

Article

Retinal Vessel Segmentation Based on a Lightweight U-Net and Reverse Attention

Fernando Daniel Hernandez-Gutierrez ¹, Eli Gabriel Avina-Bravo ^{2,3}, Mario Alberto Ibarra-Manzano ¹,
Jose Ruiz-Pinales ^{1,*}, Emmanuel Ovalle-Magallanes ⁴ and Juan Gabriel Avina-Cervantes ^{1,*}

¹ Telematics and Digital Signal Processing Research Groups (CAs), Engineering Division, Campus Irapuato-Salamanca, University of Guanajuato, Salamanca 36885, Mexico; fd.hernandezgutierrez@ugto.mx (F.D.H.-G.); ibarram@ugto.mx (M.A.I.-M.)

² Tecnológico de Monterrey, Institute of Advanced Materials for Sustainable Manufacturing, Calle del Puente 222, Tlalpan 14380, Mexico; egavinabra@tec.mx

³ Tecnológico de Monterrey, School of Engineering and Sciences, Calle del Puente 222, Tlalpan 14380, Mexico

⁴ Dirección de Investigación y Doctorado, Facultad de Ingenierías y Tecnologías, Universidad La Salle Bajío, Av. Universidad 602. Col. Lomas del Campesino, León 37150, Mexico; eovalle@lasallebajio.edu.mx

* Correspondence: pinales@ugto.mx (J.R.-P.); avina@ugto.mx (J.G.A.-C.)

Abstract

U-shaped architectures have achieved exceptional performance in medical image segmentation. Their aim is to extract features by two symmetrical paths: an encoder and a decoder. We propose a lightweight U-Net incorporating reverse attention and a preprocessing framework for accurate retinal vessel segmentation. This concept could be of benefit to portable or embedded recognition systems with limited resources for real-time operation. Compared to the baseline model (7.7 M parameters), the proposed U-Net model has only 1.9 M parameters and was tested on the DRIVE (Digital Retinal Images for Vesselness Extraction), CHASE (Child Heart and Health Study in England), and HRF (High-Resolution Fundus) datasets for vesselness analysis. The proposed model achieved Dice coefficients and IoU scores of 0.7871 and 0.6318 on the DRIVE dataset, 0.8036 and 0.6910 on the CHASE-DB1 Retinal Vessel Reference dataset, as well as 0.6902 and 0.5270 on the HRF dataset, respectively. Notably, the integration of the reverse attention mechanism contributed to a more accurate delineation of thin and peripheral vessels, which are often undetected by conventional models. The model comprised 1.94 million parameters and 12.21 GFLOPs. Furthermore, during inference, the model achieved a frame rate average of 208 FPS and a latency of 4.81 ms. These findings support the applicability of the proposed model in real-world clinical and mobile healthcare environments where efficiency and Accuracy are essential.



Academic Editor: Eva H. Dulf

Received: 27 May 2025

Revised: 19 June 2025

Accepted: 3 July 2025

Published: 5 July 2025

Citation: Hernandez-Gutierrez, F.D.; Avina-Bravo, E.G.; Ibarra-Manzano, M.A.; Ruiz-Pinales, J.; Ovalle-Magallanes, E.; Avina-Cervantes, J.G. Retinal Vessel Segmentation Based on a Lightweight U-Net and Reverse Attention. *Mathematics* **2025**, *13*, 2203.

<https://doi.org/10.3390/math13132203>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: vesselness; retinopathies; segmentation; reverse attention

MSC: 68U10

1. Introduction

Diabetes mellitus (DM) worldwide will increase to 578 million by 2030 and 700 million by 2045 [1]. The United Nations' Department of Economic and Social Affairs forecasts that the actual population is around 8.19 billion people, and by 2030 and 2045, it is predicted to be 8.55 and 9.47 billion, respectively. Meaning that 8.2% could develop DM. As a consequence of persistently high and uncontrolled levels of DM, Diabetic retinopathy (DR) is developed,

representing a significant and often irreversible consequence of some systemic disorders. It results in damage to the retinal blood vessels, as shown in Figure 1.

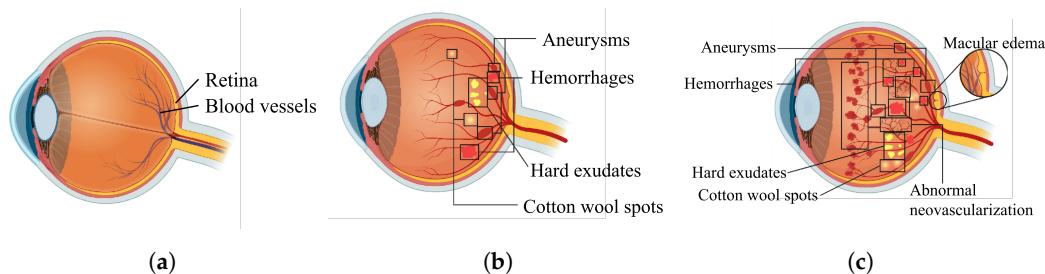


Figure 1. Diabetes-related retinopathy [2], (a) healthy retinal vasculature, (b) non-proliferative diabetes-related retinopathy (NPDR), and (c) proliferative diabetes-related retinopathy (PDR).

The classification metrics used nowadays to assess the DR severity were established in the study called Early Treatment of Diabetic Retinopathy (ETDRS) [3]. Another simplified scale was proposed by Wilkinson et al. [4], named International Clinical Diabetic Retinopathy (ICDR). In this context, Table 1 presents the comparison of these two retinopathy grades. The condition is characterized by a gradual deterioration of the retinal vasculature, which takes charge of supplying oxygen and nutrients to retinal tissues, leading to visual impairment and, in the absence of treatment, blindness [5].

Table 1. Severity Level of non-proliferative diabetes-related retinopathy (NPDR) and proliferative diabetes-related retinopathy (PDR).

| Level | ICDR | ETDRS |
|-------|----------------|------------------------|
| 10 | No retinopathy | No retinopathy |
| 20 | Mild NPDR | Very mild NPDR |
| 35 | Moderate NPDR | Mild NPDR |
| 43 | | Moderate NPDR |
| 47 | | Moderately severe NPDR |
| 53 | Severe NPDR | Severe NPDR |
| 60–61 | PDR | Mild PDR |
| 65 | | Moderate PDR |
| 71–75 | | High-risk PDR |
| 81–85 | | Advanced PDR |

At the same time, leukostasis is another medical condition that can contribute to DR, even in the early stages. It involves the accumulation of white blood cells in the retinal blood vessels, leading to vessel damage and ischemia (lack of oxygen) [6]. Additionally, in methamphetamine (METH) abusers, this condition induces retinal neovascularization [7].

The World Health Organization (WHO) places DR as one of the leading causes of blindness worldwide [8]. It ranks as the fourth leading cause. The syndrome gives rise to various clinical alterations, including the formation of microaneurysms, hemorrhages, exudates, and neovascularization [9]. The early identification of vision loss requires the utilization of regular eye examinations and contemporary imaging methodologies. The treatment options available may encompass laser therapy, intravitreal injections, and surgical procedures, contingent upon the severity and progression of the disease. Also, it is estimated that by 2045, 700 million individuals will be afflicted with diabetic retinopathy [10], which represents 89% of people with DM.

Early detection and diagnosis are crucial against visual impairment and blindness. According to the American Diabetes Association [11], it is recommended to screen for diabetic retinopathy annually for both diabetes types.

The principal study for early detection is fundus image analysis. In particular, fundus imaging is a non-invasive and high-resolution imaging technique for segmenting the retinal vessels, obtaining a contrasted view of the visible surface of the retina [9]. Hence, the ophthalmologist can review the branching angle, the retinal vessel width, and the calibrated curvature.

However, in some cases, the ophthalmologist is unable to segment the retinal vessels because of low-contrast or noisy images. Additionally, such retinal structures are generally variable from patient to patient, and other equally important human factors, such as the specialist's visual acuity, affect image interpretation, as shown in Figure 2. Therefore, this study aims to segment the veins in the fundus of the eye automatically using a low-weight neural network suitable for mobile and real-time applications.

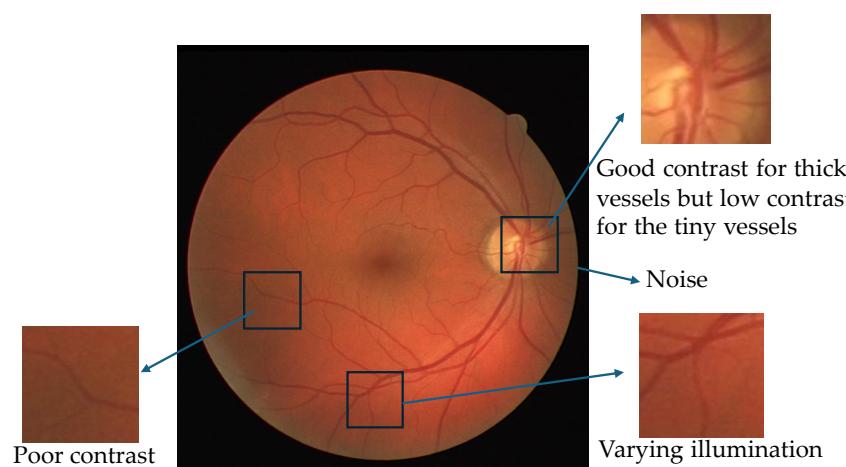


Figure 2. Fundus image presenting varying contrast conditions, emphasizing differences in vessel thickness, noise, and uneven lighting.

Currently, the importance of identifying diabetic retinopathy lies in its correlation with other medical conditions [1,12]. These include an increased risk of developing coronary artery disease, neuropathy, diabetic nephropathy, diabetic foot syndrome, sclerosis, Parkinson's disease, and Alzheimer's disease. Consequently, experts face a significant challenge in accurately delineating the retinal vessels. Under such circumstances, retinal vessel segmentation is vital for human visual health, where early signs of retinopathy, such as changes in vascular morphology like microaneurysms, neovascularization, narrowing, or hemorrhages, could prevent vision loss [13]. For this reason, an automated system is being sought as an assistive tool for specialists focused on the segmentation process. Given the aforementioned considerations, a deep learning system has been chosen, with a specific emphasis on convolutional neural network (CNN) models. Such networks have previously been used in medical image analysis with high success under controlled conditions [14].

Hence, deep learning (DL) models have been employed to recognize complex objects or fine structures in images [15]. These models can automatically enhance the salient features of an image, which are also rotationally invariant. Additionally, such models are supported by convolutional networks that work through small filters operating across the entire image. By finding a pattern similar to the convolutional filter, such networks enhance the detected region in the image [16].

The main contributions of this paper are summarized below:

1. Design a lightweight U-Net for retinal vessel segmentation, which significantly reduces the number of model parameters and computational complexity, which is particularly beneficial for applications involving small medical imaging datasets.

2. Restructure a modified and lightweight U-Net network for biomedical image segmentation over fundus images.
3. Optimize the U-Net by reducing convolutional filters for a lighter and more portable architecture.
4. Append a reverse attention module to the terminal stage of the U-Net architecture to refine the segmentation output further.
5. Evaluate systematically the GELU activation function and the AdamW optimizer, which are predominantly employed in transformer architectures.

2. Related Works

Recent advances in diabetic retinopathy (DR) research have resulted in a substantial paradigm shift in our knowledge of the disease. Contrary to long-held beliefs that DR is essentially a vascular condition, new research reveals a neurodegenerative cause [17]. This unique theory proposes that diabetic retinal neurodegeneration (DRN) not only precedes but may also be the cause of the microvascular alterations previously identified along with DR. DRN is distinguished by neural apoptosis and enhanced glial fibrillary acidic protein (GFAP) expression in retinal cells, which results in functional changes detectable using a variety of diagnostic modalities. Notably, optical coherence tomography (OCT) investigations have shown that diabetes individuals' retinal layers weaken before clinically visible DR develops [18].

Microaneurysms are typically the earliest clinically detectable signs of diabetic retinopathy. Due to the enhanced contrast in the green channel on RGB fundus images, they are significantly prominent [19]. Aslani and Sarnel [20] applied a feature extraction method to segment retinal and non-retinal vessels. They used 17 features extracted from the green channel: 13 Gabor filters, contrast-enhanced intensity, morphological top-hat transformed intensity, vesselness measure, and the B-COSFIRE filter. The feature vector was normalized to have zero mean and a unified standard deviation (i.e., z-score normalization). The green channel is used instead of intensity because its spectral light density is close to the visual system's sensitivity, avoiding the computation of intensity from RGB images. Additionally, they preprocessed each image with contrast-limited adaptive histogram equalization (CLAHE). They also processed the features' vector into a random forest classifier, which was selected because of its speed, simplicity, and information fusion capability. This method was evaluated on the DRIVE and HRF datasets.

Likewise, Aguirre-Ramos et al. [13] used the feature extraction method and a low-pass radius filter to minimize noise in the green channel. Plus, 30 Gabor filters and a Gaussian fractional derivative approach were used at various angle phases to improve the eye's structure and contour delineation. To eliminate false positives, they used a threshold, an effective bi-modal Gaussian-based threshold, field-of-view (FOV) boundary correction, artifact reduction, residual pixel reclassification, and region correction to remove artifacts. Finally, a global threshold technique found blood vessel pixels among non-blood vessel pixels.

Meanwhile, Saha Tchinda et al. [21] proposed classical feature extraction and an artificial neural network as a classifier. They considered the green channel to be the one with the best contrast to distinguish vessels from the background. In this respect, they extracted the following features: four edge detection filters, one Laplacian filter, and three morphological transformations. Subsequently, they employed a fully convolutional neural network (FCNN), a type of artificial neural network that employs a cascade of connected neurons. Additionally, additional connections are present, extending from the input layer to each subsequent layer and from each layer to all subsequent layers.

With the emergence of CNNs, feature extraction became automatic. This led to more complex models for segmenting regions. Thus, U-Net networks gave way to image segmentation, which was done with the help of their encoder-decoder. Firstly, in the encoder, they extract features from the image and reduce the image down to a latent space. After that, in the decoder, they expand the feature maps to the original size, reconstructing spatial information.

Ren et al. [22] adapted a U-Net architecture to the DRIVE dataset. They proposed a U-Net network that replaces the skip connection with a bi-directional feature pyramid network (BiFPN). Each image was divided into patches, each measuring 48×48 pixels; next, the patches fed the U-Net network. Considering preprocessing, grayscale conversion was applied to highlight the green channel, with CLAHE subsequently applied to each image.

Similarly, Liu et al. [23] designed a new U-Net architecture using a combination of U-Net, SegNet, and HardNet architectures. They denoted that structure as TP-UNET. To improve the segmentation contours of the tiny vessels, they considered the synergy of these structures in the segmentation of the DRIVE and CHASE datasets.

Recently, Ding et al. [24] presented a U-Net, RCAR-UNET, based on the channel attention mechanism and rough neurons. They used rough neurons due to the uncertainty of the retinal vessels. Shortly after, Chu et al. [25] implemented a dual-layer nested U-Net, known as the U2Net model. The U2-Net architecture features a hierarchical cascade of recursively nested U-Net subnetworks. Each U-Net module embeds another U-Net within itself, enabling multiscale feature extraction across progressively refined representations. The model employs the same skip connection to extract features at different levels. Hence, they introduced a dual attention mechanism in each skip connection. However, they indicated that excess information may cause the network to overlook relevant details.

Zhang et al. [26] proposed a lightweight network, the multiscale feature refinement network (LMFR-Net), which utilizes a dual-decoding structure to enhance retinal vessel segmentation. Additionally, an enhanced or inception convolution block (ICB) was introduced to efficiently extract fundamental features. Kande et al. [27] introduced a U-Net-based model that integrates a multiscale feature extractor within the encoder and decoder. Rather than relying on a conventional skip connection in the bottleneck, they enhanced the architecture by including squeeze-and-excitation and spatial attention blocks. Furthermore, the input images underwent morphological opening (MO), CLAHE, and shade correction as a preprocessing step before being fed into the network.

Luo et al. [28] presented a hybrid model by integrating a transformer architecture. They introduced the lightweight parallel transformer (LPT), which was designed to effectively capture long-range dependencies, thereby preventing the disruption of slender retinal vessels. Additionally, they developed an adaptive vascular feature fusion module aimed at improving the recognition of microvessels. Jian et al. [29] introduced an improved U-Net based on dual attention, DAU-NET. The enhancement was localized in each level of the U-Net, using a transformer-like model as a backbone. For this purpose, the skip connection was replaced with local-global attention (LGA) and cross-fusion attention (CFA). They considered that LGA would accentuate vessel-related features while suppressing irrelevant background information, while CFA addresses potential information loss during feature extraction and interaction between the encoder and decoder paths.

In brief, Table 2 presents an overview of the most relevant methodologies and recent hybrid models used in the retinal fundus image segmentation literature.

Table 2. Datasets and methods used in the literature for retinal vessel segmentation.

| Article | Dataset | Method | Year |
|---------------------------|--|---|------|
| Aguirre-Ramos et al. [13] | DRIVE | Gabor filters, fractional derivatives, and expectation maximization | 2018 |
| Saha Tchinda et al. [21] | DRIVE, CHASE, and STARE | Edge detection filters | 2021 |
| Ren et al. [22] | DRIVE | Bi-FPN network | 2022 |
| Liu et al. [23] | DRIVE and CHASE | TP-UNET | 2023 |
| Ding et al. [24] | DRIVE, STARE, and CHASE | RCAR-UNET | 2024 |
| Chu et al. [25] | DRIVE, CHASE, and HRF | U2Net | 2025 |
| Zhang et al. [26] | DRIVE, CHASE, and STARE | LMFR-Net | 2025 |
| Kande et al. [27] | DRIVE, STARE, CHASE, HRF, and DR HAGIS | MSMA Net | 2025 |
| Luo et al. [28] | DRIVE, CHASE, STARE, and HRF | PA-Net | 2025 |
| Jian et al. [29] | DRIVE, CHASE, and STARE | DAUNET | 2025 |

3. Mathematical Foundations

This section presents the mathematical background used to develop the proposed framework.

3.1. Convolutional Layer

A convolutional layer in a U-Net model receives an input feature map X and uses a set of learnable filters W and biases b , producing an output feature map Y . The mathematical operation for a single output channel can be expressed as

$$Y_{i,j,c} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{d=0}^{D-1} W_{m,n,d,c} \cdot X_{i+m, j+n, d} + b_c, \quad (1)$$

where $X_{i+m, j+n, d}$ represents the input feature at position $(i+m, j+n)$ and depth d , $W_{m,n,d,c}$ is the convolutional kernel of size (M, N, D) , b_c is the bias associated with the output channel c , and $Y_{i,j,c}$ denotes the resulting output at location (i, j) in channel c .

Figure 3 shows how each filter is designed to emphasize distinct features of the image throughout the convolution process. By selectively enhancing specific patterns, textures, or edges, these filters enable the model to extract meaningful representations, contributing to the overall feature learning and hierarchical abstraction within the network.

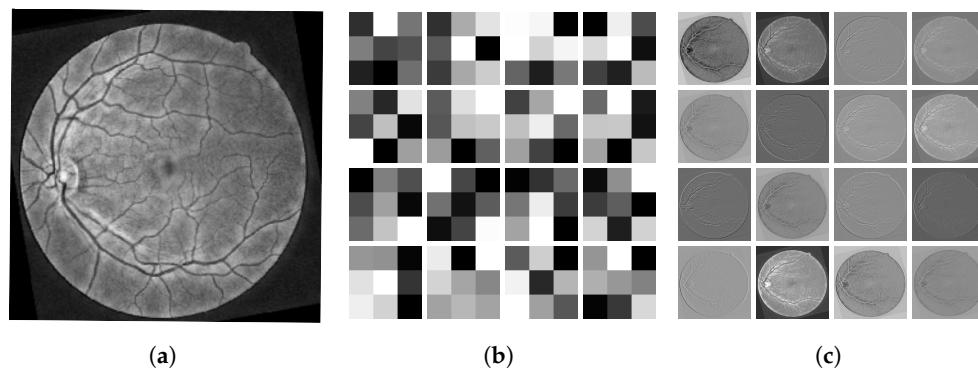


Figure 3. Illustration of a convolutional layer. (a) Input image, (b) 16 kernels, and (c) 16 features maps.

3.2. Data Augmentation

Given the dataset's characteristics, data augmentation plays a crucial role in training deep learning models, especially when dealing with limited image availability. Notably, this approach has been proven to be highly effective, even for large-scale datasets, as it significantly improves model performance and generalization capabilities [30].

Common data augmentation techniques in fundus imaging include rotation, shearing, flipping, and translation [31]. The main linear transformations of data augmentation are presented below. The first includes a rotation matrix, defined as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (2)$$

where (x, y) are the original coordinates, (x', y') are the transformed coordinates, and θ is the rotation angle.

Shearing introduces a distortion in either the horizontal or vertical direction. In the case of horizontal shearing, the transformation is given by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & s_x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (3)$$

where s_x is the shear factor along the horizontal axis. Similarly, vertical shearing is described by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ s_y & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (4)$$

where s_y represents the vertical shear factor.

Flipping reverses the image along a specific axis. The transformation for horizontal flipping is defined by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (5)$$

while vertical flipping follows the transformation matrix given by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (6)$$

Similarly, in digital image segmentation, data augmentation techniques are applied to the corresponding mask images to ensure consistency and improve the model's generalization performance. Figure 4 presents a visualization of four sample images from the DRIVE dataset, each corresponding to the green channel in the original image.



Figure 4. Cont.

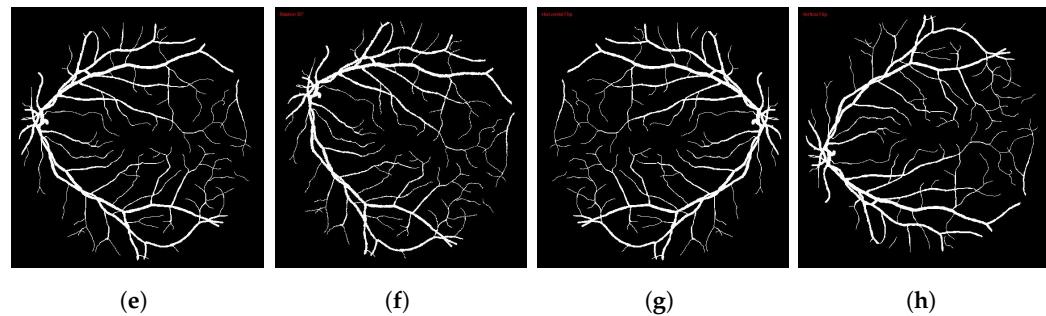


Figure 4. (a) Original image (green channel), (b) image rotated 30° , (c) image horizontally flipped, (d) image vertically flipped, (e) image mask, (f) mask rotated 30° , (g) mask horizontally flipped, and (h) mask vertically flipped.

3.3. Function Loss

In medical imaging, specific datasets present a significant challenge, as the mask regions to be segmented are substantially smaller compared to the overall image background [32]. This issue, commonly known as class imbalance, can adversely impact the model's learning process, leading to biased predictions and suboptimal segmentation performance. Two main loss functions are commonly utilized in semantic segmentation, particularly in binary segmentation. The first and most widely adopted is Binary Cross-Entropy (BCE), which measures the pixel-wise discrepancy between the predicted and ground-truth segmentation masks [33]. BCE is particularly effective when dealing with balanced datasets; however, its performance can degrade in scenarios characterized by a significant class imbalance, as it treats each pixel independently without accounting for the relative proportions of foreground and background regions. The BCE is formally defined by

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (7)$$

where N represents the total number of samples, y_i denotes the ground-truth label, and \hat{y}_i is the predicted probability of the corresponding sample belonging to the positive class.

The Dice index, also referred to as the Dice Similarity Coefficient (DSC) when employed for image segmentation, is the most widely utilized metric for evaluating segmentation in class imbalanced datasets [34]. The Dice loss \mathcal{L}_{DSC} can be defined as

$$\mathcal{L}_{\text{Dice}} = 1 - 2 \frac{\sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i}, \quad (8)$$

where N represents the total number of pixels in the image, y_i denotes the ground-truth label, and \hat{y}_i corresponds to the predicted probability for each pixel.

Additionally, this study proposed and evaluated a hybrid loss function that linearly combines Binary Cross-Entropy and Dice loss functions; it is formally described by

$$\mathcal{L}_{\text{Hybrid}} = \alpha \mathcal{L}_{\text{BCE}} + (1 - \alpha) \mathcal{L}_{\text{Dice}}, \quad (9)$$

where $\alpha = 0.5$.

3.4. Inverse Gamma Correction

Gamma correction is a non-linear transformation applied to enhance image luminance in a manner that aligns with human visual perception. This technique is particularly advantageous in retinal fundus imaging in highlighting minute structures such as blood vessels and pathological features that may be obscured in image regions characterized by low light intensity. Thus, Gamma correction involves redistributing pixel values, amplifying

or reducing perceptual differences in these regions without significantly affecting brighter areas. This is particularly significant in medical imaging, where preserving subtle details is crucial for accurate diagnosis. Inverse gamma correction is defined by

$$I_{out} = I_{in}^{\frac{1}{\gamma}}, \quad (10)$$

where I_{in} and I_{out} are the input and output intensity values (normalized between 0 and 1) and γ is the gamma exponent. Values of $\gamma < 1$ enhance dark areas, and $\gamma > 1$ brightens the image, as shown in Figure 5.

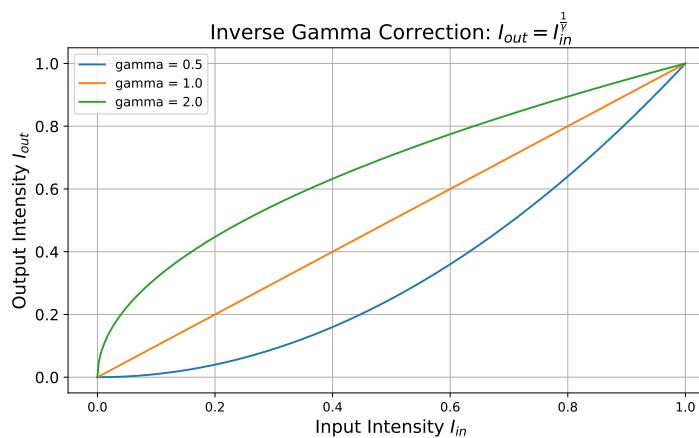


Figure 5. Inverse gamma correction curves for different values of γ .

3.5. Adaptive Histogram Equalization

Due to contrast variability in the image, low contrast in an overall image can cause the extraction of thin vessels to disappear. Therefore, adaptive local enhancement was used to enhance small regions of the image, as shown in Figure 6. For a given pixel intensity $I(x, y)$, the CLAHE-corrected intensity is obtained as follows:

$$I_{CLAHE}(x, y) = \frac{CDF_{clip}(I(x, y)) - CDF_{clip}(I_{min})}{CDF_{clip}(I_{max}) - CDF_{clip}(I_{min})} \times (I_{max} - I_{min}) + I_{min}, \quad (11)$$

where $CDF_{clip}(\cdot)$ is the clipped cumulative distribution function of the local histogram, and I_{min} and I_{max} represent the minimum and maximum intensities within the corresponding local region.

In this formulation, the histogram is first clipped at a specific threshold to control the contribution of any intensity values that might otherwise dominate the enhancement process. After that, the clipped CDF value at $I(x, y)$ is normalized by subtracting $CDF_{clip}(I_{min})$ and dividing by the dynamic range of the clipped CDF, namely $CDF_{clip}(I_{max}) - CDF_{clip}(I_{min})$. This normalization step rescales the pixel's intensity to the interval $[0, 1]$.

The expression is then scaled by $(I_{max} - I_{min})$ so that it spans the target intensity range and finally shifted by I_{min} to ensure that the output intensities remain valid within the interval $[I_{min}, I_{max}]$. By combining histogram clipping and local neighborhood processing, CLAHE avoids excessive contrast stretching and noise amplification, providing a more balanced enhancement compared to traditional global histogram equalization. Figure 6 shows the substantial improvement in the contrast and visibility of small structures in retinal fundus images.

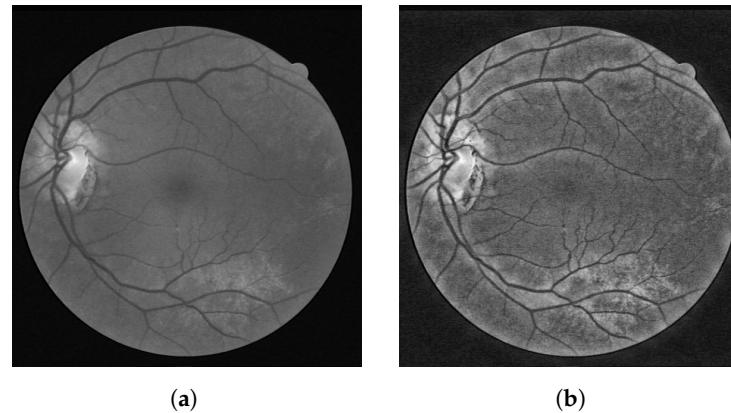


Figure 6. (a) Original fundus image (green channel) and (b) fundus image processed with the CLAHE method to enhance contrast and the small regions' visibility.

3.6. AdamW Optimizer

In recent research, AdamW [35] has emerged as a popular optimizer. It achieves rapid convergence in deep learning models and demonstrates superior performance compared to the conventional stochastic gradient descent (SGD) method, which uses a single learning rate for all gradient coordinates. The AdamW optimization process is ruled by the following equations:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t}, \\ \theta_t &= \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} (\hat{m}_t + \lambda \theta_{t-1}), \end{aligned} \quad (12)$$

where m_t and v_t denote the first- and second-moment estimates of the gradient g_t , respectively; β_1 and β_2 are the decay rates for these moments; \hat{m}_t and \hat{v}_t are their bias-corrected versions; α is the learning rate; ϵ is a small constant ensuring numerical stability; and λ corresponds to the weight-decay term. Unlike the original Adam optimizer with L2 regularization, AdamW explicitly incorporates weight decay in the parameter update, leading to more robust performance by decoupling the effect of weight decay from the gradient-based update.

3.7. Gaussian Error Linear Units (GELUs)

One of the GELU's main features is its smoothing on the network weights, its differentiability, and its ability to approximate the ReLU function [36]. As Figure 7 demonstrates, the GELU activation curve exhibits a smooth transition near zero.

GELU is a variant of the ReLU activation function and has been used in transformer models. The GELU function is defined as

$$\text{GELU}(x) = x \Phi(x), \quad (13)$$

where $\Phi(z)$ is the cumulative distribution function (CDF) of the standard normal distribution,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt. \quad (14)$$

Additionally, using the derivative rule for the product $x\Phi(x)$, a quite interesting property is found:

$$\frac{d}{dx} [x \Phi(x)] = \Phi(x) + \frac{x}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \quad (15)$$

Using this mathematical perspective, Lee [36] provided an alternative approximation of the GELU function:

$$\text{GELU}(x) \approx 0.5 \left(1 - \tanh \left(\sqrt{\frac{2}{\pi}} (x + 0.0444715 x^3) \right) \right) x. \quad (16)$$

Other activation functions, such as tanh and sigmoid, are smooth and differentiable; however, they frequently saturate for large positive or negative inputs, resulting in vanishing gradients. The GELU offers a middle ground with a gentler gradient behavior, making it an appealing choice in modern deep learning models.

All in all, the GELU activation function combines the best aspects of ReLU-like gating, preserving positive values and dampening negatives with a smoother and fully differentiable transition near zero.

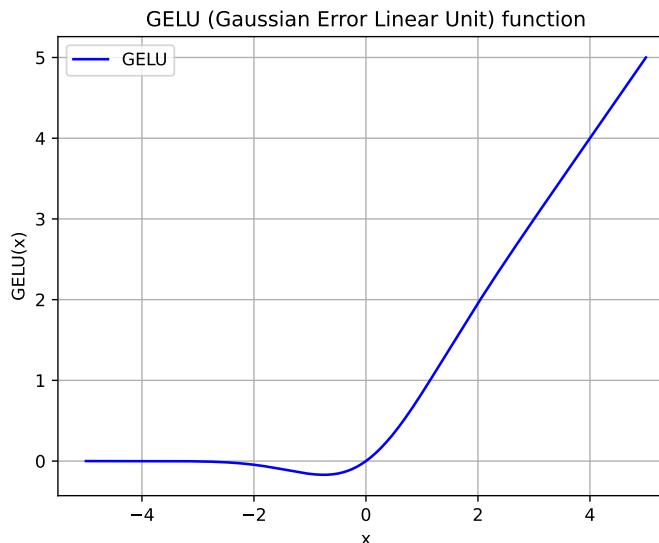


Figure 7. Gaussian error linear unit (GELU) function.

3.8. Reverse Attention

Using dropout or pooling layers might reduce the spatial resolution, complicating the detection and segmentation of small-scale regions [37]. Reverse attention seeks to overcome this fundamental challenge [38].

Figure 8 shows the internal architecture of the reverse attention (RA) module. In this module, the input is a feature map, and the sigmoid function is subsequently applied, yielding an attention mask. Then, the inverse is applied to this attention mask, thereby inverting the pixel values. Thus, the model focuses on previously missed or poorly segmented regions.

The main operators in the RA module are mathematically defined by

$$\begin{aligned} M_k^{RA} &= s(up(F_{k+1})), \\ F_k^{RA} &= \tilde{M}_k^{RA} \otimes F_k, \\ \tilde{M}_k^{RA} &= \mathbf{1} - M_k^{RA}, \end{aligned} \quad (17)$$

where $s(\cdot)$ is a sigmoid activation function and \otimes is element-wise multiplication. M_k^{RA} represents the foreground probability map, and its binary complement \tilde{M}_k^{RA} redirects the

network's attention toward the less-confident regions. Here, the upsampling function $up(\cdot)$ is used to match the resolution of the deep-level feature map F_{k+1} to the shallow-level resolution F_k . The proposed reverse attention module, configured with 16 input channels, comprises 4672 trainable parameters.

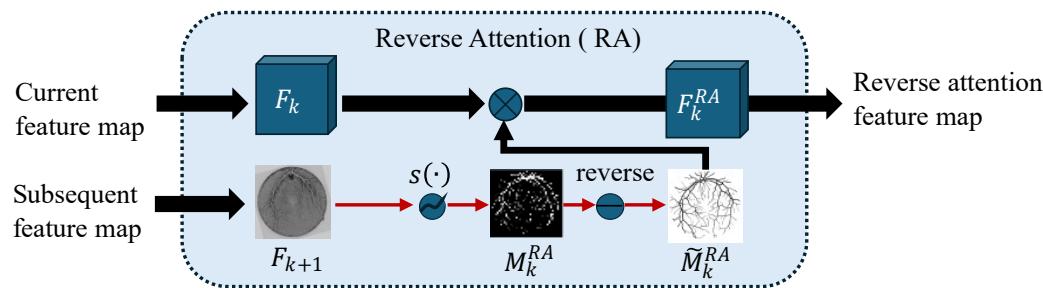


Figure 8. Reverse attention module [39]. $s(\cdot)$ is the sigmoid function. \otimes represents element-wise multiplication. F_k represents the feature map at stage k , and F_{k+1} is the feature map subsequently generated at stage $k + 1$. F_k^{RA} is the reverse-attention-filtered feature map.

3.9. U-Net Architecture

The U-Net [40] architecture constitutes the baseline for image segmentation, otherwise referred to as vanilla U-Net. Modifications have been made to this baseline architecture for the most specific task. The two-path model, comprising an encoder and a decoder, constitutes the primary functionality of the architecture.

On the one hand, the encoder path performs feature extraction across five consecutive stages. Two 2D convolution blocks perform feature extraction in each stage, with batch normalization applied after every convolution. Subsequently, a downsampling step halves the spatial dimensions of the resulting feature maps. Following the encoder's final stage is a layer commonly called the “bottleneck.” This depth layer encapsulates the extracted features at their highest level of abstraction, providing a compact yet information-rich representation before subsequent upsampling and refinement.

After the bottleneck, the decoder path comprises five stages that expand in a manner analogous to the encoder. Each stage incorporates a 2D convolution, with the notable difference that every stage is interconnected at the same level via skip connections, thereby preserving essential feature information. Meanwhile, the decoder performs an upsampling operation at the end of each stage, progressively restoring the spatial dimensions until they match the original input size of the encoder.

4. Materials and Methods

4.1. Datasets

The following databases were used in this study: The Digital Retinal Images for Vessel Extraction (DRIVE), CHASE_DB1, and High-Resolution Fundus (HRF) datasets. On the one hand, DRIVE consists of 40 color fundus images divided into 20 training and 20 test images, each with its segmented image (ground truth). On the other hand, the HRF dataset consists of 45 equally sized (3304×2336) color fundus images with their ground truth provided. Finally, CHASE_DB1 contains 28 color retinal images of 999×960 , collected from the left and right eyes of 14 school children. Table 3 presents a concise overview of the datasets employed in this study.

Table 3. A comprehensive comparative analysis thoroughly examines each database's structural framework and its respective image dimensions.

| Dataset | Number of Images | Image Size | Training [%]–Testing [%] |
|----------------|------------------|-------------|--------------------------|
| DRIVE [41] | 40 | 565 × 584 | 70–30 |
| HRF [42] | 20 | 3304 × 2336 | 70–30 |
| CHASE_DB1 [43] | 28 | 999 × 960 | 70–30 |

4.2. Overall Framework

Figure 9 shows the functional blocks used to implement the proposed model in the images from the DRIVE, CHASE, and HRF datasets. Such images are preprocessed for better image conditioning to continue with data normalization and data augmentation. After that, 70% of the data were randomly chosen for training, and the remainder 30% were used for validation. Finally, the lightweight U-Net focused on retinal vessel segmentation was tested and evaluated.

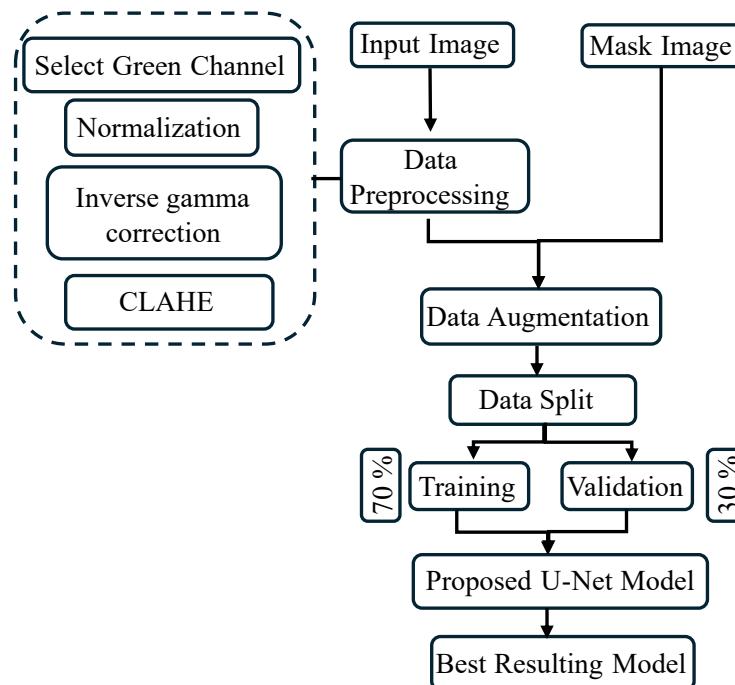
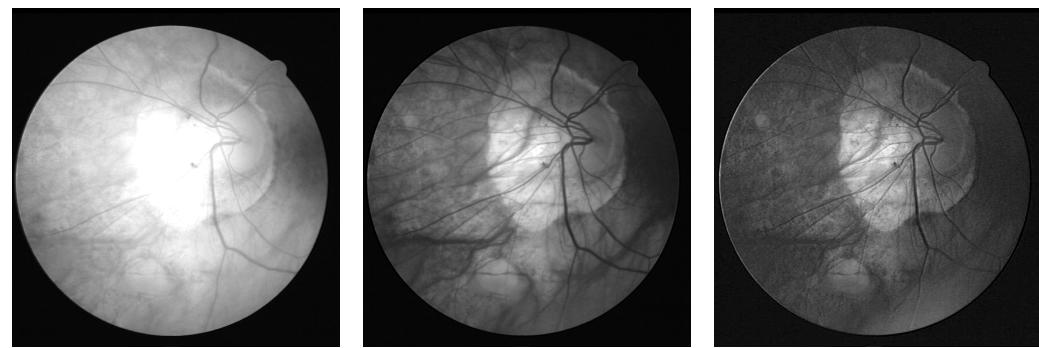


Figure 9. General description of the proposed method. The dashed line represents a sub-process in the data preprocessing block.

4.3. Preprocessing

First, the color image was processed. This was performed for each channel of the image, resulting in the best values for the metrics for the green channel, as indicated by the state of the art [44]. Figure 10 shows a visual comparison of each RGB channel of the original fundus image. Once the green channel was selected, adaptive histogram equalization was performed.

Subsequently, each image was resized to 512×512 pixels using bilinear interpolation to ensure uniform dimensions across the dataset, despite slight distortion. This resizing operation standardizes all images to the exact resolution, simplifying further processing and analysis steps, reducing computational complexity, and making comparing results across different models or techniques easier.



(a) Red channel.

(b) Green channel.

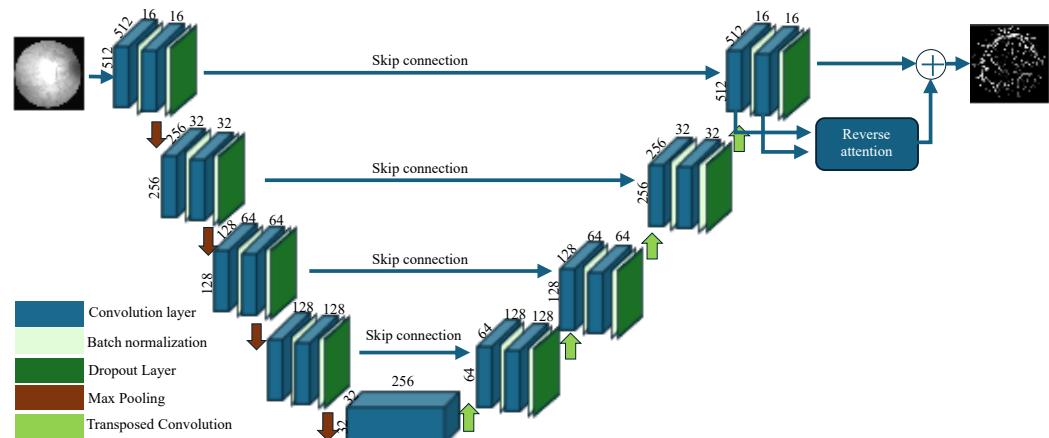
(c) Blue channel.

Figure 10. Visualization of three sample images from the DRIVE dataset, each representing a different color channel from the RGB space.

4.4. Proposed Model

Figure 11 shows the proposed model. The model architecture comprises five levels of convolutional blocks. In the first four levels, each block consists of two convolutional modules. Following each convolutional operation, batch normalization is applied to maintain a stable distribution of filter weights, thereby improving training stability and convergence.

At the end of each level, a max-pooling operation with a 2×2 kernel is employed to progressively downsample the feature maps, reducing spatial dimensions while preserving essential feature representations. Additionally, a dropout layer is incorporated between each convolutional pair and its corresponding batch normalization to enhance generalization and mitigate overfitting. The number of convolutional filters doubles at each level, starting with 16 and increasing sequentially to 32, 64, 128, and ultimately 256 filters, allowing the network to capture increasingly complex hierarchical features.

**Figure 11.** U-Net network architecture modified from the original.

At the output of the U-Net network, reverse attention was placed to refine the segmentation by focusing on the regions that the model predicted incorrectly. This is because the network's output contains the feature maps at the final resolution; otherwise, Accuracy could be lost on lower-resolution feature maps. The code is available at <https://github.com/fdfernandezgutierrez/RVS> (accessed on 18 June 2025).

Table 4 compares the baseline U-Net with the proposed lightweight variant in terms of total and trainable parameters, thereby quantifying the parameter reduction achieved by the new architecture.

Table 4. Comparison table of hyperparameters between the baseline and the proposed lightweight U-Net model.

| Network Architecture | Total Parameters | Trainable Parameters |
|-------------------------------------|------------------|----------------------|
| Baseline U-Net (32 initial filters) | 7,771,681 | 7,765,601 |
| Proposed lightweight U-Net | 1,946,897 | 1,943,857 |

4.5. Implementation Details

The proposed model was implemented on a high-performance workstation with an Intel Core i7 processor, 32 GB of RAM, and an NVIDIA RTX3070Ti GPU featuring 8 GB of VRAM. All training and evaluation procedures were conducted within the PyTorch (version 2.5.1) framework. During the training process, 32-bit floating point precision (FP32) was used for all operations.

The dataset underwent comprehensive preprocessing to facilitate robust segmentation, including gamma correction and contrast-limited adaptive histogram equalization (CLAHE). By employing a gamma value of 1.2 and a CLAHE clip limit of 5.0 with a tile grid size of 32×32 , these methods effectively enhanced the visibility of delicate vascular structures, thereby improving the model's ability to capture subtle features.

The experimental process employed the AdamW optimizer with a learning rate of 0.001 and a weight decay of 0.001 to enhance convergence stability and mitigate overfitting. The model was trained for 1000 epochs, ensuring comprehensive optimization. The images were processed with a batch size of 4. To avoid overfitting, the best model was saved based on its performance on the validation set, and an early stopping criterion with a patience value of 20 epochs was applied to terminate training once validation performance ceased to improve, thereby mitigating the risk of overfitting.

The experiments were conducted to evaluate the performance of several loss functions in semantic segmentation. The findings of this analysis indicate that the Dice Loss proved to be the most effective in improving the Intersection over Union (IoU) metric, demonstrating superior performance in scenarios with class imbalance.

5. Numerical Results

5.1. Evaluations Metrics

The proposed lightweight U-Net model is evaluated using several binary classification metrics, including the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Sensitivity, Specificity, and Accuracy (Acc). These metrics are fundamental to understanding the model's behavior and reliability regarding the databases analyzed and the learning architecture's generalization. It is worth noting that a single unique metric is insufficient to evaluate recognition systems, especially in the integral validation step or in the use of imbalanced datasets.

5.1.1. Dice Similarity Coefficient

The Dice Similarity Coefficient (DSC) is a well-liked similarity binary metric used to quantify the overlap between two sets of data, particularly in image segmentation [45]. Its value ranges from 0 to 1, where the value 1 corresponds to the best or perfect segmentation. DSC is formally defined as

$$DSC = 2 \sum \frac{A \cap B}{A \cup B}, \quad (18)$$

where A represents the predicted segmentation image and B is the corresponding manually delineated ground truth. In highly imbalanced classes, this metric could be biased.

5.1.2. Intersection over Union

Intersection over Union (IoU) is a binary metric widely used in object detection and image segmentation [46]. It measures the degree of overlap between the predicted segmentation and the ground-truth mask. This metric is also scale-invariant [47], which is clinically relevant, particularly in tumor and anatomical human structure detection for early treatment. It is formally defined as

$$\text{Mean IoU} = \sum \frac{A \cap B}{A \cup B}. \quad (19)$$

Here, A corresponds to the ground-truth image, while B corresponds to the predicted segmentation. The intersection of these two images is represented by $A \cap B$, while the union of these images is represented by $A \cup B$.

5.1.3. Sensitivity

Sensitivity (the recall score or true positive rate) is a metric that quantifies the ability of a model to predict true positives. In image segmentation, Sensitivity measures the ability of a model to predict the true positive pixels or regions [48]. Therefore, a highly sensitivity coefficient minimizes the appearance of false negatives. Such a metric is defined as

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad (20)$$

This metric is also seen as the ratio of the number of true positives over all sick positive individuals.

5.1.4. Specificity

Specificity (or the true negative rate) is a metric used to assess the ability of a model to predict true negatives [48]. It is defined by

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (21)$$

where TN is the number of pixels correctly classified as negative and FP is the number of pixels misclassified as positive. This ratio is also seen as the number of true negatives over all healthy or negative elements in the population.

5.2. Accuracy

Accuracy is a machine learning metric that quantifies the percentage of correct predictions (e.g., correctly classified pixels) made by a model. The metric tends to be misleading, especially in unbalanced databases, inducing inaccurate model conclusions when used alone. Mathematically, it is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP}, \quad (22)$$

where TP represents the number of pixels correctly classified as positive and TN is the number of pixels correctly classified as negative. Conversely, FP is the number of pixels misclassified as positive, and FN is the number of pixels misclassified as negative.

5.3. Segmentation Results

This section provides a comprehensive visual comparison of segmentation results across the DRIVE, CHASE, and HRF datasets obtained using both the baseline and proposed models. Additionally, a detailed analysis of each figure highlights the qualitative differences between the methods. Finally, an ablation study is included to validate the

impact of the proposed improvements and confirm the optimal model configuration. Figure 12 displays the segmented fundus images for the DRIVE dataset.

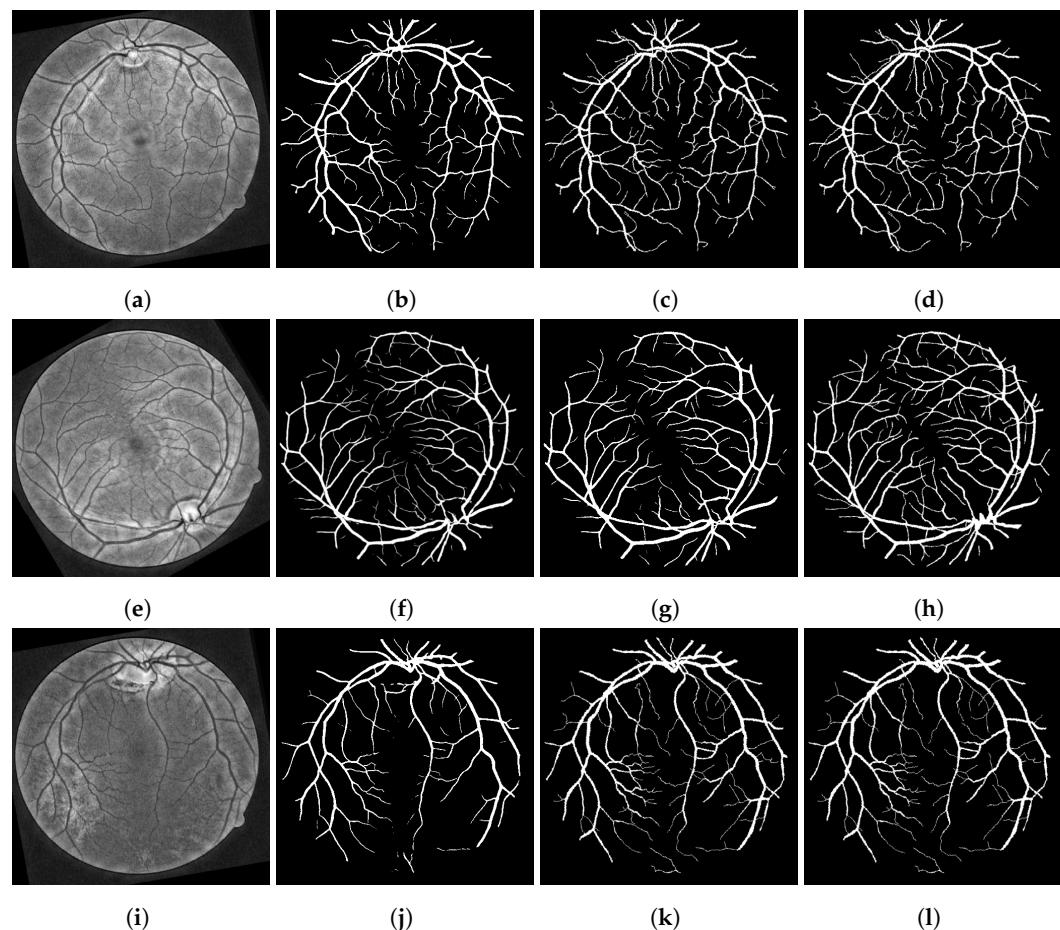


Figure 12. Example of three DRIVE fundus images. The first column (a,e,i) represents the preprocessed images. The second column (b,f,j) is the baseline U-Net’s results. The third column (c,g,k) shows the improved lightweight U-Net results. The final column (d,h,l) presents the ground truth.

The images in the first column (Figure 12a,e,i) belong to the input images; these are the preprocessed images that previously underwent data augmentation in the green channel and applied gamma correction and CLAHE. The next column, column 2 (Figure 12b,f,j), belongs to the images segmented by the baseline U-Net model; in the third column (Figure 12c,g,k), the images are segmented by the improved lightweight U-Net model. The last column (Figure 12d,h,l) shows the images belonging to the ground truth.

Figure 13 illustrates the model’s performance on the DRIVE dataset using a boxplot, providing insights into the experiment’s repeatability of five times. Similarly, a 5-fold cross-validation approach was used. The metrics displayed are the Dice Similarity Coefficient, Intersection over Union, Accuracy, Sensitivity, and Specificity. These have been abbreviated to DSC, mIoU, Acc, Sen, and Spec, respectively. Figure 13a presents the box plot corresponding to the modified U-Net architecture, whereas Figure 13b depicts the baseline U-Net. The results indicate that the modified model shown in Figure 13a outperforms the baseline across all evaluated metrics.

The comparison between the modified U-Net model and the baseline U-Net model demonstrates notable improvements across several key performance metrics, indicating the positive impact of the modifications. The modified U-Net exhibits a substantial increase in the DSC, with a 95% confidence interval ranging from 0.771 to 0.775, compared to the baseline model, which has a lower DSC interval of 0.737 to 0.741. Similarly, the mIoU for

the modified model ranges from 0.627 to 0.633. At the same time, the baseline exhibits a lower mIoU range of 0.583 to 0.589, reflecting an enhanced ability of the modified U-Net to generate more accurate segmentations.

The sensitivity of the modified model, with a confidence interval of 0.797–0.821, is slightly lower than that of the baseline, which ranges from 0.866 to 0.885, indicating that the baseline model is more sensitive in detecting true positives. However, the specificity of the modified U-Net, with a confidence interval of 0.974 to 0.978, surpasses the baseline, which ranges from 0.952 to 0.955, indicating better performance in correctly identifying true negatives. Additionally, the accuracy of the modified model, with a confidence interval ranging from 0.911 to 0.913, is slightly higher than the baseline's accuracy range of 0.894 to 0.895, indicating improved overall performance.

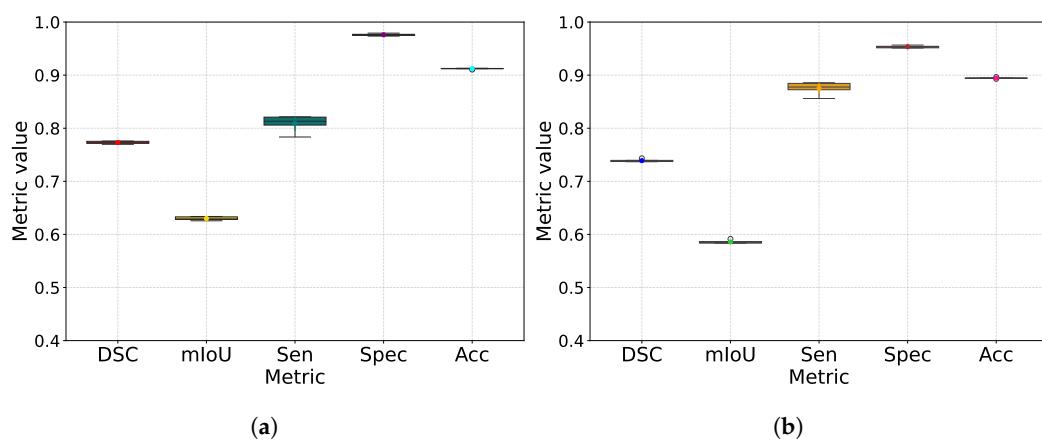


Figure 13. Using the improved lightweight U-Net on the DRIVE boxplot distribution dataset resulted in an increase in key performance metrics when considering a 5-fold cross-validation approach with five replicates: (a) improved lightweight U-Net and (b) baseline U-Net.

This improvement is particularly enunciated in the mean Intersection over Union (mIoU), which measures the overlap between the predicted segmentation and the ground-truth mask, and in Sensitivity, which is defined as the proportion of correctly detected vessels. Table 5 compares the retinal vessel segmentation methods on the DRIVE dataset. While specific studies avoid one or more metrics, each approach demonstrates distinct strengths: high Sensitivity in Li et al. [49] or high DSC in Kande et al. [27]. In this light, the improved lightweight U-Net stands out with a balanced performance, with a DSC of 0.7871, an mIoU of 0.6318, a Sensitivity of 0.7421, a Specificity of 0.9837, and an Accuracy of 0.9113, garnering its robustness in correctly identifying vessel pixels while minimizing false positives. The findings underscore the significance of multi-metric assessments in comprehensively evaluating segmentation quality.

Table 5. Performance comparison of vesselness segmentation using the DRIVE dataset.

| Method | Year | DSC | mIoU | Sensitivity | Specificity | Accuracy |
|----------------------------|------|--------|--------|-------------|-------------|----------|
| Saha Tchinda et al. [21] | 2021 | - | - | 0.7352 | 0.9775 | 0.9480 |
| Li et al. [49] | 2021 | - | - | 0.9896 | 0.7931 | 0.9698 |
| Toptaş and Hanbay [50] | 2021 | 0.7609 | 0.6148 | 0.8400 | 0.9716 | 0.9618 |
| Aurangzeb et al. [51] | 2022 | - | - | 0.8491 | 0.9774 | 0.9659 |
| Liu et al. [23] | 2023 | 0.8291 | - | 0.8184 | 0.9773 | 0.9571 |
| Ding et al. [24] | 2024 | - | 0.6732 | 0.7487 | 0.9836 | 0.9537 |
| Kande et al. [27] | 2025 | 0.8692 | - | 0.8792 | 0.9886 | 0.9827 |
| Proposed lightweight U-Net | 2025 | 0.7871 | 0.6318 | 0.7421 | 0.9837 | 0.9113 |

Figure 14 shows the images segmented from the CHASE dataset using the baseline U-Net and the lightweight U-Net. The images within the initial column are the input images that have undergone preprocessing. The subsequent columns present the images that have

been segmented by the lightweight U-Net and the baseline U-Net, respectively. The final column displays the images that serve as the ground truth.

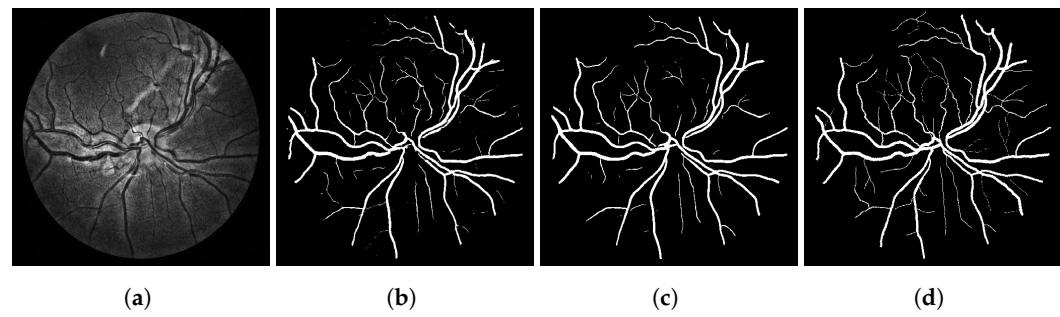


Figure 14. (a) Input image, (b) baseline U-Net, (c) improved lightweight U-Net, and (d) ground truth.

Figure 15 presents the metric for each model, which displays the box plot using 5-fold cross-validation and demonstrates its behavior. Figure 15a shows the metrics that demonstrate relatively tight clustering, with DSC and mIoU values ranging from 0.79 to 0.80 and from 0.65 to 0.67, respectively, as well as uniformly high Specificity (0.98 to 0.99) and Accuracy (0.97 to 0.98). This consistent performance suggests that the lightweight U-Net achieves a favorable trade-off between precision (0.78–0.80) and Sensitivity (0.79–0.83) without over-segmentation. In contrast, Figure 15b exhibits a marginally lower DSC (0.76–0.78) and mIoU (0.62–0.64), accompanied by significantly lower Specificity (0.76–0.78) and Accuracy (0.68–0.72). However, the baseline U-Net model has high Sensitivity (0.97–0.98) and competitive precision (0.85–0.87). The aforementioned patterns serve to emphasize the fundamental trade-off inherent in segmentation tasks. Specifically, while baseline U-Net's high recall capacity can result in over-segmentation (and consequently lower Specificity), the lightweight U-Net attains a more balanced and stable performance across all metrics.

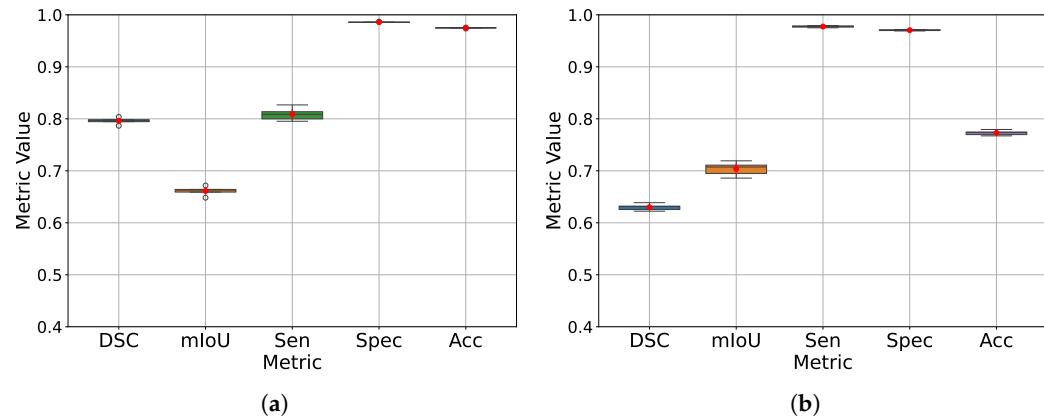


Figure 15. Box plot of the model replicates obtained using a 5-fold cross-validation approach with the CHASE dataset and five replicates (a): improved lightweight U-Net and (b) baseline U-Net.

The comparison between the modified U-Net model and the baseline U-Net model in Figure 15 reveals significant improvements in key performance metrics, highlighting the effectiveness of the implemented modifications. The modified U-Net exhibits a marked increase in the DSC, with a 95% confidence interval ranging from 0.791 to 0.801, indicating superior segmentation Accuracy compared to the baseline, which has a lower DSC interval of 0.625 to 0.635. Additionally, the mIoU for the modified model ranges from 0.654 to 0.668. In contrast, the baseline exhibits a higher mIoU, ranging from 0.693 to 0.714, demonstrating that the modified model achieves better overall segmentation quality.

The sensitivity of the modified model, with a confidence interval between 0.799 and 0.818, is slightly lower than that of the baseline (ranging from 0.976 to 0.979), suggesting that the baseline model is more sensitive to true positives. However, the Specificity and Accuracy of the modified model show remarkable improvement, with Specificity ranging from 0.985 to 0.986 and Accuracy from 0.974 to 0.975, compared to the baseline's Specificity (between 0.970 and 0.971) and Accuracy (between 0.769 and 0.777). These results demonstrate that the modifications have significantly enhanced the model's overall robustness, resulting in improved generalization and performance in segmentation tasks.

Table 6 presents the proposed improved lightweight U-Net model, which achieves the highest mIoU (0.6910) and Specificity (0.9843) metrics. This is in comparison to the baseline U-Net and the methods by Saha Tchinda et al. [21] and Liu et al. [23], as well as by Ding et al. [24]. Despite exhibiting slightly lower Sensitivity (0.8220) and Accuracy (0.9718) metrics than the baseline U-Net model, the proposed framework demonstrates competitive performance in contemporary state-of-the-art methodologies.

Table 6. Performance comparison of vesselness segmentation using the CHASE_DB1 dataset.

| Method | Year | DSC | mIoU | Sensitivity | Specificity | Accuracy |
|----------------------------|------|--------|--------|-------------|-------------|----------|
| Saha Tchinda et al. [21] | 2021 | - | - | 0.7279 | 0.9658 | 0.9452 |
| Aurangzeb et al. [51] | 2022 | - | - | 0.8607 | 0.9806 | 0.9731 |
| Liu et al. [23] | 2023 | - | - | 0.8242 | 0.9805 | 0.9664 |
| Ding et al. [24] | 2024 | - | 0.5983 | 0.7475 | 0.9798 | 0.9566 |
| Baseline U-Net | 2015 | 0.7821 | 0.6836 | 0.8427 | 0.9833 | 0.9738 |
| Proposed lightweight U-Net | 2025 | 0.7946 | 0.6910 | 0.8220 | 0.9843 | 0.9718 |

As can be seen in Figure 16, the images from the HRF dataset are presented. The initial column shows the preprocessed images, with subsequent columns depicting the images segmented by the enhanced lightweight U-Net model, the baseline U-Net model, and the corresponding mask images.

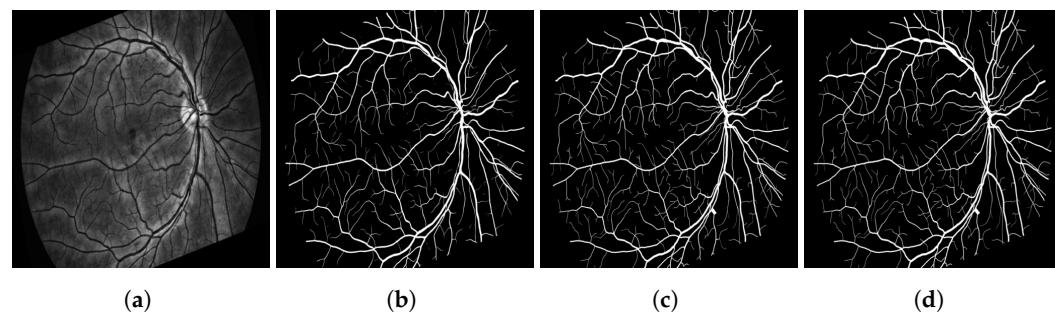


Figure 16. (a) Input image preprocessed, (b) segmented image from baseline U-Net, (c) segmented image from improved lightweight U-Net, and (d) ground-truth image.

Figure 17 shows the boxplots of DSC, mIoU, Sensitivity, Specificity, and Accuracy, highlighting the enhanced and more consistent performance of the lightweight U-Net (lightweight U-Net, LU-Net) in comparison to the baseline U-Net on the HRF dataset. LU-Net demonstrates higher median values and narrower interquartile ranges, signifying stronger repeatability and robustness. The elevated DSC and mIoU indices demonstrate superior overlap with the ground-truth masks, while increased Sensitivity and Specificity indices reveal the effective capture of fine vessel structures alongside a lower rate of false positives. In addition, LU-Net's Accuracy exceeds that of the baseline U-Net, indicating a greater proportion of correctly classified pixels overall. Thus, the findings demonstrate that LU-Net achieves not only superior average performance but also reduced variability, making it particularly suitable for clinical applications demanding consistent and reliable segmentation results.

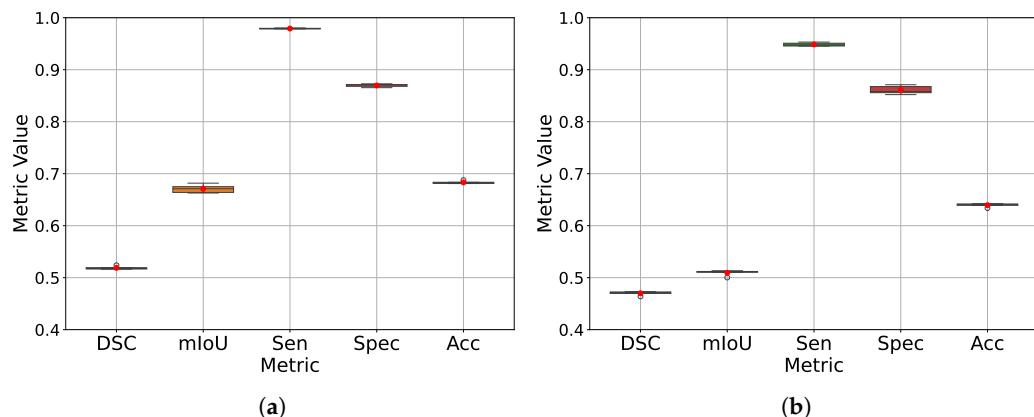


Figure 17. Using the improved lightweight U-Net on the HRF violin distribution dataset resulted in an increase in key performance metrics when considering a 5-fold cross-validation approach with 5 replicates: (a) improved lightweight U-Net and (b) baseline U-Net.

The comparison between the modified U-Net model and the baseline U-Net model in Figure 17 reveals significant improvements in various performance metrics, particularly in segmentation Accuracy and Sensitivity. The modified U-Net demonstrates an improvement in the DSC, with a 95% confidence interval ranging from 0.516 to 0.521, compared to the baseline model, which shows a lower DSC interval between 0.467 and 0.473. Similarly, the mean mIoU for the modified model ranges from 0.665 to 0.677. By contrast, the U-Net baseline has a lower mIoU range of 0.505 to 0.514, indicating that the modified U-Net is more effective at achieving precise segmentations.

The sensitivity of the modified model, with a confidence interval of 0.979–0.980, is higher than that of the baseline, which ranges from 0.946 to 0.951, indicating an enhanced ability of the modified model to detect true positives accurately. However, the Specificity of the modified U-Net, with a 95% confidence interval between 0.867 and 0.872, is only slightly higher than that of the baseline, which ranges from 0.855 to 0.867, indicating a comparable performance in correctly identifying true negatives. Finally, the Accuracy of the modified model, with a confidence interval ranging from 0.681 to 0.685, surpasses the baseline's Accuracy range of 0.637 to 0.642, indicating a stronger overall performance.

Table 7 offers a comparative analysis of the proposed LU-Net's performance with other methods on the HRF dataset about retinal vessel segmentation. While LU-Net demonstrated higher Dice Similarity Coefficients (0.6902 vs. 0.6417) and mean Intersection over Union values (0.5270 vs. 0.4725), there was a marginal decline in Sensitivity (0.8161 vs. 0.8559) and Accuracy (0.8437 vs. 0.8710).

Table 7. Performance comparison of vesselness segmentation using the HRF dataset.

| Method | Year | DSC | mIoU | Sensitivity | Specificity | Accuracy |
|----------------------------|------|--------|--------|-------------|-------------|----------|
| Chu et al. [25] | 2025 | - | - | 0.7840 | 0.9820 | 0.9640 |
| Kande et al. [27] | 2025 | - | - | 0.8612 | 0.9883 | 0.9825 |
| Luo et al. [28] | 2025 | - | - | 0.8497 | - | - |
| Baseline U-Net [40] | 2015 | 0.6417 | 0.4725 | 0.8559 | 0.9531 | 0.8710 |
| Proposed lightweight U-Net | 2025 | 0.6902 | 0.5270 | 0.8161 | 0.9707 | 0.8437 |

However, LU-Net exhibited an enhanced Specificity (0.9707 vs. 0.9531). Several studies have reported sensitivities ranging from 0.7840 to 0.8612, with accuracies frequently exceeding 0.96. However, direct comparisons have been hindered by inconsistencies in the reported metrics (for example, DSC and mIoU were omitted). Nevertheless, the gains achieved by LU-Net in Dice and mIoU underscore the production of more spatially coherent segmentation masks, suggesting that its balance of Specificity and Sensitivity may be

advantageous in clinical applications where minimizing false positives is prioritized while maintaining effective vascular detection.

Table 8 presents the results of an ablation study for the DRIVE, CHASE, and HRF dataset. The table comprehensively compares the modules and their impact on the metrics presented. The modification of the loss function from BCE to DICE loss initially resulted in enhanced mIoU metrics. Dice Loss optimizes the overlap between predicted and ground-truth regions, making it particularly effective in handling class imbalance and improving segmentation performance in small or sparse structures.

Table 8. Ablation study of different parameters on the DRIVE, CHASE, and HRF datasets.

| Dataset | Method | Optimizer | BCE Loss | Dice Loss | RA | DSC | mIoU | Sensitivity | Specificity | Accuracy |
|---------|-------------------|-----------|----------|-----------|----|--------|--------|-------------|-------------|----------|
| DRIVE | Baseline U-Net | Adam | ✓ | ✗ | ✗ | 0.7373 | 0.5839 | 0.8687 | 0.9515 | 0.8931 |
| | Baseline U-Net | Adam | ✓ | ✓ | ✗ | 0.7402 | 0.5876 | 0.8381 | 0.9630 | 0.9077 |
| | Lightweight U-Net | Adam | ✓ | ✗ | ✗ | 0.7298 | 0.5764 | 0.8631 | 0.9501 | 0.8906 |
| | Lightweight U-Net | Adam | ✓ | ✓ | ✗ | 0.7541 | 0.6053 | 0.7827 | 0.9749 | 0.9119 |
| | Lightweight U-Net | AdamW | ✓ | ✗ | ✗ | 0.6948 | 0.5324 | 0.9231 | 0.8792 | 0.8970 |
| | Lightweight U-Net | AdamW | ✓ | ✓ | ✗ | 0.7546 | 0.6059 | 0.7513 | 0.9782 | 0.9053 |
| | Lightweight U-Net | AdamW | ✗ | ✓ | ✗ | 0.7334 | 0.5791 | 0.6811 | 0.9825 | 0.9004 |
| | Lightweight U-Net | Adam | ✓ | ✗ | ✓ | 0.7402 | 0.5876 | 0.8764 | 0.9512 | 0.8936 |
| | Lightweight U-Net | Adam | ✗ | ✓ | ✓ | 0.7677 | 0.6230 | 0.7672 | 0.9770 | 0.9032 |
| | Lightweight U-Net | Adam | ✓ | ✓ | ✓ | 0.7633 | 0.6172 | 0.7699 | 0.9778 | 0.9109 |
| | Lightweight U-Net | AdamW | ✗ | ✓ | ✓ | 0.7871 | 0.6318 | 0.7421 | 0.9837 | 0.9113 |
| | Lightweight U-Net | AdamW | ✓ | ✓ | ✓ | 0.7741 | 0.6315 | 0.7678 | 0.9782 | 0.9034 |
| CHASE | Baseline U-Net | Adam | ✓ | ✗ | ✗ | 0.7830 | 0.6440 | 0.7702 | 0.9864 | 0.9725 |
| | Baseline U-Net | Adam | ✓ | ✓ | ✗ | 0.7945 | 0.6591 | 0.7997 | 0.9864 | 0.9749 |
| | Lightweight U-Net | Adam | ✓ | ✗ | ✗ | 0.7796 | 0.6396 | 0.7932 | 0.9833 | 0.9711 |
| | Lightweight U-Net | Adam | ✓ | ✓ | ✗ | 0.8051 | 0.6740 | 0.8402 | 0.9795 | 0.9688 |
| | Lightweight U-Net | AdamW | ✓ | ✗ | ✗ | 0.7838 | 0.6451 | 0.8065 | 0.9826 | 0.9713 |
| | Lightweight U-Net | AdamW | ✗ | ✓ | ✗ | 0.7732 | 0.6309 | 0.7299 | 0.9890 | 0.9723 |
| | Lightweight U-Net | AdamW | ✓ | ✓ | ✗ | 0.7442 | 0.5928 | 0.7199 | 0.9821 | 0.9619 |
| | Lightweight U-Net | Adam | ✓ | ✗ | ✓ | 0.7917 | 0.6556 | 0.7943 | 0.9853 | 0.9729 |
| | Lightweight U-Net | Adam | ✗ | ✓ | ✓ | 0.7925 | 0.6570 | 0.7628 | 0.9888 | 0.9742 |
| | Lightweight U-Net | Adam | ✓ | ✓ | ✓ | 0.8169 | 0.6905 | 0.8028 | 0.9868 | 0.9703 |
| | Lightweight U-Net | AdamW | ✗ | ✓ | ✓ | 0.7946 | 0.6598 | 0.7795 | 0.9873 | 0.9739 |
| | Lightweight U-Net | AdamW | ✓ | ✓ | ✓ | 0.7857 | 0.6474 | 0.7899 | 0.9843 | 0.9714 |
| HRF | Baseline U-Net | Adam | ✓ | ✗ | ✗ | 0.6417 | 0.4725 | 0.8559 | 0.9531 | 0.8710 |
| | Baseline U-Net | Adam | ✓ | ✓ | ✗ | 0.7179 | 0.5599 | 0.7740 | 0.9737 | 0.8526 |
| | Lightweight U-Net | Adam | ✓ | ✗ | ✗ | 0.7745 | 0.6327 | 0.7760 | 0.9796 | 0.8447 |
| | Lightweight U-Net | AdamW | ✓ | ✗ | ✗ | 0.7757 | 0.6343 | 0.8372 | 0.9721 | 0.8499 |
| | Lightweight U-Net | AdamW | ✗ | ✓ | ✗ | 0.7757 | 0.6344 | 0.7534 | 0.9827 | 0.8448 |
| | Lightweight U-Net | AdamW | ✓ | ✓ | ✗ | 0.7023 | 0.5412 | 0.7112 | 0.9789 | 0.8563 |
| | Lightweight U-Net | Adam | ✓ | ✗ | ✓ | 0.7729 | 0.6306 | 0.7922 | 0.9755 | 0.8348 |
| | Lightweight U-Net | Adam | ✗ | ✓ | ✓ | 0.7751 | 0.6335 | 0.7817 | 0.9794 | 0.8496 |
| | Lightweight U-Net | Adam | ✓ | ✓ | ✓ | 0.660 | 0.4993 | 0.6363 | 0.9810 | 0.8475 |
| | Lightweight U-Net | AdamW | ✗ | ✓ | ✓ | 0.7756 | 0.6342 | 0.7794 | 0.9793 | 0.8474 |
| | Lightweight U-Net | AdamW | ✓ | ✓ | ✓ | 0.6887 | 0.5253 | 0.6742 | 0.9792 | 0.8377 |

Furthermore, the table presents a comparison of a hybrid function that combines the BCE and DICE loss functions, each weighted at 0.5. Additionally, integrating reverse attention (RA) yielded notable gains in two of the three evaluated datasets. On DRIVE, RA raised the Dice score by 6.8% points and the IoU metric by 8.3%, while using CHASE_DB1, the proposed method improved these metrics by 2.7% and 4.3%, respectively. A non-significant measurable benefit was observed when using the HRF dataset.

Similarly, the experimental findings demonstrate that LU-Net outperforms the baseline on both the CHASE and HRF datasets, although it should be noted that each dataset

responds to different training choices. For CHASE, the optimal segmentation is achieved by pairing AdamW with Dice loss, as evidenced by metrics such as Dice = 0.7946 and mIoU = 0.6598. This outcome underscores the efficacy of weight-decay regularization and synthetic variability in facilitating the recovery of thin vessels. The HRF model, which exhibits larger and more uniform vessels, is predominantly influenced by AdamW and Dice loss, with minimal gains from reverse attention. The optimal configuration achieves Dice = 0.7756 and mIoU = 0.6342. Across both datasets, AdamW consistently enhances overlap metrics; Dice loss improves Specificity at a negligible cost to Sensitivity; and the sub-two-million-parameter architecture provides competitive performance suitable for real-time, resource-constrained retinal imaging.

Table 9 presents the LU-Net model demonstrating remarkable computational efficiency, requiring only 1.94 million parameters and 12.21 GFLOPs. A comparison with the baseline U-Net reveals a 75% reduction in parameters and an 86% decrease in floating-point operations while maintaining or exceeding the performance of the MSMA Net and TP-UNET models. The proposed architecture's substantial savings make it well-suited for deployment on hardware with limited resources and for large-scale clinical workflows where the speed of inference and the amount of memory used are critical.

Table 9. Computational efficiency compared with other methods.

| Model | Parameters | | Inference (Mean ± Std) | |
|----------------------------|---------------------|-----------------------------|------------------------|--------------------|
| | M = 10 ⁶ | G = 10 ⁹ | FPS | Latency (ms/Image) |
| MSMA Net [50] | 4.092 M | - | - | - |
| TP-UNET [23] | 15.6 M | 132.58 G | - | - |
| Aurangzeb et al. [51] | 5.0 M | - | - | - |
| Baseline U-Net [40] | 7.77 M | 84.50 G (322.63 M/pixel) | 105.28 ± 15.84 | 9.86 ± 2.46 |
| Proposed Lightweight U-Net | 1.94 M | 12.21 G (46.58 M/pixel) | 208.00 ± 10.95 | 4.81 ± 0.28 |

6. Discussion

The proposed pipeline combines a lightweight optimized U-Net with a tailored preprocessing routine that applies contrast-limited adaptive histogram equalization, gamma correction, and affine data augmentation. This dual strategy tackles two central challenges in retinal vessel analysis: (i) low contrast between thin vessels and background and (ii) restricted hardware capacity in point-of-care settings. Relative to the baseline, the new architecture lifts the Dice coefficient on CHASE_DB1 from 0.7830 to 0.7946, on HRF from 0.6417 to 0.7757, and on DRIVE from 0.7373 to 0.7871 while keeping the parameter count below two million. These gains confirm that GELU activation and the AdamW optimizer, with its decoupled weight decay, are decisive for convergence and generalization in highly imbalanced medical images.

The comparative evaluation reveals clear dataset-dependent behavior. On DRIVE, coupling AdamW with Dice loss and reverse attention raises the mIoU from 0.5839 to 0.6318. It improves Specificity to 0.9837, underscoring the importance of synthetic variability and regularization for reliably detecting thin vessels. A similar trend is observed on CHASE_DB1, where reverse attention yields a roughly one-percentage-point improvement in overlap metrics. In contrast, HRF images already contain diverse illumination and larger vessels, so additional reverse attention offers negligible benefits. We detected that the HRF database was built using a mydriatic medicament to dilate the pupil. This process provokes vasoconstriction (narrowing of blood vessels), which obviously affects the detection quality of the proposed method. These observations highlight the need for dataset-specific hyperparameter tuning rather than a single, universal recipe.

Although efficient, the framework has limitations. The present loss design sacrifices a small amount of Sensitivity to gain higher Specificity, which may lead to the under-detection

of extremely narrow capillaries. Furthermore, only three public datasets were examined, so performance across broader demographic or pathological variations remains to be validated. Future work will explore hybrid loss functions that better balance recall and precision, self-supervised pre-training to exploit unlabeled data, and quantization or pruning techniques to reduce the model size.

7. Conclusions

The proposed lightweight U-Net achieves accurate retinal vessel segmentation with a markedly reduced computational footprint. Its efficiency stems from two targeted enhancements to the baseline U-Net: (i) substituting ReLU with smoother GELU activation, which improves gradient flow in shallow feature extractors, and (ii) adopting the AdamW optimizer, whose decoupled weight decay accelerates convergence and limits overfitting. Experiments on the CHASE_DB1 and HRF datasets confirm that the resulting architecture, which contains fewer than two million parameters, attains Dice scores of 0.7946 and 0.7756, respectively, matching or surpassing considerably larger counterparts while remaining deployable on resource-constrained hardware such as embedded GPUs and edge devices.

According to Ajani et al. [52], the memory available on such devices ranges from 32 kilobytes (KB) to 2 megabytes (MB) and consumes from 4 to 80 mA on average. The current model has a size of 7.5 MB. Although there are a few embedded architectures that can have a memory up to the gigabytes (e.g., Raspberry PI), an optimization stage will be needed to fit the model due to embedded hardware requirements. The forthcoming tasks will be used to benchmark the model geared on low-power embedded platforms (e.g., NVIDIA Jetson Nano, and Raspberry Pi CM4) to evaluate inference speed and energy efficiency for point-of-care applications. Additionally, a padding-based resizing strategy will be implemented to further reduce noise-induced distortion. Reverse attention improved Dice/IoU by 6.8/8.3 percentage points (p.p.t.) on DRIVE and 2.7/4.3 p.p.t. on CHASE_DB1 while showing no measurable gain on HRF. The proposed lightweight U-Net delivers markedly superior computational efficiency: it reduces the model size by approximately 75% (1.94 M vs. 7.77 M parameters) and cuts FLOPs by approximately 85% (12.21 G vs. 84.50 G) while nearly doubling its throughput (208 ± 11 FPS) and halving its inference latency (4.81 ± 0.28 ms per image) relative to the baseline U-Net. These findings demonstrate the framework's suitability for real-time, point-of-care ophthalmic applications and offer a practical blueprint for energy-efficient medical image analysis.

Author Contributions: Conceptualization, J.R.-P. and E.O.-M.; data curation, M.A.I.-M.; formal analysis, E.G.A.-B.; investigation, F.D.H.-G. and J.G.A.-C.; methodology, M.A.I.-M., E.O.-M., and J.G.A.-C.; software, F.D.H.-G., E.G.A.-B., and J.R.-P.; validation, M.A.I.-M.; visualization, F.D.H.-G. and J.G.A.-C.; writing—original draft, F.D.H.-G.; writing—review and editing, E.G.A.-B., J.R.-P., E.O.-M., and J.G.A.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the University of Guanajuato under project CIIC-UG 163/2025 and Grant NUA 143745. It was partially funded by the Secretary of Science, Humanities, Technology and Innovation (SECIHTI) Grant 838509/1080385.

Data Availability Statement: The data and codes presented in this study are available (accessed on 18 June 2025) in RVS at <https://github.com/fd hernandezgutierrez/RVS>. These data were derived from the following resources available in the public domain: DRIVE <https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction>. HRF <https://www5.cs.fau.de/research/data/fundus-images/>. CHASE-DB1 <https://www.kaggle.com/datasets/khoongweihao/chasedb1>.

Acknowledgments: The authors thank the University of Guanajuato for the facilities and support given to develop this project.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study; in data collection, analyses, and interpretation; and in manuscript writing or the decision to publish the results.

References

1. Kropp, M.; Golubnitschaja, O.; Mazurakova, A.; Koklesova, L.; Sargheini, N.; Vo, T.T.K.S.; de Clerck, E.; Polivka, J.; Potuznik, P.; Stetkarova, I.; et al. Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—Risks and mitigation. *EPMA J.* **2023**, *14*, 21–42. [[CrossRef](#)] [[PubMed](#)]
2. Srejovic, J.V.; Muric, M.D.; Jakovljevic, V.L.; Srejovic, I.M.; Sreckovic, S.B.; Petrovic, N.T.; Todorovic, D.Z.; Bolevich, S.B.; Sarenac Vulovic, T.S. Molecular and Cellular Mechanisms Involved in the Pathophysiology of Retinal Vascular Disease—Interplay Between Inflammation and Oxidative Stress. *Int. J. Mol. Sci.* **2024**, *25*, 11850. [[CrossRef](#)]
3. Group, E.T.D.R.S.R. Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs—An Extension of the Modified Airlie House Classification: ETDRS Report Number 10. *Ophthalmology* **1991**, *98*, 786–806. [[CrossRef](#)]
4. Wilkinson, C.; Ferris, F.L.; Klein, R.E.; Lee, P.P.; Agardh, C.D.; Davis, M.; Dills, D.; Kampik, A.; Pararajasegaram, R.; Verdaguer, J.T. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* **2003**, *110*, 1677–1682. [[CrossRef](#)] [[PubMed](#)]
5. Gegundez-Arias, M.E.; Marin-Santos, D.; Perez-Borrero, I.; Vasallo-Vazquez, M.J. A new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and modified U-Net model. *Comput. Methods Programs Biomed.* **2021**, *205*, 106081. [[CrossRef](#)] [[PubMed](#)]
6. Wang, W.; Lo, A.C.Y. Diabetic Retinopathy: Pathophysiology and Treatments. *Int. J. Mol. Sci.* **2018**, *19*, 1816. [[CrossRef](#)]
7. Lee, M.; Leskova, W.; Eshaq, R.S.; Harris, N.R. Retinal hypoxia and angiogenesis with methamphetamine. *Exp. Eye Res.* **2021**, *206*, 108540. [[CrossRef](#)]
8. Ilesanmi, A.E.; Ilesanmi, T.; Gbotoso, G.A. A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks. *Healthc. Anal.* **2023**, *4*, 100261. [[CrossRef](#)]
9. Radha, K.; Karuna, Y. Modified Depthwise Parallel Attention UNet for Retinal Vessel Segmentation. *IEEE Access* **2023**, *11*, 102572–102588. [[CrossRef](#)]
10. Teo, Z.L.; Tham, Y.C.; Yu, M.; Chee, M.L.; Rim, T.H.; Cheung, N.; Bikbov, M.M.; Wang, Y.X.; Tang, Y.; Lu, Y.; et al. Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis. *Ophthalmology* **2021**, *128*, 1580–1591. [[CrossRef](#)]
11. Association, A.D. Standards of Care in Diabetes—2023 Abridged for Primary Care Providers. *Clin. Diabetes* **2023**, *41*, 4–31. [[CrossRef](#)] [[PubMed](#)]
12. Burlina, P.; Galdran, A.; Costa, P.; Cohen, A.; Campilho, A. Chapter 18—Artificial intelligence and deep learning in retinal image analysis. In *Computational Retinal Image Analysis*; Trucco, E., MacGillivray, T., Xu, Y., Eds.; The Elsevier and MICCAI Society Book Series; Academic Press: Cambridge, MA, USA, 2019; pp. 379–404. [[CrossRef](#)]
13. Aguirre-Ramos, H.; Avina-Cervantes, J.G.; Cruz-Aceves, I.; Ruiz-Pinales, J.; Ledesma, S. Blood vessel segmentation in retinal fundus images using Gabor filters, fractional derivatives, and Expectation Maximization. *Appl. Math. Comput.* **2018**, *339*, 568–587. [[CrossRef](#)]
14. Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 113. [[CrossRef](#)]
15. Anaya-Isaza, A.; Mera-Jiménez, L.; Zequera-Díaz, M. An overview of deep learning in medical imaging. *Inform. Med. Unlocked* **2021**, *26*, 100723. [[CrossRef](#)]
16. Chen, C.; Mat Isa, N.A.; Liu, X. A review of convolutional neural network based methods for medical image classification. *Comput. Biol. Med.* **2025**, *185*, 109507. [[CrossRef](#)]
17. Lynch, S.K.; Abràmoff, M.D. Diabetic retinopathy is a neurodegenerative disorder. *Vis. Res.* **2017**, *139*, 101–107. [[CrossRef](#)]
18. Vujosevic, S.; Midena, E. Retinal layers changes in human preclinical and early clinical diabetic retinopathy support early retinal neuronal and Müller cells alterations. *J. Diabetes Res.* **2013**, *2013*, 905058. [[CrossRef](#)]
19. Indumathi, G.; Sathananthavathi, V. Chapter 5—Microaneurysms Detection for Early Diagnosis of Diabetic Retinopathy Using Shape and Steerable Gaussian Features. In *Telemedicine Technologies*; Hemanth, D.J., Balas, V.E., Eds.; Academic Press: Cambridge, MA, USA, 2019; pp. 57–69. [[CrossRef](#)]
20. Aslani, S.; Sarnel, H. A new supervised retinal vessel segmentation method based on robust hybrid features. *Biomed. Signal Process. Control* **2016**, *30*, 1–12. [[CrossRef](#)]
21. Saha Tchinda, B.; Tchiotsop, D.; Noubom, M.; Louis-Dorr, V.; Wolf, D. Retinal blood vessels segmentation using classical edge detection filters and the neural network. *Inform. Med. Unlocked* **2021**, *23*, 100521. [[CrossRef](#)]
22. Ren, K.; Chang, L.; Wan, M.; Gu, G.; Chen, Q. An improved U-net based retinal vessel image segmentation method. *Heliyon* **2022**, *8*, e11187. [[CrossRef](#)]

23. Liu, R.; Pu, W.; Nan, H.; Zou, Y. Retina image segmentation using the three-path Unet model. *Sci. Rep.* **2023**, *13*, 22579. [[CrossRef](#)] [[PubMed](#)]
24. Ding, W.; Sun, Y.; Huang, J.; Ju, H.; Zhang, C.; Yang, G.; Lin, C.T. RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism. *Inf. Sci.* **2024**, *657*, 120007. [[CrossRef](#)]
25. Chu, B.; Zhao, J.; Zheng, W.; Xu, Z. (DA-U)²Net: Double attention U²Net for retinal vessel segmentation. *BMC Ophthalmol.* **2025**, *25*, 86. [[CrossRef](#)]
26. Zhang, W.; Qu, S.; Feng, Y. LMFR-Net: Lightweight multi-scale feature refinement network for retinal vessel segmentation. *Pattern Anal. Appl.* **2025**, *28*, 44. [[CrossRef](#)]
27. Kande, G.B.; Nalluri, M.R.; Manikandan, R.; Cho, J.; Veerappampalayam Easwaramoorthy, S. Multi scale multi attention network for blood vessel segmentation in fundus images. *Sci. Rep.* **2025**, *15*, 3438. [[CrossRef](#)]
28. Luo, X.; Peng, L.; Ke, Z.; Lin, J.; Yu, Z. PA-Net: A hybrid architecture for retinal vessel segmentation. *Pattern Recognit.* **2025**, *161*, 111254. [[CrossRef](#)]
29. Jian, M.; Xu, W.; Nie, C.; Li, S.; Yang, S.; Li, X. DAU-Net: A novel U-Net with dual attention for retinal vessel segmentation. *Biomed. Phys. Eng. Express* **2025**, *11*, 025009. [[CrossRef](#)]
30. Islam, T.; Hafiz, M.S.; Jim, J.R.; Kabir, M.M.; Mridha, M. A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. *Healthc. Anal.* **2024**, *5*, 100340. [[CrossRef](#)]
31. Goceri, E. Medical image data augmentation: Techniques, comparisons and interpretations. *Artif. Intell. Rev.* **2023**, *56*, 12561–12605. [[CrossRef](#)]
32. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [[CrossRef](#)]
33. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nat. Commun.* **2024**, *15*, 654. [[CrossRef](#)] [[PubMed](#)]
34. Nguyen, Q.D.; Thai, H.T. Crack segmentation of imbalanced data: The role of loss functions. *Eng. Struct.* **2023**, *297*, 116988. [[CrossRef](#)]
35. Zhou, P.; Xie, X.; Lin, Z.; Yan, S. Towards Understanding Convergence and Generalization of AdamW. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 6486–6493. [[CrossRef](#)] [[PubMed](#)]
36. Lee, M. Mathematical Analysis and Performance Evaluation of the GELU Activation Function in Deep Learning. *J. Math.* **2023**, *2023*, 4229924. [[CrossRef](#)]
37. Chen, S.; Tan, X.; Wang, B.; Lu, H.; Hu, X.; Fu, Y. Reverse Attention-Based Residual Network for Salient Object Detection. *IEEE Trans. Image Process.* **2020**, *29*, 3763–3776. [[CrossRef](#)]
38. Wang, Z.; Xie, X.; Yang, J.; Song, X. RA-Net: Reverse attention for generalizing residual learning. *Sci. Rep.* **2024**, *14*, 12771. [[CrossRef](#)]
39. Lee, G.E.; Cho, J.; Choi, S.I. Shallow and reverse attention network for colon polyp segmentation. *Sci. Rep.* **2023**, *13*, 15243. [[CrossRef](#)]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer: Cham, Switzerland, 2015; pp. 234–241.
41. Staal, J.; Abramoff, M.; Niemeijer, M.; Viergever, M.; van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [[CrossRef](#)]
42. Odstrcilik, J.; Kolar, R.; Budai, A.; Hornegger, J.; Jan, J.; Gazarek, J.; Kubena, T.; Cernosek, P.; Svoboda, O.; Angelopoulou, E. Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database. *IET Image Process.* **2013**, *7*, 373–383. [[CrossRef](#)]
43. Fraz, M.M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.R.; Owen, C.G.; Barman, S.A. An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2538–2548. [[CrossRef](#)]
44. Yue, K.; Zhan, L.; Wang, Z. Unsupervised domain adaptation teacher–student network for retinal vessel segmentation via full-resolution refined model. *Sci. Rep.* **2025**, *15*, 2038. [[CrossRef](#)] [[PubMed](#)]
45. Zou, K.H.; Warfield, S.K.; Bharatha, A.; Tempany, C.M.; Kaus, M.R.; Haker, S.J.; Wells III, W.M.; Jolesz, F.A.; Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index: Scientific reports. *Acad. Radiol.* **2004**, *11*, 178–189. [[CrossRef](#)] [[PubMed](#)]
46. Zanddizi, H.; Nguyen, N.; Zeinali, B.; Chang, J.M. A new preprocessing approach to improve the performance of CNN-based skin lesion classification. *Med. Biol. Eng. Comput.* **2021**, *59*, 1123–1131. [[CrossRef](#)]
47. Rezatofighi, H.; Tsai, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666. [[CrossRef](#)]

48. Müller, D.; Soto-Rey, I.; Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res. Notes* **2022**, *15*, 210. [[CrossRef](#)]
49. Li, Z.; Jia, M.; Yang, X.; Xu, M. Blood Vessel Segmentation of Retinal Image Based on Dense-U-Net Network. *Micromachines* **2021**, *12*, 1478. [[CrossRef](#)] [[PubMed](#)]
50. Toptaş, B.; Hanbay, D. Retinal blood vessel segmentation using pixel-based feature vector. *Biomed. Signal Process. Control* **2021**, *70*, 103053. [[CrossRef](#)]
51. Aurangzeb, K.; Alharthi, R.S.; Haider, S.I.; Alhussein, M. An Efficient and Light Weight Deep Learning Model for Accurate Retinal Vessels Segmentation. *IEEE Access* **2023**, *11*, 23107–23118. [[CrossRef](#)]
52. Ajani, T.S.; Imoize, A.L.; Atayero, A.A. An Overview of Machine Learning within Embedded and Mobile Devices—Optimizations and Applications. *Sensors* **2021**, *21*, 4412. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.