Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 25, No. 3, July 2026, pp. 551∼564
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 10/C/C3/DT.05.00/2025
DOI: 10.30812/matrik.v25i3.6373

Retinal Blood Vessel Segmentation Using Attention V-Net with Scale-Aware Evaluation

Hendra Wijaya[1], **Erwin Erwin**[1], **Annisa Darmawahyuni**[1], **Sinta Bella Agustina**[2]
[1]Universitas Sriwijaya, Palembang, Indonesia
[2]Politeknik Prasetya Mandiri, Indonesia

Article Info

Article history:

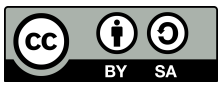
Received May 05, 2026
Revised June 05, 2026
Accepted June 18, 2026

Keywords:

Attention Gate;
Retinal Vessels;
Small Vessels;
V-Net.

ABSTRACT

Retinal blood vessel segmentation remains a significant challenge, especially for small blood vessels with diameters less than 3 pixels in the DRIVE dataset and less than 4 pixels in the STARE dataset, owing to their low contrast and narrow structures. The aim of this study is to improve small retinal blood vessel segmentation performance through an Attention V-Net architecture that integrates attention-gating mechanisms into the skip connections of a V-Net backbone to strengthen the feature representation of thin vascular structures. The research method involves training and evaluating the proposed model on the DRIVE and STARE datasets using a scale-aware evaluation framework based on pixel-pitch calibration, classifying blood vessels into small and large categories, and measuring performance using accuracy, sensitivity, specificity, precision, Dice coefficient, and IoU. The results show that for small vessel segmentation, the method achieves sensitivities of 0.7033 and 0.6984, Dice scores of 0.4720 and 0.4699, and IoUs of 0.3096 and 0.3079 on the DRIVE and STARE datasets, respectively. For large vessels, sensitivities of 0.9219 and 0.8851, Dice scores of 0.8031 and 0.8179, and IoUs of 0.6719 and 0.6933 are obtained. Global evaluation yields accuracies of 0.9475 and 0.9602, sensitivities of 0.8727 and 0.8719, and Dice scores of 0.8080 and 0.8268. In conclusion, Attention V-Net demonstrates consistent segmentation performance across vessel scales, and the scale-aware evaluation framework effectively reveals the performance gap between small and large vessel segmentation, providing a more clinically relevant assessment than conventional global evaluation for early diagnosis of retinal diseases.



Corresponding Author:

Hendra Wijaya, +62 852-6821-9643,
Faculty of Computer Science, master's program in computer science,
Universitas Sriwijaya, Palembang, Indonesia,
Email: hendrawijaya150896@gmail.com

How to Cite:
H. Wijaya, E. Erwin, A. Darmawahyuni, and S. B. Agustina, "Retinal Blood Vessel Segmentation Using Attention V-Net with Scale-Aware Evaluation", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 25, No. 3, pp. 551-564, July, 2026.

1. INTRODUCTION

Retinal blood vessels provide a unique opportunity for non-invasive observation of the microcirculation [1]. Pathological alterations in these vessels are strongly linked to a range of ocular diseases, including diabetic retinopathy, glaucoma, retinal artery occlusion, and retinal detachment, many of which can lead to significant visual impairment or blindness [2, 3]. Consequently, the analysis of retinal blood vessels is essential for elucidating the onset and progression of eye diseases and for facilitating early diagnosis [4]. Imaging modalities such as color fundus photography and optical coherence tomography are commonly employed to evaluate retinal blood vessel health [5]. Nevertheless, the quality of retinal images is frequently compromised by factors including diffraction, defocus, low contrast, and uneven illumination, which can result in suboptimal vessel visualization and hinder accurate, consistent assessment [6, 7]. These needs underscore the necessity for an automated, objective, and robust retinal blood vessel segmentation that performs reliably under diverse image quality [8, 9].

Recent advancements in deep learning have substantially advanced retinal blood vessel segmentation. Agustina et al. [10] introduced CNN VV-Net, which incorporates image quality enhancement via grayscale conversion, median filtering, and CLAHE preprocessing prior to segmentation. When evaluated on the DRIVE, STARE, and CHASEDB_1 datasets, this method achieved 97.04% accuracy, 83.55% sensitivity, and 66.40% Jaccard score on CHASEDB_1, indicating competitive performance across multiple benchmarks. Ramadhani et al. [11] presented ResVNet, a hybrid architecture that eliminates the U-Net's bridge component, establishes a V-Net backbone, and integrates ResNet-based residual connections into the encoder to enhance deep feature extraction. On the DRIVE and STARE datasets, ResVNet achieved 96.57% accuracy and 82.28% sensitivity on DRIVE; however, sensitivity and Jaccard score on STARE were limited, highlighting the need for further architectural refinement. Hernandez-Gutierrez et al. [12] developed a lightweight U-Net with a reverse attention mechanism and a dedicated preprocessing framework, achieving a Dice coefficient of 0.7871 and an IoU of 0.6318 on DRIVE, while utilizing only 1.94 million parameters at 208 FPS. The reverse attention mechanism improved segmentation of thin and peripheral vessels, which are commonly missed by conventional models. Radha et al. [13] combined an Attention U-Net with the Unfolded Deep Kernel Estimation (UDKE) method to improve semantic segmentation performance under limited training data. Experiments on the DRIVE, STARE, and CHASE_DB datasets yielded an accuracy of 96.71% and a Dice score of 0.7228, though sensitivity remained at 75.91%, indicating ongoing challenges in detecting low-contrast vascular structures. Despite these advancements, a common limitation is that performance evaluation is typically conducted on blood vessels, without distinguishing between small and large vessels. This practice may obscure model deficiencies in detecting the finest vascular structures.

Although deep learning-based methods such as CNN VV-Net, ResVNet, lightweight U-Net with reverse attention, and UDKE-based Attention U-Net have demonstrated strong overall segmentation performance, accurately segmenting small blood vessels remains a significant challenge. Small blood vessels typically exhibit low contrast and very narrow diameters, often less than 65 μm [14], which correspond to approximately 3 pixels in the DRIVE dataset and 4 pixels in the STARE dataset after pixel-pitch calibration [15]. These properties render small vessels particularly vulnerable to information loss during downsampling and feature compression in encoder–decoder architectures. Clinically, the integrity of small vessels is critical for the early detection of conditions such as diabetic retinopathy, where microangiopathy initially appears in the finest vascular structures. Moreover, global performance metrics commonly used in previous studies often obscure model deficiencies in detecting small vessels, as the dominant contribution of large vessels inflates overall scores. To address these limitations, this study introduces a scale-aware deep learning framework based on the Attention V-Net architecture to enhance segmentation sensitivity for small blood vessels, while maintaining sensitivity for large blood vessels in the DRIVE and STARE datasets.

This study aims to improve retinal blood vessel segmentation, especially for small vessels that are often overlooked in standard evaluation methods. To achieve this, the present study introduces an Attention V-Net architecture that integrates attention-gating mechanisms into skip connections to better capture small, low-contrast vessels. A scale-aware evaluation framework is also proposed, which separates blood vessels into small and large categories using pixel-pitch calibration based on clinical definitions of vessel diameter. The main contributions of this study are threefold. First, the proposed Attention V-Net architecture preserves the fine structural details of small blood vessels, which are often lost during encoding and downsampling. Second, the scale-aware evaluation framework enables a more detailed and clinically meaningful assessment of segmentation performance for both small and large vessels, providing a more informative alternative to conventional global metrics. Third, the proposed framework offers a clinically meaningful evaluation strategy for vessel-scale-sensitive retinal segmentation and contributes to the development of more reliable automated tools for early diagnosis of retinal diseases such as diabetic retinopathy and glaucoma, ultimately supporting clinical decision-making and improving patient outcomes.

2. RESEARCH METHOD

This study employs a deep learning method and a quantitative experimental approach to segment retinal blood vessels in fundus images. The method is based on a Convolutional Neural Network (CNN), namely the Attention V-Net architecture, which is a development of V-Net that applies an attention gate mechanism to its skip connections. The attention gate mechanism allows the model to focus on relevant features of small blood vessels while suppressing irrelevant background information, thereby improving segmentation performance, especially for vessels with very small pixel sizes.

The case study in this study is retinal blood vessel segmentation in fundus images. Two publicly available benchmark datasets are used: DRIVE and STARE, both widely adopted in retinal vessel segmentation research and containing cases with varying levels of pathological complexity. A detailed description of each dataset, including image specifications, acquisition conditions, and annotation procedures, is provided in Subsection 2.1 Data Acquisition. The stages of the blood vessel segmentation methodology in retinal images are shown in Figure 1.

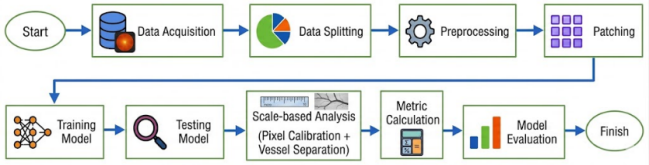


Figure 1. Diagram of the stages of retinal blood vessel segmentation

2.1. Data Acquisition

The first dataset used in this study is the DRIVE (Digital Retinal Images for Vessel Extraction) dataset, which is publicly available at the DRIVE repository [16]. Images were acquired with a Canon CR5 fundus camera in the Netherlands, using a 45^o field of view (FOV). The DRIVE dataset comprises 40 color retinal images [17], divided equally into 20 images for training and 20 for testing [18]. All images are stored in TIF format with 8-bit color depth and a resolution of 768 $\times$ 584 pixels [9]. Approximately 25% of the images exhibit pathological signs, thereby increasing the complexity of retinal vessel segmentation [18]. The inclusion of pathological cases renders DRIVE challenging and relevant for assessing the robustness and generalization of segmentation algorithms in realistic clinical scenarios. The second dataset used in this study is the Structured Analysis of the Retina (STARE), which is publicly available in the STARE repository [19]. This dataset was originally developed by Michael Goldbaum, MD, at the University of California, San Diego, with support from the National Institutes of Health (NIH) and in collaboration with the Shiley Eye Center and the Veterans Affairs Medical Center. The STARE dataset comprises 20 digital retinal images acquired with a Topcon TRV-50 fundus camera and a 35^o field of view (FOV) [20]. Each image has a resolution of 605 $\times$ 700 pixels, with an effective field of view of approximately 650 $\times$ 550 pixels. The dataset includes images from patients with diverse pathological conditions, thereby introducing greater clinical variability compared to non-pathological datasets. All images are preprocessed to retain only the relevant retinal region, and each image is independently annotated by two experts, yielding ground truth vessel segmentations [21].

2.2. Data Splitting

Data were partitioned at the file level using a stratified random seed to ensure reproducibility and maintain the distribution of data characteristics, including pathological variations, across subsets. The datasets were divided into training and validation sets at an 80:20 ratio, with 80% allocated to training and 20% to validation [22]. This approach reduces overfitting and enables the model to learn statistically meaningful features rather than memorizing the training data [23]. To prevent data leakage, patch extraction was performed after the data split, ensuring that patches originated only from images designated for training or validation. Table 1 outlines the data-splitting strategy used in this study. The DRIVE and STARE datasets are each divided into training and validation subsets. Specifically, the DRIVE datasets include 16 training and 4 validation images, whereas the STARE datasets include 12 training and 4 validation images. This approach facilitates robust model training and supports reliable evaluation on previously unseen data.

Table 1. Data Splitting

z	Training	Validation
DRIVE	16 Images	4 Images
STARE	12 Images	4 Images

2.3. Preprocessing

Preprocessing is a critical step in retinal vessel segmentation from fundus images, as it directly enhances image quality and improves segmentation performance [24]. In the present study, the preprocessing stage was conducted as follows. Green channel extraction is commonly employed in retinal vessel segmentation because it provides higher contrast between blood vessels and the retinal background than the red and blue channels [25]. Additionally, the green channel exhibits lower noise levels and reduced susceptibility to oversaturation, thereby facilitating the preservation of fine and low-contrast vascular structures [18, 26]. These properties enhance its ability to highlight fine, low-contrast vascular structures, thereby enabling clearer separation between vessels and the background as expressed in Equation 1.

$$g = \frac{G}{R + G + B} \tag{1}$$

Here g denotes the normalized green intensity, and R, G, and B represent the red, green, and blue channel intensities, respectively. This normalization improves robustness to illumination variations. After green channel extraction, contrast enhancement was conducted using Contrast Limited Adaptive Histogram Equalization (CLAHE), which is effective for medical images exhibiting non-uniform intensity distributions [27]. In contrast to global histogram equalization, CLAHE enhances contrast locally by applying histogram equalization to small image tiles. This approach preserves fine retinal vascular structures under uneven illumination conditions [28]. To mitigate noise amplification, a clip limit parameter was implemented [29]. The average intensity within each tile was calculated as expressed in Equation 2.

$$N_{avg} = \frac{N_{rX} N_{rY}}{N_{gray}} \tag{2}$$

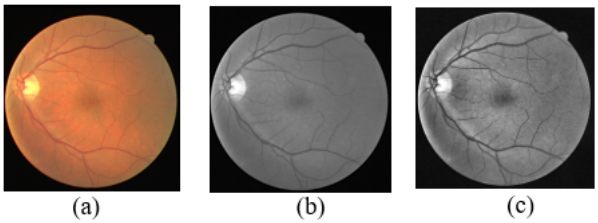
where N_{rX} and N_{rY} denote the tile dimensions, and N_{gray} represents the number of gray level intensities. The clip limit is then calculated as expressed in Equation 3

$$N_{CL} = N_{clip} \times N_{avg} \tag{3}$$

where N_CLdenotes the maximum allowable histogram count per tile. In this study, a clip limit of 2.0 and an 8×8 tile grid are applied to balance local contrast enhancement and noise suppression. After CLAHE, Min–max normalization is used to enhance the contrast of retinal images, thereby improving the distinguishability of blood vessels from the background. This procedure is essential for achieving accurate segmentation, particularly in low-contrast images, and helps reduce misclassification errors [30, 31]. The effectiveness of min–max normalization in enhancing segmentation performance has been extensively documented in previous studies [29]. The normalization process is defined where N_{rX} and N_{rY} denote the tile dimensions, and N_{gray} represents the number of gray level intensities. The clip limit is then calculated as expressed in Equation 4.

$$I_{norm}(x, y) = \frac{I(x, y) - I_{min}}{I_{max} - I_{min}} \tag{4}$$

Where, I(x,y) denotes the original pixel intensity at coordinates (x,y), while I_{max} and I_{min} represent the minimum and maximum intensity values in the image, respectively. The normalized pixel value at (x,y) is then scaled to the range [0,1]. Figure 2 shows the preprocessing results of (a) the original image, (b) the green channel, and (c) CLAHE.



(a) (b) (c)

Figure 2. Preprocessing results are presented for (a) the original image, (b) the green channel, and (c) the CLAHE

2.4. Patching

After preprocessing, a patch-based approach was used to train the model on retinal images, enabling it to learn local vascular structures without relying on full-resolution images. Random patch extraction increases data variability and stabilizes training, particularly for limited datasets [32], while preserving computational efficiency and segmentation performance [33]. In this study, random patches measuring 256 × 256 pixels were extracted from each image in both the training and validation sets, with 400 patches per image. Patches were sampled within the region of interest (ROI), defined as areas with an average ROI value greater than 0.75, to ensure adequate coverage of the retinal field of view. To address class imbalance and ensure sufficient vessel representation, each selected patch was required to contain at least 200 vessel pixels. Table 2 presents the patching results for each dataset. The DRIVE datasets yielded 6,400 training patches and 1,600 validation patches, whereas the STARE datasets yielded 4,800 training patches and 1,600 validation patches. The variation in the number of training patches reflects the differing numbers of original images across datasets. The number of validation patches was held constant to facilitate fair and consistent model evaluation.

Table 2. Patching results of DRIVE and STARE training and validation data

No	Datasets	Training	Validation
1	DRIVE	6400 patches	1600 patches
2	STARE	4800 patches	1600 patches

Figure 3 presents the patch extraction process from a retinal image. The left panel shows the original retinal image, which displays the complete retinal visual field. The yellow box delineates the region of interest (ROI) selected for patch extraction. The red arrow demonstrates the correspondence between the ROI in the original image and the extracted patch. Patch 1 and Patch 2 on the right are representative examples of smaller images extracted from distinct locations within the ROI.

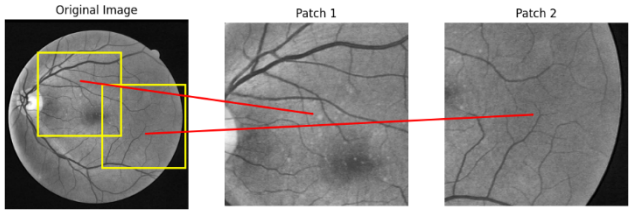


Figure 3. Patching results

2.5. Attention V-Net Architecture

The standard U-Net architecture comprises three main components: an encoder, a bridge, and a decoder, forming a U-shaped structure. Reyes-Figueroa and Rivera [34] introduced a modification that removed the bridge component, resulting in an encoder–decoder structure that visually resembles the letter V. This variant, termed V-Net, connects the encoder directly to the decoder via skip connections, without an intermediate bridge layer. This approach has been adopted in retinal vessel segmentation research, including ResVNet [11] by Ramadhani et al. and VV-Net [12] by Agustina et al [10]. Building on this foundation, this study proposes Attention V-Net, which integrates attention-gating mechanisms into V-Net's skip connections to improve the feature representation of thin, low-contrast retinal blood vessels. Figure 4 presents the proposed Attention V-Net architecture.

The model accepts a 256×256×1 patch as input within a fully convolutional framework (None×None×1), allowing flexible inference across different spatial dimensions. The encoder consists of three sequential feature extraction stages. Stage E1 applies two Conv2D layers (64 filters, 3×3 kernels, same padding, He normal initialization), each followed by Batch Normalization and ReLU activation. Spatial Dropout (rate = 0.3) is applied after the first layer. The stage concludes with 2×2 max-pooling (stride 2), resulting in an E1 feature map of 256×256×64. Stage E2 uses the same structure with 128 filters and 2×2 max-pooling, producing an E2 feature map of 128×128×128. Stage E3 maintains the convolutional structure with 256 filters but omits pooling, yielding an E3 feature map of 64×64×256. This two-downsampling design prevents excessive feature compression and preserves fine structural details of small blood vessels. In accordance with the V-Net definition by Reyes-Figueroa and Rivera [34], no bridge component is included; E3 functions both as the decoder input via upsampling and as the gating signal for the first attention gate.

The decoder consists of two attention-gated upsampling stages. In the first stage (D2), E3 is upsampled using 2×2 UpSampling2D to 128×128×128. An attention gate (Att2) refines the E2 skip connection with E3 as the gating signal. Specifically, E2

is projected via a 2×2 convolution with stride 2 (theta_x, 128 filters), while E3 is projected via a 1×1 convolution (phi_g, 128 filters). The summed projections are processed through ReLU, a 1×1 convolution (psi, 1 filter), and sigmoid activation to generate an attention coefficient map. After 2×2 upsampling, this map is applied element-wise to the E2 features, yielding the refined attention feature att2. The upsampled E3 and att2 are concatenated (128×128×384) and processed by two Conv2D layers (128 filters, 3×3, Batch Normalization, ReLU, Spatial Dropout 0.3), producing the D2 feature map of 128×128×128. In the second stage (D1), D2 is upsampled to 256×256×128, and an attention gate (Att1) refines the E1 skip connection using D2 as the gating signal, following the same procedure as Att2 but with 64 filters for both theta_x and phi_g projections. The upsampled D2 and refined att1 are concatenated (256×256×192) and processed by two Conv2D layers (64 filters, 3×3, Batch Normalization, ReLU, Spatial Dropout 0.3), yielding a D1 feature map of 256×256×64. A final 1×1 convolution with sigmoid activation generates a binary segmentation probability map of 256×256×1.

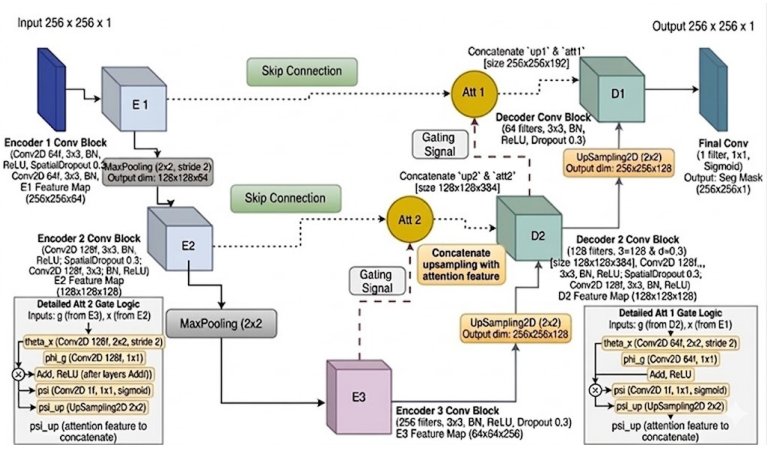


Figure 4. Proposed Attention V-Net Architecture

2.6. Pixel Pitch Calibration

Retinal vessel classification based on size requires pixel-pitch calibration. This process converts vessel diameters from micrometers (μm) to pixel units. In clinical practice, small vessels are defined as those with diameters less than 65 µm [14]. To ensure consistency between clinical definitions and digital image representation, pixel pitch is estimated for each dataset using the field of view (FOV) and anatomical retinal dimensions. In the DRIVE datasets, with a 45^o FOV, the pixel pitch is about 22.40 μm. As a result, a 65 µm vessel corresponds to roughly 3 pixels. In the STARE datasets, the FOV is 35^o, and the pixel pitch is about 16.08 μm. This indicates that a 65 µm vessel spans approximately 4 pixels. Following this calibration, vessels smaller than 3 pixels in the DRIVE datasets and vessels smaller than 4 pixels in the STARE datasets are classified as small. All other vessels are considered large. This method grounds vessel classification in physiologically consistent retinal proportions instead of arbitrary pixel thresholds.

2.7. Blood Vessel Separation

Vessel separation is achieved by classifying vascular structures as small or large vessels based on local diameter estimation. Diameter estimation uses a distance transform applied to the binary vessel image, which measures the distance from each vessel pixel to its nearest background pixel. The maximum distance along the vessel centerline defines the local radius, and the vessel diameter is determined as twice this value. Skeletonization is applied to reduce vessels to a single-pixel width while preserving their topology, thereby producing a stable centerline. In junction regions, points with more than 2 connections are removed to prevent bias in diameter estimation, yielding independent vessel segments. The average diameter of each segment is subsequently calculated from the distance-transform values along the skeleton.

Vessel segments are then classified as small or large using thresholds determined by pixel pitch calibration: less than 3 pixels for the DRIVE datasets and less than 4 pixels for the STARE datasets. To minimize fragmentation of large vessels due to local diameter variations, morphological closing is applied to preserve structural continuity. Area-based filtering is also used to eliminate small objects that are likely to represent noise. This methodology enables vessel separation based on geometric and morphological characteristics, resulting in a classification that is more anatomically consistent than intensity-based methods.

2.8. Experimental Setup

The implementation was written in Python on Google Colab, using an NVIDIA A100 GPU to accelerate training of deep learning models. Experiments were performed using Python 3 and the TensorFlow framework, with support from additional libraries such as Albumentations, OpenCV, NumPy, Pandas, Matplotlib, and Scikit-learn for data preprocessing, analysis, and visualization. In addition to the cloud-based environment, initial processes such as data processing, code management, and documentation were performed on a personal computer (PC) equipped with a 12th-generation Intel Core i7-1260P processor (2.10 GHz), 16 GB of RAM, and Windows 11 Pro 64-bit.

During model training, a batch size of 32 and a maximum of 100 epochs were used. Several optimization strategies were implemented to maintain model performance, including ModelCheckpoint to save the best model based on the validation Dice coefficient, ReduceLROnPlateau to adaptively adjust the learning rate when validation loss stagnated, and EarlyStopping to automatically terminate training if no performance improvement was observed over a specified number of epochs.

2.9. Evaluation Metrics

Performance assessment of medical image segmentation models is essential for verifying the reliability and accuracy of retinal vascular structure delineation. In these tasks, each pixel is assigned to one of two classes: blood vessels (foreground) or background. Segmentation performance is quantified using a confusion matrix, which summarizes classification outcomes as true positives, true negatives, false positives, and false negatives. Six primary evaluation metrics are utilized in this study. Accuracy quantifies the overall correctness of predictions by calculating the proportion of correctly classified pixels across the entire image [35]. Sensitivity measures the model's ability to detect vessel regions by comparing correctly identified vessel pixels to all true vessel pixels [13]. Precision indicates the reliability of vessel predictions by evaluating how many predicted vessel pixels correspond to the ground truth [13]. Specificity evaluates how effectively the model distinguishes background regions from vessel structures, expressed as the proportion of correctly classified background pixels [12]. The Dice Coefficient evaluates spatial overlap between the predicted segmentation and the reference annotation [12]. Intersection over Union (IoU) measures overlap between predicted and reference masks by comparing their intersection and union [35]. The combined use of these metrics enables a comprehensive evaluation, particularly for addressing class imbalance in retinal images, where background pixels greatly outnumber vessel pixels. The evaluation metrics employed in this study are calculated using Equations 5, 6, 7, 8, 9, and 10.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{9}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{10}$$

3. RESULT AND ANALYSIS

3.1. Training Process

The training process is evaluated by analyzing learning curves that depict the progression of loss and Dice coefficient on the DRIVE and STARE datasets. Figure 5(a–d) presents the training and validation loss and Dice coefficient curves for the proposed Attention V Net on the DRIVE and STARE datasets. For the DRIVE datasets (Figures 5(a) and 5(b)), both training and validation losses decrease rapidly during the initial epochs and stabilize after approximately 15 epochs. The Dice coefficient increases steadily and converges, with only a minimal gap between the two curves, which indicates stable training and strong generalization. In contrast,

for the STARE datasets (Figures 5(c) and 5(d)), the training loss is lower than the validation loss, and the training Dice coefficient is higher than the validation Dice. The more pronounced gap between training and validation curves reflects the greater complexity and structural variability of retinal vessels in the STARE datasets. However, the validation curves remain stable and do not diverge, suggesting that the proposed model maintains reliable generalization performance even on more challenging data.

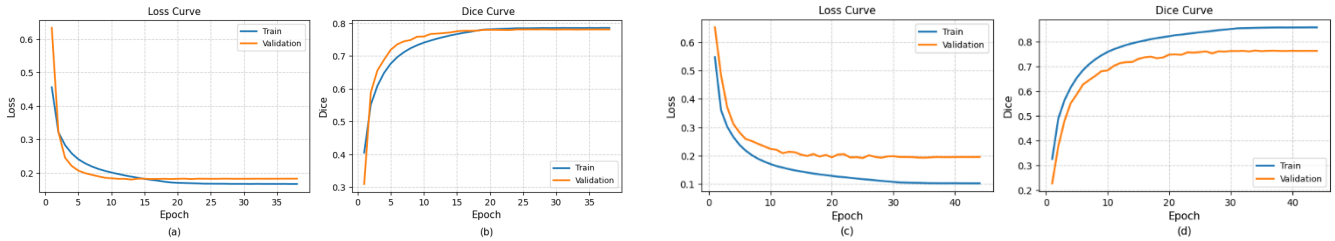


Figure 5. Training curves on the DRIVE and STARE datasets: (a) Loss DRIVE, (b) Dice DRIVE, (c) Loss STARE, (d) Dice STARE

3.2. Blood Vessel Separation Result

Retinal vessels are classified into small and large categories to facilitate a more detailed assessment of segmentation performance. This classification is achieved by estimating local vessel diameter using a distance transform applied to the binary vessel image, followed by projection onto the centerline derived from the skeleton. Diameter thresholds are established using pixel-pitch calibration, with thresholds set at 3 pixels or fewer for the DRIVE datasets and 4 pixels or fewer for the STARE datasets. Figure 6 presents the categorization of retinal vessels into large and small groups for the DRIVE and STARE datasets. The ”Original Ground Truth” displays the reference annotations that include all vascular structures. This separation is conducted exclusively for analytical purposes, allowing vessels to be grouped by size rather than indicating direct outputs of the proposed model. The ”Large Vessels” and ”Small vessels” present the results of dividing the ground truth into two main categories. The overlay offers a combined visualization, with small vessels highlighted in red and large vessels in white. This approach enhances the visualization of the spatial distribution and structural differences among vessel types. Large vessels are visually thicker and more continuous, while small vessels are thinner, more dispersed, and exhibit greater structural complexity.

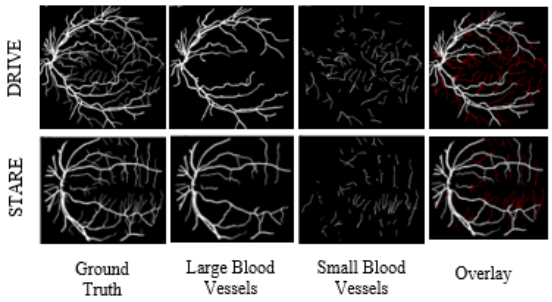


Figure 6. The result of separating large and small blood vessels

3.3. Global Segmentation Result on the STARE and DRIVE Datasets

Global segmentation performance was evaluated for all vascular structures using the DRIVE and STARE datasets. As shown in Table 3, Attention V Net demonstrates consistent segmentation performance across both datasets. On the DRIVE datasets, the model achieves an accuracy of 0.9475, sensitivity of 0.8727, specificity of 0.9587, a Dice score of 0.8080, and an IoU of 0.6781, reflecting a balanced trade off between vessel detection and background suppression. For the STARE datasets, accuracy increases to 0.9602, with corresponding improvements in the Dice score (0.8268) and IoU (0.7065), indicating more stable segmentation performance on data with greater structural variability. These performance differences primarily reflect inherent datasets characteristics, while the overall results confirm consistent model behavior in global evaluation.

Table 3. Global DRIVE and STARE Evaluation Results Attention V-Net

No	Datasets	Acc	Sen	Spe	Pre	Dice	IoU
1	DRIVE	0.9475	0.8727	0.9587	0.7557	0.8080	0.6781
2	STARE	0.9602	0.8719	0.9707	0.7984	0.8268	0.7065

Figure 7 presents retinal vessel segmentation results generated by Attention V-Net on the DRIVE and STARE datasets. The predicted segmentations demonstrate strong concordance with the ground truth, especially in the preservation of major vessels and well-defined branches. Minor discontinuities persist in very thin vessels, highlighting persistent challenges in segmenting small-diameter structures. In summary, the visual comparison suggests that Attention V-Net achieves consistent vessel representation across both datasets without evident loss in segmentation quality.

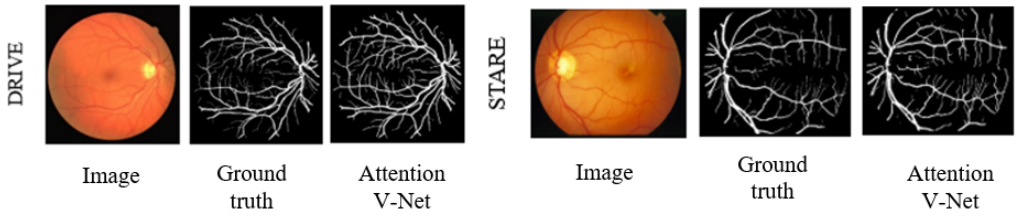


Figure 7. Global attention V-Net segmentation results

3.4. Large Segmentation Result on the STARE and DRIVE Datasets

The evaluation was extended to large vessels to assess the proposed model's ability to segment dominant vascular structures, with particular attention to continuity and structural clarity. The results presented in Table 4 indicate that large vessel segmentation generated by Attention V-Net closely matches the ground truth across both datasets. For the DRIVE datasets, a Dice score of 0.8031 and an Intersection over Union (IoU) of 0.6719 indicate substantial spatial similarity between the predicted large vessels and the reference annotations. The precision of 0.7149 suggests a relatively low false-positive rate. In the STARE datasets, higher Dice (0.8179) and IoU (0.6933), along with improved precision (0.7741), indicate that large-vessel segmentation aligns more closely with the ground truth. These observed differences primarily reflect inherent characteristics of the dataset. Overall, the results confirm the consistent accuracy of Attention V-Net in delineating large-vessel structures.

Table 4. Large DRIVE and STARE Evaluation Results of Attention V-Net

No	Datasets	Acc	Sen	Spe	Pre	Dice	IoU
1	DRIVE	0.9546	0.9219	0.9585	0.7149	0.8031	0.6719
2	STARE	0.9622	0.8851	0.9706	0.7741	0.8179	0.6933

Figure 8 presents representative large-vessel segmentation results generated by Attention V-Net on the DRIVE and STARE datasets. The predicted segmentations demonstrate strong concordance with the ground truth annotations, particularly for major vessels and well-defined branches. Vessel contours are generally smooth and continuous, with only minor discrepancies near certain boundary regions. These results suggest that Attention V-Net accurately represents large-vessel regions with high similarity to the ground truth, and prediction errors are limited and do not substantially affect the overall segmentation outcome.

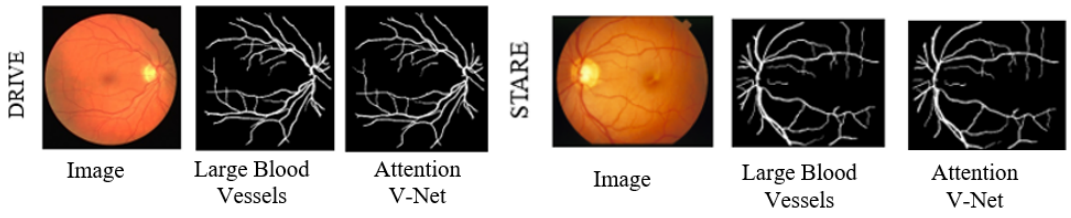


Figure 8. Large Blood Vessel Segmentation Results of Attention V-Net

3.5. Small Segmentation Result on the STARE and DRIVE Datasets

The results for small vessels segmentation in Table 5 underscore the inherent challenges of this task across both datasets. For DRIVE and STARE, Dice scores (0.4720 and 0.4699) and IoU values (0.3096 and 0.3079) are markedly lower than those reported for global and large-vessel evaluations, which reflects the thin, fragmented, and low-contrast characteristics of small vessels. However, sensitivities approaching 0.70 indicate that Attention V-Net can detect a substantial proportion of small vessels, though with limited precision. These results indicate that scale-aware evaluation offers a more realistic and informative assessment of segmentation performance and establish small vessels analysis as the primary contribution of this study.

Table 5. DRIVE and STARE Small Vessels Evaluation Results

No	Datasets	Acc	Sen	Spe	Pre	Dice	IoU
1	DRIVE	0.9505	0.7033	0.9585	0.3595	0.4720	0.3096
2	STARE	0.9504	0.6984	0.9586	0.3586	0.4699	0.3079

Figure 9 presents representative small-vessel segmentation results from Attention V-Net on the DRIVE and STARE datasets, with magnified regions highlighting thin vessel structures. The predicted segmentations capture a subset of low-contrast, small vessels and generally follow the ground-truth patterns, although discontinuities and fragmentation persist in certain areas. These findings underscore the increased challenge of segmenting small vessels compared to larger ones. The visual results align with the scale-aware quantitative evaluation and indicate that emphasizing small vessels yields a more realistic assessment of model performance, which represents the primary contribution of the proposed approach.

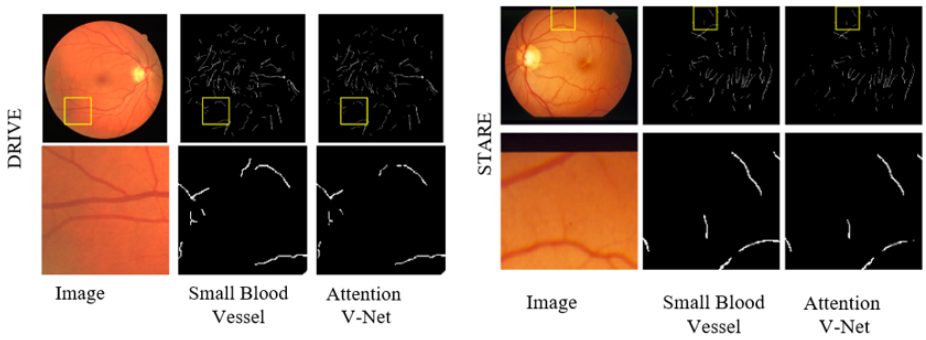


Figure 9. Small Blood Vessel Segmentation Result of Attention

3.6. Discussion

Experimental results indicate that Attention V-Net achieves stable, competitive performance on the DRIVE and STARE datasets. Although overall improvements are modest, scale-based analysis reveals that the most significant differences occur in small-vessel segmentation, which is particularly challenging due to the thinness and low contrast of these structures. In contrast, performance on large vessels remains similar. Comparisons with recent methods further indicate that the proposed model achieves competitive results on key metrics, despite variations in experimental configurations.

Table 6 presents a comparison of small-vessel segmentation performance across the DRIVE datasets. The proposed method shows a higher sensitivity (0.7033) than UN-LPCOS (0.6757), indicating enhanced detection of small-vessel structures. Additionally, the method exhibits strong performance in accuracy (0.9505), specificity (0.9585), Dice coefficient (0.4720), and Intersection over Union (IoU) (0.3096), although these metrics are not reported for UN-LPCOS. While direct comparison is limited, these results indicate that the proposed method effectively captures fine vascular structures.

Table 6. Comparison of DRIVE small vessels evaluation results with other models

No	Model	Acc	Sen	Spe	Pre	Dice	IoU
1	UN-LPCOS [14]	-	0.6757	-	-	-	-
2	Proposed	**0.9505**	**0.7033**	**0.9585**	**0.3595**	**0.4720**	**0.3096**

A comprehensive comparison using the DRIVE datasets Table 7 demonstrates that the proposed Attention V Net achieves competitive results across key evaluation metrics. The model yields Dice and IoU scores marginally higher than those of Performance Aware U Net and Lightweight U Net, while maintaining accuracy comparable to BVU Net and U Net. Minor variations in accuracy and specificity among the methods reflect a trade off between vessel detection and background suppression, aligning with the design objective of enhancing vessel representation. Although experimental settings and metric reporting vary across studies, these findings indicate that the proposed approach delivers balanced and competitive performance in retinal vessel segmentation.

Table 7. Comparison of The Evaluation Results of All DRIVE Vessels With Other Methods

No	Model	Acc	Sen	Spe	Pre	Dice	IoU
1	Lightweight U-Net and Reverse Attention [12]	0.9113	0.7421	**0.9837**	-	0.7871	0.6318
2	BVU-Net [35]	0.9636	0.7871	0.9810	-	**-**	0.65
3	Edge-Aware U-Net [36]	**0.9701**	0.7719	0.9799	-	0.8021	0.6513
4	VV-Net [10]	0.9627	0.8438	-	**0.7595**	-	0.6628
5	ResV-Net [11]	0.9657	0.8228	-	0.7957	-	0.6761
6	Proposed	0.9475	**0.8727**	0.9587	0.7557	**0.8080**	**0.6781**

Table 8 presents a comparison of global segmentation performance on the STARE datasets. The proposed method shows competitive results, achieving a Dice score of 0.8268 and an IoU of 0.7065, surpassing methods such as Edge-Aware U-Net (0.8021) and RCAR U-Net (0.7850). Additionally, it achieves the highest sensitivity (0.8719), which reflects improved detection of vascular structures. Although some methods do not report all metrics, these findings indicate that the proposed model offers balanced performance in both segmentation accuracy and detection capability.

Table 8. Comparison of The Results of The Evaluation of All STARE Vessels With Other Methods

No	Model	Acc	Sen	Spe	Pre	Dice	IoU	AUC
1	Unfolded Deep Kernel Estimation–Attention U-Net [13]	0.9671	0.7591	0.9837	0.7264	0.7228	0.5945	-
2	BVU-Net [35]	0.9639	0.7935	0.9852	-	0.7716	0.63	-
3	RCAR-Unet [37]	0.9594	0.6979	0.9905	-	0.7850	0.6461	0.9708
4	VV-Net [10]	0.9658	0.8278	-	0.7673	-	0.6538	-
5	Edge-Aware U-Net [36]	**0.9691**	0.6912	**0.9911**	-	0.8021	0.6513	-
6	ResV-Net [11]	0.9671	0.7944	-	0.7944	-	0.6505	-
7	Proposed	0.9602	**0.8719**	0.9707	**0.7984**	**0.8268**	**0.7065**	**-**

4. CONCLUSION

The proposed Attention V-Net model demonstrates consistent and competitive performance in retinal vessel segmentation on the DRIVE and STARE datasets. The model achieves robust results in global and large-vessel segmentation, with Dice scores exceeding 0.80 and Intersection over Union (IoU) values up to 0.70, reflecting accurate delineation of major vascular structures. In contrast, segmentation of small vessels remains challenging, as indicated by lower Dice and IoU values, although a sensitivity approaching 0.70 suggests reasonable detection capability. These results underscore that scale-aware evaluation offers a more comprehensive assessment of model performance, particularly by elucidating limitations in small-vessel segmentation. Future research will aim to enhance the continuity of small vessels and assess the proposed approach on additional datasets to further evaluate generalization.

5. ACKNOWLEDGEMENTS

Thanks are extended to all colleagues and organizations who have provided support, assistance, and contributions during the implementation and writing of this research.

6. DECLARATIONS

AI USAGE STATEMENT

During the preparation of this work, the authors used AI tools (e.g., Grammarly, Copilot) to refine language and enhance clarity. No AI software was used to generate the research data or the research findings.

AUTHOR CONTIBUTION

Hendra Wijaya: Conceptualization, Methodology, Writing Original Draft. Erwin: supervision. Annisa Darmawahyuni: supervisor. Sinta Bella Agustina: supervisor.

FUNDING STATEMENT

No specific grant from public, commercial, or not-for-profit funding agencies was received for this research.

COMPETING INTEREST

The authors declare no financial or non-financial conflicts of interest that could have influenced the results or interpretation of the data presented in this article.

REFERENCES

[1] A. Pece, F. Fossataro, and E. V. Longhi, "Vascular and Degenerative Retinal Diseases," in *Managing Psychosexual Consequences in Chronic Diseases*, E. V. Longhi, Ed. Cham: Springer International Publishing, December, 2023, pp. 339–349, https://doi.org/10.1007/978-3-031-31307-3_28.

[2] X. Deng and J. Ye, "A retinal blood vessel segmentation based on improved D-MNet and pulse-coupled neural network," *Biomedical Signal Processing and Control*, vol. 73, p. 103467, Mar. 2022, https://doi.org/10.1016/j.bspc.2021.103467.

[3] J. Dang, P. Parida, and R. Rout, *The Role of Smart Devices for Effective Retinal Disease Diagnosis*, 1st ed. Boca Raton: CRC Press, Oct. 2025, pp. 133–151, https://doi.org/10.1201/9781003560296-9.

[4] N. Chen, Z. Zhu, W. Yang, and Q. Wang, "Progress in clinical research and applications of retinal vessel quantification technology based on fundus imaging," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1329263, Feb. 2024, https://doi.org/10.3389/fbioe.2024.1329263.

[5] N. U. Häner, C. Dysli, and M. R. Munk, "Imaging in retinal vascular disease: A review," *Clinical & Experimental Ophthalmology*, vol. 51, no. 3, pp. 217–228, Apr. 2023, https://doi.org/10.1111/ceo.14203.

[6] R. Kumar and A. K. Bhandari, "Noise reduction deep CNN-based retinal fundus image enhancement using recursive histogram," *Neural Computing and Applications*, vol. 36, no. 27, pp. 17 221–17 243, Sep. 2024, https://doi.org/10.1007/s00521-024-09996-1.

[7] L. Zhao, H. Chi, T. Zhong, and Y. Jia, "Perception-oriented generative adversarial network for retinal fundus image super-resolution," *Computers in Biology and Medicine*, vol. 168, p. 107708, Jan. 2024, https://doi.org/10.1016/j.compbiomed.2023.107708.

[8] S. Iqbal, T. M. Khan, K. Naveed, S. S. Naqvi, and S. J. Nawaz, "Recent trends and advances in fundus image analysis: A review," *Computers in Biology and Medicine*, vol. 151, p. 106277, Dec. 2022, https://doi.org/10.1016/j.compbiomed.2022.106277.

[9] A. A. Abdulsahib, M. A. Mahmoud, H. Aris, S. S. Gunasekaran, and M. A. Mohammed, "An Automated Image Segmentation and Useful Feature Extraction Algorithm for Retinal Blood Vessels in Fundus Images," *Electronics*, vol. 11, no. 9, p. 1295, Apr. 2022, https://doi.org/10.3390/electronics11091295.

[10] S. B. Agustina, E. Erwin, A. Desiani, and T. Saputra, "Blood Vessel Segmentation in Retinal Images Using Convolutional Neural Network VV-Net Method," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 1, pp. 201–209, Feb. 2024, https://doi.org/10.52436/1.jutif.2024.5.1.1723.

[11] S. D. Ramadhani, E. Erwin, A. Desiani, and S. Bella Agustina, "Blood Vessel Segmentation in Retinal Images Using Resvnet Architecture," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 1139–1147, Aug. 2024, https://doi.org/10.52436/1.jutif.2024.5.4.2637.

[12] F. D. Hernandez-Gutierrez, E. G. Avina-Bravo, M. A. Ibarra-Manzano, J. Ruiz-Pinales, E. Ovalle-Magallanes, and J. G. Avina-Cervantes, "Retinal Vessel Segmentation Based on a Lightweight U-Net and Reverse Attention," *Mathematics*, vol. 13, no. 13, p. 2203, Jul. 2025, https://doi.org/10.3390/math13132203.

[13] K. Radha, K. Yepuganti, S. Saritha, C. Kamireddy, and D. P. Bavirisetti, "Unfolded deep kernel estimation-attention UNet-based retinal image segmentation," *Scientific Reports*, vol. 13, no. 1, p. 20712, Nov. 2023, https://doi.org/10.1038/s41598-023-48039-y.

[14] X. Kuang, X. Xu, L. Fang, E. Kozegar, H. Chen, Y. Sun, F. Huang, and T. Tan, "Improved fully convolutional neuron networks on small retinal vessel segmentation using local phase as attention," *Frontiers in Medicine*, vol. 10, p. 1038534, Mar. 2023, https://doi.org/10.3389/fmed.2023.1038534.

[15] T. Long, Y. Xu, H. Zou, L. Lu, T. Yuan, Z. Dong, J. Dong, X. Ke, S. Ling, and Y. Ma, "A Generic Pixel Pitch Calibration Method for Fundus Camera via Automated ROI Extraction," *Sensors*, vol. 22, no. 21, p. 8565, Nov. 2022, https://doi.org/10.3390/s22218565.

[16] "DRIVE - Grand Challenge," https://drive.grand-challenge.org/.

[17] J. Almotiri, K. Elleithy, and A. Elleithy, "Retinal Vessels Segmentation Techniques and Algorithms: A Survey," *Applied Sciences*, vol. 8, no. 2, p. 155, Jan. 2018, https://doi.org/10.3390/app8020155.

[18] T. A. Soomro, A. Ali, N. A. Jandan, A. J. Afifi, M. Irfan, S. Alqhtani, A. Glowacz, A. Alqahtani, R. Tadeusiewicz, E. Kantoch, and L. Zheng, "Impact of Novel Image Preprocessing Techniques on Retinal Vessel Segmentation," *Electronics*, vol. 10, no. 18, p. 2297, Sep. 2021, https://doi.org/10.3390/electronics10182297.

[19] M. Goldbaum, "STructured Analysis of the Retina," https://cecas.clemson.edu/˜ahoover/stare/, 1975.

[20] G. Gojić, V. B. Petrović, D. Dragan, D. B. Gajić, D. Mišković, V. Džinić, Z. Grgić, J. Pantelić, and A. Oros, "Comparing the Clinical Viability of Automated Fundus Image Segmentation Methods," *Sensors*, vol. 22, no. 23, p. 9101, Nov. 2022, https://doi.org/10.3390/s22239101.

[21] M. Mehmood, S. Iqbal, T. M. Khan, I. Spence, and M. Fahim, "LVS-Net: A Lightweight Vessels Segmentation Network for Retinal Image Analysis," 2024, https://doi.org/10.48550/ARXIV.2412.05968.

[22] V. R. Joseph, "Optimal ratio for data splitting," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 4, pp. 531–538, Aug. 2022, https://doi.org/10.1002/sam.11583.

[23] H. R. Maier, F. Zheng, H. Gupta, J. Chen, J. Mai, D. Savic, R. Loritz, W. Wu, D. Guo, A. Bennett, A. Jakeman, S. Razavi, and J. Zhao, "On how data are partitioned in model development and evaluation: Confronting the elephant in the room to enhance model generalization," *Environmental Modelling & Software*, vol. 167, p. 105779, Sep. 2023, https://doi.org/10.1016/j.envsoft.2023.105779.

[24] A. Alsayat, M. Elmezain, S. Alanazi, M. Alruily, A. M. Mostafa, and W. Said, "Multi-Layer Preprocessing and U-Net with Residual Attention Block for Retinal Blood Vessel Segmentation," *Diagnostics*, vol. 13, no. 21, p. 3364, Nov. 2023, https://doi.org/10.3390/diagnostics13213364.

[25] L. Cao and H. Li, "Detail-richest-channel based enhancement for retinal image and beyond," *Biomedical Signal Processing and Control*, vol. 69, p. 102933, Aug. 2021, https://doi.org/10.1016/j.bspc.2021.102933.

[26] A. Kalsi, S. Chawla, and N. Kaur, "Impact of Color Channels on Performance of U-Net for Blood Vessel Segmentation," in *Soft Computing: Theories and Applications*, R. Kumar, A. K. Verma, O. P. Verma, and J. Rajpurohit, Eds. Singapore: Springer Nature Singapore, September, 2025, vol. 1344, pp. 145–155, https://doi.org/10.1007/978-981-96-5958-6_13.

[27] M. S. Z. Ahmad, N. A. A. Aziz, H. S. Lim, A. K. Ghazali, and A. A. Latiff, "Impact of Image Enhancement Using Contrast-Limited Adaptive Histogram Equalization (CLAHE), Anisotropic Diffusion, and Histogram Equalization on Spine X-Ray Segmentation with U-Net, Mask R-CNN, and Transfer Learning," *Algorithms*, vol. 18, no. 12, p. 796, Dec. 2025, https://doi.org/10.3390/a18120796.

[28] K. Alshamrani, H. A. Alshamrani, F. F. Alqahtani, and B. S. Almutairi, "Enhancement of Mammographic Images Using Histogram-Based Techniques for Their Classification Using CNN," *Sensors*, vol. 23, no. 1, p. 235, Dec. 2022, https://doi.org/10.3390/s23010235.

[29] B. Obuchowicz, J. Zarzecka, M. Strzelecki, M. Jakubowska, R. Obuchowicz, A. Piórkowski, E. Zarzecka-Francica, and J. Lasek, "Improving Endodontic Radiograph Interpretation with TV-CLAHE for Enhanced Root Canal Detection," *Journal of Clinical Medicine*, vol. 14, no. 15, p. 5554, Aug. 2025, https://doi.org/10.3390/jcm14155554.

[30] S. Bashir, K. Rohail, F. Sadak, M. Hadi, A. Muneer, M. Ragab, M. Awais, and R. Qureshi, "Exploring the Impact of Preprocessing Techniques on Retinal Blood Vessel Segmentation Using a Study Group Learning Scheme," in *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. Philadelphia, PA, USA: IEEE, Dec. 2023, pp. 1–6, https://doi.org/10.1109/SPMB59478.2023.10372702.

[31] V. Bhutnal and N. R. Moparthi, "Diabetic retinopathy classification using lightweight retinal features extraction from fundus images," *Multimedia Tools and Applications*, vol. 84, no. 42, pp. 50 827–50 848, Oct. 2025, https://doi.org/10.1007/s11042-025-21106-4.

[32] Z. Liu, M. S. Sunar, T. S. Tan, and W. H. W. Hitam, "Deep learning for retinal vessel segmentation: A systematic review of techniques and applications," *Medical & Biological Engineering & Computing*, vol. 63, no. 8, pp. 2191–2208, Aug. 2025, https://doi.org/10.1007/s11517-025-03324-y.

[33] T. Zhang, K. Kasichainula, Y. Zhuo, B. Li, J.-S. Seo, and Y. Cao, "Patch-based Selection and Refinement for Early Object Detection," September, 2023, https://doi.org/10.48550/ARXIV.2311.02274.

[34] A. Reyes-Figueroa and M. Rivera, "W–net: A Convolutional Neural Network for Retinal Vessel Segmentation," vol. 12725, pp. 355–368, June, 2021, https://doi.org/10.1007/978-3-030-77004-4_34.

[35] A. Desiani, E. Erwin, B. Suprihatin, Y. Wahyudi, E. S. Cahyono, and M. Arhami, "BVU-Net: A U-Net Modification by VGG-Batch Normalization for Retinal Blood Vessel Segmentation," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 6, pp. 303–314, Dec. 2022, https://doi.org/10.22266/ijies2022.1231.29.

[36] Y. Zhang, J. Fang, Y. Chen, and L. Jia, "Edge-aware U-net with gated convolution for retinal vessel segmentation," *Biomedical Signal Processing and Control*, vol. 73, p. 103472, Mar. 2022, https://doi.org/10.1016/j.bspc.2021.103472.

[37] W. Ding, Y. Sun, J. Huang, H. Ju, C. Zhang, G. Yang, and C.-T. Lin, "RCAR-UNet: Retinal vessel segmentation network algorithm via novel rough attention mechanism," *Information Sciences*, vol. 657, p. 120007, Feb. 2024, https://doi.org/10.1016/j.ins.2023.120007.