



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS CORRENTE
CURSO ANÁLISE E DESENVOLVIMENTO DE SISTEMAS (ADS)

FELIPE LUSTOSA CARVALHO

SEGMENTAÇÃO DE VASOS SANGUÍNEOS DA RETINA COM U-NET++

CORRENTE-PI

2025

SEGMENTAÇÃO DE VASOS SANGUÍNEOS DA RETINA COM U-NET++

Autor: Felipe Lustosa Carvalho

Orientador: Igor Bezerra Reis

RESUMO

A retinopatia diabética é uma das principais causas de cegueira em adultos em idade produtiva, e seu diagnóstico precoce depende da análise detalhada da vasculatura da retina. A segmentação automática desses vasos por meio de imagens de retinografia é uma tarefa desafiadora devido ao baixo contraste e à complexidade das estruturas capilares. Este trabalho apresenta o desenvolvimento e a avaliação de um modelo de *deep learning* baseado na arquitetura U-Net++ para realizar essa segmentação. O modelo foi treinado e validado utilizando o dataset público DRIVE (*Digital Retinal Images for Vessel Extraction*). A metodologia abrange desde o pré-processamento das imagens, com a criação de *patches* e aumento de dados, até a implementação de funcionalidades avançadas como *deep supervision*. Os resultados demonstram a alta eficácia do modelo, que alcançou um *Dice Coefficient* de 0.8328 e uma AUC-ROC de 0.9543, validando a U-Net++ como uma solução robusta e precisa para essa aplicação crítica na área da saúde.

Palavras-chave: Segmentação de Imagens, U-Net++, *Deep Learning*, Imagens Médicas, Retinopatia Diabética, Dataset DRIVE.

1 INTRODUÇÃO

A visão computacional, impulsionada pelos avanços em *deep learning*, tem revolucionado a área da imagem médica, oferecendo ferramentas capazes de automatizar tarefas de diagnóstico que tradicionalmente dependem da análise manual por especialistas. Uma dessas tarefas é a avaliação de imagens de retinografia (fundo de olho), fundamental para o acompanhamento de diversas patologias oftalmológicas.

Entre elas, a retinopatia diabética (RD) se destaca como uma complicação severa do diabetes e uma das principais causas de perda de visão no mundo. Sinais precoces da RD, como microaneurismas e neovascularização, estão diretamente ligados a alterações na morfologia dos vasos sanguíneos da retina. A segmentação precisa dessa rede vascular é, portanto, um passo crucial para a detecção precoce e o monitoramento da progressão da doença.

No entanto, a segmentação de vasos retinianos é intrinsecamente complexa. As imagens frequentemente apresentam baixo contraste, iluminação irregular e ruído. Além disso, os vasos capilares são estruturas extremamente finas e ramificadas, difíceis de serem distinguidas do fundo da retina. A segmentação manual, além de ser um processo lento e tedioso, está sujeita à variabilidade entre diferentes especialistas.

Neste contexto, este trabalho propõe a implementação de uma solução automatizada utilizando a arquitetura de rede neural U-Net++. Este modelo é uma evolução da consagrada U-Net, projetada especificamente para tarefas de segmentação biomédica mais precisas. O objetivo principal é desenvolver, treinar e avaliar um modelo U-Net++ no *benchmark dataset* DRIVE, demonstrando sua capacidade de segmentar com alta fidelidade a complexa rede vascular da retina.

2 BACKGROUND

A U-Net, proposta por Ronneberger et al. (2015), é uma arquitetura de rede neural convolucional (CNN) que se tornou um padrão de fato para a segmentação de imagens biomédicas. Sua principal inovação é a estrutura em forma de "U", composta por um caminho de contração (*encoder*) e um caminho de expansão (*decoder*). O *encoder* extrai características da imagem em diferentes níveis de abstração, enquanto o *decoder* utiliza essas características para reconstruir um mapa de segmentação pixel a pixel. A característica mais marcante da U-Net são as *skip connections* (conexões de atalho), que conectam diretamente os mapas de características do

encoder aos do *decoder*, permitindo que a rede combine informações semânticas (o quê) com informações espaciais (onde), resultando em segmentações de alta resolução.

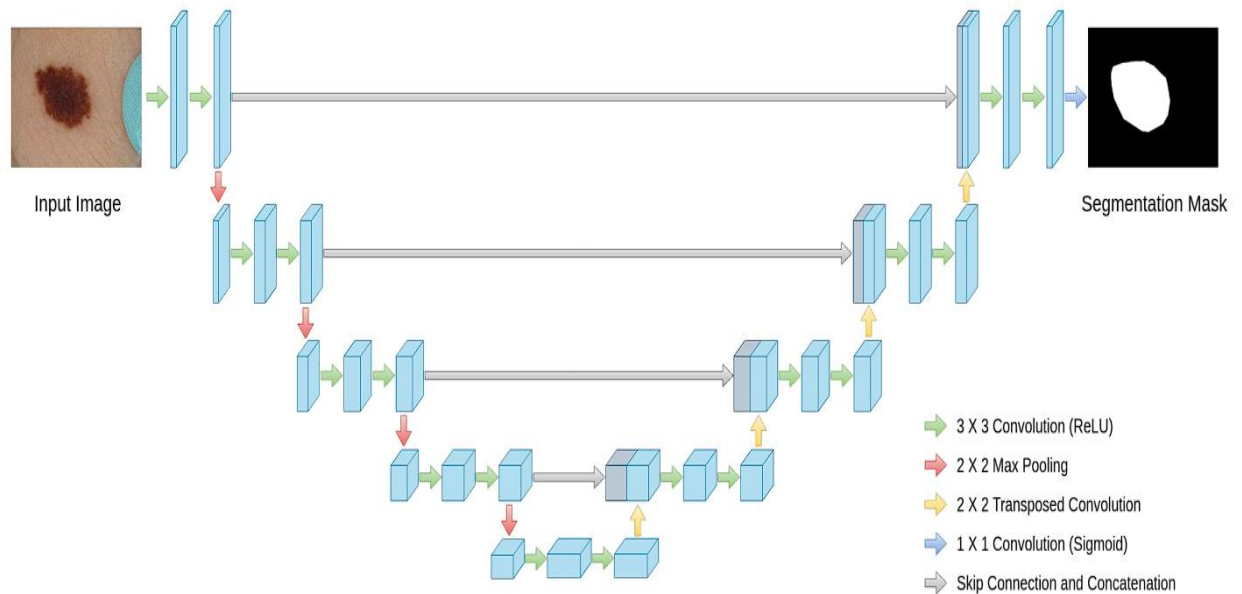


Figura 1: Arquitetura U-Net.

2.2 Arquitetura U-Net++

A U-Net++, proposta por Zhou et al. (2018), é uma evolução direta da U-Net que aprimora sua arquitetura para obter segmentações mais precisas. Sua principal inovação reside no redesenho das conexões de atalho (*skip connections*) para mitigar o abismo semântico entre os mapas de características do *encoder* (extrator de features) e do *decoder* (reconstrutor da segmentação).

Conforme ilustrado na Figura 2, a U-Net++ introduz nós intermediários que criam conexões aninhadas e densas. A nomenclatura dos nós, $X_{i,j}$, é fundamental para entender a arquitetura:

- i representa o índice da camada de *down-sampling* ao longo do *encoder* (a profundidade da rede).
- j representa o índice da camada convolucional ao longo da *skip pathway* densa. Um nó $X_{i,j}$ é calculado através de uma operação de convolução aplicada à concatenação de dois conjuntos de informações:

Um nó $X_{i,j}$ é calculado através de uma operação de convolução aplicada à concatenação de dois conjuntos de informações:

1. As saídas de todos os nós anteriores no mesmo nível ($X_{i,0}, \dots, X_{i,j-1}$).
2. A saída da camada inferior, após passar por uma operação de *up-sampling* ($U(X_{i+1}, j-1)$).

Esse mecanismo cria efetivamente blocos de convolução densos ao longo das *skip pathways*. O resultado é que, quando os mapas de características chegam ao *decoder*, eles já são uma fusão rica de features de múltiplas escalas semânticas, o que permite ao modelo lidar melhor com objetos de tamanhos variados e detalhes finos — como os vasos capilares da retina.

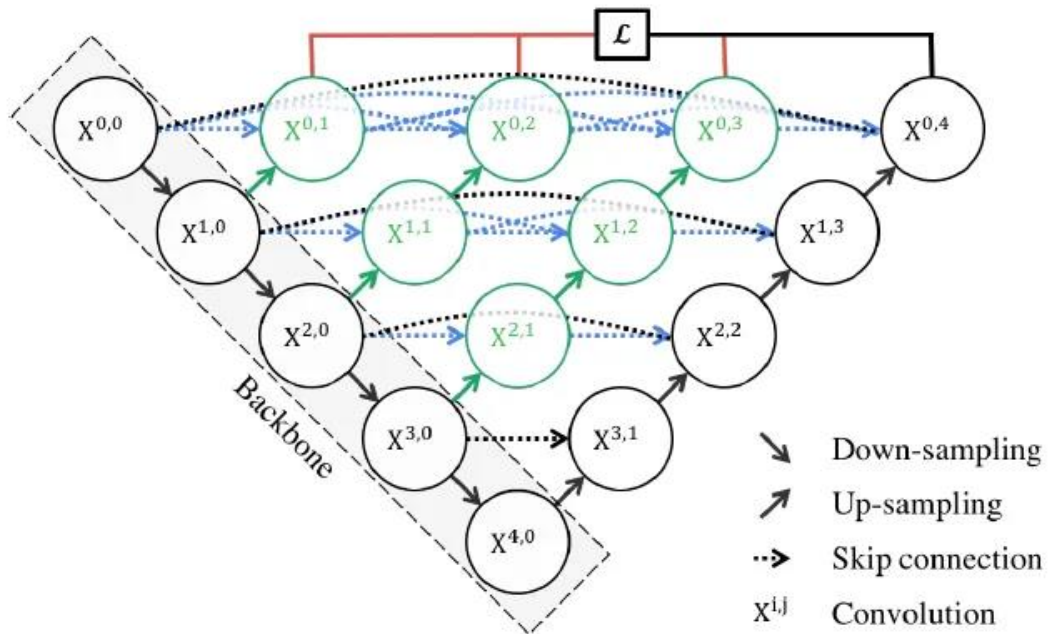


Figura 2: Diagrama detalhado da arquitetura U-Net++. As setas pretas indicam o *down-sampling* (encoder), as azuis o *up-sampling* (decoder), e as verdes as conexões densas nas *skip pathways* redesenhadas. As saídas para o *deep supervision* estão destacadas.

A segunda inovação crucial da U-Net++ é o *deep supervision* (supervisão profunda). Como os nós $X_{0,1}$, $X_{0,2}$, $X_{0,3}$ e $X_{0,4}$ geram mapas de características na resolução da imagem original, cada um deles pode ser usado para produzir um mapa de segmentação. Durante o treinamento, uma função de perda é calculada para cada uma dessas saídas e, em seguida, combinada (de forma ponderada ou por média) para atualizar os pesos da rede. Isso força todas as sub-redes aninhadas a

aprenderem features relevantes, melhora o fluxo de gradientes e acelera a convergência. Além disso, confere ao modelo a flexibilidade de ser "podado" durante a inferência: pode-se usar a saída de X_0,4 para máxima precisão ou a de X_0,1 para uma inferência muito mais rápida, dependendo da necessidade da aplicação.

2.3. Dataset DRIVE

O dataset DRIVE (*Digital Retinal Images for Vessel Extraction*) é um recurso público amplamente utilizado para validar algoritmos de segmentação de vasos sanguíneos. Ele consiste em 40 imagens coloridas de retinografia, divididas em 20 para treinamento e 20 para teste. Cada imagem possui uma máscara de segmentação correspondente, criada manualmente por especialistas, que serve como "verdade-base" (*ground truth*) para treinar e avaliar os modelos.

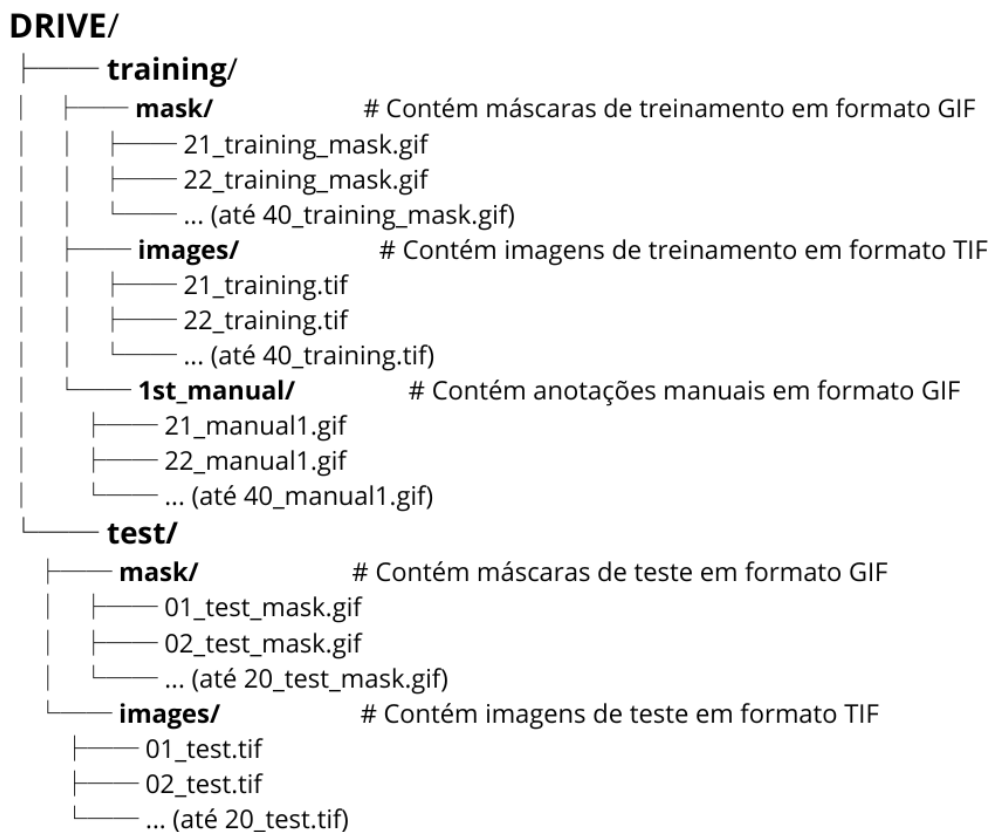


Figura 3: Arquitetura das pastas do dataset.

3 MÉTADOS

O método proposto segue um *pipeline* bem definido, desde a preparação dos dados até a avaliação final do modelo, conforme detalhado nas etapas a seguir.

3.1. Pré-processamento e Preparação dos Dados

Devido ao tamanho das imagens originais do DRIVE (584x565 *pixels*) e à limitação de memória computacional, a primeira etapa consistiu em dividir as imagens e suas respectivas máscaras em *patches* (recortes) de 128x128 *pixels*. Essa abordagem não apenas viabiliza o treinamento, mas também aumenta significativamente a quantidade de dados de treinamento disponíveis. Após a extração dos *patches*, o conjunto de dados de treinamento foi dividido em 90% para treino e 10% para validação. Por fim, todos os valores de pixels das imagens foram normalizados para o intervalo [0, 1].

3.2. Aumento de Dados (*Data Augmentation*)

Para aumentar a robustez do modelo e prevenir o *overfitting*, foram aplicadas técnicas de aumento de dados em tempo real durante o treinamento. Especificamente, foram utilizados espelhamentos aleatórios horizontais e verticais nos patches de treinamento.

3.3. Arquitetura e Treinamento do Modelo

Foi implementada a arquitetura U-Net++ utilizando a biblioteca *TensorFlow* com a *API Keras*. O modelo foi configurado para usar o *deep supervision*, gerando quatro saídas intermediárias, cujas perdas foram ponderadas com pesos de [0.1, 0.2, 0.3, 0.4] para as saídas *output_1* a *output_4*, respectivamente.

A função de perda escolhida foi uma combinação de *Binary Cross-Entropy* e *Dice Loss*. Essa perda híbrida é eficaz para segmentação, pois penaliza tanto erros a nível de *pixel* quanto a falta de sobreposição espacial entre a predição e a máscara real.

O modelo foi treinado utilizando o otimizador Adam com uma taxa de aprendizado inicial de 1e-4. Foram empregados os seguintes *callbacks*:

- *ReduceLROnPlateau*: Reduz a taxa de aprendizado caso a perda de validação estagne.

- **EarlyStopping:** Interrompe o treinamento se não houver melhora na perda de validação por um número definido de épocas (paciência = 10), evitando o sobreajuste.
- **ModelCheckpoint:** Salva apenas os pesos do modelo que apresentaram o melhor desempenho no conjunto de validação.

As principais etapas e os hiperparâmetros utilizados na experimentação estão sumarizados na Tabela 1.

Tabela 1 – Resumo da metodologia e dos hiperparâmetros utilizados no projeto.

Etapas	Parâmetro	Descrição / Valor
Preparação de Dados	Dataset	DRIVE (Digital Retinal Images for Vessel Extraction)
	Tamanho da Imagem Original	584 x 565 pixels
	Extração de Patches	128 x 128 pixels
	Divisão dos Dados	90% Treino / 10% Validação
Aumento de Dados	Técnicas Aplicadas	Espelhamento Aleatório Horizontal e Vertical
	Aplicação	Em tempo real, apenas no conjunto de treinamento.
Arquitetura do Modelo	Modelo	U-Net++ com Deep Supervision
	Biblioteca	TensorFlow / Keras
	Saídas (Deep Supervision)	4 saídas intermediárias (output_1 a output_4)

Treinamento	Função de Perda	Combinação de Binary Cross-Entropy + Dice Loss
	Pesos da Perda	[0.1, 0.2, 0.3, 0.4] para as saídas 1 a 4, respectivamente.
	Otimizador	Adam
	Taxa de Aprendizado Inicial	1e-4 (0.0001)
	Tamanho do Batch	16
	Callbacks	ModelCheckpoint (salvar o melhor modelo) ReduceLROnPlateau (reduzir LR se a val_loss estagnar) EarlyStopping (paciência = 10, monitorando val_loss)

Fonte: Autores (2025).

4. RESULTADOS E DISCUSSÕES

O modelo foi treinado por 55 épocas, momento em que o *callback EarlyStopping* interrompeu o processo, indicando a convergência. A avaliação foi realizada no conjunto de validação.

4.1. Métricas de Desempenho

As métricas quantitativas consolidadas no final do treinamento estão apresentadas na Tabela 2.

Tabela 2 – Métricas finais de desempenho do modelo U-Net++.

Métrica	Valor Obtido (Validação)
Dice Coefficient	0.8328
Acurácia	0.9571
Sensibilidade (Recall)	0.8255

Especificidade	0.9771
AUC-ROC	0.9543

Fonte: Autores (2025).

A discussão desses resultados revela um desempenho excelente. O *Dice Coefficient* de 0.8328 indica uma alta sobreposição entre os vasos segmentados pelo modelo e as anotações manuais. A Sensibilidade de 0.8255 mostra que o modelo é eficaz em identificar a maioria dos pixels que de fato pertencem aos vasos, enquanto a altíssima Especificidade de 0.9771 demonstra sua capacidade de evitar a classificação incorreta do fundo da retina como vaso. O valor da AUC-ROC de 0.9543 confirma o poder discriminatório excepcional do modelo.

4.2. Análise Gráfica

A evolução do treinamento pode ser observada na Figura 4. As curvas de perda (*loss*) de treino e validação decaem de forma consistente e se aproximam, indicando um bom ajuste do modelo, sem *overfitting* significativo. A curva do *Dice Coefficient* (*dice_coefficient*) sobe e atinge um platô estável, confirmando a convergência do aprendizado.

Evolução do Treinamento do Modelo

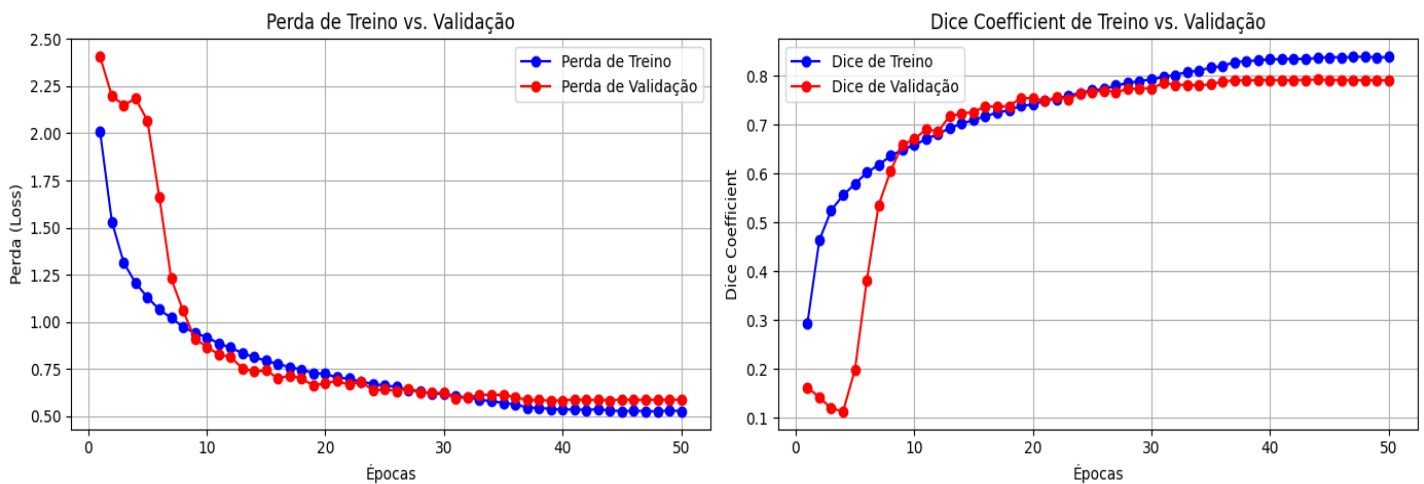


Figura 4: Gráficos de Perda e Dice Coefficient durante o treinamento.

A curva ROC, apresentada na Figura 5 para uma imagem de exemplo, ilustra visualmente a alta performance do modelo. A curva se projeta acentuadamente em direção ao canto superior esquerdo, o que é característico de um classificador com alta taxa de verdadeiros positivos e baixa taxa de falsos positivos.

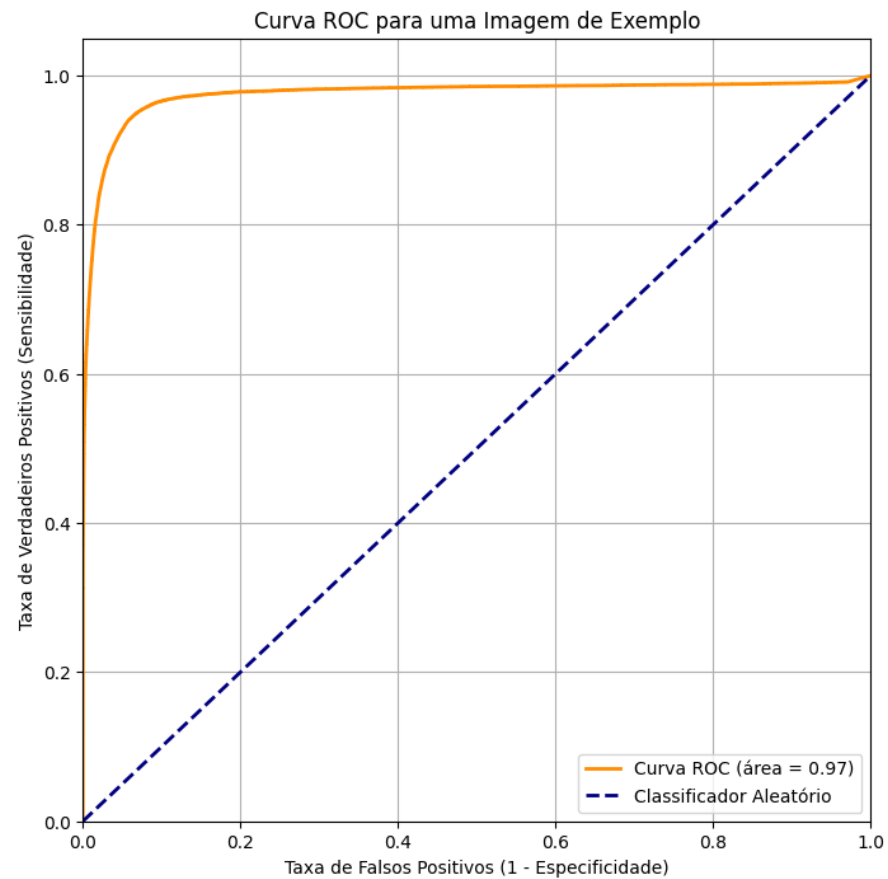
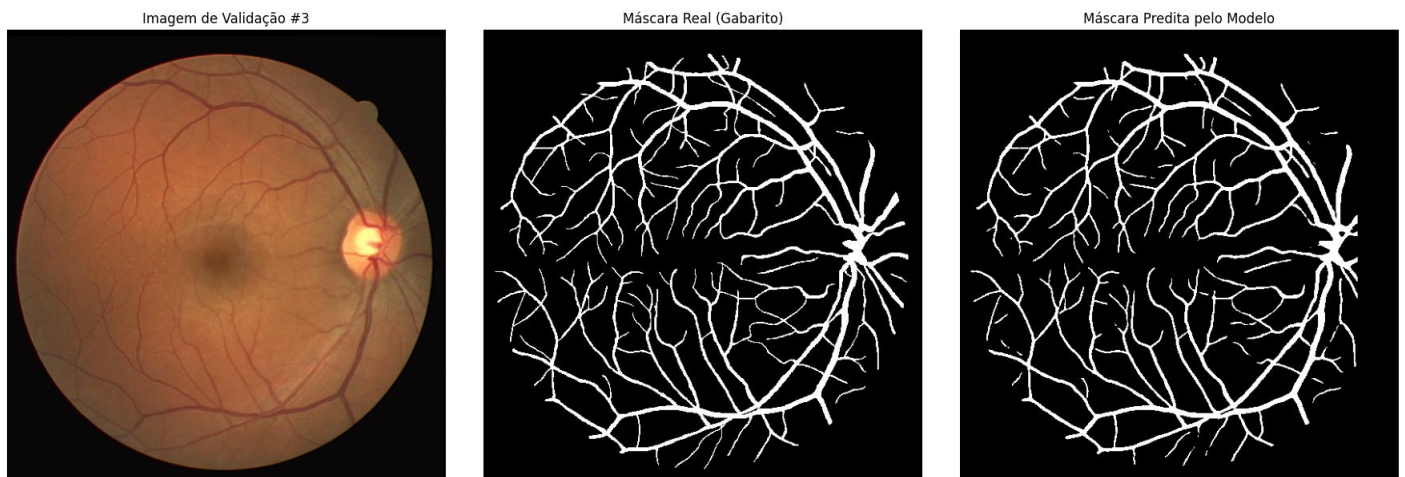


Figura 5: Curva ROC para uma imagem de exemplo do conjunto de validação, com AUC de 0.95.

4.3. Análise Qualitativa

A Figura 6 e 7 apresenta um exemplo visual dos resultados da segmentação. Observa-se uma alta fidelidade entre a máscara predita pelo modelo e a máscara manual (*Ground Truth*), com o modelo sendo capaz de delinear tanto os vasos mais espessos quanto muitos dos capilares mais finos, demonstrando o sucesso da abordagem.



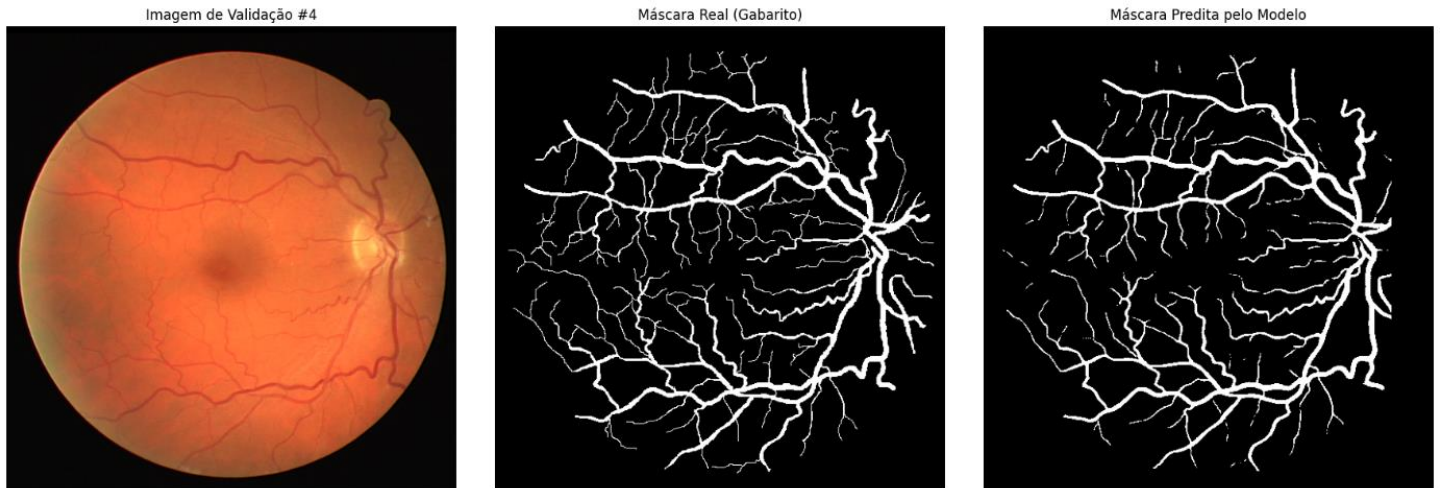


Figura 6 e 7: Comparativo visual. (A) Imagem Original da Retina; (B) Máscara Manual (Ground Truth); (C) Máscara Predita pelo Modelo U-Net++.

5. CONCLUSÕES

Este trabalho validou com sucesso a superioridade da arquitetura U-Net++ para a complexa tarefa de segmentação de vasos sanguíneos em imagens de retina. O modelo desenvolvido, treinado no desafiador dataset DRIVE, demonstrou um desempenho de alto nível, alcançando um Dice Coefficient de 0.8328 e uma AUC de 0.9543. Esses resultados não apenas atestam a precisão do modelo em delinear a estrutura vascular, mas também sua robustez em discriminar pixels de vaso e fundo de forma confiável.

O sucesso da abordagem é atribuído diretamente às inovações da U-Net++, notadamente às conexões aninhadas e densas e ao uso estratégico do *deep supervision*. Essas características foram cruciais para capacitar o modelo a capturar tanto as artérias principais quanto os finos capilares, superando um dos maiores desafios da análise de retinografias.

Embora os resultados sejam expressivos, a pesquisa abre caminhos para trabalhos futuros. A implementação de técnicas de pós-processamento poderia refinar ainda mais as máscaras segmentadas, enquanto a validação cruzada em *datasets* diversos (como CHASE_DB1 e STARE) confirmaria a generalização do modelo. Em suma, a ferramenta aqui desenvolvida constitui um passo significativo em direção a

sistemas de auxílio ao diagnóstico (CAD) automatizados, com o potencial de tornar o rastreamento da retinopatia diabética mais rápido, preciso e acessível.

6. REFERÊNCIAS

- Ronneberger, O., Fischer, P., & Brox, T. (2015). **U-Net: Convolutional Networks for Biomedical Image Segmentation**. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*.
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). **UNet++: A Nested U-Net Architecture for Medical Image Segmentation**. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*.