UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação

Dissertação

Uma nova abordagem baseada em modelos do tipo *encoder-decoder* para segmentação semântica de lesões de fundo associadas à retinopatia diabética

Alejandro da Silva Pereira

Pelotas, 2025

Alejandro da Silva Pereira

Uma nova abordagem baseada em modelos do tipo *encoder-decoder* para segmentação semântica de lesões de fundo associadas à retinopatia diabética

Dissertação apresentada ao Programa de Pós-Graduação em Computação do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Marilton Sanchotene de Aguiar
Coorientador: Prof. Dr. Carlos Alexandre Silva dos Santos

Pelotas, 2025

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação da Publicação

P436n Pereira, Alejandro da Silva

Uma nova abordagem baseada em modelos do tipo encoder-decoder para segmentação semântica de lesões de fundo associadas à retinopatia diabética [recurso eletrônico] / Alejandro da Silva Pereira ; Marilton Sanchotene de Aguiar, orientador ; Carlos Alexandre Silva dos Santos, coorientador. — Pelotas, 2025.
92 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2025.

1. Retinopatia diabética. 2. Imagens de fundo. 3. Segmentação de lesões. 4. Aprendizado profundo. I. Aguiar, Marilton Sanchotene de, orient. II. Santos, Carlos Alexandre Silva dos, coorient. III. Título.

CDD 005

Elaborada por Maria Inez Figueiredo Figas Machado CRB: 10/1612

Alejandro da Silva Pereira

Uma nova abordagem baseada em modelos do tipo *encoder-decoder* para segmentação semântica de lesões de fundo associadas à retinopatia diabética

Dissertação aprovada, como requisito parcial, para obtenção do grau de Mestre em Ciência da Computação, Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas.

Data da Defesa: 7 de julho de 2025

Banca Examinadora:

Prof. Dr. Marilton Sanchotene de Aguiar (orientador)
Doutor em Computação pela Universidade Federal do Rio Grande do Sul.

Prof. Dr. Carlos Alexandre Silva dos Santos (coorientador)
Doutor em Computação pela Universidade Federal de Pelotas.

Prof. Dr. Lucas Ferrari de Oliveira
Doutor em Ciências Médicas pela Universidade de São Paulo.

Prof. Dr. Ulisses Brisolara Corrêa
Doutor em Computação pela Universidade Federal de Pelotas.

Dedico esta dissertação à minha mãe, à minha namorada Renata e, especialmente, à minha querida avó Marly, que, infelizmente, não pôde estar presente neste momento tão importante da minha vida, mas seus ensinamentos e valores continuam me acompanhando todos os dias.

AGRADECIMENTOS

Primeiramente, agradeço a Deus, por iluminar meus passos, guiar meu caminho e me dar a sabedoria necessária para enfrentar todos os desafios desta jornada. Sua presença foi fundamental para que eu pudesse superar os obstáculos e concluir este trabalho.

À minha namorada, Renata, minha companheira de vida, agradeço profundamente por ter estado ao meu lado em cada etapa, desde a graduação até aqui. Você tornou cada caminho mais leve e me ajudou a superar todos os desafios, até chegar a este momento tão especial. Obrigado por dividir os dias comigo. Você é minha parceira, minha inspiração e meu porto seguro.

Também agradeço a toda a minha família, especialmente à minha mãe, Alessandra, pela dedicação, amparo e afeto que me concedeu ao longo de toda a minha trajetória. Você sempre me orientou pelos caminhos certos e foi um exemplo de vida para mim. Sou imensamente grato por ter uma mãe tão especial como você.

À minha irmã, Nauana, que, mesmo distante fisicamente, sempre esteve disponível para me ouvir, aconselhar e dar apoio em todos os momentos em que precisei. Nossa conexão como irmãos é muito importante para mim e me fortaleceu em cada etapa dessa caminhada. Muito obrigado por tudo.

Aos meus amigos e colegas, que sempre torceram por mim e estiveram ao meu lado, tornando os dias mais leves, meu muito obrigado. Tenho imensa gratidão por poder contar com vocês.

Aos meus orientadores, professores Marilton Sanchotene de Aguiar e Carlos Alexandre Silva dos Santos, pela confiança depositada em mim desde a graduação, por compartilharem seu conhecimento ao longo de toda essa trajetória e pelo apoio constante. Chegar até aqui só foi possível graças a vocês. Guardarei sempre minha gratidão e a lembrança de tudo que fizeram por mim.

Agradeço ao professor Daniel Welfer, do Departamento de Computação Aplicada da Universidade Federal de Santa Maria, pelo apoio e colaboração desde as pesquisas iniciais da graduação até a conclusão deste trabalho.

Também agradeço ao professor Ulisses Brisolara Corrêa e ao Hub de Inovação em Inteligência Artificial da Universidade Federal de Pelotas, pelo compartilhamento dos recursos computacionais utilizados na realização dos experimentos.

Por fim, deixo meu sincero agradecimento a todos os professores que fizeram parte da minha trajetória desde o ensino médio até os dias atuais, pelo conhecimento, atenção e incentivo que sempre me ofereceram.

A nossa maior glória não reside no fato de nunca cairmos, mas sim em levantarmo-nos sempre depois de cada queda.

— OLIVER GOLDSMITH

RESUMO

PEREIRA, Alejandro da Silva. **Uma nova abordagem baseada em modelos do tipo *encoder-decoder* para segmentação semântica de lesões de fundo associadas à retinopatia diabética**. Orientador: Marilton Sanchotene de Aguiar. 2025. 92 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2025.

O diabetes é uma doença crônica que compromete a capacidade do corpo de produzir ou utilizar insulina de forma adequada. Atualmente, estima-se que aproximadamente 537 milhões de adultos em todo o mundo sejam afetados pela doença, o que representa 10,5% da população entre 20 e 79 anos, com projeção de atingir 783 milhões até 2045. Nesse contexto global, o Brasil ocupa a quarta posição em prevalência, com cerca de 13,7 milhões de casos. A Retinopatia Diabética (RD) é uma das principais complicações microvasculares do diabetes e constitui uma das principais causas de cegueira em adultos em idade produtiva. Como o diagnóstico da RD é geralmente realizado por meio da análise de imagens do fundo de olho – nas quais se identificam lesões como Exsudatos Duros, Hemorragias, Exsudatos Algodonosos e Microaneurismas –, é possível empregar abordagens baseadas em aprendizado profundo para a segmentação dessas lesões. O objetivo deste trabalho é propor uma nova abordagem para aprimorar a precisão na segmentação dessas lesões, contribuindo para um diagnóstico mais eficaz da doença e facilitando o processo de anotação de imagens em novos conjuntos de dados de RD. A abordagem proposta tem como base as arquiteturas U-Net++, Attention U-Net e R2U-Net, integradas por meio de *ensembles*. Além disso, foram aplicadas etapas de pré-processamento, aumento de dados e pré-treinamento com o objetivo de melhorar a extração das características das lesões. Os experimentos utilizaram dois conjuntos de dados públicos de RD. O conjunto DDR foi empregado para treinar, ajustar e avaliar os modelos que compõem a abordagem proposta, alcançando mDC de 63,75% e $mIoU$ de 83,18% na etapa de validação, e mDC de 66,47% e $mIoU$ de 81,15% na etapa de teste. O conjunto IDRiD foi utilizado para avaliar a capacidade de generalização da abordagem proposta em diferentes bases de dados de RD, apresentando $mIoU$ de 84,36% na etapa de validação e 82,84% na etapa de teste. Os resultados obtidos superaram os apresentados em trabalhos relacionados, evidenciando que a integração de *ensembles* de modelos do tipo *encoder-decoder*, combinada às etapas de pré-processamento, aumento de dados e pré-treinamento, constitui uma estratégia eficaz para a segmentação de lesões no fundo de olho.

Palavras-chave: retinopatia diabética; imagens de fundo; segmentação de lesões; aprendizado profundo.

ABSTRACT

PEREIRA, Alejandro da Silva. **A novel encoder-decoder model-based approach for semantic segmentation of fundus lesions associated with diabetic retinopathy**. Advisor: Marilton Sanchotene de Aguiar. 2025. 92 f. Dissertation (Masters in Computer Science) – Technology Development Center, Federal University of Pelotas, Pelotas, 2025.

Diabetes is a chronic disease that affects the body's ability to produce or use insulin properly. Researchers estimate that approximately 537 million adults worldwide are affected by the disease, representing 10.5% of the population between 20 and 79 years old, and they project this number will reach 783 million by 2045. In this global context, Brazil occupies the fourth position in prevalence, with approximately 13.7 million cases. Diabetic Retinopathy (DR) is one of the main microvascular complications of diabetes and one of the leading causes of blindness in adults of working age. Since specialists usually diagnose DR by analyzing fundus images and identifying lesions such as Hard Exudates, Hemorrhages, Soft Exudates, and Microaneurysms, researchers can use deep learning approaches to segment these lesions. This work aims to propose a new approach to improve the accuracy in the segmentation of these lesions, which will contribute to a more effective diagnosis of the disease and facilitate the process of annotating images in new DR datasets. The proposed approach combined the U-Net++, Attention U-Net, and R2U-Net architectures as ensembles. This work applied pre-processing, data augmentation, and pre-training steps to improve the extraction of lesion features. The experiments used two public RD datasets. The DDR dataset supported training, adjustment, and evaluation of the models composing the proposed approach, achieving an mDC of 63.75%, an $mIoU$ of 83.18% in the validation stage, and an mDC of 66.47% and an $mIoU$ of 81.15% in the testing stage. To evaluate the generalization capacity of the proposed approach on different DR datasets, this study employed the IDRiD dataset, reaching an $mIoU$ of 84.36% in the validation stage and 82.84% in the testing stage. The results obtained were superior to those presented in related works, evidencing that integrating ensembles of encoder-decoder type models, combined with pre-processing, data augmentation, and pre-training steps, represents an effective strategy for segmenting fundus lesions.

Keywords: diabetic retinopathy; fundus image; lesions segmentation; deep learning.

LISTA DE FIGURAS

Figura 1 Exemplo de imagem de fundo do olho com a presença de Microaneurisma. 24
Figura 2 Exemplo de imagem de fundo do olho com a presença de Hemorragia. 25
Figura 3 Exemplo de imagem de fundo do olho com Exsudato Duro. 26
Figura 4 Exemplo de imagem de fundo do olho com Exsudato Algodonoso. 26
Figura 5 Exemplos de tarefas de Visão Computacional: Classificação de imagens, detecção de objetos e segmentação de imagens. 28
Figura 6 Tipos de segmentação de imagem. Em (a), a imagem original; (b) Segmentação semântica; (c) Segmentação de instância; e, (d) Segmentação panóptica. 29
Figura 7 Exemplo de modelo estrutural de uma Rede Neural Convolucional composta por camadas de Convolução, *Pooling* e pela Camada Totalmente Conectada. 30
Figura 8 Exemplo de convolução entre uma matriz de entrada de dimensões $3\times3\times3$ e um filtro de tamanho $2\times2\times3$. A operação consiste na multiplicação dos elementos do filtro pelos elementos correspondentes no campo receptivo da matriz de entrada, e a soma dos produtos gerados resulta no valor da saída da convolução. 31
Figura 9 Exemplo de aplicação de *Max pooling* sobre uma matriz de dimensões $4\times4\times3$. O filtro de *pooling* de tamanho 2×2 é aplicado em cada canal da matriz de entrada, com um *stride* de 2, realizando a redução da dimensionalidade. 32
Figura 10 Vetor coluna obtido pela operação de *flattening*, a partir de um mapa de características agrupados pela camada de *pooling*. 33
Figura 11 Diagrama de arquitetura da U-Net. 34
Figura 12 Diagrama de arquitetura da U-Net++. 36
Figura 13 Diagrama de arquitetura da Attention U-Net. 36
Figura 14 Diagrama de arquitetura da R2U-Net. 38
Figura 15 Diagrama de blocos da abordagem proposta para segmentação de lesões de fundo. Inicialmente, as imagens passam pelo bloco de pré-processamento. Em seguida, são transferidas para o bloco de aumento de dados. Na etapa seguinte, os modelos são inicializados com pesos pré-treinados no conjunto de dados ImageNet. Por fim, os modelos treinados são integrados por meio de *ensembles* para realizar as predições das lesões. 43

Figura 16 Gráfico distribuição das resolução das imagens com anotações a nível de píxel do conjunto DDR. 45
Figura 17 Exemplo de imagem de fundo do conjunto de dados DDR e diferentes anotações em nível de píxel. Em (a), a imagem de fundo original; (b) anotações a nível de píxel dos Exsudatos Duros; (c) anotações a nível de píxel das Hemorragias; (d) anotações a nível de píxel dos Exsudatos Algodonosos; e, (e) anotações a nível de píxel dos Microaneurismas. 46
Figura 18 Exemplo de imagem de fundo do conjunto de dados DDR com as anotações de caixa delimitadora das lesões. Em (a), a imagem de fundo original; e, (b) a imagem com suas anotações de caixa delimitadora. As anotações em amarelo representam Exsudatos Duros; as anotações em vermelho representam Hemorragias; as anotações em azul representam Exsudatos Algodonosos; e, as anotações em verde representam Microaneurismas. 46
Figura 19 Exemplo de imagem de fundo do conjunto de dados IDRiD e diferentes anotações em nível de píxel. Em (a) a imagem de fundo original; (b) anotações a nível de píxel dos Exsudatos Duros; (c) anotações a nível de píxel das Hemorragias; (d) anotações a nível de píxel dos Exsudatos Algodonosos; (e) anotações a nível de píxel dos Microaneurismas; e, (f) a anotação do Disco Óptico. 47
Figura 20 Exemplo de imagem de fundo do conjunto de dados IDRiD com as anotações de caixa delimitadora das lesões. Em (a) a imagem de fundo original; e, (b) a imagem com suas anotações de caixa delimitadora. As anotações em amarelo representam Exsudatos Duros; as anotações em vermelho representam Hemorragias; as anotações em azul representam Exsudatos Algodonosos; e, as anotações em verde representam Microaneurismas. 48
Figura 21 *Pipeline* de pré-processamento da abordagem proposta aplicado a uma imagem de fundo do conjunto de dados DDR. Em (a) Imagem original; (b) Filtro de mediana; (c) CLAHE; (d) *Cropping*; e, (e) SAHI. 49
Figura 22 Resultado aplicação do filtro de mediana de tamanho 5×5 em imagens de fundo do conjunto de dados DDR com a introdução artificial de ruído *Salt & Pepper*. Em (a) Imagens com ruído *Salt & Pepper*; e, (b) Imagens após aplicação do filtro de mediana 5×5. 50
Figura 23 O método CLAHE segmenta a imagem em blocos de 8×8 e realiza a equalização do histograma individualmente em cada uma dessas regiões. Durante esse processo, os valores de píxel que excedem o limiar predefinido são redistribuídos, resultando em um novo histograma com contraste aprimorado. 52
Figura 24 Resultado aplicação do CLAHE com tamanho de bloco de 8×8 e clip limit igual a 6, em imagens de fundo do conjunto de dados DDR com a aplicação do filtro de mediana. Em (a) Imagens com filtro de mediana; e, (b) Imagens após aplicação CLAHE. 54
Figura 25 Transformação da circunferência da retina em seu retângulo equivalente por meio das coordenadas dos cantos superior esquerdo $(x_{\text{min}}, y_{\text{min}})$ e inferior direito $(x_{\text{max}}, y_{\text{max}})$. 57

Figura 26 *Pipeline* para realização do *cropping* do fundo preto das imagens de fundo. Em (a) Imagem original; (b) Imagem suavizada com filtro de mediana; (c) Imagem em tons de cinza; (d) Detecção de bordas com filtro de Sobel; (e) Demarcação do contorno da retina com a transformada de Hough; e, (f) Imagem após realização do *cropping*. 57
Figura 27 Exemplo de execução do método COCO *slice* realizado pelo *framework* SAHI, com blocos de tamanho 1.024×1.024 píxeis e taxa de sobreposição de 20%. Em (a) Imagem de fundo original; (b) *Ground truth* dos EX; (c) *Ground truth* dos HE; (d) *Ground truth* dos SE; e, (e) *Ground truth* dos MA. 59
Figura 28 Exemplo de aumento de dados em uma imagem de fundo pré-processada do conjunto DDR. Em (a) Imagem pré-processada, (b) *HorizontalFlip*, (c) *VerticalFlip*, (d) *ElasticTransform*, (e) *GridDistortion*, e (f) *OpticalDistortion*. 60
Figura 29 Diagrama da relação entre os conceitos de VP, VN, FP e FN, utilizados como base para o cálculo das métricas de avaliação do trabalho proposto. 63
Figura 30 Comparação entre métodos de redimensionamento das máscaras de segmentação. Em (a) Bloco predito (256×256); (b) Redimensionamento convencional (1.024×1.024); (c) Redimensionamento com Real-ESRGAN (1.024×1.024); e, (d) Bloco original gerado pelo SAHI (1.024×1.024). 65
Figura 31 Processo para a reconstrução do mapa de segmentação completo para cada um dos tipos de lesões de fundo. Em (a) Reconstrução do mapa de segmentação dos EX; (b) Reconstrução do mapa de segmentação dos HE; (c) Reconstrução do mapa de segmentação dos SE; e, (d) Reconstrução do mapa de segmentação dos MA. . . 67
Figura 32 Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta e o *ground truth* em imagens do conjunto de teste do *dataset* DDR. (a) Imagens pré-processadas, (b) *Ground truth*, e (c) Abordagem proposta. 71
Figura 33 Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta com os modelos que a compõe em imagens do conjunto de teste do *dataset* DDR. (a) Imagens pré-processadas, (b) *Ground truth*, (c) U-Net++, (d) Attention U-net, (e) R2U-Net, e (f) Abordagem proposta. 72
Figura 34 Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta e o *ground truth* em imagens do conjunto de teste do *dataset* IDRiD. (a) Imagens pré-processadas, (b) *Ground truth*, e (c) Abordagem proposta. 74
Figura 35 Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta com os modelos que a compõe em imagens do conjunto de teste do *dataset* IDRiD. (a) Imagens pré-processadas, (b) *Ground truth*, (c) U-Net++, (d) Attention U-net, (e) R2U-Net, e (f) Abordagem proposta. 75

Figura 36 Exemplo de imagem de fundo do novo conjunto de dados de RD e as anotações em nível de píxel geradas pelos modelos que compõem a abordagem proposta. Em (a), a imagem de fundo original; (b) anotações a nível de píxel dos Exsudatos Duros; (c) anotações a nível de píxel das Hemorragias; (d) anotações a nível de píxel dos Exsudatos Algodonosos; e, (e) anotações a nível de píxel dos Microaneurismas. 76
Figura 37 Exemplo de imagem de fundo do novo conjunto de dados de RD com as anotações de caixa delimitadora das lesões geradas pelos modelos que compõem a abordagem proposta. Em (a), a imagem de fundo original; e, (b) a imagem com suas anotações de caixa delimitadora. As anotações em amarelo representam Exsudatos Duros; as anotações em vermelho representam Hemorragias; as anotações em azul representam Exsudatos Algodonosos; e, as anotações em verde representam Microaneurismas. 76

LISTA DE TABELAS

Tabela 1 Análise comparativa dos trabalhos relacionados na segmentação das lesões retinianas. 41

Tabela 2 Número de amostras de cada classe nos conjuntos de Treino, Validação e Teste. 44

Tabela 3 Número de imagens contidas em cada conjunto de dados após aplicação do método *Hold-out validation*. 48

Tabela 4 Valores obtidos com a métrica $PSNR$ após a remoção de ruídos do tipo *Salt & Pepper* por meio da aplicação do filtro de mediana com janela de 5$\times$5 em imagens de fundo do conjunto de dados DDR. . . 51

Tabela 5 Valores obtidos com as métrica EME e E após a aplicação do CLAHE em imagens de fundo do conjunto de dados DDR previamente processadas com o filtro de mediana. 55

Tabela 6 Valores de parâmetros usados ao aplicar métodos de aumento de dados nas imagens de fundo de olho. 60

Tabela 7 Hiperparâmetros utilizados nos modelos da abordagem proposta. . 62

Tabela 8 Pesos atribuídos às arquiteturas para cada tipo de lesão após a otimização com o algoritmo de Evolução Diferencial. 62

Tabela 9 Resultados obtidos na segmentação de Exsudatos Duros, Hemorragias, Exsudatos Algodonosos e Microaneurismas comparados com outros modelos utilizando as métricas Acc, Sen, Pre, DC e IoU no conjunto de validação do *dataset* DDR. O símbolo $-$ indica que a métrica correspondente não foi reportada no trabalho original do respectivo modelo. 69

Tabela 10 Resultados obtidos na segmentação de Exsudatos Duros, Hemorragias, Exsudatos Algodonosos e Microaneurismas comparados com outros modelos utilizando as métricas Acc, Sen, Pre, DC e IoU no conjunto de teste do *dataset* DDR. O símbolo $-$ indica que a métrica correspondente não foi reportada no trabalho original do respectivo modelo. 70

Tabela 11 Resultados obtidos na segmentação de lesões de fundo de olho comparados aos demais modelos utilizando as métricas IoU e $mIoU$ no conjunto de validação do *dataset* IDRiD. 73

Tabela 12 Resultados obtidos na segmentação de lesões de fundo de olho comparados aos demais modelos utilizando as métricas IoU e $mIoU$ no conjunto de teste do *dataset* IDRiD. 73

Tabela 13 Classificação Internacional da Retinopatia Diabética (a) e do Edema Macular Diabético (b) com base na oftalmoscopia com dilatação. . . 90

Tabela 14 Características Clínicas da Retinopatia Diabética. 92

LISTA DE ABREVIATURAS E SIGLAS

21HT	*Hough Gradient 2-1 Hough Transform*
Acc	*Accuracy*
AFMA	*Across Feature Map Attention*
AM	Aprendizado de Máquina
AP	*Average Precision*
CBAM	*Convolutional Block Attention Module*
CHASEDB1	*Child Heart and Health Study in England Database 1*
CLAHE	*Contrast-Limited Histogram Adaptive Equalization*
CNN	*Convolutional Neural Network*
DC	*Dice Coefficient*
DDR	*Dataset for Diabetic Retinopathy*
DenseNet	*Densely Connected Convolutional Network*
DIARETDB1	*Diabetic Retinopathy Database 1*
DO	Disco Óptico
DRIVE	*Digital Retinal Images for Vessel Extraction*
E	Entropia
EMD	Edema Macular Diabético
EME	*Measure of Enhancement*
EX	Exsudatos Duros
FN	Falsos Negativos
FP	Falsos Positivos
GAN	*Generative Adversarial Networks*
gt	*ground truth*
HE	Hemorragias
HT	Transformada de Hough
IA	Inteligência Artificial

IBGE	Instituto Brasileiro de Geografia e Estatística
ICO	*International Council of Ophthalmology*
IDRiD	*Indian Diabetic Retinopathy Image Dataset*
IoU	*Intersection Over Union*
JPG	*Joint Photographic Experts Group*
LAB	Luminância, A (eixo verde-vermelho) e B (eixo azul-amarelo)
MA	Microaneurismas
mAP	*mean Average Precision*
mDC	*mean Dice Coefficient*
$mIoU$	*mean Intersection Over Union*
MSE	*Mean Squared Error*
NALoss	*Noise Adjusted Loss*
ODConv	*Omni-Dimensional Dynamic Convolution*
PDI	Processamento Digital de Imagens
pm	*predicted mask*
Pre	*Precision*
$PSNR$	*Peak Signal-to-Noise Ratio*
RAM	*Random Access Memory*
RCL	*Recurrent Convolutional Layers*
RCNN	*Recurrent Convolutional Neural Network*
RD	Retinopatia Diabética
RDNP	Retinopatia Diabética Não-Proliferativa
RDP	Retinopatia Diabética Proliferativa
ReLU	*Rectified Linear Unit*
ResNet	*Deep Residual Network*
SAHI	*Slicing Aided Hyper Inference*
SE	Exsudatos Algodonosos
Sen	*Sensitivity*
SUS	Sistema Único de Saúde
TIFF	*Tagged Image File Format*
VC	Visão Computacional
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos
VRAM	*Video Random Access Memory*

SUMÁRIO

1 INTRODUÇÃO 19
1.1 Problema de Pesquisa 20
1.2 Hipóteses 21
1.3 Objetivos 21
1.4 Organização do Texto 21

2 REFERENCIAL TEÓRICO E TECNOLÓGICO 23
2.1 Lesões de Fundo 23
2.1.1 Microaneurismas 23
2.1.2 Hemorragias 24
2.1.3 Exsudatos Duros 25
2.1.4 Exsudatos Algodonosos 25
2.2 Segmentação Semântica de Objetos com Redes Neurais Profundas . 27
2.2.1 Redes Neurais Convolucionais 28
2.2.2 Arquiteturas *encoder-decoder* 32
2.3 Trabalhos Relacionados 38

3 ABORDAGEM PROPOSTA 42
3.1 Metodologia 42
3.1.1 Conjunto de Dados 43
3.1.2 Pré-processamento 48
3.1.3 Aumento de Dados 58
3.1.4 Treinamento e Integração dos Modelos 60
3.1.5 Métricas de Avaliação 63
3.1.6 Reconstrução do Mapa de Segmentação 64
3.2 Experimentos, Resultados e Discussões 68

4 CONCLUSÃO 77

REFERÊNCIAS 80

ANEXO A CLASSIFICAÇÃO DA RETINOPATIA DIABÉTICA E EDEMA MACULAR 89

ANEXO B CARACTERÍSTICAS CLÍNICAS DA RETINOPATIA DIABÉTICA . 91

1 INTRODUÇÃO

O diabetes é uma doença crônica grave que ocorre quando a produção de insulina é insuficiente ou quando a insulina produzida não é adequadamente utilizada pelo organismo. Estima-se que cerca de 537 milhões de adultos, com idades entre 20 e 79 anos, sejam afetados pela doença em todo o mundo, o que representa 10,5% da população nessa faixa etária. Esse número deve aumentar para 643 milhões até 2030 e para 783 milhões até 2045 (Kumar et al., 2024). O Brasil ocupa a quarta posição entre os países com maior prevalência de diabetes, com aproximadamente 13,7 milhões de brasileiros afetados, muitos dos quais ainda não foram diagnosticados. Um estudo de 2013, realizado pelo Ministério da Saúde em parceria com o Instituto Brasileiro de Geografia e Estatística (IBGE), revelou que 6,2% da população brasileira com mais de 18 anos era portadora da doença (Pereira et al., 2020).

Os altos níveis de glicose no sangue podem causar danos significativos aos órgãos, aumentando a probabilidade de diversas complicações – em especial a Retinopatia Diabética (RD), uma das principais complicações microvasculares do diabetes e uma causa relevante de cegueira entre adultos em idade produtiva. Estima-se que cerca de 20% dos pacientes diabéticos desenvolvam essa condição, e o número de casos de perda de visão decorrente da RD continua a crescer globalmente (Yu et al., 2023). Diante desse cenário, a identificação precoce da RD é fundamental para ampliar a eficácia dos tratamentos e prevenir a progressão para formas mais graves da doença (Mendes, 2017).

O diagnóstico da RD geralmente é realizado por meio da análise de imagens do fundo de olho, nas quais são identificadas lesões como Microaneurismas (MA), Hemorragias (HE), Exsudatos Duros (EX) e Exsudatos Algodonosos (SE). A RD pode ser classificada em Retinopatia Diabética Não-Proliferativa (RDNP) ou Retinopatia Diabética Proliferativa (RDP). A RDNP é subdividida em estágios que variam de leve a moderado e grave, conforme a progressão das lesões na retina. Essas lesões se iniciam com MA isolados e podem evoluir para pequenas HE, além de apresentarem sinais indiretos de hiperpermeabilidade vascular.

A presença de EX e SE reflete o impacto da doença nos vasos sanguíneos retinianos (Sivaprasad; Sen; Cunha-vaz, 2023). No estágio grave da RDNP, há uma elevada probabilidade de progressão rápida para a RDP, caracterizada pela formação de neovasos. Esses novos vasos sanguíneos são mais frágeis, o que aumenta o risco de ruptura e consequente sangramento no humor vítreo, podendo resultar em perda de visão. Detalhes adicionais sobre a classificação e as características da RD podem ser consultados nos Anexos A e B.

Recentemente, abordagens baseadas em aprendizado profundo têm demonstrado grande eficácia na segmentação das lesões mencionadas anteriormente. Em especial, arquiteturas do tipo *encoder-decoder*, como a U-Net (Ronneberger; Fischer; Brox, 2015), têm obtido resultados promissores em aplicações de segmentação de imagens médicas (Wang et al., 2023).

Atualmente, existem diversos estudos na literatura cujo foco é a segmentação de lesões no fundo de olho associadas à RD, como os de Li et al. (2019); Raudonis et al. (2023); Nair et al. (2023); Anand et al. (2023); Pereira et al. (2024); Xu et al. (2024); Jiang; Zhao (2024); Zhang; Xu; Zhai (2025). Contudo, apesar do uso de modelos de redes neurais profundas nesses trabalhos, os resultados ainda apresentam limitações relevantes.

Tais limitações se devem, principalmente, à escassez de lesões rotuladas e à baixa qualidade das imagens disponíveis nos conjuntos de dados públicos de RD. Além disso, a limitada representatividade das características extraídas durante o treinamento dos modelos – agravada pelo tamanho reduzido de algumas lesões, como os MA – também constitui um desafio significativo para a segmentação precisa.

Neste contexto, propõe-se uma nova abordagem para a segmentação das lesões no fundo de olho, com o objetivo de aumentar a precisão desse processo. Tal aprimoramento pode contribuir significativamente para um diagnóstico mais eficaz da RD, além de facilitar a anotação das lesões e viabilizar a criação de novos conjuntos de dados voltados a essa condição. Esses aspectos constituem os principais motivadores do problema de pesquisa abordado neste trabalho.

1.1 Problema de Pesquisa

Este trabalho tem como objetivo responder à seguinte questão de pesquisa:

1. Como aprimorar modelos baseados em aprendizado profundo para detectar lesões retinianas – especialmente Microaneurismas – diante da escassez de imagens com anotações a nível de píxel nos conjuntos de dados públicos de Retinopatia Diabética?

1.2 Hipóteses

As seguintes hipóteses foram estabelecidas nesta Dissertação:

1. Arquiteturas do tipo *encoder-decoder* apresentam maior precisão na segmentação de lesões associadas à RD quando comparadas a detectores de estágio único e de dois estágios.

2. A combinação de modelos *encoder-decoder*, aliada à aplicação de etapas de pré-processamento, aumento de dados e pré-treinamento, resulta em maior precisão na segmentação de lesões no fundo de olho.

1.3 Objetivos

O objetivo geral deste trabalho é propor uma nova abordagem para aprimorar a precisão na segmentação de lesões no fundo de olho associadas à Retinopatia Diabética. A proposta consiste na criação de *ensembles* de diferentes arquiteturas *encoder-decoder*, em conjunto com etapas de pré-processamento, aumento de dados e pré-treinamento, com o intuito de auxiliar no diagnóstico precoce da doença e facilitar a criação de novos conjuntos de dados.

Para atingir o objetivo geral, foram definidos os seguintes objetivos específicos:

1. Desenvolvimento de uma nova abordagem para segmentação de lesões de fundo com base em arquiteturas *encoder-decoder*;

2. Definição da metodologia para execução das etapas de pré-processamento, aumento de dados e pré-treinamento;

3. Avaliação das arquiteturas para composição da abordagem proposta;

4. Integração dos modelos treinados na forma de *ensembles*;

5. Avaliação da eficácia da abordagem proposta em comparação com trabalhos relacionados disponíveis na literatura; e

6. Automatização da anotação de um novo conjunto de dados de imagens de fundo utilizando a abordagem desenvolvida.

1.4 Organização do Texto

A estrutura desta Dissertação está organizada da seguinte forma: no Capítulo 2, apresenta-se o referencial teórico e tecnológico, abordando Lesões de Fundo (Seção 2.1), Segmentação Semântica de Objetos com Redes Neurais Profundas (Seção 2.2)

e Trabalhos Relacionados (Seção 2.3). O Capítulo 3 descreve os materiais e métodos utilizados na abordagem proposta, bem como os experimentos realizados, os resultados obtidos e as respectivas discussões. Por fim, o Capítulo 4 apresenta as considerações finais do trabalho.

2 REFERENCIAL TEÓRICO E TECNOLÓGICO

Este capítulo tem como objetivo apresentar os conceitos fundamentais necessários para a compreensão do trabalho desenvolvido nesta Dissertação. Inicialmente, são descritas as principais lesões de fundo abordadas no estudo. Em seguida, são introduzidos os conceitos relacionados à segmentação semântica de objetos com o uso de Redes Neurais Profundas. Por fim, é realizada uma revisão da literatura, com foco em trabalhos cujo objetivo é a segmentação de lesões no fundo de olho.

2.1 Lesões de Fundo

A retina é a camada mais interna do globo ocular e tem a função de captar as imagens, convertendo-as em impulsos elétricos que são transmitidos ao cérebro por meio do nervo óptico (Pedro, 2022). Em pacientes diabéticos, essa estrutura pode ser comprometida pela RD, que ocorre quando substâncias anormais se depositam nas paredes dos vasos sanguíneos da retina, especialmente na região conhecida como fundo do olho. Esse processo pode levar ao estreitamento ou até ao bloqueio desses vasos, além de provocar o enfraquecimento de suas paredes e o surgimento de diversas lesões retinianas.

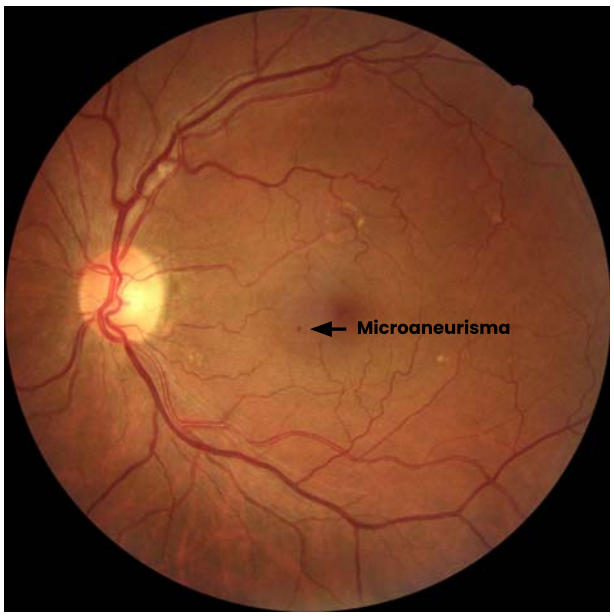
O exame de fundo de olho é essencial para a identificação dessas lesões e para o diagnóstico de doenças oftalmológicas que afetam a retina (Santos, 2023). As Seções a seguir apresentam as principais lesões abordadas neste trabalho.

2.1.1 Microaneurismas

Os Microaneurismas são os primeiros sinais clínicos da RD e manifestam-se como pequenas lesões arredondadas e avermelhadas no fundo da retina. São pontos isolados de diferentes tamanhos, resultantes do enfraquecimento das paredes dos vasos sanguíneos, o que leva ao seu inchaço e consequente dilatação (Soares; Castelobranco; Pinheiro, 2023). Geralmente, apresentam diâmetro reduzido, variando entre 10 e 100 μm.

Embora não impactem diretamente a visão, sua presença indica danos iniciais nos vasos sanguíneos da retina, funcionando como um sinal de alerta para a progressão da doença (Souza, 2023). A ocorrência isolada de Microaneurismas é indicativa de um estágio leve de Retinopatia Diabética Não Proliferativa (RDNP), podendo refletir tentativas malsucedidas de neovascularização ou enfraquecimento das paredes capilares (International Council of Ophthalmology, 2017). A Figura 1 apresenta um exemplo de Microaneurisma em uma imagem de fundo de olho.

Figura 1 – Exemplo de imagem de fundo do olho com a presença de Microaneurisma.

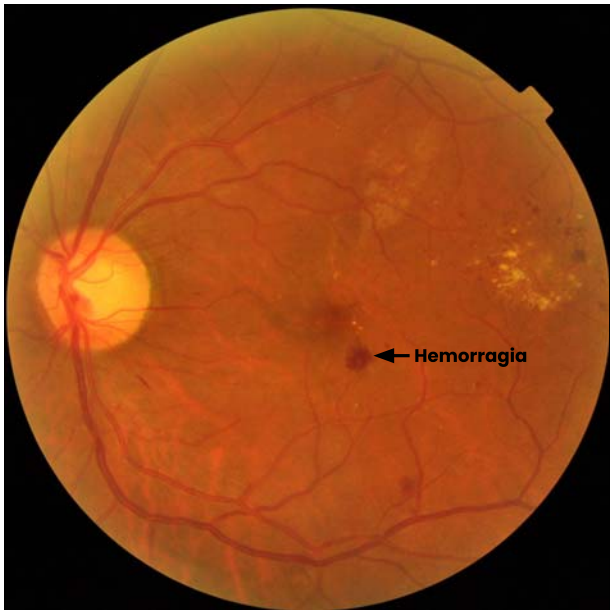


Fonte: Adaptado do conjunto de dados DDR (Li et al., 2019).

2.1.2 Hemorragias

As Hemorragias surgem a partir da evolução dos Microaneurismas, manifestando-se como manchas vermelhas decorrentes do acúmulo de sangue em áreas da retina, causado pelo rompimento de veias e artérias retinianas (Costa, 2019). Os principais tipos são as Hemorragias em mancha e as Hemorragias puntiformes. As Hemorragias em mancha ocorrem nas camadas mais superficiais da retina, sendo formadas em regiões onde há obstrução de aglomerados de vasos capilares, resultando em manchas avermelhadas com margens e densidade irregulares. Por outro lado, as Hemorragias puntiformes ocorrem em camadas mais profundas e, devido à semelhança visual com os Microaneurismas, sua diferenciação nem sempre é possível – ainda que apresentem tamanhos distintos (International Council of Ophthalmology, 2017). A Figura 2 apresenta um exemplo de imagem de fundo de olho com presença de Hemorragia, cujo tamanho geralmente ultrapassa os 125 μm.

Figura 2 – Exemplo de imagem de fundo do olho com a presença de Hemorragia.



Fonte: Adaptado do conjunto de dados DDR (Li et al., 2019).

2.1.3 Exsudatos Duros

Os Exsudatos Duros formam-se quando as junções entre as células do epitélio pigmentar da retina são rompidas, provocando o vazamento de fluido rico em lipídeos e proteínas a partir dos capilares. Essas lesões apresentam coloração amarelada, contornos irregulares e, geralmente, localizam-se na camada plexiforme externa da retina. Podem estar associadas à hiperlipidemia (Flaxel et al., 2020). São considerados marcadores de Edema Macular, atual ou pregresso, sendo que o Edema Macular Diabético (EMD) ocorre quando os vasos sanguíneos da retina retêm líquido em excesso, promovendo o inchaço da região macular e comprometendo sua função. O EMD representa uma das principais causas de cegueira entre pacientes diabéticos (International Council of Ophthalmology, 2017).

Adicionalmente, os Exsudatos Duros costumam localizar-se ao redor ou nas proximidades da mácula e apresentam características refrativas nos exames de retinografia (Cardoso, 2019). Podem ocorrer de diferentes formas: em agrupamentos, como anéis parciais ou completos, de forma isolada ou em trilhas confluentes. A Figura 3 apresenta um exemplo de Exsudato Duro em uma imagem de fundo de olho.

2.1.4 Exsudatos Algodonosos

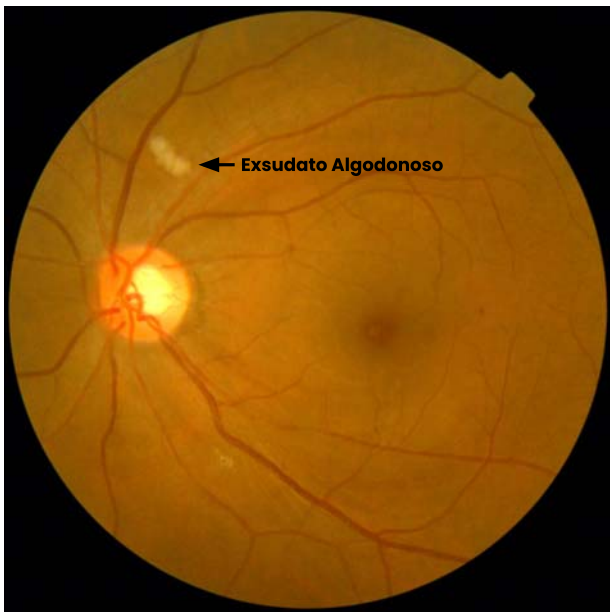
Os Exsudatos Algodonosos apresentam aspecto pálido, geralmente em tonalidade branco-acinzentada (Cardoso, 2019). Essas lesões surgem como resposta à isquemia retiniana, sendo comuns durante a fase pré-proliferativa da Retinopatia Diabética, caracterizada por infarto retiniano, isquemia progressiva e anormalidades vasculares

Figura 3 – Exemplo de imagem de fundo do olho com Exsudato Duro.

Fonte: Adaptado do conjunto de dados DDR (Li et al., 2019).

intrarretinianas (Francelino et al., 2023). Também conhecidos como Exsudatos Moles ou manchas algodonosas, recebem esse nome devido à sua semelhança visual com pedaços de algodão. Essas lesões costumam surgir nos estágios mais avançados da doença e não apresentam bordas bem definidas (Basha; Prasad, 2008). A Figura 4 apresenta um exemplo de Exsudato Algodonoso em uma imagem de fundo de olho.

Figura 4 – Exemplo de imagem de fundo do olho com Exsudato Algodonoso.



Fonte: Adaptado do conjunto de dados DDR (Li et al., 2019).

2.2 Segmentação Semântica de Objetos com Redes Neurais Profundas

O Aprendizado de Máquina (AM) é um ramo da Inteligência Artificial (IA) que permite que sistemas aprendam a partir de dados, sem a necessidade de serem programados explicitamente para executar tarefas. Por meio de algoritmos, as máquinas podem ser treinadas para realizar ações complexas e tomar decisões de forma autônoma, com pouca ou nenhuma intervenção humana (Charleaux; Toledo, 2025).

No entanto, alguns problemas exigem etapas adicionais de pré-processamento dos dados, como normalização, padronização e extração de características. Dentre essas, a extração de características se destaca por exigir conhecimento prévio sobre o domínio da aplicação por parte do desenvolvedor, que deve identificar os atributos mais relevantes para representar o problema e permitir que o modelo aprenda de maneira eficaz (Lecun; Bengio; Hinton, 2015).

O Aprendizado Profundo, por sua vez, é uma subárea do AM que permite aos computadores aprender a partir da experiência, organizando as informações em múltiplos níveis de abstração. Como esse processo de aprendizado ocorre de forma autônoma, não é necessário que um operador humano especifique previamente todo o conhecimento necessário para o funcionamento do sistema (Goodfellow et al., 2016).

A Visão Computacional (VC) é uma subárea da Inteligência Artificial (IA) voltada à análise e automação de tarefas de percepção visual realizadas por computadores. Ela abrange uma ampla gama de aplicações, com destaque para a área da saúde, onde tem transformado processos clínicos e ampliado as possibilidades de diagnóstico, monitoramento e intervenção. A notável capacidade dos computadores de identificar elementos em imagens complexas é viabilizada por cálculos matemáticos avançados e pelo treinamento de redes neurais, que aprendem a reconhecer padrões visuais. Dessa forma, simula-se o processo de aprendizagem humana, possibilitando a interpretação automatizada e precisa de dados visuais, como imagens médicas e vídeos cirúrgicos (Tavares et al., 2024).

Entre as tarefas mais comuns em VC, destacam-se a classificação de imagens, a detecção de objetos e a segmentação de imagens. Conforme ilustrado na Figura 5, a classificação de imagens consiste em atribuir um rótulo predefinido a uma imagem, como, por exemplo, determinar se ela contém ou não um determinado objeto. A detecção de objetos, por sua vez, vai além da classificação, identificando e localizando múltiplos objetos na imagem por meio de caixas delimitadoras. Já a segmentação de imagens representa o nível mais detalhado de compreensão visual, atribuindo um rótulo a cada píxel da imagem, o que permite a separação precisa entre diferentes regiões pertencentes aos objetos (Kotra, 2023).

Figura 5 – Exemplos de tarefas de Visão Computacional: Classificação de imagens, detecção de objetos e segmentação de imagens.

Fonte: Adaptado de Kotra (2023).

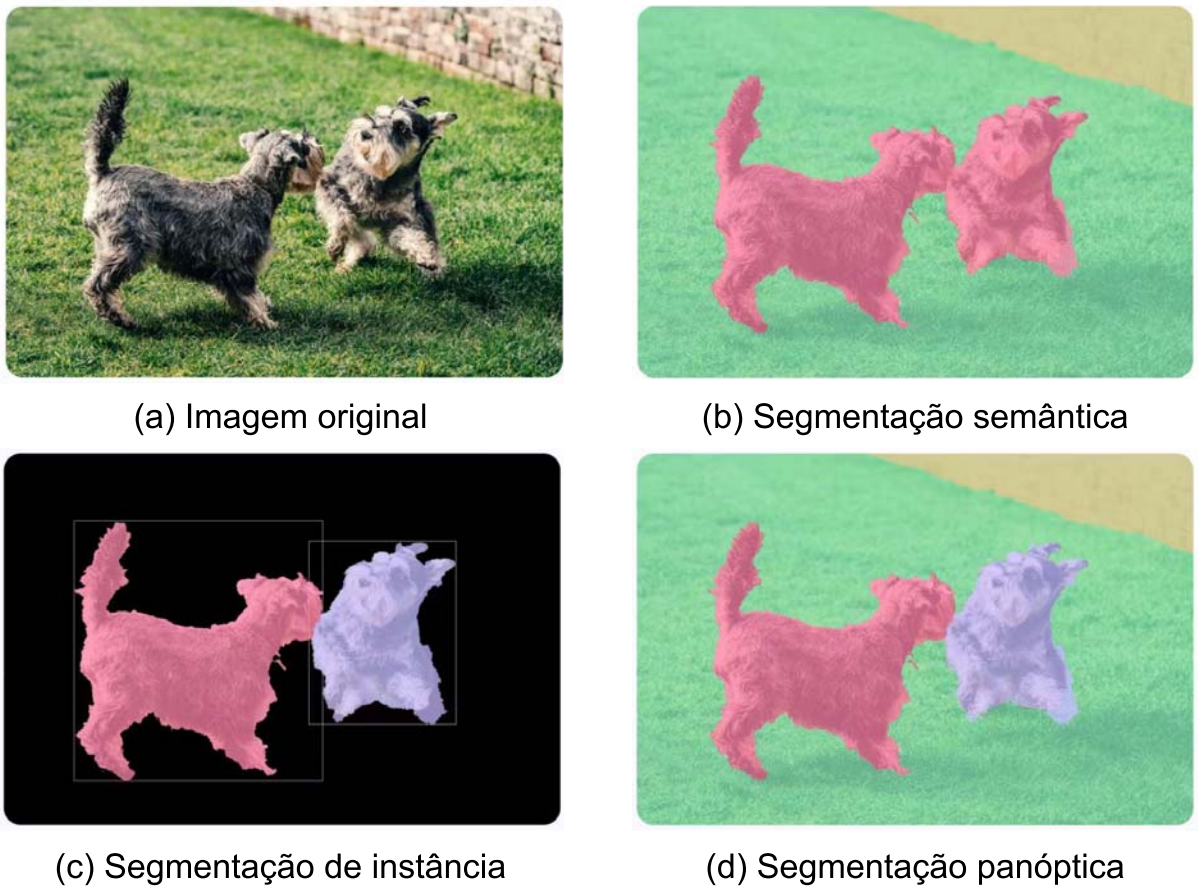
Dentre as tarefas de Visão Computacional, a segmentação de imagens recebe atenção especial neste trabalho, uma vez que é empregada na identificação de lesões no fundo de olho associadas à RD. A segmentação de imagens pode ser classificada em diferentes tipos, a depender do nível de granularidade e dos objetivos da aplicação. Conforme ilustrado na Figura 6, a segmentação semântica atribui um rótulo a cada píxel da imagem com base na classe a que pertence, sem, no entanto, distinguir instâncias individuais de um mesmo objeto. Já a segmentação por instância vai além da classificação por píxel, diferenciando múltiplas ocorrências de um mesmo objeto dentro da imagem. Por fim, a segmentação panóptica combina as abordagens anteriores, identificando e rotulando todas as regiões da imagem, separando instâncias individuais e distinguindo objetos de fundo (Acharya, 2024).

Para que a segmentação semântica produza resultados precisos – especialmente em contextos complexos, como o das imagens médicas – é essencial a utilização de modelos capazes de extrair e interpretar características visuais de forma eficiente. Nesse cenário, as Redes Neurais Convolucionais (em inglês, *Convolutional Neural Networks* – CNNs) desempenham um papel central, servindo como base para diversas abordagens voltadas à segmentação de imagens.

2.2.1 Redes Neurais Convolucionais

As Redes Neurais Convolucionais foram desenvolvidas com base na inspiração biológica do funcionamento do córtex visual, especialmente na forma como diferentes áreas desse córtex correspondem a regiões específicas do campo visual. Cada uma dessas regiões, conhecidas como campos receptivos, é responsável pela ativação de determinados neurônios, apresentando, muitas vezes, sobreposição com os campos

Figura 6 – Tipos de segmentação de imagem. Em (a), a imagem original; (b) Segmentação semântica; (c) Segmentação de instância; e, (d) Segmentação panóptica.



(a) Imagem original (b) Segmentação semântica

(c) Segmentação de instância (d) Segmentação panóptica

Fonte: Adaptado de Acharya (2024).

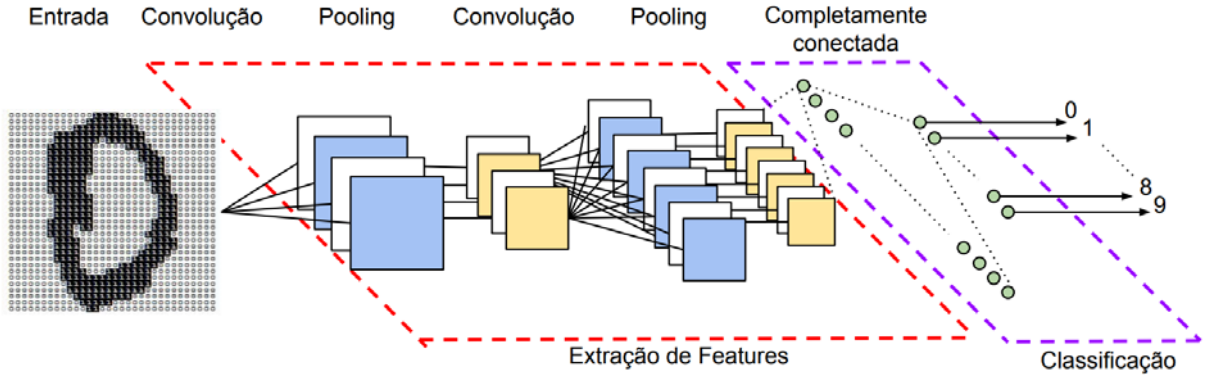
de neurônios vizinhos. Essa organização do sistema visual humano foi essencial para o desenvolvimento de métodos computacionais capazes de extrair características relevantes diretamente dos dados de entrada (Oliveira; Alves; Malqui, 2017).

As CNNs podem ser compreendidas como uma extensão das Redes Neurais Perceptrons Multicamadas (Parlos et al., 1994), projetadas especificamente para aplicar filtros a dados visuais, preservando a estrutura espacial e as relações de vizinhança entre os píxeis durante o processamento (Vargas; Paes; Vasconcelos, 2016). Suas camadas são organizadas sequencialmente, de modo que a saída de uma camada serve como entrada para a camada subsequente.

A Figura 7 ilustra a estrutura de uma Rede Neural Convolucional (CNN). Diferentemente dos métodos tradicionais de extração de características – que dependem de técnicas manuais para identificar e selecionar padrões específicos nas imagens –, as CNNs realizam essa tarefa de forma automatizada por meio de operações de convolução (Li et al., 2022). Essas redes são compostas por múltiplas camadas, cada uma com funções específicas durante o processo de aprendizado e classificação de

imagens, incluindo a Camada de Convolução, a Camada de *Pooling* e a Camada Totalmente Conectada (em inglês, *Fully Connected*) (Yamashita et al., 2018).

Figura 7 – Exemplo de modelo estrutural de uma Rede Neural Convolucional composta por camadas de Convolução, *Pooling* e pela Camada Totalmente Conectada.

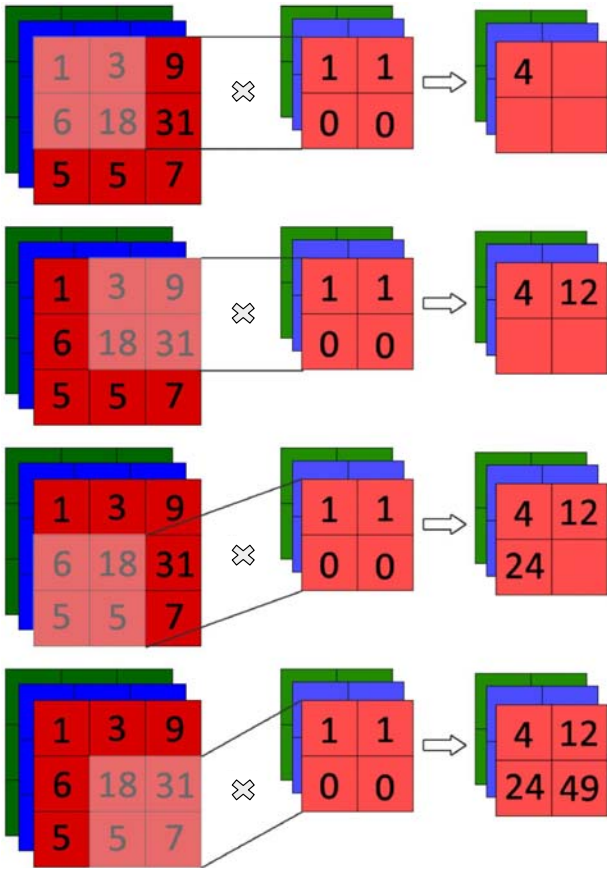


Fonte: Vargas; Paes; Vasconcelos (2016).

A camada de convolução atua sobre imagens representadas como matrizes tridimensionais, considerando suas dimensões espaciais e a quantidade de canais de cor. Cada filtro, estruturado como uma matriz de pesos treináveis, é aplicado localmente a pequenas regiões da imagem, conhecidas como campos receptivos, respeitando a estrutura espacial dos dados (Silva, 2018). A operação de convolução consiste na aplicação desses filtros sobre regiões específicas da imagem de entrada, realizando multiplicações ponto a ponto entre os valores do filtro e os valores correspondentes da região da imagem. O resultado dessas multiplicações é somado, gerando um único valor que compõe o chamado mapa de características (em inglês, *Feature Map*), preservando as informações espaciais essenciais para as etapas subsequentes da rede, conforme ilustrado na Figura 8 (Ingole, 2023). Esse processo permite que a rede extraia padrões relevantes, mantendo as relações entre píxeis vizinhos e aprendendo representações significativas a partir de diferentes partes da imagem.

A camada de *Pooling* desempenha um papel fundamental na redução da dimensionalidade dos dados dentro da rede neural, contribuindo para acelerar o treinamento e aumentar a capacidade do modelo de reconhecer padrões, mesmo quando estes aparecem em posições ligeiramente diferentes. O processo de *Pooling* consiste no agrupamento de informações em regiões específicas da matriz de entrada. Para isso, a matriz é dividida em sub-regiões, chamadas janelas, das quais se extrai um único valor representativo. Existem diferentes estratégias para essa extração, sendo a mais comum a função de máximo (em inglês, *Max pooling*), que seleciona o maior valor dentro de cada vizinhança. A Equação 1 apresenta a formulação matemática da operação de *Max pooling*. Por meio desse processo, a dimensionalidade dos mapas de características é reduzida, resultando na diminuição da altura e largura dos mapas,

Figura 8 – Exemplo de convolução entre uma matriz de entrada de dimensões 3×3×3 e um filtro de tamanho 2×2×3. A operação consiste na multiplicação dos elementos do filtro pelos elementos correspondentes no campo receptivo da matriz de entrada, e a soma dos produtos gerados resulta no valor da saída da convolução.



Fonte: Adaptado de Pacheco (2019).

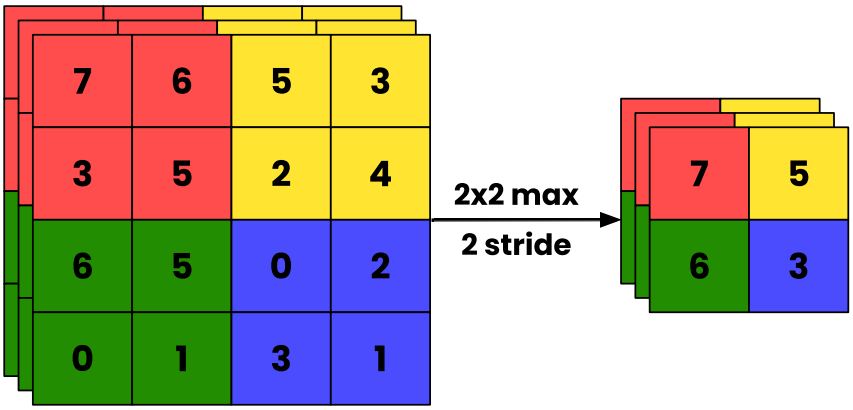
enquanto a profundidade é preservada (Vargas; Paes; Vasconcelos, 2016), conforme ilustrado na Figura 9.

$$f(x) = \max(x, 0) \tag{1}$$

A função de ativação desempenha um papel crucial no processo de treinamento das redes neurais, permitindo a introdução de não linearidades no modelo. Uma das funções de ativação não lineares mais utilizadas é a *Rectified Linear Unit* (ReLU), na qual todos os valores inferiores a zero são transformados em zero, enquanto os valores iguais ou superiores a zero são mantidos. A adoção da ReLU tem impacto significativo na velocidade de treinamento das redes neurais, especialmente quando comparada a outras funções de ativação, como a função logística (Rocha, 2020).

A Camada Totalmente Conectada – também chamada de camada densa – é responsável por integrar as características extraídas pelas camadas anteriores, a fim de gerar a saída final do modelo, que normalmente corresponde a uma classificação

Figura 9 – Exemplo de aplicação de *Max pooling* sobre uma matriz de dimensões 4×4×3. O filtro de *pooling* de tamanho 2×2 é aplicado em cada canal da matriz de entrada, com um *stride* de 2, realizando a redução da dimensionalidade.



Fonte: Adaptado de Pacheco (2019).

ou predição. Nessa etapa, cada neurônio é conectado a todos os neurônios da camada anterior, permitindo ao modelo combinar e relacionar as informações previamente aprendidas. Antes que os dados sejam alimentados na Camada Totalmente Conectada, é necessário convertê-los de uma estrutura multidimensional para um vetor unidimensional – processo conhecido como *flattening* (Massucatto, 2018).

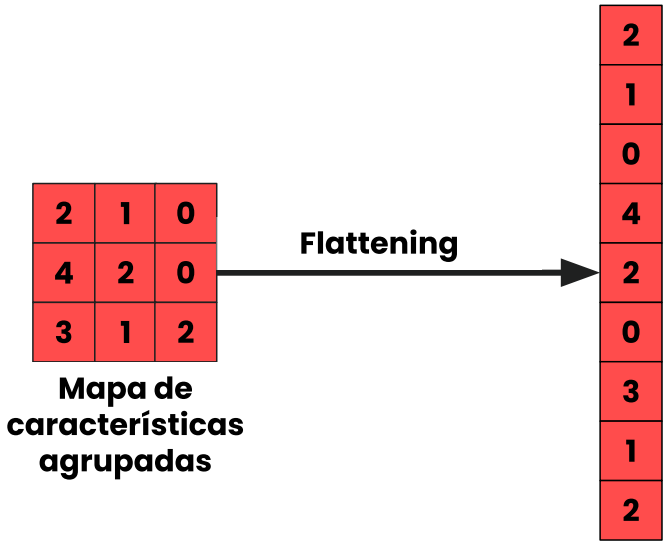
Essa transformação reorganiza os dados em uma estrutura linear, facilitando seu processamento pelas camadas densas da rede neural. O vetor resultante dessa operação – conforme ilustrado na Figura 10 – representa a conversão do mapa de características em um vetor unidimensional. A principal função dessa etapa é possibilitar que a rede combine de forma eficiente todas as características extraídas da imagem, auxiliando na tomada de decisão, como a identificação de objetos ou padrões visuais. Adicionalmente, o processo de aprendizado envolve a atualização contínua dos pesos dessas conexões durante o treinamento, com o objetivo de minimizar o erro da saída do modelo (Steinheiser, 2020).

Embora as CNNs tenham se mostrado eficazes na extração automática de características e no aprendizado de representações espaciais em imagens, sua aplicação em tarefas de segmentação exige arquiteturas mais especializadas. Nesse contexto, destacam-se as arquiteturas do tipo *encoder-decoder*, que possibilitam lidar com entradas e saídas de tamanhos variáveis, facilitando a modelagem de problemas que envolvem mapeamento entre diferentes domínios.

2.2.2 Arquiteturas *encoder-decoder*

As arquiteturas do tipo *encoder-decoder* foram desenvolvidas para resolver problemas em que é necessário transformar uma entrada em uma saída com estrutura igualmente complexa, como ocorre na segmentação semântica de imagens. Diferen-

Figura 10 – Vetor coluna obtido pela operação de *flattening*, a partir de um mapa de características agrupados pela camada de *pooling*.



Fonte: Adaptado de Massucatto (2018).

temente das redes convolucionais tradicionais – que realizam tarefas de classificação a partir de uma representação reduzida da entrada – essas arquiteturas são projetadas para preservar e reconstruir informações espaciais detalhadas (Badrinarayanan; Kendall; Cipolla, 2017).

A estrutura típica de uma arquitetura *encoder-decoder* é composta por duas partes principais: o codificador (*encoder*) e o decodificador (*decoder*). O codificador é responsável por extrair representações compactas e discriminativas dos dados de entrada, reduzindo progressivamente sua dimensão espacial por meio de operações de convolução e *pooling*. O decodificador, por sua vez, atua no sentido inverso, expandindo essas representações para restaurar a resolução espacial original ou gerar uma saída no formato desejado, utilizando operações de *upsampling* ou convoluções transpostas (Minaee et al., 2022).

Um dos principais desafios enfrentados pelas arquiteturas do tipo *encoder-decoder* é a preservação das informações espaciais finas, muitas vezes perdidas durante o processo de codificação. Para mitigar essa limitação, foram propostos mecanismos como as conexões de atalho (em inglês, *skip connections*), que conectam diretamente camadas correspondentes do codificador e do decodificador, permitindo a recuperação de detalhes essenciais (Ronneberger; Fischer; Brox, 2015).

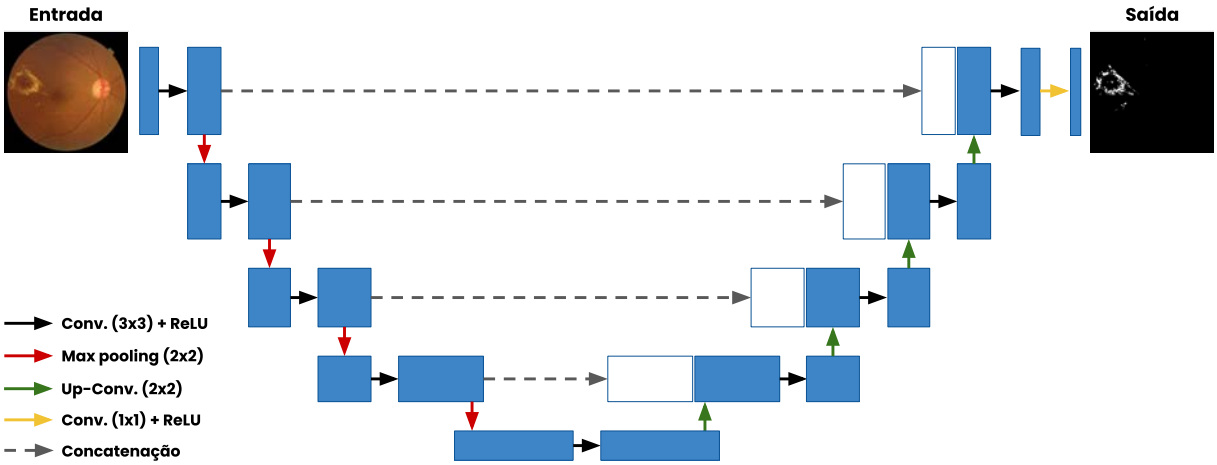
Nos últimos anos, diversas variantes das arquiteturas *encoder-decoder* têm sido desenvolvidas com o objetivo de aprimorar a eficiência e a precisão em tarefas de segmentação. Entre essas variantes, destacam-se modelos como a U-Net, U-Net++, Attention U-Net e R2U-Net, os quais serão descritos a seguir.

2.2.2.1 *U-Net*

A U-Net é uma arquitetura amplamente utilizada em tarefas de segmentação de imagens biomédicas. Projetada originalmente para segmentar estruturas celulares em imagens de microscopia, sua eficácia levou à ampliação de seu uso para diversas aplicações médicas, incluindo a segmentação de lesões retinianas em imagens de fundo de olho (Ronneberger; Fischer; Brox, 2015).

A arquitetura é composta por duas fases principais – codificação e decodificação – organizadas em uma estrutura simétrica em forma de "U", conforme ilustrado na Figura 11. A fase de codificação é responsável por extrair características progressivamente mais abstratas da imagem, por meio de blocos que combinam duas camadas de convolução seguidas de funções de ativação ReLU, intercaladas com operações de *Max pooling* para redução da resolução espacial.

Figura 11 – Diagrama de arquitetura da U-Net.



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

Na fase de decodificação, a U-Net utiliza operações de convolução transposta – também conhecidas como *up-convolution* – para aumentar a resolução das representações e reconstruir o mapa de segmentação. Um dos principais diferenciais da U-Net é o uso de conexões de atalho entre camadas simétricas do codificador e do decodificador, que realizam a concatenação direta das características de baixo nível com aquelas de nível mais alto. Essa estratégia contribui para a preservação de detalhes espaciais importantes, aprimorando a definição dos contornos nas máscaras segmentadas.

A etapa final aplica uma convolução 1×1 para reduzir a dimensionalidade do mapa de características ao número de classes da segmentação. De forma geral, a U-Net equilibra profundidade e resolução espacial, tornando-se uma base sólida para o desenvolvimento de diversas variantes, como as arquiteturas U-Net++, Attention U-Net e R2U-Net.

2.2.2.2 U-Net++

A U-Net++ foi desenvolvida como uma evolução da arquitetura U-Net, com o objetivo de reduzir a diferença semântica entre os mapas de características gerados na fase de codificação – que contêm informações mais abstratas – e aqueles processados na fase de decodificação, que exigem maior riqueza de detalhes para a reconstrução precisa da imagem. A principal inovação introduzida pela U-Net++ é o uso de conexões densas e hierárquicas entre os níveis intermediários da rede, permitindo uma transição mais gradual e uma melhor preservação das informações semânticas ao longo dos diferentes estágios da arquitetura.

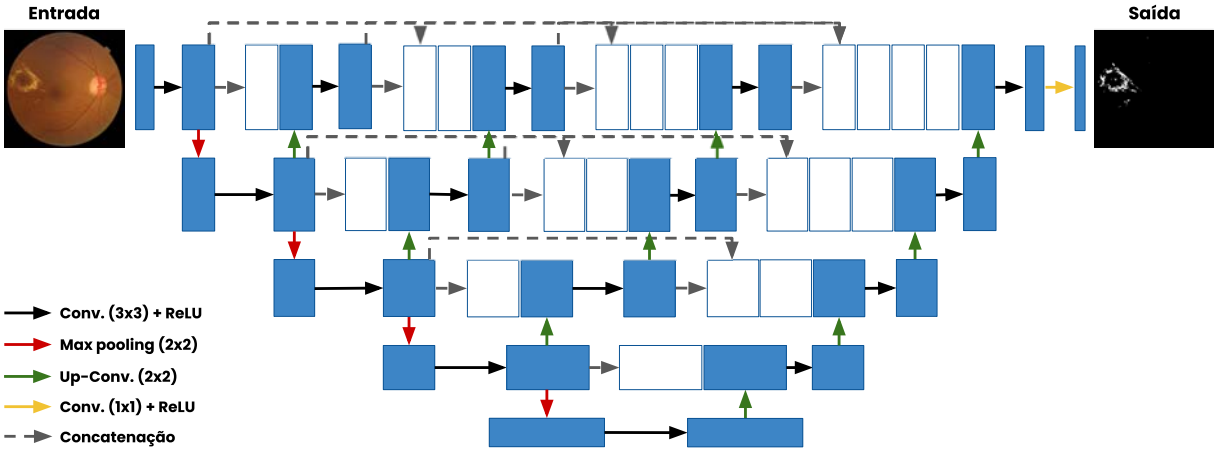
Na prática, a U-Net++ reorganiza as conexões de atalho tradicionais da U-Net, substituindo os enlaces diretos por caminhos intermediários compostos por múltiplas camadas convolucionais adicionais. Essa abordagem promove uma fusão mais eficaz entre características extraídas em diferentes profundidades, fornecendo ao decodificador informações mais consistentes e detalhadas. Como resultado, a U-Net++ apresenta maior precisão em tarefas de segmentação, especialmente em contextos que envolvem estruturas morfológicas complexas (Zhou et al., 2018).

Uma característica adicional relevante da U-Net++ é a possibilidade de aplicar supervisão profunda (em inglês, *deep supervision*) – técnica que consiste em supervisionar não apenas a saída final da rede, mas também as saídas intermediárias geradas ao longo dos diferentes caminhos de decodificação. Essa abordagem oferece dois benefícios principais: primeiramente, incentiva os blocos intermediários a capturar representações discriminativas mais eficazes, ou seja, padrões que permitem distinguir melhor as características relevantes para a tarefa; em segundo lugar, contribui para um fluxo de gradientes mais estável, acelerando a convergência durante o treinamento. A supervisão profunda pode ser ativada ou desativada conforme a necessidade, possibilitando um equilíbrio entre desempenho e custo computacional.

A Figura 12 ilustra a organização da U-Net++. De modo geral, a arquitetura preserva a estrutura fundamental da U-Net – com codificação baseada em convoluções e *Max pooling*, e decodificação por meio de convoluções transpostas –, mas adiciona caminhos de conexão compostos por múltiplas operações convolucionais (representadas pelos blocos adicionais entre os caminhos horizontais da rede). Esses caminhos favorecem o reuso de características e ampliam a capacidade de generalização da rede, ao mesmo tempo em que facilitam o fluxo de gradientes durante o processo de treinamento.

Assim como outras variações da U-Net, a U-Net++ busca aprimorar o desempenho da segmentação sem causar aumentos excessivos no número de parâmetros do modelo. Sua estrutura densa e hierárquica contribui de forma significativa para a melhoria da precisão em tarefas de segmentação médica (Zhou et al., 2018).

Figura 12 – Diagrama de arquitetura da U-Net++.

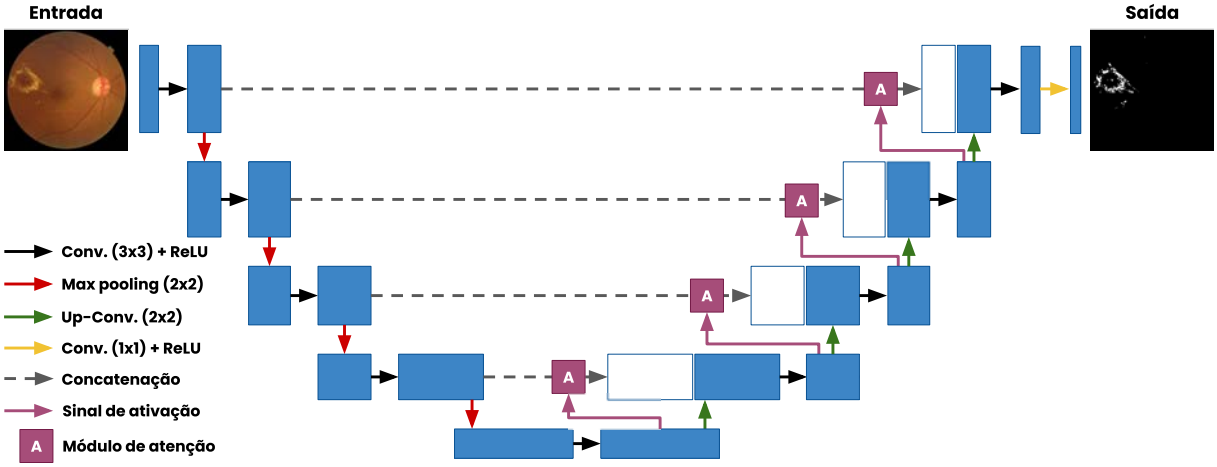


Fonte: Adaptado de Huo et al. (2024).

2.2.2.3 Attention U-Net

A Attention U-Net é uma variação da arquitetura U-Net que incorpora mecanismos de atenção com o objetivo de aprimorar a qualidade da segmentação. Conforme ilustrado na Figura 13, sua estrutura é composta por fases de codificação e decodificação, assim como na U-Net original. A principal inovação introduzida por essa arquitetura consiste na inserção de módulos de atenção ao longo das conexões de atalho entre o codificador e o decodificador.

Figura 13 – Diagrama de arquitetura da Attention U-Net.



Fonte: Adaptado de Oktay et al. (2018).

Esses módulos atuam como filtros adaptativos, permitindo que a rede foque nas regiões mais relevantes da imagem. Para isso, utilizam informações provenientes de ambas as fases – codificação e decodificação – para calcular mapas de atenção que indicam o grau de importância de cada região da imagem. Dessa forma, regiões ir-

relevantes ou com baixa relevância para a tarefa de segmentação são atenuadas, enquanto áreas de maior interesse são realçadas. Esse mecanismo é especialmente vantajoso em imagens médicas, onde as estruturas-alvo frequentemente são pequenas, sutis ou pouco contrastadas (Oktay et al., 2018).

O sinal de ativação gerado em cada etapa do decodificador é utilizado como guia para o módulo de atenção correspondente. Esse sinal atua como um mecanismo de controle, ajustando os mapas de características provenientes do codificador por meio do módulo de atenção antes da concatenação. Esse processo favorece uma reconstrução mais precisa das regiões de interesse, pois permite ao modelo concentrar-se dinamicamente nas áreas mais informativas da imagem durante a decodificação. Tais modificações contribuem para o aprimoramento da capacidade discriminativa do modelo, sem implicar aumentos substanciais no número de parâmetros.

2.2.2.4 R2U-Net

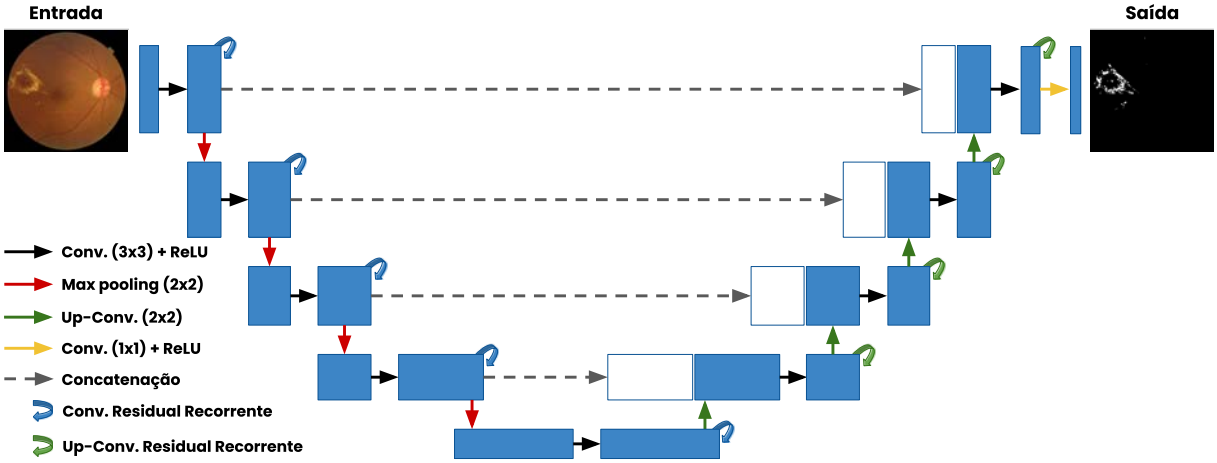
A R2U-Net foi proposta a partir da integração dos conceitos das arquiteturas *Deep Residual Network* (ResNet) (He et al., 2016), *Recurrent Convolutional Neural Network* (RCNN) (Liang; Hu, 2015) e U-Net (Ronneberger; Fischer; Brox, 2015), com o objetivo de combinar os pontos fortes de cada uma dessas abordagens em aprendizado profundo. Entre as principais inovações que a diferenciam da U-Net original, destacam-se: (1) a incorporação de Camadas Convolucionais Recorrentes (em inglês, *Recurrent Convolutional Layers* – RCL) e RCL com conexões residuais; (2) o uso de um mecanismo eficiente de agregação de características nas unidades RCL; e (3) a substituição da tradicional operação de corte e cópia da U-Net por operações de concatenação.

Mais especificamente, a substituição das camadas convolucionais convencionais por RCLs com conexões residuais proporciona à arquitetura uma estrutura mais profunda e eficiente. O novo método de acumulação de características implementado nas unidades RCL aprimora a capacidade de convergência da rede durante o treinamento e a validação, favorecendo a extração de representações mais discriminativas. Além disso, a remoção da operação de corte e cópia – em favor do uso exclusivo de concatenações – simplifica o fluxo de dados dentro da arquitetura. Essas alterações visam potencializar o desempenho da R2U-Net, explorando avanços contemporâneos para maximizar sua eficácia, especialmente em aplicações de segmentação de imagens médicas (Alom et al., 2018).

A estrutura arquitetural da R2U-Net é ilustrada na Figura 14. A arquitetura é organizada em duas seções principais: a fase de codificação, representada pelos blocos à esquerda, e a fase de decodificação, composta pelos blocos à direita. Adicionalmente, incorpora operações recorrentes e conexões residuais. As operações recorrentes permitem a reutilização de informações ao longo de múltiplos estágios da rede, favorecendo a captura de contextos mais amplos. Já as conexões residuais intro-

duzem atalhos entre camadas, mitigando problemas de degradação de desempenho típicos em redes profundas. Resultados experimentais reportados na literatura indicam que essas modificações beneficiam tanto o processo de treinamento quanto o desempenho durante a inferência, sem implicar aumentos significativos no número de parâmetros do modelo (Pereira et al., 2024).

Figura 14 – Diagrama de arquitetura da R2U-Net.



Fonte: Adaptado de Pereira et al. (2024).

2.3 Trabalhos Relacionados

Esta seção tem como objetivo revisar as principais abordagens disponíveis na literatura voltadas à segmentação de lesões no fundo de olho associadas à RD.

Em Li et al. (2019), foi proposto um novo conjunto de dados para RD, denominado *Dataset for Diabetic Retinopathy* (DDR). Além disso, os autores avaliaram modelos de aprendizado profundo – HED (Xie; Tu, 2015) e DeepLab-v3+ (Chen et al., 2018) – aplicados à segmentação de lesões retinianas. Os resultados foram apresentados por meio das métricas *Average Precision* (AP), *mean Average Precision* (mAP), *Intersection over Union* (IoU) e *mean Intersection over Union* ($mIoU$). Foram obtidos valores de mAP de 15,87% com o modelo HED (Xie; Tu, 2015) e 30,10% com o DeepLab-v3+ (Chen et al., 2018), enquanto o $mIoU$ foi de 8,83% e 17,91%, respectivamente. Esses resultados demonstram que os modelos apresentaram dificuldades na segmentação de lesões, especialmente os MA, para os quais foram observados valores de IoU de apenas 1,10% com o HED e 3,25% com o DeepLab-v3+, no conjunto de teste.

Em Raudonis et al. (2023), foi proposta uma nova abordagem para a segmentação de MA em imagens de fundo de olho. O método consiste em uma etapa inicial de pré-processamento, na qual as imagens são divididas em múltiplas janelas sobrepostas,

com o objetivo de aumentar a visibilidade das lesões. A segmentação foi realizada por meio de um *ensemble* de três redes neurais profundas: U-Net (Ronneberger; Fischer; Brox, 2015), ResNet34-UNet (Abedalla et al., 2020) e U-Net++ (Zhou et al., 2018). Com essa abordagem foram obtidos valores de *Dice Coefficient* (DC) de 95% e IoU de 91%. Embora promissora, a abordagem apresenta duas limitações importantes: (1) segmenta apenas um tipo de lesão associada à RD; e (2) não utiliza conjuntos de dados públicos nos experimentos, o que dificulta a reprodutibilidade e a comparação com outros estudos.

Em Nair et al. (2023), foi apresentada a MesU-Net, uma nova arquitetura de rede neural que combina elementos da U-Net (Ronneberger; Fischer; Brox, 2015) e da MesNet (Guo et al., 2021) para segmentação de imagens retinianas. O estudo concentrou-se na segmentação de EX, HE, MA e vasos sanguíneos. Os modelos foram treinados utilizando conjuntos de dados públicos, sendo o *Indian Diabetic Retinopathy Image Dataset* (IDRiD) empregado para lesões de fundo de olho e o *Digital Retinal Images for Vessel Extraction* (DRIVE) para segmentação de vasos. Os resultados foram apresentados principalmente por meio da métrica de acurácia (*Accuracy* – Acc). Para as lesões retinianas, os valores de Acc alcançados foram de 97,60% para EX, 98,10% para HE e 99,20% para MA, com média de 98,30%. Já para os vasos sanguíneos, a Acc obtida foi de 83,70%.

Por sua vez, Anand et al. (2023) propuseram modificações na arquitetura básica da U-Net com o objetivo de aumentar a precisão na segmentação de vasos sanguíneos, EX e MA. O estudo utilizou diversos conjuntos de dados públicos, incluindo IDRiD, *Diabetic Retinopathy Database 1* (DIARETDB1), *Structured Analysis of the Retina* (STARE), *Child Heart and Health Study in England Database 1* (CHASEDB1), DRIVE e *High-Resolution Fundus* (HRF). Os resultados também foram reportados por meio da métrica Acc, com destaque para os valores de 99,86% e 99,9% obtidos no *dataset* IDRiD na segmentação de EX e MA, respectivamente. Embora o trabalho não tenha abordado a segmentação de HE e SE, os resultados ressaltam o potencial significativo da arquitetura U-Net na segmentação de lesões retinianas.

Em Pereira et al. (2024), foi proposta uma abordagem para a segmentação de EX, SE, HE e MA, utilizando a arquitetura R2U-Net (Alom et al., 2018), combinada com técnicas de aumento de dados. Os resultados obtidos indicam o potencial dessas técnicas para aprimorar a precisão do modelo, atingindo 21,60% de *mean Dice Coefficient* (mDC) e 57,91% de $mIoU$ na etapa de teste do *dataset* DDR. No entanto, devido ao tamanho reduzido de algumas lesões, a abordagem ainda enfrenta desafios, especialmente na segmentação de MA, para os quais foram observados valores de DC de 10,99% e IoU de 53,36%.

Já Xu et al. (2024) propuseram a HACDR-Net, uma rede convolucional voltada à segmentação de múltiplas lesões (EX, SE, HE e MA). A arquitetura busca lidar com

a variabilidade em localização, forma e tamanho das lesões, além do desafio representado pela grande proporção de fundo nas imagens. A HACDR-Net incorpora convoluções heterogêneas e um módulo de agregação adaptativo para capturar regiões lesionadas com maior precisão. Para mitigar o desequilíbrio entre fundo e lesões, os autores introduziram a função de perda *Noise Adjusted Loss* (NALoss). Os experimentos demonstraram que a HACDR-Net superou outras abordagens, atingindo um mDC de 49,30% e um $mIoU$ de 33,70%.

Em Jiang; Zhao (2024), foi proposta uma estrutura para segmentação de EX, HE, SE e MA, baseada na arquitetura U-Net++, combinada com uma ResNet50 (He et al., 2016) como *backbone* para extração de características. Os autores integraram um módulo de atenção híbrido ao bloco residual, com o objetivo de aprimorar a capacidade de extração de representações discriminativas. Além disso, introduziram e adaptaram o módulo *Across Feature Map Attention* (AFMA), visando melhorar a precisão na segmentação de lesões de pequenas dimensões.

Os experimentos foram realizados nos conjuntos de dados IDRiD e DDR, alcançando mDC de 63,57% e 43,90%, respectivamente. Apesar dos resultados gerais promissores, os valores de DC obtidos para a segmentação de MA – 44,20% no IDRiD e 28,37% no DDR – evidenciam dificuldades persistentes na segmentação de lesões de menor tamanho.

Já em Zhang; Xu; Zhai (2025), foi proposta uma abordagem híbrida para segmentação de lesões retinianas, integrando as arquiteturas *Densely Connected Convolutional Network* (DenseNet) (Huang et al., 2017) e U-Net (Ronneberger; Fischer; Brox, 2015). A DenseNet substitui o codificador da U-Net, aprimorando a propagação de informações e a extração de características. A principal inovação da proposta está na utilização da *Omni-Dimensional Dynamic Convolution* (ODConv), que substitui as convoluções tradicionais, aumentando a adaptabilidade da rede e a precisão da segmentação.

Além disso, os autores incorporaram o *Convolutional Block Attention Module* (CBAM) na fase de decodificação, o que aprimora a captura de regiões relevantes e a definição dos contornos das lesões. Os experimentos realizados nos conjuntos de dados DDR e IDRiD apresentaram melhorias nas métricas mDC e $mIoU$. No *dataset* DDR, os valores alcançados foram de 46,03% para mDC e 30,41% para $mIoU$, representando aumentos de 0,46% e 0,14%, respectivamente. Já no *dataset* IDRiD, os resultados foram de 65,68% para mDC e 50,48% para $mIoU$, correspondendo a incrementos de 0,53% e 1,12% em relação às abordagens comparadas.

A Tabela 1 apresenta uma síntese comparativa dos trabalhos relacionados, considerando critérios como os conjuntos de dados utilizados, os modelos empregados, os principais resultados obtidos e as limitações identificadas em cada abordagem. Embora diferentes estudos tenham explorado redes neurais profundas para a seg-

mentação de lesões retinianas em imagens de fundo de olho, os resultados ainda apresentam limitações significativas. Essas dificuldades decorrem, em grande parte, da escassez de anotações de lesões e da qualidade heterogênea das imagens disponíveis nos conjuntos de dados públicos de RD. Além disso, a extração de características em regiões muito pequenas da imagem – particularmente no caso de lesões do tipo MA – representa um desafio adicional durante o processo de treinamento dos modelos.

Tabela 1 – Análise comparativa dos trabalhos relacionados na segmentação das lesões retinianas.

Artigo	*Dataset*	Modelos	Resultados	Limitações
Li et al. (2019)	DDR	HED DeepLab-v3+	$mIoU$ = 0,0883 $mIoU$ = 0,1791	Baixa precisão na segmentação de MA.
Raudonis et al. (2023)	PerDiRe Project	U-Net ResNet34-UNet U-Net++	DC = 0,9500	Conjunto de dados privado. Segmentação apenas de MA.
Nair et al. (2023)	IDRiD	MesU-Net	Acc = 0,9830	Avaliação apenas com Acc. SE não segmentados.
Anand et al. (2023)	IDRiD DIARETDB1	U-Net	Acc = 0,9992	HE e SE não segmentados.
Pereira et al. (2024)	DDR IDRiD	R2U-Net	DC = 0,2160	Baixa precisão na segmentação de MA. Ausência etapa de pré-processamento.
Xu et al. (2024)	DDR IDRiD	HACDR-Net	DC = 0,4930	Baixa precisão na segmentação de MA.
Jiang; Zhao (2024)	DDR IDRiD	U-Net++	DC = 0,4390	Ausência etapa de aumento de dados. Baixa precisão na segmentação MA.
Zhang; Xu; Zhai (2025)	DDR IDRiD	DenseCUNet	DC = 0,4603	Ausência etapa de aumento de dados. Etapa de pré-processamento não descrita. Baixa precisão na segmentação de SE e MA.

Fonte: Elaborada pelo autor.

3 ABORDAGEM PROPOSTA

Este capítulo tem como objetivo apresentar a abordagem proposta para a segmentação semântica de lesões de fundo associadas à RD. São descritos os materiais, técnicas e métodos empregados na construção dos modelos, bem como as estratégias adotadas para o pré-processamento e aumento de dados. Além disso, discute-se a condução dos experimentos e os resultados obtidos com a aplicação da abordagem proposta.

3.1 Metodologia

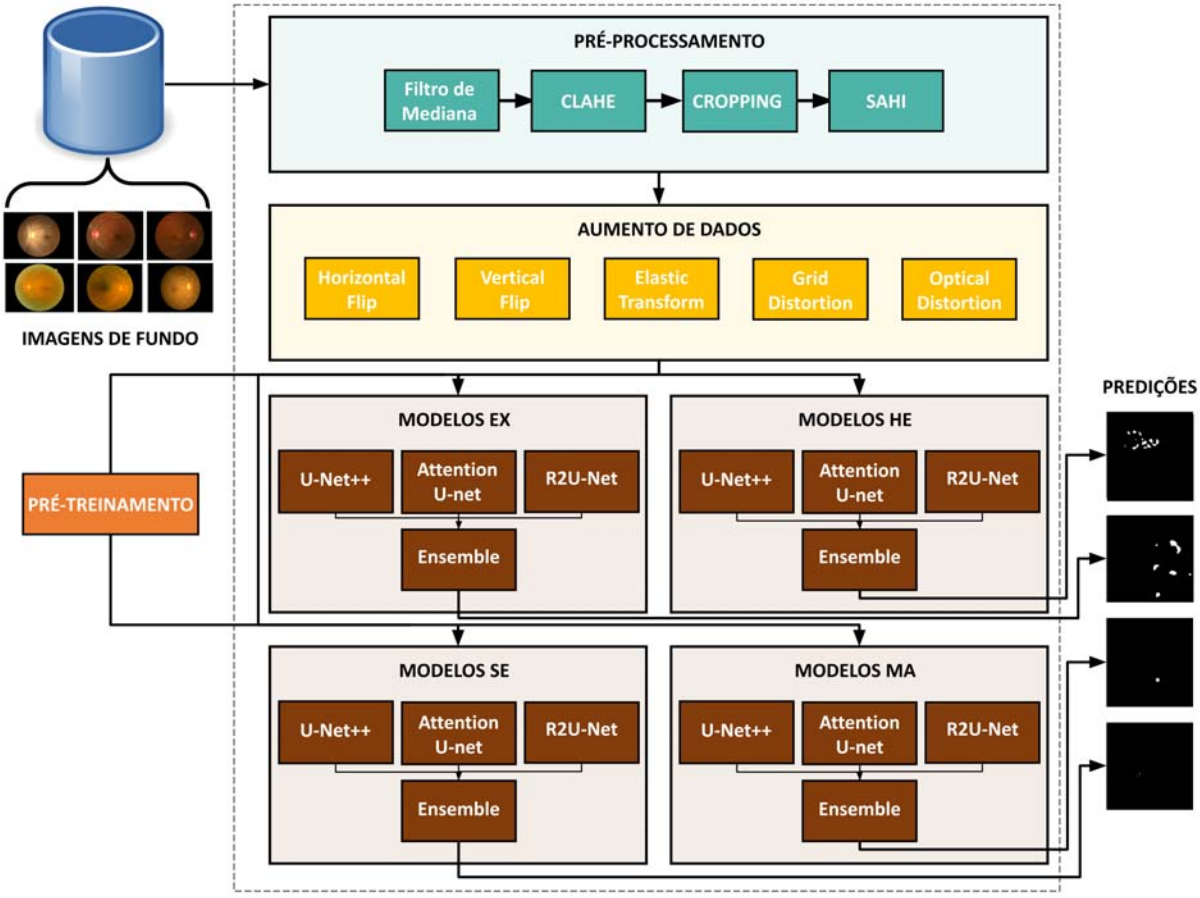
Os experimentos foram conduzidos em um ambiente de *hardware* composto por um processador AMD Ryzen 9 5900X (12 núcleos, 24 threads, @ 3,7GHz), 128GB de memória *Random Access Memory* (RAM) e uma GPU NVIDIA RTX 4090 com 24 GB de *Video Random Access Memory* (VRAM). A abordagem proposta é composta pelas arquiteturas U-Net++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018) e R2U-Net (Alom et al., 2018), integradas na forma de *ensembles* para realizar a segmentação das lesões de fundo de olho. Adicionalmente, foram aplicadas etapas de pré-processamento, aumento de dados e pré-treinamento, com o objetivo de aprimorar a precisão da segmentação.

Conforme ilustrado no diagrama de blocos da Figura 15, o processo inicia-se com o bloco de pré-processamento, no qual é aplicado um filtro de mediana para remoção de ruídos, seguido da técnica *Contrast-Limited Adaptive Histogram Equalization* (CLAHE) para aprimoramento do contraste. Na sequência, realiza-se a remoção do fundo preto das imagens. Em seguida, aplica-se o *framework Slicing Aided Hyper Inference* (SAHI) (Akyon; Altinuc; Temizel, 2022), por meio do método COCO *slice*, com o intuito de dividir as imagens em sub-regiões menores, facilitando a detecção de lesões de pequeno porte.

Após o pré-processamento, as imagens são encaminhadas ao bloco de aumento de dados, com o objetivo de gerar novas amostras artificiais que ampliem a diversidade do conjunto de treinamento e contribuam para a generalização dos modelos.

Em seguida, os modelos da abordagem proposta são treinados utilizando as imagens já pré-processadas e aumentadas. Adicionalmente, é realizada uma etapa de pré-treinamento, na qual os pesos iniciais dos modelos são ajustados com base no conjunto de dados ImageNet (Russakovsky et al., 2015), com o intuito de acelerar a convergência e melhorar a extração de características. As seções a seguir descrevem em detalhe as etapas que compõem a metodologia adotada.

Figura 15 – Diagrama de blocos da abordagem proposta para segmentação de lesões de fundo. Inicialmente, as imagens passam pelo bloco de pré-processamento. Em seguida, são transferidas para o bloco de aumento de dados. Na etapa seguinte, os modelos são inicializados com pesos pré-treinados no conjunto de dados ImageNet. Por fim, os modelos treinados são integrados por meio de *ensembles* para realizar as predições das lesões.



Fonte: Elaborada pelo autor.

3.1.1 Conjunto de Dados

O conjunto de dados DDR foi utilizado para o treinamento, ajuste e avaliação dos modelos que compõem a abordagem proposta. O *dataset* DDR está disponível pu-

blicamente no Baidu Drive[1] e no Google Drive[2], sendo composto por 13.673 imagens coloridas da retina, no formato *Joint Photographic Experts Group* (JPG).

As imagens foram capturadas entre 2016 e 2018 em 147 hospitais na China (Li et al., 2019), abrangendo 9.598 pacientes com idades entre 1 e 100 anos e média de 54,13 anos. Do total de pacientes, 48,23% eram do sexo masculino e 51,77%, do sexo feminino. As imagens foram adquiridas com o uso de 42 modelos distintos de câmeras, todas com campo de visão de 45°.

Para os experimentos, as imagens foram divididas de forma aleatória em três subconjuntos: 50% para treinamento, 20% para validação e 30% para teste. O conjunto DDR apresenta três tipos de anotações: (1) classificação da RD no nível da imagem; (2) anotações no nível de píxel; e (3) anotações em formato de caixas delimitadoras (*bounding boxes*) para as lesões retinianas associadas à RD. As imagens foram classificadas por oftalmologistas de acordo com a Classificação Internacional da Retinopatia Diabética, sendo divididas em seis categorias: sem RD, RDNP leve, RDNP moderada, RDNP grave, RDP e não classificável. A Tabela 2 apresenta a distribuição das amostras por classe nos conjuntos de treinamento, validação e teste.

Tabela 2 – Número de amostras de cada classe nos conjuntos de Treino, Validação e Teste.

	Treino	**Validação**	**Teste**	**Total**
Sem RD	3133	1253	1880	6266
RDNP leve	315	126	189	630
RDNP moderada	2238	895	1344	4477
RDNP grave	118	47	71	236
RDP	456	182	275	913
Não classificável	575	230	346	1151
Total	6835	2733	4105	13673

Fonte: Elaborada pelo autor.

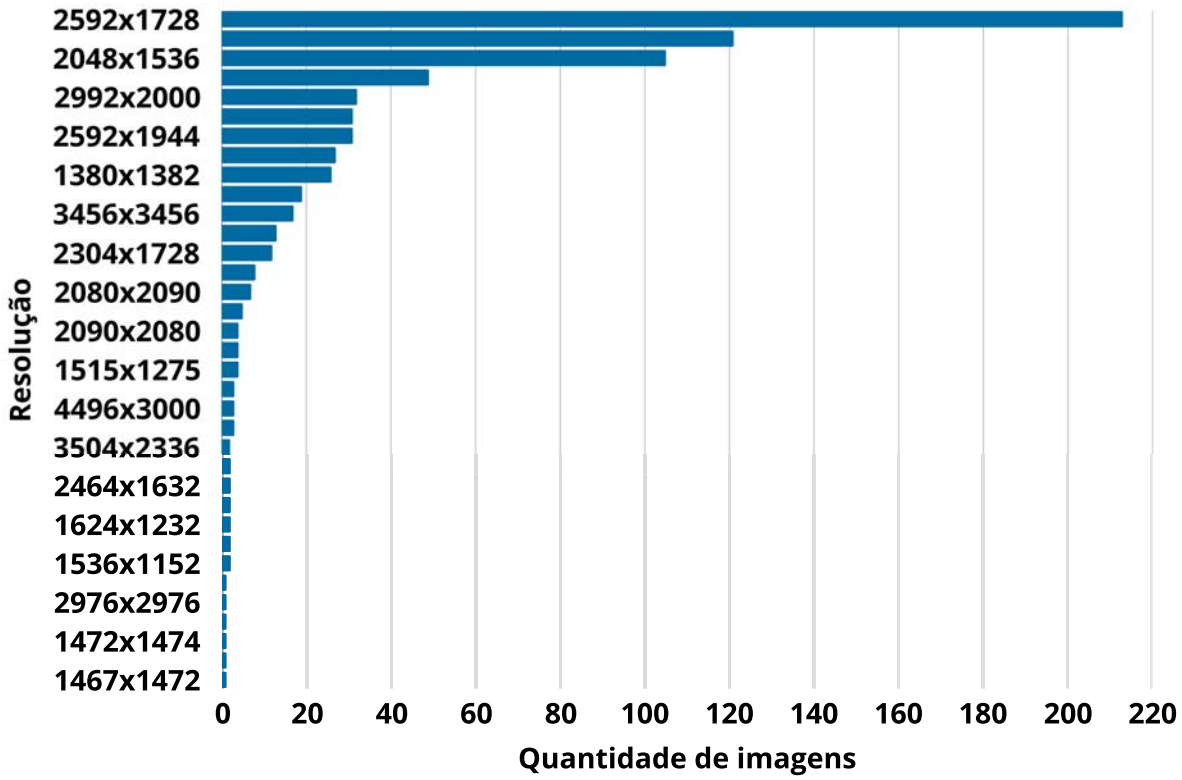
Além disso, um total de 757 imagens foi selecionado para receber anotações no nível de píxel e caixas delimitadoras das lesões. Essas imagens apresentam resoluções variadas. Conforme ilustrado na Figura 16, observa-se que a maioria está concentrada nas resoluções 2.592×1.728, 2.736×1.824 e 2.048×1.536. A Figura 17 apresenta um exemplo de imagem com anotações de lesões no nível de píxel, enquanto a Figura 18 mostra uma imagem de fundo do olho com anotações em caixas delimitadoras.

As anotações das imagens foram realizadas por seis anotadores, utilizando uma ferramenta com capacidade de identificação automática de lesões retinianas, incluindo Exsudatos Duros, Exsudatos Algodonosos, Hemorragias e Microaneurismas. Apesar do suporte automatizado, os anotadores foram responsáveis por verificar a precisão

[1] `https://pan.baidu.com/s/1560JK2pzxTN9Ny1TcmNasQ`

[2] `https://drive.google.com/drive/folders/1z6tSFmxW_aNayUqVxx6h6bY4kwGzUTEC`

Figura 16 – Gráfico distribuição das resolução das imagens com anotações a nível de píxel do conjunto DDR.



Fonte: Elaborada pelo autor.

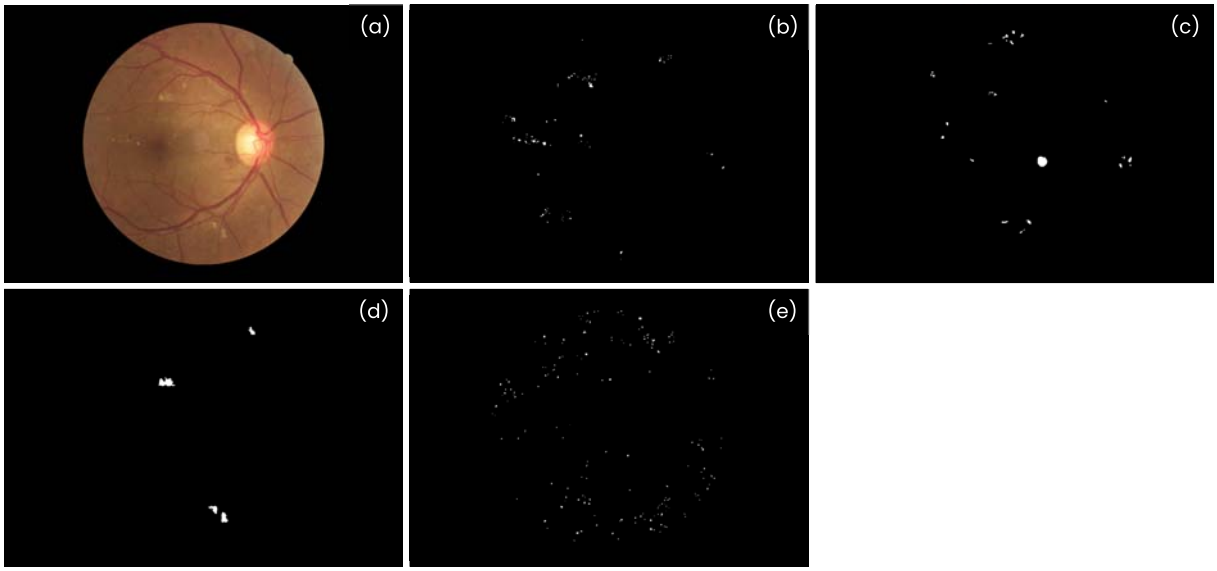
das detecções, realizando a correção de anotações incorretas e adicionando manualmente as lesões não identificadas pelo sistema.

Como forma de avaliar a qualidade do conjunto, os autores realizaram o treinamento e a validação de diversos modelos baseados em aprendizado profundo. Dentre os modelos utilizados nos experimentos, cinco foram treinados para classificação do grau de RD, três para detecção de lesões e dois para segmentação semântica das estruturas retinianas. Os resultados obtidos indicaram baixa precisão, especialmente na segmentação de lesões pequenas, como os Microaneurismas.

O conjunto de dados IDRiD foi utilizado para avaliar a capacidade de generalização da abordagem proposta em diferentes conjuntos de dados de Retinopatia Diabética. O *dataset* IDRiD está disponível publicamente no *IEEE Dataport Repository*[3] e é composto por 516 imagens coloridas do fundo do olho, com resolução de 4.288×2.848 píxeis, armazenadas no formato JPG. As imagens foram subdivididas pelos autores em dois subconjuntos: 80% destinadas ao treinamento e os 20% restantes reservadas para teste.

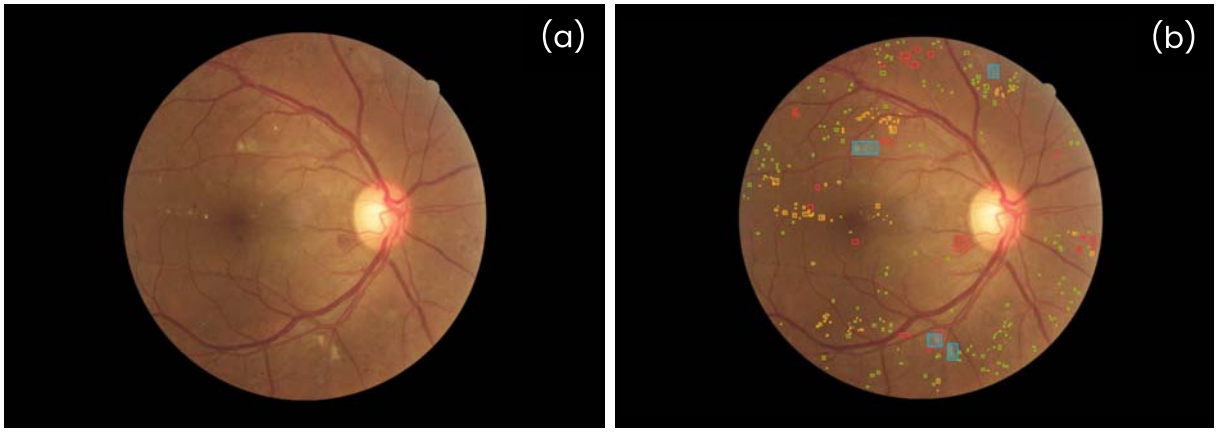
[3] `http://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid`

Figura 17 – Exemplo de imagem de fundo do conjunto de dados DDR e diferentes anotações em nível de píxel. Em (a), a imagem de fundo original; (b) anotações a nível de píxel dos Exsudatos Duros; (c) anotações a nível de píxel das Hemorragias; (d) anotações a nível de píxel dos Exsudatos Algodonosos; e, (e) anotações a nível de píxel dos Microaneurismas.



Fonte: Elaborada pelo autor.

Figura 18 – Exemplo de imagem de fundo do conjunto de dados DDR com as anotações de caixa delimitadora das lesões. Em (a), a imagem de fundo original; e, (b) a imagem com suas anotações de caixa delimitadora. As anotações em amarelo representam Exsudatos Duros; as anotações em vermelho representam Hemorragias; as anotações em azul representam Exsudatos Algodonosos; e, as anotações em verde representam Microaneurismas.



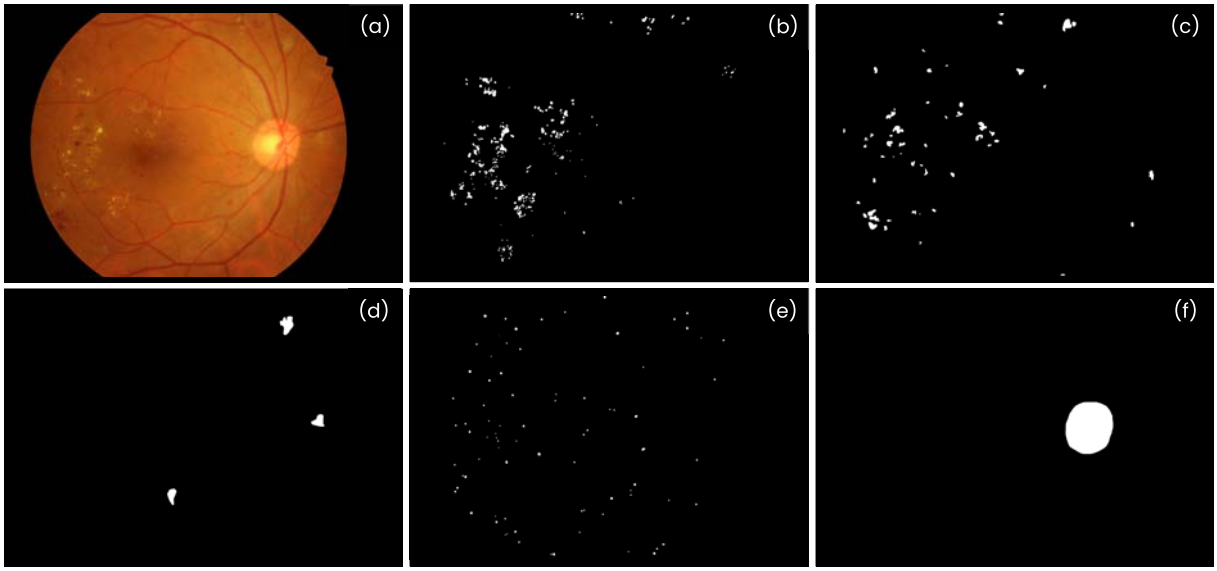
Fonte: Elaborada pelo autor.

As imagens foram obtidas a partir de exames realizados em uma clínica oftalmológica localizada em Nanded, Índia. A aquisição foi realizada com uma câmera Kowa VX-10, equipada com campo de visão de 50°. Além disso, antes da captura, os pacientes tiveram suas pupilas dilatadas com Tropicamida, um agente midriático que induz a dilatação da pupila e a paralisia da acomodação visual (Porwal et al., 2020).

Esse conjunto de dados disponibiliza três tipos de anotações: (1) anotações no nível de píxel, (2) classificações de RD no nível da imagem e (3) coordenadas do Disco Óptico (DO) e do centro da fóvea. Todas as 516 imagens foram classificadas quanto ao grau de RD, conforme a Escala Internacional de Retinopatia Diabética Clínica, sendo distribuídas em cinco categorias: sem RD, RDNP leve, RDNP moderada, RDNP grave e RDP. Adicionalmente, o risco de Edema Macular Diabético (EMD) foi avaliado em uma escala de 0 (sem EMD) a 2 (EMD grave).

Para as anotações no nível de píxel, foram selecionadas 81 imagens com sinais de RD. Nessas imagens, foram realizadas anotações das lesões EX, HE, SE e MA, bem como do DO. As máscaras binárias correspondentes às anotações foram fornecidas no formato de arquivo *Tagged Image File Format* (TIFF). A Figura 19 apresenta um exemplo de imagem de fundo do conjunto IDRiD, juntamente com as respectivas máscaras binárias das lesões e do DO. Já a Figura 20 exibe um exemplo de imagem com as anotações em caixas delimitadoras das lesões de fundo.

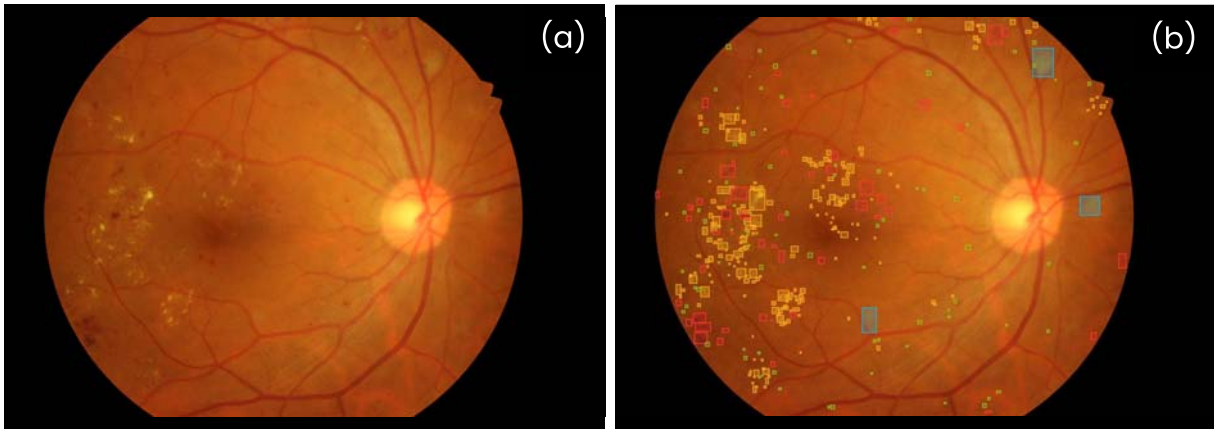
Figura 19 – Exemplo de imagem de fundo do conjunto de dados IDRiD e diferentes anotações em nível de píxel. Em (a) a imagem de fundo original; (b) anotações a nível de píxel dos Exsudatos Duros; (c) anotações a nível de píxel das Hemorragias; (d) anotações a nível de píxel dos Exsudatos Algodonosos; (e) anotações a nível de píxel dos Microaneurismas; e, (f) a anotação do Disco Óptico.



Fonte: Elaborada pelo autor.

Apesar das imagens do conjunto IDRiD apresentarem alta qualidade, sua principal limitação está na quantidade reduzida de amostras disponibilizadas. Essa restrição torna-se ainda mais evidente quando comparada ao conjunto DDR, que contém 13.673 imagens com anotações do grau de RD, enquanto o IDRiD disponibiliza apenas 516. Além disso, o DDR conta com 757 imagens anotadas em nível de píxel, ao passo que o IDRiD possui apenas 81 imagens com esse tipo de anotação.

Figura 20 – Exemplo de imagem de fundo do conjunto de dados IDRiD com as anotações de caixa delimitadora das lesões. Em (a) a imagem de fundo original; e, (b) a imagem com suas anotações de caixa delimitadora. As anotações em amarelo representam Exsudatos Duros; as anotações em vermelho representam Hemorragias; as anotações em azul representam Exsudatos Algodonosos; e, as anotações em verde representam Microaneurismas.



Fonte: Elaborada pelo autor.

Para a realização dos experimentos, aplicou-se o método de *Hold-out validation* (Lee; Liong; Jemain, 2018) em ambos os conjuntos, utilizando apenas as imagens com anotações em nível de píxel das lesões. Com o objetivo de padronizar a divisão entre os subconjuntos, adotou-se a proporção de 50% para treinamento, 20% para validação e 30% para teste, conforme realizado em Li et al. (2019); Santos (2023); Pereira et al. (2024). A Tabela 3 apresenta a quantidade de imagens em cada subconjunto após a divisão.

Tabela 3 – Número de imagens contidas em cada conjunto de dados após aplicação do método *Hold-out validation*.

Dataset	**Treinamento**	**Validação**	**Teste**	**Total**
DDR	383	149	225	757
IDRiD	42	17	22	81

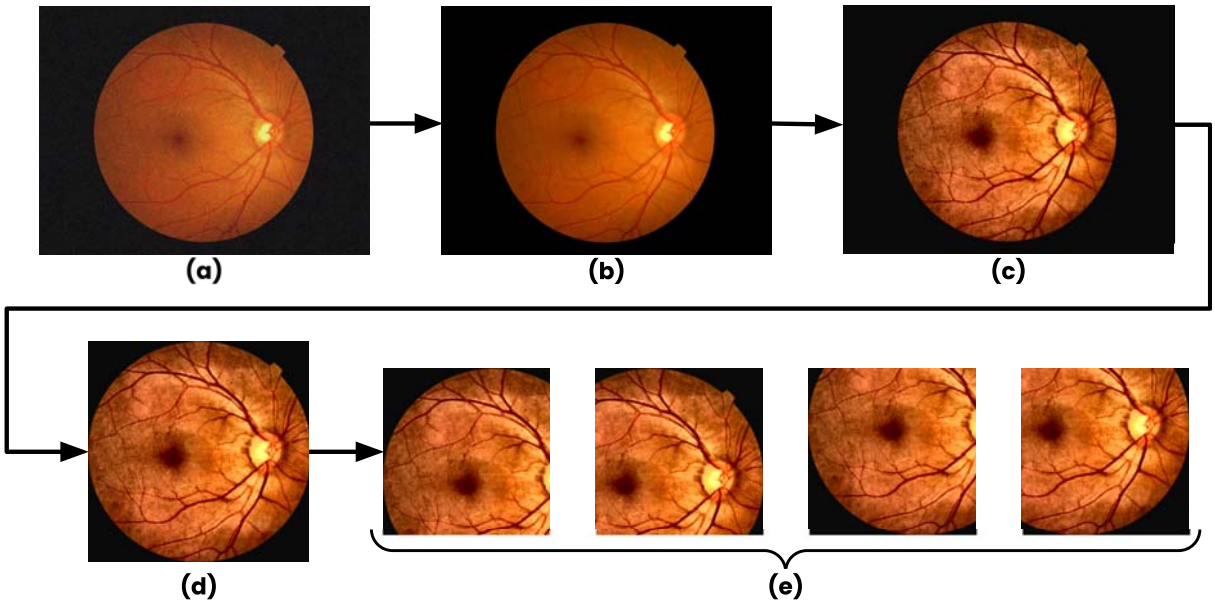
Fonte: Elaborada pelo autor.

3.1.2 Pré-processamento

O bloco de pré-processamento da abordagem proposta é composto por quatro etapas e tem como objetivo aprimorar a qualidade das imagens de fundo, além de minimizar os desafios associados à detecção de microlesões. A Figura 21 ilustra o *pipeline* completo de pré-processamento aplicado a uma imagem de fundo do conjunto DDR. A aplicação desse *pipeline* permitiu uma ampliação significativa dos conjuntos de dados utilizados nos experimentos, elevando o número de imagens do DDR de 757

para 4.570, e do IDRiD de 81 para 1.324. A seguir, cada uma das etapas realizadas é descrita em detalhes.

Figura 21 – *Pipeline* de pré-processamento da abordagem proposta aplicado a uma imagem de fundo do conjunto de dados DDR. Em (a) Imagem original; (b) Filtro de mediana; (c) CLAHE; (d) *Cropping*; e, (e) SAHI.



Fonte: Elaborada pelo autor.

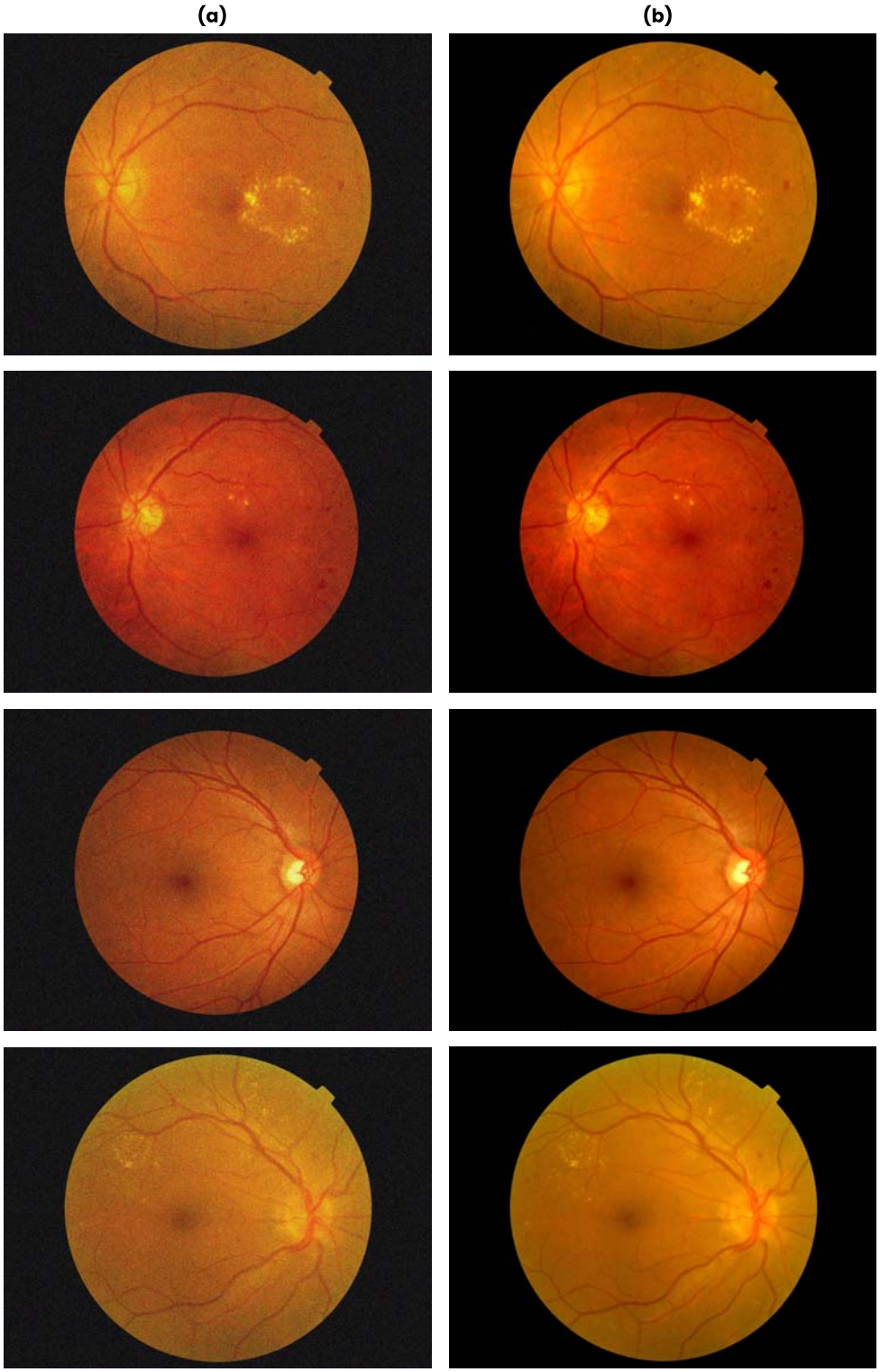
3.1.2.1 Filtro de Mediana

A primeira etapa de pré-processamento realizada nas imagens de fundo consiste na aplicação de um filtro de mediana, com o objetivo de remover possíveis ruídos gerados durante a captura das imagens de fundo do olho. O filtro de mediana é um método não linear eficaz na eliminação de ruídos impulsivos, como o ruído *Salt & Pepper*. Ele é amplamente utilizado no processamento de imagens por sua capacidade de reduzir ruídos preservando as bordas e os detalhes relevantes. Além disso, pode ser empregado na suavização de informações e na remoção de valores atípicos.

Seu funcionamento baseia-se na substituição de cada valor de píxel pela mediana dos valores de seus vizinhos, o que permite reduzir ruídos sem distorcer significativamente os dados originais (Sheeba; Raj, 2022). A operação é realizada por meio de uma janela deslizante que percorre toda a imagem. Para cada píxel, calcula-se a mediana dos valores contidos na janela, substituindo o valor original pelo resultado obtido (Mishiba, 2023).

Neste trabalho, foi utilizada uma janela de tamanho 5×5. A Figura 22 apresenta os resultados da aplicação do filtro em imagens de fundo do olho do conjunto DDR, com ruído *Salt & Pepper* inserido artificialmente.

Figura 22 – Resultado aplicação do filtro de mediana de tamanho 5×5 em imagens de fundo do conjunto de dados DDR com a introdução artificial de ruído *Salt & Pepper*. Em (a) Imagens com ruído *Salt & Pepper*; e, (b) Imagens após aplicação do filtro de mediana 5×5.



Fonte: Elaborada pelo autor.

A fim de avaliar a eficácia da remoção de ruído nas imagens filtradas, foi utilizada a métrica *Peak Signal-to-Noise Ratio* ($PSNR$), conforme adotado em Santos (2023),

que analisaram a qualidade da filtragem de ruído em imagens de fundo utilizando essa métrica. O $PSNR$ é expresso em decibéis (dB) e indica a razão entre o sinal de pico e o ruído presente na imagem. Valores mais altos de $PSNR$ indicam melhor qualidade da imagem processada em relação à original.

Para o cálculo do $PSNR$, é necessário inicialmente determinar o *Mean Squared Error* (MSE), que representa o erro quadrático médio entre a imagem original e a imagem filtrada. A Equação 2 apresenta essa métrica, em que M e N correspondem, respectivamente, ao número de linhas e colunas das imagens.

$$MSE = \frac{\sum_{m,n} [I_1(m,n) - I_2(m,n)]^2}{M * N} \tag{2}$$

A partir do MSE, o valor de $PSNR$ é calculado conforme apresentado na Equação 3, em que R representa o maior valor possível para o tipo de dado da imagem (por exemplo, 255 para imagens com codificação em 8 bits).

A Tabela 4 apresenta os valores de $PSNR$ obtidos após a aplicação do filtro de mediana com janela 5×5, visando à remoção de possíveis ruídos gerados durante a captura das imagens de fundo do conjunto de dados DDR. Os resultados mostram um aumento considerável nos valores de $PSNR$ após o filtro, o que indica uma redução efetiva do ruído e uma melhora na qualidade visual das imagens processadas em comparação às imagens ruidosas.

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \tag{3}$$

Tabela 4 – Valores obtidos com a métrica $PSNR$ após a remoção de ruídos do tipo *Salt & Pepper* por meio da aplicação do filtro de mediana com janela de 5×5 em imagens de fundo do conjunto de dados DDR.

Imagem	$PSNR$ **imagem com ruído**	$PSNR$ **imagem com filtro de mediana 5×5**
007-3820-200	15,9170	33,3309
007-4433-200	16,2415	36,0555
20170509100155767	16,0634	38,2022
20170516092409004	15,7926	38,5344

Fonte: Elaborada pelo autor.

3.1.2.2 CLAHE

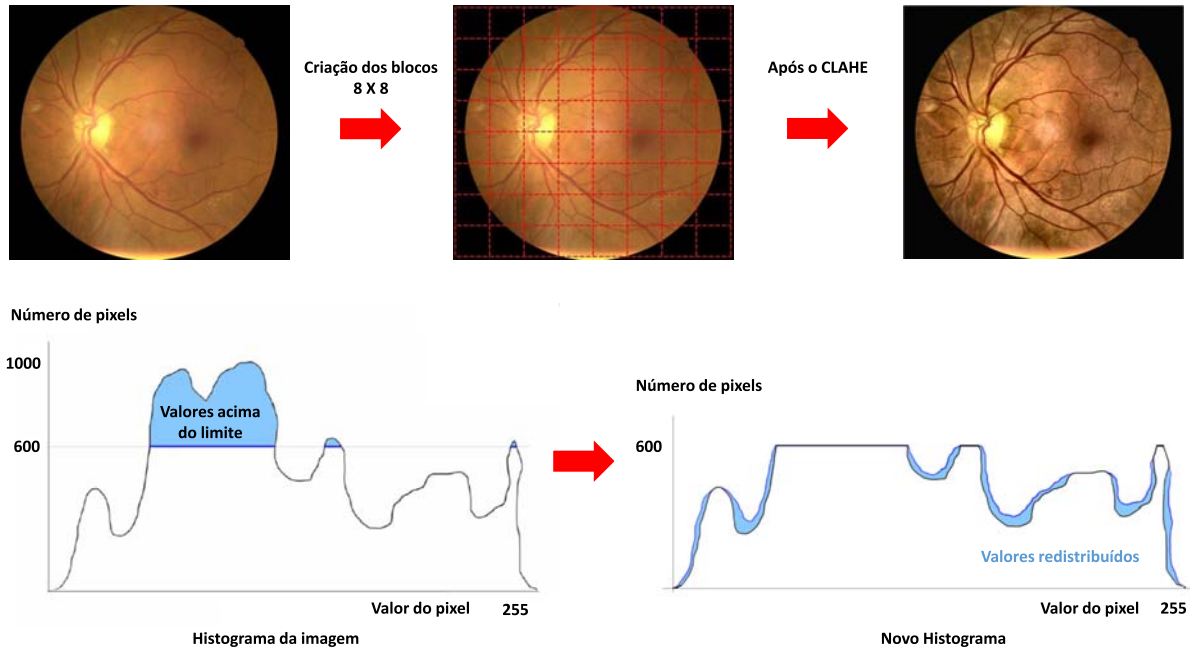
A próxima etapa do pré-processamento consiste na aplicação do método CLAHE, com o objetivo de melhorar o contraste das imagens. Diferentemente da equalização global do histograma, o CLAHE atua de forma local, o que o torna particularmente eficaz em imagens com variações de iluminação ou regiões com baixo contraste (Kuran; Kuran, 2021).

Inicialmente, as imagens são convertidas para o espaço de cores LAB, no qual o componente de luminância (L) é separado dos componentes de cromaticidade (A e B). Essa separação permite que a manipulação do contraste ocorra exclusivamente no canal de luminância, evitando a introdução de distorções de cor na imagem resultante.

Conforme ilustrado na Figura 23, o método subdivide a imagem em pequenos blocos, sobre os quais a equalização do histograma é aplicada individualmente. Para evitar a amplificação excessiva de ruído em regiões homogêneas, o CLAHE impõe um *clip limit* à altura dos picos do histograma (Chen et al., 2023). Neste trabalho, adotou-se um *clip limit* de 6 e um tamanho de bloco de 8×8 – valores amplamente utilizados na literatura por proporcionarem um equilíbrio entre o aprimoramento do contraste e a preservação dos detalhes da imagem (D. D. Silva; B. P. Carneiro; F. S. Cardoso, 2018; Alyoubi; Abulkhair; Shalash, 2021; Santos, 2023).

Após a aplicação do CLAHE, os blocos são interpolados a fim de garantir uma transição suave entre as regiões processadas, minimizando a ocorrência de artefatos visuais.

Figura 23 – O método CLAHE segmenta a imagem em blocos de 8×8 e realiza a equalização do histograma individualmente em cada uma dessas regiões. Durante esse processo, os valores de píxel que excedem o limiar predefinido são redistribuídos, resultando em um novo histograma com contraste aprimorado.



Fonte: Santos (2023).

Na sequência, os componentes de cromaticidade (A e B) são recombinados com o componente de luminância aprimorado, e a imagem é convertida novamente para o espaço de cores original. Por fim, é aplicado um filtro bilateral para realizar a su-

avização da imagem e minimizar possíveis artefatos gerados após o realce. Foram utilizados os mesmos parâmetros adotados em Santos (2023), com diâmetro do filtro $d = 9$ para realizar uma filtragem mais intensa dos ruídos e $\sigma = 75$ a fim de evitar que as lesões presentes nas imagens de fundo fossem desfocadas.

Essa abordagem tem sido amplamente adotada em aplicações recentes de processamento de imagens médicas, com o objetivo de aumentar o contraste e revelar detalhes sutis que podem não ser perceptíveis a olho nu (Chen et al., 2023). A Figura 24 apresenta os resultados da aplicação do CLAHE em imagens do conjunto de dados DDR, previamente submetidas ao filtro de mediana.

Para avaliar a qualidade do contraste obtido nas imagens após o realce, foram utilizadas as métricas *Measure of Enhancement* (EME) e Entropia (E). A métrica EME é empregada para quantificar o grau de realce de contraste em imagens, sendo seu cálculo baseado na divisão da imagem em blocos menores, nos quais se determina a razão entre os valores máximos e mínimos de intensidade. Essa razão é então utilizada para estimar o contraste local em cada bloco. Quanto maior o valor de EME, maior é o contraste percebido na imagem.

Assim como em Silva (2018) e Santos (2023), neste trabalho adotou-se o tamanho de bloco igual a 8×8. A Equação 4 descreve o cálculo do EME, em que K_1 representa a quantidade de linhas na matriz de subimagens; K_2, a quantidade de colunas; I^W_{MAX}, a intensidade máxima de nível de cinza na subimagem W; e I^W_{MIN}, a intensidade mínima na mesma subimagem.

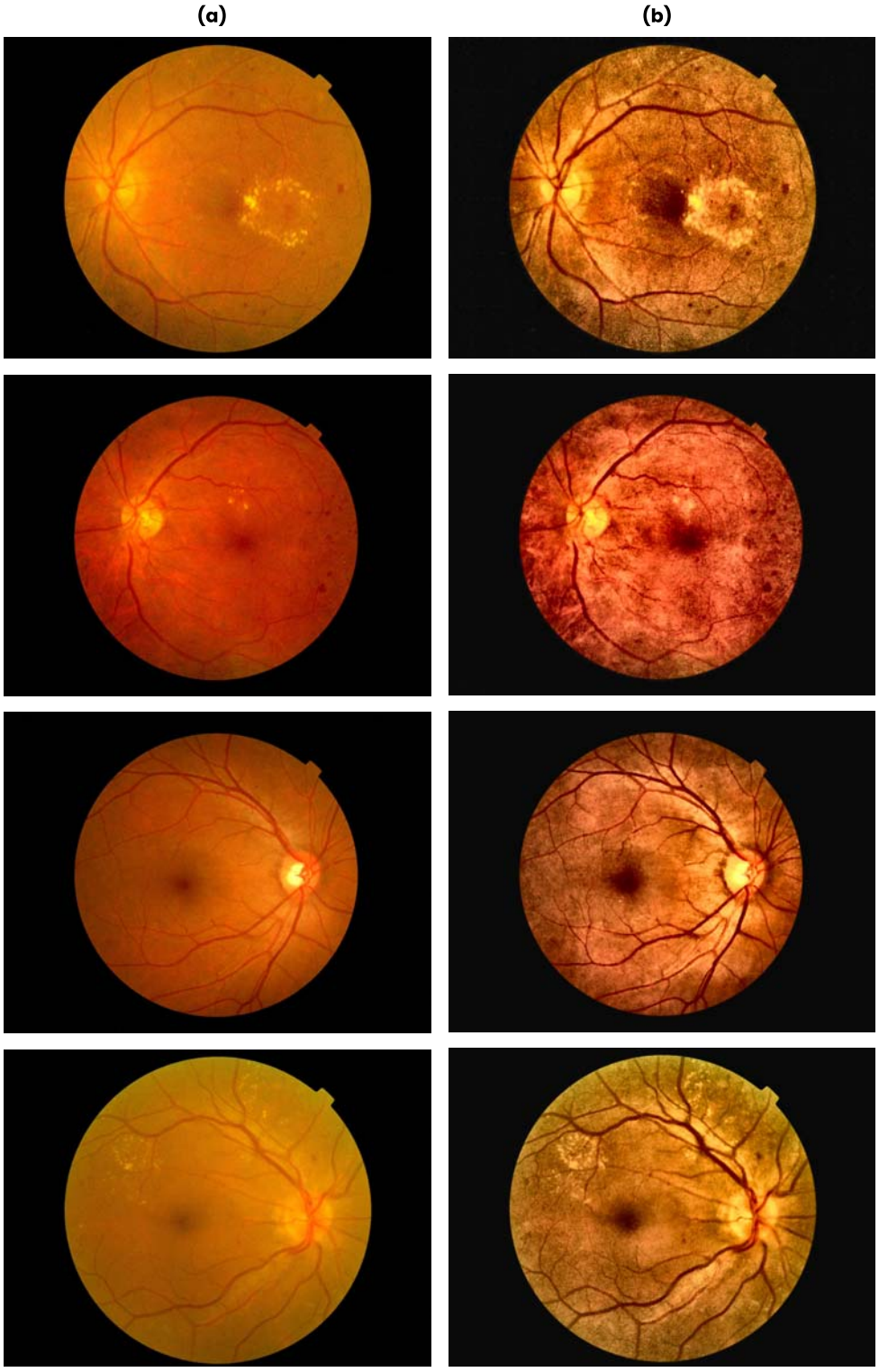
$$EME = \frac{1}{K_1 K_2} \sum_{l=1}^{K_2} \sum_{k=1}^{K_1} 20 \log \left(\frac{I^W_{MAX}; k, l}{I^W_{MIN}; k, l} \right) \tag{4}$$

Já a métrica E corresponde a uma medida estatística da quantidade de informação presente na imagem, sendo frequentemente utilizada para avaliar sua complexidade estrutural. Seu cálculo baseia-se na distribuição de probabilidade dos níveis de intensidade de cinza. Dessa forma, imagens com valores mais elevados de E geralmente apresentam maior diversidade tonal e riqueza de detalhes.

A Equação 5 expressa o cálculo da entropia E, onde $P(x_i)$ representa a probabilidade de ocorrência do nível de cinza i. Na Tabela 5, são apresentados os valores de EME e E obtidos após a aplicação do CLAHE em imagens de fundo do conjunto de dados DDR previamente processadas com o filtro de mediana. Os resultados, aliados à avaliação qualitativa das imagens, indicam um aumento significativo no contraste e na percepção visual dos detalhes.

$$E(x) = -\sum_{i=0}^{255} P(x_i) \log P(x_i) \tag{5}$$

Figura 24 – Resultado aplicação do CLAHE com tamanho de bloco de 8×8 e clip limit igual a 6, em imagens de fundo do conjunto de dados DDR com a aplicação do filtro de mediana. Em (a) Imagens com filtro de mediana; e, (b) Imagens após aplicação CLAHE.



Fonte: Elaborada pelo autor.

Tabela 5 – Valores obtidos com as métrica EME e E após a aplicação do CLAHE em imagens de fundo do conjunto de dados DDR previamente processadas com o filtro de mediana.

Imagem	EME		E	
	Sem CLAHE	**Com CLAHE**	**Sem CLAHE**	**Com CLAHE**
007-3820-200	2,8887	17,9546	4,5975	5,6006
007-4433-200	3,3868	15,9167	3,9825	4,7426
20170509100155767	2,8895	15,7881	4,0451	4,6904
20170516092409004	2,1867	16,1031	4,4526	5,0083

Fonte: Elaborada pelo autor.

3.1.2.3 Cropping

A próxima etapa do pré-processamento consiste na aplicação da técnica de *cropping* para remoção do fundo preto das imagens. Essa abordagem, também adotada em diversos trabalhos anteriores (Alyoubi; Abulkhair; Shalash, 2021; Santos, 2023; Pereira et al., 2023; Dias et al., 2023; Jiang; Zhao, 2024), tem como principal objetivo reduzir a ocorrência de Falsos Positivos (FP) durante a segmentação de lesões retinianas, por meio da eliminação de regiões compostas por píxeis sem informações relevantes. Em imagens do fundo do olho, apenas a área correspondente à retina contém dados significativos, enquanto as regiões externas são consideradas plano de fundo.

A primeira etapa do *pipeline* de *cropping* consiste na aplicação de um filtro de mediana com janela de tamanho 31×31. Essa operação tem como finalidade suavizar a imagem e eliminar detalhes irrelevantes para a detecção da circunferência da retina. Em seguida, as imagens são convertidas para tons de cinza, com o intuito de destacar as regiões de interesse e facilitar a detecção das bordas. Na sequência, aplica-se o filtro de *Sobel* sobre a imagem em escala de cinza, a fim de identificar o contorno da retina.

O filtro de *Sobel* é uma técnica de detecção de bordas baseada no cálculo da primeira derivada das intensidades dos píxeis. Seu principal objetivo é identificar componentes de alta frequência na imagem, os quais resultam de variações abruptas de intensidade entre píxeis adjacentes – geralmente indicativas da presença de contornos. Para isso, o método estima o vetor gradiente da imagem a partir das derivadas parciais nos eixos horizontal e vertical, permitindo a identificação precisa das regiões de transição, como os limites entre a retina e o fundo preto (Millon; Frati; Rucci, 2020).

Posteriormente, a Transformada de Hough (HT) foi empregada para demarcar o contorno da retina, por ser uma técnica amplamente utilizada no Processamento Digital de Imagens (PDI) para detectar formas geométricas, como linhas e círculos, em imagens digitais. Seu principal diferencial reside na capacidade de representar essas

formas por meio de equações matemáticas conhecidas, o que permite identificá-las mesmo em ambientes com ruído ou estruturas complexas (Conde; Santos; Santos, 2025).

No caso da retina, cuja borda geralmente se aproxima de uma circunferência, a HT permite identificar essa estrutura ao mapear os pontos da imagem para um espaço paramétrico, no qual cada ponto contribui com votos para possíveis centros e raios de círculos. O acúmulo desses votos indica a presença de uma circunferência na imagem original. Assim como em Santos (2023), utilizou-se o algoritmo *Hough Gradient 2-1 Hough Transform* (21HT).

O 21HT executa a HT em dois estágios: no primeiro, uma transformada é acumulada para identificar as coordenadas dos centros candidatos; no segundo, é gerado um histograma de raios para cada centro identificado na etapa anterior. Embora o 21HT apresente baixo desempenho na detecção de formas circulares pequenas, essa limitação não compromete a identificação da retina. Além disso, sua escolha justifica-se pelo bom desempenho geral do método e pela baixa exigência de espaço de armazenamento (Santos, 2023).

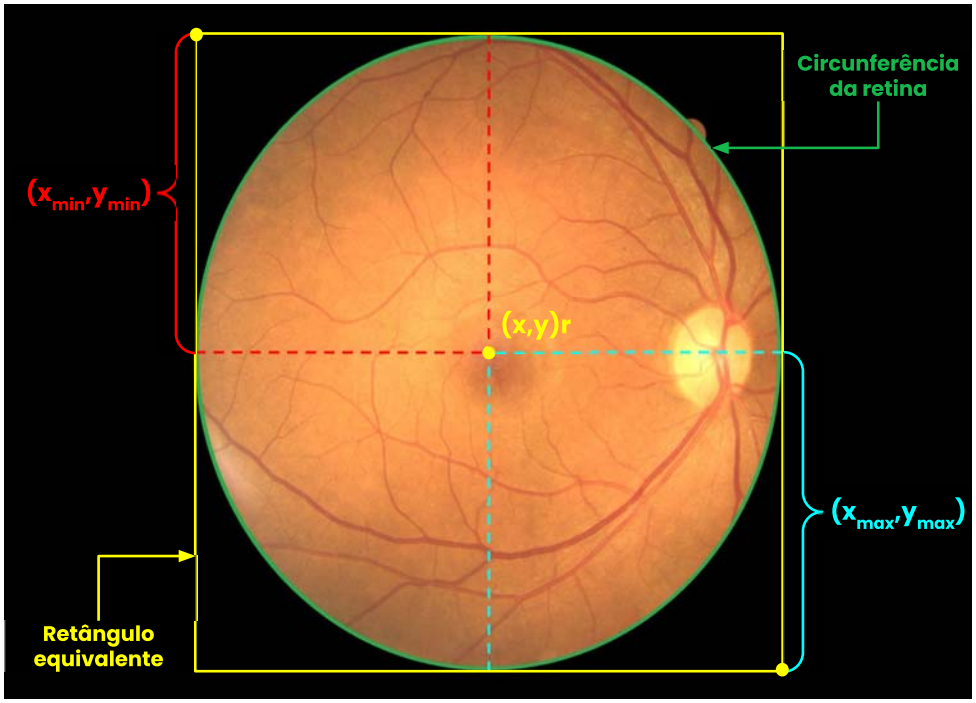
Conforme ilustrado na Figura 25, uma vez identificada a retina, a circunferência delimitadora é substituída por seu retângulo equivalente, definido a partir das coordenadas dos cantos superior esquerdo $(x_{\text{min}}, y_{\text{min}})$ e inferior direito $(x_{\text{max}}, y_{\text{max}})$ da imagem. Dessa forma, a área situada fora desse retângulo corresponde à região que será removida com o *cropping*. Por fim, a Figura 26 apresenta o *pipeline* completo para a realização do *cropping* do fundo preto das imagens retinianas.

3.1.2.4 SAHI

A última etapa do pré-processamento consiste na utilização do *framework* de código aberto SAHI (Akyon; Altinuc; Temizel, 2022). Essa ferramenta adota uma abordagem baseada na divisão da imagem em múltiplos blocos, com o objetivo de aprimorar o desempenho de inferência e o ajuste fino em tarefas de detecção e segmentação. O principal propósito dessa técnica é superar as limitações enfrentadas por modelos computacionais na identificação de objetos pequenos em imagens de alta resolução. Essa estratégia tem se mostrado especialmente eficaz na mitigação dos desafios relacionados à detecção de microlesões retinianas, como os MA, cuja área reduzida dificulta a extração eficiente de características pelas redes neurais. Essa limitação compromete a propagação da informação ao longo das camadas da rede, favorecendo a ocorrência do problema conhecido como dissipação do gradiente.

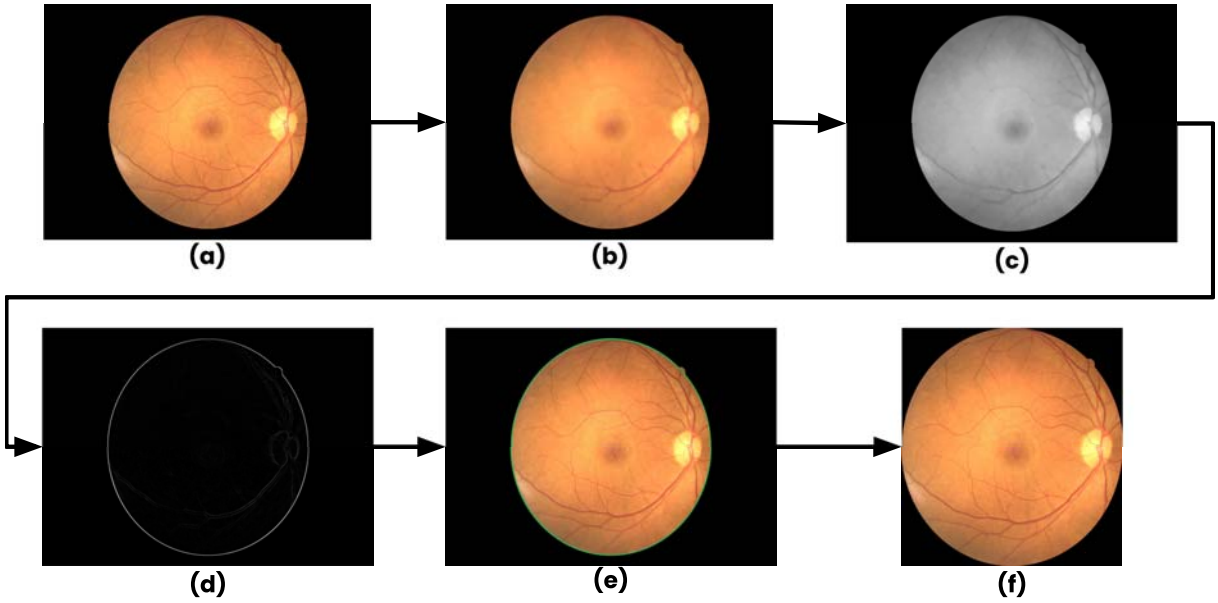
Empregado em estudos recentes, como os de Raudonis et al. (2023); Pereira et al. (2023); Danko et al. (2024), o método surge como uma alternativa ao tradicional *tiling*, utilizado por Santos (2023), com o objetivo de aumentar a precisão na detecção de microlesões retinianas. O processo adotado pelo *framework* consiste, basicamente,

Figura 25 – Transformação da circunferência da retina em seu retângulo equivalente por meio das coordenadas dos cantos superior esquerdo $(x_{\text{min}}, y_{\text{min}})$ e inferior direito $(x_{\text{max}}, y_{\text{max}})$.



Fonte: Elaborada pelo autor.

Figura 26 – *Pipeline* para realização do *cropping* do fundo preto das imagens de fundo. Em (a) Imagem original; (b) Imagem suavizada com filtro de mediana; (c) Imagem em tons de cinza; (d) Detecção de bordas com filtro de Sobel; (e) Demarcação do contorno da retina com a transformada de Hough; e, (f) Imagem após realização do *cropping*.



Fonte: Elaborada pelo autor.

na divisão da imagem de entrada em blocos sobrepostos, o que amplia a visibilidade de objetos de pequenas dimensões. Vale destacar que tanto as dimensões dos blocos quanto a taxa de sobreposição entre eles podem ser parametrizadas. Na abordagem proposta, foi adotada uma taxa de sobreposição de 20%, conforme em Pereira et al. (2023).

Adicionalmente, utilizaram-se blocos de 1.024$\times$1.024 píxeis, definidos com base em limitações computacionais que inviabilizaram o uso de blocos menores. No entanto, conforme destacado por Akyon; Altinuc; Temizel (2022), a divisão em blocos menores resulta em um aumento proporcional do tamanho relativo dos objetos dentro da região analisada, favorecendo a detecção de estruturas de pequenas dimensões. Assim, é razoável inferir que a adoção de blocos com dimensões reduzidas poderia proporcionar uma identificação ainda mais precisa das microlesões, uma vez que estas passariam a ocupar uma maior proporção da área analisada pelo detector.

Conforme ilustrado na Figura 27, cada bloco identificado resulta na geração de uma nova imagem. Esse processo é aplicado tanto às máscaras binárias contendo o *ground truth* (gt) das lesões quanto à imagem de fundo original. Dessa forma, as lesões presentes nas imagens geradas por essa etapa passam a ocupar uma área proporcionalmente maior do que nas imagens originais, que são redimensionadas na entrada das redes neurais. Esse aumento efetivo na resolução relativa das lesões favorece uma extração de características mais detalhada e precisa, refletindo-se diretamente na melhora da performance na segmentação.

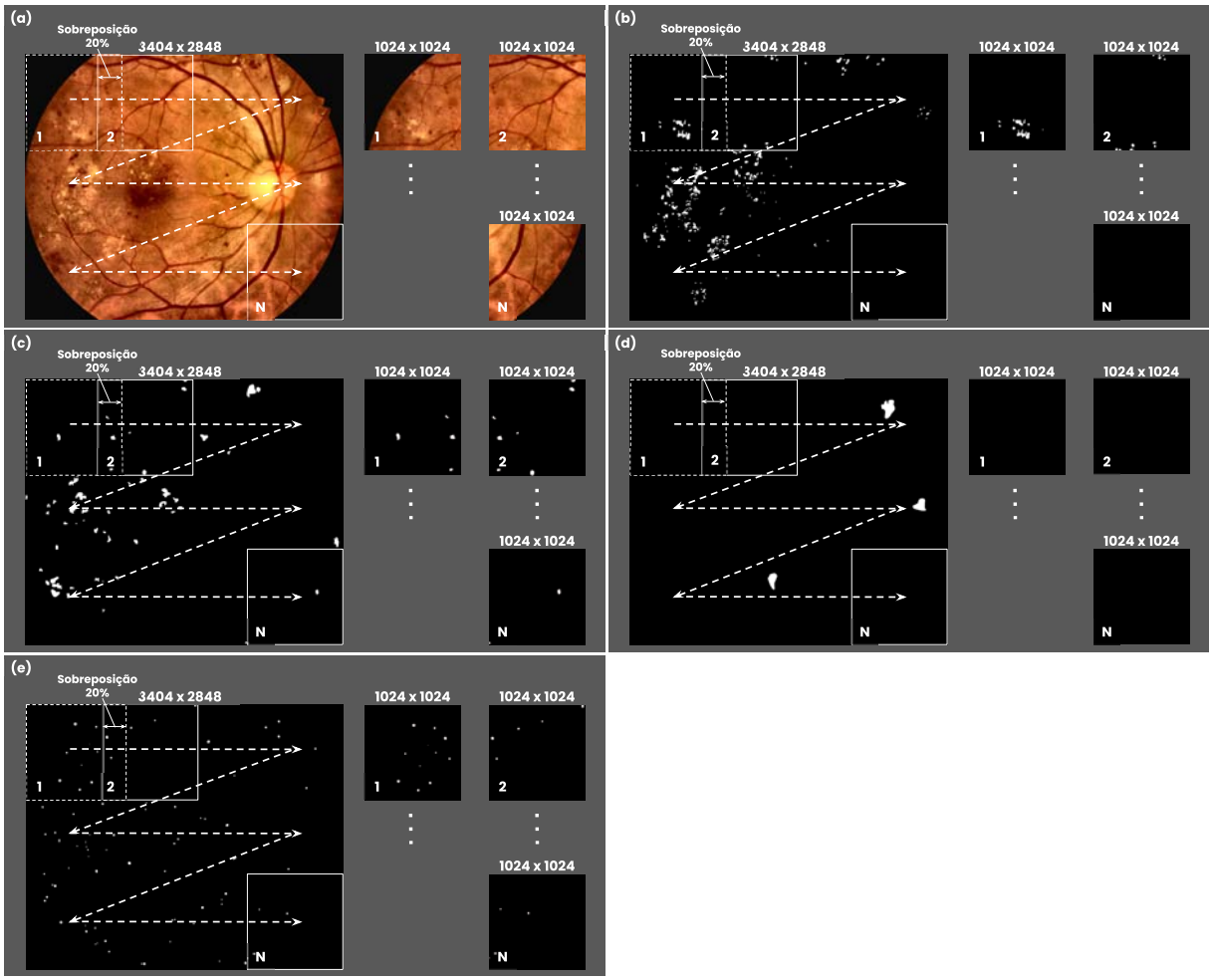
3.1.3 Aumento de Dados

Dentre os principais desafios enfrentados na construção de modelos computacionais para a segmentação de lesões de fundo associadas à RD, destaca-se a escassez de lesões rotuladas nas imagens da retina. Essa limitação é recorrente nos conjuntos de dados públicos de RD, incluindo o conjunto DDR utilizado no treinamento dos modelos da abordagem proposta neste trabalho. Para mitigar esse problema, foi aplicada uma etapa de aumento de dados sobre as imagens do conjunto de treino, com o objetivo de ampliar a variabilidade dos dados e aprimorar a eficácia dos modelos na segmentação das lesões retinianas.

Nesta etapa, foram empregadas diversas técnicas disponíveis na biblioteca *Albumentations: Fast and Flexible Image Augmentations* (Buslaev et al., 2020). Os seguintes métodos foram utilizados: *HorizontalFlip*, *VerticalFlip*, *ElasticTransform*, *GridDistortion* e *OpticalDistortion*. Os métodos *HorizontalFlip* e *VerticalFlip* foram aplicados como forma de reduzir vieses posicionais nos dados de entrada, realizando espelhamentos da imagem ao longo dos eixos y e x, respectivamente.

As transformações *ElasticTransform* e *GridDistortion* foram aplicadas com o intuito de simular variações estruturais e distorções nas imagens de fundo do olho. O *Elas-*

Figura 27 – Exemplo de execução do método COCO *slice* realizado pelo *framework* SAHI, com blocos de tamanho 1.024×1.024 píxeis e taxa de sobreposição de 20%. Em (a) Imagem de fundo original; (b) *Ground truth* dos EX; (c) *Ground truth* dos HE; (d) *Ground truth* dos SE; e, (e) *Ground truth* dos MA.



Fonte: Elaborada pelo autor.

ticTransform introduz deformações complexas e não lineares, semelhantes a ondas, modificando a geometria dos objetos. Já o *GridDistortion* altera a distribuição dos píxeis, criando distorções que simulam variações topográficas, como se a imagem fosse observada em perspectiva.

Adicionalmente, o método *OpticalDistortion* foi utilizado para simular imperfeições ópticas similares às causadas por lentes, enriquecendo o conjunto com variações realistas. A exposição do modelo a essas transformações visa melhorar sua capacidade de generalização frente às diferentes aparências e estruturas morfológicas das lesões retinianas. A Tabela 6 apresenta os parâmetros adotados em cada uma das transformações.

A Figura 28 ilustra o processo de aumento de dados aplicado a uma imagem de fundo, utilizando as técnicas previamente descritas. Cada técnica foi aplicada de forma

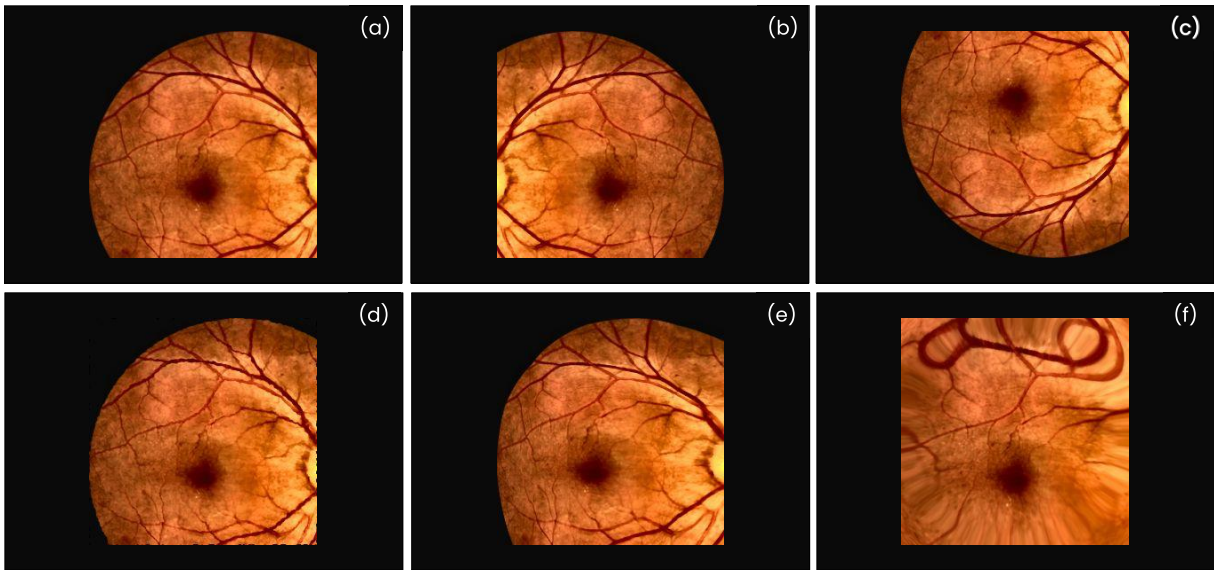
Tabela 6 – Valores de parâmetros usados ao aplicar métodos de aumento de dados nas imagens de fundo de olho.

Tipo de aumento	Parâmetros	Valores
HorizontalFlip	*probability*	1
VerticalFlip	*probability*	1
ElasticTransform	*alpha, alpha affine, sigma, probability*	120; 3,6; 6; 1
GridDistortion	*probability*	1
OpticalDistortion	*distort limit, shift limit, probability*	2; 0,5; 1

Fonte: Elaborada pelo autor.

independente, ou seja, as transformações foram realizadas uma de cada vez, sem combinações entre elas. Como resultado, foi possível gerar cinco novas variações a partir de cada imagem pré-processada do conjunto de treino do *dataset* DDR, totalizando 13.800 imagens disponíveis para o treinamento dos modelos que compõem a abordagem proposta.

Figura 28 – Exemplo de aumento de dados em uma imagem de fundo pré-processada do conjunto DDR. Em (a) Imagem pré-processada, (b) *HorizontalFlip*, (c) *VerticalFlip*, (d) *ElasticTransform*, (e) *GridDistortion*, e (f) *OpticalDistortion*.



Fonte: Elaborada pelo autor.

3.1.4 Treinamento e Integração dos Modelos

As arquiteturas U-Net++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018) e R2U-Net (Alom et al., 2018) foram selecionadas para compor os *ensembles* responsáveis pela segmentação semântica das lesões de fundo. Cada rede foi estruturada com cinco níveis hierárquicos, utilizando 64, 128, 256, 512 e 1024 filtros, respectiva-

mente do nível mais superficial ao mais profundo. O processo de *downsampling* e *upsampling* da resolução das imagens foi realizado em dois estágios. Além disso, os blocos residuais recorrentes da R2U-Net foram configurados com duas iterações internas. Por fim, para extração de características foi utilizado como *backbone* a arquitetura VGG-16 (Simonyan; Zisserman, 2014).

Para o treinamento dos modelos, todas as imagens foram redimensionadas para o tamanho de entrada das redes, configuradas para receber imagens com dimensões de $256 \times 256 \times 3$. Para cada tipo de lesão, foram treinados três modelos distintos, cada um baseado em uma das arquiteturas mencionadas, resultando em um conjunto de três modelos por lesão. Cada modelo foi treinado com taxa de aprendizado de 0,001, tamanho de lote igual a 4 e por um total de 50 épocas.

A função de ativação utilizada no processo de treinamento dos modelos foi a *ReLU*, combinada com o otimizador *Adam*. Para aprimorar a capacidade de generalização dos modelos, todos foram inicializados com pesos pré-treinados no *dataset* ImageNet (Russakovsky et al., 2015). O processo de pré-treinamento, amplamente adotado em redes neurais profundas, consiste em treinar os modelos previamente em um grande conjunto de dados genérico, como o ImageNet, que contém milhões de imagens rotuladas em diversas categorias. Os pesos resultantes desse treinamento aprendem representações visuais genéricas, como bordas, texturas e formas, que são reutilizáveis em diferentes tarefas de *Visão Computacional* (VC). Ao adotar esses pesos como ponto de partida, os modelos podem convergir mais rapidamente durante o treinamento específico para a segmentação das lesões, além de apresentarem melhor desempenho, mesmo em cenários com quantidade limitada de dados anotados (Kashiparekh et al., 2019).

Os pesos utilizados estão disponíveis publicamente no *Google Cloud Storage*[4]. Adicionalmente, os hiperparâmetros empregados no treinamento dos modelos da abordagem proposta estão apresentados na Tabela 7. Esses valores foram selecionados com base em Pereira et al. (2024), com o objetivo de permitir uma comparação direta entre modelos individuais e combinações de modelos integrados na forma de *ensembles*.

Os modelos treinados foram integrados por meio de *ensembles* ponderados, organizados individualmente para cada tipo de lesão. Em cada ensemble, as predições das arquiteturas U-Net++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018) e R2U-Net (Alom et al., 2018) foram combinadas por meio de uma média ponderada das máscaras de segmentação geradas por cada modelo. Os pesos atribuídos a cada arquitetura foram otimizados com o objetivo de maximizar a métrica *Dice Coefficient* (DC) no conjunto de validação do *dataset* DDR, utilizando o algoritmo de Evolução

[4] `https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5`

Tabela 7 – Hiperparâmetros utilizados nos modelos da abordagem proposta.

Hiperparâmetro	**Valor**	**Descrição**
Tamanho da Entrada	(256, 256, 3)	Dimensão das imagens de entrada esperadas pela rede (altura, largura, canais de cor).
Tamanho do lote	4	Número de amostras processadas antes de cada atualização dos pesos da rede.
Taxa de aprendizado	0,001	Controla o passo das atualizações de pesos durante o treinamento.
Número de épocas	50	Número de vezes que todos os dados de treinamento são mostrados à rede durante o treinamento.
Função de ativação	ReLU	Permite que o modelo aprenda padrões mais complexos, ao introduzir não linearidade entre as camadas.
Ativação de saída	Sigmoid	Transforma a saída para o intervalo [0, 1], adequada para segmentação binária.
Otimizador	Adam	Algoritmo que ajusta os pesos com base nos gradientes, combinando *momentum* e adaptação de taxa de aprendizado.
Normalização em lote	*True*	Normaliza a saída das camadas intermediárias, acelerando o treinamento e melhorando a estabilidade.

Fonte: Elaborada pelo autor.

Diferencial (Storn; Price, 1997). A Tabela 8 apresenta os pesos atribuídos às arquiteturas para cada tipo de lesão após a otimização.

Tabela 8 – Pesos atribuídos às arquiteturas para cada tipo de lesão após a otimização com o algoritmo de Evolução Diferencial.

Lesão	**Peso**		
	U-Net++	Attention U-Net	R2U-Net
Exsudatos Duros	0,4024	0,0627	0,5349
Hemorragias	0,2192	0,2601	0,5207
Exsudatos Algodonosos	0,1131	0,4042	0,4827
Microaneurismas	0,4047	0,4633	0,1320

Fonte: Elaborada pelo autor.

Para realizar a combinação das predições, utilizou-se a função `tensordot`, disponibilizada pela biblioteca NumPy (Harris et al., 2020). Essa função permitiu aplicar os pesos otimizados às máscaras geradas pelos modelos. As máscaras foram organizadas em um arranjo tridimensional, em que o primeiro eixo representa os diferentes modelos e os dois eixos subsequentes correspondem à altura e à largura da imagem.

A operação de combinação foi realizada ao longo do primeiro eixo, resultando em uma média ponderada das máscaras.

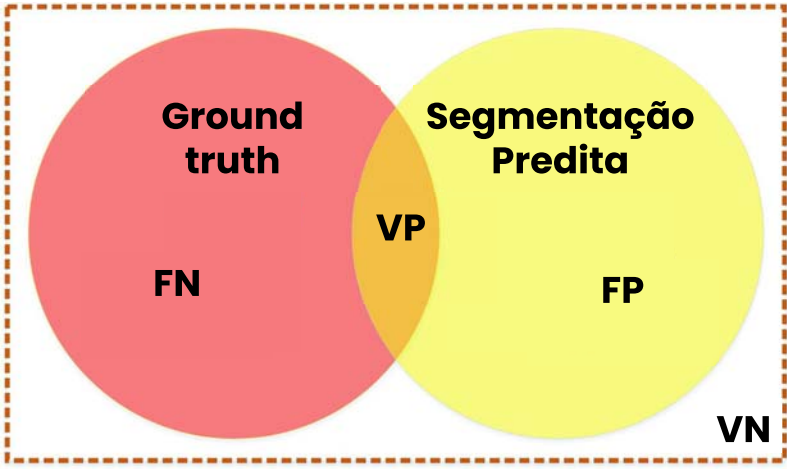
A predição final de cada píxel foi obtida a partir da aplicação de um limiar de 0,5 sobre os valores da média ponderada, gerando assim uma máscara binária combinada. Essa máscara final representa o resultado da segmentação e foi utilizada para o cálculo das métricas de avaliação da abordagem proposta.

3.1.5 Métricas de Avaliação

A avaliação da abordagem proposta foi conduzida por meio das métricas *Accuracy* (Acc), *Sensitivity* (Sen), *Precision* (Pre), *Dice Coefficient* (DC), *mean Dice Coefficient* (mDC), *Intersection over Union* (IoU) e *mean Intersection over Union* ($mIoU$). Para uma compreensão mais aprofundada dessas métricas, é essencial compreender os conceitos de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN).

Conforme ilustrado na Figura 29, os VP referem-se aos casos em que o modelo classifica corretamente uma instância como positiva. Os VN representam os casos em que a classificação correta ocorreu para instâncias negativas. Já os FP ocorrem quando o modelo classifica incorretamente uma instância negativa como positiva, enquanto os FN indicam falhas na detecção de instâncias verdadeiramente positivas.

Figura 29 – Diagrama da relação entre os conceitos de VP, VN, FP e FN, utilizados como base para o cálculo das métricas de avaliação do trabalho proposto.



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

A Acc é uma métrica amplamente utilizada que avalia a precisão geral do modelo, indicando a proporção de predições corretas em relação ao total de predições. Essa métrica é importante para se ter uma visão geral do desempenho do modelo, considerando tanto os VP quanto os VN. Já a métrica Sen mensura a proporção de casos

positivos que foram adequadamente identificados pelo modelo, sendo especialmente útil em situações em que a detecção de positivos é crucial.

A Pre é definida pela razão entre a quantidade de exemplos classificados corretamente como positivos e o total de exemplos classificados como positivos, atribuindo maior ênfase aos erros por FP. Por sua vez, o DC é uma métrica de validação estatística empregada para avaliar o desempenho e a precisão da sobreposição espacial entre duas amostras. É amplamente utilizado na análise de imagens e serve como indicador relevante para medir a concordância e a qualidade da segmentação, destacando a similaridade entre o gt e a predição do modelo.

Outra métrica comumente utilizada para avaliação de modelos na tarefa de segmentação de instâncias é o IoU. Essa métrica quantifica o nível de sobreposição entre as máscaras do gt e as máscaras previstas pelo modelo. Ela é expressa em uma escala de 0 a 1, onde 0 indica ausência total de sobreposição entre as máscaras, enquanto 1 indica uma sobreposição perfeita entre a máscara predita (pm) pelo modelo e a máscara do gt. As métricas de avaliação são calculadas conforme as equações abaixo:

$$Acc = \frac{\textit{(VP + VN)}}{\textit{(VP + FP + VN + FN)}} \tag{6}$$

$$Sen = \frac{\textit{VP}}{\textit{(VP + FN)}} \tag{7}$$

$$Pre = \frac{\textit{VP}}{\textit{(VP + FP)}} \tag{8}$$

$$DC = \frac{\textit{2VP}}{\textit{(2VP + FN + FP)}} \tag{9}$$

$$mDC = \frac{1}{N}\sum_{i=1}^{N} DC_i \tag{10}$$

$$IoU = \frac{\textit{Área de Sobreposição}}{\textit{Área de União}} = \frac{\textit{Área } (gt \cap pm)}{\textit{Área } (gt \cup pm)} \tag{11}$$

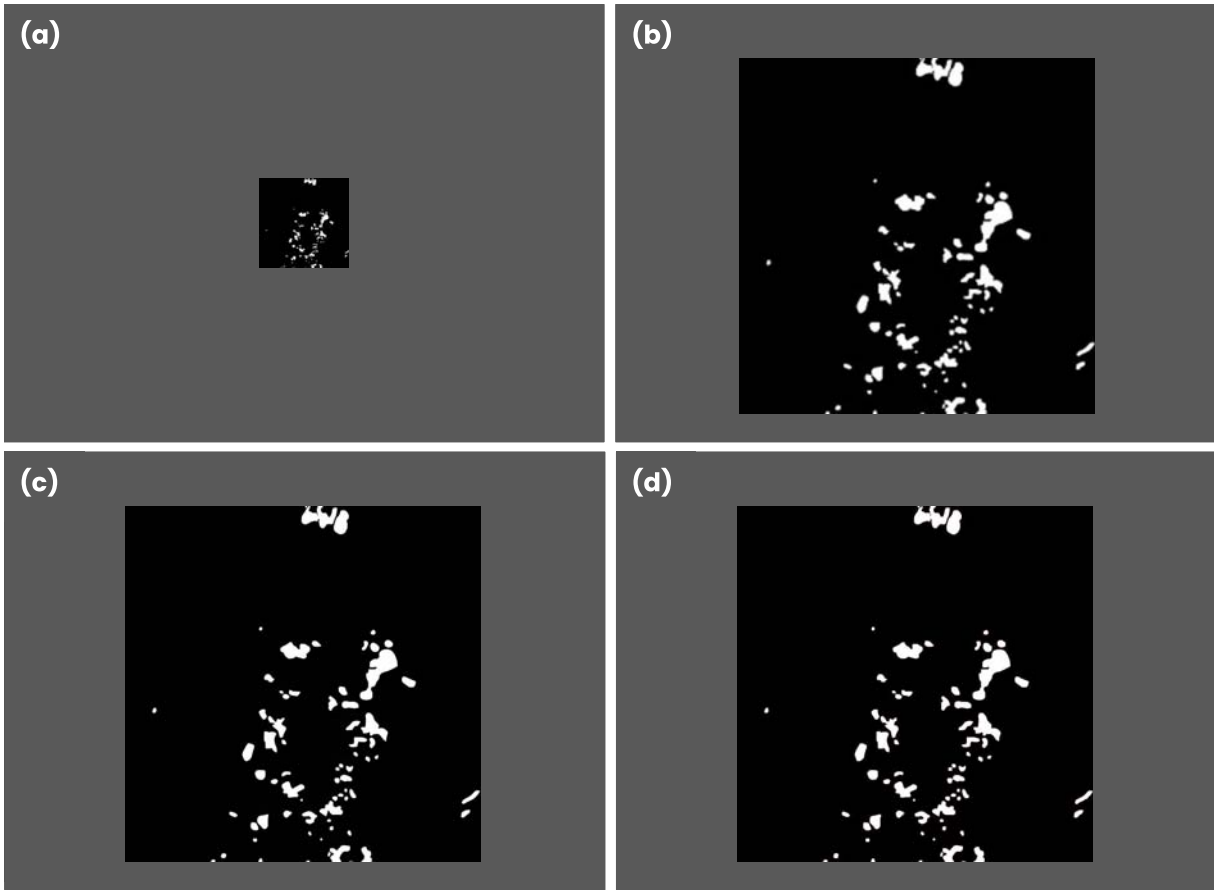
$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \tag{12}$$

3.1.6 Reconstrução do Mapa de Segmentação

Após a segmentação individual de cada bloco gerado pelo SAHI, utilizando a abordagem proposta – composta pelas arquiteturas U-Net++, Attention U-Net e R2U-Net integradas na forma de *ensembles* –, é necessário reconstruir o mapa completo de

segmentação da imagem original para cada um dos tipos de lesões de fundo. Como os modelos de segmentação treinados neste trabalho preveem máscaras na resolução de 256×256, é necessário realizar o redimensionamento dessas máscaras para 1.024×1.024, que corresponde ao tamanho original dos blocos. Esse redimensionamento foi realizado com a ferramenta Upscayl[5], que utiliza Redes Adversariais Generativas (em inglês, *Generative Adversarial Networks* – GAN) para o *upscaling*. O modelo empregado foi o Real-ESRGAN (Wang et al., 2021), reconhecido por sua eficácia na melhoria da resolução de imagens médicas (Aghelan; Rouhani, 2024; Rashid; Shakibapour; Ebrahimi, 2022; Nandal et al., 2024).

Figura 30 – Comparação entre métodos de redimensionamento das máscaras de segmentação. Em (a) Bloco predito (256×256); (b) Redimensionamento convencional (1.024×1.024); (c) Redimensionamento com Real-ESRGAN (1.024×1.024); e, (d) Bloco original gerado pelo SAHI (1.024×1.024).



Fonte: Elaborada pelo autor.

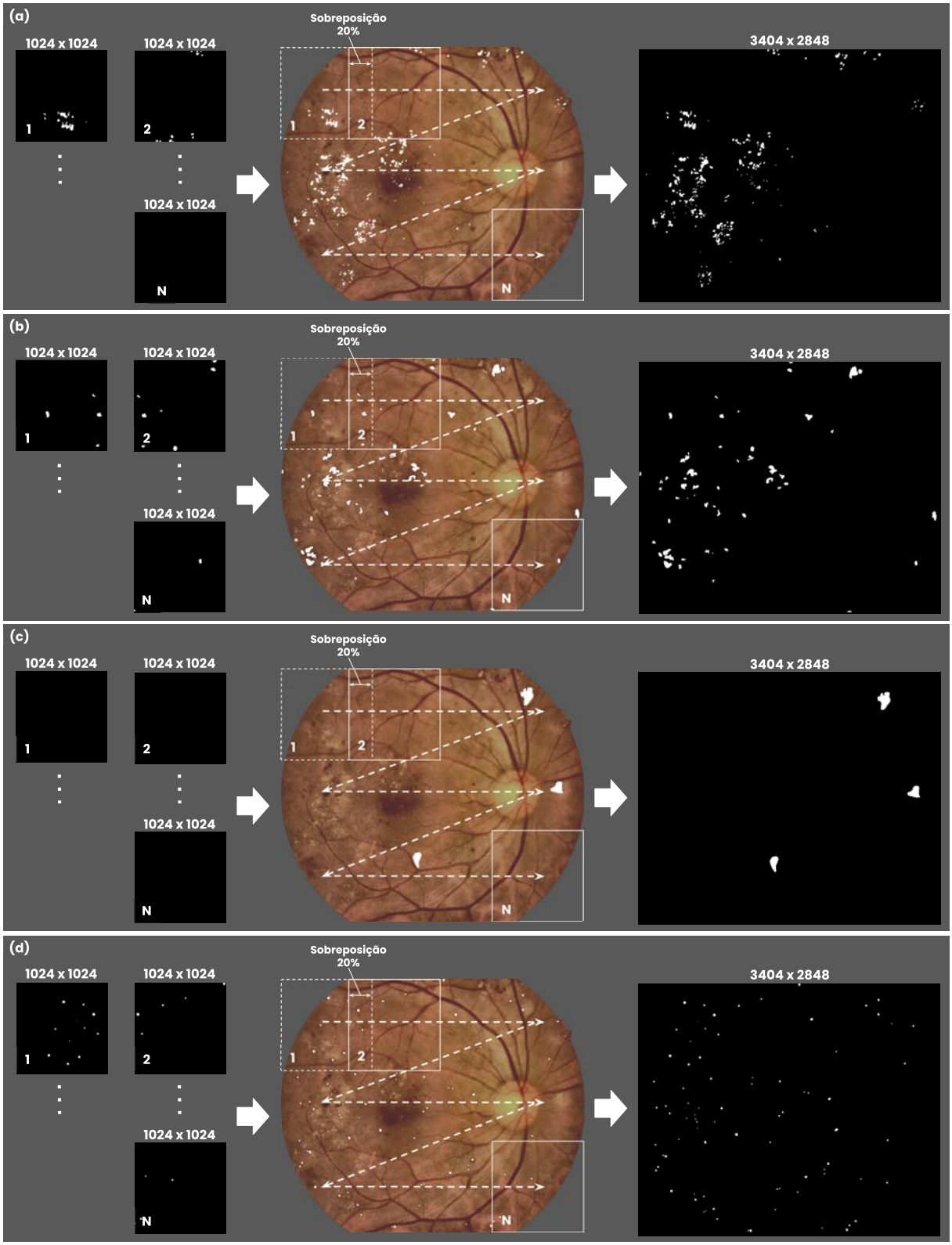
A Figura 30 apresenta uma comparação entre o redimensionamento convencional e uma abordagem de redimensionamento baseada em IA aplicada às máscaras de segmentação. Em (a), observa-se a máscara predita na resolução de 256×256. Em (b), é apresentado o resultado do redimensionamento convencional

[5] https://github.com/upscayl/upscayl

para 1.024$\times$1.024. Visualmente, é possível perceber algumas alterações nos contornos, o que também se reflete nas métricas, com IoU de 0,9389 e DC de 0,9685 em relação à máscara original. Já o redimensionamento realizado com o modelo Real-ESRGAN, mostrado em (c), apresentou IoU de 0,9624 e DC de 0,9808, indicando uma maior similaridade com a máscara original em relação ao redimensionamento convencional. Por fim, (d) exibe o bloco original gerado pelo SAHI, utilizado como referência para as comparações.

Na sequência, conforme ilustrado pela Figura 31, as predições redimensionadas com o modelo Real-ESRGAN são reposicionadas em suas respectivas localizações na imagem original, nos mesmos pontos exatos de onde os blocos foram extraídos. Como há sobreposição entre os blocos, algumas regiões da imagem possuem múltiplas predições. Nesses casos, os valores de probabilidade de cada píxel são somados e divididos pelo número de vezes em que aquele píxel foi avaliado, resultando em uma média. Por fim, aplica-se um limiar de 0,5 aos mapas preditos. Os píxeis com valor superior ao limiar são classificados como pertencentes à respectiva classe de lesão, enquanto os demais são considerados como fundo.

Figura 31 – Processo para a reconstrução do mapa de segmentação completo para cada um dos tipos de lesões de fundo. Em (a) Reconstrução do mapa de segmentação dos EX; (b) Reconstrução do mapa de segmentação dos HE; (c) Reconstrução do mapa de segmentação dos SE; e, (d) Reconstrução do mapa de segmentação dos MA.



Fonte: Elaborada pelo autor.

3.2 Experimentos, Resultados e Discussões

A abordagem proposta foi comparada com os modelos obtidos nos experimentos iniciais, realizados com o objetivo de selecionar as melhores arquiteturas para serem integradas na forma de *ensembles* para a segmentação das lesões retinianas. Nesse processo, foram avaliadas as seguintes arquiteturas: (1) U-Net (Ronneberger; Fischer; Brox, 2015), (2) U-Net++ (Zhou et al., 2018), (3) Attention U-Net (Oktay et al., 2018) e (4) R2U-Net (Alom et al., 2018). Cada uma dessas redes foi inicialmente analisada em sua configuração *baseline*, sem a aplicação de etapas adicionais. Em seguida, foram treinadas com diferentes combinações das etapas que compõem a abordagem proposta, incluindo o uso de pré-processamento, aumento de dados e pré-treinamento.

Adicionalmente, a abordagem proposta também foi comparada com modelos desenvolvidos em trabalhos relacionados, sendo eles: (1) HED (Li et al., 2019), (2) DeepLab-v3+ (Li et al., 2019), (3) R2U-Net (Pereira et al., 2024), (4) HACDR-Net (Xu et al., 2024), (5) U-Net++ (Jiang; Zhao, 2024) e (6) DenseCUNet (Zhang; Xu; Zhai, 2025).

A Tabela 9 apresenta os resultados obtidos na segmentação de EX, HE, SE e MA, utilizando as métricas Acc, Sen, Pre, DC e IoU no conjunto de validação do *dataset* DDR. Os valores em negrito indicam os melhores desempenhos para cada métrica e tipo de lesão. Valores representados com o símbolo $-$ indicam que a métrica correspondente não foi reportada no trabalho em questão.

Ao analisar os resultados, observa-se que a abordagem proposta superou os demais modelos em todas as métricas, com exceção da Acc. Essa diferença pode estar relacionada ao fato de que a abordagem proposta remove o fundo preto das imagens e realiza a segmentação a partir de blocos gerados pelo método slice do SAHI, reduzindo a presença de regiões sem lesões. Nos demais modelos, essas regiões são mantidas e, por serem facilmente classificadas como negativas, contribuem para o aumento dos valores de Acc.

Apesar disso, os resultados nas demais métricas evidenciam a eficácia da abordagem proposta, com destaque para os valores de DC de 0,7619, 0,6051, 0,7663 e 0,4167 para as classes EX, HE, SE e MA, respectivamente, além de um mDC igual a 0,6375. Os valores correspondentes de IoU foram 0,8904, 0,7975, 0,8778 e 0,7615, com $mIoU$ de 0,8318.

Além disso, desconsiderando os resultados da abordagem proposta, observa-se que os modelos U-Net++, Attention U-Net e R2U-Net, quando combinados com as etapas de pré-processamento, aumento de dados e pré-treinamento, apresentaram os melhores desempenhos nas métricas avaliadas. Esses resultados evidenciam a eficácia dessas arquiteturas quando utilizadas em conjunto com a metodologia desenvolvida neste trabalho.

Tabela 9 – Resultados obtidos na segmentação de Exsudatos Duros, Hemorragias, Exsudatos Algodonosos e Microaneurismas comparados com outros modelos utilizando as métricas Acc, Sen, Pre, DC e IoU no conjunto de validação do *dataset* DDR. O símbolo $-$ indica que a métrica correspondente não foi reportada no trabalho original do respectivo modelo.

Modelos	**EX**					**HE**				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
U-Net (Ronneberger; Fischer; Brox, 2015)	0,9994	0,1666	0,3064	0,1218	0,5429	0,9984	0,0000	0,0000	0,0000	0,4990
U-Net + aumento de dados	0,9994	0,2029	0,4357	0,1778	0,5628	0,9983	0,0217	0,1540	0,0186	0,5047
U-Net + pré-treinamento	0,9994	0,4639	0,2295	0,2119	0,5757	0,9205	0,0690	0,4838	0,0020	0,0035
U-Net + pré-treinamento + aumento de dados	0,9995	0,3101	0,3343	0,2277	0,5843	0,9987	0,3587	0,4838	0,3308	0,6176
U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9992	0,8192	0,6853	0,5922	0,7726	0,9978	0,6722	0,7051	0,5414	0,7407
U-Net++ (Zhou et al., 2018)	0,9995	0,0000	0,0000	0,0000	0,4995	0,9984	0,0000	0,0000	0,0000	0,4990
U-Net++ + aumento de dados	0,9995	0,0001	0,0666	0,0002	0,4996	0,9979	0,0003	0,0286	0,0003	0,4988
U-Net++ + pré-treinamento	0,9995	0,0000	0,0000	0,0000	0,4995	0,9984	0,0000	0,0000	0,0000	0,4990
U-Net++ + pré-treinamento + aumento de dados	0,9995	0,0004	0,3056	0,0007	0,4997	0,9983	0,0000	0,0000	0,0000	0,4990
U-Net++ + pré-treinamento + aumento de dados + pré-processamento	0,9988	0,7321	0,8765	0,6845	0,8637	0,9971	0,5323	0,8837	0,5100	0,7711
Attention U-net (Oktay et al., 2018)	0,9995	0,1434	0,3737	0,1279	0,5448	0,9984	0,0000	0,0000	0,0000	0,4990
Attention U-net + aumento de dados	0,9995	0,2903	0,2932	0,1848	0,5651	0,9982	0,0443	0,1725	0,0431	0,5122
Attention U-net + pré-treinamento	0,9993	0,4197	0,1966	0,1842	0,5644	0,9982	0,0247	0,6314	0,0412	0,5118
Attention U-net + pré-treinamento + aumento dados	0,9995	0,2242	0,3653	0,1810	0,5627	0,9983	0,1514	0,4428	0,1495	0,5500
Attention U-net + pré-treinamento + aumento de dados + pré-processamento	0,9992	0,8040	0,7224	0,6107	0,7889	0,9977	0,6193	0,6126	0,4815	0,7070
R2U-Net (Alom et al., 2018)	0,9992	0,3631	0,2948	0,2073	0,5720	0,9962	0,0740	0,1270	0,0509	0,5137
R2U-Net + aumento de dados (Pereira et al., 2024)	**0,9997**	0,3234	0,4809	0,2705	0,6053	**0,9988**	0,2284	0,7585	0,2991	0,6094
R2U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9992	0,7830	0,7552	0,6698	0,8050	0,9976	0,6470	0,8021	0,5758	0,7722
HED (Li et al., 2019)	-	-	-	-	0,0948	-	-	-	-	0,2183
DeepLab-v3+ (Li et al., 2019)	-	-	-	-	0,2910	-	-	-	-	0,2819
Abordagem proposta	0,9992	**0,8405**	**0,9442**	**0,7619**	**0,8904**	0,9978	**0,6810**	**0,9101**	**0,6051**	**0,7975**

Modelos	**SE**					**MA**				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
U-Net (Ronneberger; Fischer; Brox, 2015)	0,9990	0,0000	0,0000	0,0000	0,4992	**0,9997**	0,0000	0,0000	0,0000	0,4999
U-Net + aumento de dados	0,9991	0,1336	0,4984	0,1469	0,5531	**0,9997**	0,0450	0,3728	0,0662	0,5237
U-Net + pré-treinamento	**0,9992**	0,2656	0,7786	0,3054	0,6184	**0,9997**	0,0319	0,2575	0,0465	0,5158
U-Net + pré-treinamento + aumento de dados	**0,9992**	0,4068	0,6603	0,4062	0,6588	0,9996	0,1192	0,1024	0,0894	0,5251
U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9988	0,7779	0,9141	0,7341	0,8514	0,9996	0,4136	0,6946	0,3798	0,6581
U-Net++ (Zhou et al., 2018)	0,9990	0,0000	0,0000	0,0000	0,4992	**0,9997**	0,0000	0,0000	0,0000	0,4999
U-Net++ + aumento de dados	0,9985	0,0197	0,0463	0,0167	0,5044	**0,9997**	0,0000	0,0000	0,0000	0,4999
U-Net++ + pré-treinamento	0,9990	0,0000	0,0000	0,0000	0,4992	**0,9997**	0,0000	0,0000	0,0000	0,4999
U-Net++ + pré-treinamento + aumento de dados	0,9989	0,0001	0,0666	0,0002	0,4992	**0,9997**	0,0000	0,0000	0,0000	0,4999
U-Net++ + pré-treinamento + aumento de dados + pré-processamento	0,9983	0,7148	0,9298	0,6775	0,8382	0,9996	0,2933	0,9110	0,2860	0,7278
Attention U-net (Oktay et al., 2018)	0,9990	0,0149	0,5928	0,0238	0,5066	**0,9997**	0,0000	0,0000	0,0000	0,4999
Attention U-net + aumento de dados	0,9981	0,2930	0,3111	0,2009	0,5721	**0,9997**	0,0375	0,2875	0,0532	0,5185
Attention U-net + pré-treinamento	**0,9992**	0,2357	0,7889	0,2912	0,6091	**0,9997**	0,0115	0,1909	0,0177	0,5052
Attention U-net + pré-treinamento + aumento de dados	**0,9992**	0,3411	0,7097	0,3843	0,6505	**0,9997**	0,0045	0,0830	0,0071	0,5017
Attention U-net + pré-treinamento + aumento de dados + pré-processamento	0,9987	0,7834	0,8954	0,7335	0,8566	0,9996	0,4130	0,6603	0,3608	0,6632
R2U-Net (Alom et al., 2018)	0,9991	0,1313	0,7759	0,1682	0,5601	**0,9997**	0,0408	0,5528	0,0629	0,5200
R2U-Net + aumento de dados (Pereira et al., 2024)	**0,9992**	0,2230	0,8604	0,2796	0,6072	**0,9997**	0,2298	0,3418	0,2185	0,5656
R2U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9987	0,7661	0,8970	0,7352	0,8583	0,9996	0,3461	0,7363	0,3587	0,6730
HED (Li et al., 2019)	-	-	-	-	0,0379	-	-	-	-	0,0204
DeepLab-v3+ (Li et al., 2019)	-	-	-	-	0,2756	-	-	-	-	0,0429
Abordagem proposta	0,9987	**0,7851**	**0,9703**	**0,7663**	**0,8778**	0,9996	**0,4315**	**0,9183**	**0,4167**	**0,7615**

Fonte: Elaborada pelo autor.

Dessa forma, foi possível explorar suas complementaridades e aprimorar ainda mais a precisão na segmentação das lesões de fundo, por meio da seleção desses modelos para compor os *ensembles* da abordagem proposta.

Na Tabela 10, são apresentados os resultados da segmentação das lesões de fundo EX, HE, SE e MA no conjunto de teste do *dataset* DDR, utilizando as métricas Acc, Sen, Pre, DC e IoU. Nesse conjunto, a abordagem proposta foi comparada com trabalhos relacionados do estado da arte (Pereira et al., 2024; Xu et al., 2024; Jiang; Zhao, 2024; Zhang; Xu; Zhai, 2025).

Assim como no conjunto de validação, a abordagem proposta superou os modelos comparados na maioria das métricas, atingindo valores de DC iguais a 0,6683, 0,5900, 0,8868 e 0,5138, e de IoU iguais a 0,7722, 0,6838, 0,9426 e 0,8476 para EX, HE, SE e MA, respectivamente. Além disso, foram obtidos valores de mDC igual a 0,6647 e de $mIoU$ igual a 0,8115.

Tabela 10 – Resultados obtidos na segmentação de Exsudatos Duros, Hemorragias, Exsudatos Algodonosos e Microaneurismas comparados com outros modelos utilizando as métricas Acc, Sen, Pre, DC e IoU no conjunto de teste do *dataset* DDR. O símbolo $-$ indica que a métrica correspondente não foi reportada no trabalho original do respectivo modelo.

Modelos	**EX**					**HE**				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
U-Net (Ronneberger; Fischer; Brox, 2015)	0,9976	0,1745	0,5455	0,1936	0,5668	0,9951	0,0001	0,5000	0,0001	0,4973
U-Net + aumento de dados	0,9976	0,2072	0,6625	0,2415	0,5831	0,9941	0,0074	0,1765	0,0068	0,4989
U-Net + pré-treinamento	0,9978	0,4868	0,4611	0,3791	0,6312	0,9110	0,0640	0,0052	0,0079	0,4577
U-Net + pré-treinamento + aumento de dados	0,9979	0,3357	0,5816	0,3354	0,6238	**0,9959**	0,2555	0,5553	0,2824	0,5939
U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9960	0,6509	0,6688	0,4837	0,6855	0,9924	0,4506	0,6841	0,3676	0,6469
U-Net++ (Zhou et al., 2018)	0,9972	0,0000	0,0000	0,0000	0,4983	0,9951	0,0000	0,0000	0,0000	0,4973
U-Net++ + aumento de dados	0,9972	0,0057	0,6601	0,0090	0,5014	0,9943	0,0008	0,0929	0,0007	0,4971
U-Net++ + pré-treinamento	0,9972	0,0000	0,0000	0,0000	0,4983	0,9951	0,0000	0,0000	0,0000	0,4973
U-Net++ + pré-treinamento + aumento de dados	0,9972	0,0072	0,6550	0,0084	0,5009	0,9951	0,0001	0,0117	0,0001	0,4973
U-Net++ + pré-treinamento + aumento de dados + pré-processamento	0,9944	0,4584	0,7763	0,5263	0,7343	0,9912	0,2770	0,8253	0,5591	0,6423
Attention U-net (Oktay et al., 2018)	0,9976	0,1474	0,5935	0,1693	0,5576	0,9951	0,0011	0,0356	0,0007	0,4975
Attention U-net + aumento de dados	0,9976	0,2666	0,5429	0,2660	0,5934	0,9951	0,0155	0,2620	0,0219	0,5032
Attention U-net + pré-treinamento	0,9977	0,4441	0,4214	0,3400	0,6206	0,9950	0,0200	0,6546	0,0340	0,5083
Attention U-net + pré-treinamento + aumento de dados	0,9979	0,3042	0,6012	0,3160	0,6145	0,9951	0,1000	0,5248	0,1257	0,5385
Attention U-net + pré-treinamento + aumento de dados + pré-processamento	0,9959	0,6329	0,7151	0,5068	0,7064	0,9923	0,3870	0,6593	0,5716	0,6416
R2U-Net (Alom et al., 2018)	0,9976	0,4077	0,4872	0,3314	0,6143	0,9902	0,0910	0,1586	0,0597	0,5137
R2U-Net + aumento de dados (Pereira et al., 2024)	**0,9980**	0,3566	0,6960	0,3777	0,6466	**0,9959**	0,1686	0,8138	0,2319	0,5778
R2U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9959	0,5735	0,7109	0,6566	0,7024	0,9915	0,4222	0,7334	0,5730	0,6634
HED (Li et al., 2019)	-	-	-	-	0,1874	-	-	-	-	0,0524
DeepLab-v3+ (Li et al., 2019)	-	-	-	-	0,3118	-	-	-	-	0,1425
HACDR-Net (Xu et al., 2024)	-	-	-	0,6124	0,4413	-	-	-	0,5396	0,3695
U-Net++ + ResNet50 (*Backbone*) + AFMA (Jiang; Zhao, 2024)	-	-	-	0,5664	0,3951	-	-	-	0,4774	0,3136
DenseCUNet (Zhang; Xu; Zhai, 2025)	-	-	-	0,5943	0,4181	-	-	-	0,4925	0,3181
Abordagem proposta	0,9956	**0,6942**	**0,9126**	**0,6683**	**0,7722**	0,9924	**0,4576**	**0,8879**	**0,5900**	**0,6838**

Modelos	**SE**					**MA**				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
U-Net (Ronneberger; Fischer; Brox, 2015)	0,9997	0,0001	0,3333	0,0002	0,4994	**0,9998**	0,0000	0,0000	0,0000	0,4998
U-Net + aumento de dados	0,9997	0,0777	0,1747	0,0516	0,5186	**0,9998**	0,0369	0,1457	0,0272	0,5082
U-Net + pré-treinamento	0,9997	0,1994	0,2619	0,1322	0,5533	**0,9998**	0,0077	0,0485	0,0077	0,5026
U-Net + pré-treinamento + aumento de dados	0,9997	0,3396	0,2117	0,1527	0,5593	0,9996	0,0824	0,0509	0,0382	0,5099
U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9996	0,9096	0,8907	0,8053	0,9053	0,9997	0,6277	0,4908	0,3107	0,6563
U-Net++ (Zhou et al., 2018)	0,9997	0,0000	0,0000	0,0000	0,4993	**0,9998**	0,0000	0,0000	0,0000	0,4998
U-Net++ + aumento de dados	0,9992	0,0429	0,0187	0,0108	0,5027	**0,9998**	0,0000	0,0000	0,0000	0,4998
U-Net++ + pré-treinamento	0,9997	0,0000	0,0000	0,0000	0,4993	**0,9998**	0,0000	0,0000	0,0000	0,4998
U-Net++ + pré-treinamento + aumento de dados	0,9997	0,0265	0,0841	0,0168	0,5048	**0,9998**	0,0000	0,0000	0,0000	0,4998
U-Net++ + pré-treinamento + aumento de dados + pré-processamento	0,9994	0,8980	0,8994	0,8140	0,9071	0,9997	0,5870	0,7953	0,4556	0,7886
Attention U-net (Oktay et al., 2018)	0,9997	0,0104	0,1224	0,0094	0,5025	**0,9998**	0,0000	0,0000	0,0000	0,4998
Attention U-net + aumento de dados	0,9983	0,2948	0,0515	0,0441	0,5138	**0,9998**	0,0499	0,1134	0,0299	0,5090
Attention U-net + pré-treinamento	0,9997	0,1840	0,2126	0,0998	0,5378	**0,9998**	0,0029	0,0531	0,0039	0,5018
Attention U-net + pré-treinamento + aumento de dados	0,9997	0,3043	0,2384	0,1508	0,5591	**0,9998**	0,0080	0,0616	0,0067	0,5015
Attention U-net + pré-treinamento + aumento de dados + pré-processamento	0,9995	0,9162	0,8868	0,8183	0,9071	0,9996	0,6219	0,4916	0,3308	0,6692
R2U-Net (Alom et al., 2018)	0,9997	0,0805	0,2077	0,0619	0,5215	**0,9998**	0,0430	0,2287	0,0381	0,5135
R2U-Net + aumento de dados (Pereira et al., 2024)	**0,9998**	0,1704	0,3953	0,1446	0,5584	**0,9998**	0,1899	0,1801	0,1099	0,5336
R2U-Net + pré-treinamento + aumento de dados + pré-processamento	0,9995	0,9118	0,8548	0,7976	0,8961	0,9996	0,5655	0,5508	0,3520	0,6777
HED (Li et al., 2019)	-	-	-	-	0,0782	-	-	-	-	0,0110
DeepLab-v3+ (Li et al., 2019)	-	-	-	-	0,2295	-	-	-	-	0,0325
HACDR-Net (Xu et al., 2024)	-	-	-	0,5564	0,3854	-	-	-	0,2634	0,1517
U-Net++ + ResNet50 (*Backbone*) + AFMA (Jiang; Zhao, 2024)	-	-	-	0,4286	0,2728	-	-	-	0,2837	0,1653
DenseCUNet (Zhang; Xu; Zhai, 2025)	-	-	-	0,4646	0,3066	-	-	-	0,2898	0,1736
Abordagem proposta	0,9996	**0,9186**	**0,9671**	**0,8868**	**0,9426**	0,9997	**0,6302**	**0,9036**	**0,5138**	**0,8476**

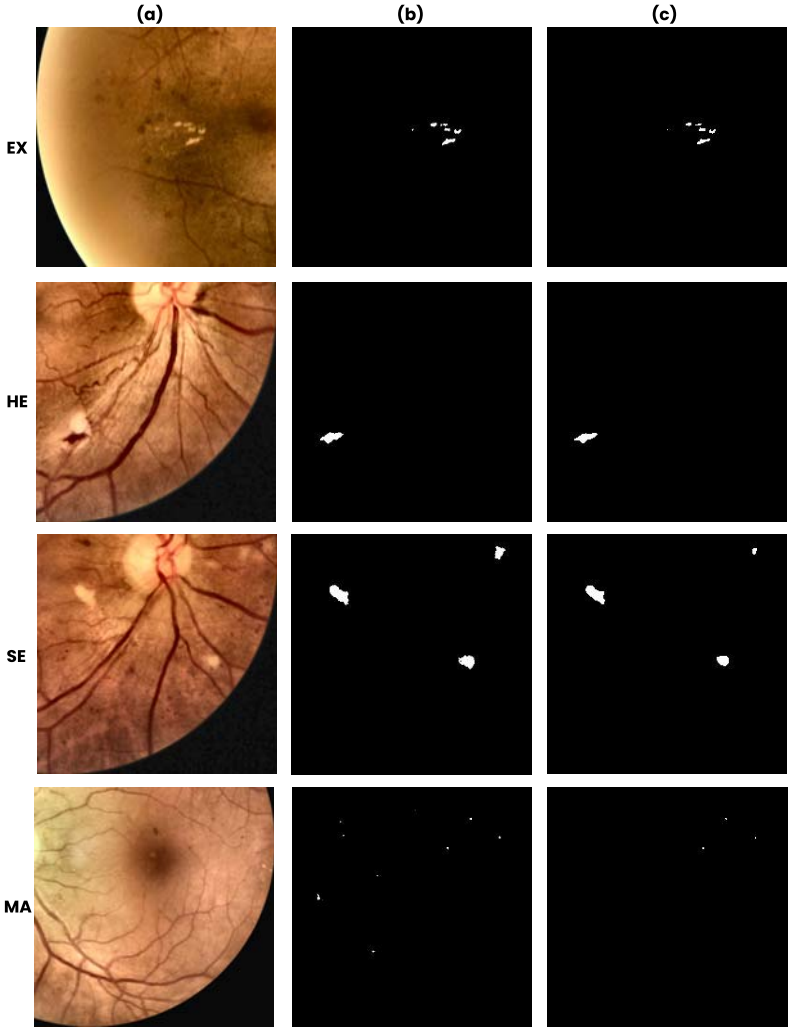
Fonte: Elaborada pelo autor.

A Figura 32 apresenta uma comparação entre as predições realizadas pela abordagem proposta e o *ground truth* (gt) das imagens do conjunto de teste do *dataset* DDR. Ao comparar as colunas (b) e (c), observa-se que, embora algumas lesões não tenham sido identificadas – principalmente no caso dos MA –, a abordagem proposta conseguiu segmentar com precisão a maioria das instâncias presentes nas imagens de fundo.

Além disso, a Figura 33 apresenta uma comparação visual entre a abordagem proposta e os modelos individuais que a compõem, no conjunto de teste do *dataset* DDR, evidenciando as melhorias obtidas com a estratégia de combinação.

Ao observar as predições realizadas, percebe-se que, embora os modelos utilizados individualmente consigam identificar parte das lesões, em alguns casos es-

Figura 32 – Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta e o *ground truth* em imagens do conjunto de teste do *dataset* DDR. (a) Imagens pré-processadas, (b) *Ground truth*, e (c) Abordagem proposta.



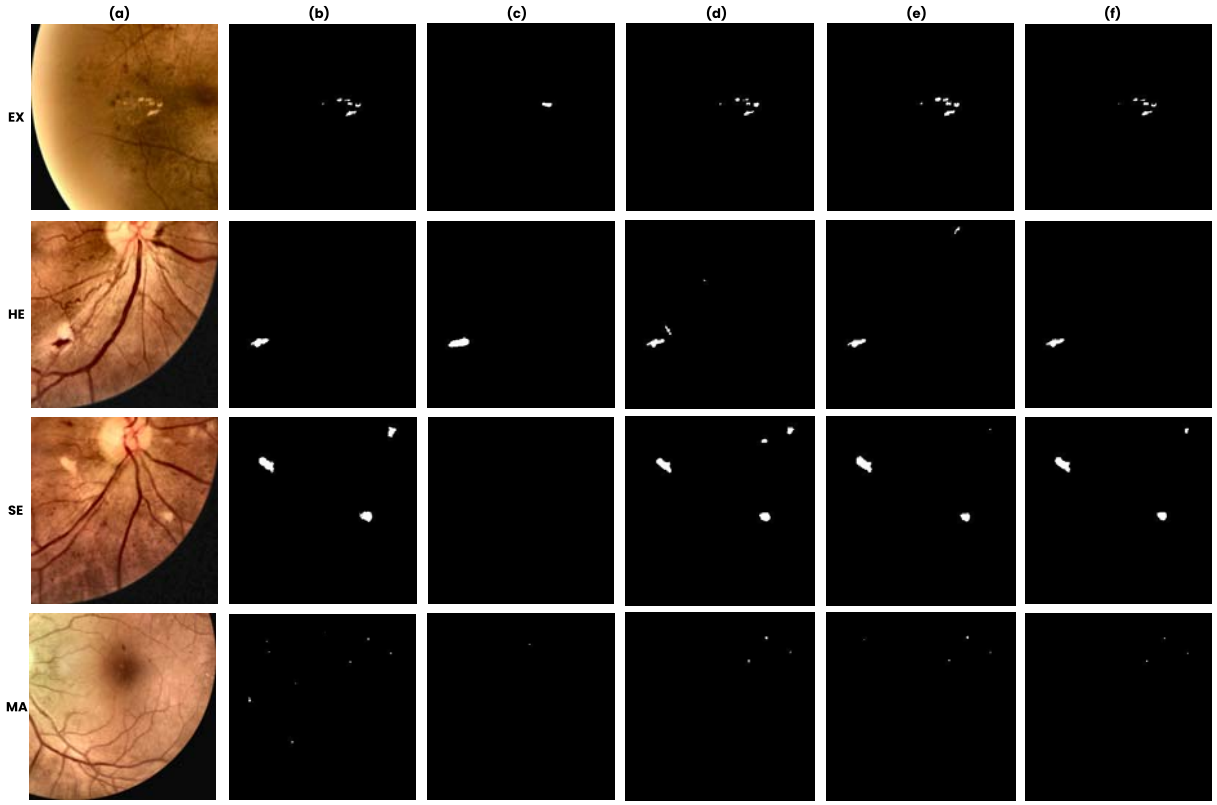
Fonte: Elaborada pelo autor.

pecíficos uma determinada arquitetura falha na segmentação de certas instâncias. No entanto, a abordagem proposta demonstra maior robustez ao integrar os resultados de múltiplos modelos, compensando eventuais erros individuais e resultando em uma segmentação mais precisa, conforme evidenciado pelos valores apresentados nas métricas anteriores.

Visando avaliar a capacidade de generalização da abordagem proposta em diferentes conjuntos de dados de RD, foram realizados experimentos adicionais utilizando o *dataset* IDRiD. Nessa etapa, os modelos previamente treinados no conjunto DDR, que compõem a abordagem proposta, foram empregados para realizar predições no conjunto IDRiD.

A Tabela 11 apresenta os resultados obtidos no conjunto de validação, com base nas métricas IoU e $mIoU$. A abordagem proposta alcançou um $mIoU$ de 0,8436.

Figura 33 – Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta com os modelos que a compõe em imagens do conjunto de teste do *dataset* DDR. (a) Imagens pré-processadas, (b) *Ground truth*, (c) U-Net++, (d) Attention U-net, (e) R2U-Net, e (f) Abordagem proposta.



Fonte: Elaborada pelo autor.

Além disso, foram obtidos valores de IoU iguais a 0,7829, 0,7093, 0,9745 e 0,9075 para as classes EX, HE, SE e MA, respectivamente.

Embora tenha apresentado desempenho ligeiramente inferior ao modelo U-Net++ com as etapas de pré-treinamento, aumento de dados e pré-processamento na segmentação de HE, o presente trabalho demonstrou resultados superiores nas demais classes e no desempenho médio.

No conjunto de teste do *dataset* IDRiD, também avaliando as métricas IoU e $mIoU$, conforme mostrado na Tabela 12, os resultados foram semelhantes aos obtidos no conjunto de validação. A abordagem proposta superou os demais modelos em todas as classes, com exceção da classe HE. Os valores de IoU alcançados foram de 0,7780, 0,6538, 0,9793 e 0,9024 para as classes EX, HE, SE e MA, respectivamente, e um $mIoU$ de 0,8284.

Observa-se que, no conjunto de teste, os valores de IoU foram ligeiramente inferiores aos da validação, mas ainda assim indicam a capacidade de generalização da abordagem proposta na segmentação de lesões de fundo associadas à RD.

Tabela 11 – Resultados obtidos na segmentação de lesões de fundo de olho comparados aos demais modelos utilizando as métricas IoU e $mIoU$ no conjunto de validação do *dataset* IDRiD.

Modelos	IoU				$mIoU$
	EX	**HE**	**SE**	**MA**	
U-Net (Ronneberger; Fischer; Brox, 2015)	0,4987	0,4999	0,4999	0,0000	0,3746
U-Net + aumento de dados	0,4984	0,4986	0,4994	0,4999	0,4991
U-Net + pré-treinamento	0,4975	0,4595	0,4994	0,4999	0,4890
U-Net + pré-treinamento + aumento de dados	0,4982	0,4987	0,4996	0,4998	0,4991
U-Net + pré-treinamento + aumento de dados + pré-processamento	0,6394	0,5870	0,9219	0,7662	0,7286
U-Net++ (Zhou et al., 2018)	0,0000	0,0000	0,0000	0,0000	0,0000
U-Net++ + aumento de dados	0,4989	0,4970	0,4987	0,0000	0,3736
U-Net++ + pré-treinamento	0,0000	0,0000	0,0000	0,0000	0,0000
U-Net++ + pré-treinamento + aumento de dados	0,4991	0,0000	0,4984	0,0000	0,2494
U-Net++ + pré-treinamento + aumento de dados + pré-processamento	0,7529	**0,7384**	0,9418	0,9047	0,8344
Attention U-net (Oktay et al., 2018)	0,4989	0,4999	0,4995	0,0000	0,3745
Attention U-net + aumento de dados	0,4985	0,4997	0,4977	0,4999	0,4989
Attention U-net + pré-treinamento	0,4978	0,4997	0,4996	0,4999	0,4992
Attention U-net + pré-treinamento + aumento de dados	0,4984	0,4994	0,4993	0,4999	0,4993
Attention U-net + pré-treinamento + aumento de dados + pré-processamento	0,6957	0,5707	0,9382	0,7934	0,7495
R2U-Net (Alom et al., 2018)	0,4963	0,4956	0,4994	0,4999	0,4978
R2U-Net + aumento de dados (Pereira et al., 2024)	0,4983	0,4993	0,4992	0,4998	0,4992
R2U-Net + pré-treinamento + aumento de dados + pré-processamento	0,7094	0,6423	0,9291	0,7716	0,7631
Abordagem proposta	**0,7829**	0,7093	**0,9745**	**0,9075**	**0,8436**

Fonte: Elaborada pelo autor.

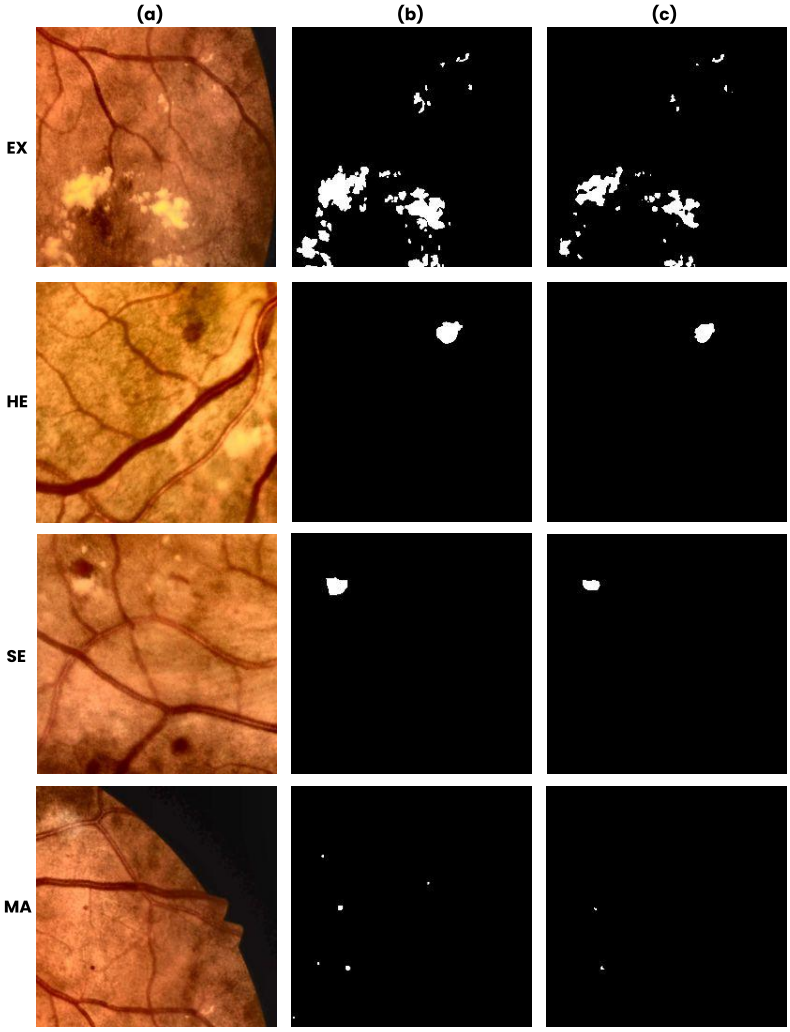
Tabela 12 – Resultados obtidos na segmentação de lesões de fundo de olho comparados aos demais modelos utilizando as métricas IoU e $mIoU$ no conjunto de teste do *dataset* IDRiD.

Modelos	IoU				$mIoU$
	EX	**HE**	**SE**	**MA**	
U-Net (Ronneberger; Fischer; Brox, 2015)	0,4986	0,4995	0,4999	0,0000	0,3745
U-Net + aumento de dados	0,4980	0,4969	0,4994	0,4999	0,4985
U-Net + pré-treinamento	0,4966	0,4608	0,4996	0,4999	0,4892
U-Net + pré-treinamento + aumento de dados	0,4970	0,4979	0,4994	0,4997	0,4985
U-Net + pré-treinamento + aumento de dados + pré-processamento	0,6811	0,5444	0,9160	0,7746	0,7290
U-Net++ (Zhou et al., 2018)	0,0000	0,0000	0,0000	0,0000	0,0000
U-Net++ + aumento de dados	0,4989	0,4970	0,4985	0,0000	0,3736
U-Net++ + pré-treinamento	0,0000	0,0000	0,0000	0,0000	0,0000
U-Net++ + pré-treinamento + aumento de dados	0,4947	0,0000	0,4999	0,0000	0,2487
U-Net++ + pré-treinamento + aumento de dados + pré-processamento	0,7274	**0,7154**	0,9339	0,9015	0,8195
Attention U-net (Oktay et al., 2018)	0,4988	0,4999	0,4998	0,0000	0,3746
Attention U-net + aumento de dados	0,4980	0,4991	0,4975	0,4999	0,4986
Attention U-net + pré-treinamento	0,4966	0,4994	0,4997	0,4999	0,4989
Attention U-net + pré-treinamento + aumento de dados	0,4978	0,4993	0,4995	0,4999	0,4991
Attention U-net + pré-treinamento + aumento de dados + pré-processamento	0,7021	0,5236	0,9215	0,7856	0,7332
R2U-Net (Alom et al., 2018)	0,4954	0,4933	0,4995	0,4999	0,4970
R2U-Net + aumento de dados (Pereira et al., 2024)	0,4978	0,4992	0,4990	0,4998	0,4990
R2U-Net + pré-treinamento + aumento de dados + pré-processamento	0,7202	0,6026	0,9160	0,7554	0,7486
HACDR-Net (Xu et al., 2024)	0,6461	0,4735	0,5621	0,2831	0,4912
U-Net++ + ResNet50 (*Backbone*) + AFMA (Jiang; Zhao, 2024)	0,6613	0,4768	0,4915	0,2837	0,4783
DenseCUNet (Zhang; Xu; Zhai, 2025)	0,6769	0,5055	0,5127	0,3241	0,5048
Abordagem proposta	**0,7780**	0,6538	**0,9793**	**0,9024**	**0,8284**

Fonte: Elaborada pelo autor.

A Figura 34 ilustra as predições realizadas pela abordagem proposta na segmentação das lesões EX, HE, SE e MA, desta vez no conjunto de teste do *dataset* IDRiD. A comparação entre as colunas (b) e (c) demonstra que a maioria das instâncias foi

Figura 34 – Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta e o *ground truth* em imagens do conjunto de teste do *dataset* IDRiD. (a) Imagens pré-processadas, (b) *Ground truth*, e (c) Abordagem proposta.



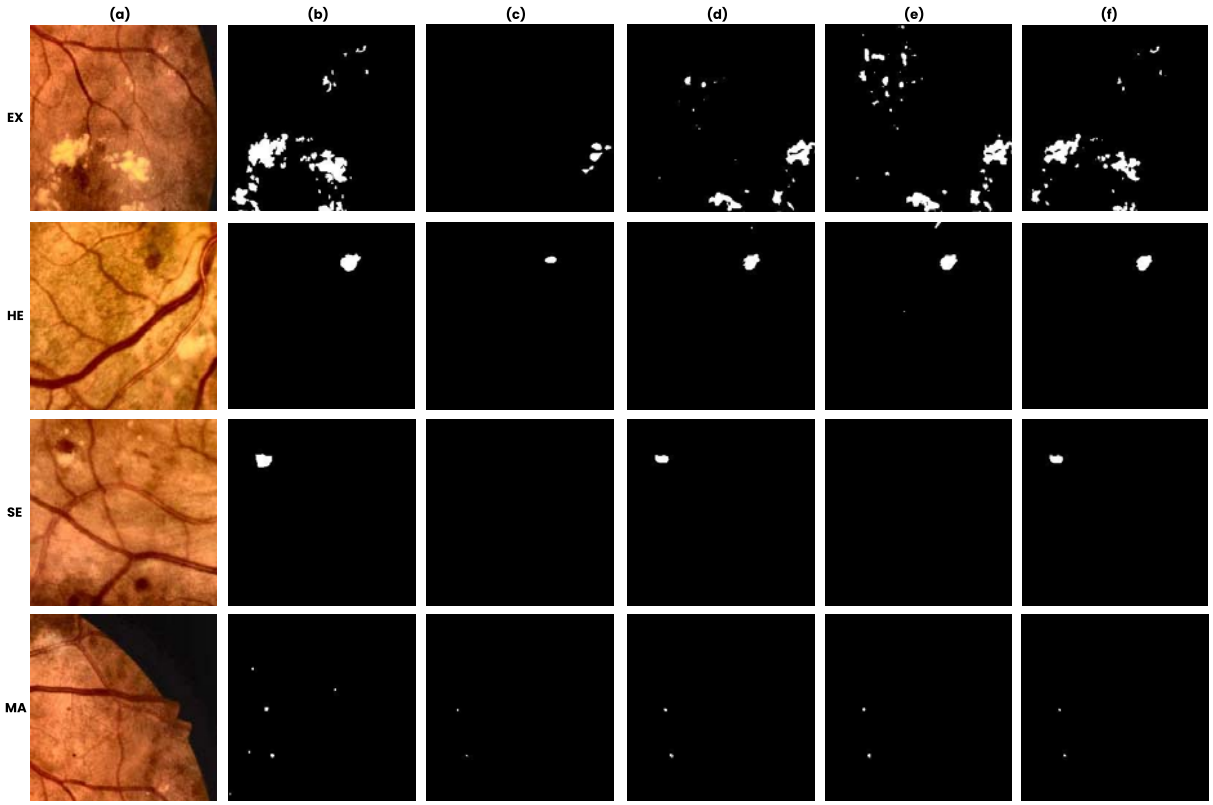
Fonte: Elaborada pelo autor.

corretamente identificada, o que está de acordo com os valores de IoU apresentados. Contudo, é notável que, em alguns casos, a segmentação não abrange completamente a área das lesões, principalmente nas instâncias de EX. Além disso, assim como observado no conjunto DDR, a abordagem proposta não conseguiu identificar algumas instâncias de MA.

Na Figura 35 é apresentada a comparação visual entre as segmentações realizadas pela abordagem proposta e aquelas produzidas individualmente pelos modelos U-Net++, Attention U-Net e R2U-Net, no conjunto de teste do *dataset* IDRiD. De forma análoga ao observado no conjunto DDR, nota-se que os modelos individuais, embora eficazes em diversos cenários, em algumas situações deixam de detectar certas instâncias de lesões ou as identificam incorretamente. A abordagem proposta, por outro lado, ao combinar os resultados dos diferentes modelos, compensa essas in-

consistências pontuais. Isso evidencia o potencial do uso de *ensembles* para produzir segmentações mais consistentes e próximas ao gt das lesões de fundo.

Figura 35 – Comparação visual das segmentações de instâncias de lesões de fundo realizadas pela abordagem proposta com os modelos que a compõe em imagens do conjunto de teste do *dataset* IDRiD. (a) Imagens pré-processadas, (b) *Ground truth*, (c) U-Net++, (d) Attention U-net, (e) R2U-Net, e (f) Abordagem proposta.



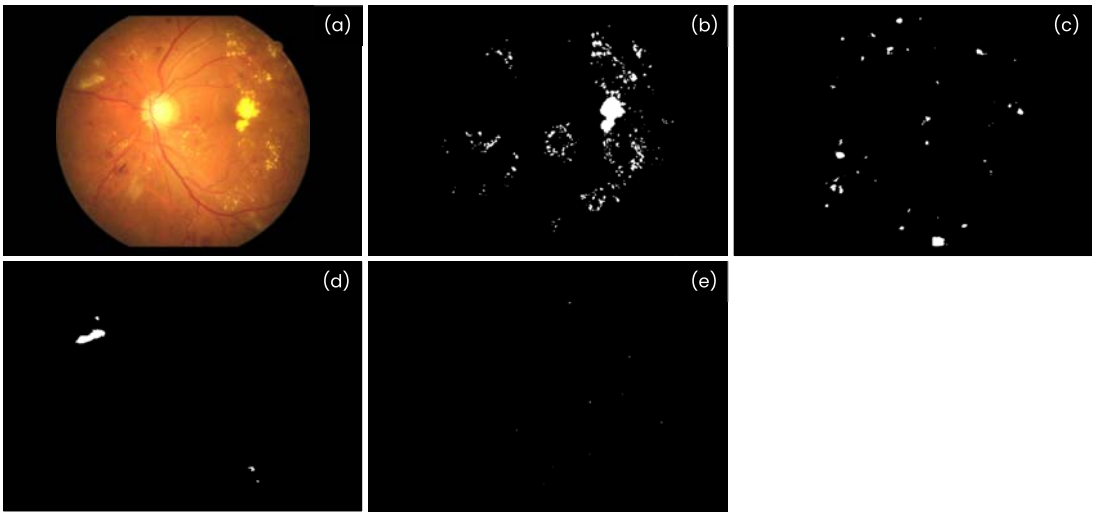
Fonte: Elaborada pelo autor.

Por fim, os modelos que compõem a abordagem proposta foram utilizados como etapa inicial no processo de anotação das lesões de fundo durante a criação de um novo conjunto de dados de RD, composto por imagens de fundo de olho coletadas entre 2012 e 2024 em uma clínica oftalmológica na cidade de Pelotas, no estado do Rio Grande do Sul, Brasil. Como resultado, foram geradas máscaras para as lesões do tipo EX, HE, SE e MA em 100 imagens, que posteriormente foram validadas por um médico especialista, com o objetivo de corrigir eventuais erros nas anotações.

As Figuras 36 e 37 ilustram, respectivamente, as anotações em nível de píxel e as anotações de caixas delimitadoras das lesões geradas pelos modelos que compõem a abordagem proposta em uma imagem de fundo do novo conjunto de RD. As principais correções realizadas pelo especialista durante o processo de validação envolveram a inclusão de lesões não identificadas automaticamente, com destaque para os MA, o que está em conformidade com as comparações visuais das segmentações apresentadas nas Figuras 32, 33, 34 e 35. Ainda assim, grande parte das lesões foi

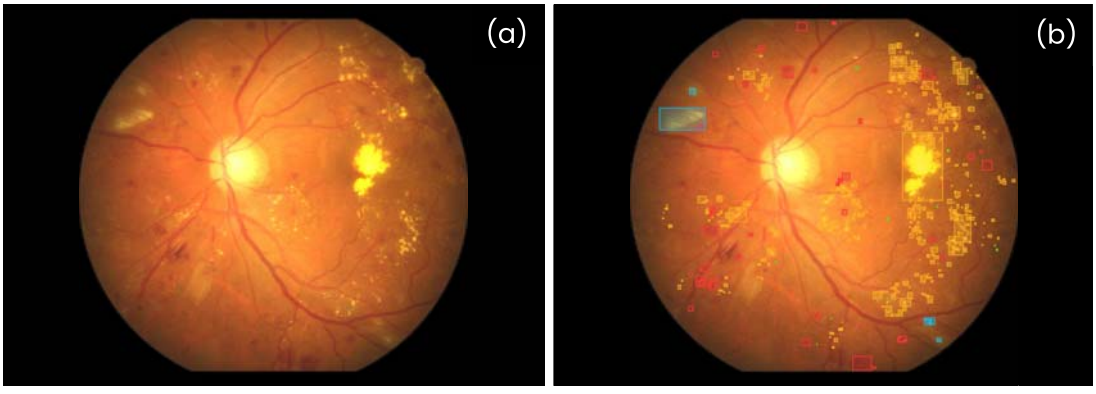
corretamente identificada de forma automatizada, o que reforça o desempenho dos modelos propostos neste trabalho.

Figura 36 – Exemplo de imagem de fundo do novo conjunto de dados de RD e as anotações em nível de píxel geradas pelos modelos que compõem a abordagem proposta. Em (a), a imagem de fundo original; (b) anotações a nível de píxel dos Exsudatos Duros; (c) anotações a nível de píxel das Hemorragias; (d) anotações a nível de píxel dos Exsudatos Algodonosos; e, (e) anotações a nível de píxel dos Microaneurismas.



Fonte: Elaborada pelo autor.

Figura 37 – Exemplo de imagem de fundo do novo conjunto de dados de RD com as anotações de caixa delimitadora das lesões geradas pelos modelos que compõem a abordagem proposta. Em (a), a imagem de fundo original; e, (b) a imagem com suas anotações de caixa delimitadora. As anotações em amarelo representam Exsudatos Duros; as anotações em vermelho representam Hemorragias; as anotações em azul representam Exsudatos Algodonosos; e, as anotações em verde representam Microaneurismas.



Fonte: Elaborada pelo autor.

4 CONCLUSÃO

Esta Dissertação apresentou uma nova abordagem para a segmentação semântica de lesões de fundo associadas à RD. Nos experimentos, foram utilizados dois conjuntos de dados públicos: o conjunto DDR, empregado no treinamento, ajuste e avaliação dos modelos que compõem a abordagem proposta; e o conjunto IDRiD, utilizado para avaliar a capacidade de generalização da abordagem em diferentes bases de dados de RD. Para a realização dos experimentos, ambos os conjuntos foram particionados utilizando o método de *Hold-out validation*, nas proporções de 50% para treinamento, 20% para validação e 30% para teste.

A abordagem proposta teve como base as arquiteturas U-Net++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018) e R2U-Net (Alom et al., 2018), integradas por meio de *ensembles*. Além disso, foram incorporadas etapas de pré-processamento, aumento de dados e pré-treinamento. Na etapa de validação com o conjunto DDR, os principais resultados obtidos pela abordagem proposta foram valores de DC iguais a 76,19%, 60,51%, 76,63% e 41,67%, para EX, HE, SE e MA, respectivamente, com mDC de 63,75%. Os respectivos valores de IoU foram de 89,04%, 79,75%, 87,78% e 76,15%, com $mIoU$ de 83,18%.

Além disso, na etapa de validação com o conjunto de dados IDRiD, foi obtido um $mIoU$ de 84,36%, com valores de IoU iguais a 78,29%, 70,93%, 97,45% e 90,75% para EX, HE, SE e MA, respectivamente, evidenciando a capacidade de generalização da abordagem proposta. Ademais, na etapa de teste, os resultados obtidos foram superiores aos apresentados nos trabalhos relacionados do estado da arte (Pereira et al., 2024; Xu et al., 2024; Jiang; Zhao, 2024; Zhang; Xu; Zhai, 2025), com destaque para os valores de mDC = 66,47% e $mIoU$ = 81,15% no *dataset* DDR, bem como o $mIoU$ de 82,84% no *dataset* IDRiD.

Os resultados indicam um grande potencial no uso de modelos arquiteturais do tipo *encoder-decoder* combinados em forma de *ensembles* para a segmentação de lesões de fundo. Além disso, a aplicação das etapas de pré-processamento, aumento de dados e pré-treinamento contribuiu para uma extração mais eficaz das características das lesões, impactando diretamente no aumento da precisão da segmentação – es-

pecialmente com o uso do *framework* SAHI, que minimizou os desafios associados à identificação de microlesões. Esses resultados reforçam a viabilidade da utilização da abordagem proposta em cenários reais, como no Sistema Único de Saúde (SUS), particularmente no apoio à triagem automatizada da RD. Dessa forma, este trabalho pode representar uma ferramenta valiosa para auxiliar no diagnóstico médico, contribuindo para a identificação precoce de pacientes com a doença.

Adicionalmente, a abordagem proposta foi utilizada como apoio no processo de anotação de lesões de fundo em um novo conjunto de dados brasileiro de RD. As máscaras geradas serviram como referência inicial para a marcação das lesões, sendo posteriormente revisadas por um especialista. Embora grande parte das lesões tenha sido corretamente identificada, foram necessárias correções pontuais – especialmente relacionadas aos Microaneurismas não identificados. Esses resultados reforçam o potencial da abordagem para facilitar o processo de anotação de imagens, além de evidenciar sua aplicabilidade prática na criação de novos conjuntos de dados de RD.

No entanto, é necessário aprimorar a precisão da abordagem, especialmente no que se refere ao diagnóstico precoce da doença, por meio da identificação dos MA. Nesse sentido, como parte dos trabalhos futuros, pretende-se aplicar novas técnicas de aumento de dados e pré-processamento, bem como explorar arquiteturas mais avançadas, com ênfase em modelos baseados em *transformers* voltados ao processamento multimodal.

Além disso, busca-se implementar modificações nas arquiteturas utilizadas, incorporando mecanismos de atenção com o objetivo de potencializar a extração de características associadas às microlesões. Ademais, pretende-se utilizar o *framework* SAHI com blocos do mesmo tamanho da entrada das redes, visando evitar a perda de informação causada pela redução do tamanho das imagens.

Outro objetivo consiste na integração da abordagem proposta em um sistema funcional, que permita o envio de imagens de fundo e apresente como saída a segmentação correspondente às lesões identificadas. Por fim, planeja-se disponibilizar publicamente a ferramenta desenvolvida, com o intuito de apoiar profissionais da saúde no processo de identificação das lesões de fundo e no diagnóstico da RD.

A abordagem proposta para a segmentação das lesões de fundo associadas à Retinopatia Diabética, desenvolvida ao longo desta pesquisa de mestrado, foi submetida para publicação em periódicos e conferências da área.

Artigos publicados:

Título: An Improved Approach for Semantic Segmentation of Fundus Lesions using R2U-Net
Autores: Alejandro Pereira, Carlos Santos, Marilton Aguiar, Daniel Welfer, Marcelo Dias, Rafaela de Menezes, Paulo Roberto Ferreira, Fábio Rossi, Marcos d'Ornellas, Carlos Jesus Haygert, Juliano Kazienko, Fábio Comim, Aurélio Hoppe, Juliano Weber, Leandro Tavares, Leonardo Guths, Rodrigo Guerra
Conferência: Engineering in Medicine and Biology Society (EMBC)
Ano: 2024
DOI: http://dx.doi.org/10.1109/EMBC53108.2024.10782763
Qualis: A1

Título: Enhanced Semantic Segmentation of Retinal Microlesions through R2U-Net Architecture
Autores: Alejandro Pereira, Carlos Santos, Marilton Aguiar, Daniel Welfer, Marcelo Dias, Rafaela de Menezes, Douglas Santana
Conferência: Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)
Ano: 2024
DOI: http://dx.doi.org/10.5753/sbcas.2024.1737
Qualis: A4

Artigos submetidos:

Título: Enhancing Retinal Lesion Segmentation in Diabetic Retinopathy Using a U-Net++-Based Approach Inference
Autores: Alejandro Pereira, Carlos Santos, Marilton Aguiar, Daniel Welfer, Marcelo Dias
Periódico: IEEE Access
Ano: 2025
Qualis: A3

Título: Deep Neural Networks for Classification, Segmentation and Lesion Detection in a New Brazilian Diabetic Retinopathy Image Database
Autores: Carlos Santos, Alejandro Pereira, Marcelo Dias, Marilton Aguiar, Iasmin Dembinski, João Zanata, Daniel Welfer
Periódico: IEEE Access
Ano: 2025
Qualis: A3

REFERÊNCIAS

ABEDALLA, A. et al. 2ST-UNet: 2-Stage Training Model using U-Net for Pneumothorax Segmentation in Chest X-Rays. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2020., 2020, Glasgow, UK. **Proceedings...** IEEE, 2020. p.1–6.

ACHARYA, A. **Instance Segmentation in Computer Vision: A Comprehensive Guide**. 2024. Acesso em: 23 fev. 2025. Disponível em: <https://encord.com/blog/instance-segmentation-guide-computer-vision/>.

AGHELAN, A.; ROUHANI, M. Fine-Tuned Generative Adversarial Network-Based Model for Medical Image Super-Resolution. In: INTERNATIONAL CONFERENCE ON COMPUTER AND KNOWLEDGE ENGINEERING (ICCKE), 2024., 2024, Mashhad, Iran. **Proceedings...** IEEE, 2024. p.174–181.

AKYON, F. C.; ALTINUC, S. O.; TEMIZEL, A. Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP), 2022, Bordeaux, France. **Proceedings...** IEEE, 2022.

ALOM, M. Z. et al. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. **CoRR**, Ithaca, NY, USA, v.abs/1802.06955, 2018.

ALYOUBI, W. L.; ABULKHAIR, M. F.; SHALASH, W. M. Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning. **Sensors**, Basel, Switzerland, v.21, n.11, 2021.

ANAND, M. et al. **Channel and Spatial Attention Aware UNet Architecture for Segmentation of Blood Vessels, Exudates and Microaneurysms in Diabetic Retinopathy**. 2023. Durham, NC: Research Square, 2023. Disponível em: <https://doi.org/10.21203/rs.3.rs-3323155/v1>.

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. **IEEE Transactions on Pat-**

tern Analysis and Machine Intelligence, Piscataway, NJ, USA, v.39, n.12, p.2481–2495, 2017.

BASHA, S. S.; PRASAD, K. S. Automatic detection of hard exudates in diabetic retinopathy using morphological segmentation and fuzzy logic. **International Journal of Computer Science and Network Security**, Seoul, Korea, v.8, n.12, p.211–218, 2008.

BUSLAEV, A. et al. Albumentations: Fast and flexible image augmentations. **Information**, Basel, Switzerland, v.11, n.2, 2020.

CARDOSO, C. d. F. d. S. **Segmentação automática do disco óptico e de vasos sanguíneos em imagens de fundo de olho**. 2019. 149p. Tese de Doutorado (Curso de Doutorado em Ciências) — Universidade Federal de Uberlândia, Uberlândia, MG, Brasil.

CHARLEAUX, L.; TOLEDO, V. **Machine Learning**: o que é, como funciona e quais são os tipos de aprendizado de máquina. 2025. Acesso em: 20 mar. 2025. Disponível em: <https://tecnoblog.net/responde/machine-learning-o-que-e-como-funciona-e-quais-sao-os-tipos-de-aprendizado-de-maquina/>.

CHEN, L.-C. et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2018, Munich, Germany. **Proceedings...** Springer, 2018. p.801–818.

CHEN, R.-C. et al. Contrast Limited Adaptive Histogram Equalization for Recognizing Road Marking at Night Based on Yolo Models. **IEEE Access**, Piscataway, NJ, USA, v.11, p.92926–92942, 2023.

CONDE, A. L. d. S.; SANTOS, H. G. d.; SANTOS, F. S. d. GEORREFERENCIAMENTO DE CARREADORES: ABORDAGEM COM MICROSSERVIÇOS, TRANSFORMADA DE HOUGH E OTIMIZAÇÃO DE MENOR VALOR ORDENADO. **Revista Contemporânea**, São José dos Pinhais, PR, Brasil, v.5, n.4, p.e7848, abr. 2025.

COSTA, A. E. M. F. **Detecção automática de exsudatos e microaneurismas em imagens de retinopatia diabética utilizando o algoritmo bioinspirado na colônia artificial de abelhas e processamento digital de imagens**. 2019. 71p. Dissertação de Mestrado (Engenharia Elétrica) — Instituto Federal da Paraíba, João Pessoa, PB, Brasil.

D. D. Silva, A.; B. P. Carneiro, M.; F. S. Cardoso, C. Realce De Microaneurimas Em Imagens De Fundo De Olho Utilizando Clahe. In: V CONGRESSO BRASILEIRO DE ELETROMIOGRAFIA E CINESIOLOGIA E X SIMPóSIO DE ENGENHARIA BIOMéDICA, 2018, Uberlândia, MG, Brasil. **Anais...** Even3, 2018. p.772–775.

DANKO, T. et al. Small Object Detection in Fundus Images. In: INTERNATIONAL SYMPOSIUM ELMAR, 2024., 2024, Zadar, Croatia. **Proceedings...** IEEE, 2024. p.105–108.

DIAS, M. et al. Um novo método baseado em detector de dois estágios para segmentação de instância de lesões retinianas usando o modelo Mask R-CNN e a biblioteca Detectron2. In: L SEMINáRIO INTEGRADO DE SOFTWARE E HARDWARE, 2023, Porto Alegre, RS, Brasil. **Anais...** SBC, 2023. p.1–12.

FLAXEL, C. J. et al. Diabetic retinopathy preferred practice pattern®. **Ophthalmology**, Amsterdam, Netherlands, v.127, n.1, p.P66–P145, 2020.

FRANCELINO, G. A. R. et al. Retinopatia Diabética-um relato de caso. **Revista Pró-UniverSUS**, Vassouras, RJ, Brasil, v.14, n.3, p.38–41, 2023.

GOODFELLOW, I. et al. **Deep learning**. Cambridge, MA, USA: MIT press Cambridge, 2016. v.1, n.2.

GUO, F. et al. MES-Net: a new network for retinal image segmentation. **Multimedia Tools and Applications**, New York, NY, USA, v.80, n.10, p.14767–14788, 2021.

HARRIS, C. R. et al. Array programming with NumPy. **Nature**, New York, NY, USA, v.585, n.7825, p.357–362, Sept. 2020.

HE, K. et al. Deep Residual Learning for Image Recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016., 2016, Las Vegas, NV, USA. **Proceedings...** IEEE, 2016. p.770–778.

HUANG, G. et al. Densely Connected Convolutional Networks. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2017., 2017, Honolulu, HI, USA. **Proceedings...** IEEE, 2017. p.2261–2269.

HUO, F. et al. Research on cuttings image segmentation method based on improved MultiRes-Unet++ with attention mechanism. **Signal, Image and Video Processing**, New York, NY, USA, v.18, n.Suppl 1, p.799–808, 2024.

INGOLE, S. **An Extensive Guide to Convolution Neural Network (2023)**. 2023. Acesso em: 27 abr. 2025. Disponível em: <https://medium.com/@singole/an-extensive-guide-to-convolution-neural-network-2023-84872b16bd78>.

International Council of Ophthalmology. Updated 2017 ICO Guidelines for Diabetic Eye Care. **ICO Guidelines for Diabetic Eye Care**, Brussels, Belgium, p.1–33, 2017.

JIANG, H.; ZHAO, J. Multi-lesion segmentation of fundus images using improved unet+. **IAENG International Journal of Computer Science**, Hong Kong, v.51, n.10, 2024.

KASHIPAREKH, K. et al. ConvTimeNet: A Pre-trained Deep Convolutional Neural Network for Time Series Classification. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019., 2019, Budapest, Hungary. **Proceedings...** IEEE, 2019. p.1–8.

KOTRA, N. **Image Classification vs Image Segmentation vs Object Detection**. 2023. Acesso em: 22 fev. 2025. Disponível em: <https://www.kaggle.com/discussions/getting-started/463662>.

KUMAR, A. et al. Prevalence of diabetes in India: A review of IDF diabetes atlas 10th edition. **Current diabetes reviews**, Sharjah, UAE, v.20, n.1, p.105–114, 2024.

KURAN, U.; KURAN, E. C. Parameter selection for CLAHE using multi-objective cuckoo search algorithm for image contrast enhancement. **Intelligent Systems with Applications**, Amsterdam, Netherlands, v.12, p.200051, 2021.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, London, UK, v.521, n.7553, p.436–444, 2015.

LEE, L. C.; LIONG, C.-Y.; JEMAIN, A. A. Validity of the best practice in splitting data for hold-out validation strategy as performed on the ink strokes in the context of forensic science. **Microchemical Journal**, Amsterdam, Netherlands, v.139, p.125–133, 2018.

LI, T. et al. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. **Information Sciences**, Amsterdam, Netherlands, v.501, p.511–522, 2019.

LI, Z. et al. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. **IEEE Transactions on Neural Networks and Learning Systems**, Piscataway, NJ, USA, v.33, n.12, p.6999–7019, 2022.

LIANG, M.; HU, X. Recurrent convolutional neural network for object recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2015., 2015, Boston, MA, USA. **Proceedings...** IEEE, 2015. p.3367–3375.

LIU, X. et al. Plant leaf veins coupling feature representation and measurement method based on DeepLabV3+. **Frontiers in Plant Science**, Lausanne, Switzerland, v.13, p.1043884, 2022.

MASSUCATTO, J. D. P. **Aplicação de conceitos de redes neurais convolucionais na classificação de imagens de folhas**. 2018. 68p. Trabalho de Conclusão (Curso

de Engenharia de Computação) — Universidade Federal do Paraná, Pato Branco, PR, Brasil.

MENDES, C. C. **Classificação de imagens de retina para auxílio a diagnóstico de retinopatia diabética via detecção de exsudatos duros, com base em classificadores robustos, máquinas de vetores de suporte e algoritmos com treinamento de dicionários**. 2017. Dissertação de Mestrado (Engenharia Biomédica) — Universidade de Brasília, Brasília, DF, Brasil.

MILLON, R.; FRATI, F. E.; RUCCI, E. Implementación de Filtro de Detección de Bordes Sobel en SoC usando Síntesis de Alto Nivel. In: CONGRESO ARGENTINO DE SISTEMAS EMBEBIDOS (CASE), 2020, Buenos Aires, Argentina. **Anais...** ACSE, 2020. p.73–75.

MINAEE, S. et al. Image Segmentation Using Deep Learning: A Survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, NJ, USA, v.44, n.7, p.3523–3542, 2022.

MISHIBA, K. Fast Guided Median Filter. **IEEE Transactions on Image Processing**, Piscataway, NJ, USA, v.32, p.737–749, 2023.

NAIR, A. T. et al. Segmentation of Retinal Images Using Improved Segmentation Network, MesU-Net. **International Journal of Online and Biomedical Engineering (iJOE)**, Fairfax, VA, EUA, v.19, n.15, p.pp. 77–91, Oct. 2023.

NANDAL, P. et al. Super-Resolution of Medical Images Using Real ESRGAN. **IEEE Access**, Piscataway, NJ, USA, v.12, p.176155–176170, 2024.

OKTAY, O. et al. Attention U-Net: Learning Where to Look for the Pancreas. **CoRR**, Ithaca, NY, USA, v.abs/1804.03999, 2018.

OLIVEIRA, R.; ALVES, E.; MALQUI, C. Redes Neurais Convolucionais Aplicadas à Preensão Robótica. In: XIII CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL, 2017, Curitiba, PR, Brasil. **Anais...** ABRICOM, 2017. p.1–11.

PACHECO, A. G. C. Classificação de espécies de peixe utilizando redes neurais convolucional. **CoRR**, Ithaca, NY, USA, v.abs/1905.03642, 2019.

PARLOS, A. et al. An accelerated learning algorithm for multilayer perceptron networks. **IEEE Transactions on Neural Networks**, Piscataway, NJ, USA, v.5, n.3, p.493–497, 1994.

PEDRO, D. C. d. S. **Avaliação do impacto da exposição às fontes de luz artificiais na retina**: revisão sistemática da literatura. 2022. 40p. Dissertação de Mestrado (Segurança e Higiene do Trabalho) — Instituto Politécnico de Lisboa, Escola Superior de Tecnologia da Saúde de Lisboa, Lisboa, Portugal.

PEREIRA, A. et al. Detection of retinal microlesions through YOLOR-CSP architecture and image slicing with the SAHI algorithm. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2023., 2023, Queensland, Australia. **Proceedings...** IEEE, 2023. p.1–8.

PEREIRA, A. et al. Enhanced Semantic Segmentation of Retinal Microlesions through R2U-Net Architecture. In: XXIV SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE, 2024, Porto Alegre, RS, Brasil. **Anais...** SBC, 2024. p.13–24.

PEREIRA, J. A. et al. Atualizações sobre retinopatia diabética: uma revisão narrativa. **Revista Eletrônica Acervo Saúde**, Ouro Fino, MG, Brasil, n.49, p.e3428–e3428, 2020.

PORWAL, P. et al. IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge. **Medical Image Analysis**, Amsterdam, Netherlands, v.59, 2020.

RASHID, S. I.; SHAKIBAPOUR, E.; EBRAHIMI, M. Single MR image super-resolution using generative adversarial network. **CoRR**, Ithaca, NY, USA, v.abs/2207.08036, 2022.

RAUDONIS, V. et al. Automatic detection of microaneurysms in fundus images using an ensemble-based segmentation method. **Sensors**, Basel, Switzerland, v.23, n.7, p.3431, 2023.

ROCHA, R. d. L. **Redes neurais convolucionais aplicadas à inspeção de componentes do vagão ferroviário**. 2020. 78p. Dissertação de Mestrado (Computação Aplicada) — Universidade Federal do Pará, Tucuruí, PA, Brasil.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION – MICCAI 2015, 2015, Munich, Germany. **Proceedings...** Springer, 2015. p.234–241.

RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. **International journal of computer vision**, New York, NY, USA, v.115, p.211–252, 2015.

SANTOS, C. A. S. d. **Desenvolvimento de Abordagens Baseadas em Redes Neurais Profundas para Detecção e Segmentação de Instância de Lesões Retinianas**.

2023. 216p. Tese de Doutorado (Curso de Doutorado em Ciência da Computação) — Universidade Federal de Pelotas, Pelotas, RS, Brasil.

SHEEBA, T.; RAJ, S. A. A. Analysis of noise removal techniques on retinal optical coherence tomography images. **International Journal of Advanced Computer Science and Applications**, West Yorkshire, UK, v.13, n.9, 2022.

SHIM, J.-H. et al. Evaluation of U-Net models in automated cervical spine and cranial bone segmentation using X-ray images for traumatic atlanto-occipital dislocation diagnosis. **Scientific Reports**, London, UK, v.12, n.1, p.21438, 2022.

SILVA, R. E. V. d. **Um estudo comparativo entre redes neurais convolucionais para a classificação de imagens**. 2018. 51p. Trabalho de Conclusão (Curso de Sistemas de Informação) — Universidade Federal do Ceará, Quixadá, CE, Brasil.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **CoRR**, Ithaca, NY, USA, v.abs/1409.1556, 2014.

SIVAPRASAD, S.; SEN, S.; CUNHA-VAZ, J. Perspectives of diabetic retinopathy—challenges and opportunities. **Eye**, London, UK, v.37, n.11, p.2183–2191, 2023.

SOARES, I.; CASTELO-BRANCO, M.; PINHEIRO, A. Microaneurysms detection in retinal images using a multi-scale approach. **Biomedical Signal Processing and Control**, Amsterdam, Netherlands, v.79, p.104184, 2023.

SOUZA, V. B. d. **Segmentação de microaneurismas em imagens de retina usando aprendizado de máquina**. 2023. 44p. Trabalho de Conclusão (Curso de Sistemas de Informação) — Instituto Federal Goiano, Urutaí, GO, Brasil.

STEINHEISER, P. W. **Estudo do desempenho da detecção de objetos em exames de AgNOR**. 2020. 78p. Trabalho de Conclusão (Curso de Sistemas de Informação) — Universidade Federal do Paraná, Florianópolis, SC, Brasil.

STORN, R.; PRICE, K. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. **Journal of global optimization**, New York, NY, USA, v.11, p.341–359, 1997.

TAVARES, A. R. et al. Visão computacional na saúde: revisão de métodos e desafios educacionais para integração multidisciplinar. **Cuadernos de Educación y Desarrollo**, Paiva Castle, Portugal, v.16, n.13, p.e7088, dez. 2024.

VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: OF THE XXIX CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 2016, São José dos Campos, SP, Brasil. **Proceedings...** IEEE, 2016. v.1, n.4.

WANG, H. et al. MCA-UNet: multi-scale cross co-attentional U-Net for automatic medical image segmentation. **Health Information Science and Systems**, New York, NY, USA, v.11, n.1, p.10, 2023.

WANG, X. et al. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In: IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION WORKSHOPS (ICCVW), 2021., 2021, Montreal, BC, Canada. **Proceedings...** IEEE, 2021. p.1905–1914.

XIE, S.; TU, Z. Holistically-Nested Edge Detection. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 2015, Santiago, Chile. **Proceedings...** IEEE, 2015.

XU, Q. et al. HACDR-Net: Heterogeneous-Aware Convolutional Network for Diabetic Retinopathy Multi-Lesion Segmentation. In: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2024, Washington, DC, EUA. **Proceedings...** AAAI Press, 2024. v.38, n.6, p.6342–6350.

YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. **Insights into imaging**, New York, NY, USA, v.9, n.4, p.611–629, 2018.

YU, W. et al. Diabetic Retinopathy and Cardiovascular Disease: A Literature Review. **Diabetes, Metabolic Syndrome and Obesity**, Macclesfield, UK, v.16, p.4247–4261, 2023. PMID: 38164419.

ZHANG, Y.; XU, Y.; ZHAI, J. Diabetic Retinopathy Image Segmentation Method Based on Fusion DenseNet and U-Net Network. **Engineering Letters**, Hong Kong, v.33, n.2, 2025.

ZHOU, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In: DEEP LEARNING IN MEDICAL IMAGE ANALYSIS AND MULTIMODAL LEARNING FOR CLINICAL DECISION SUPPORT, 2018, Granada, Spain. **Proceedings...** Springer International Publishing, 2018. p.3–11.

Anexos

ANEXO A – Classificação da Retinopatia Diabética e Edema Macular

O Conselho Internacional de Oftalmologia (em inglês, *International Council of Ophthalmology* – ICO) criou diretrizes específicas para o cuidado com os olhos de pessoas com diabetes, com foco principal na prevenção e no tratamento da RD. Essas orientações ajudam profissionais da área da saúde a saber quando examinar os pacientes, como acompanhar os casos e quais tratamentos considerar, dependendo da situação. Um ponto importante do material é a importância do trabalho conjunto entre oftalmologistas e outros profissionais de saúde, como clínicos e endocrinologistas. As diretrizes foram pensadas para serem úteis em diferentes contextos, inclusive em locais com poucos recursos, e passam por atualizações periódicas para acompanhar os avanços na área. De acordo com o ICO, os estágios da RD podem ser classificados com base na Classificação Internacional de Retinopatia Diabética, conforme apresentado na Tabela 13 (International Council of Ophthalmology, 2017).

Tabela 13 – Classificação Internacional da Retinopatia Diabética (a) e do Edema Macular Diabético (b) com base na oftalmoscopia com dilatação.

Retinopatia Diabética	**Sinais clínicos na oftalmoscopia dilatada**
Sem RD	Nenhuma anormalidade
RDNP leve	Apenas Microaneurismas
RDNP moderada	Microaneurismas e outros sinais (ex.: Hemorragias, Exsudatos Duros e Exsudatos Algodonosos), mas menos que na forma grave
RDNP grave	RDNP moderada com qualquer um dos seguintes: • $\geq$ 20 Hemorragias intrarretinianas em cada quadrante; • Dilatação venosa proeminente em 2 quadrantes; • Anormalidades microvasculares intrarretinianas em 1 quadrante; • Sem sinais de RDP.
RDP	RDNP grave com um ou mais dos seguintes: • Neovascularização; • Hemorragia vítrea ou pré-retiniana.

(a)

Edema Macular Diabético	**Sinais clínicos na oftalmoscopia dilatada**
Sem EMD	Nenhum espessamento retiniano ou Exsudato Duro na mácula
EMD não central	Espessamento retiniano na mácula que não envolve a zona central de 1mm
EMD central	Espessamento retiniano na mácula que envolve a zona central de 1mm

(b)

Fonte: International Council of Ophthalmology (2017).

ANEXO B – Características Clínicas da Retinopatia Diabética

A RD apresenta uma variedade de sinais clínicos que refletem a complexidade da fisiopatologia da doença. Esses sinais podem ser observados por meio da oftalmoscopia, e seu reconhecimento adequado é essencial para o diagnóstico e o acompanhamento da progressão da doença. A Tabela 14 apresenta uma descrição detalhada das principais características da RD, com observações relevantes para sua avaliação clínica (International Council of Ophthalmology, 2017).

Tabela 14 – Características Clínicas da Retinopatia Diabética.

Característica	**Descrição**	**Considerações para Avaliação**
Microaneurismas	Pontos vermelhos esféricos isolados de tamanho variável. Podem representar uma tentativa falha de formação de novos vasos ou fraqueza da parede capilar.	Mais facilmente visualizados por angiografia com fluoresceína.
Hemorragias puntiformes	Semelhantes aos Microaneurismas, mas com tamanhos variados.	O termo Hemorragia de Ponto/Microaneurisma (H/Ma) é normalmente usado.
Hemorragias de mancha	Formadas onde aglomerados de capilares obstruem, levando à formação de Hemorragias de mancha intrarretinianas.	A lesão pode ser visualizada na camada plexiforme externa na angiografia com fluoresceína, onde não oculta o leito capilar subjacente, ao contrário das Hemorragias puntiformes e em chama, que estão localizadas mais superficialmente na retina.
Pontos algodonosos	Representam as extremidades inchadas de axônios interrompidos, onde ocorre o acúmulo de fluxo axoplasmático na borda do infarto	Essas características não são exclusivas de RD e não parecem aumentar, por si mesmas, o risco de formação de novos vasos. Por exemplo, podem ocorrer em hipertensão de HIV/AIDS.
Anomalias Microvasculares Intrarretinianas	São remanescente capilares dilatados seguindo fechamentos extensos de redes capilares entre arteríola e vênula. Características relacionadas incluem: • Perolização venosa (foco de proliferação de célula endotelial venosa que falhou em se desenvolver em novos vasos); • Reduplicação venosa (rara); • Circuitos venosos (desenvolvem-se devido a pequenas oclusões vasculares e abertura de circulação alternativa); • Palidez da retina e vasos brancos.	São os mais fáceis de serem observados em angiografia fluorescente.
Alterações maculares na RDNP – Edema macular – Doença macrovascular	Espessamento da retina acontece devido à acumulação de fluidos de exsudatos originados na barreira sangue-retina exterior danificada (edema extracelular) ou como resultado de hipoxia, levando à acumulação de fluido dentro de células retinianas individuais (edema intracelular). Pode ser focal ou difuso. Hemorragia em chama e formação de ponto algodonoso. Pode ocorrer devido a oclusão arteriolar, sem oclusão capilar, o que frequentemente afeta a camada horizontal de fibras nervosas da retina.	A aparência de edema macular pode ser observada em exame estereoscópico ou inferida pela presença de Exsudato Intrarretiniano.
Alterações no disco óptico	Discos ópticos inchados podem, ocasionalmente, ser vistos (papilopatia diabética) em pacientes diabéticos.	Em papilopatia diabética, a visão, geralmente, não é significativamente prejudicada.

Fonte: International Council of Ophthalmology (2017).