

Encoder Selection and Segmentation Quality in U-Net-based Retinal Vessel Analysis

Jegan Sivaraman¹, Arjun Paramarthalingam², Arulnancy Thirunavukkarasu³, Regan Rajendran⁴

¹Lecturer, ^{2,4}Assistant Professor, Department of Computer Science and Engineering, University College of Engineering Villupuram, Tamilnadu, India.

³Assistant Professor, Department of Electronics and Communication Engineering, VRS College of Engineering and Technology, Tamilnadu. India.

E-mail: ¹Jeganeucev@gmail.com, ²arjun_ucev@ymail.com, ³joynandyjoy@gmail.com, ⁴rregan85@gmail.com

Abstract

Retinal vessel segmentation through automation is a very significant area in retinal image processing, which can help in diagnosing many different ocular and systemic diseases. In this research study, a comparative study of the encoder-based U-Net models is presented for retinal vessel segmentation. Five different models including ResNet18+U-Net, ResNet34+U-Net, Traditional U-Net, VGG16+U-Net, and VGG19+U-Net have been developed and trained on the retinal images belonging to the DRIVE and CHASE_DB1 databases. The proposed methodology involves image pre-processing, image normalization, image augmentation, feature extraction from image using the encoder architecture, and U-Net decoding with the skip connection for pixel-wise segmentation of retinal vessels. Among all models evaluated in this paper, the VGG19+U-Net model gives the highest value of Jaccard score of 81.83%, F1 score of 89.53%, Precision of 89.47%, Accuracy of 98.41%, and Specificity of 99.13%, whereas the highest Sensitivity score of 90.32% is achieved by the VGG16+U-Net model.

Keywords: Retinal Vessel Segmentation, U-Net, Encoder Architectures, Deep Learning, Fundus Image Analysis, Transfer Learning, Medical Image Segmentation.

1. Introduction

Vessel segmentation is a part of the automated retinal image analysis process and, therefore, the characteristics of blood vessels may include important diagnostic information

about diseases related to ocular system and even general health conditions [8-10]. The process of blood vessels segmentation is one of the main components in computer-assisted retinal image analysis. There are a lot of traditional methods based on the use of such techniques as matched filtering, morphology, active contours, and machine learning [3, 12-16]. The use of these techniques, however, might be restricted due to some handcrafted features and the selected approach to process. It is possible to make an end-to-end pixel-wise segmentation of retinal blood vessels now due to deep learning [21, 22]. It is possible to use the U-Net architecture for that because of its encoder-decoder structure with skip connections to combine spatial and contextual information. The modern studies of retinal vessel segmentation include the use of dense and dilated features [2], deep supervision [19], multiscale feature integration [20], deformable neural networks [17], attention mechanism [5, 6], structural priors [7].

The role of various encoder architectures inside the same U-Net architecture needs further comparison. Hence, this paper attempts to compare ResNet18 + U-Net, ResNet34 + U-Net, U-Net, VGG16 + U-Net, and VGG19 + U-Net for retinal blood vessels segmentation using the DRIVE and CHASE_DB1 datasets. These models are assessed based on Jaccard similarity coefficient, F1 Score, Sensitivity, Precision, Accuracy, and Specificity.

2. Literature Review

Retinal vessel segmentation plays a key role in retinal image analysis since features of blood vessels may help diagnose ocular and systemic disorders, including glaucoma, hypertension, dementia, and stroke [8–10]. The early methods relied mainly on image processing and machine learning using manually crafted algorithms. Literature surveys of traditional methods for vessel segmentation by Almotiri et al. [3] and Moccia et al. [12] pointed out several problems in this task, such as the varying width of vessels, contrast and noise issues, and the complexity of retinal background images. Typical methods are matched filtering [13], hybrid active contours [14], multiscale morphology with seed point detection [15], and support vector machine classification [16].

Deep learning-based segmentation provided remarkable progress in automated blood vessel extraction. Fully convolutional network models were used for establishing end-to-end pixel-level predictions for semantic segmentation [21], whereas Oliveira et al. [22] utilized these models for retinal blood vessel segmentation. The usage of UNet-based approaches became

popular due to the encoder-decoder architecture and skip connections, which made feature extraction as well as recovery of spatial information easier. More advanced approaches involved dense and dilated feature extraction [2], deep supervision [19], cross connected multi-scale features [20], deformable convolutions [17], and full resolution processing with dual threshold refinement [4].

Attention-based learning and structure-aware learning have added more value to vessel representation. The hard attention techniques [5] and the fully attention-based networks [6] focus more on informative retinal regions and feature representation, and the use of skeletal priors along with contrastive loss exploits vascular structure [7]. ELEMENT [1] has shown the usage of multimodal information through coupled region growing and machine learning, and the recent optimization of U-Net [18] focuses on the same area of interest.

However, despite the above, with the increase in the complexity of the architecture of segmentations, it becomes important to choose the right encoder. ResNet encoders give a residual learning of features while VGG architectures give deeper levels of convolution. Hence, the comparison between ResNet18+U-Net, ResNet34+U-Net, Traditional U-Net, VGG16+U-Net and VGG19+U-Net will help in determining the effect of the depth and representation of features in encoder and its effect on the segmentation of retinal vessels. The current research is therefore based on the comparison of these U-Nets with various segmentation metrics.

2.1 Research Gap and Motivation

Literature review shows a considerable advancement in segmentation of the retinal vessels via traditional filters and machine learning algorithms [3,12-16], fully convolutional networks [21,22], multiscale and densely connected networks [2,19,20], deformable networks [17], attention-based models [5,6], full resolution approaches [4], and structure-aware learning [7]. On the other hand, the increased model complexity makes it hard to choose an encoder architecture that is best balanced regarding segmentation accuracy and complexity.

As a result, the current research centers on an examination of ResNet18+U-Net, ResNet34+U-Net, Traditional U-Net, VGG16+U-Net, and VGG19+U-Net approaches in automated retinal vessel segmentation through a controlled comparative analysis. Instead of implementing various architectural changes, the current study seeks to examine the effect of the usage of various encoders in the U-Net architecture. More specifically, the research is focused

on the comparison of the VGG19-based approach as its deeper architecture allows extracting features better than the VGG16-based one does, but, at the same time, its performance can be directly compared to the performance of both the shallower VGG and residual encoders.

3. Proposed Work

3.1 Overall Segmentation Framework

The proposed framework implements automatic retinal blood vessel segmentation from color fundus images based on the encoder-decoder segmentation model architecture based on U-Net. The experimental framework presented in Figure 1 comprises five main steps: data preparation, image preprocessing, augmentation of the training set, feature extraction via the encoder, and pixelwise vessel segmentation. Five segmentation models are experimentally compared: ResNet18+U-Net, ResNet34+U-Net, Traditional U-Net, VGG16+U-Net, and VGG19+U-Net.

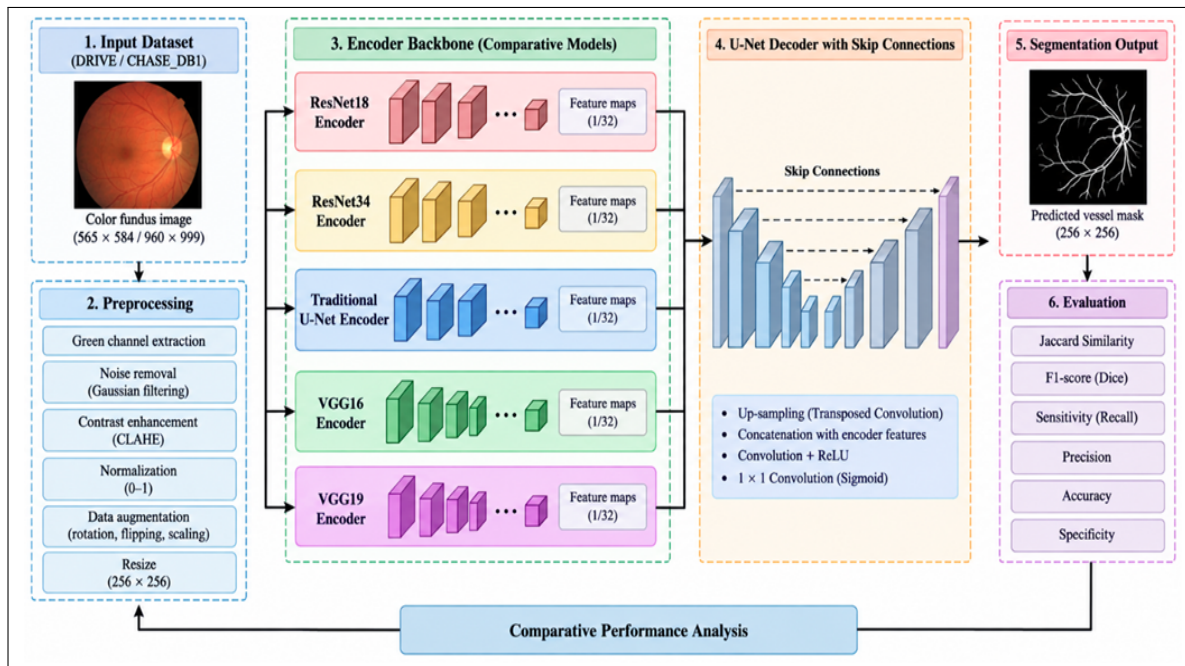


Figure 1. Proposed framework for automated retinal vessel segmentation

The input retinal images are first standardized to a common spatial resolution and subjected to intensity normalization. Training images are subsequently augmented using geometric and intensity-based transformations to increase sample diversity. The encoder extracts hierarchical representations from the retinal images, while the decoder progressively restores

spatial resolution to generate a binary vessel-segmentation mask. Skip connections transfer spatially detailed information from the encoder to the corresponding decoder stages.

3.2 Retinal Vessel Datasets

Two publicly established retinal fundus-image datasets are considered: DRIVE (Digital Retinal Images for Vessel Extraction) [23] and CHASE_DB1 [24]. The datasets provide retinal fundus images together with manually generated vessel-segmentation annotations for supervised learning.

3.2.1 DRIVE Dataset

Figure 2 shows an example of a sample fundus image and the VGG19+U-Net vessel segmentation result of the same image. The fundus images in the DRIVE data set consist of 40 color fundus images that have been taken mainly in a diabetic retinopathy screening campaign in the Netherlands. These images have been taken using a Canon CR5 non-mydratic 3CCD camera with a field of view of 45°.

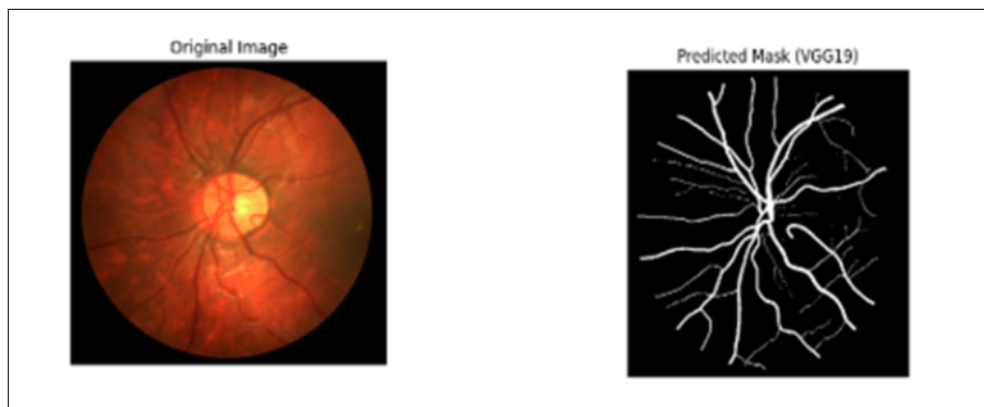


Figure 2. Sample dataset image

Table 1. Images distribution for training, validation and testing after data augmentation (CHASE-DB) – fundus

DRIVE	No. of Images	No. of Labels
Training	20	20
Validation	13	13
Testing	7	7

3.2.2 CHASE_DB1 Dataset

CHASE_DB1[24] contains 28 high-resolution retinal images obtained from the left and right eyes of 14 school children. The images have an original resolution of 999×960 pixels. Each retinal image is associated with vessel annotations generated independently by two human experts, providing manually segmented reference masks for supervised vessel-segmentation evaluation.

Table 2. Images distribution for training, validation and testing after data augmentation (DRIVE) – fundus

CHASE DB	No. of Images	No. of Labels
Training	28	28
Validation	18	18
Testing	10	10

3.2.3 Dataset Partitioning

The datasets were divided into training, validation, and testing subsets for model development and performance evaluation. The training subset was used for model optimization and data augmentation, while the validation subset was used to monitor model performance during training. The testing subset was reserved for the final evaluation of the trained models. Data augmentation was applied only to the training subset and did not alter the number of independent images in the original datasets. The exact image distribution among the training, validation, and testing subsets should be reported consistently with the experimental protocol used for model development and testing.

3.3 Image Preprocessing

Preprocessing is performed to standardize the retinal images before their input to the segmentation networks. Since the fundus images are sourced from various datasets, each having varying spatial resolutions, the size of the images and their respective vessel masks are scaled to a common size of 512×512 pixels. Since the same transformation is applied to both the images and the ground truth masks, pixel-by-pixel correlation is maintained.

Furthermore, image intensity normalization is performed to ensure that the distribution

of the image intensities is standardized. Therefore, the image preprocessing process can be summarized as: Acquisition of retinal fundus images and corresponding vessel masks; Image-mask pairing; Spatial resizing to 512×512 pixels; Intensity normalization; Organization of processed samples for model training and evaluation

3.4 Training Data Augmentation

Data augmentation is employed to increase the diversity of training samples and improve model generalization. The article specifies the use of horizontal flipping, vertical flipping, rotation, scaling/zooming, and brightness adjustment.

Geometric transformations are applied simultaneously to the retinal image and its corresponding vessel mask. For brightness modification, the transformation is applied only to the retinal image because the vessel annotation represents a binary anatomical label rather than an intensity-valued image.

The augmentation process consists of:

- Horizontal flipping: changes the lateral orientation of the retinal image.
- Vertical flipping: introduces additional spatial variation.
- Rotation: changes the orientation of vascular structures.
- Scaling/zooming: introduces variation in the apparent scale of vessel structures.
- Brightness adjustment: introduces intensity variation that can improve robustness to imaging differences.

Augmentation is performed during the training stage to increase the effective diversity of training samples. The validation and testing images are evaluated without stochastic augmentation so that the reported performance reflects the models' behavior on the original evaluation samples.

3.5 Modified U-Net Segmentation Architecture

The segmentation framework is based on a modified U-Net encoder-decoder architecture. U-Net is adopted because its contracting path can extract hierarchical contextual representations while its expanding path reconstructs spatially detailed segmentation maps. Skip connections connect corresponding encoder and decoder stages to retain fine-grained spatial information that may be lost during downsampling.

The architecture consists of four encoder blocks, A central bottleneck, four decoder blocks, skip connections between corresponding encoder and decoder levels, A final 1×1 convolutional layer for generating the segmentation output.

Each encoder layer consists of two convolutional layers that are then followed by batch normalization and ReLU activation functions. This is followed by the application of max pooling operation which is responsible for the down sampling of the feature maps. With the increase in the depth of the encoder network, the spatial resolution of the feature maps reduces, however, the level of abstraction of the features increases.

The bottleneck is the deepest layer in the sense of feature extraction and creates very abstract representations of the retinal image. The decoder layer up-samples the spatial resolution using transposed convolutions. In the decoder layer, the up-sampled layer is concatenated with the respective encoder layer feature map through skip connections.

3.6 Encoder Architectures

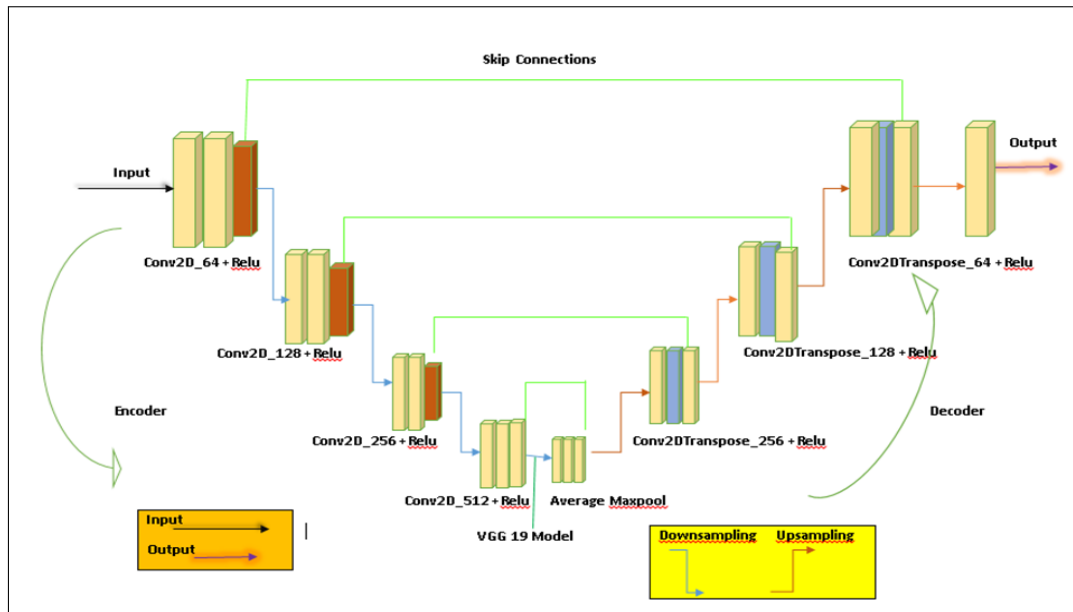


Figure 3. Proposed VGG19 architecture

To investigate the influence of encoder depth and feature-extraction capability, four pretrained encoders are incorporated into the U-Net framework: ResNet18, ResNet34, VGG16, and VGG19. A conventional U-Net is retained as the baseline architecture. The use of such neural network models as ResNet18, ResNet34, Traditional U-Net, VGG16, and VGG19 in the designed solution is quite reasonable and justified in view of optimizing the automation process

of retinal vessel segmentation due to their different abilities and characteristics.

ResNet18-U-Net- ResNet18 acts as the encoder of U-Net architecture. Residual connections help to learn deeper features without facing the problems of optimization due to the growing network depth. Thus, the designed ResNet18-U-Net architecture provides an encoder of relatively small size while maintaining hierarchical feature extraction.

ResNet34-U-Net- ResNet34 is chosen to build the deep residual feature extractor compared to ResNet18 due to its higher depth and therefore better representational power.

Traditional U-Net- The conventional U-Net serves as the baseline segmentation architecture. It employs a symmetric encoder–decoder structure with skip connections, allowing both contextual and spatial information to contribute to vessel-mask reconstruction.

VGG16-U-Net- VGG16 is incorporated as an alternative convolutional encoder. Its sequence of convolutional layers with small receptive fields enables hierarchical representation learning from the fundus images. The extracted features are transferred to the U-Net decoder through the corresponding skip connections.

VGG19-U-Net- VGG19 represents the deepest VGG-based encoder considered in the study. Compared with VGG16, its additional convolutional depth provides increased capacity for hierarchical feature representation. Within the proposed segmentation framework, VGG19 is used to extract increasingly complex retinal features before their reconstruction through the U-Net decoder.

3.7 Transfer Learning and Fine-Tuning

Transfer learning is employed for the ResNet18, ResNet34, VGG16, and VGG19 encoders. The pretrained networks are initialized using weights learned from the ImageNet dataset. This initialization provides previously learned convolutional representations that are subsequently adapted to retinal vessel segmentation.

During fine-tuning, the pretrained feature-extraction layers are adapted to the characteristics of retinal fundus images. The hierarchical representations learned by the encoders are therefore refined according to vessel-specific visual patterns, including local vessel structures and larger-scale contextual information.

3.8 Model Parameterization

The reported parameter counts demonstrate the increasing model complexity associated with the different encoder configurations

Table 3. Parameter configuration of the evaluated architectures

Model Name	Total Parameter	Trainable Parameter
ResNet18 + U-Net	1,83,12,809	1,83,12,809
ResNet34 + U-Net	2,84,20,969	2,84,20,969
Traditional U-Net	3,10,43,521	3,10,43,521
VGG16 + U-Net	4,18,96,833	4,18,96,833
VGG19 + U-Net	4,72,09,089	4,72,09,089

3.9 Training Configuration

The segmentation task is formulated as binary pixel-level classification, with one output class corresponding to the vessel segmentation task. The training section specifies the Adam optimizer with an initial learning rate of 1×10^{-4} .

The model parameters are updated through backpropagation. For each mini-batch, the predicted vessel map is compared with the corresponding ground-truth mask through the selected segmentation loss. The resulting gradients are propagated through the decoder and encoder, and the Adam optimizer updates the trainable parameters.

Table 4. Training and architectural configuration

Parameter	Configuration
Input size	512×512
Task	Binary semantic segmentation
Output channels	1
Optimizer	Adam
Initial learning rate	1×10^{-4}
Number of epochs	30
Weight initialization	ImageNet pretrained

Encoder models	ResNet18, ResNet34, U-Net, VGG16, VGG19
Decoder	U-Net decoder
Encoder blocks	4
Decoder blocks	4
Pooling	Max pooling
Upsampling	Transposed convolution
Normalization	Batch normalization
Activation	ReLU
Output layer	1×1 convolution

3.10 Performance Evaluation Metrics

The segmentation performance is evaluated using Jaccard similarity, F1-score, sensitivity, precision, accuracy, and specificity, as reported.

Sensitivity measures the proportion of actual vessel pixels correctly identified by the model:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

A higher sensitivity indicates that fewer actual vessel pixels are missed.

Precision measures the proportion of predicted vessel pixels that correspond to actual vessel pixels:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Higher precision indicates fewer non-vessel pixels being incorrectly classified as vessels.

Pixel-level accuracy is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

It measures the proportion of correctly classified pixels over the complete evaluated image region.

Specificity measures the proportion of actual background pixels correctly classified as non-vessel:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

Higher specificity indicates improved suppression of false vessel predictions in background regions.

The Jaccard similarity coefficient is used to quantify the overlap between the predicted vessel mask and the ground-truth mask:

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

The F1-score for the binary segmentation task can be expressed as:

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (6)$$

The F1/Dice measure provides a complementary overlap assessment by jointly considering false-positive and false-negative vessel pixels.

4. Results and Discussion

4.1 Training and Validation Performance

The five investigated architectures, namely ResNet18+U-Net, ResNet34+U-Net, Traditional U-Net, VGG16+U-Net, and VGG19+U-Net, were trained for 30 epochs using the experimental configuration described in Section 3. The training and validation loss and accuracy were monitored throughout the training process to assess model convergence and generalization behavior. The reported learning curves show a progressive reduction in loss accompanied by an increase in training and validation accuracy for all evaluated architectures.

The ResNet18+U-Net model exhibits a continuous reduction in both training and validation loss during training. The corresponding accuracy curves show a progressive increase as the model learns discriminative representations of retinal vessel structures. The close trend between training and validation curves indicates consistent optimization during the training

process. The first five reported epochs show a reduction in training loss from 0.9489 to 0.6560 and validation loss from 0.9654 to 0.6747, while training and validation accuracy increase from 0.0683 and 0.0421 to 0.3550 and 0.3345, respectively.

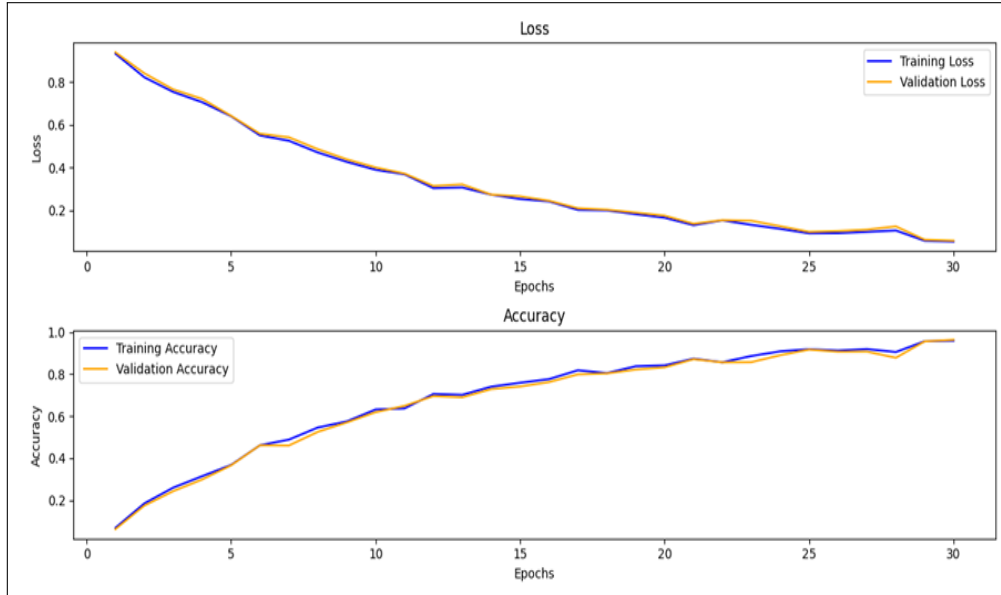


Figure 4. Training and validation curve of modified U-Net with ResNet18

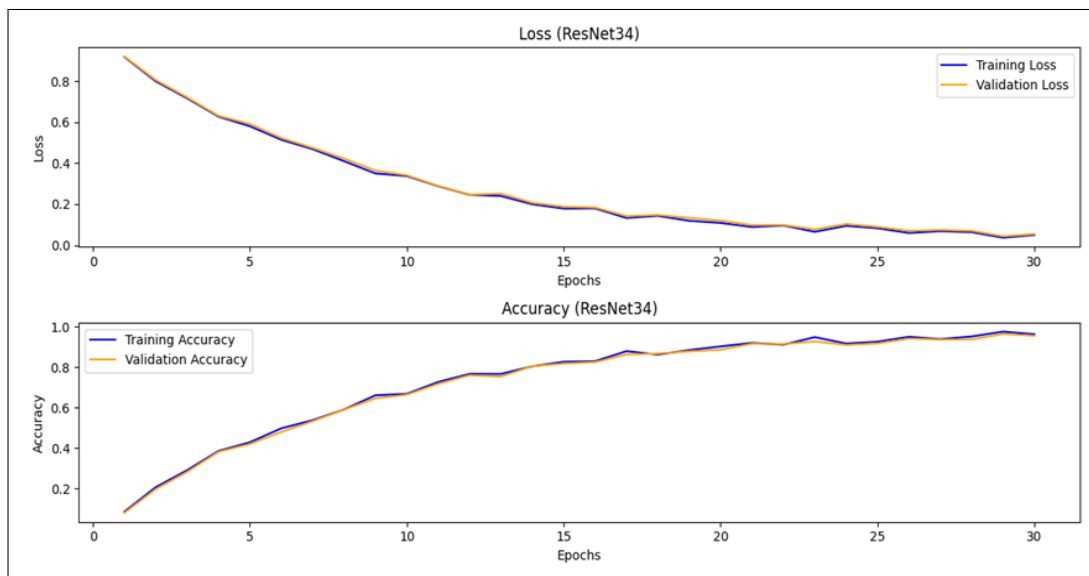


Figure 5. Training and validation curve of modified U-Net with ResNet34

ResNet34+U-Net demonstrates a similar convergence pattern. Training loss decreases from 0.9224 to 0.5502 over the first five reported epochs, while validation loss decreases from 0.9326 to 0.5589. Training accuracy increases from 0.0778 to 0.4558, accompanied by an

increase in validation accuracy from 0.0758 to 0.4484. The relatively similar training and validation trajectories indicate stable optimization during the reported training period.

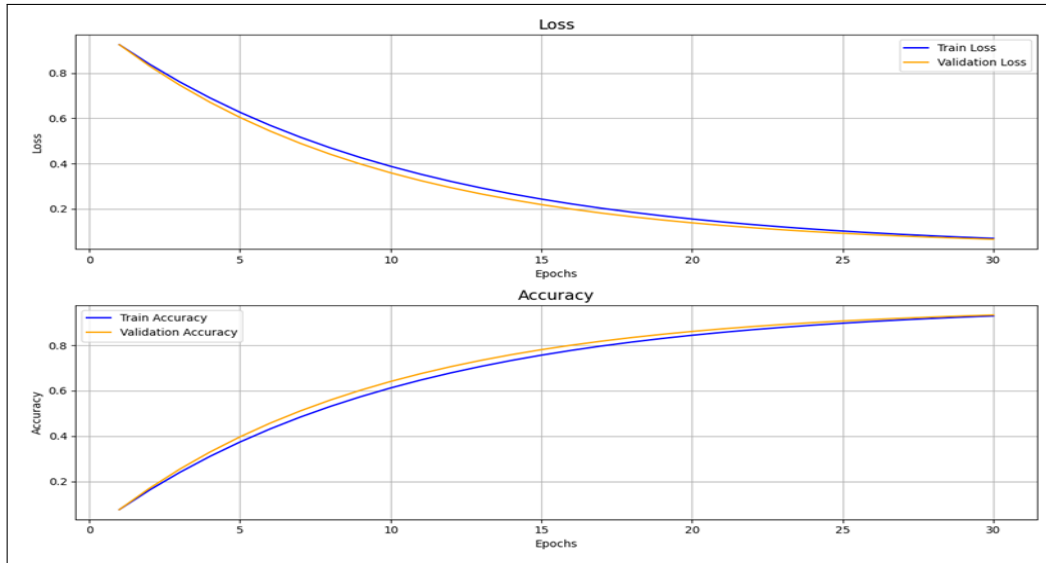


Figure 6. Training and validation curve of traditional U-Net

The Traditional U-Net also demonstrates stable convergence. The reported final training loss is 0.069787, while the final validation loss is 0.065674. The corresponding training and validation accuracies are 0.930213 and 0.934326, respectively. The close final loss and accuracy values indicate that the baseline U-Net successfully learns the vessel-segmentation task without a substantial divergence between training and validation performance in the reported experiment.

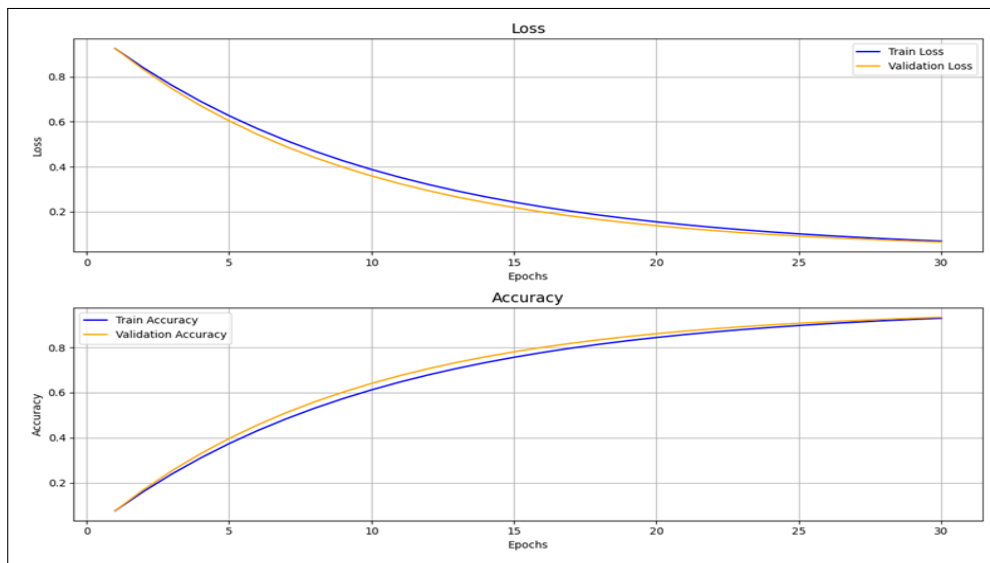


Figure 7. Training and validation curve of modified U-Net with VGG16

VGG16+U-Net exhibits a further improvement in the reported training behavior. The final training loss and validation loss are 0.033518 and 0.038316, respectively, while training and validation accuracy reach 0.966482 and 0.961684. The reduction in loss and corresponding increase in accuracy indicate effective adaptation of the VGG16-based encoder–decoder architecture to the retinal vessel segmentation task.

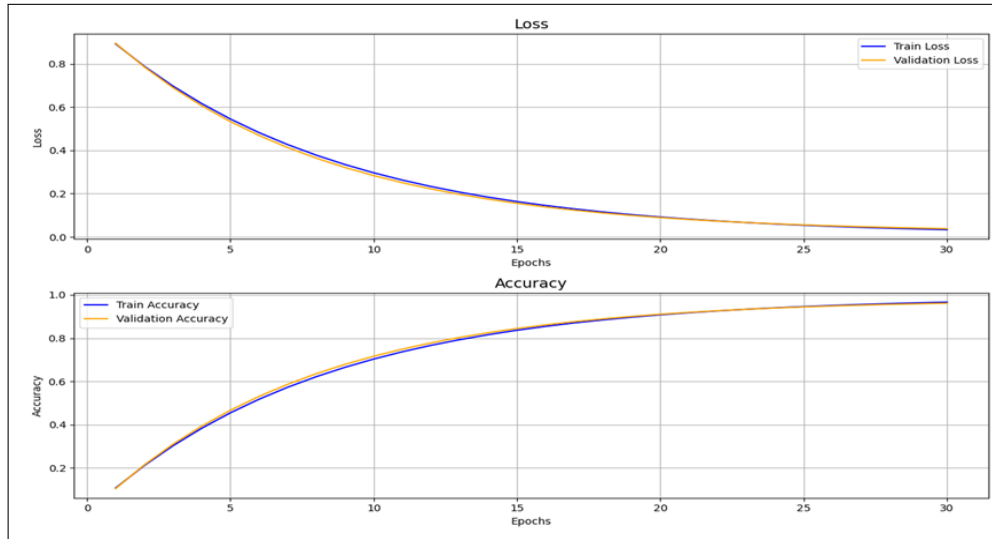


Figure 8. Training and validation curve of modified U-Net with VGG19

The VGG19+U-Net model shows a similar progressive convergence behavior, with training and validation loss decreasing throughout the training process and accuracy increasing toward the later epochs. The curves do not exhibit a pronounced divergence between training and validation behavior in the reported plots. This indicates that the VGG19-based architecture was able to progressively learn retinal vessel representations during the 30-epoch training process. Overall, the training curves demonstrate that all five architectures successfully optimized their respective segmentation objectives. However, because retinal vessel pixels constitute only a portion of the complete fundus image, accuracy alone is insufficient for determining segmentation quality. Therefore, the final comparison is based on overlap and classification metrics obtained from the reported testing results.

4.2 Quantitative Segmentation Results

The quantitative testing performance of the five architectures is presented in Table 4. The evaluation considers Jaccard similarity, F1-score, sensitivity, precision, accuracy, and specificity. These metrics provide complementary information regarding region overlap, vessel detection,

false-positive suppression, and overall pixel classification performance

Table 5. Performance comparison of the evaluated models

Model	Jaccard Similarity	F1-Score	Sensitivity	Precision	Accuracy	Specificity
ResNet18 + U-Net	0.6897	0.8159	0.8053	0.8286	0.9742	0.9871
ResNet34 + U-Net	0.7068	0.8274	0.8066	0.8513	0.9760	0.9799
Traditional U-Net	0.7634	0.8638	0.8684	0.8610	0.9799	0.9890
VGG16 + U-Net	0.8131	0.8921	0.9032	0.8825	0.9833	0.9899
VGG19 + U-Net	0.8183	0.8953	0.8969	0.8947	0.9841	0.9913

The results show a progressive improvement in the principal overlap measures from ResNet18+U-Net to VGG19+U-Net. ResNet18+U-Net obtains a Jaccard similarity of 0.6897 and F1-score of 0.8159, whereas ResNet34+U-Net increases these values to 0.7068 and 0.8274, respectively. Traditional U-Net further improves the Jaccard similarity to 0.7634 and F1-score to 0.8638. The VGG-based configurations produce the strongest overall results. VGG16+U-Net achieves a Jaccard similarity of 0.8131 and F1-score of 0.8921, while VGG19+U-Net obtains the highest reported values of 0.8183 and 0.8953, respectively.

4.3 Comparison of ResNet18+U-Net and ResNet34+U-Net

The comparison between the two residual encoder configurations indicates that increasing the encoder depth from ResNet18 to ResNet34 improves the principal overlap measures. ResNet18+U-Net achieves a Jaccard similarity of 0.6897, F1-score of 0.8159, sensitivity of 0.8053, and precision of 0.8286. ResNet34+U-Net increases these values to 0.7068, 0.8274, 0.8066, and 0.8513, respectively. Accuracy also increases from 0.9742 to 0.9760. The improvement in Jaccard and F1-score indicates that the deeper ResNet34 encoder provides more effective feature representations for vessel segmentation than ResNet18 under the reported experimental configuration. However, specificity decreases from 0.9871 for ResNet18+U-Net to

0.9799 for ResNet34+U-Net. This indicates that the improvement in vessel-region overlap is accompanied by a reduction in background classification specificity. Thus, increasing encoder depth alone does not guarantee improvement across every metric. The results demonstrate the importance of jointly considering vessel detection and background suppression when comparing retinal vessel segmentation architectures.

4.4 Comparison with Traditional U-Net

Conventional U-Net outperforms the other two methods in terms of segmentation performance. The Jaccard similarity score of 0.7634 is superior to the other two methods, and so is the F1-score of 0.8638. The model has obtained sensitivity, precision, accuracy, and specificity scores of 0.8684, 0.8610, 0.9799, and 0.9890, respectively. It can be seen that the conventional U-Net encoder-decoder architecture is still efficient in retinal vessel segmentation. Its skip connections enable the decoder to recover spatial details from earlier feature representations while the deeper encoder stages provide contextual information.

The design of the feature-transfer mechanism between encoder and decoder is equally important for preserving fine vascular structures.

4.5 Comparison of VGG16+U-Net and VGG19+U-Net

The VGG-based configurations provide the highest overall segmentation performance among the five evaluated architectures. VGG16+U-Net achieves a Jaccard similarity of 0.8131, F1-score of 0.8921, sensitivity of 0.9032, precision of 0.8825, accuracy of 0.9833, and specificity of 0.9899. VGG19+U-Net further increases the Jaccard similarity to 0.8183 and the F1-score to 0.8953. Precision increases from 0.8825 to 0.8947, while accuracy increases from 0.9833 to 0.9841. Specificity also improves from 0.9899 to 0.9913. However, VGG16+U-Net obtains a higher sensitivity of 0.9032, compared with 0.8969 for VGG19+U-Net. Therefore, VGG16+U-Net detects a slightly larger proportion of the reference vessel pixels, whereas VGG19+U-Net provides better precision and specificity. This difference is important because retinal vessel segmentation requires an appropriate balance between detecting thin and complex vessels and avoiding false-positive vessel regions. The VGG19+U-Net results indicate a more balanced performance according to the combined overlap, precision, and specificity measures.

4.6 Overall Model Comparison

Comparison of the results indicates the obvious progress in the performance achieved by the tested models. The Jaccard similarity is improved from 0.6897 in case of ResNet18+U-Net to 0.8183 in the case of VGG19+U-Net. The same concerns the F1-score, which increased from 0.8159 to 0.8953. Respectively, the highest value of sensitivity was achieved by the VGG16+U-Net with the figure of 0.9032. Therefore, despite the fact that VGG19+U-Net proved to be the best performing architecture according to most criteria, VGG16+U-Net is still competitive, especially when the sensitivity matters.

In addition, the improvements are not identical while comparing models. The difference between VGG16+U-Net and VGG19+U-Net is quite small as compared to the improvement from ResNet18+U-Net to V.

Table 6. Testing result of proposed system

Model	Jaccard Similarity	F1-Score	Sensitivity	Precision	Accuracy	Specificity
ResNet18 + U-Net	0.6897	0.8159	0.8053	0.8286	0.9742	0.9871
ResNet34 + U-Net	0.7068	0.8274	0.8066	0.8513	0.9760	0.9799
Traditional U-Net	0.7634	0.8638	0.8684	0.8610	0.9799	0.9890
VGG16 + U-Net	0.8131	0.8921	0.9032	0.8825	0.9833	0.9899
VGG19 + U-Net	0.8183	0.8953	0.8969	0.8947	0.9841	0.9913

4.7 Qualitative Segmentation Results

Figure 9 presents representative segmentation outputs for qualitative comparison. The predicted masks demonstrate that all investigated architectures are capable of identifying the major vascular structures in the retinal fundus images. The comparison also illustrates the ability of the models to reconstruct the branching structure of retinal vessels.

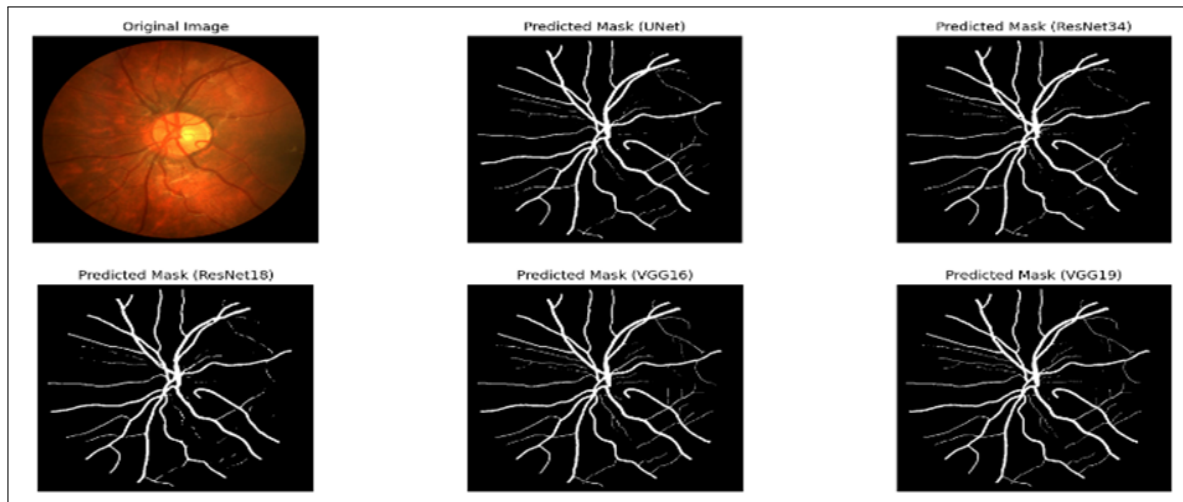


Figure 9. Predicted images comparison

The qualitative results complement the quantitative evaluation by providing a visual assessment of vessel continuity and spatial localization. In particular, the comparison of the predicted masks with the reference segmentation provides an indication of how effectively each architecture preserves the overall vascular topology.

The VGG16+U-Net and VGG19+U-Net outputs are consistent with their higher quantitative scores, showing comparatively accurate reconstruction of the major vessel structures. Nevertheless, quantitative metrics remain necessary for evaluating differences that may not be visually apparent, particularly for thin vessels and small false-positive regions.

4.8 Model Complexity and Performance Trade-off

The performance comparison should also be interpreted in relation to model complexity. The reported parameter counts increase from 18,312,809 parameters for ResNet18+U-Net to 47,209,089 parameters for VGG19+U-Net. VGG16+U-Net contains 41,896,833 parameters, while Traditional U-Net, ResNet34+U-Net, and ResNet18+U-Net contain 31,043,521, 28,420,969, and 18,312,809 parameters, respectively. The VGG19+U-Net architecture therefore provides the highest reported segmentation performance at the cost of the largest parameter count. The relatively small improvement from VGG16+U-Net to VGG19+U-Net suggests that the additional model capacity should be considered alongside computational requirements. This observation is important when selecting a model for practical deployment. If maximum overall segmentation performance is prioritized, VGG19+U-Net is the preferred configuration based on the reported results. If computational complexity or sensitivity is given greater importance, VGG16+U-Net or a lighter architecture may provide a more appropriate trade-off.

5. Conclusion

This work provides a comparative analysis of five architectures based on the U-Net architecture: ResNet18+U-Net, ResNet34+U-Net, Traditional U-Net, VGG16+U-Net, and VGG19+U-Net. The models are analyzed based on a set of segmentation criteria: Jaccard similarity, F1-score, sensitivity, precision, accuracy, and specificity. This comparative analysis proves that the choice of encoder greatly affects the performance of the U-Net in segmentation tasks. In terms of the proposed architectures, the best performance was shown by VGG19+U-Net, which reached the following values of the metrics: Jaccard score = 0.8183, F1-score = 0.8953, precision = 0.8947, accuracy = 0.9841, and specificity = 0.9913. The highest sensitivity = 0.9032 was shown by VGG16+U-Net, which implies its ability to detect a slightly higher proportion of pixels with vessels. The performance of the traditional U-Net architecture was quite good and proved that the classical encoder-decoder architecture is efficient in vessel segmentation tasks. Comparing to the above models, ResNet18+U-Net and ResNet34+U-Net showed worse results in terms of the main segmentation metrics. However, the performance gain achieved by VGG19 over VGG16 is not that significant and must therefore be taken in conjunction with the rise in model complexity. In future research, the results obtained in the current paper need to be corroborated using larger retinal image databases, proper data splitting, and efficiency testing of the architectures.

References

- [1] Rodrigues, Erick O., Aura Conci, and Panos Liatsis. "ELEMENT: Multi-modal Retinal Vessel Segmentation based on a Coupled Region Growing and Machine Learning Approach." *IEEE Journal of Biomedical and Health Informatics* 24, no. 12 (2020): 3507-3519.
- [2] Mou, Lei, Li Chen, Jun Cheng, Zaiwang Gu, Yitian Zhao, and Jiang Liu. "Dense Dilated Network with Probability Regularized Walk for Vessel Detection." *IEEE Transactions on Medical Imaging* 39, no. 5 (2019): 1392-1403.
- [3] Almotiri, Jasem, Khaled Elleithy, and Abdelrahman Elleithy. "Retinal Vessels Segmentation Techniques and Algorithms: a Survey." *Applied Sciences* 8, no. 2 (2018): 155.
- [4] Liu, Wentao, Huihua Yang, Tong Tian, Zhiwei Cao, Xipeng Pan, Weijin Xu, Yang Jin, and Feng Gao. "Full-Resolution Network and Dual-Threshold Iteration for Retinal Vessel and Coronary

- Angiograph Segmentation." *IEEE Journal of Biomedical and Health Informatics* 26, no. 9 (2022): 4623-4634.
- [5] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi and Y. Tao, "Hard Attention Net for Automatic Retinal Vessel Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3384-3396, Dec. 2020.
 - [6] Li, Kaiqi, Xingqun Qi, Yiwen Luo, Zeyi Yao, Xiaoguang Zhou, and Muyi Sun. "Accurate Retinal Vessel Segmentation in Color Fundus Images via Fully Attention-based Networks." *IEEE Journal of Biomedical and Health Informatics* 25, no. 6 (2020): 2071-2081.
 - [7] Tan, Yubo, Kai-Fu Yang, Shi-Xuan Zhao, and Yong-Jie Li. "Retinal Vessel Segmentation with Skeletal Prior and Contrastive Loss." *IEEE Transactions on Medical Imaging* 41, no. 9 (2022): 2238-2251.
 - [8] Cheung, Carol Yim-lui, M. Kamran Ikram, Christopher Chen, and Tien Yin Wong. "Imaging Retina to Study Dementia and Stroke." *Progress in Retinal and Eye Research* 57 (2017): 89-107.
 - [9] Fraser-Bell, Samantha, Richard Symes, and Anagha Vaze. "Hypertensive eye Disease: a Review." *Clinical & Experimental Ophthalmology* 45, no. 1 (2017): 45-53.
 - [10] Akil, Handan, Alex S. Huang, Brian A. Francis, Sirinivas R. Sadda, and Vikas Chopra. "Retinal Vessel Density from Optical Coherence Tomography Angiography to Differentiate Early Glaucoma, Pre-Perimetric Glaucoma and Normal Eyes." *PloS one* 12, no. 2 (2017): e0170476.
 - [11] L Srinidhi, Chetan, P. Aparna, and Jeny Rajan. "Recent Advancements in Retinal Vessel Segmentation." *Journal of Medical Systems* 41, no. 4 (2017): 70.
 - [12] Moccia, Sara, Elena De Momi, Sara El Hadji, and Leonardo S. Mattos. "Blood Vessel Segmentation Algorithms—Review of Methods, Datasets and Evaluation Metrics." *Computer Methods and Programs in Biomedicine* 158 (2018): 71-91.
 - [13] Singh, Nagendra Pratap, and Rajeev Srivastava. "Retinal Blood Vessels Segmentation by using Gumbel Probability Distribution Function based Matched Filter." *Computer Methods and Programs in Biomedicine* 129 (2016): 40-50.
 - [14] Karn, Prakash Kumar, Birendra Biswal, and Subhransu Ranjan Samantaray. "Robust Retinal Blood Vessel Segmentation using Hybrid Active Contour Model." *IET Image Processing* 13, no. 3 (2019): 440-450.
 - [15] Wang, Weihua, Jingzhong Zhang, Wen Yuan Wu, and Shuang Zhou. "An Automatic Approach for Retinal Vessel Segmentation by Multi-scale Morphology and Seed Point Tracking." *Journal of*

- Medical Imaging and Health Informatics 8, no. 2 (2018): 262-274.
- [16] Tuba, Eva, Lazar Mrkela, and Milan Tuba. "Retinal Blood Vessel Segmentation by Support Vector Machine Classification." In 2017 27th International Conference Radioelektronika (RADIOELEKTRONIKA), pp. 1-6. IEEE, 2017.
- [17] Jin, Qiangguo, Zhaopeng Meng, Tuan D. Pham, Qi Chen, Leyi Wei, and Ran Su. "DUNet: A Deformable Network for Retinal Vessel Segmentation." Knowledge-based systems 178 (2019): 149-162.
- [18] Sivaraman, Jegan, Arjun Paramarthalingam, Arulnancy Thirunavukkarasu, Asokan Vasudevan, Soon Eu Hui, and Duraimurugan Samiayya. "Optimized U-Net Model for Precise Retinal Blood Vessel Segmentation from Colour Fundus Images." Scientific Reports (2026).
- [19] Guo, Song, Kai Wang, Hong Kang, Yujun Zhang, Yingqi Gao, and Tao Li. "BTS-DSN: Deeply Supervised Neural Network with Short Connections for Retinal Vessel Segmentation." International journal of Medical Informatics 126 (2019): 105-113.
- [20] Feng, Shouting, Zhongshuo Zhuo, Daru Pan, and Qi Tian. "CcNet: A Cross-Connected Convolutional Network for Segmenting Retinal Vessels using Multi-Scale Features." *Neurocomputing* 392 (2020): 268-276.
- [21] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440. 2015.
- [22] Oliveira, Américo, Sergio Pereira, and Carlos A. Silva. "Retinal Vessel Segmentation based on Fully Convolutional Neural Networks." Expert Systems with Applications 112 (2018): 229-242.
- [23] Larxel. "DRIVE Digital Retinal Images for Vessel Extraction." Kaggle, 2019. <https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction>.
- [24] Rasha Sarhan Alharthi, "Chase-DB1," Kaggle, Accessed September 18, 2026. <https://chatgpt.com/c/6aacbd68-88ec-83ee-b22e-57654ae85333>