A Comprehensive Review of Deep Learning and Hierarchical Modeling Techniques for Retinal Blood Vessel Segmentation in Fundus Images

Padma Yaramala[1]*, P.Lachi Reddy[2]

Department of ECE, Sree Vahini Institute of Science and Technology, Tiruvuru, AP.

Department of ECE, Lakireddy Bali Reddy College of Engineering. Vijayawada, AP.

Abstract

Retinal blood vessel segmentation is a fundamental step in the automated diagnosis of ocular diseases, such as diabetic retinopathy, glaucoma, and hypertension. The accurate extraction of vascular structures from fundus images enables early clinical intervention and supports decision-making in ophthalmology. Over the years, segmentation techniques have evolved from traditional image processing methods to advanced deep-learning approaches. Convolutional Neural Networks (CNNs) have significantly improved local feature extraction, whereas recent transformer-based models have enhanced the understanding of global contextual relationships. Hybrid architectures that combine CNNs and transformers have further improved segmentation performance. Despite these advancements, persistent challenges remain, particularly in detecting fine and low-contrast vessels and ensuring model generalization across diverse data sets. Many existing approaches struggle to maintain vessel continuity and perform consistently under varying imaging conditions. This review presents a comprehensive analysis of recent developments in retinal vessel segmentation, categorizing the methods based on their architectural innovations and training strategies. Additionally, it highlights emerging research directions, with a particular focus on hierarchical image warping frameworks, which offer promising potential for improving segmentation accuracy and robustness in complex retinal-imaging scenarios.

Keywords: Retinal Vessel Segmentation, Fundus Imaging, Deep Learning, Transformer, Hierarchical Modeling, Image Warping

1. Introduction

Retinal image analysis has become an essential component of modern ophthalmology, particularly for the early detection and monitoring of diseases such as diabetic retinopathy, glaucoma, and hypertensive retinopathy[1]. Among various tasks, blood vessel segmentation plays a vital role, as the morphology of the retinal vasculature often reflects underlying pathological changes[2]. Fundus imaging provides a noninvasive and cost-effective means of capturing retinal structures, making it widely used in both clinical practice and research[3]. However, extracting precise vascular information from these images is challenging. One of the primary difficulties is the low contrast between blood vessels and the surrounding retinal background, particularly in pathological cases. In addition, the presence of thin and fragmented vessels makes it difficult for segmentation models to capture fine details without losing continuity[4]. Noise, illumination variations, and imaging artifacts further complicate this process, often leading to incomplete or inaccurate segmentation results. These challenges highlight the need for more robust and adaptive computational models to address these issues.

Over the past two decades, segmentation techniques have undergone significant transformations. Early approaches relied on traditional image processing methods, such as thresholding, filtering, and morphological operations[5]. Although these methods are computationally simple, they lack robustness in complex scenarios. This led to the adoption of machine learning techniques, in which handcrafted features were used for classification. Although these methods improve performance to some extent, their dependency on feature engineering limits their scalability[6].

The advent of deep learning has marked a breakthrough. Convolutional Neural Networks (CNNs), particularly architectures such as U-Net, have demonstrated remarkable capabilities in learning hierarchical features directly from data[7]. Recently, transformer-based models have been introduced to capture long-range dependencies and global contexts, addressing some of the limitations of CNNs[8]. Hybrid models that combine CNNs and transformers have further enhanced segmentation accuracy. Despite these advancements, challenges such as poor generalization and weak detection of fine vessels persist in the literature. Recent studies have emphasized hierarchical and multiscale modeling as effective methods for improving segmentation accuracy. By integrating spatial transformations and multilevel feature representations, emerging approaches aim to better capture both global vessel structures and intricate capillary details, paving the way for more reliable and clinically applicable solutions.

2. Review Methodology

This review adopts a structured approach to identify and analyze recent advancements in retinal blood vessel segmentation using fundus image analysis. Relevant research articles were systematically collected from well-established scientific databases, including IEEE Xplore, SpringerLink, Elsevier ScienceDirect, and the Nature Publishing Group. These sources were selected to ensure the inclusion of high-quality peer-reviewed studies with significant contributions to the field. The time frame for the literature selection spanned from 2020 to 2026, with particular emphasis on studies published between 2023 and 2026, as they reflect the most recent developments in deep learning and medical image analysis[9][10][11]. The inclusion criteria were carefully defined to maintain relevance and focus of the study. Only studies that specifically addressed fundus image-based vessel segmentation and employed deep learning methodologies, such as convolutional neural networks, transformers, or hybrid architectures, were considered.

From the collected pool, important representative papers were selected based on their novelty, methodological strength, and impact on improving segmentation accuracy[4][6][1][5][2]. These studies collectively provide insights into current trends, including multiscale modeling, attention mechanisms, and advanced training strategies. The selected works form the foundation for analyzing existing limitations and identifying potential directions for future research, particularly in the context of hierarchical image-warping approaches. The overall review process followed in this study is illustrated in **Figure 1**, which presents a structured workflow for selecting, analyzing, and synthesizing the relevant literature.

2.1. Datasets Used in Literature

Retinal vessel segmentation studies mainly rely on a few standard datasets. The most used datasets are DRIVE, STARE, CHASE_DB1, HRF, and FIVES. The datasets differ in size, resolution, and diversity. This variation directly affects the model performance.

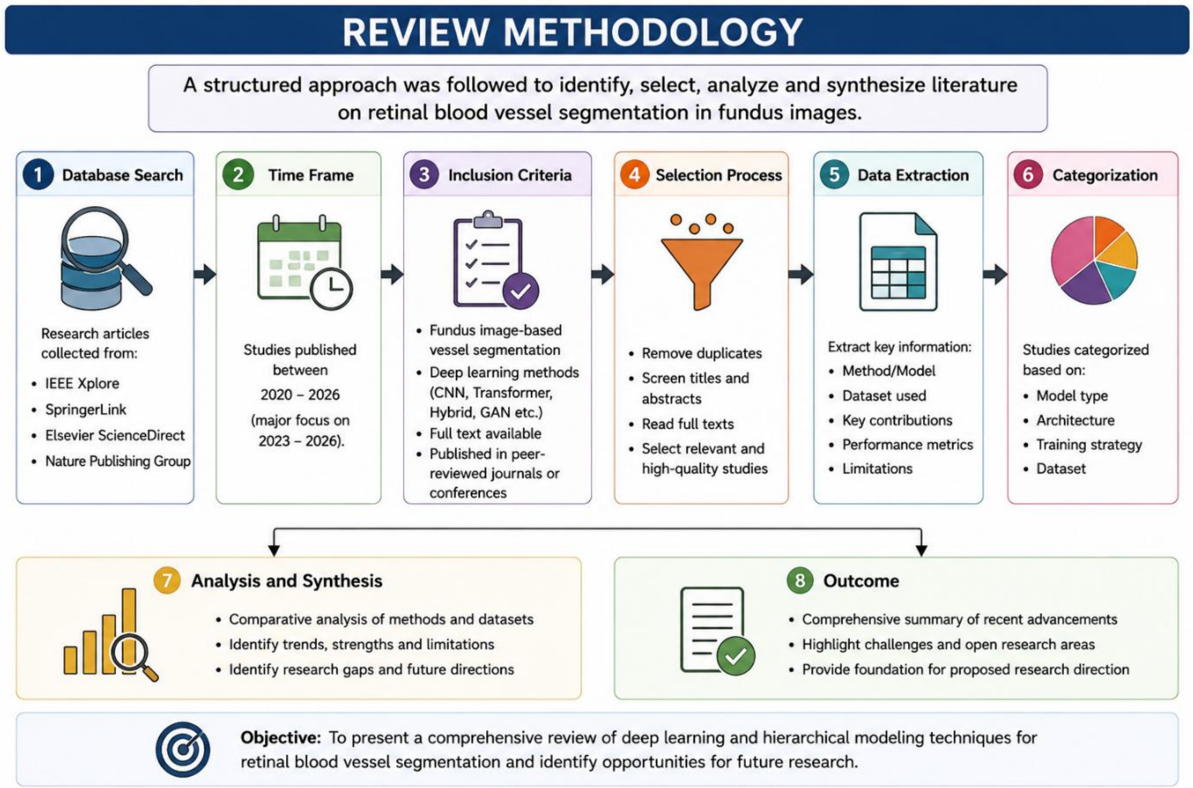

Fig. 1 Structured Workflow for Review Methodology in Retinal Vessel Segmentation Studies

The DRIVE dataset contains 40 fundus images. It is widely used for initial training and benchmarking[12]. The STARE dataset includes 20 images of different retinal conditions. This helps evaluate the model robustness[13]. The CHASE_DB1 dataset contains 28 images collected from children. It provides two manual annotations that support validation[14]. The HRF dataset contains 45 high-resolution images. This is useful for analyzing fine vessel details [15]. The FIVES dataset contains approximately 800 high-resolution images with expert labels. This is suitable for training deep learning models[6]. Most early datasets were small. This limits the variability. Models trained on these data often overfit. They perform well on known data but fail to perform well on new images. Differences in annotations also affect consistency. To overcome these dataset limitations, large-scale datasets such as FIVES are increasingly being used. They provide more samples and better diversity than the previous datasets. This supports stronger training. It also improves generalization, which is essential for real-world clinical applications.

2.2 Various types of Existing Methods used for Retinal Blood Vessel Segmentation in Fundus Images

i. **Traditional Methods:** Early retinal vessel segmentation methods relied on simple image-processing techniques. These include thresholding and morphological operations, respectively. They are easy to implement and require less computation time. However, these methods depend heavily on image quality. Their performance decreases for noisy or low-contrast images. They also struggle with thin vessels and complex backgrounds.

ii. **CNN-based Methods:** With CNN-based models, such as U-Net and ResNet-based segmentation, have become popular with deep learning. These models directly learn features from the data. They perform well in extracting the local patterns. The edges and textures of the vessels were effectively captured. However, CNNs mainly focus on local regions. They often miss the global structures of vessels. This can lead to broken or disconnected vessel predictions in the model.

iii. **Transformer-Based Methods:** Recent approaches use transformer architectures, such as the Swin Transformer and TransUNet. These models capture the long-range dependencies in images. They help to maintain vessel continuity across larger regions. This improves the segmentation of complex vascular networks. However, they require more data and incur higher computational costs.

iv. **GAN-Based Methods:** GAN-based models, such as GANVesselNet, focus on improving image quality and segmentation. They enhance low-contrast vessels and recover the missing details. This is useful for challenging retinal images. However, training GANs is unstable and time-consuming.

v. **Multi-Scale and Hybrid Models:** Hybrid models combine CNNs and transformers. Attention mechanisms were also included. These models operate at multiple scales. They capture both the global and local features. Consequently, they performed better in complex cases. Multiscale feature extraction improves the detection of both large vessels and fine capillaries.

2.3 Training Strategies in Recent Research

Recent studies on retinal vessel segmentation have focused not only on model design but also on effective training strategies. These strategies help improve accuracy and generalization, especially when the data are limited.

i. **Transfer Learning:** Transfer learning is widely used in medical imaging applications. Models pretrained on large datasets, such as ImageNet, are fine-tuned on fundus images. This reduces the training time and improves performance. This is especially useful when labeled medical data are limited.

ii. Self-supervised learning has gained attention in recent years. This allows the models to learn useful representations from unlabeled data. Subsequently, the model was fine-tuned using a smaller labeled dataset. This approach reduces the dependency on manual annotations and improves robustness.

iii. **Transformer Training:** Transformer-based training focuses on capturing the global context. Unlike CNNs, transformers can model long-range relationships within images. This helps maintain vessel continuity and improve the segmentation of complex structures. However, these models often require additional computational resources.

iv. **Data Augmentation, GAN, and Diffusion Methods:** Data augmentation is commonly used to increase dataset diversity. The techniques included rotation, flipping, and contrast adjustment. GAN-based methods generate realistic synthetic images to enhance the training data. Recently, diffusion models have been explored for generating high-quality samples. These approaches help to reduce overfitting and improve model generalization.

2,4 Performance Metrics

Evaluating retinal vessel segmentation models requires clear and reliable evaluation metrics. Each metric reflects a different aspect of the performance. Together, they provide a complete understanding of the model behavior.

Accuracy measures the overall correctness of a model. This shows the number of correctly classified pixels. However, this may be misleading when the number of vessel pixels is much smaller than that of the background pixels. Sensitivity focuses on the detection of true vessels. It measures the accuracy with which the model identifies actual blood vessels. High sensitivity means fewer missed vessels, which is important for clinical analyses. Specificity measures the ability of the model to identify non-vessel regions. This ensures that the background pixels are not incorrectly classified as vessels. A good balance between sensitivity and specificity is essential. Dice Score evaluates the overlap between predicted vessels and ground truth. It is widely used for segmentation tasks. A higher Dice score indicates better alignment and more accurate segmentation of the images. The Area Under Curve (AUC) reflects the model's ability to distinguish between vessel and non-vessel pixels across thresholds. It provides a more comprehensive evaluation than single-point metrics. Evaluation is typically performed using metrics such as accuracy, sensitivity, specificity, the Dice coefficient, and AUC. These measures help to compare the models fairly and highlight their strengths and weaknesses. Using multiple metrics ensures that both the detection quality and structural consistency of the vessels are properly assessed.

3. Comparative Analysis of Existing Methods

A comparative view of recent methods helps in understanding their strengths and limitations. It also shows how different models perform on the datasets. Table 2 presents a summary of the selected studies. Traditional methods exhibit stable but limited performance. They fail in complex cases. CNN-based models improve accuracy and feature-learning. However, they focused more on local patterns. Transformer-based and hybrid models achieved better results. They preserve vessel continuity and capture the global context. However, many models depend on small datasets. This may lead to overfitting. Some methods also require high computational resource. These limitations highlight the need for more efficient and generalizable approaches, such as hierarchical and multiscale models.

Table 2: Comparative Analysis of Retinal Vessel Segmentation Methods

Model	Dataset	Accuracy	Limitations
Matched Filter + Thresholding [13]	DRIVE	Moderate (~0.94)	Struggles with noise and thin vessels. Sensitive to the image quality.
Multiscale Method [15]	STARE	Moderate (~0.95)	Limited adaptability. Weak performance for complex structures.
AGC-UNet (Fu & [14]	DRIVE, STARE	High (~0.96–0.97)	It focuses on local features. It may miss global vessel continuity.

TransUNext[5]	DRIVE, HRF	Very High (~0.97–0.98)	They have a high computational cost. Requires large training data.
TCDDU-Net[1]	DRIVE, CHASE_DB1	Very High (~0.98)	Complex architecture. Training time is high.
Multi-scale Fusion Model [4]	DRIVE, STARE	Very High (~0.98)	May overfits small datasets. Needs careful tuning.
GANVesselNet [2]	DRIVE	High (~0.97)	Training instability. GAN convergence issues.

3.1. Research Gaps

Despite notable progress, several gaps remain in the retinal vessel segmentation. Many models still struggle with detecting thin vessels, particularly in low-contrast regions. These fine structures are often missed or are fragmented. Another key issue is the poor generalization across datasets. Models trained on one dataset do not perform well on others because of variations in image quality and annotations. The use of hierarchical spatial modeling is also limited. Most approaches do not fully capture the relationships between large vessels and fine capillaries. In addition, warping or spatial alignment techniques are rarely integrated, which can improve structural consistency. Existing models often fail to maintain vessel continuity under varying conditions.

4. Future Research Directions

Future research should focus on more structured and adaptive models. Hierarchical image warping approaches can capture vessel patterns better at different scales. They can help preserve both the major vessels and fine capillaries. Combining CNNs with transformers and spatial alignment techniques is a promising approach. This integration can improve both local detail extraction and global context comprehension. There is also a need for multi-dataset training in this regard. Training on diverse datasets can improve the robustness of the model and reduce bias. In addition, self-supervised learning with large-scale data reduces the dependence on manual annotations. This can support more reliable and scalable solutions for real-world clinical applications.

5. Conclusion

Retinal vessel segmentation has significantly progressed over the years. Early methods relied on simple image-processing techniques. However, these methods are limited under complex conditions. Machine learning has introduced better feature handling but requires manual design. Deep learning, especially CNN-based models, improves performance by learning features directly from data. Recently, transformer and hybrid models have enhanced the global understanding of vessel continuity. Despite these improvements, key challenges remain. The detection of thin vessels is difficult. Models often struggle to generalize to different datasets. This highlights the need for more robust and adaptable approaches. Future studies should focus on hierarchical modeling to capture both large and fine vessel structures. Advanced training strategies can further improve the performance and stability. This study moves in that direction by proposing a hierarchical-image warping framework. It aims to combine multi-scale feature learning with spatial alignment. This approach has the potential to improve the accuracy and support reliable clinical applications.

References

[1] N. Lv, L. Xu, Y. Chen, W. Sun, J. Tian, and S. Zhang, "TCDDU-Net: combining transformer and convolutional dual-path decoding U-Net for retinal vessel segmentation," *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-77464-w.

[2] B. Almarri, B. Naveen Kumar, H. Aditya Pai, S. Bhatia Khan, F. Asiri, and T. R. Mahesh, "Redefining retinal vessel segmentation: empowering advanced fundus image analysis with the potential of GANs," *Front. Med.*, vol. 11, no. October, pp. 1–14, 2024, doi: 10.3389/fmed.2024.1470941.

[3] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 9992–10002, 2021, doi: 10.1109/ICCV48922.2021.00986.

[4] J. Zhou, G. Ma, H. He, S. Li, and G. Zhang, "A multi-scale feature extraction and fusion-based model for retinal vessel segmentation in fundus images," *Med. Biol. Eng. Comput.*, vol. 63, no. 2, pp. 595–608, 2025, doi: 10.1007/s11517-024-03223-8.

[5] X. Li, M. Liu, and L. Duan, "TransUNext: towards a more advanced U-shaped framework for automatic vessel segmentation in the fundus image," 2024, [Online]. Available: http://arxiv.org/abs/2411.02724

[6] K. Jin *et al.*, "FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation," *Sci. Data*, vol. 9, no. 1, pp. 1–8, 2022, doi: 10.1038/s41597-022-01564-3.

[7] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," no. Midl, 2018, [Online]. Available: http://arxiv.org/abs/1804.03999

[8] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[9] L. Song, K. Bao, S. Ke, C. Li, J. Zhang, and Y. Zheng, "Efficient Spatio-Temporal Randomly Wired Neural Networks for Traffic Forecasting," in *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2022, pp. 1–8. doi: 10.1109/ICDMW58026.2022.00140.

[10] O. O. Olanrele, S. O. Ismaila, O. A. Adeaga, O. A. Adeyemi, and A. S. Akintaro, "Managing Uncertainty in Production Planning for Fast-Moving Consumer Goods: A Linear Programming and Monte Carlo Simulation Framework," in *2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)*, 2023, pp. 1–8. doi: 10.1109/SEB-SDG57117.2023.10124567.

[11] M. S. Haleem *et al.*, "A Self-Attention Deep Neural Network Regressor for real time blood glucose estimation in paediatric population using physiological signals," *Biomed. Signal Process. Control*, vol. 92, no. February, p. 106065, 2024, doi: 10.1016/j.bspc.2024.106065.

[12] T. A. Qureshi, M. Habib, A. Hunter, and B. Al-Diri, “A manually-labeled, artery/vein classified benchmark for the DRIVE dataset,” in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 485–488. doi: 10.1109/CBMS.2013.6627847.

[13] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 203–210, 2000, doi: 10.1109/42.845178.

[14] X. Fu and N. Zhao, “AGC-UNet:A Global Context Feature Fusion Method Based On U-Net for Retinal Vessel Segmentation,” in *2022 IEEE 2nd International Conference on Information Communication and Software Engineering (ICICSE)*, 2022, pp. 94–99. doi: 10.1109/ICICSE55337.2022.9828894.

[15] A. Budai, J. Hornegger, and G. Michelson, “Multiscale Approach for Blood Vessel Segmentation on Retinal Fundus Images,” *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 13, p. 325, Apr. 2009.