Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

# ResASPP-UNet: A Modified U-NET Using ResNeT ASPP for Retinal Blood Vessels Segmentation

**Salma Salsabila[1)*], Erwin[2)], Anita Desiani[3)]**
[1)2)3)] Sriwijaya University, Indonesia
[1)]salmaslsma@gmail.com, [2)]erwin@unsri.ac.id, [3)]anita_desiani@unsri.ac.id

**Submitted** :Apr 4, 2026 | **Accepted** : Apr 10, 2026 | **Published** : Apr 17, 2026

**Abstract:** In medical imaging, segmenting retinal blood vessels is a crucial task. A ResNet encoder, an ASPP module for multiscale feature extraction, and a UNet decoder comprise the modified U-Net architecture that this work suggests for retinal vascular segmentation. The suggested model extracts the green channel and applies CLAHE during data processing to segment retinal blood vessels. Accuracy, sensitivity, specificity, the Dice coefficient, and Intersection over Union (IoU) are used to assess performance. According to the experimental results, the proposed model obtains an Accuracy, sensitivity, specificity, the Dice coefficient, and IoU of 0.9554, 0.7294, 0.9771, 0.7408, and 0.5884 on the DRIVE dataset and an accuracy of 0.9170 on the DRIVE dataset. Meanwhile, 0.9556, 0.7902, 0.9702, 0.7386, and 0.5865 on the STARE dataset, respectively.

**Keywords:** ASPP, ResNet, Retinal Blood Vessels Segmentation, U-Net.

## INTRODUCTION

The World Health Organization stated that in 2023, more than 1 billion people had visual impairments, 8 million had macular degeneration, 7.7 million had glaucoma, and 3.9 million had diabetic retinopathy (Steinmetz et al., 2021; World Health Organization, 2023). These diseases can be easily diagnosed due to morphological changes in the retinal blood vessels that can be seen from fundus images (Guo et al., 2020). However, early detection and screening still heavily rely on manual examination by trained specialists, time-consuming and difficult to perform (Sule, 2022), and caused human errors due to the complexity and low contrast of fundus images (C. Chen et al., 2021). One approach that can be used to address these limitations is to perform retinal blood vessels segmentation.

Image segmentation of retinal blood vessels aims to separate the blood vessels from the background of the retinal image, thereby enabling the recognition of the boundaries, objects, and structures of the blood vessels in the retinal image (Abdulsahib et al., 2021). Deep learning architecture that performs exceptionally well and capable recognize patterns in medical images is U-Net (Ronneberger et al., 2015).

U-Net is an architecture with a symmetric encoder-decoder structure; it is capable of preserving spatial details, as well as accurately capturing and mapping objects (Ronneberger et al., 2015). The encoder component of U-Net consists of convolutional layers and max-pooling. The encoder functions to reduce resolution and increase feature depth to extract information. In addition to the encoder, there is a decoder consisting of upsampling the feature map followed by concatenation and convolution. The decoder increases image resolution through upsampling. However, the U-Net architecture has shallow learning layers, which sometimes makes feature learning difficult to optimize.(Erwin et al., 2022). This can lead to issues such as gradient vanishing during downsampling (Dash et al., 2024). CNN encoder capable of preventing gradient vanishing is ResNet (L.-K. Huang et al., 2023).

ResNet is able to prevent gradients from vanishing during training because it features residual skip connections that transmit features from previous layers to subsequent layers (Borawar & Kaur, 2023), thereby preventing information loss during downsampling in the encoder. Several studies have used ResNet to mitigate performance degradation (K.-W. Huang et al., 2023; Nisa & Ismail, 2022). However, the downsampling process can result in the loss of some feature information (K.-W. Huang et al., 2023). During downsampling, the network retains only a portion of the feature representation and discards unselected local details during max pooling (Saeedan et al., 2018). To address this, a module that can enhance feature learning capabilities without reducing resolution is the ASPP module (L. Hu et al., 2024; Vijay et al., 2023).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

ASPP (Atrous Spatial Pyramid Pooling) is a multi-scale feature extraction module consisting of atrous convolution blocks at various scales without increasing the number of parameters (L.-C. Chen et al., 2018). By using dilated convolution, atrous convolution blocks enlarge the receptive field by adjusting the kernel spacing without increasing the size or number of parameters (Wei et al., 2018). ASPP is used to expand the receptive field and capture contextual information without reducing spatial resolution (L.-C. Chen et al., 2018). By expanding the receptive field, the network is able to capture broader information and objects (Luo et al., 2016). Several studies have used ASPP to extract features from objects of various sizes simultaneously (Duan et al., 2025; T. Hu et al., 2025; Y. Liu et al., 2024; Shi et al., 2021), enhancing the model's ability to recognize complex and large-scale structures (Tsai et al., 2025), and also to help improve model performance (Cai et al., 2024).

This study proposes a model architecture that utilizes a modified U-Net architecture, which employs ResNet as an encoder to prevent gradient loss during downsampling and feature reduction, an ASPP module to enrich and enlarge features so that object details can be reconstructed effectively, and a U-Net decoder. This study will perform segmentation on retinal blood vessels by extracting the green channel and applying CLAHE during data processing, then evaluating performance using accuracy, sensitivity, specificity, the Dice Coefficient, and IoU. This study uses DRIVE and STARE datasets, which can be accessed at https://drive.grand-challenge.org/ and https://cecas.clemson.edu/~ahoover/stare/.

## LITERATURE REVIEW

Numerous deep learning studies have been conducted to improve evaluation results by modifying various architectures for segmentation in medical images, particularly fundus images. Several studies on the DRIVE and STARE datasets that utilize U-Net modifications with the ASPP module are summarized.

**Table 1** Previous studies using U-Net modifications

| Reference | Method | Results | Limitations |
|---|---|---|---|
| Han et al., 2022 | NoL-UNet: improved U-Net with NoL-Block attention mechanism | DRIVE:<br>Sen: 0.8489<br>AUC: 0.9861<br>Acc: 0.9697 | Need to higher-resolution inputs and improving robustness to color ambiguity |
| Panchal & Kokare, 2024 | Residual Multi-kernel U-Net (ResMU-Net) | DRIVE:<br>Acc: 0.9685<br>Dice: 0.8149<br>AUC: 0.9842 | sensitivity in segmentation still needs to be improved |
| Hussain et al., 2022 | DilUnet: U-Net architecture using improved skip connection and dilated convolution | DRIVE: STARE:<br>Acc: 0.9680 Acc: 0.9694<br>Sen: 0.8837 Sen: 0.8263<br>IoU: 0.8698 IoU: 0.7951 | Specificity needs to be improved and enhancing robustness to severe noise |
| Y. Liu et al., 2023 | a residual DO-conv U-Net (ResDO-UNet) | DRIVE: STARE:<br>Acc: 0.9561 Acc: 0.9635<br>Sen: 0.7985 Sen: 0.8039<br>Spe:0.9791 Spe: 0.9836<br>Dice: 0.8229 IoU: 0.8315 | refining post-processing to reduce false positives |

Modifications of the U-Net design have been thoroughly investigated in recent breakthroughs in retinal vessel segmentation to solve enduring issues such low vessel-to-background contrast, class imbalance, and the detection of small microvascular structures. In order to improve feature representation, a number of research have implemented architectural improvements, such as attention mechanisms, residual connections, and dilated convolutions, as shown in Table 1. Although these methods show gains in metrics like accuracy and AUC, but still need to be improved at segmenting narrow vessels. The very small increases in the Dice coefficient, in particular, indicate that these models have difficulty reliably maintaining vessel continuity and overlap, particularly in areas with low contrast or complex structures.

As shown in Table 2, some research has integrated Atrous Spatial Pyramid Pooling (ASPP) modules into U-Net-based designs to overcome the challenge of gathering multi-scale contextual information. By using parallel dilated convolutions with different receptive fields, ASPP makes multi-scale feature extraction possible and enhances the model's capacity to capture both local and global context. Despite these benefits, boundary preservation and fine structure reconstruction are two areas where ASPP-based techniques frequently fall short. The reported Dice and IoU scores very competitive, but do not consistently beat current techniques, suggesting that multi-scale context by itself is insufficient for precisely defining microvascular borders.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

**Table 2.** Previous studies using U-Net modifications with ASPP modules

| Reference | Method | Results | Limitations |
|---|---|---|---|
| Duan et al., 2025 | Deformable U-Net with Atrous-Convolution Feature Pyramid | DRIVE:<br>Acc: 0.9572<br>Dice: 0.8298 | The STARE dataset was not used as the research subject, and computation remains complex |
| Shi et al., 2021 | Multi-scale dense network (MD-Net) with an additional atrous spatial pyramid pooling residual module | DRIVE: STARE:<br>Acc: 0.9676 Acc: 0.9732<br>Dice: 0.8099 Dice: 0.8411<br>Sen: 0.8065 Sen: 0.8290<br>Spec: 0.9826 Spec: 0.9866 | The architecture used still requires high computational costs |
| Shi et al., 2020 | Multi-Scale U-Net with an additional ASPP module | DRIVE: STARE:<br>Acc: 0.9667 Acc: 0.9732<br>Sen: 0.8159 Sen: 0.8272<br>Spec: 0.9805 Spec: 0.9866<br>Dice: 0.8059 Dice: 0.8400 | Model performance needs to be improved |
| Y. Liu et al., 2024 | MobileViTv2-ResUNet with an additional ASPP module | DRIVE:<br>Acc: 79.67<br>IoU: 69.57<br>F1-score: 82.06<br>Recall: 79.67 | Accuracy in segmentation still needs to be improved |

Research that applies ResNet as part of the encoder can be seen in Table 3.

**Table 3.** Previous studies using Residual Networks in the encoder

| Reference | Method | Results | Limitations |
|---|---|---|---|
| Khan et al., 2020 | Residual Connection-Based Encoder Decoder Network (RCED-Net) for Retinal Vessel Segmentation | DRIVE: STARE:<br>Sen: 0.8252 Sen: 0.9780<br>Spec: 0.8440 Spec: 0.9830<br>Acc: 0.8397 Acc: 0.9810 | Binary masks are not available for the STARE dataset |
| Lian et al., 2021 | GLUE: Global and Local Enhanced Residual U-Net | DRIVE: STARE:<br>Acc: 0.9692 Acc: 0.9740<br>Sen: 0.8278 Sen: 0.8342<br>Spec: 0.9861 Spec: 0.9916 | This method does not utilize local patch information. |
| Ramadhani et al., 2024 | a modification of V-Net with a ResNet architecture in the encoder section | DRIVE: STARE :<br>Acc: 0.9657 Acc: 0.9671<br>Sen: 0.8228 Sen: 0.7944<br>Pre: 0.7957 Pre: 0.7944<br>Jaccard: 0.6761 Jaccard: 0.6505 | Model performance needs to be improved |
| Z. Liu et al., 2020 | U-Net with ResNet as the encoder | DOM:<br>mIoU: 0.84 | Model performance needs to be improved |
| Nisa & Ismail, 2022 | Dual U-Net with ResNet Encoder | Kvasir-SEG:<br>Dice: 0.8715<br>IoU: 0.8042 | Accuracy in segmentation still needs to be improved |

As seen in Table 3, additional enhancements have been investigated by including Residual Networks (ResNet) as encoder backbones. By reducing vanishing gradient problems and enabling deeper feature extraction, residual connections increase generalization and training stability. However, ResNet-based encoders are not specifically geared to capture the elongated and tubular shapes typical of retinal arteries, but built for generic feature extraction. Then, these models frequently fall short of accurately simulating fine-grained vessel architecture, especially in low-contrast areas. Furthermore, some methods fail to effectively utilize local spatial information, which is essential for accurate small vessel segmentation.

More significantly, microvascular continuity is still not preserved by current methods, especially in low-contrast areas and substantially unbalanced contexts. This suggests that deep feature extraction using ResNet or multi-scale context modeling using ASPP are insufficient to handle the intrinsic complexity of retinal vascular segmentation.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This paper suggests a hybrid design that combines a ResNet-based encoder, an ASPP module, and a U-Net-style decoder to overcome these problems. By combining multi-scale context modeling, accurate spatial reconstruction, and deep hierarchical feature extraction, this method improves the segmentation of thin and low-contrast vessels.

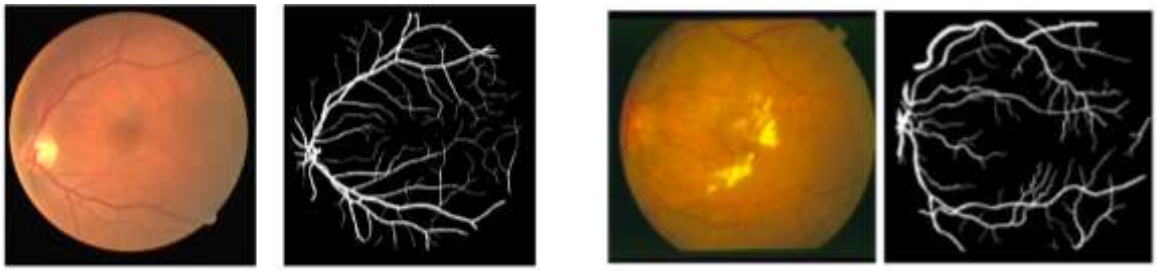
## METHOD

This study was conducted using a Personal Computer (PC) with processor Intel(R) Core (TM) i7-8665U CPU @ 1.90 GHz  2.11 GHz, RAM 48.0 GB. In addition to using this hardware, this study utilized a web-based computing platform, namely Google Colab (NVIDIA Tesla T4 GPU) and Python programming language.

This study was conducted in several stages, starting with a literature review, data collection and preparation, data preprocessing, architecture design, data training and testing, and model performance evaluation.

### Data Collection and Preparation

This study used secondary data, where datasets consisting of fundus photographs or retinal blood vessel images were used for segmentation. The datasets used were the DRIVE and STARE datasets.



(a) (b)

**Fig 1**. (a) DRIVE dataset (b) STARE dataset. RGB image (left) and manually segmented image (right)

### DRIVE Dataset

The DRIVE dataset (Digital Retinal Image for Vessel Extraction) is one of the datasets used for retinal blood vessel segmentation. This dataset contains 40 fundus images compressed into JPEG format. The images in this dataset have dimensions of 564x584 pixels and are divided into two parts: 20 images each for training and testing. This dataset includes masks and manual segmentation performed by ophthalmologists. Fig. 1(a) shows one of the fundus images and the manually segmented image

### STARE Dataset

The STARE dataset, or Structured Analysis of the Retina, is also one of the datasets used in this study for retinal blood vessel image segmentation. This dataset consists of 20 fundus images and 20 manually labeled images. The images in portable pixmap (PPM) format and have a resolution of 700x605 pixels. The images are divided into two parts: 16 images each for training and and 4 images for testing.  One of the images in the dataset is shown in Fig. 1(b).

### Data Preprocessing

This stage is the initial step before the dataset containing RGB fundus images and blood vessel masks undergo the training process. The fundus images in the DRIVE and STARE datasets are converted to RGB format, the green channel is extracted, and image quality is enhanced using CLAHE.

Next, the data was augmented by applying image variations with several visual transformations—including flipping, rotation, and brightness-contrast adjustments—to increase the dataset size. The augmented images were then combined with the original dataset and loaded into the data loader.

### Architecture Design

In this study, the ResASPP-Unet architecture is used to perform segmentation on retinal blood vessel images. This architecture is designed to combine the strengths of a residual network acting as an encoder, Atrous Spatial Pyramid Pooling (ASPP) acting as a bridge, and a decoder based on the U-Net architecture. This architecture is expected to improve segmentation performance, particularly in processes requiring multi-scale context and smooth boundary delineation. Fig. 2 illustrates the ResASPP-Unet architecture used in this study.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

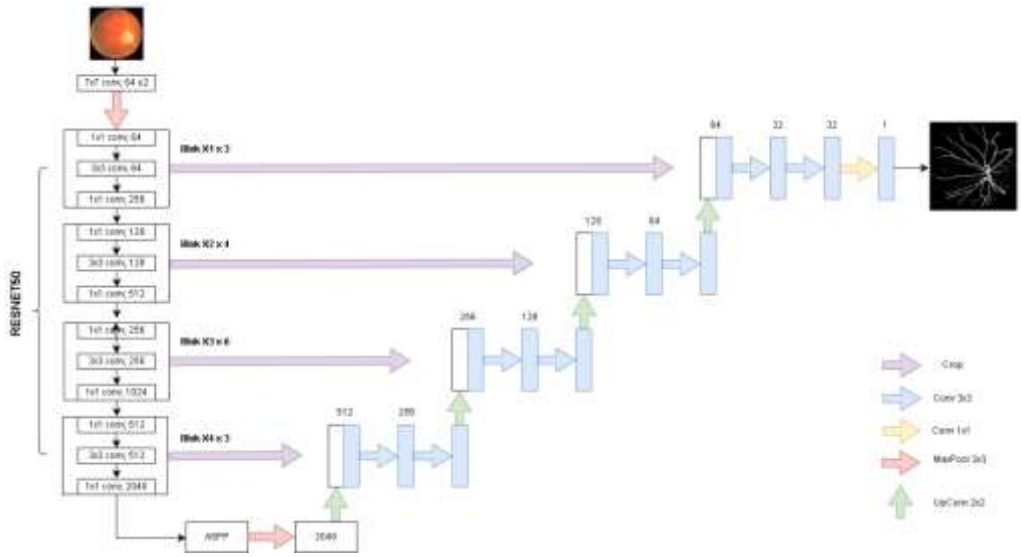



**Figure 2** ResASPP-UNet Architecture

The residual network consists of 4 stages comprising 3 (block x1), 4 (block x2), 6 (block x3), and 3 (block x4) block lengths, where each block consists of a convolutional layer. Each block produces different features, starting from 256, 512, 1024, and 2048. The output from each of these blocks will also be used as a skip connection in the decoder section to reconstruct information regarding image features.

Afterward, the 2048-channel features from the encoder are processed through the ASPP. In this process, the ASPP module performs 1x1 convolutions, followed by dilated convolutions with dilations of 6, 12, and 18 to capture context at various scales by expanding the receptive field without reducing the features in the channels. Image-level pooling is then performed, and all paths are combined via concatenation. The output from the ASPP is subsequently downsampled to 256 channels for further processing in the decoder.

In the decoder section, there are 4 upsampling stages, with each stage, feature maps are upsampled and fused with corresponding encoder features via skip connections. Each block in the decoder consists of two 3x3 convolutional layers and a ReLU activation function. Throughout all decoding phases, this procedure is performed hierarchically, thereby improving spatial resolution while decreasing semantic abstraction. Subsequently, a 1x1 convolution is performed in the final stage to produce a single channel, and a sigmoid activation function is applied to yield binary segmentation results.

**Data Training and Testing**

In this study using the ResASPP-UNet architecture for training. During the training process, the parameters used were a batch size of 2, a learning rate of 0.0001, 100 epochs, the Adam optimizer, and the Hybrid Loss function which consist of Dice, Boundary, and Topology loss. These parameters were applied to each dataset, namely the DRIVE and STARE datasets. Then, evaluate generalization using cross-dataset validation, and apply 10-fold cross-validation on the DRIVE dataset, and 5-fold cross-validation on the STARE dataset to evaluate model performance.

**Model Performance Evaluation**

After training and testing the data on retinal blood vessel images, the next step is to evaluate the performance results of the ResASPP-UNet model. This performance evaluation is calculated from the confusion matrix obtained during the training and testing processes. The confusion matrix values are used to measure accuracy using equation (1), sensitivity using equation (2), specificity using equation (3), Dice using equation (4), and IoU using equation (5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Spesificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (3)$$

$$\text{Dice} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (4)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (5)$$

The model performance evaluation results are analyzed to assess how well the proposed architecture performs. These results will then be compared with previous studies to evaluate the model's effectiveness and performance.

## RESULT

### Data Description

This study was conducted on retinal blood vessel images using secondary datasets, namely the DRIVE and STARE datasets. The DRIVE dataset is a dataset used for retinal blood vessel segmentation, available at https://drive.grand-challenge.org/. This dataset consists of 40 color fundus images with a resolution of 565 × 584 pixels, divided into two subsets: 20 images for training and 20 images for testing.

In addition to the DRIVE dataset, there is the STARE dataset which is also used for retinal blood vessel segmentation. The STARE dataset can be accessed via the website https://cecas.clemson.edu/~ahoover/stare/ and consists of 20 retinal images with a resolution of 700 × 605 pixels. The images are divided into two parts: 16 images each for training and and 4 images for testing.

### Data Preprocessing Steps

Before the model training process, the retinal images undergo data preprocessing and augmentation. Preprocessing is performed to improve image quality and clarify the structure of the blood vessels. In this study, the data was processed using green channel extraction and Contrast Limited Adaptive Histogram Equalization (CLAHE), as shown in Fig. 3.

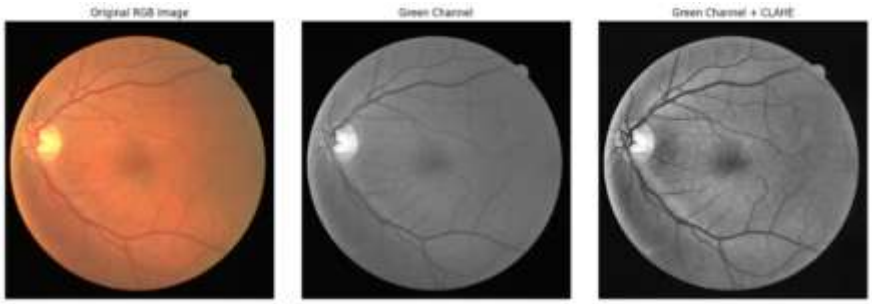



**Figure 3** Original image in RGB format and image after green channel extraction and CLAHE

Green channel extraction, which has the clearest contrast compared to the red and blue channels, is performed to help the model better identify blood vessel structures. Next, contrast enhancement was performed using the CLAHE method. This method works by increasing local contrast in the image, making small blood vessel structures more clearly visible. The application of this preprocessing aims to help the model extract blood vessel features more effectively, especially for very small blood vessels.

After preprocessing, data augmentation was performed using flipping (horizontal and vertical), rotation, and adjustments to brightness and contrast, as seen in Fig. 4. This data augmentation was carried out by applying variations of the image with different visual orientations to the data to increase the dataset size.

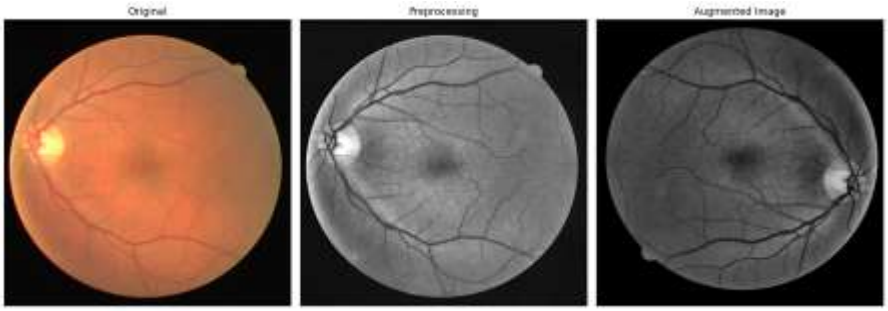



**Figure 4** Fundus photos before and after the data preprocessing stage

### Retinal Blood Vessels Segmentation Using the ResASPP-U-Net Architecture

In this study, segmentation was performed by modifying the U-Net architecture, which is frequently used in medical image segmentation. In this study, modifications were made to the encoder section using ResNet, the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

addition of an ASPP module at the beginning of the bridge, and followed by a U-Net-based decoder. This model is named ResASPP-UNet.

The training process using ResASPP-UNet begins with a ResNet50-based encoder acting as a multilevel feature extractor with hierarchical outputs from layer 1 to layer 4. Subsequently, ASPP functions to capture multi-scale context through variations of distorted convolutions. The features generated by ASPP are then gradually reconstructed by the U-Net decoder through an upsampling mechanism and connected via skip connections. This helps preserve spatial details during retinal blood vessel image segmentation.

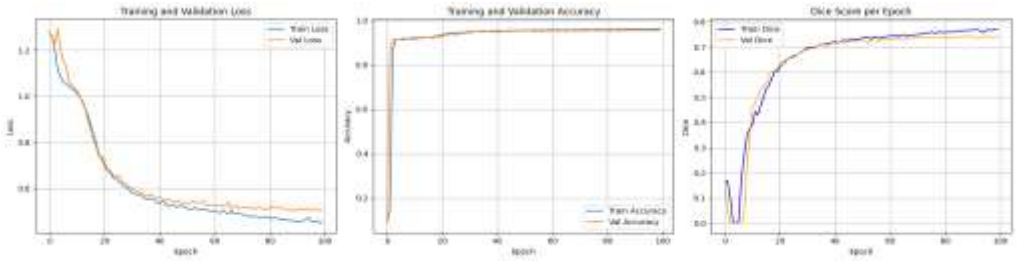
Regarding the configuration used during the training process, the loss function employed is the Hybrid Loss function which consist of Dice, Boundary, and Topology loss. The optimizer is the Adam optimizer with a learning rate of 0.0001, and the training process runs for 100 epochs. During the training process, evaluate generalization using cross-dataset validation and the results of retinal blood vessel segmentation using ResASPP-U-Net were obtained in evaluation metrics, such as accuracy, sensitivity, specificity, Dice score, and IoU.

**Experimental Results**

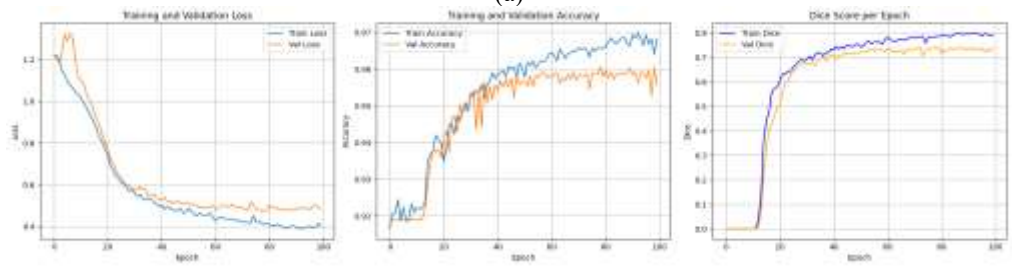
The experimental results during training can be observed through the training loss curve, validation loss curve, training accuracy curve, validation accuracy curve, and Dice score curve, as seen in Fig. 5. These curves were obtained for each epoch during the training process. Fig. 5(a) and Fig. 5(b) show the results obtained on the DRIVE and STARE datasets

Both training and validation losses for the DRIVE dataset (Fig. 5(a)) decline gradually and converge without appreciable divergence, suggesting stable optimization and favorable convergence behavior. During the early epochs, the accuracy grows quickly and stabilizes above 0.94, indicating rapid convergence. Although there is a slight difference between the training and validation curves, suggesting modest overfitting, the Dice score likewise steadily improves. This disparity, however, is still under control, suggesting that the model learns significant vascular structures while maintaining appropriate generalization.

The model shows a similar but marginally more dynamic behavior for the STARE dataset (Fig. 5(b)). The loss curves exhibit slight variations during the validation stage, indicating sensitivity to data variability, but they consistently converge. When compared to DRIVE, the accuracy attains higher values, suggesting increased learning efficacy. Notably, improved segmentation performance is reflected in the Dice score, which rises more quickly and stabilizes at a higher level. Even though there is still a little discrepancy between training and validation metrics, this suggests high stability with little overfitting. Overall, the model performs better on the STARE dataset and exhibits quick convergence, stable training dynamics, and consistent generalization across both datasets.



(a)



(b)

**Figure 5** Training and validation loss, training and validation accuracy, and Dice score per epoch for (a) the DRIVE dataset (b) the STARE dataset

After the training process was completed, several metrics commonly used in medical image segmentation research were obtained. The metrics used were accuracy, sensitivity, specificity, Dice score, and IoU. In this study, two preprocessing configurations were used: one without and one using green channel extraction and CLAHE.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Based on Table 4, the use of green channel preprocessing and CLAHE can improve the performance of the model applied to the DRIVE dataset. The sensitivity increased to 0.7294 and Dice score increased to 0.7408, indicating a good match between the model's segmentation results and the ground truth. Additionally, the sensitivity value increased significantly, indicating that the model can detect retinal blood vessels more effectively. This improvement indicates that the preprocessing used is capable of enhancing the contrast of the blood vessels, making them easier for the model to recognize.

**Table 4.** Evaluation results for segmentation using ResASPP-UNet on the DRIVE dataset

| Preprocessing | Dataset | Accuracy | Sensitivity | Specificity | Dice | IoU |
|---|---|---|---|---|---|---|
| × | DRIVE | 0.9545 | 0.6820 | 0.9806 | 0.7238 | 0.5674 |
| Green channel + CLAHE | DRIVE | 0.9554 | 0.7294 | 0.9771 | 0.7408 | 0.5884 |

**Table 5.** Evaluation results in segmentation evaluation metrics using ResASPP-UNet on the STARE dataset

| Preprocessing | Dataset | Accuracy | Sensitivity | Specificity | Dice | IoU |
|---|---|---|---|---|---|---|
| × | STARE | 0.9572 | 0.7586 | 0.9747 | 0.7344 | 0.5822 |
| Green channel + CLAHE | STARE | 0.9556 | 0.7902 | 0.9702 | 0.7386 | 0.5865 |

The experimental results on the STARE dataset shown in Table 5 indicate that the proposed model achieved a sensitivity of 0.7902 and Dice score of 0.7386 on the STARE dataset. These values demonstrate that the model can produce fairly accurate blood vessel segmentation and exhibits high agreement with the ground truth.

**Comparison with State-of-the-Art Research**

Retinal blood vessel segmentation has been performed in previous studies using various methods. A comparison of the results of previous studies with the results of this study using the proposed model can be seen in the Table 6 and Table 7.

**Table 6.** Comparison of segmentation results from previous studies on the DRIVE dataset

| References | Method | Accuracy | Sensitivity | Specificity | Dice | IoU |
|---|---|---|---|---|---|---|
| Duan et al., 2025 | Deformable U-Net with Atrous-Convolution Feature Pyramid | 0.9572 | - | - | 0.8298 | - |
| Shi et al., 2021 | Multi-scale dense network (MD-Net) with an additional atrous spatial pyramid pooling residual module | 0.9676 | 0.8065 | 0.9826 | 0.8099 | - |
| Y. Liu et al., 2024 | MobileViTv2-ResUNet with an additional ASPP module | 0.7967 | 0.7967 | - | 0.8206 | 0.6957 |
| Khan et al., 2020 | Residual Connection-Based Encoder Decoder Network (RCED-Net) for Retinal Vessel- -Segmentation | 0.8397 | 0.8252 | 0.8440 | - | - |
| Lian et al., 2021 | GLUE: Global and Local Enhanced Residual U-Net | 0.9692 | 0.8278 | 0.9861 | | -- |
| Ramadhani et al., 2024 | V-Net modified with ResNet architecture in the encoder section | 0.9657 | 0.8228 | - | - | 0.6761 |
| Proposed Model | U-Net modification using ResNet and ASPP modules | 0.9554 | 0.7294 | 0.9771 | 0.7408 | 0.5884 |

**Table 7.** Comparison of segmentation results from previous studies on the STARE dataset

| References | Method | Accuracy | Sensitivity | Specificity | Dice | IoU |
|---|---|---|---|---|---|---|
| Shi et al., 2021 | Multi-scale dense network (MD-Net) with an additional atrous spatial pyramid pooling residual module | 0.9732 | 0.8290 | 0.9866 | 0.8411 | - |
| Khan et al., 2020 | Residual Connection-Based Encoder Decoder Network (RCED-Net) for Retinal Vessel Segmentation | 0.9810 | 0.9780 | 0.9830 | - | - |
| Lian et al., 2021 | GLUE: Global and Local Enhanced Residual U-Net | 0.9740 | 0.8342 | 0.9916 | - | - |
| Ramadhani et al., 2024 | V-Net modified with ResNet architecture in the encoder section | 0.9671 | 0.7944 | - | - | 0.6505 |

*name of corresponding author

| | | | | | | |
|---|---|---|---|---|---|---|
| Proposed Model | U-Net modification using ResNet and ASPP modules | 0.9556 | 0.7902 | 0.9702 | 0.7386 | 0.5865 |

According to the findings shown in Tables 6 and 7, the suggested ResASPP-UNet performs consistently throughout the DRIVE and STARE datasets, obtaining accuracy values of 0.9554 and 0.9556, respectively. Additionally, the model achieves high specificity (0.9771 on DRIVE and 0.9702 on STARE), demonstrating its remarkable ability to minimize false positives and accurately detect background regions. Nonetheless, the sensitivity on the STARE dataset (0.7902) is higher than that on the DRIVE dataset (0.7294), indicating that the model is more effective at identifying vessels in the STARE dataset.

Despite these advantages, the suggested model produces comparatively lower Dice and IoU scores when compared to State-of-the-art, suggesting that the overlap between the ground truth and anticipated segmentation is still not ideal. The model's a reduced ability to capture detailed vessel architecture is probably the cause of this limit. Although ResASPP-UNet exhibits robustness and high background detection overall, more work is needed to increase segmentation precision, especially for small and complex vessel architectures.

**Ablation Study**

An ablation study was conducted to compare various model configurations. This process began with U-Net, Res-UNet, and then ResASPP-U-Net.

**Table 8.** Comparison of evaluation results using the model with and without modifications on the DRIVE dataset

| Model | Dataset | Accuracy | Sensitivity | Specificity | Dice | IoU |
|---|---|---|---|---|---|---|
| UNet | DRIVE | 0.9517 | 0.6727 | 0.9784 | 0.7083 | 0.5485 |
| Res-UNet | DRIVE | 0.9500 | 0.7196 | 0.9720 | 0.7153 | 0.5569 |
| ResASPP-UNet | DRIVE | 0.9554 | 0.7294 | 0.9771 | 0.7408 | 0.5884 |

**Table 9.** Comparison of evaluation results using the model with and without modifications on the STARE dataset

| Model | Dataset | Accuracy | Sensitivity | Specificity | Dice | IoU |
|---|---|---|---|---|---|---|
| UNet | STARE | 0.9460 | 0.7209 | 0.9650 | 0.6730 | 0.5080 |
| Res-UNet | STARE | 0.9554 | 0.7402 | 0.9736 | 0.7199 | 0.5632 |
| ResASPP-UNet | STARE | 0.9556 | 0.7902 | 0.9702 | 0.7386 | 0.5865 |

The results as seen in Table 8 and Table 9. shows that the ResNet encoder improves performance compared to the standard U-Net. The addition of the ASPP module provides a more significant improvement in performance, indicating that multi-scale features are crucial in retinal blood vessel segmentation.

In addition to the evaluation metrics, Fig. 6 present visual segmentation results by comparing the original image, ground truth, and model predictions in detecting retinal blood vessel structures on the DRIVE and STARE datasets using ResASPP-U-Net.

A qualitative comparison of the original retinal images, ground truth masks, and the expected segmentation outcomes from the suggested ResASPP-UNet model on the DRIVE and STARE datasets is shown in Figure 6. The model can accurately depict the major blood vessels and represent the major vascular features with good continuity. Particularly in thicker vessel regions and principal branches, the projected masks exhibit a high degree of closeness to the ground reality.

However, there are some differences in the segmentation of finer vessels, where some low-contrast, thin structures are either overlooked or only partially identified. Furthermore, a few isolated false positives show up in the background areas. The model shows a significant ability to maintain the overall vascular topology and structural consistency in spite of these drawbacks. These findings show that the suggested ResASPP-UNet is useful for segmenting retinal arteries, particularly large veins, but more work is required to accurately capture small vascular features.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

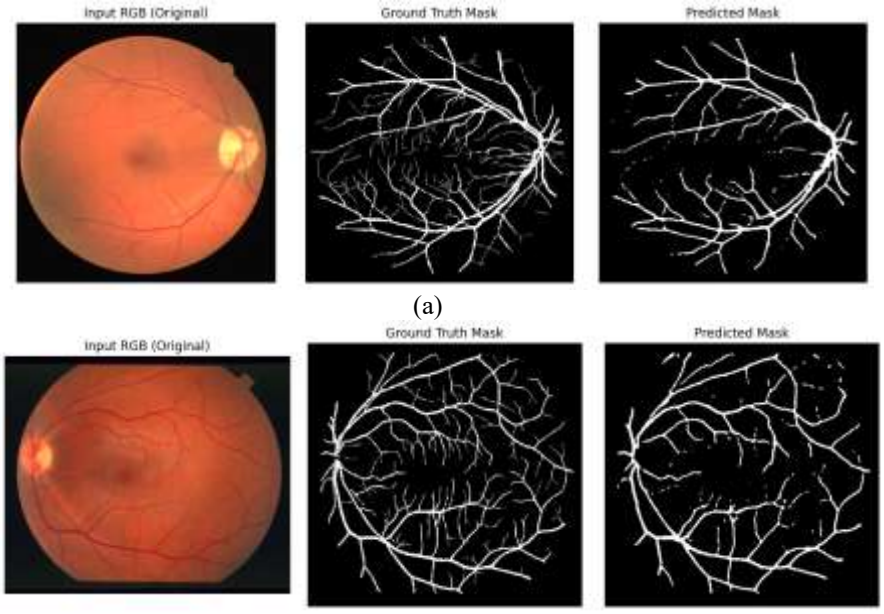



**Figure 6** Original image segmentation results, ground truth, and prediction results using ResASPP-UNet on the (a) DRIVE dataset (b) STARE dataset

## DISCUSSIONS

According to the experimental findings, ResASPP-UNet performs better in Dice, sensitivity, and IoU than the conventional U-Net. This improvement is a reflection of both the architectural design's efficacy and empirical gains. Improved Dice and IoU point to a more balanced segmentation performance, while increased sensitivity shows improved thin vessel recognition by lowering false negatives. By using residual connections to increase gradient flow and facilitate deeper feature learning, the encoder ResNet50 improves feature representation.

Furthermore, the ASPP module uses dilated convolutions to expand the receptive field, strengthening multi-scale feature extraction. This makes it possible for the model to incorporate both local features and global context, which is crucial for segmenting vessels of different sizes. The suggested model offers a more explicit multi-scale representation than U-Net, which mostly uses downsampling. However, low contrast, loss of fine details during downsampling, and class imbalance continue to be obstacles to the detection of micro-vessels. ASPP does not particularly handle fine boundary details, even though it enhances contextual comprehension. Overall, the findings show that ResASPP-UNet enhances segmentation performance; nevertheless, other improvements, such as boundary-aware techniques or attention mechanisms, are required to better catch extremely small vessels.

## CONCLUSION

This study presented a modified U-Net architecture for retinal vascular segmentation that included an Atrous Spatial Pyramid Pooling (ASPP) module and a ResNet encoder. The experimental findings show that the suggested model performs well and consistently on the STARE and DRIVE datasets. The suggested model obtained an accuracy of 0.9170, sensitivity of 0.8798, specificity of 0.9857, Dice coefficient of 0.8332, and IoU of 0.7144 on the DRIVE dataset. In the meantime, the model obtained an accuracy of 0.9272, sensitivity of 0.8868, specificity of 0.9922, Dice coefficient of 0.8550, and IoU of 0.7469 on the STARE dataset. It introduced ResASPP-UNet, which integrates a ResNet encoder and ASPP module for retinal vascular segmentation. This suggests that major issues like low contrast and class imbalance are only partially addressed by the approach. The obtained Dice and IoU scores, however, indicate that the model does not yet substantially outperform cutting-edge techniques. While ASPP improves multi-scale context, a critical study shows that ResNet is not specifically made for tubular structure modeling, and ASPP is constrained in maintaining fine boundaries.

As a result, the model continues to have issues with micro-vessel continuity and generates false positives in background areas. These results demonstrate that fine-grained segmentation requires more than just better feature representation. Future research should concentrate on larger or more diverse retinal datasets and structure-aware techniques including transformer-based encoders, boundary refinement, and topology-aware loss, as well as more thorough evaluation that includes statistical analysis and thorough failure scenario assessment and to further optimize performance.

*name of corresponding author

## REFERENCES

Abdulsahib, A. A., Mahmoud, M. A., Mohammed, M. A., Rasheed, H. H., Mostafa, S. A., & Maashi, M. S. (2021). Comprehensive review of retinal blood vessel segmentation and classification techniques: intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *10*(1), 20. https://doi.org/10.1007/s13721-021-00294-7

Borawar, L., & Kaur, R. (2023). ResNet: Solving vanishing gradient in deep networks. *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022*, 235–247.

Cai, J., Shi, J., Leau, Y.-B., Meng, S., Zheng, X., & Zhou, J. (2024). Res50-SimAM-ASPP-Unet: A Semantic Segmentation Model for High-resolution Remote Sensing Images. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2024.3519260

Chen, C., Chuah, J. H., Ali, R., & Wang, Y. (2021). Retinal Vessel Segmentation Using Deep Learning: A Review. *IEEE Access*, *9*, 111985–112004. https://doi.org/10.1109/ACCESS.2021.3102176

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Dash, D., Mahapatra, S., & Agrawal, S. (2024). A Novel Enhanced Res-Unet Model for Retinal Blood Vessels Segmentation. *International Conference on Intelligent Computing and Advances in Communication*, 57–68.

Duan, Y., Yang, R., Zhao, M., Qi, M., & Peng, S.-L. (2025a). DAF-UNet: Deformable U-Net with Atrous-Convolution Feature Pyramid for Retinal Vessel Segmentation. *Mathematics*, *13*(9), 1454. https://doi.org/10.3390/math13091454

Duan, Y., Yang, R., Zhao, M., Qi, M., & Peng, S.-L. (2025b). DAF-UNet: Deformable U-Net with Atrous-Convolution Feature Pyramid for Retinal Vessel Segmentation. *Mathematics*, *13*(9), 1454. https://doi.org/10.3390/math13091454

Erwin, A. D., Wahyudi, B. S. Y., Cahyono, E. S., & Arhami, M. (2022). BVU-Net: A U-net modification by VGG-batch normalization for retinal blood vessel segmentation. *Int. J. Intell. Eng. Syst.*

Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2020). SA-UNET: Spatial attention U-net for retinal vessel segmentation. *Proceedings - International Conference on Pattern Recognition*, 1236–1242. https://doi.org/10.1109/ICPR48806.2021.9413346

Han, J., Wang, Y., & Gong, H. (2022). Fundus retinal vessels image segmentation method based on improved U-Net. *IRBM*, *43*(6), 628–639.

Hu, L., Zhou, X., Ruan, J., & Li, S. (2024). ASPP+-LANet: A Multi-Scale Context Extraction Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *Remote Sensing*, *16*(6), 1036.

Hu, T., Ding, J., Liu, Y., Zhang, Y., & Yang, L. (2025). DAA-UNet: A Dense Connectivity and Atrous Spatial Pyramid Pooling Attention UNet Model for Retinal Optical Coherence Tomography Fluid Segmentation. *IET Software*, *2025*(1). https://doi.org/10.1049/sfw2/6006074

Huang, K.-W., Yang, Y.-R., Huang, Z.-H., Liu, Y.-Y., & Lee, S.-H. (2023). Retinal Vascular Image Segmentation Using Improved UNet Based on Residual Module. *Bioengineering*, *10*(6), 722. https://doi.org/10.3390/bioengineering10060722

Huang, L.-K., Tseng, H.-T., Hsieh, C.-C., & Yang, C.-S. (2023). Deep learning based text detection using resnet for feature extraction. *Multimedia Tools and Applications*, *82*(30), 46871–46903. https://doi.org/10.1007/s11042-023-15449-z

Hussain, S., Guo, F., Li, W., & Shen, Z. (2022). DilUnet: A U-net based architecture for blood vessels segmentation. *Computer Methods and Programs in Biomedicine*, *218*, 106732.

Khan, T. M., Alhussein, M., Aurangzeb, K., Arsalan, M., Naqvi, S. S., & Nawaz, S. J. (2020). Residual Connection-Based Encoder Decoder Network (RCED-Net) for Retinal Vessel Segmentation. *IEEE Access*, *8*, 131257–131272. https://doi.org/10.1109/ACCESS.2020.3008899

Lian, S., Li, L., Lian, G., Xiao, X., Luo, Z., & Li, S. (2021). A Global and Local Enhanced Residual U-Net for Accurate Retinal Vessel Segmentation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *18*(3), 852–862. https://doi.org/10.1109/TCBB.2019.2917188

Liu, Y., Fu, Y., & Wen, T. (2024). MobileViTv2-ResUNet: U-shaped transformer architecture for precise retinal vessel segmentation. In Y. Yue (Ed.), *International Conference on Optics, Electronics, and Communication Engineering (OECE 2024)* (p. 13). SPIE. https://doi.org/10.1117/12.3048337

*name of corresponding author

Liu, Y., Shen, J., Yang, L., Bian, G., & Yu, H. (2023). ResDO-UNet: A deep residual network for accurate retinal vessel segmentation from fundus images. *Biomedical Signal Processing and Control*, *79*, 104087.

Liu, Z., Chen, B., & Zhang, A. (2020). Building segmentation from satellite imagery using U-Net with ResNet encoder. *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 1967–1971. https://doi.org/10.1109/ICMCCE51767.2020.00431

Luo, W., Li, Y., Urtasun, R., & Zemel, R. (2016). Understanding the effective receptive field in deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *29*.

Nisa, S. Q., & Ismail, A. R. (2022). Dual U-Net with Resnet Encoder for Segmentation of Medical Images. *International Journal of Advanced Computer Science and Applications*, *13*(12). https://doi.org/10.14569/IJACSA.2022.0131265

Panchal, S., & Kokare, M. (2024). ResMU-Net: Residual Multi-kernel U-Net for blood vessel segmentation in retinal fundus images. *Biomedical Signal Processing and Control*, *90*, 105859.

Ramadhani, S. D., Erwin, E., Desiani, A., & Bella Agustina, S. (2024). BLOOD VESSEL SEGMENTATION IN RETINAL IMAGES USING RESVNET ARCHITECTURE. *Jurnal Teknik Informatika (Jutif)*, *5*(4), 1139–1147. https://doi.org/10.52436/1.jutif.2024.5.4.2637

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Saeedan, F., Weber, N., Goesele, M., & Roth, S. (2018). Detail-preserving pooling in deep networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9108–9116.

Shi, Z., Wang, T., Huang, Z., Xie, F., Liu, Z., Wang, B., & Xu, J. (2021). MD-Net: A multi-scale dense network for retinal vessel segmentation. *Biomedical Signal Processing and Control*, *70*, 102977. https://doi.org/10.1016/j.bspc.2021.102977

Shi, Z., Wang, T., Xie, F., Huang, Z., Zheng, X., & Zhang, W. (2020). MSU-Net. *Proceedings of the 1st International Symposium on Artificial Intelligence in Medical Sciences*, 177–181. https://doi.org/10.1145/3429889.3430295

Steinmetz, J. D., Bourne, R. R. A., Briant, P. S., Flaxman, S. R., Taylor, H. R. B., Jonas, J. B., Abdoli, A. A., Abrha, W. A., Abualhasan, A., Abu-Gharbieh, E. G., Adal, T. G., Afshin, A., Ahmadieh, H., Alemayehu, W., Alemzadeh, S. A. S., Alfaar, A. S., Alipour, V., Androudi, S., Arabloo, J., … Vos, T. (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *The Lancet Global Health*, *9*(2), e144–e160. https://doi.org/10.1016/S2214-109X(20)30489-7

Sule, O. O. (2022). A Survey of Deep Learning for Retinal Blood Vessel Segmentation Methods: Taxonomy, Trends, Challenges and Future Directions. *IEEE Access*, *10*, 38202–38236. https://doi.org/10.1109/ACCESS.2022.3163247

Tsai, C.-M., Qiu, S.-H., & Chang, C.-W. (2025). Retinal Vessel Segmentation Using an Attention-Enhanced U-Net Architecture. *International Conference on Multimedia Information Technology and Applications*, 3–15.

Vijay, S., Guhan, T., Srinivasan, K., Vincent, P. M. D. R., & Chang, C.-Y. (2023). MRI brain tumor segmentation using residual Spatial Pyramid Pooling-powered 3D U-Net. *Frontiers in Public Health*, *11*, 1091850.

Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., & Huang, T. S. (2018). Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7268–7277.

World Health Organization. (2023). *Blindness and vision impairment*. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

*name of corresponding author