

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Marcelo Colares da Silva

**Deteccção da Retinopatia Diabética Através da Classificação de
Retinografias: Uma Abordagem Baseada em *Deep Learning* e
Inteligência Artificial Explicável.**

Fortaleza – CE

7 de março de 2025

Marcelo Colares da Silva

**Detecção da Retinopatia Diabética Através da
Classificação de Retinografias: Uma Abordagem Baseada
em *Deep Learning* e Inteligência Artificial Explicável.**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Auxílio ao diagnóstico médico

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho

Coorientador: Prof. Dr. Róger Moura Sarmiento

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

7 de março de 2025

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

- S586d Silva, Marcelo Colares da.
Detecção da Retinopatia Diabética Através da Classificação de Retinografias: Uma Abordagem Baseada em Deep Learning e Inteligência Artificial Explicável / Marcelo Colares da Silva. - 2025.
82 f.
- Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2025.
Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho.
Coorientação: Prof. Dr. Róger Moura Sarmento.
1. Deep learning. 2. Retinopatia diabética. 3. Inteligência artificial explicável. I. Título.
CDD 004
-



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
Av. 13 de maio, 2081 - Bairro Benfica - CEP 60040-531 - Fortaleza - CE - www.ifce.edu.br

FOLHA DE APROVAÇÃO

Processo: 23256.000734/2025-59

Interessado: Marcelo Colares da Silva

MARCELO COLARES DA SILVA

Detecção da Retinopatia Diabética Através da Classificação de Retinografias: Uma Abordagem Baseada em Deep Learning e Inteligência Artificial Explicável

Dissertação apresentada ao Curso de Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) Campus de Fortaleza, como requisito parcial para obtenção do Título de Mestre em Ciência da Computação.

Aprovada em: 29/01/2025.

BANCA EXAMINADORA

Prof. Dr. Pedro Pedrosa Rebouças Filho (Orientador)

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Prof. Dr. Róger Moura Sarmento (Coorientador)

Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE

Profa. Dra. Suane Pires Pinheiro da Silva

Centro Universitário Farias Brito

Prof. Dr. Solon Alves Peixoto

Universidade Federal do Ceará - UFC



Documento assinado eletronicamente por **Roger Moura Sarmento, Professor do Ensino Básico, Técnico e Tecnológico**, em 30/01/2025, às 10:19, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Pedro Pedrosa Rebouças Filho, Professor do Ensino Básico, Técnico e Tecnológico**, em 30/01/2025, às 10:22, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Solon Alves Peixoto, Usuário Externo**, em 30/01/2025, às 18:04, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Suane Pires Pinheiro da Silva, Usuário Externo**, em 31/01/2025, às 08:19, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade do documento pode ser conferida no site https://sei.ifce.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 informando o código verificador **6997348** e o código CRC **D9C9BDF1**.

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

Agradecimentos

A realização desta dissertação só foi possível graças ao apoio e à dedicação de pessoas que estiveram ao meu lado nessa trajetória.

Agradeço, em primeiro lugar, ao meus orientador, Prof. Dr. Pedro Pedrosa Rebouças Filho e ao meu coorientador Prof. Dr. Róger Moura Sarmento, que desempenharam um papel fundamental no desenvolvimento deste trabalho.

Aos meus pais, sou eternamente grato por serem minha base em todos os momentos. Sua confiança e apoio incondicional foram a força que me manteve firme nos desafios enfrentados.

À minha namorada, Maria Clara Alves Sena, meu sincero agradecimento por sua paciência, compreensão e por sempre me motivar a dar o meu melhor.

Não poderia deixar de mencionar meu colega, Caio Marques, cuja parceria foi indispensável em diversas etapas deste trabalho. Também agradeço à Prof^a Dr^a Suane Pires Pinheiro da Silva por todo o apoio prestado nas revisões dos artigos produzidos ao longo do mestrado.

“Seu tempo é limitado, então não o desperdice vivendo a vida de outra pessoa.”
(Steve Jobs)

Resumo

A retinopatia diabética (RD) é uma complicação ocular comum em pessoas com diabetes, causada por danos nos vasos sanguíneos da retina. Ela pode levar à perda de visão se não for tratada, sendo classificada em estágios que vão desde leve até proliferativa. A detecção de RD é geralmente realizada por meio de exames de fundo de olho, fundoscopia ou imagens de retina, com auxílio de técnicas como a segmentação de lesões para identificar alterações na retina. Os conjuntos de dados IDRiD e APTOS-2019 são utilizados para a segmentação de lesões provenientes da RD e detecção de retinopatia diabética, respectivamente. O conjunto de dados IDRiD possui imagens retinográficas capturadas de pacientes com RD, assim como as anotações pixel a pixel para lesões específicas, como microaneurismas, hemorragias, exsudatos duros e exsudatos moles. O conjunto de dados APTOS-2019 contém imagens de fundo de olho rotuladas com cinco categorias. Esse conjunto é amplamente utilizado para treinar modelos de aprendizado profundo, com o objetivo de detectar e classificar diferentes estágios da doença. Foi utilizada a arquitetura de rede neural convolucional U-Net para realizar a segmentação das lesões presentes nas imagens. Em seguida, as lesões segmentadas foram classificadas utilizando diversas arquiteturas de CNNs. Para analisar e interpretar as decisões do modelo, foi aplicado o Grad-CAM, uma técnica de inteligência artificial explicável que permite a visualização das áreas das imagens com maior peso na tomada de decisão da rede neural. A CNN EfficientNetB1 obteve uma acurácia de 99,45% e um F1-Score de 99,20%. Quando combinada com a técnica Grad-CAM, foi possível identificar corretamente as regiões responsáveis pela classificação do tipo de RD analisada. Dessa forma, essa abordagem não apenas demonstrou alta precisão na detecção dos diferentes tipos de RD, mas também proporcionou uma explicabilidade das decisões tomadas pela arquitetura utilizada.

Palavras-chave: *Deep learning*. Retinopatia Diabética. Inteligência Artificial Explicável.

Abstract

Diabetic retinopathy (DR) is a common ocular complication in people with diabetes, caused by damage to the blood vessels in the retina. It can lead to vision loss if untreated and is classified into stages ranging from mild to proliferative. The detection of DR is generally performed through eye fundus exams, fundoscopy, or retinal imaging, using techniques such as lesion segmentation to identify changes in the retina. The IDRiD and APTOS-2019 datasets are used for lesion segmentation caused by DR and for diabetic retinopathy detection, respectively. The IDRiD dataset contains retinal images captured from DR patients, as well as pixel-level annotations for specific lesions, such as microaneurysms, hemorrhages, hard exudates, and soft exudates. The APTOS-2019 dataset contains eye fundus images labeled into five categories. This dataset is widely used to train deep learning models to detect and classify different stages of the disease. The U-Net convolutional neural network architecture was used to segment the lesions present in the images. Subsequently, the segmented lesions were classified using various CNN architectures. To analyze and interpret the model's decisions, Grad-CAM was applied—a technique of explainable artificial intelligence that allows the visualization of the areas in the images with the most influence on the neural network's decision-making process. The EfficientNetB1 CNN achieved an accuracy of 99.45% and an F1-Score of 99.20%. When combined with the Grad-CAM technique, it was possible to correctly identify the regions responsible for the classification of the analyzed type of DR. Thus, this approach not only demonstrated high precision in detecting the different types of DR but also provided explainability for the decisions made by the utilized architecture.

Keywords: Deep Learning. Diabetic Retinopathy. Explainable Artificial Intelligence.

Lista de ilustrações

Figura 1 – Exemplos de lesões retinianas: microaneurismas, hemorragias, exsudatos moles e exsudatos duros.	27
Figura 2 – Realização de retinografia para detecção de lesões oculares.	28
Figura 3 – Representação de neurônio RNA.	30
Figura 4 – Representação de uma RNA.	31
Figura 5 – Distinção entre <i>deep learning</i> e <i>machine learning</i>	34
Figura 6 – Ilustração de um filtro convolucional.	35
Figura 7 – <i>Max pooling</i> e <i>average pooling</i>	36
Figura 8 – Arquitetura U-net.	39
Figura 9 – Construção do mapa de ativação do Grad-CAM.	40
Figura 10 – Diagrama da abordagem proposta.	42
Figura 11 – Amostras de lesões e suas respectivas anotações do conjunto de dados IDRiD.	43
Figura 12 – Amostras de cada classe do conjunto de dados APTOS-2019.	44
Figura 13 – Esquema de subdivisão da retinografia e da anotação em subimagens.	45
Figura 14 – Distribuição de RD no Conjunto de Dados Analisado.	46
Figura 15 – Exemplos de transformações aplicadas: rotações, flip Horizontal, Zoom e ajuste de brilho/contraste.	47
Figura 16 – Resultado da aplicação do pré-processamento nos conjuntos de dados IDRiD e APTOS-2019.	48
Figura 17 – Comparação entre a imagem original, as anotações manuais e as predições do modelo para microaneurismas.	58
Figura 18 – Comparação entre a imagem original, as anotações manuais e as predições do modelo para hemorragias.	58
Figura 19 – Comparação entre a imagem original, as anotações manuais e as predições do modelo para exsudatos moles.	58
Figura 20 – Comparação entre a imagem original, as anotações manuais e as predições do modelo para exsudatos duros.	59
Figura 21 – Todas os tipos de lesões: Imagem original, anotação e predição, respectivamente.	59
Figura 22 – Curvas de obtidas durante o processo de segmentação tipo de lesão.	60
Figura 23 – Resultados da aplicação do teste de Tukey.	65
Figura 24 – Matriz de confusão final do conjunto de testes utilizando a EfficientNetB1.	67
Figura 25 – Retinografia sem RD e sobreposição com o mapa de ativação.	68
Figura 26 – Retinografia com RD leve e sobreposição com o mapa de ativação.	69
Figura 27 – Retinografia com RD leve e sobreposição com o mapa de ativação.	69

Figura 28 – Retinografia com RD proliferativa e sobreposição com o mapa de ativação.	70
Figura 29 – Mapa de ativação gerado pela VGG-16.	72
Figura 30 – Mapa de ativação gerado pela DRNet.	73

Lista de tabelas

Tabela 1 – Configuração para busca de hiperparâmetros da U-Net.	49
Tabela 2 – Arquiteturas de <i>CNN</i> , configurações utilizadas e tamanho de imagem de entrada.	50
Tabela 3 – Configuração para busca de hiperparâmetros das <i>CNNs</i>	50
Tabela 4 – Métricas de desempenho de segmentação.	56
Tabela 5 – Tempos de treinamento, teste e quantidade de épocas para cada lesão.	57
Tabela 6 – Valores de IoU obtidos para a segmentação de cada tipo de lesão.	60
Tabela 7 – Comparação dos resultados obtidos durante o processo de segmentação pelo método proposto em relação a outras abordagens.	61
Tabela 8 – Métricas de avaliação obtidas para cada <i>CNN</i>	63
Tabela 9 – Tempos de treinamento e teste para cada <i>CNN</i>	64
Tabela 10 – <i>p</i> -valor obtido para cada <i>CNN</i>	65
Tabela 11 – Métricas de avaliação para cada classe obtidas pela EfficientNetB1.	66
Tabela 12 – Comparação dos resultados obtidos pelo método proposto com outras abordagens.	70

Lista de abreviaturas e siglas

RD	Retinopatia diabética
DM	Diabetes melitus
CAD	<i>Computer-aided-diagnosis</i>
PDI	<i>Processamento digital de imagens</i>
RP	Reconhecimento de padrões
ROI	<i>Region of interest</i>
IA	Inteligência artificial
VC	Visão computacional
RGB	<i>Red-green-blue</i>
CNN	<i>Convolutional neural network</i>
XAI	<i>Explainable artificial intelligence</i>
Grad-CAM	<i>Gradient-weighted class activation mapping</i>
RNA	Rede neural artificial
IoU	<i>Insertion into the Union</i>
IDRiD	<i>Indian diabetic retinopathy</i>
APTOS	<i>Asia Pacific Tele-Ophthalmology Society</i>

Lista de símbolos

$f(x)$	Função de ativação
w	Peso da rede neural
E	Erro entre a predição e a classificação
$\frac{\partial E}{\partial w}$	Derivada da função de erro em relação ao peso
H	Entropia cruzada
p_l	Probabilidade predita
$L_{\text{Grad-CAM}}^c$	Mapa de ativação
$ReLU$	Função de ativação ReLU
c	classe
y^c	Pontuação da classe
A^k	Mapa de características
Z	Número total de elementos no mapa de características
y_i, c	Valor binário que indica se a amostra i pertence à classe
$p_{i, c}$	Probabilidade prevista pelo modelo para a classe
A	Representa a região da previsão
B	Representa a região do rótulo verdadeiro
A	Região da previsão
B	Região do rótulo verdadeiro
W	Estatística W
s^2	Soma dos desvios das amostras em relação à média
p	p -value
H_0	Hipótese nula
H_1	Hipótese alternativa

α	Nível de confiabilidade
$\hat{C}$	Contraste
$\hat{\mu}$	Média estimada
VP	Verdadeiros positivos
VF	Verdadeiros negativos
FP	Falsos positivos
FN	Falsos negativos

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	19
1.2	Descrição do Problema	20
1.3	Objetivos	20
1.3.1	Objetivos Específicos	20
1.4	Desafios	20
1.5	Estado da Arte	21
1.6	Produção Científica	22
1.7	Organização da Dissertação	23
2	RETINOPATIA DIABÉTICA	24
2.1	Diabetes Melitus	24
2.2	Retinopatia Diabética	26
2.3	Manifestações Clínicas da Retinopatia Diabética	26
2.4	Diagnóstico e Tratamento	28
3	INTELIGÊNCIA ARTIFICIAL	30
3.1	Redes Neurais Artificiais	30
3.1.1	Aprendizagem Profunda	33
3.1.2	Redes Neurais Convolucionais	34
3.1.2.1	Camada Convolucional	35
3.1.2.2	Camada de <i>Pooling</i>	35
3.1.2.3	Camada Totalmente Conectada	36
3.1.2.4	<i>Dropout</i>	36
3.1.2.5	<i>Arquiteturas de Redes Convolucionais</i>	37
3.2	Inteligência Artificial Explicável	40
3.2.1	Grad-CAM	40
4	METODOLOGIA	42
4.1	Conjuntos de Dados	43
4.1.1	<i>Data Augmentation</i>	44
4.1.1.1	Segmentação	44
4.1.1.2	Classificação	45
4.2	Pré-Processamento	47
4.3	Segmentação de Lesões	48
4.4	Classificação	49

4.5	Métricas de Avaliação	50
4.5.1	Testes Estatísticos	52
4.5.1.1	Teste de Shapiro-Wilk	52
4.5.1.2	Teste ANOVA	53
4.5.1.3	Teste de Tukey	54
5	RESULTADOS E DISCUSSÕES	55
5.1	Análise dos Resultados do Processo de Segmentação	55
5.1.1	Comparação com Trabalhos Relacionados	61
5.2	Análise dos Resultados do Processo de Classificação	63
5.2.1	Resultados do Grad-CAM	68
5.2.2	Comparação com trabalhos relacionados	70
6	CONCLUSÕES	74
6.1	Trabalhos Futuros	74
	REFERÊNCIAS	76
7	TABELA F PARA $\alpha = 0.05$.	81

1 Introdução

Os sistemas computacionais desempenham um papel essencial em praticamente todos os aspectos da vida moderna, influenciando significativamente a sociedade atual e se integrando de forma profunda em nosso cotidiano. Esses sistemas permitem o armazenamento eficiente de grandes volumes de dados, o processamento rápido de informações e a aplicação de técnicas de Inteligência Artificial (IA) para replicar o raciocínio de especialistas humanos. Graças a essas capacidades, eles encontram aplicação em uma vasta gama de áreas, incluindo engenharia, gerenciamento e armazenamento de dados, planejamento e organização, redes, finanças e saúde.

Nas últimas décadas, o campo dos sistemas computacionais baseados em conhecimento tem despertado um interesse crescente, devido à sua capacidade de oferecer soluções eficazes e versáteis para diversos problemas que frequentemente não podem ser resolvidos por métodos tradicionais. Entre as áreas de maior destaque para a aplicação dessas tecnologias está a imagiologia médica. A criação de sistemas computacionais integrados ao conhecimento médico oferece uma ferramenta valiosa para apoiar tanto o diagnóstico de doenças quanto o monitoramento de seus respectivos tratamentos (REBOUÇAS FILHO, 2013).

Dentre as várias doenças que podem ser diagnosticadas com o uso de técnicas de Inteligência Artificial, tem-se a retinopatia diabética, que é diagnosticada por meio da realização do exame de fundo de olho. Esse exame pode ser feito utilizando um retinógrafo, especialmente quando o objetivo é capturar imagens detalhadas da retina. O retinógrafo é um equipamento avançado que combina um sistema de iluminação com uma câmera de alta resolução para registrar imagens do fundo do olho. Essas imagens são frequentemente utilizadas para análise, diagnóstico e acompanhamento de condições oculares, sendo denominadas retinografias.

Para realizar o diagnóstico, certos aspectos das imagens de retinografias podem ser analisados para evidenciar a presença de RD. O médico especialista geralmente realiza uma segmentação visual dessas imagens, buscando identificar áreas que possam apresentar alterações patológicas. No entanto, as mudanças observadas nas retinografias podem ser sutis, especialmente nos estágios iniciais da doença, tornando-se, em alguns casos, quase imperceptíveis a uma análise visual simples (MESSAY; HARDIE; ROGERS, 2010). Sistemas computacionais baseados no conhecimento médico, aliados a técnicas de Visão Computacional (VC), possibilitam o processamento de imagens com a realce de detalhes que antes não poderiam ser identificados por um observador humano. Isso não só acelera a análise dos exames, como também aumenta a precisão no diagnóstico, oferecendo suporte

ao médico especialista na interpretação dos resultados. Considerando a complexidade das neuroimagens, a urgência na identificação do tipo de lesão e a eventual falta temporária de um radiologista disponível, os sistemas de Diagnóstico Auxiliado por Computador (CAD) tornam-se extremamente valiosos.

De maneira geral, os sistemas CAD fazem uso de técnicas de VC, IA, Processamento Digital de Imagens (PDI) e Reconhecimento de Padrões (RP) com o objetivo de detectar, segmentar, reconhecer/identificar e classificar uma Região de Interesse (ROI) em uma imagem, de forma automática ou semiautomática. Esses sistemas podem ser divididos em quatro etapas principais: (i) aquisição das imagens, (ii) pré-processamento, (iii) segmentação, (iv) classificação (GONZALEZ, 2009), como etapa adicional, proposta por esta dissertação, tem-se a explicabilidade dos resultados.

Na etapa de aquisição, a imagem é capturada por um retinógrafo e posteriormente é convertida em uma imagem colorida. A qualidade da imagem pode ser afetada por condições ambientais ou do equipamento, podendo apresentar ruído devido o tipo de câmera ou scanner utilizado para capturar a retinografia pode gerar ruído devido à baixa qualidade do sensor ou de outros componentes. O pré-processamento visa melhorar a qualidade da imagem, aplicando técnicas para realçar características importantes e corrigir problemas como ruído, contraste ou brilho, utilizando métodos de Processamento Digital de Imagens (GONZALEZ, 2009). A segmentação tem a função de identificar e delimitar as regiões de interesse na imagem onde estão presente lesões características da RD, dividindo-a com base em similaridades ou detectando bordas. Erros nesta etapa afetam diretamente as fases seguintes (NIXON; AGUADO, 2019).

Na extração de atributos, são analisadas as características da imagem segmentada, como forma, tamanho e textura, para gerar dados numéricos que descrevem a Região de Interesse. Esses dados são usados para a etapa de classificação e reconhecimento. Por fim, na classificação e reconhecimento, os objetos da imagem são identificados automaticamente, agrupando-os em classes com base em suas semelhanças, usando técnicas de Reconhecimento de Padrões e Inteligência Artificial. A explicabilidade dos resultados de classificação de retinografias utilizando técnicas de Inteligência artificial explicável (XAI) é tem um importante papel para aumentar a confiança dos médicos e especialistas nos modelos de inteligência artificial aplicados ao diagnóstico (LI et al., 2022). Ao aplicar XAI, é possível não apenas obter a previsão do modelo, mas também compreender os motivos por trás de suas decisões, como quais características da imagem influenciaram a classificação e por quê. Isso permite que os profissionais de saúde validem os resultados de maneira mais assertiva, identificando áreas de interesse na retina que podem indicar patologias, e assegura maior transparência nos sistemas de diagnóstico assistido por computador, reduzindo a "caixa preta" associada às técnicas de IA.

1.1 Motivação

Este estudo visa explorar como técnicas de aprendizagem profunda podem auxiliar no diagnóstico médico da retinopatia diabética, uma das principais causas de cegueira evitável no mundo. A RD consiste em uma das principais complicações do diabetes, afetando assim os vasos sanguíneos da retina, podendo levar à perda progressiva da visão se não for diagnosticada e tratada precocemente. Diante do aumento da incidência do diabetes e da limitação de recursos médicos para atender a toda a demanda de pacientes, soluções baseadas em inteligência artificial têm se mostrado promissoras para auxiliar no diagnóstico e na tomada de decisão clínica.

O foco deste trabalho está na aplicação de *CNNs* para segmentar lesões características da RD e classificar o grau da doença em diferentes estágios de gravidade. Além disso, busca-se integrar técnicas de inteligência artificial explicável, que fornecem interpretações visuais sobre as regiões da imagem mais relevantes para a predição do modelo, possibilitando que oftalmologistas compreendam melhor os fatores que influenciam a decisão algorítmica. Essa abordagem não apenas aumenta a confiabilidade dos sistemas baseados em IA, mas também auxilia no processo de aceitação no ambiente clínico, onde a transparência das decisões é um fator crítico para a adoção de novas tecnologias.

Mesmo os recentes avanços tecnológicos no diagnóstico por imagem, a RD continua sendo uma condição subdiagnosticada, especialmente em regiões com poucos especialistas. Isso ocorre porque a análise manual de exames oftalmológicos, como retinografias, pode ser um processo demorado e sujeito à variabilidade interpessoal. Além disso, fatores como fadiga e sobrecarga de trabalho podem comprometer a acurácia das avaliações médicas, resultando em diagnósticos tardios ou imprecisos. Nesse contexto, a implementação de modelos baseados em aprendizagem profunda pode tornar o diagnóstico mais preciso e consistente, permitindo que exames sejam analisados automaticamente com alta eficiência.

A combinação entre redes neurais profundas e métodos explicáveis representa um avanço significativo na área de diagnóstico assistido por IA. Além de possibilitar a detecção precoce da RD, essa abordagem permite que os médicos realizem uma triagem preliminar mais eficaz, otimizando o tempo de atendimento e reduzindo a probabilidade de erros diagnósticos. Dessa forma, esta dissertação tem como principal motivação apresentar uma metodologia para a detecção automatizada da RD, ao mesmo tempo em que destaca as regiões da imagem que foram mais relevantes para a tomada de decisão do modelo. Espera-se que os resultados obtidos contribuam para o desenvolvimento de ferramentas computacionais que auxiliem na triagem da RD e na melhoria da qualidade do atendimento oftalmológico, beneficiando tanto os profissionais de saúde quanto os pacientes.

1.2 Descrição do Problema

A RD é uma complicação comum do diabetes, caracterizada por danos progressivos aos vasos sanguíneos da retina, podendo levar à cegueira se não diagnosticada e tratada precocemente. Dada a importância de identificar a RD em seus estágios iniciais, esta dissertação aborda o desenvolvimento de um sistema computacional avançado que utiliza técnicas de segmentação de imagens e classificação para a detecção automática da RD. Além disso, serão aplicadas técnicas de interpretabilidade, como Grad-CAM, para destacar regiões relevantes das imagens analisadas, fornecendo maior transparência e confiabilidade às decisões do modelo, e auxiliando profissionais da saúde na validação dos resultados. Este trabalho visa contribuir para a melhoria da precisão diagnóstica e a redução do impacto dessa doença na qualidade de vida dos pacientes.

1.3 Objetivos

O objetivo principal deste trabalho é desenvolver uma abordagem para a segmentação e classificação da retinopatia diabética utilizando *Deep Learning* e inteligência artificial explicável, visando auxiliar oftalmologistas no processo de diagnóstico.

1.3.1 Objetivos Específicos

De modo mais específico, este trabalho possui os seguintes objetivos específicos:

- Segmentar lesões provenientes da RD;
- Explorar diversas arquiteturas de CNNs para realizar a detecção e classificação do grau de RD;
- Aplicar técnicas de inteligência artificial explicável de modo a tornar mais transparentes as regiões que foram consideradas na decisão da classificação;
- Analisar a relevância da abordagem proposta utilizando as métricas: acurácia, sensibilidade, precisão, F1-Score e *Log Loss*;
- Comparar os resultados obtidos com os estudos já publicados para validação da abordagem proposta.

1.4 Desafios

O desenvolvimento de sistemas computacionais para a detecção e análise da RD apresenta uma série de desafios significativos.

- Primeiramente, a qualidade e a variabilidade das imagens retinais representam um obstáculo, pois fatores como iluminação inadequada, ruídos, e variações anatômicas podem afetar a precisão dos modelos. Além disso, a segmentação

precisa de estruturas relevantes, como vasos sanguíneos, microaneurismas e exsudatos, exige técnicas robustas e bem treinadas para lidar com a complexidade dessas características;

- Outro desafio está na necessidade de interpretar os resultados do modelo, uma vez que as decisões tomadas devem ser confiáveis e compreensíveis para os profissionais de saúde, especialmente ao aplicar Grad-CAM para justificar as classificações;
- Por fim, garantir a generalização dos modelos em diferentes populações e cenários clínicos é essencial, demandando conjuntos de dados diversificados e um rigoroso processo de validação. Esses desafios destacam a complexidade do problema e a importância de abordagens inovadoras e bem fundamentadas para superá-los.

1.5 Estado da Arte

O estudo de (KOU et al., 2020) sobre segmentação de microaneurismas alcançou 98% de acurácia utilizando a ERU-Net, uma U-Net residual aprimorada com três caminhos U, que melhora a fusão de características e captura de detalhes. A presença de um bloco residual na arquitetura otimizou a extração de características, resultando em uma sensibilidade de 92,90%. O trabalho conduzido por (NAZIR et al., 2020) utilizou a Faster R-CNN, alcançando 95,20% de acurácia para segmentação de hemorragias, com sensibilidade de 95,10% e F1-Score de 94,30%.

A abordagem proposta por (HARIS et al., 2019), dedicado à segmentação de exsudatos moles, foi obtida uma acurácia de 90,46% e uma sensibilidade de 88,86%. A pesquisa utilizou a SOP-Net, uma rede de segmentação semântica com arquitetura encoder-decoder, projetada para realizar a delimitação simultânea de patologias retinais. O estudo de (SAMBIAL et al., 2020) propôs uma abordagem baseada na U-Net para a segmentação de lesões de exsudatos duros. Foram utilizados os conjuntos de dados IDRiD e e-optha (DECENCIERE et al., 2013), alcançando uma acurácia de 99% no primeiro e uma sensibilidade de 99,85% no segundo. No estudo conduzido por (FURTADO, 2020), é apresentado um novo método de segmentação de lesões provenientes da retinopatia diabética. Nesse método, todas as lesões são segmentadas por uma única arquitetura, DeepLabv3. Durante a realização dos experimentos, foi atingida uma acurácia de 90% e um IoU de 0,86.

Em (EL-ATEIF; IDRI, 2022), os autores compararam o desempenho de abordagens de aprendizado profundo para a classificação automática de Retinopatia Diabética (RD). Eles utilizaram fusão conjunta, combinando os conjuntos de dados APTOS-2019 e Messidor-2 para classificar imagens de retinografia, alcançando 97,49% de acurácia. Em (KHALIFA et al., 2019), modelos de *transfer learning* foram aplicados para a detecção de RD usando o

conjunto de dados APTOS-2019. Redes como AlexNet, ResNet18, SqueezeNet, GoogleNet, VGG16 e VGG19 foram escolhidas por sua menor complexidade em relação a modelos maiores, como DenseNet. Técnicas de aumento de dados foram empregadas para reduzir *overfitting*, com a AlexNet alcançando a melhor acurácia, 97,9%. O estudo utilizou CNNs para classificação de RD, similar à abordagem deste trabalho.

Em (TAUFIQURRAHMAN et al., 2020) é proposto o uso de deep learning para classificação de RD. O modelo foi treinado no conjunto de dados APTOS-2019, utilizando pesos pré-treinados do *ImageNet*. Além disso, foi otimizado com a combinação de um classificador SVM, formando o modelo híbrido MobileNetV2-SVM, atingindo uma acurácia de 85,00%. O estudo de (ALGHAMDI, 2022) abordou a detecção de RD utilizando aprendizado profundo, avaliando modelos em duas abordagens: classificação binária e multiclasse, conforme os níveis de gravidade da doença. Entre os modelos testados, o VGG-16 apresentou o melhor desempenho, alcançando uma acurácia de 73,04%. Além disso, foi utilizado o método Grad-CAM para trazer explicabilidade às decisões dos modelos.

O estudo de (MURUGAPPAN et al., 2022) propôs um método inovador para diagnóstico e classificação de RD utilizando *Few-Shot Learning*, permitindo o treinamento eficiente com poucos dados e resolvendo problemas de sobreajuste. A abordagem é baseada na DRNet, uma rede prototípica com mecanismos de atenção que utiliza aprendizado episódico para criar representações mais ricas e focadas. Com uma acurácia de 98,18%, a DRNet demonstrou alto desempenho ao destacar regiões relevantes das imagens de fundo de olho, como microaneurismas e exsudatos. O Grad-CAM foi empregado para explicar as decisões do modelo.

Em (RAHMAN et al., 2023), os autores apresentam uma metodologia para detecção e localização de câncer cerebral utilizando métodos de Inteligência Artificial Explicável. Com o uso de uma Rede Neural Convolucional, combinada com o Grad-CAM, o sistema analisa imagens de Ressonância Magnética para superar as limitações de precisão dos modelos existentes. No trabalho realizado por (SAMPAIO; CORDEIRO, 2022), os autores u Os autores fizeram uso de modelos de aprendizado supervisionado para a detecção de massas em imagens de mamografia, reduzindo a necessidade de anotações detalhadas em nível de pixel. Foram utilizados diferentes métodos de mapas de ativação (CAM, GradCAM, GradCAM++, XGrad-CAM e LayerCAM). Os resultados indicam que estratégias variadas de mapas de ativação durante as fases de treinamento e teste melhoram o desempenho do modelo.

1.6 Produção Científica

Durante este projeto de mestrado, os seguintes trabalhos científicos foram aceitos e publicados:

- **Qualis A3** - **Silva, M. C.**; Silva, C. M.; Matos, A. G.; Silva, S. P. P.; Sarmiento, R. M.; Filho, P. P. R., 2024. *A New Diabetic Retinopathy Classification Approach Based on Normalizer Free Network*. Symposium on Computer-Based Medical Systems, 39, pp. 164-169.
- **Qualis A1** - **Silva, M. C.** ; Silva, C. M. ; Silva, S. P. P. ; Rebouças, E. S. ; Cunha, C. C. H. ; Benevides, C. A. ; Sarmiento, R. M. ; Song, H. H. ; Rebouças Filho, P. P. *Health Prediction: A Comprehensive IoT-Driven Health Monitoring System with Machine Learning Analysis and XAI Insights*. Conference on Neural Information Processing Systems, 38.
- **Qualis A3** - Silva, C. M.; **Silva, M. C.**; Silva, S. P. P.; Rebouças Filho, P. P.; Nascimento, N. M. M., 2024. Computer Vision for Brain Tumor Classification: A Novel Approach Based on Zernike Moments. Symposium on Computer-Based Medical Systems, 39, pp. 94-99.

1.7 Organização da Dissertação

Esta dissertação está organizada em cinco capítulos, sendo este primeiro destinado à introdução. Os Capítulos 2 e 3 apresentam os fundamentos teóricos utilizados nesta pesquisa acerca da RD, aprendizagem profunda e inteligência artificial explicável. O Capítulo 4, por sua vez, descreve em mais detalhes a metodologia utilizada. Em seguida, no Capítulo 5, são apresentados os resultados e as discussões dos resultados gerados, bem como a comparação com outros trabalhos relacionados. Por fim, nos Capítulos 6 e 6.1, apresentam-se as conclusões obtidas a partir dos resultados, além de prospecções para trabalhos futuros.

2 Retinopatia Diabética

Este capítulo é destinado à apresentação dos conceitos fundamentais para a compreensão da RD. Inicialmente, será apresentada uma visão geral sobre a doença, abordando sua definição, causas e fatores de risco. Em seguida, serão discutidos os sinais e sintomas da condição. Por fim, serão exploradas as principais abordagens diagnósticas e os métodos de tratamento mais utilizados. Com esta finalidade, esta seção foi estruturada por quatro seções:

2.1 Diabetes Melitus

O diabetes mellitus (DM) é caracterizado como um distúrbio metabólico de diversas origens, destacando-se pela hiperglicemia crônica e por alterações no metabolismo de carboidratos, lipídios e proteínas, resultantes de falhas na secreção ou na ação da insulina. Os efeitos do DM podem causar danos a longo prazo, incluindo disfunções e insuficiências em vários órgãos. Estima-se que, em 2021, mais de 537 milhões de adultos estavam vivendo com diabetes, e esse número deve atingir 783 milhões até 2045, de acordo com a Federação Internacional de Diabetes (COUNTRY, 2022). O Brasil é o 5º país em incidência de diabetes no mundo, com 16,8 milhões de doentes adultos (20 a 79 anos), perdendo apenas para China, Índia, Estados Unidos e Paquistão (Brasil, Ministério da Saúde, 2024). A estimativa da incidência da doença em 2030 chega a 21,5 milhões. A doença é responsável por várias complicações, como retinopatia diabética, que afeta até 40% dos diabéticos e pode levar à perda de visão se não for tratada adequadamente.

Os sintomas mais comuns incluem sede excessiva, aumento da frequência urinária, problemas de visão e perda de peso (BYTZER et al., 2002). Em casos mais graves, pode ocorrer cetoacidose ou um estado hiperosmolar não cetótico, levando à letargia, coma e, sem tratamento adequado, à morte. Em muitos pacientes, o DM muitas vezes não apresenta sintomas ou eles são muito leves, o que pode dificultar o diagnóstico precoce (MENDOLA et al., 2018). Existem vários tipos de diabetes, sendo os principais (BANTING; BEST et al., 1922):

- **Diabetes Tipo 1:** Consistem em uma condição autoimune em que o sistema imunológico destrói as células do pâncreas responsáveis pela produção de insulina, resultando em uma deficiência da insulina no corpo. Geralmente diagnosticado em crianças ou jovens, embora também possa ocorrer em adultos. Pacientes com diabetes tipo 1 precisam administrar insulina para controlar os níveis de glicose no sangue. Os sintomas incluem sede excessiva, urinação frequente, perda de

peso inexplicada, cansaço e visão embaçada. Caso não seja tratado, pode levar a complicações como a cetoacidose diabética;

- **Diabetes Tipo 2:** Neste tipo, o corpo desenvolve resistência à insulina ou não consegue produzir insulina suficiente. É mais comum em adultos, especialmente em pessoas com excesso de peso, mas está se tornando cada vez mais comum também em jovens. Pode ser controlado com mudanças na alimentação, prática de atividades físicas e, em alguns casos, com medicamentos orais. Em estágios mais avançados, pode ser necessário o uso de insulina. Os sinais são semelhantes aos do tipo 1, mas geralmente mais sutis e podem não ser notados por um longo tempo. Fatores como obesidade e sedentarismo aumentam o risco de desenvolvimento;
- **Diabetes Gestacional:** Bastante comum durante a gestação, quando o corpo não consegue produzir insulina suficiente para atender às necessidades aumentadas durante a gravidez, levando ao aumento da glicose no sangue. Costuma desaparecer após o parto, mas mulheres que a desenvolvem têm maior risco de desenvolver diabetes tipo 2 no futuro. Geralmente, não apresenta sintomas óbvios, sendo identificada por meio de exames de glicose realizados durante a gestação;
- **Diabetes Monogênica:** Resulta de uma alteração genética em um único gene, o que causa falhas na produção de insulina. Esse tipo é raro e pode se manifestar em jovens, frequentemente de forma leve. Embora se manifeste em idades mais precoces, a evolução pode ser mais suave que o Tipo 1. Os sintomas são semelhantes aos do Tipo 1 ou 2, mas com características genéticas distintas.

O tratamento da diabetes é de vital importância para o controle glicêmico e a prevenção de complicações associadas à doença. Esse manejo compreende intervenções no estilo de vida, como a adoção de uma dieta equilibrada, rica em fibras e com baixo índice glicêmico, associada à prática regular de exercícios físicos e à manutenção de um peso corporal saudável.

Para muitos pacientes, o uso de medicamentos antidiabéticos orais ou insulina é indispensável, sendo ajustado de acordo com o tipo de diabetes e as necessidades individuais. O acompanhamento clínico periódico permite monitorar a progressão da doença e avaliar a eficácia das intervenções terapêuticas, promovendo ajustes personalizados. Além disso, a educação em saúde é um pilar central no tratamento, capacitando os pacientes a compreenderem sua condição e a adotarem medidas proativas para minimizar riscos, como complicações microvasculares (retinopatia diabética, nefropatia, neuropatia) e macrovasculares (doenças cardiovasculares).

2.2 Retinopatia Diabética

Caracterizada como uma das principais complicações da DM e se caracteriza por lesões na retina resultantes de altos níveis de glicose, podendo levar à cegueira (GUPTA et al., 2018). Segundo (Brazilian Council of Ophthalmology, 2024), a RD é responsável por aproximadamente 4,8% dos casos de cegueira no mundo. Embora seja uma doença que possua tratamento, uma vez que a visão tenha sido perdida, não pode ser totalmente restaurada.

Inicialmente, a retinopatia diabética pode ser assintomática, o que significa que os pacientes podem não perceber nenhum sintoma (CANDRILLI et al., 2007). No entanto, à medida que a doença avança, pode ocorrer o aparecimento de microaneurismas (pequenos vasos sanguíneos dilatados), hemorragias, edema macular (inchaço na área central da retina) e a formação de neovasos (novos vasos sanguíneos anormais) que podem romper e causar sangramentos. Em estágios mais avançados, esses problemas podem levar a complicações graves, como o descolamento da retina e glaucoma.

A RD é classificada em diferentes estágios, desde a retinopatia diabética não proliferativa (menos grave) até a proliferativa (mais avançada), que é associada à formação de novos vasos sanguíneos danosos. A fase proliferativa pode resultar em cegueira irreversível, caso não seja tratada de forma adequada.

Fatores como o controle inadequado da glicemia, hipertensão arterial, duração do diabetes e a presença de outras condições, como dislipidemia, aumentam o risco de desenvolver retinopatia diabética (YAU et al., 2012). A detecção precoce é essencial para prevenir a progressão da doença e evitar danos permanentes à visão. Exames oftalmológicos regulares, como a fundoscopia e a tomografia de coerência óptica (OCT), são fundamentais para monitorar e diagnosticar a doença em seus estágios iniciais.

2.3 Manifestações Clínicas da Retinopatia Diabética

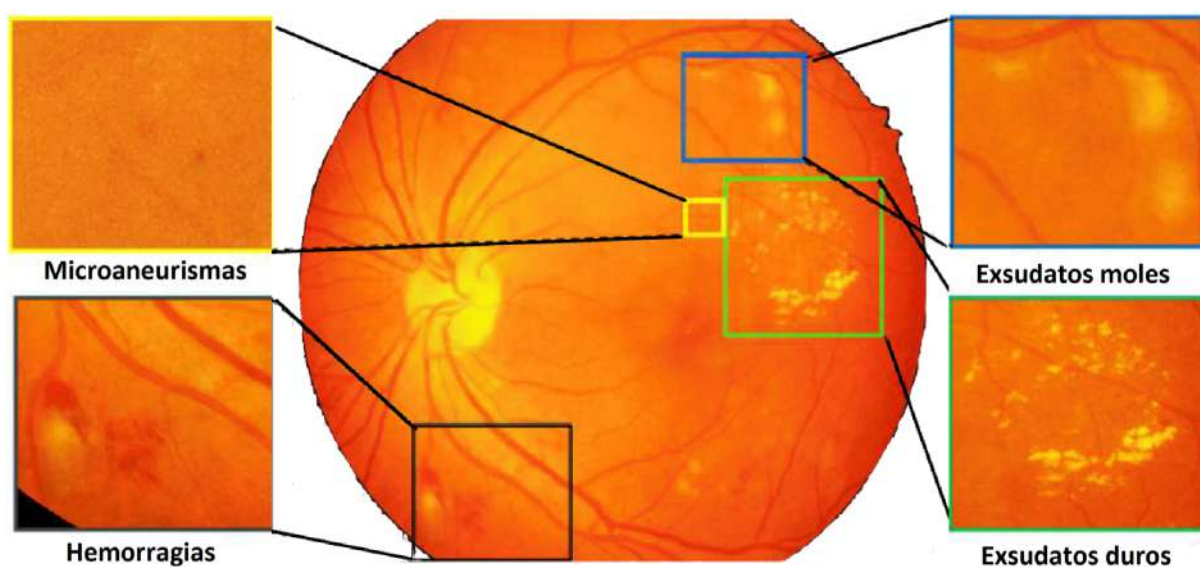
A RD progride gradualmente, com a gravidade determinada por lesões retinianas, como acúmulo de líquido, microaneurismas, hemorragias e exsudatos. A Figura 1 ilustra exemplos de como essas lesões se manifestam na região da retina. Essas lesões possuem as seguintes características (GUPTA et al., 2018):

- **Microaneurismas:** São as primeiras lesões visíveis da retinopatia diabética. Aparecem como pequenos pontos vermelhos e são dilatações microscópicas dos capilares da retina. Essas lesões indicam o rompimento da parede dos vasos sanguíneos, permitindo o vazamento de fluido para dentro da retina;
- **Hemorragias:** Ocorrem quando os vasos sanguíneos rompem, liberando sangue na retina. As hemorragias podem aparecer como manchas escuras ou áreas aver-

melhadas e estão associadas a estágios mais avançados da retinopatia diabética. Podem ser classificadas como intrarretinianas ou sub-hialoides, dependendo da profundidade do sangramento;

- **Exsudatos duros:** São depósitos lipídicos que se acumulam na retina devido ao vazamento de fluidos ricos em gordura dos capilares comprometidos. Esses exsudatos aparecem como áreas amareladas ou brancas e indicam dano vascular contínuo, sendo um marcador de retinopatia diabética severa;
- **Exsudatos moles:** São causados pela isquemia da retina, resultando em áreas de infarto das fibras nervosas. Aparecem como áreas esbranquiçadas e opacas e indicam uma interrupção no suprimento sanguíneo local, sendo um sinal de complicações avançadas da retinopatia.

Figura 1 – Exemplos de lesões retinianas: microaneurismas, hemorragias, exsudatos moles e exsudatos duros.



Fonte: Imagem adaptada de (PORWAL et al., 2020).

De acordo com (WILKINSON C. P., 2003), RD pode ser classificada em diferentes tipos com base na gravidade da doença. Nesse sentido, foi criada a Escala Internacional de Gravidade da Doença para a RD. Assim, cinco graus de gravidade são diferenciados, que, por sua vez, são agrupados em dois tipos principais: retinopatia diabética não proliferativa, associada aos estágios iniciais da doença, e retinopatia diabética proliferativa, associada a estágios mais avançados (WU L., 2024). Portanto, os graus de gravidade da retinopatia diabética foram divididos da seguinte forma:

- **Leve:** Início da RD com a presença de alguns microaneurismas na retina, marcando o início das alterações associadas à doença;
- **Moderada:** Há aumento de microaneurismas, hemorragias intrarretinianas, exsudatos duros e, em alguns casos, edema macular leve. Podem exsudatos

moles devido à isquemia da retina. Esses sinais indicam maior comprometimento vascular, com risco de progressão para estágios graves;

- **Severa:** Há um aumento significativo de hemorragias intrarretinianas e microaneurismas, além de extensa isquemia, evidenciada por múltiplas manchas algodonosas. Surgem anormalidades venosas, como dilatações e loops vasculares, indicando um risco iminente de progressão para a RD proliferativa;
- **Proliferativa:** Ocorre crescimento anormal de novos vasos sanguíneos na retina (neovascularização), resultante da isquemia crônica. Esses vasos são frágeis e podem romper, causando hemorragias vítreas. Exsudatos moles, fibrose e descolamento de retina também podem ocorrer, levando a perda significativa da visão. Uma outra característica dessa fase da doença é a degeneração da retina, uma vez que a retina sofre alterações graves, como o crescimento anormal de novos vasos sanguíneos, conhecidos como neovasos. Esses vasos são frágeis e podem vaziar fluido, causando hemorragias e cicatrizes. Esse processo pode levar à degeneração da retina, resultando em perda de visão significativa.

2.4 Diagnóstico e Tratamento

Atualmente, a RD pode ser detectada por meio da realização de uma retinografia do olho do paciente, que consiste em um exame feito com a ajuda de um equipamento denominado retinógrafo (Figura 2), o qual fotografa a retina, a camada do olho responsável pela visão. Durante o exame, o paciente posiciona o rosto no aparelho, e o retinógrafo emite um feixe de luz para capturar imagens nítidas da parte interna do olho. Em alguns casos, o médico aplica colírios para dilatar as pupilas, facilitando uma visualização mais ampla da retina. Atualmente, esse é considerado o padrão ouro para o diagnóstico de RD (GROUP et al., 1991).

Figura 2 – Realização de retinografia para detecção de lesões oculares.



Fonte: Imagem adaptada de (MACCUISH, 1993).

Essas imagens permitem que o médico oftalmologista ou sistemas de inteligência artificial identifiquem sinais precoces de RD, como alterações nos vasos sanguíneos, hemorragias e depósitos de exsudatos. O procedimento é rápido, não invasivo e indolor, sendo

amplamente utilizado em clínicas e hospitais para detectar lesões oculares relacionadas ao diabetes.

O início precoce do tratamento da RD traz inúmeras vantagens, sendo crucial para preservar a visão e melhorar a qualidade de vida dos pacientes (NATHAN; GROUP, 2014). Quando diagnosticada e tratada nas fases iniciais, a doença tem maiores chances de ser controlada, prevenindo complicações graves como edema macular, hemorragia vítrea e descolamento de retina. Além disso, intervenções precoces, como o controle rigoroso dos níveis glicêmicos e o uso de terapias específicas, podem retardar ou até interromper a progressão da retinopatia, reduzindo a necessidade de tratamentos invasivos no futuro. Esse manejo também diminui o risco de cegueira, possibilitando que os pacientes mantenham suas atividades diárias e autonomia. Portanto, o diagnóstico e o tratamento precoce são essenciais para otimizar os resultados clínicos e minimizar os impactos socioeconômicos associados à perda de visão.

O tratamento depende do estágio da doença e do grau de dano à retina, sendo essencial para prevenir a progressão e preservar a visão (STITT et al., 2016). A base do manejo está no controle rigoroso do diabetes, incluindo a manutenção dos níveis de glicose, pressão arterial e colesterol dentro dos limites recomendados, além de mudanças no estilo de vida, como uma alimentação equilibrada, prática de exercícios físicos e abandono do tabagismo. Nos casos mais avançados, podem ser necessários tratamentos específicos, como a fotocoagulação a laser, que sela vasos sanguíneos anormais para prevenir hemorragias, ou as injeções intravítreas de medicamentos anti-VEGF ou corticosteroides, que reduzem o inchaço da mácula e inibem o crescimento de vasos anômalos (GUPTA et al., 2013). Em situações mais graves, como hemorragias vítreas ou descolamento de retina, a cirurgia de vitrectomia é indicada para restaurar a transparência visual. Além dessas abordagens, o acompanhamento regular com exames oftalmológicos é imprescindível para o diagnóstico precoce e intervenções oportunas. Tecnologias avançadas, como inteligência artificial, têm contribuído para melhorar o diagnóstico e personalizar o tratamento, oferecendo novas possibilidades para enfrentar a retinopatia diabética.

3 Inteligência Artificial

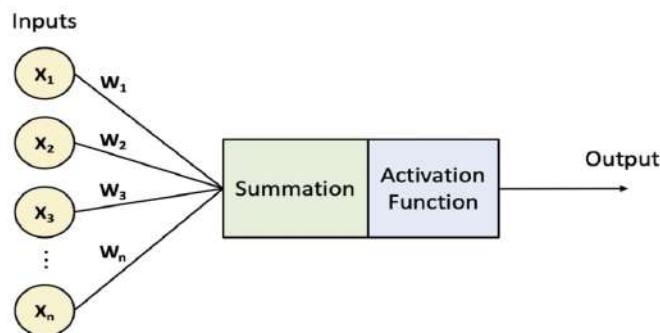
Este capítulo é destinado à apresentação dos conceitos fundamentais para a compreensão dos métodos de aprendizagem profunda empregados nesta dissertação. Inicialmente, será apresentada uma visão geral sobre redes neurais artificiais. Em seguida, será explorada a inteligência artificial explicável, mais particularmente o método Grad-CAM. Com essa finalidade, esta seção foi estruturada em duas partes:

3.1 Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são métodos computacionais inteligentes que se baseiam no funcionamento dos neurônios biológicos. Uma RNA é composta por várias unidades de processamento, denominadas neurônios, que realizam operações relativamente simples. Essas unidades são interligadas por meio de canais de comunicação que possuem pesos associados (SILVA et al., 2016).

As entradas dos neurônios são ajustadas por esses pesos e combinadas utilizando uma função de ativação, que define se a saída do neurônio será ativada com base em um valor de referência determinado pela própria função de ativação. As entradas dos neurônios são ajustadas por pesos e processadas de acordo com a função de ativação. Essa função tem a responsabilidade de combinar de forma linear os valores de entrada, ponderados pelos pesos, e determinar se a saída do neurônio será ativada com base em um valor de referência. Nas redes neurais artificiais, os neurônios podem ser replicados e organizados em diferentes camadas, onde cada camada pode utilizar a mesma função de ativação. A combinação dessas camadas de neurônios cria redes neurais que possibilitam o ajuste dos pesos sinápticos ao longo do processo de treinamento. A Figura 3 ilustra uma representação de um neurônio de RNA.

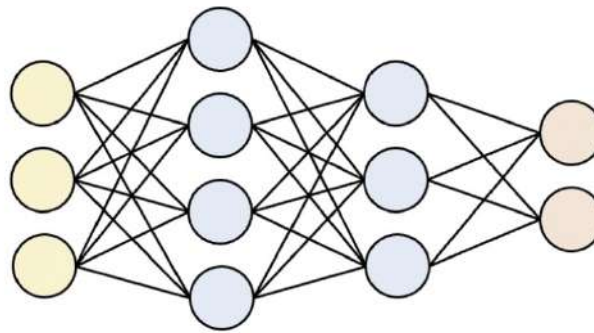
Figura 3 – Representação de neurônio RNA.



Fonte: Imagem adaptada de (DINGLI; FARRUGIA, 2023).

Os neurônios em uma RNA podem ser combinados e replicados, formando várias camadas, em que cada uma utiliza a mesma função de ativação. Essas conexões entre neurônios criam redes neurais capazes de ajustar os pesos sinápticos ao longo do processo de treinamento. No aprendizado supervisionado, a rede aprende a partir de um conjunto de dados de entrada com saídas previamente conhecidas, ajustando seus parâmetros conforme necessário. Já no aprendizado não supervisionado, as saídas são classificadas em um determinado número de grupos, definido pela quantidade de neurônios presentes na camada de saída. A Figura 4 ilustra uma representação de uma RNA.

Figura 4 – Representação de uma RNA.



Fonte: Imagem adaptada de (DINGLI; FARRUGIA, 2023).

Existe uma grande variedade de funções de ativação que podem ser empregadas, desde que sejam monotonicamente crescentes, contínuas e diferenciáveis. Esses requisitos são impostos pelo método de descida de gradiente para o ajuste dos pesos sinápticos dos neurônios nas camadas ocultas. As equações (2.1 -2.5) apresentam as principais funções de ativação.

Função Linear:

$$f(x) = x \quad (2.1)$$

Função Limiar:

$$f(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \quad (2.2)$$

Função Tangente Hiperbólica:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.4)$$

Função Sigmoid:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.5)$$

Função Softmax:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (2.5)$$

As Redes Neurais Artificiais têm diversas vantagens, tornando-as úteis em muitas aplicações. Uma das principais é que elas conseguem tomar decisões com base em dados históricos, sem que o usuário precise ter conhecimento técnico avançado. Além disso, elas são bastante robustas, o que significa que continuam funcionando bem mesmo quando os dados fornecidos têm inconsistências. Outra vantagem é que, uma vez que a RNA está devidamente treinada para uma tarefa específica, ela pode ser utilizada em tempo real, sem precisar de ajustes constantes em sua estrutura.

Por outro lado, as RNAs também apresentam algumas desvantagens. Por exemplo, elas podem chegar a soluções que não seguem regras ou teorias estabelecidas, o que pode ser problemático em algumas situações. Além disso, o processo de treinamento de uma RNA pode ser bastante demorado e exigir muita capacidade computacional. Outro ponto é que, mesmo após o treinamento, não há garantias de que a RNA vá encontrar a melhor solução possível.

O processo de aprendizagem de uma RNA, inicialmente, se dá por uma etapa denominada *feedforward*, a qual envolve a propagação dos dados de entrada pela rede, camada por camada, até a saída. Cada neurônio calcula uma soma ponderada de suas entradas, aplica uma função de ativação e repassa o valor para a próxima camada. A segunda etapa do processo de aprendizagem é denominada *backpropagation*. Após gerar a saída, os pesos são ajustados para minimizar o erro da previsão. O erro é calculado por meio de uma função de custo (como o erro quadrático médio) e propagado da saída para as camadas anteriores. O ajuste dos pesos w_{ij} é feito utilizando o gradiente do erro em relação a cada peso, conforme a equação 3.1:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial E}{\partial w_{ij}} \quad (3.1)$$

onde η é a taxa de aprendizagem, um hiperparâmetro que controla a magnitude da atualização, e $\frac{\partial E}{\partial w_{ij}}$ é o gradiente do erro em relação ao peso w_{ij} . Na propagação de erros durante o processo de *backpropagation* em uma Rede Neural Artificial, o erro é uma medida de quão longe a saída da rede está da saída esperada. Este erro é frequentemente quantificado usando uma função de custo, como o erro quadrático médio (MSE), que é definida como:

$$E = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3.2)$$

onde N é o número total de amostras, y_i é o valor real e $\hat{y}_i$ é a saída prevista pela rede. Durante a fase de *backpropagation*, o erro é retropropagado para as camadas

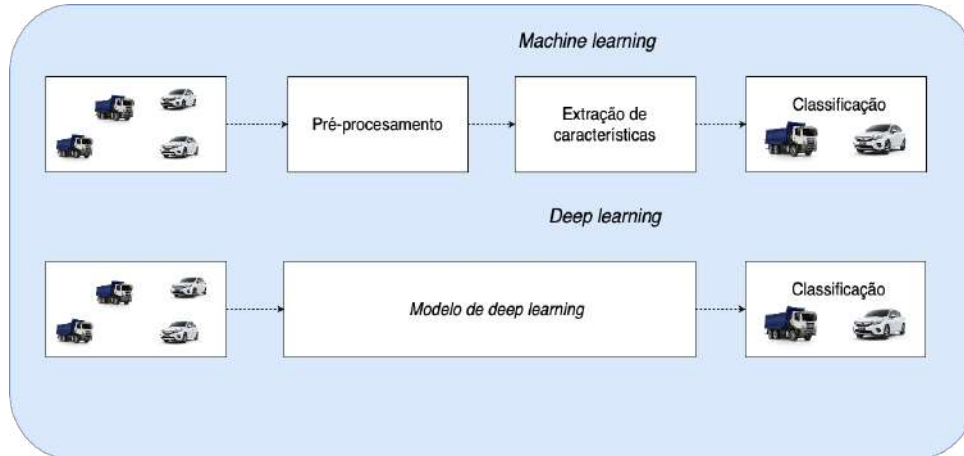
anteriores, permitindo que cada neurônio calcule a contribuição do seu peso para o erro total. O gradiente do erro é então utilizado para ajustar os pesos através da equação mencionada anteriormente:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial E}{\partial w_{ij}} \quad (3.3)$$

O objetivo é minimizar o erro da rede, possibilitando que ela aprenda padrões nos dados de entrada. A taxa de aprendizagem η desempenha um papel crucial, pois determina a velocidade com que os pesos são atualizados; uma taxa muito alta pode resultar em oscilações e instabilidade, enquanto uma taxa muito baixa pode levar a um aprendizado lento e a ficar preso em mínimos locais.

3.1.1 Aprendizagem Profunda

Aprendizagem profunda ou *deep learning* consiste em uma subárea do campo da inteligência artificial que utiliza redes neurais de múltiplas camadas. Essas redes são projetadas para aprender representações complexas dos dados, permitindo que máquinas executem tarefas complexas, como reconhecimento de imagens, interpretação de linguagem natural entre outras. O *deep learning* se destaca pela sua habilidade de lidar com grandes quantidades de dados e identificar padrões sutis, frequentemente superando abordagens tradicionais. Com o uso de técnicas como retropropagação e algoritmos de otimização avançados, as redes neurais profundas são treinadas em extensos conjuntos de dados, resultando em modelos com alta capacidade de generalização (LECUN; BENGIO; HINTON, 2015). As abordagens tradicionais de aprendizado de máquina exigem uma série de etapas, como o pré-processamento de imagens, a extração e a seleção de características, além do aprendizado e da classificação. Em contrapartida, abordagens baseadas em *deep learning* se destacam por sua capacidade de automatizar o aprendizado de características a partir de grandes conjuntos de dados, permitindo que o processo de aprendizado e a classificação ocorram simultaneamente. A Figura 5 apresenta, de forma conceitual, as principais diferenças entre os métodos tradicionais de *machine learning* e *deep learning*.

Figura 5 – Distinção entre *deep learning* e *machine learning*.

Fonte: Autor.

Técnicas baseadas em *deep learning* são caracterizadas por sua robustez, com recursos otimizados que podem ser aprendidos de forma automatizada em relação à tarefa desejada. Além disso, sua alta capacidade de generalização permite que se adaptem a diferentes cenários e variações, mantendo a saída correta.

3.1.2 Redes Neurais Convolucionais

No ramo de *deep learning*, a Rede Neural Convolucional (*CNN*) é o algoritmo conceitualmente mais famoso e mais utilizado em diversas aplicações (LI et al., 2021; KRIZHEVSKY; SUTSKEVER; HINTON, 2017). A principal vantagem apresentada por uma *CNN* em comparação com seus antecessores é que ela identifica automaticamente os recursos relevantes sem a necessidade de qualquer supervisão humana (GU et al., 2018). Além disso, possui a capacidade de realizar a extração das características. As *CNNs* têm sido amplamente aplicadas em diversos campos, incluindo reconhecimento facial (BALABAN, 2015), processamento de voz (ZHANG et al., 2018), entre outros.

As Redes Neurais Convolucionais baseiam-se na visão humana, e seu conceito é fundamentado em uma operação matemática fundamental, denominada convolução. A convolução consiste no somatório do produto entre duas funções ao longo da região em que elas se sobrepõem, em razão do deslocamento existente entre elas, conforme apresentado na equação 3.4.

$$(f * g)(x, y) = \sum_{m=-M}^M \sum_{n=-N}^N f(x - m, y - n) \cdot g(m, n) \quad (3.4)$$

Onde:

- $f(x, y)$ é a função de entrada;
- $g(m, n)$ é o kernel ou filtro;

- $*$ representa a operação de convolução;
- (x, y) são as coordenadas da saída;
- M e N definem o tamanho do kernel.

As redes neurais convolucionais são compostas por cinco camadas principais: a camada convolucional, a camada de *pooling*, a camada totalmente conectada (*Fully Connected*), e a camada de *Dropout*. Além dessas, a inclusão de funções de ativação é de vital importância. Abaixo, explico o papel de cada uma dessas camadas em uma rede.

3.1.2.1 Camada Convolucional

A camada convolucional é uma das mais importantes em uma rede neural convolucional, pois ela tem a função de extrair características das imagens de entrada. Isso é feito através de filtros (também chamados de *kernels*), que se deslocam sobre a imagem, identificando padrões como bordas, texturas e formas. Esses filtros ajudam a simplificar a imagem, preservando as informações mais relevantes para as camadas seguintes da rede.

Figura 6 – Ilustração de um filtro convolucional.

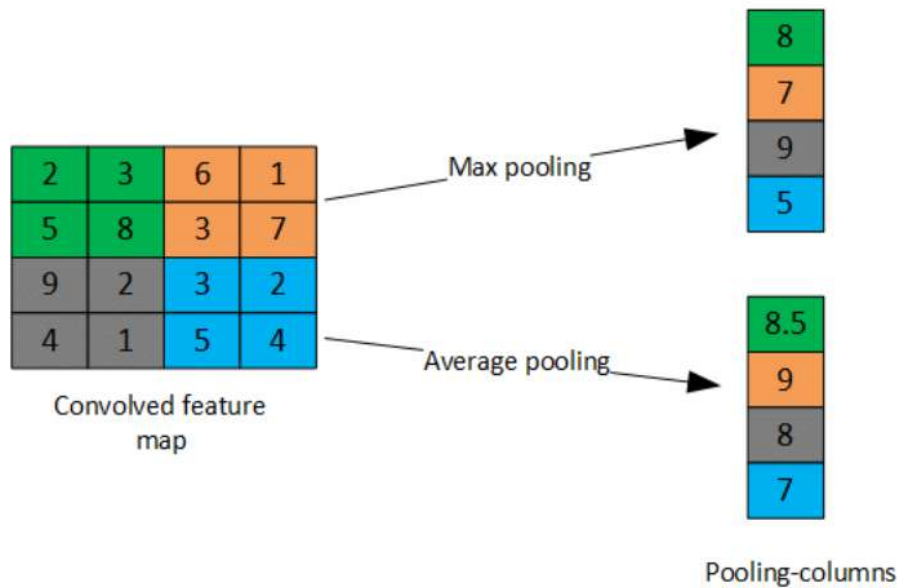
Input		Kernel		Output																																													
<table border="1" style="border-style: dashed; border-width: 2px;"> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>2</td><td>0</td></tr> <tr><td>0</td><td>3</td><td>4</td><td>5</td><td>0</td></tr> <tr><td>0</td><td>6</td><td>7</td><td>8</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>	0	0	0	0	0	0	0	1	2	0	0	3	4	5	0	0	6	7	8	0	0	0	0	0	0	*	<table border="1"> <tr><td>0</td><td>1</td></tr> <tr><td>2</td><td>3</td></tr> </table>	0	1	2	3	=	<table border="1"> <tr><td>0</td><td>3</td><td>8</td><td>4</td></tr> <tr><td>9</td><td>19</td><td>25</td><td>10</td></tr> <tr><td>21</td><td>37</td><td>43</td><td>16</td></tr> <tr><td>6</td><td>7</td><td>8</td><td>0</td></tr> </table>	0	3	8	4	9	19	25	10	21	37	43	16	6	7	8	0
0	0	0	0	0																																													
0	0	1	2	0																																													
0	3	4	5	0																																													
0	6	7	8	0																																													
0	0	0	0	0																																													
0	1																																																
2	3																																																
0	3	8	4																																														
9	19	25	10																																														
21	37	43	16																																														
6	7	8	0																																														

Fonte: Imagem adaptada de (ZHANG et al., 2021).

A saída dessa camada (ver Figura 6) é denominada mapa de recursos. Este mapa possui características sobre a imagem, como cantos e bordas e serve como entrada para outras camadas com a finalidade de aprender vários outros recursos da imagem de entrada (GUO et al., 2016).

3.1.2.2 Camada de *Pooling*

A camada de pooling em uma *CNN* é fundamental para realizar a redução da dimensionalidade das características extraídas, o que ajuda a diminuir a complexidade computacional e a evitar o *overfitting*. Esse processo é realizado por meio de operações como *max pooling* e *average pooling* (SIMONYAN; ZISSERMAN, 2014). A Figura 7 ilustra o processo de *max pooling* e *average pooling*.

Figura 7 – *Max pooling* e *average pooling*.

Fonte: Imagem adaptada de (ES-SABERY et al., 2021).

Essa camada condensa informações relevantes ao selecionar os valores máximos ou médios de áreas específicas da imagem, preservando apenas as características mais significativas. Essa técnica não só aumenta a eficiência do modelo, mas também proporciona invariância a pequenas alterações, como deslocamentos e distorções, tornando o reconhecimento de padrões em tarefas de visão computacional mais robusto e preciso.

3.1.2.3 Camada Totalmente Conectada

A camada totalmente conectada, conhecida como *Fully Connected* (FC), é uma parte essencial das redes neurais convolucionais e costuma aparecer nas etapas finais da rede. Diferente das camadas convolucionais, que focam em áreas específicas da imagem, a camada totalmente conectada estabelece ligações entre todos os neurônios da camada anterior e da próxima camada. Essa densidade de conexões permite que a rede combine as características extraídas anteriormente, como bordas e texturas, e faça previsões. Cada neurônio da camada FC processa informações globais da imagem, ajudando a rede a aprender relações complexas entre os dados e as classes de saída. O resultado é um vetor que indica a probabilidade de a imagem pertencer a uma determinada classe. Normalmente, essa camada é acompanhada por uma função de ativação, como a softmax, que converte os resultados em probabilidades, facilitando a interpretação final da rede, especialmente em tarefas de classificação (SIMONYAN; ZISSERMAN, 2014).

3.1.2.4 Dropout

Quando as camadas estão conectadas a uma camada totalmente conectada, significa que cada entrada possui uma associação com todos os neurônios da camada seguinte.

Isso pode acabar resultando em *overfitting*, que ocorre quando o modelo se adapta muito bem aos dados de treinamento, mas falha ao generalizar para novos dados, levando a um desempenho ruim em outras imagens de entrada. O *overfitting* prejudica a capacidade do modelo de lidar com dados desconhecidos, comprometendo sua eficácia em diferentes aplicações.

Para evitar esse problema, é utilizada a camada de *dropout*. Essa camada remove aleatoriamente alguns neurônios durante o treinamento, tornando o modelo mais leve. Ao eliminar temporariamente neurônios, o *dropout* força a rede a aprender com diferentes combinações de neurônios, melhorando sua capacidade de generalização. Um *dropout* de 0,2 implica que, para cada neurônio ativado, há uma probabilidade de 20% de ele ser zerado durante o treinamento. Quando a rede estiver em modo de teste, o *dropout* não é aplicado, mas os pesos das conexões são escalados para compensar essa omissão durante o treinamento.

3.1.2.5 Arquiteturas de Redes Convolucionais

Diversas arquiteturas de redes neurais se destacaram ao longo da história, tornando-se amplamente reconhecidas e frequentemente mencionadas devido à sua relevância. A seguir, serão apresentados e contextualizados alguns dos modelos mais representativos.

AlexNet. A AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) foi lançada em 2012, essa arquitetura inovou a área de visão computacional, vencendo a competição ImageNet com uma grande margem. Sua estrutura, composta por cinco camadas convolucionais seguidas de três camadas totalmente conectadas, foi inovadora na época. A AlexNet também popularizou o uso de GPUs para acelerar o treinamento de redes neurais profundas e introduziu técnicas como ReLU e *dropout* para melhorar o desempenho e a generalização. Ela foi um marco que impulsionou a adoção em larga escala de *CNNs* no reconhecimento de imagens.

VGG. A VGG16 e a VGG19 são arquiteturas desenvolvidas pelo *Visual Geometry Group* (VGG) da Universidade de Oxford (SIMONYAN; ZISSERMAN, 2014). Essas redes têm como principal destaque a simplicidade de sua estrutura, empregando pequenas camadas convolucionais compostas por filtros 3x3 dispostas em profundidade. A VGG16 contém 16 camadas treináveis, ao passo que a VGG19 possui 19, sendo a principal diferença o número de camadas convolucionais entre elas. Ambas as arquiteturas terminam com três camadas totalmente conectadas. O principal objetivo dessas redes é explorar a profundidade, o que facilita a extração de características visuais mais detalhadas, tornando-as altamente eficazes em tarefas de classificação de imagens, como no desafio do ImageNet.

Inception. A Inception (SZEGEDY et al., 2015) também conhecida como GoogLeNet, essa arquitetura foi lançada pela Google em 2014, introduzindo o módulo Inception, que inovou a maneira como as redes processam múltiplas escalas de características simulta-

neamente. Esse módulo faz uso de diferentes tamanhos de filtros em paralelo, permitindo desse modo que a rede faça a extração de informações de diferentes resoluções de forma eficiente. Além disso, a Inception reduziu o número de parâmetros sem sacrificar a precisão, tornando-a uma das arquiteturas mais eficientes da época e estabelecendo um novo padrão para redes profundas com alta capacidade de generalização.

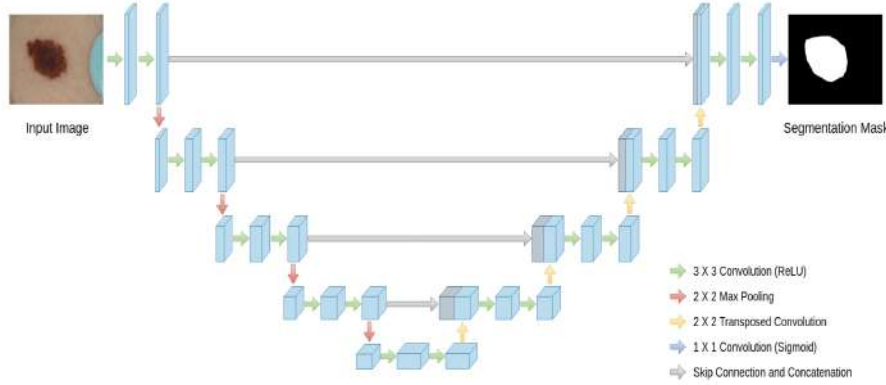
ResNet. A ResNet (*Residual Network*) (HE et al., 2016), foi introduzida pela Microsoft em 2015, a qual apresentou uma inovação ao incorporar blocos residuais, que permitem o fluxo direto de informações entre as camadas, facilitando o treinamento de redes muito profundas. Isso resolve o problema de degradação do gradiente, que antes limitava o desempenho de redes com muitas camadas. A ResNet foi apresentada a capacidade de treinar redes com mais de mil camadas, estabelecendo novos padrões de acurácia e influenciando a criação de várias outras arquiteturas baseadas em seu conceito.

EfficientNet. A EfficientNet (TAN; LE, 2019) é uma arquitetura desenvolvida pela Google em 2019, centrada em otimizar tanto a acurácia quanto a eficiência computacional. Ela faz uso de uma técnica chamada denominada *compound scaling*, que ajusta a profundidade, largura e a resolução da rede de forma equilibrada, permitindo um uso mais eficiente dos recursos em comparação com outras arquiteturas. Essa abordagem faz com que a EfficientNet se destaque por alcançar alta acurácia em *benchmarks* como o ImageNet, ao mesmo tempo em que reduz o número de operações de ponto flutuante necessárias para treinar e inferir a rede.

U-Net. A arquitetura U-Net é uma arquitetura de rede neural convolucional voltada para tarefas de segmentação de imagens biomédicas (IBTEHAZ; RAHMAN, 2020). Sua principal vantagem consiste na capacidade de produzir mapas de segmentação precisos, mesmo quando há uma quantidade limitada de dados disponíveis. A estrutura da U-Net é simétrica e consiste em dois caminhos principais: um de compressão, denominado *encoder*, que captura o contexto da imagem por meio de convoluções e operações de pooling, e outro de expansão, denominado *decoder*, que restaura os detalhes e a resolução espacial, utilizando camadas de *upsampling* e convoluções transpostas. O formato em "U" permite combinar informações de diferentes escalas, essencial para segmentar regiões pequenas e complexas em imagens. A Figura 8 ilustra a arquitetura completa da U-net. Uma das inovações fundamentais da U-Net é o uso de conexões de *skip* entre as camadas correspondentes do encoder e do decoder, o que facilita a recuperação de detalhes espaciais que poderiam ser perdidos durante o processo de compressão. Isso aprimora o desempenho em tarefas de segmentação, especialmente em áreas com diferenças sutis entre objetos e o fundo da imagem. A U-Net tem sido amplamente utilizada não apenas em imagens médicas, mas também em outros domínios da visão computacional, consolidando-se como um modelo de referência em segmentação de imagens.

Outra característica dessa arquitetura é sua função de ativação, que pode ser

Figura 8 – Arquitetura U-net.



Fonte: Imagem adaptada de (IBTEHAZ; RAHMAN, 2020).

representada pelo termo da função de perda ponderada para tarefas de segmentação, considerando a energia de ativação baseada na entropia cruzada, conforme descrito na Equação 3.5.

$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(x)) \quad (3.5)$$

Onde Ω é o conjunto de todos os pixels da imagem, x representa as coordenadas de um pixel, $w(x)$ é o peso aplicado ao pixel x , e $p_{l(x)}(x)$ é a probabilidade predita para o pixel x pertencente à classe $l(x)$.

Embora tenham um design simples, essas redes demandam elevado poder computacional e mais memória devido ao grande número de parâmetros. Ainda assim, a VGG16 e a VGG19 são amplamente utilizadas como bases para inúmeras aplicações em visão computacional, sendo frequentemente empregadas em transfer learning como redes pré-treinadas. Elas demonstraram que o aumento da profundidade pode melhorar significativamente o desempenho, tornando-se um marco na evolução de *CNNs* mais avançadas e eficazes.

NFNet. A NFNet (*Normalized-Free Network*) (BROCK et al., 2021) é uma arquitetura que dispensa a necessidade de normalização por lotes. Em vez disso, a NFNet faz uso de uma técnica denominada *Scaled Weight Standardization* para garantir a estabilidade no treinamento, mesmo sem a normalização tradicional. Isso torna a rede mais eficiente em termos de memória e velocidade, especialmente para lotes pequenos.

A NFNet foi desenvolvida para tarefas de visão computacional, como a classificação de imagens, e mostrou resultados de alto desempenho, superando modelos populares como a EfficientNet no *benchmark* ImageNet. Além disso, suas versões escaláveis permitem ajustar o equilíbrio entre acurácia e custo computacional, tornando-a adequada para diferentes cenários de aplicação.

3.2 Inteligência Artificial Explicável

A Inteligência Artificial Explicável (XAI) é um ramo que tem como objetivo tornar os modelos de inteligência artificial mais compreensíveis e transparentes para os humanos, respondendo à complexidade dos sistemas que frequentemente operam como "caixas-pretas" (ADADI; BERRADA, 2018).

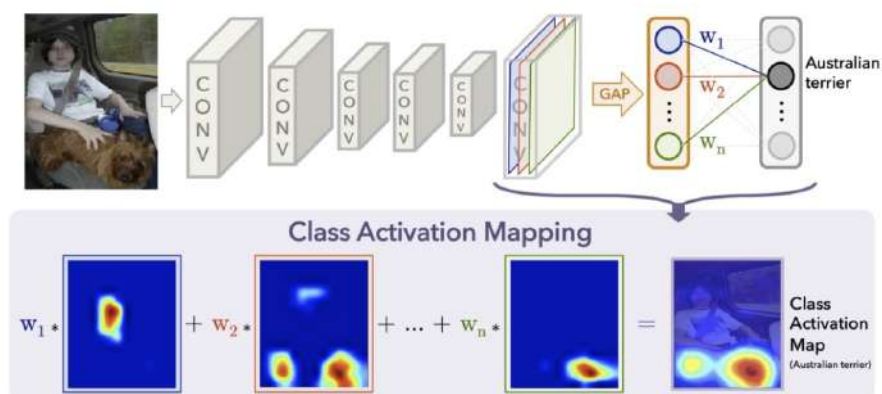
Sua importância consiste em aumentar a confiança dos usuários, promover a responsabilidade nas decisões automatizadas e ajudar a identificar e mitigar vieses nos dados. Técnicas como modelos interpretáveis e visualizações são utilizadas para oferecer explicações sobre como os modelos tomam decisões. Seu uso ganha destaque principalmente quando utilizado em aplicações médicas, uma vez que esses métodos aumentam a confiança dos médicos ao permitir que eles compreendam como o modelo chegou a determinadas conclusões, o que é essencial para a aceitação da tecnologia em ambientes clínicos (TJOA; GUAN, 2020).

3.2.1 Grad-CAM

O Grad-CAM (*Gradient-weighted Class Activation Mapping*) é uma técnica utilizada para tornar as decisões das CNNs mais interpretáveis (SELVARAJU et al., 2017). Essa técnica cria mapas de ativação que indicam quais regiões de uma imagem são mais importantes para a rede ao realizar uma classificação.

Para gerar esses mapas, o Grad-CAM utiliza os gradientes da função de perda em relação às ativações das últimas camadas convolucionais, conforme ilustrado na Figura 9. Esses gradientes são usados para calcular a relevância de cada neurônio, destacando as áreas da imagem que influenciaram mais a decisão da rede.

Figura 9 – Construção do mapa de ativação do Grad-CAM.



Fonte: Imagem adaptada de (ZHOU et al., 2016).

O resultado final é um mapa de calor que pode ser sobreposto à imagem original, auxiliando na interpretação visual. O Mapa de ativação é descrito pela Equação 3.6:

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad (3.6)$$

Onde A^k representa o mapa de características da k -ésima camada convolucional, e α_k^c são os pesos de importância para a classe c , calculados como a média dos gradientes sobre o mapa de características representados na Equação 3.7:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (3.7)$$

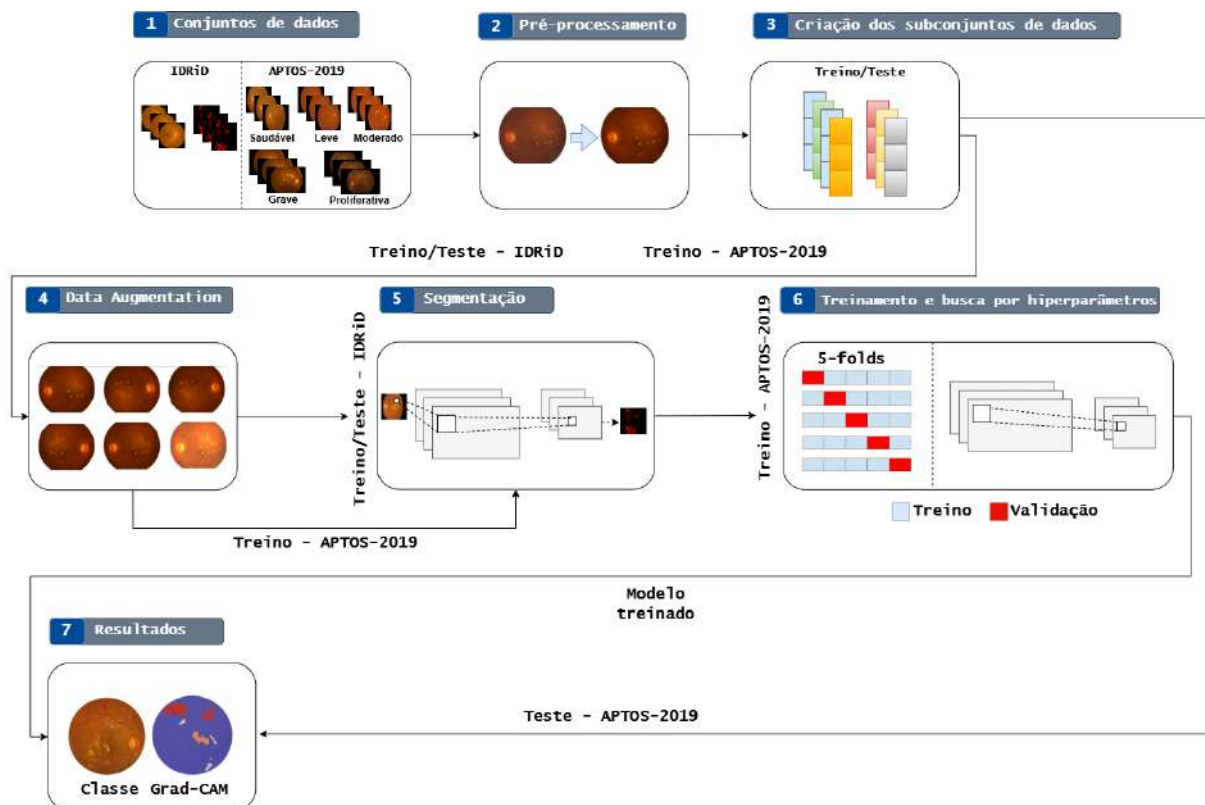
Sendo y^c a pontuação da classe c e Z o número total de elementos no mapa de características A^k . Em imagens médicas, é uma técnica fundamental para interpretar as decisões de CNN, permitindo a visualização das regiões mais relevantes de uma imagem para uma predição. Ele gera mapas de calor que destacam áreas de interesse, como lesões, tumores ou anomalias, que influenciaram o diagnóstico da rede.

Em aplicações como a detecção de RD, o Grad-CAM pode mostrar, por exemplo, quais áreas da retina foram consideradas ao identificar exsudados ou hemorragias, facilitando a validação e melhorando a confiança no modelo treinado.

4 Metodologia

Este capítulo apresenta a abordagem proposta para a segmentação das lesões decorrentes da RD e sua subsequente classificação. Inicialmente, é realizada uma breve explanação sobre os conjuntos de dados IDRiD (*Indian Diabetic Retinopathy Image Dataset*) (PORWAL et al., 2020) e APTOS-2019 (*Asian Pacific Tele-Ophthalmology Society*) (KAGGLE, 2019), que são utilizados para as tarefas de segmentação e classificação, respectivamente. Posteriormente, a etapa de pré-processamento é detalhada. Em seguida, realiza-se o treinamento da U-Net no conjunto IDRiD para segmentação de lesões, considerando abordagens individuais e conjuntas. Após o treinamento, as imagens do APTOS-2019 são segmentadas, e divididas em subconjuntos de treino e teste. Os resultados são avaliados utilizando métricas selecionadas, e a melhor arquitetura é escolhida com base em testes estatísticos para garantir uma comparação robusta e confiável. A metodologia utilizada é ilustrada na Figura 10.

Figura 10 – Diagrama da abordagem proposta.



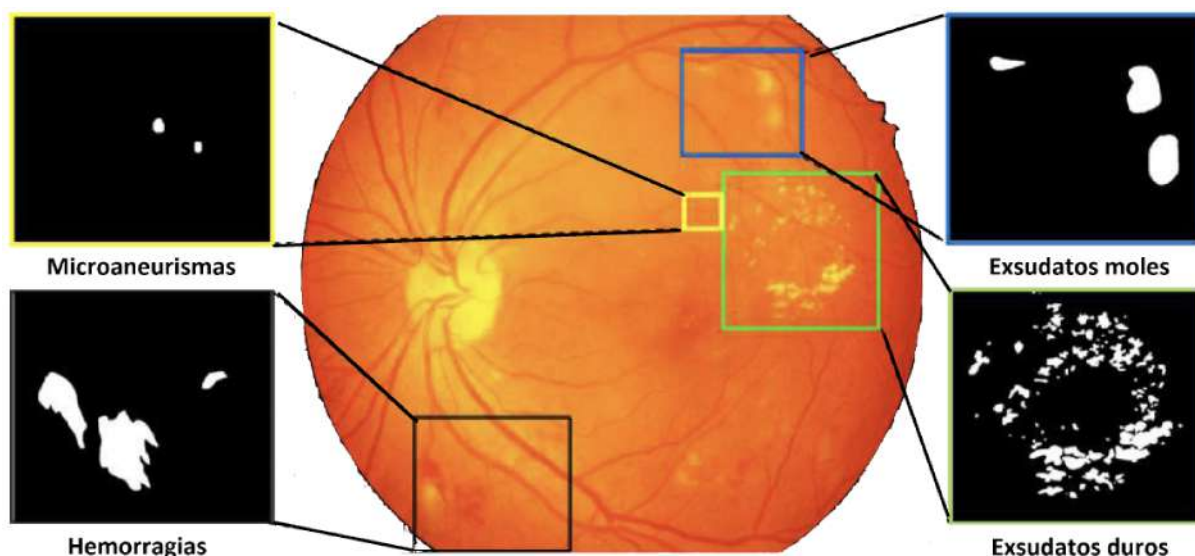
Fonte: Autor.

Este capítulo está dividido da seguinte forma: a Seção 4.1 descreve os conjunto de dados utilizados, a Seção 4.2 apresenta as etapas de pré-processamento aplicadas, a Seção 4.3 detalha a etapa de segmentação das lesões, a Seção 4.4 expõe o processo de classificação, e a Seção 4.5 apresenta as métricas empregadas para avaliar os resultados das segmentações e classificações das retinografias bem como os respectivos testes estatísticos empregados.

4.1 Conjuntos de Dados

Para realizar a segmentação das lesões, foi utilizado o conjunto de dados IDRiD. Esse conjunto de dados contém retinografias, capturadas usando uma câmera de fundo de olho, e suas respectivas marcações de lesões, como microaneurismas, hemorragias, exsudatos moles e exsudatos duros que foram feitas de forma manual por oftalmologistas. As imagens desse conjunto possuem uma resolução de 4288 pixels por 2848 pixels, com 3 canais (RGB). O conjunto de dados foi inicialmente dividido em conjunto de treino e teste, com 54 e 27 imagens, respectivamente.

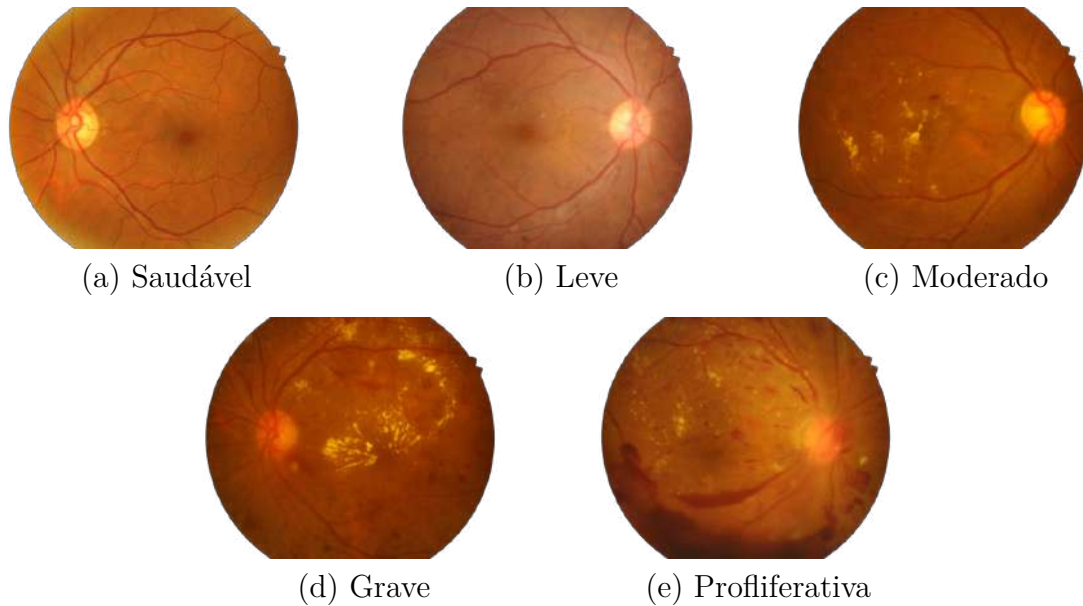
Figura 11 – Amostras de lesões e suas respectivas anotações do conjunto de dados IDRiD.



Fonte: Imagem adaptada de (PORWAL et al., 2020).

O conjunto de dados APTOS-2019 é o conjunto de dados aberto composto por 3662 imagens rotuladas de retinas de pacientes diabéticos tratados no Departamento de Oftalmologia do Hospital da Universidade de Firat, na Turquia. A equipe de médica do hospital classificou as imagens em cinco categorias distintas: 1805 saudáveis (49,20%); 370 com RD leve (10,10%); 999 com RD moderada (27,30%); 193 com RD severa (5,30%); e 295 com RD proliferativa (8,10%). As imagens apresentam resoluções variadas. As imagens desse conjunto de dados foram geradas por uma câmera de fundo de olho. Na Figura 12, é possível observar amostras de cada classe do conjunto de dados APTOS-2019.

Figura 12 – Amostras de cada classe do conjunto de dados APTOS-2019.



Fonte: Autor.

Conforme é possível verificar, existem classes que possuem elementos sutis que caracterizam, por exemplo, a classe de RD com grau leve, assim como existem elementos bastante visíveis, como nas classes graves e proliferativas, que apresentam uma grande concentração de lesões.

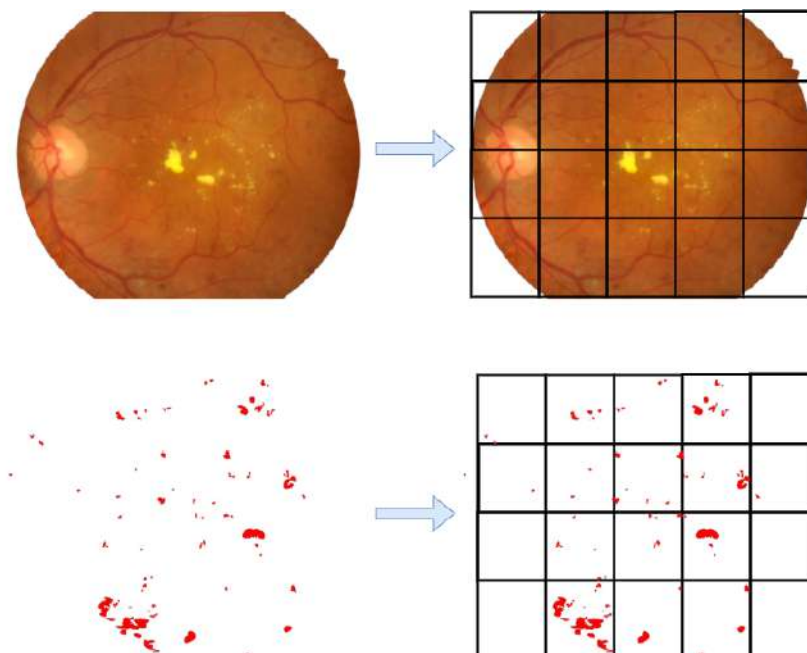
4.1.1 *Data Augmentation*

O *data augmentation* é uma técnica bastante utilizada em aprendizado de máquina para aumentar a diversidade dos dados de treinamento aplicando transformações como rotações, espelhamentos e ajustes de brilho. Isso melhora a generalização dos modelos, reduz o *overfitting* e é especialmente útil em áreas como visão computacional e saúde, onde dados rotulados podem ser limitados. Ambos os conjuntos de dados utilizados necessitavam da aplicação de *data augmentation*. A seguir, serão detalhadas as abordagens aplicadas a cada conjunto de dados.

4.1.1.1 Segmentação

Inicialmente, como a quantidade de imagens presente no conjunto de dados utilizado é bastante reduzida e como a resolução de 4288 pixels $\times$ 2848 pixels $\times$ 3 canais (RGB) é alta, as imagens e suas respectivas máscaras foram divididas em subimagens com resolução de 512 pixels $\times$ 512 pixels por 3 canais para a rede U-Net. A Figura 13 ilustra o processo de divisão aplicado nas imagens e em suas respectivas anotações.

Figura 13 – Esquema de subdivisão da retinografia e da anotação em subimagens.



Fonte: Autor.

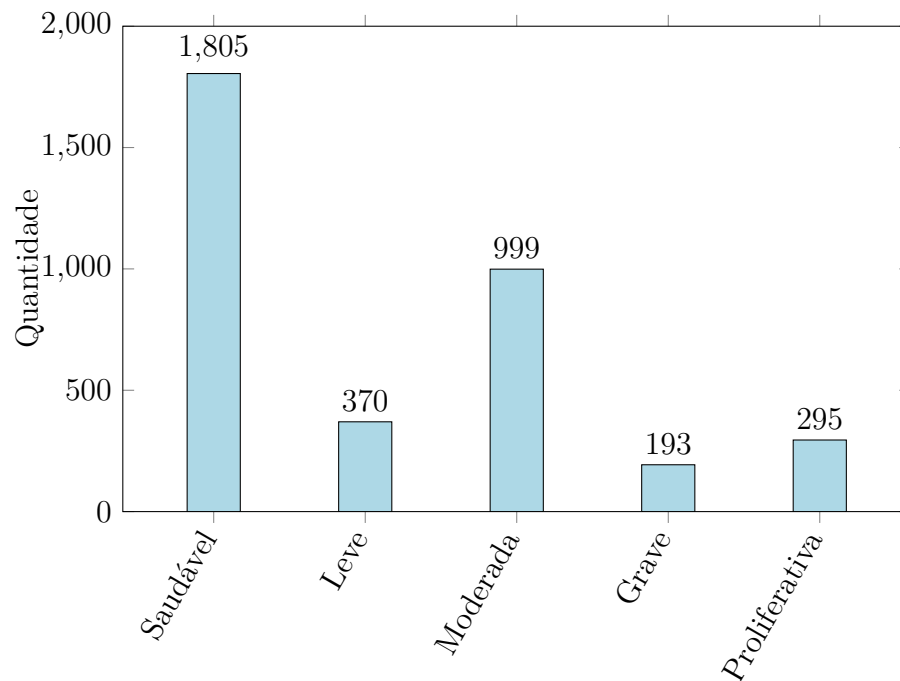
A divisão de uma única imagem em subimagens menores gerou imagens que não continham informações relevantes para o processo de aprendizagem, tanto na imagem original quanto na máscara, resultando em matrizes nulas. Para otimizar o processo de treinamento, as subimagens e suas respectivas máscaras foram descartadas, mantendo apenas aquelas com informações significativas. Como resultado, o conjunto inicial de 54 imagens foi ampliado para um total de 2004 novas imagens, enquanto o conjunto de teste, que consistia em 27 imagens, foi aumentado para 1005 imagens.

4.1.1.2 Classificação

Antes da realização da etapa de treinamento das *CNNs* para realizar a classificação, foi aplicado *data augmentation* no conjunto de classificação, uma vez que ele possui classes com um certo desbalanceamento. A Figura 14 apresenta a quantidade de imagens em cada classe.

Para balancear o conjunto de dados composto por imagens saudáveis e suas classes seus respectivos graus de RD, foi aplicada uma série de técnicas de *data augmentation* (TIAN et al., 2018) objetivando equilibrar o número de amostras entre as classes. Esse processo foi aplicado apenas no conjunto destinado ao treinamento, ou seja, usando 80% do total de imagens do conjunto de dados (a divisão entre treinamento e teste é melhor explicada em Seção 5.2). Inicialmente, a classe Saudável possuía 1444 imagens, enquanto as classes RD Leve, RD Moderada, RD Severa e RD Proliferativa tinham 296, 799, 154 e 236 imagens, respectivamente. Como a classe saudável já era a mais numerosa, seu número de imagens sem alterações e sua quantidade total foram definidos como o total alvo para as

Figura 14 – Distribuição de RD no Conjunto de Dados Analisado.



Fonte: Autor.

demais classes, utilizando *data augmentation* para elevar cada uma até aproximadamente 1444 imagens.

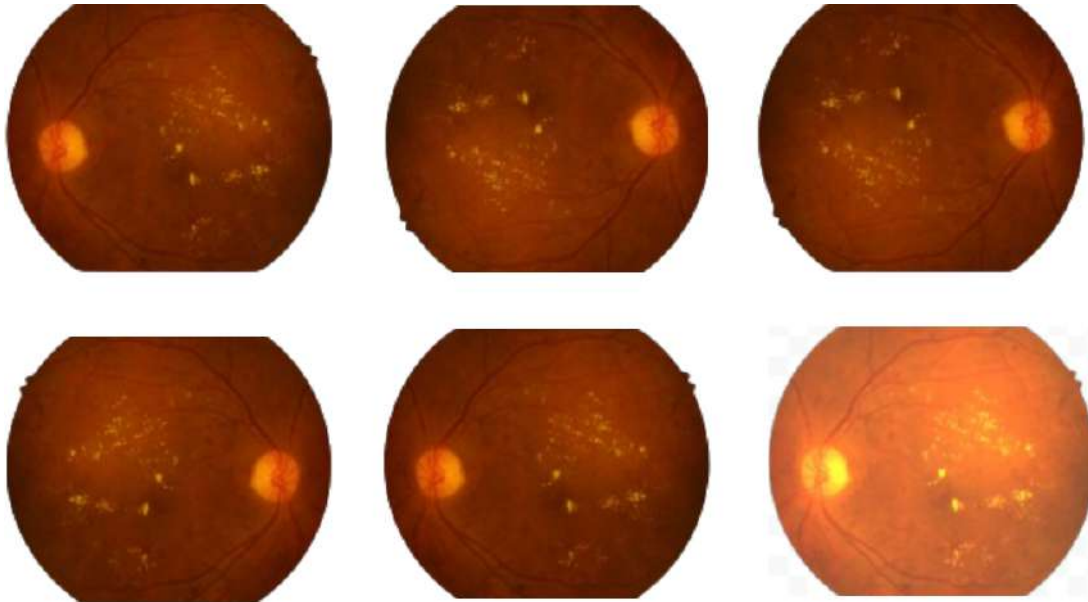
Para a classe RD Leve, com 296 imagens iniciais, foram aplicadas transformações para gerar aproximadamente 1148 novas imagens. As transformações incluíram rotações dentro do intervalo de ± 10 a ± 20 graus, representando 40% das imagens aumentadas, flips horizontais em 25% das amostras, variações de zoom de 0,9x a 1,1x em 15%, e alterações de brilho e contraste de $\pm 15\%$ e $\pm 10\%$, respectivamente, para os 20% restantes. A classe RD Moderada, com 799 imagens iniciais, recebeu 644 novas amostras distribuídas entre rotações de ± 10 a ± 20 graus em 30% das imagens, flips horizontais e verticais em outros 30%, e ajustes de brilho ($\pm 15\%$) e contraste ($\pm 10\%$) em 20%, além de zooms leves entre 0,9x e 1,1x aplicados em 20% das novas imagens.

Para a classe RD Severa, que contava com apenas 154 imagens, o conjunto foi aumentado para 1289 novas imagens. Aplicando rotações mais amplas, de ± 20 a ± 30 graus, em 40% das imagens aumentadas, flips horizontais e verticais em 30%, variações de brilho de $\pm 20\%$ e de contraste de $\pm 15\%$ em 15%, e zooms entre 0,85x e 1,15x para os 15% finais. Finalmente, para a classe DR Proliferativa, com 236 imagens iniciais, resultando em 1208 novas imagens usando rotações entre ± 10 e ± 15 graus para 35% das imagens, flips horizontais em 25%, ajustes de brilho ($\pm 15\%$) e contraste ($\pm 10\%$) para 20% das novas amostras, e zooms leves de 0,9x a 1,1x aplicados nos 20% restantes.

Com essa estratégia, o conjunto de imagens foi balanceado para que cada classe contivesse aproximadamente 1444 imagens, resultando em um conjunto de dados balanceado

de 7229 imagens no total. A Figura 15 ilustra exemplos de imagens resultantes após a aplicação das transformações citadas.

Figura 15 – Exemplos de transformações aplicadas: rotações, flip Horizontal, Zoom e ajuste de brilho/contraste.



Fonte: Autor.

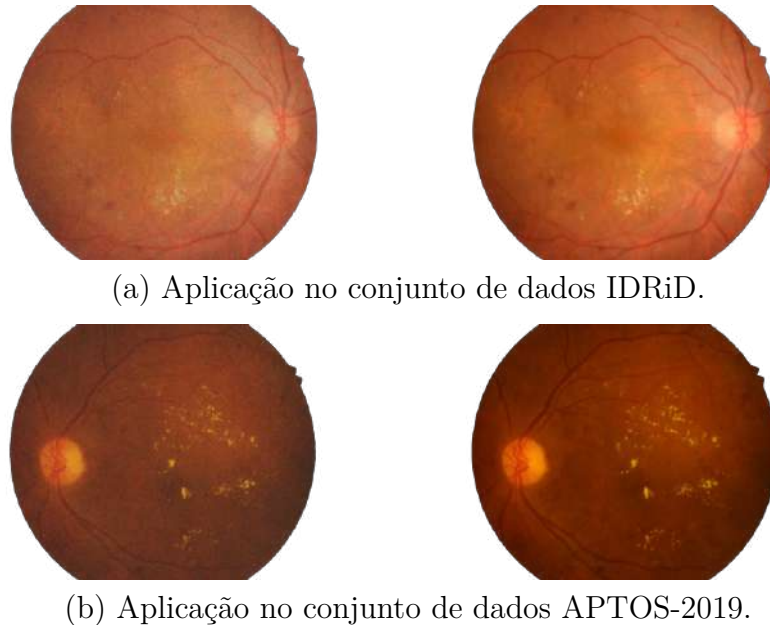
O uso desse método permitiu uma representação mais equitativa das diferentes classes de RD, mitigando o problema de desbalanceamento e proporcionando um conjunto mais robusto para o treinamento do modelo.

4.2 Pré-Processamento

Imagens médicas podem ser contaminadas por ruído devido a falhas nos equipamentos ou condições ambientais. Esse ruído compromete a qualidade da imagem, dificultando a identificação de padrões importantes e afetando a precisão de diagnósticos (SILVA; SOUZA; COSTA, 2014). Nesse contexto, antes de submeter as imagens às *CNNs*, é de vital importância aplicar um pré-processamento nas imagens. Para solucionar esse problema, foi utilizado o filtro da média com *kernel* 5×5 . Posteriormente, as imagens foram normalizadas entre 0 e 1. A Figura 16 apresenta o resultado do pré-processamento em imagens de ambos os conjuntos de dados.

O filtro da média é amplamente utilizado para reduzir ruídos aleatórios em imagens, como pontos brancos ou pretos causados por imperfeições de sensores ou compressão. A remoção de ruídos melhora a qualidade das imagens, permitindo que a *CNN* foque em padrões mais relevantes, como bordas e texturas, em vez de ruídos não desejados. Além disso, reduzir o ruído e suavizar a imagem ajuda a *CNN* a extrair características mais consistentes (CHEN; XU; ZHANG, 2021). Isso pode levar a uma melhor generalização do

Figura 16 – Resultado da aplicação do pré-processamento nos conjuntos de dados IDRiD e APTOS-2019.



Fonte: Autor.

modelo para dados não vistos, já que as variações desnecessárias na textura e no brilho são minimizadas.

A normalização das imagens garante que os dados de entrada estejam em uma faixa controlada, como $[0, 1]$, o que mantém os gradientes em uma faixa mais estável (YIN et al., 2019), evitando esses problemas e permitindo atualizações de pesos mais eficientes, isso proporciona aceleração na convergência do treinamento das *CNNs*, essa normalização é comumente adotada por diversos autores por convenção.

A realização desses passos é de vital importância para o sucesso da realização das tarefas de segmentação e classificação das imagens retinográficas, evitando erros no processo de aprendizagem devido à presença de elementos indesejados. Além disso, o processo de treinamento é otimizado graças à normalização aplicada.

4.3 Segmentação de Lesões

O processo de segmentação foi realizado em três etapas: a primeira consiste no treinamento da U-Net, seguindo pela etapa de testes, por fim, ocorre a repetição das etapas anteriores. O processo de treinamento e teste do modelo foi realizado em três etapas principais. Na etapa de treinamento do modelo, o conjunto de treino e o conjunto de teste obtidos após a aplicação do *data augmentation* são unidos em um único conjunto, sendo que 80% dos dados foram utilizados para essa finalidade. Para definir os hiperparâmetros ideais, foi realizada uma validação cruzada de 5-*folds* durante o treinamento. A função de

perda utilizada foi a *Cross-Entropy* (BISHOP, 2006). A Tabela 1 apresenta um resumo dos hiperparâmetros considerados. Na etapa de teste do modelo, os 20% restantes do subconjunto de dados foram utilizados para testar os modelos salvos, e as métricas de desempenho foram calculadas nesta fase.

Tabela 1 – Configuração para busca de hiperparâmetros da U-Net.

Parâmetro	Configuração
Taxa de Aprendizado	0,001 - 0,1
Função de Ativação	<i>sigmoid</i> / <i>softmax</i>
Tamanho do Lote	16, 32 e 64
Otimizador	Adam e Adamax
Número de Épocas	Usando a técnica de <i>Early Stopping</i>

Fonte: Autor.

Para a técnica de *Early Stopping*, quando não houver uma redução significativa na função de perda durante 5 épocas consecutivas, o processo de treinamento será interrompido, e os melhores pesos obtidos até o momento serão considerados. Por fim, os processos de treinamento e teste foram repetidos, com os subconjuntos de dados sendo divididos aleatoriamente em novos conjuntos de treinamento e teste, gerando novos dados para análise. Essa repetição foi realizada dez vezes, contemplando todas as etapas descritas anteriormente.

4.4 Classificação

As arquiteturas de *CNN* utilizadas foram pré-treinadas e inicializadas com o conjunto de dados de imagens ImageNet (RUSSAKOVSKY et al., 2015), que contém 1000 classes com mais de 1 milhão de amostras. As *CNNs* foram adaptadas para o problema substituindo a última camada totalmente conectada por uma nova camada totalmente conectada correspondente ao número de classes presentes no conjunto de dados utilizado. Na Tabela 2, são apresentadas as arquiteturas de *CNNs* utilizadas e suas respectivas configurações.

A classificação é realizada em três etapas: a primeira etapa consiste no treinamento das arquiteturas. Nesta etapa, utilizamos 80% do subconjunto de dados para treinamento. Consideramos as configurações de hiperparâmetros apresentadas na Tabela 3 para determinar a configuração ideal dos classificadores. Para encontrar os hiperparâmetros ideais, realizamos uma busca aleatória com 10 iterações. Para definir os hiperparâmetros ideais, foi realizada uma validação cruzada de 5-*folds* durante o treinamento. A função de perda utilizada foi a *Cross-Entropy*.

Para a técnica de *Early Stopping*, quando não houver uma redução significativa na função de perda durante 5 épocas consecutivas, o processo de treinamento será interrompido,

Tabela 2 – Arquiteturas de *CNN*, configurações utilizadas e tamanho de imagem de entrada.

Arquitetura	Configuração	Tamanho da Imagem de Entrada
VGG	VGG16	224x224
	VGG19	224x224
EfficientNet	EfficientNetB0	224x224
	EfficientNetB1	224x224
DenseNet121	DenseNet121	224x224
	DenseNet201	224x224
NFNet	NFNet	224x224
ResNet	ResNet50	224x224
Inception	InceptionV3	299x299
AlexNet	AlexNet	227x227

Fonte: Autor.

Tabela 3 – Configuração para busca de hiperparâmetros das *CNNs*

Parâmetro	Configuração
Taxa de Aprendizado	0,001-0,1
Função de Ativação	<i>softmax</i>
Tamanho do Lote	32-64
Otimizador	Adam, RMSprop e Adamax
Número de Épocas	Usando a Técnica <i>Early Stopping</i>

Fonte: Autor.

e os melhores pesos obtidos até o momento serão considerados. Na segunda etapa, é realizado o processo de teste com os 20% restantes do subconjunto de dados, utilizando as arquiteturas treinadas. Nessa fase, são calculadas as métricas de avaliação de cada arquitetura. Por fim, os subconjuntos de dados são novamente divididos aleatoriamente em novos conjuntos de treinamento e teste, gerando novos subconjuntos. Em seguida, são realizadas dez repetições das etapas anteriores.

4.5 Métricas de Avaliação

A saída de cada *CNN* é avaliada usando a matriz de confusão, da qual são derivadas as métricas: acurácia, sensibilidade, precisão e F1-Score. Além disso, foi utilizado o *Log loss*. Para a avaliação específica dos resultados provenientes da segmentação, foi utilizada a métrica IoU (Interseção sobre a União).

A acurácia descreve a proporção de predições corretas feitas pelo modelo em relação ao total de predições realizadas. Em termos simples, é uma métrica que indica a eficiência geral do modelo em classificar as imagens corretamente em suas respectivas

classes. A expressão 4.1 descreve como essa métrica é calculada:

$$Acurácia(\%) = \frac{VP + VN}{VN + FN + VP + FP} \times 100 \quad (4.1)$$

A sensibilidade é uma métrica de avaliação de desempenho de um modelo de classificação, que indica a capacidade do modelo em identificar corretamente os casos positivos. Em outras palavras, a sensibilidade mede a habilidade do modelo em detectar corretamente as instâncias de RD, matematicamente, a sensibilidade é definida pela expressão 4.2:

$$Sensibilidade = \frac{VP}{VP + FN} \quad (4.2)$$

A precisão consiste em uma métrica de avaliação de desempenho de um modelo de classificação, que mede a capacidade do modelo em identificar corretamente as instâncias positivas entre todas as instâncias que ele classificou como positivas. Em outras palavras, a precisão indica, das vezes que o modelo previu que a condição estava presente, quantas vezes essa previsão estava correta. A expressão 4.3 descreve como essa métrica é calculada:

$$Precisão = \frac{VP}{VP + FP} \quad (4.3)$$

O F1-Score pode ser descrito como a média harmônica entre Sensibilidade e Precisão; essa métrica fornece um número que sugere a qualidade geral da abordagem, matematicamente, a sensibilidade é definida pela expressão 4.4:

$$F1 - Score(\%) = 2 \times \frac{Sensibilidade \times Precisão}{Sensibilidade + Precisão} \times 100 \quad (4.4)$$

O *log loss* é uma métrica crucial para avaliar modelos de classificação, medindo a disparidade entre as previsões do modelo e os rótulos verdadeiros dos dados. Essa métrica é calculada como o logaritmo negativo da probabilidade atribuída à classe correta pelo modelo. A função logarítmica utilizada penaliza previsões incorretas feitas com maior confiança de forma mais severa. Um valor de *Log Loss* mais baixo indica um desempenho superior do modelo. Na expressão 4.5, N representa o número total de amostras, K denota o número total de classes, $y_{i,c}$ é um valor binário que indica se a amostra i pertence à classe c , e $p_{i,c}$ é a probabilidade prevista pelo modelo para a classe c .

$$Log Loss = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^K y_{i,c} \log(p_{i,c}) \quad (4.5)$$

Para analisar os resultados do processo de segmentação, utilizou-se também a métrica Interseção sobre a União (IoU). Seu uso é amplamente difundido em tarefas de segmentação de imagens para avaliar o desempenho de um modelo. Ela mede a sobreposição entre duas áreas: a região prevista pelo modelo (previsão) e a região real (rótulo verdadeiro). O valor do IoU varia de 0 (sem sobreposição) a 1 (sobreposição perfeita). Essa métrica é

útil para avaliar quão bem o modelo conseguiu prever a forma e a localização dos objetos ou regiões em uma imagem. Na equação, A representa a região da previsão. B representa a região do rótulo verdadeiro. $|A \cap B|$ é a área da interseção entre as duas regiões e $|A \cup B|$ é a área da união das duas regiões.

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (4.6)$$

Além das métricas mencionadas anteriormente, é importante analisar os tempos de treinamento e teste ao avaliar uma *CNN*. O tempo de treinamento refere-se ao intervalo necessário desde o início do processo de treinamento da *CNN* até que o modelo esteja pronto para realizar a classificação. Por outro lado, o tempo de teste é o período que o classificador leva para prever a classe de uma imagem após receber seu vetor de características. Assim, o tempo de treinamento desempenha um papel crucial na fase de construção do modelo, sendo essencial para sua eficiência e desempenho.

4.5.1 Testes Estatísticos

Como forma de avaliar os resultados produzidos pelas *CNNs* e verificar aquela que apresentou o melhor desempenho com base em sua acurácia, considerando as iterações realizadas, são utilizados os testes estatísticos de Shapiro-Wilk, ANOVA e Tukey. A realização desses testes permite verificar se os resultados obtidos para a métrica de acurácia seguem uma distribuição que pode ser considerada normal, se as *CNNs* têm desempenhos semelhantes e, posteriormente, quais apresentam desempenhos estatisticamente superiores ou inferiores.

4.5.1.1 Teste de Shapiro-Wilk

O teste de Shapiro-Wilk (SHAPIRO; WILK, 1965) é um teste de normalidade nas estatísticas frequentistas, o qual testa a hipótese nula de que uma amostra y é proveniente de uma população normalmente distribuída. A estatística W é definida como:

$$W = \frac{B^2}{s^2} = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.7)$$

Em que y_i é a variável aleatória observada e a_i são coeficientes tabelados. Para calcular W , de uma amostra aleatória de tamanho n , dada por $y_1, y_2, \dots, y_n$, procede-se da seguinte forma: primeiramente, ordenam-se as observações em ordem crescente, ou seja, $y_1 \leq y_2 \leq \dots \leq y_n$. Em seguida, calcula-se a soma dos desvios das amostras em relação à média, representada por s^2 . Caso o número de observações n seja par, ou seja, $n = 2k$, calcula-se B pela expressão:

$$B = \sum_{i=1}^n a_{n-1+i} (y_{n-1+i} - y_i) \quad (4.8)$$

Se n for ímpar, ou seja, $n = 2k + 1$, os cálculos permanecem os mesmos, com a única exceção de que $a_{k+1} = 0$. Os valores de a_{n-i+1} são tabelados. Por fim, calcula-se W .

A hipótese nula desse teste é que a população está normalmente distribuída. Assim, se o valor de p for menor que o nível escolhido, então a hipótese nula é rejeitada e há evidências de que os dados testados não são normalmente distribuídos. Por outro lado, se o valor de p for maior do que o nível alfa escolhido, a hipótese nula (de que os dados vieram de uma população normalmente distribuída) não pode ser rejeitada (por exemplo, para um nível de confiabilidade de 0.05, um conjunto de dados com um valor de p inferior a 0.05 rejeita a hipótese nula de que os dados são de uma população normalmente distribuída).

4.5.1.2 Teste ANOVA

O teste de ANOVA (STIGLER, 1990) é bastante útil para realizar a comparação entre três ou mais grupos com amostras independentes. Nesse tipo de teste é possível utilizar variáveis quantitativas ou qualitativas e como requisito para sua aplicação tem-se que os dados devem pertencer a uma distribuição normal. Durante o processo é comparada a variação entre os grupos e dentro deles. Assim como no teste de hipótese tem-se duas possibilidades as quais são: H_0 : não há diferença estatística e H_1 : existe diferença estatística. Dada uma quantidade M de grupos com uma quantidade de amostras N , a aplicação do teste consiste na realização dos seguintes passos:

- Inicialmente calcula-se a soma dos quadrados de cada grupo (SSG) em relação a média geral;
- O passo seguinte consiste em calcular os graus de liberdade (DFG), dado pela quantidade de grupos de menos uma unidade;
- Em seguida calcula-se o erro quadrático em relação à média de cada grupo (SSE);
- O passo seguinte consiste em calcular o grau de liberdade para o grupo (DFE), dado por $DFE = (N - 1) \times M$;

Com todos os valores calculados, por fim, calcula-se o valor de F por meio da seguinte equação:

$$F = \frac{\frac{SSG}{DFG}}{\frac{SSE}{DFE}}. \quad (4.9)$$

Esse valor representa a relação entre a variabilidade explicada pelo modelo (entre grupos) e a não explicada (dentro dos grupos). Após calcular o F -valor, ele é comparado

com o valor crítico obtido a partir da tabela F (ver Apêndice 7), considerando os graus de liberdade entre os grupos e dentro dos grupos.

A interpretação é feita com base no p -valor:

- Se $p \leq \alpha$ (geralmente $\alpha = 0.05$), rejeita-se a hipótese nula (H_0), indicando diferenças significativas entre os grupos.
- Caso contrário, se $p > \alpha$, não há evidência suficiente para rejeitar H_0 .

4.5.1.3 Teste de Tukey

Conforme citado em (ABDI; WILLIAMS, 2010), a análise da variância permite verificar se existe alguma diferença significativa entre as médias entre os grupos de dados. Para o caso em que há pelo menos um par de médias com diferença significativa faz-se necessário o uso de algum teste que possibilite comparar médias. Nesse contexto, é possível utilizar o teste de Tukey que permite então comparar a totalidade dos contrastes entre duas médias. Testes comparativos desse tipo baseiam-se na dms (diferença mínima significativa), representando o menor valor que a estimativa de um contraste, diferença entre duas médias, deve apresentar para que se possa considerá-lo como significativo. Um teste é dito mais conservador que outro quando a dms dele é maior, pois ele tende a "conservar" a hipótese de igualdade entre as médias como verdadeira.

De forma a realizar o teste a um nível α de significância é necessário que inicialmente, sejam anunciadas as hipóteses H_0 e H_1 . Em seguida, obtêm-se as estimativas dos contrastes, dadas por $\hat{C} = \hat{\mu}_i - \hat{\mu}_j$. Após isso, calcula-se a diferença mínima significativa (dms). Por fim, se $|\hat{C}| \geq \text{dms}$, rejeita-se H_0 ; caso contrário, não se rejeita H_0 .

O teste de Tukey é válido para a totalidade dos contrastes para duas médias, exigindo um balanceamento dos dados. É considerado um teste exato para testar a maior diferença, nos demais casos é conservador. Por tratar-se de um teste rigoroso, geralmente, é aplicado um nível $\alpha = 0.05$.

5 Resultados e discussões

Nesta seção, são discutidos os resultados obtidos ao longo desta dissertação, que teve como foco a segmentação de lesões retinianas (exsudados moles, exsudados duros, hemorragias e microaneurismas) e a subsequente classificação dos tipos de retinopatia diabética. Além disso, foi aplicado o método Grad-CAM para avaliar a interpretabilidade dos modelos desenvolvidos, buscando compreender melhor o processo de tomada de decisão do modelo de classificação.

A apresentação dos resultados deste capítulo está organizada em duas seções principais. Na primeira seção, são apresentados e discutidos os resultados da segmentação das retinografias, acompanhados de comparações com trabalhos relacionados na literatura. Na segunda seção, são detalhados os resultados da classificação das retinografias, obtidos por meio de CNNs. Nesta seção, também são avaliadas as métricas de desempenho, os tempos de processamento assim como a discussão dos resultados obtidos pelo método Grad-CAM, identificando as regiões das imagens consideradas relevantes para as classificações realizadas e verificando se essas áreas realmente apresentam relevância clínica para uma classificação precisa, por fim, são realizadas comparações com outros trabalhos relacionados.

A metodologia aplicada foi implementada em um computador com sistema operacional macOS, munido com um processador M1 e 8 GB de RAM. A implementação utilizou a linguagem de programação Python v3.12.5, juntamente com as bibliotecas Keras v2.3.1, Scikit-Learn v1.5.1 e OpenCV v4.1.0.

5.1 Análise dos Resultados do Processo de Segmentação

Nesta seção, são apresentados os resultados da segmentação das estruturas de interesse: exsudatos duros, exsudatos moles, microaneurismas e hemorragias. Os resultados incluem tanto a segmentação individual de cada estrutura quanto a segmentação conjunta, na qual todas as estruturas são segmentadas simultaneamente. Para a análise quantitativa, foram avaliadas as métricas citadas na Seção 4.5.

A Tabela 4 apresenta os valores médios obtidos para as métricas de avaliação consideradas. Observando os resultados obtidos, verifica-se que a lesão segmentada com maior sucesso é a dos exudatos duros, com uma acurácia de 99,14% na identificação de pixels correspondentes a essa classe de lesão, uma vez que ela se caracteriza por apresentar áreas significativas. A sensibilidade obtida de 99,12% indica que o modelo consegue detectar uma grande parte dos exudatos duros presentes na imagem, o que é de grande valia em aplicações médicas, visto que é importante não deixar casos de pacientes afetados por

Tabela 4 – Métricas de desempenho de segmentação.

Lesão	Acurácia (%)	Sensibilidade (%)	Precisão (%)	F1-Score (%)
Hemorragias	94,83 ± 0,24	94,04 ± 0,31	91,32 ± 0,25	92,66 ± 0,25
Microaneurismas	93,53 ± 0,18	93,34 ± 0,58	93,59 ± 0,56	93,96 ± 0,58
Exsudatos moles	93,92 ± 0,45	93,76 ± 0,49	93,87 ± 0,45	93,28 ± 0,51
Exsudatos duros	99,14 ± 0,25	99,12 ± 0,34	98,56 ± 0,53	99,15 ± 0,29
Todas	93,84 ± 0,12	93,82 ± 0,31	93,35 ± 0,24	93,24 ± 0,27

Fonte: Autor.

alguma doença serem perdidos durante o processo de detecção. A precisão de 98,56% aponta que o modelo tem uma baixa taxa de falsos positivos, sugerindo que, quando o modelo indica a presença de um exudato duro em um determinado pixel, é altamente provável que a indicação esteja correta. Nesse contexto, são evitados falsos diagnósticos positivos. Essa harmonia entre essas duas métricas é representada pelo valor de F1-Score obtido de 99,15%.

O segundo tipo de lesão que apresentou melhores resultados no processo de segmentação foram as hemorragias. Assim como os exsudatos duros, esse tipo de lesão é caracterizado por afetar regiões consideráveis da retina. Os resultados apontaram uma acurácia de 94,84%, indicando uma boa capacidade na distinção entre pixels afetados por essa lesão e pixels saudáveis. Os valores de sensibilidade e precisão obtidos foram de 94,04% e 91,32%, respectivamente, indicando que o modelo tem uma boa efetividade em detectar pixels realmente acometidos por hemorragias, além de evitar falsos positivos. Isso é reforçado pelo valor do F1-Score obtido, que foi de 92,66%.

O tipo de lesão que apresentou os piores resultados durante o processo de segmentação foi a dos microaneurismas. Isso se deve ao fato de que, apesar de se apresentarem em grande quantidade, os microaneurismas são caracterizados por pequenos pontos e têm uma baixa proporção de ocorrência nas imagens. Esse desequilíbrio entre as classes torna sua detecção mais desafiadora. Ainda assim, os resultados obtidos indicaram uma acurácia de 93,53%, demonstrando uma boa capacidade relativa de segmentar esse tipo de lesão. Em termos de sensibilidade e precisão, alcançaram-se valores de 93,34% e 93,59%, respectivamente, evidenciando um equilíbrio entre essas duas métricas. Esse equilíbrio é reforçado pelo valor do F1-Score, que foi de 93,96%.

No último experimento, todas as lesões foram agrupadas em uma única classe, resultando em uma acurácia de 93,84%, sensibilidade de 93,82% e precisão de 93,35%. Esses resultados indicam que o modelo de segmentação para múltiplas lesões possui uma capacidade considerável de identificar pixels com lesão. No entanto, seu desempenho foi inferior ao observado nos modelos de segmentação individual. Isso pode ser explicado pelo fato de que lesões diferentes podem ocorrer simultaneamente nas mesmas regiões

da retina, tornando difícil para o modelo distinguir entre elas. Por exemplo, hemorragias e microaneurismas podem surgir próximos ou até sobrepostos, o que pode resultar em segmentações imprecisas quando o modelo tenta classificar cada pixel em uma única categoria.

Com base nos resultados obtidos, a decisão entre utilizar modelos individuais ou um único modelo que considera todas as lesões foi fundamentada na análise das segmentações geradas e no custo computacional envolvido. A Tabela 5 apresenta o número de épocas necessárias para a convergência de cada modelo, bem como os tempos de treinamento e teste correspondentes, baseados nas 10 iterações realizadas, os menores tempos estão destacados em verde. Ao analisar os tempos de treinamento, observamos que a segmentação de microaneurismas exigiu mais tempo para a convergência da rede, enquanto a segmentação de exsudatos duros demandou o menor tempo de treinamento. Em relação ao tempo de teste, os exsudatos duros apresentaram o menor tempo, com 0,801ms, enquanto as hemorragias tiveram um tempo ligeiramente maior, de 0,891ms.

Tabela 5 – Tempos de treinamento, teste e quantidade de épocas para cada lesão.

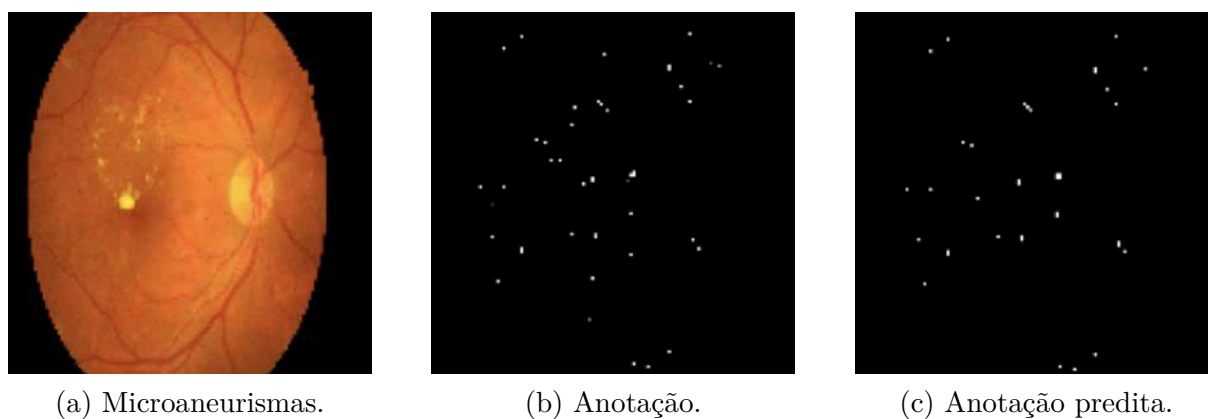
Lesão	Tempo de treinamento (s)	Tempo de teste (s)	Épocas
Hemorragias	1024,06±1,43	0,891±0,005	110±1, 33
Microaneurismas	1124,24±1,12	0,811 ± 0,028	103±1, 44
Exsudatos moles	1314,13±1,52	0,861±0,001	125±1, 33
Exsudatos duros	907, 11 ± 1, 52	0,801 ± 0,005	82±1, 31
Todas	898, 13 ± 1, 51	0,841±0,005	90±1, 34

Fonte: Autor.

Ao comparar o tempo de segmentação das lesões utilizando modelos individuais presentes na Tabela ??, observou-se que o tempo médio necessário é de 3,364 s. Em contraste, o modelo que agrupa todas as lesões em uma única classe apresenta um tempo médio de apenas 0,841s, o que representa uma redução de cerca de 4 vezes no tempo de processamento.

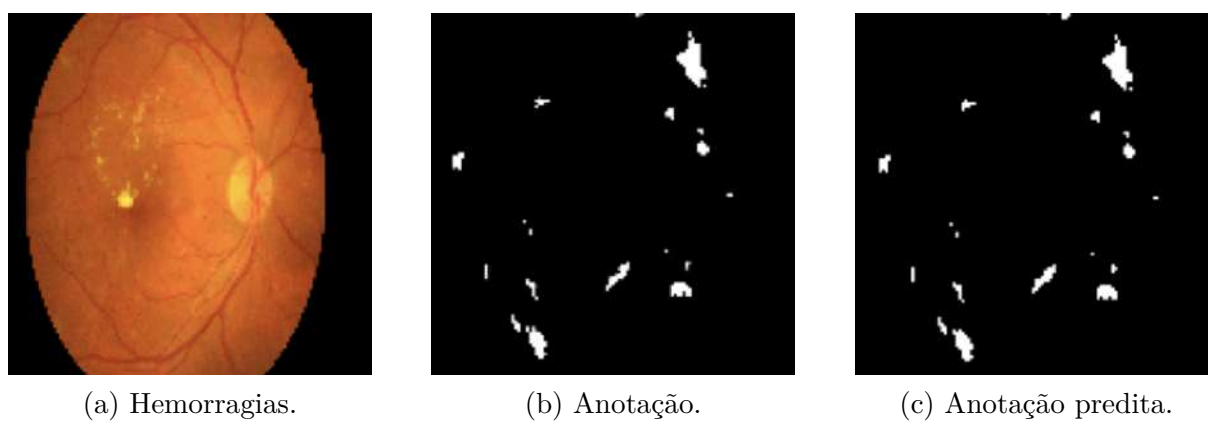
Embora o modelo que considera todas as lesões como uma única classe tenha um desempenho inferior, a significativa economia de tempo o torna uma opção mais vantajosa em contextos onde a eficiência é crucial. Para uma análise mais detalhada, foram avaliadas as segmentações fornecidas por cada modelo, com o objetivo de verificar a qualidade das segmentações fornecidas por cada abordagem, conforme ilustrado nas Figuras 17 - 21.

Figura 17 – Comparação entre a imagem original, as anotações manuais e as previsões do modelo para microaneurismas.



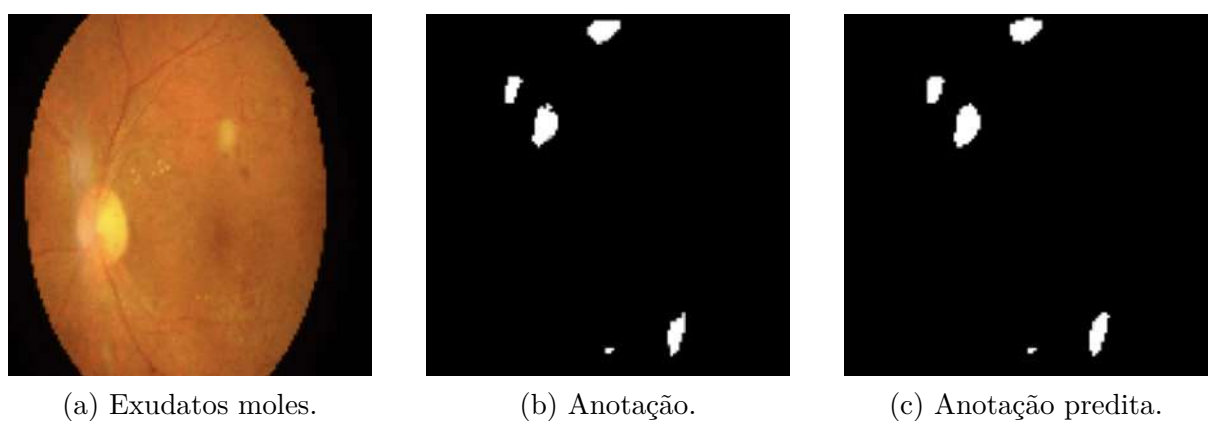
Fonte: Autor.

Figura 18 – Comparação entre a imagem original, as anotações manuais e as previsões do modelo para hemorragias.



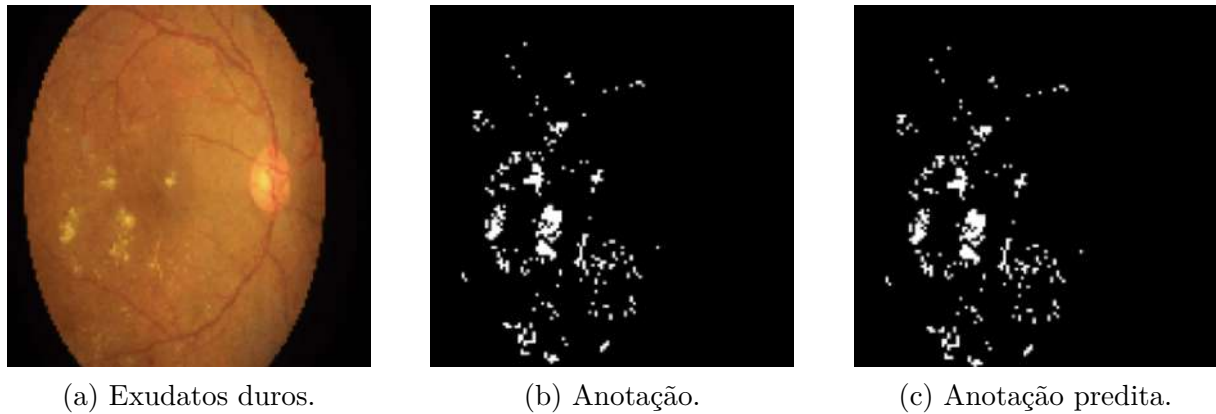
Fonte: Autor.

Figura 19 – Comparação entre a imagem original, as anotações manuais e as previsões do modelo para exudatos moles.



Fonte: Autor.

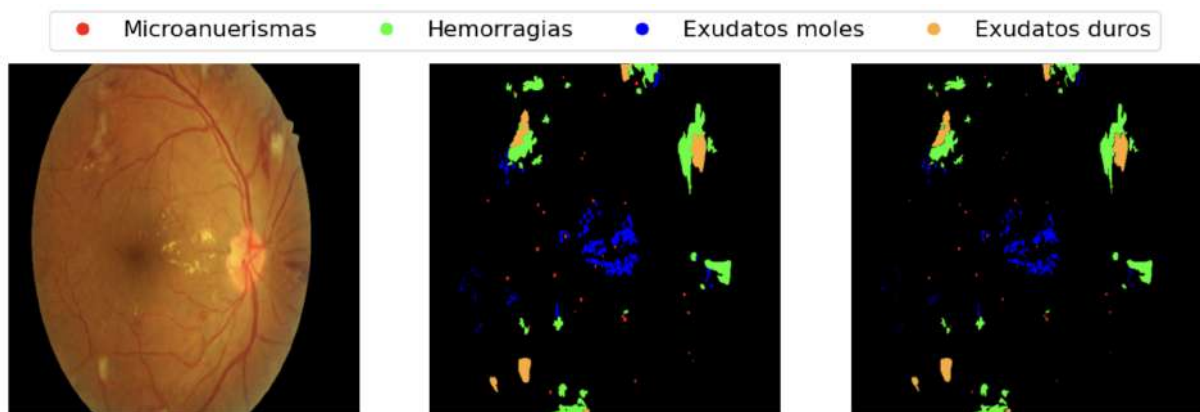
Figura 20 – Comparação entre a imagem original, as anotações manuais e as predições do modelo para exudatos duros.



Fonte: Autor.

A Figura 21 apresenta o resultado do processo de segmentação utilizando um modelo multiclasse, o qual considera cada tipo de lesão como uma classe distinta. A figura exhibe os microaneurismas em vermelho, as hemorragias em verde, e os exsudatos moles e duros nas cores azul e laranja, respectivamente.

Figura 21 – Todas os tipos de lesões: Imagem original, anotação e predição, respectivamente.



Fonte: Autor.

Ao realizar uma inspeção visual nos resultados fornecidos pelo modelo treinado, é possível verificar uma certa correspondência com as segmentações reais das respectivas imagens. Para realizar uma avaliação mais quantitativa das previsões realizadas, foram observados os valores de interseção sobre União obtidos para cada uma das imagens segmentadas. A Tabela 6 apresenta os resultados obtidos.

Conforme observado, as lesões do tipo exsudato duro e hemorragia apresentaram os maiores valores de IoU, o que indica que as segmentações geradas possuem uma elevada correspondência entre as áreas segmentadas pelo modelo e as regiões de interesse nas máscaras de referência. Em contrapartida, lesões como microaneurismas demonstraram

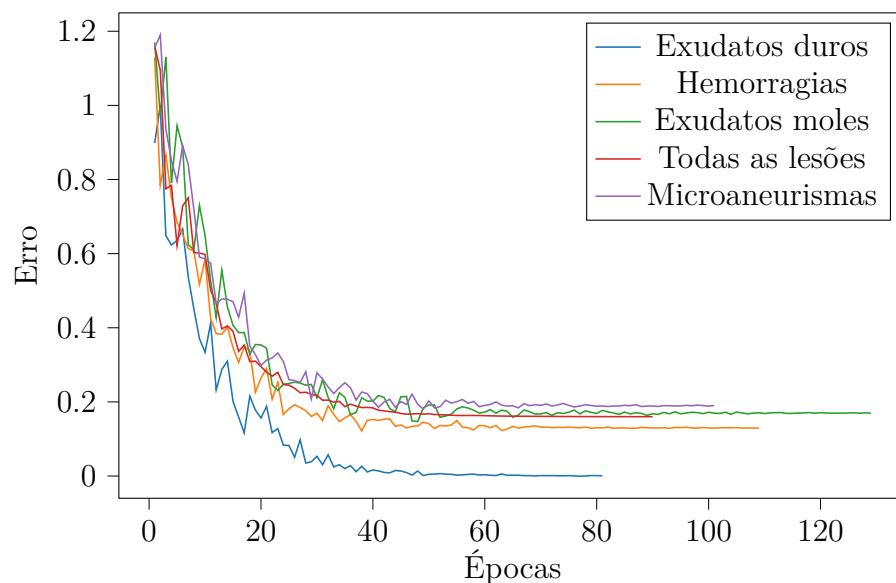
Tabela 6 – Valores de IoU obtidos para a segmentação de cada tipo de lesão.

Lesão	IoU
Microaneurismas	$0,68 \pm 0,024$
Hemorragias	$0,96 \pm 0,032$
Exudatos moles	$0,92 \pm 0,052$
Exudatos duros	$0,97 \pm 0,047$
Todas as lesões	$0,87 \pm 0,045$

Fonte: Autor.

menor sobreposição, refletindo um desempenho mais modesto na segmentação dessas áreas. Ao analisar a segmentação simultânea de todos os tipos de lesão, obteve-se um valor de IoU de 0,87, considerado intermediário em relação aos resultados individuais. Dessa forma, o uso de um modelo único revelou-se vantajoso, uma vez que oferece um significativo ganho de eficiência no tempo de predição. Uma vez que as anotações preditas por esse modelo apresentam certa proximidade com as anotações reais, além de manter um desempenho consistente na segmentação. Portanto, este modelo foi considerado apropriado para a segmentação das imagens a serem utilizadas no processo de classificação. A Figura 22 mostra as curvas de erro obtidas para os dados de validação, considerando a média das 10 iterações realizadas para cada tipo de lesão. Observa-se que a curva de erro para os exsudatos duros exigiu um menor número de épocas e convergiu para o menor patamar de erro entre todas as curvas analisadas. Em contraste, a curva de erro para as lesões do tipo microaneurismas demonstra uma estabilidade em um patamar de erro mais elevado.

Figura 22 – Curvas de obtidas durante o processo de segmentação tipo de lesão.



Fonte: Autor.

Esses resultados corroboram as observações feitas na Tabela 4, reforçando a dificul-

dade associada à segmentação de microaneurismas em comparação com outras lesões.

5.1.1 Comparação com Trabalhos Relacionados

Os resultados da segmentação de imagens utilizando *CNNs* são comparados com trabalhos da literatura que utilizam o mesmo conjunto de dados. A Tabela 7 apresenta os resultados obtidos durante o processo de segmentação de cada tipo de lesão, comparando-os com outros trabalhos relacionados, os melhores resultados estão destacados em verde.

Tabela 7 – Comparação dos resultados obtidos durante o processo de segmentação pelo método proposto em relação a outras abordagens.

Lesão	Autor	Resultados			
		Acurácia (%)	Sensibilidade (%)	F1-Score (%)	IoU
Microaneurismas	(KOU et al., 2020)	98,00	92,90	–	–
	Método proposto	93,53	93,34	93,96	0,68
Hemorragias	(NAZIR et al., 2020)	95,20	95,10	94,30	–
	Método proposto	94,83	94,04	92,66	0,96
Exsudatos moles	(HARIS et al., 2019)	90,46	88,86	–	–
	Método proposto	93,92	93,76	93,87	0,92
Exsudatos duros	(SAMBYAL et al., 2020)	99,00	99,85	–	–
	Método proposto	99,14	99,12	99,15	0,97
Todas	(FURTADO, 2020)	90,00	–	–	0,86
	Método proposto	93,84	93,82	93,24	0,87

Fonte: Autor.

Ao comparar os resultados, o trabalho realizado por (KOU et al., 2020) na segmentação de microaneurismas alcançou uma acurácia de 98%. Essa pesquisa empregou uma U-Net residual aprimorada (ERU-Net), que apresenta uma estrutura de três caminhos U, incluindo três caminhos de upsampling e um caminho de downsampling. Esse design inovador melhora a fusão de características e a captura de detalhes em imagens de fundo. Além disso, a presença de um bloco residual na ERU-Net contribui para uma extração de características mais eficaz. Essa modificação na arquitetura resultou em uma acurácia superior à da abordagem proposta anteriormente. Em termos de sensibilidade, a ERU-Net obteve uma pontuação de 92,90%, que é levemente inferior à pontuação de 93,34% alcançada pelo presente estudo.

Para a segmentação de lesões do tipo hemorragias, o trabalho conduzido por (NAZIR et al., 2020) utilizou uma Rede Neural Convolutiva de Região Rápida (*Faster*

R-CNN), atingindo uma acurácia de 95,20%. Esse resultado é levemente próximo ao obtido na abordagem proposta, que foi de 94,83%. Observando as métricas de sensibilidade e F1-Score, esse trabalho obteve 95,10% e 94,30%, respectivamente. Os resultados superiores atingidos por essa abordagem podem ser explicados pela arquitetura utilizada pelos autores, uma vez que a U-Net consiste em uma arquitetura projetada principalmente para segmentação semântica, onde o foco está em pixel por pixel. Ela possui uma estrutura de codificador-decodificador que pode ser eficaz para a segmentação de imagens, mas pode não capturar detalhes contextuais tão bem quanto a *Faster R-CNN* em algumas situações.

Para a tarefa de segmentação de exsudatos moles, o trabalho realizado por (HARIS et al., 2019) atingiu 90,46% de acurácia e 88,86% de sensibilidade, utilizando uma rede de segmentação semântica baseada em encoder-decoder, chamada SOP-Net (Segmentação de Patologias Oculares Usando Rede Neural Convolutiva Profunda), para a delimitação simultânea de patologias retiniais. Ao comparar com os resultados obtidos pela abordagem proposta, que alcançou 93,92% de acurácia e 93,76% de sensibilidade, observa-se um desempenho relativamente superior, demonstrando que a rede proposta pelos autores, apesar dos bons resultados, apresentou um desempenho inferior ao fornecido pela abordagem proposta utilizando a U-Net, indicando que a mesma possui uma maior capacidade de detectar pixels da imagem afetados por esse tipo de lesão.

O trabalho realizado por (SAMBYAL et al., 2020) apresenta uma metodologia utilizando a U-Net para segmentar lesões do tipo exsudato duro. Nesse estudo, foram utilizados dois conjuntos de dados, o IDRiD e o e-optha (DECENCIERE et al., 2013), atingindo uma acurácia de 99% e uma sensibilidade de 99,85%, respectivamente. A metodologia adotada é bastante similar à proposta neste trabalho; entretanto, os autores utilizaram uma quantidade maior de imagens. Nesse contexto, é observada uma maior capacidade de detecção dos pixels afetados por esse tipo de lesão, produzindo resultados ligeiramente superiores aos obtidos durante os experimentos, com 99,14% de acurácia e 99,12% de sensibilidade.

No estudo conduzido por (FURTADO, 2020), é apresentado um novo método de segmentação de lesões provenientes da retinopatia diabética. Nesse método, todas as lesões são segmentadas por uma única arquitetura, DeepLabv3. Durante a realização dos experimentos, foi atingida uma acurácia de 90% e um IoU de 0,86. Esses resultados são relativamente próximos aos obtidos por este estudo, onde a acurácia para a segmentação de todas as lesões foi de 93,43% e o IoU de 0,87. Desse modo, o modelo de segmentação treinado no presente estudo apresentou capacidades similares às do modelo desenvolvido pelo autor, sendo capaz de detectar pixels de cada tipo de lesão específica, segmentando as regiões afetadas com boa precisão.

5.2 Análise dos Resultados do Processo de Classificação

Nesta seção, serão apresentados os resultados dos experimentos realizados para a classificação das retinografias, utilizando as *CNNs* mencionadas no Capítulo 4. Os resultados são uma média de 10 iterações realizadas. A Tabela 8 apresenta os resultados obtidos com as melhores pontuações destacadas na cor verde.

Tabela 8 – Métricas de avaliação obtidas para cada *CNN*.

<i>CNN</i>	Acurácia (%)	Sensibilidade (%)	Precisão (%)	F1-Score (%)	<i>Log loss</i>
EfficientNetB1	99.45±0.80	99.16±0.16	99.23±0.19	99.20±0.22	0.125±0.013
EfficientNetB0	98.30±0.50	94.29±0.18	97.50±0.21	95.87±0.24	0.143±0.015
DenseNet201	98.10±0.10	94.88±0.14	97.30±0.17	96.07±0.29	0.152±0.012
DenseNet121	97.60±0.30	93.59±0.17	96.90±0.20	95.10±0.23	0.161±0.014
ResNet50	98.77±0.20	94.86±0.01	97.20±0.02	96.40±0.21	0.166±0.015
NFNet	98.23±0.20	94.54±0.01	97.12±0.02	96.45±0.21	0.166±0.015
InceptionV3	94.50±0.30	90.38±0.17	93.60±0.20	91.50±0.23	0.233±0.015
VGG19	97.30±0.40	92.37±0.19	96.50±0.22	94.30±0.25	0.175±0.016
VGG16	96.70±0.20	92.58±0.16	95.80±0.19	93.60±0.22	0.193±0.015
AlexNet	96.50±0.60	95.74±0.15	95.40±0.18	95.57±0.02	0.253±0.012

Fonte: Autor.

Observando os resultados obtidos, a EfficientNetB1 apresentou um desempenho superior em termos de acurácia, atingindo 99,45%, seguida pela EfficientNetB0 e DenseNet201, respectivamente. Isso indica que a EfficientNetB1 apresentou uma capacidade maior de classificar as classes corretamente. Em termos da métrica de sensibilidade, destacaram-se a EfficientNetB1, DenseNet121 e EfficientNetB0, atingindo 99,16%, 94,88% e 94,29%, respectivamente. Para essa métrica, a EfficientNetB1 apresentou o melhor resultado, indicando uma alta capacidade em detectar tanto casos positivos de retinopatia diabética quanto pacientes que eventualmente sejam saudáveis. Uma sensibilidade alta é de vital importância em aplicações médicas, pois indica que há uma baixa taxa de falsos negativos, o que evita que pacientes acometidos pela doença sejam classificados como pacientes saudáveis e, por consequência, não realizem o tratamento adequado.

Em termos de precisão, destacou-se a EfficientNetB1 com 99,23%, seguida pela EfficientNetB0 e DenseNet201 com 97,50% e 97,30%, respectivamente. Isso indica que o modelo apresenta um baixo número de falsos positivos, o que significa que poucas vezes um paciente é classificado como portador da doença sem realmente tê-la, evitando preocupações e tratamentos desnecessários.

Observando a métrica de F1-Score, a EfficientNetB1 apresentou a melhor pontuação, atingindo 99,20%, seguida pela EfficientNetB0 e DenseNet201. Isso indica um bom compromisso entre sensibilidade e precisão, desse modo, a EfficientNetB1 apresenta poucos falsos positivos e falsos negativos, realizando, portanto, a classificação de forma correta para cada classe. A análise das métricas anteriores é reforçada pela métrica de *log*

loss, obtendo-se os valores de 0,125, 0,143 e 0,152 para a EfficientNetB1, EfficientNetB0 e DenseNet201, respectivamente. Isso indica que a EfficientNetB1 tem um bom desempenho em prever a classe correta e atribuir probabilidades adequadas a essas previsões, desse modo, suas predições apresentam altas probabilidades de estarem corretas.

A Tabela 9 apresenta os tempos de treinamento e teste para cada *CNN*, os melhores resultados estão destacados em verde. A análise do tempo de teste é de vital importância, pois isso pode influenciar a escolha da *CNN* em uma aplicação, uma vez que é desejável obter previsões corretas com a maior velocidade possível.

Tabela 9 – Tempos de treinamento e teste para cada *CNN*.

<i>CNN</i>	Tempo de treino (s)	Tempo de teste (ms)
EfficientNetB1	6438, 33±35, 27	0, 519±0, 121
EfficientNetB0	5238, 33±66.23	0, 453±0.154
DenseNet201	4238, 33±66.23	0, 674±0, 009
DenseNet121	3920, 23±38.44	0, 553±0, 054
ResNet50	2724, 13±16.44	0, 293±0.011
InceptionV3	2359, 32±36, 43	0, 557±0, 046
NFNet	2158, 33±18, 42	0, 583±0, 018
VGG19	2134, 15±12, 12	0, 474±0, 008
VGG16	1920, 43±12, 49	0, 572±0, 090
AlexNet	1331, 16±29, 33	0, 159±0, 098

Fonte: Autor.

Analizando os resultados, as *CNNs* AlexNet, VGG16 e VGG19 apresentaram os menores tempos de treinamento, com 1331,16s, 1920,43s e 2134,15s, respectivamente. Como essas *textitCNNs* possuem arquiteturas relativamente mais simples do que as demais avaliadas, isso influenciou no tempo de treinamento. No que diz respeito ao tempo de teste, ou seja, o tempo que a arquitetura demora para realizar uma classificação, a AlexNet apresentou o menor tempo, completando essa tarefa em 0,159 ms, seguida pelas *textitCNNs* EfficientNetB0 e ResNet50, que realizaram a classificação em 0,453 ms e 0,293 ms, respectivamente.

Com o objetivo de verificar a existência de diferenças estatisticamente significativas entre os desempenhos das *textitCNNs*, foi aplicado o teste de Tukey nos valores de acurácia obtidos em cada iteração realizada, precedido pelo uso do teste ANOVA e do teste de normalidade Shapiro-Wilk. Os testes foram conduzidos considerando um nível de confiança de 95% (isto é, $\alpha = 0.05$). A Tabela a seguir apresenta os valores de *p-value* obtidos durante a realização dos testes. A Tabela 10 indica os resultados da aplicação desse teste.

Uma vez que todos os valores obtidos são superiores ao *p*-valor adotado, então para todas as *CNNs* a hipótese nula é aceita. Desse modo, todas as distribuições podem

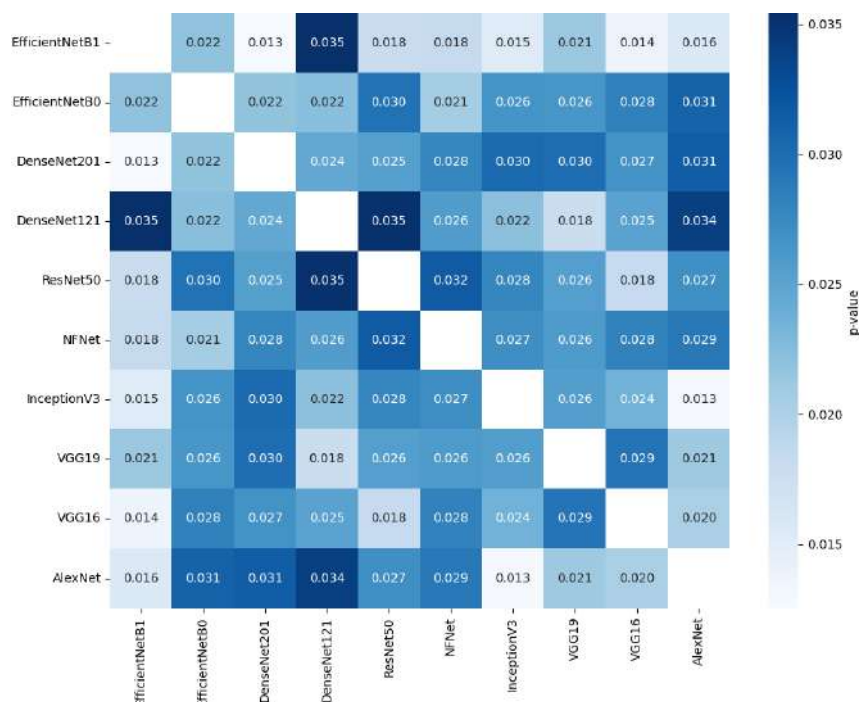
Tabela 10 – p -valor obtido para cada CNN.

CNN	p -valor
EfficientNetB1	0,12
EfficientNetB0	0,14
DenseNet201	0,17
DenseNet121	0,24
ResNet50	0,21
NFNet	0,33
InceptionV3	0,15
VGG19	0,18
VGG16	0,12
AlexNet	0,17

Fonte: Autor.

ser consideradas como normais. Após a verificação da condição prévia, o teste ANOVA foi aplicado, com o objetivo de avaliar se os resultados obtidos pelos classificadores apresentam diferenças estatisticamente significativas entre si. O valor de p -valor obtido foi de $2,62 \times 10^{-3}$. Assim, considerando o nível de significância adotado de 95%, conclui-se que há evidências para afirmar que os resultados dos classificadores diferem estatisticamente, uma vez que a hipótese nula é rejeitada. Por fim, para determinar quais pares de médias são significativamente diferentes entre si. A Figura 23 ilustra o p -valor obtido para a comparação entre os pares de cada CNN.

Figura 23 – Resultados da aplicação do teste de Tukey.



Fonte: Autor.

Observando os valores obtidos, verifica-se que todos os pares avaliados apresentaram resultados que levam à rejeição da hipótese nula, indicando que há diferenças estatisticamente significativas entre eles. Ao analisar os valores apresentados na Tabela 8, nota-se que a maior média foi obtida pela EfficientNetB1. Essa CNN demonstrou ser estatisticamente superior às demais, evidenciando um desempenho consistentemente maior entre as avaliadas.

Diante dos resultados observados na Tabela 8 e com base no resultado proveniente do teste estatístico, além de considerar os tempos de treinamento e teste da Tabela 9, a EfficientNetB1 apresenta a melhor capacidade de realizar a classificação de retinografias de forma consistente. Embora seu tempo de teste tenha sido superior ao das demais textitCNNs.

Após a realização dos testes estatísticos, a CNN EfficientNetB1 foi verificada como a mais eficiente para realizar a detecção e classificação da RD. A Tabela 11 ilustra a análise das métricas adotadas nessa dissertação, porém, dispostas para cada classe.

Tabela 11 – Métricas de avaliação para cada classe obtidas pela EfficientNetB1.

Classe	Acurácia (%)	Sensibilidade (%)	Precisão (%)	F1-Score (%)	Log loss
Saudável	99,14±0,31	98,98±0,34	98,96±0,25	98,97±0,45	0,125±0,02
Leve	99,01±0,23	98,95±0,11	98,92±0,23	98,93±0,24	0,125±0,09
Moderado	99,76±0,56	99,60±0,53	99,62±0,31	99,62±0,31	0,125±0,08
Grave	99,75±0,11	99,35±0,03	99,25±0,42	99,40±0,13	0,125±0,07
Proliferativa	99,59±0,32	98,93±0,34	99,15±0,12	99,04±0,35	0,125±0,08

Fonte: Autor.

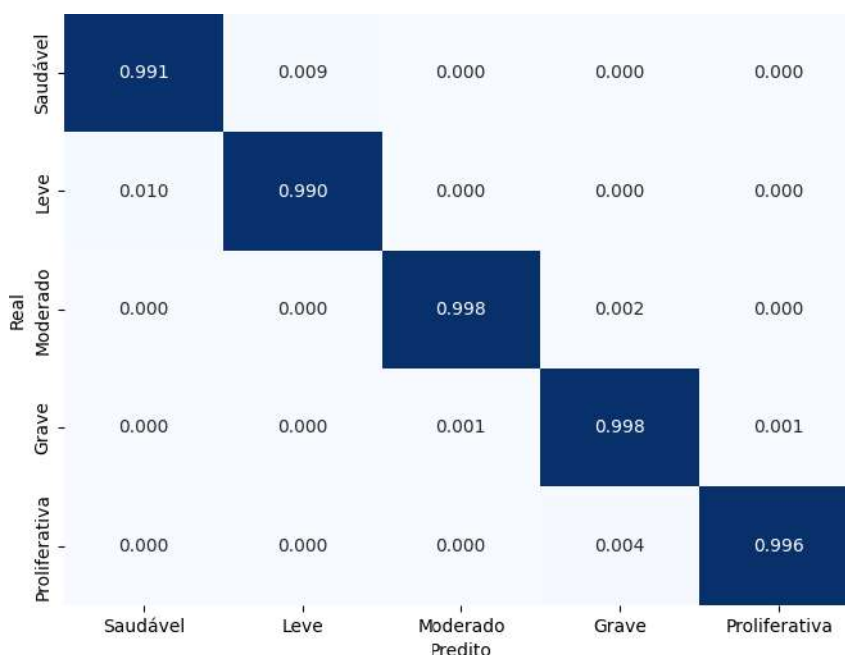
Ao analisar os resultados obtidos, observou-se que o grau leve de RD apresentou a menor acurácia entre as classes avaliadas, alcançando 99,01%. Esse desempenho deve-se ao fato de que, nesse estágio inicial da doença, são observadas apenas lesões do tipo microaneurismas, que, devido ao seu pequeno tamanho, apresentam certa dificuldade de visualização. Quanto às métricas de sensibilidade e precisão, essa classe obteve 98,98% e 98,96%, respectivamente, indicando que a rede treinada possui boa capacidade de identificar corretamente os casos de RD leve, com uma quantidade relativamente baixa de falsos negativos e falsos positivos. Esse equilíbrio é corroborado pela métrica F1-Score, que apresentou um valor de 98,93%. Por fim, o *log loss* obtido demonstra que as previsões realizadas possuem boa confiabilidade.

A classe saudável apresentou o segundo melhor desempenho, com uma acurácia de 99,14%. Para essa classe, tem-se como objetivo principal minimizar o número de falsos positivos, ou seja, evitar que casos de pacientes com a presença de RD sejam rotulados como saudáveis, e como objetivo secundário, evitar casos de falsos negativos, isto é, rotular casos saudáveis como portadores de RD. Pelos resultados obtidos, esse equilíbrio foi alcançado, conforme observado nas métricas de precisão e recall atingidas.

Para as classes Moderado e Grave, houve uma certa equivalência nos resultados encontrados, pois esses graus de RD são mais facilmente detectáveis em virtude das lesões presentes. Para a classe Proliferativa, a acurácia atingida foi de 99,59%. Observando a precisão, foi atingida uma pontuação de 98,93%, indicando que a rede treinada apresentou uma boa capacidade de detecção desse grau de RD. Analisando os valores de precisão e F1-Score, reforça-se a capacidade de detectar casos positivos para essa classe de RD, evitando falsos positivos e falsos negativos. A leve queda de desempenho para essa classe, em comparação com as classes Moderado e Grave, deve-se ao fato de que há uma sobreposição com outras classes, especialmente com a classe Grave. Em ambos os estágios (Grave e Proliferativa), as lesões podem apresentar sinais visuais semelhantes, como hemorragias extensas e áreas de neovascularização, o que dificulta para o modelo diferenciar com precisão.

A Figura 24 apresenta a matriz de confusão final obtida pela EfficientNetB1, sendo uma soma das 10 matrizes de confusão obtidas a partir das 10 iterações realizadas. Observando os resultados, os principais erros ocorreram na distinção das classes Saudável e RD Leve, devido à dimensão das lesões presentes nesse grau de RD.

Figura 24 – Matriz de confusão final do conjunto de testes utilizando a EfficientNetB1.



Fonte: Autor.

As classes moderado e grave apresentaram pequenos erros de classificação entre si, o que é razoável, considerando que são estágios consecutivos da RD. Para a classe proliferativa, ocorreram confusões com a classe grave, possivelmente devido à sobreposição das lesões em muitos casos, o que pode dificultar a distinção entre esses estágios. De modo geral, os erros observados ocorreram entre classes vizinhas, o que reforça a capacidade da

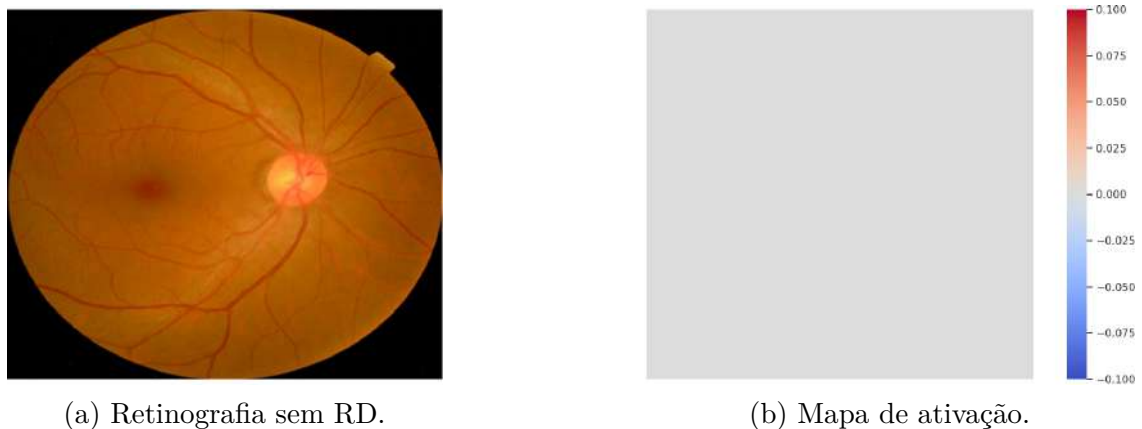
rede de detectar RD sem cometer erros mais graves de classificação, como atribuir um rótulo de RD grave a um paciente saudável.

5.2.1 Resultados do Grad-CAM

Nesta seção, serão apresentados os resultados obtidos com a aplicação do Grad-CAM em retinografias do conjunto de teste, utilizando a EfficientNetB1. A análise buscou identificar as áreas de maior relevância para as decisões do modelo, destacando casos específicos para avaliar tanto acertos quanto erros.

Foram avaliados quatro casos principais: (i) retinografia classificada como saudável de forma correta, (ii) retinografia com RD leve corretamente classificada, (iii) retinografia com RD leve classificada de forma incorreta, e (iv) retinografia com RD proliferativa, a forma mais grave da doença. Para cada caso, os mapas de ativação gerados pelo Grad-CAM foram sobrepostos às imagens originais, permitindo identificar as regiões mais importantes utilizadas pelo modelo para tomar a decisão. A Figura 25 ilustra o mapa de ativação resultante ao aplicar o Grad-CAM em uma retinografia com ausência de RD:

Figura 25 – Retinografia sem RD e sobreposição com o mapa de ativação.

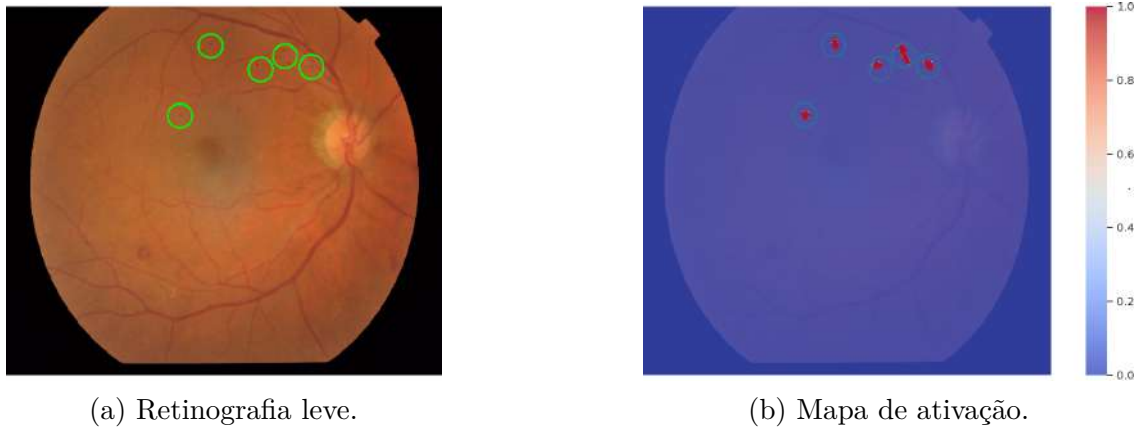


Fonte: Autor.

Conforme verificado, o mapa resultante não apresenta as regiões relevantes para a classificação de RD, gerando um mapa totalmente uniforme com valores próximos a zero, uma vez que não há lesões características da RD na retinografia analisada. A Figura 26 ilustra uma retinografia representativa de RD de grau leve, acompanhada do mapeamento gerado pelo Grad-CAM, evidenciando as regiões de maior importância para a decisão do modelo:

Observando as regiões com maior importância, destaca-se a presença de áreas associadas a microaneurismas (destacadas em verde na imagem original), que são características desse tipo de RD. Em contrapartida, as demais regiões da retina nas quais não há presença de lesões características da RD não apresentam contribuição significativa para o processo de classificação.

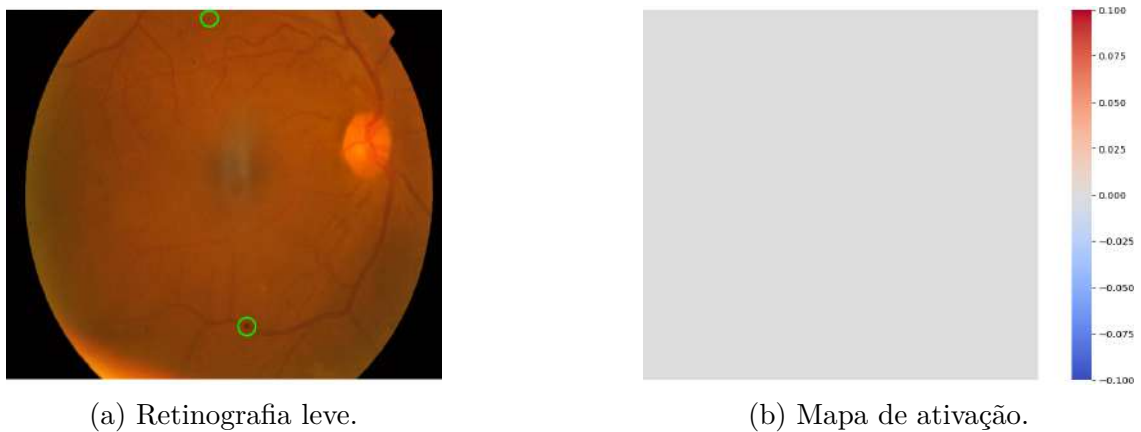
Figura 26 – Retinografia com RD leve e sobreposição com o mapa de ativação.



Fonte: Autor.

Em contrapartida, analisar as retinografias classificadas de forma errada com Grad-CAM é de vital importância para entender as decisões do modelo e identificar padrões que não estão sendo corretamente detectados. Essa análise é fundamental para destacar pontos de melhoria na abordagem de detecção da RD. A Figura 27 ilustra a análise do resultado produzido por uma retinografia com a presença de RD leve, mas que foi erroneamente classificada como saudável:

Figura 27 – Retinografia com RD leve e sobreposição com o mapa de ativação.

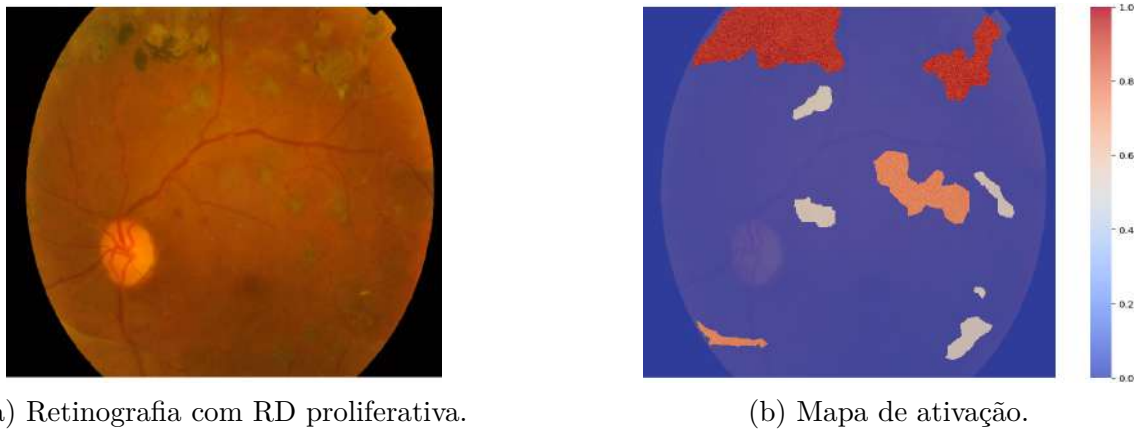


Fonte: Autor.

A classificação equivocada ocorreu devido à não detecção dos microaneurismas na retinografia, o que representa uma falha na etapa de segmentação. Nesse contexto, a U-Net treinada para o conjunto de dados utilizado na classificação não foi capaz de segmentar adequadamente os microaneurismas, que estão presentes em pequenas quantidades (conforme destacado em verde na imagem original). Observando o mapa de ativação, tem-se um mapa totalmente uniforme com valores próximos a zero, uma vez que a imagem fornecida para classificação consiste em uma imagem composta apenas por pixels pretos. A Figura 28 apresenta o resultado ao aplicar uma imagem com RD com

grau proliferativo:

Figura 28 – Retinografia com RD proliferativa e sobreposição com o mapa de ativação.



Fonte: Autor.

Observando as regiões com maior peso (tonalidade mais avermelhada), percebe-se que essas áreas apresentam degeneração da retiniana sendo uma das características desse estágio da doença. As demais regiões com menor intensidade caracterizam-se pela presença de áreas com menor grau de degeneração retiniana. Além disso, também são destacadas regiões que contêm pequenos exsudatos e microaneurismas, que são lesões que, embora estejam presentes em graus iniciais da doença, tendem a permanecer com a progressão da doença.

5.2.2 Comparação com trabalhos relacionados

Os resultados da classificação de retinografias são comparados com outros estudos da literatura que utilizam o mesmo conjunto de dados. A Tabela 12 apresenta um breve resumo dos trabalhos comparados.

Tabela 12 – Comparação dos resultados obtidos pelo método proposto com outras abordagens.

Autor	Método	Resultados				Grad-CAM
		Acurácia (%)	Sensibilidade (%)	Precisão (%)	F1-Score (%)	
Método proposto	U-Net e EfficientNetB1	99,45	99,16	99,23	99,20	Sim
(EL-ATEIF; IDRI, 2022)	VGG19	97,49	96,24	97,50	96,87	Não
(KHALIFA et al., 2019)	AlexNet	97,90	95,42	96,23	95,82	Não
(TAUFIQURRAHMAN et al., 2020)	MobileNetV2 e SVM	85,00	74,26	71,78	74,00	Não
(ALGHAMDI, 2022)	VGG-16	73,04	52,00	87,00	65,00	Sim
(MURUGAPPAN et al., 2022)	DRNet	98,18	97,41	99,55	98,46	Sim

Fonte: Autor.

O estudo conduzido por (EL-ATEIF; IDRI, 2022), os autores compararam o desempenho de abordagens de aprendizado profundo para a classificação automática de Retinopatia Diabética (RD). Eles utilizaram fusão conjunta, combinando os conjuntos de dados APTOS-2019 e Messidor-2 para classificar imagens de retinografia, alcançando

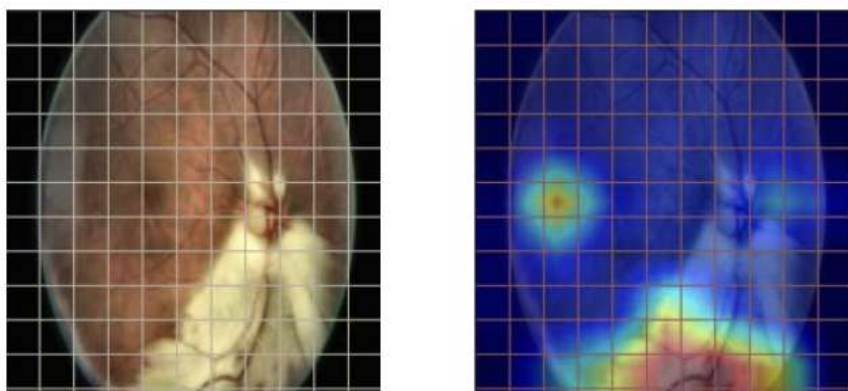
97,49% de acurácia. Esses resultados superiores podem ser atribuídos à segmentação prévia, uma vez que esse processo seleciona as regiões que de fato são importantes para a detecção de RD. Na metodologia adotada, os autores combinaram as imagens originais com as imagens submetidas a um fundo de desfoque gaussiano ponderado, compondo duas modalidades para formar o conjunto de dados final e realizar o processo de treinamento. Tal abordagem pode melhorar os resultados; entretanto, para o processo de treinamento, adiciona mais complexidade, uma vez que as imagens originais e as versões com desfoque gaussiano ponderado (WGBF) podem conter informações redundantes, levando à sobrecarga durante a extração de características e aumentando o risco de *overfitting*.

Em (KHALIFA et al., 2019), Os autores utilizaram modelos de *deep learning*, os quais foram treinados e testados com o conjunto de dados APTOS-2019, um dos primeiros estudos a utilizar esse conjunto de dados, publicado no segundo trimestre de 2019. Os modelos escolhidos — AlexNet, ResNet18, SqueezeNet, GoogleNet, VGG16 e VGG19 — foram selecionados por possuírem menos camadas em comparação a modelos maiores, como DenseNet e InceptionResNet. Técnicas de aumento de dados foram aplicadas para aumentar a robustez e mitigar o *overfitting*. A AlexNet obteve a maior acurácia, atingindo 97,90%. Assim como no presente trabalho, os autores utilizaram CNNs para realizar a classificação de RD e, além disso, aplicaram aumento de dados para aumentar a robustez do modelo. Nesta dissertação, a acurácia obtida utilizando a AlexNet foi de 96,50%, sendo levemente inferior à obtida pelos autores. Comparando com a EfficientNetB1, que obteve o melhor desempenho em termos de acurácia, atingindo 99,45%, observa-se um desempenho superior quando é analisada a acurácia de classificação por classe. Por exemplo, o presente trabalho atingiu uma acurácia de 99,01% na classificação de RD com grau leve, ao passo que os autores atingiram 98%. O desempenho da EfficientNetB1 foi superior para todas as classes, exceto para a classe saudável, que teve uma acurácia de 99,14% contra 99,7%. Os melhores resultados obtidos nesta dissertação podem ser explicados pela etapa prévia de segmentação e também pelo fato de a EfficientNetB1 ser mais robusta do que a AlexNet utilizada pelos autores.

Na abordagem apresentada por (TAUFIQURRAHMAN et al., 2020) propõe o uso de *deep learning* para classificação de RD. Diferente de modelos baseados em arquiteturas grandes de CNNs, que exigem muitos dados e alta capacidade computacional, foi utilizada a MobileNetV2. O modelo foi treinado no conjunto de dados APTOS-2019, utilizando pesos pré-treinados do *ImageNet*. Além disso, foi otimizado com a combinação de um classificador SVM, formando o modelo híbrido MobileNetV2-SVM, atingindo uma acurácia de 85,00%. Observando a matriz de confusão final apresentada pelos autores, é evidente que ocorreram erros de classificação entre todas as classes. Um exemplo significativo é o caso de imagens de RD de grau leve sendo classificadas erroneamente como proliferativas. A ausência de técnicas de balanceamento dos dados e a não aplicação de segmentação podem ter contribuído para o desempenho inferior ao observado na presente dissertação.

No método proposto por (ALGHAMDI, 2022), Os autores focaram na detecção de RD utilizando aprendizado profundo. Modelos de *deep learning* foram avaliados em duas abordagens: classificação binária e classificação multiclasse, de acordo com os níveis de gravidade da RD. Os experimentos destacaram o desempenho superior do modelo VGG-16, que obteve 73,04% de acurácia. Além disso, foi aplicado o método Grad-CAM para fornecer explicabilidade às decisões dos modelos. A Figura 29 apresenta a comparação entre os resultados obtidos pelo Grad-CAM e o método utilizado pelos autores.

Figura 29 – Mapa de ativação gerado pela VGG-16.



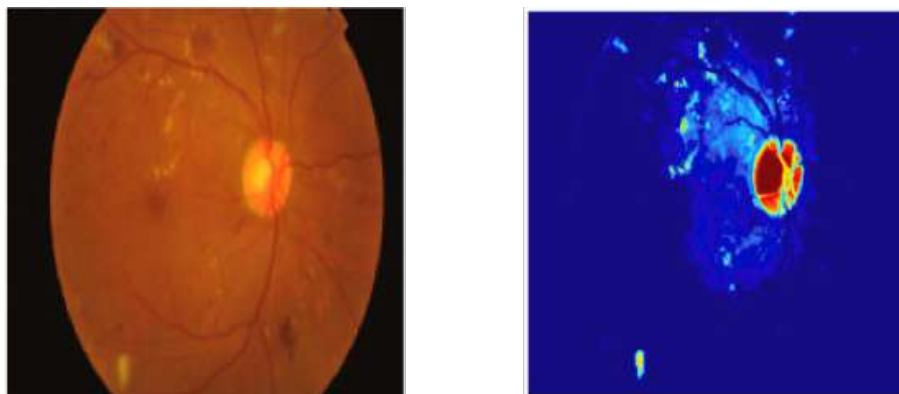
Fonte: Adaptado de (ALGHAMDI, 2022).

A análise de explicabilidade realizada com Grad-CAM revelou que o modelo nem sempre identifica corretamente as lesões relacionadas à RD. Isso foi evidenciado na imagem analisada pelos autores, na qual a região contendo exsudatos não foi completamente considerada como relevante para a classificação. Esse comportamento sugere que o modelo treinado não conseguiu identificar de forma plena os padrões nas imagens que correspondem às áreas afetadas pela RD. Tal limitação pode estar relacionada a fatores como a ausência de pré-processamentos adequados ou a falta de aplicação de técnicas de segmentação, que poderiam auxiliar no reconhecimento mais preciso das regiões afetadas pela RD.

Em (MURUGAPPAN et al., 2022) é apresentado um método inovador para diagnóstico e classificação da RD utilizando *Few-Shot Learning*, uma abordagem que permite treinar modelos eficazmente com poucos dados de treinamento, resolvendo problemas como sobreajuste e má generalização em pequenos conjuntos de dados. O método proposto é baseado na DRNet, uma rede prototípica com mecanismos de atenção desenvolvida para detecção e classificação da RD. A DRNet utiliza aprendizado episódico para treinar a rede em tarefas de classificação de poucos exemplos, empregando transformações agregadas e ativações de gradiente para projetar representações de imagens mais ricas e focadas. Como resultado, foi atingida uma acurácia de 98,18% sendo a maior dentre os estudos comparados, sendo essa metodologia bastante robusta uma vez que o uso do *Few-Shot Learning* permite que o modelo aprenda efetivamente com um número limitado de exemplos, além disso, a DRNet incorpora um mecanismo de atenção que foca seletivamente nas regiões

mais relevantes das imagens de fundo de olho, destacando características críticas para a detecção de RD, como microaneurismas, hemorragias ou exsudatos. Foi empregado uso do Grad-CAM para realizar a explicabilidade do modelo, a Figura 30 ilustra o resultado após a aplicação do método.

Figura 30 – Mapa de ativação gerado pela DRNet.



Fonte: Adaptado de (MURUGAPPAN et al., 2022).

O mapa de ativação gerado pelo modelo mostrou que ele conseguiu identificar as regiões afetadas pela RD. No entanto, observou-se que o modelo atribuiu um peso significativo ao disco óptico, o que pode indicar um erro de atenção. Esse comportamento pode ser consequência de um viés presente no conjunto de dados ou de padrões incorretos aprendidos durante o treinamento. Por exemplo, o conjunto de dados pode conter imagens em que o disco óptico aparece mais destacado devido a variações de iluminação, contraste ou anomalias visuais. Além disso, se as anotações incluírem casos graves de RD com alterações próximas ao disco óptico, o modelo pode ter aprendido, de forma equivocada, a associar a presença de RD a essa região específica.

6 Conclusões

Esta dissertação apresenta uma abordagem para segmentar e classificar retinografias em cinco classes, além de destacar as regiões importantes para a classificação por meio de técnicas de inteligência artificial explicável. A abordagem é dividida em três etapas principais: i) Segmentação das lesões utilizando a U-Net, ii) Classificação utilizando diversas arquiteturas de *CNNs* e iii) Apresentação das regiões mais relevantes para a classificação.

Para a tarefa de segmentação de lesões provenientes da RD, foram analisadas as abordagens de segmentação de cada lesão de forma individual e de forma conjunta utilizando a U-Net, sendo esta última abordagem a que se destacou, atingindo uma acurácia de 93,84% e uma F1-Score de 93,24%, com um tempo de teste de 0,841 s. Desse modo, essa abordagem se mostrou eficiente para segmentar as lesões provenientes da RD, mantendo um compromisso entre performance e tempo de teste.

Para realizar a tarefa de classificação, foram avaliadas dez arquiteturas de *CNNs*. Os resultados mostram que a arquitetura EfficientNet, utilizando a configuração EfficientNetB1, apresentou uma acurácia de 99,45% e um F1-Score de 99,20%, indicando alta capacidade de classificar corretamente as retinografias, com um tempo de teste de 0,519 ms. Esses resultados podem ser atribuídos à sua arquitetura, que equilibra profundidade, largura e resolução de entrada de forma otimizada. Portanto, este método mostrou-se eficiente na classificação das retinografias, apresentando um bom desempenho tanto na detecção quanto na classificação do grau de RD.

Em relação aos resultados obtidos pelo Grad-CAM, verificou-se que os mapas de ativação apresentados representam com precisão as áreas correspondentes à presença da RD, justificando a classificação atribuída pela arquitetura. Sendo uma ferramenta capaz de indicar as regiões relevantes para o processo de classificação.

6.1 Trabalhos Futuros

Para trabalhos futuros, pretende-se explorar outras arquiteturas para segmentação, como variantes da U-Net, para aprimorar a segmentação das lesões da RD. Também pretende-se explorar outras arquiteturas de *CNNs* para realizar a tarefa de classificação. Além disso, busca-se validar a abordagem proposta em diferentes conjuntos de dados, a fim de avaliar a robustez e a capacidade de generalização dos modelos em diferentes distribuições de imagens.

No contexto da interpretabilidade, outras técnicas de Inteligência Artificial Explicável podem ser avaliadas para fornecer uma compreensão mais detalhada das decisões

do modelo, aumentando sua confiabilidade no ambiente clínico. Além disso, a integração desses métodos em sistemas hospitalares pode facilitar a adoção da tecnologia por parte dos especialistas, auxiliando profissionais de saúde no diagnóstico e acompanhamento da doença.

Referências

- ABDI, H.; WILLIAMS, L. J. Newman-keuls test and tukey test. *Encyclopedia of research design*, Sage Thousand Oaks, CA, p. 1–11, 2010. Citado na página 54.
- ADADI, A.; BERRADA, M. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, IEEE, v. 6, p. 52138–52160, 2018. Citado na página 40.
- ALGHAMDI, H. S. Towards explainable deep neural networks for the automatic detection of diabetic retinopathy. *Applied Sciences*, MDPI, v. 12, n. 19, p. 9435, 2022. Citado 3 vezes nas páginas 22, 70 e 72.
- BALABAN, S. Deep learning and face recognition: the state of the art. *Biometric and surveillance technology for human and activity identification XII*, SPIE, v. 9457, p. 68–75, 2015. Citado na página 34.
- BANTING, F. G.; BEST, C. H. et al. The internal secretion of the pancreas. University Library: Pub. by the librarian, 1922. Citado na página 24.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006. Citado na página 49.
- Brasil, Ministério da Saúde. *Dia Nacional do Diabetes*. 2024. Acesso em: 5 dez. 2024. Disponível em: <<https://bvsms.saude.gov.br/26-6-dia-nacional-do-diabetes-4/#:~:text=O%20Brasil%20%C3%A9%20o%205%C2%BA,chege%20a%2021%2C5%20milh%C3%B5es>>. Citado na página 24.
- Brazilian Council of Ophthalmology. *Brazilian Council of Ophthalmology - Official Website*. 2024. Accessed on: December 19, 2024. Disponível em: <<https://www.cbo.net.br>>. Citado na página 26.
- BROCK, A. et al. High-performance large-scale image recognition without normalization. In: PMLR. *International conference on machine learning*. [S.l.], 2021. p. 1059–1071. Citado na página 39.
- BYTZER, P. et al. Gi symptoms in diabetes mellitus are associated with both poor glycemic control and diabetic complications. *Official journal of the American College of Gastroenterology/ ACG*, LWW, v. 97, n. 3, p. 604–611, 2002. Citado na página 24.
- CANDRILLI, S. D. et al. Prevalence and the associated burden of illness of symptoms of diabetic peripheral neuropathy and diabetic retinopathy. *Journal of Diabetes and its Complications*, Elsevier, v. 21, n. 5, p. 306–314, 2007. Citado na página 26.
- CHEN, Z.; XU, L.; ZHANG, Y. Image denoising using attention-residual convolutional neural networks. *arXiv preprint arXiv:2101.07713*, 2021. Accessed: 2024-12-22. Disponível em: <<https://arxiv.org/abs/2101.07713>>. Citado na página 47.
- COUNTRY, G. P. by. World population review, 2022. : <https://worldpopulationreview.com/countryrankings/freedom-index-by-country>, 2022. Citado na página 24.

- DECENCIERE, E. et al. Teleophta: Machine learning and image processing methods for teleophthalmology. *Irbm*, Elsevier, v. 34, n. 2, p. 196–203, 2013. Citado 2 vezes nas páginas 21 e 62.
- DINGLI, A.; FARRUGIA, D. *Neuro-Symbolic AI: Design transparent and trustworthy systems that understand the world as you do*. [S.l.]: Packt Publishing Ltd, 2023. Citado 2 vezes nas páginas 30 e 31.
- EL-ATEIF, S.; IDRI, A. Single-modality and joint fusion deep learning for diabetic retinopathy diagnosis. *Scientific African*, Elsevier, v. 17, p. e01280, 2022. Citado 2 vezes nas páginas 21 e 70.
- ES-SABERY, F. et al. Sentence-level classification using parallel fuzzy deep learning classifier. *IEEE Access*, PP, p. 1–1, 01 2021. Citado na página 36.
- FURTADO, P. Multi-class segmentation of diabetic retinopathy lesions: effects of metrics, improvements and loss. In: IEEE. *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. [S.l.], 2020. p. 1410–1417. Citado 3 vezes nas páginas 21, 61 e 62.
- GONZALEZ, R. C. *Digital image processing*. [S.l.]: Pearson education india, 2009. Citado na página 18.
- GROUP, E. T. D. R. S. R. et al. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airle house classification: Etdrs report number 10. *Ophthalmology*, Elsevier, v. 98, n. 5, p. 786–806, 1991. Citado na página 28.
- GU, J. et al. Recent advances in convolutional neural networks. *Pattern recognition*, Elsevier, v. 77, p. 354–377, 2018. Citado na página 34.
- GUO, Y. et al. Deep learning for visual understanding: A review. *Neurocomputing*, Elsevier, v. 187, p. 27–48, 2016. Citado na página 35.
- GUPTA, A. et al. Diabetic retinopathy: A review. *International Journal of Retina and Vitreous*, Springer, v. 4, n. 1, p. 1–19, 2018. Citado na página 26.
- GUPTA, N. et al. Diabetic retinopathy and vegf. *The open ophthalmology journal*, Bentham Science Publishers, v. 7, p. 4, 2013. Citado na página 29.
- HARIS, M. et al. Segmentation of ocular pathologies using deep convolutional neural network. In: IEEE. *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. [S.l.], 2019. p. 1–5. Citado 3 vezes nas páginas 21, 61 e 62.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 38.
- IBTEHAZ, N.; RAHMAN, M. S. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks*, Elsevier, v. 121, p. 74–87, 2020. Citado 2 vezes nas páginas 38 e 39.
- KAGGLE. Aptos 2019 blindness detection. 2019. Disponível em: <<https://www.kaggle.com/c/aptos2019-blindness-detection/data>>. Citado na página 42.

- KHALIFA, N. E. M. et al. Deep transfer learning models for medical diabetic retinopathy detection. *Acta Informatica Medica*, The Academy of Medical Sciences of Bosnia and Herzegovina, v. 27, n. 5, p. 327, 2019. Citado 3 vezes nas páginas 21, 70 e 71.
- KOU, C. et al. An enhanced residual u-net for microaneurysms and exudates segmentation in fundus images. *IEEE Access*, IEEE, v. 8, p. 185514–185525, 2020. Citado 2 vezes nas páginas 21 e 61.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012. Citado na página 37.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. Citado na página 34.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado na página 33.
- LI, G. et al. Efficient densely connected convolutional neural networks. *Pattern Recognition*, Elsevier, v. 109, p. 107610, 2021. Citado na página 34.
- LI, X. et al. Interpretdl: explaining deep models in paddlepaddle. *Journal of Machine Learning Research*, v. 23, n. 197, p. 1–6, 2022. Citado na página 18.
- MACCUISH, A. Early detection and screening for diabetic retinopathy. *Eye*, Nature Publishing Group, v. 7, n. 2, p. 254–259, 1993. Citado na página 28.
- MENDOLA, N. D. et al. Prevalence of total, diagnosed, and undiagnosed diabetes among adults: United states, 2013-2016. US Department of Health and Human Services, Centers for Disease Control and . . . , 2018. Citado na página 24.
- MESSAY, T.; HARDIE, R. C.; ROGERS, S. K. A new computationally efficient cad system for pulmonary nodule detection in ct imagery. *Medical image analysis*, Elsevier, v. 14, n. 3, p. 390–406, 2010. Citado na página 17.
- MURUGAPPAN, M. et al. A novel few-shot classification framework for diabetic retinopathy detection and grading. *Measurement*, Elsevier, v. 200, p. 111485, 2022. Citado 4 vezes nas páginas 22, 70, 72 e 73.
- NATHAN, D. M.; GROUP, D. R. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes care*, Am Diabetes Assoc, v. 37, n. 1, p. 9–16, 2014. Citado na página 29.
- NAZIR, T. et al. Diabetic retinopathy lesions detection using faster-rcnn from retinal images. In: . [S.l.: s.n.], 2020. Citado 2 vezes nas páginas 21 e 61.
- NIXON, M.; AGUADO, A. *Feature extraction and image processing for computer vision*. [S.l.]: Academic press, 2019. Citado na página 18.
- PORWAL, P. et al. Idrid: Diabetic retinopathy–segmentation and grading challenge. *Medical image analysis*, Elsevier, v. 59, p. 101561, 2020. Citado 3 vezes nas páginas 27, 42 e 43.

- RAHMAN, A. et al. Neuroxai++: An efficient x-ai intensive brain cancer detection and localization. In: IEEE. *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*. [S.l.], 2023. p. 1–6. Citado na página 22.
- REBOUÇAS FILHO, P. P. Métodos de contornos ativos crisp adaptativo 2d e 3d aplicados na segmentação dos pulmões em imagens de tomografia computadorizada do tórax. 2013. Citado na página 17.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015. Citado na página 49.
- SAMBYAL, N. et al. Modified u-net architecture for semantic segmentation of diabetic retinopathy images. *Biocybernetics and Biomedical Engineering*, Elsevier, v. 40, n. 3, p. 1094–1109, 2020. Citado 3 vezes nas páginas 21, 61 e 62.
- SAMPAIO, V.; CORDEIRO, F. A study on class activation map methods to detect masses in mammography images using weakly supervised learning. In: *Anais do XIX Encontro Nacional de Inteligência Artificial e Computacional*. Porto Alegre, RS, Brasil: SBC, 2022. p. 437–448. ISSN 2763-9061. Disponível em: <<https://sol.sbc.org.br/index.php/eniac/article/view/22802>>. Citado na página 22.
- 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. Citado na página 40.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, Oxford University Press (OUP), v. 52, n. 3-4, p. 591–611, dec 1965. Disponível em: <<https://doi.org/10.1093/biomet/52.3-4.591>>. Citado na página 52.
- SILVA, E.; SOUZA, A.; COSTA, F. Análise entre técnicas para redução de ruído em imagens de ressonância magnética. *CBEB 2014 - Congresso Brasileiro de Engenharia Biomédica*, 2014. Disponível em: <https://www.canal6.com.br/cbeb/2014/artigos/cbeb2014_submission_344.pdf>. Citado na página 47.
- SILVA, I. N. et al. Artificial neural networks: a practical course. *Springer*, 2016. Citado na página 30.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado 3 vezes nas páginas 35, 36 e 37.
- STIGLER, S. M. *The history of statistics: The measurement of uncertainty before 1900*. [S.l.]: Harvard University Press, 1990. Citado na página 53.
- STITT, A. W. et al. The progress in understanding and treatment of diabetic retinopathy. *Progress in retinal and eye research*, Elsevier, v. 51, p. 156–186, 2016. Citado na página 29.
- SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9. Citado na página 37.
- TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International conference on machine learning*. [S.l.], 2019. p. 6105–6114. Citado na página 38.

- TAUFIQURRAHMAN, S. et al. Diabetic retinopathy classification using a hybrid and efficient mobilenetv2-svm model. In: IEEE. *2020 IEEE REGION 10 CONFERENCE (TENCON)*. [S.l.], 2020. p. 235–240. Citado 3 vezes nas páginas 22, 70 e 71.
- TIAN, Y. et al. Training and testing object detectors with virtual images. *IEEE/CAA Journal of Automatica Sinica*, IEEE, v. 5, n. 2, p. 539–546, 2018. Citado na página 45.
- TJOA, E.; GUAN, C. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems*, IEEE, v. 32, n. 11, p. 4793–4813, 2020. Citado na página 40.
- WILKINSON C. P., F. F. L. K. R. E. L. P. P. A. C. D. D. M. D. D. K. A. P. R. V. J. T. . L. F. roposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *HIGHER TECHNICAL SCHOOL OF TELECOMMUNICATION ENGINEERS*, Proposed international clinical diabetic, v. 9, p. 1677–1682, 2003. Citado na página 27.
- WU L., F.-L. P. S. J. H.-B. E. . M. M. Classification of diabetic retinopathy and diabetic macular edema. *World Journal of Diabetes*, 2013, v. 4, p. 290, 2024. Citado na página 27.
- YAU, J. W. et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes care*, Am Diabetes Assoc, v. 35, n. 3, p. 556–564, 2012. Citado na página 26.
- YIN, B. et al. Localnorm: Robust image classification through dynamically regularized normalization. *arXiv preprint arXiv:1902.06550*, 2019. Accessed: 2024-12-22. Disponível em: <<https://arxiv.org/abs/1902.06550>>. Citado na página 48.
- ZHANG, A. et al. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021. Citado na página 35.
- ZHANG, Z. et al. Deep learning for environmentally robust speech recognition: An overview of recent developments. *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM New York, NY, USA, v. 9, n. 5, p. 1–28, 2018. Citado na página 34.
- ZHOU, B. et al. Learning deep features for discriminative localization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 2921–2929. Citado na página 40.

7 Tabela F para $\alpha = 0.05$.

Critical Values of the F -Distribution: $\alpha = 0.05$

Denom. d.f.	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
31	4.160	3.305	2.911	2.679	2.523	2.409	2.323	2.255	2.199	2.153
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244	2.189	2.142
33	4.139	3.285	2.892	2.659	2.503	2.389	2.303	2.235	2.179	2.133
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225	2.170	2.123
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209	2.153	2.106
37	4.105	3.252	2.859	2.626	2.470	2.356	2.270	2.201	2.145	2.098
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194	2.138	2.091
39	4.091	3.238	2.845	2.612	2.456	2.342	2.255	2.187	2.131	2.084
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
41	4.079	3.226	2.833	2.600	2.443	2.330	2.243	2.174	2.118	2.071
42	4.073	3.220	2.827	2.594	2.438	2.324	2.237	2.168	2.112	2.065
43	4.067	3.214	2.822	2.589	2.432	2.318	2.232	2.163	2.106	2.059
44	4.062	3.209	2.816	2.584	2.427	2.313	2.226	2.157	2.101	2.054
45	4.057	3.204	2.812	2.579	2.422	2.308	2.221	2.152	2.096	2.049
46	4.052	3.200	2.807	2.574	2.417	2.304	2.216	2.147	2.091	2.044
47	4.047	3.195	2.802	2.570	2.413	2.299	2.212	2.143	2.086	2.039
48	4.043	3.191	2.798	2.565	2.409	2.295	2.207	2.138	2.082	2.035
49	4.038	3.187	2.794	2.561	2.404	2.290	2.203	2.134	2.077	2.030
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951
90	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043	1.986	1.938
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910
140	3.909	3.061	2.669	2.436	2.279	2.164	2.076	2.005	1.947	1.899
180	3.894	3.046	2.655	2.422	2.264	2.149	2.061	1.990	1.932	1.884
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878
∞	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831