Segmentation of Retinal Blood Vessels in Fundus Images Using Attention Mechanisms and Deep Supervised Networks

Nikhil Satyakumar[1*], Ramaswamy Karthikeyan Balasubramanian[2], and Manoj Ravindra Phirke[3]
[1,2]Department of Electronics and Communication Engineering,
M. S. Ramaiah University of Applied Sciences
Bangalore, India 560058
[3]Imaging and Robotics Lab, HCL Technologies Ltd
Bangalore, India 562106
Email: [1]ss.nikhil.s@gmail.com, [2]karthikeyan.ec.et@msruas.ac.in, [3]manoj.p@hcl.com

***Abstract*—Retinal blood vessel segmentation is crucial for detecting and monitoring retinal disorders such as diabetic retinopathy, age-related macular degeneration and glucoma. Automating the segmentation of blood vessels leads to a reduction in the time and cost of manual segmentation, enables large-scale clinical studies, improves accuracy, ensures consistency, allows for real-time analysis, and facilitates early disease detection. The research examines the performance of 7 semantic segmentation architectures, each combined with 10 pre-trained backbones, on 5 publicly available fundus image datasets. Models are trained on a NVIDIA GeForce GTX 1080 Graphics Processing Unit (GPU), and key hyperparameters, such as batch size, optimizers, and learning rate schedulers, are systematically optimized. Intersection over Union (IoU), accuracy, sensitivity, and computational time are used as key performance indicators. Approximately 97 experiments are conducted to achieve state-of-the-art accuracies of 97.72%, 98.23%, 97.62%, 97.83%, and 98.42%, along with IoU scores of 67.82%, 66.29%, 63.89%, 71.34%, and 78.45% on the DRIVE, STARE, HRF, HEI-MED-1, and HEI-MED-2 datasets, respectively. The best performance is achieved using the U-Net++ architecture with ResNeSt backbone, RAdam optimizer, and Cosine Annealing scheduler. This combination leverages deep supervision, attention mechanisms, and bottleneck architectures to enhance multi-scale feature learning, localization, robustness to image variability, and model generalization. Although the models demonstrate strong performance, challenges remain in addressing dataset imbalance and ensuring generalization to unseen patient populations.**

***Index Terms*—Retinal Blood Vessels, Diabetic Retinopathy, Deep Supervision, Semantic Segmentation, Attention Mechanism**

Received: March 07, 2025; received in revised form: Oct. 20, 2025; accepted: Oct. 20, 2025; available online: April 15, 2026.
*Corresponding Author

I. Introduction

SEGMENTATION of retinal blood vessels is important in the field of medical imaging and ophthalmology. It is useful for various applications like: 1) detection of different retinal diseases, such as diabetic retinopathy, age-related macular degeneration and glucoma, 2) assessment and monitoring of risks related to disease progression, 3) automated screening programs, tele-medicine and remote monitoring to enable periodic screening and early disease detection for susceptible patients, and 4) treatment planning with personalized care. Overall, segmentation of retinal blood vessels has a broad range of applications that contribute to diagnosis, treatment, and research related to various eye diseases as described by previous research [1].

Automated segmentation of blood vessels has many advantages. First, it reduces time and cost for manual segmentation. Second, it enables large clinical studies. Third, it improves the accuracy of blood vessel segmentation, using sophisticated computer vision and deep learning algorithms. Fourth, it ensures consistency and reduces the risk of human error associated with fatigue, distractions and variability in human judgment. Fifth, it allows real-time analysis of retinal images, especially useful in emergencies or during surgery. Sixth, it facilitates early disease detection and intervention [2].

Fundus cameras are the most popular devices for large-scale population screening of retinal images because they are cost-effective, user-friendly, consistent, accurate, and non-invasive. Open-source retinal fundus datasets, such as DRIVE, STARE, HRF and HEI-MED, have been used for model training, evaluation and benchmarking. These datasets contain both the

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

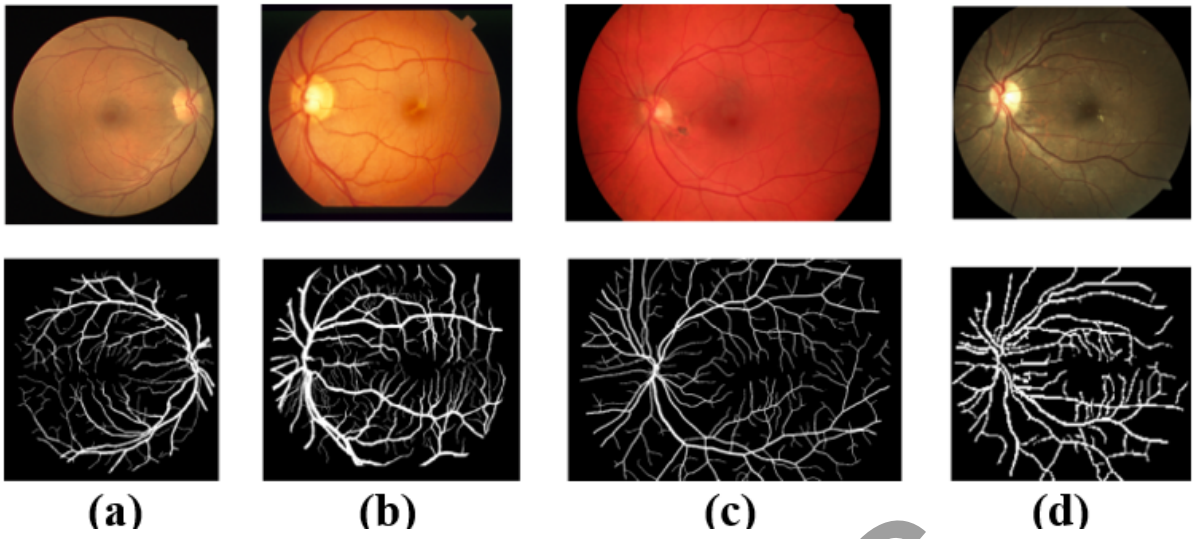

Fig. 1. Sample images and masks for vessel segmentation from different datasets: (a) DRIVE, (b) STARE, (c) HRF, and (d) HEI-MED dataset.

input retinal images as well as binary masks, as shown in Fig. 1. Automated segmentation of the blood vessels is performed using semantic segmentation techniques, where a semantic label is assigned to each pixel in an image, effectively dividing the image into different meaningful segments or regions. In semantic segmentation, the model predicts as well as localizes the blood vessels present in the image. Each pixel of the input image is mapped to a specific class. It provides a detailed understanding of the retinal images, allowing clinicians to comprehend the visual context of an image for diagnosis [3].

Conventional deep learning based semantic segmentation approaches uses sliding window-based methods, but it causes a lot of overhead because the model has to process all the pixels over and over again. Fully Cconvolutional Networks (FCN) proposes an encoder and decoder-based architecture for image segmentation rather than using fully connected networks. There are other variants of FCN like U-Net, U-Net++, MA-Net, LinkNet, DeepLab++ and Feature Pyramid Network (FPN).

U-Net becomes widely adopted due to its skip connections and efficient feature localization. However, the standard U-Net struggles with class imbalance and thin vessel segmentation. U-Net++ improves upon this with deep supervision, a technique that introduces auxiliary outputs at intermediate decoder stages to mitigate vanishing gradients and guide training, resulting in better convergence and finer vessel detection. Attention-based models like ResNeSt backbones further enhance focus on small or complex vascular structures [4]. Despite these advances, no single architecture consistently delivers robust performance across datasets, and challenges remain with generalization, image variability, and optimal hyperparameter selection. This motivates the need for a systematic comparative study. To address the above gaps, the research:

- Benchmarks 7 semantic segmentation architectures and 10 pre-trained backbone networks across 5 open-source retinal datasets.
- Identifies the best-performing architecture-backbone combination for vessel segmentation.
- Conducts an extensive ablation study for hyperparameter fine-tuning conducted with 5 optimizers, 8 learning rate schedulers, and 4 different batch sizes.
- Analyzes how architectural components, such as skip connections, deep supervision, and attention mechanisms, influence performance.
- Provides a comprehensive evaluation using metrics like IoU, accuracy, sensitivity, and computational efficiency.

Overall, around 97 experiments are conducted to determine the optimum architecture and methods for the segmentation of blood vessels. U-Net++ architecture, ResNeSt backbone, RAdam optimizer, Cosine annealing learning rate scheduler, with a batch size of 32 and learning rate of 0.0001, are determined as the optimum settings for blood vessel segmentation. U-Net++ architecture and ResNeSt backbone enable deep supervision and attention mechanism, respectively, thereby leading to multiple advantages for blood vessel segmentation. The multiple advantages are as follows:

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

- Multi-scale feature learning captures information at multiple image resolutions, thereby improving the model's ability to detect vessels of various thicknesses and lengths
- Improved localization and boundary detection leads to precise localization of blood vessels by preserving finer details and accurately delineating vessel boundaries by allowing the model to attend to a specific region of interest
- Handling class imbalance improves the network's ability to focus on the minority class during training, by directing the model's attention to vessel regions. It is especially useful for vessel segmentation tasks, where the number of pixels related to blood vessels is significantly smaller than non-vessel pixels.
- Effective training and convergence mitigates issues related to vanishing gradient during model training, thereby reducing the need for extensive hyperparameter tuning
- Robustness to image variability and enhanced generalizations makes the model robust to imaging conditions, patient demographics and acquisition devices, illumination, noise and other artifacts commonly found in retinal images
- Contextual information integration enables the model to consider global and local contextual information, facilitating a better understanding of the vessel structures within the broader context of retinal images

A. Related Work

Before the widespread adoption of fully convolutional networks, vessel segmentation is seen as a pixel-by-pixel classification process, as proposed by previous research [5–7]. To address the issue of structured prediction, different semantic segmentation-based networks like FCN, SegNet, DeepLab, Mask-RCNN, and U-Net for the segmentation of blood vessels are found in the literature. These methods improve the ability of the model to capture spatial context and preserve structural information of vessels. However, their performance still depends on factors, such as network design, feature extraction capability, and the ability to handle variations in vessel thickness and image quality.

Many studies [8–10] have proposed an FCN network where the input image is downsampled and upsampled to result in the segmentation map. Similarly, other studies [11, 12] have proposed a Stationary Wavelet Transform (SWT) along with an FCN network to increase the number of channels in the image to cope with varying width and direction of the vessel structure. Then, some studies [13, 14] have proposed the SegNet architecture, which is a variant of FCN, where the pooling operations are replaced by strided convolutions. DeepLab model is an extension to the FCN network, which is proposed by different authors [15–17], but the model is computationally heavy.

Next, previous studies [18–24] have proposed a U-Net architecture with various modifications like a local degression scheme to generate additional labels, segment-level loss along with pixel-wise loss, data-aware deep supervision to focus on thin vessels, ResNet and DenseNet-based feature extractors. Similarly, other research [25–30] have used multi-scale networks based on U-Net architecture, where the first branch takes the input image to get the preliminary results. The subsequent branch is used to refine the results. The overall network contains a cascade of U-Net-based modules to improve the overall accuracy of the segmentation masks. Moreover, previous research [31, 32] has utilized MA-Net, which is a variant of U-Net with the addition of an attention mechanism to skip connections. Last, previous research [33] has proposed a Capsule Network with Inception architecture, while another research [34] have used unsupervised ensemble learning to combine multiple segmentation results.

The research addresses the specific issues in the existing literature. First, it analyzes multiple semantic segmentation architectures and selects the most suitable architecture for vessel segmentation. Second, it examines the performance of multiple feature extractor backbones and identifies the most appropriate network for the different semantic segmentation architectures. Third, it determines the optimum hyperparameters for the selected architecture and feature extractor. Fourth, it benchmarks the identified architectures, models and methods on multiple opensource datasets. Fifth, it identifies deep learning based architectural components, which influence the performance of semantic segmentation.

II. Research Method

This section contains a detailed description of the method used. First, it shows different deep learning semantic segmentation model architectures and the various backbones networks. Second, it describes retinal fundus datasets that have been considered for model evaluation and benchmarking. Third, it has metrics that have been used to benchmark the performance of the model. Last, it shows architecture of the overall model proposed for blood vessel segmentation.

Seven semantic segmentation architectures are selected to ensure diversity in benchmarking different deep learning components, such as Atrous convolutions, encoder-decoder architecture, skip connections,

scale variations, deep supervision, attention mechanisms, and optimizations. For a robust evaluation, 10 encoder models are chosen to encompass a variety of approaches. For example, there are classical Convolutional Neural Network (CNN) networks, skip connections, feature concatenation, inception module, depthwise convolutions, attention mechanism, and search space optimizations.

A. Semantic Segmentation Architectures

Seven semantic segmentation techniques are evaluated in the research. Models are selected based on literature review and performance on various open-source datasets. DeepLab addresses limitations of deep convolutional neural networks like reduced feature resolution, the existence of objects at multiple scales and reduced localization accuracy using techniques like Atrous convolutions, Atrous Spatial Pyramid Pooling (ASPP) and Conditional Random Field (CRF), respectively [35]. Atrous convolution and ASPP are applied on the input image to result in a coarse score map, which is four times smaller than the input image. Hence, bilinear interpolation and CRF are used to yield the final segmentation mask. U-Net is a fully convolutional neural network, which contains an encoder, a decoder and skip connections. The encoder (backbone) is used for feature extraction and the decoder for precise localization. The decoder is symmetrical to the encoder network. Features are transferred from encoder to decoder through skip connections. The final layer of the decoder is a 1×1 convolution. The spatial dimension of output is the same as the resolution of the input image, while the number of channels represents the classes [36].

DeepLab++ proposed by previous research [37] is a variant of DeepLab and U-Net. Encoder in DeepLab++ contains multiple Atrous convolutions to reduce the spatial dimension of the input and increase the number of channels. Unlike U-Net, the decoder in DeepLab++ contains very few layers. U-Net++ in another research [38] is similar to U-Net, but with additional intermediate convolutional nodes (nested nodes) at skip connections to bridge the semantic gap between the feature maps of encoder and decoder. The results in dense skip connections, thereby ensuring smooth transfer of information from one layer to another. Instead of concatenating feature maps from skip connections as in U-Net, LinkNet, proposed by previous research [39], uses elementwise summation to aggregate features from multiple paths, reducing the computational overhead. The MA-Net architecture in previous research [40] is like U-Net++, with the addition of an attention mechanism. Here, self-attention is used to capture channel and spatial dependencies and fuse low and high-level semantic features.

B. Backbone Networks for Segmentation

The encoder network for the semantic segmentation techniques is also called a backbone network. These networks are similar to classification models, but the final fully connected layers will be removed. Around 10 different state-of-the-art backbone networks are experimented with 7 different semantic segmentation architectures in this paper.

VGG16 is a 16-layer CNN classification model, where the number of channels increases as the network goes deeper. ResNet introduces the concept of skip connections, where the activations from previous layers are added to layers that are much deeper in the network, thereby enriching the feature representation and improving the accuracy of the model. DenseNet is similar to ResNet, but the features from different layers are concatenated through skip connections [41]. The Dual Path Network (DPN) in previous research [42] combines ResNet and DenseNet, both of which have skip connections. It reduces feature redundancy by residual paths and explores new features by dense paths. InceptionNet, proposed by previous research [43], consists of three convolutional layers with different-sized filters (1×1, 3×3, and 5×5), and there is max-pooling. The output of these four operations is concatenated. Inception-ResNet architecture is a variant of Inception blocks with skip residual connections. SENet in previous research [44] uses an attention network to find channel-wise dependencies among feature channels.

MobilenetV3 in previous research [45] introduces depth-wise convolutional kernels and squeeze and excitation operations. A large number of 1×1 kernels and a smaller number of 3×3 kernels are considered. GeNet, also called GPU-efficient networks, is proposed by previous study [46]. A lightweight Neural Architecture Search (NAS) algorithm is used, based on the idea that full and depth-wise convolution is preferred in initial and final layers, respectively. The ResNeSt network architecture, as described by previous study [47], uses channel-wise multipath attention on various network branches to capitalise on their ability to identify cross-feature interaction and acquire varied representations. A multi-path network structure (InceptionNet) is integrated with the channel-wise attention (SeNet) method. RegNet in previous research [48], focuses on searching the design space rather than a specific configuration. Here, the neural network can find the best design space.

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

TABLE I
DATASET DISTRIBUTION.

Dataset	Original Resolution	Total Images	Total Images (256×256)
DRIVE	565×584	40	80
STARE	700×605	40	80
HRF	504×2336	45	276
HEI-MED 1	549×490	169	338
HEI-MED 2	768×576	169	676
Total Images		463	1,450

C. Dataset and Evaluation Metrics

To reduce the bias caused by training the model on a specific dataset, five different datasets are considered for model training and evaluation in the research. Therefore, it leads to model generalization. Images from these datasets are randomly shuffled and distributed to training, validation, and test folders. To mitigate the effects of model overfitting or bias, augmentation techniques can be used, improving the performance of the model.

DRIVE, STARE, HRF and HEI-MED datasets are popular datasets used for blood vessel segmentation. These open-source datasets contain input images and annotations of blood vessels in the form of mask images. It can be observed from Table I that HRF has the highest resolution of input images, and the HEI-MED 2 dataset has the largest number of images. After combining the images from all datasets, the overall number of images is 463, which can be used to train and evaluate the model. The original input resolution of the fundus images is high, so the images are cropped to a resolution of 256×256. The cropped images are filtered to ignore less relevant images. Hence, overall, there are 1,450 images from 5 datasets with a resolution of 256×256.

There are various evaluation metrics in the literature, like False Positive (FP), True Positive (TP), False Negative (FN), True Negative (TN), sensitivity, Precision, specificity, accuracy, and Intersection over Union (IoU). IoU is the preferred metric for evaluation of image segmentation-based models due to its ability to capture both localization and spatial overlap, making it suitable for a wide range of segmentation scenarios, especially when dealing with imbalanced classes or complex segmentation tasks. IoU is defined as $\frac{\text{Area of Overlap}}{\text{Area of Union}}$. For benchmarking the algorithm with state-of-the-art, apart from IoU, other metrics like sensitivity and accuracy are also used. Sensitivity can be defined as $\frac{TP}{(TP+FN)}$. Accuracy is defined as $\frac{(TP+TN)}{\text{(Total number of pixels in image)}}$.

D. Diabetic Retinopathy Detection Model Architecture

Blood vessel segmentation consists of two classes: blood vessels and background. For training the semantic segmentation model, all five datasets are merged into a single dataset, and the images are shuffled and distributed into training, validation, and test datasets. As the resolution of the original input images is high, the images are cropped into smaller images of size 256×256. Filtering techniques are applied to exclude input images where the proportion of retinal pixels is below a predefined threshold relative to background pixels. Specifically, images are discarded if fewer than 25% of their pixels have grayscale intensity values greater than 100, effectively removing regions that do not contain retina. Both the input images and the corresponding mask images are streamed to the semantic segmentation architecture, in batches of 32.

Around 7 different segmentation architectures are considered with 10 different backbones (encoders) in the research. Then, 5 different optimizers are evaluated along with 8 learning rate schedulers. In Fig. 2, the U-Net++ architecture with ResNest backbone is shown as an example.

The models considered as backbones are pre-trained on the ImageNet dataset. The decoder network is designed to be symmetric to the backbone. The weights of all the layers of the backbone and decoder are fine-tuned using the cropped images from five datasets. Skip connections are used to connect the backbone with the decoder. The output of the segmentation architecture is passed through a Sigmoid activation to result in a mask output. During the model training phase, RAdam is used as the optimizer with a Learning Rate (LR) of 0.0001, and a Cosine annealing LR scheduler is used to alter the LR. The predicted mask image is compared with the manually labeled mask image (ground truth), and loss is estimated using the Dice loss function. Based on the estimated loss, the weights of the model are updated over multiple epochs. The models are trained for 30 epochs and evaluated for different metrics like IoU, sensitivity, accuracy and computation time required for training and inferring from the model over a batch of data.

III. RESULTS AND DISCUSSION

This section contains several explanations. First, it shows detailed analysis of the results when the 7 different semantic segmentation architectures and 10 different backbones are fine-tuned on 5 different datasets. Second, it has extensive ablation study to select the optimum methods and hyper-parameters including batch size, optimizer and learning rate decay function. Last, it includes comprehensive performance

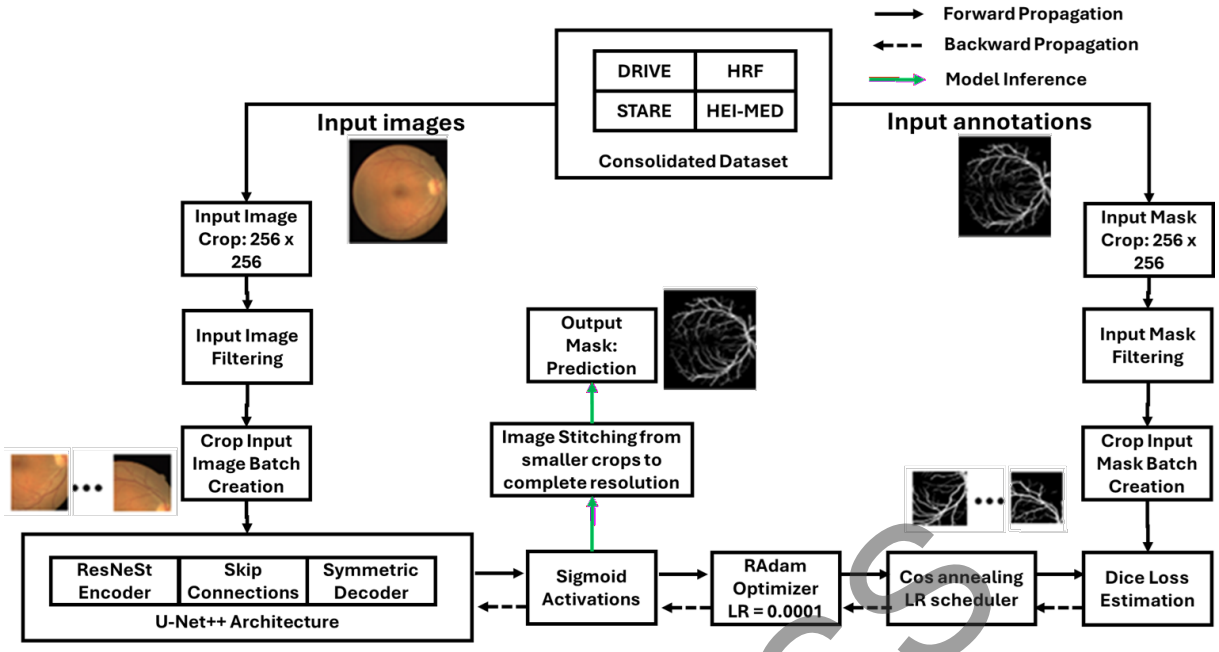

Fig. 2. Architecture of the proposed vessel segmentation network.

analysis of the selected models on open source datasets and benchmarking with techniques proposed in literature.

A. Performance Analysis of Different Semantic Segmentation Architectures and Backbones

For vessel segmentation, images from all five datasets are merged for training as well as for validation. Seven semantic segmentation architectures were evaluated using ten different backbone networks. The experiments are conducted with a batch size of 16, employing the Adam optimizer with a learning rate of 0.0001, a multi-step learning rate scheduler, and Jaccard Loss. The goal is to identify the most effective combination of segmentation architecture and backbone, with the results summarized in Table II and a plot available in Fig. 3.

Usage of CRF-based post-processing and absence of encoder-decoder based architecture in DeepLab leads to relatively lower IoU and higher model training time. In DeepLab++, the encoder and the decoder modules are not symmetric, and there are only two layers in the decoder. Hence, the IoU score of the DeepLab++ architecture is lower than that of U-Net-based architectures. U-Net achieves a good performance because of several reasons. First, it is weighted loss function. It is especially useful when the background is more compared to the vessels. Second, skip connections prevent vanishing and exploding gradients. Third, absence of fully connected layers leads to lower computation time. A disadvantage of the U-Net is that skip connections do not contain any dense layers or attention networks. Unlike U-Net, where the feature maps between the skip connections are concatenated, in LinkNet, the feature maps are added. It reduces the computational overhead in LinkNet, but it compromises IoU. The attention mechanism in MA-Net makes the network slower but more accurate compared to the U-Net architecture. U-Net++ incorporates several dense connections between the encoder and decoder, resulting in more detailed feature maps and achieving the highest IoU score among the evaluated architectures. However, this improvement comes at the cost of increased computational complexity. The FPN architecture, which utilizes feature maps at multiple scales, achieves better IoU scores compared to the standard U-Net architecture. This multi-scale processing, however, leads to longer training times than the U-Net. Additionally, the absence of dense connections between the encoder and decoder in FPN results in a lower IoU score relative to U-Net++.

Table III presents the performance of the U-Net++ architecture with different backbones, reporting IoU alongside computational metrics, including training and inference times. VGG16 has the least IoU score among all the different backbones, as the model does not have skip connections to prevent vanishing or exploding gradients. It is also slow because of a large number of parameters and computations. ResNet-based model achieves better performance compared

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

TABLE II
COMPARISON OF VARIOUS SEMANTIC SEGMENTATION ARCHITECTURES ACROSS MULTIPLE BACKBONE NETWORKS.

Batch 16, Adam, LR 10^{-4}, Multi-Step LR Scheduler, Jaccard Loss								
	Backbone	Semantic segmentation architectures: IoU for vessels						
		DeepLab	DeepLab++	U-Net	LinkNet	MA-Net	FPN	U-Net++
Vessel IoU	VGG-16	56.16	58.54	60.48	59.95	62.75	61.59	63.33
	ResNet	57.52	60.68	61.29	60.74	63.28	62.19	64.14
	Inception-ResNet	58.36	60.53	61.77	61.75	63.87	62.72	64.62
	DenseNet	61.53	63.12	64.84	64.10	66.36	65.67	67.69
	DPN	59.64	61.56	63.08	63.13	64.81	63.73	65.93
	MobileNet-V3	58.55	61.22	62.80	62.96	64.23	62.96	65.65
	GeNet	58.32	60.97	62.17	62.05	63.46	62.23	65.02
	SENet	62.20	64.62	65.86	65.85	67.22	66.86	68.71
	RegNet	62.52	65.73	66.38	66.30	67.86	66.23	69.23
	ResNeSt	64.31	66.23	67.78	66.75	68.49	68.12	**70.63**
	ResNeSt IoU (Background)	94.47	95.84	95.94	95.65	96.22	95.53	**96.46**
	ResNeSt Train time (min/epoch)	9.64	6.57	5.47	**5.11**	7.36	6.32	10.42

Note: Bold values indicate the best performance. It has Learning Rate (LR) and Feature Pyramid Network (FPN).

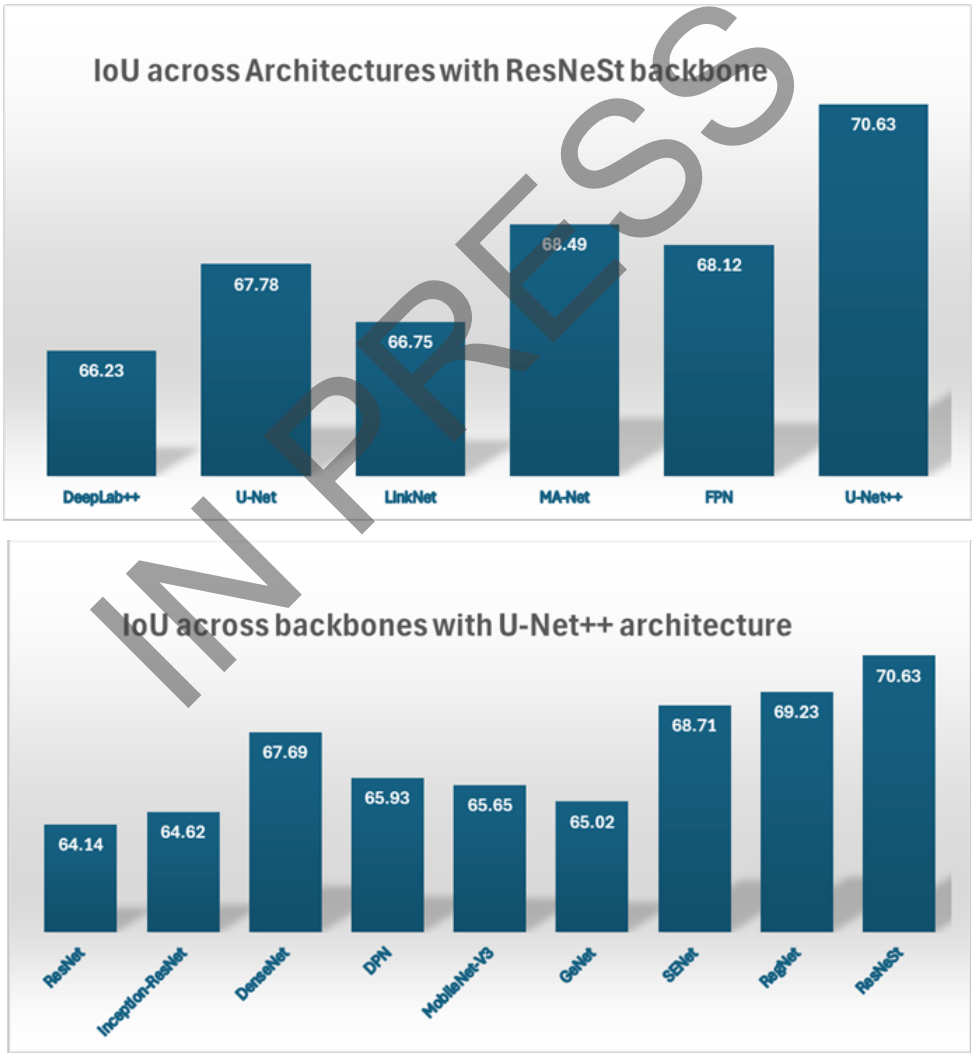

Fig. 3. Vessel Intersection over Union (IoU) comparison of different semantic segmentation architectures across multiple backbone networks.

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

TABLE III
PERFORMANCE METRICS OF U-NET++ WITH DIFFERENT BACKBONE NETWORKS: BATCH 16, ADAM, LEARNING RATE (LR) OF 10^{-4}, MULTI-STEP LR SCHEDULER, AND JACCARD LOSS.

Backbone	U-Net++ IoU		U-Net++ Training Time (min)/epoch)	U-Net++ Inference Time (ms)
	Vessels	Background		
VGG16	63.33	93.63	4.87	87
ResNet	64.14	93.37	3.83	84
Inception-ResNet	64.62	93.51	4.64	76
DenseNet	66.66	94.87	4.41	79
DPN	65.93	94.27	11.91	89
MobileNet-V3	65.65	94.11	**1.41**	**73**
GeNet	65.02	93.72	2.49	75
SENet	68.71	95.64	10.57	95
RegNet	69.23	96.26	11.83	91
ResNeSt	**70.63**	**96.46**	10.42	95

Note: Bold values indicate the best performance.

to VGG16, because of the skip connections. It is also faster than other attention-based networks, densely connected models and inception-based networks. Different filter sizes used in InceptionResNet improves the IoU of the model by enabling it to handle different sizes of blood vessels in the image. Hence, the IoU of InceptionResNet is better than ResNet, but slower compared to other backbones. In DenseNet, as the features from the skip connections are concatenated, it results in strengthened feature propagation and increased feature redundancy. It results in higher IoU, but larger model training and inference time. In DPN, the feature maps from DenseNet and ResNet are added and then split randomly. The size of the feature map results in higher computation, and randomness results in lower IoU. The attention mechanism used in SeNet highlights the significant features like vessels and biomarkers, leading to a higher IoU score, compared to non-attention-seeking networks. However, SeNet is slower compared to other models.

MobileNetV3 architecture takes the least training and inference time, because of features like depth-wise convolutions. As the model uses an attention mechanism on top of InceptionResnet, the IoU of MobileNetV3 is better than that of InceptionResnet. GeNet is the second fastest model, but with a lower IoU because of the removal of dropout and attention mechanism. RegNet model searches the best design space to identify the most suitable model. The IoU of RegNet is the second highest, but this also leads to longer training time. ResNeSt is a combination of InceptionNet and SeNet, so it achieves the highest IoU. The attention module results in an increase in computation time for model training and inference.

Among the evaluated backbones, ResNeSt and MobileNetV3 are selected for further ablation studies based on their superior IoU and computational efficiency, respectively. U-Net and U-Net++ are chosen for additional benchmarking due to their strong performance across different backbone networks. This choice helps to balance between accuracy and computational cost, instead of focusing only on one aspect. It also allows analysis of how different architectural components like attention and lightweight convolutions influence the performance. At the same time, limiting the combinations reduces the experimental complexity while still covering the most relevant design variations.

B. Hyper-Parameter Tuning and Ablation Study

Based on the obtained IoU scores from Tables II and III, U-Net++ with ResNeSt is employed to identify the most suitable optimizers, learning rate schedulers, batch sizes, loss functions, and learning rates for vessel segmentation. The initial parameters considered in the ablation study include a batch size of 16, Adam optimizer, a learning rate of 1e-4, a multi-step learning rate scheduler, and a Jaccard loss function. Using these baseline settings, the optimal combinations of optimizers, learning rate schedulers, batch sizes, loss functions, and learning rates are determined, as summarized in Tables IV to VII. Subsequently, an integrated model is trained using the best-performing hyperparameters detailed in Table VIII.

The IoU score for vessel segmentation with different optimizers is shown in Table IV. Adam [49], builds upon the concepts of RMSProp [50], by incorporating both adaptive learning rates and momentum to accelerate convergence. As a result, Adam generally performs better than RMSProp. Further variants, including NAdam [51], AdamW [52], and RAdam [53], extend Adam through modifications such as Nesterov momentum, decoupled weight decay, and rectified adaptive learning rates, respectively, thereby improving stability and achieving higher IoU performance. Nesterov momentum in NAdam leads to better IoU compared to Adam. In AdamW, the weight decay is

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

TABLE IV
INTERSECTION OVER UNION (IOU) FOR BLOOD VESSELS USING DIFFERENT OPTIMIZERS.

U-Net++, ResNeSt, Batch 16, Learning Rate (LR) 10^{-4}, Multi-Step LR Scheduler, and Jaccard Loss				
RMS-prop	Adam (Original)	NAdam	AdamW	RAdam
70.37	70.63	70.81	71.01	**71.42**

Note: Bold value indicates the best IoU.

TABLE V
PERFORMANCE OF U-NET++ WITH RESNEST UNDER DIFFERENT LEARNING RATE SCHEDULERS.

U-Net++, ResNeSt, Batch 16, Adam, Learning Rate (LR) 10^{-4}, and Jaccard Loss			
LR Scheduler	Vessel IoU	LR Scheduler	Vessel IoU
Constant	55.65	Step	70.16
Multiplicative	56.78	Multistep (Original)	70.63
Linear	67.28	Cos anneal-ing	**71.25**
Exponential	67.79	Cos annealing, warm restart	70.90

Note: Bold value indicates the best Intersection over Union (IoU).

TABLE VI
EFFECT OF BATCH SIZE ON U-NET++ WITH RESNEST PERFORMANCE.

U-Net++, ResNeSt, Adam, Learning Rate (LR) 10^{-4}, Multi-Step LR scheduler, and Jaccard Loss				
Parameters	IoU for Vessels			
	Batch 32	Batch 16 (Original)	Batch 8	Batch 4
Vessel IoU	**71.94**	70.63	70.02	69.58

Note: Bold value indicates the best Intersection over Union (IoU).

decoupled from the gradient-based update, leading to better regularization and better model generalization. One of the main problems with the Adam optimizer is bad convergence because of an undesirably large variance in the LR during the early stages of model training due to the usage of a limited amount of training samples.

In early training stages, the adaptive learning rates can fluctuate wildly, especially with small batch sizes or sparse gradients for the Adam and NAdam optimizers for binary segmentation tasks. RAdam introduces a rectification term that adapts the variance of the learning rate based on how many training steps have occurred. This leads to more stable learning, less overfitting, and better convergence from the start. Blood vessel segmentation is a binary segmentation that involves a blood vessel class that has fewer pixels compared to the background pixels (non-blood vessels). RAdam's adaptive learning rate with rectified variance helps to maintain gradient stability even when the positive class (blood vessel) is underrepresented. It is critical when the loss is highly sensitive to a few pixels (e.g., Dice loss or Focal loss). RAdam is especially effective for binary image segmentation because it introduces rectified updates that stabilize early training, making it well-suited for blood vessel segmentation, where the dataset is relatively small, and it also has class imbalance.

Different LR schedulers are experimented with U-Net++ and ResNeSt architecture, as shown in Table V. The constant LR scheduler has no LR decay and leads to suboptimal convergence for binary segmentation tasks. It results in the lowest IoU score for the segmentation of blood vessels. Multiplicative, linear, and exponential LR schedulers are similar, where they introduce simple decay, which may decay too fast if it is not configured well, leading to underfit and lower IoU scores. Linear LR scheduler leads to a gradual reduction of LR, but it is not optimal unless used with warm-up strategies. Step LR scheduler leads to a sudden drop of LR, which is required especially during the plateaus. Multistep LR drops LR at specific epochs, which can be tuned to task-specific learning phases. It is especially useful when the dataset is imbalanced, like blood vessel segmentation datasets. IoU score of Multi-step LR is higher than Step LR scheduler, but lower than cosine annealing LR schedulers. Cosine annealing LR scheduler allows for a smoother decay of LR, thereby allowing finer convergence without aggressive LR drops. Cosine annealing LR with warm restarts helps escape local minima by periodically increasing the LR. It is especially useful for long training where avoiding overfitting is the key. Cosine annealing LR schedulers result in the highest IoU compared to all the other schedulers.

Different batch sizes are experimented with, as shown in Table VI. A batch size of 32 is the maximum value that can be considered because of the GPU capacity and the input image resolution. The other batch sizes considered for evaluation are 16, 8, and 4. It can be observed from the results that a batch size of 32 leads to the highest accuracy. The IoU of the algorithm increases when the batch size increases too. When a larger number of examples are considered as a batch, the model generalizes better, thereby leading to a smoother gradient curve and faster model training.

Loss functions are used to measure the overlap between the predicted segmentation map and the ground truth. The Dice and Jaccard loss functions used [54] are evaluated under different learning rates. The results are summarized in Table VII.

In blood vessel segmentation, the foreground (blood

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

TABLE VII
COMPARISON OF LOSS FUNCTIONS AND LEARNING RATE (LR) FOR U-NET++ WITH RESNEST.

U-Net++, ResNeSt, Batch 16, Adam, and Multi-Step LR Scheduler						
Parameters (IoU)	Jaccard Loss: IoU for Vessels			Dice Loss: IoU for Vessels		
	10^{-4} (Original)	10^{-3}	10^{-5}	10^{-4}	10^{-3}	10^{-5}
Vessel	70.63	43.89	68.05	**71.35**	49.52	67.39

Note: Bold value indicates the best Intersection over Union (IoU).

TABLE VIII
RESULTS OF ABLATION EXPERIMENTS.

Hyper-Parameters	Architecture + Backbone	Vessel IoU	Background IoU	Train (Min/Epoch)
Batch 16 + Adam + 10^{-4}, Multi-Step, and Jaccard Loss	U-Net + MobileNetV3	62.80	93.19	0.74
	U-Net + ResNeSt	67.78	95.28	5.78
	U-Net++ + ResNeSt	70.63	96.46	10.42
Batch 32 + RAdam + 10^{-4}, Cos Annealing, and Dice Loss	U-Net + ResNeSt	68.55	96.11	3.20
	U-Net++ + MobileNetV3	65.83	94.78	1.21
	U-Net++ + ResNeSt	**72.76**	**97.84**	8.97

Note: Bold values indicate the best performance.

vessels) is represented by very few pixels compared to the background (retina). Dice loss handles this better because it emphasizes true positives. In contrast, Jaccard loss gives equal penalty to false positives and false negatives, making it less forgiving when tiny positive regions are missed. Dice loss produces smoother gradients, especially early in training, which helps the network to converge faster and more stably. Jaccard loss has sharper transitions, which can make optimization unstable in cases with sparse positives. In vessel segmentation, missing thin vessels is a bigger issue than detecting some background as a vessel. Dice loss, by prioritizing recall through its weighting of true positives, helps the model to detect finer structures. In summary, Dice loss is more effective than Jaccard loss for binary, sparse, and highly imbalanced segmentation tasks, such as blood vessel segmentation in fundus images. It can also be observed that a learning rate of 0.0001 yields better IoU compared to other learning rates.

Optimum hyperparameters, identified in previous experiments, are used for the ablation study, as shown in Table VIII. A few of the observations are: 1) changing the backbone from MobileNetV3 to ResNest improves the IoU significantly and 2) changing the architecture from U-Net to U-Net++ and fine-tuning the hyperparameters improves the IoU. It can be noted that U-Net++ with ResNest architecture, with a batch size of 32, RAdam optimizer, Dice loss, and Cosine annealing LR scheduler results in the best IoU score.

In the research, there is an attempt to train and evaluate the model on each dataset separately, but it results in a lower IoU. The results in Table IX suggest that, to achieve a higher IoU, the model has to be trained on the consolidated dataset containing the training samples from all datasets and evaluated on each dataset separately.

The main intent of the research is to identify the key architectural components of deep learning that influence the performance of the semantic segmentation model by extensively experimenting with various semantic segmentation architectures, feature extractor backbones, and hyperparameters. It can be observed from the results that skip connections (U-Net architecture) improve the IoU by around 1.5%, while deep supervision (U-Net++ architecture) improves the IoU by a further 3%. Encoder based on attention mechanisms (ResNeSt) further upgrades the performance by around 2.5%. In summary, a combination of skip connections, deep supervision, attention mechanism, and hyper-parameter tuning results in state-of-the-art performance on multiple fundus datasets.

C. Comparison with State-of-the-Art

Most of the state-of-the-art papers have reviewed and used sensitivity and accuracy to benchmark their models. Hence, sensitivity and accuracy are calculated along with IoU for the models developed in this research. Table X indicates the results of benchmarking of the proposed U-Net++ model with methods proposed in the literature on the DRIVE and STARE datasets.

Previous research has used a CNN-based architecture for the segmentation of blood vessels, leading to lower accuracy and sensitivity [6]. Another research has used a DeepLab semantic segmentation model where CRF-based post-processing and the absence of

Cite this article as: N. Satyakumar, R. K. Balasubramanian, and M. R. Phirke, "Segmentation of retinal blood vessels in fundus images using attention mechanisms and deep supervised networks", CommIT Journal 20(1), 183–196, 2026.

TABLE IX
PERFORMANCE OF THE MODEL WHEN TRAINED ON INDIVIDUAL DATASETS VS CONSOLIDATED DATASET.

U-Net++ + ResNeSt + Batch 32 + RAdam Learning Rate (LR) 10^{-4} + Cosine Annealing LR Scheduler + Dice Loss			
Dataset	Train Time (Min/Epoch)	Vessel IoU (Test Data)	
		Train and Test on Each Dataset Separately	Train on Consolidated Dataset, Test on Each Dataset
DRIVE	0.27	65.12	67.82
STARE	0.37	64.48	66.29
HRF	1.50	62.43	63.89
HEI-MED-1	1.94	69.12	71.34
HEI-MED-2	4.11	77.66	78.45

Note: Consolidated training improves vessel Intersection over Union (IoU) across datasets.

TABLE X
COMPARISON WITH METHODS PROPOSED IN THE LITERATURE.

Proposed Model from Table VIII: U-Net++, ResNeSt, Batch 32, RAdam, LR 10^{-4}, Cosine annealing scheduler, Dice loss, and Intersection over Union (IoU) for blood vessel segmentation: 72.76%

Research	DRIVE		STARE	
	Sensitivity	Accuracy	Sensitivity	Accuracy
[6]: Custom CNN	78.37	96.13	-	-
[16]: DeepLab	74.12	95.85	71.30	94.89
[9]: Feature Pyramid Network (FCN)	77.09	96.33	-	-
[20]: U-Net	75.81	96.12	76.33	96.10
[30]: U-Net	79.63	96.72	80.03	96.88
[24]: U-Net + Custom CNN	84.33	-	-	-
[28]: U-Net + VGG	-	-	79.48	96.48
Proposed model	**85.25**	**97.72**	**83.87**	**98.23**

Note: Bold values indicate the best performance.

an encoder-decoder architecture reduces the performance of the model [16]. Then, previous research has proposed FCN for segmentation, where skip connections are not used. Therefore, it degrades the performance of the model [9]. Similarly, another study has proposed a U-Net model based on a DenseNet encoder [20]. Then, previous research has used U-Net architecture-based cascades with ResNet and InceptionNet backbones, respectively [30]. Then, U-Net-based architecture with custom CNN and VGG backbones is proposed [24, 28]. However, the U-Net architecture does not use deep supervision and an attention mechanism. Hence, the U-Net++ architecture and ResNeSt backbone identified in this research comfortably outperform methods proposed in the literature for all experimented datasets.

IV. CONCLUSION

Automated vessel segmentation is essential for the detection and monitoring of retinal diseases like diabetic retinopathy. In the research, an attempt has been made to train, evaluate, benchmark and analyze the performance of 7 different semantic segmentation architectures with 10 state-of-the-art pre-trained backbones on 5 different open-source fundus datasets for the segmentation of retinal blood vessels. A comprehensive analysis is conducted to identify the optimum hyperparameters like optimizers, schedulers, and batch sizes. Overall, around 97 different experiments are conducted. U-Net++ segmentation architecture, ResNeSt backbone, RAdam optimizer, Cosine annealing learning rate scheduler, Dice Loss and a batch size of 32 are identified as the best settings for vessel segmentation. The proposed techniques result in state-of-the-art accuracy of 97.72%, 98.23%, 97.62%, 97.83% and 98.42%, along with IoU scores of 67.82%, 66.29%, 63.89%, 71.34%, and 78.45% for DRIVE, STARE, HRF, HEI-MED-1 and HEI-MED-2 datasets, respectively.

In conclusion, the research provides a systematic benchmark of segmentation architectures, backbones, and hyperparameter choices for retinal vessel segmentation and identifies the combination of U-Net++ with ResNeSt as the most effective. By integrating deep supervision and attention mechanisms, the research advances the state of the art in automated retinal vessel analysis. The researchers believe these findings contribute to the broader field of medical image analysis and offer a strong foundation for developing robust clinical decision support systems for retinal disease screening and monitoring.

Despite the strong performance, certain limitations remain. The research relies on publicly available datasets, which may not fully capture the diversity of real-world clinical populations, and dataset imbalance

continues to be a challenge. Imaging artifacts, illumination variability, and differences across acquisition devices may also affect generalization. In addition, training on higher-resolution images or larger datasets is limited by computational resources, and external validation with unseen hospital datasets is not performed. Addressing these challenges opens directions for future work, such as fine-tuning models for biomarker segmentation in diabetic retinopathy and age-related macular degeneration, fusing vessel segmentation with biomarker and disease classification models, employing generative augmentation techniques to mitigate data imbalance, implementing cascaded segmentation models on more powerful GPUs, and optimizing deployment for mobile and point-of-care devices.

Author Contribution

Conceived and designed the analysis, R. K. B.; Collected the data, R. K. B.; Contributed data or analysis tools, N. S., R. K. B., and M. R. P.; Performed the analysis, N. S.; Wrote the paper, N. S.; and Reviewed and gave feedback, R. K. B. and M. R. P.

Data Availability

The data that support the findings of the research are available in https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction, https://cecas.clemson.edu/~ahoover/stare/, and https://www5.cs.fau.de/research/data/fundus-images/.

References

[1] A. E. Fayed, M. J. Menten, L. Kreitner, J. C. Paetzold, D. Rueckert, S. M. Bassily, R. R. Fikry, A. M. Hagag, and S. Sivaprasad, "Retinal vasculature of different diameters and plexuses exhibit distinct vulnerability in varying severity of diabetic retinopathy," *Eye*, vol. 38, no. 9, pp. 1762–1769, 2024.

[2] 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, 2025.

[3] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical image segmentation: A comprehensive review of deep learning-based methods," *Tomography*, vol. 11, no. 5, pp. 1–45, 2025.

[4] L. Ming and L. Qi, "DMSU-Net++: A dual multiscale retinal vessel segmentation method based on improved U-Net++," *PLoS One*, vol. 20, no. 7, pp. 1–13, 2025.

[5] M. Zhang, F. Yu, J. Zhao, L. Zhang, and Q. Li, "BEFD: Boundary enhancement and feature denoising for vessel segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Lima, Peru: Springer, Oct. 4–8, 2020, pp. 775–785.

[6] L. Yu, Z. Qin, T. Zhuang, Y. Ding, Z. Qin, and K. K. R. Choo, "A framework for hierarchical division of retinal vascular networks," *Neurocomputing*, vol. 392, pp. 221–232, 2020.

[7] V. Cherukuri, V. K. Bg, R. Bala, and V. Monga, "Deep retinal image segmentation with regularization under geometric priors," *IEEE Transactions on Image Processing*, vol. 29, pp. 2552–2567, 2019.

[8] J. Ding, Z. Zhang, J. Tang, and F. Guo, "A multichannel deep neural network for retina vessel segmentation via a fusion mechanism," *Frontiers in Bioengineering and Biotechnology*, vol. 9, pp. 1–14, 2021.

[9] Z. Feng, J. Yang, and L. Yao, "Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation," in *2017 IEEE International Conference on Image Processing (ICIP)*. Beijing, China: IEEE, Sep. 17–20, 2017, pp. 1742–1746.

[10] Y. Jiang, F. Wang, J. Gao, and W. Liu, "Efficient BFCN for automatic retinal vessel segmentation," *Journal of Ophthalmology*, vol. 2020, no. 1, pp. 1–14, 2020.

[11] Y. Zhong, T. Chen, D. Zhong, and X. Liu, "Wavelet-guided network with fine-grained feature extraction for vessel segmentation." *The Visual Computer*, vol. 41, no. 6, pp. 4377–4392, 2025.

[12] Y. Liu, J. Shen, L. Yang, H. Yu, and G. Bian, "Wave-Net: A lightweight deep network for retinal vessel segmentation from fundus images," *Computers in Biology and Medicine*, vol. 152, 2023.

[13] J. Ryu, M. U. Rehman, I. F. Nizami, and K. T. Chong, "SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation," *Computers in Biology and Medicine*, vol. 163, 2023.

[14] N. Mukkapati and M. S. Anbarasi, "Brain tumor classification based on enhanced CNN model," *Revue d'Intelligence Artificielle*, vol. 36, no. 1, pp. 125–130, 2022.

[15] Y. Bai, J. Li, L. Shi, Q. Jiang, B. Yan, and Z. Wang, "DME-DeepLabV3+: A lightweight

model for diabetic macular edema extraction based on DeepLabV3+ architecture," *Frontiers in Medicine*, vol. 10, pp. 1–11, 2023.

[16] H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, and J. Liu, "Deepvessel: Retinal vessel segmentation via deep learning and conditional random field," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer, Oct. 17–21, 2016, pp. 132–139.

[17] H.-J. Kim, H. Eesaar, and K. T. Chong, "Transformer-enhanced retinal vessel segmentation for diabetic retinopathy detection using attention mechanisms and multi-scale fusion," *Applied Sciences*, vol. 14, no. 22, pp. 1–19, 2024.

[18] S. Kushwaha, J. Boga, B. S. S. Rao, S. N. Taqui, R. G. Vidhya, and J. Surendiran, "Machine learning method for the diagnosis of retinal diseases using convolutional neural network," in *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*. Chennai, India: IEEE, Dec. 21–23, 2023, pp. 1–6.

[19] Z. Li, M. Jia, X. Yang, and M. Xu, "Blood vessel segmentation of retinal image based on dense-U-Net network," *Micromachines*, vol. 12, no. 12, pp. 1–11, 2021.

[20] Y. Wu, Y. Xia, Y. Song, D. Zhang, D. Liu, C. Zhang, and W. Cai, "Vessel-Net: Retinal vessel segmentation under multi-path supervision," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen, China: Springer, Oct. 13–17, 2019, pp. 264–272.

[21] Z. Huang, Y. Fang, H. Huang, X. Xu, J. Wang, and X. Lai, "Automatic retinal vessel segmentation based on an improved U-Net approach," *Scientific Programming*, vol. 2021, no. 1, pp. 1–15, 2021.

[22] H. Hu and Z. Liu, "Retinal vessel segmentation based on recurrent convolutional skip connection U-Net," in *2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS)*. Wuhan, China: IEEE, May 14–16, 2021, pp. 65–71.

[23] S. Mishra, D. Z. Chen, and X. S. Hu, "A data-aware deep supervised method for retinal vessel segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Iowa City, IA, USA: IEEE, April 3–7, 2020, pp. 1254–1257.

[24] H. Zhao, H. Li, and L. Cheng, "Improving retinal vessel segmentation with joint local loss by matting," *Pattern Recognition*, vol. 98, 2020.

[25] X. Du, J. Wang, and W. Sun, "Densely connected U-Net retinal vessel segmentation algorithm based on multi-scale feature convolution extraction," *Medical Physics*, vol. 48, no. 7, pp. 3827–3841, 2021.

[26] C. Liu, P. Gu, and Z. Xiao, "Multiscale U-Net with spatial positional attention for retinal vessel segmentation," *Journal of Healthcare Engineering*, vol. 2022, no. 1, pp. 1–10, 2022.

[27] X. Yang, L. Liu, and T. Li, "MR-UNet: An UNet model using multi-scale and residual convolutions for retinal vessel segmentation," *International Journal of Imaging Systems and Technology*, vol. 32, no. 5, pp. 1588–1603, 2022.

[28] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "Cc-Net: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features," *Neurocomputing*, vol. 392, pp. 268–276, 2020.

[29] X. Li, J. Ding, J. Tang, and F. Guo, "Res2Unet: A multi-scale channel attention network for retinal vessel segmentation," *Neural Computing and Applications*, vol. 34, no. 14, pp. 12 001–12 015, 2022.

[30] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "NFN+: A novel network followed network for retinal vessel segmentation," *Neural Networks*, vol. 126, pp. 153–162, 2020.

[31] Y. Wu, A. Liu, L. Chen, D. Zhao, H. Zhou, and Q. Zheng, "Multi-scale attention net for retina blood vessel segmentation," in *Proceedings of the 2020 4th International Conference on Computer Science and Artificial Intelligence*. Zhuhai. China: Association for Computing Machinery, Dec. 11–13, 2020, pp. 86–90.

[32] G. Wang, Y. Huang, K. Ma, Z. Duan, Z. Luo, P. Xiao, and J. Yuan, "Automatic vessel crossing and bifurcation detection based on multi-attention network vessel segmentation and directed graph search," *Computers in Biology and Medicine*, vol. 155, pp. 1–11, 2023.

[33] C. Kromm and K. Rohr, "Inception capsule network for retinal blood vessel segmentation and centerline extraction," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1223–1226.

[34] K. B. Khan, M. S. Siddique, M. Ahmad, and M. Mazzara, "A hybrid unsupervised approach for retinal vessel segmentation," *BioMed Research International*, vol. 2020, no. 1, pp. 1–20, 2020.

[35] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image

segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[36] Q. Zhao, J. Cao, J. Ge, Q. Zhu, X. Chen, and W. Liu, "Multi-UNet: An effective Multi-U convolutional networks for semantic segmentation," *Knowledge-Based Systems*, vol. 309, 2025.

[37] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with Atrous separable convolution for semantic segmentation," in *Computer Vision – ECCV 2018*, Munich, Germany, Sep. 8–14, 2018.

[38] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *International Workshop on Deep Learning in Medical Image Analysis*. Granada, Spain: Springer, Sep. 20, 2018, pp. 3–11.

[39] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. St. Petersburg, FL, USA: IEEE, Dec. 10–13, 2017, pp. 1–4.

[40] T. Fan, G. Wang, Y. Li, and H. Wang, "Ma-Net: A multi-scale attention network for liver and tumor segmentation," *IEEE Access*, vol. 8, pp. 179 656–179 665, 2020.

[41] M. Hasal, M. Pecha, J. Nowaková, D. Hernández-Sosa, V. Snášel, and J. Timkovič, "Retinal vessel segmentation by U-Net with VGG-16 backbone on patched images with smooth blending," in *International Conference on Intelligent Networking and Collaborative Systems*. Springer, 2023, pp. 465–474.

[42] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," vol. 30, 2017, pp. 4467–4475.

[43] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1. San Francisco, USA: Association for the Advancement of Artificial Intelligence, Feb. 4–9, 2017.

[44] X. Jin, Y. Xie, X. S. Wei, B. R. Zhao, Z. M. Chen, and X. Tan, "Delving deep into spatial pooling for squeeze-and-excitation networks," *Pattern Recognition*, vol. 121, 2022.

[45] S. Qian, C. Ning, and Y. Hu, "MobileNetV3 for image classification," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. Nanchang, China: IEEE, March 26–28, 2021, pp. 490–497.

[46] M. Lin, H. Chen, X. Sun, Q. Qian, H. Li, and R. Jin, "Neural architecture design for GPU-efficient networks," 2020. [Online]. Available: https://arxiv.org/abs/2006.14090

[47] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. Smola, "ResNeSt: Split-attention networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, Louisiana, 2022, pp. 2736–2746.

[48] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 428–10 436.

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[50] A. Graves, "Generating sequences with recurrent neural networks," 2013. [Online]. Available: https://arxiv.org/abs/1308.0850

[51] T. Dozat, "Incorporating Nesterov momentum into Adam," 2016. [Online]. Available: https://openreview.net/forum?id=OM0jvwB8jIp57ZJjtNEZ

[52] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017. [Online]. Available: https://arxiv.org/abs/1711.05101

[53] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," 2019. [Online]. Available: https://arxiv.org/abs/1908.03265

[54] T. P. T. Armand, S. Bhattacharjee, H.-K. Choi, and H. C. Kim, "Transformers effectiveness in medical image segmentation: A comparative analysis of UNet-based architectures," in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. Osaka, Japan: IEEE, Feb. 19–22, 2024, pp. 238–242.