maior eficiência computacional mantendo a expressividade da rede.

A MobileNetV2 representa um marco na conciliação entre eficiência e capacidade de extração de atributos. Seu bloco de resíduo invertido com *bottleneck* linear combina expansão, convolução *depthwise* e projeção linear para reduzir operações (*MAdds*) e consumo de memória, preservando padrões morfológicos relevantes do DO e da escavação. O uso de ReLU6 favorece a quantização em baixa precisão, tornando a arquitetura adequada a dispositivos móveis e ambientes clínicos com restrições de *hardware*. Essa concepção possibilita aplicações de triagem em larga escala para o glaucoma, estabelecendo um paradigma de eficiência sem perda significativa de precisão e influenciando modelos biomédicos recentes.

3.6.4.4 Integração no Método Proposto

O método proposto foi concebido para mitigar os efeitos do *domain shift*, fenômeno em que discrepâncias nas distribuições de treinamento e teste reduzem o desempenho do modelo

Figura 19 – Arquitetura da MobileNetV2 e bloco *bottleneck* invertido.



Fonte: Adaptado de Tragoudaras *et al.* (2022).

em ambientes reais. Esse problema é recorrente em aplicações biomédicas devido a diferenças de dispositivos, protocolos de captura, iluminação e características populacionais, tornando a generalização para domínios não vistos um desafio crítico (WANG *et al.*, 2023).

Para enfrentar esse cenário, a metodologia integra duas estratégias principais:

1. **Arquiteturas complementares:** utilizam-se três CNNs pré-treinadas com pesos congelados — EfficientNet-B0, MobileNetV2 e ResNet-50 — para extrair vetores de características com diferentes *inductive biases*. A combinação de modelos heterogêneos reduz a variância e aumenta a consistência frente a variações de domínio, conforme suportado por teorias de decomposição viés-variância e diversidade em *ensembles* (DONG *et al.*, 2020; GANAIE *et al.*, 2022).
2. **Data augmentation direcionada:** aplicaram-se transformações geométricas e fotométricas (rotações, flips horizontais, ajustes de brilho e contraste) para simular variações interdo-

mínio e aumentar a diversidade do conjunto de treinamento, mitigando o sobreajuste em cenários de dados limitados (SHORTEN; KHOSHGOFTAAR, 2019; PEREZ; WANG, 2017).

Essa integração, representada no fluxograma metodológico da Figura 23, busca equilibrar sensibilidade e especificidade, fornecendo um modelo mais estável diante da heterogeneidade interdomínios e maximizando sua capacidade de generalização.

3.6.5 Mecanismos de Atenção

O uso de módulos de atenção tem se mostrado eficiente na detecção de glaucoma, ao direcionar seletivamente o foco da rede para regiões de maior relevância diagnóstica, reduzindo a influência de áreas irrelevantes e aumentando a sensibilidade nos estágios iniciais da doença (DAS; NAYAK; PACHORI, 2024).

Dentre os mais empregados, destacam-se:

- *Attention Gate (AG)*: aplicado em arquiteturas do tipo U-Net, direciona o fluxo de informação para regiões específicas do DO e da escavação, suprimindo ruídos de fundo e preservando detalhes críticos (SHYAMALEE *et al.*, 2024).
- *Channel and Spatial Attention Module (CSAM)*: integra atenção de canal e espacial, ponderando simultaneamente a importância relativa de cada mapa de características e suas localizações, favorecendo a identificação de padrões morfológicos sutis (DAS; NAYAK; PACHORI, 2024).
- *Enhanced Spatial Attention Module (ESAM)*: amplia a resolução da atenção espacial, fortalecendo a discriminação entre regiões normais e glaucomatosas, especialmente em estágios precoces da doença (DAS; NAYAK; PACHORI, 2024).

A operação fundamental de um módulo de atenção pode ser descrita por:

$$F' = \alpha(F) \odot F \quad (3.19)$$

em que $F \in \mathbb{R}^{H \times W \times C}$ representa o mapa de características, $\alpha(F) \in [0, 1]^{H \times W \times C}$ é o mapa de atenção aprendido e $\odot$ denota a multiplicação elemento a elemento.

O mapa $\alpha(F)$ é calculado por uma função parametrizada (tipicamente convoluções seguidas de sigmoide), que aprende a realçar regiões relevantes como o contorno da escavação e os vasos adjacentes, alinhando-se ao objetivo clínico de estimar com precisão a CDR (DAS; NAYAK; PACHORI, 2024).

De forma comparativa, o AG é mais indicado para arquiteturas de segmentação, o CSAM beneficia classificadores globais ao combinar atenção de canal e espacial, e o ESAM destaca-se em cenários que exigem padrões espaciais de alta resolução, como nos estágios iniciais do glaucoma.

Embora o método proposto não utilize módulos de atenção, compreender esses mecanismos permite contextualizar avanços recentes e evidencia oportunidades para futuras melhorias, sobretudo na priorização de regiões clínicas relevantes e na redução de falsos negativos em bases heterogêneas.

3.6.6 *Análise Comparativa e Limitações*

A literatura recente evidencia contrastes marcantes entre diferentes abordagens para detecção de glaucoma em imagens de fundo de olho, tanto em termos de desempenho quanto de requisitos computacionais:

- **Modelos com segmentação prévia** (ex., U-Net+AG (SHYAMALEE *et al.*, 2024)): oferecem interpretabilidade e permitem extração explícita de biomarcadores como a CDR. Em cenários controlados, alcançam AUC acima de 0,95; entretanto, sua acurácia decai significativamente quando a segmentação é degradada por ruídos ou artefatos. Além disso, apresentam elevado custo computacional, com arquiteturas U-Net ultrapassando 30 milhões de parâmetros.
- **Classificação direta via CNN** (ex., MobileNetV2 (CIVIT-MASOT *et al.*, 2020)): eliminam a dependência da segmentação, reduzindo o tempo de processamento (inferência inferior a 30ms em GPU de médio porte) e a complexidade ($\approx 3,4$ milhões de parâmetros, ≈ 300 MFLOPs). Por outro lado, podem perder sensibilidade a alterações morfológicas discretas, resultando em quedas de até 5–10% no AUC em estágios iniciais da doença.
- **Modelos com mecanismos de atenção** (ex., AES-Net (DAS; NAYAK; PACHORI, 2024)): ampliam a sensibilidade em regiões-chave do DO e da escavação, sobretudo em casos iniciais, com ganhos médios de 2–4% no AUC em relação à CNN base. Todavia, frequen-

temente apresentam alta complexidade computacional (ResNet-50 + ESAM $\approx$ 25 milhões de parâmetros, $\approx$ 4 GFLOPs), dificultando a adoção em dispositivos de baixa capacidade ou aplicações *point-of-care*.

Além das limitações de generalização interbases, a literatura destaca fragilidades recorrentes:

1. **Dependência de imagens de alta qualidade:** muitos métodos assumem condições ideais de aquisição e falham em bases clínicas diversas.
2. **Alto custo computacional:** inviabiliza o uso em ambientes com restrição de *hardware*.
3. **Sensibilidade a dados desbalanceados:** favorece a classe majoritária, reduzindo a detecção de casos positivos raros.
4. **Dependência de pré-processamentos complexos:** aumenta o risco de falhas quando aplicados a imagens atípicas.

Essas limitações fundamentaram o desenvolvimento do método proposto, que combina arquiteturas complementares (EfficientNet-B0, MobileNetV2 e ResNet-50) em um *ensemble* com votação ponderada. Essa estratégia busca equilibrar generalização (via diversidade arquitetural), eficiência computacional (uso de modelos leves) e resiliência a variações interbases e a dados desbalanceados, ao mesmo tempo em que preserva sensibilidade a padrões morfológicos críticos para o rastreio de glaucoma.

3.6.7 Tendências em Abordagens Híbridas

Estudos recentes mostram que a combinação de múltiplas CNNs pré-treinadas com classificadores tradicionais pode superar limitações recorrentes de abordagens isoladas, como o *overfitting* a um único estilo de aquisição e a baixa sensibilidade a variações morfológicas sutis. Ao integrar arquiteturas complementares, como EfficientNet-B0, MobileNetV2 e ResNet-50, é possível ampliar a diversidade de representações extraídas, mitigando vieses arquiteturais e aumentando a capacidade de generalização frente à heterogeneidade das imagens (HE *et al.*, 2016; SHAFIQ; GU, 2022).

Em bases desbalanceadas, a literatura destaca o uso de variantes ponderadas da BCE, aliadas a estratégias de *data augmentation* para aumentar a sensibilidade em casos minoritários (YU *et al.*, 2020; HO; WOOKEY, 2020). Além disso, a combinação de redes leves e profundas

permite equilibrar eficiência computacional e precisão diagnóstica, favorecendo a adaptação a diferentes condições de aquisição (HOWARD *et al.*, 2017; SZEGEDY *et al.*, 2016).

Tais configurações híbridas têm se mostrado particularmente promissoras para aplicações clínicas, oferecendo equilíbrio entre acurácia, custo computacional e viabilidade de implementação em ambientes com recursos limitados (DAS; NAYAK; PACHORI, 2024; GÓMEZ-VALVERDE *et al.*, 2019).

Tabela 8 – Análise comparativa de arquiteturas CNN para detecção de glaucoma

Referência	Arquitetura	AUC	Parâm. (M)	FLOPs (G)	Base	Limitações
Gómez-Valverde <i>et al.</i> (2019)	VGG19	0,927	143	19,6	DRISHTI-GS	Queda de 15% no AUC em bases pequenas.
Shoukat <i>et al.</i> (2023)	ResNet-50	0,970	25,6	4,1	RIM-ONE	-8,3% no AUC em imagens de baixa qualidade.
Tan e Le (2019)	EfficientNetB1	0,983	7,8	0,7	ORIGA	Sensível a desbalanceamento.
Das, Nayak e Pachori (2024)	AES-Net	0,949	18,9	7,07	REFUGE	AUC cai para 0,82 quando CDR<0,7.
Das <i>et al.</i> (2024)	CDAM-Net	0,992	1,9	0,35	ACRIMA	AUC cai para 0,87 em cross-dataset.
Shyamalee <i>et al.</i> (2024)	U-Net+AG+ResNet-50	0,990	25	4,3	Múltiplas	Alto custo computacional.

Fonte: Autor (2024).

3.7 Classificadores de aprendizado de máquina

A classificação é uma tarefa de aprendizado supervisionado que busca prever variáveis-alvo discretas (classes), diferindo da regressão, que prediz valores contínuos (PEREIRA; MITCHELL; BOTVINICK, 2009). Um classificador utiliza características extraídas (variáveis preditoras) para determinar a qual classe um exemplo pertence.

No contexto deste trabalho, aplicado à detecção de glaucoma em imagens de fundo de olho, os atributos preditores correspondem a descritores extraídos das imagens, enquanto o atributo de classe define a categoria diagnóstica (*glaucoma* ou *não glaucoma*). A relação entre esses elementos é fundamental para que o modelo discrimine corretamente padrões associados à patologia (SIDDIQUI *et al.*, 2020).

Após o treinamento em um conjunto de dados rotulado, o classificador é avaliado em um conjunto de teste contendo amostras não utilizadas no treino. Esse processo permite verificar se o modelo aprendeu padrões discriminativos e generalizáveis, garantindo desempenho consistente na predição de novos casos (SIDDIQUI *et al.*, 2020).

3.7.1 Bagging e Boosting

Bagging (*Bootstrap Aggregating*) consiste em treinar repetidamente o mesmo algoritmo em subconjuntos amostrais gerados por *bootstrap*, combinando os resultados finais por média (regressão) ou votação majoritária (classificação). Seu principal benefício é reduzir a variância do modelo e aumentar a estabilidade preditiva (WEN; HUGHES, 2020; TR; REDDY; ACHARYA, 2023).

Formalmente, dado um conjunto de classificadores $h_1(x), h_2(x), \dots, h_T(x)$, a predição final no *bagging* é:

$$H(x) = \text{mode}\{h_t(x)\}_{t=1}^T \quad (3.20)$$

No *boosting*, os modelos são treinados de forma sequencial, cada um corrigindo os erros do anterior. Amostras de difícil classificação recebem maior peso a cada iteração, e a decisão final é obtida por uma soma ponderada dos classificadores:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right) \quad (3.21)$$

onde α_t reflete a importância do classificador h_t (WARDHANI *et al.*, 2019), e as diferenças principais entre esses métodos estão resumidas na Tabela 10.

Embora complementares, ambos têm limitações: o *bagging* pode ser pouco eficaz em bases muito pequenas ou com forte desbalanceamento, enquanto o *boosting* tende a sobreajustar em presença de ruído e requer maior cuidado com regularização e ajuste de hiperparâmetros.

3.7.2 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte (*Support Vector Machine* — SVM) é um método supervisionado de classificação e regressão cujo objetivo central é encontrar um hiperplano que separe as classes com a maior margem possível. Essa margem corresponde à distância entre o hiperplano e os pontos mais próximos a ele, denominados vetores de suporte, conforme ilustrado na Figura 20. Maximizar essa margem aumenta a capacidade de generalização do modelo, favorecendo o desempenho em dados inéditos (ENGELN; HOOS, 2020; MOHAMMADI *et al.*, 2021).

Formalmente, a função de decisão de uma SVM linear é definida por:

Tabela 10 – Resumo dos atributos principais do *boosting* e *bagging*.

Aspecto	<i>Bagging</i>	<i>Boosting</i>
Abordagem	Modelos treinados independentemente.	Modelos treinados sequencialmente.
Objetivo	Reduzir a variância (estabilidade).	Reduzir viés e variância (precisão).
Amostragem	Subconjuntos criados por <i>bootstrap</i> .	Usa o conjunto completo, ajustando pesos.
Interação	Não há interação entre modelos base.	Modelos interagem corrigindo erros.
Convergência	Depende do número de modelos combinados; árvores independentes não ajustam erros anteriores.	Converge gradualmente para corrigir erros, com melhoria contínua a cada iteração.
Combinação de resultados	Votação majoritária ou média.	Combinação ponderada pelas performances.
Modelos típicos	<i>Random Forest</i> , classificador de <i>bagging</i> .	<i>AdaBoost</i> , <i>Gradient Boost</i> , <i>XGBoost</i> .
Risco de <i>Overfitting</i>	Menor risco de <i>overfitting</i> .	Maior risco de <i>overfitting</i> , se não regularizado.

Fonte: Autor (2024).

$$f(x) = w \cdot x + b \quad (3.22)$$

em que w representa o vetor de pesos, b o viés e x o vetor de entrada. A largura da margem ótima é dada por:

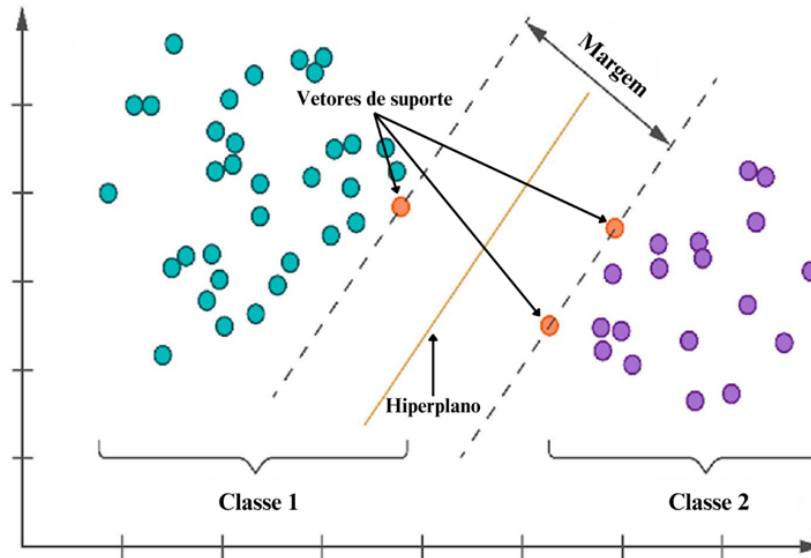
$$\text{Margem} = \frac{2}{\|w\|} \quad (3.23)$$

As SVMs podem ser:

- **Lineares:** adequadas quando os dados são separáveis por um hiperplano em seu espaço original, maximizando a margem entre as classes (KURANI *et al.*, 2023).
- **Não lineares:** utilizadas quando não há separação linear possível. Nesse caso, funções de *kernel* transformam os dados para um espaço de maior dimensão, onde se torna viável construir fronteiras de decisão mais complexas (MOHAMMADI *et al.*, 2021).

Apesar de sua eficácia, as SVMs apresentam limitações como alto custo computacional em bases muito grandes, sensibilidade a desbalanceamentos de classe (tendência a favorecer

Figura 20 – Margem de separação em uma SVM, destacando vetores de suporte e o hiperplano.



Fonte: Adaptado de Engelen e Hoos (2020).

a classe majoritária) e forte dependência da escolha do *kernel*, cujo ajuste inadequado pode comprometer a generalização.

3.7.3 *Random Forest*

Random Forest (RF) é uma técnica de aprendizado de conjunto amplamente utilizada para tarefas de classificação e regressão. Essa abordagem combina múltiplos modelos individuais, no caso, árvores de decisão, para aumentar a precisão das previsões. A agregação das saídas das árvores é realizada por meio de votação, em problemas de classificação, ou cálculo da média, em problemas de regressão (SHEYKHMUSA *et al.*, 2020; RIGATTI, 2017).

Matematicamente, a saída do RF pode ser expressa como:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (3.24)$$

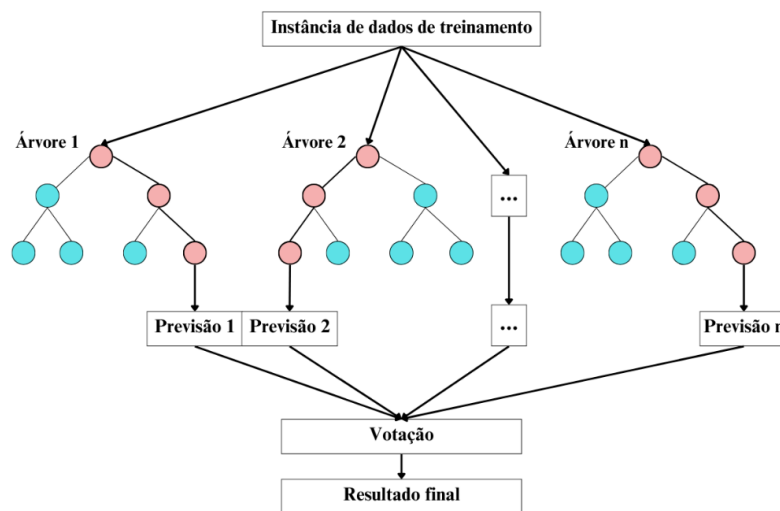
onde T é o número total de árvores no conjunto e $h_t(x)$ é a predição da t -ésima árvore.

A randomização ocorre de duas maneiras principais. Primeiro, o conjunto de dados é submetido à amostragem com substituição, conhecida como amostras *bootstrap*, criando subconjuntos que são utilizados no treinamento de cada árvore de decisão. Esse processo, denominado *bagging*, contribui para reduzir a variância do modelo e aumentar sua capacidade de generalização. A segunda randomização acontece nos nós de decisão, onde um subconjunto de preditores é selecionado aleatoriamente. Para um conjunto com p preditores, é comum escolher $\sqrt{p}$ atributos por nó, embora esse valor possa ser ajustado conforme o problema (RIGATTI,

2017).

A construção do modelo RF, conforme ilustrado na Figura 21, estabelece o número de árvores de decisão a serem geradas. Nesse processo, cada árvore é treinada individualmente a partir de uma amostra coletada dos dados originais com substituição. Aproximadamente $2/3$ dos dados são utilizados para o treinamento de cada árvore (*in-bag*), enquanto o $1/3$ restante (*out-of-bag*) é empregado na avaliação da precisão do modelo de forma independente ao conjunto de treinamento (SAHOO *et al.*, 2024).

Figura 21 – Representação do algoritmo Random Forest.



Fonte: Adaptado de Zeini *et al.* (2023).

Durante o desenvolvimento de cada árvore, o RF seleciona aleatoriamente um subconjunto de atributos em cada nó para determinar a melhor divisão, em vez de considerar todos os atributos, como ocorre em árvores de decisão tradicionais. O processo continua até que a árvore esteja totalmente construída ou atinja os critérios definidos, como profundidade máxima. Por fim, o erro calculado com os dados *out-of-bag* fornece uma estimativa confiável do desempenho geral, eliminando a necessidade de um conjunto de validação separado (SAHOO *et al.*, 2024).

O RF pode demandar alto custo de memória e processamento quando o número de árvores é elevado, além de perder interpretabilidade em comparação a modelos individuais. Em bases com muitos atributos irrelevantes, o desempenho do RF pode se degradar.

3.7.4 Gradient Boosting

O Gradient Boosting (GB) é uma técnica de aprendizado de conjunto baseada na construção sequencial de modelos fracos, tipicamente árvores de decisão, onde cada novo modelo

corrige os erros dos anteriores, combinando aprendizes simples em um comitê mais poderoso (BARRIGA *et al.*, 2022; BENTÉJAC; CSÖRGŐ; MARTÍNEZ-MUÑOZ, 2021; BIAU; CADRE; ROUVIÈRE, 2019).

Formalmente, o objetivo é encontrar uma aproximação da função que mapeia as entradas para seus valores de saída, minimizando o valor esperado de uma dada função de perda. Essa aproximação é construída de forma aditiva:

$$F_m(X) = F_{m-1}(X) + \rho_m h_m(X) \quad (3.25)$$

onde ρ_m é o peso atribuído ao modelo base $h_m(X)$ na m -ésima iteração.

O processo se inicia com uma aproximação inicial constante:

$$F_0(X) = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, \alpha) \quad (3.26)$$

Em seguida, cada iteração ajusta um novo modelo $h_m(X)$ buscando minimizar:

$$(\rho_m, h_m(X)) = \arg \min_{\rho, h} \sum_{i=1}^N L(y_i, F_{m-1}(X_i) + \rho h(X_i)) \quad (3.27)$$

Para isso, o algoritmo utiliza os pseudo-resíduos:

$$r_{mi} = \left[\frac{\partial L(y_i, F_{m-1}(X))}{\partial F_{m-1}(X)} \right]_{F(X)=F_{m-1}(X)} \quad (3.28)$$

que representam a direção de maior redução da função de perda. Esses resíduos são usados para treinar cada novo modelo base, ajustando iterativamente o comitê (BENTÉJAC; CSÖRGŐ; MARTÍNEZ-MUÑOZ, 2021; ZHANG *et al.*, 2019).

Apesar de poderoso, o GB é sensível a ruídos, propenso ao *overfitting* em bases pequenas sem regularização (ex.: *shrinkage*, *subsampling*) e apresenta maior custo computacional que o RF devido ao treinamento sequencial.

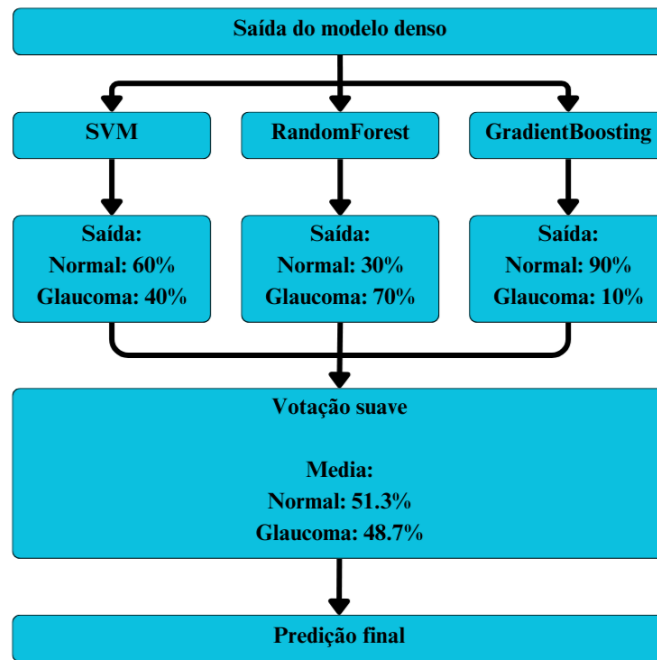
3.8 Classificadores em *ensemble*

Convencionalmente, um classificador individual, como uma SVM, RF ou GB, é treinado utilizando um conjunto de dados devidamente rotulados. No entanto, dependendo da complexidade e da distribuição dos padrões nos dados, um único modelo pode não conseguir capturar

todas as variações relevantes, resultando em desempenho insatisfatório no conjunto de teste. Para lidar com esse problema, utiliza-se uma abordagem alternativa: treinar múltiplos classificadores e combinar suas previsões para melhorar tanto a acurácia quanto a generalização (RAHMAN; TASNIM, 2014).

Os classificadores em *ensemble* combinam múltiplos modelos base para explorar a diversidade entre eles, conforme ilustrado na Figura 22. Quando os modelos são complementares e relativamente independentes, o erro de previsão tende a ser reduzido, pois erros individuais podem ser compensados pela votação ou pela média das saídas. Essa ideia é frequentemente associada ao princípio da "sabedoria coletiva", no qual múltiplas opiniões resultam em uma decisão mais consistente (KOTU; DESHPANDE, 2019; LATHA; JEEVA, 2019).

Figura 22 – Representação dos classificadores em *ensemble*.



Fonte: Autor (2024).

No presente trabalho, selecionaram-se três classificadores com perfis distintos, visando complementaridade:

- **SVM:** adequado para margens de separação bem definidas, mesmo em alta dimensionalidade (KURANI *et al.*, 2023);
- **RF:** menos sensível a ruídos e variações amostrais, reduzindo a variância do modelo (SAHOO *et al.*, 2024);

- **GB**: capaz de capturar padrões complexos e corrigir erros de forma sequencial (BENTÉJAC; CSÖRGŐ; MARTÍNEZ-MUÑOZ, 2021).

Essa combinação permite equilibrar estabilidade, capacidade de generalização e sensibilidade, características essenciais em aplicações clínicas.

Apesar da eficácia, *ensembles* demandam maior custo computacional e apresentam interpretabilidade reduzida. Em bases pequenas, podem ainda sofrer sobreajuste sem regularização e validação adequadas.

Em síntese, este capítulo apresentou os fundamentos clínicos e computacionais que sustentam a detecção de glaucoma assistida por inteligência artificial. Foram discutidos aspectos anatômicos e clínicos relevantes, as características das imagens de fundo de olho, técnicas de pré-processamento, bem como os princípios das redes neurais convolucionais e dos classificadores de aprendizado de máquina. Também foram detalhadas estratégias de *ensemble* e suas potenciais vantagens no contexto de bases heterogêneas e desbalanceadas. Esses conceitos estabelecem a base necessária para o próximo capítulo, no qual serão descritos os conjuntos de dados empregados, os procedimentos de pré-processamento adotados e a metodologia proposta para a construção e avaliação do sistema de classificação de imagens glaucomatosas.

5 RESULTADOS

Este capítulo apresenta e discute os resultados obtidos pelo modelo proposto para a classificação de imagens glaucomatosas e normais, além de compará-los com aqueles reportados em trabalhos da literatura. No *pipeline* desenvolvido, a etapa central consiste em identificar se a imagem de fundo corresponde a um caso glaucomatoso ou normal.

A região do DO desempenha papel fundamental nesse processo, uma vez que a CDR e a razão ISNT figuram entre os principais critérios utilizados para discriminar imagens suspeitas de glaucoma. Entretanto, a eficácia dessas técnicas é fortemente impactada pela qualidade das imagens, que podem apresentar heterogeneidades de iluminação, variações de cor, contraste e saturação, além de diferenças de resolução e de qualidade dos equipamentos de captura.

5.1 Análise de hiperparametrização

Inicialmente, foram conduzidos experimentos variando a *taxa de aprendizagem* (10^{-3} , 10^{-4} e 10^{-5}) e o tamanho do lote (16, 32, 64, 128), mantendo fixo o número de épocas de treinamento em 50. Os resultados de acurácia para cada configuração nos diferentes conjuntos de dados são apresentados nas Tabelas 18 a 22.

5.1.1 ACRIMA

Os dados da Tabela 18 mostram que a combinação tamanho do lote = 32 e taxa de aprendizagem = 10^{-3} atingiu a acurácia máxima (96,23%) no banco ACRIMA. Para tamanhos de lote mais elevados (64 e 128), observou-se queda no desempenho, sugerindo que atualizações excessivamente agressivas dos pesos podem induzir instabilidades no gradiente. Esse comportamento corrobora a premissa de que lotes de tamanho intermediário equilibram a variabilidade estocástica e a estabilidade da otimização, favorecendo a generalização do modelo.

Tabela 18 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto ACRIMA.

Tamanho do lote / Taxa de aprendizagem	10^{-3}	10^{-4}	10^{-5}
16	95,28	94,81	94,81
32	96,23	94,81	94,81
64	95,28	95,28	95,28
128	93,86	94,81	94,34

Fonte: Autor (2024).

No conjunto ACRIMA, a melhor configuração foi obtida com lote = 32 e taxa de aprendizagem = 10^{-3} , alcançando 96,23% de acurácia. Em contraste, a pior configuração ocorreu para lote = 128 com a mesma taxa, resultando em 93,86%, diferença absoluta de 2,37 pontos percentuais (p.p.).

Quando analisadas por taxa de aprendizagem, as médias de desempenho foram: 10^{-3} (95,16%), 10^{-4} (94,93%) e 10^{-5} (94,81%). Esses resultados evidenciam a superioridade de taxas moderadas em relação a valores excessivamente baixos, que não trouxeram ganhos adicionais e, em alguns casos, reduziram o desempenho.

É importante destacar que, embora essa configuração tenha apresentado resultados superiores no ACRIMA, para garantir consistência entre bases nos experimentos comparativos foi adotada a configuração global definida na Subseção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso): tamanho do lote = 32 e taxa de aprendizagem = 10^{-4} . Essa escolha reflete o equilíbrio entre estabilidade e desempenho médio observado nos diferentes conjuntos. Além disso, a adoção de uma configuração única está diretamente alinhada ao objetivo central deste trabalho: assegurar a capacidade de generalização do modelo em cenários clínicos heterogêneos, em que a consistência entre bases é mais relevante do que a otimização pontual em um único conjunto.

5.1.2 *Drishti-GSI*

Conforme os dados da Tabela 19, a melhor configuração foi obtida com tamanho do lote = 32 e taxa de aprendizagem = 10^{-3} , alcançando 88,24% de acurácia. As piores configurações ocorreram para lote = 128 com taxa de aprendizagem = 10^{-5} e para lote = 32 com taxa de aprendizagem = 10^{-4} , ambas resultando em 84,31%. Isso representa uma diferença absoluta de 3,93 p.p. em relação ao melhor cenário.

Tabela 19 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto *Drishti-GSI*.

Tamanho do lote / Taxa de aprendizagem	10^{-3}	10^{-4}	10^{-5}
16	86,27	86,27	86,27
32	88,24	84,31	86,27
64	86,27	86,27	86,27
128	86,27	86,27	84,31

Fonte: Autor (2024).

Quando analisadas isoladamente, as taxas de aprendizagem apresentaram os seguintes

desempenhos médios: 10^{-3} (86,76%), 10^{-4} (85,78%) e 10^{-5} (85,78%). Esses resultados confirmam a vantagem de taxas mais moderadas, enquanto valores excessivamente baixos não proporcionaram ganhos adicionais. A variação relativamente alta entre as configurações (até 3,93 p.p.) indica que o *Drishti-GSI* é mais sensível a ajustes de hiperparâmetros, possivelmente em função do seu tamanho reduzido e da maior heterogeneidade intrínseca.

Embora a melhor configuração local tenha sido lote = 32 e taxa = 10^{-3} , para garantir consistência entre bases nos experimentos comparativos foi adotada a configuração global definida na Subseção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso): lote = 32 e taxa de aprendizagem = 10^{-4} . Apesar de a configuração local (10^{-3}) ter apresentado uma acurácia 3,93 p.p. superior à pior configuração testada, a configuração global adotada (10^{-4}) ainda assim manteve desempenho consistente, atingindo 86,27% de acurácia e preservando estabilidade em outros conjuntos de dados. Essa padronização evita resultados inflados por ajustes específicos de base e está alinhada ao objetivo central deste trabalho: garantir a capacidade de generalização do modelo em cenários clínicos heterogêneos, em que a consistência entre bases é mais importante do que a otimização pontual.

5.1.3 ORIGA

A análise dos resultados de hiperparâmetros do conjunto ORIGA (Tabela 20) indica que a melhor configuração foi alcançada com tamanho do lote = 16 e taxa de aprendizagem = 10^{-3} , atingindo 80,00% de acurácia. Em contraste, o pior desempenho ocorreu com tamanho do lote = 64 e taxa de aprendizagem = 10^{-5} , resultando em 77,44%. A diferença absoluta de 2,56 p.p. entre os cenários extremos evidencia apenas uma influência moderada da escolha de hiperparâmetros neste conjunto, sugerindo que o ruído e a complexidade inerentes aos dados do ORIGA podem limitar os ganhos obtidos por ajustes finos.

Tabela 20 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto ORIGA.

Tamanho do lote / Taxa de aprendizagem	10^{-3}	10^{-4}	10^{-5}
16	80,00	77,95	77,95
32	78,97	78,97	77,95
64	78,46	79,49	77,44
128	79,49	77,95	77,95

Fonte: Autor (2024).

Quando analisadas isoladamente, as taxas de aprendizagem apresentaram os seguintes

desempenhos médios: 10^{-3} (79,23%), 10^{-4} (78,59%) e 10^{-5} (77,82%). Observa-se que taxas mais altas favoreceram a convergência, especialmente em lotes menores, ao passo que taxas muito baixas induziram subtreinamento. A variação de até 2,56 p.p. reforça que o ORIGA é sensível a ajustes de hiperparâmetros, possivelmente em função de sua heterogeneidade e variabilidade de aquisição.

Embora a melhor configuração local tenha sido lote = 16 e taxa = 10^{-3} , para garantir consistência entre bases nos experimentos comparativos foi adotada a configuração global definida na Subseção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso): lote = 32 e taxa de aprendizagem = 10^{-4} . Ainda que a configuração local tenha superado em 2,56 p.p. a pior, a configuração global (10^{-4}) se mostrou estável, atingindo 78,97% de acurácia e mantendo desempenho consistente em outros conjuntos de dados. Essa padronização evita vieses decorrentes de ajustes específicos de base e está alinhada ao objetivo central deste trabalho: garantir a capacidade de generalização em cenários clínicos heterogêneos, nos quais a estabilidade é mais relevante do que a otimização pontual.

5.1.4 REFUGE

Os resultados de hiperparâmetros do conjunto REFUGE (Tabela 21) indicam que o melhor desempenho (93,33% de acurácia) foi alcançado com múltiplas configurações: tamanhos de lote 16 ou 128 combinados com taxas de aprendizagem 10^{-3} ou 10^{-4} . Em contraste, o pior desempenho (92,50%) ocorreu com a configuração lote = 32 e taxa = 10^{-3} . A diferença absoluta de apenas 0,83 p.p. entre os cenários extremos evidencia a baixa sensibilidade do modelo a variações de hiperparâmetros neste conjunto, sugerindo estabilidade de desempenho no banco REFUGE.

Tabela 21 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto REFUGE.

Tamanho do lote / Taxa de aprendizagem	10^{-3}	10^{-4}	10^{-5}
16	93,33	93,33	93,33
32	92,50	93,33	93,33
64	92,50	93,33	93,33
128	93,33	93,33	92,50

Fonte: Autor (2024).

Quando analisadas isoladamente, todas as taxas de aprendizagem apresentaram desempenho médio semelhante (aproximadamente 93%). Essa baixa variabilidade confirma que o

REFUGE é o conjunto mais estável entre os avaliados, com diferença inferior a 1 p.p. entre a melhor e a pior configuração. Esse comportamento indica que, em um domínio mais padronizado, o modelo se mostra pouco sensível a ajustes de hiperparâmetros, reforçando sua aplicabilidade em cenários clínicos desbalanceados.

Embora a melhor configuração local tenha atingido 93,33% de acurácia, a configuração global definida na Subseção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso) — lote = 32 e taxa de aprendizagem = 10^{-4} apresentou o mesmo valor, demonstrando estabilidade. Essa equivalência prática reforça que a padronização adotada não compromete o desempenho e, ao contrário, contribui para o objetivo central deste trabalho: garantir generalização estável em múltiplos domínios clínicos, priorizando consistência em relação a variações pontuais.

5.1.5 RIM-ONE

A análise dos resultados de hiperparâmetros do conjunto RIM-ONE (Tabela 22) indica que o melhor desempenho (97,87% de acurácia) foi alcançado com tamanho do lote = 64 e taxa de aprendizagem = 10^{-4} . Em contraste, o pior desempenho (96,81%) ocorreu com tamanho do lote = 16 e taxa = 10^{-3} . A diferença absoluta de apenas 1,06 p.p. entre os cenários extremos evidencia a baixa sensibilidade do modelo a variações de hiperparâmetros, confirmando a estabilidade de desempenho no conjunto RIM-ONE.

Tabela 22 – Variação de desempenho (acurácia em %) em diferentes tamanhos de lote e taxas de aprendizagem no conjunto RIM-ONE.

Tamanho do lote / Taxa de aprendizagem	10^{-3}	10^{-4}	10^{-5}
16	96,81	97,16	97,16
32	97,16	96,81	97,16
64	97,16	97,87	97,16
128	97,16	97,16	96,81

Fonte: Autor (2024).

Quando analisadas isoladamente, as taxas de aprendizagem apresentaram os seguintes desempenhos médios: 10^{-4} (97,25%), 10^{-3} (97,07%) e 10^{-5} (97,07%). Essa proximidade confirma que o RIM-ONE é um conjunto estável e homogêneo, no qual diferentes combinações de hiperparâmetros produzem resultados elevados e próximos entre si.

Embora a melhor configuração local tenha sido lote = 64 e taxa = 10^{-4} , a configuração global definida na Subseção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso) — lote = 32 e taxa de aprendizagem = 10^{-4} — também se mostrou estável, atingindo 96,81% de acurácia.

A diferença de apenas 1,06 p.p. em relação à melhor configuração local (97,87%) é marginal e reforça a consistência do modelo neste domínio. Essa padronização não compromete o desempenho e, ao contrário, reforça o objetivo central deste trabalho: garantir que o modelo mantenha resultados previsíveis e generalizáveis em cenários clínicos reais, mesmo diante de variações de parâmetros.

5.1.6 Síntese e Análise Comparativa dos Resultados de Hiperparametrização

De forma consolidada, os experimentos de hiperparametrização evidenciam três tendências principais entre as cinco bases avaliadas:

- **Tamanhos de lote intermediários (32 ou 64)** apresentaram os melhores resultados, equilibrando a variabilidade estocástica (que auxilia na evasão de mínimos locais) e a estabilidade na estimativa do gradiente. Esse padrão foi observado no ACRIMA (96,23%, lote 32) e no RIM-ONE (97,87%, lote 64).
- **Taxas de aprendizagem moderadas (10^{-3} e 10^{-4})** mostraram maior eficácia, superando consistentemente taxas excessivamente baixas (10^{-5}), que não proporcionaram ganhos adicionais e, em alguns casos, resultaram em desempenho inferior por convergência lenta ou subtreinamento. Esse efeito foi mais evidente no *Drishti-GSI* e no ORIGA.
- **Estabilidade depende da homogeneidade da base de dados.** Conjuntos mais homogêneos e de maior qualidade, como RIM-ONE e REFUGE, apresentaram baixa variabilidade entre as combinações testadas (diferenças absolutas de apenas 1,06 p.p. e 0,83 p.p., respectivamente). Já bases mais heterogêneas e desbalanceadas, como ORIGA e *Drishti-GSI*, apresentaram maior sensibilidade, com variações de até 2,56 p.p. e 3,93 p.p., respectivamente.

A síntese comparativa dos resultados de hiperparâmetros, apresentada na Tabela 23, evidencia a melhor e a pior configuração para cada conjunto de dados, bem como a diferença absoluta de desempenho em pontos percentuais.

Esses resultados demonstram que, embora cada base apresente uma configuração localmente ótima, a escolha da configuração global — lote = 32 e taxa de aprendizagem = 10^{-4} — representa o melhor compromisso entre desempenho e consistência entre conjuntos de dados. Esta decisão estratégica evita a superotimização em bases específicas, assegura a comparabilidade direta entre os resultados dos diferentes conjuntos de dados e, mais importante, está diretamente

Tabela 23 – Síntese comparativa dos resultados de hiperparametrização por base de dados.

Base	Melhor Configuração	Pior Configuração	Δ p.p.
ACRIMA	Lote 32, 10^{-3} (96,23%)	Lote 128, 10^{-3} (93,86%)	2,37
<i>Drishti-GS1</i>	Lote 32, 10^{-3} (88,24%)	Lote 32, 10^{-4} ou Lote 128, 10^{-5} (84,31%)	3,93
ORIGA	Lote 16, 10^{-3} (80,00%)	Lote 64, 10^{-5} (77,44%)	2,56
REFUGE	Lote 16 ou 128, 10^{-3} ou 10^{-4} (93,33%)	Lote 32, 10^{-3} (92,50%)	0,83
RIM-ONE	Lote 64, 10^{-4} (97,87%)	Lote 16, 10^{-3} (96,81%)	1,06

Fonte: Autor (2024).

alinhada ao objetivo central deste trabalho: garantir que o modelo mantenha capacidade de generalização consistente em cenários clínicos heterogêneos. Em aplicações reais, onde a diversidade de equipamentos, protocolos e populações é a regra, a previsibilidade e a estabilidade do desempenho frente a diferentes domínios são qualidades operacionais indiscutivelmente mais relevantes do que ganhos marginais obtidos em ajustes locais e específicos.

Encerrada a análise de hiperparametrização, pode-se concluir que a configuração global definida na Subseção 4.6.2 (Ajuste de Hiperparâmetros — Modelo Denso) (lote = 32, taxa de aprendizagem = 10^{-4}) representa o ponto de equilíbrio ideal entre desempenho e consistência entre bases. A partir desta definição, todos os experimentos subsequentes foram conduzidos sob este mesmo protocolo, garantindo comparabilidade direta entre bases e eliminando vieses decorrentes de otimizações locais. Com os hiperparâmetros fixados, a próxima etapa de investigação concentra-se na avaliação dos classificadores individuais (SVM, RF, GB e modelo denso) e, posteriormente, na análise crítica do *ensemble*, de modo a verificar em que medida a combinação dos modelos contribui para aumentar a estabilidade, a consistência e a confiabilidade da detecção de glaucoma.

5.2 Classificadores Individuais vs Ensemble

Para avaliar criticamente a contribuição de cada componente do sistema e validar a estratégia de *ensemble* proposta no Capítulo 4 (Materiais e Métodos), realizou-se uma análise comparativa abrangente entre os classificadores individuais — modelo denso, SVM, RF e GB — e o *ensemble* que combina suas predições por média ponderada.

Os resultados foram obtidos a partir da média e do desvio padrão calculados nos cinco

folds da validação cruzada estratificada, utilizando a configuração otimizada de hiperparâmetros (Seção 4.6). Os resultados sintetizados na Tabela 24 evidenciam o contraste entre os classificadores individuais e a estratégia de *ensemble*, destacando que, embora o Modelo Denso alcance as maiores médias de acurácia, o *ensemble* se diferencia pela redução da variabilidade e pela maior previsibilidade de desempenho.

Tabela 24 – Desempenho comparativo (acurácia média $\pm$ desvio padrão) entre os classificadores. Valores em porcentagem.

Base	Modelo Denso	SVM	RF	GB	Ensemble
ACRIMA	99,79 $\pm$ 0,19	96,81 $\pm$ 0,52	93,48 $\pm$ 0,60	95,53 $\pm$ 0,37	98,98 $\pm$ 0,26
<i>Drishiti-GSI</i>	98,89 $\pm$ 1,95	91,09 $\pm$ 3,90	86,72 $\pm$ 3,26	93,07 $\pm$ 2,81	96,36 $\pm$ 1,67
ORIGA	94,18 $\pm$ 2,69	78,71 $\pm$ 0,78	78,38 $\pm$ 0,68	80,42 $\pm$ 0,73	85,65 $\pm$ 2,19
REFUGE	98,58 $\pm$ 1,96	95,97 $\pm$ 0,48	94,05 $\pm$ 0,51	95,51 $\pm$ 0,42	96,87 $\pm$ 0,76
RIM-ONE	97,71 $\pm$ 1,05	88,68 $\pm$ 1,24	82,20 $\pm$ 1,78	86,03 $\pm$ 1,53	93,36 $\pm$ 1,43
TOTAL	96,48 $\pm$ 3,43	81,68 $\pm$ 1,83	79,27 $\pm$ 1,50	82,76 $\pm$ 1,73	90,22 $\pm$ 1,46

Fonte: Autor (2024).

A análise revela dois padrões principais que validam a arquitetura do ensemble. O primeiro diz respeito ao cenário TOTAL, o mais heterogêneo e clinicamente relevante por refletir a diversidade de equipamentos, protocolos e populações. Nesse conjunto, o desvio padrão caiu de 3,43 p.p. (Modelo Denso) para 1,46 p.p. (*Ensemble*), correspondendo a uma redução relativa de 57,4%. Essa diminuição drástica da variabilidade traduz-se em ganho de confiabilidade, com desempenho previsível frente à heterogeneidade inerente a dados reais. Em contextos de triagem, tal previsibilidade sobrepõe-se a ganhos marginais em acurácia máxima.

O segundo padrão emerge em bases desafiadoras e desbalanceadas, como o ORIGA, onde o *ensemble* (85,65%) superou o melhor classificador individual — GB (80,42%) — por 5,23 p.p. A margem expressiva confirma o princípio de que a complementaridade de vieses indutivos — SVM (margens), RF (*bagging*), GB (*boosting*) e Modelo Denso (redes neurais) — permite compensar erros idiossincráticos e capturar padrões complexos, assegurando resiliência preditiva mesmo em cenários difíceis.

Para quantificar a contribuição individual de cada modelo, conduziu-se um estudo de ablação sistemático, cujos resultados são apresentados na Tabela 25. A remoção sequencial de cada classificador resultou em degradação consistente do desempenho, confirmando que todos os componentes agregam valor único à arquitetura proposta.

O Modelo Denso mostrou-se o contribuidor mais significativo (queda de 3,07 p.p. quando

Tabela 25 – Análise de ablação mostrando o impacto da remoção de cada modelo no desempenho do *ensemble* no conjunto TOTAL.

Configuração do <i>Ensemble</i>	Acurácia Média (%)
Completo (todos os modelos)	90,22
Sem Modelo Denso	87,15
Sem SVM	89,43
Sem RF	89,64
Sem GB	88,37

Fonte: Autor (2024).

excluído), seguido pelo GB (queda de 1,85 p.p.), refletindo sua capacidade complementar de capturar padrões não lineares complexos. Entretanto, mesmo a exclusão do modelo menos impactante (RF, $-0,58$ p.p.) causou redução mensurável, evidenciando que a diversidade de abordagens — de redes neurais a métodos baseados em árvores — é essencial para o desempenho ótimo.

Estes resultados empiricamente validam a sinergia entre os quatro paradigmas de aprendizado adotados, onde a combinação de vieses indutivos diversos (baseados em margens, *bagging*, *boosting* e redes neurais) cria um sistema mais resiliente do que qualquer modelo isolado. A análise corrobora que o valor do *ensemble* transcende a soma de suas partes, justificando a complexidade adicional através de ganhos tangíveis em estabilidade e capacidade generalizante.

Em síntese, os resultados obtidos respondem de forma consistente à questão central desta investigação: a estratégia de *ensemble* apresenta-se como uma vantagem estratégica — não apenas como mero *overhead* computacional. Embora o Modelo Denso apresente maior acurácia, sua alta dispersão ($\pm 3,43$ p.p.) compromete a confiabilidade. O *ensemble* reduz essa variabilidade, oferecendo desempenho mais estável ($\pm 1,46$ p.p.). No cenário agregado (TOTAL), posiciona-se como a solução metodologicamente mais alinhada com as demandas de um sistema de diagnóstico assistido: generalizabilidade, previsibilidade e segurança preditiva. O custo computacional extra justifica-se pelo ganho em confiabilidade, atributo indispensável para aplicações clínicas, nos quais as decisões dependem da consistência mais do que do pico de acurácia.

Essa abordagem confirma o valor de arquiteturas híbridas em cenários médicos complexos, onde a incerteza precisa ser gerida de forma explícita. Uma vez estabelecida a relevância do *ensemble* em termos de estabilidade e generalização, uma investigação mais granular sobre seu comportamento discriminativo por classe torna-se necessária. A próxima seção analisa matrizes de confusão e métricas por categoria (glaucoma vs. normal), identificando padrões de erro,

vieses preditivos e os desafios remanescentes em bases específicas.

5.3 Análise Discriminativa e Padrões de Erro a partir das Matrizes de Confusão

Esta seção apresenta uma análise detalhada do desempenho discriminativo do modelo proposto, estratificada por classe (glaucomatosa vs. normal) e por base de dados. Os resultados, consolidados na Tabela 26 e nas matrizes de confusão da Figura 27, são criticamente discutidos à luz das características específicas de cada conjunto, com foco na identificação de padrões de erro e suas possíveis causas.

5.3.1 Métricas de Desempenho por Classe

Os dados resumidos na Tabela 26 detalham o desempenho do modelo para cada classe em todos os conjuntos de dados avaliados, evidenciando variações significativas entre conjuntos homogêneos e heterogêneos. Essas diferenças refletem não apenas variações de desempenho, mas também implicações clínicas relevantes. Padrões semelhantes são ilustrados visualmente na Figura 27, que apresenta a distribuição de acertos e erros (VP, FN, FP, VN) por base de dados.

Tabela 26 – Desempenho discriminativo completo por classe: comparação entre casos glaucomatosos e normais em conjuntos públicos de imagens de retina (valores em %).

Conjunto de dados	Classe	Acurácia (Global)	Precisão	Sensibilidade (Especificidade)	F1-score
ACRIMA	Glaucoma	96,23	95,12	98,32	96,69
	Normal		97,75	93,55	95,60
Drishti-GS1	Glaucoma	88,24	86,36	100,00	92,68
	Normal		100,00	53,85	70,00
ORIGA	Glaucoma	80,00	68,97	40,00	50,63
	Normal		81,93	93,79	87,46
REFUGE	Glaucoma	93,33	100,00	33,33	50,00
	Normal		93,10	100,00	96,43
RIM-ONE	Glaucoma	97,87	96,00	97,96	96,97
	Normal		98,90	97,83	98,36
TOTAL	Glaucoma	90,23	87,95	85,17	86,54
	Normal		91,50	93,19	92,34

Nota: Para a classe Normal, a métrica apresentada como “Sensibilidade” corresponde à **Especificidade** (proporção de verdadeiros negativos). A acurácia é reportada apenas como medida global, não decomposta por classe.
Fonte: Autor (2024).

5.3.1.1 Bases Equilibradas e de Alta Performance (ACRIMA e RIM-ONE)

Os conjuntos ACRIMA e RIM-ONE exibiram o desempenho mais consistente e balanceado da avaliação. No RIM-ONE, todas as métricas permaneceram acima de 97%, com sensibilidade (97,96%) e especificidade (97,83%) em patamares praticamente equivalentes. Esse desempenho consistente confirma a capacidade do modelo em generalizar para bases homogêneas caracterizadas por protocolos de aquisição padronizados, boa iluminação e alto contraste das imagens, corroborando achados de estudos prévios que também reportaram estabilidade nesta base (LI *et al.*, 2019; SHYAMALEE; MEEDENIYA, 2022).

No ACRIMA, observou-se um desempenho elevado, ainda que levemente assimétrico, com sensibilidade (98,32%) superando ligeiramente a especificidade (93,55%). A precisão se manteve acima de 95%. Esse alto desempenho é visualmente confirmado na Figura 27, onde se observa a predominância de valores na diagonal principal da matriz de confusão, com contagens absolutas de falsos negativos e falsos positivos muito baixas.

Clinicamente, esse equilíbrio representa um cenário de alta confiabilidade para triagem, onde as taxas de ambos os erros — subdiagnóstico (falsos negativos) e encaminhamentos desnecessários (falsos positivos) — são minimizadas, potencialmente tornando o modelo adequado para um primeiro filtro automatizado no fluxo de diagnóstico oftalmológico.

5.3.1.2 Bases com Desempenho Assimétrico (Drishti-GS1, ORIGA e REFUGE)

A análise revela desafios críticos impostos pela heterogeneidade e pelo desbalanceamento dos conjuntos. O *Drishti-GS1* apresentou sensibilidade máxima (100%), mas especificidade substancialmente reduzida (53,85%), conforme ilustrado na Figura 27. Este padrão indica que o modelo desenvolveu viés pronunciado pela classe majoritária, classificando incorretamente 6 dos 13 casos normais como falsos positivos. Em contexto clínico, isso se traduziria em excesso de encaminhamentos desnecessários, sobrecarregando o sistema de saúde. Estratégias como ajuste de limiar de decisão ou amostragem preferencial da classe minoritária durante o treinamento poderiam mitigar este viés.

O REFUGE apresentou padrão oposto: especificidade perfeita (100%) mas sensibilidade insuficiente para uso clínico (33,33%). O modelo mostrou-se excessivamente conservador, falhando em detectar 20 dos 30 casos de glaucoma. Este comportamento, potencialmente causado pelo forte desbalanceamento do conjunto, representa risco clínico grave de subdiagnóstico. Técnicas como loss functions ponderadas ou *data augmentation* direcionada para a classe

positiva são recomendadas para aplicações futuras.

O conjunto ORIGA apresentou o cenário mais crítico, com sensibilidade limitada (40,00%) apesar da alta especificidade (93,79%). O modelo falhou em identificar 30 dos 50 casos positivos, indicando dificuldade em capturar padrões sutis de glaucoma inicial ou sensibilidade a artefatos de aquisição. A estratégia de *data augmentation* utilizada mostrou-se insuficiente, sugerindo necessidade de abordagens mais sofisticadas como aprendizado por transferência com conjuntos mais diversificados ou normalização baseada em domínio.

Em síntese, bases desbalanceadas como *Drishti-GS1*, REFUGE e ORIGA expuseram vulnerabilidades distintas do modelo: tendência a superdiagnóstico, subdiagnóstico ou insensibilidade a variações sutis. Tais padrões reforçam que, para aplicação clínica consistente, é indispensável incorporar estratégias de balanceamento e adaptação por domínio.

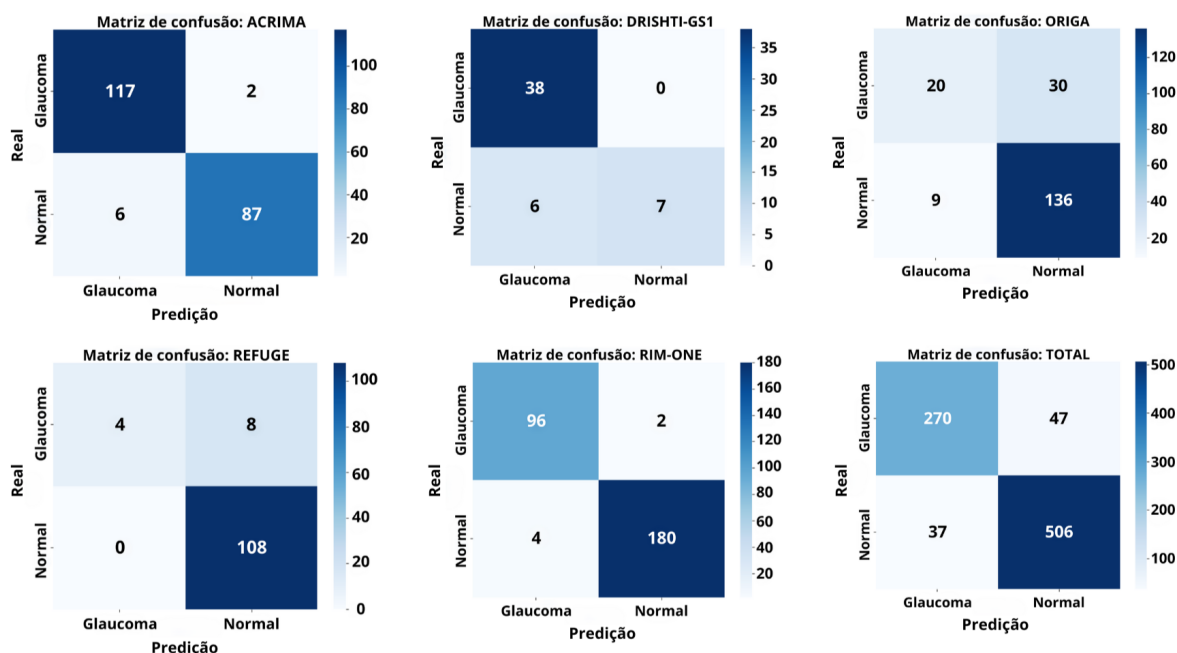
5.3.2 Análise Qualitativa de Erros a partir das Matrizes de Confusão

A visualização apresentada na Figura 27 complementa de forma essencial a análise numérica da Tabela 26, destacando os padrões de erro do modelo. Para garantir clareza na interpretação, adotam-se os seguintes termos: Verdadeiros Positivos (VP), casos de glaucoma corretamente identificados; Falsos Negativos (FN), casos de glaucoma erroneamente classificados como normais; Falsos Positivos (FP), casos normais erroneamente classificados como glaucoma; e Verdadeiros Negativos (VN), casos normais corretamente identificados.

No ACRIMA, a matriz revela 117 VP, apenas 2 FN e 6 FP, valores que corroboram a alta sensibilidade (98,32%) e especificidade (93,55%) reportadas. De forma semelhante, o RIM-ONE exibiu 96 VP, 2 FN e apenas 4 FP, traduzindo-se em métricas superiores a 97% em todas as categorias. Estes padrões visuais confirmam que, em bases homogêneas e de alta qualidade, os erros são minimizados e o desempenho do modelo é consistentemente alto.

Em contraste, bases com heterogeneidade ou desbalanceamento revelaram padrões assimétricos críticos. No *Drishti-GS1*, todos os 38 casos glaucomatosos foram corretamente identificados (VP), mas 6 dos 13 casos normais foram classificados como glaucoma (FP), resultando em especificidade substancialmente reduzida (53,85%). De forma oposta, no REFUGE, observou-se o padrão inverso: 20 dos 30 casos glaucomatosos foram erroneamente classificados como normais (FN), enquanto todos os casos normais foram corretamente identificados (especificidade de 100%). O conjunto ORIGA apresentou o cenário mais crítico, com 30 FN em 50 casos glaucomatosos (sensibilidade de 40,00%), apesar de alta especificidade (93,79%).

Figura 27 – Matrizes de confusão de ACRIMA, *Drishiti-GSI*, ORIGA, REFUGE, RIM-ONE e TOTAL.



Fonte: Autor (2024).

O conjunto TOTAL, que melhor representa um cenário clínico realista, apresentou 270 VP e 506 VN, mas também 47 FN e 37 FP. Esta distribuição evidencia um compromisso (*trade-off*) entre sensibilidade e especificidade, com um viés moderado em favor da redução de falsos positivos - reduzindo encaminhamentos desnecessários, mas ainda com uma taxa relevante de casos não detectados.

A distribuição diferenciada de erros conforme as características das bases tem implicações clínicas diretas:

- Em bases homogêneas (ACRIMA e RIM-ONE), a minimização equilibrada de ambos os erros (FN e FP) sugere que o modelo é adequado para triagem inicial nestes contextos específicos.
- Em bases desbalanceadas (*Drishiti-GSI*, REFUGE e ORIGA), prevalecem vieses assimétricos que comprometem ou a sensibilidade ou a especificidade, limitando a aplicabilidade clínica direta sem ajustes e/ou adaptação por domínio.
- No conjunto TOTAL, o equilíbrio observado mantém-se dentro de parâmetros aceitáveis para aplicação clínica inicial, mas a presença de 47 falsos negativos destaca a necessidade crítica de estratégias adicionais para reduzir subdiagnósticos sem elevar excessivamente os encaminhamentos desnecessários.

5.3.2.1 Falsos Negativos (FN) e Impacto Clínico

A ocorrência de Falsos Negativos (FN) constitui a falha mais crítica em sistemas de triagem oftalmológica, uma vez que representa casos de glaucoma que seriam incorretamente liberados durante um processo de rastreamento.

A análise dos padrões de erro apresentados na Figura 27 revela que os conjuntos ORIGA (30 FN) e TOTAL (47 FN) registraram as maiores quantidades absolutas de falsos negativos. Quando normalizados pelo tamanho das bases, esses valores correspondem a 60% dos casos positivos no ORIGA e 14,8% no TOTAL, evidenciando um risco acentuado de subdiagnóstico no primeiro caso.

As implicações clínicas são diretas: a progressão não detectada da doença em pacientes que recebem um falso resultado de triagem. Uma análise qualitativa dos casos específicos que resultaram em FN — não contemplada no escopo deste estudo, mas recomendada para trabalhos futuros — poderia investigar fatores determinantes, como:

1. estágios iniciais da doença com alterações morfológicas sutis,
2. artefatos de aquisição de imagem, ou
3. variações anatômicas que confundem a detecção automática.

Do ponto de vista da implementação clínica, a redução de falsos negativos deve ser uma prioridade. Estratégias como o uso de loss functions ponderadas (atribuindo maior peso à classe positiva), *data augmentation* direcionada para exemplos glaucomatosos, ou o ajuste fino dos limiares de decisão são caminhos promissores para mitigar este problema. É importante notar que tais ajustes podem implicar um aumento considerado clinicamente tolerável de falsos positivos, balanceando o risco entre sobrediagnóstico e subdiagnóstico.

5.3.2.2 Falsos Positivos (FP) e Especificidade

Embora clinicamente menos críticos que os falsos negativos, os falsos positivos acarretam implicações operacionais relevantes, incluindo custos desnecessários com exames complementares, ansiedade desencadeada em pacientes e sobrecarga do sistema de saúde devido a encaminhamentos desnecessários para investigação especializada.

O conjunto *Drishti-GSI* destacou-se com o maior impacto proporcional de FPs, registrando 6 falsos positivos em uma base de apenas 13 imagens normais - equivalente a uma taxa

de 46,2% de alarmes falsos nesta classe. Este resultado está alinhado com a baixa especificidade (53,85%) reportada na Tabela 26, sugerindo que as características aprendidas pelo modelo não foram suficientemente discriminativas para a classe normal, possivelmente em função do desbalanceamento intrínseco deste conjunto.

Para mitigar FPs, técnicas de calibração pós-treinamento, ajuste de limiares de decisão ou mesmo a incorporação de etapas de pré-processamento mais agressivas para normalização de domínio mostram-se como alternativas viáveis. É importante notar que tais ajustes devem ser realizados com cuidado para não comprometer a sensibilidade do modelo, mantendo o equilíbrio adequado entre os dois tipos de erro de acordo com o contexto de aplicação clínica pretendido.

5.3.2.3 Conjunto TOTAL: Um Cenário Realista

A matriz de confusão do conjunto TOTAL apresentada na Figura 27 representa o cenário mais fiel às condições de aplicação clínica real, agregando a heterogeneidade inerente a diferentes fontes de dados e protocolos de aquisição. O modelo demonstrou solidez ao classificar corretamente a maioria das amostras (776 acertos entre 860, correspondendo a 90,23% de acurácia global), com 270 verdadeiros positivos e 506 verdadeiros negativos, além de 47 falsos negativos e 37 falsos positivos.

Entretanto, a presença de 47 falsos negativos (5,4% do total de imagens) e 37 falsos positivos (4,3% do total de imagens) revela oportunidades de refinamento. A distribuição assimétrica desses erros - com predomínio de falsos negativos - indica um viés conservador do modelo, que prioriza a especificidade (93,2%) em detrimento da sensibilidade (85,2%), resultando em um *trade-off* clinicamente relevante. Este compromisso operacional, embora beneficie a redução de encaminhamentos desnecessários, eleva o risco de subdiagnóstico em contextos de triagem.

Para implementação clínica, estes resultados sintetizam os padrões identificados nas bases individuais: a tendência a subdiagnóstico observada em ORIGA e REFUGE (evidenciada pelos 47 falsos negativos) e a propensão a superdiagnóstico registrada no *Drishti-GS1* (refletida nos 37 FPs). Assim, o TOTAL confirma-se como o cenário de avaliação mais realista, consolidando tanto os pontos fortes quanto as vulnerabilidades do modelo frente à heterogeneidade clínica esperada. Esta análise indica claramente a necessidade de:

1. ajustes específicos para aumentar a sensibilidade sem comprometer excessivamente a especificidade;

2. validação prospectiva em ambientes multicêntricos; e
3. desenvolvimento de limiares diferenciados conforme o perfil de risco da população-alvo.

5.3.3 Síntese e Implicações

Os resultados desta análise estratificada permitem concluir que:

- O modelo proposto demonstra alta eficácia e equilíbrio em bases de dados homogêneas e bem equilibradas (RIM-ONE, ACRIMA), com métricas consistentes acima de 95% e baixa variabilidade, indicando confiabilidade para aplicação em contextos controlados.
- O desbalanceamento de classes é um fator crítico que induz vieses de classificação pronunciados, levando o modelo a supervalorizar sistematicamente a classe majoritária - como evidenciado pela alta sensibilidade/baixa especificidade no *Drishti-GS1* (100%/53,85%) e pelo padrão inverso no REFUGE (33,33%/100%). Estratégias de aumento de dados e ponderação de classes mostraram-se benéficas, porém insuficientes, requerendo ajustes específicos por domínio.
- O desempenho limitado no conjunto ORIGA (sensibilidade de 40,00%) sugere que fatores além do balanceamento - como variabilidade na aquisição das imagens, heterogeneidade de protocolos ou estágios iniciais da doença - impõem desafios adicionais à generalização do modelo. Estudos anteriores corroboram a complexidade inerente a este conjunto, indicando que abordagens de aprendizado por domínio podem ser necessárias.
- O desempenho no conjunto TOTAL (acurácia: 90,23%, sensibilidade: 85,17%, especificidade: 93,19%) valida a utilidade do modelo em cenários realistas, funcionando como um *proxy* multicêntrico que consolida as vulnerabilidades observadas nas bases individuais. Os resultados evidenciam o *trade-off* entre sensibilidade e especificidade, com viés em favor da especificidade. Este balanceamento, ainda que benéfico para reduzir custos operacionais, pode ser clinicamente limitante em contextos de triagem onde falsos negativos têm consequências graves.

Em síntese, os resultados confirmam o potencial do modelo para triagem automatizada de glaucoma, mas destacam a importância crítica de:

1. estratégias específicas para mitigação de vieses em bases desbalanceadas;

- 2. técnicas de adaptação de domínio para melhor generalização; e
- 3. customização essencial dos limiares operacionais para alinhar o desempenho do modelo a diferentes perfis de risco populacional e contextos clínicos específicos.

5.4 Desempenho Geral do Modelo Proposto na Detecção de Glaucoma em Imagens de Fundo de Olho

5.4.1 Análise Consolidada de Desempenho

Nesta subseção, precisão, sensibilidade e *F1-score* são reportadas como *weighted average* por classe (ponderação pelo número de amostras), mitigando o impacto do desbalanceamento entre as classes positiva e negativa e preservando a comparabilidade entre conjuntos heterogêneos. A acurácia é reportada como métrica global de acertos, enquanto a AUC corresponde à área sob a curva ROC, refletindo a capacidade discriminativa do modelo em múltiplos limiares de decisão. Os valores numéricos detalhados estão na Tabela 27, enquanto a visualização na Figura 28 favorece a identificação de padrões, tendências e *trade-offs* entre os diferentes conjuntos e métricas.

Tabela 27 – Desempenho do modelo proposto em conjuntos públicos de imagens de fundo de olho (valores em %, exceto AUC).

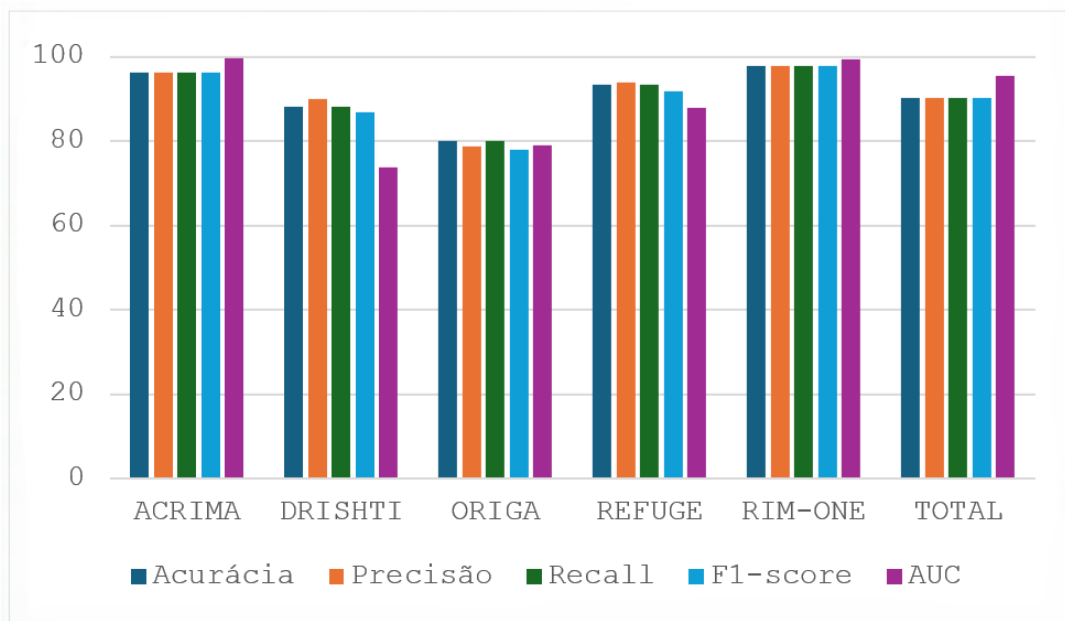
Conjunto de dados	Acurácia	Precisão	Sensibilidade	<i>F1-score</i>	AUC
ACRIMA	96,23	96,28	96,23	96,22	0,996
<i>Drishti-GS1</i>	88,24	89,84	88,24	86,90	0,737
ORIGA	80,00	78,60	80,00	78,02	0,789
REFUGE	93,33	93,79	93,33	91,79	0,879
RIM-ONE	97,87	97,89	97,87	97,88	0,995
TOTAL	90,23	90,19	90,23	90,20	0,954

Fonte: Autor (2024).

A análise conjunta dos dados da Tabela 27 e da Figura 28 revela padrões de desempenho que transcendem os valores absolutos:

- *Cluster* de Alta Performance: a visualização gráfica confirma de forma imediata que RIM-ONE e ACRIMA formam um grupo distinto de excelência, com todas as métricas consistentemente acima de 96%. A proximidade visual das barras para cada base indica um

Figura 28 – Desempenho comparativo do modelo proposto nos diferentes bancos de dados.



Fonte: Autor (2024).

equilíbrio ideal entre precisão e sensibilidade, corroborado por valores de AUC próximos à perfeição (0,995 e 0,996).

- **Identificação Visual de Desafios:** a visualização na Figura 28 evidencia de forma clara a diferença de desempenho dos conjuntos ORIGA e *Drishiti-GS1*. A altura consistentemente menor das barras e a maior variabilidade entre métricas — em especial a diferença acentuada entre Precisão e *F1-score* no *Drishiti-GS1* — indicam os desafios impostos pela qualidade da imagem e pelo desbalanceamento extremo.
- **Análise de *Trade-offs*:** enquanto a tabela fornece os números, o gráfico facilita a análise de compensações. Por exemplo, observa-se que o REFUGE atinge um equilíbrio notável entre Precisão (93,79%) e Sensibilidade (93,33%), com barras de altura quase idêntica na visualização, um feito considerando o desbalanceamento da base.
- **Performance no Cenário Realista (TOTAL):** o conjunto TOTAL, apesar de agregar heterogeneidade, mantém um perfil de barras altas e equilibradas, situando-se próximo do *cluster* de alta performance (RIM-ONE, ACRIMA). Este é também o cenário mais próximo da aplicação clínica real, validando visualmente a consistência do modelo em contextos práticos (AUC = 0,954).

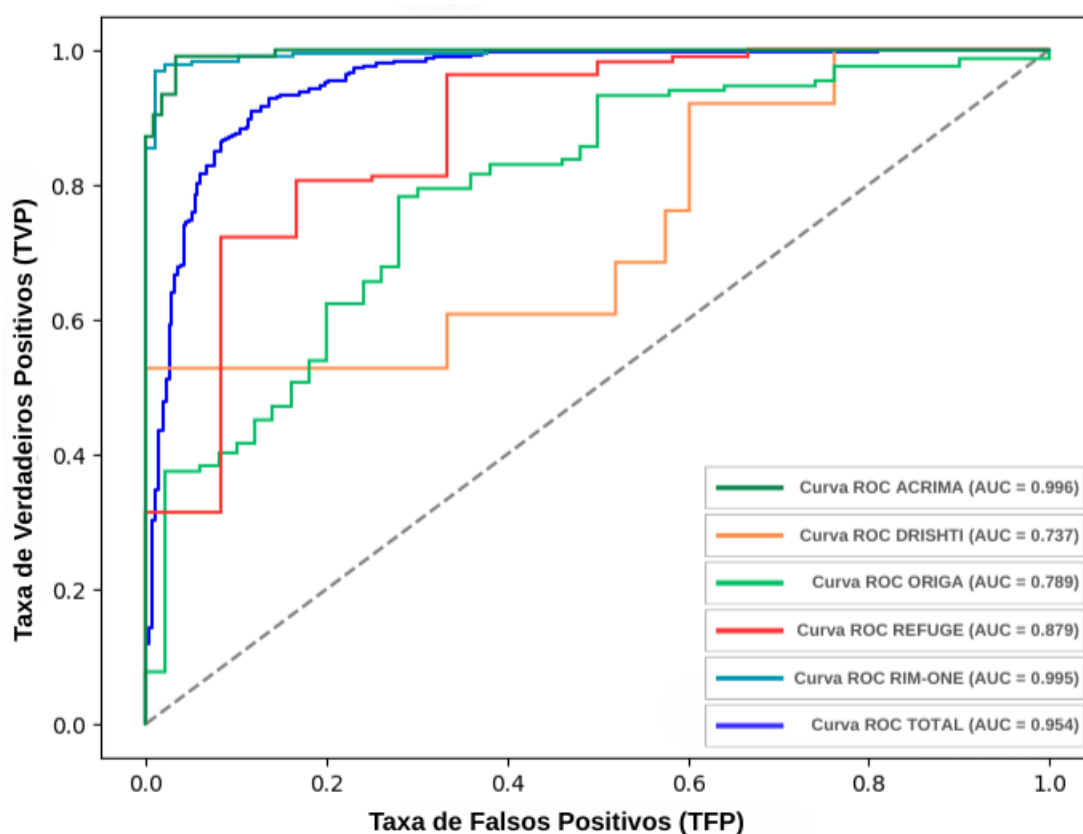
Em resumo, a análise consolidada demonstra que o modelo atinge desempenho excepcional em bases homogêneas (ACRIMA, RIM-ONE), desempenho sólido e equilibrado em um

cenário agregado e realista (TOTAL, REFUGE), e encontra desafios significativos em bases com ruídos, artefatos ou desbalanceamento crítico (ORIGA, *Drishti-GS1*), padrões que são imediatamente identificáveis através da análise visual oferecida pela Figura 28.

5.4.2 Análise das Curvas ROC por Base de Dados

A avaliação das curvas *Receiver Operating Characteristic* (ROC) possibilita uma interpretação abrangente do comportamento discriminativo do modelo em diferentes limiares de decisão. As curvas de todos os conjuntos de dados, apresentadas na Figura 29, permitem uma análise visual direta do *trade-off* entre sensibilidade e especificidade. A quantificação numérica do desempenho geral, expressa pelos valores de AUC, está sintetizada na Tabela 27, complementando a ilustração gráfica.

Figura 29 – Curvas ROC do modelo proposto nos diferentes conjuntos de dados (a AUC quantifica a capacidade discriminativa).



Fonte: Autor (2024).

5.4.2.1 Análise Visual das Curvas ROC

A análise das curvas ROC na Figura 29 revela padrões distintos de desempenho discriminativo:

A curva do conjunto ACRIMA ($AUC = 0,996$) exibe uma trajetória com inclinação acentuada no quadrante inferior esquerdo, atingindo rapidamente altos valores de sensibilidade ($> 90\%$) com elevação mínima da taxa de falsos positivos ($< 10\%$). Este padrão, próximo ao canto superior esquerdo ideal, é característico de classificadores com excelente capacidade de separação entre classes.

Similarmente, a curva do RIM-ONE ($AUC = 0,995$) demonstra aderência quase perfeita ao eixo superior esquerdo do gráfico, mantendo-se distante da linha de não-discriminação em todos os pontos. Este comportamento confirma a consistência do modelo neste domínio, alcançando alta sensibilidade mesmo sob limiares restritivos que privilegiam a especificidade.

Para o conjunto REFUGE ($AUC = 0,879$), observa-se uma curva com crescimento constante, porém menos acentuado que nos melhores casos. Incrementos graduais na sensibilidade são acompanhados por aumentos proporcionais na taxa de falsos positivos, resultando em uma área sob a curva solidamente acima da média, mas sem atingir a excelência dos conjuntos anteriores.

A curva do ORIGA ($AUC = 0,789$) apresenta inclinação moderada com tendência a platôs em determinados intervalos, refletindo a dificuldade do modelo em discriminar consistentemente entre classes nestes limiares. A proximidade relativa com a linha de base em certos segmentos é compatível com a maior variabilidade e desafios específicos deste conjunto.

O conjunto *Drishti-GSI* ($AUC = 0,737$) mostra a curva com maior aderência à diagonal de não-discriminação, particularmente em intervalos intermediários. Este padrão visual confirma os desafios quantificados pelo AUC mais baixo, indicando dificuldades significativas na separação das classes mesmo com ajustes no critério de decisão.

Finalmente, a curva do conjunto TOTAL ($AUC = 0,954$) demonstra um excelente compromisso entre os comportamentos observados, com trajetória bem acima da diagonal e formato convexo acentuado. Esta performance visual corrobora a consistência global do modelo frente à diversidade de características dos diferentes conjuntos de dados, validando sua aplicabilidade em cenários clínicos heterogêneos.

A análise integrada das curvas ROC confirma e expande os *insights* obtidos através das métricas pontuais, destacando a importância de considerar múltiplos pontos de operação. Em

particular, em contextos de triagem clínica, a possibilidade de ajustar o limiar para privilegiar a sensibilidade — mesmo com leve perda em especificidade — é um fator decisivo para reduzir o risco de subdiagnóstico.

5.4.3 *Análise Crítica e Limitações*

A análise integrada do desempenho revela padrões importantes que delineiam claramente os pontos fortes e as limitações circunscritas do modelo proposto. O entendimento dessas limitações é crucial para contextualizar a aplicabilidade clínica e orientar desenvolvimentos futuros.

- **Variação Interdomínio (Alta Sensibilidade a Características do Conjunto):** a disparidade de desempenho (até 17,87 p.p., diferença entre RIM-ONE e ORIGA) não é aleatória, mas sistemática e diretamente correlacionada com a homogeneidade e qualidade dos dados. Conjuntos como RIM-ONE e ACRIMA, adquiridos sob protocolos padronizados, permitem performance superior. Em contraste, a heterogeneidade do ORIGA (originado de múltiplas fontes e equipamentos) expõe uma limitação fundamental: a dificuldade do modelo em generalizar para domínios não vistos durante o treinamento. Esta é uma barreira conhecida para a tradução clínica de modelos de IA (STACKE *et al.*, 2021). A avaliação multibase, portanto, não é apenas uma métrica de desempenho, mas um teste de estresse necessário para validar a consistência do modelo.
- **Tendência Conservadora e Viés em Bases Desbalanceadas:** estratégia conservadora no REFUGE de privilegiar a classe majoritária (normal), priorizando a especificidade, ainda que à custa da sensibilidade, é um artefato direto do seu desbalanceamento extremo. O modelo aprende que prever a classe majoritária (normal) é a estratégia de menor risco para otimizar a função de perda padrão. Isso resulta em um número clinicamente inaceitável de falsos negativos. A AUC moderado (0,879) revela que este é um problema intrínseco ao equilíbrio do modelo, não resolvido apenas pelo ajuste de limiar. Estratégias de *loss* ponderada ou amostragem durante o treinamento são, portanto, não apenas sugeridas, mas necessárias para uma implantação equitativa.
- **Desempenho Subótimo em Casos Limítrofes e de Difícil Diagnóstico:** a baixa AUC no *Drishiti-GS1* (0,737) vai além de um simples desbalanceamento. Ele sugere uma falha do modelo em aprender características invariantes que capturem os aspectos sutis do

glaucoma inicial ou que sejam resilientes a artefatos de aquisição. O modelo pode estar superajustado a características de fácil aprendizado (ex., escavações muito pronunciadas) e falhar em casos onde o diagnóstico até para especialistas é desafiador. Isso limita sua utilidade para triagem de população geral, onde a detecção precoce é o objetivo primordial.

- **Sensibilidade a Artefatos e Baixa Qualidade de Imagem:** a performance variável é um *proxy* para a sensibilidade do modelo à qualidade da imagem. Artefatos comuns em conjuntos do "mundo real" como ORIGA (baixo contraste, desfoque, iluminação irregular) degradam significativamente a performance. Isso indica que o modelo processa a imagem de forma holística, sem uma etapa consistente de normalização ou extração de características invariantes a essas condições. Um *pipeline* de pré-processamento adaptativo constitui um pré-requisito para a implantação em ambientes clínicos não controlados.

Em síntese crítica, as limitações identificadas apontam que a principal barreira para a implantação não é a precisão bruta em conjuntos ideais, mas a fragilidade do modelo frente à heterogeneidade do mundo real. A validação em múltiplos conjuntos não serve apenas para mostrar onde o modelo funciona, mas para diagnosticar precisamente como e por que ele falha em contextos específicos, fornecendo um roteiro claro para engenharia e validação futuras.

5.4.4 Implicações Práticas e Recomendações

Com base na análise crítica das limitações, as seguintes estratégias são recomendadas para tradução clínica e desenvolvimentos futuros, priorizando a capacidade de generalização do modelo:

- **Calibração de Limiares Específica por Domínio e Tarefa:** em vez de um limiar universal, implementar múltiplos pontos de operação otimizados para diferentes contextos (YOU *et al.*, 2019):
 - **Triagem populacional:** limiar de alta sensibilidade ($> 95\%$) para maximizar a detecção de casos, mesmo à custa de um aumento tolerável de falsos positivos.
 - **Confirmação diagnóstica:** limiar de alta especificidade ($> 90\%$) para auxiliar em ambientes especializados, minimizando encaminhamentos desnecessários.
 - **Contextos de alta prevalência:** limiar que maximize o valor preditivo positivo.

- **Pré-processamento Adaptativo e Específico:** desenvolver um pipeline de pré-processamento que inclua etapas explícitas para mitigar variabilidade de domínio (ZHU *et al.*, 2020):
 - Normalização de histograma baseada em correspondência de histogramas para reduzir diferenças de contraste e iluminação entre centros.
 - Detecção e rejeição automática de imagens com artefatos graves (ex., baixo foco, reflexos) que ultrapassem um limiar de qualidade definido.
 - Aumento de dados orientado ao domínio, simulando artefatos comuns de diferentes equipamentos durante o treinamento.
- **Monitoramento Contínuo e Detecção de Mudança de Domínio:** implementar, em produção, um sistema de monitoramento que alerte para mudanças de distribuição em tempo real, baseado em:
 - Análise de características de entrada das imagens (distribuição de intensidade, contraste).
 - Discordância entre a confiança do modelo e a distribuição esperada de predições.

Isso permite a recalibração proativa do modelo antes que a performance degrade clinicamente.

- **Abordagens Híbridas e Combinação por Domínio:** explorar arquiteturas que combinem a força do modelo proposto com técnicas clássicas para aumentar a confiabilidade:
 - Um esquema de combinação onde modelos treinados especificamente em domínios particulares (ex., apenas imagens de certo equipamento) têm seu voto ponderado pela similaridade da imagem de entrada com aquele domínio.
 - Integração de características manuais (ex., razão escavação-disco - CDR) como entrada complementar ao modelo de aprendizado profundo, potencialmente mais estáveis a variações de estilo.

5.4.5 Conclusão do Desempenho Geral

O modelo proposto demonstra um desempenho global sólido e competitivo para a detecção automática de glaucoma, validando o potencial da abordagem. Os resultados permitem concluir que:

- **Alta Proficiência em Ambientes Controlados:** o modelo atinge desempenho de excelência ($AUC \approx 0,99$) em conjuntos homogêneos e de alta qualidade (ACRIMA, RIM-ONE), posicionando-o como uma ferramenta altamente confiável para contextos com protocolos de aquisição padronizados.
- **Consistência em Cenários Heterogêneos:** a performance estável no cenário agregado (TOTAL, $AUC = 0,954$) é um indicador positivo de capacidade de generalização, sugerindo que o modelo mantém boa performance em uma diversidade considerável de casos reais, um requisito fundamental para a implantação clínica.
- **Resiliência a Desafios Operacionais:** a manutenção de um desempenho adequado frente a desbalanceamento moderado e variabilidade indica que a arquitetura possui mecanismos intrínsecos de adaptação, embora insuficientes para casos extremos.

Contudo, o estudo também evidencia de forma clara os limites atuais da generalização em IA aplicada à saúde. A variabilidade de desempenho entre domínios não é uma falha, mas um diagnóstico preciso do desafio central a ser superado: a sensibilidade a características específicas dos dados de treinamento.

As limitações identificadas — especialmente em casos limítrofes e em domínios com grande variabilidade — fornecem, em vez disso, um roteiro claro para pesquisa e desenvolvimento futuros. Elas destacam que o caminho para a implantação não é por meio de modelos universalmente perfeitos, mas através de sistemas adaptativos e conscientes de seu contexto operacional.

Em última análise, este trabalho contribui não apenas com um modelo de alto desempenho, mas com uma estrutura de avaliação crítica e multidomínio que expõe as nuances do problema real. Os resultados reforçam que a métrica final de sucesso é a confiabilidade e a segurança ao longo do tempo e através da diversidade inerente aos sistemas de saúde.

5.5 Análise comparativa

Esta seção apresenta uma comparação direta e justa (*fair comparison*) entre os resultados do modelo proposto e os demais trabalhos da literatura. Para garantir a comparabilidade entre os métodos, adotou-se, para cada base de dados, a mesma divisão de treino/teste descrita nos estudos de referência. Especificamente, nos conjuntos ACRIMA e RIM-ONE utilizou-se a divisão 70/15/15 reportada por Shyamalee e Meedeniya (2022). Para as bases *Drishti-GS1* e ORIGA,

seguiu-se o protocolo de Elangovan e Nath (2021), que define divisão 70/30 (treino/teste). Para a base REFUGE, seguiu-se o protocolo de Govindan *et al.* (2024), igualmente utilizando a divisão 70/30 (treino/teste). Quando o protocolo original incluía subconjunto de validação (ex., 70/15/15), essa partição foi apenas respeitada por reprodutibilidade, mas não utilizada para ajuste de hiperparâmetros (fixados na Subseção 4.6.2).

Manteve-se o pré-processamento próprio e os hiperparâmetros globais definidos na Subseção 4.6.2 (lote = 32, taxa de aprendizagem = 10^{-4}), replicando apenas as proporções de divisão de cada protocolo e reportando as mesmas métricas quando disponíveis (acurácia, precisão, sensibilidade, *F1-score* e AUC), de modo a isolar o efeito da metodologia de classificação. Quando não havia descrição completa do particionamento em algum estudo, manteve-se a divisão padrão 70/15/15 e os resultados publicados foram utilizados apenas como referência.

Esta abordagem elimina um viés crítico comum em comparações entre estudos — onde diferenças na divisão dos dados podem inflacionar ou mascarar o desempenho real de um modelo — e permite afirmar, com maior confiança, que as diferenças de desempenho aqui observadas são devidas, de fato, à eficácia dos diferentes métodos de classificação. No *pipeline* proposto, a etapa mais crucial consiste em determinar se a imagem de entrada representa um caso de glaucoma ou uma condição normal, utilizando a arquitetura descrita na Seção 4.3 (Modelo proposto para a classificação de imagens glaucomatosas).

O modelo proposto supera as abordagens comparadas treinadas no conjunto de dados ACRIMA, conforme demonstrado pelos resultados da Tabela 28. Quando avaliado sob exatamente a mesma divisão de dados utilizada por Shyamalee e Meedeniya (2022), o modelo proposto demonstra uma superioridade abrangente. A significativa vantagem de quase 4 p.p. em acurácia (96,23% vs. 92,64%), replicada em precisão, *F1-score* e AUC, é um resultado expressivo que valida a eficácia de nossa abordagem. É notável que este ganho de desempenho foi alcançado mantendo-se uma sensibilidade comparável (96,23% vs. 97,08%), indicando que a melhoria não ocorreu às custas da capacidade de detectar casos positivos.

A análise sob o protocolo de Elangovan e Nath (2021) confirma o *Drishti-GS1* como um conjunto desafiador. Conforme os dados da Tabela 29, o modelo proposto atingiu a maior acurácia (88,24%), mas a análise das demais métricas revela *trade-offs* importantes. A AUC modesto (0,737) corrobora a complexidade inerente a esta base. A sensibilidade de 100% reportada por MK e Kalra (2024) é um resultado notável, mas geralmente indica um ponto de operação que prioriza a classe positiva, elevando a taxa de falsos positivos e, conseqüentemente, reduzindo a acurácia global (85,70%). O modelo proposto, portanto, oferece um ponto de

Tabela 28 – Comparação do método proposto para o conjunto de dados ACRIMA.

Referência	Acurácia	Precisão	Sensibilidade	<i>F1-score</i>	AUC
Devecioglu <i>et al.</i> (2021)	80,80	84,40	85,20	84,80	—
Diaz-Pinto <i>et al.</i> (2019)	70,21	—	68,93	—	0,767
Mahalakshmi <i>et al.</i> (2024)	89,80	—	91,87	94,10	—
Shyamalee e Meedeniya (2022)	92,64	93,45	97,08	95,23	0,940
Proposto	96,23	96,28	96,23	96,22	0,996

Fonte: Autor (2024).

operação mais equilibrado para este domínio específico.

Tabela 29 – Comparação do método proposto para o conjunto de dados *Drishti-GS1*.

Referência	Acurácia	Precisão	Sensibilidade	<i>F1-score</i>	AUC
Diaz-Pinto <i>et al.</i> (2019)	75,25	—	74,19	—	0,804
Elangovan e Nath (2021)	86,62	92,38	92,31	—	—
MK e Kalra (2024)	85,70	—	100,00	—	0,810
Parashar e Agrawal (2021)	85,20	—	84,67	—	0,858
Shyamalee <i>et al.</i> (2024)	80,95	91,66	83,33	87,29	0,830
Velpula e Sharma (2023)	77,23	82,19	85,71	—	0,827
Proposto	88,24	89,84	88,24	86,90	0,737

Fonte: Autor (2024).

Utilizando a mesma configuração experimental dos trabalhos comparados, o modelo proposto alcançou a maior acurácia e sensibilidade para o conjunto ORIGA, conforme apresentado na Tabela 30. Esse resultado tem relevância prática significativa. A sensibilidade de 80,00% é notavelmente superior à dos outros métodos (que variaram de 48,70% a 79,90%), indicando maior capacidade de identificar corretamente casos de glaucoma e reduzir falsos negativos, fator crítico em um contexto de triagem. Obtido sob condições de comparação justa, esse desempenho evidencia a consistência do modelo mesmo em bases complexas e desbalanceadas.

Tabela 30 – Comparação do método proposto para o conjunto de dados ORIGA.

Referência	Acurácia	Precisão	Sensibilidade	<i>F1-score</i>	AUC
Elangovan e Nath (2021)	78,32	84,36	58,06	—	—
Guo <i>et al.</i> (2020) — Método 1	69,30	—	48,70	—	0,808
Guo <i>et al.</i> (2020) — Método 2	73,20	—	52,60	—	0,866
Guo <i>et al.</i> (2018)	76,90	—	79,90	—	0,831
Proposto	80,00	78,60	80,00	78,02	0,789

Fonte: Autor (2024).

Os resultados no REFUGE, obtidos sob o mesmo protocolo, demonstram a eficácia do modelo proposto, que obteve a melhor acurácia (93,33%) e precisão (93,79%), conforme apresentado na Tabela 31. O trabalho de Shyamalee *et al.* (2024), no entanto, obteve a AUC ligeiramente superior (0,930 vs. 0,879) e *F1-score* (93,74% vs. 91,79%). Esta disparidade, mesmo sob a mesma divisão de dados, sugere uma diferença fundamental no *design* dos modelos. Enquanto o modelo proposto busca otimizar o equilíbrio entre acurácia e precisão, o modelo concorrente pode ter uma curva ROC mais favorável em certos limiares ou ser mais adepto a capturar as características específicas do REFUGE. Ainda assim, a alta acurácia do modelo proposto o mantém como uma opção extremamente competitiva.

Tabela 31 – Comparação do método proposto para o conjunto de dados REFUGE.

Referência	Acurácia	Precisão	Sensibilidade	<i>F1-score</i>	AUC
Arsyan e Maki (2022) — Método 1	89,50	87,82	92,50	91,10	—
Arsyan e Maki (2022) — Método 2	90,00	87,82	92,50	90,10	—
Govindan <i>et al.</i> (2024)	91,20	—	88,40	91,40	—
Shyamalee <i>et al.</i> (2024)	91,11	93,75	93,75	93,74	0,930
Proposto	93,33	93,79	93,33	91,79	0,879

Fonte: Autor (2024).

O desempenho no RIM-ONE é o mais expressivo. Quando comparado diretamente com Shyamalee e Meedeniya (2022) sob o mesmo protocolo, o modelo proposto apresenta uma melhoria de desempenho de quase 3 p.p. em acurácia (97,87% vs. 94,95%) e uma vantagem decisiva na AUC (0,995 vs. 0,950), conforme evidenciado na Tabela 32. Este resultado, obtido em uma base conhecida por sua alta qualidade e homogeneidade, é uma forte evidência do poder discriminativo superior da arquitetura proposta. O modelo não apenas classifica corretamente mais imagens, mas também apresenta uma separação quase ideal entre as classes.

Tabela 32 – Comparação do método proposto para o conjunto de dados RIM-ONE.

Referência	Acurácia	Precisão	Sensibilidade	<i>F1-score</i>	AUC
Devecioglu <i>et al.</i> (2021)	68,00	78,60	51,30	62,10	—
Li <i>et al.</i> (2020)	85,20	—	84,80	—	0,916
MK e Kalra (2024)	88,80	—	89,00	—	0,890
Parashar e Agrawal (2021)	92,47	—	93,20	—	0,924
Shyamalee e Meedeniya (2022)	94,95	95,37	97,98	96,65	0,950
Proposto	97,87	97,89	97,87	97,88	0,995

Fonte: Autor (2024).

Em conclusão, a análise comparativa realizada sob protocolos experimentais idênticos permitiu uma avaliação justa e criteriosa. Os resultados demonstram, de forma inequívoca, que o modelo proposto é consistentemente superior em acurácia nos cinco conjuntos de dados públicos avaliados. Trata-se de um feito notável que transcende diferenças metodológicas e atribui o mérito do desempenho à eficácia do método proposto, com vantagem particularmente pronunciada em ACRIMA e RIM-ONE.

A análise também evidencia *trade-offs* inerentes ao desenho de qualquer modelo. Em conjuntos mais complexos, como *Drishti-GS1* e REFUGE, outros métodos podem sobressair-se em métricas específicas (AUC ou sensibilidade), possivelmente por otimizações direcionadas ao domínio ou por escolhas distintas no balanceamento entre sensibilidade e especificidade. Ainda assim, a consistência do modelo proposto entre domínios diversos — liderando em acurácia e mantendo equilíbrio adequado entre precisão, sensibilidade, *FI-score* e AUC — constitui seu principal diferencial, reforçando a confiabilidade para aplicação em cenários clínicos reais.

Esses resultados, portanto, demandam uma análise crítica mais aprofundada que investigue as causas subjacentes a tais padrões, contextualize o *trade-off* entre generalização e desempenho pontual e avalie as implicações clínicas reais dos pontos fortes e fracos identificados. A discussão a seguir integrará essas evidências para responder não apenas se o modelo é preciso, mas sob quais condições ele é confiável, quais são seus limites de aplicação e como se posiciona no cenário mais amplo do desenvolvimento de ferramentas de IA para apoio diagnóstico em oftalmologia.

5.6 Discussão dos Resultados

Os resultados obtidos demonstram o potencial do modelo proposto, mas uma análise crítica revela limitações fundamentais que devem ser escrutinizadas. Esta discussão vai além da mera descrição de desempenho, concentrando-se nas causas profundas dos erros, nos *trade-offs* inerentes à arquitetura escolhida e nas reais barreiras para uma implantação clínica eficiente.

5.6.1 Análise Crítica da Arquitetura: Leveza vs. Performance

A opção por uma arquitetura sem segmentação prévia — frequentemente caracterizada como "leve" por eliminar etapas complexas de pré-processamento — configurou-se como uma estratégia de compensação (*trade-off*). Por um lado, conferiu agilidade ao *pipeline* e reduziu a dependência da acurácia de modelos de segmentação, notoriamente propensos a falhas em

imagens de baixa qualidade. Por outro lado, essa escolha arquitetural transferiu integralmente a carga de aprendizado para as redes neurais e classificadores, exigindo que eles aprendessem simultaneamente a localizar regiões de interesse (como o DO) e a classificá-las.

Nossos resultados indicam que esta estratégia foi eficaz em bases homogêneas (ex., RIM-ONE e ACRIMA), mas atingiu limites evidentes em conjuntos com alta variabilidade morfológica ou artefatos pronunciados (ex., ORIGA, *Drishti-GSI*). Essa observação é consistente com Shyamalee e Meedeniya (2022), cujo *pipeline* adota módulos de atenção em uma U-Net para segmentação do DO e da escavação antes da classificação. Nesse arranjo, a etapa de segmentação atua como guia anatômico (DO/escavação óptica), contribuindo para maior consistência em imagens heterogêneas e obtendo altas acurácias de segmentação e classificação. Contudo, ao comparar os resultados sob o mesmo protocolo de avaliação, observa-se que, mesmo com esse reforço, o modelo de Shyamalee e Meedeniya (2022) não superou o desempenho global do *pipeline* proposto. Portanto, a "leveza" obtida pela ausência de segmentação tem um custo mensurável: uma maior sensibilidade à qualidade intrínseca das imagens.

5.6.2 Impacto do Desbalanceamento

O desbalanceamento de classes revelou-se um dos fatores mais determinantes para o desempenho global do modelo. Em bases com proporção equilibrada, como ACRIMA e RIM-ONE, observou-se simultaneamente alta sensibilidade e especificidade, evidenciando aprendizado consistente entre as classes.

Por outro lado, em bases com desproporção acentuada entre classes — como REFUGE e ORIGA — o classificador desenvolveu um viés em favor da classe majoritária (normal), reduzindo a sensibilidade e elevando a taxa de falsos negativos. As matrizes de confusão confirmam esse padrão: na ORIGA, apenas 20 casos glaucomatosos foram corretamente classificados, enquanto 30 foram rotulados como normais, resultando em sensibilidade de apenas 40%. No REFUGE, 8 dos 12 glaucomas foram classificados incorretamente como normais, comprometendo o potencial de triagem. Esses achados mostram que a acurácia global, isoladamente, pode mascarar falhas críticas na detecção da classe minoritária, tornando indispensável o uso de métricas para uma avaliação clínica realista.

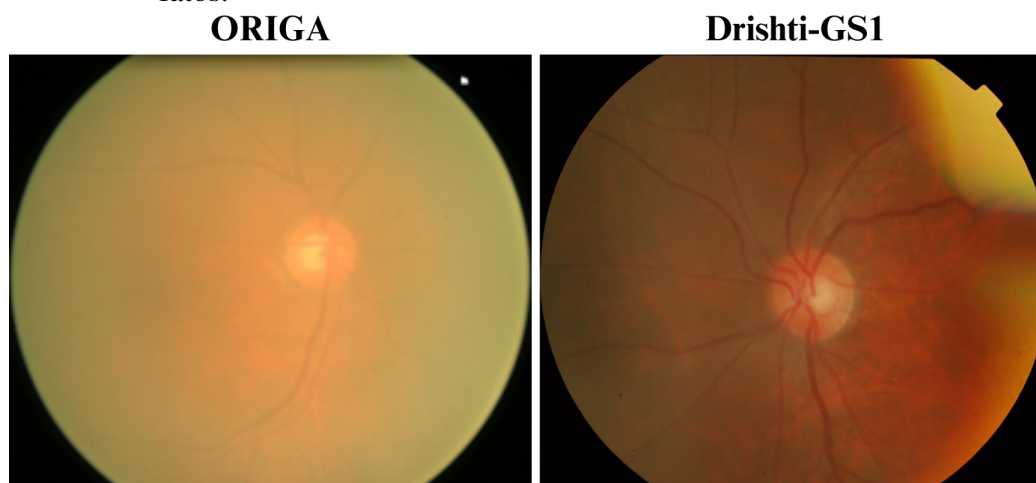
Esse comportamento não é exclusivo do método proposto, mas constitui um desafio estrutural amplamente relatado na literatura. Shyamalee e Meedeniya (2022) observaram que, mesmo com pré-processamento intensivo (CLAHE e filtro da mediana) e segmentação com

Attention U-Net, o desempenho caiu em bases desbalanceadas. De modo semelhante, Elangovan e Nath (2021) reportaram degradação superior a 15 p.p. na acurácia ao migrar de ACRIMA para ORIGA, mesmo após aplicar rotação como técnica de aumento de dados. Esses resultados, aliados à análise da qualidade das imagens e à queda de desempenho em bases externas, reforçam que o desbalanceamento potencializa outros desafios discutidos a seguir, como a sensibilidade à qualidade das imagens e a dificuldade de generalização para domínios não vistos.

5.6.3 Impacto da Qualidade das Imagens

Além do desbalanceamento, a qualidade intrínseca das imagens mostrou-se um fator crítico para a generalização. Artefatos como iluminação irregular, saturação inconsistente e baixo contraste atuaram como ruído dominante, degradando a capacidade discriminativa do modelo. A Figura 30 ilustra esse impacto: no ORIGA, a baixa nitidez obscurece os contornos do DO, elevando a taxa de falsos negativos (sensibilidade de apenas 40%), enquanto no *Drishti-GS1*, reflexos e saturação irregular resultaram em sensibilidade artificialmente elevada (100%), às custas de uma especificidade severamente reduzida (53,85%).

Figura 30 – Exemplos representativos das variações de qualidade entre as imagens dos conjuntos ORIGA (esquerda) e *Drishti-GS1* (direita). Notam-se desafios como baixo contraste, iluminação irregular e presença de artefatos.



Fonte: Autor (2024).

Apesar do uso de técnicas de *data augmentation* geométrico e fotométrico (rotações, zoom, translações, espelhamentos e ajustes de brilho/contraste), a heterogeneidade visual de bases como ORIGA e *Drishti-GS1* continuou comprometendo o desempenho. Por sua vez, Elangovan e Nath (2021) utilizaram apenas rotações de até 180° como aumento de dados, o

que possivelmente limitou a diversidade sintética e contribuiu para o sobreajuste em bases mais complexas. A estratégia mais abrangente de *augmentation* adotada neste trabalho explica parcialmente o ganho de desempenho observado em ACRIMA e RIM-ONE, mas ainda se mostrou insuficiente para compensar a heterogeneidade extrema dessas bases.

O pré-processamento minimalista adotado — restrito a redimensionamento e normalização de intensidade — buscou preservar informações anatômicas relevantes (CDR, padrão vascular), evitando distorções introduzidas por métodos mais invasivos de realce. No entanto, essa escolha deixou o modelo mais vulnerável a variabilidades de iluminação e contraste. Em contraste, o *pipeline* de Shyamalee e Meedeniya (2022) empregou uma etapa explícita de segmentação com Attention U-Net, que provavelmente reduziu a variância e aumentou a estabilidade em bases homogêneas, mas à custa de maior complexidade computacional e dependência de máscaras — algumas construídas manualmente. Para cenários clínicos reais, soluções híbridas podem ser necessárias: *pipelines* adaptativos de pré-processamento e estratégias de *augmentation* orientadas por domínio (simulando artefatos clínicos como reflexos e desfoque) devem ser considerados para preparar o modelo às condições adversas de aquisição.

5.6.4 *Overfitting e Generalização: Uma Conquista Parcial*

A estratégia de *ensemble* e o uso de *data augmentation* mitigaram, mas não eliminaram, o risco de *overfitting*. A queda de desempenho em bases externas (como a sensibilidade de apenas 40% no ORIGA) é um sintoma claro do *domain shift* — fenômeno em que a distribuição de dados de teste difere substancialmente da de treinamento — e não apenas uma falha do modelo em si. A generalização para bases clínicas reais, não curadas, com distribuições imprevisíveis, permanece como o desafio central não resolvido. A AUC de 0,954 no conjunto TOTAL é um indicador positivo, mas fortemente influenciada pelo desempenho superior nos domínios de ACRIMA e RIM-ONE. A verdadeira generalização, capaz de manter alta acurácia em contextos clínicos novos e não vistos, ainda não foi plenamente alcançada.

Esse cenário contrasta com os resultados de Govindan *et al.* (2024), cujo modelo híbrido ResNet50 + InceptionV3 atingiu acurácia de 99,4% e *F1-score* de 99,2% no REFUGE sob a mesma divisão 70/30. Embora esses números representem o estado da arte intra-base, o estudo não avaliou generalização para domínios externos, o que impede concluir se o desempenho seria mantido em cenários de maior heterogeneidade. Esse aspecto ressalta que métricas quase perfeitas em um único conjunto de dados não necessariamente se traduzem em consistência

clínica ampla.

Fenômeno semelhante foi observado por Elangovan e Nath (2021), que reportaram queda superior a 15 p.p. ao migrar de ACRIMA para ORIGA, e por Shyamalee e Meedeniya (2022), que apesar de não terem identificado sinais de *overfitting* em suas curvas de treino/validação, relataram variação relevante de desempenho entre bases — com acurácia superior a 94% em ACRIMA e RIM-ONE (bases mais homogêneas), mas queda perceptível em ORIGA e *Drishti-GSI* (mais heterogêneas). Esses achados corroboram que a variabilidade entre bases é um desafio estrutural na detecção de glaucoma, exigindo soluções que combinem robustez estatística, estratégias de normalização de domínio e protocolos de avaliação que incluam validação cruzada entre bases ou dados multi-institucionais para garantir aplicabilidade clínica.

5.6.5 Limitações Computacionais do Ensemble

O *ensemble* proposto, composto por quatro modelos, conferiu maior consistência e estabilidade, como pretendido. No entanto, essa escolha introduziu uma limitação computacional evidente. O custo de treinamento e inferência é aproximadamente quadruplicado em comparação a uma arquitetura individual. Em um cenário de triagem em larga escala, onde velocidade e recursos computacionais são frequentemente limitantes, este *trade-off* entre confiabilidade e eficiência deve ser criteriosamente ponderado. Embora o modelo seja "leve" por não precisar de segmentação, torna-se "pesado" pela necessidade de executar múltiplos classificadores. Esta é uma limitação prática crucial para implantação em sistemas embarcados ou ambientes com recursos computacionais restritos.

Ainda assim, a redução de 57% no desvio padrão obtida pelo *ensemble* no conjunto TOTAL evidencia que o ganho em estabilidade compensa a perda de eficiência, sobretudo em contextos clínicos nos quais a previsibilidade é tão importante quanto a acurácia média. Em síntese, os resultados revelam que a ausência de segmentação garante leveza arquitetural, mas amplia a vulnerabilidade a desbalanceamentos, à qualidade das imagens e ao *domain shift*. Embora o *ensemble* tenha mitigado parcialmente esses efeitos, desafios estruturais permanecem e devem ser considerados para a transposição do modelo a cenários clínicos reais.

6 CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS

Este trabalho representou uma investigação abrangente no desenvolvimento e aplicação de técnicas avançadas de aprendizado de máquina e visão computacional para o diagnóstico automatizado de glaucoma através de imagens de fundo de olho. O glaucoma, uma doença silenciosa e de progressão irreversível, demanda métodos diagnósticos precisos e acessíveis, especialmente em estágios iniciais. O sistema proposto, baseado em uma arquitetura híbrida que combina CNNs pré-treinadas com métodos de *ensemble learning*, demonstrou desempenho excepcional em múltiplos bancos de dados públicos, estabelecendo novos parâmetros de excelência para esta aplicação médica crítica.

6.1 Síntese das Conclusões e Contribuições

A investigação desenvolvida nesta dissertação confirmou a hipótese de que uma arquitetura híbrida, combinando redes neurais convolucionais pré-treinadas (EfficientNet-B0, MobileNetV2 e ResNet-50) com classificadores tradicionais (Support Vector Machine, Random Forest e Gradient Boosting), integrada por meio de *ensemble* ponderado, mostrou-se superior às abordagens unimodais na detecção de glaucoma. O sistema proposto alcançou resultados expressivos em múltiplos bancos públicos, com destaque para o conjunto RIM-ONE (97,87% de acurácia; $AUC = 0,995$), e manteve desempenho consistente no conjunto TOTAL, formado pela combinação de cinco bases heterogêneas, simulando com maior fidelidade um cenário clínico real. Nesse contexto, obteve-se equilíbrio relevante entre sensibilidade (85,17%) e especificidade (93,19%), reforçando sua aplicabilidade prática em triagens oftalmológicas.

Dessa forma, os resultados obtidos respondem diretamente às lacunas destacadas no Capítulo 2 (Trabalhos correlatos), ao eliminar a dependência de etapas manuais de segmentação, validar o sistema em cinco bases públicas heterogêneas e demonstrar uma estabilidade interdomínio superior àquelas apresentadas pelas abordagens unimodais descritas na literatura.

Uma análise crítica revelou um *trade-off* significativo entre desempenho máximo e estabilidade. Embora o modelo denso isolado tenha alcançado a maior acurácia média ($96,48\% \pm 3,43$) no conjunto combinado (TOTAL, Tabela 24), ele exibiu maior variabilidade em relação ao *ensemble* ($90,22\% \pm 1,46$). O *ensemble* ponderado, por sua vez, embora com acurácia média ligeiramente inferior em alguns cenários, demonstrou superioridade incontestável em estabilidade, reduzindo a variabilidade do desempenho (desvio padrão) em aproximadamente 57,4%, como mostra a análise de estabilidade apresentada na Seção 5.2 (Classificadores Individuais vs

Ensemble). Esse resultado valida a estratégia do *ensemble* como a mais adequada para aplicações clínicas, nas quais a confiabilidade e a previsibilidade do modelo são atributos tão importantes quanto a acurácia absoluta.

Entre os avanços alcançados, destacam-se:

1. a eliminação da dependência de segmentação manual ou algoritmos de pré-processamento complexos, tradicionalmente apontados como gargalos metodológicos na literatura;
2. a utilização de um *pipeline* minimalista de padronização das imagens, que demonstrou ser suficiente para a extração de características discriminativas, preservando a integridade clínica dos exames;
3. a definição automática de pesos do *ensemble* via validação cruzada, garantindo maior previsibilidade e confiabilidade dos resultados; e
4. a aplicação de técnicas de *data augmentation*, fundamentais para lidar com o desbalanceamento entre classes e ampliar a variabilidade das amostras de treinamento.

Do ponto de vista científico, este trabalho contribui ao consolidar uma abordagem híbrida de classificação que alia diversidade de modelos a mecanismos de integração adaptativa e inteligente, validada em múltiplos domínios. Do ponto de vista tecnológico, demonstra viabilidade mesmo em cenários com restrições de infraestrutura computacional, pavimentando o caminho para aplicações clínicas descentralizadas. Por fim, do ponto de vista clínico, fornece uma ferramenta promissora para triagem automatizada, capaz de reduzir variabilidade inter-observador e apoiar decisões médicas em contextos de grande volume de pacientes.

Todos os objetivos específicos delineados no Capítulo 1 (Introdução) foram integralmente contemplados, abrangendo a utilização de CNNs pré-treinadas como extratoras de características, a aplicação sistemática de técnicas de *data augmentation*, a proposição e validação de um modelo denso, a integração de múltiplos classificadores em um *ensemble* ponderado por validação cruzada, bem como a avaliação abrangente do sistema em cinco bases públicas heterogêneas e em seu conjunto combinado, aproximando-se das condições reais de aplicação clínica.

6.2 Limitações e Considerações sobre a Aplicabilidade

Apesar dos avanços, a pesquisa evidencia limitações relevantes que merecem consideração. Os bancos de dados públicos utilizados no treinamento concentram-se predominantemente

em imagens de populações específicas (asiáticas e caucasianas), o que pode restringir a performance do modelo em grupos étnicos sub-representados. Esse viés potencial reforça a necessidade de diversificação dos dados de treinamento em futuras iterações do sistema. Soma-se a isso a interpretabilidade limitada das decisões de classificadores em *ensemble*, característica comum em modelos complexos, que exige cautela em contextos médicos regulados, onde a transparência diagnóstica é essencial. O desenvolvimento de técnicas explicativas complementares, como mapas de atenção e métodos baseados em gradiente, poderá mitigar essa limitação e ampliar a confiabilidade clínica do sistema.

Outra restrição identificada refere-se à variabilidade na qualidade das imagens, especialmente em conjuntos como *Drishti-GS1* e ORIGA, nos quais o modelo apresentou desequilíbrios acentuados (sensibilidade de 100% contra especificidade de 53,85% no *Drishti-GS1*, e sensibilidade de apenas 40% no ORIGA). Esses resultados ilustram o impacto negativo de fatores como diferenças nos protocolos de aquisição, equipamentos utilizados e critérios de anotação entre os bancos públicos. Tais flutuações reforçam a necessidade de estratégias mais refinadas de normalização interbases, bem como de iniciativas colaborativas de padronização e anotação que promovam maior consistência e representatividade dos conjuntos de dados empregados em pesquisas futuras.

6.3 Trabalhos Futuros e Perspectivas

As implicações clínicas deste trabalho são significativas, pois a automação do rastreamento de glaucoma pode representar um avanço decisivo no enfrentamento da cegueira evitável. Para que esse potencial se converta em realidade prática, destacam-se os seguintes eixos de pesquisa:

- **Validação clínica e implementação:** realização de estudos prospectivos em ambientes hospitalares e multicêntricos, avaliando a integração do modelo em fluxos de trabalho reais e seu impacto em desfechos clínicos relevantes, como tempo de encaminhamento e taxa de detecção precoce.
- **Explicabilidade (XAI):** incorporação de mecanismos de atenção espacial e de técnicas de interpretação, como Grad-CAM e LIME, para gerar visualizações que evidenciem regiões decisivas para o diagnóstico. Tais recursos podem ampliar a confiabilidade clínica e favorecer a adesão de profissionais de saúde.
- **Otimização e expansão:** exploração de arquiteturas emergentes e mais eficientes, como

Vision Transformers otimizados ou modelos compactos para dispositivos móveis, visando reduzir o custo computacional e ampliar a acessibilidade. Além disso, a expansão da tarefa para classificação de estágios da doença (leve, moderado, severo) pode aumentar a utilidade clínica do sistema.

- **Diversidade de dados:** ampliação do conjunto de treinamento com maior diversidade étnica, demográfica e de equipamentos de aquisição. O uso de imagens sintéticas geradas por técnicas como GANs pode contribuir para mitigar vieses e aumentar a capacidade de generalização entre domínios.

A transição bem-sucedida dessa tecnologia para a prática clínica exigirá colaborações interdisciplinares contínuas entre engenheiros, oftalmologistas e gestores de saúde, assegurando que avanços técnicos sejam acompanhados de validação prática, padronização regulatória e aceitação no contexto médico.

6.4 Considerações Finais

Em síntese, este trabalho demonstrou a viabilidade de um sistema automatizado para detecção de glaucoma, superando limitações identificadas no estado da arte. Mais do que um avanço técnico, representa um passo concreto na democratização do diagnóstico oftalmológico, oferecendo um referencial metodológico sólido para futuras pesquisas. As direções apontadas buscam não apenas aprimorar a arquitetura proposta, mas sobretudo assegurar que seus benefícios se traduzam em melhorias tangíveis e equitativas para os sistemas de saúde e para os pacientes na prática clínica ao redor do mundo.

Em última análise, sua principal implicação clínica reside na possibilidade concreta de reduzir casos de glaucoma não diagnosticados em estágios iniciais — condição crítica que, sem detecção precoce, evolui inevitavelmente para a progressão irreversível da cegueira.

REFERÊNCIAS

- ABRAMOFF, M.; KAY, C. N. Chapter 6 - image processing. In: RYAN, S. J.; SADDA, S. R.; HINTON, D. R.; SCHACHAT, A. P.; SADDA, S. R.; WILKINSON, C.; WIEDEMANN, P.; SCHACHAT, A. P. (Ed.). **Retina (Fifth Edition)**. Fifth edition. London: W.B. Saunders, 2013. p. 151–176. ISBN 978-1-4557-0737-9. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9781455707379000060>>.
- ADAM, K. D. B. J. *et al.* A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, v. 1412, n. 6, 2014.
- ADEM, K.; KILIÇARSLAN, S. Performance analysis of optimization algorithms on stacked autoencoder. In: **2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)**. [S.l.: s.n.], 2019. p. 1–4.
- AHMED, T.; SABAB, N. H. N. Classification and understanding of cloud structures via satellite images with efficientnet. **SN Computer Science**, Springer, v. 3, n. 1, p. 99, 2022.
- AHN, J. M.; KIM, S.; AHN, K.-S.; CHO, S.-H.; LEE, K. B.; KIM, U. S. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. **PloS one**, Public Library of Science San Francisco, CA USA, v. 13, n. 11, p. e0207982, 2018.
- ALI, S.; HUSSAIN, A.; AICH, S.; PARK, M. S.; CHUNG, M. P.; JEONG, S. H.; SONG, J. W.; LEE, J. H.; KIM, H. C. A soft voting ensemble-based model for the early prediction of idiopathic pulmonary fibrosis (ipf) disease severity in lungs disease patients. **Life**, v. 11, n. 10, 2021. ISSN 2075-1729. Disponível em: <<https://www.mdpi.com/2075-1729/11/10/1092>>.
- ALLISON, K.; PATEL, D.; ALABI, O. Epidemiology of glaucoma: the past, present, and predictions for the future. **Cureus**, Cureus Inc., v. 12, n. 11, 2020.
- ALOM, M. Z.; TAHA, T. M.; YAKOPCIC, C.; WESTBERG, S.; SIDIKE, P.; NASRIN, M. S.; HASAN, M.; ESSEN, B. C. V.; AWWAL, A. A. S.; ASARI, V. K. A state-of-the-art survey on deep learning theory and architectures. **Electronics**, v. 8, n. 3, 2019. ISSN 2079-9292. Disponível em: <<https://www.mdpi.com/2079-9292/8/3/292>>.
- ALOMAR, K.; AYSEL, H. I.; CAI, X. Data augmentation in classification and segmentation: A survey and new strategies. **Journal of Imaging**, v. 9, n. 2, 2023. ISSN 2313-433X. Disponível em: <<https://www.mdpi.com/2313-433X/9/2/46>>.
- ARCHANA, R.; JEEVARAJ, P. E. Deep learning models for digital image processing: a review. **Artificial Intelligence Review**, Springer, v. 57, n. 1, p. 11, 2024.
- ARIYO, O.; AKTER, F. Anatomy and physiology of the neuron. **Neuroscience for Neurosurgeons**, Cambridge University Press, p. 72, 2024.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, Amer. Statist. Assoc., the Bernoulli Soc., the Inst. Math. Statist., and the Statist. Soc. Canada, v. 4, n. none, p. 40 – 79, 2010. Disponível em: <<https://doi.org/10.1214/09-SS054>>.
- ARSYAN, A.; MAKI, W. F. A. Classification of glaucoma using invariant moment methods on k-nearest neighbor and random forest models. **Building of Informatics, Technology and Science (BITS)**, v. 3, n. 4, p. 466–472, Mar. 2022. Disponível em: <<https://ejurnal.seminar-id.com/index.php/bits/article/view/1244>>.

ASAM, J. S.; POLZER, M.; TAFRESHI, A.; HIRNSCHALL, N.; FINDL, O. Anterior segment oct. **High resolution imaging in microscopy and ophthalmology: new frontiers in biomedical optics**, Springer, p. 285–299, 2019.

ATCHISON, D. A. Optics of the human eye. In: GUENTHER, B. D.; STEEL, D. G. (Ed.). **Encyclopedia of Modern Optics (Second Edition)**. Second edition. Oxford: Elsevier, 2018. p. 43–63. ISBN 978-0-12-814982-9. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128035818097733>>.

BAJWA, M. N.; SINGH, G. A. P.; NEUMEIER, W.; MALIK, M. I.; DENGEL, A.; AHMED, S. G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection. In: **2020 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2020. p. 1–7.

BALAKRISHNAN, U. Ndc-ivm: An automatic segmentation of optic disc and cup region from medical images for glaucoma detection. **Journal of Innovative Optical Health Sciences**, World Scientific, v. 10, n. 03, p. 1750007, 2017.

BARRIGA, J. A.; BLANCO-CIPOLLONE, F.; TRIGO-CÓRDOBA, E.; GARCÍA-TEJERO, I.; CLEMENTE, P. J. Crop-water assessment in citrus (*citrus sinensis* l.) based on continuous measurements of leaf-turgor pressure using machine learning and iot. **Expert Systems with Applications**, v. 209, p. 118255, 2022. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417422013987>>.

BASHA, S. S.; DUBEY, S. R.; PULABAIGARI, V.; MUKHERJEE, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. **Neurocomputing**, v. 378, p. 112–119, 2020. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231219313803>>.

BATES, S.; HASTIE, T.; TIBSHIRANI, R. Cross-validation: What does it estimate and how well does it do it? **Journal of the American Statistical Association**, Informa UK Limited, v. 119, n. 546, p. 1434–1445, maio 2023. ISSN 1537-274X. Disponível em: <<http://dx.doi.org/10.1080/01621459.2023.2197686>>.

BATISTA, F. J. F.; DIAZ-ALEMAN, T.; SIGUT, J.; ALAYON, S.; ARNAY, R.; ANGEL-PEREIRA, D. Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. **Image Analysis and Stereology**, v. 39, n. 3, p. 161–167, Nov. 2020. Disponível em: <<https://www.ias-iss.org/ojs/IAS/article/view/2346>>.

BENARDOS, P.; VOSNIAKOS, G.-C. Optimizing feedforward artificial neural network architecture. **Engineering Applications of Artificial Intelligence**, v. 20, n. 3, p. 365–382, 2007. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197606001072>>.

BENTÉJAC, C.; CSÖRGŐ, A.; MARTÍNEZ-MUÑOZ, G. A comparative analysis of gradient boosting algorithms. **Artificial Intelligence Review**, Springer, v. 54, p. 1937–1967, 2021.

BERLIN, M. S.; STEIN, H. A.; LENT-SCHOCHET, D.; QUACH, C.; CAI, Z. Chapter 25 - glaucoma. In: STEIN, H. A.; STEIN, R. M.; FREEMAN, M. I. (Ed.). **The Ophthalmic Assistant (Tenth Edition)**. Tenth edition. London: Elsevier, 2018. p. 436–474. ISBN 978-0-323-39477-2. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780323394772000257>>.

BERNARDES, R.; SERRANHO, P.; LOBO, C. Digital ocular fundus imaging: a review. **Ophthalmologica**, S. Karger AG Basel, Switzerland, v. 226, n. 4, p. 161–181, 2011.

BIAU, G.; CADRE, B.; ROUVIÈRE, L. Accelerated gradient boosting. **Machine learning**, Springer, v. 108, p. 971–992, 2019.

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.

CANONACO, G.; BERGAMASCO, A.; MONGELLUZZO, A.; ROVERI, M. Adaptive federated learning in presence of concept drift. In: **2021 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2021. p. 1–7.

CARMONA, E. J.; RINCÓN, M.; GARCÍA-FEIJÓ, J.; CASA, J. M. M. de-la. Identification of the optic nerve head with genetic algorithms. **Artificial Intelligence in Medicine**, v. 43, n. 3, p. 243–259, 2008. ISSN 0933-3657. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0933365708000547>>.

CIVIT-MASOT, J.; DOMÍNGUEZ-MORALES, M. J.; VICENTE-DÍAZ, S.; CIVIT, A. Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction. **IEEE Access**, IEEE, v. 8, p. 127519–127529, 2020.

COAN, L. J.; WILLIAMS, B. M.; ADITHYA, V. K.; UPADHYAYA, S.; ALKAFRI, A.; CZANNER, S.; VENKATESH, R.; WILLOUGHBY, C. E.; KAVITHA, S.; CZANNER, G. Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review. **Survey of ophthalmology**, Elsevier, v. 68, n. 1, p. 17–41, 2023.

COLLERAN, I.; VIGNARAJAN, J.; FROST, S.; SAHA, S. Evaluating the performance of different convolutional neural networks in glaucoma detection. In: **2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)**. [S.l.: s.n.], 2022. p. 1–5.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

COSTA, V. P.; LEUNG, C. K.; KOOK, M. S.; LIN, S. C.; AL-ASWAD, L. A.; ARAIE, M.; BAUDOUIN, C.; COUPAL, D. J.; FECHTNER, R.; KHAW, P. T. *et al.* Clear lens extraction in eyes with primary angle closure and primary angle-closure glaucoma. **Survey of Ophthalmology**, Elsevier, v. 65, n. 6, p. 662–674, 2020.

DAS, D.; CEN, Y.; WANG, J.; FONG, X. Bilayer-skyrmion-based design of neuron and synapse for spiking neural network. **Phys. Rev. Appl.**, American Physical Society, v. 19, p. 024063, Feb 2023. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevApplied.19.024063>>.

DAS, D.; NAYAK, D. R.; BHANDARY, S. V.; ACHARYA, U. R. Cdam-net: Channel shuffle dual attention based multi-scale cnn for efficient glaucoma detection using fundus images. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 133, p. 108454, 2024.

DAS, D.; NAYAK, D. R.; PACHORI, R. B. Ca-net: A novel cascaded attention-based network for multistage glaucoma classification using fundus images. **IEEE Transactions on Instrumentation and Measurement**, v. 72, p. 1–10, 2023.

DAS, D.; NAYAK, D. R.; PACHORI, R. B. Aes-net: An adapter and enhanced self-attention guided network for multi-stage glaucoma classification using fundus images. **Image and Vision Computing**, Elsevier, v. 146, p. 105042, 2024.

DEVECIOGLU, O. C.; MALIK, J.; INCE, T.; KIRANYAZ, S.; ATALAY, E.; GABBOUJ, M. Real-time glaucoma detection from digital fundus images using self-onns. **Ieee Access**, IEEE, v. 9, p. 140031–140041, 2021.

DIAZ-PINTO, A.; MORALES, S.; NARANJO, V.; KÖHLER, T.; MOSSI, J. M.; NAVEA, A. Cnns for automatic glaucoma assessment using fundus images: an extensive validation. **Biomedical engineering online**, Springer, v. 18, p. 1–19, 2019.

DONG, X.; YU, Z.; CAO, W.; SHI, Y.; MA, Q. A survey on ensemble learning. **Frontiers of Computer Science**, Springer, v. 14, n. 2, p. 241–258, 2020.

DORAIRAJ, S.; TSAI, J. C.; GRIPPO, T. M. Changing trends of imaging in angle closure evaluation. **International Scholarly Research Notices**, Wiley Online Library, v. 2012, n. 1, p. 597124, 2012.

ELANGO VAN, P.; NATH, M. K. Glaucoma assessment from color fundus images using convolutional neural network. **International Journal of Imaging Systems and Technology**, v. 31, n. 2, p. 955–971, 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22494>>.

ELFWING, S.; UCHIBE, E.; DOYA, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. **Neural Networks**, v. 107, p. 3–11, 2018. ISSN 0893-6080. Special issue on deep reinforcement learning. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0893608017302976>>.

ELMOUFIDI, A.; SKOUTA, A.; JAI-ANDALOUSSI, S.; OUCHETTO, O. Deep multiple instance learning for automatic glaucoma prevention and auto-annotation using color fundus photography. **Progress in Artificial Intelligence**, Springer, v. 11, n. 4, p. 397–409, 2022.

ELSAYED, N.; MAIDA, A. S.; BAYOUMI, M. Empirical activation function effects on unsupervised convolutional lstm learning. In: **2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)**. [S.l.: s.n.], 2018. p. 336–343.

ENGELN, J. E. V.; HOOS, H. H. A survey on semi-supervised learning. **Machine learning**, Springer, v. 109, n. 2, p. 373–440, 2020.

FAJAR, A.; SARNO, R.; FATICHAH, C.; SUSILO, R. I.; PANGESTU, G. Cyclical learning rate optimization on deep learning model for brain tumor segmentation. **IEEE Access**, v. 11, p. 119802–119810, 2023.

FAN, R.; ALIPOUR, K.; BOWD, C.; CHRISTOPHER, M.; BRYE, N.; PROUDFOOT, J. A.; GOLDBAUM, M. H.; BELGHITH, A.; GIRKIN, C. A.; FAZIO, M. A. *et al.* Detecting glaucoma from fundus photographs using deep learning without convolutions: transformer for improved generalization. **Ophthalmology science**, Elsevier, v. 3, n. 1, p. 100233, 2023.

FOGEL-LEVIN, M.; SADDI, S. R.; ROSENFELD, P. J.; WAHEED, N.; QUERQUES, G.; FREUND, B. K.; SARRAF, D. Advanced retinal imaging and applications for clinical practice: A consensus review. **Survey of Ophthalmology**, v. 67, n. 5, p. 1373–1390, 2022. ISSN 0039-6257. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0039625722000352>>.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001.

FUMERO, F.; ALAYON, S.; SANCHEZ, J. L.; SIGUT, J.; GONZALEZ-HERNANDEZ, M. Rim-one: An open retinal image database for optic nerve evaluation. In: **2011 24th International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.: s.n.], 2011. p. 1–6.

GANAI, M.; HU, M.; MALIK, A.; TANVEER, M.; SUGANTHAN, P. Ensemble deep learning: A review. **Engineering Applications of Artificial Intelligence**, v. 115, p. 105151, 2022. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S095219762200269X>>.

GARBIN, C.; ZHU, X.; MARQUES, O. Dropout vs. batch normalization: an empirical study of their impact to deep learning. **Multimedia tools and applications**, Springer, v. 79, n. 19, p. 12777–12815, 2020.

GHANE, E.; FAGERSTRÖM, M.; MIRKHALAF, M. Recurrent neural networks and transfer learning for predicting elasto-plasticity in woven composites. **European Journal of Mechanics - A/Solids**, v. 107, p. 105378, 2024. ISSN 0997-7538. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S099775382400158X>>.

GÓMEZ-VALVERDE, J. J.; ANTÓN, A.; FATTI, G.; LIEFERS, B.; HERRANZ, A.; SANTOS, A.; SÁNCHEZ, C. I.; LEDESMA-CARBAYO, M. J. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. **Biomedical optics express**, Optica Publishing Group, v. 10, n. 2, p. 892–913, 2019.

GOVINDAN, M.; DHAKSHNAMURTHY, V. K.; SREERANGAN, K.; NAGARAJAN, M. D.; RAJAMANICKAM, S. K. A framework for early detection of glaucoma in retinal fundus images using deep learning. **Engineering Proceedings**, v. 62, n. 1, 2024. ISSN 2673-4591. Disponível em: <<https://www.mdpi.com/2673-4591/62/1/3>>.

GULHANE, M.; KUMAR, S. Deep learning based heart diseases detection using resnet50 architecture. In: **2023 Intelligent Methods, Systems, and Applications (IMSA)**. [S.l.: s.n.], 2023. p. 193–198.

GUO, F.; LI, W.; TANG, J.; ZOU, B.; FAN, Z. Automated glaucoma screening method based on image segmentation and feature extraction. **Medical & Biological Engineering & Computing**, Springer, v. 58, p. 2567–2586, 2020.

GUO, F.; MAI, Y.; ZHAO, X.; DUAN, X.; FAN, Z.; ZOU, B.; XIE, B. Yanbao: A mobile app using the measurement of clinical parameters for glaucoma screening. **IEEE Access**, v. 6, p. 77414–77428, 2018.

GUPTA, N.; GARG, H.; AGARWAL, R. A robust framework for glaucoma detection using clahe and efficientnet. **The Visual Computer**, Springer, p. 1–14, 2022.

HAGIWARA, Y.; KOH, J. E. W.; TAN, J. H.; BHANDARY, S. V.; LAUDE, A.; CIACCIO, E. J.; TONG, L.; ACHARYA, U. R. Computer-aided diagnosis of glaucoma using fundus images: A review. **Computer methods and programs in biomedicine**, Elsevier, v. 165, p. 1–12, 2018.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2015. p. 1026–1034.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HO, Y.; WOOKEY, S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. **IEEE Access**, v. 8, p. 4806–4813, 2020.

HOLM, S.; RUSSELL, G.; NOURRIT, V.; MCLOUGHLIN, N. Dr hagsis—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. **Journal of Medical Imaging**, Society of Photo-Optical Instrumentation Engineers, v. 4, n. 1, p. 014503–014503, 2017.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.

HUANG, F.; DASHTBOZORG, B.; ROMENY, B. M. t. H. Artery/vein classification using reflection features in retina fundus images. **Machine Vision and Applications**, Springer, v. 29, n. 1, p. 23–34, 2018.

IMRIE, C.; TATHAM, A. J. Glaucoma: The patient’s perspective. **British Journal of General Practice**, British Journal of General Practice, v. 66, n. 646, p. e371–e373, 2016.

JANSSENS, A. C. J. W.; MARTENS, F. K. Reflection on modern methods: Revisiting the area under the roc curve. **International Journal of Epidemiology**, v. 49, n. 4, p. 1397–1403, 01 2020. ISSN 0300-5771. Disponível em: <<https://doi.org/10.1093/ije/dyz274>>.

JIANG, J.; YE, C.; ZHANG, C.; LIN, Z.; TANG, Y.; YE, W.; XU, X.; ZHANG, S.; LU, F.; LIANG, Y. The patterns of visual field defects in primary angle-closure glaucoma compared to high-tension glaucoma and normal-tension glaucoma. **Ophthalmic Research**, S. Karger AG, v. 66, n. 1, p. 928–936, 2023.

JOSHI, S.; PARTIBANE, B.; HATAMLEH, W. A.; TARAZI, H.; YADAV, C. S.; KRAH, D. Glaucoma detection using image processing and supervised learning for classification. **Journal of Healthcare Engineering**, Wiley Online Library, v. 2022, n. 1, p. 2988262, 2022.

JUNEJA, M.; THAKUR, N.; THAKUR, S.; UNİYAL, A.; WANI, A.; JINDAL, P. Gc-net for classification of glaucoma in the retinal fundus image. **Machine Vision and Applications**, Springer, v. 31, p. 1–18, 2020.

KAUSU, T.; GOPI, V. P.; WAHID, K. A.; DOMA, W.; NIWAS, S. I. Combination of clinical and multiresolution features for glaucoma detection and its classification using fundus images. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 38, n. 2, p. 329–341, 2018.

KEANE, P. A.; SADDA, S. R. Retinal imaging in the twenty-first century: state of the art and future directions. **Ophthalmology**, Elsevier, v. 121, n. 12, p. 2489–2500, 2014.

KELLY, E.; WEN, Q.; HADDAD, D.; O’BANION, J. Effects of an aging population and racial demographics on eye disease prevalence: projections for georgia through 2050. **American journal of ophthalmology**, Elsevier, v. 210, p. 35–40, 2020.

KESKAR, N. S.; MUDIGERE, D.; NOCEDAL, J.; SMELYANSKIY, M.; TANG, P. T. P. On large-batch training for deep learning: Generalization gap and sharp minima. **arXiv preprint arXiv:1609.04836**, 2016.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KOTU, V.; DESHPANDE, B. Chapter 2 - data science process. In: KOTU, V.; DESHPANDE, B. (Ed.). **Data Science (Second Edition)**. Second edition. Morgan Kaufmann, 2019. p. 19–37. ISBN 978-0-12-814761-0. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128147610000022>>.

KROGH, A. What are artificial neural networks? **Nature biotechnology**, Nature Publishing Group US New York, v. 26, n. 2, p. 195–197, 2008.

KROGH, A.; HERTZ, J. A simple weight decay can improve generalization. **Advances in neural information processing systems**, v. 4, 1991.

KUMAR, J. H.; SEELAMANTULA, C. S.; KAMATH, Y. S.; JAMPALA, R. Rim-to-disc ratio outperforms cup-to-disc ratio for glaucoma prescreening. **Scientific reports**, Nature Publishing Group UK London, v. 9, n. 1, p. 7099, 2019.

KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. **Machine learning**, Springer, v. 51, n. 2, p. 181–207, 2003.

KURANI, A.; DOSHI, P.; VAKHARIA, A.; SHAH, M. A comprehensive comparative study of artificial neural network (ann) and support vector machines (svm) on stock forecasting. **Annals of Data Science**, Springer, v. 10, n. 1, p. 183–208, 2023.

KYARI, F.; ENTEKUME, G.; RABIU, M.; SPRY, P.; WORMALD, R.; NOLAN, W.; MURTHY, G. V.; GILBERT, C. E.; BLINDNESS, N. N.; GROUP, V. I. S. A population-based survey of the prevalence and types of glaucoma in nigeria: results from the nigeria national blindness and visual impairment survey. **BMC ophthalmology**, Springer, v. 15, p. 1–15, 2015.

LAIOS, K.; MOSCHOS, M. M.; ANDROUTSOS, G. Glaucoma and the origins of its name. **Journal of Glaucoma**, LWW, v. 25, n. 5, p. e507–e508, 2016.

LATHA, C. B. C.; JEEVA, S. C. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. **Informatics in Medicine Unlocked**, v. 16, p. 100203, 2019. ISSN 2352-9148. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S235291481830217X>>.

LATIF, J.; TU, S.; XIAO, C. *et al.* **ODGNet: a deep learning model for automated optic disc localization and glaucoma classification using fundus images**. **SN Appl Sci** 4: 98. 2022.

LATIF, J.; TU, S.; XIAO, C.; BILAL, A.; REHMAN, S. U.; AHMAD, Z. Enhanced nature inspired-support vector machine for glaucoma detection. **Computers, Materials & Continua**, v. 76, n. 1, 2023.

LAUDANI, A.; LOZITO, G. M.; FULGINEI, F. R.; SALVINI, A. On training efficiency and computational costs of a feed forward neural network: A review. **Computational Intelligence and Neuroscience**, v. 2015, n. 1, p. 818243, 2015. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1155/2015/818243>>.

LEFFLER, C. T.; SCHWARTZ, S. G.; HADI, T. M.; SALMAN, A.; VASUKI, V. The early history of glaucoma: the glaucous eye (800 bc to 1050 ad). **Clinical Ophthalmology**, Taylor & Francis, p. 207–215, 2015.

- LI, J.; HAN, Y.; ZHANG, M.; LI, G.; ZHANG, B. Multi-scale residual network model combined with global average pooling for action recognition. **Multimedia Tools and Applications**, Springer, v. 81, n. 1, p. 1375–1393, 2022.
- LI, L.; XU, M.; LIU, H.; LI, Y.; WANG, X.; JIANG, L.; WANG, Z.; FAN, X.; WANG, N. A large-scale database and a cnn model for attention-based glaucoma detection. **IEEE transactions on medical imaging**, IEEE, v. 39, n. 2, p. 413–424, 2019.
- LI, L.; XU, M.; LIU, H.; LI, Y.; WANG, X.; JIANG, L.; WANG, Z.; FAN, X.; WANG, N. A large-scale database and a cnn model for attention-based glaucoma detection. **IEEE Transactions on Medical Imaging**, v. 39, n. 2, p. 413–424, 2020.
- LIN, M.; CHEN, Q.; YAN, S. **Network In Network**. 2014. Disponível em: <<https://arxiv.org/abs/1312.4400>>.
- LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLÁR, P. Focal loss for dense object detection. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2017. p. 2980–2988.
- MAHALAKSHMI, K.; DINESHKUMAR, D.; KAMESH, S.; KISHORE, S. Glaucoma detection from retinal fundus images. In: **2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)**. [S.l.: s.n.], 2024. p. 1–6.
- MANCONI, A.; ARMANO, G.; GNOCCHI, M.; MILANESI, L. A soft-voting ensemble classifier for detecting patients affected by covid-19. **Applied Sciences**, v. 12, n. 15, 2022. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/12/15/7554>>.
- MANDAL, B.; OKEUKWU, A.; THEIS, Y. Masked face recognition using resnet-50. **arXiv preprint arXiv:2104.08997**, 2021.
- MARCHINI, G.; CHEMELLO, F.; BERZAGHI, D.; ZAMPIERI, A. New findings in the diagnosis and treatment of primary angle-closure glaucoma. **Progress in Brain Research**, Elsevier, v. 221, p. 191–212, 2015.
- MARTINS, J.; CARDOSO, J. S.; SOARES, F. Offline computer-aided diagnosis for glaucoma detection using fundus images targeted at mobile devices. **Computer Methods and Programs in Biomedicine**, Elsevier, v. 192, p. 105341, 2020.
- MASTERS, D.; LUSCHI, C. **Revisiting Small Batch Training for Deep Neural Networks**. 2018. Disponível em: <<https://arxiv.org/abs/1804.07612>>.
- MAYYA, V.; S, S. K.; KULKARNI, U.; SURYA, D. K.; ACHARYA, U. R. An empirical study of preprocessing techniques with convolutional neural networks for accurate detection of chronic ocular diseases using fundus images. **Applied Intelligence**, Springer, v. 53, n. 2, p. 1548–1566, 2023.
- MICROSOFT. **DirectML (DirectMachine Learning)**. 2023. <<https://learn.microsoft.com/en-us/windows/ai/directml/directml>>. Acessado em 21 de agosto de 2025. Disponível em: <<https://learn.microsoft.com/en-us/windows/ai/directml/directml>>.
- MIKOŁAJCZYK, A.; GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In: **2018 International Interdisciplinary PhD Workshop (IIPhDW)**. [S.l.: s.n.], 2018. p. 117–122.

MISHRA, C.; TRIPATHY, K. **Fundus Camera**. StatPearls Publishing, Treasure Island (FL), 2023. Disponível em: <<http://europepmc.org/books/NBK585111>>.

MISHRA, V. K.; KUMAR, S.; SHUKLA, N. Image acquisition and techniques to perform image acquisition. **SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology**, v. 9, n. 01, p. 21–24, 2017.

MITRA, A.; ROY, S.; ROY, S.; SETUA, S. K. Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract. **Computer Methods and Programs in Biomedicine**, v. 156, p. 169–178, 2018. ISSN 0169-2607. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169260717315389>>.

MK, S. B.; KALRA, M. Leveraging cnn and fundus imaging for enhanced glaucoma detection. **SN Computer Science**, Springer, v. 5, n. 8, p. 1–7, 2024.

MOHAMMADI, M.; RASHID, T. A.; KARIM, S. H.; ALDALWIE, A. H. M.; THO, Q. T.; BIDAKI, M.; RAHMANI, A. M.; HOSSEINZADEH, M. A comprehensive survey and taxonomy of the svm-based intrusion detection systems. **Journal of Network and Computer Applications**, v. 178, p. 102983, 2021. ISSN 1084-8045. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1084804521000102>>.

MOLČAN, S.; SMIEŠKOVÁ, M.; BACHRATÝ, H.; BACHRATÁ, K.; NOVOTNÝ, P. Classification of red blood cells using time-distributed convolutional neural networks from simulated videos. **Applied Sciences**, v. 13, n. 13, 2023. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/13/13/7967>>.

MORGAN, J. E.; BOURTSOUKLI, I.; RAJKUMAR, K. N.; ANSARI, E.; CUNLIFFE, I. A.; NORTH, R. V.; WILD, J. M. The accuracy of the inferior> superior> nasal> temporal neuroretinal rim area rule for diagnosing glaucomatous optic disc damage. **Ophthalmology**, Elsevier, v. 119, n. 4, p. 723–730, 2012.

MRAD, Y.; ELLOUMI, Y.; AKIL, M.; BEDOUI, M. H. A fast and accurate method for glaucoma screening from smartphone-captured fundus images. **Irbm**, Elsevier, v. 43, n. 4, p. 279–289, 2022.

MUKHERJEE, N.; SENGUPTA, S. Application of deep learning approaches for classification of diabetic retinopathy stages from fundus retinal images: a survey. **Multimedia Tools and Applications**, Springer, v. 83, n. 14, p. 43115–43175, 2024.

MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. **Array**, v. 16, p. 100258, 2022. ISSN 2590-0056. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2590005622000911>>.

MURTHY, G. J.; ARIGA, M.; SINGH, M.; GEORGE, R.; SARMA, P.; DUBEY, S.; CHOUDHRY, R. M.; PARIKH, R.; PANDAY, M. A deep dive into the latest european glaucoma society and asia-pacific glaucoma society guidelines and their relevance to india. **Indian journal of ophthalmology**, Medknow, v. 70, n. 1, p. 24–35, 2022.

NIRTHIKA, R.; MANIVANNAN, S.; RAMANAN, A.; WANG, R. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. **Neural Computing and Applications**, Springer, v. 34, n. 7, p. 5321–5347, 2022.

NOLAN, W.; YIP, J. L. Prevalence and geographical variations. In: **Glaucoma**. [S.l.]: Elsevier, 2015. p. 1–10.

NUGROHO, B.; YUNIARTI, A. Performance of root-mean-square propagation and adaptive gradient optimization algorithms on covid-19 pneumonia classification. In: **2022 IEEE 8th Information Technology International Seminar (ITIS)**. [S.l.: s.n.], 2022. p. 333–338.

ODSTRCILIK, J.; KOLAR, R.; BUDAI, A.; HORNEGGER, J.; JAN, J.; GAZAREK, J.; KUBENA, T.; CERNOSEK, P.; SVOBODA, O.; ANGELOPOULOU, E. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. **IET Image Processing**, Wiley Online Library, v. 7, n. 4, p. 373–383, 2013.

ORLANDO, J. I.; FU, H.; Barbosa Breda, J.; van Keer, K.; BATHULA, D. R.; DIAZ-PINTO, A.; FANG, R.; HENG, P.-A.; KIM, J.; LEE, J.; LEE, J.; LI, X.; LIU, P.; LU, S.; MURUGESAN, B.; NARANJO, V.; PHAYE, S. S. R.; SHANKARANARAYANA, S. M.; SIKKA, A.; SON, J.; van den Hengel, A.; WANG, S.; WU, J.; WU, Z.; XU, G.; XU, Y.; YIN, P.; LI, F.; ZHANG, X.; XU, Y.; BOGUNOVIĆ, H. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. **Medical Image Analysis**, v. 59, p. 101570, 2020. ISSN 1361-8415. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1361841519301100>>.

PANWAR, N.; HUANG, P.; LEE, J.; KEANE, P. A.; CHUAN, T. S.; RICHHARIYA, A.; TEOH, S.; LIM, T. H.; AGRAWAL, R. Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide healthcare. **Telemedicine and e-Health**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 22, n. 3, p. 198–208, 2016.

PARASHAR, D.; AGRAWAL, D. K. Automatic classification of glaucoma stages using two-dimensional tensor empirical wavelet transform. **IEEE Signal Processing Letters**, v. 28, p. 66–70, 2021.

PARK, J.; YI, D.; JI, S. A novel learning rate schedule in optimization for neural networks and it's convergence. **Symmetry**, v. 12, n. 4, 2020. ISSN 2073-8994. Disponível em: <<https://www.mdpi.com/2073-8994/12/4/660>>.

PAVLOVIC, V.; SCHONFELD, D.; FRIEDMAN, G. Stochastic noise process enhancement of hopfield neural networks. **IEEE Transactions on Circuits and Systems II: Express Briefs**, v. 52, n. 4, p. 213–217, 2005.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. *et al.* Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011.

PEREIRA, F.; MITCHELL, T.; BOTVINICK, M. Machine learning classifiers and fmri: A tutorial overview. **NeuroImage**, v. 45, n. 1, Supplement 1, p. S199–S209, 2009. ISSN 1053-8119. Mathematics in Brain Imaging. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1053811908012263>>.

PEREZ, L.; WANG, J. **The Effectiveness of Data Augmentation in Image Classification using Deep Learning**. 2017. Disponível em: <<https://arxiv.org/abs/1712.04621>>.

PETERS, H. J. M. Perceptrons. In: _____. **Artificial Neural Networks: An Introduction to ANN Theory and Practice**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995. p. 67–81. ISBN 978-3-540-49283-2. Disponível em: <<https://doi.org/10.1007/BFb0027023>>.

PICTON, P. Perceptrons. **Introduction to Neural Networks**, Springer, p. 25–45, 1994.

POON, L. Y.-C.; VALLE, D. S.-D.; TURALBA, A. V.; FALKENSTEIN, I. A.; HORSLEY, M.; KIM, J. H.; SONG, B. J.; TAKUSAGAWA, H. L.; WANG, K.; CHEN, T. C. The isn't rule: how often does it apply to disc photographs and retinal nerve fiber layer measurements in the normal population? **American journal of ophthalmology**, Elsevier, v. 184, p. 19–27, 2017.

RAGHAVENDRA, U.; GUDIGAR, A.; BHANDARY, S. V.; RAO, T. N.; CIACCIO, E. J.; ACHARYA, U. R. A two layer sparse autoencoder for glaucoma identification with fundus images. **Journal of medical systems**, Springer, v. 43, p. 1–9, 2019.

RAHMAN, A.; TASNIM, S. Ensemble classifiers and their applications: a review. **arXiv preprint arXiv:1404.4088**, 2014.

RAJARAMAN, S.; ANTANI, S. K.; POOSTCHI, M.; SILAMUT, K.; HOSSAIN, M. A.; MAUDE, R. J.; JAEGER, S.; THOMA, G. R. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. **PeerJ**, PeerJ Inc., v. 6, p. e4568, 2018.

REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-validation. In: _____. **Encyclopedia of Database Systems**. Boston, MA: Springer US, 2009. p. 532–538. ISBN 978-0-387-39940-9. Disponível em: <https://doi.org/10.1007/978-0-387-39940-9_565>.

RIGATTI, S. J. Random forest. **Journal of Insurance Medicine**, v. 47, n. 1, p. 31–39, 01 2017. ISSN 0743-6661. Disponível em: <<https://doi.org/10.17849/in-sm-47-01-31-39.1>>.

ROSENBLATT, M.; TEJAVIBULYA, L.; JIANG, R.; NOBLE, S.; SCHEINOST, D. Data leakage inflates prediction performance in connectome-based machine learning models. **Nature Communications**, Nature Publishing Group UK London, v. 15, n. 1, p. 1829, 2024.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986.

SABA, T.; BOKHARI, S. T. F.; SHARIF, M.; YASMIN, M.; RAZA, M. Fundus image classification methods for the detection of glaucoma: A review. **Microscopy research and technique**, Wiley Online Library, v. 81, n. 10, p. 1105–1121, 2018.

SAHOO, P. K.; MISHRA, S.; DAS, M.; SWAIN, S. C. Enhanced incipient fault identification in underground distribution cable implementing random forest classifier. **Microsystem Technologies**, Springer, p. 1–11, 2024.

SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 4510–4520.

SANGHAVI, J.; KURHEKAR, M. An efficient framework for optic disk segmentation and classification of glaucoma on fundus images. **Biomedical Signal Processing and Control**, v. 89, p. 105770, 2024. ISSN 1746-8094. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S174680942301203X>>.

SCHREUDER, A.; BOSMAN, A.; ENGELBRECHT, A.; CLEGHORN, C. Training feedforward neural networks with bayesian hyper-heuristics. **Information Sciences**, v. 686, p. 121363, 2025. ISSN 0020-0255. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025524012775>>.

SCHUSTER, A. K.; ERB, C.; HOFFMANN, E. M.; DIETLEIN, T.; PFEIFFER, N. The diagnosis and treatment of glaucoma. **Deutsches Ärzteblatt International**, Deutscher Ärzte-Verlag GmbH, v. 117, n. 13, p. 225, 2020.

SHABBIR, A.; RASHEED, A.; SHEHRAZ, H.; SALEEM, A.; ZAFAR, B.; SAJID, M.; ALI, N.; DAR, S. H.; SHEHRYAR, T. Detection of glaucoma using retinal fundus images: A comprehensive review. **Mathematical Biosciences and Engineering**, v. 18, n. 3, p. 2033–2076, 2021.

SHAFIQ, M.; GU, Z. Deep residual learning for image recognition: A survey. **Applied Sciences**, v. 12, n. 18, 2022. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/12/18/8972>>.

SHALLUE, C. J.; LEE, J.; ANTOGNINI, J.; SOHL-DICKSTEIN, J.; FROSTIG, R.; DAHL, G. E. Measuring the effects of data parallelism on neural network training. **Journal of Machine Learning Research**, v. 20, n. 112, p. 1–49, 2019.

SHARMA, S. K.; MUDULI, D.; PRIYADARSHINI, R.; KUMAR, R. R.; KUMAR, A.; PRADHAN, J. An evolutionary supply chain management service model based on deep learning features for automated glaucoma detection using fundus images. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 128, p. 107449, 2024.

SHEN, Z.; FU, H.; SHEN, J.; SHAO, L. Modeling and enhancing low-quality retinal fundus images. **IEEE transactions on medical imaging**, IEEE, v. 40, n. 3, p. 996–1006, 2020.

SHEYKHMUSA, M.; MAHDIANPARI, M.; GHANBARI, H.; MOHAMMADIMANESH, F.; GHAMISI, P.; HOMAYOUNI, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 6308–6325, 2020.

SHINDE, R. Glaucoma detection in retinal fundus images using u-net and supervised machine learning algorithms. **Intelligence-Based Medicine**, Elsevier, v. 5, p. 100038, 2021.

SHOBA, S. G.; THERESE, A. B. Detection of glaucoma disease in fundus images based on morphological operation and finite element method. **Biomedical Signal Processing and Control**, Elsevier, v. 62, p. 101986, 2020.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of big data**, Springer, v. 6, n. 1, p. 1–48, 2019.

SHORTEN, C.; KHOSHGOFTAAR, T. M.; FURHT, B. Text data augmentation for deep learning. **Journal of big Data**, Springer, v. 8, n. 1, p. 101, 2021.

SHOUKAT, A.; AKBAR, S.; HASSAN, S. A.; IQBAL, S.; MEHMOOD, A.; ILYAS, Q. M. Automatic diagnosis of glaucoma from retinal images using deep learning approach. **Diagnostics**, MDPI, v. 13, n. 10, p. 1738, 2023.

SHYAMALEE, T.; MEEDENIYA, D. Glaucoma detection with retinal fundus images using segmentation and classification. **Machine Intelligence Research**, Springer, v. 19, n. 6, p. 563–580, 2022.

SHYAMALEE, T.; MEEDENIYA, D.; LIM, G.; KARUNARATHNE, M. Automated tool support for glaucoma identification with explainability using fundus images. **IEEE Access**, v. 12, p. 17290–17307, 2024.

SIDDIQUI, M. K.; MORALES-MENENDEZ, R.; HUANG, X.; HUSSAIN, N. A review of epileptic seizure detection using machine learning classifiers. **Brain informatics**, Springer, v. 7, n. 1, p. 5, 2020.

SINGH, L. K.; POOJA; GARG, H.; KHANNA, M. Deep learning system applicability for rapid glaucoma prediction from fundus images across various data sets. **Evolving Systems**, Springer, v. 13, n. 6, p. 807–836, 2022.

SIVASWAMY, J.; KRISHNADAS, S. R.; JOSHI, G. D.; JAIN, M.; TABISH, A. U. S. Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. In: **2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)**. [S.l.: s.n.], 2014. p. 53–56.

SKOUTA, A.; ELMOUFIDI, A.; JAI-ANDALOUSSI, S.; OUCHETTO, O. Advances in the application of convolutional neural networks for glaucoma diagnosis. In: **2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)**. [S.l.: s.n.], 2023. p. 1–8.

SONG, L.; FAN, J.; CHEN, D.-R.; ZHOU, D.-X. Approximation of nonlinear functionals using deep relu networks. **Journal of Fourier Analysis and Applications**, Springer, v. 29, n. 4, p. 50, 2023.

SOOMRO, K.; ZAMIR, A. R.; SHAH, M. **UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild**. 2012. Disponível em: <<https://arxiv.org/abs/1212.0402>>.

SOUDY, A. H.; SAYED, O.; TAG-ELSER, H.; RAGAB, R.; MOHSEN, S.; MOSTAFA, T.; ABOHANY, A. A.; SLIM, S. O. Deepfake detection using convolutional vision transformers and convolutional neural networks. **Neural Computing and Applications**, Springer, v. 36, n. 31, p. 19759–19775, 2024.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

STAAL, J.; ABRAMOFF, M.; NIEMEIJER, M.; VIERGEVER, M.; GINNEKEN, B. van. Ridge-based vessel segmentation in color images of the retina. **IEEE Transactions on Medical Imaging**, v. 23, n. 4, p. 501–509, 2004.

STACKE, K.; EILERTSEN, G.; UNGER, J.; LUNDSTRÖM, C. Measuring domain shift for deep learning in histopathology. **IEEE Journal of Biomedical and Health Informatics**, v. 25, n. 2, p. 325–336, 2021.

SUN, X.; DAI, Y.; CHEN, Y.; YU, D.-Y.; CRINGLE, S. J.; CHEN, J.; KONG, X.; WANG, X.; JIANG, C. Primary angle closure glaucoma: what we know and what we don't know. **Progress in retinal and eye research**, Elsevier, v. 57, p. 26–45, 2017.

SUZUKI, K. **Artificial neural networks: methodological advances and biomedical applications**. [S.l.]: BoD–Books on Demand, 2011.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 2818–2826.

TALEBI, H.; MILANFAR, P. **Learning to Resize Images for Computer Vision Tasks**. 2021. Disponível em: <<https://arxiv.org/abs/2103.09950>>.

TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. **International conference on machine learning**. [S.l.], 2019. p. 6105–6114.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. 2020. Disponível em: <<https://arxiv.org/abs/1905.11946>>.

THAKUR, A.; VENU, S.; GURUSAMY, M. An extensive review on agricultural robots with a focus on their perception systems. **Computers and Electronics in Agriculture**, v. 212, p. 108146, 2023. ISSN 0168-1699. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0168169923005343>>.

THAM, Y.-C.; LI, X.; WONG, T. Y.; QUIGLEY, H. A.; AUNG, T.; CHENG, C.-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. **Ophthalmology**, Elsevier, v. 121, n. 11, p. 2081–2090, 2014.

THONGINNETRA, O.; GREENSTEIN, V. C.; CHU, D.; LIEBMANN, J. M.; RITCH, R.; HOOD, D. C. Normal versus high tension glaucoma: a comparison of functional and structural defects. **Journal of glaucoma**, LWW, v. 19, n. 3, p. 151–157, 2010.

TR, J.; REDDY, N. S.; ACHARYA, U. D. Modeling daily reference evapotranspiration from climate variables: Assessment of bagging and boosting regression approaches. **Water Resources Management**, Springer, v. 37, n. 3, p. 1013–1032, 2023.

TRAGOUDARAS, A.; STOIKOS, P.; FANARAS, K.; TZIOUVARAS, A.; FLOROS, G.; DIMITRIOU, G.; KOLOMVATSOS, K.; STAMOULIS, G. Design space exploration of a sparse mobilenetv2 using high-level synthesis and sparse matrix techniques on fpgas. **Sensors**, v. 22, n. 12, 2022. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/22/12/4318>>.

TRAN, K.; MENDEL, T. A.; HOLBROOK, K. L.; YATES, P. A. Construction of an inexpensive, hand-held fundus camera through modification of a consumer “point-and-shoot” camera. **Investigative ophthalmology & visual science**, The Association for Research in Vision and Ophthalmology, v. 53, n. 12, p. 7600–7607, 2012.

VALDEZ-RODRÍGUEZ, J. E.; FELIPE-RIVERÓN, E. M.; CALVO, H. Optic disc preprocessing for reliable glaucoma detection in small datasets. **Mathematics**, v. 9, n. 18, 2021. ISSN 2227-7390. Disponível em: <<https://www.mdpi.com/2227-7390/9/18/2237>>.

VAROQUAUX, G.; RAAMANA, P. R.; ENGEMANN, D. A.; HOYOS-IDROBO, A.; SCHWARTZ, Y.; THIRION, B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. **NeuroImage**, Elsevier, v. 145, p. 166–179, 2017.

VELPULA, V. K.; SHARMA, L. D. Automatic glaucoma detection from fundus images using deep convolutional neural networks and exploring networks behaviour using visualization techniques. **SN Computer Science**, Springer, v. 4, n. 5, p. 487, 2023.

VERNAZZA, S.; TIRENDI, S.; BASSI, A. M.; TRAVERSO, C. E.; SACCÀ, S. C. Neuroinflammation in primary open-angle glaucoma. **Journal of clinical medicine**, MDPI, v. 9, n. 10, p. 3172, 2020.

WALCZAK, S. Artificial neural networks. In: **Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction**. [S.l.]: IGI Global Scientific Publishing, 2019. p. 40–53.

WANG, J.; LAN, C.; LIU, C.; OUYANG, Y.; QIN, T.; LU, W.; CHEN, Y.; ZENG, W.; YU, P. S. Generalizing to unseen domains: A survey on domain generalization. **IEEE Transactions on Knowledge and Data Engineering**, v. 35, n. 8, p. 8052–8072, 2023.

WANG, S.-C. Artificial neural network. In: _____. **Interdisciplinary Computing in Java Programming**. Boston, MA: Springer US, 2003. p. 81–100. ISBN 978-1-4615-0377-4. Disponível em: <https://doi.org/10.1007/978-1-4615-0377-4_5>.

WANG, Y.; LI, Y.; SONG, Y.; RONG, X. The influence of the activation function in a convolution neural network model of facial expression recognition. **Applied Sciences**, v. 10, n. 5, 2020. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/10/5/1897>>.

WARDHANI, N. W. S.; ROCHAYANI, M. Y.; IRIANY, A.; SULISTYONO, A. D.; LESTANTYO, P. Cross-validation metrics for evaluating classification performance on imbalanced data. In: **2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)**. [S.l.: s.n.], 2019. p. 14–18.

WEN, L.; HUGHES, M. Coastal wetland mapping using ensemble learning algorithms: A comparative study of bagging, boosting and stacking techniques. **Remote Sensing**, v. 12, n. 10, 2020. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/12/10/1683>>.

WILLOUGHBY, C. E.; PONZIN, D.; FERRARI, S.; LOBO, A.; LANDAU, K.; OMIDI, Y. Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function—a review. **Clinical & Experimental Ophthalmology**, Wiley Online Library, v. 38, p. 2–11, 2010.

WONG, T.-T.; YEH, P.-Y. Reliable accuracy estimates from k-fold cross validation. **IEEE Transactions on Knowledge and Data Engineering**, v. 32, n. 8, p. 1586–1594, 2020.

WOSTYN, P.; KILLER, H. E. Normal-tension glaucoma: A glymphopathy? **Eye and Brain**, Taylor & Francis, p. 37–44, 2023.

WRIGHT, C.; TAWFIK, M. A.; WAISBOURD, M.; KATZ, L. J. Primary angle-closure glaucoma: an update. **Acta ophthalmologica**, Wiley Online Library, v. 94, n. 3, p. 217–225, 2016.

WU, Y.-c.; FENG, J.-w. Development and application of artificial neural network. **Wireless Personal Communications**, Springer, v. 102, p. 1645–1656, 2018.

XIONG, Z.; GUO, Q.; LIU, M.; LI, A. Pan-sharpening based on convolutional neural network by using the loss function with no-reference. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 897–906, 2021.

YADAV, A.; YADAV, K.; Anirbid Sircar. Feedforward neural network for joint inversion of geophysical data to identify geothermal sweet spots in gandhar, gujarat, india. **Energy Geoscience**, v. 2, n. 3, p. 189–200, 2021. ISSN 2666-7592. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666759221000019>>.

YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, Springer, v. 9, p. 611–629, 2018.

YANG, T.; YING, Y. AUC maximization in the era of big data and AI: A survey. **ACM computing surveys**, ACM New York, NY, v. 55, n. 8, p. 1–37, 2022.

YAQUB, M.; FENG, J.; ZIA, M. S.; ARSHID, K.; JIA, K.; REHMAN, Z. U.; MEHMOOD, A. State-of-the-art cnn optimizer for brain tumor segmentation in magnetic resonance images. **Brain Sciences**, v. 10, n. 7, 2020. ISSN 2076-3425. Disponível em: <<https://www.mdpi.com/2076-3425/10/7/427>>.

YOU, K.; LONG, M.; CAO, Z.; WANG, J.; JORDAN, M. I. Universal domain adaptation. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2019. p. 2720–2729.

YU, R.; WANG, Y.; ZOU, Z.; WANG, L. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. **Transportation Research Part C: Emerging Technologies**, v. 119, p. 102740, 2020. ISSN 0968-090X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0968090X20306549>>.

ZAFAR, A.; AAMIR, M.; NAWI, N. M.; ARSHAD, A.; RIAZ, S.; ALRUBAN, A.; DUTTA, A. K.; ALMOTAIRI, S. A comparison of pooling methods for convolutional neural networks. **Applied Sciences**, v. 12, n. 17, 2022. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/12/17/8643>>.

ZAIDI, S. S. A.; ANSARI, M. S.; ASLAM, A.; KANWAL, N.; ASGHAR, M.; LEE, B. A survey of modern deep learning based object detection models. **Digital Signal Processing**, v. 126, p. 103514, 2022. ISSN 1051-2004. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1051200422001312>>.

ZEINI, H. A.; AL-JEZNAWI, D.; IMRAN, H.; BERNARDO, L. F. A.; AL-KHAFAJI, Z.; OSTROWSKI, K. A. Random forest algorithm for the strength prediction of geopolymer stabilized clayey soil. **Sustainability**, v. 15, n. 2, 2023. ISSN 2071-1050. Disponível em: <<https://www.mdpi.com/2071-1050/15/2/1408>>.

ZHANG, Y.; ZHANG, Y.; PENG, L.; QUAN, L.; ZHENG, S.; LU, Z.; CHEN, H. Base-2 softmax function: Suitability for training and efficient hardware implementation. **IEEE Transactions on Circuits and Systems I: Regular Papers**, v. 69, n. 9, p. 3605–3618, 2022.

ZHANG, Z.; YIN, F. S.; LIU, J.; WONG, W. K.; TAN, N. M.; LEE, B. H.; CHENG, J.; WONG, T. Y. 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**. [S.l.: s.n.], 2010. p. 3065–3068.

ZHANG, Z.; ZHAO, Y.; CANES, A.; STEINBERG, D.; LYASHEVSKA, O. *et al.* Predictive analytics with gradient boosting in clinical medicine. **Annals of translational medicine**, AME Publications, v. 7, n. 7, 2019.

ZHAO, X.; WANG, L.; ZHANG, Y.; HAN, X.; DEVECI, M.; PARMAR, M. A review of convolutional neural networks in computer vision. **Artificial Intelligence Review**, Springer, v. 57, n. 4, p. 99, 2024.

ZHOU, Y.; WANG, X.; ZHANG, M.; ZHU, J.; ZHENG, R.; WU, Q. Mpce: A maximum probability based cross entropy loss function for neural network classification. **IEEE Access**, v. 7, p. 146331–146341, 2019.

ZHU, J.; SHEN, Y.; ZHAO, D.; ZHOU, B. In-domain gan inversion for real image editing. In: SPRINGER. **European conference on computer vision**. [S.l.], 2020. p. 592–608.

ZHUANG, F.; QI, Z.; DUAN, K.; XI, D.; ZHU, Y.; ZHU, H.; XIONG, H.; HE, Q. A comprehensive survey on transfer learning. **Proceedings of the IEEE**, v. 109, n. 1, p. 43–76, 2021.

ŞEN, S. Y.; ÖZKURT, N. Convolutional neural network hyperparameter tuning with adam optimizer for ecg classification. In: **2020 Innovations in Intelligent Systems and Applications Conference (ASYU)**. [S.l.: s.n.], 2020. p. 1–6.