

# Wavelet Skip U-Net: Realce de Alta Frequência para Segmentação de Lesões de Retinopatia Diabética

Lucas Araújo Gonçalves<sup>1</sup>, Geraldo Braz Junior<sup>1</sup>, João Dallyson Sousa de Almeida<sup>1</sup>,  
Marcos Melo Ferreira<sup>1</sup>

<sup>1</sup>Núcleo de Ciência Aplicada – Universidade Federal do Maranhão (UFMA)  
CEP 65.080-805 – São Luís – MA – Brasil

{gerald, araujo.lucas, jdallyson}@nca.ufma.br  
marcos.melo@ifma.edu.br

**Abstract.** *Diabetic Retinopathy (DR) is one of the leading causes of vision loss in diabetic patients, making the early detection of lesions such as exudates and hemorrhages essential. This work proposes a deep learning architecture based on U-Net, incorporating a pre-trained EfficientNet-B4 encoder and a Wavelet Enhancement Module integrated into the first skip connection. The proposed approach uses two-dimensional discrete wavelet decomposition (2D DWT—Haar) to capture edge and texture details, thereby facilitating the segmentation of small lesions. Experiments conducted on the IDRiD dataset demonstrated that including of the wavelet module improved the Dice coefficient for the hemorrhage class from 62,3% to 65,1% on the test set, outperforming classical architectures such as the standard U-Net and DeepLabV3+.*

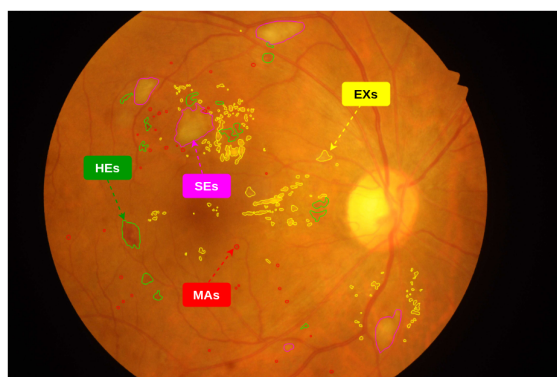
**Resumo.** *A retinopatia diabética (RD) é uma das principais causas de perda de visão em pacientes diabéticos, tornando essencial a detecção precoce de lesões, como exsudatos e hemorragias. Este trabalho propõe uma arquitetura de aprendizado profundo baseada em U-Net, que utiliza o encoder EfficientNet-B4 pré-treinado no ImageNet e um Módulo de Realce Wavelet na primeira skip connection. O objetivo é explorar a decomposição wavelet (DWT 2D - Haar) para capturar detalhes de bordas e texturas que facilitam a segmentação de pequenas lesões. Os experimentos realizados no dataset IDRiD demonstraram que a inclusão do módulo wavelet aumentou o coeficiente Dice de 62,3% para 65,1% na classe de hemorragia, no conjunto de teste, superando arquiteturas clássicas de segmentação (U-Net e DeepLabV3+)*

## 1. Introdução

A Retinopatia Diabética (RD) é uma patologia ocular de alta prevalência mundial, acometendo majoritariamente pacientes diagnosticados com *diabetes mellitus*. Caracterizada por danos microvasculares progressivos da retina, a RD representa uma das principais causas de cegueira irreversível quando não tratada em seus estágios iniciais [Chagas et al. 2023]. No cenário brasileiro, dados de 2024 indicam que 12,9% da população adulta nas capitais apresenta diagnóstico de diabetes, refletindo um crescimento de 135% em 18 anos [Ministério da Saúde 2025]. Com reflexo direto dessa expansão, estudos de 2023 identificaram uma prevalência de RD de 36,28% entre indivíduos com diabetes no Brasil, valor que supera a média global de 22,27% e representa cerca de

5,7 milhões de brasileiros afetados pela patologia [Chagas et al. 2023]. Tal cenário evidencia que o monitoramento contínuo e o diagnóstico preciso são fundamentais para a gestão da saúde ocular desses pacientes [Chagas et al. 2023].

As manifestações patológicas da RD na retina ocorrem por meio de lesões microvasculares, comumente divididas entre lesões vermelhas e brancas. O grupo das lesões vermelhas é composto por microaneurismas (MAs), definidos como pequenas protuberâncias vasculares, e por hemorragias (HEs), que consistem em manchas maiores decorrentes de sangramentos internos. Em contrapartida, as lesões brancas englobam os exsudatos duros (EXs), formados pelo acúmulo de lipoproteínas, e os exsudatos moles (SEs), também conhecidos como manchas de algodão, que indicam danos em fibras nervosas ou isquemia. A identificação precisa desses sinais clínicos em imagens de fundo de olho é fundamental para graduar a gravidade da patologia e prevenir o comprometimento visual irreversível [Almattar et al. 2023].



**Figura 1. Amostra do conjunto de dados IDRiD [Porwal et al. 2018], adaptada para visualização das máscaras de referência (ground truth) das lesões: microaneurismas (MAs), hemorragias (HEs), exsudatos duros (EXs) e exsudatos moles (SEs).**

A fotografia colorida de fundo de olho (*Color Fundus Photography* — CFP) é considerada o padrão-ouro para o rastreamento da doença, contribuindo para o Sistema Internacional de Classificação da Retinopatia Diabética (ICDR) e viabilizando a visualização de biomarcadores que caracterizam seus diferentes estágios de severidade [Tan and Wong 2023]. Essas imagens são analisadas por especialistas, que as utilizam para examinar a retina e classificá-la conforme escalas de gravidade estabelecidas. O crescimento contínuo da população diabética implica diretamente no aumento do volume de exames a serem avaliados, tornando essa tarefa progressivamente mais demandante para os profissionais de saúde. Trata-se de uma atividade complexa, sujeita a variações anatômicas, condições de iluminação inconsistentes e subjetividade interpretativa, fatores que podem comprometer tanto a precisão quanto a reprodutibilidade dos diagnósticos [Wong et al. 2021].

Diante desses desafios, têm sido desenvolvidas ferramentas baseadas em inteligência artificial (IA) para auxiliar na análise de imagens retinianas, algumas das quais já estão aprovadas para uso clínico na detecção de retinopatia diabética (RD). Em particular, métodos fundamentados em aprendizagem profunda (*deep learning*) têm apresentado desempenho promissor tanto em tarefas de classificação quanto de segmentação automática

das lesões associadas à doença. Nesse cenário, a arquitetura U-Net e suas variantes, como a *Residual U-Net*, destacam-se por integrar mecanismos que preservam informações espaciais locais e estratégias de aprendizagem residual que favorecem a propagação de gradientes e a otimização de redes mais profundas [Gaikwad and Gautam 2024].

Apesar dos avanços obtidos, a delimitação precisa de bordas e a representação de texturas associadas a lesões de pequenas dimensões ainda constituem um desafio relevante para redes neurais convolucionais convencionais. Nesse contexto, este estudo apresenta como principal contribuição a proposição de uma arquitetura baseada em U-Net com *encoder EfficientNet-B4* [Tan and Le 2019], investigando a incorporação de um módulo de decomposição wavelet bidimensional (DWT 2D, base Haar) na conexão de atalho (*skip connection*) da saída da stem [Wu et al. 2025]. Parte-se da hipótese de que o realce de componentes de frequência espacial pode favorecer a segmentação de exsudatos e hemorragias ao preservar informações de alta frequência frequentemente atenuadas durante os processos usuais de codificação. O desempenho da arquitetura proposta é avaliado em comparação com modelos de referência descritos na literatura, com o objetivo de contribuir para o aprimoramento de métodos computacionais de apoio ao diagnóstico da RD.

## 2. Trabalhos Relacionados

A segmentação automática de lesões em imagens de fundo de olho evoluiu substancialmente com o uso de Redes Neurais Convolucionais (CNNs), sendo a arquitetura U-Net o padrão para tarefas médicas por integrar caminhos de contração e expansão para localização precisa [Ronneberger et al. 2015]. No entanto, modelos tradicionais, como a FCN-8s e a DeepLabV3+, demonstram limitações significativas ao lidar com a variabilidade e a escala reduzida de lesões de Retinopatia Diabética (RD), como microaneurismas e pequenos exsudatos [Gaikwad and Gautam 2024].

Com o objetivo de mitigar o problema do fluxo de informações em redes profundas, foi proposta a arquitetura Residual U-Net, a qual emprega o princípio da aprendizagem residual [Gaikwad and Gautam 2024]. Complementarmente, a integração de blocos Squeeze and Excite (SE) permitiu a recalibração dinâmica de mapas de características, possibilitando que a rede priorizasse canais informativos e suprimisse ruídos menos relevantes, aumentando consideravelmente o poder discriminativo do modelo.

No contexto específico da detecção de RD, diversas variantes da U-Net foram propostas para superar desafios como o baixo contraste das imagens. Sambyal et al. (2020) apresentaram uma U-Net modificada que combina convolução subpixel para o escalonamento das imagens, visando a preservação de detalhes espaciais na segmentação de microlesões. Outra abordagem relevante é a Attention U-Net, que utiliza *attention gates* para focar em regiões anatômicas específicas, técnica também explorada no modelo ARU-GD para segmentação de tumores [Maji et al. 2022]. Adicionalmente, Xue et al. (2019) desenvolveram sistemas de membranas dinâmicas para a segmentação multitarefa de lesões, enquanto Ullah et al. (2023) propuseram a SSMD-U-Net, uma rede com múltiplos decoders que utiliza aprendizado semi-supervisionado para melhorar a generalização em bases de dados limitadas.

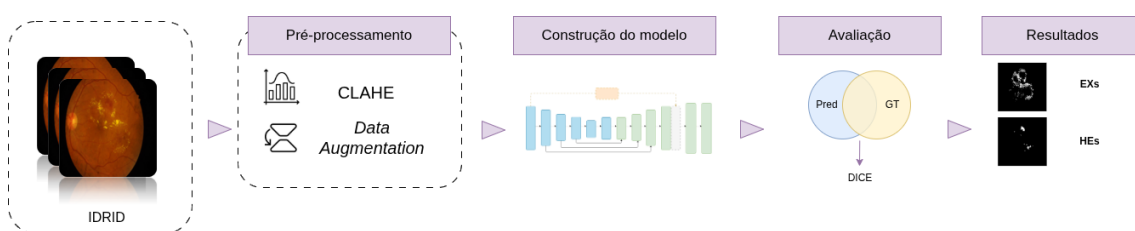
Em termos de desempenho comparativo no *dataset IDRiD*, a literatura aponta que a Residual U-Net alcança um coeficiente de Dice de aproximadamente 72,6% para exsudatos (EXs), superando modelos como a U-Net padrão (29,7%) e a DeepLabV3+ (40,1%)

[Ullah et al. 2023]. Embora mecanismos de atenção espacial e de canal (como os blocos SE) melhorem a calibração da rede, a preservação de detalhes de alta frequência permanece um desafio crítico. Diversas abordagens têm explorado o impacto da resolução de entrada e de novas arquiteturas para mitigar a perda de detalhes. Moreira et al. (2020) demonstraram que o aumento da resolução de análise permite capturar lesões em segmentos reduzidos da imagem, resultando em ganhos de precisão. No campo da segmentação de instâncias, Dias et al. (2023) propuseram o uso da Mask R-CNN, destacando a dificuldade de detectar objetos com baixa cobertura de *pixels*. Adicionalmente, Pereira et al. (2024) investigaram a arquitetura R2U-Net para segmentação semântica de microlesões, visando mitigar a escassez de dados anotados, enquanto Kaizer et al. (2025) reforçaram a eficácia do pré-processamento com CLAHE (*Contrast Limited Adaptive Histogram Equalization*) aliado a redes residuais.

Este trabalho se diferencia das abordagens anteriores ao introduzir a decomposição wavelet (DWT 2D, base Haar) diretamente na *skip connection*. Essa estratégia permite o realce explícito de componentes de borda e de texturas finas, oferecendo uma fundamentação matemática que complementa as convoluções tradicionais para capturar a morfologia complexa de exsudatos e hemorragias.

### 3. Metodologia Proposta

Nesta seção, apresenta-se a metodologia proposta para a segmentação de lesões de RD em imagens de fundo de olho. Na primeira etapa, houve a seleção de dados de repositórios públicos especializados. Em seguida, as imagens passam por pré-processamento e por aumento de dados para mitigar ruídos e ampliar a variabilidade do treinamento. A terceira etapa detalha a arquitetura do modelo, que integra redes neurais profundas a módulos de processamento de frequência. Por fim, estabelece-se o protocolo de avaliação, utilizando métricas quantitativas e métodos de validação cruzada para assegurar a robustez estatística dos resultados. As quatro etapas desenvolvidas são apresentadas na Figura 2.



**Figura 2.** Visão geral da metodologia proposta: pré-processamento (CLAHE e *Data Augmentation*), construção do modelo (U-Net com encoder EfficientNet-B4 e Módulo de Realce Wavelet na *skip connection* da camada *stem*), avaliação por coeficiente Dice e análise dos resultados. A arquitetura completa é detalhada na Figura 3 e o fluxo interno do módulo wavelet na Figura 4.

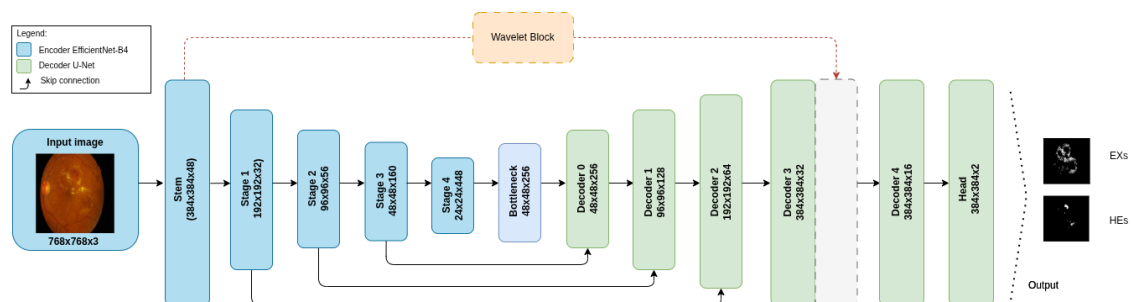
#### 3.1. Pré-Processamento e Aumento de Dados

Nesta etapa, as imagens são redimensionadas para uma resolução de 768x768 *pixels* para compatibilidade com a rede. Aplicou-se a técnica de CLAHE especificamente no

canal L do espaço de cores Lab, com um *clip limit* de 2,0, visando mitigar variações de iluminação e realçar a morfologia das lesões. Para a mitigação de ruídos de alta frequência, empregou-se um filtro Gaussiano com núcleo (kernel) de  $3 \times 3$  e desvio padrão ( $\sigma$ ) de 0,5. Para expandir a diversidade do treinamento e mitigar o *overfitting*, empregou-se uma estratégia de *Data Augmentation* baseada na biblioteca Albumentations, onde cada técnica possui uma probabilidade de aplicação de  $p = 0,5$ . Foram utilizadas inversões horizontais e verticais (*Horizontal/Vertical Flip*), rotações aleatórias de  $90^\circ$  e transformações combinadas de translação (*shift*) e escala (*scale*) com limites de 10% (0,1), além de rotações aleatórias no intervalo de  $\pm 45^\circ$ .

## 3.2. Construção do Modelo

A arquitetura proposta baseia-se na estrutura U-Net, utilizando o EfficientNet-B4 como backbone do encoder. Essa escolha justifica-se por experimentos preliminares que demonstraram a superioridade do B4 frente às versões B1, B2 e B3 da mesma família, oferecendo o melhor equilíbrio entre a capacidade de extração de características profundas e a preservação da resolução espacial necessária para lesões pequenas. A wavelet Haar foi selecionada por sua simplicidade computacional e alta sensibilidade a variações abruptas de intensidade. Optou-se por apenas um nível de decomposição para garantir que os mapas de detalhe (LH, HL, HH) mantenham uma resolução espacial elevada, evitando a perda de informação morfológica que ocorreria em níveis de decomposição mais profundos. A principal inovação introduzida é o Módulo de Realce Wavelet, integrado à primeira *skip connection*, conforme ilustrado na Figura 3.



**Figura 3. Visão geral da arquitetura proposta composta por um encoder EfficientNet-B4, o módulo de realce por wavelets na skip connection e um decoder U-Net para segmentação multi-rótulo de lesões de RD.**

### 3.2.1. Bloco Wavelet DWT 2D (Haar)

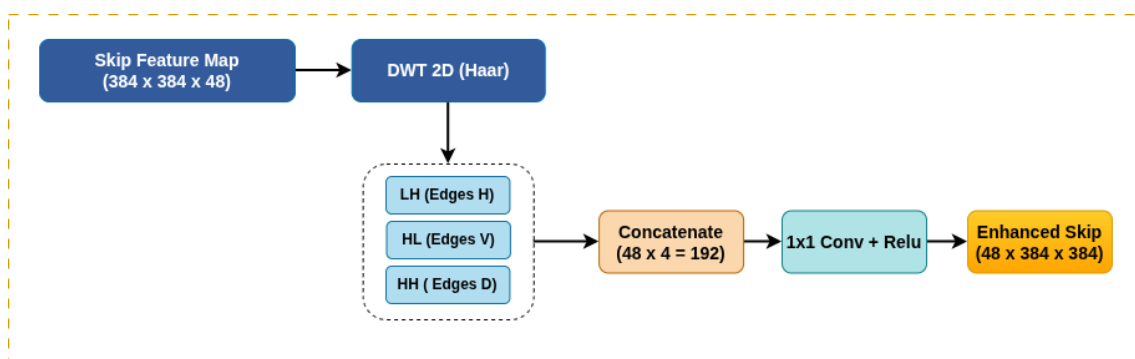
Neste módulo, os mapas de características provenientes do encoder, especificamente da saída *stem*, com 48 canais e resolução de  $384 \times 384$ , são submetidos à Decomposição Wavelet Discreta 2D (DWT 2D, base Haar). Este processo gera quatro sub-bandas de frequência: LL (aproximação), LH (bordas horizontais), HL (bordas verticais) e HH (bordas diagonais). Para garantir a compatibilidade espacial na integração, essas sub-bandas são redimensionadas para a resolução original de  $384 \times 384$  pixels.

Nesta implementação, a sub-banda LL é deliberadamente descartada. A justificativa para essa exclusão reside no fato de que a componente LL representa apenas uma

versão reduzida e suavizada da imagem original com baixa frequência, o que seria redundante, visto que o fluxo principal da U-Net já preserva informações espaciais de baixa frequência.

Ao selecionar apenas as sub-bandas de detalhe (LH, HL e HH), o módulo foca exclusivamente nas informações de alta frequência, essenciais para delimitar as bordas sutis e as texturas finas de exsudatos e pequenas hemorragias, frequentemente perdidas em camadas convolucionais profundas.

Como resultado, realiza-se a concatenação entre o mapa de características original do *stem* e as três sub-bandas de detalhe processadas, totalizando um tensor de 192 canais, conforme detalhado na Figura 4. Para reintegrar essa informação ao fluxo do decoder de forma eficiente, aplica-se uma convolução 1x1 com ativação ReLU, reduzindo a dimensionalidade de volta para 48 canais e criando o *Enhanced Skip*. Esta conexão fornece guias estruturais de alta fidelidade diretamente ao processo de reconstrução da imagem.



**Figura 4. Fluxo de processamento do módulo de realce, ilustrando a aplicação da DWT 2D (Haar), a concatenação das sub-bandas de detalhe (LH, HL, HH) e o refinamento da dimensionalidade através de uma convolução 1x1 com ativação ReLU.**

### 3.3. Avaliação

A métrica de avaliação adotada foi o coeficiente de Dice [Fleiss et al. 1981], amplamente utilizado em problemas de segmentação. Esse coeficiente combina precisão e sensibilidade em uma única medida, capturando o equilíbrio entre as predições corretas e os erros do modelo. Dado um mapa de segmentação predito  $\hat{Y}$  e a máscara de referência (*ground truth*)  $Y$ , o coeficiente é definido por:

$$\text{Dice}(\hat{Y}, Y) = \frac{2 \text{ TP}}{2 \text{ TP} + \text{ FP} + \text{ FN}} \quad (1)$$

onde TP (*true positives*) representa os *pixels* corretamente classificados como lesão, FP (*false positives*) os *pixels* de fundo classificados incorretamente como lesão, e FN (*false negatives*) os *pixels* de lesão não detectados. O coeficiente varia entre 0 (sem sobreposição) e 1 (concordância perfeita).

## 4. Resultados e Discussão

Nesta seção, são apresentados os resultados dos experimentos realizados. O método foi aplicado em Python, com a biblioteca PyTorch e aceleração por GPU. A placa de vídeo

utilizada foi a NVIDIA GeForce GTX 4090. As imagens foram pré-processadas e redimensionadas para  $768 \times 768$  pixels.

Para o desenvolvimento e avaliação do modelo, utilizou-se primordialmente o *dataset IDRiD* (Indian Diabetic Retinopathy Image Dataset), que disponibiliza imagens de alta resolução ( $4288 \times 2848$ px) com anotações especializadas para hemorragias (HEs) e exsudatos duros (EXs), exsudatos moles (SEs) e microaneurismas (MAs). Apesar da diversidade de lesões catalogadas no *dataset*, este trabalho foca exclusivamente na segmentação de HEs e EXs, por se tratar de biomarcadores críticos e fundamentais para a detecção de estágios iniciais da doença. A base é composta por 54 imagens para o conjunto de treinamento e 27 para o conjunto de teste isolado [Porwal et al. 2018].

O treinamento foi conduzido por 100 épocas utilizando uma função de perda combinada (Dice Loss + Focal Loss [Lin et al. 2017]), com o escalonador *OneCycleLR* e taxa máxima de aprendizado de  $1 \times 10^{-3}$ , visando mitigar o desequilíbrio de classes entre os pixels de lesão e o fundo. Aplicou-se uma validação cruzada com 5  *folds* sobre as 54 imagens iniciais, com uma divisão de 43 imagens para treinamento e 11 para validação em cada iteração, adotando a separação por paciente para evitar o vazamento de dados (*data leakage*).

Para a avaliação final, utilizou-se a técnica de *ensemble* (média das predições) dos cinco modelos gerados na validação cruzada para processar o conjunto de teste de 27 imagens, assegurando maior robustez estatística aos resultados. A Tabela 1 apresenta os valores do coeficiente de Dice obtidos durante as iterações da validação cruzada.

**Tabela 1. Resultados dos valores de Dice obtidos na validação cruzada.**

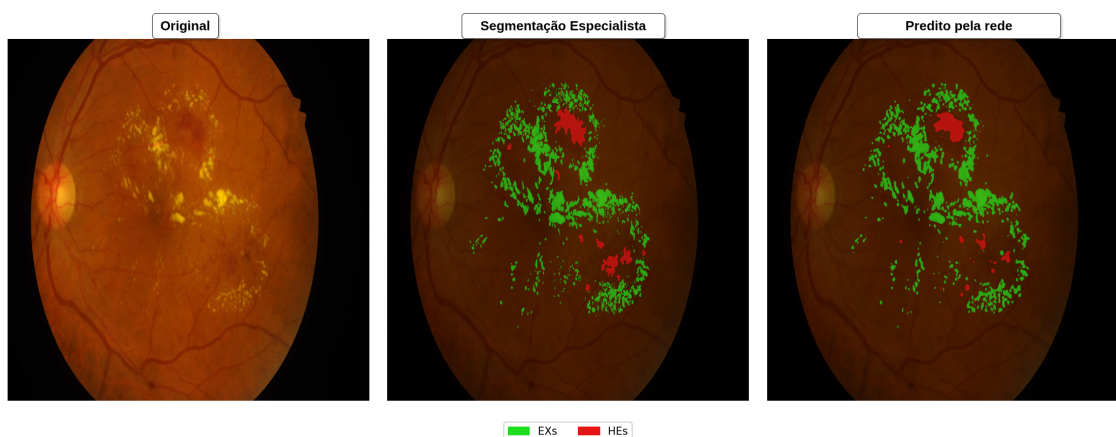
Iteração	Dice HEs	Dice EXs
1	62,8%	73,1%
2	57,5%	71,4%
3	50,4%	65,8%
4	64,8%	69,4%
5	53,4%	65,3%
<b>Média</b>	57,8%	69,0%
<b>Desvio padrão</b>	5,5%	2,7%

Os resultados experimentais foram obtidos por meio de uma avaliação sistemática que envolveu a seleção do melhor encoder, a otimização de hiperparâmetros e a validação do impacto do módulo wavelet. Inicialmente, comparou-se o desempenho de diferentes *backbones* da família EfficientNet integrados à U-Net. O modelo EfficientNet-B4 demonstrou superioridade como extrator de características, superando as versões B1, B2 e B3 em testes preliminares. O treinamento por 100 épocas com imagens de  $768 \times 768$  pixels permitiram estabelecer um baseline robusto (definido como a U-Net com backbone EfficientNet-B4, porém sem o módulo wavelet proposto), atingindo coeficientes Dice de 73,5% para EXs e 62,3% para HEs no conjunto de teste. É importante ressaltar que esses valores de baseline representam o desempenho do modelo base para comparação e são detalhados posteriormente na Tabela 3.

A introdução do Bloco Wavelet representou o avanço mais significativo na arquitetura, elevando o desempenho final para um Dice de 73,8% para exsudatos (EXs) e 65,1% para hemorragias (HEs) no conjunto de teste. Esse ganho é atribuído à capacidade da decomposição wavelet (DWT 2D, Haar) de isolar características de borda e detalhes de alta frequência, essenciais para a delimitação de lesões pequenas e de baixo contraste. O incremento de 2,8% no Dice de hemorragias em relação ao *baseline* confirma que a integração de wavelets na *skip connection* atua como um mecanismo eficaz de preservação morfológica, superando limitações de arquiteturas clássicas de segmentação (U-Net e DeepLabV3+).

Analisando os resultados qualitativos das segmentações obtidas, observa-se que a rede demonstra aprendizado consistente ao analisar o caso 055 da base IDRiD, mostrado na Figura 5. Na primeira coluna, observa-se a retinografia original, seguida pela segmentação manual do médico com EXs (verde), HEs (vermelho) e, à direita, a segmentação gerada pela Wavelet Skip U-Net, que mostra boa correspondência com essas estruturas anatômicas.

A análise do caso 055 evidencia o desempenho do modelo na segmentação simultânea de exsudatos e hemorragias. Para os EXs, a rede demonstra precisão na delimitação de bordas e texturas, mantendo consistência mesmo em lesões de pequenas dimensões. Para as HEs, a localização espacial foi preservada, porém com sensibilidade variável em lesões de baixo contraste, o que corrobora a variação observada nas métricas quantitativas da Tabela 1. A preservação morfológica das patologias reforça a eficácia do Bloco Wavelet em capturar detalhes de alta frequência que seriam atenuados em arquiteturas convencionais.



**Figura 5. Estudo de caso 055 da base IDRiD: segmentação simultânea de EXs e HEs pela Wavelet Skip U-Net.**

Ao comparar os resultados da Tabela 2, observa-se que o modelo proposto superou as arquiteturas de referência da literatura (U-Net padrão e DeepLabV3+). Em relação à segmentação de exsudatos (EXs), o modelo atingiu um Dice de 73,8%, superando ligeiramente a Residual U-Net (72,6%). No entanto, para hemorragias (HEs), o desempenho de 65,1% permaneceu abaixo da Residual U-Net (73,8%) e da SSMD-U-Net (87,3%), sugerindo que a variabilidade textural das hemorragias ainda representa um desafio para o realce via wavelets de Haar.

**Tabela 2. Comparação quantitativa do desempenho (Coeficiente Dice) entre o modelo proposto (*Wavelet Skip U-Net*) e arquiteturas da literatura, utilizando o conjunto de teste do *dataset* IDRiD.**

<b>Métodos</b>	<b>Dice HEs</b>	<b>Dice EXs</b>
U-Net [Ullah et al. 2023]	25,1%	29,7%
DeepLabV3+ [Ullah et al. 2023]	38,6%	40,1%
FCN [Ullah et al. 2023]	45,3%	45,9%
SSMD-U-Net [Ullah et al. 2023]	87,3%	91,7%
Residual U-Net [Gaikwad and Gautam 2024]	73,8%	72,6%
<b>Metodologia Proposta</b>	<b>65,1%</b>	<b>73,8%</b>

Adicionalmente, a Tabela 3 compara o desempenho do modelo proposto com o *baseline*, definido como a arquitetura EfficientNet-B4 integrada ao decodificador U-Net sem o Módulo de Realce Wavelet e treinado sob as mesmas condições experimentais, com ganhos nas classes de hemorragia (HEs) e exsudato (EXs).

**Tabela 3. Resultados dos valores de Dice pela metodologia proposta, comparados com o *baseline*.**

<b>Modelos</b>	<b>Dice HEs</b>	<b>Dice EXs</b>
<i>Baseline</i>	62,3%	73,5%
Método proposto	<b>65,1%</b>	<b>73,8%</b>

Os resultados confirmam que a adição do Bloco Wavelet trouxe melhorias em relação ao *baseline*, com incrementos de 2,8% no Dice de HEs e 0,3% no Dice de EXs. Embora o ganho em exsudatos seja marginal, o avanço em hemorragias é mais expressivo, reforçando a hipótese de que a decomposição wavelet de Haar na primeira *skip connection* atua como um mecanismo eficaz de extração de características de alta frequência, favorecendo a delimitação de lesões de bordas irregulares.

Essa variabilidade confirma que a integração da *Enhanced Skip* ao *encoder* EfficientNet-B4 favorece a captura de detalhes de alta frequência, tornando o modelo particularmente robusto à morfologia dos exsudatos, enquanto hemorragias sutis permanecem um desafio para a sensibilidade da rede. O desempenho inferior observado em amostras com essas características, que podem apresentar índices de Dice abaixo de 40%, sugere que estratégias complementares, como aumento de dados específico para essa classe ou funções de perda com maior penalização de falsos negativos, podem ser exploradas em trabalhos futuros.

Um ponto fundamental na avaliação do modelo proposto é a sua eficiência em ambientes de escassez de dados (*low-data*). Enquanto modelos de alto desempenho na literatura, como a SSMD-UNet [Ullah et al. 2023], baseiam-se em aprendizagem semi-supervisionada e utilizam bases massivas como o EyePACS (com aproximadamente 80.000 imagens) para pré-treinamento, a arquitetura proposta foi treinada exclusivamente com os dados limitados do IDRiD. Semelhantemente, trabalhos que integram blocos Squeeze-and-Excitation (SE) à Residual U-Net frequentemente recorrem a bases exter-

nas, como o FGADR, antes do ajuste fino no IDRiD [Gaikwad and Gautam 2024]. Em contraste, o presente trabalho utilizou apenas 43 imagens para treino e 11 para validação, totalizando 54 amostras, com avaliação em 27 imagens de teste.

Mesmo com volume de dados inferior, o Wavelet Skip U-Net demonstrou robustez ao atingir Dice de 73,8% para EXs, superando a Residual U-Net (72,6%) e modelos, como o U-Net padrão (29,7%) e o DeepLabV3+ (40,1%). Embora a SSMD-U-Net apresente métricas superiores (91,7% em EXs), deve-se considerar que sua generalização é impulsionada por uma escala de dados ordens de magnitude maior, o que torna a comparação direta desfavorável ao presente método do ponto de vista dos recursos disponíveis.

Esses resultados corroboram a hipótese de que a decomposição wavelet de Haar na *skip connection* compensa a ausência de grandes volumes de dados, ao facilitar a captura de detalhes morfológicos essenciais para a segmentação de lesões pequenas. Assim, a arquitetura proposta mostra-se uma alternativa viável e eficiente para aplicações clínicas em que a disponibilidade de dados anotados por especialistas é restrita.

## 5. Conclusão

Este trabalho propôs uma arquitetura baseada em U-Net com *encoder* EfficientNet-B4 e um Módulo de Realce Wavelet integrado à primeira *skip connection* para a segmentação de lesões de Retinopatia Diabética em imagens de fundo de olho. A abordagem explorou a decomposição wavelet discreta 2D (DWT, base Haar) para preservar informações de alta frequência, como bordas e texturas finas, frequentemente atenuadas pelas camadas convolucionais profundas, mitigando assim um dos principais desafios na delimitação de exsudatos e hemorragias de pequenas dimensões.

Os resultados obtidos no dataset IDRiD demonstram a eficácia da Wavelet Skip U-Net, que alcançou coeficientes Dice de 73,8% para exsudatos e 65,1% para hemorragias no conjunto de teste. A arquitetura proposta superou integralmente modelos consolidados, como a U-Net padrão e a DeepLabV3+, em ambas as classes, além de apresentar um desempenho superior ao da Residual U-Net, especificamente na segmentação de exsudatos. Um diferencial relevante da metodologia proposta é sua eficiência em regime de escassez de dados: enquanto modelos de maior desempenho na literatura, como a SSMD-U-Net, recorrem a bases com dezenas de milhares de imagens para pré-treinamento, a arquitetura aqui desenvolvida foi treinada exclusivamente com o *dataset* do IDRiD, reforçando seu potencial para aplicações clínicas em que dados anotados por especialistas são escassos.

Apesar dos resultados promissores, a segmentação de hemorragias ainda apresenta maior variabilidade, especialmente em lesões de baixo contraste, o que indica que essa classe permanece um desafio para o realce baseado em wavelets. Como perspectivas futuras, sugere-se investigar famílias wavelet com maior suavidade, como *Daubechies* ou *Symlets*, combiná-las com mecanismos de atenção *Squeeze-and-Excitation* para recalibração dinâmica de canais, e expandir o treinamento para bases de dados mais extensas. Esses avanços podem consolidar a arquitetura como uma ferramenta viável no auxílio ao diagnóstico precoce da retinopatia diabética em cenários clínicos de larga escala.

## 6. Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código Financeiro 001, ao Conselho Nacional de Desenvolvimento Científico

e Tecnológico (CNPq), à Empresa Brasileira de Serviços Hospitalares (Ebserh) (Concesso número 409593/2021-4) e à Fundação de Amparo à Pesquisa Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) pelo apoio financeiro.

## Referências

- Almattar, W., Luqman, H., and Khan, F. A. (2023). Diabetic retinopathy grading review: Current techniques and future directions. *Image and Vision Computing*, 139:104821.
- Chagas, T. A., Dos Reis, M. A., Leivas, G., Santos, L. P., Gossenheimer, A. N., Melo, G. B., Malerbi, F. K., and Schaan, B. D. (2023). Prevalence of diabetic retinopathy in brazil: a systematic review with meta-analysis. *Diabetology & Metabolic Syndrome*, 15(1):34.
- Dias, M., Santos, C., Aguiar, M., Welfer, D., Pereira, A., and Ribeiro, M. (2023). 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 *Seminário Integrado de Software e Hardware (SEMISH)*, pages 1–12. SBC.
- Fleiss, J. L., Levin, B., Paik, M. C., et al. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23.
- Gaikwad, A. and Gautam, A. (2024). Segmentation of hard exudates and hemorrhages from diabetic retinopathy images using residual u-net with squeeze and excite blocks. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 3064–3069. IEEE.
- Kaizer, A. C. and Machado, A. M. C. (2025). Aplicação de técnicas de aprendizado profundo na classificação da retinopatia diabética. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 49–54. SBC.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.
- Maji, D., Sigedar, P., and Singh, M. (2022). Attention res-unet with guided decoder for semantic segmentation of brain tumors. *Biomedical Signal Processing and Control*, 71:103077.
- Ministério da Saúde (2025). Vigitel Brasil 2006-2024: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal entre 2006 e 2024. Technical report, Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde e Ambiente. Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis, Brasília. Acesso em: 5 mar. 2026.
- Moreira, F., Schaan, B., Schneiders, J., Reis, M., Serpa, M., and Navaux, P. (2020). Impacto da resolução na detecção de retinopatia diabética com uso de deep learning. 494–499.
- Pereira, A., Santos, C., Aguiar, M., Welfer, D., Dias, M., de Menezes, R., and Santana, D. (2024). Enhanced semantic segmentation of retinal microlesions through r2u-net

- architecture. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 13–24. SBC.
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., and Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research. *Data*, 3(3):25.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Sambyal, N., Saini, P., Syal, R., and Gupta, V. (2020). Modified u-net architecture for semantic segmentation of diabetic retinopathy images. *Biocybernetics and Biomedical Engineering*, 40(3):1094–1109.
- Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.
- Tan, T.-E. and Wong, T. Y. (2023). Diabetic retinopathy: Looking forward to 2030. *Frontiers in endocrinology*, 13:1077669.
- Ullah, Z., Usman, M., Latif, S., Khan, A., and Gwak, J. (2023). Ssm-d-unet: Semi-supervised multi-task decoders network for diabetic retinopathy segmentation. *Scientific Reports*, 13(1):9087.
- Wong, D. C. S., Kiew, G., Jeon, S., and Ting, D. (2021). Singapore eye lesions analyzer (selena): the deep learning system for retinal diseases. In *Artificial intelligence in ophthalmology*, pages 177–185. Springer.
- Wu, J., Li, J., Yang, J., and Mei, S. (2025). Wavelet-integrated deep neural networks: A systematic review of applications and synergistic architectures. *Neurocomputing*, page 131648.
- Xue, J., Yan, S., Qu, J., Qi, F., Qiu, C., Zhang, H., Chen, M., Liu, T., Li, D., and Liu, X. (2019). Deep membrane systems for multitask segmentation in diabetic retinopathy. *Knowledge-Based Systems*, 183:104887.