

**UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS**

**Alexandre Moretti Bernardo**

**Auxílio ao Diagnóstico da Retinopatia Diabética por  
Segmentação Semântica utilizando rede neural *U-Net* em  
Imagens de Retina**

**São Carlos**

**2022**



**Alexandre Moretti Bernardo**

**Auxílio ao Diagnóstico da Retinopatia Diabética por  
Segmentação Semântica utilizando rede neural *U-Net* em  
Imagens de Retina**

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências - Programa de Pós-Graduação em Engenharia Elétrica.

Área de concentração: Processamento de Sinais e Instrumentação

Orientador: Prof. Dr. Evandro Luís Linhari Rodrigues

**São Carlos**

**2022**

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS  
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da  
EESC/USP com os dados inseridos pelo(a) autor(a).

B518a Bernardo, Alexandre Moretti  
Auxílio ao Diagnóstico da Retinopatia Diabética  
por Segmentação Semântica utilizando rede neural U-Net  
em Imagens de Retina / Alexandre Moretti Bernardo;  
orientador Evandro Luis Linhari Rodrigues. São Carlos,  
2022.

Dissertação (Mestrado) - Programa de  
Pós-Graduação em Engenharia Elétrica e Área de  
Concentração em Processamento de Sinais e  
Instrumentação -- Escola de Engenharia de São Carlos da  
Universidade de São Paulo, 2022.

1. Retinopatia diabética. 2. Processamento de  
imagens de retina. 3. Segmentação de microaneurismas,  
hemorragias e exsudatos na retina. 4. Aprendizado  
Profundo. 5. Aprendizado Supervisionado. 6. Rede Neural  
U-Net. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

## FOLHA DE JULGAMENTO

Candidato: Engenheiro **ALEXANDRE MORETTI BERNARDO**.

Título da dissertação: "Auxílio ao diagnóstico da retinopatia diabética por segmentação semântica utilizando rede neural UNET em imagens de retina".

Data da defesa: 29/04/2022.

### Comissão Julgadora

### Resultado

Prof. Associado **Evandro Luis Linhari Rodrigues**  
**(Orientador)**  
(Escola de Engenharia de São Carlos – EESC/USP)

Aprovado

Profa. Titular **Agma Juci Machado Traina**  
(Instituto de Ciências Matemáticas e de Computação/ICMC-USP)

Aprovado

Prof. Dr. **Marcio Merino Fernandes**  
(Universidade Federal de São Carlos/UFSCar)

Aprovado

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:  
Prof. Associado **João Bosco Augusto London Junior**

Presidente da Comissão de Pós-Graduação:  
Prof. Titular **Murilo Araujo Romero**



*A Deus, criador de tudo.*





## AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, por sua presença em minha vida, por ser meu guia e me mostrar o real sentido de tudo. Por Ele me fazer mais forte durante esse percurso e diante dos desafios.

À minha família, por todo suporte, amor e compreensão durante todo esse tempo, no qual, mesmo distante, se fizeram presentes em minha vida. Agradeço à minha mãe Vitalina ao meu pai Luiz por todo amor, carinho e cuidado. Também agradeço ao meu irmão, Samuel, por seu exemplo de dedicação, pelo apoio e momentos de partilha sobre os desafios acadêmicos.

Em especial, agradeço à minha namorada e futura esposa, Mariângela, por tanta paciência, compreensão e companheirismo. Por todos os momentos de convívio, os quais incluem missas, orações, jantares, passeios, filmes assistidos, entre outros. Agradeço à ela pelos aprendizados, dedicação, amor e principalmente por partilhar nosso sonho de construir uma família, um futuro juntos.

Ao meu amigo e professor orientador Evandro, por todos ensinamentos, conversas e por confiar em mim, enxergando um potencial maior do que eu mesmo conseguia ver. Desse modo, sou grato a ele por todo o auxílio a mim oferecido para que este trabalho fosse realizado.

Aos professores que participaram do exame de qualificação deste trabalho, professor Dr. Marcelo Andrade da Costa Vieira e professor Dr. Adilson Gonzaga, pelas valiosas contribuições.

À professora Dra. Agma Juci Machado Traina e ao professor Dr. Márcio Merino Fernandes pelo olhar cuidadoso e dedicado a este trabalho.

Ao meu amigo Nicolas por ter dividido comigo sua experiência e conhecimento acadêmico e, também, pelo bom convívio desde o tempo da graduação. Reconheço sua paciência, dedicação, ajuda e amizade dispensadas a mim durante esses anos.

Aos amigos da DMC Equipamentos que estiveram presentes durante todo percurso do mestrado, pela compreensão das privações necessárias durante a elaboração deste trabalho.

A todos os meus amigos que estiveram presentes no meu percurso de mestrado, pela amizade, companheirismo, ajuda e orações.

Por fim, agradeço a todos aqueles que contribuíram direta ou indiretamente para a realização deste trabalho.



*“Conhecendo-te, tu te humilharás ao perceber que, por ti mesmo, nada és”*  
*Santa Catarina de Sena*



## RESUMO

Bernardo, AM **Auxílio ao Diagnóstico da Retinopatia Diabética por Segmentação Semântica utilizando rede neural *U-Net* em Imagens de Retina.** 2022. 132p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2022.

O diabetes mellitus é uma doença caracterizada pela disfunção na produção ou ação da insulina no organismo. Esta disfunção pode ser causada por diversos fatores, incluindo fatores genéticos, ambientais, nutricionais, podendo estes serem agravados pelo sedentarismo. Esta doença pode ocasionar complicações microvasculares na retina, caracterizando a retinopatia diabética. Esta, por sua vez, é a principal causa de cegueira evitável em pessoas com idade ativa no mundo. Aproximadamente uma em cada três pessoas portadoras do diabetes mellitus tem algum grau de retinopatia diabética e estima-se que até 2045 haja um crescimento de 46% nos casos dessa doença. Dessa forma, agilizar, facilitar e padronizar o diagnóstico desta condição é de grande importância para reduzir estes altos índices e projeções. O uso de processamento de imagens médicas tem auxiliado o diagnóstico de diversas anomalias difíceis de serem detectadas sem tratamento computacional. Este trabalho, portanto, propõe a aplicação de processamento de imagens de retina para a segmentação automática de regiões de interesse como, exsudatos duros, exsudatos moles, microaneurismas e hemorragias. Foi proposta a avaliação das regiões de interesse isoladamente em seus canais de cores RGB e suas combinações, bem como a combinação das diferentes estruturas de retinopatia diabética em uma mesma imagem. Cada imagem, com seu respectivo *ground truth*, foi dividida em subimagens, as quais foram utilizadas como entradas da rede neural *U-Net* proposta. O melhor método combinado obteve acurácia, sensibilidade, precisão e especificidade de 97,51%, 37,49%, 58,86% e 99,24% respectivamente, para todas as regiões de interesse somadas na mesma imagem e utilizando os canais RGB. Dessa maneira, o método proposto não apresentou o melhor resultado da literatura, porém, apresentou, comparativamente, uma análise individual e associada das regiões de interesse da retinopatia diabética. Pré-processar as imagens separando seus canais de cores permitiu avaliar especificamente, para a rede *U-Net*, a combinação que potencializa os resultados das estruturas características da retinopatia diabética, não havendo resultados comparativos presentes na literatura. Obter um método que permita segmentar todas as regiões de interesse combinadas é alvo de interesse por propiciar maior agilidade no processamento das imagens, neste caso, uma única rede neural será otimizada para esta tarefa.

**Palavras-chave:** Retinopatia diabética. Processamento de imagens de retina. Segmentação de microaneurismas, hemorragias e exsudatos na retina. Aprendizado Profundo.

Aprendizado Supervisionado. Rede Neural *U-Net*.

## ABSTRACT

Bernardo, AM **Diagnosing Diabetic Retinopathy by Semantic Segmentation using *U-Net* neural network in Retina Images**. 2022. 132p. Dissertation (Master) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2022.

Diabetes mellitus is a disease characterized by dysfunction in the production or action of insulin in the body. This dysfunction may result from a variety of factors, including result of genetic, environmental and nutritional factors aggravated by the sedentary lifestyle. Diabetes mellitus can cause specific microvascular complications in the retina, characterizing diabetic retinopathy. Diabetic retinopathy is the leading cause of preventable blindness in people of working age in the world. Approximately one in three people with diabetes mellitus have some degree of Diabetic Retinopathy and it is estimated that by 2045 there will be a growth of 46 % in the cases of this disease. Streamlining, facilitating and standardizing the diagnosis of retinal examination is therefore the target of studies to reduce its growth. The use of medical imaging has aided the diagnosis of several anomalies difficult to detect without computational treatment. This work proposes the application of retinal image processing for the automatic segmentation of regions of interest such as microaneurysms, hemorrhages and exudates. To detect such regions, the IDRiD database was used with retinal images previously evaluated by specialists. Using this database, it was proposed to evaluate the regions of interest separately in their RGB color channels and their combinations, as well as the combination of different structures of diabetic retinopathy in the same image. Each image, with its respective ground truth, was divided into subimages. These, in turn, were used as inputs to the proposed U-Net neural network. The best combined method obtained accuracy, sensitivity, precision and specificity of 97.51%, 37.49%, 58.86% and 99.24% respectively, for all regions of interest summed in the same image and using RGB channels. Thus, the proposed method did not present the best result in the literature, however, it comparatively presents an individual and a combined analysis of the regions of interest for diabetic retinopathy. Preprocessing the images by separating their color channels made it possible to specifically evaluate, for the U-net network, the combination that potentiates the results of the characteristic structures of diabetic retinopathy, with no comparative results present in the literature. Obtaining a method that presents good results from the combined regions is relevant because it allows greater agility in the processing of the images, since a single neural network segments the regions of interest without the need for prior distinction.

**Keywords:** Diabetic retinopathy. Processing of retinal images. Detection of microaneurysms, haemorrhages and exudates in the retina. Deep Learning. Supervised Learning.

Convolutional Neural Network Unet.



## LISTA DE FIGURAS

Figura 1 – Mapa do cenário mundial da RD em 2021 e sua projeção para 2045. . . . .	26
Figura 2 – Histórico de crescimento do diabetes no mundo. . . . .	27
Figura 3 – Representação de uma retina saudável e sua respectiva imagem. . . . .	27
Figura 4 – Representação de uma retina com retinopatia diabética e sua respectiva imagem. . . . .	27
Figura 5 – Imagem de retina apresentando os pontos de interesse (microaneurismas, hemorragias, exsudatos duros e exsudatos moles). . . . .	28
Figura 6 – Imagem de retina apresentando o <i>ground truth</i> dos pontos de interesse (microaneurismas, hemorragias, exsudatos duros e exsudatos moles). . . . .	28
Figura 7 – Estágios clínicos e principais eventos patogênicos da retinopatia diabética.	35
Figura 8 – Sistema representativo de um neurônio em RNA. . . . .	39
Figura 9 – Sistema representativo de uma rede neural artificial. . . . .	40
Figura 10 – Diferença entre RNA e <i>Deep learning</i> . . . . .	42
Figura 11 – Diferença entre aprendizado profundo e aprendizado de máquina tradi- cional. . . . .	42
Figura 12 – Ilustração de um Filtro Convolutacional. . . . .	44
Figura 13 – Três diferentes tipos de operação <i>pooling</i> . . . . .	45
Figura 14 – Publicações de artigos relacionados a <i>U-Net</i> . . . . .	49
Figura 15 – Diagrama de blocos <i>U-Net</i> . . . . .	50
Figura 16 – Aplicações da <i>U-Net</i> em diferentes imagens médicas . . . . .	51
Figura 17 – Métrica particular utilizada em (GUO <i>et al.</i> , 2020) . . . . .	55
Figura 18 – Resultados da segmentação de microaneurismas. . . . .	58
Figura 19 – Resultados da segmentação de hemorragias. . . . .	59
Figura 20 – Ilustração da subdivisão da imagem original da retina em subimagens.	62
Figura 21 – Ilustração da subdivisão da imagem original do <i>ground truth</i> em sub- imagens. . . . .	63
Figura 22 – Diagrama de blocos da rede simplificada implementada. . . . .	66
Figura 23 – Canal individual: vermelho (R), verde (G) e azul (B), respectivamente.	71
Figura 24 – Canal duplo: vermelho e verde (RG), vermelho e azul (RB) e verde e azul (GB), respectivamente. . . . .	72
Figura 25 – Imagem original com os três canais de cores (RGB) e <i>ground truth</i> com todos os pontos de interesse somados, respectivamente. . . . .	72
Figura 26 – Canal vermelho e respectivo <i>ground truth</i> . . . . .	73
Figura 27 – Canal verde e respectivo <i>ground truth</i> . . . . .	73
Figura 28 – Canal azul e respectivo <i>ground truth</i> . . . . .	73
Figura 29 – Canais vermelho e verde na mesma imagem e respectivo <i>ground truth</i> . . . . .	74

Figura 30 – Canais vermelho e azul na mesma imagem e respectivo <i>ground truth</i> . . .	74
Figura 31 – Canais verde e azul na mesma imagem e respectivo <i>ground truth</i> . . . .	74
Figura 32 – Canais vermelho e respectivo <i>ground truth</i> . . . . .	75
Figura 33 – Canal verde e respectivo <i>ground truth</i> . . . . .	75
Figura 34 – Canal azul e respectivo <i>ground truth</i> . . . . .	76
Figura 35 – Canais vermelho e verde na mesma imagem e respectivo <i>ground truth</i> . .	76
Figura 36 – Canais vermelho e azul na mesma imagem e respectivo <i>ground truth</i> . .	76
Figura 37 – Canais verde e azul na mesma imagem e respectivo <i>ground truth</i> . . . .	77
Figura 38 – Canal vermelho e respectivo <i>ground truth</i> . . . . .	77
Figura 39 – Canal verde e respectivo <i>ground truth</i> . . . . .	78
Figura 40 – Canal azul e respectivo <i>ground truth</i> . . . . .	78
Figura 41 – Canais vermelho e verde na mesma imagem e respectivo <i>ground truth</i> . .	78
Figura 42 – Canais vermelho e azul na mesma imagem e respectivo <i>ground truth</i> . .	79
Figura 43 – Canais verde e azul na mesma imagem e respectivo <i>ground truth</i> . . . .	79
Figura 44 – Canal vermelho e respectivo <i>ground truth</i> . . . . .	80
Figura 45 – Canal verde e respectivo <i>ground truth</i> . . . . .	80
Figura 46 – Canal azul e respectivo <i>ground truth</i> . . . . .	80
Figura 47 – Canais vermelho e verde na mesma imagem e respectivo <i>ground truth</i> . .	81
Figura 48 – Canais vermelho e azul na mesma imagem e respectivo <i>ground truth</i> . .	81
Figura 49 – Canais verde e azul na mesma imagem e respectivo <i>ground truth</i> . . . .	81
Figura 50 – Curva ROC - Exsudatos Duros com seus respectivos canais de cores . .	82
Figura 51 – Curva ROC - Exsudatos Moles com seus respectivos canais de cores . .	83
Figura 52 – Curva ROC - Hemorragias com seus respectivos canais de cores . . . .	83
Figura 53 – Curva ROC - Microaneurismas com seus respectivos canais de cores . .	84
Figura 54 – Curva ROC: Três melhores resultados de exsudatos duros. . . . .	88
Figura 55 – Curva ROC: Três melhores resultados de exsudatos moles. . . . .	88
Figura 56 – Curva ROC: Três melhores resultados de hemorragias. . . . .	89
Figura 57 – Curva ROC: Três melhores resultados de microaneurismas. . . . .	89
Figura 58 – Curva ROC: Três melhores resultados de exsudatos duros e exsudatos moles. . . . .	90
Figura 59 – Curva ROC: Três melhores resultados de exsudatos duros e hemorragias.	90
Figura 60 – Curva ROC: Três melhores resultados de exsudatos duros e microaneu- rismas. . . . .	91
Figura 61 – Curva ROC: Três melhores resultados de hemorragias e exsudatos moles.	91
Figura 62 – Curva ROC: Três melhores resultados de hemorragias e microaneurismas.	92
Figura 63 – Curva ROC: Três melhores resultados de microaneurismas e exsudatos moles. . . . .	92
Figura 64 – Curva ROC: Três melhores resultados de todas as estruturas somadas.	93

Figura 65 – Exsudatos duros: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	93
Figura 66 – Exsudatos moles: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	94
Figura 67 – Hemorragias: Imagem original, <i>ground truth</i> e predição, respectivamente.	94
Figura 68 – Microaneurismas: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	94
Figura 69 – Exsudatos duros e moles: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	95
Figura 70 – Exsudatos duros e hemorragias: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	95
Figura 71 – Exsudatos duros e microaneurismas: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	95
Figura 72 – Hemorragias e exsudatos moles: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	96
Figura 73 – Hemorragias e microaneurismas: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	96
Figura 74 – Microaneurismas e exsudatos moles: Imagem original, <i>ground truth</i> e predição, respectivamente. . . . .	96
Figura 75 – Diagrama apresentando os melhores resultados de acurácia das estruturas originais e suas combinações . . . . .	98
Figura 76 – Diagrama apresentando os melhores resultados de sensibilidade das estruturas originais e suas combinações . . . . .	98
Figura 77 – Diagrama apresentando os melhores resultados de precisão das estruturas originais e suas combinações . . . . .	99
Figura 78 – Diagrama apresentando os melhores resultados de especificidade das estruturas originais e suas combinações . . . . .	99
Figura 79 – Diagrama apresentando os melhores resultados de AUC das estruturas originais e suas combinações . . . . .	100
Figura 80 – Diagrama apresentando os melhores resultados de IoU das estruturas originais e suas combinações . . . . .	100



## LISTA DE TABELAS

Tabela 1 – Combinações de regiões de interesse avaliadas . . . . .	65
Tabela 2 – Parâmetros da rede neural <i>U-Net</i> utilizada. . . . .	66
Tabela 3 – Descrição das camadas da arquitetura U-Net . . . . .	67
Tabela 4 – Computador e sistema utilizado . . . . .	68
Tabela 5 – Resultado da AUC para os componentes de canais de cores e suas combinações . . . . .	84
Tabela 6 – Resultados das estruturas separadas otimizando a métrica IoU . . . . .	85
Tabela 7 – Resultados das estruturas combinadas otimizando a métrica IoU . . . . .	85
Tabela 8 – Resultados das estruturas separadas otimizando a métrica acurácia . . . . .	85
Tabela 9 – Resultados das estruturas combinadas otimizando a métrica acurácia . . . . .	85
Tabela 10 – Resultados das estruturas separadas otimizando a métrica sensibilidade . . . . .	86
Tabela 11 – Resultados das estruturas combinadas otimizando a métrica sensibilidade . . . . .	86
Tabela 12 – Resultados das estruturas separadas otimizando a métrica precisão . . . . .	86
Tabela 13 – Resultados das estruturas combinadas otimizando a métrica precisão . . . . .	86
Tabela 14 – Resultados das estruturas separadas otimizando a métrica especificidade . . . . .	86
Tabela 15 – Resultados das estruturas combinadas otimizando a métrica especificidade . . . . .	87
Tabela 16 – AUC para resultados apresentados nos gráficos - Três melhores resultados . . . . .	87
Tabela 17 – Resultados comparativos para cada estrutura individual. . . . .	97
Tabela 18 – Treinamentos exsudatos duros . . . . .	120
Tabela 19 – Treinamentos exsudatos duros . . . . .	121
Tabela 20 – Treinamentos exsudatos moles . . . . .	122
Tabela 21 – Treinamentos hemorragias . . . . .	123
Tabela 22 – Treinamentos hemorragias . . . . .	124
Tabela 23 – Treinamentos Microaneurismas . . . . .	124
Tabela 24 – Treinamentos exsudatos duros e moles . . . . .	125
Tabela 25 – Treinamento exsudatos duros e hemorragias . . . . .	126
Tabela 26 – Treinamento exsudatos e microaneurismas . . . . .	127
Tabela 27 – Treinamento exsudatos moles e hemorragias . . . . .	127
Tabela 28 – Treinamento: exsudatos moles e microaneurismas . . . . .	128
Tabela 29 – Treinamentos: Hemorragias e microaneurismas . . . . .	128
Tabela 30 – Treinamento: Todas as imagens somadas . . . . .	129



## LISTA DE ABREVIATURAS E SIGLAS

AUC	<i>Area Under the Curve</i>
BN	<i>Batch Normalization</i> – Normalização em Lotes;
CNN	<i>Convolutional Neural Network</i> – Rede Neural; Convolucional;
CUDA	<i>Computer Unified Device Architecture</i> – Arquitetura Unificada de Dispositivo de Computador;
CPU	<i>Central Process Unit</i> – Unidade Central de Processamento;
DM	Diabete Mellitus
DME	<i>Diabetic Macular Edema</i> – Edema Macular Diabético;
ED	Exsudatos Duros;
EM	Exsudatos Moles;
FC	<i>Fully Connected</i> – Totalmente Conectada;
FCN	<i>Fully Connected Network</i> – Rede Totalmente Conectada;
FP	Falso Positivo;
FN	Falso Negativo;
GPU	<i>Graphics Processing Unit</i> – Unidade de Processamento Gráfico;
HDD	<i>Hard Disk Drive</i> – Unidade de Disco Rígido;
HE	Hemorragias;
IA	Inteligência Artificial;
IDRiD	<i>Indian Diabetic Retinopathy Image Dataset</i> – Conjunto de dados Indiano de Imagens de Retinopatia Diabética;
IEEE	<i>Institute of Electrical and Electronic Engineers</i> – Instituto de Engenheiros Eletricistas e Eletrônicos;
IoU	<i>Intersection over Union</i> – Interceção sobre União.
ISBI	<i>International Symposium on Biomedical Imaging</i> – Simpósio Internacional de Imagem Biomédica;

MA	Microaneurismas;
NN	<i>Nearest Neighbour</i> ;
RD	Retinopatia Diabética;
ReLU	<i>Rectified Linear Unit</i> – Unidade Linear Retificada;
RGB	<i>Red, Green and Blue</i> – Vermelho, Verde e Azul;
RNA	Rede Neural Artificial;
RM	Ressonância Magnética;
MLP	<i>Multilayer Perceptron</i> – Perceptron Multicamadas;
ROC	<i>Receiver Operating Characteristic</i> ;
SSD	<i>Solid State Drive</i> – Unidade de Estado Sólido;
SVM	<i>Support Vector Machine</i> – Máquina de vetores de suporte;
TC	Tomografia Computadorizada;
VP	Verdadeiro Positivo;
VN	Verdadeiro Negativo.



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>25</b>
<b>1.1</b>	<b>Retinopatia Diabética</b>	<b>25</b>
<b>1.2</b>	<b>Processamento de Imagens Médicas</b>	<b>26</b>
<b>1.3</b>	<b>Aprendizado de Máquina</b>	<b>28</b>
<b>1.4</b>	<b><i>Deep learning</i></b>	<b>29</b>
<b>1.5</b>	<b>Proposta</b>	<b>30</b>
<b>1.6</b>	<b>Objetivo</b>	<b>30</b>
<b>1.7</b>	<b>Organização do Texto</b>	<b>31</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>33</b>
<b>2.1</b>	<b>Diabetes Mellitus</b>	<b>33</b>
<b>2.2</b>	<b>Retinopatia Diabética</b>	<b>34</b>
2.2.1	Estágios clínicos e principais eventos patogênicos da RD.	34
<b>2.3</b>	<b>Etapas do processamento de imagens</b>	<b>36</b>
<b>2.4</b>	<b>Inteligência Artificial na Medicina</b>	<b>36</b>
<b>2.5</b>	<b>Aprendizado de Máquina</b>	<b>37</b>
<b>2.6</b>	<b>Aprendizado de Representação</b>	<b>38</b>
<b>2.7</b>	<b>Redes Neurais Artificiais</b>	<b>39</b>
<b>2.8</b>	<b><i>Deep learning</i></b>	<b>41</b>
2.8.1	Aplicações	43
2.8.2	Redes Neurais Convolucionais	43
2.8.2.1	Camada Convolutiva	44
2.8.2.2	Camadas de Agrupamento ( <i>Pooling</i> )	44
2.8.2.3	Camada totalmente conectada	45
2.8.2.4	<i>Dropout</i>	46
2.8.2.5	Rectified Linear Unit (ReLU)	46
2.8.2.6	Topologias de Redes Convolucionais Conhecidas	46
2.8.2.7	Processo de <i>data augmentation</i>	50
2.8.3	Métricas de Avaliação	51
<b>2.9</b>	<b>Considerações sobre o uso da técnica</b>	<b>53</b>
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>55</b>
<b>3.1</b>	<b>Deteção de Retinopatia Diabética por Sistema de Diagnóstico por Computador (CAD)</b>	<b>55</b>
<b>3.2</b>	<b>Segmentação de exsudatos moles e duros</b>	<b>56</b>
<b>3.3</b>	<b>Segmentação de microaneurismas</b>	<b>57</b>

3.4	<b>Segmentação de Hemorragias</b>	57
3.5	<b>Considerações</b>	58
4	<b>MÉTODOS</b>	61
4.1	<b>Base de dados</b>	61
4.2	<b>Pré-processamento</b>	62
4.2.1	Primeira etapa: divisão das imagens originais	62
4.2.2	Segunda etapa: Divisão dos canais RGB	64
4.2.3	Terceira etapa: Combinações das estruturas de interesse	64
4.3	<b>Definições da rede neural</b>	64
4.4	<b>Métrica de avaliação</b>	69
5	<b>APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS</b>	71
5.0.1	Primeira etapa: divisão das imagens originais	71
5.0.2	Segunda etapa: Divisão dos canais RGB	71
5.0.3	Terceira etapa: Análise individual e combinações das estruturas de interesse	87
6	<b>CONCLUSÕES</b>	101
7	<b>TRABALHOS FUTUROS</b>	105
	<b>REFERÊNCIAS</b>	107
	<b>APÊNDICES</b>	117
	<b>APÊNDICE A – RESULTADOS DOS MELHORES TREINAMENTOS</b>	119
	<b>ANEXOS</b>	131

# 1 INTRODUÇÃO

## 1.1 Retinopatia Diabética

A Retinopatia Diabética (RD) é a principal causa da perda de visão em adultos em idade economicamente ativa (20 a 74 anos) (CHEUNG; MITCHELL; WONG, 2010). Aproximadamente uma em cada três pessoas portadoras do diabetes mellitus (DM) tem algum grau de RD e estima-se que uma a cada dez perderá a visão. (YAU *et al.*, 2012).

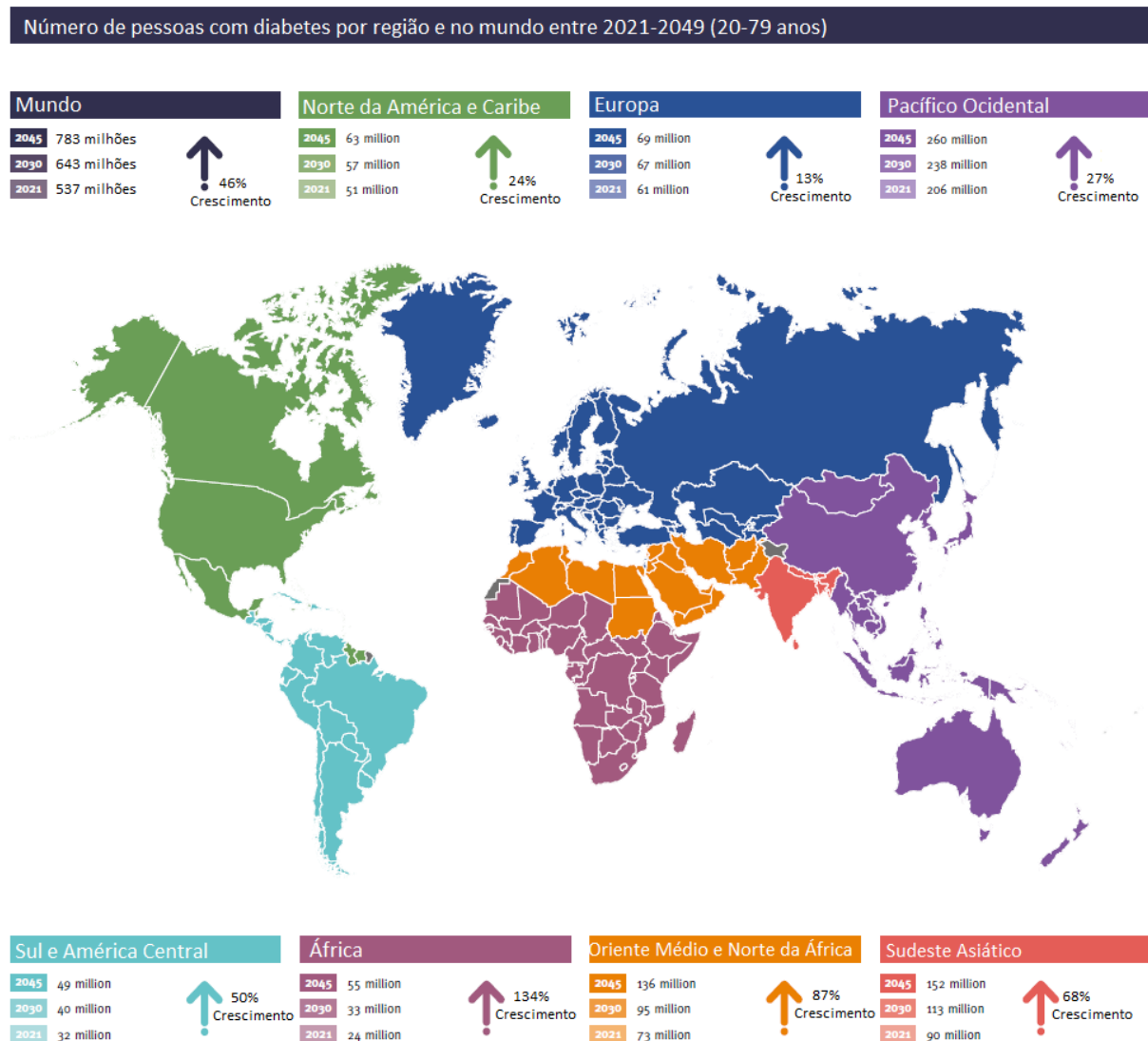
O *Vision Loss Expert Group* estimou que, em 2015, cerca de 1,07% da cegueira mundial foi atribuída à Retinopatia Diabética (FLAXMAN *et al.*, 2017). É previsto que esse número aumente à medida que a prevalência global de diabetes se mantenha em crescimento. A Figura 1 apresenta a distribuição de pessoas com diabetes por região e suas respectivas projeções de crescimento para o ano de 2045. A Figura 2, por sua vez, apresenta a curva histórica de crescimento da doença desde o ano 2000 até o ano de 2021, confirmando a tendência prevista de crescimento da doença.

Sendo assim, a identificação e o tratamento precoce podem prevenir quase toda a cegueira causada pela RD. No entanto, muitas pessoas que possuem diabetes desconhecem a importância e até mesmo a necessidade de examinarem sua visão anualmente. Além disso, estas pessoas são assintomáticas durante os estágios iniciais da doença e/ou não conseguem acesso à atendimento no momento oportuno. Quando ocorre alteração na visão, a RD já não está mais no estágio inicial, o que dificulta o tratamento e, conseqüentemente, agrava a tendência da perda de visão (YAU *et al.*, 2012). A Figura 3 apresenta uma imagem formada em uma retina saudável, já a Figura 4 representa a imagem formada em uma retina com RD em estágio avançado.

Na maioria dos países subdesenvolvidos e sem recursos, a tecnologia e os sistemas de assistência necessários para identificar e tratar de forma eficaz a RD são frequentemente inexistentes. Portanto, investimentos significativos são imprescindíveis para evitar o aumento da perda de visão causada pela RD (FLAXMAN *et al.*, 2017).

Para detecção da RD, deve-se realizar o exame de fundo de olho (retina). Este exame, não invasivo, é realizado após a dilatação da pupila (midríase) a fim de permitir a visualização de toda a retina e, assim, inferir a presença e o grau da RD. Para sua realização, é aplicado um corante chamado fluoresceína na corrente sanguínea, que circula pelos vasos sanguíneos. Ao atingir os vasos sanguíneos dos olhos, são realizadas fotografias permitindo identificar anormalidades na retina. Cabe ressaltar que as alterações típicas vistas precocemente na RD (não proliferativa) são: microaneurismas (MA), hemorragias (HE), exsudatos duros (ED) e exsudatos moles (EM) (WONG *et al.*, 2016), conforme apresentado nas Figuras 5 e 6.

Figura 1 – Mapa do cenário mundial da RD em 2021 e sua projeção para 2045.

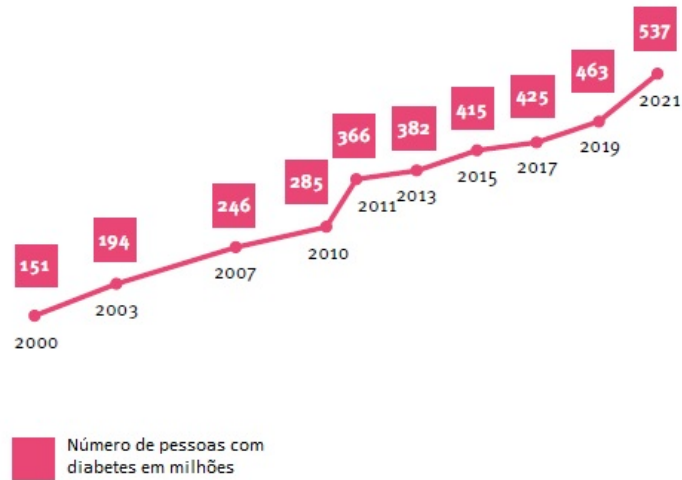


Fonte: Imagem adaptada do Atlas Internacional de diabetes - Décima Edição 2021 (FEDERATION, 2021)

## 1.2 Processamento de Imagens Médicas

Com o crescente uso de sistemas de imagens digitais para diagnósticos médicos, o processamento digital de imagens torna-se cada vez mais importante na área da saúde (DESERNO, 2011). Além dos métodos digitais, como Tomografia Computadorizada (TC) ou Imagem por Ressonância Magnética (RM), as modalidades iniciais de imagens analógicas, como endoscopia ou radiografia, são atualmente equipadas com sensores digitais. Nesse sentido, pode-se afirmar que toda imagem médica pode ser processada com o intuito de auxiliar o diagnóstico médico, possibilitando maior precisão e padronização dos métodos (DESERNO, 2011).

Figura 2 – Histórico de crescimento do diabetes no mundo.



Fonte: Imagem adaptada do Atlas Internacional de diabetes - Décima Edição 2021 (FEDERATION, 2021)

Figura 3 – Representação de uma retina saudável e sua respectiva imagem.



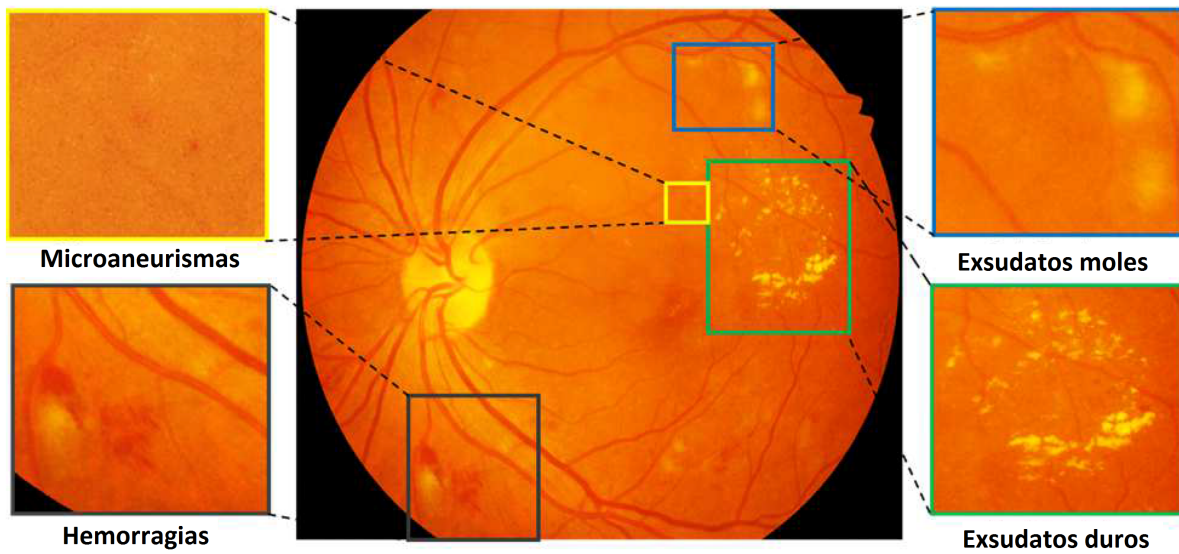
Fonte: Imagem adaptada de North-Valley (2019).

Figura 4 – Representação de uma retina com retinopatia diabética e sua respectiva imagem.



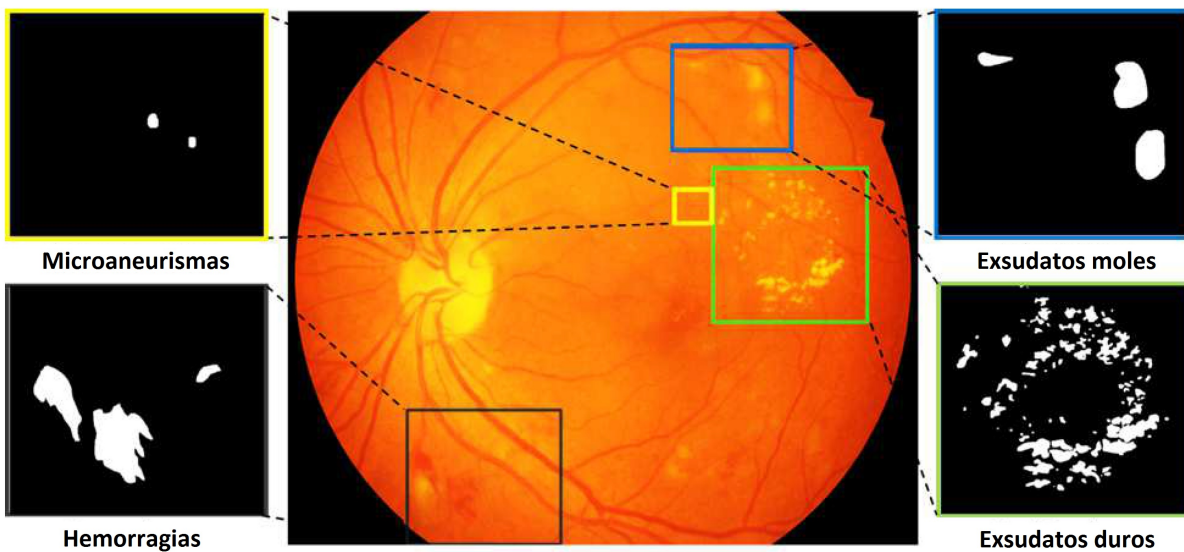
Fonte: Imagem adaptada de North-Valley (2019).

Figura 5 – Imagem de retina apresentando os pontos de interesse (microaneurismas, hemorragias, exsudatos duros e exsudatos moles).



Fonte: Imagem adaptada de Porwal *et al.* (2018a).

Figura 6 – Imagem de retina apresentando o *ground truth* dos pontos de interesse (microaneurismas, hemorragias, exsudatos duros e exsudatos moles).



Fonte: Imagem adaptada de Porwal *et al.* (2018a).

### 1.3 Aprendizado de Máquina

A digitalização de imagens médicas por computadores modernos possibilitou o desenvolvimento de sistemas para análise automatizada. Inicialmente, durante as décadas de 1970 a 1990, a análise de imagens médicas foi realizada com aplicação sequencial de processamento de pixel de baixo nível (filtros de detecção de bordas e linhas, crescimento de regiões) e modelagem matemática (linhas de ajuste, círculos e elipses) para construir

composições baseadas em regras e sistemas que resolviam tarefas específicas. Contudo, tais sistemas especializados eram frequentemente frágeis e suscetíveis a erros.

No final da década de 1990, as técnicas supervisionadas, as quais utilizam os dados de treinamento para desenvolver um sistema, estavam se tornando cada vez mais populares na análise de imagens médicas (LITJENS *et al.*, 2017). Exemplos incluem modelos de formas ativas (para segmentação), métodos atlas (nos quais os atlas adequados aos novos dados formam os dados de treinamento) com conceito de extração de características e uso de classificadores estatísticos (para detecção e diagnóstico auxiliados por computador). Essa abordagem de reconhecimento de padrões ou aprendizado de máquina ainda é muito popular e, por isso, constitui a base de muitos sistemas de análise de imagens médicas comercialmente disponíveis. Sendo assim, há uma mudança de sistemas que são completamente projetados por seres humanos para sistemas que são treinados por computadores. Estes sistemas extraem vetores de características e determinam o limite de decisão ideal no espaço de características de alta dimensão. (LITJENS *et al.*, 2017).

#### 1.4 *Deep learning*

Seguindo esta linha de pesquisa, as próximas etapas evolutivas permitiram que os computadores aprendessem as características que representam os dados para o problema em questão. Este conceito está na base de muitos algoritmos de aprendizagem profunda: modelos (redes) compostos de muitas camadas que transformam dados de entrada (por exemplo, imagens) em saídas (por exemplo, doença presente/ausente) enquanto aprendem características de nível cada vez mais alto. O tipo mais bem-sucedido de modelos para análise de imagens, atualmente, são as redes neurais convolucionais (CNNs) (LITJENS *et al.*, 2017). As CNNs contêm muitas camadas que transformam sua entrada em mapas de características por meio de filtros convolucionais. As primeiras pesquisas com CNNs foram propostas nos anos de 1980 por Fukushima (1980). Em 1990, foi apresentado o primeiro aplicativo de sucesso para reconhecimento de dígitos escritos à mão. Apesar desses sucessos iniciais, o uso de CNNs não ganhou força até que várias novas técnicas foram desenvolvidas para o treinamento eficiente de redes profundas. O uso da CNN obteve destaque em dezembro de 2012 ao ganhar o desafio *ImageNet* com resultados muito superiores aos demais. Nos anos subsequentes, foram alcançados mais progressos ao utilizar arquiteturas relacionadas, mas mais profundas (maior número de camadas) (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

Apesar da comunidade de pesquisa em análise de imagens médicas ter acompanhado esses desenvolvimentos, a transição de sistemas que utilizam características manuais para sistemas que aprendem diretamente dos dados aconteceu de forma gradual. Antes do avanço da *AlexNet*, muitas técnicas diferentes para aprender características eram populares. Por exemplo, Bengio (2013) introduziu CNNs que eram totalmente treinadas somente no

final de sua revisão, em uma seção intitulada “*Global training of deep models*”.

As aplicações de aprendizagem profunda na análise de imagens médicas surgiram inicialmente em *workshops* e conferências e, depois, em periódicos. O número de artigos cresceu rapidamente em 2015 e 2016. O tema se tornou dominante nas principais conferências e uma primeira edição especial apareceu na *IEEE Transaction on Medical Imaging* em maio de 2016 (GREENSPAN; GINNEKEN; SUMMERS, 2016).

No contexto de identificação da RD, diversas técnicas têm sido empregadas, principalmente para a classificação binária da doença em imagens de retina (GULSHAN *et al.*, 2016). A utilização de redes neurais profundas para esta tarefa é adequada por apresentar grande capacidade de extrair características representativas da doença, generalização e robustez das predições. Entretanto, apresentam algumas limitações. Logo, identificar a melhor arquitetura para ajustar os pesos sinápticos é uma tarefa complexa. Outro fator de relevância é o alto custo computacional para a fase de treinamento, pois esse tipo de sistema também requer uma grande base de imagens para treinamento, o que nem sempre é acessível.

## 1.5 Proposta

Este trabalho tem como motivação utilizar técnicas de aprendizagem profunda para auxílio médico na identificação da RD. Isso inclui a utilização de rede neural profunda que permite a identificação de padrões das regiões de interesse. Esta técnica pode convergir para identificação de padrões não usuais que apresentam alto grau de assertividade. Pode-se afirmar que, mesmo com os avanços tecnológicos, a RD é uma doença que apresenta grande perspectiva de crescimento de casos, levando o paciente à perda completa da visão. Por esta razão, ao utilizar técnicas de aprendizagem profunda torna-se possível que o diagnóstico médico seja mais preciso e padronizado. Assim, aplica-se o tratamento médico adequado a fim de que o efeito da cegueira causado no paciente seja evitado. Dessa forma, aliar uma poderosa ferramenta computacional ao auxílio do diagnóstico médico pode gerar uma contribuição científica e social.

## 1.6 Objetivo

Este trabalho tem como objetivo implementar uma rede neural profunda, aperfeiçoando a topologia da rede convolucional já conhecida, *U-Net*, dentro do contexto de aprendizado supervisionado para a tarefa de segmentação semântica em imagens de RD. As regiões de interesse que se deseja segmentar em imagens de retina são: microaneurismas, hemorragias, exsudatos duros e exsudatos moles. A rede neural treinada deve ser capaz de aprender as regiões de interesse e, posteriormente, segmentar estas potenciais anomalias para auxiliar o diagnóstico médico. Por meio do banco de dados IDRiD (*Indian Diabetic Retinopathy Image Dataset*), foi avaliado o desempenho da rede neural com diferentes



parâmetros e com as combinações de seus canais cores (RGB) para cada estrutura de interesse. Também foram realizados treinamentos com as combinações (associações) das regiões de interesse, analisando comparativamente os seus resultados.

## **1.7 Organização do Texto**

O texto deste trabalho está organizado em seis seções, sendo esta primeira destinada à sua introdução. A Seção 2 é composta pelos conceitos teóricos utilizados nesta pesquisa, mais precisamente, os principais fundamentos acerca da RD e da aprendizagem profunda. A Seção 3, por sua vez, contempla os principais trabalhos presentes na literatura que também estudam o tema aqui proposto. Em seguida, na Seção 4 são evidenciados os métodos utilizados para a realização desta pesquisa. Na Seção 5, são apresentados os resultados obtidos ao se empregar os métodos propostos. Por fim, na Seção 6, apresenta-se a discussão e a sumarização dos resultados apresentados neste trabalho com o intento de destacar os aspectos e os resultados mais relevantes. Nesta seção também são apresentadas as direções futuras de pesquisa.



## 2 FUNDAMENTAÇÃO TEÓRICA

Esta seção é destinada à apresentação dos conceitos fundamentais para a compreensão do trabalho realizado. Dessa maneira, serão descritos os procedimentos presentes na literatura para o desenvolvimento de uma rede neural convolucional - *U-Net* - para o processamento de imagens. Com esta finalidade, esta seção foi estruturada por nove subseções.

A subseção 2.1 será composta pelos conceitos teóricos acerca da doença DM. Em seguida, a subseção 2.2 abordará os conceitos da manifestação da doença na retina, bem como os exames por imagem para identificação da mesma. Para realizar o processamento destas imagens, serão necessárias algumas etapas específicas, as quais serão descritas na subseção 2.3. Uma vez processadas, estas imagens poderão ser aplicadas em sistemas com inteligência artificial, conforme será abordado na subseção 2.4. Mais especificamente, serão utilizados modelos de *machine learning*, conforme será apresentado na subseção 2.5. Esta questão é recorrente do tema de aprendizado de representação, no qual será discutido na subseção 2.6. Em seguida, a subseção 2.7 discorrerá sobre redes neurais artificiais e seus conceitos básicos de funcionamento. Dentro da área de *machine learning*, este trabalho está inserido em RNA, conforme será abordado na subseção 2.7. A subseção 2.8 contextualizará o desenvolvimento das CNNs, suas arquiteturas, aplicações e métricas de avaliação utilizadas nesse trabalho. Por fim, a subseção 2.9 apresentará os limites e desafios da técnica proposta.

### 2.1 Diabetes Mellitus

O termo diabetes mellitus (DM) corresponde a um distúrbio metabólico de múltiplas etiologias caracterizadas por hiperglicemia crônica com distúrbios de carboidratos, gordura e metabolismo proteico resultantes de defeitos na secreção ou ação da insulina. Os efeitos do DM incluem danos a longo prazo, disfunção e insuficiência de vários órgãos. Sintomas característicos incluem sede, poliúria, perda de visão e perda de peso. Em suas formas mais graves, a cetoacidose ou um estado hiperosmolar não cetótico pode ser desenvolvido e levar a um estado de letargia, coma e, na ausência de tratamento efetivo, a morte. Frequentemente, os sintomas do DM não são graves ou até mesmo podem ser ausentes (assintomático). Esta característica assintomática pode estar presente por um longo período antes que o diagnóstico seja feito.

Os efeitos a longo prazo do DM incluem o desenvolvimento progressivo das complicações da retinopatia como: perda de visão, nefropatia que pode levar à insuficiência renal e/ou neuropatia com risco de úlceras nos pés, amputação e disfunções no sistema nervoso autônomo, incluindo disfunção sexual. Pessoas com diabetes possuem um risco aumentado

de desenvolver doenças cardiovasculares, vasculares periféricas e cérebro-vascular. Além disso, diversos processos patogênicos estão envolvidos no desenvolvimento do diabetes, incluindo a destruição de células *beta* do pâncreas com consequente deficiência de insulina e, também, outros que se convertem em resistência à ação da insulina (ALBERTI; ZIMMET, 1998).

## 2.2 Retinopatia Diabética

A RD é uma complicação microvascular comum e específica do diabetes e continua a ser a principal causa de cegueira evitável em pessoas com idade economicamente ativa. A perda completa da visão é identificada em um terço das pessoas com diabetes e está associada ao aumento do risco de complicações vasculares sistêmicas potencialmente fatais, as quais incluem acidente vascular cerebral, doença coronariana e insuficiência cardíaca (LEE; WONG; SABANAYAGAM, 2015).

No estágio inicial da RD, as alterações microvasculares são frequentemente encontradas na região temporal e área central da mácula (Figura 7), sendo causadas pela integridade vascular alterada e perda de pericitos. No estágio avançado da RD, podem ser desenvolvidas anomalias microvasculares intra-retinianas, as quais representam os capilares retiniais visíveis e indicam possível risco de neovascularização e retinopatia diabética proliferativa (NENTWICH; ULBIG, 2015).

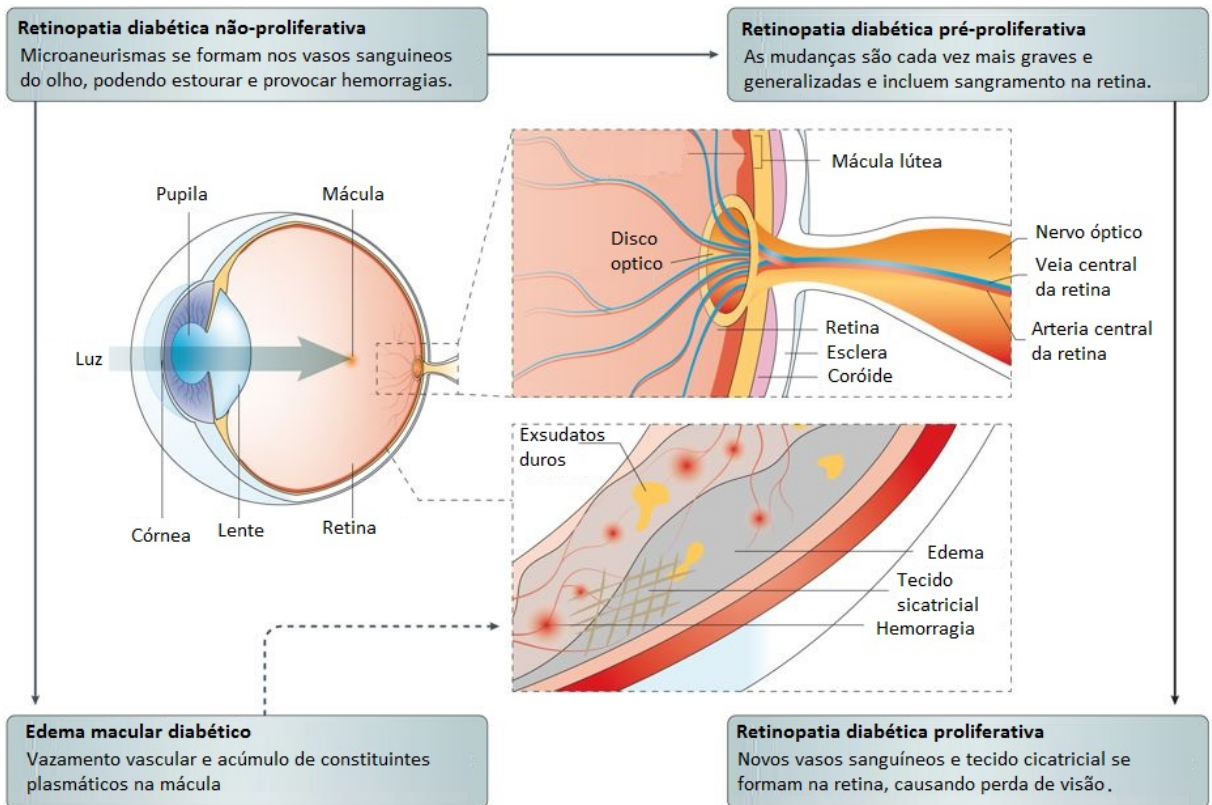
### 2.2.1 Estágios clínicos e principais eventos patogênicos da RD.

O primeiro estágio patológico identificado no exame de retina é a RD não proliferativa, que por sua vez é dividida em leve, moderada ou grave. O estágio em que há formação de edema macular diabético pode ser considerado um subtipo particular de RD não proliferativo, no qual o vazamento vascular envolve a mácula. A RD não proliferativa grave é caracterizada por fechamentos capilares e áreas não perfundidas.

O segundo estágio patológico é a RD proliferativa, na qual a hipóxia tem um papel essencial e é o estágio final da RD. Suas principais características são o desenvolvimento da neovascularização, o haemovitreo (sangramento na cavidade vítrea) e o descolamento de retina, sendo estas, portanto, complicações avançadas de RD proliferativa, como pode ser observado na Figura 7.

O diagnóstico precoce e, conseqüentemente, o tratamento médico da RD podem prevenir a perda da visão (ABRAMOFF; GARVIN; SONKA, 2010). Neste caso, os pacientes diabéticos são encaminhados para um acompanhamento anual ou bianual regular, além das consultas frequentes para o rastreamento de sua retina (JELINEK; CREE, 2009). Destaca-se que a eliminação da deficiência visual evitável depende não apenas do grupo de médicos especialistas como também das infraestruturas básicas de cuidados de saúde essenciais para o tratamento do olho (JELINEK; CREE, 2009; JONES; EDWARDS,

Figura 7 – Estágios clínicos e principais eventos patogênicos da retinopatia diabética.



Fonte: Imagem adaptada de Wong *et al.* (2016)

2010). Por exemplo, no subcontinente indiano, a proporção populacional de especialistas e pacientes é de 1:107.000, em várias regiões essa proporção é de 1:9000, enquanto em outras partes há apenas 1:608.000 habitantes (RAMAN *et al.*, 2016; PORWAL *et al.*, 2018b). Devido ao grande número de pessoas que necessitam de acompanhamento contínuo e à escassez de oftalmologistas, faz-se necessária a utilização de ferramentas de diagnóstico auxiliada por computador que permitam avaliar o histórico de evolução da doença, bem como obter um diagnóstico padronizado (TING; CHEUNG; WONG, 2016; WALTER *et al.*, 2002).

Os recentes avanços tecnológicos em computação, sistemas de comunicação e técnicas de aprendizado de máquina oferecem oportunidades aos engenheiros biomédicos e cientistas da computação para atender aos requisitos da prática clínica (SHORTLIFFE; BLOIS, 2006; PATTON *et al.*, 2006). As imagens processadas auxiliam a comunidade científica para o desenvolvimento, a validação, a comparação e a melhoria de algoritmos de detecção e segmentação de lesões de RD usados na aplicação clínica (TRUCCO *et al.*, 2013a). Anotações de nível de pixel precisas de anormalidades associadas à RD, como microaneurismas, exsudatos moles, exsudatos duros e hemorragias, são recursos valiosos para avaliação de desempenho de técnicas de segmentação de lesões individuais.

As informações confiáveis sobre o nível de gravidade da RD e edema macular diabético (DME), bem como a presença das referidas estruturas, são úteis no desenvolvimento, análise e detecção precoce da doença (PORWAL *et al.*, 2018a).

### 2.3 Etapas do processamento de imagens

O termo “processamento de imagem biomédica” se refere ao fornecimento de processamento de imagem digital para ciências biomédicas (DESERNO, 2011). Este processamento requer algumas etapas:

- Formação de imagem: inclui todas as etapas, desde a captura da imagem até a formação de uma matriz de imagem digital.
- Visualização de imagem: refere-se a todos os tipos de manipulação dessa matriz, resultando em uma saída otimizada da imagem.
- Análise de imagem: inclui etapas de análise, utilizando medidas quantitativas ou qualitativas em imagens biomédicas. Essa etapa exige um conhecimento prévio da natureza e do conteúdo das imagens, as quais devem ser integradas aos algoritmos em um alto nível de abstração. Assim, o processo de análise de imagens depende de algoritmos especializados.
- Gerenciamento de imagens: resume todas as técnicas que fornecem armazenamento, comunicação, transmissões, arquivamento e acesso (recuperação) de dados de imagem.
- Segmentação semântica: é a tarefa de classificar cada um dos pixels de uma imagem em classes (WANG *et al.*, 2018).

Em contraste com a análise de imagens, que muitas vezes também é o processamento de imagem de alto nível, o processamento de baixo nível denota técnicas, que podem ser realizadas sem um conhecimento prévio sobre o conteúdo específico das imagens (DESERNO, 2011).

O trabalho proposto abrange as etapas de visualização, análise e gerenciamento de imagens de alto e baixo nível, utilizando o banco de dados IDRiD com imagens avaliadas por especialistas.

### 2.4 Inteligência Artificial na Medicina

Inteligência Artificial (IA) é um termo geral que define o uso de um computador para modelar o comportamento inteligente com o mínimo de intervenção humana (ZHANG; LU, 2021). A IA, descrita como a ciência e engenharia de fabricação de máquinas inteligentes, nasceu oficialmente em 1956. O termo é aplicável a uma ampla gama de itens da medicina,

tais como: robótica, diagnóstico médico, estatística médica e biologia humana. A aplicação da IA na medicina possui dois ramos principais: físico e virtual (HAMET; TREMBLAY, 2017).

No ramo físico da medicina, se destacam os sofisticados dispositivos médicos e os robôs que participam da prestação de cuidados (*carebots*) aos seres humanos (CORNET, 2013). Um exemplo é o robô companheiro, utilizado para a população idosa com declínio cognitivo ou mobilidade limitada. Outros exemplos são os robôs utilizados em cirurgias, denominados como cirurgiões assistentes (LARSON; JOHNSON; BHAYANI, 2014).

No ramo virtual da medicina, ganham destaque os algoritmos matemáticos que, por meio de treinamento, aperfeiçoam seus resultados. Este processo é denominado aprendizado de máquina (*machine learning*) e será tratado na subseção 2.5.

## 2.5 Aprendizado de Máquina

O Aprendizado de Máquina (*Machine Learning*) é um subcampo da inteligência artificial que utiliza algoritmos para análise de dados e reconhecimento de padrões. Em sua maioria, estes sistemas computacionais podem aprender com seus erros e fazer previsões. Para isso, são necessários treinamentos que permitam a identificação destes padrões e a busca pela redução da função de custos (MAHESH, 2020).

Existem três métodos de treinamentos em aprendizado de máquina: (i) não supervisionado (capacidade de encontrar padrões), (ii) supervisionado (algoritmos de classificação e previsão baseados em exemplos anteriores) e (iii) aprendizado por reforço (uso de sequências de recompensas e punições para formar uma estratégia para operação em um espaço de problema específico) (DASGUPTA; NATH, 2016).

Os métodos supervisionados são aqueles que possuem uma saída bem definida, sendo, portanto, possível medir diretamente seu resultado a cada treinamento. Neste caso, o algoritmo aprende as regras que mapeiam entradas em saídas. Para aplicações de segmentação de imagens médicas, o treinamento supervisionado equivale a encontrar parâmetros que melhor prevejam as regiões desejadas com base em uma função de custo. Dessa maneira, essa função de custo permite atualizar os pesos de cada camada, convergindo a entrada na saída desejada (BELTHANGADY; ROYER, 2019).

No aprendizado não supervisionado, não há saída rotulada e, portanto, o algoritmo deve ser capaz de classificar os dados de acordo com possíveis padrões que possam existir. Neste caso, o conjunto de treinamento possui apenas as entradas, não existindo uma referência de modelo a seguir ou aprender. Por conseguinte, muitas vezes os resultados precisarão ser analisados por especialistas ou por outros métodos a fim de que seja identificada se a convergência do treinamento resultou em informações válidas. A respeito das imagens médicas, pode-se utilizar esse método para encontrar estruturas ou padrões

sem conhecimento prévio. Um exemplo de sucesso na medicina são os algoritmos de interação proteína-proteína não supervisionados que levaram à inéditas descobertas de novos alvos terapêuticos (THEOFILATOS *et al.*, 2015).

No aprendizado por reforço, o algoritmo interage dinamicamente com o ambiente com objetivos específicos. Neste caso, o algoritmo ajusta os pesos por meio de repetidas tentativas e erros com objetivo de reduzir sua função de custo. (THEOFILATOS *et al.*, 2015).

Neste estudo aqui desenvolvido, utiliza-se redes neurais convolucionais, sub-área de aprendizado de representação, treinadas de maneira supervisionada com o propósito de identificar e segmentar pontos de interesse em imagens de retina. Ressalta-se que esses pontos de interesse foram previamente classificados por médicos experientes, os quais indicaram qual é a saída esperada. Assim, após o treinamento, espera-se que o sistema seja capaz de prever e segmentar os pontos de interesse para avaliação da RD.

## 2.6 Aprendizado de Representação

O aprendizado de máquina pode ser aplicado em diversas áreas devido à sua capacidade de encontrar as melhores representações para as suas entradas durante o treinamento. Entre as possíveis aplicações, destacam-se o uso em segmentação de dados, reconhecimento de padrões, inferência de ações, classificação e tomadas de decisão, dentre outros. (GOODFELLOW; BENGIO; COURVILLE, 2016).

Os métodos de aprendizado profundo são métodos de representação/aprendizado com múltiplos níveis, obtidos pela composição de módulos não lineares, que transformam a entrada primária em uma representação em um nível mais abstrato. Com a composição de tais transformações, funções muito complexas podem ser aprendidas.

Para tarefas de classificação, camadas mais altas de representação amplificam aspectos da entrada que são importantes para a discriminação e suprimem variações irrelevantes. Uma imagem, por exemplo, vem na forma de uma matriz de valores de pixel e as características aprendidas na primeira camada são tipicamente as presenças ou as ausências de bordas. A segunda camada, por sua vez, geralmente detecta arranjos particulares de bordas, independentemente de pequenas variações nas posições. A terceira camada pode agrupar combinações maiores que correspondem a partes de objetos familiares, enquanto que as camadas subsequentes detectam objetos com combinações dessas partes. Dessa maneira, o aspecto principal do aprendizado profundo são as camadas de características não projetadas por humanos, mas identificadas a partir da imagem original utilizando ajustes de pesos sinápticos, no processo de treinamento, identificando padrões para diversos fins (BENGIO; COURVILLE; VINCENT, 2013).

Alguns trabalhos utilizam algoritmos que se assemelham aos sistemas de similari-



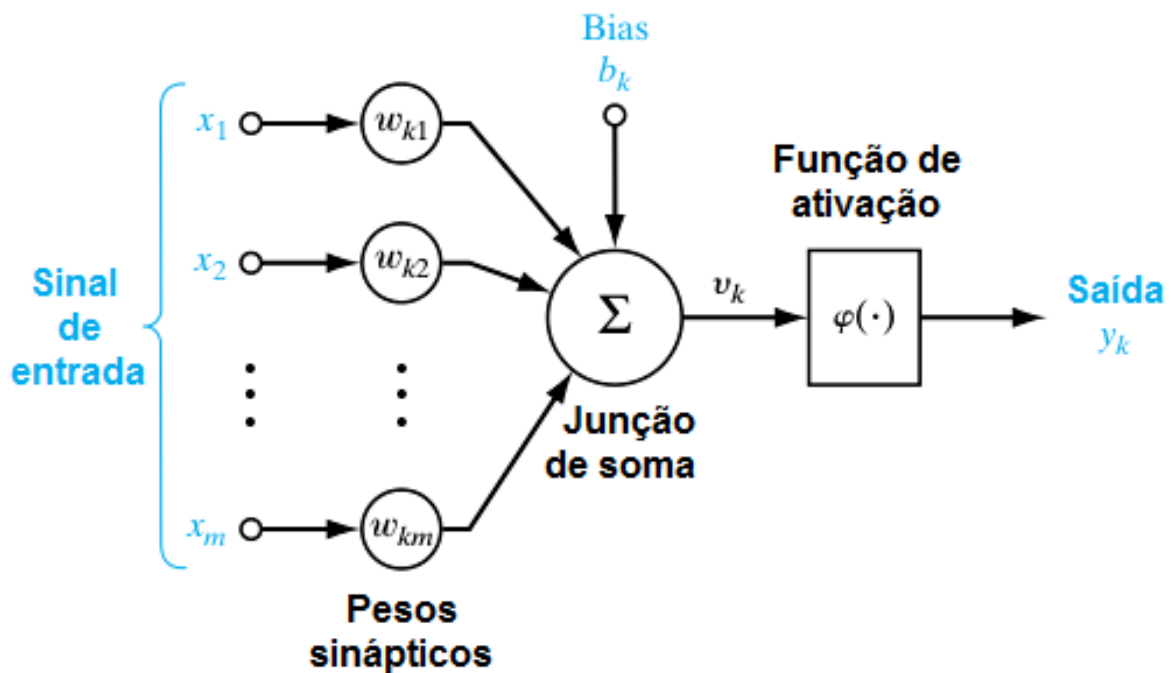
dade encontrados no contexto biológico, mais especificamente, apresentam neurônios com funções semelhantes às dos neurônios do córtex visual. (LEE; EKANADHAM; NG, 2008; LEE *et al.*, 2009).

## 2.7 Redes Neurais Artificiais

As RNAs são técnicas computacionais inteligentes inspiradas nos neurônios biológicos. Uma RNA é composta por várias unidades de processamento (nós ou neurônios) que possuem operação relativamente simples. Essas unidades, geralmente, são conectadas por canais de comunicação que possuem um peso associado (SILVA *et al.*, 2016).

As entradas dos neurônios são ponderadas por pesos e associadas de acordo com a função de ativação. Essa função é responsável por combinar linearmente os valores de entrada, ponderados pelos pesos e avaliar se sua saída será ativada em função do valor de referência. A representação de um neurônio em RNA poderá ser observada na Figura 8.

Figura 8 – Sistema representativo de um neurônio em RNA.



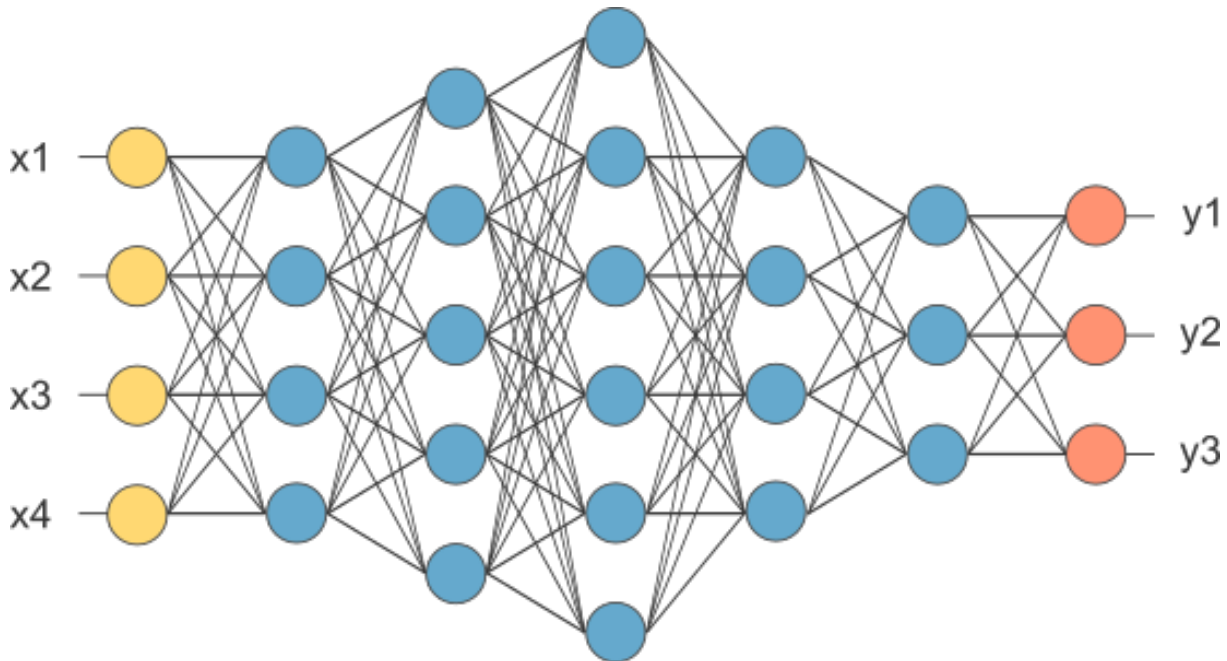
Fonte: Imagem adaptada de Krogh (2008)

Os neurônios em RNA, podem ser repetidos e combinados, formando diferentes camadas, em que cada camada tem a mesma função de ativação. Estas associações de neurônios formam redes neurais com funções que permitem o ajuste dos pesos sinápticos durante o processo de treinamento (SILVA *et al.*, 2016).

No aprendizado supervisionado, a rede utiliza como aprendizado um conjunto de entradas com suas respectivas saídas ideais conhecidas para ajustar os parâmetros. Já o

aprendizado não supervisionado agrupa as saídas em um número de grupos determinado pela quantidade de neurônios na camada de saída. A Figura 9 apresenta um exemplo de rede neural com cinco camadas ocultas (KROGH, 2008).

Figura 9 – Sistema representativo de uma rede neural artificial.



Fonte: Imagem adaptada de Krogh (2008)

Diferentes funções de ativação podem ser empregadas, desde que sejam monotonicamente crescentes, contínuas e diferenciáveis. Estes requisitos são impostos pelo método de descida de gradiente, para ajuste dos pesos sinápticos dos neurônios das camadas escondidas (FELDMAN; ROJAS, 2013). As funções mais utilizadas são: linear (2.1), limiar (2.2), sigmoide (2.5), tangente hiperbólica (2.4) e *softmax* (2.5) (BISHOP, 2006), apresentadas a seguir.

**Linear:**

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

**Limiar:**

$$f(x) = \begin{cases} 0, & \text{se } x \geq 0,5 \\ 1, & \text{c.c.} \end{cases} \quad (2.2)$$

**Sigmoide:**

$$f_c(x) = \frac{1}{1 + e^{-cx}}, \quad f : \mathbb{R} \mapsto (0, 1) \quad (2.3)$$

**Tangente hiperbólica:**

$$f(x) = \tanh(x), \quad f : \mathbb{R} \mapsto (-1, 1) \quad (2.4)$$

***Softmax:***

$$\hat{P}_{Y|X}(j|x) = y_j(x) = \frac{e^{Z_j}}{\sum_j e^{Z_j}}, \quad Z_j = w_j^T \mathbf{X} + b_j \quad (2.5)$$

As principais vantagens das RNAs incluem: não ser necessário conhecimento especializado para tomada de decisões, pois são baseadas em dados históricos; Não sofrem degradação, mesmo que os dados sejam inconsistentes, garantindo robustez; ao construir uma RNA precisa para uma aplicação específica, ela pode ser empregada em tempo real sem ter que mudar sua arquitetura a cada atualização. Entretanto, algumas desvantagens das RNAs são: as redes podem convergir para soluções inconsistentes com regras e teorias estabelecidas; treinar uma RNA pode ser computacionalmente custoso e demorado; são métodos que não garantem a obtenção do menor valor possível da função de custo (mínimo global) para o sistema proposto; é necessário um grande conjunto de dados para que uma RNA possa ser corretamente treinada (SILVA *et al.*, 2016).

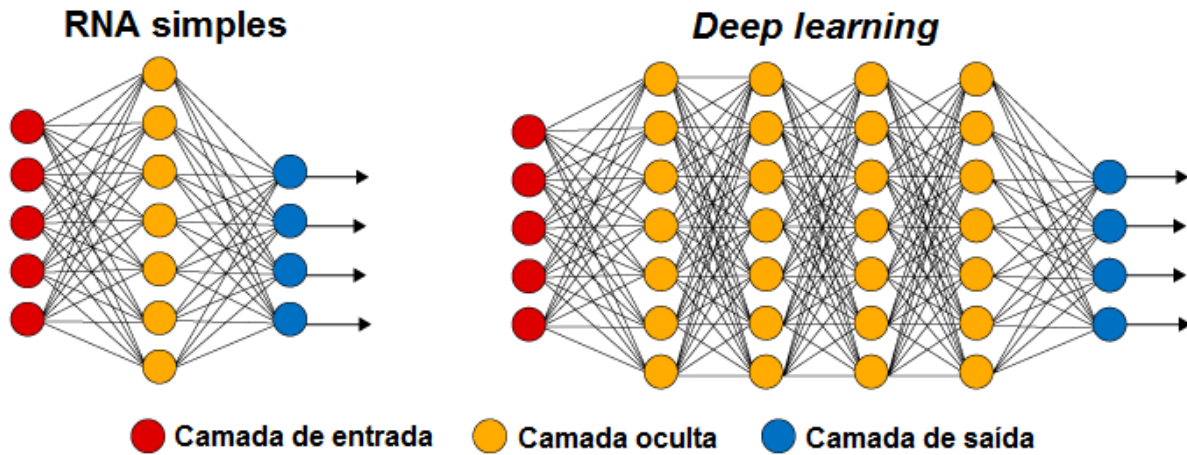
A forma mais simples de uma rede neural é a Rede Neural Artificial (RNA), que consiste em 3 camadas de neurônios: uma camada de entrada, uma camada oculta e uma camada final de saída. Essas arquiteturas são conhecidas como Redes Neurais Rasas (*Feed-Forward*), porque possuem apenas uma camada oculta. Em contraste, uma Rede Neural Profunda (*Feed-Forward*) (DNN) consiste em mais de duas camadas ocultas e serão descritas na subseção 2.8. Estas duas formas possuem como entrada uma matriz unidimensional e, portanto, não podem ser usadas diretamente com dados de imagem. (TSIKNAKIS *et al.*, 2021a)

## 2.8 *Deep learning*

O aprendizado profundo (*Deep learning*) é um subconjunto do aprendizado de máquina e é inspirado nos padrões de processamento de informações encontrados no cérebro humano. O *Deep learning* não requer nenhuma regra projetada por humanos para operar. Em vez disso, usa uma grande quantidade de dados para mapear a entrada fornecida para rótulos específicos. Seu algoritmo é projetado usando várias camadas de neurônios, sendo que, cada uma fornece uma interpretação diferente dos dados que chegaram a elas (LECUN; BENGIO; HINTON, 2015; ZHANG; CUI; ZHU, 2020). A Figura 10 apresenta a diferença estrutural entre uma rede neural simples e uma *deep learning*, sendo possível evidenciar o grande número de camadas ocultas em *Deep learning*.

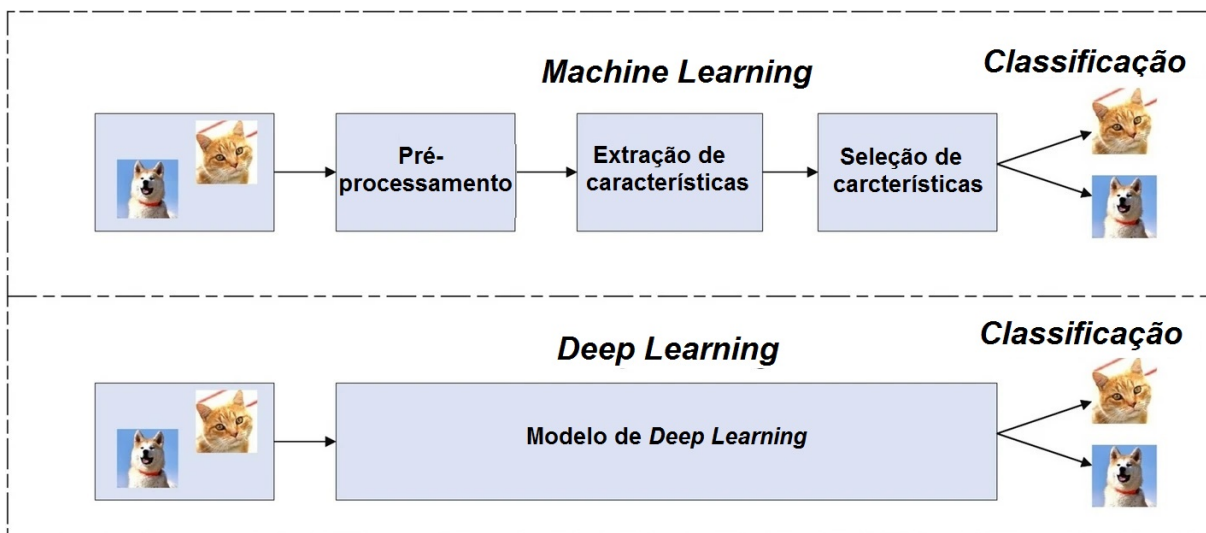
Técnicas tradicionais de aprendizado de máquina requerem etapas sequenciais, como extração de recursos, seleção inteligente, aprendizado e classificação. Como vantagens, *deep learning* tem a capacidade de automatizar o aprendizado de conjuntos de recursos para várias tarefas, permitindo que o aprendizado e a classificação sejam alcançados diretamente. A Figura 11 exemplifica conceitualmente a diferença entre *machine learning* e *deep learning*.

Figura 10 – Diferença entre RNA e *Deep learning*.



Fonte: Imagem adaptada de Vázquez (2017).

Figura 11 – Diferença entre aprendizado profundo e aprendizado de máquina tradicional.



Fonte: Imagem adaptada de Alzubaidi *et al.* (2021).

O sucesso do *deep learning* decorre, dentre outros fatores:

- **Robustez:** Recursos otimizados podem ser aprendidos de forma automatizada em relação à tarefa desejada. Dessa forma, variações usuais nos dados de entrada não são degenerados na saída.
- **Generalização:** Diferentes tipos de dados ou diferentes aplicativos podem usar a mesma técnica de *deep learning*.
- **Escalabilidade:** Capacidade de se adaptar em diferentes cenários e variações de forma a manter a saída correta.

- Abordagem de aprendizado universal: Capacidade de funcionar em, aproximadamente, todos os domínios de aplicação (ALZUBAIDI *et al.*, 2021).

### 2.8.1 Aplicações

Com relação às aplicações possíveis de *deep learning*, vale destacar que esta técnica possui as mesmas características presentes em arquiteturas clássicas de redes neurais artificiais (SILVA *et al.*, 2016), as quais permitem que os modelos atuem como/em:

- Aproximador de Funções;
- Controle de Processos;
- Classificação de Padrões;
- Agrupamento de Dados;
- Sistemas de Previsão;
- Otimização de Sistemas;
- Memórias Associativas.

### 2.8.2 Redes Neurais Convolucionais

Na área de *deep learning*, a Rede Neural Convolucional (CNN) é o algoritmo mais famoso e comumente empregado (KRIZHEVSKY; SUTSKEVER; HINTON, 2017; ZHOU, 2020; AL-AZZAWI *et al.*, 2020; WANG *et al.*, 2020; LI *et al.*, 2021). O principal benefício da CNN em comparação com seus antecessores é que ela identifica automaticamente os recursos relevantes sem qualquer supervisão humana (GU *et al.*, 2018). As CNNs têm sido amplamente aplicadas em diversos campos, incluindo visão computacional (FANG *et al.*, 2020), processamento de fala (PALAZ; MAGIMAI-DOSS; COLLOBERT, 2019), reconhecimento facial (LI *et al.*, 2020), entre outros.

As Redes Neurais Convolucionais foram inspiradas na visão humana e seu conceito é baseado em uma operação matemática fundamental, chamada de “convolução” (MENDES, 2017). A convolução é uma operação de somatório do produto entre duas funções, ao longo da região em que elas se sobrepõem, em razão do deslocamento existente entre elas, conforme equação 2.6.

$$(f * g)(x, y) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} f(i, j)g(x - i, y - j) \quad (2.6)$$

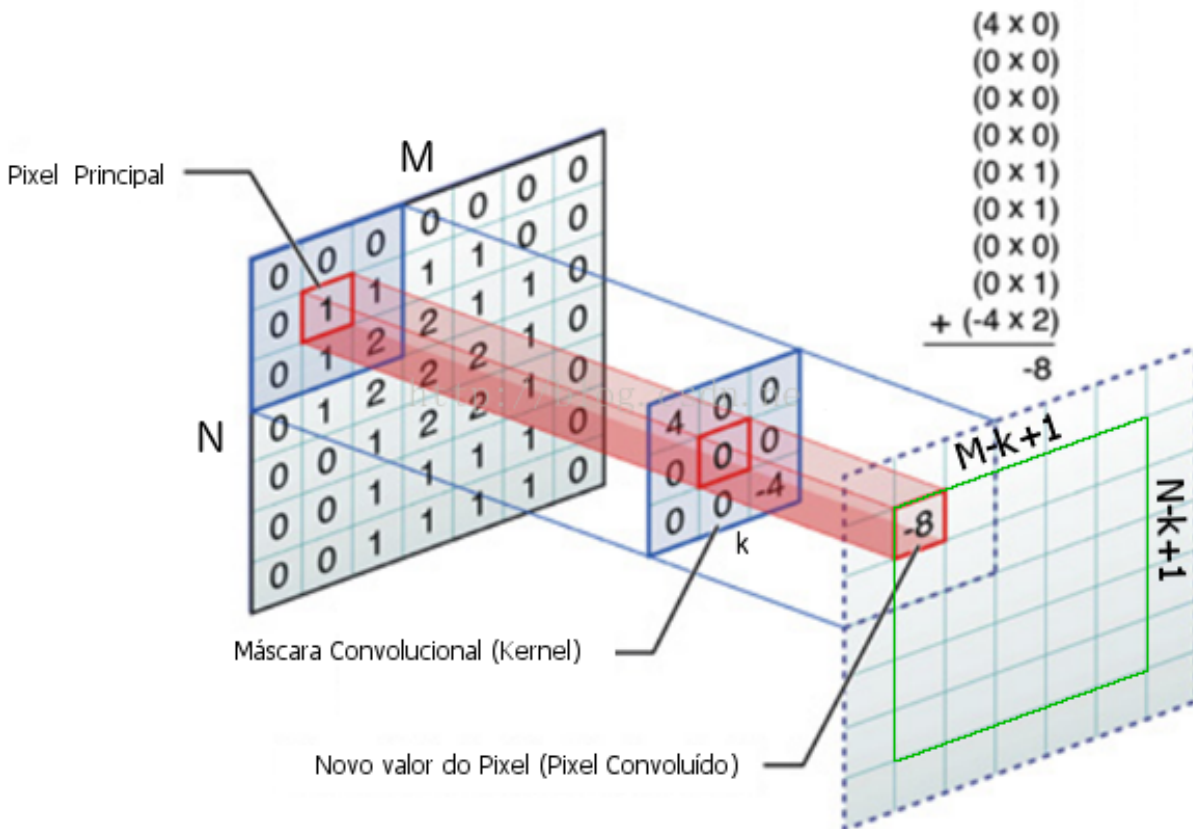
As cinco principais camadas de uma rede neural convolucional incluem: camada convolucional, de agrupamento (*Pooling*), totalmente conectada (*fully connected layer*),

*Dropout*. Além dessas camadas é necessário o uso de um funções de ativação (VERNWAL, 2017). A seguir apresenta-se as funções das diferentes camadas de uma rede.

### 2.8.2.1 Camada Convolutiva

A camada convolutiva é a primeira camada usada para extrair os vários recursos das imagens de entrada. Nesta camada, a operação matemática de convolução é realizada entre a imagem de entrada e um filtro de um determinado tamanho  $k \times k$ . Ao deslizar o filtro sobre a imagem de entrada, o produto escalar é obtido entre o filtro e as partes da imagem de entrada em relação ao tamanho do filtro (NOH; HONG; HAN, 2015), conforme pode ser observado na Figura 12.

Figura 12 – Ilustração de um Filtro Convolutiva.



Fonte: Imagem adaptada de Apple (2016).

A saída é denominada como mapa de recursos. Este mapa fornece informações 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).

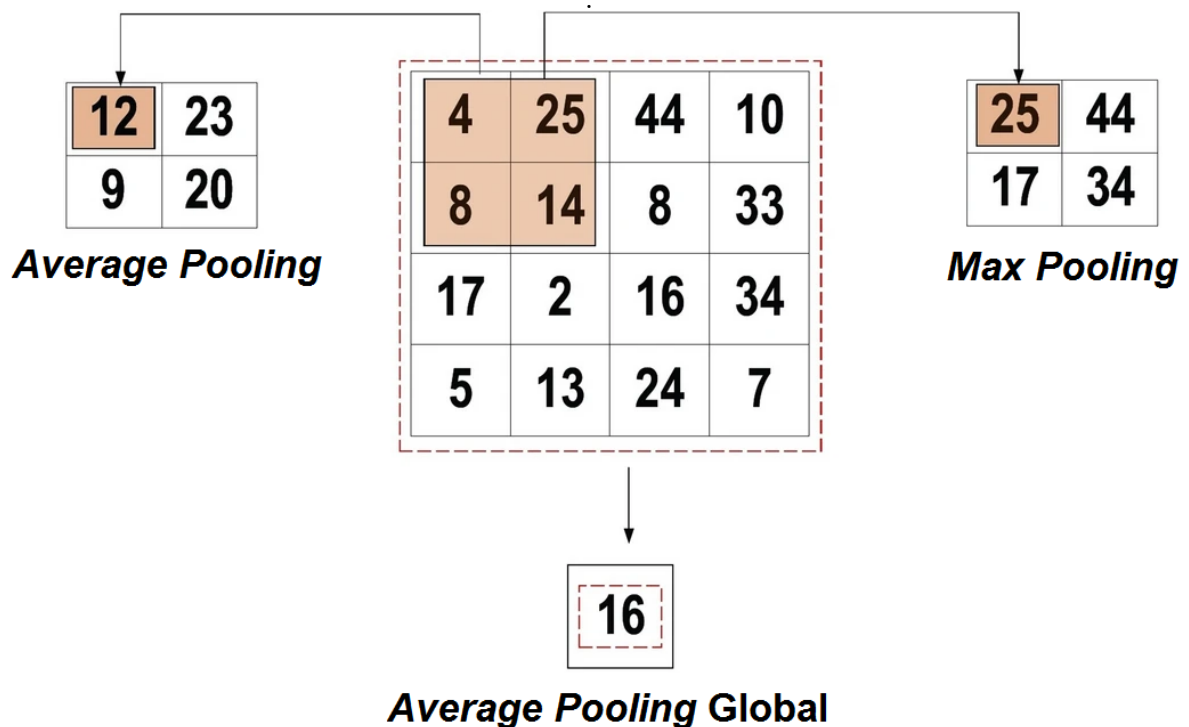
### 2.8.2.2 Camadas de Agrupamento (*Pooling*)

Na maioria dos casos, uma camada convolutiva é seguida por uma camada de *Pooling*. O principal objetivo desta camada é diminuir o tamanho do *feature map* sem

alterar as características mais importantes. Isso é realizado diminuindo as conexões entre as camadas de forma independente em cada mapa de recursos. Dependendo do método utilizado, existem vários tipos de operações de *pooling* (SIMONYAN; ZISSERMAN, 2015).

No *Max Pooling*, o maior elemento é retirado do mapa de recursos. Já no *Average Pooling* é calculada a média dos elementos em uma seção de imagem de tamanho predefinido. No *Global Average Pooling* é calculada a média de todos os elementos na seção predefinida, no qual se diferencia do *Average Pooling* por resultar em um elemento por posição do kernel. Na Figura 13, poderão ser observados os três tipos de operação *pooling* citados. A camada de *pooling*, usualmente, funciona como interface entre a camada convolucional e a camada totalmente conectada (*Fully Connected layer*).

Figura 13 – Três diferentes tipos de operação *pooling*



Fonte: Imagem adaptada de Alzubaidi *et al.* (2021).

### 2.8.2.3 Camada totalmente conectada

A camada totalmente conectada (FC) são responsáveis por conectar todas as entradas de uma camada em cada unidade de ativação da camada seguinte. Em geral, também, são utilizadas antes da camada de saída, formando as últimas camadas de uma arquitetura CNN. Nesta etapa, o processo de classificação começa a ocorrer (SIMONYAN; ZISSERMAN, 2015).

#### 2.8.2.4 Dropout

Quando as camadas estão conectados à camada FC, significa que uma mesma entrada está conectada em todos os neurônios da camada seguinte, como consequência podem ocorrer *overfitting* no conjunto de dados de treinamento. O *overfitting* ocorre quando um modelo se especializa nos dados de treinamento e passa a apresentar erros em outras imagens de entrada, ou seja, a rede perde sua capacidade de generalização causando um impacto negativo no conjunto de aplicações.

Para superar esse problema, uma camada de *dropout* é utilizada. Nesta camada alguns neurônios são descartados da rede neural durante o processo de treinamento, resultando em tamanho reduzido do modelo. Este processo melhora a generalização por permitir que o mesmo conceito seja aprendido com diferentes neurônios. Um *dropout* de 0,3 implica no descarte aleatório de 30% dos nós da rede neural.

#### 2.8.2.5 Rectified Linear Unit (ReLU)

Um dos parâmetros mais importantes do modelo CNN é a função de ativação (DUMOULIN; VISIN, 2016). Essas funções são usados para aprender e aproximar qualquer tipo de relação contínua e complexa entre variáveis da rede, em outras palavras, decide quais informações do modelo devem se propagar na rede e quais devem ser retidas.

Existem várias funções de ativação comumente usadas, como as funções *ReLU*, *Softmax*, *tanH* e *Sigmoid*. Cada uma dessas funções tem um uso específico. Para um modelo CNN de classificação binária, as funções *sigmoid* e *softmax* ganharam destaque, porém não mitigam a dificuldade apresentada quando os pesos sinápticos de uma rede neural ficam demasiadamente pequenos. Este problema é conhecido como desaparecimento do gradiente (*Vanishing gradient problem*). Assim, a função de ativação que mais ganhou destaque foi a função *ReLU*, expressa pela equação 2.7, proposta por Hahnloser *et al.* (2000), provando superior às demais por apresentar efeito de esparsidade na topologia da rede que permitiram atingir os menores pesos sinápticos (GLOROT; BORDES; BENGIO, 2011).

$$f(x) = x^+ = \max(0, x) \tag{2.7}$$

#### 2.8.2.6 Topologias de Redes Convolucionais Conhecidas

Algumas topologias de redes neurais ganharam grande destaque histórico e, por sua representatividade, são modelos amplamente citados e reconhecidos por seus nomes. A seguir, alguns deles serão apresentados e contextualizados.

**LeNet.** A *LeNet* é uma estrutura de rede neural convolucional proposta por LeCun *et al.* (1998) em 1989. Neste trabalho, foi utilizado uma arquitetura de rede neural



---

convolucional para reconhecer caracteres manuscritos e impressos por computador. Cabe destacar que esta rede ganhou grande destaque pelos resultados e por sua arquitetura simples e direta.

**AlexNet.** A arquitetura *AlexNet* alcançou resultados inovadores nas áreas de reconhecimento e classificação de imagens (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). A *AlexNet* competiu no *ImageNet Large Scale Visual Recognition Challenge* em 30 de setembro de 2012. Esta rede reduziu o erro top-5 de 26.1% para 15.3%. O trabalho proposto por Krizhevsky, Sutskever e Hinton (2012) foi responsável por aumentar a popularidade das CNNs em aplicações de visão computacional, mas destacou que a profundidade do modelo era essencial para seu alto desempenho, tendo como desvantagem o alto custo computacional. Apesar disso, a arquitetura era viabilizada devido à utilização de unidades de processamento gráfico (GPUs) durante o treinamento (GOODFELLOW; BENGIO; COURVILLE, 2016).

**ZF Net.** Eleita a técnica campeã da *ILSVRC 2013 (ImageNet Large Scale Visual Recognition Challenge)*. A *ILSVRC* é uma competição de referência mundial onde são apresentados, testados e comparados os melhores algoritmos e sistemas de processamento de imagens (YANG *et al.*, 2021). A topologia *ZF Net* foi desenvolvida por Zeiler e Fergus (2014). Esta rede se constitui em um aprimoramento da *AlexNet*, uma vez que os competidores ajustaram os hiperparâmetros da arquitetura. Mais especificamente, foi aumentado o tamanho das camadas convolucionais intermediárias e diminuído o tamanho e o *stride* dos filtros aplicados às primeiras camadas.

**VGGNet.** Depois que a CNN foi reconhecida pela sua eficácia no campo de reconhecimento de imagem, um princípio de design fácil e eficiente para a CNN foi proposto por Simonyan e Zisserman (2015). Esse design inovador foi chamado de *Visual Geometry Group (VGG)*. Um modelo multicamadas, com dezenove camadas a mais que *ZefNet* e *AlexNet*, para simular as relações da capacidade representacional da rede em profundidade.

**GoogLeNet/Inception.** Foi a arquitetura vencedora da *ILSVRC 2014*, proposta por Szegedy *et al.* (2015) do Google. Esta topologia utilizou o módulo *inception*, e de forma inovadora, aplicou camada de *average pooling*, permitindo a redução considerável dos número de parâmetros livres.

**MobileNets.** Apresenta uma classe de modelos eficientes para aplicativos de visão móvel e integrados. Os *MobileNets* se baseiam em uma arquitetura simplificada que utiliza convoluções separáveis em profundidade para construir redes neurais profundas e leves. Todas as camadas são seguidas por uma não-linearidade do tipo *ReLU*, com exceção da camada final que é totalmente conectada e linear (HOWARD *et al.*, 2017).

**ResNet.** O trabalho proposto por He *et al.* (2016) foi a arquitetura vencedora da

*ILSVRC 2015*. Esta rede utiliza conexões de salto ou atalhos para pular algumas camadas. Os modelos típicos do *ResNet* são implementados com saltos de camada dupla ou tripla que contêm não linearidades (*ReLU*) e *Batch Normalization*. Estes saltos evitam o problema de desaparecimento do gradiente e mitigam o problema de degradação, no qual adicionar mais camadas a um modelo profundo leva a um maior erro de treinamento.

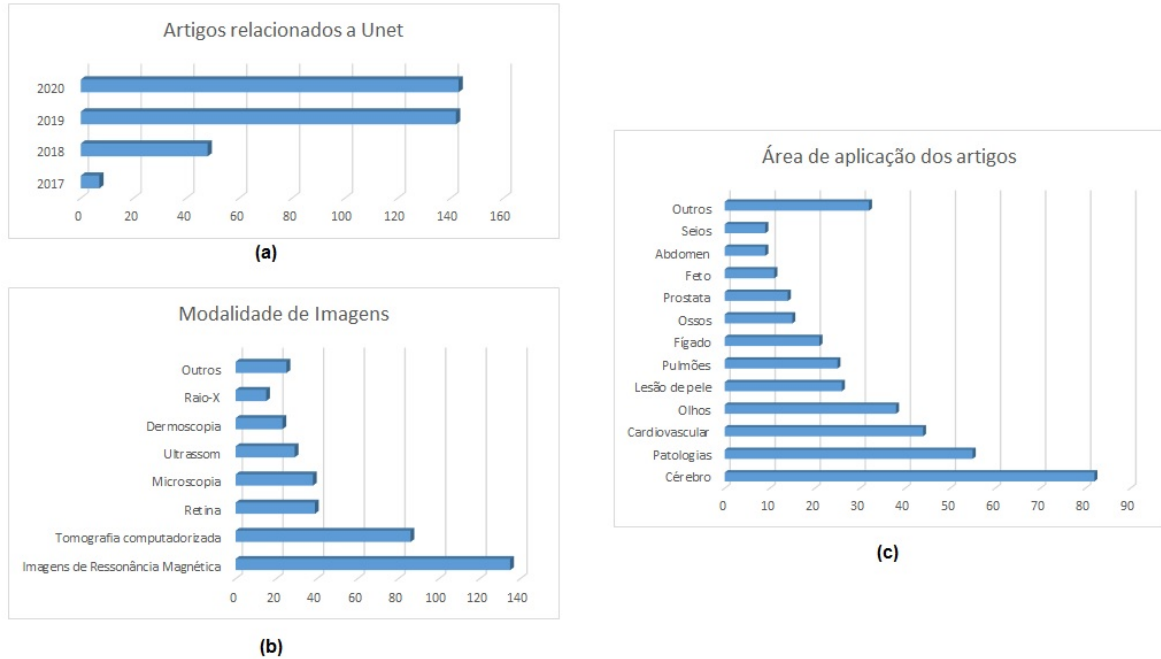
**U-Net.** A *U-Net* é uma arquitetura de rede neural projetada principalmente para segmentação de imagens (RONNEBERGER; FISCHER; BROX, 2015). A estrutura básica de uma arquitetura *U-Net* consiste em dois caminhos. O primeiro é o caminho de contração, também conhecido como codificador ou caminho de análise, o qual é semelhante a uma rede de convolução regular e fornece informações de classificação. O segundo é um caminho de expansão, também conhecido como decodificador ou caminho de síntese, constituindo-se por convoluções ascendentes e concatenações com recursos do caminho de contração. Essa expansão permite que a rede aprenda informações de classificação localizadas. Além disso, este segundo caminho também aumenta a resolução da saída, podendo então passar para uma camada convolucional final para criar uma imagem totalmente segmentada. A rede resultante é uma arquitetura quase simétrica, dando-lhe uma forma de “U”.

A principal tarefa canônica realizada pela maioria das redes convolucionais é a de classificar toda a imagem em um único rótulo. No entanto, as redes de classificação não fornecem informações contextuais em nível de pixel, o que é de vital importância na análise de imagens médicas. Embora tenha havido tentativas anteriores de tarefas de segmentação, somente com a implementação proposta por Ronneberger, Fischer e Brox (2015) que ocorreu uma melhora significativa no desempenho da segmentação de imagens médicas. É válido ressaltar que a rede *U-Net* foi desenvolvida baseada nos trabalhos de Long, Shelhamer e Darrell (2015), utilizando redes totalmente convolucionais. Sua implementação alcançou melhor desempenho histórico no desafio *ISBI 2012 e 2015 (International Symposium on Biomedical Imaging)*, superando o estado da arte na época por uma margem considerável.

O que torna a *U-Net* particularmente útil é a criação de mapas de segmentação altamente detalhados, utilizando amostras limitadas. Isso é obtido por meio da deformação elástica aleatória nos dados de treinamento, o que permite que a rede aprenda essas variações sem exigir novos dados rotulados (RONNEBERGER; FISCHER; BROX, 2015). Esta característica é de grande importância na comunidade de imagens médicas, devido a baixa disponibilidade de imagens rotuladas para treinamento. Outra vantagem desta topologia é separar objetos tocantes da mesma classe, o que é resolvido aplicando uma função de perda ponderada que penaliza o modelo caso não consiga separar os dois objetos. Na etapa de treinamento, a *U-Net* também ganha destaque por ser muito mais rápida quando comparada com a maioria dos outros modelos de segmentação, devido ao seu aprendizado baseado em contexto.

Desde a sua criação em 2015, a *U-Net* ganhou grande repercussão no uso em imagens médicas. E, naturalmente, houve muitos avanços nesta arquitetura por pesquisadores que implementaram novos métodos ou incorporam outros métodos de imagem na *U-Net*.

Figura 14 – Publicações de artigos relacionados a *U-Net*



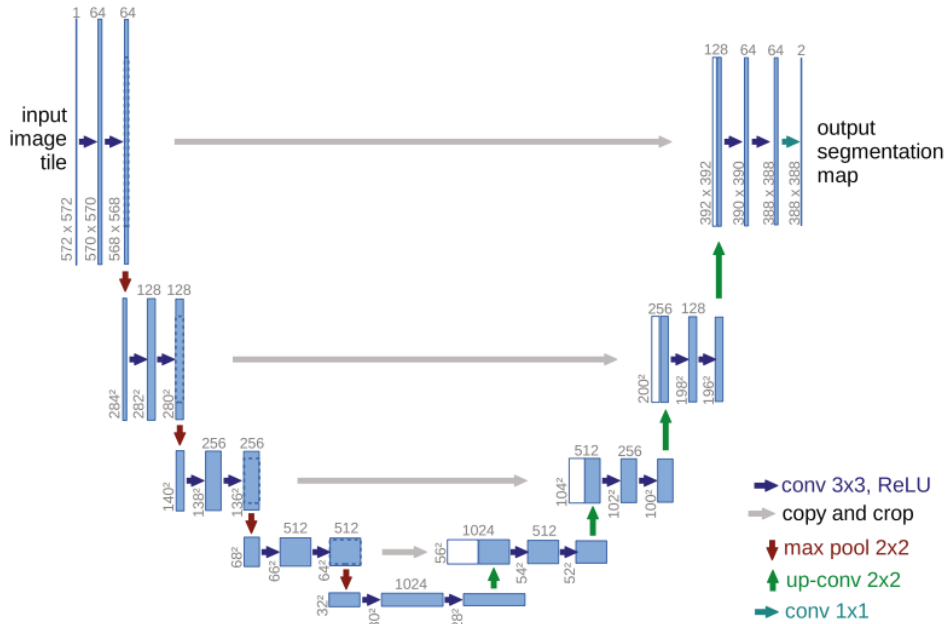
Distribuição de (a) artigos relacionados a *U-Net* por ano de publicação a partir de 2017, (b) modalidade de imagem em artigos relacionados a *U-Net* e (c) área de aplicação em artigos relacionados a *U-Net*. Fonte: Imagem adaptada de Siddique *et al.* (2021)

A arquitetura de rede é ilustrada na Figura 15, a qual consiste em um caminho de contração (lado esquerdo) e um caminho expansivo (lado direito). Pode-se afirmar que o caminho de contração segue a arquitetura típica de uma rede convolucional e consiste na aplicação repetida de duas convoluções  $3 \times 3$ , cada uma seguida por uma unidade de linearização (*ReLU*) e uma operação de *max pooling*  $2 \times 2$  com passo 2 para *downsampling*. Em cada etapa de redução da amostra, dobra-se o número de canais de recursos. Cada passo no caminho expansivo consiste em um *upsampling* do mapa de recursos seguido por uma convolução  $2 \times 2$ , reduzindo pela metade o número de canais de recursos. Vale ressaltar que o recorte é necessário devido à perda de pixels de borda em cada convolução. Na fase final, mais uma convolução  $1 \times 1$  é aplicada para reduzir o mapa de características ao número necessário de canais e produzir a imagem segmentada.

A função de energia para rede é dada por:

$$E = \sum w(x) \log(p_{k(x)}(x)) \quad (1)$$

Sendo  $p_k$  a função *SoftMax* aplicada sobre o mapa de recursos final definido como:

Figura 15 – Diagrama de blocos *U-Net*

Arquitetura *U-Net* (exemplo para 32x32 pixels na resolução mais baixa). Cada caixa azul corresponde a um mapa de recursos multicanal. O número de canais é indicado na parte superior da caixa. O tamanho x-y é fornecido na borda inferior esquerda da caixa. As caixas brancas representam mapas de recursos copiados. As setas indicam as diferentes operações. (RONNEBERGER; FISCHER; BROX, 2015)

$$p_k = \exp(a_k(x)) / \sum_{k'=1}^K \exp(a_{k'}(x)) \quad (2)$$

e  $a_k$  denota a ativação no canal  $k$ .

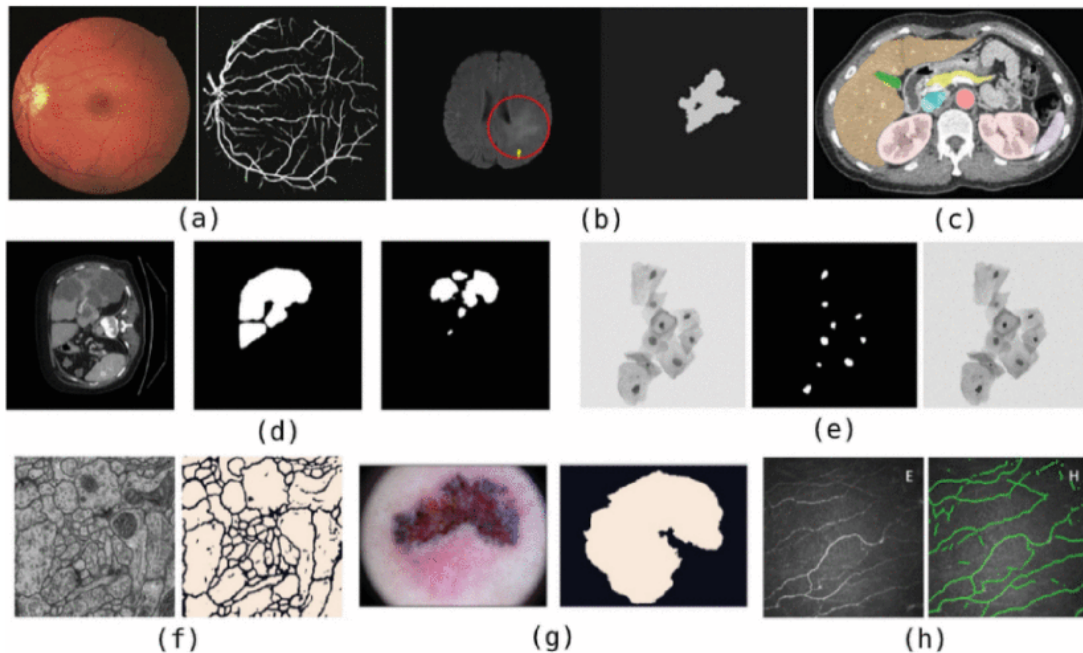
A segmentação é a principal tarefa dos modelos *U-Net* (RONNEBERGER; FISCHER; BROX, 2015; SIDDIQUE *et al.*, 2021). O objetivo das tarefas de segmentação é identificar e destacar objetos diferentes ao invés de classificar toda a imagem. Isso é de particular importância na comunidade de imagens médicas, pois o diagnóstico das condições requer uma análise cuidadosa das regiões específicas em uma imagem. Por exemplo, o diagnóstico de tumores cerebrais requer a separação dos tumores do resto das estruturas cerebrais (SIDDIQUE *et al.*, 2021).

A Figura 16 ilustra algumas aplicações da *U-Net* em diferentes áreas.

### 2.8.2.7 Processo de *data augmentation*

A capacidade de uma CNN classificar objetos de forma robusta, mesmo que seja colocada em diferentes orientações, é a invariância. Mais especificamente, uma CNN pode ser invariante à rotação, espelhamento, tamanho ou iluminação (ou ainda combinações

Figura 16 – Aplicações da *U-Net* em diferentes imagens médicas



Exemplos de aplicações *U-Net*: (a) segmentação de vasos da retina; (b) Detecção e segmentação de tumor cerebral; (c) Segmentação abdominal de múltiplos órgãos (fígado; baço; rins esquerdo e direito; pâncreas; vesícula biliar; aorta; e veia cava inferior) em TC; (d) Segmentação do tumor hepático, da esquerda para a direita: imagem original de TC, imagem da segmentação do fígado e imagem da segmentação da lesão; (e) Predição de núcleos, da esquerda para a direita: imagens de células originais, predição de núcleos, marcação de núcleos nas imagens originais; (f) Segmentação celular; (g) Segmentação de lesão de pele; e (h) segmentação do nervo da córnea. Imagem adaptada de (SIDDIQUE *et al.*, 2021).

dos anteriores) (EIGEN; PUHRSCHE; FERGUS, 2014). Esta é essencialmente a premissa do *data augmentation* (GOODFELLOW; BENGIO; COURVILLE, 2016).

Em situações das quais não se dispõe de grande variedade de imagens, pode-se aumentar esse número de imagens sinteticamente por intermédio de alterações e transformações das imagens originais como: diferentes orientações, localizações, escalas e brilho. Pode-se também inserir distorções geométricas não lineares ou perturbações randômicas. Mesmo em conjuntos de dados suficientemente grandes, o *data augmentation* é considerada uma boa prática por aumentar a capacidade de generalização da rede (GOODFELLOW; BENGIO; COURVILLE, 2016).

### 2.8.3 Métricas de Avaliação

A análise qualitativa (visual) dos resultados obtidos de uma CNN pode ser tendenciosa e não ser suficiente para dizer que uma abordagem é melhor que outra. Desse modo, a maioria dos trabalhos utilizam as métricas que serão apresentadas a seguir para avaliar seus métodos, podendo assim compará-los com outros artigos presentes na literatura (EIGEN;

PUHRSCH; FERGUS, 2014; LI *et al.*, 2015; CAO; WU; SHEN, 2016).

- Acurácia:

É a probabilidade do teste em fornecer resultados corretos, ou seja, ser positivo nas regiões com presença da doença e negativo nas regiões de ausência da doença. Expresso de outra forma é a probabilidade dos verdadeiros positivos (VP) e verdadeiros negativos (VN) como uma proporção de todos os resultados - falso positivo (FP) e falso negativo (FN) e é calculado como:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

- Sensibilidade:

É a capacidade que o teste apresenta de detectar os indivíduos VP, ou seja, de diagnosticar corretamente os pontos de interesse. Em termos matemáticos:

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

- Precisão:

É a capacidade que o teste possui em detectar os VP dentre todas as classificações positivas. Em termos matemáticos:

$$Precisão = \frac{VP}{(VP + FP)}$$

- Especificidade:

É a capacidade que o teste apresenta em detectar os VN, isto é, de diagnosticar corretamente as regiões de VN da imagem. Em termos matemáticos:

$$Especificidade = \frac{VN}{(FP + VN)}$$

- IoU:

A técnica de avaliação *Intersection over Union* é uma métrica utilizada para medir a precisão de um detector de objetos em um conjunto de dados específico. Sendo assim, esta métrica é invariante em função da proporção espacial da região a ser avaliada. Logo, se a predição estiver completamente correta,  $IoU = 1$ . Quanto menor a IoU, pior o resultado da previsão.

$$IoU = \frac{Intersecção}{União}$$

- Curva ROC (*Receiver Operating Characteristic Curve*) e AUC (*Area Under the Curve*):

Para classificações binárias é necessário escolher uma regra de predição que determine quando a amostra é classificada com lesão ou não, estabelecendo um ponto de corte entre o valor 0 e 1. Uma maneira utilizada para determinar esse ponto é o emprego da análise da curva ROC. Esta curva é construída plotando-se a sensibilidade em função de “1-especificidade” para todos os possíveis pontos de corte (*threshold*). Após a determinação desta curva, é possível calcular a AUC, a qual fornece uma informação importante, representando a probabilidade de um par aleatório de amostras com lesão e sem lesão ser classificado corretamente (HANLEY; MCNEIL, 1982). Como a curva representa um gráfico com eixos de 0 a 1, sua área máxima é um quadrado de área 1, representando o melhor resultado (resultado perfeito) (PRATI *et al.*, 2008). Desse modo, pode-se avaliar o resultado comparando o formato da curva e as frações da área obtida.

## 2.9 Considerações sobre o uso da técnica

É possível afirmar que a habilidade das redes neurais profundas de extrair características cada vez mais representativas, bem como a sua capacidade de interpretar conceitos mais complexos, a torna uma das técnicas de aprendizado de máquina mais promissoras. Isto pode ser comprovado ao se observar os resultados/*benchmarks*, os quais tem superado técnicas clássicas em diferentes áreas de pesquisa e suas aplicações (ALZUBAIDI *et al.*, 2021). Entretanto, ao mesmo tempo que estas redes são adequadas para diversas aplicações, ainda são computacionalmente custosas, requerem ajustes de hiperparâmetros não determinísticos e uma elevada quantidade de dados para serem (ALZUBAIDI *et al.*, 2021). Em vista disso, muitos trabalhos são apresentados com o intuito de superar estes desafios.

Na próxima Seção, serão abordados os melhores resultados para segmentação das regiões de interesse em exsudatos duros, exsudatos moles, hemorragias e microaneurismas.





### 3 REVISÃO BIBLIOGRÁFICA

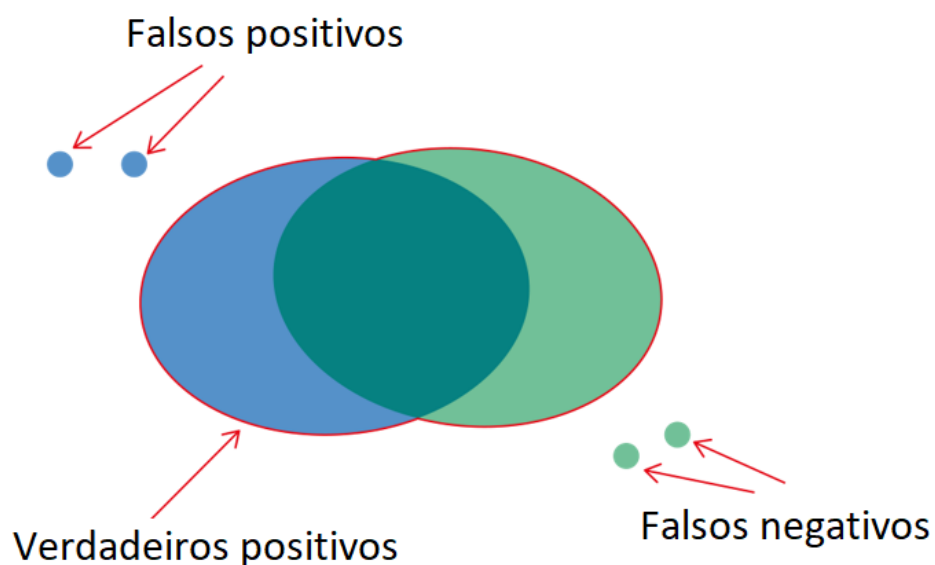
#### 3.1 Detecção de Retinopatia Diabética por Sistema de Diagnóstico por Computador (CAD)

O objetivo do diagnóstico assistido por computador é reconhecer pontos de interesse na imagem como áreas de exsudatos, microaneurismas, veias, texturas e hemorragias características (TRUCCO *et al.*, 2013b; MOOKIAH *et al.*, 2013).

Para a identificação da retinopatia diabética, diversos trabalhos foram propostos, empregando diferentes técnicas (USHER *et al.*, 2004), (SINTHANAYOTHIN *et al.*, 2003), (APTEL *et al.*, 2008), (REZA; ESWARAN, 2011), (GARDNER *et al.*, 1996), (KAHAI; NAMUDURI; THOMPSON, 2006), (OSAREH *et al.*, 2002) e (QUELLEC; RUSSELL; ABRAMOFF, 2011).

A seguir, serão listados alguns trabalhos que possuem resultados relevantes, conforme as suas métricas. Vale ressaltar que alguns trabalhos apresentam perspectivas próprias de análise de dados, como Guo *et al.* (2020), em que é considerado verdadeiro positivo toda região segmentada, se ao menos um pixel estiver contido na região demarcada no *ground truth*, conforme Figura 17. Esta perspectiva foi utilizada para uma aplicação onde o interesse era identificar a existência de uma estrutura e não necessariamente identificar seu formato.

Figura 17 – Métrica particular utilizada em (GUO *et al.*, 2020)



Fonte: Imagem adaptada de Guo *et al.* (2020)

### 3.2 Segmentação de exsudatos moles e duros

Os exsudatos são um dos principais sinais de RD que podem ser prevenidos com um processo de avaliação e detecção precoce. Serão descritos, portanto, alguns trabalhos que realizam a segmentação destas estruturas.

Um dos primeiros métodos usados para segmentação de exsudatos é por limiar. Os primeiros trabalhos propuseram um algoritmo que consiste em correção das cores, aprimoramento de contraste e nitidez, baseado em um valor de limiar global (WARD; TOMLIIVSON; TAYLOR, 1989). A desvantagem desse método é que o valor do limite é selecionado manualmente e, portanto, depende do banco de dados. Phillips, Forrester e Sharp (1993) surgiram com um método semelhante, mas utilizando uma combinação de limiar local e global. Frame *et al.* (1998) empregaram médicos especialistas para identificar os exsudatos e utilizaram a análise de textura para estudá-los. Wang *et al.* (2000) utilizaram ajuste de brilho, um método de classificação estatística e uma estratégia de verificação baseada em janela local para detectar exsudatos. A abordagem de crescimento de região foi utilizada por Ege *et al.* (2000). Em seguida, um filtro de mediana (31 x 31 pixels) foi utilizado para permitir que o limiar fosse definido para segmentar lesões parcialmente brilhantes. Uma classificação baseada no vizinho mais próximo (*Nearest Neighbors*, NN) foi adotada para distinguir os exsudatos. Osareh *et al.* (2001) empregaram normalização de cor e realce de contraste local na fase de pré-processamento. Em seguida, foi adotado o agrupamento *fuzzy* de *c-means* e classificador de redes neurais. Li e Chutatape (2003) combinaram métodos de crescimento de região e detecção de bordas para localizar os exsudatos e relataram um alto nível de sensibilidade. Sánchez *et al.* (2004) empregaram um algoritmo baseado em modos de mistura para limitar dinamicamente as imagens a fim de segmentar os exsudatos. No entanto, na fase de pós-processamento, foi utilizada uma seleção de limiares heurísticos. Noronha, Nayak e Bhat (2006) identificaram uma variação de alto nível de pixels em exsudatos e determinaram seus contornos por meio de reconstrução morfológica.

O sucesso de classificadores de aprendizado profundo como *AlexNet*, *VGGNet*, *GoogleNet* e *ResNet* é muito evidente no desafio de classificação do *ImageNet*. Essas redes forneceram bons resultados de classificação, todavia, quando estendidos para a segmentação em pixels, eles falharam em fornecer resultados igualmente bons (SAMBHYAL *et al.*, 2020).

O trabalho Sambyal *et al.* (2020) apresentam uma arquitetura *U-Net* modificada com base na rede residual e emprega embaralhamento periódico com convolução de sub-pixel inicializado com redimensionamento por NN (*Nearest Neighbour*). A arquitetura proposta foi treinada e validada para segmentação de microaneurismas e exsudato duro em dois conjuntos de dados disponíveis ao público, IDRiD e e-optha (DECENCIERE *et al.*, 2013). Para o conjunto de dados IDRiD, a rede obteve 99,88% de acurácia, 99,85% de sensibilidade e 99,95% de especificidade, sendo este os resultados recentes de maior

destaque.

### 3.3 Segmentação de microaneurismas

A segmentação de microaneurismas é muito importante, porque essas estruturas constituem a primeira característica da RD. Os primeiros relatórios que ligam essas estruturas à RD datam de 1879 (MICROANEURYSMS... , 1971). Jalli, Hellstedt e Immonen (1997) analisaram o surgimento e desaparecimento de microaneurismas em diferentes fases da angiografia fluoresceínica. Em um estudo semelhante, tanto a taxa de formação quanto o desaparecimento de microaneurismas no início da RD foram analisados (HELLSTEDT; IMMONEN, 1996). O volume de microaneurismas foi calculado a partir de imagens de fundo de retina (BERNARDES *et al.*, 2009). Ao se utilizar o método *MA-tracker* (MAHER; DHOPEHWARKAR, 2015) para contar microaneurismas, foi observado que os microaneurismas permanecem estáveis ao longo do tempo, mas apenas 29% permanecem no mesmo local.

A Figura 18 exemplifica os resultados da segmentação de microaneurismas para retinopatia normal e proliferativa. No exemplo, a componente verde da imagem RGB foi escolhida para compor a imagem de análise. Semelhante ao algoritmo de segmentação de exsudatos, foram removidas as estruturas proeminentes dentro das imagens da retina, como os vasos sanguíneos e o disco óptico. Em seguida, uma sofisticada sequência de algoritmos de processamento de imagens foram usados para determinar as áreas dentro das imagens de retina e segmentar os microaneurismas (ACHARYA *et al.*, 2009).

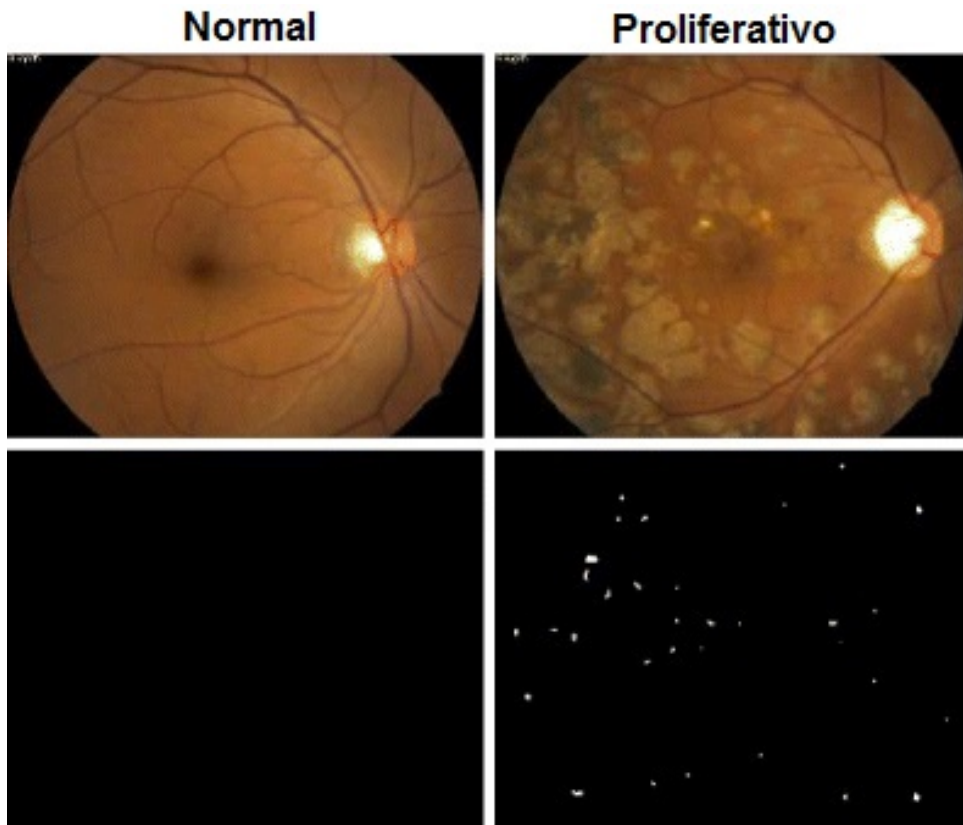
Mais recentemente, Kou *et al.* (2020a) propuseram uma rede *U-Net* especializada e aprimorada para segmentação de microaneurismas (ERU-Net). Esta rede é obtida pela adição de um bloco residual na *U-Net* com um caminho de *downsampling* e três caminhos de *upsampling*. Esta rede atingiu a acurácia de 94,20%, especificidade de 94,10% e sensibilidade de 91,78%, sendo estes resultados de destaque para segmentação de microaneurismas do banco de dados IDRiD.

### 3.4 Segmentação de Hemorragias

À medida que o grau da RD avança, as hemorragias retinianas tornam-se evidentes. Estas estruturas indicam uma retina aumentada de isquemia (perda de oxigênio). Com seu surgimento e agravamento da RD, os vasos retinianos ficam mais danificados, levando à exsudação de líquido, lipídios e proteínas. A Figura 19 apresenta o resultado da segmentação de hemorragia (ACHARYA *et al.*, 2009) e as manchas brancas indicam a hemorragia na imagem.

Algoritmos de computador totalmente automatizados foram capazes de detectar exsudatos duros, hemorragias e microaneurismas utilizando uma nova técnica, denominada

Figura 18 – Resultados da segmentação de microaneurismas.



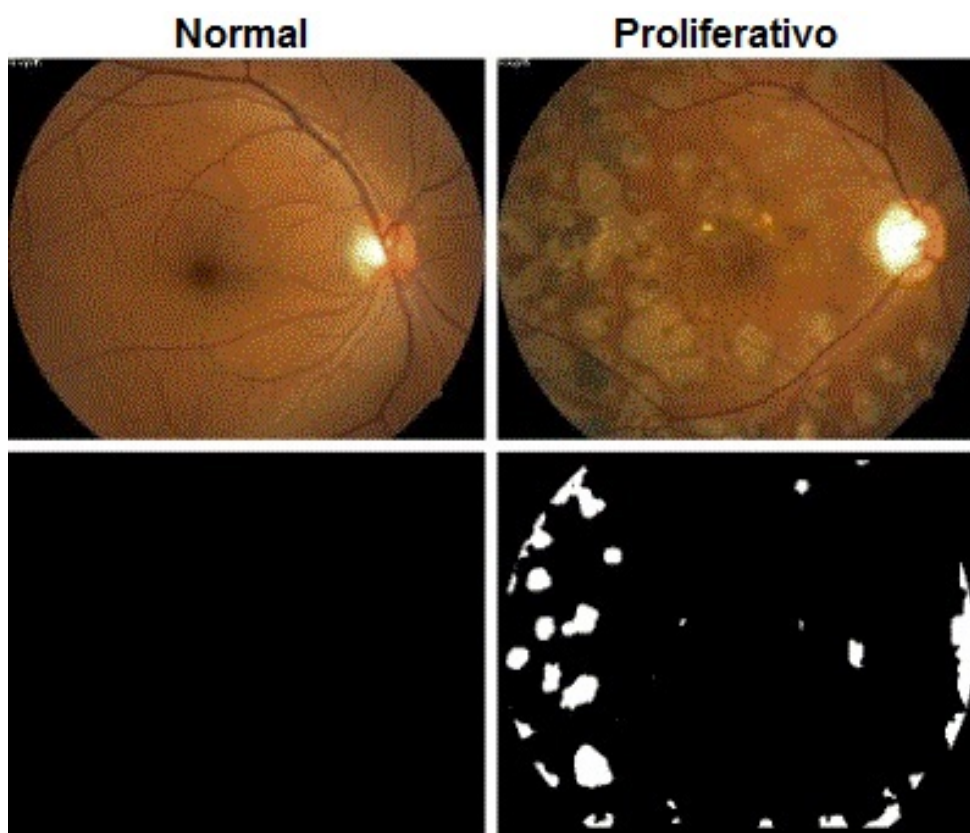
Fonte: Imagem adaptada de Faust *et al.* (2012)

*Operador de Moat*, a qual é capaz de realçar as bordas obtidas pelo contraste das lesões (SINTHANAYOTHIN *et al.*, 2002; SINTHANAYOTHIN *et al.*, 2003). A sensibilidade e especificidade para segmentação de exsudatos foram 88,5% e 99,7%, respectivamente, e o algoritmo alcançou uma sensibilidade de 77,5% e especificidade de 88,7% para a segmentação de hemorragias.

Em 2020, Nazir *et al.* (2020) alcançaram acurácia de 95,2%, sensibilidade de 95,1% e especificidade de 96,0%, sendo estes os resultados de relevância e destaque disponíveis atualmente.

### 3.5 Considerações

Figura 19 – Resultados da segmentação de hemorragias.



Fonte: Imagem adaptada de Faust *et al.* (2012)



## 4 MÉTODOS

Este trabalho compreende três análises investigativas na etapa de pré-processamento. A primeira etapa foi responsável pela divisão das imagens originais em subimagens, bem como seus respectivos *ground truth*. A segunda etapa consiste na formação de sete conjuntos de imagens com seus respectivos canais de cores (RGB) e suas associações. A terceira etapa foi composta pela associação das seguintes estruturas de interesse: exsudatos duros, exsudatos moles, hemorragias e microaneurismas. Na etapa de processamento é apresentada a definição da rede neural, suas características e justificativas de escolhas. Por fim, são apresentados diferentes métricas de avaliação para os resultados obtidos. Todas as imagens utilizadas nesse trabalho são da base de dados IDRiD (*Indian Diabetic Retinopathy Image Dataset*), a seguir suas características e justificativa de escolha são apresentadas.

### 4.1 Base de dados

A base de dados utilizada neste trabalho é a IDRiD. Esta base consiste em 81 imagens de retina com sinais de retinopatia diabética. Estas imagens possuem anotações precisas de nível de pixel, indicando anormalidades associadas a microaneurismas, exsudatos moles, exsudatos duros e hemorragias. Também são disponibilizados os respectivos *ground truth* para treinamento, testes e avaliação de desempenho de técnicas de segmentação das estruturas. Este banco de dados fornece imagens de retinas e *ground truth* coloridos e separados para cada tipo de estrutura. Os *ground truth* são, portanto, regiões identificadas por experientes profissionais e serão a base de treinamento da rede neural, assumindo que estes são os resultados perfeitos que a rede neural deve reproduzir.

As imagens disponíveis no banco de dados IDRiD são classificadas e avaliadas por médicos especialistas, os quais garantiram os seguintes aspectos: segmentação das regiões de interesse, qualidade adequada, relevância, unicidade e razoável mistura de estratificação da doença representativa da retinopatia diabética (PORWAL *et al.*, 2018a). Este conjunto de dados foi disponibilizado como parte do “Desafio da Retinopatia Diabética: Segmentação e Classificação”, organizado em conjunto com o Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) e *International Symposium on Biomedical Imaging (ISBI 2018)*, Washington DC. Este desafio foi hospedado no *Grand Challenges in Biomedical Imaging Platform* (PORWAL *et al.*, 2018a).

Cabe destacar que esta base de dados é disponibilizada com divisão do conjunto de treinamento e conjunto de testes, constituindo-se em 54 imagens para treinamento e 27 imagens para testes. Todas as imagens possuem 4288 pixels x 2848 pixels x 3 canais (RGB).

Para esta aplicação, este trabalho propõe a divisão da imagem original em imagens menores, ampliando, substancialmente, o número de imagens de treinamento e avaliação. Como forma de realizar *data augmentation offline* as imagens foram divididas em subimagens, o que amplia as possibilidades de treinamento e evita erros de generalização. Ao dividir as imagens, algumas subimagens se tornam nulas e podem ser desconsideradas do processo de treinamento, agilizando e auxiliando o mesmo.

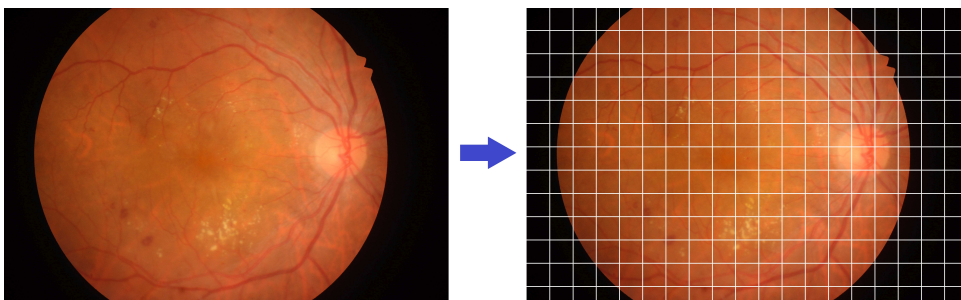
## 4.2 Pré-processamento

A etapa inicial de desenvolvimento do método foi dedicada à preparação do *dataset* para o treinamento da rede. O pré-processamento foi composto por três etapas. A primeira etapa compreendeu a divisão das imagens originais e respectivos *ground truth* em subimagens. A segunda, consistiu na formação de sete conjuntos de imagens: imagens originais (RGB), imagens com cada canal individual - vermelho (R), verde (G) e azul (B), assim como suas possíveis combinações - vermelho e verde (RG), vermelho e azul (RB) e verde e azul (GB). A terceira e última etapa foi composta pela associação das seguintes estruturas de interesse: exsudatos duros, exsudatos moles, hemorragias e microaneurismas.

### 4.2.1 Primeira etapa: divisão das imagens originais

Nesta primeira etapa, as imagens e seus respectivos *ground truth* foram divididos, formando subimagens. Isto foi necessário porque as imagens originais apresentam resolução de 4288 pixels x 2848 pixels x 3 canais (RGB) e, também, porque treinar a rede convolucional com a imagem completa, sem dividi-la, torna-se um processo computacionalmente oneroso. Como solução, as imagens foram subdivididas em 512 pixels x 512 pixels por 3 canais (RGB) para a rede *U-Net*. A Figura 20 apresenta, de maneira ilustrativa, a divisão da imagem original da retina em subimagens. A Figura 21, por sua vez, ilustra a aplicação do processo de subdivisão do *ground truth*, a qual garante a correspondência durante o processo de treinamento, testes, validação e aplicação.

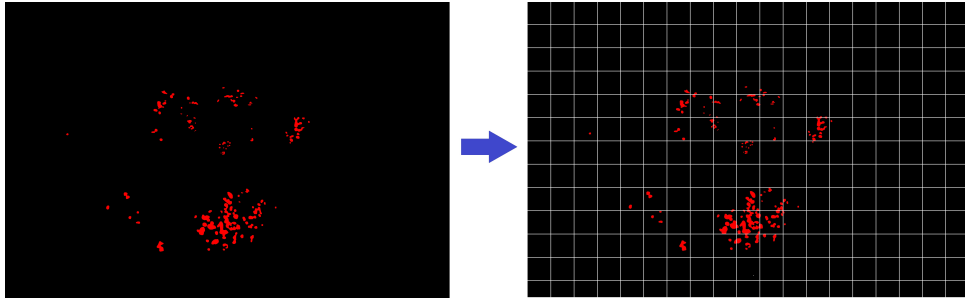
Figura 20 – Ilustração da subdivisão da imagem original da retina em subimagens.



Fonte: Elaborada pelo autor.



Figura 21 – Ilustração da subdivisão da imagem original do *ground truth* em subimagens.



Fonte: Elaborada pelo autor.

Como uma única imagem foi dividida em imagens menores, algumas subimagens passaram a não conter informações relevantes, tanto na imagem original quanto no *ground truth* resultando em matrizes nulas (conforme pode ser observado nas Figuras 20 e 21). Nesse caso, a subimagem e seu respectivo *ground truth* foram descartados com o intuito de acelerar o processo de treinamento, ou seja, manter apenas subimagens com informações/relevância.

O conjunto de imagens foi dividido levando em consideração a entrada da rede *U-Net* de 512 pixels x 512 pixels e, também, excluindo subimagens totalmente nulas. Para essa subdivisão, o conjunto inicial de treinamento de 54 imagens da base de dados IDRiD, tornou-se 2.004 subimagens e o conjunto de testes com 27 imagens, tornou-se 1.005 subimagens. Estas subdivisões e exclusões das subimagens nulas foram aplicadas no subconjunto de treinamento, testes e validação. Por conseguinte, seus resultados continuam válidos e, assim, aceleram os resultados. Este procedimento pode ser realizado no conjunto de imagens de aplicação sem perda de informação, permitindo ao final do processo reconstituir a imagem original e a imagem segmentada. Este pré-processamento configura-se como *data augmentation offline*, por permitir aumentar o número de imagens por meio da sua divisão, não havendo, portanto, perda de informações.

Das 2.004 imagens do conjunto de treinamento, 204 foram utilizadas para validação e análise do *overfitting*. Sendo assim, a proporção de treinamento-teste-validação foi de 60%-33%-7%.

Após a definição dos melhores parâmetros de decaimento para a rede neural proposta, as imagens foram alteradas com espelhamento horizontal e vertical, aleatoriamente, durante o processo de treinamento (*data augmentation* durante a execução do treinamento - *online*). Dessa maneira, por meio desses métodos, obtém-se melhores resultados ao tornar a rede com maior poder de generalização. (DAS *et al.*, 2021).

#### 4.2.2 Segunda etapa: Divisão dos canais RGB

Nesta etapa, foi proposta a avaliação das combinações dos canais de cores das imagens de entrada. Um primeiro teste foi realizado utilizando as estruturas de interesse com seus canais de cores independentes, isto é, utilizando os canais R, G e B separadamente. Em seguida, foram avaliadas as combinações RG, RB, GB e RGB, aplicando a rede neural proposta. Estas avaliações de canais de cores foram realizadas para ED, EM, HE e MA.

Trabalhos como Karasu *et al.* (2020), Zhou *et al.* (2017), Li *et al.* (2020) e Hu *et al.* (2018) utilizam os canais RGB como entrada na rede neural enquanto outros como Hassan *et al.* (2015), Fraz *et al.* (2012) e Staal *et al.* (2004) utilizam apenas o canal verde (G) como entrada na rede neural para segmentação. Sendo assim, pode-se afirmar que não há consenso sobre o melhor canal de cores a se utilizar. Dessa forma, esta segunda etapa se propõe a apresentar e discutir as diferenças de cada canal de cor RGB.

Os exsudatos duros e moles ocorrem em diferentes formas e tamanhos na superfície da retina, geralmente formados por lesões brilhantes (CHIDAMBARAM; VIJAYAN, 2018). Portanto, é possível afirmar que a rede neural apresenta melhores resultados utilizando-se dos três canais. A hemorragia se apresenta como manchas mais escuras e de maior tamanho, sua predominância ocorre no canal verde. Os microaneurismas são manchas pequenas e escuras predominantemente no canal verde e azul.

A escolha de se trabalhar com uma imagem com seus 3 canais de cores permite que a rede selecione e ocorra a convergência dos pesos sinápticos de acordo com os canais (ou combinações destes) a fim de reduzir o erro. Entretanto, um grande número de informações não utilizadas pode dificultar a convergência, despendendo maior tempo de treinamento ou resultar em mínimos locais, dificultando, assim, a obtenção de mínimos globais.

#### 4.2.3 Terceira etapa: Combinações das estruturas de interesse

Nesta etapa, foram realizados os treinamentos da rede *U-Net*, por meio das imagens individuais (exsudatos duros, exsudatos moles, hemorragias e microaneurismas) e de suas combinações, conforme referência apresentada na Tabela 1. Neste caso, os respectivos *ground truth* foram somados, compondo em uma única imagem diferentes estruturas.

Como a última camada da rede é uma função sigmóide, é possível realizar pós-processamento variando-se o *threshold* e, assim, tornar o resultado mais seletivo ou mais abrangente em função de sua saída. Dessa maneira, é possível buscar resultados priorizando diferentes métricas.

### 4.3 Definições da rede neural

A arquitetura de CNN escolhida foi a *U-Net* por ser mais adequada para segmentação semântica do que as CNNs tradicionais (TSIKNAKIS *et al.*, 2021b) devido à sua

Tabela 1 – Combinações de regiões de interesse avaliadas

Nomenclatura	Combinação
ED & EM	Exsudatos duros e exsudatos moles
ED & HE	Exsudatos duros e hemorragias
ED & MA	Exsudatos duros e microaneurismas
EM & HE	Exsudatos moles e hemorragias
EM & MA	Exsudatos moles e microaneurismas
HE & MA	Hemorragias e microaneurismas
Todas	Exsudatos duros, exsudatos moles, microaneurismas e hemorragias

capacidade de preservar a integridade estrutural da imagem. Essa arquitetura consistem em um caminho de contração para capturar o contexto relevante e um caminho de expansão simétrico, permitindo uma segmentação precisa. Além disso, uma arquitetura *U-Net* tem menos parâmetros e é mais rápida do que as CNNs tradicionais devido ao fato de processar a imagem em uma única convolução, ao invés de processar vários *patches* em uma abordagem de janela deslizante, como faria uma CNN. Por esse motivo, essas arquiteturas são chamadas de “Redes Totalmente Convolucionais” (FCN). Esta arquitetura também requer muito menos dados do que as CNNs tradicionais para realizar uma tarefa de segmentação, o que é crucial para a análise de imagens médicas, em que o número de dados disponíveis é menor do que em outros campos da visão computacional, por necessitar banco de dados com aprovação no comitê de ética e pelo envolvimento de médicos especialistas e experientes que possam fornecer *ground truth* adequado.

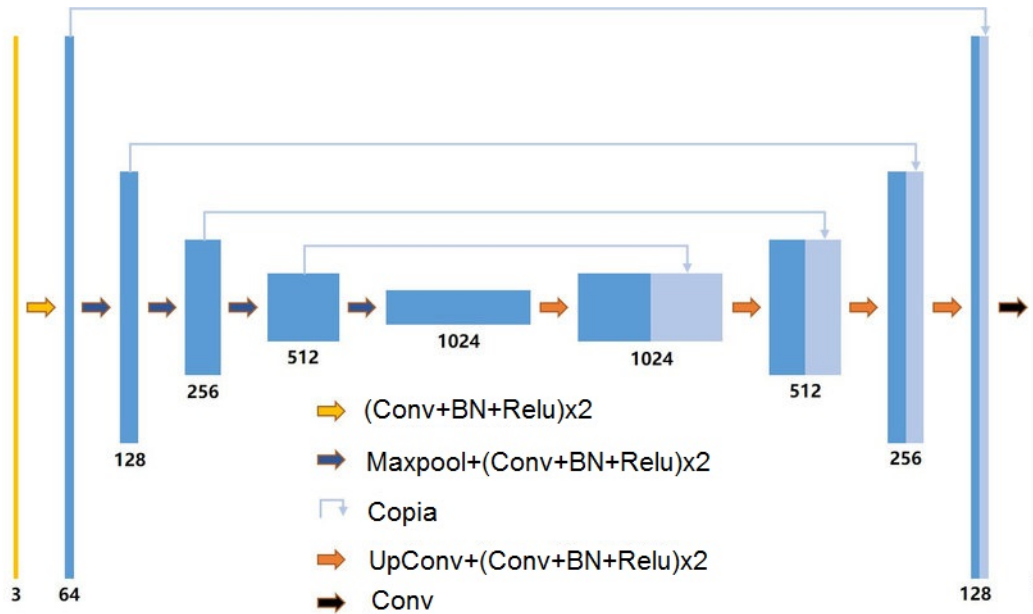
Para implementação da rede *U-Net* foi utilizada a linguagem de programação *Python* (KETKAR; SANTANA, 2017) devido à sua facilidade de aprendizado, rápida implementação e grande comunidade de usuários, permitindo suporte a linguagem quando necessário. Também foi utilizado como suporte à implementação, a biblioteca *Keras* (KETKAR; SANTANA, 2017), sendo esta uma interface de alto nível para redes neurais escritas em *Python*. A Tabela 3 apresenta a topologia de rede implementada, com seus parâmetros. Diversos treinamentos foram realizados com o intuito de encontrar mínimos globais e, dessa forma, foi utilizado decaimento com diferentes valores e *batch size* com uma ou duas imagens, limitado pela memória disponível do *hardware*. A Figura 22 apresenta o diagrama de blocos da rede neural implementada. A função de ativação utilizada na camada de saída foi a sigmóide com o intuito de gerar gradiente de intensidade das regiões segmentadas. Sendo assim, utilizar esta saída permite avaliar a curva ROC e AUC desses resultados.

O conjunto de treinamento foi processado por uma função de definição de limiar (*threshold*). Este limiar foi determinado e definido no conjunto de treinamento, por meio do incremento automático do seu valor e análise da métrica desejada. Portanto, neste trabalho, é apresentado o melhor resultado para diferentes métricas, o que implica em

diferentes valores de limiares. Após sua definição no conjunto de treinamento, este valor também foi aplicado no conjunto de testes.

Os parâmetros de treinamento que compõem a rede neural utilizada estão apresentados na Tabela 2. Sua arquitetura é descrita na Tabela 3.

Figura 22 – Diagrama de blocos da rede simplificada implementada.



Fonte: Imagem criada pelo autor.

Tabela 2 – Parâmetros da rede neural *U-Net* utilizada.

Total de parâmetros	31.032.837
Parâmetros treináveis	31.032.837
Parâmetros fixos	0
Tempo médio de treinamento por época	135 segundos
Tempo médio de predição por imagem	12 segundos
<i>Batch size</i>	2
<i>Early stopping</i>	5
<i>Data augmentation offline</i>	Divisão da imagem
<i>Data augmentation online</i>	<i>flip</i> vertical e horizontal
Taxa de aprendizado <sup>1</sup>	$32 \times 10^{-5}$
Otimizador	<i>Adam</i> ( $\beta_1 = 0,9$ , $\beta_2 = 0,99$ )

<sup>1</sup>Cada estrutura foi treinada com diferentes taxas de aprendizados, conforme Apêndice A. Para análise dos canais de cores foi utilizada mesma taxa de  $32 \times 10^{-5}$

Para o gerenciamento da atualização das variáveis do sistema, foram utilizadas as funções do *Keras* para monitoramento. Associado ao método de descida de gradiente para a otimização, foi utilizado o *momentum*, que possui uma variável de controle para as

Tabela 3 – Descrição das camadas da arquitetura U-Net

Layer (tipo)	Formato de saída	Parâmetros
input_1 (InputLayer)	(None, 256, 256, 3)	0
conv2d (Conv2D)	(None, 256, 256, 64)	1792
conv2d_1 (Conv2D)	(None, 256, 256, 64)	36928
max_p2d (MaxPooling2D)	(None, 128, 128, 64)	0
conv2d_2 (Conv2D)	(None, 128, 128, 128)	73856
conv2d_3 (Conv2D)	(None, 128, 128, 128)	147584
max_p2d_1 (MaxPooling2D)	(None, 64, 64, 128)	0
conv2d_4 (Conv2D)	(None, 64, 64, 256)	295168
conv2d_5 (Conv2D)	(None, 64, 64, 256)	590080
max_p2d_2 (MaxPooling2D)	(None, 32, 32, 256)	0
conv2d_6 (Conv2D)	(None, 32, 32, 512)	1180160
conv2d_7 (Conv2D)	(None, 32, 32, 512)	2359808
dropout (Dropout)	(None, 32, 32, 512)	0
max_p2d_3 (MaxPooling2D)	(None, 16, 16, 512)	0
conv2d_8 (Conv2D)	(None, 16, 16, 1024)	4719616
conv2d_9 (Conv2D)	(None, 16, 16, 1024)	9438208
dropout_1 (Dropout)	(None, 16, 16, 1024)	0
up_s2d (UpSampling2D)	(None, 32, 32, 1024)	0
conv2d_10 (Conv2D)	(None, 32, 32, 512)	2097664
concatenate (Concatenate)	(None, 32, 32, 1024)	0
conv2d_11 (Conv2D)	(None, 32, 32, 512)	4719104
conv2d_12 (Conv2D)	(None, 32, 32, 512)	2359808
up_s2d_1 (UpSampling2D)	(None, 64, 64, 512)	0
conv2d_13 (Conv2D)	(None, 64, 64, 256)	524544
conc_1 (Concatenate)	(None, 64, 64, 512)	0
conv2d_14 (Conv2D)	(None, 64, 64, 256)	1179904
conv2d_15 (Conv2D)	(None, 64, 64, 256)	590080
up_s2d_2 (UpSampling2D)	(None, 128, 128, 256)	0
conv2d_16 (Conv2D)	(None, 128, 128, 128)	131200
conc_2 (Concatenate)	(None, 128, 128, 256)	0
conv2d_17 (Conv2D)	(None, 128, 128, 128)	295040
conv2d_18 (Conv2D)	(None, 128, 128, 128)	147584
up_s2d_3 (UpSampling2D)	(None, 256, 256, 128)	0
conv2d_19 (Conv2D)	(None, 256, 256, 64)	32832
conc_3 (Concatenate)	(None, 256, 256, 128)	0
conv2d_20 (Conv2D)	(None, 256, 256, 64)	73792
conv2d_21 (Conv2D)	(None, 256, 256, 64)	36928
conv2d_22 (Conv2D)	(None, 256, 256, 2)	1154
conv2d_23 (Conv2D)	(None, 256, 256, 1)	3

mudanças nos parâmetros de peso da rede, prevenindo que os novos parâmetros desviem demasiadamente de sua direção de mudança anterior (SILVA; SPATTI; FLAUZINO, 2010). O *momentum* empregado foi o Adam (MA; YARATS, 2018) que, de maneira geral, leva em consideração o histórico de mudanças do parâmetro e influencia o decaimento do peso

para que decresça menos a cada passo em direção ao estado de erro mínimo (PONTI *et al.*, 2017).

Durante o treinamento, os pesos sinápticos eram atualizados a cada época e salvos se houvesse melhora dos resultados (redução da função custo). Como critério de parada (*early stopping*), foi avaliada a função de custo no conjunto de validação. Se em uma dada época este valor for registrado como valor mínimo, o sistema então é treinado por uma margem de 5 épocas adicionais para garantir que este custo não será superado por treinamentos subsequentes e, caso continue a ser o valor mínimo, os resultados da época em questão são salvos como o melhor modelo. Caso o valor seja superado, o melhor modelo é atualizado e o processo se repete. Em todos os treinamentos, foram gerados gráficos como registro e análise do decaimento, tanto do conjunto de treinamento, quando do conjunto de controle. Sendo assim, foi possível ver a atuação do *early stopping*.

A execução do programa foi obtida com o auxílio de uma *GPU* de memória de 6 GB da marca @NVIDIA. A plataforma de desenvolvimento *Tensorflow* integra a biblioteca CUDA10 (*Compute Unified Device Architecture*) em sua versão V10.1.105, que permite a aceleração da aplicação devido aos seus núcleos especializados. Cabe ressaltar que, para os resultados obtidos nas redes propostas neste trabalho, foi utilizado o computador descrito na Tabela 4.

Tabela 4 – Computador e sistema utilizado

Processador	Intel Core i5-10400 Hexa-Core 2.9Ghz (4GHz Turbo) 12MB Cache
Memória	16 GB (2x8GB DDR4 3000MHz)
<i>GPU</i>	<i>NVIDIA - RTX2060 6 GB</i>
Biblioteca ( <i>GPU</i> )	CUDA10 V10.1.105
Armazenamento	SSD Sata 6 GB/s
Sistema Operacional	Linux - Ubuntu 20.04 LTS
Programação	Python 3.8.10
Biblioteca de <i>deep learning</i>	<i>Keras 2.4.3</i>

Os hiperparâmetros são um conjunto de variáveis cujos valores são normalmente selecionados experimentalmente durante o processo de treinamento. Isso geralmente inclui taxa de aprendizado, *batch size* e número de épocas (SAMBYAL *et al.*, 2020). Os valores dos hiperparâmetros utilizados no trabalho aqui proposto foram determinados experimentalmente com a finalidade de otimizar as métricas de avaliação (IoU, acurácia, especificidade e sensibilidade). Foram testadas 10 possibilidades com incremento automático do decaimento e do *batch size*. Destes treinamentos, alguns degeneraram para matriz de predição nula. Logo, estes resultados foram desconsiderados. Por fim, foram selecionados os valores dos parâmetros que forneceram os melhores resultados para a segmentação semântica.

Neste trabalho, foi utilizada uma mesma arquitetura de CNN para diferentes estruturas com a finalidade de permitir a comparação justa entre as combinações de diferentes estruturas numa mesma imagem. Por se tratar de estruturas características de uma mesma doença RD, foram analisadas as segmentações dessas estruturas isoladas e em conjunto, ou seja, suas combinações.

#### **4.4 Métrica de avaliação**

Diferentes métricas disponíveis na literatura são comumente empregadas para avaliar a segmentação dos pontos de interesse em imagens de retina. Algumas delas apresentam perspectivas próprias em função do objetivo, como por exemplo, a métrica apresentada na Figura 17, em que Guo *et al.* (2020) consideram verdadeiro positivo toda região segmentada, caso haja ao menos um pixel como VP.

Neste trabalho, são apresentadas as métricas: acurácia, sensibilidade, precisão, especificidade, IoU e AUC, sendo estas apresentadas em sua forma padrão. É possível afirmar que a acurácia e especificidade apresentam os resultados mais altos, uma vez que as estruturas de interesse possuem tamanhos reduzidos comparativamente à imagem total. Dessa forma, o número de verdadeiros negativos tende a ser dominante em ambas as métricas de avaliação.





## 5 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Nesta seção, são apresentados os resultados das análises do pré-processamento. A primeira análise refere-se a divisão das imagens e respectivos *ground truth*. A segunda análise apresenta os resultados das estruturas de interesse (exsudatos duros, exsudatos moles, microaneurismas e hemorragias) para cada canal de cor e suas combinações. A terceira da análise das estruturas de interesse (exsudatos duros, exsudatos moles, microaneurismas e hemorragias) individuais e suas combinações. Para análise qualitativa, foi utilizada a imagem de número 59 do banco de dados IDRiD com seu respectivo canal de cor e *ground truth*. Para análise quantitativa foram apresentados o valor AUC, curva ROC, acurácia, sensibilidade, especificidade e precisão.

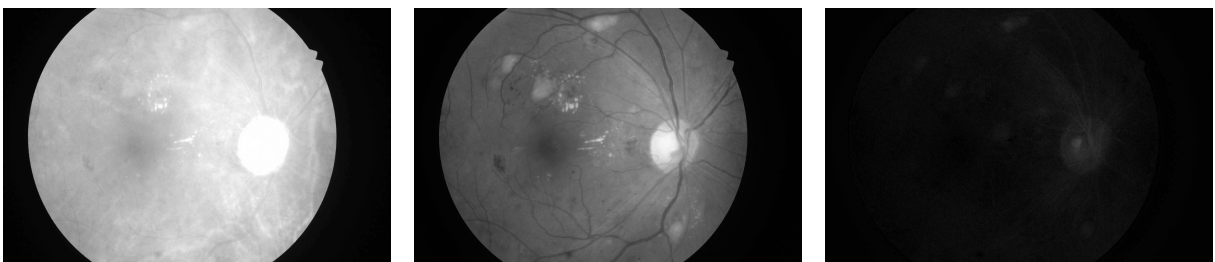
### 5.0.1 Primeira etapa: divisão das imagens originais

Ao se utilizar o método proposto, a rede *U-Net* foi implementada com entrada de 512 x 512 pixels, isto é, com quadrante reduzido e uso das diferentes opções de canais de cores. O pré-processamento foi realizado, consistindo na divisão das imagens em quadrantes menores, seguido pela exclusão dos quadrantes nulos. Realizaram-se ciclos de treinamento com vistas a otimizar o desempenho da rede de acordo com as métricas.

### 5.0.2 Segunda etapa: Divisão dos canais RGB

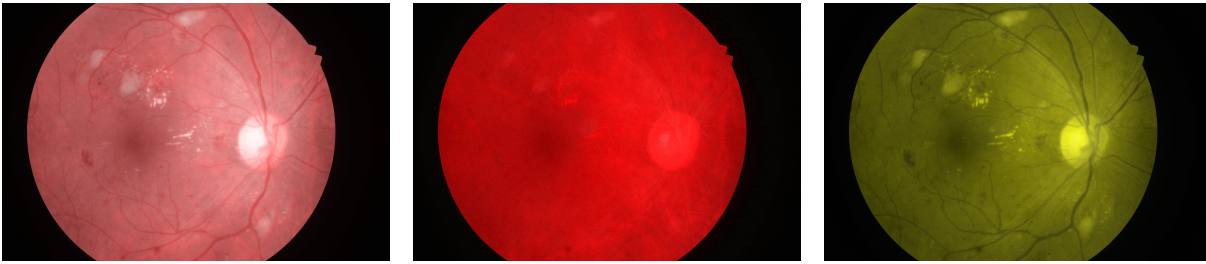
A segunda etapa do pré-processamento foi a separação das imagens em seus canais de cores. As Figuras 23 e 24 apresentam os resultados das imagens e suas combinações. Na Figura 25 é apresentado o *ground truth* com todos os pontos de interesse somados da referida imagem para análise qualitativa (visual). A imagem utilizada é a de número 59 do banco de dados IDRiD e foi escolhida por conter todas as estruturas de interesse.

Figura 23 – Canal individual: vermelho (R), verde (G) e azul (B), respectivamente.



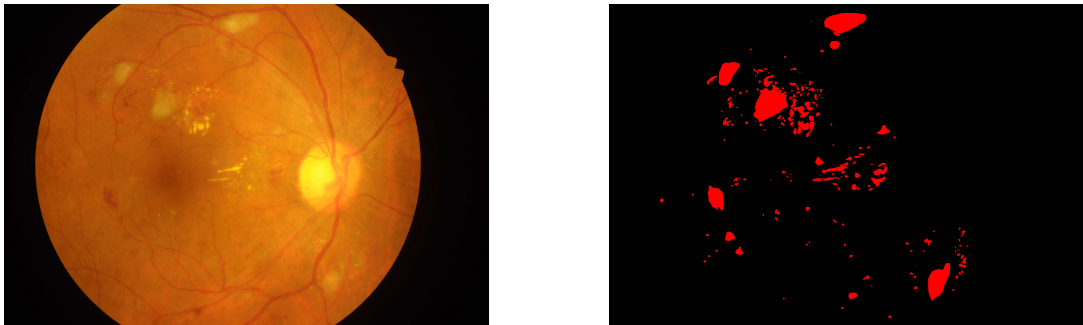
Fonte: Imagem de número 59 do banco de dados IDRiD, processada pelo autor.

Figura 24 – Canal duplo: vermelho e verde (RG), vermelho e azul (RB) e verde e azul (GB), respectivamente.



Fonte: Imagem de número 59 do banco de dados IDRiD, processada pelo autor.

Figura 25 – Imagem original com os três canais de cores (RGB) e *ground truth* com todos os pontos de interesse somados, respectivamente.



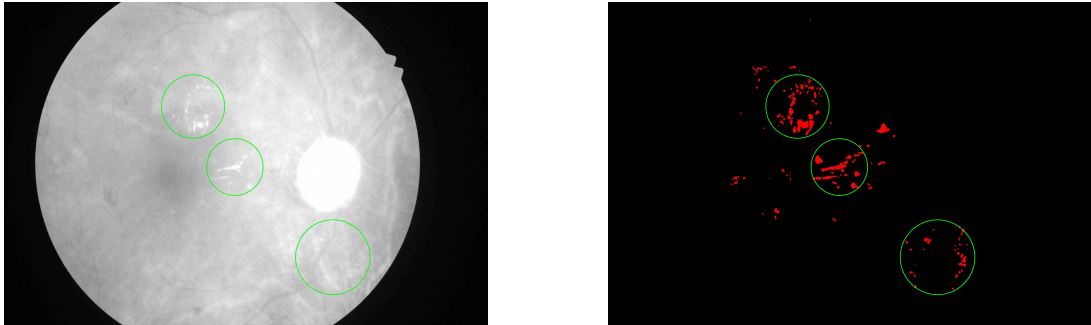
Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

A Figura 23 apresenta, comparativamente, os resultados para cada canal de cor RGB e permite observar individualmente a contribuição de cada canal para a construção da imagem e suas respectivas estruturas de interesse. A Figura 24 apresenta as possíveis combinações de canais de cores, ou seja, canais RG, RB e GB. Por fim, a Figura 25 apresenta a construção completa da imagem com todos os canais RGB e seu respectivo *ground truth*, sendo possível observar as diferentes formas de estruturas alvo de segmentação. A seguir, serão apresentadas as análises qualitativas dos resultados obtidos, destacando as estruturas de interesse em cada canal de cor.

#### **Exsudatos duros:**

- Canal vermelho: Por meio do canal vermelho é possível identificar visualmente as maiores estruturas dos exsudatos duros. Entretanto, como essas estruturas são mais claras, seu contraste é reduzido visto que toda a imagem é clara. Também é possível observar que, nesse canal de cores, o disco óptico apresenta tonalidade semelhante às estruturas de interesse, mas se diferencia pelo seu formato e maior dimensão. Numericamente, o pixel (2045, 1455), correspondente à estrutura de exsudato duro, possui valor 255, enquanto o pixel (2045, 1354), correspondente à região da retina saudável, apresenta valor 217.

Figura 26 – Canal vermelho e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal verde: Este canal apresenta maior contraste dos exsudatos duros em comparação ao canal vermelho. Nesta imagem, foi possível identificar visualmente todas as estruturas independentemente do tamanho ou formato. Numericamente, o pixel (2045, 1455), correspondente à estrutura de exsudato duro, possui valor 214, enquanto o pixel (2045, 1354), correspondente à região da retina saudável, apresenta valor 120.

Figura 27 – Canal verde e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal azul: Este canal apresenta pouca contribuição na formação da imagem, como é possível observar na Figura 28. Em uma inspeção visual, não foi possível observar nenhuma estrutura dos exsudatos duros. Numericamente, o pixel (2045, 1455), correspondente à estrutura de exsudato duro, possui valor 4, enquanto o pixel (2045, 1354), correspondente à região da retina saudável, apresenta valor 17.

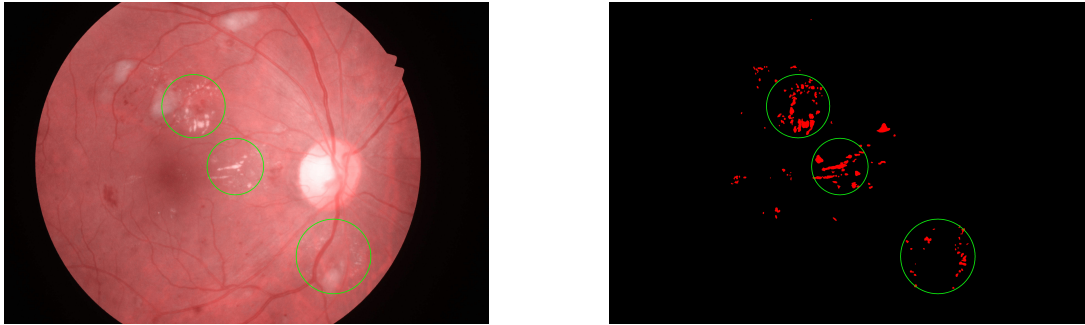
Figura 28 – Canal azul e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e verde: A associação dos canais vermelho e verde aumentou o contraste das maiores estruturas de exsudatos duros dentre as combinações avaliadas. Entretanto, reduziu o contraste das estruturas menores.

Figura 29 – Canais vermelho e verde na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e azul: Reduziu o contraste dos exsudatos duros, quando comparado ao canal vermelho. Em relação à estruturas menores, estas foram difíceis de serem localizadas por meio de inspeção visual.

Figura 30 – Canais vermelho e azul na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais verde e azul: Por meio de inspeção visual, é possível identificar todos os exsudatos independentemente de seu tamanho e formato. Entretanto, visualmente, foi prejudicado pela adição do canal azul.

Figura 31 – Canais verde e azul na mesma imagem e respectivo *ground truth*.



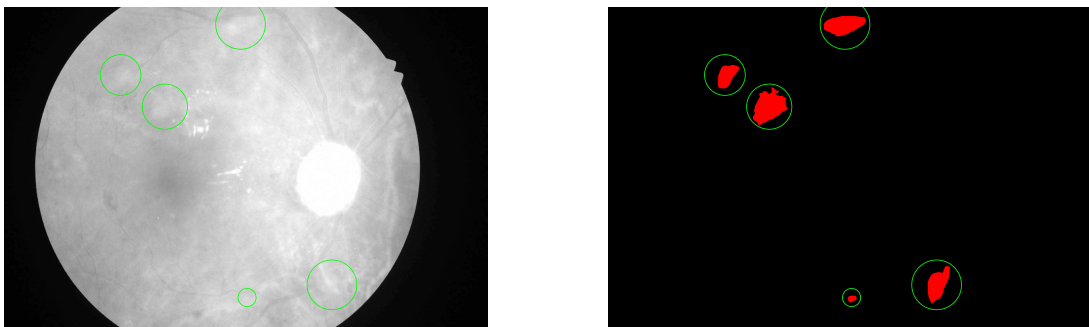
Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

As análises anteriormente descritas permitem concluir que para exsudatos duros, o canal verde apresentou, qualitativamente, por inspeção visual, o melhor resultado.

### Exsudatos moles:

- Canal vermelho: Por meio do canal vermelho, os exsudatos moles não são facilmente visualizados, seu contorno não se apresenta bem definido. Numericamente, o pixel (1050, 600), correspondente à estrutura de exsudato mole, possui valor 186, enquanto o pixel (1050, 340), correspondente à região da retina saudável, apresenta valor 167.

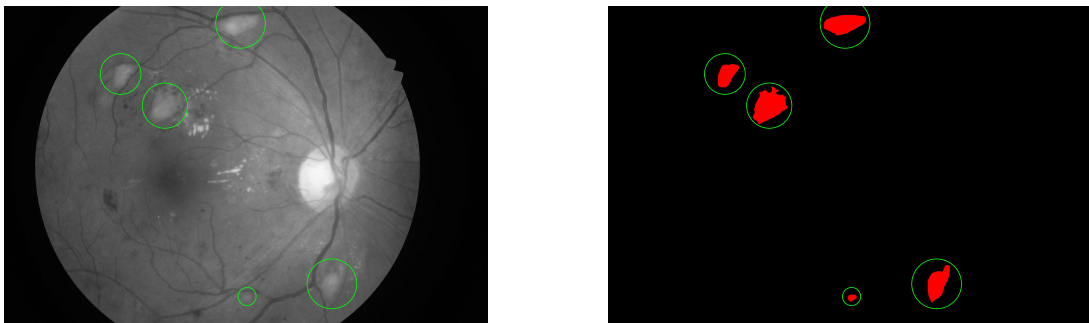
Figura 32 – Canais vermelho e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal verde: Este canal apresenta maior contraste dos exsudatos moles se comparado ao canal vermelho. Os exsudatos moles não possuem bordas bem definidas neste canal, mas todas as regiões podem ser identificadas por inspeção visual. Numericamente, o pixel (1050, 600), correspondente à estrutura de exsudato mole, possui valor 89, enquanto o pixel (1050, 340), correspondente à região da retina saudável, apresenta valor 73.

Figura 33 – Canal verde e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal azul: Apesar da pouca contribuição do canal azul na formação da imagem, os exsudatos moles são as estruturas de interesse mais visíveis nesta imagem, mas apresentam menos contraste que a imagem de canal verde. Este canal de cores pode ser utilizado, portanto, para diferenciar as estruturas. Numericamente, o pixel (1050,

600), correspondente à estrutura de exsudato mole, possui valor 12, enquanto o pixel (1050, 340), correspondente à região da retina saudável, apresenta valor 3.

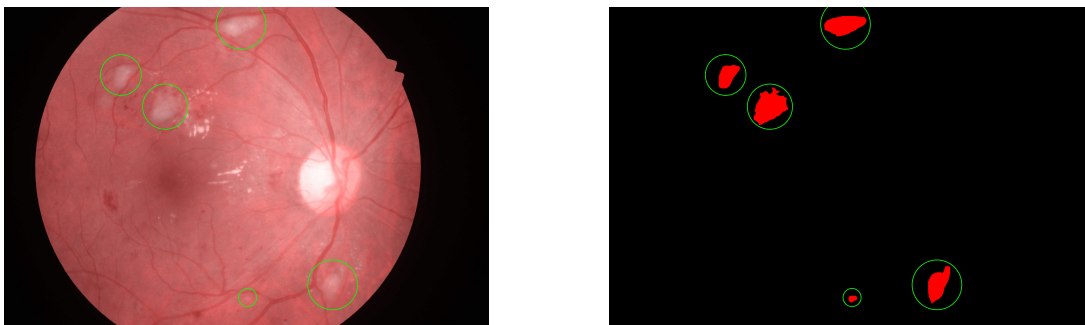
Figura 34 – Canal azul e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e verde: A associação destes canais aumentou o contraste das maiores estruturas de exsudatos moles e melhor definiu a borda destas estruturas, quando comparado com o canal vermelho. Entretanto, sua associação não representou contribuição adicional em comparação ao uso individual do canal verde.

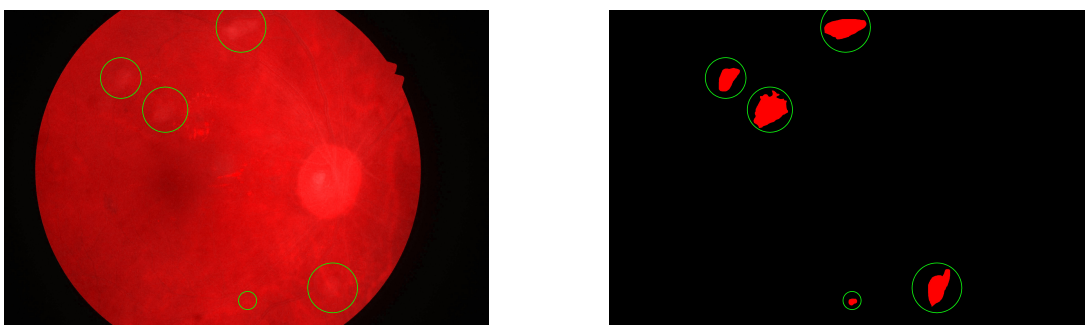
Figura 35 – Canais vermelho e verde na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e azul: Esta associação reduziu o contraste dos exsudatos moles quando comparados ao canal verde. Mesmo sendo grandes estruturas, sua localização por inspeção visual foi prejudicada.

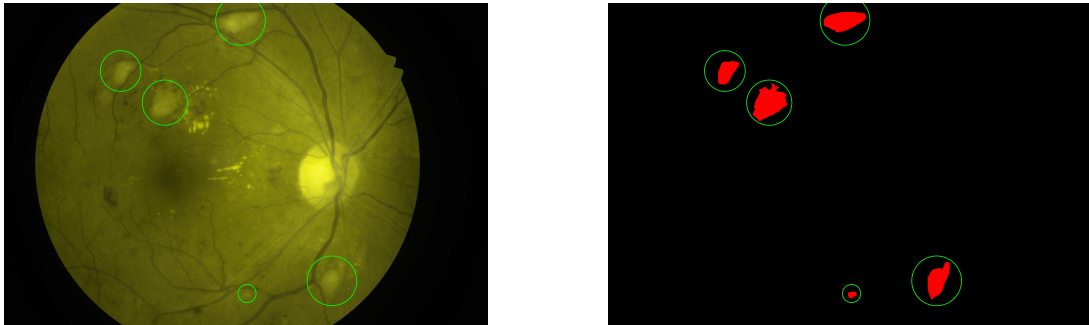
Figura 36 – Canais vermelho e azul na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais verde e azul: Por meio de inspeção visual, é possível identificar todos os exsudatos moles, independentemente de seu tamanho e formato. Esta imagem apresentou melhor definição da borda que os componentes individuais.

Figura 37 – Canais verde e azul na mesma imagem e respectivo *ground truth*.



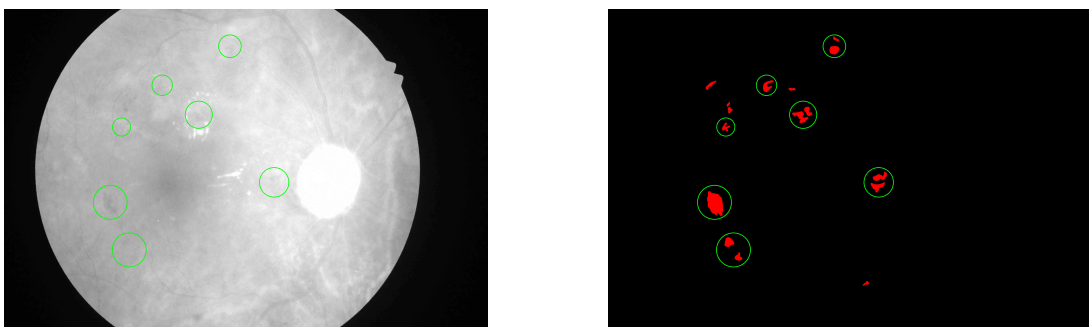
Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

As análises anteriormente descritas permitem concluir que, para exsudatos moles, o canal verde apresentou, qualitativamente, por inspeção visual, o melhor resultado.

### Hemorragias:

- Canal vermelho: Por meio deste canal, as hemorragias não são facilmente identificadas, apresentando cor mais escura que a região saudável da retina, porém, com baixo contraste. As menores estruturas tornaram-se visualmente imperceptíveis. Numericamente, o pixel (935, 1700), correspondente à região com hemorragia, possui valor 203, enquanto o pixel (935, 1500), correspondente à região da retina saudável, apresenta valor 192.

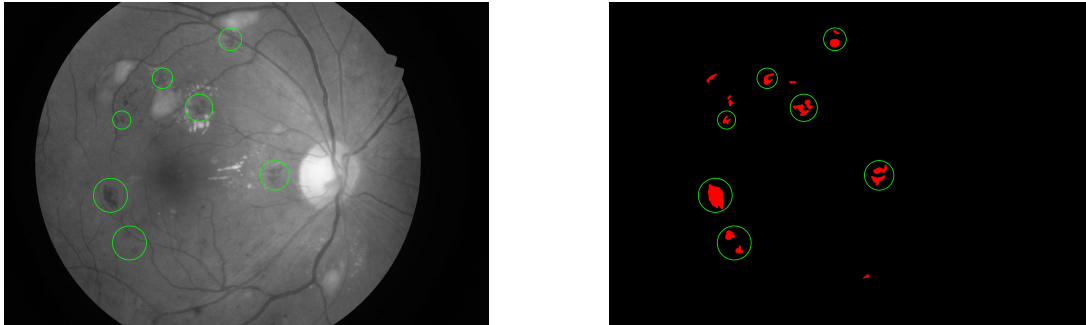
Figura 38 – Canal vermelho e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal verde: Este canal apresenta maior contraste das hemorragias em comparação ao canal vermelho. Apesar de não possuir borda bem definida, todas as regiões, neste canal, podem ser identificadas como regiões mais escuras. Numericamente, o pixel (935, 1700), correspondente à região com hemorragia, possui valor 91, enquanto o pixel (935, 1500), correspondente à região da retina saudável, apresenta valor 115.

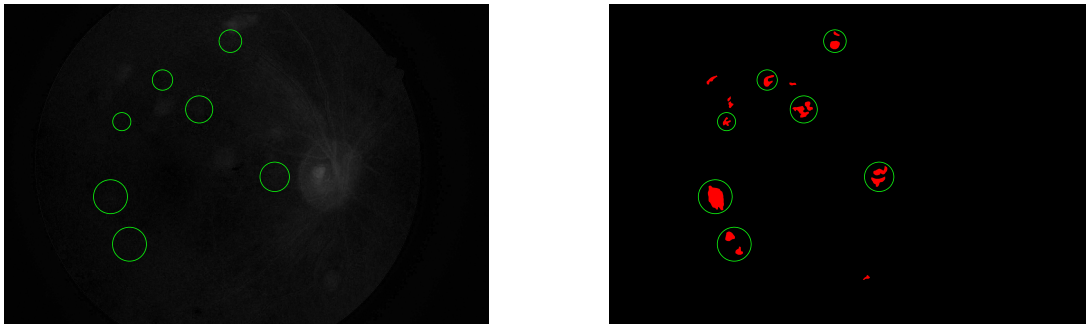
Figura 39 – Canal verde e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal azul: Este canal apresenta pouca contribuição na formação da imagem, assim como é possível observar na Figura 40. Em uma inspeção visual, não foi possível observar nenhuma estrutura de hemorragias. Numericamente, o pixel (935, 1700), correspondente à região com hemorragia, possui valor 19, enquanto o pixel (935, 1500), correspondente à região da retina saudável, apresenta valor 14.

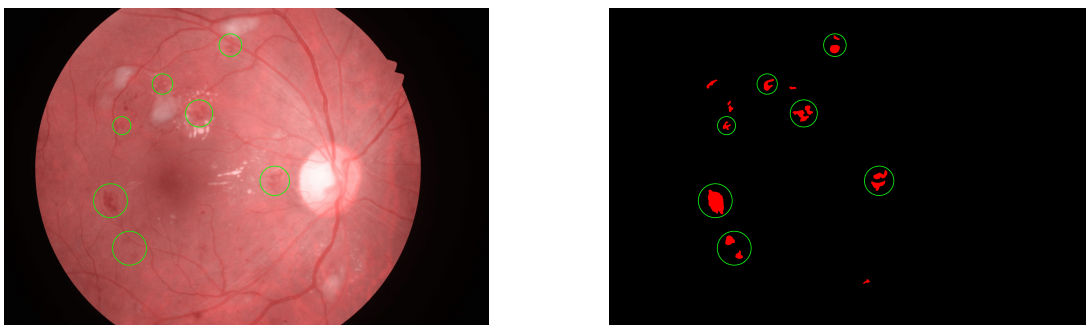
Figura 40 – Canal azul e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e verde: A associação entre estes canais reduziu o contraste das maiores estruturas quando comparado ao canal individual verde, além de dificultar a definição das bordas das estruturas de hemorragias.

Figura 41 – Canais vermelho e verde na mesma imagem e respectivo *ground truth*.

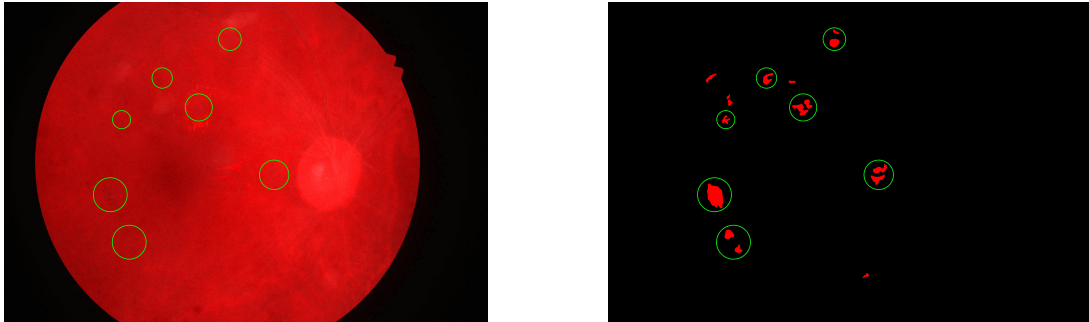


Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.



- Canais vermelho e azul: A associação entre estes canais reduziu o contraste de hemorragias em comparação ao canal vermelho. Apenas as maiores estruturas passaram a ser identificadas.

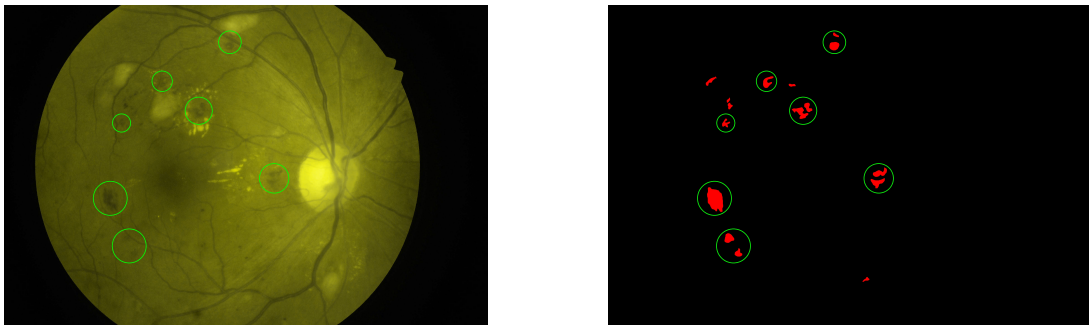
Figura 42 – Canais vermelho e azul na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais verde e azul: Por meio de inspeção visual, é possível identificar todas as hemorragias independentemente de seu tamanho e formato. Além disso, estas estruturas apresentaram melhor definição de suas bordas neste canal de cores.

Figura 43 – Canais verde e azul na mesma imagem e respectivo *ground truth*.

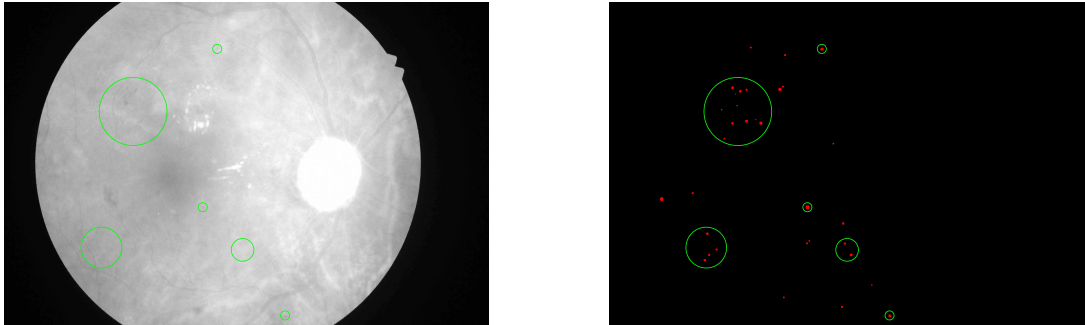


Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

### Microaneurismas:

- Canal vermelho: Por meio deste canal, os microaneurismas não são facilmente identificadas, apresentando cor mais escura que a região saudável da retina, porém, com baixo contraste. Numericamente, o pixel (1760, 1810), correspondente às estruturas de microaneurismas, possui valor 197, enquanto o pixel (1760, 1700), correspondente à região da retina saudável, apresenta valor 187.

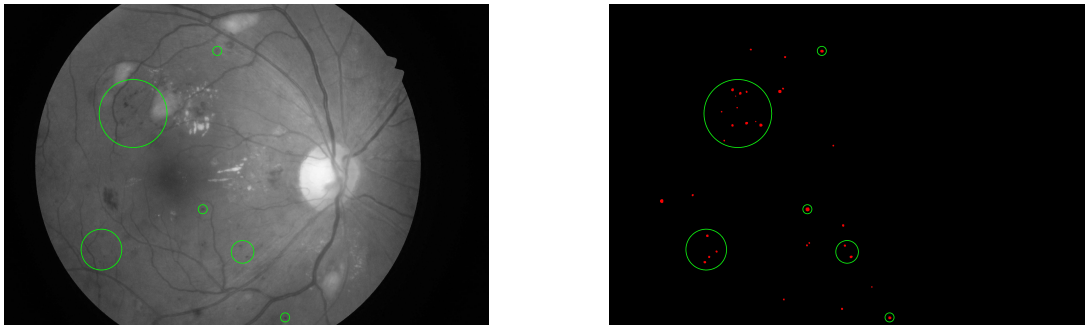
Figura 44 – Canal vermelho e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal verde: Este canal apresenta maior contraste dos microaneurismas que o canal vermelho. Todos os pontos com estas estruturas são escuros e podem ser identificados no canal de cor verde. Numericamente, o pixel (1760, 1810), correspondente às estruturas de microaneurismas, possui valor 97, enquanto o pixel (1760, 1700), correspondente à região da retina saudável, apresenta valor 90.

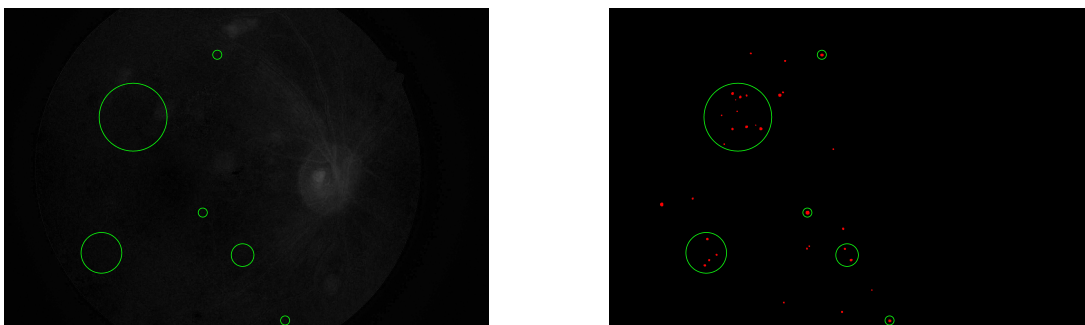
Figura 45 – Canal verde e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canal azul: Este canal apresenta pouca contribuição na formação da imagem, como é possível observar na Figura 46. Em uma inspeção visual, não foi possível observar nenhuma estrutura de microaneurisma. Numericamente, o pixel (1760, 1810), correspondente às estruturas de microaneurismas, possui valor 9, enquanto o pixel (1760, 1700), correspondente à região da retina saudável, apresenta valor 13.

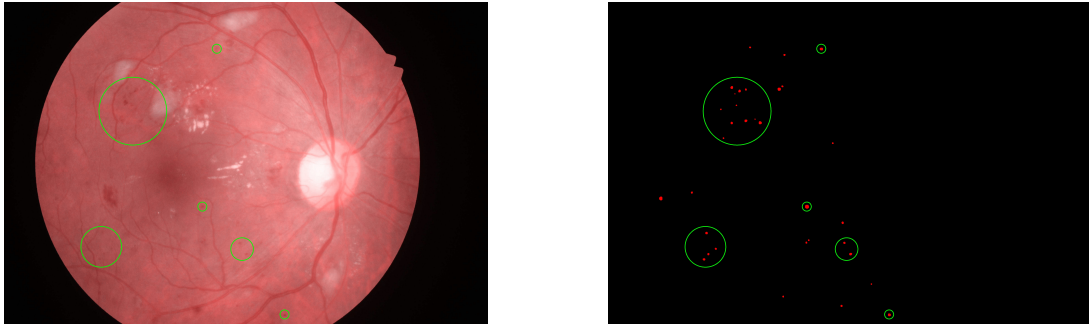
Figura 46 – Canal azul e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e verde: A associação destes canais permitiu identificar todas as estruturas. Em uma inspeção visual, houve maior facilidade em se encontrar esta estrutura em comparação ao uso individual do canal verde.

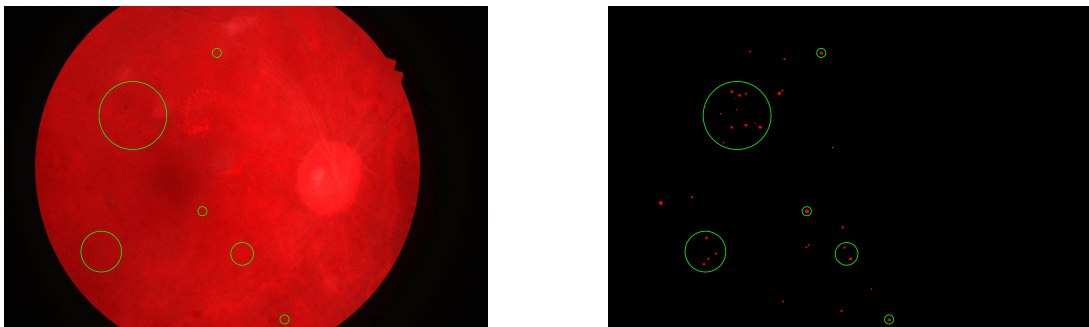
Figura 47 – Canais vermelho e verde na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais vermelho e azul: Em uma inspeção visual, não foi possível observar nenhuma estrutura de microaneurisma. Estas estruturas são escuras e, nesta combinação de canais de cores, apresentam baixa resolução.

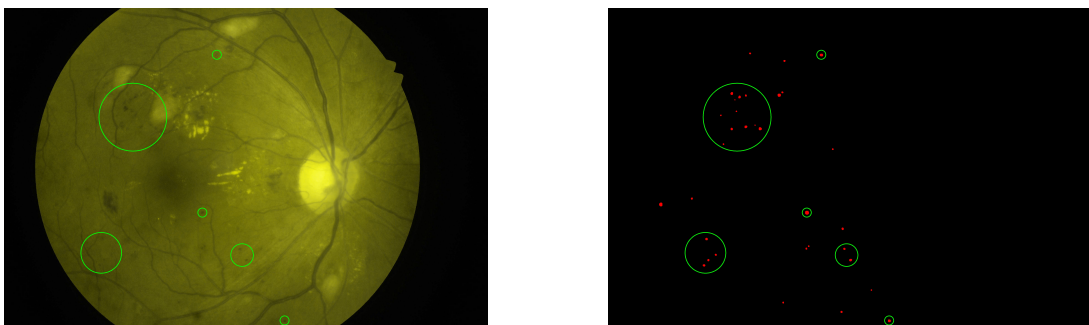
Figura 48 – Canais vermelho e azul na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

- Canais verde e azul: Por meio de inspeção visual é possível identificar a maioria dos microaneurismas. Entretanto, os microaneurismas de tamanho reduzido não se destacam nesta combinação de canais de cores.

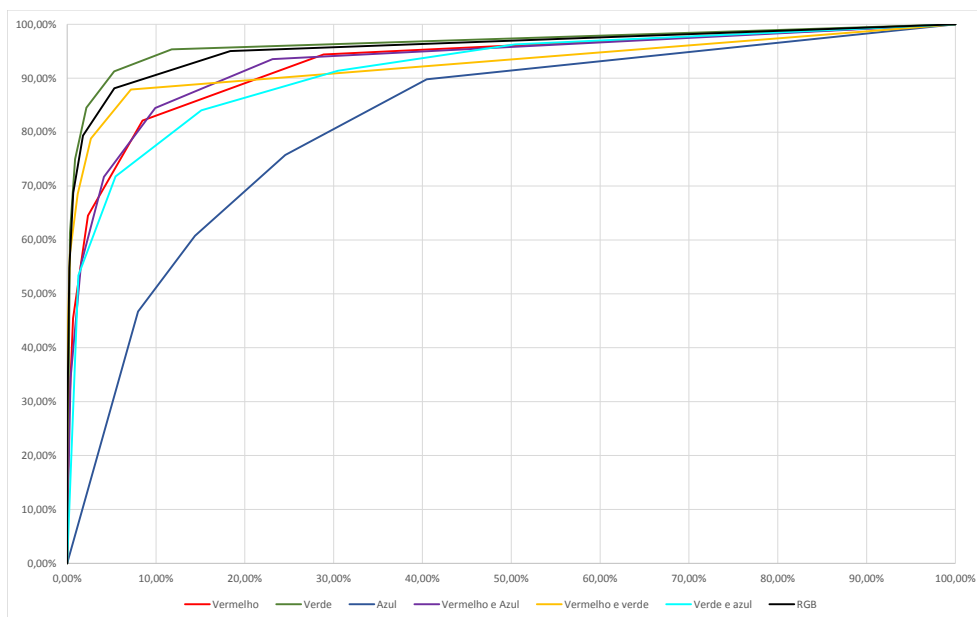
Figura 49 – Canais verde e azul na mesma imagem e respectivo *ground truth*.



Fonte: Imagem de número 59 do banco de dados IDRiD, adaptada pelo autor.

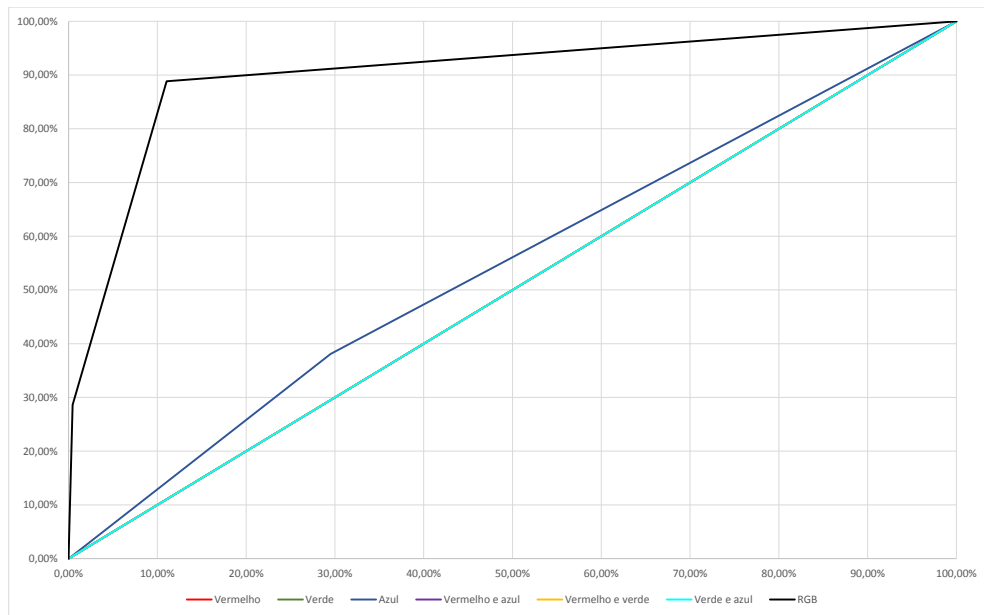
A seguir, serão apresentados os resultados quantitativos obtidos para as diferentes combinações de canais de cores. Estas combinações foram utilizadas para treinamento e testes da rede *U-Net* proposta. A métrica de avaliação utilizada foi a curva ROC e AUC.

Figura 50 – Curva ROC - Exsudatos Duros com seus respectivos canais de cores



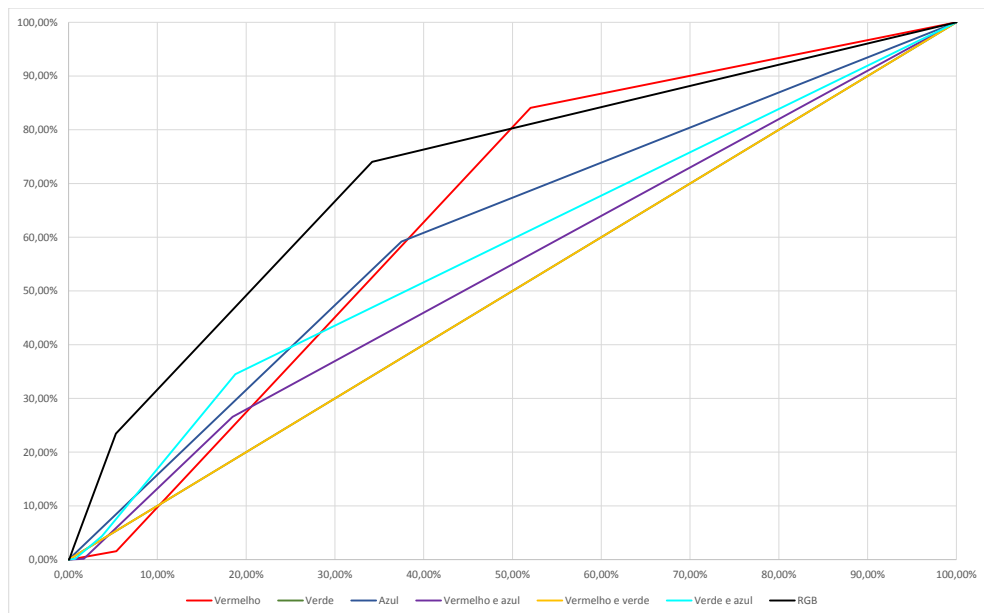
Fonte: Gráfico elaborado pelo autor.

Figura 51 – Curva ROC - Exsudatos Moles com seus respectivos canais de cores



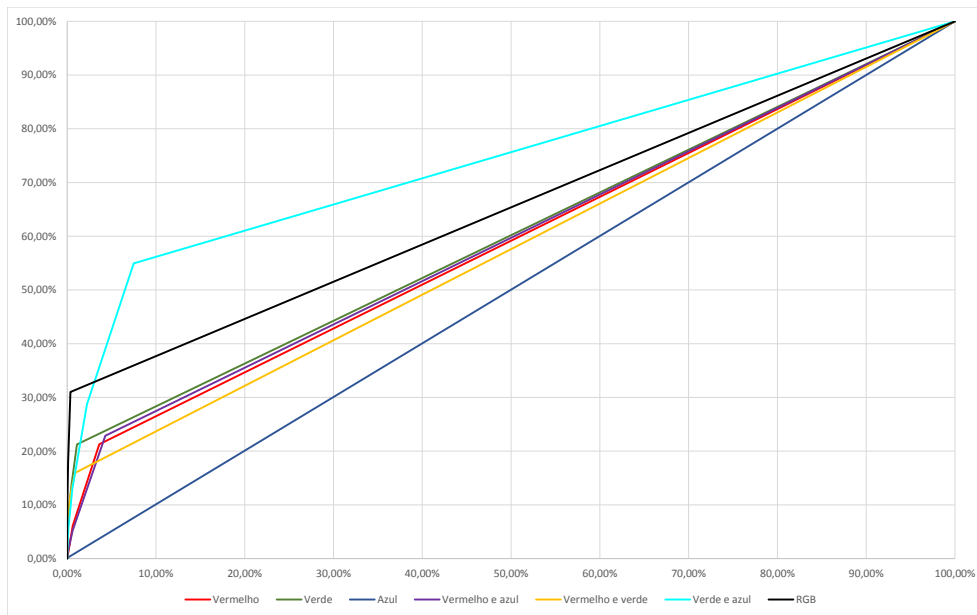
Fonte: Gráfico elaborado pelo autor.

Figura 52 – Curva ROC - Hemorragias com seus respectivos canais de cores



Fonte: Gráfico elaborado pelo autor.

Figura 53 – Curva ROC - Microaneurismas com seus respectivos canais de cores



Fonte: Gráfico elaborado pelo autor.

A Tabela 5 sumariza os resultados, apresentando os valores de AUC. Observa-se que as estruturas apresentaram o melhor resultado com diferentes canais de cores, ou seja, ED apresentou melhor resultado com o canal verde, EM com o canal azul, HE com os canais RGB e MA com os canais verde e azul. Estes resultados motivaram a utilização dos canais RGB em conjunto para obter o melhor resultado das estruturas somadas em uma mesma imagem.

Tabela 5 – Resultado da AUC para os componentes de canais de cores e suas combinações

	Vermelho	Verde	Azul	Vermelho e verde	Vermelho e azul	Verde e azul	RGB
ED	92,8%	<b>96,5%</b>	81,9%	92,8%	92,9%	91,3%	95,7%
EM	50,0%	50,0%	54,3%	50,0%	50,0%	50,0%	<b>90,3%</b>
HE	64,2%	50,0%	60,9%	53,8%	50,0%	57,6%	<b>72,0%</b>
MA	58,9%	60,1%	50,0%	57,6%	59,3%	<b>74,3%</b>	65,3%

A Tabela 6 apresenta os resultados otimizados para a métrica IoU. A Tabela 7 apresenta os resultados otimizados para as combinações das estruturas. De forma análoga, as Tabelas 8, 10, 12 e 14 apresentam os melhores resultados, buscando a maximização dos parâmetros acurácia, sensibilidade, precisão e especificidade, respectivamente. Nas Tabelas 9, 11, 13 e 15 são apresentados os melhores resultados para todas as estruturas combinadas.

Todos os resultados otimizados foram obtidos tendo como objetivo, em segunda instância, a maior média percentual dos parâmetros (acurácia, sensibilidade, precisão e especificidade) para cada estrutura. Sem esta consideração, a otimização de uma métrica poderia levar a zero outras métricas.

Tabela 6 – Resultados das estruturas separadas otimizando a métrica IoU

	<i>Threshold</i>	IoU	Acurácia	Sens.	Precisão	Especif.	Média
ED	94	54,12%	99,03%	44,76%	72,21%	99,77%	78,94%
EM	9	15,68%	99,17%	16,44%	4,23%	99,32%	54,79%
HE	3	17,95%	98,30%	31,41%	29,45%	99,10%	64,56%
MA	13	9,33%	99,83%	14,76%	20,22%	99,93%	58,69%

Tabela 7 – Resultados das estruturas combinadas otimizando a métrica IoU

	<i>Threshold</i>	IoU	Acurácia	Sens.	Precisão	Especif.	Média
ED & EM	37	59,92%	99,00%	52,00%	74,38%	99,72%	81,28%
ED & HE	36	37,71%	97,52%	34,67%	50,84%	99,14%	70,54%
ED & MA	38	53,60%	98,88%	42,66%	67,99%	99,70%	77,31%
EM & HE	4	10,26%	98,13%	23,33%	27,71%	99,16%	62,08%
EM & MA	1	2,33%	99,47%	2,48%	3,09%	99,77%	51,20%
HE & MA	1	12,46%	98,48%	23,60%	36,83%	99,47%	64,60%
Todas	15	45,01%	97,51%	37,49%	58,86%	99,24%	73,27%

Tabela 8 – Resultados das estruturas separadas otimizando a métrica acurácia

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
Exsudatos duros	94	99,03%	44,76%	72,21%	99,77%	78,94%
Exsudatos moles	54	99,75%	2,38%	5,67%	99,93%	51,93%
Hemorragias	8	98,50%	27,54%	33,39%	99,34%	64,69%
Microaneurismas	8	99,87%	9,63%	35,47%	99,98%	61,24%

Tabela 9 – Resultados das estruturas combinadas otimizando a métrica acurácia

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
ED & EM	39	99,00%	51,51%	74,99%	99,74%	81,31%
ED & HE	37	97,52%	34,67%	50,84%	99,14%	70,54%
ED & MA	38	98,88%	42,66%	67,99%	99,70%	77,31%
EM & HE	8	98,13%	23,33%	27,71%	99,16%	62,08%
EM & MA	8	99,47%	2,48%	3,09%	99,77%	51,20%
HE & MA	8	98,48%	23,60%	36,83%	99,47%	64,60%
Todas	15	97,51%	37,49%	58,86%	99,24%	73,27%

Tabela 10 – Resultados das estruturas separadas otimizando a métrica sensibilidade

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
Exsudatos duros	8	79,35%	79,75%	4,96%	79,34%	60,85%
Exsudatos moles	4	97,77%	31,03%	2,62%	97,89%	57,33%
Hemorragias	8	93,57%	40,92%	7,77%	94,20%	59,11%
Microaneurismas	8	99,81%	18,71%	18,51%	99,90%	59,23%

Tabela 11 – Resultados das estruturas combinadas otimizando a métrica sensibilidade

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
ED & EM	8	98,71%	63,43%	56,59%	99,25%	79,49%
ED & HE	8	94,27%	62,16%	24,64%	95,10%	69,04%
ED & MA	8	98,64%	50,90%	53,10%	99,34%	75,50%
EM & HE	3	81,80%	45,43%	3,42%	82,30%	53,24%
EM & MA	4	98,73%	20,21%	5,57%	98,97%	55,87%
HE & MA	8	96,08%	45,28%	15,45%	96,74%	63,39%
Todas	8	95,56%	48,27%	31,24%	96,93%	68,00%

Tabela 12 – Resultados das estruturas separadas otimizando a métrica precisão

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
Exsudatos duros	88	98,98%	36,50%	74,11%	99,83%	77,36%
Exsudatos moles	54	99,75%	2,38%	5,67%	99,93%	51,93%
Hemorragias	8	98,50%	27,54%	33,39%	99,34%	64,69%
Microaneurismas	8	99,87%	9,63%	35,47%	99,98%	61,24%

Tabela 13 – Resultados das estruturas combinadas otimizando a métrica precisão

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
ED & EM	39	99,00%	51,51%	74,99%	99,74%	81,31%
ED & HE	37	97,52%	34,67%	50,84%	99,14%	70,54%
ED & MA	38	98,88%	42,66%	67,99%	99,70%	77,31%
EM & HE	8	98,13%	23,33%	27,71%	99,16%	62,08%
EM & MA	8	99,32%	11,37%	7,61%	99,59%	54,47%
HE & MA	8	98,48%	23,60%	36,83%	99,47%	64,60%
Todas	15	97,51%	37,49%	58,86%	99,24%	73,27%

Tabela 14 – Resultados das estruturas separadas otimizando a métrica especificidade

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
Exsudatos duros	51	98,77%	14,91%	67,48%	99,90%	70,27%
Exsudatos moles	53	99,75%	2,39%	5,63%	99,93%	51,92%
Hemorragias	8	98,50%	27,54%	33,39%	99,34%	64,69%
Microaneurismas	8	99,87%	9,63%	35,47%	99,98%	61,24%



Tabela 15 – Resultados das estruturas combinadas otimizando a métrica especificidade

	<i>Threshold</i>	Acurácia	Sens.	Precisão	Especif.	Média
ED & EM	39	99,00%	51,51%	74,99%	99,74%	81,31%
ED & HE	37	97,52%	34,67%	50,84%	99,14%	70,54%
ED & MA	8	98,85%	39,56%	67,84%	99,72%	76,50%
EM & HE	8	98,13%	23,33%	27,71%	99,16%	62,08%
EM & MA	8	99,47%	2,48%	3,09%	99,77%	51,20%
HE & MA	8	98,48%	23,60%	36,83%	99,47%	64,60%
Todas	15	97,51%	37,49%	58,86%	99,24%	73,27%

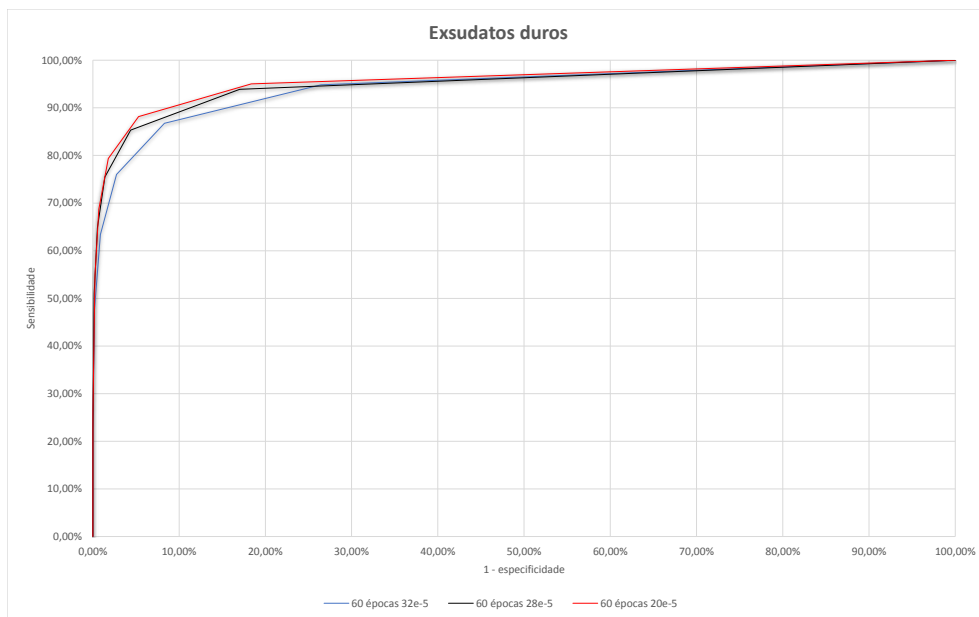
### 5.0.3 Terceira etapa: Análise individual e combinações das estruturas de interesse

Os Gráficos de 54 à 64 apresentam a curva ROC para cada estrutura analisada, apresentando os três melhores resultados para esta métrica. Na legenda, serão apresentados o número de épocas necessárias e o decaimento utilizado na rede *U-Net*. Os valores da AUC serão apresentados na Tabela 16 para os três melhores resultados, de diferentes treinamentos, abordados nos gráficos subsequentes.

Tabela 16 – AUC para resultados apresentados nos gráficos - Três melhores resultados

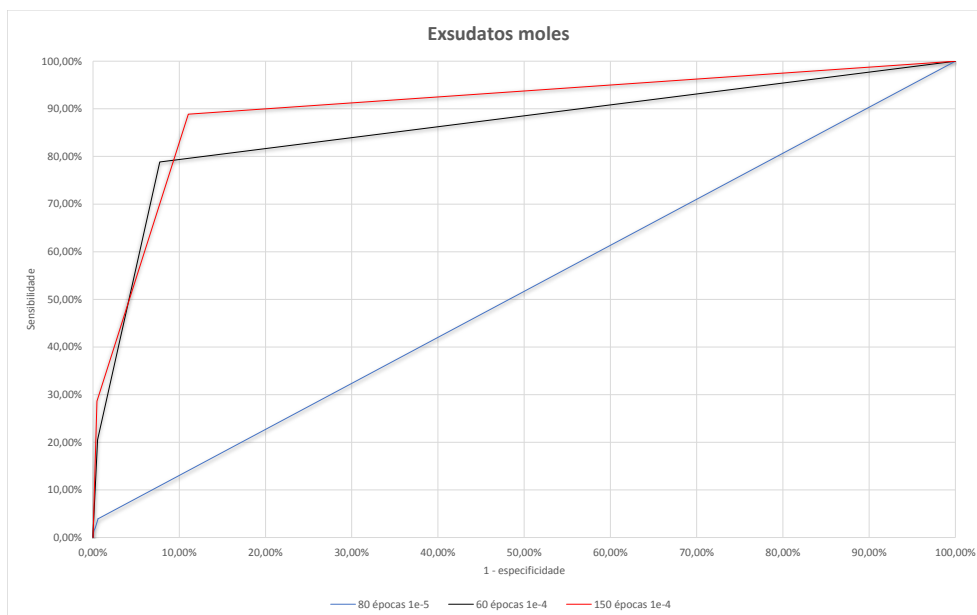
	1º AUC	2º AUC	3º AUC
ED	95,70 %	95,06 %	94,40 %
EM	90,29 %	86,13 %	51,67 %
HE	71,95 %	71,70 %	68,75 %
MA	65,34 %	62,00 %	59,74 %
ED & EM	86,37 %	85,36 %	84,60 %
ED & HE	88,58 %	80,85 %	77,66 %
ED & MA	89,44 %	82,98 %	80,33 %
EM & HE	65,49 %	62,58 %	61,88 %
EM & MA	63,10 %	62,16 %	59,60 %
HE & MA	75,17 %	70,80 %	67,22 %
Todas	78,60 %	78,35 %	77,95%

Figura 54 – Curva ROC: Três melhores resultados de exsudatos duros.



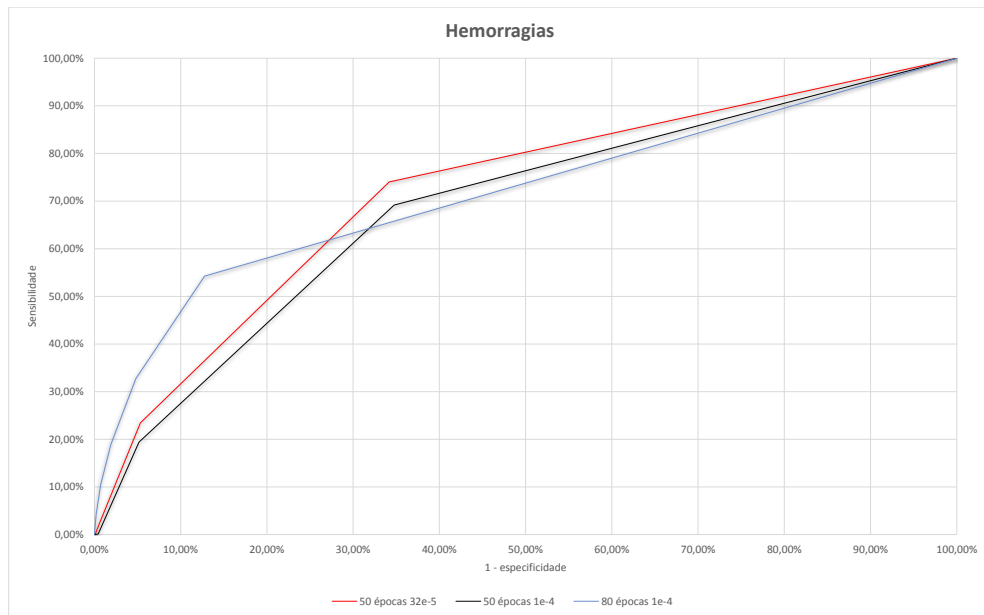
[H].

Figura 55 – Curva ROC: Três melhores resultados de exsudatos moles.



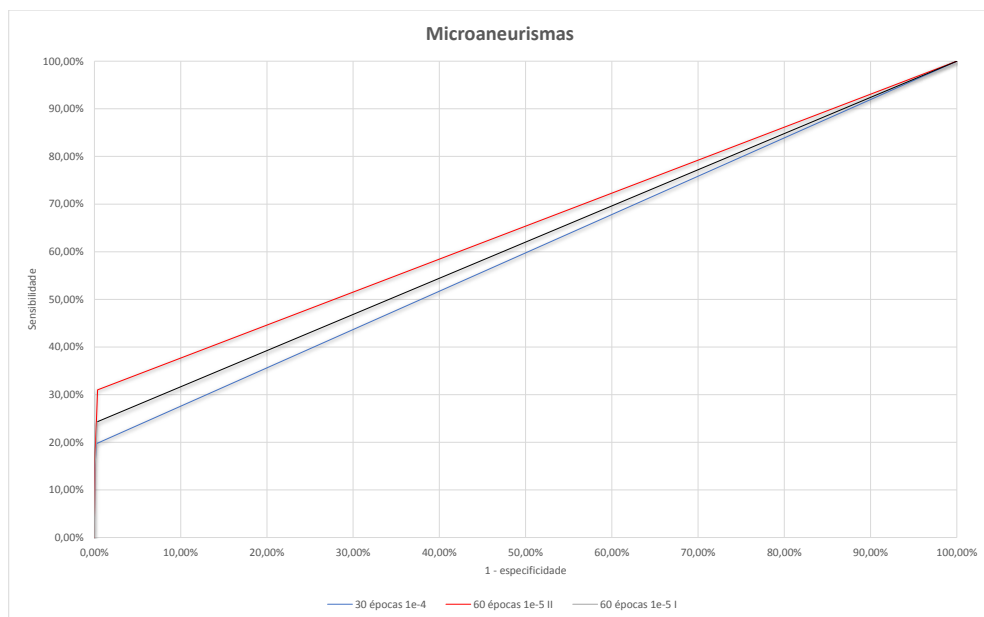
Fonte: Imagem elaborada pelo autor.

Figura 56 – Curva ROC: Três melhores resultados de hemorragias.



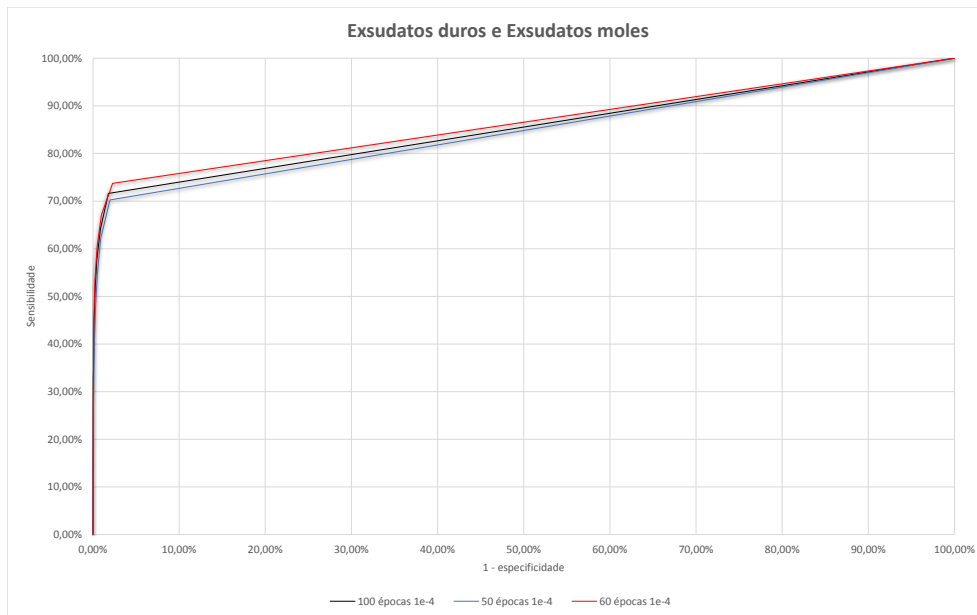
Fonte: Imagem elaborada pelo autor.

Figura 57 – Curva ROC: Três melhores resultados de microaneurismas.



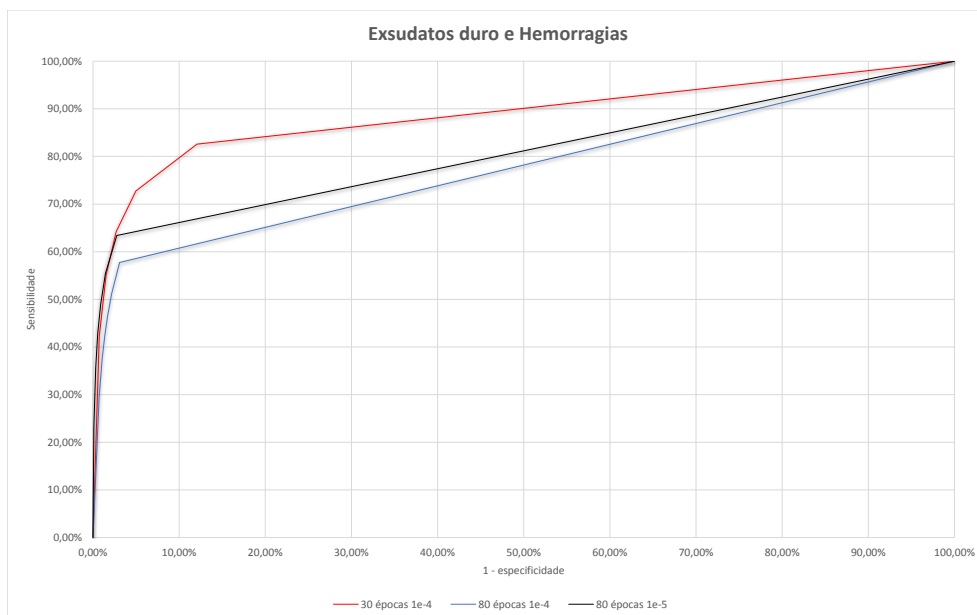
Fonte: Imagem elaborada pelo autor.

Figura 58 – Curva ROC: Três melhores resultados de exsudatos duros e exsudatos moles.



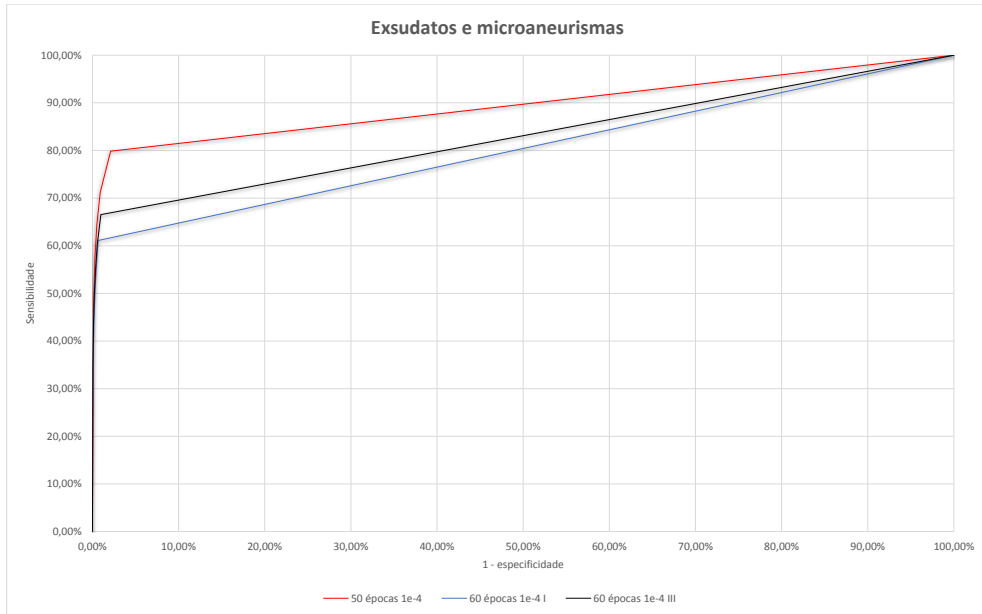
Fonte: Imagem elaborada pelo autor.

Figura 59 – Curva ROC: Três melhores resultados de exsudatos duros e hemorragias.



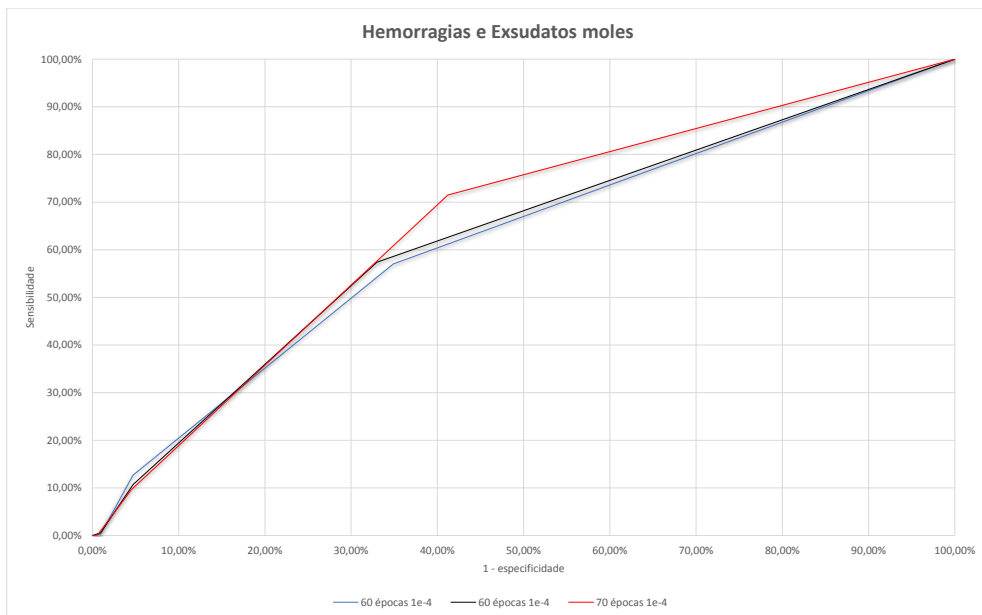
Fonte: Imagem elaborada pelo autor.

Figura 60 – Curva ROC: Três melhores resultados de exsudatos duros e microaneurismas.



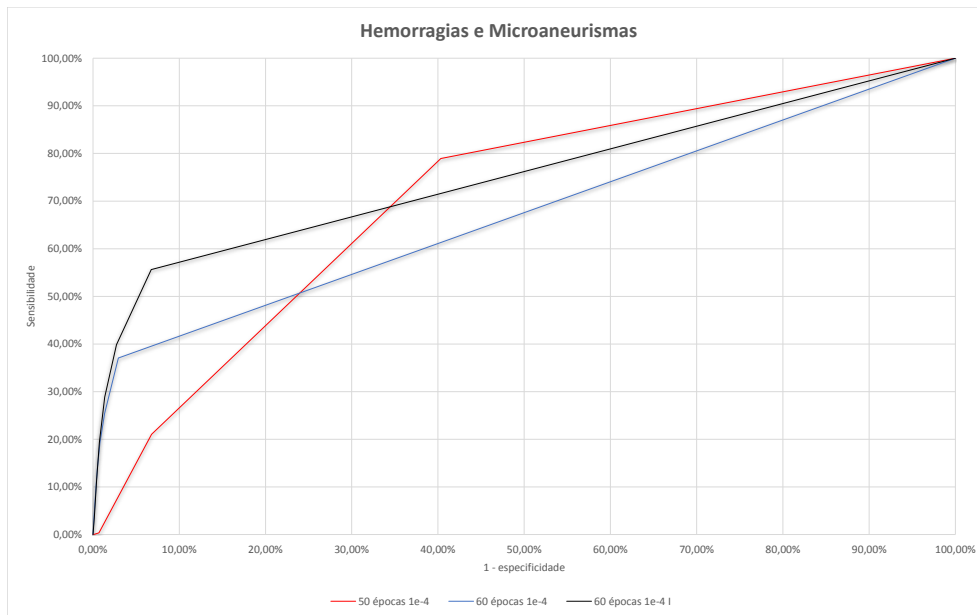
Fonte: Imagem elaborada pelo autor.

Figura 61 – Curva ROC: Três melhores resultados de hemorragias e exsudatos moles.



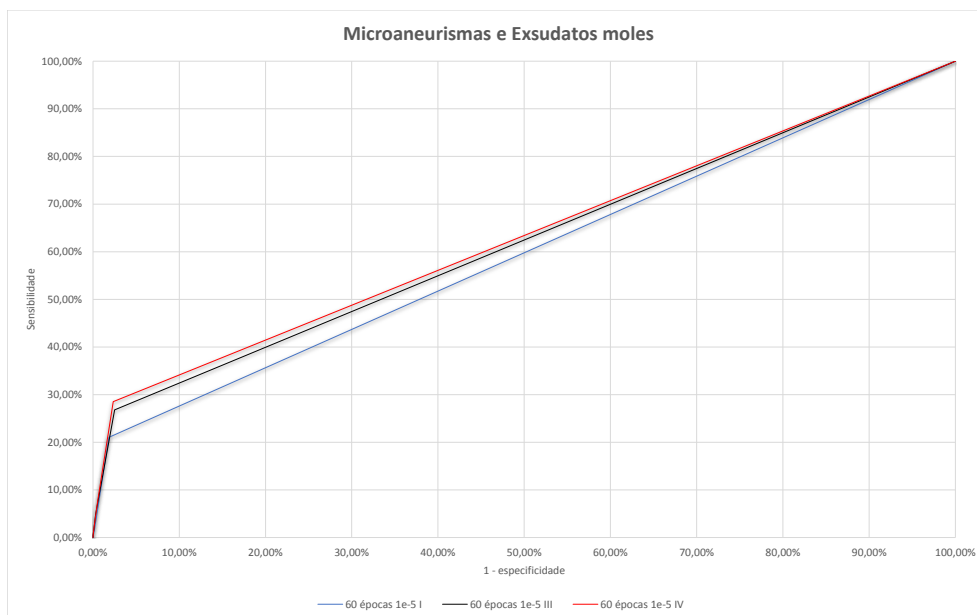
Fonte: Imagem elaborada pelo autor.

Figura 62 – Curva ROC: Três melhores resultados de hemorragias e microaneurismas.



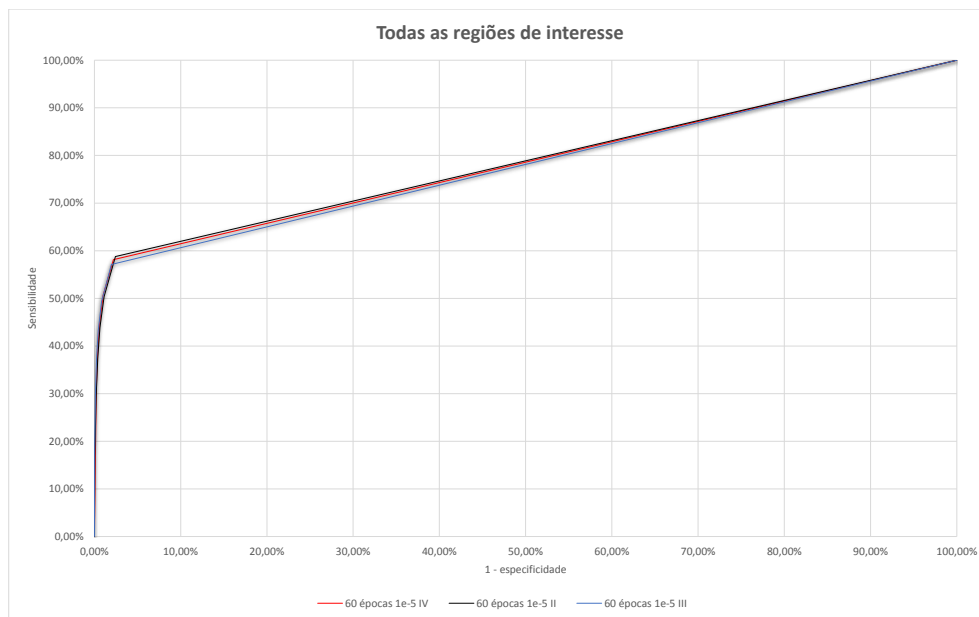
Fonte: Imagem elaborada pelo autor.

Figura 63 – Curva ROC: Três melhores resultados de microaneurismas e exsudatos moles.



Fonte: Imagem elaborada pelo autor.

Figura 64 – Curva ROC: Três melhores resultados de todas as estruturas somadas.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e ED: Dentre as estruturas analisadas separadamente, a rede teve melhor resultado ao identificar exsudatos duros. Entretanto, é possível afirmar que a presença do disco óptico prejudicou os resultados. Na predição, parte do disco óptico foi segmentado. Dessa forma, verifica-se que remover o disco óptico na etapa de pré-processamento pode auxiliar o treinamento e predição da rede *U-Net*.

Figura 65 – Exsudatos duros: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e EM: Nesta estrutura, pode-se afirmar que a predição localizou os três pontos de interesse, entretanto, seu formato e borda não foi idealmente segmentado.

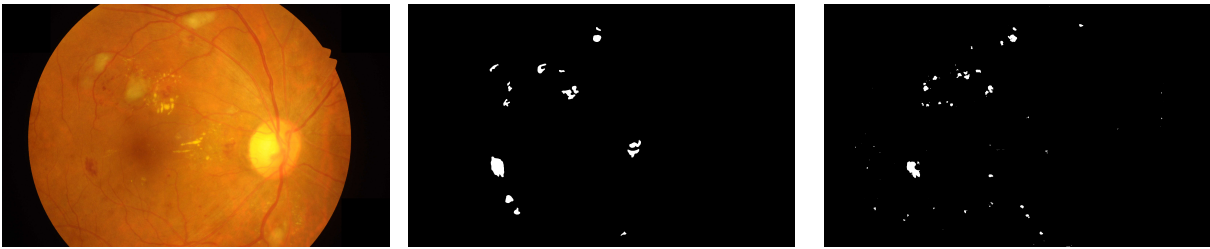
Figura 66 – Exsudatos moles: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e HE: Para detecção de hemorragias, ocorreram maior número de falsos positivos que nas demais estruturas. Nota-se que as maiores estruturas estão corretamente identificadas.

Figura 67 – Hemorragias: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e MA: Os microaneurismas são as menores estruturas. Visualmente, a predição acertou todos os pontos onde haviam estas estruturas. Entretanto, observa-se a presença de vários falsos positivos. A rede também apresentou dificuldade em determinar corretamente o contorno das estruturas.

Figura 68 – Microaneurismas: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e ED & EM: Ao associar exsudatos duros e moles, obteve-se, quantitativamente, melhor resultado do que as estruturas separadas. É possível observar que ao unir essas duas estruturas, não há segmentação do disco óptico na predição. Entretanto, os EM que se configuram como sendo estruturas maiores que ED, não foram detectados.



Figura 69 – Exsudatos duros e moles: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e ED & HE: Ao associar exsudatos duros e hemorragias, houve maior número de verdadeiros positivos, especialmente das estruturas de hemorragias. Entretanto, houve o surgimento de maior número de falsos positivos.

Figura 70 – Exsudatos duros e hemorragias: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e ED & MA: Nessa estrutura, uma pequena parte do disco óptico foi segmentada, prejudicando, assim, o resultado.

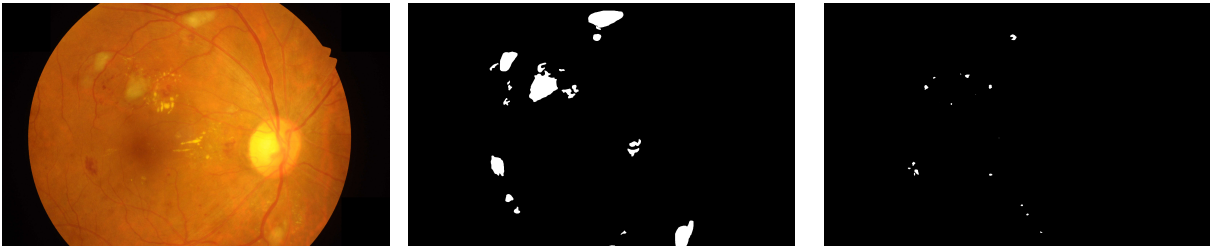
Figura 71 – Exsudatos duros e microaneurismas: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e HE & EM: Nesta combinação, a rede identificou algumas das estruturas de interesse, entretanto, suas dimensões e contornos não foram corretamente segmentadas. Observa-se maior dificuldade em segmentar as maiores estruturas.

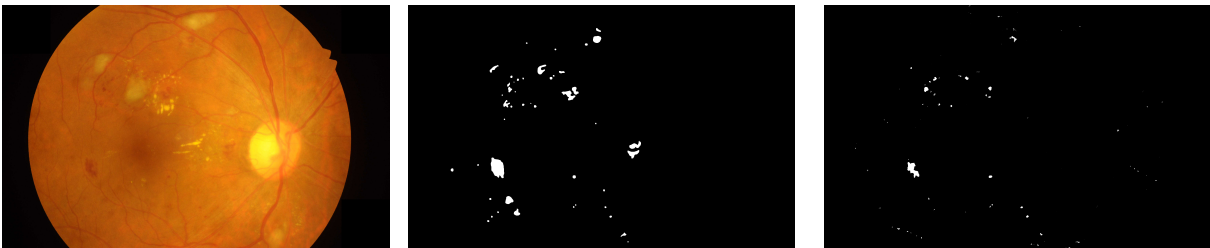
Figura 72 – Hemorragias e exsudatos moles: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e HE & MA: Nesta combinação, a rede identificou algumas das estruturas de interesse. Entretanto, suas dimensões e contornos não foram corretamente segmentadas.

Figura 73 – Hemorragias e microaneurismas: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

- Canais RGB e MA & EM: Nesta combinação a rede identificou apenas as menores estruturas de interesse, apresentando dificuldade em identificar as maiores estruturas.

Figura 74 – Microaneurismas e exsudatos moles: Imagem original, *ground truth* e predição, respectivamente.



Fonte: Imagem elaborada pelo autor.

Pode-se observar que a rede *U-Net* proposta conseguiu segmentar todos os tipos de estruturas de interesse, porém com resultados inferiores quando comparado com as estruturas individuais. Este fato decorre do treinamento simultâneo de estruturas com diferentes características, como formato, tamanho e intensidade. Os ED e EM são estruturas com características semelhantes e, desta forma, obteve-se um resultado superior quando

combinadas. Cabe ressaltar também que a subdivisão das imagens pode ter comprometido as maiores estruturas e, como consequência, prejudicado de modo considerável seu resultado nas métricas de avaliação.

A Tabela 17 apresenta os resultados obtidos buscando maximizar a métrica sensibilidade, comparativamente, ao fronte de pesquisa.

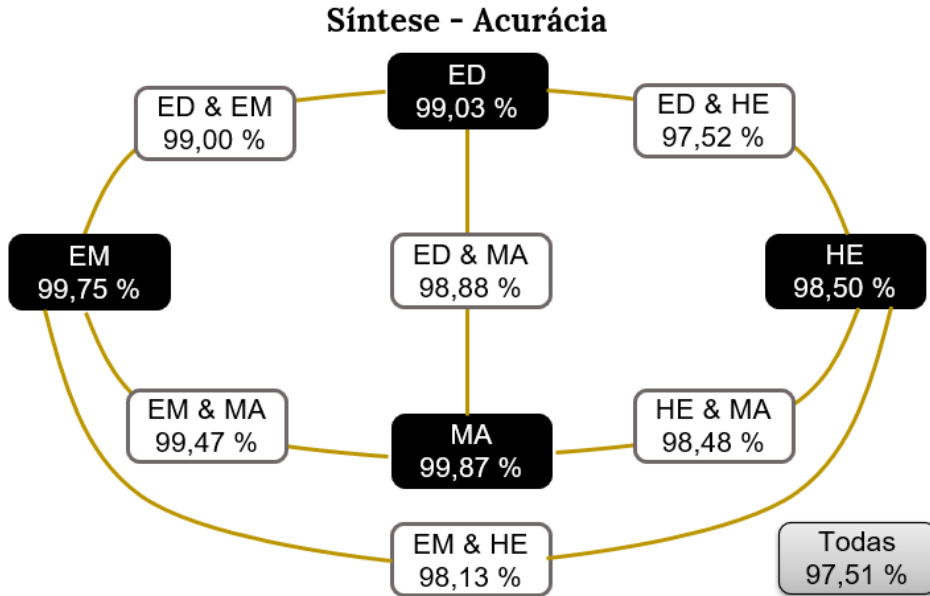
Tabela 17 – Resultados comparativos para cada estrutura individual.

Estrutura	Método	Acurácia	Especificidade	Sensibilidade
ED	Sambyal <i>et al.</i> (2020)	99,88%	99,88%	99,85%
	Método proposto	79,35%	79,34%	79,75%
EM	Haris <i>et al.</i> (2019)	90,46%	88,86%	80,57%
	Método proposto	97,77%	97,89%	31,03%
MA	Kou <i>et al.</i> (2020b)	94,20%	94,10%	91,78%
	Método proposto	99,81%	99,90%	18,71%
HE	Nazir <i>et al.</i> (2020)	95,20%	95,10%	96,0%
	Método proposto	93,57%	94,20%	40,92%

Conforme observado na Tabela 17, os resultados obtidos foram inferiores aos melhores resultados obtidos na literatura atual. Essas diferenças decorrem de melhores ajustes dos parâmetros de treinamentos e etapas de pré-processamento de maior custo computacional que auxiliaram as redes a obter melhores resultados, de acordo com as métricas apresentadas. Para as métricas de acurácia e especificidade os resultados foram comparáveis, enquanto para a métrica de sensibilidade houve um resultado consideravelmente inferior. Entretanto, os resultados obtidos neste trabalho, apesar de inferiores, permitem comparar estes com as associações das estruturas de interesse, informação esta não encontrada na literatura atual.

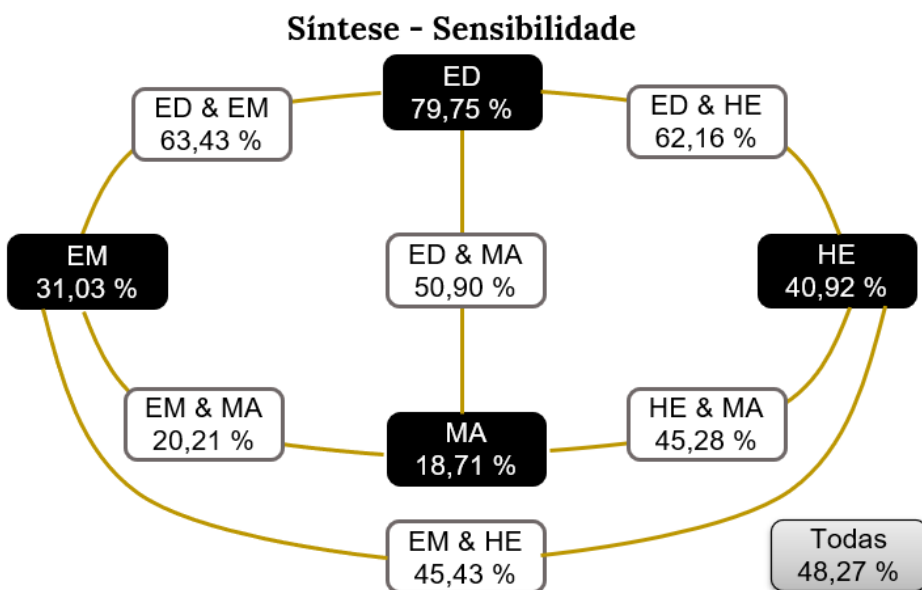
As Figuras 75, 76, 77, 78, 79 e 80 apresentam diagramas dos resultados das estruturas originais e combinações das estruturas de interesse referente as métricas de acurácia, sensibilidade, precisão, especificidade, AUC e IoU respectivamente. Nos diagramas as estruturas originais são apresentadas em destaque (ED, EM, HE e MA). Também são apresentados em destaque as combinações que geraram um resultado superior em comparação a análise das estruturas individuais. Observa-se que para a métrica de sensibilidade a associação de ED & EM gerou um resultado superior em comparação a avaliação isolada de cada estrutura, este mesmo fato ocorreu na avaliação da precisão e IoU. Ainda na precisão, a associação de HE & MA propiciou um melhor resultado que sua análise individual.

Figura 75 – Diagrama apresentando os melhores resultados de acurácia das estruturas originais e suas combinações



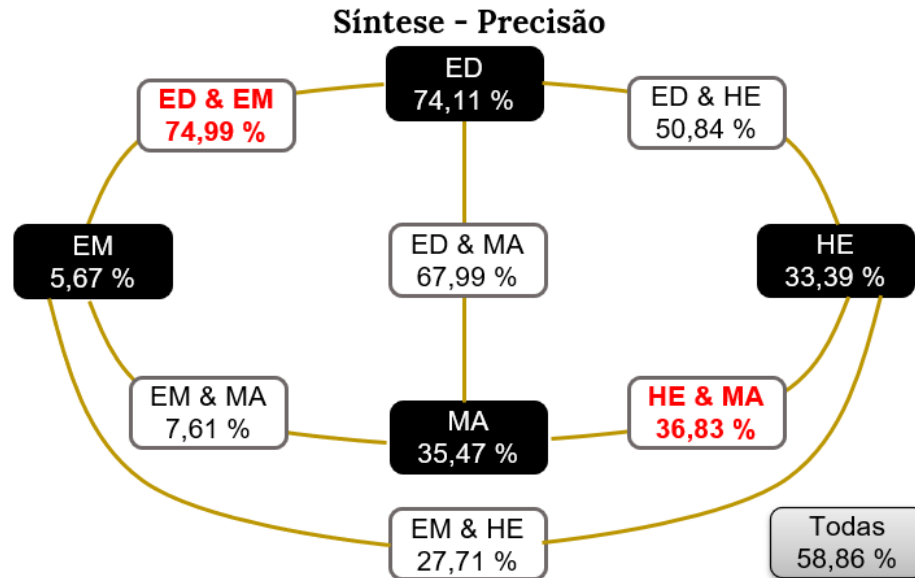
Fonte: Diagrama elaborado pelo autor.

Figura 76 – Diagrama apresentando os melhores resultados de sensibilidade das estruturas originais e suas combinações



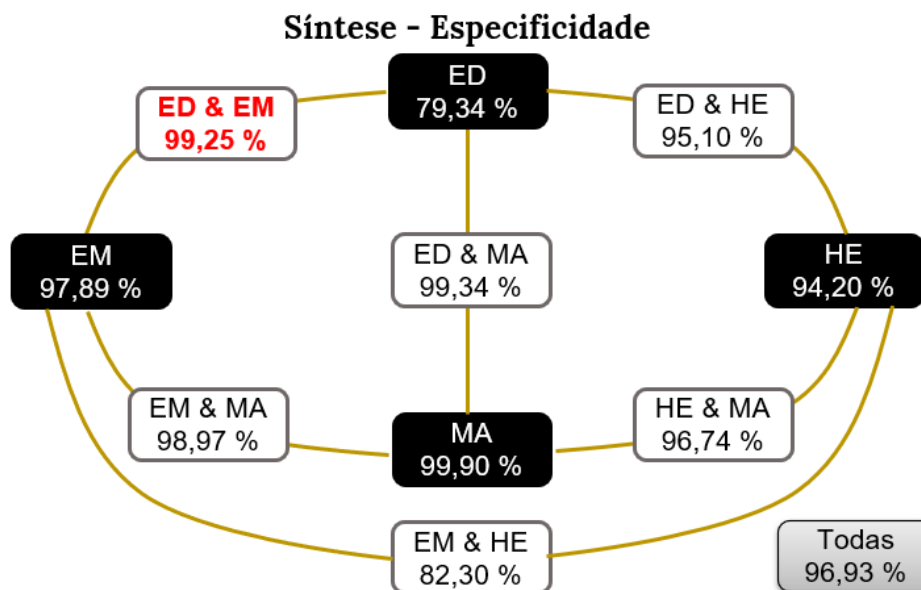
Fonte: Diagrama elaborado pelo autor.

Figura 77 – Diagrama apresentando os melhores resultados de precisão das estruturas originais e suas combinações



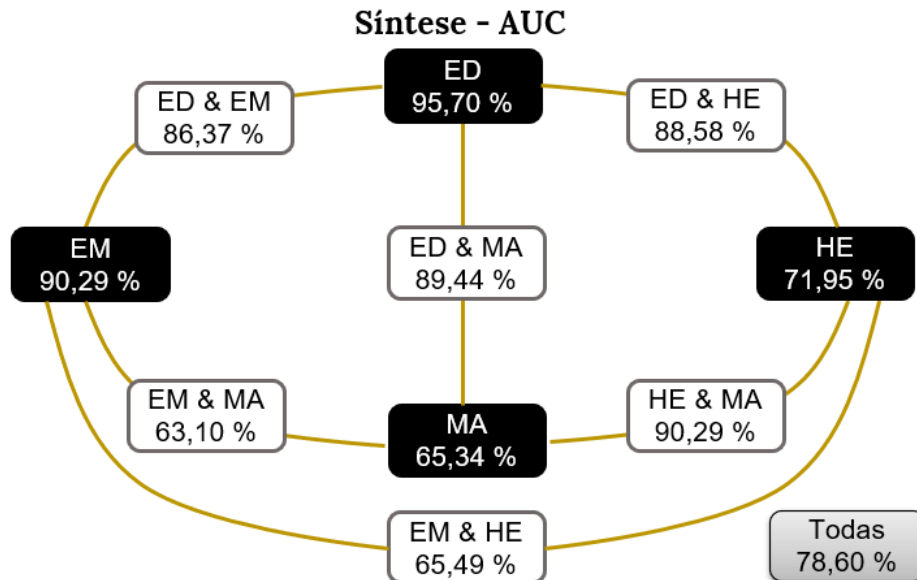
Fonte: Diagrama elaborado pelo autor.

Figura 78 – Diagrama apresentando os melhores resultados de especificidade das estruturas originais e suas combinações



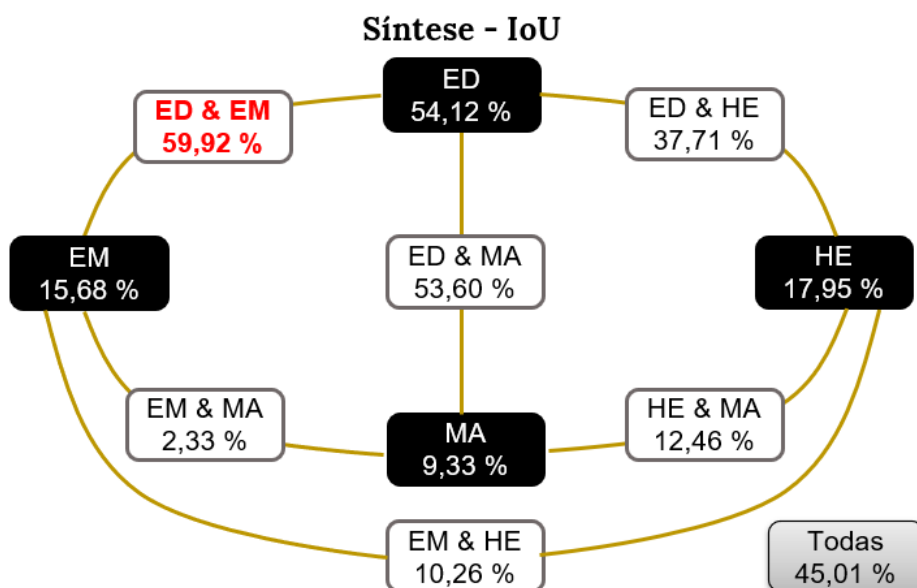
Fonte: Diagrama elaborado pelo autor.

Figura 79 – Diagrama apresentando os melhores resultados de AUC das estruturas originais e suas combinações



Fonte: Diagrama elaborado pelo autor.

Figura 80 – Diagrama apresentando os melhores resultados de IoU das estruturas originais e suas combinações



Fonte: Diagrama elaborado pelo autor.

## 6 CONCLUSÕES

A retinopatia diabética é a maior causa da cegueira evitável da população em idade economicamente ativa. Como agravante, esta doença apresenta grande perspectiva de crescimento, o que motiva o desenvolvimento de pesquisas que auxiliem a detecção e, conseqüentemente, a redução dessa perspectiva de crescimento.

Historicamente, diversas técnicas e ferramentas foram empregadas para classificar a retina como saudável ou portadora da RD, entretanto, há um número reduzido de trabalhos que buscam a segmentação de estruturas que indicam a presença da RD. Uma das principais dificuldades no diagnóstico médico é a subjetividade do especialista na análise das imagens médicas. Mais especificamente, a experiência do médico pode determinar o diagnóstico final. Métodos manuais de visualização podem ser tendenciosos, não padronizados e sujeitos a erros por parte do intérprete. Dessa forma, utilizar recursos computacionais para segmentação das regiões de interesse contribui para identificação das estruturas, bem como para a padronização do diagnóstico médico.

Os sistemas propostos na literatura não apresentam resultados perfeitos e, assim, não há garantia de um correto diagnóstico. Entretanto, segmentar as potenciais estruturas que indicam a presença da RD auxilia, de forma direta, o diagnóstico médico, reduzindo as chances de erros. Neste caso, o sistema não apresenta uma resposta conclusiva sobre a existência da RD, mas auxilia o diagnóstico médico na identificação e segmentação de regiões de risco.

Nesse sentido, almejando contribuir com as técnicas de processamento de imagens que auxiliem o diagnóstico médico, este trabalho abordou a segmentação de estruturas de interesse (exsudatos duros, exsudatos moles, hemorragias e microaneurismas) em imagens de retina para auxiliar o diagnóstico da RD. Com vistas a este propósito, foram apresentadas as características das estruturas de interesse em cada canal de cor RGB e suas possíveis combinações. Dessa forma, foi possível observar qualitativamente sua predominância nos canais de cores, bem como apresentar os resultados quantitativos de treinamentos e testes nos mesmos.

Os trabalhos presentes na literatura, em sua maioria, apresentam o desenvolvimento de técnicas com grande custo computacional de pré-processamento. A maioria desses trabalhos reduzem a resolução da imagem (*downsize*) para utilizá-la na rede *U-Net*. Como consequência, informações da imagem são perdidas, o que é prejudicial, especialmente, para as menores estruturas.

Na investigação aqui desenvolvida, foi proposto um método alternativo, em que o pré-processamento se restringiu à divisão das imagens originais. Estas divisões viabilizaram

o uso da rede neural por se adequarem ao tamanho de entrada da rede, sem perder informações e, assim, permitindo aumentar substancialmente o número de imagens (*data augmentation offline*), o que contribui para a generalização de predições.

A análise prévia dos canais de cores RGB individuais e suas combinações permitiu avaliar a contribuição de cada canal de cor para as respectivas estruturas de interesse. O canal verde apresentou melhor resultado em ED do que nos demais canais de cores ou suas combinações. Já para a estrutura de EM, o canal azul apresentou o melhor resultado dentre as combinações. Para as HE, a combinação dos canais RGB apresentaram os melhores resultados. Por fim, a combinação dos canais verde e azul geraram os melhores resultados para as estruturas de microaneurismas dentre as combinações avaliadas.

Em decorrência disso, a rede proposta foi modificada a fim de que sua entrada recebesse os três canais de cores (RGB) e, assim, pudesse, por meio dos pesos sinápticos, selecionar o canal de maior relevância ou combinação destes. Como as estruturas propostas possuem estruturas distintas, tanto em intensidade e brilho quanto em formato, foram utilizadas imagens tridimensionais (RGB) como entradas.

Diferentes métricas de avaliação dessas estruturas estão presentes em trabalhos na literatura. Na investigação desenvolvida neste trabalho, é apresentada a otimização de cada métrica comumente utilizada. Observa-se que a acurácia, apesar de sua ampla utilização, gera um resultado próximo de 100% devido às estruturas de interesse representarem diminuta parte da imagem, como consequência, o número de verdadeiros positivos é baixo e o número de verdadeiros negativos predominam. Outros trabalhos consideram como verdadeiro positivo toda região segmentada desde que esta tenha intersecção de ao menos um pixel. Essa abordagem pode levar a um resultado de acurácia próximo de 100% e, não necessariamente, corresponder a um resultado satisfatório para aplicações médicas no diagnóstico da RD.

Foi possível observar, nesta investigação, a predominância das estruturas em cada canal de cor RGB, bem como obtidos resultados numéricos e gráficos para treinamentos com cada canal isolado e suas possíveis combinações. Também foi apresentada e avaliada a combinação das diferentes estruturas de interesse. Como resultado, foi possível observar também que se obtém um melhor resultado ao segmentar exsudatos duros e moles em conjunto, em comparação com as estruturas individuais.

A rede *U-Net* utilizada neste estudo obteve como resultado, priorizando a acurácia, valores de: 99,03% para segmentação de exsudatos duros; 99,75% para segmentação de exsudatos moles; 98,50% para segmentação de hemorragias e 99,87% para segmentação de microaneurismas. Os melhores resultados apresentados na literatura são de 95,2%, 93%, 94,3% e 92,8%, respectivamente. Apesar dos valores superiores para acurácia, as sensibilidades apresentadas neste trabalho foram inferiores, sendo de 79,75%, 31,03%, 40,92% e 18,71% para segmentação de exsudatos duros, exsudatos moles, hemorragias e



---

microaneurismas, respectivamente, enquanto na literatura os resultados de destaque foram, na devida ordem, de 95,7%, 95,5%, 94,3% e 95,1%.

Pode-se afirmar que este trabalho também avaliou o resultado da soma de todas as estruturas de interesse e obteve como resultado acurácia de 97,51%, sensibilidade de 37,49%, precisão de 58,86% e especificidade de 99,24%, constituindo-se o primeiro trabalho a apresentar avaliações das estruturas em conjunto. É importante destacar que foi realizada uma abordagem inédita neste trabalho, embora se tenha obtido um resultado inferior aos que se encontram publicados na literatura até o presente momento. Nesta abordagem aplicada, busca-se obter uma única rede capaz de segmentar todas as estruturas em conjunto com a finalidade de se alcançar um resultado mais rápido e eficaz. Vale ressaltar, também, que o sistema proposto requer baixo custo computacional nas etapas de pré-processamento e aplicação ou testes.

Diante desses aspectos mencionados, conclui-se, portanto, que o objetivo deste trabalho foi atingido ao avaliar o comportamento da rede *U-Net* em diferentes configurações e propostas de pré-processamentos. Destaca-se a análise da combinação das diferentes estruturas, procedimento este não encontrado na literatura e que, portanto, pode se tornar alvo de interesse por permitir resultados mais rápidos para a identificação da retinopatia diabética.

Os resultados apresentados neste trabalho são obtidos pela divisão das imagens (*data augmentation offline* em substituição ao *downsize*), preservando todas as suas informações. Também foi utilizado técnicas de espelhamento para realizar *data augmentation* em tempo de execução e, assim, aumentar o número de imagens. É válido evidenciar que o método proposto neste estudo não apresenta alto custo computacional de aplicação uma vez que não requer pré-processamento sofisticado. Sendo assim, sua utilização se torna viável, visto que propicia maior acessibilidade do auxílio ao diagnóstico da retinopatia diabética.



## 7 TRABALHOS FUTUROS

O desenvolvimento deste trabalho permitiu a identificação de possíveis investigações futuras que possam contribuir e agregar novos resultados a esta linha de pesquisa. Cabe destacar, como trabalho futuro, a busca por diferentes bancos de dados, incluindo base de dados privadas, que possam ampliar as informações de aprendizado da rede. Aliado à obtenção de maior número de imagens, o emprego de novas técnicas de *data augmentation* pode reduzir a limitação da quantidade de imagens atualmente disponíveis.

Também podem ser realizados testes com outras arquiteturas de CNN, ajustando seus diversos parâmetros para aprimorar a performance obtida. Mais recentemente, novas técnicas têm ganhado destaque como as BNN, as quais apresentam, além da segmentação, a probabilidade bayesiana da segmentação estar correta. Esse tipo de sistema pode auxiliar o diagnóstico médico e apresentar um índice de confiabilidade do resultado obtido.

No estudo aqui desenvolvido, foi possível observar que a divisão das imagens originais apresentou melhor resultado para estruturas menores. Dessa forma, pode-se realizar dois treinamentos distintos: o primeiro com a divisão das imagens e o segundo com a redução da resolução. Por meio destes treinamentos, será possível gerar resultados complementares, os quais podem ser unidos em uma etapa de pós-processamento.

Outras investigações realizáveis seriam as avaliações dos possíveis resultados de treinamentos com a remoção do disco óptico da imagem na etapa de pré-processamento.

Por fim, cabe ressaltar que diferentes métodos de pré-processamento de imagem podem ser utilizados com o intuito de melhorar a predição da rede neural. Em relação à arquitetura da rede neural, podem ser alterados e, conseqüentemente, investigados os seguintes aspectos: número e tipos de camadas; tipo de otimizadores; números de épocas e *batch size*. Mesmo após a predição, as imagens podem ser pós-processadas com o intento de se obter resultados ainda mais precisos. Pode-se afirmar, portanto, que há muitas investigações possíveis, dependendo dos resultados encontrados e/ou do hardware disponível.



## REFERÊNCIAS

- ABRAMOFF, M. D.; GARVIN, M. K.; SONKA, M. Retinal imaging and image analysis. **IEEE Rev Biomed Eng**, v. 3, p. 169–208, 2010.
- ACHARYA, U. R. *et al.* Computer-based detection of diabetes retinopathy stages using digital fundus images. **Proceedings of the institution of mechanical engineers, part H: journal of engineering in medicine**, SAGE Publications Sage UK: London, England, v. 223, n. 5, p. 545–553, 2009.
- AL-AZZAWI, A. *et al.* Deepcryopicker: fully automated deep neural network for single protein particle picking in cryo-em. **BMC bioinformatics**, Springer, v. 21, n. 1, p. 1–38, 2020.
- ALBERTI, K. G. M. M.; ZIMMET, P. f. Definition, diagnosis and classification of diabetes mellitus and its complications. part 1: diagnosis and classification of diabetes mellitus. provisional report of a who consultation. **Diabetic medicine**, Wiley Online Library, v. 15, n. 7, p. 539–553, 1998.
- ALZUBAIDI, L. *et al.* Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. **Journal of big Data**, Springer, v. 8, n. 1, p. 1–74, 2021.
- APPLE. **Performing Convolution Operations**. 2016. Disponível em: <https://developer.apple.com/library/content/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html>.
- APTEL, F. *et al.* Screening of diabetic retinopathy: effect of field number and mydriasis on sensitivity and specificity of digital fundus photography. **Diabetes Metab.**, v. 34, n. 3, p. 290–293, Jun 2008.
- BELTHANGADY, C.; ROYER, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. **Nature methods**, Nature Publishing Group, v. 16, n. 12, p. 1215–1225, 2019.
- BENGIO, Y. **Deep Learning of Representations: Looking Forward**. arXiv, 2013. Disponível em: <https://arxiv.org/abs/1305.0445>.
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 8, p. 1798–1828, 2013. ISSN 01628828.
- BERNARDES, R. *et al.* Computer-assisted microaneurysm turnover in the early stages of diabetic retinopathy. **Ophthalmologica**, Karger Publishers, v. 223, n. 5, p. 284–291, 2009.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. [*S.l.*: *s.n.*]: Springer, 2006. ISBN 978-0387-31073-2.
- CAO, Y.; WU, Z.; SHEN, C. Estimating Depth from Monocular Images as Classification Using Deep Fully Convolutional Residual Networks. **CoRR**, abs/1605.0, p. 1–9, 2016. Disponível em: <http://arxiv.org/abs/1605.02305>.

CHEUNG, N.; MITCHELL, P.; WONG, T. Y. Diabetic retinopathy. **Lancet**, v. 376, n. 9735, p. 124–136, Jul 2010.

CHIDAMBARAM, N.; VIJAYAN, D. Detection of exudates in diabetic retinopathy. *In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.: s.n.], 2018. p. 660–664.

CORNET, G. Robot companions and ethics: A pragmatic approach of ethical design. **Journal international de bioéthique**, ESKA, v. 24, n. 4, p. 49–58, 2013.

DAS, S. *et al.* Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. **Biomedical Signal Processing and Control**, v. 68, p. 102600, 2021. ISSN 1746-8094. Disponível em: <https://www.sciencedirect.com/science/article/pii/S174680942100197X>.

DASGUPTA, A.; NATH, A. Classification of machine learning algorithms. **International Journal of Innovative Research in Advanced Engineering (IJIRAE)**, v. 3, n. 3, p. 6–11, 2016.

DECENCIERE, E. *et al.* Teleophtha: Machine learning and image processing methods for teleophthalmology. **Irbm**, Elsevier, v. 34, n. 2, p. 196–203, 2013.

DESERNO, T. M. **Biomedical image processing**. [S.l.: s.n.]: Springer Science & Business Media, 2011.

DUMOULIN, V.; VISIN, F. A guide to convolution arithmetic for deep learning. p. 1–31, 2016. ISSN 16113349. Disponível em: <http://arxiv.org/abs/1603.07285>.

EGE, B. M. *et al.* Screening for diabetic retinopathy using computer based image analysis and statistical classification. **Computer methods and programs in biomedicine**, Elsevier, v. 62, n. 3, p. 165–175, 2000.

EIGEN, D.; PUHRSCHE, C.; FERGUS, R. Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. **NIPS**, p. 1–9, 2014. ISSN 10495258. Disponível em: <http://arxiv.org/abs/1406.2283>.

FANG, W. *et al.* Computer vision for behaviour-based safety in construction: A review and future directions. **Advanced Engineering Informatics**, Elsevier, v. 43, p. 100980, 2020.

FAUST, O. *et al.* Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. **Journal of medical systems**, Springer, v. 36, n. 1, p. 145–157, 2012.

FEDERATION, I. D. [www.diabetesatlas.org](http://www.diabetesatlas.org). 2021.

FELDMAN, J.; ROJAS, R. **Neural Networks: A Systematic Introduction**. [S.l.: s.n.]: Springer Berlin Heidelberg, 2013. ISBN 9783642610684.

FLAXMAN, S. R. *et al.* Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. **The Lancet Global Health**, Elsevier, v. 5, n. 12, p. e1221–e1234, 2017.

FRAME, A. J. *et al.* A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms. **Computers in biology and medicine**, Elsevier, v. 28, n. 3, p. 225–238, 1998.

FRAZ, M. M. *et al.* Blood vessel segmentation methodologies in retinal images—a survey. **Computer methods and programs in biomedicine**, Elsevier, v. 108, n. 1, p. 407–433, 2012.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, v. 36, n. 4, p. 193–202, 1980. ISSN 03401200.

GARDNER, G. G. *et al.* Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. **Br J Ophthalmol**, v. 80, n. 11, p. 940–944, Nov 1996.

GLOROT, X.; BORDES, A.; BENGIO, Y. Deep sparse rectifier neural networks. **AISTATS '11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics**, v. 15, p. 315–323, 2011. ISSN 15324435.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [*S.l.: s.n.*]: MIT Press, 2016.

GREENSPAN, H.; GINNEKEN, B. V.; SUMMERS, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. **IEEE Transactions on Medical Imaging**, IEEE, v. 35, n. 5, p. 1153–1159, 2016.

GU, J. *et al.* Recent advances in convolutional neural networks. **Pattern Recognition**, Elsevier, v. 77, p. 354–377, 2018.

GULSHAN, V. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. **Jama**, American Medical Association, v. 316, n. 22, p. 2402–2410, 2016.

GUO, S. *et al.* Bin loss for hard exudates segmentation in fundus images. **Neurocomputing**, v. 392, p. 314–324, 2020. ISSN 0925-2312. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231219304783>.

GUO, Y. *et al.* Deep learning for visual understanding: A review. **Neurocomputing**, v. 187, p. 27–48, 2016. ISSN 18728286.

HAHNIOSE, R. H. *et al.* Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. **Nature**, v. 405, n. 6789, p. 947–951, 2000. ISSN 00280836.

HAMET, P.; TREMBLAY, J. Artificial intelligence in medicine. **Metabolism**, Elsevier, v. 69, p. S36–S40, 2017.

HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. **Radiology**, v. 143, n. 1, p. 29–36, 1982.

HARIS, M. *et al.* Segmentation of ocular pathologies using deep convolutional neural network. *In*: **2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)**. [*S.l.: s.n.*], 2019. p. 1–5.

HASSAN, G. *et al.* Retinal blood vessel segmentation approach based on mathematical morphology. **Procedia Computer Science**, Elsevier, v. 65, p. 612–622, 2015.

HE, K. *et al.* Deep Residual Learning for Image Recognition. *In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [*S.l.: s.n.*], 2016. p. 770–778. ISBN 978-1-4673-8851-1. ISSN 1664-1078. Disponível em: <http://ieeexplore.ieee.org/document/7780459/>.

HELLSTEDT, T.; IMMONEN, I. Disappearance and formation rates of microaneurysms in early diabetic retinopathy. **British journal of ophthalmology**, BMJ Publishing Group Ltd, v. 80, n. 2, p. 135–139, 1996.

HOWARD, A. G. *et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.

HU, K. *et al.* Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. **Neurocomputing**, Elsevier, v. 309, p. 179–191, 2018.

JALLI, P.; HELLSTEDT, T. J.; IMMONEN, I. Early versus late staining of microaneurysms in fluorescein angiography. **Retina (Philadelphia, Pa.)**, v. 17, n. 3, p. 211–215, 1997.

JELINEK, H.; CREE, M. J. **Automated image detection of retinal pathology**. [*S.l.: s.n.*]: Crc Press, 2009.

JONES, S.; EDWARDS, R. Diabetic retinopathy screening: a systematic review of the economic evidence. **Diabetic medicine**, Wiley Online Library, v. 27, n. 3, p. 249–256, 2010.

KAHAI, P.; NAMUDURI, K. R.; THOMPSON, H. A decision support framework for automated screening of diabetic retinopathy. **Int J Biomed Imaging**, v. 2006, p. 45806, 2006.

KARASU, S. *et al.* A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. **Energy**, Elsevier, v. 212, p. 118750, 2020.

KETKAR, N.; SANTANA, E. **Deep learning with Python**. [*S.l.: s.n.*]: Springer, 2017. v. 1.

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.

KOU, C. *et al.* An enhanced residual u-net for microaneurysms and exudates segmentation in fundus images. **IEEE Access**, v. 8, p. 185514–185525, 2020.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. **Advances In Neural Information Processing Systems**, p. 1–9, 2012. ISSN 10495258.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional. **Neural Netw**, v. 60, p. 84–90, 2017.



- 
- KROGH, A. What are artificial neural networks? **Nature biotechnology**, Nature Publishing Group, v. 26, n. 2, p. 195–197, 2008.
- LARSON, J. A.; JOHNSON, M. H.; BHAYANI, S. B. Application of surgical safety standards to robotic surgery: five principles of ethics for nonmaleficence. **Journal of the American College of Surgeons**, LWW, v. 218, n. 2, p. 290–293, 2014.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LECUN, Y. *et al.* Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278–2323, 1998. ISSN 00189219.
- LEE, H.; EKANADHAM, C.; NG, A. Y. Sparse deep belief net model for visual area V2. *In: Advances in Neural Information Processing Systems 20*. [*S.l.: s.n.*], 2008. p. 873—880. ISBN 160560352X | 9781605603520.
- LEE, H. *et al.* Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *In: Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. [*S.l.: s.n.*], 2009. p. 1–8. ISBN 9781605585161. ISSN 02643294. Disponível em: <http://portal.acm.org/citation.cfm?doid=1553374.1553453>.
- LEE, R.; WONG, T. Y.; SABANAYAGAM, C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. **Eye and vision**, BioMed Central, v. 2, n. 1, p. 17, 2015.
- LI, B. *et al.* Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs. *In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. [*S.l.: s.n.*], 2015. v. 07-12-June, p. 1119–1127. ISBN 9781467369640. ISSN 10636919.
- LI, G. *et al.* Efficient densely connected convolutional neural networks. **Pattern Recognition**, Elsevier, v. 109, p. 107610, 2021.
- LI, H.; CHUTATAPE, O. A model-based approach for automated feature extraction in fundus images. *In: IEEE*. **null**. [*S.l.: s.n.*], 2003. p. 394.
- LI, X. *et al.* Lightweight attention convolutional neural network for retinal vessel image segmentation. **IEEE Transactions on Industrial Informatics**, IEEE, v. 17, n. 3, p. 1958–1967, 2020.
- LITJENS, G. *et al.* A survey on deep learning in medical image analysis. **Medical image analysis**, Elsevier, v. 42, p. 60–88, 2017.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. [*S.l.: s.n.*], 2015. p. 3431–3440.
- MA, J.; YARATS, D. Quasi-hyperbolic momentum and adam for deep learning. **arXiv preprint arXiv:1810.06801**, 2018.

MAHER, S. K. R.; DHOPEHWARKAR, D. M. Review of automated detection for diabetes retinopathy using fundus images. **International Journal of Advanced Research in Computer Science and Software Engineering**, v. 5, n. 3, 2015.

MAHESH, B. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR)**. [Internet], v. 9, p. 381–386, 2020.

MENDES, C. C. T. **Navigability estimation for autonomous vehicles using machine learning**. 2017. 96 p. Tese (Tese de Doutorado) — Universidade de São Paulo, 2017. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-25092017-102021/pt-br.php>.

MICROANEURYSMS In Diabetic Retinopathy. **The British Medical Journal**, BMJ, v. 3, n. 5774, p. 548–548, 1971. ISSN 00071447. Disponível em: <http://www.jstor.org/stable/25415740>.

MOOKIAH, M. R. *et al.* Computer-aided diagnosis of diabetic retinopathy: a review. **Comput. Biol. Med.**, v. 43, n. 12, p. 2136–2155, Dec 2013.

NAZIR, T. *et al.* Diabetic retinopathy lesions detection using faster-rcnn from retinal images. *In: 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*. [S.l.: s.n.], 2020. p. 38–42.

NENTWICH, M. M.; ULBIG, M. W. Diabetic retinopathy - ocular complications of diabetes mellitus. **World J Diabetes**, v. 6, n. 3, p. 489–499, Apr 2015.

NOH, H.; HONG, S.; HAN, B. Learning deconvolution network for semantic segmentation. **Proceedings of the IEEE International Conference on Computer Vision**, v. 2015 Inter, p. 1520–1528, may 2015. ISSN 15505499. Disponível em: <http://arxiv.org/abs/1505.04366>.

NORONHA, K.; NAYAK, J.; BHAT, S. Enhancement of retinal fundus image to highlight the features for detection of abnormal eyes. *In: IEEE. TENCON 2006-2006 IEEE Region 10 Conference*. [S.l.: s.n.], 2006. p. 1–4.

NORTH-VALLEY, E. C. **Diabetic Retinopathy**. North Valley Eye Care, 2019. Disponível em: [https://www.northvalleyeyecare.com/Content/EyeConditions/what\\_is\\_diabetic\\_retinopathy.aspx?popup=true](https://www.northvalleyeyecare.com/Content/EyeConditions/what_is_diabetic_retinopathy.aspx?popup=true).

OSAREH, A. *et al.* Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks. *In: CITESEER. Proc. Medical Image Understanding Analysis Conf*. [S.l.: s.n.], 2001. v. 3, p. 49–52.

OSAREH, A. *et al.* Classification and localisation of diabetic-related eye disease. *In: HEYDEN, A. et al. (ed.). Computer Vision — ECCV 2002*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p. 502–516. ISBN 978-3-540-47979-6.

PALAZ, D.; MAGIMAI-DOSS, M.; COLLOBERT, R. End-to-end acoustic modeling using convolutional neural networks for hmm-based automatic speech recognition. **Speech Communication**, Elsevier, v. 108, p. 15–32, 2019.

PATTON, N. *et al.* Retinal image analysis: concepts, applications and potential. **Progress in retinal and eye research**, Elsevier, v. 25, n. 1, p. 99–127, 2006.

PHILLIPS, R.; FORRESTER, J.; SHARP, P. Automated detection and quantification of retinal exudates. **Graefe's archive for clinical and experimental ophthalmology**, Springer, v. 231, n. 2, p. 90–94, 1993.

PONTI, M. A. *et al.* Everything you wanted to know about deep learning for computer vision but were afraid to ask. *In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*. [S.l.: s.n.], 2017. p. 17–41.

PORWAL, P. *et al.* Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. **Data**, Multidisciplinary Digital Publishing Institute, v. 3, n. 3, p. 25, 2018.

PORWAL, P. *et al.* Automatic retinal image analysis for the detection of diabetic retinopathy. *In: Biomedical Signal and Image Processing in Patient Care*. [S.l.: s.n.]: IGI Global, 2018. p. 146–161.

PRATI, R. *et al.* Curvas roc para avaliação de classificadores. **Revista IEEE América Latina**, v. 6, n. 2, p. 215–222, 2008.

QUELLEC, G.; RUSSELL, S. R.; ABRAMOFF, M. D. Optimal filter framework for automated, instantaneous detection of lesions in retinal images. **IEEE Trans Med Imaging**, v. 30, n. 2, p. 523–533, Feb 2011.

RAMAN, R. *et al.* Diabetic retinopathy: An epidemic at home and around the world. **Indian journal of ophthalmology**, Wolters Kluwer–Medknow Publications, v. 64, n. 1, p. 69, 2016.

REZA, A. W.; ESWARAN, C. A decision support system for automatic screening of non-proliferative diabetic retinopathy. **J Med Syst**, v. 35, n. 1, p. 17–24, Feb 2011.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. *In: SPRINGER. International Conference on Medical image computing and computer-assisted intervention*. [S.l.: s.n.], 2015. p. 234–241.

SAMBYAL, N. *et al.* Modified u-net architecture for semantic segmentation of diabetic retinopathy images. **Biocybernetics and Biomedical Engineering**, v. 40, n. 3, p. 1094–1109, 2020. ISSN 0208-5216. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0208521620300747>.

SÁNCHEZ, C. I. *et al.* Retinal image analysis to detect and quantify lesions associated with diabetic retinopathy. *In: IEEE. The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. [S.l.: s.n.], 2004. v. 1, p. 1624–1627.

SHORTLIFFE, E. H.; BLOIS, M. S. The computer meets medicine and biology: emergence of a discipline. *In: Biomedical Informatics*. [S.l.: s.n.]: Springer, 2006. p. 3–45.

SIDDIQUE, N. *et al.* U-net and its variants for medical image segmentation: A review of theory and applications. **IEEE Access**, v. 9, p. 82031–82057, 2021.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para Engenharia e ciências aplicadas**. [S.l.: s.n.]: ARTLIBER, 2010. ISBN 9788588098534.

SILVA, I. N. da *et al.* **Artificial Neural Networks: A Practical Course**. [*S.l.: s.n.*]: Springer International Publishing, 2016. ISBN 9783319431628.

SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. **International Conference on Learning Representations (ICRL)**, p. 1–14, 2015. ISSN 09505849. Disponível em: <http://arxiv.org/abs/1409.1556>.

SINTHANAYOTHIN, C. *et al.* Automated detection of diabetic retinopathy on digital fundus images. **Diabetic medicine**, Wiley Online Library, v. 19, n. 2, p. 105–112, 2002.

SINTHANAYOTHIN, C. *et al.* Automated screening system for diabetic retinopathy. *In: IEEE. 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the.* [*S.l.: s.n.*], 2003. v. 2, p. 915–920.

STAAL, J. *et al.* Ridge-based vessel segmentation in color images of the retina. **IEEE transactions on medical imaging**, IEEE, v. 23, n. 4, p. 501–509, 2004.

SZEGEDY, C. *et al.* Going deeper with convolutions. *In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* [*S.l.: s.n.*], 2015. v. 07-12-June, p. 1–9. ISBN 9781467369640. ISSN 10636919.

THEOFILATOS, K. *et al.* Predicting protein complexes from weighted protein–protein interaction graphs with a novel unsupervised methodology: evolutionary enhanced markov clustering. **Artificial intelligence in medicine**, Elsevier, v. 63, n. 3, p. 181–189, 2015.

TING, D. S. W.; CHEUNG, G. C. M.; WONG, T. Y. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. **Clinical & experimental ophthalmology**, Wiley Online Library, v. 44, n. 4, p. 260–277, 2016.

TRUCCO, E. *et al.* Validating retinal fundus image analysis algorithms: issues and a proposal. **Investigative ophthalmology & visual science**, The Association for Research in Vision and Ophthalmology, v. 54, n. 5, p. 3546–3559, 2013.

TRUCCO, E. *et al.* Validating retinal fundus image analysis algorithms: issues and a proposal. **Invest. Ophthalmol. Vis. Sci.**, v. 54, n. 5, p. 3546–3559, May 2013.

TSIKNAKIS, N. *et al.* Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. **Computers in Biology and Medicine**, Elsevier, v. 135, p. 104599, 2021.

TSIKNAKIS, N. *et al.* Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. **Computers in Biology and Medicine**, v. 135, p. 104599, 2021. ISSN 0010-4825. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0010482521003930>.

USHER, D. *et al.* Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. **Diabet. Med.**, v. 21, n. 1, p. 84–90, Jan 2004.

VERNWAL, R. **How should I start with CNN? – Becoming Human: Artificial Intelligence Magazine**. 2017. Disponível em: <https://becominghuman.ai/how-should-i-start-with-cnn-c62a3a89493b>.

---

VÁZQUEZ, F. **Deep Learning made easy with Deep Cognition**. 2017. Disponible em: <https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351>.

WALTER, T. *et al.* A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. **IEEE transactions on medical imaging**, IEEE, v. 21, n. 10, p. 1236–1243, 2002.

WANG, H. *et al.* An effective approach to detect lesions in color retinal images. *In: IEEE. Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*. [S.l.: s.n.], 2000. v. 2, p. 181–186.

WANG, P. *et al.* Understanding convolution for semantic segmentation. *In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2018. p. 1451–1460.

WANG, T. *et al.* A hybrid method for heartbeat classification via convolutional neural networks, multilayer perceptrons and focal loss. **PeerJ Computer Science**, PeerJ Inc., v. 6, p. e324, 2020.

WARD, N. P.; TOMLIIVSON, S.; TAYLOR, C. J. Image analysis of fundus photographs: the detection and measurement of exudates associated with diabetic retinopathy. **Ophthalmology**, Elsevier, v. 96, n. 1, p. 80–86, 1989.

WONG, T. Y. *et al.* Diabetic retinopathy. **Nature Reviews Disease Primers**, v. 2, p. 16012, 03 2016.

YANG, K. *et al.* A study of face obfuscation in imagenet. **arXiv preprint arXiv:2103.06191**, 2021.

YAU, J. W. *et al.* Global prevalence and major risk factors of diabetic retinopathy. **Diabetes Care**, v. 35, n. 3, p. 556–564, Mar 2012.

ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. *In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [S.l.: s.n.], 2014. v. 8689 LNCS, n. PART 1, p. 818–833. ISBN 9783319105895. ISSN 16113349.

ZHANG, C.; LU, Y. Study on artificial intelligence: The state of the art and future prospects. **Journal of Industrial Information Integration**, Elsevier, v. 23, p. 100224, 2021.

ZHANG, Z.; CUI, P.; ZHU, W. Deep learning on graphs: A survey. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, 2020.

ZHOU, D.-X. Theory of deep convolutional neural networks: Downsampling. **Neural Networks**, Elsevier, v. 124, p. 319–327, 2020.

ZHOU, L. *et al.* Improving dense conditional random field for retinal vessel segmentation by discriminative feature learning and thin-vessel enhancement. **Computer methods and programs in biomedicine**, Elsevier, v. 148, p. 13–25, 2017.



## APÊNDICES





## **APÊNDICE A – RESULTADOS DOS MELHORES TREINAMENTOS**

Tabela 18 – Treinamentos exsudatos duros

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
50	1,00E-03	122	21,52%	8	92,08%	44,92%	7,69%	92,72%	59,35%
				121	98,33%	19,23%	30,02%	99,39%	61,74%
				122	98,33%	19,15%	30,18%	99,40%	61,77%
50	2,80E-04	88	53,13%	8	96,82%	67,69%	24,68%	97,21%	71,60%
				88	98,98%	36,50%	74,11%	99,83%	77,36%
50	2,00E-04	94	54,12%	8	95,86%	78,48%	21,36%	96,10%	72,95%
				94	99,03%	44,76%	72,21%	99,77%	78,94%
				93	99,03%	45,00%	71,94%	99,76%	78,93%
50	3,20E-04	111	46,92%	8	94,17%	70,78%	14,79%	94,49%	68,56%
				110	98,93%	32,34%	71,96%	99,83%	75,77%
				111	98,93%	32,17%	72,24%	99,83%	75,79%
50	3,00E-04	9	49,82%	8	98,89%	42,78%	62,38%	99,65%	75,93%
				9	98,91%	42,07%	63,47%	99,67%	76,03%
50	3,00E-04	33	46,68%	8	98,52%	54,51%	45,32%	99,11%	74,37%
				33	98,93%	44,16%	64,23%	99,67%	76,75%
				8	91,33%	58,32%	8,74%	91,78%	62,54%
50	5,00E-04	131	36,88%	130	98,75%	19,25%	60,01%	99,83%	69,46%
				131	98,75%	19,12%	60,32%	99,83%	69,51%
50	8,00E-04	129	26,16%	8	91,74%	54,77%	8,71%	92,24%	61,87%
				128	98,63%	13,83%	45,56%	99,78%	64,45%
				129	98,63%	13,68%	45,94%	99,78%	64,51%
80	1,00E-04	55	53,75%	8	98,68%	57,85%	50,50%	99,23%	76,57%
				54	98,98%	43,81%	67,95%	99,72%	77,61%
				55	98,98%	43,64%	68,14%	99,72%	77,62%
80	3,00E-04	59	54,24%	8	98,34%	56,60%	41,08%	98,90%	73,73%
				58	98,98%	41,84%	69,92%	99,76%	77,62%
				59	98,99%	41,67%	70,19%	99,76%	77,65%

Tabela 19 – Treinamentos exsudatos duros

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
80	3,00E-04	59	54,24%	8	98,34%	56,60%	41,08%	98,90%	73,73%
						41,84%	69,92%	99,76%	77,62%
						41,67%	70,19%	99,76%	77,65%
20	3,20E-04	51	30,92%	8	97,11%	43,58%	21,37%	97,83%	64,97%
						14,91%	67,48%	99,90%	70,27%
						51,20%	53,00%	99,39%	75,58%
50	1,00E-04	45	51,31%	8	98,74%	41,71%	69,47%	99,75%	77,48%
						41,55%	69,72%	99,76%	77,50%
						45,10%	41,90%	99,16%	71,15%
50	3,00E-04	42	47,55%	8	98,43%	34,33%	65,83%	99,76%	74,70%
						34,16%	66,22%	99,76%	74,76%
						79,75%	4,96%	79,34%	60,85%
60	3,20E-04	148	35,05%	8	79,35%	23,85%	53,67%	99,72%	68,99%
						50,85%	54,87%	99,44%	75,99%
						40,99%	71,00%	99,77%	77,69%
80	1,00E-04	63	54,53%	8	98,79%	40,88%	71,16%	99,78%	77,70%
						47,52%	27,42%	98,30%	67,72%
						29,98%	54,01%	99,66%	70,59%
80	3,00E-04	37	42,48%	8	97,62%	28,98%	55,69%	99,69%	70,78%
						49,59%	56,38%	99,48%	76,07%
						40,78%	72,60%	99,79%	78,04%
100	1,00E-04	55	57,51%	8	98,82%	40,67%	72,76%	99,79%	78,06%

Tabela 20 – Treinamentos exsudatos moles

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
120	1,00E-04	53	1,41%	8	99,62%	3,76%	3,27%	99,80%	51,61%
<b>150</b>	<b>1,00E-04</b>	<b>9</b>	<b>15,68%</b>	8	99,03%	18,99%	4,03%	99,17%	55,31%
				<b>9</b>	<b>99,17%</b>	<b>16,44%</b>	<b>4,23%</b>	<b>99,32%</b>	<b>54,79%</b>
50	1,00E-05	3	1,89%	8	99,37%	3,01%	1,19%	99,54%	50,78%
60	1,00E-04	4	11,22%	2	98,45%	10,48%	1,36%	98,61%	52,22%
				3	98,80%	7,84%	1,37%	98,96%	51,74%
				8	99,00%	15,34%	3,20%	99,15%	54,17%
				4	97,77%	31,03%	2,62%	97,89%	57,33%
80	1,00E-05	34	1,42%	5	98,32%	25,05%	2,88%	98,45%	56,18%
				8	99,54%	5,90%	3,55%	99,71%	52,17%
				34	99,70%	3,24%	4,51%	99,87%	51,83%

Tabela 21 – Treinamentos hemorragias

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
80	1,00E-04	40	8,13%	8	93,57%	40,92%	7,77%	94,20%	59,11%
				40	97,23%	20,70%	11,72%	98,14%	56,95%
				39	97,19%	21,01%	11,64%	98,10%	56,98%
60	1,00E-04	40	8,13%	8	96,59%	32,01%	12,65%	97,36%	59,65%
				9	94,78%	31,08%	11,45%	98,30%	58,90%
				40	94,40%	25,60%	9,37%	98,50%	56,97%
80	1,00E-05	5	17,29%	8	98,27%	30,46%	28,42%	99,08%	64,06%
				5	98,19%	31,98%	27,30%	98,98%	64,12%
				8	98,50%	27,54%	33,39%	99,34%	64,69%
<b>60</b>	<b>1,00E-05</b>	<b>3</b>	<b>17,95%</b>	<b>3</b>	<b>98,30%</b>	<b>31,41%</b>	<b>29,45%</b>	<b>99,10%</b>	<b>64,56%</b>
				5	98,41%	29,39%	31,48%	99,24%	64,63%
				2	98,20%	32,96%	27,89%	98,98%	64,51%
60	2,00E-04	9	5,37%	8	94,46%	26,10%	6,19%	95,28%	55,51%
				9	95,06%	23,32%	6,38%	95,92%	55,17%
				10	95,55%	20,89%	6,54%	96,44%	54,86%
60	1,00E-04	5	4,82%	8	94,43%	23,94%	5,70%	95,27%	54,84%
				5	92,01%	33,54%	5,21%	92,71%	55,87%
				7	93,79%	26,71%	5,56%	94,59%	55,16%
60	1,00E-03	5	5,42%	8	95,66%	21,09%	6,80%	96,55%	55,02%
				5	92,42%	35,89%	5,84%	93,10%	56,81%
				7	94,95%	25,02%	6,62%	95,79%	55,59%
50	3,00E-05	8	13,62%	8	98,21%	23,78%	24,01%	99,10%	61,28%
				7	98,19%	24,19%	23,68%	99,07%	61,28%
				6	98,15%	24,66%	23,31%	99,03%	61,29%
50	1,00E-04	10	4,25%	8	92,33%	26,60%	4,41%	93,12%	54,11%
				10	94,33%	20,70%	4,90%	95,21%	53,78%
				12	95,58%	16,29%	5,30%	96,52%	53,42%

Tabela 22 – Treinamentos hemorragias

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
50	3,20E-04	16	4,79%	8	91,30%	28,41%	4,09%	92,05%	53,96%
				16	96,49%	14,67%	6,46%	97,46%	53,77%
				17	96,75%	13,68%	6,74%	97,74%	53,73%
60	1,00E-04	10	10,21%	8	96,59%	32,01%	12,65%	97,36%	59,65%
				6	96,29%	34,11%	12,05%	97,03%	59,87%
				9	96,70%	31,17%	12,89%	97,49%	59,56%

Tabela 23 – Treinamentos Microaneurismas

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
30	1,00E-04	24	9,08%	8	99,81%	17,36%	17,84%	99,91%	58,73%
				24	99,85%	12,99%	23,09%	99,95%	58,97%
				25	99,85%	12,83%	23,31%	99,95%	58,99%
60	1,00E-05	26	9,10%	8	99,81%	18,71%	18,51%	99,90%	59,23%
				26	99,86%	12,28%	25,89%	99,96%	59,50%
				27	99,86%	12,09%	26,19%	99,96%	59,52%
<b>60</b>	<b>1,00E-05</b>	<b>13</b>	<b>9,33%</b>	8	99,81%	17,04%	17,39%	99,91%	58,54%
				<b>13</b>	<b>99,83%</b>	<b>14,76%</b>	<b>20,22%</b>	<b>99,93%</b>	<b>58,69%</b>
				8	99,87%	9,63%	35,47%	99,98%	61,24%
80	1,00E-04	1	9,22%	1	99,86%	11,72%	30,20%	99,97%	60,44%

Tabela 24 – Treinamentos exsudatos duros e moles

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
100	1,00E-05	24	57,86%	8	98,73%	58,91%	57,89%	99,34%	78,72%
				23	98,90%	51,00%	68,56%	99,64%	79,53%
				24	98,91%	50,68%	69,01%	99,65%	79,56%
50	1,00E-04	28	53,88%	8	98,53%	56,04%	51,47%	99,19%	76,31%
				27	98,76%	47,48%	61,72%	99,55%	76,88%
				28	98,76%	47,19%	62,04%	99,56%	76,89%
<b>60</b>	<b>1,00E-04</b>	<b>37</b>	<b>59,92%</b>	8	98,71%	63,43%	56,59%	99,25%	79,49%
				<b>37</b>	<b>99,00%</b>	<b>52,00%</b>	<b>74,38%</b>	<b>99,72%</b>	<b>81,28%</b>
				39	99,00%	51,51%	74,99%	99,74%	81,31%
60	1,00E-04	6	54,00%	8	98,84%	47,69%	66,18%	99,63%	78,08%
				4	98,78%	50,90%	61,79%	99,52%	77,75%
				6	98,82%	49,07%	64,34%	99,58%	77,95%

Tabela 25 – Treinamento exsudatos duros e hemorragias

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
30	1,00E-04	9	24,87%	8	94,27%	62,16%	24,64%	95,10%	69,04%
				9	94,50%	61,30%	25,36%	95,35%	69,13%
				10	94,68%	60,51%	26,02%	95,57%	69,19%
				14	95,23%	57,86%	28,15%	96,19%	69,36%
50	1,00E-04	7	31,13%	8	96,75%	40,45%	36,65%	98,20%	68,01%
				7	96,70%	41,08%	36,13%	98,13%	68,01%
				9	96,79%	39,89%	37,10%	98,26%	68,01%
80	1,00E-05	36	37,71%	8	97,09%	42,42%	42,11%	98,50%	70,03%
				36	97,51%	34,83%	50,67%	99,13%	70,53%
				37	97,52%	34,67%	50,84%	99,14%	70,54%
80	1,00E-04	30	29,92%	8	96,08%	43,33%	30,35%	97,44%	66,80%
				29	96,49%	38,40%	32,94%	97,99%	66,45%
				30	96,50%	38,26%	33,01%	98,00%	66,44%
<b>80</b>	<b>1,00E-05</b>	<b>7</b>	<b>38,85%</b>	<b>8</b>	<b>97,47%</b>	<b>33,57%</b>	<b>49,49%</b>	<b>99,12%</b>	<b>69,91%</b>
				6	97,41%	34,87%	47,89%	99,02%	69,80%
				7	97,44%	34,18%	48,75%	99,07%	69,86%



Tabela 26 – Treinamento exsudatos e microaneurismas

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
50	1,00E-04	38	53,60%	8	98,64%	50,90%	53,10%	99,34%	75,50%
				37	98,87%	42,83%	67,72%	99,70%	77,28%
				38	98,88%	42,66%	67,99%	99,70%	77,31%
60	1,00E-03	3	46,36%	8	98,85%	39,56%	67,84%	99,72%	76,50%
				2	98,76%	44,47%	59,92%	99,56%	75,68%
				3	98,80%	43,14%	62,19%	99,61%	75,93%
60	1,00E-04	3	40,65%	8	98,73%	38,95%	59,59%	99,61%	74,22%
				3	98,63%	41,50%	53,56%	99,47%	73,29%
				5	98,69%	40,22%	56,64%	99,55%	73,77%
60	1,00E-04	3	46,14%	8	98,73%	42,33%	58,77%	99,56%	74,85%
				3	98,62%	45,08%	52,87%	99,41%	74,00%

Tabela 27 – Treinamento exsudatos moles e hemorragias

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
60	1,00E-05	3	14,92%	8	98,13%	23,33%	27,71%	99,16%	62,08%
				3	97,85%	27,54%	24,35%	98,82%	62,14%
				4	97,94%	26,29%	25,31%	98,93%	62,12%
60	1,00E-04	5	4,06%	8	93,42%	21,46%	5,04%	94,42%	53,58%
				4	85,72%	39,37%	3,83%	86,36%	53,82%
				5	89,15%	33,08%	4,34%	89,92%	54,12%
60	1,00E-05	3	11,39%	8	98,02%	18,34%	22,25%	99,12%	59,43%
				2	97,60%	22,48%	18,57%	98,64%	59,32%
				3	97,73%	21,29%	19,54%	98,79%	59,34%
60	1,00E-04	4	3,66%	8	93,04%	21,49%	4,74%	94,03%	53,33%
				3	81,80%	45,43%	3,42%	82,30%	53,24%
				4	86,11%	38,20%	3,83%	86,77%	53,73%
70	1,00E-04	8	3,99%	8	93,21%	20,14%	4,59%	94,22%	53,04%
				7	92,03%	23,17%	4,36%	92,98%	53,14%

Tabela 28 – Treinamento: exsudatos moles e microaneurismas

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
<b>60</b>	<b>1,00E-05</b>	<b>5</b>	<b>4,79%</b>	8	99,32%	11,37%	7,61%	99,59%	54,47%
				<b>4</b>	<b>99,05%</b>	<b>15,64%</b>	<b>6,33%</b>	<b>99,30%</b>	<b>55,08%</b>
				5	99,15%	14,22%	6,72%	99,41%	54,87%
60	1,00E-05	1	2,66%	8	99,47%	2,48%	3,09%	99,77%	51,20%
				1	97,95%	18,66%	3,01%	98,19%	54,45%
60	1,00E-05	3	3,36%	8	99,27%	7,08%	4,50%	99,55%	52,60%
				3	98,68%	15,32%	4,12%	98,93%	54,26%
				2	98,23%	19,90%	3,75%	98,46%	55,08%
60	1,00E-05	4	4,58%	8	99,14%	14,88%	6,83%	99,39%	55,06%
				4	98,73%	20,21%	5,57%	98,97%	55,87%
				6	98,99%	17,02%	6,28%	99,24%	55,38%

Tabela 29 – Treinamentos: Hemorragias e microaneurismas

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
<b>150</b>	<b>1,00E-04</b>	<b>1</b>	<b>17,48%</b>	<b>8</b>	<b>98,48%</b>	<b>23,60%</b>	<b>36,83%</b>	<b>99,47%</b>	<b>64,60%</b>
					98,33%	27,19%	32,81%	99,27%	64,40%
50	1,00E-04	10	4,06%	8	91,25%	26,71%	4,25%	92,10%	53,58%
					92,50%	23,46%	4,46%	93,40%	53,46%
					93,49%	20,68%	4,66%	94,44%	53,32%
60	1,00E-04	7	14,63%	8	97,41%	34,03%	20,29%	98,24%	62,49%
					97,14%	36,67%	18,91%	97,93%	62,66%
					97,47%	33,36%	20,64%	98,32%	62,45%
60	1,00E-04	29	14,64%	8	96,08%	45,28%	15,45%	96,74%	63,39%
					97,39%	34,32%	20,20%	98,22%	62,53%
					97,42%	33,99%	20,34%	98,25%	62,50%
80	1,00E-04	19	14,98%	8	97,23%	36,82%	19,65%	98,02%	62,93%
					97,51%	33,78%	21,17%	98,35%	62,70%
					97,53%	33,56%	21,27%	98,37%	62,68%

Tabela 30 – Treinamento: Todas as imagens somadas

Épocas	Decaimento	Threshold	IoU (100)	Threshold	Acurácia	Sensibilidade	Precisão	Especificidade	Percentual médio
50	2,00E-04	30	19,96%	8	87,39%	39,43%	9,23%	88,78%	56,21%
				29	96,64%	11,87%	27,52%	99,10%	58,78%
				30	96,70%	11,42%	28,36%	99,17%	58,91%
<b>60</b>	<b>1,00E-05</b>	<b>13</b>	<b>43,59%</b>	8	97,41%	40,10%	55,41%	99,07%	73,00%
				12	97,49%	38,21%	58,05%	99,20%	73,24%
				<b>13</b>	<b>97,50%</b>	<b>37,82%</b>	<b>58,57%</b>	<b>99,23%</b>	<b>73,28%</b>
60	3,00E-05	13	37,22%	8	97,00%	38,31%	46,06%	98,70%	70,02%
				12	97,10%	36,41%	48,04%	98,86%	70,10%
				13	97,12%	36,02%	48,44%	98,89%	70,12%
60	3,00E-05	14	42,02%	8	97,20%	40,76%	50,36%	98,84%	71,79%
				13	97,28%	39,03%	52,26%	98,97%	71,89%
				14	97,29%	38,75%	52,56%	98,99%	71,90%
60	1,00E-05	19	41,11%	8	97,26%	42,10%	51,46%	98,85%	72,42%
				18	97,43%	37,75%	56,55%	99,16%	72,72%
				19	97,44%	37,44%	56,88%	99,18%	72,74%
60	1,00E-04	29	29,81%	8	95,56%	48,27%	31,24%	96,93%	68,00%
				28	96,20%	41,76%	35,20%	97,78%	67,74%
				29	96,22%	41,56%	35,31%	97,80%	67,72%
60	1,00E-05	15	45,01%	8	97,38%	40,40%	54,60%	99,03%	72,85%
				14	97,49%	37,82%	58,39%	99,22%	73,23%
				15	97,51%	37,49%	58,86%	99,24%	73,27%
60	1,00E-04	23	34,62%	8	96,68%	43,81%	41,47%	98,21%	70,04%
				22	96,96%	39,45%	45,37%	98,63%	70,10%
				23	96,97%	39,24%	45,55%	98,64%	70,10%
50	2,00E-04	30	19,96%	8	87,39%	39,43%	9,23%	88,78%	56,21%
				29	96,64%	11,87%	27,52%	99,10%	58,78%
				30	96,70%	11,42%	28,36%	99,17%	58,91%



## **ANEXOS**

