Research Article GU J Sci, Part A, 11(4): 742-758 (2024) 10.54287/gujsa.1575986

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

UKnow-Net: Knowledge-Enhanced U-Net for Improved Retinal Vessel Segmentation

Zeki KUŞ[1*]

[1] *Fatih Sultan Mehmet Vakif University, Department of Computer Engineering, Istanbul, Türkiye*

Keywords

Retinal Vessel Segmentation

Knowledge Distillation and Enhancement

Semi-supervised Learning

Abstract

Retinal vessel segmentation plays a critical role in diagnosing and managing ophthalmic and systemic diseases, as abnormalities in retinal vasculature can indicate disease progression. Traditional manual segmentation by expert ophthalmologists is time-consuming, labor-intensive, and prone to variability, underscoring the need for automated methods. While deep learning approaches like U-Net have advanced retinal vessel segmentation, they often struggle to generalize across diverse datasets due to differences in image acquisition techniques, resolutions, and patient demographics. To address these challenges, I propose UKnow-Net, a knowledge-enhanced U-Net architecture designed to improve retinal vessel segmentation across multiple datasets. UKnow-Net employs a multi-step process involving knowledge distillation and enhancement techniques. First, I train four specialized teacher networks separately on four publicly available retinal vessel segmentation datasets—DRIVE, CHASE_DB1, DCA1, and CHUAC—allowing each to specialize in the unique features of its respective dataset. These teacher networks generate pseudo-labels representing their domain-specific knowledge. We then train a student network using the ensemble of pseudo-labels from all teacher networks, effectively distilling the collective expertise into a unified model capable of generalizing across different datasets. Experiments demonstrate that UKnow-Net outperforms traditional handcrafted networks (such as U-Net, UNet++, and Attention U-Net) and several state-of-the-art models in key performance metrics, including sensitivity, specificity, F1 score, and Intersection over Union (IoU). Specifically, our two variants, UKnowNet-A and UKnowNet-B, show well performance; UKnowNet-A, trained solely on pseudo-labels, achieved higher sensitivity across all datasets, indicating a superior ability to detect true positives, while UKnowNet-B, which combines pseudo-labels with ground truth annotations, achieved balanced precision and recall, leading to higher F1 scores and IoU metrics. The integration of pseudo-labels effectively transfers the collective expertise of the teacher networks to the student network, enhancing generalization and robustness. I aim to ensure fair comparison and reproducibility in future research by publicly sharing our source code and model weights.

Cite

Kuş, Z. (2024). UKnow-Net: Knowledge-Enhanced U-Net for Improved Retinal Vessel Segmentation. *GU J Sci, Part A*, *11*(4), 742-758. doi:10.54287/gujsa.1575986

Author ID (ORCID Number)

0000-0001-8762-7233 Zeki KUŞ

Article Process

Submission Date 30.10.2024
Revision Date 11.11.2024
Accepted Date 14.11.2024
Published Date 30.12.2024

1. INTRODUCTION

Retinal vessel segmentation is a crucial task in medical image analysis, vital for diagnosing and managing eye-related and systemic conditions like diabetic retinopathy, glaucoma, hypertension, and cardiovascular diseases (Abràmoff et al., 2010). The retinal vasculature reflects the body's microcirculation state, and vessel morphology abnormalities can indicate disease progression (Patton et al., 2006; Fraz et al., 2012). Accurate segmentation of retinal vessels enables clinicians to quantify vascular changes, assess disease severity, and monitor treatment efficacy.

*Corresponding Author, e-mail: zkus@fsm.edu.tr

Traditionally, expert ophthalmologists have performed retinal vessel segmentation manually, which is a process that is time-consuming, work-intensive, and prone to variability between different observers. (Niemeijer et al., 2004). Manual feature extraction, while precise, is not feasible for large-scale screenings or real-time applications due to the high costs and resource requirements. Consequently, there is a pressing need for automated segmentation methods that can provide rapid, reliable, and reproducible analysis of retinal images. Early computational approaches employed handcrafted features and classical image processing techniques, such as matched filters (Chaudhuri et al., 1989), morphological operations (Zana & Klein, 2001), and multiscale analysis (Mendonca & Campilho, 2006). These methods aimed to enhance vessel-like structures and suppress background noise. However, they frequently encountered difficulties due to differences in image quality, lighting, and patient anatomy, which restricted their ability to generalize across various datasets and imaging modalities.

The emergence of deep learning has transformed medical image segmentation by allowing models to learn hierarchical feature representations directly from samples (Litjens et al., 2017). Convolutional neural networks (CNNs), particularly fully convolutional networks and encoder-decoder architectures, have demonstrated remarkable success in biomedical segmentation tasks. U-Net (Ronneberger et al., 2015) introduced a symmetric encoder-decoder structure with skip connections, effectively fusing semantic and spatial information. U-Net has been successfully applied to brain tumor segmentation in MRI scans, demonstrating its capability to handle complex structures and improve diagnostic accuracy (Isensee et al., 2018). The model has also been used in dermatology for skin lesion segmentation, aiding in the early detection of melanoma by accurately segmenting lesions from surrounding skin (Anand et al., 2022). U-Net variants have been adapted for object detection tasks, providing precise localization and classification in images (Jaeger et al., 2019). The architecture has been instrumental in analyzing satellite imagery for land cover classification, urban planning, and environmental monitoring (Amritesh et al., 2023). In agriculture, U-Net has been applied to segment leaves for plant phenotyping and disease detection, contributing to advancements in crop management (Mu et al., 2024).

Variants of U-Net and other deep learning architectures have since been applied to retinal vessel segmentation with notable improvements in accuracy (Fu et al., 2016; Liskowski & Krawiec, 2016; Zhou et al., 2021; Liu et al., 2022; Qu et al., 2023). Mou et al. (2019) present CS-Net, a novel network architecture designed to segment curvilinear structures within various medical imaging modalities. CS-Net incorporates channel and spatial attention mechanisms to enhance feature extraction and improve segmentation accuracy, outperforming existing state-of-the-art methods across multiple datasets. Attention Guided Network (AG-Net) (Zhang et al., 2019) enhances retinal image segmentation by preserving structural information through a novel attention-guided filter. This filter integrates an attention mechanism to reduce background noise and improve the accuracy of segmenting retinal structures such as blood vessels and optic discs/cups. Wang et al. (2020) introduce RVSeg-Net, a new and efficient network for segmenting retinal vessels. It tackles challenges like different sizes, small blood vessels, and complex image structures. The network employs a feature pyramid cascade (FPC) module to capture multi-scale features and a multi-frequency convolution (MFC) module to reduce redundancy and improve efficiency, thereby overcoming overfitting issues. SCS-Net (Wu et al., 2021) is a novel scale and context-sensitive network designed for retinal vessel segmentation, addressing challenges such as large-scale variations and complex anatomical contexts. Zhou et al. (2021) present a Study Group Learning (SGL) framework designed to enhance the robustness of retinal vessel segmentation models trained with noisy labels. It introduces a novel method for synthesizing noisy labels and demonstrates improved performance on the DRIVE and CHASE DB1 datasets using a K-fold cross-validation-inspired training scheme. FR-UNet (Liu et al., 2022), combined with a dual-threshold iterative algorithm (DTI) for improved vessel segmentation images. It enhances vessel connectivity and sensitivity by maintaining full image resolution and integrating multiscale feature maps, outperforming state-of-the-art methods on several datasets. Qu et al. (2023) introduce TP-Net, a novel two-path network designed for refined retinal vessel segmentation, addressing challenges in segmenting thin and low-contrast vessels. TP-Net comprises a main-path for detecting vessel trunks, a sub-path for capturing edge information, and a multi-scale feature aggregation module (MFAM) to combine predictions, enhancing segmentation accuracy. Despite these advancements, developing models that generalize well across diverse retinal datasets remains challenging. Differences in image acquisition techniques, resolutions, and patient demographics can hinder a model's performance when applied

to unseen data sources. Addressing this issue requires strategies that enhance a model's ability to learn robust and transferable features.

Knowledge distillation is an effective method for transferring knowledge from a larger, complex teacher model to a smaller, more efficient student model (Hinton et al., 2015). By learning to mimic the teacher's outputs, the student model can achieve competitive performance with reduced complexity. Knowledge enhancement extends this concept by leveraging the expertise of multiple teacher models, allowing the student model to capture a broader range of features and improve generalization (Wang & Yoon, 2021). Several studies have explored knowledge distillation and enhancement in medical imaging. These techniques have been explored to improve model efficiency and performance. For instance, Qin et al. (2021) employed knowledge distillation to enhance the segmentation of medical scans, achieving state-of-the-art results with reduced computational complexity. Shen et al. (2019) utilized knowledge distillation to improve lesion detection in mammography by transferring knowledge from a teacher model trained on a large dataset to a student model with fewer parameters. Zhang and Lu (2024) proposed a knowledge distillation framework for skin lesion segmentation, transferring knowledge from teacher network to enhance performance.

In this study, we propose UKnow-Net, a knowledge-enhanced U-Net architecture designed to improve retinal vessel segmentation across multiple datasets. UKnow-Net addresses the challenges of variability and complexity in retinal images by employing a multi-step process that utilizes the strengths of multiple teacher networks. Each teacher network is trained separately on different datasets—DRIVE, CHASE_DB1, DCA1, and CHUAC—allowing them to specialize in the unique features and patterns specific to their respective datasets. This specialization ensures that each teacher network becomes an expert in its domain, which is crucial for accurately capturing the diverse characteristics of retinal images. A key aspect of UKnow-Net is using pseudo-labels generated by the teacher networks. These pseudo-labels represent the distilled knowledge and expertise of the teacher networks, serving as a rich source of information for training the student network. Using these pseudo-labels effectively distills the domain-specific knowledge from the teacher networks into the student network. This process allows the student network to learn from the teacher networks' collective expertise, enhancing its ability to generalize across different datasets. The experimental studies conducted demonstrate that UKnow-Net outperforms traditional handcrafted networks and state-of-the-art models in several key metrics, including sensitivity, specificity, F1 score, and intersection over union (IoU). UKnowNet-A and UKnowNet-B, the two variants of the proposed model, show superior performance in detecting true positives and achieving balanced precision and recall across different datasets. For instance, UKnowNet-A achieved a sensitivity of 85.37% on the DRIVE dataset, surpassing the handcrafted networks. On the CHASE_DB1 dataset, UKnowNet-A excelled in sensitivity with 89.84%, significantly outperforming the handcrafted networks. The integration of pseudo-labels and ground truths in UKnowNet-B provides a more robust learning framework, allowing it to generalize better across different datasets. Our contributions are as follows:

1. **Unified Model for Multi-Dataset Segmentation:** I develop a single student model capable of segmenting retinal vessels in four publicly available datasets (DRIVE, CHASE_DB1, DCA1, and CHUAC), addressing the challenge of variability across data sources.
2. **Knowledge Enhancement via Teacher Networks:** I train separate teacher networks on each dataset to specialize in their respective domains. By generating pseudo-labels from these specialized networks, we capture diverse feature representations.
3. **Integration of Pseudo-Labels for Student Training:** The student network is trained using the ensemble of pseudo-labels from all teacher networks, effectively distilling knowledge from multiple experts and enhancing generalization.
4. **Comprehensive Evaluation:** We evaluate UKnow-Net on test sets from all four datasets, demonstrating improved segmentation performance and robustness compared to baseline models and existing methods.
5. **Fair Comparison and Reproducibility:** I have publicly shared the source code on GitHub and the best model weight on the same platform.

The rest of this paper is structured as follows: Section 2 outlines the architecture of UKnow-Net and details the experimental setup. Section 3 presents the results and provides a comparative analysis with current state-of-the-art methods. Lastly, Section 4 concludes the study and proposes possible directions for future research.

2. MATERIAL AND METHOD

2.1. UKnow-Net: Knowledge-Enhanced U-Net

U-Net is a convolutional neural network architecture proposed for biomedical image segmentation (Ronneberger et al., 2015). It has a U-shaped structure and consists of encoder and decoder blocks. Encoder networks extract helpful information and features from a given image, followed by maximum pooling layers that decrease the image size by half. After performing encoder blocks, a latent representation of the image is obtained. Then, the image is reconstructed in decoder blocks using deconvolution operations, doubling the image size. It also has skip connections, which allow features from earlier layers to be reused in later layers. This design allows U-Net to effectively utilize both high-level and low-level features, making it particularly well-suited for applications in medical imaging, such as tumor detection and organ segmentation. Its ability to work with relatively small datasets and produce high-quality segmentation maps has led to its widespread adoption in various fields beyond medical imaging, including satellite image analysis and autonomous driving.

Knowledge distillation is a machine learning method where a smaller, more efficient model, called the student, learns from a larger, more complex model, known as the teacher. The aim is to improve the student model's performance by leveraging the insights, patterns, and features learned by the teacher model. This allows the student model to reach a similar level of accuracy and effectiveness as the teacher while being more efficient and suitable for deployment in resource-limited environments. The process involves training the student model to mimic the teacher model's outputs by reducing the discrepancy between their predictions. Knowledge enhancement is the objective of distillation. Knowledge enhancement aims to train student networks to develop more generalized feature representations by leveraging the expertise of specialized teacher networks. This approach enables the student networks to consistently deliver strong performance across various tasks.

In this study, I have proposed UKnow-Net to improve the segmentation performance of the U-Net model for different retinal vessel segmentation problems. Figure 1 illustrates the general structure of the proposed method. The proposed method includes three steps: (1) Four teacher networks are trained separately on different datasets: DRIVE, CHASEDB1, DCA1, and CHUAC. Each teacher network specializes in learning the features and patterns specific to its respective dataset, ensuring that it becomes an expert in that particular domain. I have selected the original U-Net structure (Ronneberger et al., 2015) as the teacher network for each dataset. (2) Step 2 involves using these pre-trained teacher networks to generate pseudo-labels. Each pre-trained teacher network (trained at Step 1) produces pseudo-labels for its corresponding dataset, which are essentially the predictions made by the teacher networks. These pseudo-labels serve as a form of distilled knowledge that will be used to train the student network. (3) In Step 3, the student network is trained using the combination of pseudo-labels generated by all the teacher networks. This approach allows the student network to learn from the collective expertise of all the teacher networks, potentially enhancing its ability to generalize across different datasets. Finally, the student network is evaluated on test sets from all four datasets (DRIVE, CHASEDB1, DCA1, CHUAC) to evaluate its performance and generalization capabilities. This method leverages the strengths of multiple teacher networks to create a robust student model, aiming for improved performance on unseen data.

UKnow-Net addresses the challenges of variability and complexity in retinal images by employing a multi-step process that utilizes the strengths of multiple teacher networks. Each of these teacher networks is trained separately on different datasets—DRIVE, CHASEDB1, DCA1, and CHUAC—allowing them to specialize in the unique features and patterns specific to their respective datasets. This specialization ensures that each teacher network becomes an expert in its domain, which is crucial for accurately capturing the diverse characteristics of retinal images. A key aspect of UKnow-Net is using pseudo-labels generated by the teacher networks. These pseudo-labels represent the distilled knowledge and expertise of the teacher networks, serving as a rich source of information for training the student network. Using these pseudo-labels effectively distills the domain-specific knowledge from the teacher networks into the student network. This process allows the student network to learn from the teacher networks' collective expertise, enhancing its ability to generalize

across different datasets. The robustness and versatility of UKnow-Net are further highlighted by its ability to create a robust student model that performs well across various conditions and datasets. By training the student network on the combination of pseudo-labels from all teacher networks, the method ensures that the student network can generalize effectively to unseen data. This collective learning approach improves the model's performance on test sets from all four datasets and makes it a versatile tool for retinal vessel segmentation. Overall, UKnow-Net represents a significant step forward in developing models that can handle the complexities of retinal images, ultimately aiming for improved performance and generalization capabilities.

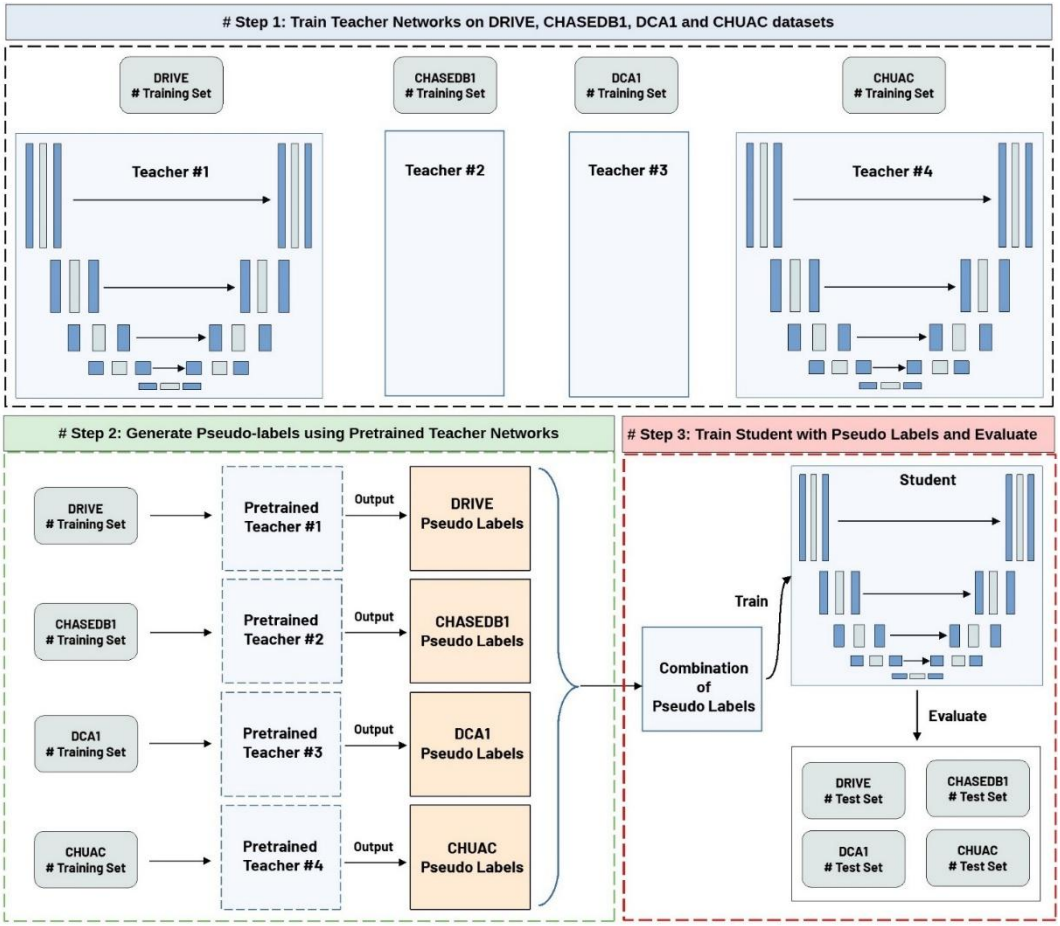

Figure 1. *Overview of the UKnow-Net*

2.2. Experimental Design

2.2.1. Datasets

I have selected the most commonly used retinal vessel segmentation datasets, DRIVE, CHASE_DB1, DCA1, and CHUAC, to evaluate the proposed method. These datasets are widely recognized and frequently used in the field of retinal vessel segmentation (Liu et al., 2022; Kuş & Kiraz, 2023; Qu et al., 2023). They offer a diverse range of imaging modalities (CHUAC-DCA1 vs DRIVE-CHASE_DB1), resolutions, and patient demographics, which are crucial for evaluating the generalization capabilities of our proposed model.

Figure 2 presents the exemplary two images and corresponding ground truths for each dataset.

Digital Retinal Images for Vessel Extraction (DRIVE) (Staal et al., 2004**)**: It is a valuable resource for medical image analysis, specifically for retinal vessel segmentation. It consists of high-resolution retinal images from 40 subjects collected as part of a diabetic retinopathy screening program in the Netherlands. Each image is manually annotated by experts to identify the blood vessels. Researchers and practitioners use this dataset to develop and benchmark automated methods for detecting and analyzing retinal vasculature, crucial in diagnosing and monitoring conditions like diabetic retinopathy, hypertension, and cardiovascular diseases.

I have selected annotations created by the first expert. The dataset contains 40 images, each sized 565×584, and the first 20 images are used for training, while the rest are used for testing, as in other studies (Liu et al., 2022).

CHUAC (Carballal et al., 2018): The dataset originates from the University Hospital Complex of a Coruña in Spain. It is a collection of coronary angiography retinal images gathered for research focused on ophthalmic diseases. It is a valuable resource for developing and testing algorithms for the early detection and assessment of conditions such as diabetic retinopathy and age-related macular degeneration. The dataset contains 30 samples with 512×512 sized images. The first 20 images are used to train the model, and others are used for evaluation (Samuel & Veeramalai, 2021).

The Child Heart and Health Study in England Database 1 (CHASE_DB1) (Carballal et al., 2018): It is a specialized dataset consisting of retinal images collected from a pediatric population. Specifically, the images are obtained from children aged between 7 and 8 years old as part of a study investigating cardiovascular risk factors in early life. CHASE_DB1 includes high-quality retinal photographs along with expert annotations of the blood vessel network. This dataset is essential for researchers focusing on the development of image processing algorithms for vessel segmentation and analysis in children's retinal images. It consists of 28 colored images sized at 999×960 sized. The images are annotated by two different experts, and I have used the annotations made by the first expert. I have used the first 20 images out of 28 for training and the remaining for testing (Liu et al., 2022; Li et al., 2023)

DCA1 (Cervantes-Sanchez et al., 2019): The DCA1 dataset, which stands for Digital Retinal Images for Vessel Extraction, is a widely used collection of retinal images in the medical imaging community for developing and evaluating algorithms. The dataset contains color fundus images showing various retinal conditions, which are crucial for researchers working on automated detection and analysis of retinal diseases. Each image in DCA1 is often accompanied by expert annotations, including segmentation maps of retinal vessels and identification of pathological features. Utilizing the DCA1 dataset enables the advancement of computer-aided diagnostic systems, contributing to improved screening processes and patient care in ophthalmology. It includes 134 samples sized at 300x300. I have chosen 100 images for model training and the remaining 34 for model evaluation (Kuş & Kiraz, 2023).

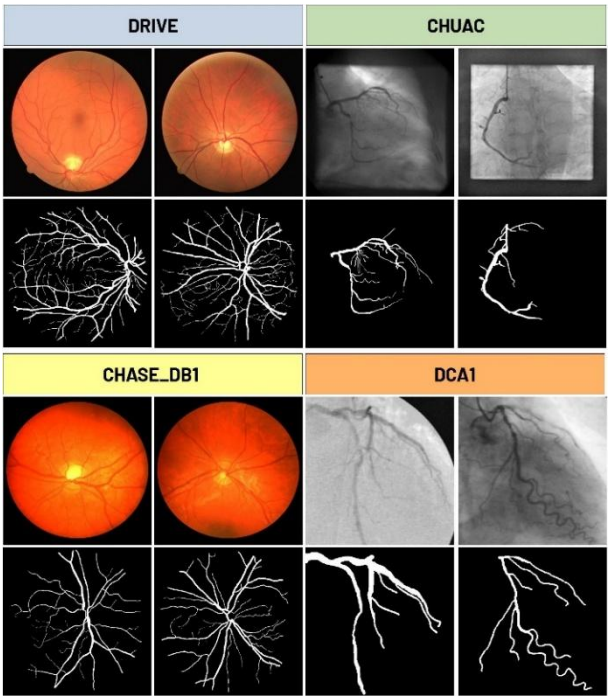

Figure 2. *Exemplary images from DRIVE, CHUAC, CHASE_DB1, and DCA1 datasets. In the figure, the first row represents the input image, and the second row represents the corresponding ground truth image.*

2.2.2. Pre-processing

In the training phases, I employ several preprocessing steps. First, I transform colored images into grayscale and normalize them to standardize the input data. To increase the volume of training data, we extract image patches using a sliding window of size 48×48 pixels with a stride of 48 pixels; overlapping patches are excluded to maintain data independence. I further enhance the diversity of the training dataset by applying data augmentation techniques, including random horizontal and vertical flips and rotations at angles of 90°, 180°, and 270° (Kuş & Kiraz, 2023). Notably, I have not performed any augmentation steps or patch generation to samples in the test set. I have used raw test images to evaluate the actual performance of the model.

2.2.3. Implementation Details

Pytorch library is used for all implementations. During both the training and test phases, each network is trained for 200 epochs. For all experiments, the Adam optimizer and Binary Cross Entropy (BCE) loss are used. I have set the learning rate to 1×10^{-3} and used the batch size of 128. The experiments are executed on a system equipped with a Ryzen 5600X processor, RTX 3060 GPU with 12 GB of memory, and 16 GB of RAM. The source codes are shared on Github.

2.2.4. Performance Measures

I have used several performance measures, including accuracy, sensitivity, specificity, intersection over union (IoU), and F1-score, to evaluate the proposed methods. Sensitivity and specificity measure the model's ability to correctly identify true positives and true negatives, respectively, which are critical for medical applications. The F1 score balances precision and recall, offering insight into the model's overall accuracy. IoU evaluates the spatial accuracy of the segmentation, which is essential for assessing the overlap between predicted and actual vessel structures. Each measure offers a unique perspective on the methods' effectiveness, ensuring a comprehensive assessment across different evaluation criteria. I have also selected these measures to make a fair comparison with reported results in recent literature. Accuracy indicates the overall correctness of the methods. It is the proportion of all predictions that were correct. While useful, accuracy alone might not be sufficient, especially if the data is imbalanced, with some classes being much more common than others. Sensitivity, also known as recall or true positive rate, measures how effectively the methods identify positive cases. It quantifies the percentage of true positives that are accurately identified. This metric shows the ability of the methods to detect the presence of a condition or feature. Specificity, or true negative rate, measures how effectively the methods identify negative cases. It calculates how many actual negatives are accurately classified. Intersection over Union (IoU) evaluates the overlap between the predicted results and the ground truth. Calculated as the area of overlap divided by the area of union between the prediction and the actual result, IoU provides insight into the spatial accuracy of the methods. The F1-score, which calculates their harmonic mean, offers a balance between recall and precision. This metric is especially useful when the data has an imbalanced class distribution, as it accounts for both false positives and false negatives.

3. RESULTS AND DISCUSSION

In this section, I have compared UKnowNet with handcrafted networks, other U-shaped networks, and neural architecture search studies in terms of Sensitivity, Specificity, F1 Score, and Intersection over Union measures. Tables 1 and 2 present the results DRIVE-CHASE_DB1 and DCA1-CHUAC datasets separately. In these tables, red and blue colors highlight the two best performance results obtained for each metric.

3.1. Comparison with Handcrafted Networks

The handcrafted networks, namely U-Net, UNet++, and Attention U-Net, have been foundational in medical image segmentation, particularly for tasks like retinal vessel segmentation. These models are characterized by their U-shaped architectures, which facilitate the capture of both local and global features through a series of downsampling and upsampling layers. Despite their simplicity and effectiveness, these models often rely heavily on the availability of labeled data for training.

In contrast, UKnowNet-A and UKnowNet-B represent more recent advancements that leverage both pseudo-labels and ground truths from multiple datasets, including DRIVE, CHASE_DB1, CHUAC, and DCA.

UKnowNet-A, trained solely on pseudo-labels, demonstrates the potential of semi-supervised learning to achieve competitive performance without extensive labeled data. This approach is particularly advantageous in scenarios where obtaining ground truth annotations is challenging or resource-intensive. UKnowNet-B, on the other hand, combines pseudo-labels with ground truths, likely enhancing its performance by utilizing the strengths of both semi-supervised and supervised learning paradigms.

Table 1. *Performance evaluation of retinal vessel segmentation methods on DRIVE and CHASE_DB1 datasets. The comparison includes traditional handcrafted networks (U-Net, UNet++, Attention U-Net), state-of-the-art models, neural architecture search studies (MedUNAS GA, ODE) and our proposed UKnowNet variants. Metrics include Sensitivity (SEN), Specificity (SPE), F1 Score (F1), and Intersection over Union (IOU). The checkmark (✓) indicates U-shaped architecture. Red indicates the best performance for each metric across the evaluated methods, and blue signifies the second-best performance for each metric.*

Methods	U shape	DRIVE				CHASE_DB1			
		SEN	SPE	F1	IOU	SEN	SPE	F1	IOU
U-Net(Ronneberger et al., 2015)	✓	80.57	98.33	81.41	68.64	76.50	**98.84**	78.98	65.26
UNet++(Zhou et al., 2018)	✓	78.91	98.50	81.14	68.27	83.57	98.32	80.15	66.88
Attention U-Net(Oktay et al., 2018)	✓	79.06	98.31	80.39	67.21	83.84	98.20	79.64	66.17
HRNet(Sun et al., 2019)	✗	80.40	**98.64**	82.65	**70.43**	84.43	98.47	81.48	**68.75**
CS-Net(Mou et al., 2019)	✓	81.70	98.54	80.39	70.17	84.00	98.32	80.42	67.25
AG-Net(Zhang et al., 2019)	✓	81.00	98.48	-	69.65	81.86	98.48	-	66.69
RVSeg-Net(Wang et al., 2020)	✓	81.07	98.45	-	-	80.69	98.36	-	-
SCS-Net(Wu et al., 2021)	✓	82.89	98.38	-	-	83.65	98.39	-	-
SGL(Zhou et al., 2021)	✓	83.80	98.34	83.16	-	86.90	98.43	82.71	-
VSSCNet(Samuel& Veeramalai, 2021)	✗	78.27	98.21	-	-	72.33	**98.65**	-	-
RV-GAN(Kamran et al., 2021)	✗	79.27	**99.69**	**86.90**	-	81.99	98.06	**89.57**	-
FR-UNet(Liu et al., 2022)	✓	83.56	98.37	83.16	**71.20**	**87.98**	98.14	81.51	**68.82**
TP-Net(Qu et al., 2023)	✗	**87.49**	97.58	**85.69**	-	86.00	98.41	**85.18**	-
MedUNAS GA(Kuş & Kiraz, 2023)	✓	84.54	**98.64**	82.06	69.59	86.52	98.44	79.50	66.03
MedUNAS ODE(Kuş & Kiraz, 2023)	✓	83.41	98.36	82.18	69.77	84.50	98.60	80.22	67.01
UKnowNet-A	✓	**85.37**	97.87	82.10	69.66	**89.84**	97.40	78.57	64.76
UKnowNet-B	✓	85.12	98.04	82.65	70.44	87.38	97.94	80.14	66.90

When comparing the performance metrics, UKnowNet-A and UKnowNet-B generally outperform the handcrafted networks across several key metrics. For instance, UKnowNet-A achieves a sensitivity of 85.37% on the DRIVE dataset, surpassing the handcrafted networks, which range from 78.91% to 80.57%. Similarly, UKnowNet-B shows improved specificity and F1 scores, indicating better overall accuracy and balance between precision and recall. This suggests that the integration of pseudo-labels and ground truths in UKnowNet-B provides a more robust learning framework, allowing it to generalize better across different

datasets. On the DRIVE dataset, UKnowNet-A and UKnowNet-B both demonstrate superior sensitivity compared to the handcrafted networks, with UKnowNet-A achieving 85.37% and UKnowNet-B 85.12%, significantly higher than the 80.57% of U-Net, the best among the handcrafted models. This indicates that the UKnowNet models are more effective at correctly identifying true positives. In terms of specificity, the handcrafted networks maintain a slight edge, particularly UNet++ with 98.50%, while UKnowNet-B closely follows at 98.04%, suggesting a minor trade-off in specificity for the increased sensitivity seen in the UKnowNet models. For the F1 score, which balances precision and recall, UKnowNet-B leads with 82.65%, surpassing the handcrafted networks, with U-Net at 81.41% being the highest among them. This suggests that UKnowNet-B offers a more balanced performance. The Intersection over Union (IOU) metric further highlights UKnowNet-B's superior segmentation accuracy, achieving 70.44%, compared to U-Net's 68.64%, the best among the handcrafted models.

On the CHASE_DB1 dataset, UKnowNet-A excels in sensitivity with 89.84%, significantly outperforming the handcrafted networks, where Attention U-Net achieves the highest at 83.84%. This highlights UKnowNet-A's strong ability to detect true positives, likely due to its training with pseudo-labels. However, in terms of specificity, the handcrafted networks, particularly U-Net with 98.84%, outperform the UKnowNet models, which show a slight decrease, with UKnowNet-B at 97.94%. This suggests a trade-off where the UKnowNet models prioritize sensitivity over specificity. The F1 score for UKnowNet-B on CHASE_DB1 is 80.14%, matching UNet++ and indicating a balanced performance between precision and recall. In terms of IOU, UKnowNet-B slightly surpasses UNet++ with 66.90%, reflecting its improved segmentation accuracy. Overall, the UKnowNet models, particularly UKnowNet-B, demonstrate the benefits of integrating pseudo-labels and ground truths, achieving higher sensitivity and balanced F1 scores, while the handcrafted networks maintain high specificity, especially on the CHASE_DB1 dataset. This comparison underscores the effectiveness of modern training strategies in enhancing model performance across different datasets.

Table 2. *Performance evaluation of retinal vessel segmentation methods on DCA1 and CHUAC datasets. The comparison includes traditional handcrafted networks (U-Net, UNet++, Attention U-Net), state-of-the-art models, neural architecture search studies (MedUNAS GA, ODE) and our proposed UKnowNet variants. Metrics include Sensitivity (SEN), Specificity (SPE), F1 Score (F1), and Intersection over Union (IOU). The checkmark (✓) indicates U-shaped architecture. Red indicates the best performance for each metric across the evaluated methods, and blue signifies the second-best performance for each metric.*

Methods	U shape	DCA1				CHUAC			
		SEN	SPE	F1	IOU	SEN	SPE	F1	IOU
U-Net(Ronneberger et al., 2015)	✓	78.16	98.66	77.35	63.07	58.81	**99.40**	67.68	51.15
UNet++(Zhou et al., 2018)	✓	79.54	98.62	77.86	63.75	66.87	**99.37**	73.23	57.77
Attention U-Net(Oktay et al., 2018)	✓	79.86	98.53	77.48	63.24	65.26	99.13	71.54	55.69
HRNet(Sun et al., 2019)	✗	80.07	98.76	**79.19**	**65.54**	74.56	99.06	75.26	60.33
CS-Net(Mou et al., 2019)	✓	78.95	98.67	77.90	63.80	67.35	99.18	71.71	55.89
FR-UNet(Liu et al., 2022)	✓	82.48	98.75	**80.22**	**67.08**	**81.71**	98.68	**76.01**	**61.51**
VSSCNet(Samuel& Veeramalai, 2021)	✗	77.28	98.09	-	-	78.92	97.97	-	-
MedUNAS GA(Kuş & Kiraz, 2023)	✓	80.89	**99.05**	78.20	64.33	78.65	99.16	71.96	56.31
MedUNAS ODE(Kuş & Kiraz, 2023)	✓	**84.12**	**98.89**	78.22	64.37	78.29	99.12	74.56	59.54
UKnowNet-A	✓	81.49	98.58	78.30	64.47	**79.22**	98.72	**75.39**	**60.63**
UKnowNet-B	✓	83.18	98.38	78.01	64.11	73.42	99.04	74.01	58.94

On the DCA1 dataset, UKnowNet-A and UKnowNet-B demonstrate superior performance compared to the handcrafted networks (U-Net, UNet++, and Attention U-Net) across all evaluated metrics. In terms of sensitivity, UKnowNet-B achieves 83.18%, and UKnowNet-A follows with 81.49%, both surpassing the handcrafted networks, where the highest sensitivity is 79.86% by Attention U-Net. This indicates that the UKnowNet models are more effective at detecting true positives. Although the handcrafted networks maintain slightly higher specificity, with U-Net leading at 98.66%, the UKnowNet models show a trade-off, achieving slightly lower specificity but significantly higher sensitivity. For the F1 score, which balances precision and recall, UKnowNet-A and UKnowNet-B achieve 78.30% and 78.01%, respectively, outperforming the handcrafted networks, with UNet++ having the highest F1 score among them at 77.86%. In terms of IOU, both UKnowNet models also lead, with UKnowNet-A achieving 64.47% and UKnowNet-B 64.11%, compared to UNet++'s 63.75%, the best among the handcrafted models.

On the CHUAC dataset, UKnowNet-A particularly excels, achieving a sensitivity of 79.22%, which is significantly higher than the handcrafted networks, where UNet++ reaches 66.87% as the highest. This highlights UKnowNet-A's superior ability to detect true positives. In terms of specificity, the handcrafted networks perform slightly better, with U-Net achieving 99.40%, while UKnowNet-B maintains a competitive specificity of 99.04%. For the F1 score, UKnowNet-A achieves 75.39%, surpassing all handcrafted networks, with UNet++ having the highest F1 score among them at 73.23%. This suggests that UKnowNet-A offers a more balanced performance between precision and recall. In terms of IOU, UKnowNet-A again leads with 60.63%, compared to UNet++'s 57.77%, indicating superior segmentation accuracy.

Overall, the comparison highlights the evolution from traditional handcrafted networks to more sophisticated models like UKnowNet-A and UKnowNet-B, which effectively utilize both labeled and unlabeled data to enhance performance. This shift underscores the growing importance of leveraging diverse data sources and advanced training strategies in developing state-of-the-art models for medical image analysis.

3.2. Comparison with State-of-the-art Networks

When comparing UKnowNet-A and UKnowNet-B with other state-of-the-art models, it's insightful to separate the analysis based on whether the network architecture is U-shaped, as the UKnowNet models are proposed as U-shaped networks.

For the DRIVE dataset, UKnowNet-A and UKnowNet-B demonstrate strong performance among U-shaped networks. UKnowNet-A achieves a sensitivity of 85.37%, and UKnowNet-B follows closely with 85.12%, both outperforming other U-shaped models like SGL and SCS-Net, which have sensitivities of 82.89% and 83.80%, respectively. This indicates that the UKnowNet models are particularly effective at identifying true positives. In terms of specificity, UKnowNet-B achieves 98.04%, which is competitive with other U-shaped models such as CS-Net (98.54%) and AG-Net (98.48%). The F1 score for UKnowNet-B is 82.65%, surpassing CS-Net (80.39%) and closely matching FR-UNet (83.16%), suggesting a balanced performance between precision and recall. For the IOU metric, UKnowNet-B achieves 70.44%, slightly lower than FR-UNet's 71.20% but higher than CS-Net's 70.17%, reflecting strong segmentation accuracy.

On the CHASE_DB1 dataset, UKnowNet-A excels with a sensitivity of 89.84%, leading all U-shaped models and highlighting its capability in detecting true positives effectively. However, in terms of specificity, UKnowNet-B's 97.94% is slightly lower than other U-shaped models like AG-Net, which achieves 98.48%. The F1 score for UKnowNet-B is 80.14%, competitive but slightly lower than FR-UNet's 81.51%. For IOU, UKnowNet-B achieves 66.90%, close to FR-UNet's 68.82%, indicating robust segmentation performance.

In contrast, non-U-shaped models like HRNet and RV-GAN show different strengths. On the DRIVE dataset, HRNet achieves high specificity (98.64%) and IOU (70.43%), comparable to UKnowNet-B. RV-GAN excels in specificity (99.69%) and F1 score (86.90%), outperforming the UKnowNet models in these metrics. On the CHASE_DB1 dataset, VSSC Net achieves the highest specificity (98.65%) but has lower sensitivity (72.33%). RV-GAN leads in F1 score (89.57%), significantly higher than the UKnowNet models.

On the DCA1 dataset, UKnowNet-A and UKnowNet-B demonstrate superior performance among U-shaped networks. UKnowNet-B achieves a sensitivity of 83.18%, and UKnowNet-A follows closely with 81.49%,

both showing robust performance in detecting true positives. In terms of specificity, UKnowNet models have slightly lower values, with UKnowNet-A at 98.58% and UKnowNet-B at 98.38%, compared to FR-UNet's 98.75%. This suggests a trade-off where UKnowNet models prioritize sensitivity over specificity. For the F1 score, UKnowNet-A achieves 78.30%, and UKnowNet-B achieves 78.01%, both competitive but slightly lower than FR-UNet's leading score of 80.22%. In terms of IOU, UKnowNet-A achieves 64.47%, and UKnowNet-B achieves 64.11%, both slightly lower than FR-UNet's 67.08%.

On the CHUAC dataset, UKnowNet-A achieves a better sensitivity value with 79.22% than other U-shaped models (except FR-UNet). UKnowNet-B, with a sensitivity of 73.42%, also performs well, though slightly lower than some other models. In terms of specificity, UKnowNet-B achieves 99.04%, which is competitive with other U-shaped models like CS-Net, which achieves 99.18%. For the F1 score, UKnowNet-A achieves 75.39%, the second highest among U-shaped models, indicating a balanced performance between precision and recall. This surpasses FR-UNet's 76.01%, which is the highest among other models. In terms of IOU, UKnowNet-A achieves 60.63%, the second highest among U-shaped models, reflecting its strong segmentation accuracy. UKnowNet-B achieves an IOU of 58.94%, which is competitive but slightly lower than FR-UNet's 61.51%.

In contrast to U-shaped models, non-U-shaped models like HRNet and VSSC Net demonstrate distinct advantages. On the DCA1 dataset, HRNet achieves a high specificity of 98.76% and an F1 score of 79.19%, making its performance comparable to that of UKnowNet models. However, VSSC Net exhibits lower sensitivity at 77.28% compared to UKnowNet models. On the CHUAC dataset, HRNet again shows strong performance with a sensitivity of 74.56% and an F1 score of 75.26%, remaining competitive with UKnowNet models. In contrast, VSSC Net achieves a higher sensitivity of 78.92%, but its specificity is lower at 97.97%.

3.3. Comparison with Neural Architecture Search Studies

A detailed comparison of UKnowNet-A and UKnowNet-B with neural architecture search studies (MedUNAS GA and MedUNAS ODE) reveals distinct performance characteristics across different metrics and datasets. On the DRIVE dataset, UKnowNet models demonstrate superior sensitivity, with UKnowNet-A achieving 85.37% and UKnowNet-B reaching 85.12%, both surpassing MedUNAS GA (84.54%) and MedUNAS ODE (83.41%). This indicates that UKnowNet models are more effective at detecting true positives in vessel segmentation. However, in terms of specificity, MedUNAS GA leads with 98.64%, followed by MedUNAS ODE at 98.36%, while UKnowNet-B achieves 98.04% and UKnowNet-A reaches 97.87%. The F1 score comparison on the DRIVE dataset shows UKnowNet-B achieving the highest value at 82.65%, slightly outperforming both MedUNAS ODE (82.18%) and MedUNAS GA (82.06%), with UKnowNet-A following closely at 82.10%. This indicates that UKnowNet-B offers the most balanced performance between precision and recall. In terms of IOU, UKnowNet-B leads with 70.44%, demonstrating superior segmentation accuracy compared to MedUNAS ODE (69.77%) and MedUNAS GA (69.59%), while UKnowNet-A achieves 69.66%.

On the CHASE_DB1 dataset, the performance differences become more pronounced. UKnowNet-A achieves remarkable sensitivity at 89.84%, significantly outperforming MedUNAS GA (86.52%) and MedUNAS ODE (84.50%), with UKnowNet-B also showing strong performance at 87.38%. However, the specificity results show MedUNAS ODE leading at 98.60%, followed by MedUNAS GA at 98.44%, while UKnowNet-B achieves 97.94% and UKnowNet-A reaches 97.40%. This further emphasizes the sensitivity-specificity trade-off in the UKnowNet models' design. The F1 score comparison on CHASE_DB1 shows MedUNAS ODE achieving 80.22%, closely followed by UKnowNet-B at 80.14% and MedUNAS GA at 79.50%, while UKnowNet-A reaches 78.57%. For IOU, MedUNAS ODE leads with 67.01%, followed by UKnowNet-B at 66.90% and MedUNAS GA at 66.03%, with UKnowNet-A achieving 64.76%. These results suggest that while UKnowNet models excel in sensitivity, MedUNAS models maintain more balanced performance across all metrics.

On the DCA1 dataset, MedUNAS ODE demonstrates the highest sensitivity at 84.12%, followed closely by UKnowNet-B at 83.18%, while UKnowNet-A achieves 81.49% and MedUNAS GA reaches 80.89%. This indicates that MedUNAS ODE is particularly effective at detecting true positives, though UKnowNet-B maintains very competitive performance. In terms of specificity, MedUNAS models show superior

performance, with MedUNAS GA achieving the highest value at 99.05% and MedUNAS ODE following at 98.89%, while UKnowNet-A and UKnowNet-B achieve 98.58% and 98.38% respectively. This suggests that MedUNAS models are more effective at reducing false positives. The F1 score comparison reveals a close competition, with UKnowNet-A slightly leading at 78.30%, followed by MedUNAS ODE at 78.22%, MedUNAS GA at 78.20%, and UKnowNet-B at 78.01%. These nearly identical F1 scores indicate that all models maintain a similar balance between precision and recall. For the IOU metric, UKnowNet-A achieves the highest value among these models at 64.47%, slightly surpassing MedUNAS ODE (64.37%), MedUNAS GA (64.33%), and UKnowNet-B (64.11%), suggesting slightly better segmentation accuracy.

On the CHUAC dataset, UKnowNet-A leads in sensitivity with 79.22%, slightly outperforming MedUNAS GA (78.65%) and MedUNAS ODE (78.29%), while UKnowNet-B achieves 73.42%. This demonstrates UKnowNet-A's superior ability to detect true positives in this dataset. However, in terms of specificity, MedUNAS models again show stronger performance, with MedUNAS GA achieving 99.16% and MedUNAS ODE reaching 99.12%, compared to UKnowNet-B's 99.04% and UKnowNet-A's 98.72%. The F1 score comparison on CHUAC shows UKnowNet-A achieving 75.39%, followed by MedUNAS ODE at 74.56%, UKnowNet-B at 74.01%, and MedUNAS GA at 71.96%. This indicates that UKnowNet-A provides the most balanced performance between precision and recall. In terms of IOU, UKnowNet-A leads with 60.63%, followed by MedUNAS ODE at 59.54%, UKnowNet-B at 58.94%, and MedUNAS GA at 56.31%, demonstrating UKnowNet-A's superior segmentation accuracy on this dataset.

Overall, this comparison highlights the distinct strengths of each approach: UKnowNet models, particularly UKnowNet-A, outperform in sensitivity, true positive detection, and balanced performance, while MedUNAS models offer higher specificity. UKnowNet-B emerges as a strong compromise, achieving competitive results across all metrics while maintaining the sensitivity advantages of the UKnowNet architecture. These differences reflect the underlying design and training strategies of each approach, with UKnowNet models potentially being more suitable for applications where high sensitivity and balanced performance are crucial, while MedUNAS models might be preferred in scenarios requiring more specificity.

3.4. Evaluation of Pre-trained Networks with Fine-tuning

In this study, I have conducted an ablation analysis to evaluate whether conventional pre-trained convolutional neural networks—DenseNet-121, EfficientNet, ResNet-18, ResNet-34, and ResNet-50—initialized with ImageNet weights and fine-tuned over 50 epochs (200 epochs for UKnowNet-A) using identical training protocols as proposed UKnowNet-A model (see Section 2.1), could achieve performance comparable to UKnowNet-A. I have selected these pre-trained convolutional neural networks for the following reason: DenseNet-121 offers a good balance between model complexity and performance, making it suitable for medical image analysis tasks where feature reuse is beneficial. EfficientNet models are designed to achieve high performance with fewer parameters by scaling depth, width, and resolution. ResNet architectures are renowned for their residual learning framework, which helps in training deeper networks by diminishing the vanishing gradient problem. The evaluation focused on key performance metrics: Sensitivity, Specificity, F1 Score, and Intersection over Union across the four datasets. The results are shown in Figure 3. In Figure 3, Best represents the obtained results from UKnowNet-A for each metric and dataset.

For DRIVE, EfficientNet stands out with the highest Sensitivity (83.15%) and IOU (67.18%), indicating its superior ability to identify positive instances correctly and accurately overlap predictions with ground truth. ResNet-50 also performs commendably, with a SEN of 76.83% and an IOU of 66.71%. However, the "Best" results surpass all models, achieving a SEN of 85.37%, SPE of 97.87%, F1 of 82.1%, and IOU of 69.66%, demonstrating the potential of UKnowNet with pseudo-labels.

On CHASE_DB1, EfficientNet again leads with a high SEN of 87.32%, significantly outperforming other models. ResNet-34 follows with a respectable SEN of 80.04%. The IOU metrics reveal ResNet-34 as the best-performing network among the models with 62.57%, closely following the "Best" result of 64.76%. Notably, despite EfficientNet's high SEN, its F1 score (72.76%) is lower than the "Best," highlighting a trade-off between sensitivity and overall balance captured by the F1 metric.

Performance across models is more uniform in DCA1 dataset. ResNet-34 achieves the highest SEN (82.2%) and competitive IOU (61.49%), while DenseNet-121 and ResNet-18 maintain solid performances with SENs around 81%. The "Best" results continue to lead with an SEN of 81.49% and IOU of 64.47%, indicating that even marginal improvements can be crucial in specific applications. The F1 scores are relatively close among models, with the "Best" achieving the highest at 78.3%.

CHUAC dataset presents the most significant performance gaps among the models. EfficientNet achieves the highest SEN (72.95%) among the networks, yet it falls short of the "Best" result by over 6 percentage points. ResNet variants exhibit lower SENs, with ResNet-18 at 62.12%. Specificity remains high across all models, hovering around 98%, which underscores their effectiveness in correctly identifying negative instances. The "Best" results dramatically outperform individual models across all metrics, particularly in SEN (79.22%) and F1 Score (75.39%), highlighting substantial gains possible through UKnowNet and pseudo-label techniques.

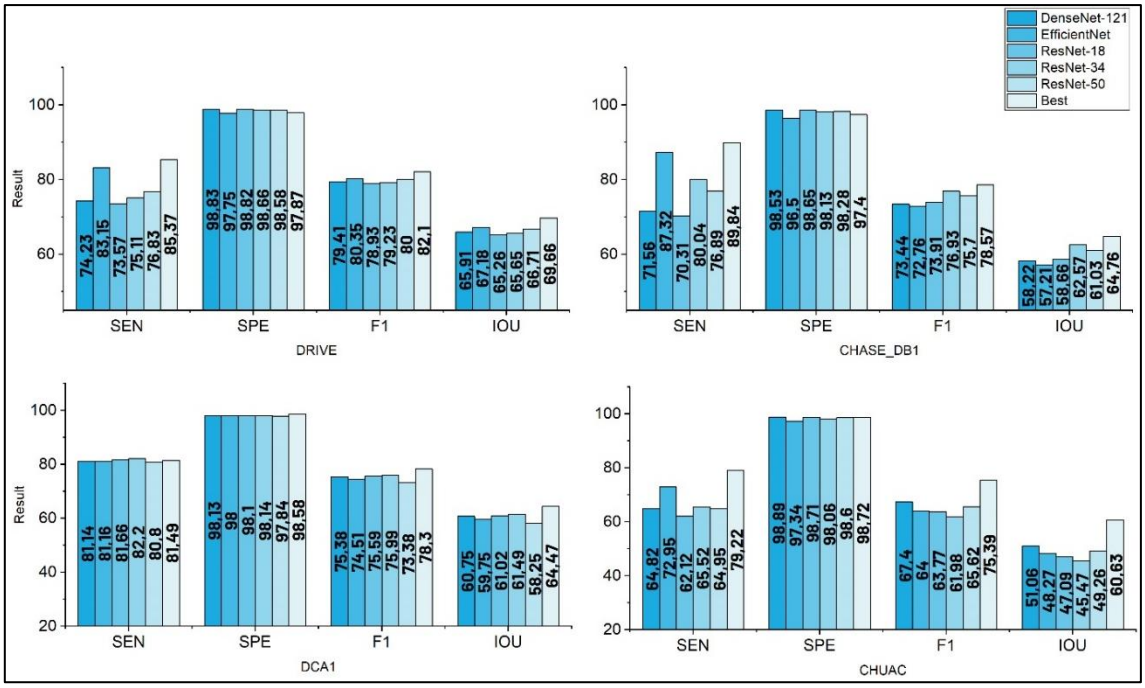

Figure 3. *Performance comparison of different neural network architectures (DenseNet-121, EfficientNet, ResNet-18, ResNet-34, ResNet-50, and Best) across four datasets (DRIVE, CHASE_DB1, DCA1, and CHUAC) using four evaluation metrics (SEN: Sensitivity, SPE: Specificity, F1: F1-Score, and IOU: Intersection over Union). Best represents the obtained results from UKnowNet-A for each metric and dataset.*

Across all evaluated datasets, EfficientNet consistently demonstrated strong sensitivity, while ResNet architectures, particularly ResNet-34 and ResNet-50, maintained high specificity and balanced performances across other metrics. DenseNet-121 also showed competitive specificity and F1 scores. However, UKnowNet-A consistently outperformed all individual pre-trained models across all metrics and datasets. This superior performance can be attributed to the use of pseudo-labels from multiple datasets, which likely enhances the model's ability to generalize and accurately segment diverse retinal images The ablation study confirms that while standard pre-trained networks provide robust baseline performances with limited fine-tuning, the comprehensive training approach employed by UKnowNet-A—utilizing multi-dataset pseudo labeling and advanced fine-tuning strategies—yields significant improvements. This confirms its efficacy in achieving state-of-the-art results in retinal image segmentation.

These findings underscore the critical importance of customized training methods in medical image segmentation tasks. While pre-trained models offer a strong baseline performance, integrating pseudo-labels from diverse sources and using specialized training techniques significantly improve segmentation accuracy and robustness. Future research could investigate the combined effects of ensemble learning and the further augmentation of training data to enhance the precision of segmentation models.

4. CONCLUSION

In this study, I have introduced UKnow-Net, a knowledge-enhanced U-Net architecture designed to improve retinal vessel segmentation across multiple datasets. By leveraging the strengths of multiple teacher networks, each trained on distinct retinal image datasets (DRIVE, CHASE_DB1, DCA1, and CHUAC), UKnow-Net addresses the challenges posed by variability and complexity in retinal images. UKnow-Net involves a multi-step process where specialized teacher networks use pseudo-labels to train a unified student network. This approach allows the student network to distill domain-specific knowledge from the teacher networks, enhancing its ability to generalize across diverse datasets. The integration of pseudo-labels effectively transfers the collective expertise of the teacher networks to the student network, resulting in improved segmentation performance.

The experimental results demonstrate that UKnow-Net outperforms traditional handcrafted networks (U-Net, UNet++, and Attention U-Net) and several state-of-the-art models in key performance metrics, including sensitivity, specificity, F1 score, and intersection over union (IoU). Specifically, UKnowNet-A, trained solely on pseudo-labels, achieved higher sensitivity across all datasets, indicating its superior ability to detect true positives. UKnowNet-B, which combines pseudo-labels with ground truth annotations, achieved a balanced performance in precision and recall, leading to higher F1 scores and IoU metrics. Our ablation study further confirms the effectiveness of the proposed knowledge enhancement approach. The comparison with pre-trained convolutional neural networks fine-tuned on the same datasets revealed that UKnow-Net consistently delivers superior performance. This underscores the importance of leveraging pseudo-labels from multiple specialized networks to improve generalization and robustness in medical image segmentation tasks. I aim to ensure fairness in comparison and reproducibility in future research by publicly sharing the source code and the best model weights. This study highlights the potential of knowledge enhancement techniques in medical imaging and encourages further exploration of multi-teacher knowledge distillation methods.

Future work could explore integrating additional datasets to enhance generalization further, applying the knowledge enhancement framework to other medical imaging tasks, and optimizing the model for real-time clinical applications.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, *3*, 169-208. https://doi.org/10.1109/rbme.2010.2084567

Amritesh, Owais, M. M., Vemula, V., Amit, A., & Natarajan, S. (2023, May 3-5). *Localised Land-Use Classification Using U-Net and Satellite Imaging*. In: A. J. Kulkarni, & N. Cheikhrouhou (Eds.), Proceedings of the 2nd International Conference on Information Science and Applications (ICISA 2023), (pp. 235-248), Pune, India. https://doi.org/10.1007/978-981-99-6984-5_15

Anand, V., Gupta, S., Koundal, D., Nayak, S. R., Barsocchi, P., & Bhoi, A. K. (2022). Modified U-net architecture for segmentation of skin lesion. *Sensors*, *22*(3), 867. https://doi.org/10.3390/s22030867

Carballal, A., Novoa, F. J., Fernandez-Lozano, C., García-Guimaraes, M., Aldama-López, G., Calviño-Santos, R., Vazquez-Rodriguez, J. M., & Pazos, A. (2018). Automatic multiscale vascular image segmentation algorithm for coronary angiography. *Biomedical Signal Processing and Control*, *46*, 1-9. https://doi.org/10.1016/j.bspc.2018.06.007

Cervantes-Sanchez, F., Cruz-Aceves, I., Hernandez-Aguirre, A., Hernandez-Gonzalez, M. A., & Solorio-Meza, S. E. (2019). Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. *Applied Sciences*, *9*(24), 5507. https://doi.org/10.3390/app9245507

Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., & Goldbaum, M. (1989). Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on Medical Imaging*, *8*(3), 263-269. https://doi.org/10.1109/42.34715

Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). Blood vessel segmentation methodologies in retinal images–a survey. *Computer Methods and Programs in Biomedicine*, *108*(1), 407-433. https://doi.org/10.1016/j.cmpb.2012.03.009

Fu, H., Xu, Y., Lin, S., Kee Wong, D. W., & Liu, J. (2016, October 17-21). *Deepvessel: Retinal vessel segmentation via deep learning and conditional random field*. In: S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, & W. Wells (Eds.), Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016), (Part II, pp. 132-139). Athens, Greece. https://doi.org/10.1007/978-3-319-46723-8_16

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the Knowledge in a Neural Network. https://doi.org/10.48550/arXiv.1503.02531

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2018, September 16). *No New-Net*. In: A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, & T. van Walsum (Eds.), Proceedings of the 4th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes 2018), (Part II, pp. 234-244), Granada, Spain. https://doi.org/10.1007/978-3-030-11726-9_21

Jaeger, P. F., Kohl, S. A. A., Bickelhaupt, S., Isensee, F., Kuder, T. A., Schlemmer, H.-P., & Maier-Hein, K. H. (2019, December 13). *Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection*. In: Proceedings of the Machine Learning for Health Workshop (ML4H), (pp. 171-183), Vancouver, Canada.

Kamran, S. A., Hossain, K. F., Tavakkoli, A., Zuckerbrod, S. L., Sanders, K. M., & Baker, S. A. (2021, September 27 - October 1). *RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network*. In: M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2021), (Part VIII, pp. 34-44), Strasbourg, France. https://doi.org/10.1007/978-3-030-87237-3_4

Kuş, Z., & Kiraz, B. (2023). Evolutionary architecture optimization for retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, *27*(2), 5895-5903 https://doi.org/10.1109/JBHI.2023.3314981

Li, J., Gao, G., Liu, Y., & Yang, L. (2023). MAGF-Net: A multiscale attention-guided fusion network for retinal vessel segmentation. *Measurement*, *206*, 112316. https://doi.org/10.1016/j.measurement.2022.112316

Liskowski, P., & Krawiec, K. (2016). Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, *35*(11), 2369-2380. https://doi.org/10.1109/TMI.2016.2546227

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60-88. https://doi.org/10.1016/j.media.2017.07.005

Liu, W., Yang, H., Tian, T., Cao, Z., Pan, X., Xu, W., Jin, Y., & Gao, F. (2022). Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. *IEEE Journal of Biomedical and Health Informatics*, *26*(9), 4623-4634. https://doi.org/10.1109/JBHI.2022.3188710

Mendonca, A. M., & Campilho, A. (2006). Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging*, *25*(9), 1200-1213. https://doi.org/10.1109/TMI.2006.879955

Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., & Liu, J. (2019, October 13–17). *CS-Net: Channel and spatial attention network for curvilinear structure segmentation*. In: D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P-T. Yap, & A. Khan (Eds.), Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2019), (Part I, pp. 721-730), Shenzhen, China. https://doi.org/10.1007/978-3-030-32239-7_80

Mu, Y., Li, K., Sun, Y., & Bao, Y. (2024). Semantic segmentation of corn leaf blotch disease images based on U-Net integrated with RFB structure and dual attention mechanism. *Agronomy (Basel, Switzerland)*, *14*(11), 2652. https://doi.org/10.3390/agronomy14112652

Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., & Abramoff, M. D. (2004, February 14-19). *Comparative study of retinal vessel segmentation methods on a new publicly available database*. In: J. M. Fitzpatrick & M. Sonka (Eds.), Proceedings of the Medical Imaging 2004: Image Processing. SPIE Proceedings, (Vol. 5370, pp. 648-656), San Diego, California, United States. https://doi.org/10.1117/12.535349

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018, July 4-6). Attention u-net: Learning where to look for the pancreas. In: Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands. https://doi.org/10.48550/arXiv.1804.03999

Patton, N., Aslam, T. M., MacGillivray, T., Deary, I. J., Dhillon, B., Eikelboom, R. H., Yogesan, K., & Constable, I. J. (2006). Retinal image analysis: concepts, applications and potential. *Progress in Retinal and Eye Research*, *25*(1), 99-127. https://doi.org/10.1016/j.preteyeres.2005.07.001

Qin, D., Bu, J.-J., Liu, Z., Shen, X., Zhou, S., Gu, J.-J., Wang, Z.-H., Wu, L., & Dai, H.-F. (2021). Efficient medical image segmentation based on knowledge distillation. *IEEE Transactions on Medical Imaging*, *40*(12), 3820-3831. https://doi.org/10.1109/TMI.2021.3098703

Qu, Z., Zhuo, L., Cao, J., Li, X., Yin, H., & Wang, Z. (2023). TP-net: Two-path network for retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, *27*(4), 1979-1990. https://doi.org/10.1109/JBHI.2023.3237704

Ronneberger, O., Fischer, P., & Brox, T. (2015, October 5-9). U-net: Convolutional networks for biomedical image segmentation. In: N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), (Part III, pp. 234-241), Munich, Germany. https://doi.org/10.1007/978-3-319-24574-4_28

Samuel, P. M., & Veeramalai, T. (2021). VSSC Net: vessel specific skip chain convolutional network for blood vessel segmentation. *Computer Methods and Programs in Biomedicine*, *198*, 105769. https://doi.org/10.1016/j.cmpb.2020.105769

Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports*, *9*(1), 12495. https://doi.org/10.1038/s41598-019-48995-4

Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, *23*(4), 501-509. https://doi.org/10.1109/TMI.2004.825627

Sun, K., Xiao, B., Liu, D., & Wang, J. (2019, June 15-20). *Deep high-resolution representation learning for human pose estimation*. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5693-5703), Long Beach, CA, USA. https://doi.org/10.1109/CVPR.2019.00584

Wang, L., & Yoon, K.-J. (2021). Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(6), 3048-3068. https://doi.ieeecomputersociety.org/10.1109/TPAMI.2021.3055564

Wang, W., Zhong, J., Wu, H., Wen, Z., & Qin, J. (2020, October 4–8). *RVSeg-net: An efficient feature pyramid cascade network for retinal vessel segmentation*. In: A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, & L. Joskowicz (Eds.), Proceedings of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020), (Part V, pp. 796-805), Lima, Peru. https://doi.org/10.1007/978-3-030-59722-1_77

Wu, H., Wang, W., Zhong, J., Lei, B., Wen, Z., & Qin, J. (2021). SCS-net: A scale and context sensitive network for retinal vessel segmentation. *Medical Image Analysis*, *70*, 102025. https://doi.org/10.1016/j.media.2021.102025

Zana, F., & Klein, J.-C. (2001). Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on Image Processing*, *10*(7), 1010-1019. https://doi.org/10.1109/83.931095

Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., Tan, M., & Xu, Y. (2019, October 13-17). *Attention guided network for retinal image segmentation*. In: D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P-T. Yap, & A. Khan (Eds.), Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2019), (Part I, pp. 797-805), Shenzhen, China. https://doi.org/10.1007/978-3-030-32239-7_88

Zhang, Z., & Lu, B. (2024). Efficient skin lesion segmentation with boundary distillation. *Medical & Biological Engineering & Computing*, *62*(9), 2703–2716. https://doi.org/10.1007/s11517-024-03095-y

Zhou, Y., Yu, H., & Shi, H. (2021, September 27–October 1). *Study Group Learning: Improving Retinal Vessel Segmentation Trained with Noisy Labels*. In: M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, & C. Essert (Eds.), Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2021), (Part I, pp. 57-67), Strasbourg, France. https://doi.org/10.1007/978-3-030-87193-2_6

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018, September 20). *Unet++: A nested u-net architecture for medical image segmentation*. In: D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, & A. Madabhushi (Eds.), Proceedings of the 4th International Workshop and 8th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA 2018, ML-CDS 2018), (pp. 3-11). Granada, Spain. https://doi.org/10.1007/978-3-030-00889-5_1