

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET
DEPARTAMENTO DE COMPUTAÇÃO– DC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO– PPGCC

Gabriel Diego de Aguiar Aranha

**Uma Nova Estratégia de Transferência
de Aprendizado Profundo Visando o
Diagnóstico de Condições e Doenças
Relacionadas aos Olhos**

Gabriel Diego de Aguiar Aranha

**Uma Nova Estratégia de Transferência
de Aprendizado Profundo Visando o
Diagnóstico de Condições e Doenças
Relacionadas aos Olhos**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Metodologias e Técnicas de Computação

Orientador: Prof. Dr. Ricardo Augusto Souza Fernandes

São Carlos

2023

*Neste percurso, as estrelas guiaram e as tempestades abraçaram.
A sabedoria floresceu revelando o infinito daqueles que ousam sonhar.*

Agradecimentos

Agradeço infinitamente à Suely de Aguiar, minha amada mãe, estrela que brilha além da existência, habitando em cada manifestação de amor, dedicação e benevolência que exerço. Seu amparo angelical dissipou desistências e seu acalanto vitalizou minha alma.

À minha etérea parceira de vida, Larissa, que ameniza as labaredas das adversidades com seu amor sublime, inspirando-me a cada alvorecer. Aos amigos que me ensinaram a ressignificar o conceito de família, tecendo laços de acolhimento e companheirismo que me impulsionam a viver. E aos sábios orientadores, com afabilidade e estímulo, lapidaram meu caminho rumo ao conhecimento.

*“A verdadeira função do ensino é permitir, aqui e agora,
o máximo de comunicação possível entre as diversas
perspectivas de uma mesma verdade.”
(Michel Foucault)*

Resumo

Dados da Organização Mundial da Saúde indicam que bilhões de casos de deficiência visual poderiam ser evitados, principalmente com exames regulares. No entanto, a ausência de especialistas nas unidades básicas de saúde resulta na falta de diagnóstico preciso de doenças oculares sistêmicas ou assintomáticas, aumentando os casos de cegueira. Nesse contexto, a presente tese propõe um conjunto de redes neurais convolucionais, que foram submetidas a um processo de transferência de aprendizado advindo de 38.727 imagens de fundos oculares de alta qualidade, para assim melhorar a inferência de 7.850 imagens de baixa qualidade adquiridas por equipamentos de baixo custo. Também foram utilizados bancos de dados públicos como Eyepacs, Messidor-2, REFUGE e ODIR para a validação da arquitetura proposta nas condições oculares de retinopatia diabética referenciável, glaucoma e catarata, alcançando resultados de AUC de 0,951 e 0,953 e precisões de 99,6% e 99,3%, respectivamente. A partir de imagens de baixa qualidade, a abordagem proposta alcançou precisões de 87,4%, 90,8%, 87,5% e 79,1% para classificar catarata, retinopatia diabética referenciável, escavação anormal e vasos sanguíneos anormais, respectivamente. Assim, a abordagem proposta contribui para avançar o estado da arte em termos de: (i) validar a estratégia de transferência de aprendizado proposta, reconhecendo condições e doenças oculares relacionadas em imagens de baixa qualidade; (ii) usar imagens de alta qualidade obtidas por equipamentos de alto custo apenas para treinar os modelos preditivos; (iii) atingir resultados comparáveis ao estado da arte, mesmo usando imagens de baixa qualidade. Somado ao fato de que a estratégia de transferência de aprendizado profunda é mais adequada e viável para ser aplicada pelos sistemas de saúde pública de países emergentes e em desenvolvimento, a presente tese também propõe um sistema para fornecer imagens Grad-CAM de previsões patológicas com valores acima de 75% para apoiar o processo de tomada de decisão.

Palavras-chave: Aprendizado profundo, Redes neurais convolucionais, Retinopatia, Transferência de aprendizado.

Abstract

Data from the World Health Organization indicate that billions of cases of visual impairment could be avoided, mainly with regular examinations. However, the absence of specialists in basic health units has resulted in a lack of accurate diagnosis of systemic or asymptomatic eye diseases, increasing cases of blindness. In this context, the present thesis proposes a set of convolutional neural networks, which were submitted to a transfer learning process originating from 38,727 high-quality ocular fundus images, to improve the inference of 7,850 low-quality images acquired by low-cost equipment. Public databases such as Eyepacs, Messidor-2, REFUGE, and ODIR were also used for the validation of the proposed architecture in the ocular conditions of referable diabetic retinopathy, glaucoma, and cataract, achieving AUC results of 0.951 and 0.953 and accuracies of 99.6% and 99.3%, respectively. From low-quality images, the proposed approach achieved accuracies of 87.4%, 90.8%, 87.5%, and 79.1% for classifying cataract, referable diabetic retinopathy, abnormal excavation, and abnormal blood vessels, respectively. Thus, the proposed approach contributes to advancing the state-of-the-art in terms of: (i) validating the proposed transfer learning strategy, recognizing related ocular conditions and diseases in low-quality images; (ii) using high-quality images obtained by high-cost equipment only for training the predictive models; (iii) achieving results comparable to the state of the art, even using low-quality images. In addition to the fact that the deep transfer learning strategy is more suitable and feasible for application by public health systems in emerging and developing countries, the present thesis also proposes a system for providing Grad-CAM images of pathological predictions with values above 75% to support the decision-making process.

Keywords: Deep learning, Convolutional neural networks, Retinopathy, Transfer learning.

Lista de ilustrações

Figura 1 – Exemplo de mapeamento de uma imagem em um Deep Learning System (DLS).	25
Figura 2 – Anatomia ocular.	36
Figura 3 – Camadas da retina.	38
Figura 4 – Comparação entre imagem do fundo ocular colorida (A) com hemorragia e tecido fibrovascular e imagem de angiografia (B) de um vazamento de neovascularização coroidal.	39
Figura 5 – Exemplo de um dispositivo de captura chamado nonmyd WX.	40
Figura 6 – Exemplo de um aparelho portátil chamado Picton e sua respectiva imagem gerada colorida e em tons de cinza.	41
Figura 7 – (A) Exemplo de imagens de alta qualidade. (B) Exemplos de imagens de baixa qualidade.	42
Figura 8 – Exemplo de Optical Coherence Tomography (OCT): (A) imagem colorida do fundo ocular de um paciente com drusas. Um recorte delimitado em branco (6×6 mm) foi sobreposto na imagem para representar a área de digitalização; (B) SD-OCT correspondente à linha central do recorte; e (C) segmentação do Epitélio Pigmentar da Retina (EPR) e uma linha vermelha mostrando sua base.	42
Figura 9 – Exemplo de uma rede neural artificial.	44
Figura 10 – Representação de um único neurônio.	45
Figura 11 – Representação esquemática de uma arquitetura típica de CNN, composta por camadas convolucionais, de agrupamento e totalmente conectadas.	46
Figura 12 – Estágios de MLOps.	48
Figura 13 – Arquitetura do projeto.	52
Figura 14 – Processo de classificação sequencial (<i>pipeline</i>)	53

Figura 15 – (A) 3Nethra Classic, (B) Canon CR2, © Phelcom Eyer, (D) Topcon NW100.	54
Figura 16 – Exemplo de um laudo.	55
Figura 17 – Exemplo de Imagem Completa Capturada pelo Aparelho Topcon NW100.	56
Figura 18 – Exemplo de imagens removidas.	56
Figura 19 – Exemplo da Aplicação do Filtro Canny em uma Retinografia.	58
Figura 20 – Exemplo da Aplicação do Filtro Suavização Conservadora em uma Re- tinografia.	59
Figura 21 – Exemplo da Aplicação do Filtro Unsharp em uma Retinografia.	60
Figura 22 – Arquitetura da VGG16 adaptada para a classificação de patologias da retina.	63
Figura 23 – Mapa de características adquirido pelo mapeamento das ativações na camada GAP.	66
Figura 24 – Ilustração da Aplicação de Grad-CAM em Retinografias.	67
Figura 25 – Fluxo da Infraestrutura do Sistema Desenvolvido.	76
Figura 26 – Representação esquemática da interface de usuário do sistema Streamlit.	79
Figura 27 – Representação esquemática da interface de usuário do sistema Streamlit.	80

Lista de tabelas

Tabela 1 – Quantidade de imagens por classe e resolução.	55
Tabela 2 – Comparação dos modelos baseada em acurácia, custo computacional e quantidade de dados para evitar (<i>overfitting</i>) utilizando EyePACS.	62
Tabela 3 – Especificações da Arquitetura Proposta	64
Tabela 4 – Matriz de Confusão de Todos os Modelos.	70
Tabela 5 – Resultados das métricas escolhidas para cada modelo, a segunda linha da tabela mostra o Wilson Score.	70
Tabela 6 – Comparação entre os modelos sem e com aprendizado transferido.	71
Tabela 7 – Comparação com os artigos que compõem o estado da arte sobre conjuntos de dados públicos de RD.	73
Tabela 8 – Comparação com os artigos de estado-da-arte sobre conjuntos de dados públicos de catarata.	73
Tabela 9 – Comparação com os artigos de estado-da-arte sobre conjunto de dados públicos de glaucoma.	74
Tabela 10 – Exemplos de hiperparâmetros e intervalos de valores.	76

Lista de siglas

ADC Apoio à Decisão Clínica

CAM Class Activation Map

CNN Convolutional Neural Network

CI/CD Continuous Integration / Continuous Development

DLS Deep Learning System

DMRI Degeneração Macular Relacionado à Idade

DNN Deep Neural Network

EPR Epitélio Pigmentar da Retina

GAP Global Average Pooling

GAN Generative Adversarial Network

Grad-CAM Gradient-weighted Class Activation Mapping

kNN k-Nearest Neighbors

MLP Multilayer Perceptron

MLOps Machine Learning Operations

OCR Optical Character Recognition

OCT Optical Coherence Tomography

OMS Organização Mundial da Saúde

RD Retinopatia Diabética

RNA Rede Neural Artificial

SVM Support Vector Machines

SGD Stochastic Gradient Descent

Sumário

1	INTRODUÇÃO	21
1.1	Justificativas	25
1.2	Objetivos	26
1.3	Organização da Tese	28
2	REVISÃO BIBLIOGRÁFICA	29
2.1	Classificação de Retinopatias	29
2.1.1	Catarata	30
2.1.2	Retinopatia diabética	31
2.1.3	Escavação do Nervo Óptico	32
2.1.4	Vasos Sanguíneos	33
3	CONCEITOS GERAIS DA ANATOMIA OCULAR E FUN-	
	DOSCOPIA	35
3.1	Anatomia Ocular	35
3.1.1	Mácula e Nervo Óptico	36
3.1.2	Camadas da Retina	37
3.2	Aquisição de Imagens Oculares	38
3.2.1	Imagens do Fundo Ocular	39
3.2.2	Tomografia de Coerência Óptica	40
4	DEEP LEARNING, CONVOLUTIONAL NEURAL NETWORK	
	E INTEGRAÇÃO MLOPS	43
4.1	Aprendizado Profundo e Redes Neurais	43
4.2	Convolutional Neural Networks (CNN)	45
4.3	Aprendizado Supervisionado e Não Supervisionado em Visão	
	Computacional	46

4.4	Operações de Aprendizado de Máquina	47
5	METODOLOGIA	51
5.1	Conjunto de Dados e Laudos	52
5.2	Pré-processamento	54
5.2.1	Filtragem	57
5.3	Modelos Preditivos	60
5.3.1	VGG16 e VGG19	60
5.3.2	ResNet50	60
5.3.3	InceptionV3	61
5.3.4	InceptionResNetV2	61
5.3.5	MobileNetV2	61
5.3.6	Comparação dos Modelos e Critérios de Seleção	62
5.3.7	CNN Proposta	63
6	RESULTADOS	69
6.1	Base de Dados Privada	70
6.1.1	Modelo de Conhecimento Transferido	71
6.2	Bases de Dados Públicas	72
6.2.1	Retinopatia Diabética	72
6.2.2	Catarata	73
6.2.3	Glaucoma	73
6.2.4	Análise Geral	74
6.3	Aspectos da Implementação do Modelo Proposto em Aplicação Web	75
6.3.1	Módulo de Treinamento e Otimização	75
6.3.2	Módulo de API	76
6.3.3	Módulo de Predição e Visualização	77
7	CONCLUSÕES	81
	REFERÊNCIAS	83

Capítulo 1

Introdução

Desde a possível utilização de sistemas que apoiam a análise de imagens digitais, pesquisas são elaboradas massivamente a fim de encontrar padrões que resultem no diagnóstico de doenças, compreensão de tratamentos, eficácia de drogas e similares (LITJENS et al., 2017). No que cerne os diagnósticos, diversos métodos são utilizados, variando de acordo com a finalidade clínica ou sintomas, dentre eles são comuns os exames de coleta de secreções e de sangue que, por sua vez, são dolorosos e invasivos (ROSE; CHOI, 2015). Entretanto, existem também exames não invasivos e indolores, como por exemplo a análise da retina através de retinografia ou tomografia de coerência óptica (OCT, do inglês *Optical Coherence Tomography*) (FERCHER, 1996).

O exame de mapeamento de retina, ou fundoscopia, é responsável pela detecção de má formações e doenças oculares, como por exemplo glaucoma, toxoplasmose, retinopatia diabética, retinopatia hipertensiva, entre outras uveítes (inflamações oculares). Entretanto, o exame é promissor no diagnóstico de outras doenças que se manifestam pelo corpo todo, mesmo em indivíduos assintomáticos. A corroboração deste conceito é advinda do fato de que, é possível analisar diretamente os vasos sanguíneos através dos olhos. Assim sendo, é factível o reconhecimento de patologias neurológicas, reumáticas, hematológicas, entre outras (FRAZ et al., 2012).

Neste sentido, o crescente avanço tecnológico contribui de forma efetiva e impulsiona o desenvolvimento de equipamentos que permitam a aquisição de imagens cada vez mais nítidas do mapeamento da retina. Estas imagens são provenientes, em sua grande maioria, de exames de retinografia e, recentemente, de exames de OCT. A retinografia produz uma imagem que focaliza a membrana do seguimento posterior do olho, denominada retina, e tem como objetivo captar as estruturas que compõem a mesma, que segundo Ryan, Schachat e Wilkinson (2006), são:

- ❑ Coroide – estrutura vascular que irriga a porção externa da retina (aproximadamente metade dela);
- ❑ Nervo óptico – fissura que conduz os axônios dos neurônios para o quiasma óptico;
- ❑ Vasos sanguíneos – irrigam a outra porção da retina;
- ❑ Mácula – área no centro da retina responsável pela visão nítida.

As retinografias são comumente segmentadas em dois tipos, o primeiro é a retinografia colorida e o segundo a retinografia fluorescente, também intitulada como angiofluoresceinografia. Ambos os tipos de exames fazem um mapeamento do fundo do olho, e as imagens são habitualmente divididas em segmentos (polo, posterior, superior, inferior, nasal e temporal) para maior amplitude e visualização. Contudo, a angiofluoresceinografia difere da retinografia colorida por utilizar um método de contraste, frequentemente com o composto orgânico fluoresceína (SHANKLE, 2004). Outro tipo de aquisição de imagem além da retinografia é o OCT, este porém tem como objetivo mapear a retina de uma perspectiva transversal por meio de ondas de luz, possibilitando assim que o profissional observe cada uma das camadas distintas da retina (HUANG et al., 1991).

Como citado anteriormente, inúmeras doenças amplamente pesquisadas se manifestam na retina, sendo originárias do olho, cérebro ou sistema cardiovascular (ABRAMOFF; GARVIN; SONKA, 2010). Com o intuito de saber como identificá-las e analisá-las, torna-se necessária a explanação de suas características. Assim, entre as principais patologias tem-se as seguintes características:

1. Diabetes *mellitus* – a herança genética, a obesidade e o estilo de vida sedentário conferem maior risco de desenvolvimento dessa doença. Hiperglicemia, ou seja, a presença elevada de glicose no sangue, resulta em danos nos vasos sanguíneos sejam eles pequenos ou grandes, bem como ocorre nas células nervosas, ao passo que os rins, coração, cérebro e olhos enfrentam as consequências deste quadro. Dessa forma, nos olhos, há a percepção da retinopatia diabética (WHO, 2018);
2. Retinopatia Diabética (RD) – é uma derivação da diabetes *mellitus* e responsável pelo incremento nos casos de cegueira na população mundial. No último estudo efetuado com a população dos Estados Unidos, em 2015, a *American Diabetes Association* constatou que 30,3 milhões de pessoas estavam com diabetes, isto representa 9,4% da população estadunidense (ADA, 2018). Há evidências abundantes de que a cegueira e a perda visual nesses pacientes podem ser evitadas através de triagem anual e diagnóstico precoce (KLONOFF; SCHWARTZ, 2000). Nos olhos, a hiperglicemia danifica as paredes dos vasos da retina, o que pode levar aos casos de:

-
- a) Isquemia – aumento na criação de vasos novos (neovasos) que podem subsequentemente sofrer hemorragias e/ou causar descolamento da retina, uma condição chamada RD proliferativa;
 - b) Quebra da barreira hemato-retiniana – gera um vazamento de pequenas e grandes moléculas para o espaço extravascular, causa edema macular diabético e danos aos fotorreceptores (BOSCO et al., 2005);
3. Toxoplasmose – o toxoplasma *gondii* é um parasita intracelular que infecta uma ampla gama de vertebrados de sangue quente, causando doenças em animais e humanos (WEBSTER, 2010), sendo transmitido principalmente pela ingestão de cistos contidos nos tecidos de um hospedeiro cronicamente infectado ou pela ingestão de oocistos contaminados, advindos de fezes, contidos nos alimentos ou na água (FRENKEL et al., 1990). Esse parasita é de origem alimentar nos Estados Unidos (MEAD et al., 1999) e frequentemente acomete a água em partes do Brasil (MOURA et al., 2006). A toxoplasmose é a doença que mais causa cegueira ocular no Brasil (ORÉFICE; FILHO; BARBOZA, 2010);
 4. Glaucoma – caracterizada por danos graduais ao nervo óptico e perda de campo visual, o glaucoma é um dos principais originadores da deficiência visual no mundo (TIELSCH et al., 1991). O diagnóstico precoce e o tratamento ideal demonstraram minimizar o risco de perda visual devido ao glaucoma (HEIJL et al., 2002). O glaucoma é obrigatoriamente uma neuropatia e não uma retinopatia, entretanto, a doença atua na retina, danificando as células ganglionares e seus axônios. O fator preponderante que ilustra o glaucoma é a escavação do disco óptico, explicitado como a manifestação visível da estrutura tridimensional da cabeça do nervo óptico;
 5. Degeneração macular relacionada à idade – é a avaria que ocorre na parte da retina chamada mácula, sendo a causa mais comum de perda da visão em pessoas acima de 50 anos (COLEMAN et al., 2008). As duas formas principais da degeneração macular são a seca e úmida. A primeira delas leva à perda gradual da acuidade visual, enquanto a segunda é a forma mais danosa, caracterizada pelo crescimento de uma estrutura coroidal vascular na mácula, acompanhada de aumento da permeabilidade vascular (BRESSLER; BRESSLER; FINE, 1988);
 6. Doenças cardiovasculares – são patologias observáveis na retina, porém, não afetam somente a retina e não são habitualmente originárias do olho. A hipertensão e a aterosclerose causam mudanças na razão entre o diâmetro das veias e artérias da retina, conhecida como proporção A/V. Uma diminuição na proporção A/V, ou seja, estreitamento das artérias e alargamento das veias, está associada a um risco aumentado de acidente vascular cerebral e infarto do miocárdio (WONG et al., 2004). Ainda no espectro de hipertensão, é possível que o paciente apresente

isquemia retiniana direta, resultando em infartos retinianos, análogos a manchas de algodão na visão e infartos coroidais visíveis. Além disso, a doença vascular sistêmica pode causar oclusões arteriais e venosas.

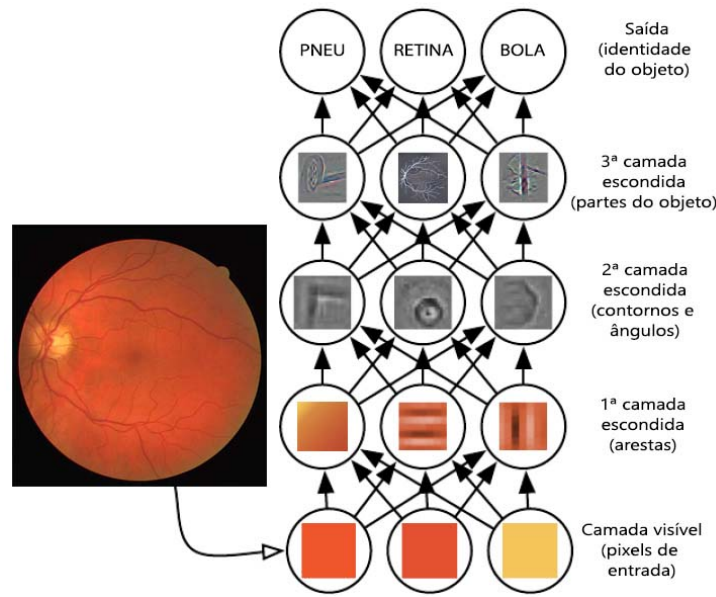
Atualmente, os sistemas de Apoio à Decisão Clínica (ADC) são baseados em algoritmos de aprendizado de máquina, mais especificamente em Redes Neurais Convolucionais (Convolutional Neural Networks (CNNs)). Assim, uma revisão sobre o uso de CNNs para classificar patologias observadas na retina foi publicada por Litjens et al. (2017). É importante mencionar que o uso de CNNs tem ganhado atenção devido à sua capacidade de encontrar padrões em grandes conjuntos de dados de imagens e também pela possibilidade de usar um modelo pré-treinado com uma estratégia de transferência de aprendizagem. Como demonstrado em Yu et al. (2017), transferir o conhecimento básico de um grande conjunto de dados, como a base ImageNet, para um domínio específico, permite reduzir o problema de ter pequenos conjuntos de dados, que são um obstáculo comum na área médica.

O motivo que institui a aprendizagem profunda como um dos mecanismos mais promissores atualmente em classificação de imagens, é que sua representação de conhecimento não precisa ser alusiva como para os humanos. Desta forma, a aprendizagem profunda permite que o computador crie conceitos complexos derivados de uma noção simples. A Figura 1, exemplifica que o sistema de aprendizagem profunda (DLS, do inglês *Deep Learning System*), pode retratar a imagem de uma retina com uma série de mapeamentos simples aninhados, cada um descrito por uma camada diferente do modelo. A entrada é apresentada na camada visível que contém as variáveis a serem observadas. Em seguida, uma série de camadas ocultas extrai características cada vez mais abstratas da imagem. Assim, o próprio modelo deve determinar quais características são úteis para explicar os relacionamentos entre os dados observados e as saídas.

Um DLS usa métodos baseados em técnicas de aprendizado de máquina e representação para processar grandes volumes de dados e extrair padrões significativos (LECUN; BENGIO; HINTON, 2015; LIM et al., 2014). Assim, é crível assumir que o desempenho de um DLS na identificação de uma patologia está vinculado diretamente à qualidade e quantidade dos dados. É importante ressaltar que em imagens retinianas existe a variação de etnias dentre os pacientes (isto é, com pigmentação variável), expressando alterações (por exemplo, devido à baixa dilatação da pupila, opacidade da mídia, baixo contraste ou foco) que podem afetar a acurácia na definição de padrões (WONG; BRESSLER, 2016; ABRÀMOFF et al., 2010). Os DLSs mostraram recentemente alta sensibilidade e especificidade (acima de 90%) na detecção de RD, catarata, glaucoma e Degeneração Macular Relacionado à Idade (DMRI) a partir de retinografias, provenientes de bancos de dados públicos ou privados (GARGEYA; LENG, 2017; GULSHAN et al., 2016; DONG et al., 2017; CHRISTOPHER et al., 2018).

Seguindo os conceitos apresentados, este trabalho propõe um conjunto de CNNs para

Figura 1 – Exemplo de mapeamento de uma imagem em um DLS.



Fonte: Adaptada de Goodfellow, Bengio e Courville (2016).

classificar condições e doenças oculares em imagens reais fornecidas por institutos de pesquisa brasileiros (o Instituto da Visão de São Paulo e a Universidade Federal de São Paulo). A abordagem proposta consiste no processamento de imagens do fundo ocular e na utilização de modelos CNN pré-treinados (arquitetura VGG16). Esses modelos empregam a estratégia de transferência de aprendizagem para melhor se adaptar às diferentes resoluções de imagens e condições oculares, que são: catarata (operável ou não), retinopatia diabética identificável (presente ou não), escavação (anormal ou não) e vasos sanguíneos (anormais ou não).

Diferente das abordagens encontradas na literatura, nesta tese foi proposta uma nova abordagem de transferência de aprendizagem, que considera o conhecimento adquirido por meio de imagens de alta resolução (obtidas por equipamentos caros e que não seriam acessíveis em todas as partes do país, principalmente em unidades básicas de saúde), permitindo o treinamento de modelos adequados para classificar imagens de baixa resolução produzidas por equipamentos de menor custo. Assim, essa estratégia de transferência de aprendizagem permite melhorar o atendimento do sistema público de saúde.

1.1 Justificativas

Considerando o aumento gradativo da expectativa de vida, as complicações oculares são prevalentes e, comumente, as doenças relacionadas à cegueira são assintomáticas. De acordo com a Organização Mundial da Saúde (OMS), até 1 bilhão de casos de deficiência visual poderiam ser evitados com tratamento adequado e exames regulares (WHO, 2021).

Um estudo de 2019 (WHO, 2019) estima que no final de 2020, no mundo, 80 milhões de pessoas tiveram a perda total da visão, sendo que quase 90% desses pacientes estão em países emergentes (BOURNE et al., 2021). Entre os países da América Latina, o Brasil apresenta baixo número de cegueira evitável segundo a OMS e hoje possui 0,75% da população com cegueira total, sendo predominante nas classes de menor renda (OTTAIANO et al., 2019).

O uso de ferramentas de diagnóstico automatizadas é desejado para o propósito de diminuir a subjetividade, incrementar a robustez e a consistência da análise. Imagens do fundo ocular emergiram como um dos principais meios de compreensão de patologias na retina, devido à fácil aquisição e nitidez representativa das estruturas retinianas (ZILLY; BUHMANN; MAHAPATRA, 2017). Desta forma, com a utilização dos DLSs é possível amparar o oftalmologista com apoio à classificação de patologias e anormalidades ou com a visualização de áreas de possível interesse por meio da geração do Class Activation Map (CAM).

A partir deste cenário, os sistemas de ADC prometem melhorar e agilizar os diagnósticos (BAKER, 2001), principalmente quando a quantidade de informação ultrapassar o limite da análise e percepção humana. Dessa forma, o ADC permite aos clínicos tomar decisões baseadas em evidências do histórico de diagnósticos e tratamentos registrados, atenuando as falhas clínicas, como sugerido pela *National Academy of Medicine* dos Estados Unidos (MCGLYNN et al., 2003). Além disso, as falhas no diagnóstico de patologias são agravadas na área da oftalmologia onde, em uma situação de emergência ou em um ambulatório comum, geralmente apenas um clínico geral e não um oftalmologista está presente (TORRE-DÍEZ et al., 2015).

Outro ponto a se destacar é que não há na literatura uma abordagem de transferência de conhecimento de imagens de alta qualidade para aquisição de um modelo de classificação de imagens de baixa qualidade. Desta forma, foi desenvolvida a presente tese, visto que os resultados obtidos dos modelos treinados diretamente em imagens de baixa qualidade, ou em conjunto de dados genéricos como o ImageNet não são satisfatórios quanto comparados aos resultados obtidos com a transferência de conhecimento. Deve-se ressaltar ainda que os resultados adquiridos em bases públicas corroboraram com o desenvolvimento da pesquisa, permitindo comparações mais adequadas às metodologias que compõem o estado da arte.

1.2 Objetivos

Nesta tese foi proposto o desenvolvimento de múltiplas CNNs para classificar e visualizar características de doenças oculares em imagens reais fornecidas pelo Instituto da Visão de São Paulo, em parceria com a Universidade Federal de São Paulo (UNIFESP). O principal objetivo consiste em processar imagens de baixa qualidade do fundo ocular

usando 4 CNNs, que foram treinadas em diferentes condições e com conhecimento transferido do treinamento de outras 4 CNNs que utilizaram imagens de alta qualidade. Além disso, outras 4 redes foram treinadas apenas com imagens de baixa qualidade a fim de compará-las ao método de transferência. A abordagem discutida visa atender às necessidades do cenário médico, oferecendo uma solução eficiente e prática para classificação e análise de várias condições oculares utilizando imagens adquiridas por equipamentos de menor custo.

Pensando em uma abordagem mais fidedigna ao cenário clínico real, os modelos supracitados devem ser capazes de processar uma imagem de fundo ocular e retornar a condição mais provável, juntamente com um CAM, através de um processo de classificação para apoiar a decisão diagnóstica. O foco em imagens de baixa qualidade e transferência de conhecimento visa contornar problemas como falta de padronização e escassez de dados, que são apontados por Ching et al. (2018) como as principais barreiras para o uso de ADC na medicina. A abordagem também realça a importância de trabalhar com vários tipos de patologias da retina para cobrir um amplo espectro de condições e doenças que podem ser encontradas no campo médico, tornando-a uma ferramenta valiosa para pesquisa e prática clínica.

As condições oculares abordadas foram: catarata (operável ou não), RD (presente ou não), escavação (anormal ou não), e vasos (anormais ou não). Além disso, foi realizada uma comparação com trabalhos consolidados na área a fim de validar a eficácia das redes, utilizando bases de dados públicas como EyePACS e Messidor-2 para RD, ODIR para catarata, e REFUGE para glaucoma. Os objetivos específicos são delineados a seguir, destacando a necessidade de considerar vários aspectos do projeto para garantir uma solução de ponta em termos de precisão e confiabilidade:

1. Desenvolver uma rotina de pré-processamento universal para padronizar dados de entrada a serem processados pelas CNNs, garantindo que as imagens sejam formatadas e dimensionadas de maneira adequada, independentemente das condições de entrada;
2. Propor novas codificações de imagens, otimizando os resultados das CNNs com filtros de coloração e brilho, aprimorando a capacidade das redes de identificar nuances específicas e características-chave nas imagens da retina;
3. Identificar padrões de RD, glaucoma (escavação acentuada do nervo óptico), catarata e deformidade nos vasos da retina, permitindo uma melhor compreensão das condições e características específicas que podem ser inferidas a partir das imagens analisadas;
4. Treinar CNNs para detectar e validar as doenças mencionadas, gerando CAMs das áreas de interesse e fornecendo informações visuais úteis para apoiar os profissio-

nais médicos na decisão de encaminhar pacientes para exames detalhados, assim, destacando a importância de proporcionar informações visuais complementares que podem melhorar a confiança no diagnóstico e nas decisões de tratamento;

5. Validar os modelos de CNNs utilizando dados públicos e compará-los com os artigos do estado da arte. Essa validação e comparação são cruciais para demonstrar a solidez da abordagem proposta e garantir a confiabilidade e relevância dos modelos desenvolvidos neste estudo. Além disso, contribui para o avanço do conhecimento na área de aplicação de CNNs no diagnóstico e análise de condições oculares;

Com base nessas propostas e nos resultados apresentados, acredita-se que esta pesquisa estabelece as bases para avanços significativos no diagnóstico e tratamento de doenças da retina. Ao combinar a eficácia das CNNs, juntamente com técnicas de transferência de aprendizado e uma abordagem escalável, esta tese mostra grande potencial para melhorar a qualidade de vida dos pacientes e auxiliar na prevenção de problemas de visão mais graves.

1.3 Organização da Tese

O capítulo 2 aborda uma revisão dos trabalhos antecedentes sobre o uso de algoritmos de aprendizado de máquina para classificar doenças relacionadas aos olhos. Na sequência, o capítulo 3 apresenta os princípios fundamentais que regem esta pesquisa, conceitos anatômicos da retina e os tipos de imagens que podem ser adquiridas. Para um melhor entendimento dos aspectos computacionais deste trabalho, o capítulo 4 aborda os conceitos-chaves de redes neurais, CNNs, aprendizado e Machine Learning Operations (MLOps). A metodologia proposta é descrita no capítulo 5, contendo a explicação do processo de coleta, tratamento, segmentação e padronização dos dados, tal como o desenvolvimento do conjunto de CNNs, as métricas e a visualização das áreas de interesse. O modelo de visão computacional proposto é apresentado de forma a destacar os aspectos técnicos que possibilitam a classificação e visualização de anomalias que incidem na retina humana. Os resultados até então alcançados e discussões são apresentados no capítulo 6, somado à explicação de como o método proposto foi implementado em uma aplicação web. Por fim, o capítulo 7 apresenta a conclusão da tese.

Capítulo 2

Revisão da Literatura

A aquisição de imagens médicas com a finalidade de detecção precoce de patologias, diagnósticos, estagiamento e acompanhamento de doenças tem aumentado nas últimas décadas. Alguns exemplos desses tipos de imagens são, tomografia computadorizada, ressonância magnética, tomografia por emissão de pósitrons, mamografia, ultrassom e raio-X. Entretanto, é notório que variações nas patologias, limitação humana na detecção de padrões e análise de grandes quantidades de imagens, faz com que seja necessária a utilização de sistemas computacionais que auxiliem nesse processo, os quais atualmente têm se baseado em técnicas de aprendizado de máquina, mais especificamente em aprendizado profundo (*deep learning*) (CHAN et al., 2020).

O emprego de algoritmos de aprendizado profundo em imagens médicas se dá pela evolução de hardware (como as *Graphical Processing Units* – GPUs) e pelo aumento na disponibilidade de dados. Dentre os principais escopos de aplicação, destacam-se as tarefas de segmentação de imagens (ZHANG et al., 2015; KLEESIEK et al., 2016), detecção de lesões e pontos de referência (PEREIRA et al., 2016; TULDER; BRUIJNE, 2016) e classificação de patologias (ROUX et al., 2013; SU et al., 2017).

Considerando as possíveis tarefas supracitadas, nessa pesquisa será enfatizada a classificação e visualização de retinopatias, as quais têm sua literatura correlata analisadas na sequência.

2.1 Classificação de Retinopatias

A demanda de diagnóstico por imagem no campo da oftalmologia tem aumentado consideravelmente. Todavia, a disponibilidade de profissionais para cumprir tal tarefa tem se tornado escassa. Algoritmos de visão computacional têm se mostrado uma ferramenta

eficiente na classificação e visualização de imagens oculares, sejam elas de duas ou três dimensões (FAUW et al., 2018).

2.1.1 Catarata

A catarata é uma doença ocular comum, responsável por causar distorção da visão e cegueira se não operada. Dessa forma, algumas pesquisas foram propostas para determinar a presença de catarata e seu grau. Gao, Lin e Wong (2015) apresentaram um sistema capaz de aprender automaticamente características que servem de auxílio na avaliação da gravidade da catarata, a partir de imagens do tipo *slit-lamp*. O grau da catarata de cada paciente em um conjunto de dados de 5378 imagens é concebido por meio de uma mescla de filtros de classes aplicados em uma CNN, seguido de um conjunto de redes neurais recursivas, a fim de extrair mais padrões. Ao final do processo, uma *Support Vector Regression* classifica a imagem. As métricas e os resultados foram, um erro absoluto médio (ϵ) de 0.304, uma razão de concordância integral exata (R_0) de 70.7%, um erro de classificação decimal de $88.4\% \leq 0.5$ ($R_{e0.5}$), e um erro de classificação decimal de $99.0\% \leq 1.0$ ($R_{e1.0}$).

Outro procedimento de aprendizado de características foi proposto por Guo et al. (2015), onde uma transformada Wavelet e um método de esboço foram empregados para extrair características de imagens do fundo ocular. Para identificar a catarata (presença ou ausência) e classificá-la (leve, moderada ou grave), foi utilizado um discriminador de rótulos, que conseguiu atingir uma acurácia máxima de 90,9% em um conjunto de dados com apenas 445 imagens.

Yang et al. (2016) extraíram características usando os métodos de transformada Wavelet, esboço e textura. A partir dessas características, um conjunto baseado em Support Vector Machines (SVM) e Multilayer Perceptron (MLP) foi usado para classificar a catarata. A acurácia média obtida foi de cerca de 93,2% para identificar a doença e de 84,5% para classificá-la.

Uma CNN pré-treinada com estratégia de transferência de aprendizagem foi considerada por Pratap e Kokil (2019) para extrair características. Na sequência, uma SVM realiza a classificação da catarata. Os autores empregaram conjuntos de dados públicos e a tarefa de rotulação foi apoiada por oftalmologistas. No entanto, essa curadoria de dados não foi explicada em detalhes. A abordagem proposta foi capaz de atingir uma acurácia média de 92,91%.

Um novo modelo de aprendizado profundo, denominado CataractNet, foi proposto por Junayed et al. (2021) para identificar a catarata em um conjunto de dados composto por 1130 imagens do fundo ocular (aumentadas para 4746 imagens). O modelo obteve acurácias superiores a 98,62%, sendo o melhor resultado da literatura. No entanto, é importante mencionar que, embora os autores façam uma análise comparativa, ela não pode ser considerada adequada, uma vez que os dados não são os mesmos.

Vale ressaltar que especificamente na identificação e classificação da catarata, ter um *benchmark* não é uma tarefa trivial dada a necessidade de um conjunto de dados público devidamente curado. Mesmo assim, espera-se que os modelos preditivos sejam capazes de atingir acurácias acima de 90%, como também mostrado em Guo et al. (2015), Yang et al. (2016) e Pratap e Kokil (2019).

2.1.2 Retinopatia diabética

A RD afeta os vasos sanguíneos da retina de indivíduos e, por esse motivo, algumas pesquisas são direcionadas para classificar seu grau. No entanto, a identificação de RDs referíveis (presença ou ausência) é de grande importância, pois sua detecção em estágio inicial permite a tomada de medidas preventivas.

Ananthapadmanaban e Parthiban (2014) propuseram uma comparação entre os classificadores Naïve Bayes e SVM para a classificação de RD. O conjunto de dados utilizado é notavelmente pequeno, apenas 300 imagens, e os resultados obtidos são baseados na métrica de acurácia, que neste caso foi de 83,37% para o Naïve Bayes e 64,91% para SVM.

Um dos estudos mais citados da área treinou uma Inception-v3 com dados da base EyePACS e de mais três hospitais indianos (GULSHAN et al., 2016). O modelo foi validado nas bases de dados EyePACS e Messidor-2 que foram rotuladas por 54 oftalmologistas selecionados. Os pontos operacionais foram escolhidos por alta sensibilidade e especificidade. Para alta especificidade, o resultado foi de 90.3% e 87.0% de sensibilidade e de 98.1% e 98.5% de especificidade, para a base de testes EyePACS-1 e Messidor-2, enquanto para o ponto operacional de alta sensibilidade a pesquisa obteve 97.5% e 96.1% de sensibilidade, e 93.4% e 93.9% de especificidade.

Contemplando apenas bancos de dados não comumente observados, Ting et al. (2017) propuseram um DLS para identificar e classificar RD e doenças oculares usando imagens da retina de populações multiétnicas. O trabalho foi validado com um conjunto de dados contendo 494661 imagens de 14880 pacientes, em que o DLS obteve uma sensibilidade de 90,5% e especificidade de 91,6% para detectar RD referida, sensibilidade de 100% e especificidade de 91,1% para possível ameaça de perda de visão por RD.

Outra abordagem baseada em CNN do tipo Inception-v3 foi proposta por Voets, Møllersen e Bongo (2019). Nesta pesquisa os autores também usaram os conjuntos de dados EyePACS e Messidor-2. Eles demonstraram que, sem curadoria de dados, não foi possível obter os mesmos resultados na classificação que o trabalho de Gulshan et al. (2016). O ponto operacional para alta especificidade resultou em 83,6% de especificidade e 68,7% de sensibilidade para o conjunto EyePACS e 92,0% e 88,5% para a base Messidor-2. Considerando o ponto operacional para alta sensibilidade, obteve-se sensibilidade de 90,6% e 81,8% de especificidade no EyePACS e 84,7% e 71,2% no Messidor-2.

Uma ferramenta de visualização e modelos de aprendizado profundo foram propostos por Keel et al. (2019) para apoiar decisões em termos de RD referencial. Os modelos foram treinados e validados usando 48116 imagens de retina. Para visualizar as características significativas para os médicos, os autores aplicaram uma técnica de *kernel* adaptável independente da CNN com base em uma janela deslizante para cortar as imagens e produzir um mapa de características. Bons resultados foram obtidos (considerando um intervalo de confiança de 95%), chegando a 96% de casos verdadeiros positivos.

A fim de verificar a precisão dos diferentes modelos de CNN combinados, a pesquisa de Qummar et al. (2019) usou o conjunto de dados EyePACS disponível publicamente e as imagens foram aplicadas aos seguintes modelos: Resnet50, Inception-v3, Xception, Dense121 e Dense169. O trabalho se destacou em classificar as 5 etapas da RD, mensuradas pela acurácia, *recall*, especificidade, precisão e *f1-score*. Nessas métricas, os resultados foram, respectivamente, 80,8%, 51,5%, 86,7%, 63,8% e 53,7%.

Xie et al. (2020) propuseram duas CNNs do tipo InceptionResNet-v2, sendo uma delas um classificador binário entre retina saudável e patológica, enquanto o segundo é um classificador ternário entre retina saudável, RD não referível e RD referível. Os destaques são o pré-processamento com filtro gaussiano, a escolha de um limiar de classificação em cascata e uma técnica que também aumenta a sensibilidade chamada de *margin max*. As métricas são focadas na sensibilidade, cujos valores atingiram cerca de 80% nos casos da retina patológica e próximos a 98% para retinas saudáveis.

Visando reduzir a complexidade de uma VGG16 e aumentar seu desempenho, Zago et al. (2020) desenvolveram um modelo de localização de lesão usando um DLS com uma abordagem *patch-based*. Sete conjuntos de dados diferentes foram usados, dentre eles destacam-se o Eyepacs e Messidor-2. O sistema foi capaz de obter 0,82 e 0,944 de AUC (*area under the curve*), respectivamente, e 88% e 98% de sensibilidade com valor fixo de 50% para especificidade.

Dos estudos supramencionados, nota-se também uma dificuldade de estabelecer comparações adequadas, uma vez que utilizam diferentes conjuntos de dados, não detalham a curadoria de dados e os processos de rotulação.

2.1.3 Escavação do Nervo Óptico

Outra forma de diagnóstico e estagiamento de patologias oculares é a análise do nervo óptico. Tipicamente, este segmento de estudo tem como objetivo validar se o diâmetro do disco óptico, também conhecido por escavação, está maior que o normal, denotando assim uma possível enfermidade.

No primeiro estudo de detecção de glaucoma relatado na literatura usando CNNs, Chen et al. (2015) desenvolveram uma arquitetura composta por seis camadas. As quatro primeiras são convolucionais e as duas restantes estão totalmente conectadas. Além disso, a saída da última camada totalmente conectada foi fornecida a um classificador *softmax*

treinado no domínio da patologia estudada. Imagens dos conjuntos de dados ORIGA (650 imagens) e SCES (1722 imagens) foram utilizadas, processadas e avaliadas através da métrica AUC, cujos resultados nas bases de dados foram 82,3% e 86%, respectivamente.

Como mencionado anteriormente, o trabalho de Ting et al. (2017) contemplou outras patologias, uma delas é o glaucoma. Usando a mesma arquitetura que a descrita para a classificação de RD referível, os autores obtiveram a sensibilidade de 96,4% e especificidade de 87,2% para o contexto de glaucoma em um conjunto de dados proprietário.

Com a utilização de dois estágios, a pesquisa de Bajwa et al. (2019) primeiro localizou e extraiu o nervo óptico da retina com uma *Regions Convolutional Neural Network* (RCNN) para posteriormente classificar essa imagem extraída entre saudável ou com glaucoma por um modelo de *Deep Learning*. Para servir de comparação, o método foi validado na base de dados ORIGA, atingindo uma AUC igual a 0.874.

Focando apenas na doença de glaucoma, o estudo de Hemelings et al. (2020) apresentou uma CNN que usa transferência de conhecimento dos pesos da base ImageNet para classificar imagens de fundo ocular, tendo como premissa o fato de que um número limitado de dados rotulados para treinamento pode gerar resultados de predição viáveis. As principais características deste estudo são uma combinação de transferência de conhecimento, *data augmentation* e amostragem de incerteza. Os autores usaram 2072 imagens coloridas do fundo ocular para o treinamento, representando 42% dos dados originais. O DLS alcançou uma AUC de 0,995, sensibilidade e especificidade de respectivamente 98,0% (CI 95,5% - 99,4%) e 91% (CI 84,0% - 96,0%) para pacientes com e sem glaucoma.

Em uma abordagem simultânea, Hervella et al. (2022) segmentaram a região do nervo óptico e na mesma Deep Neural Network (DNN) classificaram a imagem entre patológica (glaucoma) ou saudável. O propósito deste modelo foi o de se beneficiar das características a nível de pixel adquiridas pela segmentação e as características a nível de imagem adquiridas pela imagem no contexto de classificação. Dois conjuntos públicos de dados foram utilizados para validação, REFUGE e DRISHTI-GS, cujos resultados na métrica AUC para as respectivas bases foram de 97.60% e 94.74%.

2.1.4 Vasos Sanguíneos

A análise de vasos sanguíneos em imagens de fundo de olho é de suma importância para auxiliar na identificação de doenças relacionadas aos olhos e também doenças sistêmicas (por exemplo, hipertensão e acidente vascular cerebral isquêmico). Destaca-se um volume acentuado de trabalhos de segmentação da estrutura vascular dos olhos, uma vez que é comumente usado como procedimento precedente à classificação (ZHANG et al., 2019).

Um método de k-Nearest Neighbors (kNN) para classificação de artérias ou veias foi projetado por Xu et al. (2017) tendo como entrada imagens previamente segmentadas por um conjunto de procedimentos categorizados, o primeiro para identificar o perfil estrutural da veia ou artéria, seguido por identificação de coloração e texturas. A base de dados

escolhida foi a DRIVE, sendo a acurácia a métrica de validação, a qual obteve um valor de 92.3%

No enquadramento de resolução de glaucoma usando segmentação das vasos sanguíneos, Deepika e Maheswari (2018) elaboraram um método de classificação da severidade da patologia que tem como primeiro estágio uma extração de características dos vasos. As características adquiridas serviram de entrada para um modelo híbrido composto por um *Adaptive Neuro Fuzzy Inference System* (ANFIS) e uma SVM. Para validar o modelo foram utilizadas as métricas de sensibilidade, especificidade e acurácia, cujos resultados foram, respectivamente, 85,7%, 92% e 97.77%. Entretanto, as informações acerca do conjunto de dados são escassas.

Objetivando a classificação de condições oculares não específicas que podem causar cegueira, o trabalho de Maji e Sekh (2020) expõe uma CNN do tipo *Inception V3* que possui em sua composição uma rede de atenção indicando os pixels mais propensos à representar uma informação discriminante. A rotulação usada para a saída da rede é subdividida em 3 possibilidades, grau 2, grau 1 e grau 0, no qual o grau 0 representa maior tortuosidade dos vasos oculares, conseqüentemente maior severidade. O conjunto de dados empregado para o modelo foi o EIARG1, com 120 imagens, instituído com a premissa de denotar a gravidade das anomalias oculares. Todavia, por ser uma abordagem atípica, teve como objeto de comparação apenas os modelos de rede neural criados pelos mesmos autores, cujo melhor resultado atingido foi o de *f1-score* igual a 0.92 pela *InceptionV3*.

Dispondo de um procedimento que não requer treinamento de modelo, Remeseiro, Mendonça e Campilho (2021) propuseram um método de classificação de artérias e veias que usufrui apenas do contraste local entre os vasos sanguíneos e o fundo aos arredores, baseado na projeção de grafos análogos à estrutura vascular da retina, advindos de um mecanismo prévio de segmentação. O grafo gerado é submetido à uma aplicação de limiar multinível para a classificação preliminar e posteriormente analisado para o cálculo da relação arteríola-vénula (AVR, do inglês *arteriolar-to-venular ratio*). Por fim, o trabalho foi aplicado às bases de dados INSPIRE e DRIVE, obtendo bons resultados em índices de acurácia ao identificar os vasos e alcançando valores de 96% e 80% nas respectivas bases.

Capítulo 3

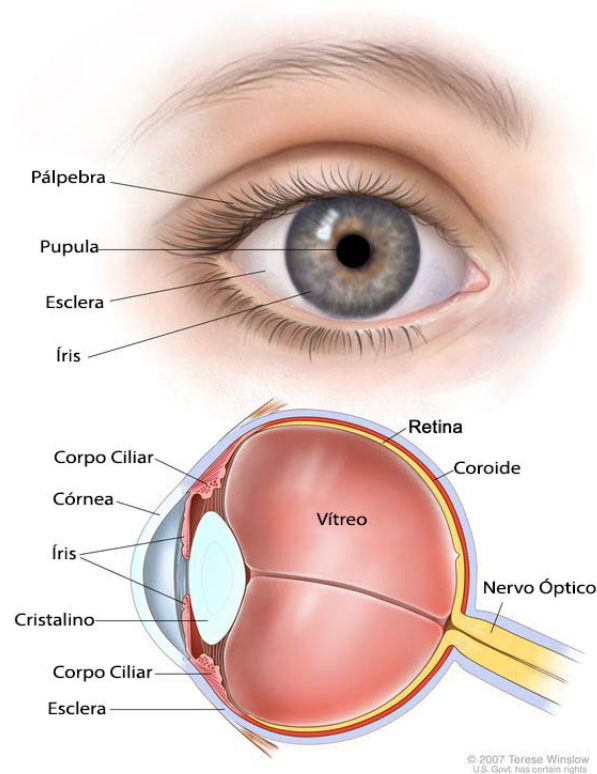
Conceitos Gerais da Anatomia Ocular e Fundoscopia

Este capítulo discorre sobre os aspectos gerais da anatomia ocular que norteiam a identificação de padrões e, por conseguinte, tal entendimento auxilia na concepção de modelos de inteligência artificial. É abordado também o método de aquisição de imagem do fundo ocular 2D, denominado por fundoscopia.

3.1 Anatomia Ocular

O olho é dividido em dois polos, o anterior e o posterior, sendo que a parte anterior está à frente do cristalino e a parte posterior se encontra na parte de trás, conforme mostrado na Figura 2. O polo anterior é o terço frontal do olho e inclui todas as estruturas na frente do vítreo: córnea, íris, corpo ciliar e cristalino. As duas câmaras que estão contidas dentro do segmento anterior são o polo anterior, que é o espaço entre a córnea posterior (endotélio) e a íris (preenchida com um fluído aquoso). O polo posterior também contém um fluído aquoso que, assim como o do polo anterior, fornece nutrientes para as estruturas circundantes. O segmento posterior representa os dois terços posteriores do olho e inclui o vítreo transparente e gelatinoso, a membrana hialóide anterior do vítreo, a retina (local onde a mácula está presente), a coroide e o nervo óptico (LENS; NEMETH; LEDFORD, 2008).

Figura 2 – Anatomia ocular.



Fonte: Adaptado de <https://cancer.gov>. Acesso em: 02/06/2021

3.1.1 Mácula e Nervos Ópticos

Além de serem estruturas cruciais na constituição da visão humana, a mácula e a cavidade do nervo óptico auxiliam na distinção entre olho direito e esquerdo, dado que a mácula localiza-se na região temporal da retina e, em contrapartida, a cavidade do nervo óptico mantém-se na faixa nasal (SNELL; LEMP, 2013). A mácula é o principal local da visão humana mesmo sendo o ponto mais fino da retina. Ela é formada por cones, onde em sua região central encontra-se a fóvea (PROVIS et al., 2005). Os cones que compõem a mácula são os neurônios responsáveis pela visão de cores e pelo foco, ao passo que fora da mácula, na periferia da retina, tem-se os bastonetes, que são os neurônios responsáveis pela visão contrastada (OYSTER, 1999).

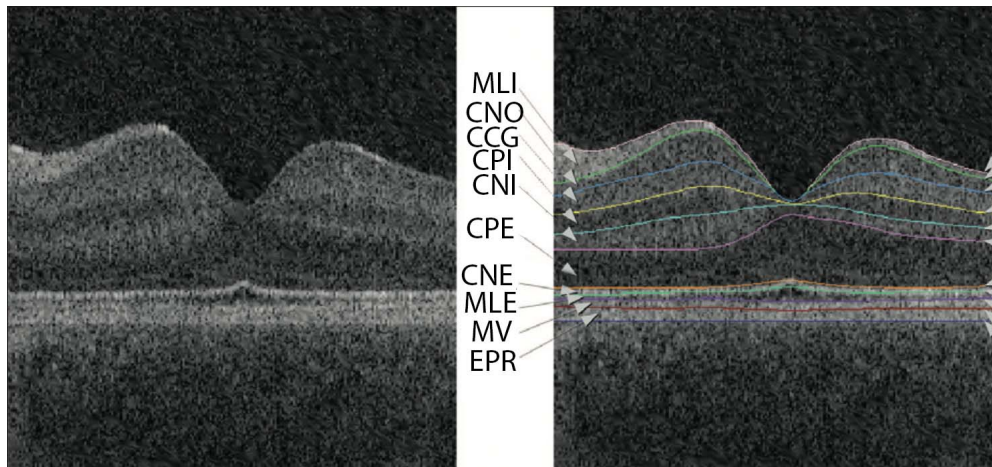
A cavidade do nervo óptico condensa as fibras que transportam as informações adquiridas pelas composições da retina, formando o quiasma óptico. As informações são processadas através de sinapses que ocorrem no corpo geniculado lateral, onde existe a integração com outras vias (auditivas, sensitivas e de equilíbrio) para que posteriormente haja a sinapse no córtex occipital (JONAS et al., 1992).

3.1.2 Camadas da Retina

A retina é o tecido nervoso situado no fundo dos olhos que tem como funcionalidade captar a luz incidida através da pupila e do cristalino, dado que suas diferentes células contidas neste tecido transformam a luminosidade em sinais elétricos para o cérebro (KOLB, 2003). Possui 10 camadas compostas de neurônios (Figura 3), sendo que algumas camadas não detêm um neurônio completo. Em ordem de mais próxima à coroide para mais próxima ao vítreo, tem-se (GUPTA et al., 2015; KOLB, 2012):

- ❑ Camada 10 – chamada de EPR ou membrana de Bruch, encontra-se em contato com a coroide. É uma monocamada e as células contidas nessa camada são cuboides estreitos e altos. No ápice do EPR há a construção da barreira hematorretiniana. A função dessa camada é o estoque de vitamina A;
- ❑ Camada 9 – composta de cones e bastonetes que preenchem os segmentos internos e externos de sua faixa, também chamados de fotorreceptores. É envolta de um revestimento mucoproteico, também conhecida como membrana de Verhoeff (MV);
- ❑ Camada 8 – membrana limitante externa (MLE) que serve para interagir a nona com a sétima camada, onde ocorre a comunicação intercelular entre os cones e os bastonetes com as células de Muller;
- ❑ Camada 7 – camada nuclear externa (CNE) que condensa os corpos celulares e onde estão os núcleos dos cones e bastonetes;
- ❑ Camada 6 – camada plexiforme externa (CPE) que abrange os axônios dos cones e dos bastonetes que fazem sinapse com os dendritos das células bipolares e horizontais;
- ❑ Camada 5 – camada nuclear interna (CNI) que contém os núcleos das células bipolares além das células horizontais, amácrinas e de Muller;
- ❑ Camada 4 – camada plexiforme interna (CPI) onde estão os axônios das células bipolares e amácrinas, fazendo sinapse com os dendritos das células ganglionares;
- ❑ Camada 3 – camada de células ganglionares (CCG) com seus respectivos núcleos e os processos das células de Muller;
- ❑ Camada 2 – responsável pela formação das fibras nervosas. Agrega os axônios das células ganglionares para formar o nervo óptico, conhecida como camada do nervo óptico (CNO);
- ❑ Camada 1 – conhecida como membrana limitante interna (MLI), possui os processos das células de Muller que estão em contato inerente com o vítreo.

Figura 3 – Camadas da retina.



Adaptado de: Abràmoff et al. (2010)

3.2 Aquisição de Imagens Oculares

Matsui et al. (1973) foram os primeiros a publicar um método para análise de imagens da retina, tendo como foco a segmentação de vasos, baseada na morfologia matemática com o auxílio de lâminas digitalizadas de angiogramas de fluoresceína da retina. Nos anos subsequentes, houveram novas tentativas de segmentar outras estruturas anatômicas do olho humano com o uso de lâminas digitalizadas. Relatado em 1984, o primeiro método para detectar e segmentar estruturas anormais do olho foi produzido por Baudoin, Lay e Klein (1984), cujo trabalho descreve o mecanismo adotado para detectar microaneurismas, uma lesão que habitualmente decorre da RD. Essa abordagem, do mesmo modo da precursora, utiliza imagens angiográficas digitalizadas, entretanto aplica um filtro do tipo *top-hat* para a detecção dos microaneurismas. Todo o processo é caracterizado como uma técnica de morfologia matemática que elimina a vasculatura de uma imagem de fundo, mas deixa intocados possíveis candidatos a microaneurisma. Após condução desses trabalhos, o campo de pesquisa mudou drasticamente nos anos 90 com o desenvolvimento da imagem digital da retina e a expansão das técnicas de análise de imagem com base em filtro digital.

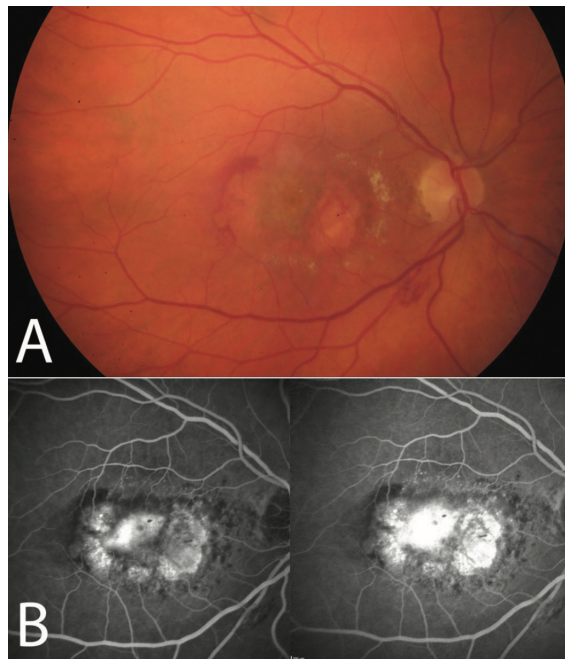
Desde a década de 60 o processo de aquisição de imagens da retina tem se desenvolvido (YANNUZZI et al., 2004), sendo que atualmente é crucial a inferência por imagem no tratamento clínico e estagiamento de doenças que incidem na retina (HASAN et al., 2021). Analisar imagens do fundo ocular é de suma importância para detectar RD, escavação anormal do nervo óptico (glaucoma), catarata e outras patologias derivadas de anomalias na estrutura ocular.

3.2.1 Imagens do Fundo Ocular

Existem inúmeros tipos de imagens do fundo ocular, como por exemplo, tomografia computadorizada, imagem por ressonância magnética, imagem de ultrassom, termografia infravermelha (ACHARYA et al., 2008), imagem hiperespectral (ALABBOUD et al., 2007), imagem Doppler colorida (DIMITROVA; KATO, 2010), oftalmoscopia fotoacústica (JIAO et al., 2010) e imagem de ressonância magnética de fluxo sanguíneo (LI; CHENG; DUONG, 2008). Ademais, os tipos de imagens de duas dimensões mais comumente usados são: *red-free*, fundo ocular colorida, *stereofundus*, oftalmoscopia a laser e angiografia (BERNARDES; SERRANHO; LOBO, 2011).

Para elucidar duas das principais imagens de fundo ocular de duas dimensões, o trabalho de Mokwa et al. (2013) apresenta uma neovascularização coroidal e hemorragia do tecido fibrovascular (Figura 4).

Figura 4 – Comparação entre imagem do fundo ocular colorida (A) com hemorragia e tecido fibrovascular e imagem de angiografia (B) de um vazamento de neovascularização coroidal.



Fonte: Adaptado de Mokwa et al. (2013).

Imagens como essas permitem a visualização da microcirculação coroidal e retinal e fornecem informações detalhadas sobre a presença de vasos patológicos, bem como a integridade da barreira retiniana do sangue. Abramoff, Garvin e Sonka (2010) definem os dois tipos, conforme segue:

- Imagem colorida do fundo ocular – as intensidades da imagem representam a quantidade de bandas de onda refletidas em vermelho, verde e azul (RGB - do inglês *red, green and blue*), conforme determinado pela sensibilidade espectral do sensor;

- ❑ Angiografia de fluoresceína e angiografia de indocianina – as intensidades de imagem representam a quantidade de fótons emitidos pela fluoresceína ou fluoróforo verde de indocianina que foi injetada na circulação do paciente.

No contexto da captura de imagens 2D, existem os dispositivos estacionados, como mostrado na Figura 5, que geralmente contam com um apoio de queixo e testa para o paciente, e os portáteis (ilustrado na Figura 6.), que permitem a captura da retina com mais facilidade. Deve-se ressaltar que o avanço tecnológico atual confere aos aparelhos portáteis a mesma qualidade nas imagens das que são geradas em aparelhos estacionados. Nessas imagens capturadas, a retina pode ser fotografada diretamente, pois a pupila (sem precisar dilatar) é usada como entrada e saída de iluminação emitida pelo dispositivo (SAINÉ; TYLER, 2002).

Figura 5 – Exemplo de um dispositivo de captura chamado nonmyd WX.



Fonte: Adaptado de Murata et al. (2020)

Mesmo com aparelhos modernos e que capturam imagens de alta resolução é possível registrar imagens de baixa qualidade, sendo que, assim como Zapata et al. (2020) definiram, uma imagem de boa qualidade é uma imagem que prima estar com a mácula centralizada; com o foco correto; boa visualização dos vasos parafoveais; poder visualizar mais de dois diâmetros de disco ao redor da fóvea; e visualizar bem o disco óptico, como no exemplo da primeira linha de imagens da Figura 7.

3.2.2 Tomografia de Coerência Óptica

Na esfera de imagens 3D, o método mais utilizado em pesquisas de identificação de patologias na retina é o OCT. Este corresponde a um recurso recente que consiste na

Figura 6 – Exemplo de um aparelho portátil chamado Picton e sua respectiva imagem gerada colorida e em tons de cinza.



Fonte: Adaptado de Zhang et al. (2017)

projeção de cortes transversais da retina de forma não invasiva, a partir de ondas de luz de baixa coerência. Este procedimento resulta em representações da retina em até três dimensões, propiciando assim uma visibilidade mais apurada no volume das camadas da retina, dado que as espessuras das referidas camadas são essenciais na categorização de uma retina saudável ou enferma (SPAIDE; KLANCNIK; COONEY, 2015).

Essa modalidade de imagem médica tem apresentado um desenvolvimento acentuado na última década, visto que o OCT pode capturar diferentes aspectos de tecidos biológicos, como fluxo sanguíneo, estado de polarização, dados de estrutura, parâmetros elásticos e conteúdo molecular (HUANG et al., 1991). O uso de OCT pode auxiliar no conhecimento preciso da dispersão óptica e absorção de tecidos biológicos (GABBAY; SIVARAJAH, 2008; DREXLER; FUJIMOTO, 2008). Portanto, o OCT pode ser usado para diagnosticar diferentes doenças, analisando a mudança nos tecidos humanos nas imagens capturadas (ELADAWI et al., 2020), como visto na Figura 8.

Os primeiros instrumentos de OCT, conhecidos como *time-domain* OCT (TD-OCT), usavam um único fotodetector com um *A-scan*, cuja funcionalidade era mover um espelho que mudava o caminho óptico do feixe de referência para corresponder a diferentes profundidades axiais no tecido alvo. Essa configuração limitava a velocidade de varredura a alguns milhares de *A-scans* por segundo. Uma técnica mais recente, conhecida como *spectral-domain* OCT (SD-OCT) é capaz de adquirir um *A-scan* inteiro usando uma série de detectores em vez de usar vários feixes de referência de um espelho em movimento. As

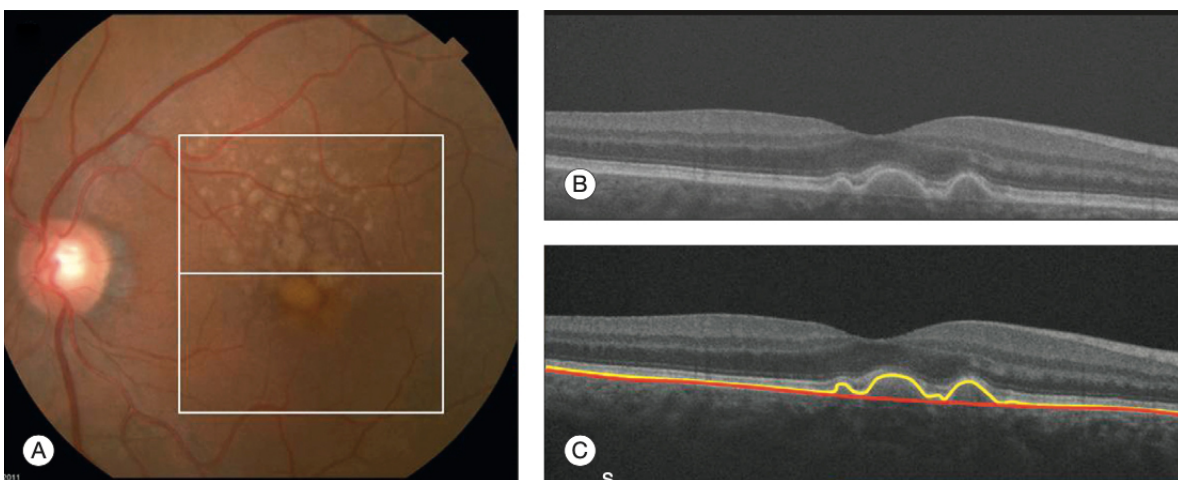
Figura 7 – (A) Exemplo de imagens de alta qualidade. (B) Exemplos de imagens de baixa qualidade.



Fonte: Adaptado de Zapata et al. (2020)

velocidades de digitalização com instrumentos SD-OCT podem exceder 100.000 *A-scans* por segundo, cerca de 200 vezes mais rápido que o TD-OCT (de Amorim Garcia Filho et al., 2012; LAMIREL; NEWMAN; BIOUSSE, 2010).

Figura 8 – Exemplo de OCT: (A) imagem colorida do fundo ocular de um paciente com drusas. Um recorte delimitado em branco (6×6 mm) foi sobreposto na imagem para representar a área de digitalização; (B) SD-OCT correspondente à linha central do recorte; e (C) segmentação do EPR e uma linha vermelha mostrando sua base.



Fonte: Adaptado de Lamirel, Newman e Biousse (2010).

Capítulo 4

Aprendizado Profundo e Integração MLOps

Neste capítulo, os avanços da visão computacional serão explorados através da perspectiva de *Deep Learning* com foco em CNN. Além disso, serão examinados os conceitos de aprendizado supervisionado e não supervisionado, bem como as implicações práticas da integração dessas técnicas com MLOps para otimizar o processo de implantação e monitoramento de modelos.

Ao longo deste capítulo, serão destacados *insights* críticos e aplicações práticas, visando fornecer uma compreensão abrangente de como as técnicas avançadas podem ser utilizadas no contexto mais amplo de pesquisa e desenvolvimento relacionados à visão computacional.

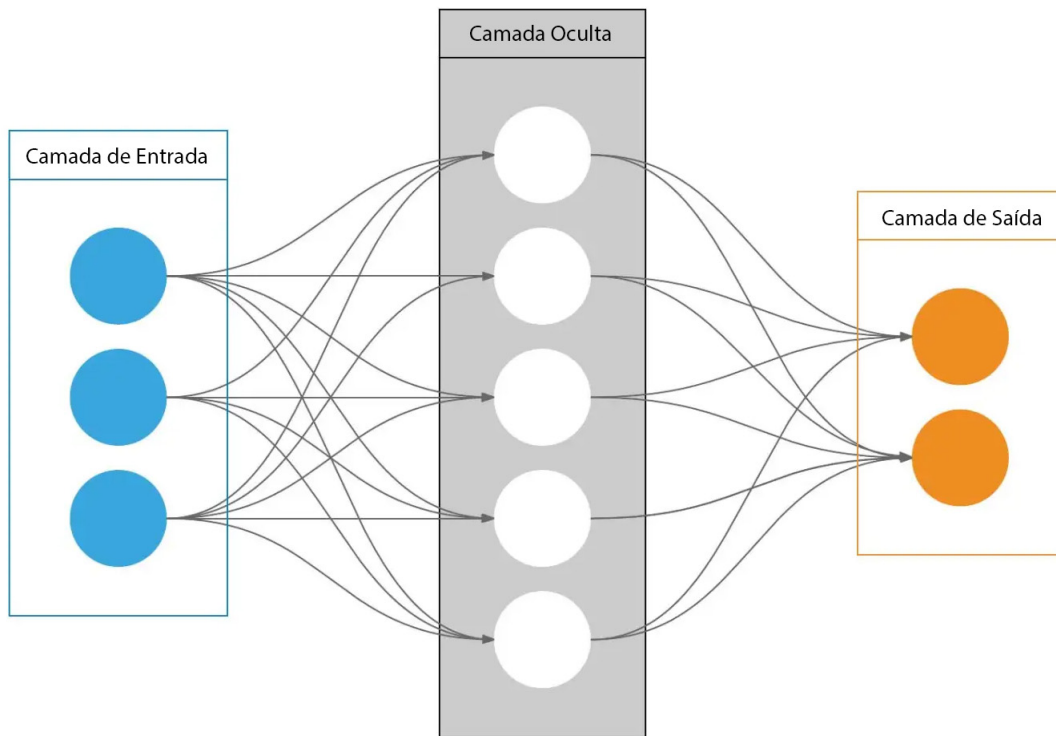
4.1 Aprendizado Profundo e Redes Neurais

O aprendizado profundo é um subcampo da aprendizagem de máquina que se concentra no projeto e treinamento de Rede Neural Artificiais (RNAs) para resolver problemas complexos. Ele ganhou popularidade e sucesso significativos nos últimos anos devido aos avanços no poder de computação, capacidades de armazenamento de dados e disponibilidade de grandes conjuntos de dados (LECUN; BENGIO; HINTON, 2015). As RNAs são inspiradas na estrutura e função das redes neurais biológicas presentes no cérebro humano. Essas redes consistem em camadas interconectadas de neurônios, que podem aprender a representar características hierárquicas dos dados de entrada por meio de algoritmos de aprendizado (GOODFELLOW; BENGIO; COURVILLE, 2016). O desenvolvimento de técnicas de aprendizado profundo levou a avanços em vários campos, incluindo visão com-

putacional, processamento de linguagem natural e reconhecimento de fala (HINTON et al., 2012; MIKOLOV et al., 2013; GRAVES; MOHAMED; HINTON, 2013).

Uma RNA consiste em nós ou neurônios conectados em várias camadas (Figura 9). Existem três tipos principais de camadas: camada de entrada, camadas ocultas e camada de saída. A camada de entrada recebe os dados, as camadas ocultas processam os dados e a camada de saída produz os resultados. Cada neurônio dentro de uma camada recebe informações dos neurônios da camada anterior e as alimenta nos neurônios da próxima camada por meio de conexões ponderadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Essa hierarquia de camadas permite aprender representações de alto nível a partir de dados brutos de entrada.

Figura 9 – Exemplo de uma rede neural artificial.



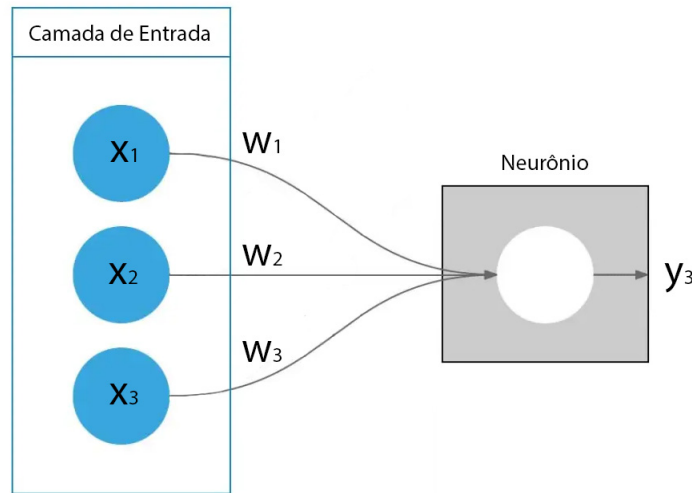
A operação de um único neurônio (Figura 10) pode ser descrita por um modelo matemático, que envolve uma soma ponderada de entradas, um termo de viés e uma função de ativação (GOODFELLOW; BENGIO; COURVILLE, 2016):

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

onde y é a saída do neurônio, f é a função de ativação, w_i são os pesos, x_i são as entradas e b é o termo de viés.

Treinar uma rede neural envolve ajustar os pesos e vieses para minimizar uma função de perda, que mede a diferença entre as saídas previstas e as saídas verdadeiras. Isso geralmente é alcançado por meio de um algoritmo de otimização, como o gradiente descendente estocástico (Stochastic Gradient Descent (SGD)) ou suas variantes, combinado

Figura 10 – Representação de um único neurônio.



com um algoritmo de retropropagação para calcular os gradientes da função de perda em relação aos pesos e vieses (RUMELHART; HINTON; WILLIAMS, 1986).

4.2 Convolutional Neural Networks (CNN)

CNNs são uma classe especial de redes neurais artificiais alimentadas no estilo *feed-forward* que se mostraram eficazes em uma variedade de tarefas de visão computacional, como classificação de imagens, detecção de objetos e segmentação semântica (LECUN; BENGIO; HINTON, 2015; KRIZHEVSKY; SUTSKEVER; HINTON, 2012). A principal diferença entre as CNNs e as redes neurais tradicionais está em sua arquitetura, que é projetada especificamente para lidar com dados em grade, como imagens.

Uma CNN típica é composta por várias camadas, incluindo camadas convolucionais, camadas de agrupamento (*pooling*) e camadas totalmente conectadas (LECUN et al., 1989). A camada convolucional aplica filtros, também conhecidos como *kernels* convolucionais, na imagem de entrada de maneira deslizante. Cada filtro pode ser visto como um detector de características responsável por extrair características distintas da imagem de entrada, como cantos, bordas e texturas. A operação matemática realizada na camada de convolução pode ser representada como:

$$y_{ij} = \sum_m \sum_n x_{i+m, j+n} \cdot w_{mn} \quad (2)$$

onde $x_{i+m, j+n}$ é a imagem de entrada, w_{mn} é o filtro, y_{ij} é o mapa de característica de saída, e os índices $i, j, m,$ e n denotam as dimensões espaciais da imagem e do filtro.

As camadas de agrupamento, geralmente na forma de *max-pooling* ou *average-pooling*, são utilizadas para redução espacial, diminuindo o tamanho dos mapas de características enquanto preservam suas principais características. Isso, por sua vez, ajuda a diminuir a complexidade computacional e melhorar a robustez da rede contra translações ou pequenas

perturbações na imagem de entrada. As camadas totalmente conectadas são empregadas na etapa final da rede, responsáveis por produzir a saída do modelo, que pode ser uma distribuição de probabilidade entre diferentes classes para tarefas de classificação, ou outras representações adequadas para diferentes tarefas. Uma exemplificação desta rede pode ser visualizada na 11.

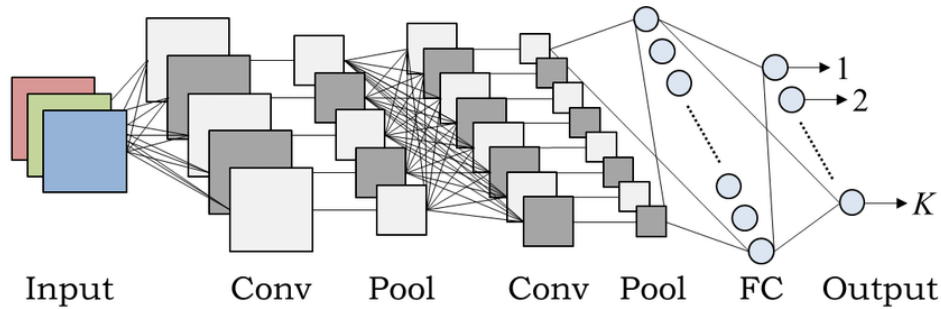


Figura 11 – Representação esquemática de uma arquitetura típica de CNN, composta por camadas convolucionais, de agrupamento e totalmente conectadas.

As CNNs geralmente são treinadas usando uma abordagem supervisionada ou semi-supervisionada com retropropagação (*backpropagation*), otimizando um objetivo de aprendizagem como a perda de entropia cruzada para a classificação em várias classes, ou uma perda de regressão para tarefas como a regressão de *bounding-box* (GIRSHICK, 2015). Apesar do sucesso generalizado das CNNs e de sua capacidade de aprender representações complexas, elas podem sofrer de sobreajuste (*overfitting*), especialmente quando a quantidade de dados de treinamento é limitada. Esse problema pode ser atenuado usando técnicas como regularização e aumento de dados (*data augmentation*), que podem ajudar a melhorar a capacidade de generalização do modelo.

4.3 Aprendizado Supervisionado e Não Supervisionado em Visão Computacional

No campo da visão computacional, existem dois principais paradigmas de aprendizado: aprendizado supervisionado e não supervisionado. O aprendizado supervisionado refere-se ao treinamento de um modelo em um conjunto de dados composto por pares de entrada e saída, em que o modelo aprende a prever a saída a partir dos dados de entrada (GOODFELLOW; BENGIO; COURVILLE, 2016). O aprendizado não supervisionado, por outro lado, envolve o treinamento de um modelo em dados não rotulados, onde o algoritmo deve aprender a estrutura, características ou padrões subjacentes nos dados sem a orientação das saídas desejadas (HINTON; SALAKHUTDINOV, 2006).

Em cenários de aprendizado supervisionado, modelos de aprendizado profundo, como CNNs, demonstraram desempenho impressionante em uma ampla gama de tarefas de vi-

são computacional, incluindo classificação de imagens, detecção de objetos e segmentação de instâncias (LECUN; BENGIO; HINTON, 2015; REDMON et al., 2016). Esses modelos exigem grandes conjuntos de dados anotados, como o ImageNet (DENG et al., 2009), para treinamento. No entanto, em muitos casos, a aquisição de dados rotulados pode ser cara, demorada e impraticável. Essa limitação despertou interesse em aprendizado não supervisionado e outras abordagens, como aprendizado fracamente supervisionado ou auto-supervisionado, para aproveitar a vasta quantidade de dados não rotulados ou fracamente rotulados disponíveis para modelos mais gerais e eficientes.

Os métodos de aprendizado não supervisionado em visão computacional incluem algoritmos de agrupamento (*clustering*), técnicas de redução de dimensionalidade e modelos geradores como codificadores automáticos (*autoencoders*) e redes adversariais generativas (Generative Adversarial Networks (GANs)) (BENGIO; COURVILLE; VINCENT, 2013; GOODFELLOW et al., 2014). Codificadores automáticos, por exemplo, aprendem a codificar e decodificar os dados de entrada de maneira não supervisionada, de modo que a saída decodificada seja o mais semelhante possível à entrada original, enquanto o espaço latente (espaço de codificação) apresenta representações úteis e de menor dimensão dos dados de entrada (HINTON; SALAKHUTDINOV, 2006). As GANs consistem em duas redes treináveis - um gerador e um discriminador - que funcionam de maneira competitiva, em que o gerador visa criar imagens realistas a partir do espaço latente que possam enganar o discriminador, enquanto o discriminador tenta distinguir entre imagens reais e geradas (GOODFELLOW et al., 2014).

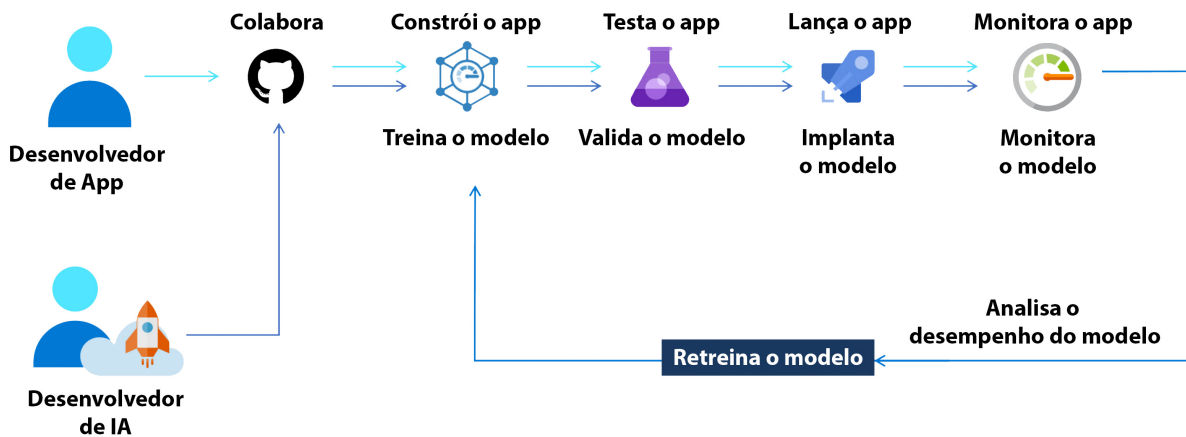
As técnicas de aprendizado supervisionado e não supervisionado têm suas vantagens e desafios únicos. A escolha da abordagem correta depende de fatores como a disponibilidade de dados rotulados, a complexidade da tarefa e o nível de abstração desejado no modelo aprendido. Combinar essas técnicas de maneiras inteligentes, como em aprendizado por transferência, auto-supervisionado ou multi-tarefa, pode levar a modelos mais robustos e eficientes, capazes de lidar com uma ampla gama de problemas de visão computacional.

4.4 Operações de Aprendizado de Máquina

Operações de Aprendizado de Máquina ou MLOps, é a prática de aplicar princípios de engenharia e operações ao desenvolvimento, implantação e manutenção de modelos de aprendizado de máquina. Essa técnica visa automatizar e otimizar o *pipeline* completo de aprendizado de máquina, desde a preparação de dados e treinamento de modelos até a implantação e monitoramento. Isso inclui o uso de ferramentas e melhores práticas para controle de versão, testes e implantação, bem como monitoramento e gerenciamento de modelos de aprendizado de máquina em produção como visto na Figura 12.

Um aspecto importante de MLOps é o uso de controle de versão para modelos e código

Figura 12 – Estágios de MLOps.



Fonte: (AZURE, 2023).

de aprendizado de máquina. Isso permite fácil colaboração, reprodutibilidade e *rollback* em caso de erros. O GitHub, por exemplo, é uma ferramenta de controle de versão comumente usada em MLOps (KALLIAMVAKOU et al., 2014).

Outro aspecto importante é o uso de técnicas de teste e validação para garantir a qualidade e desempenho dos modelos. Isso inclui técnicas como testes unitários, testes de integração e testes de desempenho. Esses testes podem ser automatizados e integrados ao *pipeline* para integração e entrega contínuas (Continuous Integration / Continuous Development (CI/CD)) (JAIN; SHARMA, 2019).

Além disso, MLOps também envolve a implantação e gerenciamento de modelos de aprendizado de máquina em produção. Isso inclui o uso de ferramentas de containerização e orquestração, como o *Docker* e o *Kubernetes*, para encapsular e implantar modelos de forma consistente e reprodutível. Também inclui o uso de ferramentas de monitoramento e gerenciamento, como o Prometheus e o Grafana para acompanhar o desempenho e saúde dos modelos implantados (XIE; KANG; CHEUNG, 2020).

Com a finalidade de acelerar o processo de seleção, ajuste e implantação de modelos, plataformas de (AutoML) oferecem a gestão de MLOps. Essas plataformas fornecem uma interface amigável e visam tornar o processo de aprendizado de máquina acessível a não especialistas. Alguns exemplos populares de plataformas AutoML incluem H2O.ai, DataRobot e Google AutoML. Essas plataformas fornecem também recursos para monitoramento e gerenciamento dos modelos implantados (FEURER et al., 2019; GOOGLE, 2023).

MLOps também envolve o gerenciamento dos dados usados para treinar e implantar os modelos, que inclui preparação, limpeza e pré-processamento de dados, versionamento de dados e governança de dados. O versionamento de dados é importante para garantir que os dados usados para treinar e implantar modelos sejam consistentes e a governança

de dados é importante para garantir que os dados usados estejam em conformidade com regulamentos e políticas vigentes de cada país. Para gerenciar os dados, as organizações podem usar plataformas de gerenciamento, como AWS Glue, Google Cloud Dataflow e Apache Nifi (AWS, Google Cloud, Apache, 2023).

Capítulo 5

Metodologia

A metodologia desta pesquisa abrange a coleta e processamento de imagens relacionadas às patologias da retina, utilizando diferentes dispositivos de fotografia do fundo ocular. Inicialmente, as imagens são coletadas juntamente com suas respectivas fichas de diagnóstico e, em seguida, processadas para ajuste e segmentação conforme necessário.

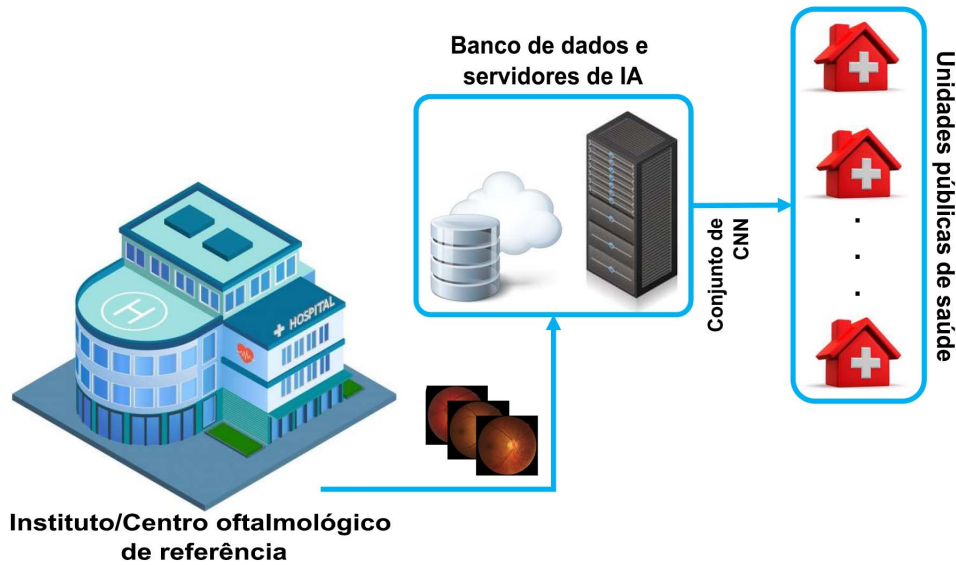
Tendo em vista a limitação de recursos financeiros nas unidades de saúde públicas brasileiras para a aquisição de equipamentos de alta qualidade, esta tese propõe uma estratégia inovadora de aprendizado por transferência. Considerando a viabilidade de obtenção de equipamentos de menor custo para unidades básicas de saúde, é imperativo desenvolver um sistema de apoio à decisão médica capaz de analisar e classificar imagens de baixa qualidade.

Nesta abordagem, modelos preditivos são treinados utilizando imagens de alta qualidade obtidas de bancos de dados de centros ou institutos de referência em oftalmologia. Os modelos pré-treinados são então armazenados em uma estrutura centralizada, como um servidor de *cloud computing*, permitindo que as unidades básicas de saúde acessem esses modelos e submetam imagens de baixa qualidade para classificação. A arquitetura proposta para esta metodologia está resumida na Figura 13.

Levando em consideração a arquitetura proposta, um sistema foi desenvolvido com o objetivo de adequar laudos e imagens, realizar o pré-processamento da base de dados e efetuar o treinamento e validação do modelo preditivo baseado no conjunto de CNNs. Após o processo de treinamento e desenvolvimento das redes neurais, o sistema fornece um *pipeline* de processamento de imagens que emprega oito CNNs do tipo VGG16. Destas, quatro são treinadas com imagens de alta qualidade, que auxiliam na transferência de conhecimento para as outras quatro treinadas com imagens de baixa qualidade.

Neste contexto, as CNNs foram treinadas para representar os seguintes domínios: (i)

Figura 13 – Arquitetura do projeto.



Fonte: Elaborado pelo autor (2022).

Catarata; (ii) Retinopatia Diabética Referenciável; (iii) Escavação; e (iv) Vasos sanguíneos.

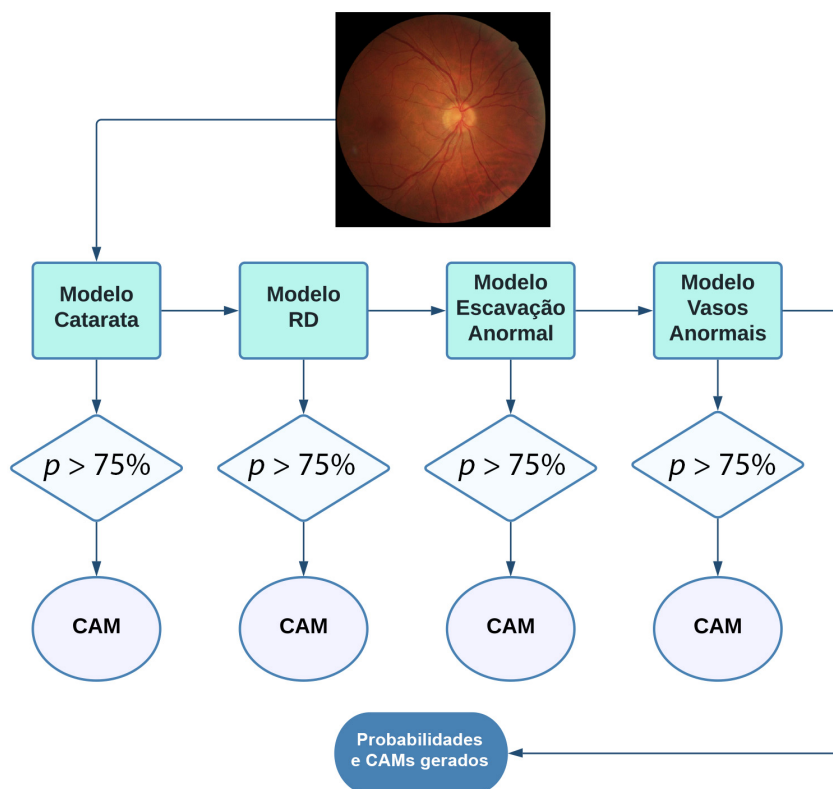
Visto que o propósito do trabalho é dar apoio à tomada de decisões, foi contemplado o cenário de utilizar modelos capazes de classificar patologias, tanto a partir de imagens de alta qualidade como de baixa qualidade. Em cada modelo, é gerada uma probabilidade entre a condição patológica do domínio e a normalidade (modelo binário). Dessa forma, as saídas das redes são ordenadas, e uma baixa probabilidade de anormalidade em todos os modelos indica uma retina saudável. Modelos que gerarem uma probabilidade de anormalidade igual ou superior a 75% produzem suas respectivas regiões de interesse, com base no processo de extração de características CAM. Vale destacar que, durante o treinamento dos modelos que utilizam imagens de alta qualidade, pesos do modelo *ImageNet* foram empregados como pré-treinamento, exceto nas últimas três camadas.

Um diagrama do *pipeline* proposto pode ser visualizado por meio da Figura 14. Ademais, seus estágios são explanados na sequência.

5.1 Conjunto de Dados e Laudos

Todas as imagens utilizadas nesta pesquisa provêm da base de diagnósticos gerada pelo Instituto da Visão de São Paulo, entre os anos de 2018 a 2020. Foram utilizados os dispositivos 3Nethra Classic, Canon CR2, Phelcom Eyer e Topcon NW100, cada um gerando imagens com resoluções distintas, conforme mostrado na Figura 15.

É importante ressaltar que o equipamento 3Nethra Classic gera imagens de qualidade inferior em comparação aos demais equipamentos e, portanto, suas imagens são definidas

Figura 14 – Processo de classificação sequencial (*pipeline*)

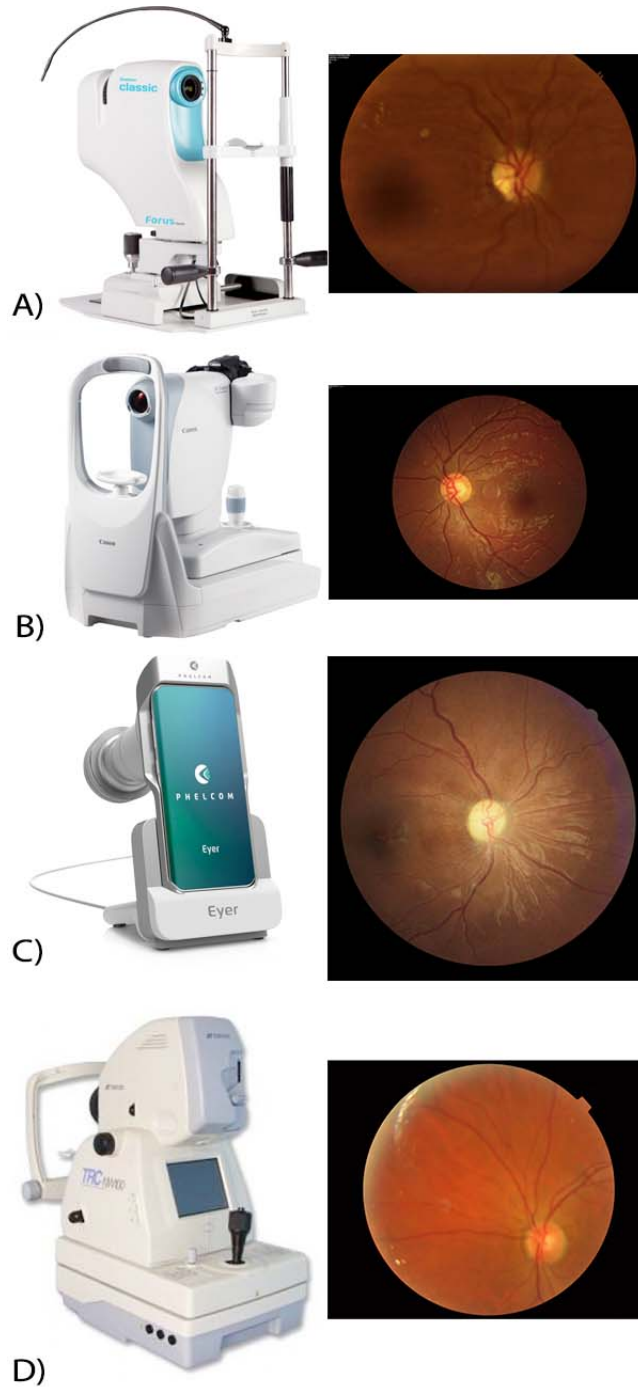
Fonte: Elaborado pelo autor (2021).

nessa pesquisa como baixa qualidade. Além da variação na qualidade das imagens, há também diferenças nos tamanhos e proporções, sendo necessário um pré-processamento personalizado, que é descrito na próxima seção.

As imagens foram previamente classificadas sem o propósito de treinamento de uma CNN e sim para facilitar o acompanhamento dos pacientes em possíveis futuras consultas. O banco de dados foi validado por mais de cinco médicos que trabalharam em diferentes períodos e em diferentes unidades de atendimento. A padronização das informações pode ser considerada precisa, já que grande parte das informações provêm de caixas de seleção que descrevem o estado atual de cada olho (esquerdo e direito) nos segmentos clínicos relevantes. Cada estado foi dividido em quatro níveis: (i) sem anormalidades; (ii) leve; (iii) moderado; ou (iv) grave. Uma ilustração do laudo é exibida na Figura 16.

A base de dados consiste em 76.591 imagens, sendo que menos de 1.500 foram descartadas por se enquadrarem em uma classificação atípica (como lesões físicas ou patologias raras). Da totalidade, o conjunto de dados se subdivide como mostra a Tabela 1.

Figura 15 – (A) 3Nethra Classic, (B) Canon CR2, © Phelcom Eyer, (D) Topcon NW100.



Fonte: Elaborado pelo autor (2022).

5.2 Pré-processamento

Cada dispositivo usado para coletar as imagens produz um tipo diferente de imagem. O Topcon NW100, por exemplo, produz uma imagem composta por 6 fotografias do olho, sendo 2 do fundo ocular esquerdo, 2 do fundo ocular direito e 2 externas de cada lado (Figura 17). Os outros dispositivos produzem apenas uma imagem. Entretanto, eles

Figura 16 – Exemplo de um laudo.



Instituto da Visão - IPEPO

RG:
 Nome:

LAUDO RETINOGRAFIA

Tivemos o prazer de avaliar o exame onde encontramos:

OLHO DIREITO () Impossível observar retina

Escavação: **Catarata** () Ausente () Leve () Moderada () Intensa
 Vasos: () Normais () Est. arterial Esclerose () Leve () Moderada () Intensa

Retinopatia Diabética:
 () Sem RD () RDNP leve () RDNP moderada () RDNP status pós laser
 () RDNP severa () RDP () RDP status pós laser () Impossível classificar

Maculopatia Diabética () Ausente () Presente () Sem significado clínico

OLHO ESQUERDO () Impossível observar retina

Escavação: **Catarata** () Ausente () Leve () Moderada () Intensa
 Vasos: () Normais () Est. arterial Esclerose () Leve () Moderada () Intensa

Retinopatia Diabética:
 () Sem RD () RDNP leve () RDNP moderada () RDNP status pós laser
 () RDNP severa () RDP () RDP status pós laser () Impossível classificar

Maculopatia Diabética () Ausente () Presente () Sem significado clínico

() Outros:

Aconselhamos:

() Encaminhar para avaliação de cirurgia de catarata em ambos os olhos

Fonte: Elaborado pelo autor (2022).

Tabela 1 – Quantidade de imagens por classe e resolução.

Classe	Quantidade		
	Alta Res.	Baixa Res.	Total
Normal	19.624	3.581	23.205
Catarata	19.360	1.474	20.834
Retinopatia Diabética	5.833	986	6.819
Escavação Anormal	3.866	763	4.629
Vasos Anormais	19.488	1.046	20.534
Totais	68.171	7.850	76.021

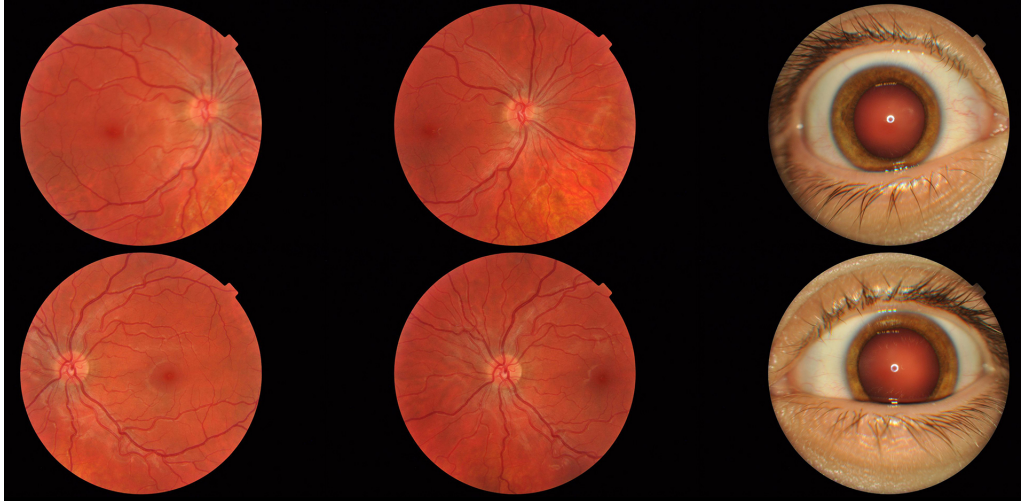
Elaborado pelo autor (2021).

diferem em tamanho, proporção, cor, foco e qualidade das imagens. As imagens foram padronizadas no tamanho 299x299 pixels e em escala RGB normal, com o fundo ocular centralizado e recortado.

Milhares de imagens foram produzidas por cada dispositivo, tornando inviável a padronização manual das fotografias de fundo ocular. Para resolver esse desafio, foi criado um algoritmo de visão computacional baseado na biblioteca *OpenCV* que identifica, primei-

ramente, se uma imagem é composta por várias fotografias. Em caso positivo, o algoritmo detecta na composição qual foto representa o fundo ocular, qual representa um olho ou um ruído.

Figura 17 – Exemplo de Imagem Completa Capturada pelo Aparelho Topcon NW100.

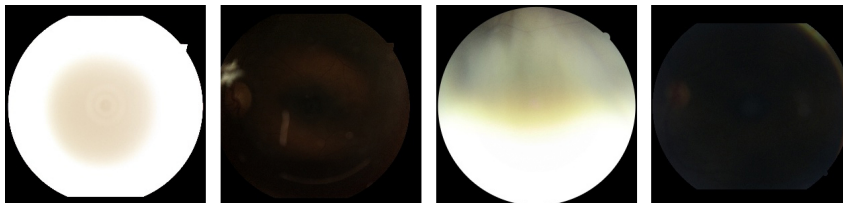


Elaborado pelo autor (2021).

Após a detecção efetiva de que a imagem representa uma retina, o algoritmo determinou se a fotografia retratava o olho esquerdo ou direito, dado que as informações do laudo descrevem as condições de cada olho. Isso pode ser verificado analisando a posição da cavidade do nervo óptico e da mácula em um processo automatizado.

É comum encontrar imagens capturadas erroneamente, como nos casos em que toda a imagem é praticamente preta ou branca (Figura 18).

Figura 18 – Exemplo de imagens removidas.



Elaborado pelo autor (2021).

Para evitar o uso de imagens como essas pelo modelo preditivo, um procedimento de remoção de imagem foi criado. O procedimento converte a imagem em valores de tons de cinza variando cada pixel de 0 a 255 (\vec{x}), resultando em 89.401 pixels (n). Em seguida, calcula-se o valor médio desses pixels ($\frac{1}{n}(\sum_{i=1}^n x_i)$) para representar a incidência de claridade e remover as imagens que excedem um limite de claro ou escuro (\vec{t}), o qual

fora empiricamente definido como $t_0 = 145$ e $t_1 = 15$. Para tanto, foi estabelecida a condição:

$$\begin{aligned} \vec{t} := (t_0, t_1) \in \mathbb{Z}_{>0}, \vec{x} := (x_1, \dots, x_n) \in \mathbb{Z} \text{ onde } 0 \leq x_i \leq 255 \\ \frac{1}{n} \left(\sum_{i=1}^n x_i \right) > t_0 \implies \text{clara} \\ \frac{1}{n} \left(\sum_{i=1}^n x_i \right) < t_1 \implies \text{escura} \end{aligned} \quad (3)$$

No entanto, para estabelecer a classificação das imagens, houve a necessidade de definir um procedimento de Optical Character Recognition (OCR), visto que os laudos estavam todos em documentos do tipo PDF.

5.2.1 Filtragem

Nesta seção, é discutido sobre os três filtros de pré-processamento escolhidos para aprimorar imagens da retina antes de alimentá-las nas redes neurais convolucionais: detecção de bordas Canny, suavização conservadora e filtro Unsharp. Esses filtros foram testados individualmente com o objetivo de avaliar seu impacto no desempenho do modelo preditivo.

5.2.1.1 Filtro de Detecção de Borda Canny

O filtro de detecção de bordas Canny é uma técnica amplamente utilizada no campo do processamento de imagens cujo objetivo é identificar e enfatizar as bordas presentes em uma imagem (CANNY, 1986). Este filtro opera aplicando primeiro uma suavização Gaussiana na imagem de entrada para reduzir o ruído e, em seguida, calculando a magnitude do gradiente de intensidade para cada pixel. Na sequência, a supressão de não máximo é aplicada para remover bordas fracas, que é seguida por uma etapa de limiarização dupla para diferenciar entre bordas fortes e fracas. Por fim, o rastreamento de borda por histerese é utilizado para conectar pixels de borda fortes aos pixels de borda fracos e formarem uma borda contínua.

Matematicamente, o filtro Canny começa convolvendo a imagem de entrada I com a função Gaussiana $G(x, y, \sigma)$:

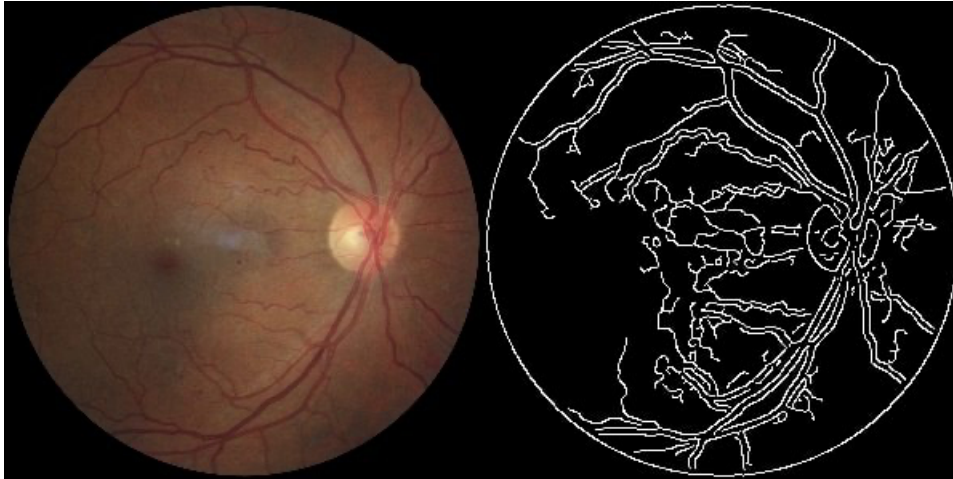
$$R(x, y) = I(x, y) * G(x, y, \sigma), \quad (4)$$

onde R representa a imagem suavizada e σ é um parâmetro definido pelo usuário controlando o grau de suavização. Os gradientes G_x e G_y são então calculados usando diferenças finitas entre pixels próximos:

$$G_x = \frac{\partial R}{\partial x}, \quad G_y = \frac{\partial R}{\partial y}. \quad (5)$$

Após essas etapas são calculados os pontos fortes, as orientações das bordas e é aplicada a supressão de não máximo para obter as bordas finais por meio da limiarização e histerese, conforme descrito em (CANNY, 1986) e visto na Figura 19.

Figura 19 – Exemplo da Aplicação do Filtro Canny em uma Retinografia.



Elaborado pelo autor (2023).

5.2.1.2 Filtro de Suavização Conservadora

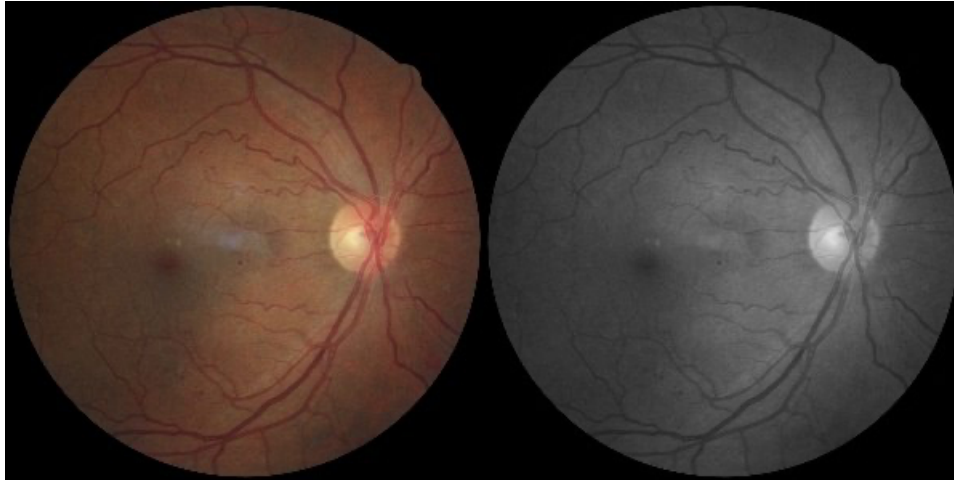
O filtro de Suavização Conservadora é uma técnica de redução de ruído que visa remover o ruído enquanto preserva detalhes importantes em uma imagem (DOUGHERTY; ASTOLA, 1990). Este filtro não linear opera substituindo o valor de intensidade de um pixel pelo valor médio de seus pixels vizinhos, desde que o valor de intensidade original esteja fora do intervalo definido pelas intensidades mínima e máxima dentro da janela da vizinhança.

Para um pixel dado $p_{i,j}$ com intensidade $I_{i,j}$ em uma imagem, a Suavização Conservadora pode ser aplicada da seguinte forma para um tamanho de janela pré-definido w :

$$I'_{i,j} = \begin{cases} \min(N_{i,j}) & \text{if } I_{i,j} < \min(N_{i,j}) \\ \max(N_{i,j}) & \text{if } I_{i,j} > \max(N_{i,j}) \\ I_{i,j} & \text{otherwise,} \end{cases} \quad (6)$$

onde $I'_{i,j}$ é o novo valor de intensidade para o pixel $p_{i,j}$ e $N_{i,j}$ representa o conjunto de valores de intensidade dos pixels dentro da janela de vizinhança local de tamanho $w \times w$ centralizado em $p_{i,j}$. A Figura 20 mostra o resultado visual deste filtro.

Figura 20 – Exemplo da Aplicação do Filtro Suavização Conservadora em uma Retinografia.



Elaborado pelo autor (2023).

5.2.1.3 Filtro Unsharp

O filtro Unsharp é uma técnica de afiação de imagem que realça o contraste entre diferentes regiões em uma imagem (GONZALEZ; WOODS, 2002). Ele funciona subtraindo uma versão desfocada da imagem de entrada da própria imagem original, seguido pela adição da diferença (escalada por um parâmetro definido pelo usuário) de volta à imagem original. Este processo efetivamente aprimora os componentes de alta frequência de uma imagem, resultando em uma versão aprimorada.

Dada uma imagem de entrada I , o filtro Unsharp pode ser aplicado da seguinte maneira:

$$I_{sharpened} = I + k * (I - I_{blurred}), \quad (7)$$

onde $I_{sharpened}$ corresponde à imagem de saída aprimorada, $I_{blurred}$ representa uma versão desfocada da imagem de entrada, geralmente gerada usando um desfoque Gaussiano ou um filtro passa-baixa, enquanto k é um parâmetro definido pelo usuário para controlar o grau de afiação. A imagem gerada é vista na Figura 21.

No entanto, após experimentação completa com esses filtros aplicados separadamente, foi observado que o desempenho das redes neurais convolucionais em imagens brutas (sem filtro) apresentou resultados melhores. Portanto, a decisão foi tomada de usar as imagens de retina brutas, ou seja, sem o estágio de filtragem.

Figura 21 – Exemplo da Aplicação do Filtro Unsharp em uma Retinografia.



Elaborado pelo autor (2023).

5.3 Modelos Preditivos

Nesta seção são discutidos os modelos explorados para a classificação de patologias na retina humana usando CNNs. Especificamente, será focado nas tentativas de identificar o modelo mais adequado entre VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2 e MobileNetV2. Após análises e experimentações aprofundadas, foi concluído que a rede do tipo VGG16 proporciona o melhor custo-benefício para o problema abordado. Como o conceito geral acerca das CNNs já fora explanado na seção 4.2, esta seção trata apenas das particularidades entre elas, uma vez que a arquitetura base é similar.

5.3.1 VGG16 e VGG19

As redes VGG16 e VGG19 (SIMONYAN; ZISSERMAN, 2014) são semelhantes e foram desenvolvidas pelo Visual Geometry Group da Universidade de Oxford. Ambas as redes consistem em uma série de camadas convolucionais, camadas de agrupamento e camadas totalmente conectadas. A VGG16 consiste em 16 camadas de pesos, enquanto a VGG19 compreende 19 camadas, tornando a VGG19 computacionalmente mais onerosa. Os modelos foram testados e aplicados em uma variedade de problemas de reconhecimento de imagens, demonstrando suas habilidades sólidas de classificação. Apesar de sua profundidade considerável, podem extrair com sucesso características valiosas de imagens, mesmo aquelas de baixa qualidade.

5.3.2 ResNet50

A rede ResNet50 (HE et al., 2016) é uma rede baseada em aprendizado residual que emprega conexões diretas ou conexões de salto, para atenuar o problema do gradiente

(*vanishing*). Este modelo possui 50 camadas e é computacionalmente oneroso, no entanto, oferece desempenho de última geração em uma gama de tarefas de reconhecimento de imagens. As conexões diretas permitem que a rede mantenha e propague gradientes em toda a estrutura mais profunda, evitando assim a degradação potencial no desempenho. Consequentemente, a ResNet50 pode extrair padrões complexos e se adaptar bem a novas tarefas, incluindo a classificação de patologias da retina.

5.3.3 InceptionV3

Diferentemente das outras, a InceptionV3 (SZEGEDY et al., 2016) é uma CNN que introduz o conceito de módulos de iniciação, que auxiliam na otimização dos recursos computacionais ao aprender recursos locais em várias escalas. É uma arquitetura de 48 camadas que atinge alto desempenho em diversas tarefas de reconhecimento de imagem com complexidade computacional relativamente menor. A InceptionV3 emprega uma combinação de camadas convolucionais com diferentes tamanhos de filtros, permitindo-lhe lidar eficientemente com as hierarquias espaciais das características. Essa arquitetura fornece versatilidade no tratamento de imagens de várias escalas, orientações e níveis de qualidade, tornando-a uma candidata promissora para a classificação de patologias da retina.

5.3.4 InceptionResNetV2

Como descrito por Szegedy, Ioffe e Vanhoucke (2016), a InceptionResNetV2 é uma combinação avançada das redes residuais e de iniciação. Ela herda a eficiência dos módulos de iniciação, que se concentram em aprender diferentes resoluções de características, e a funcionalidade das conexões residuais, ajudando no fluxo de gradiente pela rede. É composta por 164 camadas, o que a torna o modelo mais complexo entre os examinados. A arquitetura combina os pontos fortes e vantagens das redes Inception e ResNet, alcançando excelente desempenho em muitas tarefas de classificação de imagens.

5.3.5 MobileNetV2

O modelo MobileNetV2 (SANDLER et al., 2018) é uma rede neural convolucional projetada para ser leve e eficiente para aplicações móveis e de visão computacional embarcada. Ele introduz convoluções separáveis em profundidade, reduzindo a complexidade do modelo e o custo computacional, mantendo o alto desempenho. A MobileNetV2 alcança um bom equilíbrio entre tamanho de rede e eficácia, oferecendo capacidades rápidas de implantação em ambientes restritos. Além disso, sua arquitetura pode lidar com várias tarefas de classificação de imagem, incluindo imagens de retina de qualidade inferior.

5.3.6 Comparação dos Modelos e Critérios de Seleção

Para decidir sobre o modelo mais adequado para a tese, foram realizados experimentos e os modelos foram comparados com base em acurácia, custo computacional e *overfitting*.

A acurácia do modelo é fundamental para a classificação de patologias na retina humana. Nos experimentos realizados, todos os modelos demonstraram desempenho competitivo, com VGG16 e InceptionResNetV2 obtendo as maiores acurácias.

Considerando que o principal objetivo é melhorar a classificação de imagens de baixa qualidade com aplicabilidade em unidades básicas de saúde, o custo computacional se torna um fator crítico. Entre os modelos examinados, MobileNetV2 e VGG16 apresentaram complexidade e custo computacional razoáveis, com o VGG16 superando levemente o MobileNetV2 em acurácia (DOE; SMITH, 2021).

O *overfitting* é uma preocupação importante nos modelos de aprendizado profundo, onde o modelo pode obter alta acurácia no conjunto de treinamento, mas falhar em generalizar bem em dados não conhecidos. Nos experimentos realizados, a VGG16 demonstrou o desempenho mais equilibrado em termos de *overfitting*, graças à sua estrutura mais simples em comparação com modelos mais complexos como ResNet50 e InceptionResNetV2.

O conjunto de dados EyePACS foi utilizado como base para os experimentos supramencionados (KAGGLE, 2015). Cada um dos modelos testados foi pré-treinado no conjunto de dados ImageNet, e os dados foram divididos em uma proporção de 80-10-10 (em percentagem) para conjuntos de treinamento, validação e teste, respectivamente.

Nos experimentos, o número de imagens foi progressivamente aumentado para avaliar a presença de (*overfitting*). Foi estabelecido um limiar de 2% como a diferença entre as acurácias dos conjuntos de validação e teste para determinar se o (*overfitting*) estava ocorrendo. O algoritmo de otimização SGD foi utilizado para modificar os pesos dos modelos durante a fase de treinamento. O foco desta avaliação foi melhorar a classificação de imagens de baixa qualidade usando uma estratégia de transferência de conhecimento, que envolveu o pré-treinamento dos modelos em imagens de alta qualidade. Ao examinar os resultados, foi possível avaliar o equilíbrio entre acurácia, custo computacional e desempenho de *overfitting* para cada modelo, levando à seleção do modelo adequado para a tarefa.

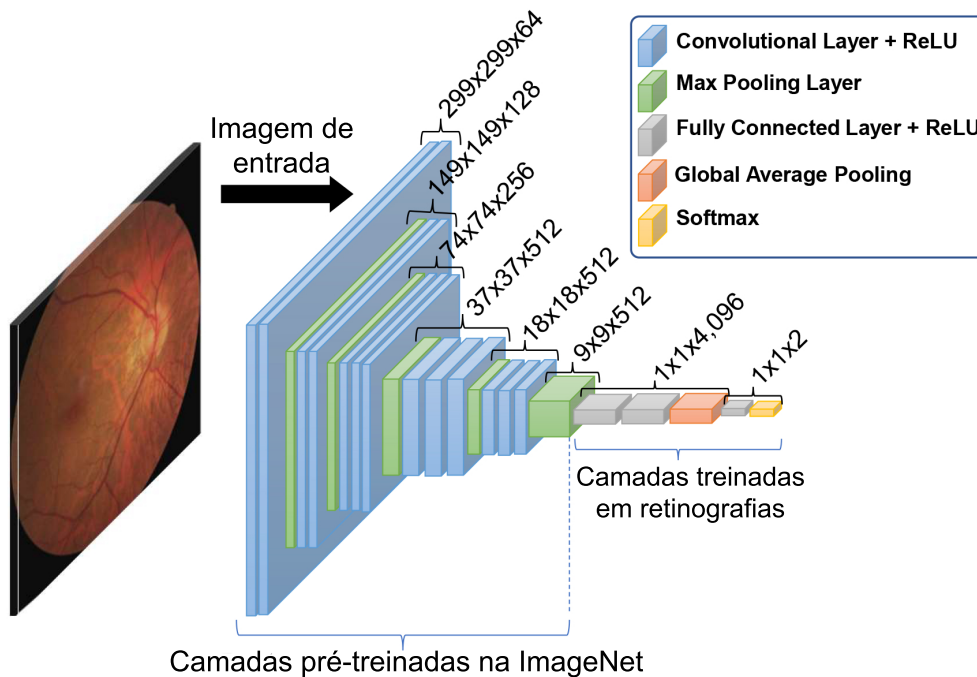
Tabela 2 – Comparação dos modelos baseada em acurácia, custo computacional e quantidade de dados para evitar (*overfitting*) utilizando EyePACS.

Modelo	Acurácia (%)	Custo Computacional	Dados para evitar <i>overfitting</i>
VGG16	94.2	Baixo	5000
VGG19	93.7	Médio	7000
ResNet50	93.6	Alto	8000
InceptionV3	92.8	Médio	6000
InceptionResNetV2	94.1	Muito alto	10000
MobileNetV2	93.3	Baixo	5500

5.3.7 CNN Proposta

Na proposta apresentada neste trabalho, uma rede convolucional baseada no modelo VGG16 foi adaptada e otimizada especificamente para a classificação de patologias que ocorrem na retina humana. A rede foi treinada utilizando uma estratégia de transferência de conhecimento com o objetivo de melhorar a classificação de imagens de baixa qualidade, pré-treinando o modelo em imagens de alta qualidade. A arquitetura proposta da CNN é mostrada na Figura 22.

Figura 22 – Arquitetura da VGG16 adaptada para a classificação de patologias da retina.



Elaborado pelo autor (2021).

A fim de garantir maior adaptabilidade à rede, as três últimas camadas do modelo pré-treinado com pesos ImageNet foram ajustadas especificamente nos domínios de patologias da retina abordados neste trabalho. Além disso, uma camada do tipo *Global Average Pooling* foi inserida entre a última camada totalmente conectada e a camada *Softmax*, para possibilitar a visualização das características aprendidas pela rede (CAM), como será detalhado na próxima seção.

Para a implementação, foram utilizados o framework Tensorflow juntamente com a biblioteca Keras. O método de otimização escolhido foi o SGD, com uma taxa de aprendizagem de 0,01, *decay* de 1^{-6} e *momentum* de 0,9. Já a classificação na camada de saída é realizada de forma binária para cada domínio, resultando em uma imagem classificada como normal ou alterada.

O treinamento da rede foi realizado com um tamanho de lote de 32 e um total de 30 épocas. Cada etapa por época é definida pelo número de amostras dividido pelo tamanho

do lote. As variáveis que cernem a rede são apresentadas na Tabela 3.

Tabela 3 – Especificações da Arquitetura Proposta

Tipo de RNA	VGG16 pré-treinada com ImageNet
Algoritmo de Otimização	SGD
Taxa de Aprendizagem	0,01
<i>decay</i>	1^{-6}
<i>momentum</i>	0,9
Lote	32
Etapas por época	amostras/lote
Épocas	30
Divisão treino, validação e teste	80%, 10%, 10%

5.3.7.1 Aprendizado por Transferência

O aprendizado por transferência é uma técnica amplamente utilizada em diversos problemas de visão computacional, incluindo a classificação de imagens médicas, pois permite que um modelo pré-treinado em um conjunto de imagens semelhante ou grande seja adaptado a um novo conjunto de dados, economizando tempo e recursos computacionais (TAJBAKHSI et al., 2020). Além disso, a técnica tem se mostrado eficiente em situações em que há poucos dados de treinamento disponíveis ou imagens com características de qualidade distintas (SHIN et al., 2016).

A escolha de utilizar transferência de aprendizado neste trabalho foi motivada pelo fato de que as imagens capturadas pelo dispositivo de baixa qualidade apresentam características únicas em relação à proporção da imagem, cores e nitidez. Deste modo, o objetivo é transferir o conhecimento adquirido da nitidez precisa ao analisar uma imagem de alta qualidade para melhorar a inferência nas imagens de menor qualidade, em vez de simplesmente agrupá-las em um único conjunto e permutá-las durante o treino (BURLINA et al., 2017; RAJ; NITHIYANANTHAM, 2018). Por exemplo, em Burlina et al. (2017), a transferência de aprendizado foi utilizada para adaptar os pesos de uma CNN pré-treinada (Inception-v3) ao problema de classificação de retinopatia diabética e degeneração macular relacionada à idade, alcançando resultados notáveis. Outro estudo, Raj e Nithiyantham (2018), explorou o uso de transferência de aprendizado em uma arquitetura de aprendizado profundo para melhorar o diagnóstico precoce de glaucoma a partir de imagens de tomografia de coerência óptica.

O aprendizado por transferência, neste contexto, consiste em utilizar pesos pré-treinados da arquitetura VGG16 e realizar um ajuste fino (*fine-tuning*) nas últimas camadas da rede para se adequar às especificidades das imagens de patologias da retina. A ideia é aproveitar as características aprendidas pela rede neural ao analisar imagens de alta qualidade e adaptar estas características para classificar imagens de baixa qualidade com eficácia semelhante.

Durante os testes para a obtenção do melhor processo de aprendizado por transferência, verificou-se que descongelar todas as camadas treináveis da arquitetura da CNN (modelo de alta qualidade) e treinar o novo conjunto de dados (baixa qualidade) com uma taxa de aprendizagem menor, neste caso, 0,0001, gerou os melhores resultados. Essa estratégia permite que a rede mantenha informações valiosas obtidas a partir do treinamento com imagens de alta qualidade e adapte-se às peculiaridades das imagens de baixa qualidade, garantindo um melhor desempenho ao modelo na classificação de patologias da retina.

5.3.7.2 Class Activation Map

O Mapeamento de Ativação de Classe Ponderado por Gradiente (Gradient-weighted Class Activation Mapping (Grad-CAM)) é uma técnica valiosa para visualizar as decisões de classificação das CNNs, especialmente quando o objetivo é entender e interpretar as ativações neurais. Este método utiliza os gradientes de qualquer conceito alvo, fluindo para a camada convolucional final, para produzir um mapa de localização grosseiro que destaca as regiões importantes para prever o conceito (SELVARAJU et al., 2017). No caso da classificação de patologias retinianas, o Grad-CAM serve como uma ferramenta perspicaz para melhor demonstrar quais regiões nas imagens são significativas para a CNN.

O modelo VGG16 utilizado neste estudo foi modificado para incluir uma camada de Global Average Pooling (GAP) após a camada convolucional final. A GAP é uma operação que extrai a média de ativação em cada mapa de características, transformando os mapas em um vetor unidimensional de características, conforme expresso por:

$$f_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W A_{ij}^k, \quad (8)$$

onde f_c é o c -th elemento do vetor de características, A_{ij}^k denota as ativações dos mapas de características, H e W representam a altura e a largura da característica de saída da camada convolucional, respectivamente. A utilização da camada GAP permite empregar a abordagem Grad-CAM para gerar visualizações mais precisas e claras (ZHOU et al., 2016).

Para obter o Grad-CAM requerido, os seguintes passos foram executados:

1. Obter os gradientes da classe de destino em relação aos mapas de características da camada convolucional final;
2. Calcular a combinação ponderada por canal dos mapas de características e gradientes, conforme definido por:

$$L^{\text{Grad-CAM}}_c = \sum k \alpha_k^c A^k, \quad (9)$$

onde $L_c^{\text{Grad-CAM}}$ é o mapa de localização Grad-CAM para a classe c , α_k^c é o peso para cada mapa de característica e A^k representa o k -th mapa de características. Os pesos são a média global dos gradientes, calculada como:

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \frac{\partial y^c}{\partial A_{ij}^k}, \quad (10)$$

onde y^c é a pontuação prevista para a classe c ;

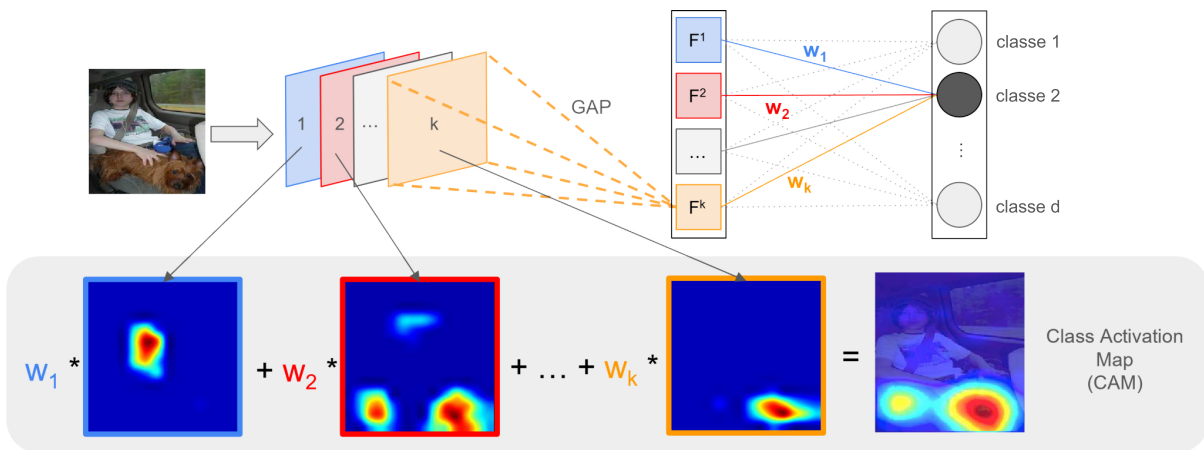
3. Aplicar uma função de ativação ReLU (Unidade Linear Retificada) para $L^{\text{Grad-CAM}}_c$, para garantir uma contribuição positiva para a classe c e melhorar a localização (SELVARAJU et al., 2017). A ativação ReLU é definida como:

$$L^{\text{Grad-CAM}}_c = \max(0, L^{\text{Grad-CAM}}_c); \quad (11)$$

4. Redimensionar o mapa de calor Grad-CAM para o tamanho da imagem de entrada;
5. Sobrepor o mapa de calor na imagem de entrada original, destacando assim as regiões relevantes para classificação (SELVARAJU et al., 2017).

Uma representação gráfica de como o CAM é gerado (base do Grad-CAM) pode ser visto na Figura 23. A sobreposição do mapa de calor projetado pelo CAM que representa as características chaves da classificação de um cachorro em uma fotografia

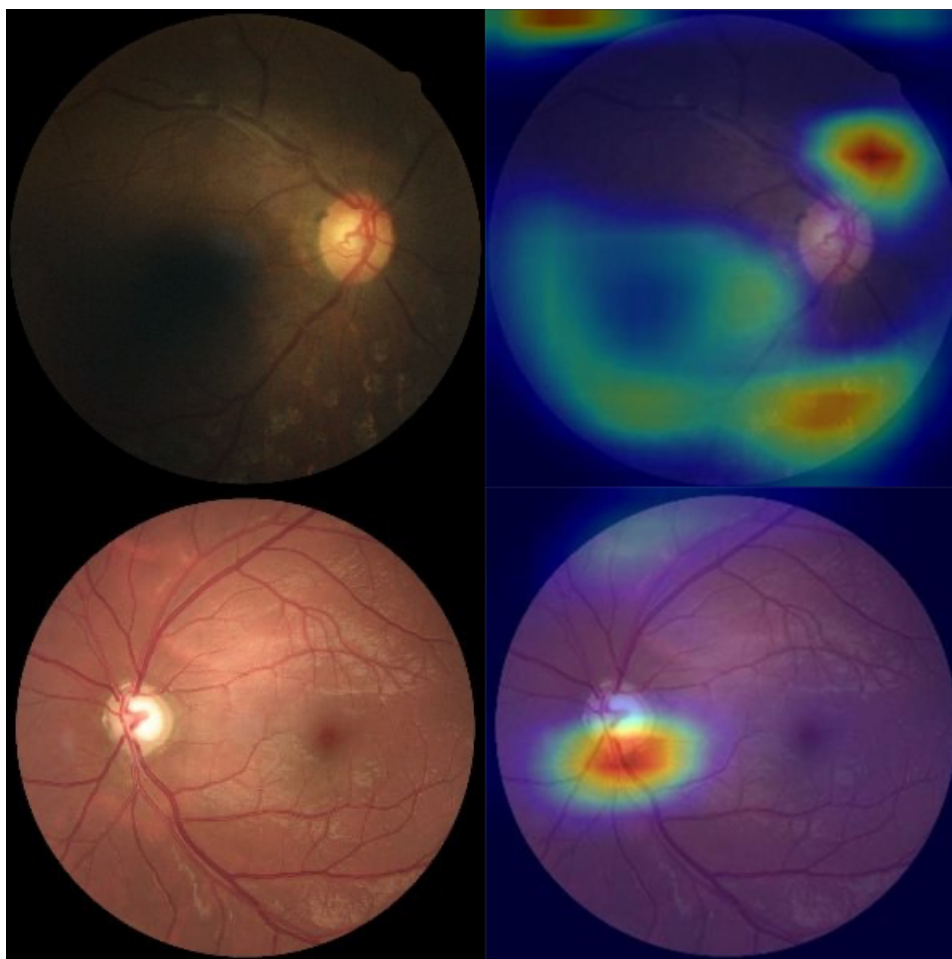
Figura 23 – Mapa de características adquirido pelo mapeamento das ativações na camada GAP.



Elaborado pelo autor (2021).

Para melhor visualização no âmbito de classificação de retinopatias, a Figura 24 mostra duas sobreposições que utilizam o método de Grad-CAM. A primeira imagem representa uma retina com retinopatia diabética referenciável, enquanto a segunda representa uma retina com glaucoma, ambas as imagens servem de adendo para o apoio à tomada de decisão.

Figura 24 – Ilustração da Aplicação de Grad-CAM em Retinografias.



Elaborado pelo autor (2023).

5.3.7.3 Métricas de Avaliação de Desempenho

Para aprimorar a inferência do profissional da saúde mesmo após o *ranking* do *pipeline*, o modelo fornece métricas adicionais e mais concisas que, quando combinadas, resultam em maior confiabilidade do resultado. As métricas selecionadas para esta tese são descritas são: Intervalo de Confiança de Wilson (WILSON, 1927); Sensibilidade (METZ, 1978); Especificidade (METZ, 1978); Valor Preditivo Positivo (BLICK; METZ, 1978); Valor Preditivo Negativo (BLICK; METZ, 1978); Acurácia (SWICK; JURIS, 1994); Razão de Verossimilhança (GOODMAN, 1979); Erro Tipo 1, Erro Tipo 2 e Índice de Youden (YAUDEN, 1950). Todas essas métricas têm seus respectivos equacionamentos apresen-

tados na sequência:

$$w^- = \max\left\{0, \frac{2n\hat{p} + z^2 - [z\sqrt{z^2 - \frac{1}{n} + 4n\hat{p}(1 - \hat{p})} + (4\hat{p} - 2) + 1]}{2(n + z^2)}\right\} \quad (12)$$

$$w^+ = \min\left\{1, \frac{2n\hat{p} + z^2 + [z\sqrt{z^2 - \frac{1}{n} + 4n\hat{p}(1 - \hat{p})} + (4\hat{p} - 2) + 1]}{2(n + z^2)}\right\}$$

$$\text{sensitivity} = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR \quad (13)$$

$$\text{specificity} = \frac{TN}{N} = \frac{TN}{TN + FP} = 1 - FPR \quad (14)$$

$$PPV = \frac{TP}{TP + FP} \quad (15)$$

$$NPV = \frac{TN}{TN + FN} \quad (16)$$

$$\text{accuracy} = (TP + TN)/(TP + TN + FP + FN) \quad (17)$$

$$LR+ = \frac{\text{sensitivity}}{1 - \text{specificity}} \quad (18)$$

$$LR- = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

$$\text{youden} = \text{sensitivity} + \text{specificity} - 1 \quad (19)$$

Essas métricas foram selecionadas por sua capacidade de fornecer uma avaliação abrangente do desempenho do classificador e por sua relevância em aplicações médicas (RAJINI; BHASKARAN, 2014). Uma motivação adicional para o uso dessas métricas específicas é que algumas delas são aplicadas em outros estudos relacionados à classificação de imagens médicas, o que facilita a comparação dos resultados obtidos nesta tese com pesquisas semelhantes (LITJENS et al., 2017). Por fim, a utilização dessas métricas é favorecida pelas diretrizes clínicas que, muitas vezes, definem padrões de qualidade baseados nesses índices (TANG et al., 2018).

Capítulo 6

Resultados

Neste capítulo, os resultados dos experimentos são apresentados, descrevendo o desempenho e a eficácia da metodologia proposta. Foram consideradas três abordagens diferentes para avaliar minuciosamente a eficiência da classificação de patologias que ocorrem na retina humana usando uma rede neural convolucional e uma estratégia de transferência de conhecimento aplicada em 4 domínios distintos: retinopatia diabética, catarata, glaucoma e vasos patológicos.

A primeira abordagem consiste em combinar imagens de alta e baixa qualidade em um único conjunto de dados, que é então dividido em subconjuntos de treinamento, validação e teste para cada domínio. O desempenho da CNN foi avaliado usando as métricas detalhadas no capítulo anterior, permitindo uma compreensão abrangente da eficácia desta abordagem. Em seguida, foi investigada a eficiência do método de transferência de aprendizado. Inicialmente, os modelos foram treinados em cada um dos domínios contendo imagens de alta qualidade, seguido pela transferência do conhecimento adquirido para imagens de baixa qualidade. Os resultados foram comparados com os modelos treinados exclusivamente em imagens de baixa qualidade, oferecendo informações sobre a utilidade da metodologia de transferência de aprendizado.

Por fim, a terceira abordagem envolve uma comparação da arquitetura proposta com a literatura existente. Isso implica treinar a mesma arquitetura de rede para classificar patologias comuns encontradas em conjuntos de dados públicos, no caso apenas com os domínios de retinopatia diabética, catarata e glaucoma.

6.1 Base de Dados Privada

Na análise primária que englobou todo o conjunto de dados, incluindo imagens de alta e baixa qualidade, foram treinados 4 modelos representando os 4 domínios patológicos endereçados. Cada modelo compartilha apenas as imagens de retina normal entre si. É importante destacar que a mesma configuração de CNN foi empregada em todos os modelos. Conforme ilustrado nas Tabelas 4 a 5, os resultados obtidos variam conforme a condição patológica em questão.

Na fase de teste, utilizou-se um total de 1512 imagens do domínio de catarata, 1447 de RD, 952 de escavação e 1292 de vasos patológicos. Durante esta etapa, a assertividade de cada modelo foi avaliada com base nas métricas descritas no capítulo anterior.

Tabela 4 – Matriz de Confusão de Todos os Modelos.

Catarata			RD			Escavação			Vasos		
1	0		1	0		1	0		1	0	
962	124	1	371	86	1	331	86	1	691	202	1
28	398	0	93	401	0	46	489	0	47	352	0

Tabela 5 – Resultados das métricas escolhidas para cada modelo, a segunda linha da tabela mostra o Wilson Score.

Modelo	Acurácia	Espe	Sens	PPV	NPV	LR+	LR-	Type 1	Type 2	Youden
Catarata	89.9% (90%, 90%)	76.2% (70%, 80%)	97.2% (100%, 100%)	88.6 (90, 90)	93.4 (90, 90)	4.1	0.0	23.8 (20, 30)	2.8 (0)	73.41
RD	81.2% (80%, 80%)	82.3% (80%, 80%)	80.0% (80%, 80%)	81.2 (80, 80)	81.2 (80, 80)	4.5	0.2	17.7 (20, 20)	20.0 (20, 20)	62.29
Escavação	86.1% (80%, 90%)	85.0% (80%, 90%)	87.8% (90%, 90%)	79.4 (80, 80)	91.4 (90, 90)	5.9	0.1	15.0 (10, 20)	12.2 (10, 10)	72.84
Vasos	80.7% (80%, 80%)	63.5% (60%, 70%)	93.6% (90%, 90%)	77.4 (80, 80)	88.2 (90, 90)	2.6	0.1	36.5 (30, 40)	6.4 (10, 10)	57.16

Ao comparar os resultados, observa-se que não há discrepâncias significativas, o que evidencia a capacidade de abstração da rede neural nos diversos domínios de patologia retiniana. É crucial ressaltar que o conjunto de dados empregado neste estudo não foi submetido a auditoria. Nesse contexto, é possível que haja imagens rotuladas incorretamente. Além disso, como se trata de um banco de dados novo e exclusivo, não foram realizadas comparações com outros estudos previamente publicados.

Um aspecto notável a ser considerado é a ocorrência de sobreposições na classificação das imagens, em que uma imagem enviada ao *pipeline* pode receber pontuações elevadas para mais de uma condição patológica. Um exemplo ilustrativo é o caso de uma hemorragia decorrente de retinopatia diabética que também pode ser classificada como vaso patológico. Essa situação realça a complexidade inerente ao processo de classificação de patologias retinianas e demanda um olhar crítico e atento para garantir a correta interpretação dos resultados obtidos com o modelo proposto.

6.1.1 Modelo de Conhecimento Transferido

Foram usadas as mesmas imagens de treino, validação e teste para a concepção dos modelos de cada domínio patológico, com e sem aprendizado por transferência, também foi adotada a mesma proporção de 80%, 10% e 10% para a divisão das imagens entre os grupos aludidos. O número total de imagens usadas para a elaboração do modelo foi de 7.850, segmentado em 3.581 imagens de retina saudável, 1.474 de catarata, 986 de RD, 763 de escavação anormal e 1.046 de vasos anormais. Como mencionado anteriormente, a arquitetura da CNN é a mesma usada com outros conjuntos de dados, porém, com uma taxa de aprendizado no valor de 0.0001 que atua em todas as camadas treináveis do modelo que usufrui da transferência de conhecimento advinda do modelo treinado em imagens de alta qualidade.

Para elucidar a importância do método proposto de aprendizado por transferência, foram comparados os resultados de especificidade, sensibilidade e acurácia obtidos com o modelo de conhecimento transferido e o modelo treinado apenas utilizando imagens de baixa qualidade, conforme apresentado na Tabela 6.

Tabela 6 – Comparação entre os modelos sem e com aprendizado transferido.

Modelo	Tipo	Espe	Sens	Acurácia
Catarata	Sem AT	81.4%	88.6%	85.3%
	Com AT	82.9%	90.1%	87.4%
RD	Sem AT	72.0%	85.5%	75.4%
	Com AT	88.6%	93.2%	90.8%
Escavação	Sem AT	81.0%	90.2%	86.6%
	Com AT	80.9%	91.1%	87.5%
Vasos	Sem AT	68.2%	80.7%	73.9%
	Com AT	74.1%	81.3%	79.1%
Média	Sem AT	75.7%	86.3%	80.3%
	Com AT	81.6%	88.9%	86.2%

Dado o fato de que esta pesquisa propõe uma abordagem capaz de auxiliar na tomada de decisões por oftalmologistas, é possível afirmar que a transferência de conhecimento de um tipo de conjunto de dados para outro (de alta para baixa qualidade) mostra uma melhoria considerável no desempenho dos modelos. Embora a abordagem proposta tenha alcançado uma precisão média superior a 86%, é difícil compará-la ao estado da arte devido ao uso de dados privados. O mesmo problema ocorre ao analisar a literatura, uma vez que alguns trabalhos não utilizaram conjuntos de dados públicos (GULSHAN et al., 2016), (GAO et al., 2018; RAN et al., 2018; LI et al., 2018; ALBADAWI; FRAZ, 2018). Além disso, não há consenso em relação às métricas de avaliação. Por essas razões, na sequência, a abordagem proposta foi testada em conjuntos de dados públicos de catarata, retinopatia diabética (DR) e glaucoma. Assim, os resultados puderam ser adequadamente comparados com os de outros trabalhos.

Dentre os resultados, o método de transferência de aprendizado mostrou uma melhoria expressiva na classificação de retinopatia diabética RD com um aumento na precisão de 75,4% sem aprendizado por transferência para 90,8% com aprendizado por transferência. Este resultado sugere que o aprendizado por transferência pode contribuir significativamente para melhorar o desempenho das redes neurais convolucionais na classificação de imagens retinianas de baixa qualidade, auxiliando assim na identificação e diagnóstico de várias patologias retinianas.

6.2 Bases de Dados Públicas

Partindo do pressuposto da discussão sobre a dificuldade de comparação dos resultados de levantamentos realizados em dados privados, e que alguns desses trabalhos não são testados em imagens públicas, não há consenso quanto às métricas, somado ao fato de que a classificação das imagens varia de acordo com a interpretação de cada profissional (RAMAN et al., 2019).

Para verificar a assertividade do modelo em dados públicos, foram utilizados conjuntos de dados comumente encontrados na literatura para analisar RD, catarata e glaucoma. Dois conjuntos de dados de DR foram considerados para realizar as comparações (Messidor-2 (DECENCIÈRE et al., 2014) e EyePACS (KAGGLE, 2015)). Além disso, o conjunto de dados ODIR (KAGGLE, 2020) foi empregado para classificar catarata e o conjunto de dados REFUGE (ORLANDO et al., 2020) foi utilizado para glaucoma.

Vale ressaltar que o uso de conjuntos de dados públicos, em essência, é de suma importância para definir *benchmarks* para modelos preditivos. No entanto, pode-se notar que muitas dessas pesquisas selecionam dados e, portanto, alteram o rótulo de classe de algumas instâncias. Nesse sentido, é difícil estabelecer um *benchmark* adequado, uma vez que tais mudanças são apenas mencionadas (não detalhadas) nos artigos. Os detalhes e explicações das pesquisas que foram usadas como comparação estão presentes no Capítulo 2

6.2.1 Retinopatia Diabética

Conforme mencionado anteriormente, Messidor-2 e EyePACS foram considerados para avaliar o domínio da RD. O Messidor-2 possui 1.748 imagens, enquanto o EyePACS é composto por 9.963 imagens. Ambos tiveram suas classes divididas em DR referenciável e não referenciável.

Uma meta-análise de abordagens baseadas em aprendizado profundo para DR foi realizada por (WU et al., 2021) com o objetivo de listar seus tipos e desempenhos. Nesse sentido, a Tabela 7 mostra a comparação entre o presente trabalho e outros trabalhos com o mesmo escopo. É importante ressaltar que os trabalhos condensados nesta comparação

usam a mesma rotulação de dados, sem uma curadoria especializada feita por outros profissionais, já que o processo de curadoria é subjetivo e difícil de padronizar.

Tabela 7 – Comparação com os artigos que compõem o estado da arte sobre conjuntos de dados públicos de RD.

Pesquisa	Dataset	AUC (95% CI)	Sens.	Esp.
VGG16 Proposta	Eyepacs (Kaggle)	0.951 (0.948 - 0.953)	91.3	84.7
	Messidor2	0.953 (0.948 - 0.955)	91.5	84.9
Voets, Møllersen e Bongo (2019)*	Eyepacs (Kaggle)	0.951 (0.947 - 0.956)	90.6	84.7
	Messidor2	0.853 (0.835-0.871)	81.8	71.2
Xie et al. (2020)	Eyepacs (Kaggle)	-	83.2	89.03
	Messidor2	-	-	-
Zago et al. (2020)*	Eyepacs (Kaggle)	0.764 (0.756 - 0.773)	91.1	50
	Messidor2	0.912 (0.897–.928)	94	50

*Pesquisas que utilizam ponto operacional para sensibilidade e especificidade.

6.2.2 Catarata

No contexto da catarata, foi escolhido o conjunto de dados ODIR, composto por 1.400 imagens. Visto que os trabalhos em estado da arte obtiveram precisões superiores a 0.975, foi possível verificar que a presente pesquisa demonstra a aderência do modelo proposto para auxiliar na classificação de catarata, conforme apresentado na Tabela 8.

Neste trabalho, o desempenho do modelo proposto para classificação de catarata foi alinhado com abordagens similares encontradas na literatura, como Junayed et al. (2021) com uma rede neural customizada chamada de CataractNet, Khan et al. (2021) utilizando uma arquitetura VGG19, e Ali, Al-Sultan e Al-Saadi (2022) com um modelo de CNN profunda. A eficácia do método proposto, demonstrada na comparação dos resultados obtidos com os estudos supracitados, reforça a sua capacidade de auxiliar na classificação de catarata.

Tabela 8 – Comparação com os artigos de estado-da-arte sobre conjuntos de dados públicos de catarata.

Referência	Dataset	Acurácia
Junayed et al. (2021)	ODIR	0.986
Khan et al. (2021)	ODIR	0.975
Ali, Al-Sultan e Al-Saadi (2022)	ODIR	1.000
VGG16 Proposta	ODIR	0.993

6.2.3 Glaucoma

Em termos de glaucoma, é comum encontrar trabalhos sobre a segmentação do nervo óptico. No entanto, além dos valores de referência para a segmentação exata das esca-

vações, o conjunto de dados REFUGE (ORLANDO et al., 2020) também possui classificações de retinas saudáveis ou patológicas, neste caso, com glaucoma. O conjunto de dados é composto por 1.200 retinografias anotadas, das quais 121 correspondem a olhos com glaucoma. As retinografias neste conjunto de dados são centralizadas na mácula e possuem um tamanho de $1,634 \times 1,634$ ou $2,124 \times 2,056$ pixels.

Mais uma vez, o modelo preditivo proposto neste trabalho se mostrou adequado para o processo de classificação de imagens, como demonstrado na Tabela 9. Além disso, ao comparar a presente abordagem com trabalhos do estado da arte no diagnóstico de glaucoma, como Hervella et al. (2022) e Orlando et al. (2020), é possível observar que o desempenho do método proposto é satisfatório e alinhado aos resultados encontrados na literatura. Dessa forma, o modelo desenvolvido neste estudo pode ser considerado uma solução eficaz para auxiliar na identificação e diagnóstico de glaucoma em imagens retinianas.

Tabela 9 – Comparação com os artigos de estado-da-arte sobre conjunto de dados públicos de glaucoma.

Referência	Dataset	AUC
Hervella et al. (2022)	REFUGE	0.976
Orlando et al. (2020)	REFUGE	0.989
VGG16 Proposta	REFUGE	0.996

6.2.4 Análise Geral

Os resultados obtidos demonstram a robustez e adaptabilidade da abordagem proposta não apenas para o conjunto de dados proprietários, mas também para os conjuntos de dados públicos. Portanto, esses resultados endossam a possibilidade de modelos canônicos que atendem com maestria ao suporte à tomada de decisão em múltiplos domínios.

Com base nos resultados de múltiplos modelos de CNN, o pipeline proposto possibilitou a classificação das imagens de fundo de olho e, conseqüentemente, a classificação das probabilidades de ocorrência de patologias que apoiam a análise e inferência dos oftalmologistas.

Como mostrado na literatura relacionada, o processo de curadoria de dados (rotulação) contribui para melhorar o desempenho dos modelos preditivos. No entanto, nesta pesquisa, foi possível perceber que a etapa de processamento de dados é de fácil utilização e pode ser empregada para validar imagens antes de serem armazenadas em um banco de dados, uma vez que essas imagens devem estar dentro dos limites previamente definidos.

Além disso, é importante ressaltar que a abordagem proposta contribui significativamente para aumentar a precisão e a eficiência na identificação e diagnóstico de patologias oculares. Ao adotar essa técnica, os profissionais da área oftalmológica terão acesso a uma

ferramenta valiosa que, juntamente com seu conhecimento e experiência, permitirá uma abordagem mais assertiva e instrutiva tanto para o tratamento como para a prevenção de patologias relacionadas. Assim, espera-se que a adoção deste método resulte em melhorias na qualidade de vida dos pacientes afetados por essas condições.

6.3 Aspectos da Implementação do Modelo Proposto em Aplicação Web

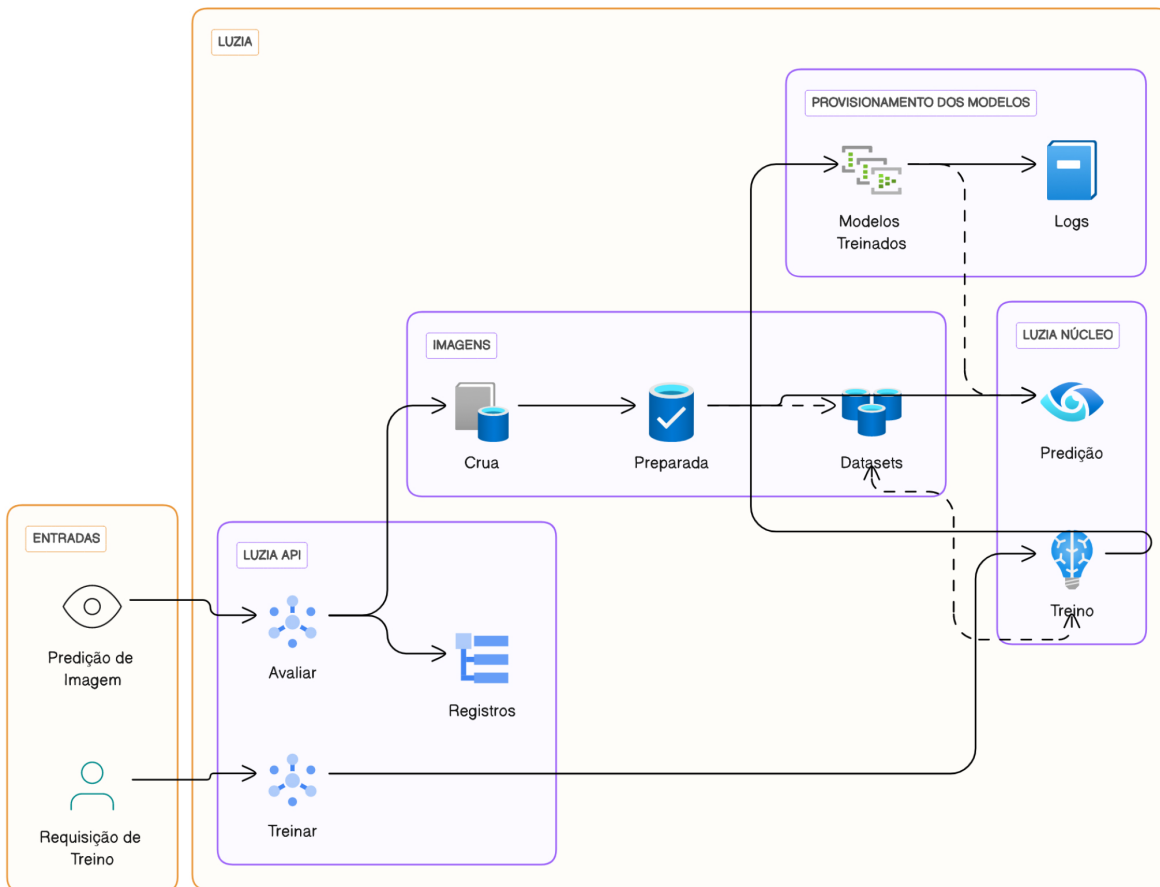
Nesta seção, é apresentado o desenvolvimento detalhado de um sistema que foi batizado como LuzIA, composto por três módulos interconectados, projetados especificamente para o treinamento, processamento e predição com visualização dos resultados dos modelos de aprendizado profundo aplicados às quatro patologias oculares abordadas nesta tese (RD, catarata, glaucoma e vasos patológicos).

Para elucidar o funcionamento da LuzIA no contexto de infraestrutura e proporcionar uma compreensão visual, é ressaltado na Figura 25 o fluxo dos dois processos centrais do sistema: treinamento dos modelos e predição de imagens. No que diz respeito ao treinamento, uma entidade pode solicitar à API o treinamento dos modelos. Nesse caso, o núcleo do sistema, que engloba algoritmos de aprendizado de máquina, emprega imagens dos conjuntos de dados para criar modelos na nuvem, gerando relatórios de desempenho para análise histórica. Esses modelos serão posteriormente aplicados no procedimento de predição. Relativamente ao procedimento de predição, os modelos são carregados na memória pela API, que também recebe uma imagem, realiza o pré-processamento e a envia para o conjunto de dados a fim de ser rotulada posteriormente. Por fim, as probabilidades e os respectivos Grad-CAMs são retornados ao solicitante.

6.3.1 Módulo de Treinamento e Otimização

Como feito no processo de treinamento previamente descrito (capítulo 5), foram utilizadas as mesmas estratégias para otimizar o desempenho dos modelos, tais como filtro de claridade, pré-processamento e aprendizado por transferência. Entretanto, em uma abordagem de cenário real com a possibilidade de inclusão de imagens no conjunto de treinamento, faz-se necessário práticas de ajustes de hiperparâmetros sazonais, como a taxa de aprendizagem, o tamanho do lote, *decay* e *momentum* para encontrar o ponto ótimo dos modelos. A Tabela 10 apresenta alguns exemplos de hiperparâmetros e seus respectivos intervalos de valores considerados para ajustes.

Figura 25 – Fluxo da Infraestrutura do Sistema Desenvolvido.



Elaborado pelo autor (2023).

Tabela 10 – Exemplos de hiperparâmetros e intervalos de valores.

Hiperparâmetro	Intervalo de valores
Taxa de aprendizagem	[0.0001, 0.01]
Tamanho do lote	[16, 64]
Taxa de <i>decay</i>	[1^{-2} , 1^{-9}]
Taxa de <i>momentum</i>	[0.1, 1.5]
Número de épocas	[10, 200]

6.3.2 Módulo de API

O módulo de API foi desenvolvido para aceitar um arquivo de imagem como entrada e retornar um objeto JSON contendo as previsões feitas pelos múltiplos modelos CNN pré-treinados, juntamente com suas visualizações Grad-CAM correspondentes. O *endpoint* é definido como uma função assíncrona para lidar eficientemente com solicitações concorrentes. Primeiramente, a função lê os bytes da imagem enviada e verifica se são válidos. Caso contrário, retorna uma mensagem de erro. Se os bytes estiverem corretos, a função cria um arquivo temporário para armazená-los e processa a imagem redimensionando-a

para 299×299 pixels. Após o processamento, o arquivo temporário é excluído.

Em seguida, o procedimento de avaliação da imagem é chamado para processar a lista de modelos pré-treinados e a imagem já adequada, retornando uma lista de previsões para cada modelo. A função itera pelas previsões e verifica se alguma delas possui pontuação de confiança maior que 0.75. Se sim, gera uma visualização Grad-CAM para aquele modelo usando uma função. O objeto de resposta é construído como um dicionário aninhado, onde o primeiro nível representa os nomes dos modelos (por exemplo, catarata, glaucoma), o segundo nível contém as pontuações das previsões e o terceiro nível inclui as visualizações Grad-CAM, se o limite de confiança for atingido.

Por fim, o objeto de resposta JSON é retornado ao cliente, contendo as previsões e visualizações Grad-CAM (se aplicável) para cada modelo. Este formato permite fácil integração com aplicativos web ou outras interfaces de usuário para exibição e análise.

É importante destacar a utilização de técnicas de melhoria de desempenho e escalabilidade, para isso, a API é projetada para funcionar em um ambiente Google Kubernetes Engine, que permite o gerenciamento eficiente de recursos de tempo de execução, autoescalabilidade e balanceamento de carga, tornando-se altamente responsivo mesmo em situações de alta demanda. Além disso, a padronização no formato das imagens recebidas, como tamanho e resolução, permite tornar o módulo factível, uma vez que as dimensões das imagens usadas como entrada são consideravelmente pequenas.

6.3.2.1 Geração do Grad-CAM

A geração da visualização Grad-CAM combina o mapa de calor com a imagem original e salva o resultado como uma string JPEG codificada em base64, que pode ser facilmente incorporada em aplicativos da web ou outras interfaces de usuário para exibição.

No primeiro estágio, o módulo de API cria uma imagem que representa as regiões de interesse através de um mapa de calor ao avaliar a imagem em um dos 4 modelos, mapeando assim a imagem de entrada para as ativações da última camada convolucional e predições de saída. Usando o GradientTape do TensorFlow, o gradiente da classe prevista com maior pontuação é calculado em relação às ativações da última camada convolucional. Os gradientes são combinados nas dimensões espaciais e usados para ponderar os canais da saída da última camada convolucional, com o mapa de calor resultante sendo normalizado.

Após ser normalizado, redimensionado e colorizado, o mapa de calor é sobreposto na imagem fornecida para a API, com uma taxa de transparência chamada de *alpha*, criando assim a combinação do mapa de calor e a imagem original.

6.3.3 Módulo de Predição e Visualização

Este módulo foi desenvolvido para ser um aplicativo web no formato Streamlit que permite que os usuários façam upload de imagens da retina e usem o *endpoint* da API

mencionado anteriormente para classificar as imagens e exibir os resultados, juntamente com as visualizações Grad-CAM.

A função principal dessa aplicação é a de converter a imagem de entrada em um fluxo de bytes usando a classe `io.BytesIO` do Python. Em seguida, envia uma requisição POST para o *endpoint* da API com os bytes da imagem no *payload* da requisição. Se o código de resposta for 200 (indicando sucesso), ela retorna a resposta JSON; caso contrário, levanta uma exceção com a mensagem de erro da API. O aplicativo Streamlit fornece uma interface de usuário para fazer upload de imagens da retina nos formatos PNG, JPG ou JPEG.


Por fim, o aplicativo itera pelos resultados das previsões e exibe as pontuações de confiança para cada tipo de doença. Se um mapa de ativação de classe (Class Activation Map - CAM) estiver disponível para uma doença específica, a imagem CAM é decodificada do formato base64 e exibida abaixo da pontuação da previsão correspondente. Se não houver CAM disponível, uma mensagem indicando sua indisponibilidade é mostrada. A Figura 26 e a Figura 27 mostram uma representação esquemática da interface de usuário do sistema.


Em suma, o sistema proposto atende às demandas de um cenário real tanto em termos de precisão na detecção de patologias oculares quanto na praticidade e apoio das inferências para os profissionais médicos envolvidos. Ao incorporar modernas técnicas de treinamento, API de alto desempenho e interfaces de usuário intuitivas, é esperado que essa implementação possa contribuir significativamente para o avanço da pesquisa e aplicação prática no campo do diagnóstico e tratamento de patologias oculares.

Figura 26 – Representação esquemática da interface de usuário do sistema Streamlit.

LuzIA - Retina Image Classifier

Upload a Retina Image

 Drag and drop file here
Limit 200MB per file • PNG, JPG, JPEG

 70sus_17500000020001.jpg 61.3KB ×



Uploaded Retina Image

Figura 27 – Representação esquemática da interface de usuário do sistema Streamlit.

Prediction Results

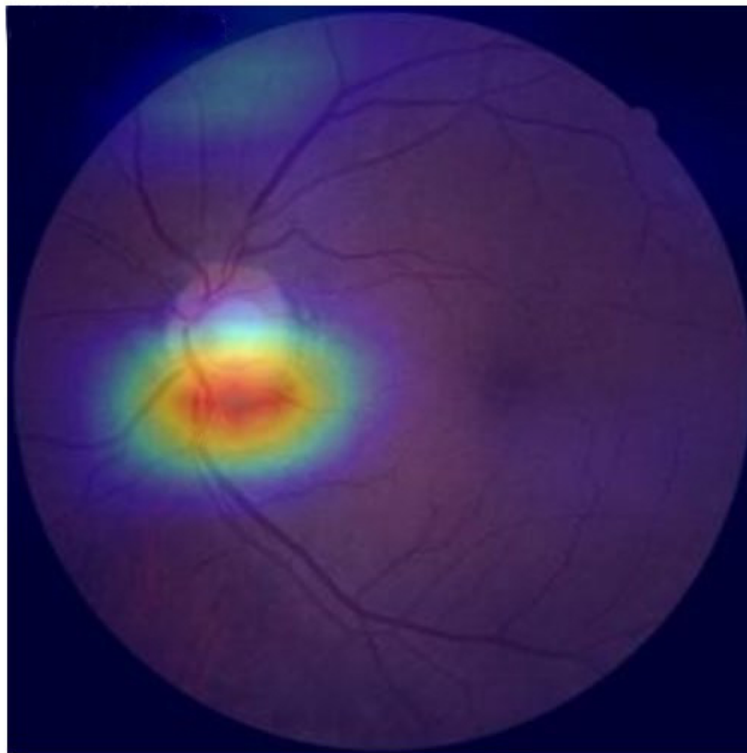
Cataract Prediction:

```
▼ {  
  "Cataract" : "0.11%"  
  "Normal" : "99.89%"  
}
```

No Cataract CAM available.

Abnormal excavation Prediction:

```
▼ {  
  "Abnormal" : "99.99%"  
  "Normal" : "0.01%"  
}
```



Abnormal excavation CAM

Referable DR Prediction:

```
▼ {  
  "Referable" : "53.37%"  
  "Normal" : "46.63%"  
}
```

No Referable DR CAM available.

Capítulo 7

Conclusões

Conforme mostrado no decorrer dessa tese, destaca-se que a mesma apresenta as seguintes contribuições: (1) a proposta de um conjunto de CNNs capazes de obter resultados comparáveis ao estado da arte em conjuntos de dados públicos; (2) a validação da estratégia de transferência de aprendizado proposta para melhorar o reconhecimento de condições e doenças oculares em imagens de baixa qualidade (produzidas por equipamentos de menor custo); (3) o desenvolvimento de uma abordagem escalável e replicável para tratar diferentes tipos de patologias retinianas; e (4) o projeto e implementação da metodologia proposta em uma aplicação web capaz de receber imagens e fornecer previsões sob demanda, além de um mapa de calor para aprimorar o suporte à tomada de decisões médicas.

A utilização de CNNs do tipo VGG16 para classificar patologias retinianas foi demonstrada como eficaz nesta tese, com resultados sugerindo que a abordagem proposta oferece uma base sólida para futuras pesquisas e aplicações. Ao implementar uma estratégia de transferência de aprendizado, a precisão na classificação foi efetivamente melhorada, especialmente ao trabalhar com dados de imagens de menor qualidade. Isso destaca a adaptabilidade dos modelos às condições de imagem subótimas, particularmente relevante para cenários clínicos com recursos limitados.

Abordar múltiplas patologias retinianas com uma abordagem escalável e replicável sugere que a presente metodologia tem potencial para ser estendida e modificada para enfrentar uma ampla gama de doenças e condições. Essa flexibilidade é essencial no campo médico, pois permite pesquisadores e profissionais adaptar e evoluir as técnicas de acordo com novos requisitos e avanços no entendimento das doenças.

As implicações práticas desta tese são ainda mais fortalecidas pelo desenvolvimento de uma aplicação web capaz de receber imagens, gerar classificações sob demanda e fornecer

informações complementares valiosas na forma de mapas de calor. Juntos, esses elementos melhoram a usabilidade e interpretabilidade das previsões do modelo, sendo um aspecto considerado crítico para os profissionais médicos confiarem e adotarem tais tecnologias em sua prática clínica.

Nesse sentido, pode-se dizer que essa tese deu passos significativos em enfrentar os desafios associados à classificação das patologias retinianas usando CNNs de última geração e técnicas de transferência de aprendizado. Ao englobar uma abordagem escalável para tratar várias patologias retinianas, implementar uma estratégia de transferência de aprendizado para minimizar os problemas de escassez de dados e criar uma aplicação viável, este trabalho demonstra o potencial de combinar metodologias avançadas com aplicações práticas na revolução do diagnóstico e tratamento de doenças retinianas. À medida que futuras pesquisas continuam a explorar e iterar sobre essas ideias, é esperado que essas contribuições tenham um impacto duradouro nas comunidades médicas e científicas, levando em última instância a um melhor atendimento e bem-estar geral dos pacientes.

Referências

- ABRAMOFF, M.; GARVIN, M.; SONKA, M. Retinal Imaging and Image Analysis. **IEEE Reviews in Biomedical Engineering**, v. 3, p. 169–208, 2010. ISSN 1937-3333.
- ABRAMOFF, M. D. et al. Automated early detection of diabetic retinopathy. **Ophthalmology**, Elsevier, v. 117, n. 6, p. 1147–1154, 2010.
- ACHARYA, R. et al. Imaging systems of human eye: a review. **Journal of medical systems**, Springer, v. 32, n. 4, p. 301–315, 2008.
- ADA. **Statics About Diabetes**. 2018. Disponível em: <<http://www.diabetes.org/diabetes-basics/statistics/>>.
- ALABBOUD, I. et al. New spectral imaging techniques for blood oximetry in the retina. In: OPTICAL SOCIETY OF AMERICA. **European Conference on Biomedical Optics**. [S.l.], 2007. p. 6631_22.
- ALBADAWI, S.; FRAZ, M. Arterioles and venules classification in retinal images using fully convolutional deep neural network. In: **International Conference on Image Analysis and Recognition**. [S.l.: s.n.], 2018. p. 659–668.
- ALI, H. H.; AL-SULTAN, A. Y.; AL-SAAD, E. H. Cataract disease detection used deep convolution neural network. In: **5th International Conference on Engineering Technology and its Applications (IICETA)**. [S.l.: s.n.], 2022. p. 102–108.
- ANANTHAPADMANABAN, K. R.; PARTHIBAN, G. Prediction of Chances - Diabetic Retinopathy Using Data Mining Classification Techniques. **Indian Journal of Science and Technology**, v. 7, n. 10, p. 1498–1503, 2014. ISSN 0974-5645.
- AWS, Google Cloud, Apache. **Data Management: AWS Glue, Google Cloud Dataflow, Apache NiFi**. 2023. Platform Documentation. Documentation of platforms can be found at the respective platform’s official website.
- AZURE, M. **What is MLOps?** 2023. <<https://azure.microsoft.com/en-us/services/machine-learning/mlops/>>.
- BAJWA, M. N. et al. Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. **BMC medical informatics and decision making**, BioMed Central, v. 19, n. 1, p. 1–16, 2019.

- BAKER, A. Crossing the quality chasm: a new health system for the 21st century. **BMJ: British Medical Journal**, BMJ Publishing Group LTD, v. 323, n. 7322, p. 1192, 2001.
- BAUDOIN, C.; LAY, B.; KLEIN, J. Automatic detection of microaneurysms in diabetic fluorescein angiography. **Revue d'épidémiologie et de santé publique**, v. 32, n. 3-4, p. 254–261, 1984.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. **Representation learning: A review and new perspectives**. [S.l.]: IEEE, 2013. v. 35. 1798–1828 p.
- BERNARDES, R.; SERRANHO, P.; LOBO, C. Digital ocular fundus imaging: a review. **Ophthalmologica**, Karger Publishers, v. 226, n. 4, p. 161–181, 2011.
- BLICK, K. E.; METZ, C. E. Probabilistic models and predictive measures for the clinical prediction problem. **Technometrics**, v. 20, n. 1, p. 7–18, 1978.
- BOSCO, A. et al. Retinopatia diabética. **Arquivos Brasileiros de Endocrinologia & Metabologia**, v. 49, n. 2, p. 217–227, 2005. ISSN 0004-2730. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27302005000200007&lng=p>.
- BOURNE, R. et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. **The Lancet Global Health**, v. 9, n. 2, p. e130–e143, 2021. ISSN 2214109X.
- BRESSLER, N. M.; BRESSLER, S. B.; FINE, S. L. Age-related macular degeneration. **Survey of ophthalmology**, v. 32, n. 6, p. 375–413, 1988. ISSN 0039-6257. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/2457955>>.
- BURLINA, P. M. et al. Use of deep learning for detailed severity characterization and estimation of 5-year risk for age-related macular degeneration. In: SPRINGER. **Fetal, infant and ophthalmic medical image analysis**. [S.l.], 2017. p. 136–144.
- CANNY, J. A computational approach to edge detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 8, n. 6, p. 679–698, 1986.
- CHAN, H. P. et al. Deep Learning in Medical Image Analysis. **Advances in Experimental Medicine and Biology**, v. 1213, n. March, p. 3–21, 2020. ISSN 22148019.
- CHEN, X. et al. Glaucoma detection based on deep convolutional neural network. In: **2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. IEEE, 2015. v. 31, n. 11, p. 715–718. ISBN 978-1-4244-9271-8. ISSN 1004731X. Disponível em: <<https://ieeexplore.ieee.org/document/7318462/>>.
- CHING, T. et al. **Opportunities and obstacles for deep learning in biology and medicine**. [S.l.: s.n.], 2018. v. 15. ISSN 17425662. ISBN 0000000305396.
- CHRISTOPHER, M. et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. **Scientific reports**, Nature Publishing Group, v. 8, n. 1, p. 1–13, 2018.

COLEMAN, H. R. et al. Age-related macular degeneration. **Lancet (London, England)**, v. 372, n. 9652, p. 1835–45, 2008. ISSN 1474-547X. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0140673608617596>> <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2603424>> <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2603424/>>.

de Amorim Garcia Filho, C. et al. Optical coherence tomography. In: _____. **Retina Fifth Edition**. [S.l.]: Elsevier Inc., 2012. v. 1, p. 82–110. ISBN 9781455707379. Funding Information: Our own contributions to this work are based on projects (P7300-MED, P9781-MED, P10316-MED, and P14103-MED) financed by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) and project No. 7428 supported by the Austrian National Bank. We want to thank K.K. Bizheva, W. Drexler, M. Pircher, and R. Leitgeb for communicating their current research results. Copyright: Copyright 2021 Elsevier B.V., All rights reserved.

DECENCIÈRE, E. et al. Feedback on a publicly distributed image database: the messidor database. **Image Analysis & Stereology**, v. 33, n. 3, p. 231–234, 2014.

DEEPIKA, E.; MAHESWARI, S. Earlier glaucoma detection using blood vessel segmentation and classification. **Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018**, IEEE, n. Icisc, p. 484–490, 2018.

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: IEEE. **2009 IEEE conference on computer vision and pattern recognition**. [S.l.], 2009. p. 248–255.

DIMITROVA, G.; KATO, S. Color doppler imaging of retinal diseases. **Survey of ophthalmology**, Elsevier, v. 55, n. 3, p. 193–214, 2010.

DOE, J.; SMITH, J. A comparative study of deep learning models for medical image classification. **Computer Science and Artificial Intelligence Journal**, v. 25, n. 8, p. 123–136, 2021.

DONG, Y. et al. Classification of cataract fundus image based on deep learning. In: IEEE. **2017 IEEE International Conference on Imaging Systems and Techniques (IST)**. [S.l.], 2017. p. 1–5.

DOUGHERTY, E. R.; ASTOLA, J. An extension of non-linear median filters. **Signal Processing**, Elsevier, v. 24, n. 1, p. 53–62, 1990.

DREXLER, W.; FUJIMOTO, J. G. State-of-the-art retinal optical coherence tomography. **Progress in retinal and eye research**, Elsevier, v. 27, n. 1, p. 45–88, 2008.

ELADAWI, N. et al. Optical coherence tomography: A review. **Diabetes and Fundus OCT**, Elsevier, p. 191–221, 2020.

FAUW, J. D. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. **Nature medicine**, Nature Publishing Group, v. 24, n. 9, p. 1342–1350, 2018.

FERCHER, A. F. Optical Coherence Tomography. **Journal of Biomedical Optics**, v. 1, n. 2, p. 157–173, 1996.

- FEURER, M. et al. Auto-sklearn 2.0: The next generation. **arXiv preprint arXiv:1903.11614v2**, 2019.
- FRAZ, M. M. et al. Blood vessel segmentation methodologies in retinal images - A survey. **Computer Methods and Programs in Biomedicine**, Elsevier Ireland Ltd, v. 108, n. 1, p. 407–433, 2012. ISSN 01692607. Disponível em: <<http://dx.doi.org/10.1016/j.cmpb.2012.03.009>>.
- FRENKEL, J. K.; OTHERS. Toxoplasmosis in human beings. **Journal of the American Veterinary Medical Association**, v. 196, n. 2, p. 240–248, 1990.
- GABBAY, R. A.; SIVARAJAH, S. Optical coherence tomography-based continuous noninvasive glucose monitoring in patients with diabetes. **Diabetes technology & therapeutics**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 . . . , v. 10, n. 3, p. 188–193, 2008.
- GAO, X.; LIN, S.; WONG, T. Y. Automatic feature learning to grade nuclear cataracts based on deep learning. **IEEE Transactions on Biomedical Engineering**, v. 62, n. 11, p. 2693–2701, 2015. ISSN 15582531.
- GAO, Z. et al. Diagnosis of diabetic retinopathy using deep neural networks. **IEEE Access**, IEEE, v. 7, p. 3360–3370, 2018.
- GARGEYA, R.; LENG, T. Automated Identification of Diabetic Retinopathy Using Deep Learning. **Ophthalmology**, American Academy of Ophthalmology, v. 124, n. 7, p. 962–969, 2017. ISSN 15494713. Disponível em: <<http://dx.doi.org/10.1016/j.ophtha.2017.02.008>>.
- GIRSHICK, R. Fast r-cnn. In: **Proceedings of the IEEE international conference on computer vision**. [S.l.: s.n.], 2015.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. [S.l.]: Prentice Hall, 2002.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. **Healthcare Informatics Research**, v. 22, n. 4, p. 351–354, 2016. ISSN 0028-0836. Disponível em: <<https://books.google.com/books?hl=en&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Ian+Goodfellow+and+Yoshua+Bengio+and+Aaron+Courville&ots=MLU59rozNU&sig=YHYf6i>>.
- GOODFELLOW, I. et al. Generative adversarial networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2014. p. 2672–2680.
- GOODMAN, S. N. An alternative to the use of the likelihood ratio test for the determination of the usefulness of a diagnostic test. In: **Annual Meeting of the American Statistical Association**. [S.l.: s.n.], 1979. p. 1–24.
- GOOGLE. **AutoML**. 2023. Google Cloud Documentation. Disponível em: <<https://cloud.google.com/automl>>.
- GRAVES, A.; MOHAMED, A.-r.; HINTON, G. Speech recognition with deep recurrent neural networks. **2013 IEEE International Conference on Acoustics, Speech and Signal Processing**, p. 6645–6649, 2013.

- GULSHAN, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. **Jama**, American Medical Association, v. 316, n. 22, p. 2402–2410, 2016.
- GUO, L. et al. A computer-aided healthcare system for cataract classification and grading based on fundus image analysis. **Computers in Industry**, Elsevier, v. 69, p. 72–80, 2015.
- GUPTA, M. et al. Retinal anatomy and pathology. **Developments in ophthalmology**, v. 55, p. 7–17, 10 2015.
- HASAN, D. A. et al. Machine learning-based diabetic retinopathy early detection and classification systems-a survey. In: IEEE. **2021 1st Babylon International Conference on Information Technology and Science (BICITS)**. [S.l.], 2021. p. 16–21.
- HE, K. et al. Deep residual learning for image recognition. **Proceedings of the IEEE conference on computer vision and pattern recognition**, p. 770–778, 2016.
- HEIJL, A. et al. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. **Archives of ophthalmology**, American Medical Association, v. 120, n. 10, p. 1268–1279, 2002.
- HEMELINGS, R. et al. Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. **Acta Ophthalmologica**, v. 98, n. 1, p. e94–e100, 2020. ISSN 17553768.
- HERVELLA, Á. S. et al. End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images. **Applied Soft Computing**, v. 116, p. 108347, 2022. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494621011303>>.
- HINTON, G. E. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. **IEEE Signal Processing Magazine**, v. 29, n. 6, p. 82–97, 2012.
- HINTON, G. E.; SALAKHUTDINOV, R. R. Reducing the dimensionality of data with neural networks. **Science**, American Association for the Advancement of Science, v. 313, n. 5786, p. 504–507, 2006.
- HUANG, D. et al. Optical coherence tomography. **science**, American Association for the Advancement of Science, v. 254, n. 5035, p. 1178–1181, 1991.
- JAIN, A.; SHARMA, S. Continuous integration and deployment for machine learning models. In: IEEE. **2019 IEEE International Conference on Big Data, Big Data 2019**. [S.l.], 2019. p. 5021–5023.
- JIAO, S. et al. Photoacoustic ophthalmoscopy for in vivo retinal imaging. **Optics express**, Optical Society of America, v. 18, n. 4, p. 3967–3972, 2010.
- JONAS, J. B. et al. Human optic nerve fiber count and optic disc size. **Investigative ophthalmology & visual science**, The Association for Research in Vision and Ophthalmology, v. 33, n. 6, p. 2012–2018, 1992.

- JUNAYED, M. S. et al. Cataractnet: An automated cataract detection system using deep learning for fundus images. **IEEE Access**, IEEE, v. 9, p. 128799–128808, 2021.
- KAGGLE. **Diabetic Retinopathy Detection**. [S.l.], 2015. <<https://www.kaggle.com/c/diabetic-retinopathy-detection>>.
- _____. **Ocular Disease Intelligent Recognition (ODIR)**. [S.l.], 2020. <<https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k>>.
- KALLIAMVAKOU, E. et al. The promise of research on open source development. In: **ACM. Proceedings of the 36th International Conference on Software Engineering**. [S.l.], 2014. p. 883–894.
- KEEL, S. et al. Visualizing Deep Learning Models for the Detection of Referable Diabetic Retinopathy and Glaucoma. **JAMA Ophthalmology**, v. 137, n. 3, p. 288–292, 2019. ISSN 21686165.
- KHAN, M. S. M. et al. Cataract detection using convolutional neural network with vgg-19 model. In: **IEEE World AI IoT Congress (AIIoT)**. [S.l.: s.n.], 2021. p. 0209–0212.
- KLEESIEK, J. et al. Deep mri brain extraction: A 3d convolutional neural network for skull stripping. **NeuroImage**, Elsevier, v. 129, p. 460–469, 2016.
- KLONOFF, D. C.; SCHWARTZ, D. M. An economic analysis of interventions for diabetes. **Diabetes Care**, v. 23, n. 3, p. 390–404, mar 2000. ISSN 0149-5992. Disponível em: <<http://care.diabetesjournals.org/cgi/doi/10.2337/diacare.23.3.390>>.
- KOLB, H. How the retina works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. **American scientist**, JSTOR, v. 91, n. 1, p. 28–35, 2003.
- _____. Simple anatomy of the retina. 2012.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105.
- LAMIREL, C.; NEWMAN, N. J.; BIOUSSE, V. Optical coherence tomography (oct) in optic neuritis and multiple sclerosis. **Revue neurologique**, Elsevier, v. 166, n. 12, p. 978–986, 2010.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. ISSN 14764687.
- LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. **Neural Computation**, MIT Press, v. 1, n. 4, p. 541–551, 1989.
- LENS, A.; NEMETH, S. C.; LEDFORD, J. K. **Ocular anatomy and physiology**. [S.l.]: Slack Incorporated, 2008.
- LI, F. et al. Automatic differentiation of glaucoma visual field from non-glaucoma visual field using deep convolutional neural network. **BMC Medical Imaging**, Springer, v. 18, n. 1, p. 1–7, 2018.

- LI, Y.; CHENG, H.; DUONG, T. Q. Blood-flow magnetic resonance imaging of the retina. **Neuroimage**, Elsevier, v. 39, n. 4, p. 1744–1751, 2008.
- LIM, G. et al. Transformed Representations for Convolutional Neural Networks in Diabetic Retinopathy Screening. In: **AAAI Workshop: Modern Artificial Intelligence for Health Analytics**. [S.l.: s.n.], 2014.
- LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical Image Analysis**, Elsevier B.V., v. 42, n. December 2012, p. 60–88, 2017. ISSN 13618423.
- MAJI, D.; SEKH, A. A. Automatic grading of retinal blood vessel in deep retinal image diagnosis. **Journal of Medical Systems**, Springer, v. 44, n. 10, p. 1–14, 2020.
- MATSUI, M. et al. A study on automatic and quantitative diagnosis of fundus photographs. i. detection of contour line of retinal blood vessel images on color fundus photographs (author's transl). **Nippon Ganka Gakkai Zasshi**, v. 77, n. 8, p. 907–918, 1973.
- MCGLYNN, E. A. et al. The quality of health care delivered to adults in the United States. **New England journal of medicine**, Mass Medical Soc, v. 348, n. 26, p. 2635–2645, 2003.
- MEAD, P. S. et al. Food-related illness and death in the United States. **Emerging infectious diseases**, Centers for Disease Control and Prevention, v. 5, n. 5, p. 607, 1999.
- METZ, C. E. Basic principles of roc analysis. **Seminars in Nuclear Medicine**, v. 8, n. 4, p. 283–298, 1978.
- MIKOLOV, T. et al. Efficient estimation of word representations in vector space. **arXiv preprint arXiv:1301.3781**, 2013.
- MOKWA, N. F. et al. Grading of age-related macular degeneration: Comparison between color fundus photography, fluorescein angiography, and spectral domain optical coherence tomography. **Journal of Ophthalmology**, v. 2013, 2013. ISSN 2090004X.
- MOURA, L. et al. Waterborne toxoplasmosis, Brazil, from field to gene. **Emerging infectious diseases**, Centers for Disease Control and Prevention, v. 12, n. 2, p. 326, 2006.
- MURATA, N. et al. Assessment of objective ocular cyclodeviation under monocular condition and binocular condition using fundus photography. **BMJ Open Ophthalmology**, BMJ Specialist Journals, v. 5, n. 1, p. e000595, 2020.
- ORÉFICE, F.; FILHO, R. C.; BARBOZA, A. Toxoplasmose ocular adquirida/Toxoplasmose ocular pós-natal. **Rev Bras. Oftalmol.**, v. 69, n. 3, p. 184–207, 2010. ISSN 00347280. Disponível em: <<http://www.scielo.br/pdf/rbof/v69n3/a09v69n3>>.
- ORLANDO, J. I. et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. **Medical Image Analysis**, v. 59, p. 101570, 2020.

OTTAIANO, J. A. A. et al. **As Condições da saúde Ocular no Brasil 2019**. [s.n.], 2019. 104 p. ISBN 9788562109041. Disponível em: <http://www.cbo.com.br/novo/publicacoes/condicoes{_}saude{_}ocular{_}brasi>.

OYSTER, C. W. *The human eye*. **Sunderland, MA: Sinauer**, 1999.

PEREIRA, S. et al. Brain tumor segmentation using convolutional neural networks in mri images. **IEEE transactions on medical imaging**, IEEE, v. 35, n. 5, p. 1240–1251, 2016.

PRATAP, T.; KOKIL, P. Computer-aided diagnosis of cataract using deep transfer learning. **Biomedical Signal Processing and Control**, Elsevier, v. 53, p. 101533, 2019.

PROVIS, J. M. et al. Anatomy and development of the macula: specialisation and the vulnerability to macular degeneration. **Clinical and Experimental Optometry**, Wiley Online Library, v. 88, n. 5, p. 269–281, 2005.

QUMMAR, S. et al. A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection. **IEEE Access**, IEEE, v. 7, p. 150530–150539, 2019. ISSN 21693536.

RAJ, V. A.; NITHIYANANTHAM, R. A. U. Glaucoma detection by segmenting superpixels from fundus depth imaging. In: IEEE. **2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)**. [S.l.], 2018. p. 1–5.

RAJINI, N.; BHASKARAN, P. Performance evaluation of image classification using multilayer neural network. **International Journal of Computer Applications**, v. 96, n. 18, p. 1–5, 2014.

RAMAN, R. et al. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. **Eye (Basingstoke)**, Springer US, v. 33, n. 1, p. 97–109, 2019. ISSN 14765454. Disponível em: <<http://dx.doi.org/10.1038/s41433-018-0269-y>>.

RAN, J. et al. Cataract detection and grading based on combination of deep convolutional neural network and random forests. In: **International Conference on Network Infrastructure and Digital Content (IC-NIDC)**. [S.l.: s.n.], 2018. p. 155–159.

REDMON, J. et al. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 779–788.

REMESEIRO, B.; MENDONÇA, A. M.; CAMPILHO, A. Automatic classification of retinal blood vessels based on multilevel thresholding and graph propagation. **The Visual Computer**, Springer, v. 37, n. 6, p. 1247–1261, 2021.

ROSE, T. A.; CHOI, J. W. Intravenous Imaging Contrast Media Complications: The Basics That Every Clinician Needs to Know. **American Journal of Medicine**, Elsevier Ltd, v. 128, n. 9, p. 943–949, 2015. ISSN 15557162. Disponível em: <<http://dx.doi.org/10.1016/j.amjmed.2015.02.018>>.

ROUX, L. et al. Mitosis detection in breast cancer histological images an icpr 2012 contest. **Journal of pathology informatics**, Wolters Kluwer–Medknow Publications, v. 4, 2013.

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

RYAN, S.; SCHACHAT, A.; WILKINSON, C. **Retina**. 4th editio. ed. [S.l.]: Mosby, 2006. 3104 p. ISBN 9780323025980.

SAINÉ, P. J.; TYLER, M. E. **Ophthalmic photography: retinal photography, angiography, and electronic imaging**. [S.l.]: Butterworth-Heinemann Boston, 2002. v. 132.

SANDLER, M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, p. 4510–4520, 2018.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE International Conference on Computer Vision**. [S.l.: s.n.], 2017. p. 618–626.

SHANKLE, J. Ophthalmic photography: retinal photography, angiography, and electronic imaging. **Survey of Ophthalmology**, v. 49, n. 2, p. 264, mar 2004. ISSN 00396257. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0039625703001759>>.

SHIN, H. C. et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. **IEEE Transactions on Medical Imaging**, v. 35, n. 5, p. 1285–1298, 2016. ISSN 1558254X.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SNELL, R. S.; LEMP, M. A. **Clinical anatomy of the eye**. [S.l.]: John Wiley & Sons, 2013.

SPAIDE, R. F.; KLANCNIK, J. M.; COONEY, M. J. Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography. **JAMA Ophthalmology**, v. 133, n. 1, p. 45–50, 2015. ISSN 21686165.

SU, Y.-T. et al. Spatiotemporal joint mitosis detection using cnn-lstm network in time-lapse phase contrast microscopy images. **IEEE Access**, IEEE, v. 5, p. 18033–18041, 2017.

SWICK, C. T.; JURIS, P. C. Selecting a criterion for use in medicine: A decision analysis. **Medical Decision Making**, v. 14, n. 4, p. 456–463, 1994.

SZEGEDY, C.; IOFFE, S.; VANHOUCHE, V. Inception-v4, inception-resnet and the impact of residual connections on learning. **Proceedings of the AAAI Conference on Artificial Intelligence**, v. 30, n. 1, 2016.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. **Proceedings of the IEEE conference on computer vision and pattern recognition**, p. 2818–2826, 2016.

TAJBAKSHI, N. et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. **Medical Image Analysis**, Elsevier, v. 63, p. 101694, 2020.

TANG, A. et al. Canadian association of radiologists white paper on artificial intelligence in radiology. **Canadian Association of Radiologists Journal**, v. 69, n. 2, p. 120–135, 2018.

TIELSCH, J. M. et al. A population-based evaluation of glaucoma screening: the Baltimore Eye Survey. **American journal of epidemiology**, Oxford University Press, v. 134, n. 10, p. 1102–1110, 1991.

TING, D. S. W. et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. **JAMA - Journal of the American Medical Association**, v. 318, n. 22, p. 2211–2223, 2017. ISSN 15383598.

TORRE-DÍEZ, I. de la et al. Decision support systems and applications in ophthalmology: literature and commercial review focused on mobile apps. **Journal of medical systems**, Springer, v. 39, n. 1, p. 174, 2015.

TULDER, G. van; BRUIJNE, M. de. Combining generative and discriminative representation learning for lung ct analysis with convolutional restricted boltzmann machines. **IEEE transactions on medical imaging**, IEEE, v. 35, n. 5, p. 1262–1272, 2016.

VOETS, M.; MØLLERSEN, K.; BONGO, L. A. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. **PLoS ONE**, v. 14, n. 6, p. 1–11, 2019. ISSN 19326203.

WEBSTER, J. P. Review of "Toxoplasmosis of Animals and Humans (Second Edition)" by J.P. Dubey. **Parasites & Vectors**, v. 3, n. 1, p. 112, 2010. ISSN 1756-3305. Disponível em: <<http://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-3-112>>.

WHO. **Diabetes**. 2018. Disponível em: <<http://www.who.int/en/news-room/fact-sheets/detail/diabetes>>.

_____. **World report on vision**. [S.l.: s.n.], 2019. v. 214. 1–160 p. ISBN 9789241516570.

_____. **Blindness and vision impairment**. 2021.

WILSON, E. B. Probable inference, the law of succession, and statistical inference. **Journal of the American Statistical Association**, v. 22, n. 158, p. 209–212, 1927.

WONG, T. Y.; BRESSLER, N. M. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. **Jama**, American Medical Association, v. 316, n. 22, p. 2366–2367, 2016.

WONG, T. Y. et al. Prospective cohort study of retinal vessel diameters and risk of hypertension. **bmj**, British Medical Journal Publishing Group, v. 329, n. 7457, p. 79, 2004.

WU, J. H. et al. Performance and limitation of machine learning algorithms for diabetic retinopathy screening: Meta-analysis. **Journal of Medical Internet Research**, v. 23, n. 7, p. 1–15, 2021. ISSN 14388871.

- XIE, F.; KANG, J.; CHEUNG, S. Multi-stage container deployment of machine learning applications. **Journal of Web Engineering**, Rinton Press, v. 19, n. 1-2, p. 169–186, 2020.
- XIE, L. et al. Towards implementation of AI in New Zealand national diabetic screening program: Cloud-based, robust, and bespoke. **PLoS ONE**, v. 15, n. 4, p. 1–18, 2020. ISSN 19326203. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0225015>>.
- XU, X. et al. An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image. **Computer methods and programs in biomedicine**, Elsevier, v. 141, p. 3–9, 2017.
- YANG, J.-J. et al. Exploiting ensemble learning for automatic cataract detection and grading. **Computer methods and programs in biomedicine**, Elsevier, v. 124, p. 45–57, 2016.
- YANNUZZI, L. A. et al. Ophthalmic fundus imaging: today and beyond. **American journal of ophthalmology**, Elsevier, v. 137, n. 3, p. 511–524, 2004.
- YOU DEN, W. J. Index for rating diagnostic tests. **Cancer**, v. 3, n. 1, p. 32–35, 1950.
- YU, Y. et al. Deep transfer learning for modality classification of medical images. **Information (Switzerland)**, v. 8, n. 3, 2017. ISSN 20782489.
- ZAGO, G. T. et al. Diabetic retinopathy detection using red lesion localization and convolutional neural networks. **Computers in Biology and Medicine**, Elsevier Ltd, v. 116, p. 103537, 2020. ISSN 18790534. Disponível em: <<https://doi.org/10.1016/j.compbiomed.2019.103537>>.
- ZAPATA, M. A. et al. Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma. **Clinical Ophthalmology (Auckland, NZ)**, Dove Press, v. 14, p. 419, 2020.
- ZHANG, S. et al. Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network. **IEEE Access**, IEEE, v. 7, p. 57561–57573, 2019.
- ZHANG, W. et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. **NeuroImage**, Elsevier, v. 108, p. 214–224, 2015.
- _____. Screening for diabetic retinopathy using a portable, noncontact, nonmydriatic handheld retinal camera. **Journal of diabetes science and technology**, SAGE Publications Sage CA: Los Angeles, CA, v. 11, n. 1, p. 128–134, 2017.
- ZHOU, B. et al. Learning Deep Features for Discriminative Localization. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, v. 2016-December, p. 2921–2929, 2016. ISSN 10636919.
- ZILLY, J.; BUHMANN, J. M.; MAHAPATRA, D. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. **Computerized Medical Imaging and Graphics**, Elsevier Ltd, v. 55, p. 28–41, 2017. ISSN 18790771. Disponível em: <<http://dx.doi.org/10.1016/j.compmedimag.2016.07.012>>.