Dense-PMSFNet: DenseNet Pyramidal Multi-Scale Fusion Network for Retinal Vasculature and FAZ Segmentation in OCTA Images

Nisan Pranavah Raja, Srivatsan Sarvesan, Anju Thomas, and Varun P. Gopi, *Senior Member, IEEE*

***Abstract*—Diabetic Retinopathy (DR) is a leading cause of vision loss, particularly among working-age populations in developing countries, due to prolonged high blood sugar levels. This condition damages the retina, and its progression can result in permanent blindness. Early diagnosis and monitoring are critical to preventing vision impairment, with Optical Coherence Tomography Angiography (OCTA) offering advanced imaging to observe retinal blood vessels. Accurate segmentation of retinal structures, including capillaries, arteries, veins, and the Foveal Avascular Zone (FAZ), is essential for effective diagnosis and treatment planning. We propose the DenseNet Pyramidal Multi-Scale Fusion Network (Dense-PMSFNet), a deep learning-based model for retinal vessel and zone segmentation from OCTA images. The model integrates a DenseNet Encoder to capture local semantic context, a Multi-Scale Pyramidal Fusion Module (MSPFM) for feature merging across scales, and a Deep Fusion technique to enhance decoder outputs. Evaluated on four segmentation tasks—capillaries, arteries, veins, and FAZ—Dense-PMSFNet achieves state-of-the-art results, outperforming existing methods in Dice Coefficient (Dice) and Intersection over Union (IoU) metrics. Specifically, the model achieves Dice scores of 87.51% and 90.88% for capillaries, 88.16% and 90.21% for arteries, and 87.38% and 89.23% for veins. FAZ segmentation yields exceptional results of 94.84% and 97.92% on the OCTA_6mm and OCTA_3mm subsets, respectively. These outcomes underscore the robustness and accuracy of Dense-PMSFNet, offering a practical, reliable solution for clinicians in DR diagnosis and treatment.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9439

***Index Terms*—Artificial Intelligence, Artery, Capillary, Deep Fusion, Deep Learning, Densenet Encoder, Foveal Avascular Zone, Multiscale Pyramidal Fusion Module, Optical Coherence Tomography Angiography, Vein.**

I. INTRODUCTION

DIABETIC retinopathy (DR) is a leading cause of blindness worldwide, particularly among the working-age population in developing countries, where the rising prevalence of Diabetes Mellitus (DM) disproportionately impacts public health [1], [2]. It is estimated that 80% of people with diabetes reside in developing countries, and the number of individuals with DR has reached 93 million globally, with a projected 69% increase in DM cases in these regions by 2030 [2]. Unfortunately, blindness due to DR, which is largely preventable through early detection and treatment, remains a significant burden in developing nations due to delayed diagnosis and limited access to effective healthcare interventions [2]. Optical Coherence Tomography Angiography (OCTA) is a non-invasive imaging modality that provides detailed retinal blood vessel images, offering disease insights [3], [4]. Deep learning has emerged as a powerful tool for retinal vasculature segmentation in OCTA images, essential for detecting lesions, monitoring disease progression, and guiding treatment decisions [5]. Recent studies use deep learning to segment capillaries, arteries, veins, and the Foveal Avascular Zone (FAZ). Traditional techniques such as thresholding and filtering often fail due to device variability, producing artifacts and noise [6]. Deep learning, on the other hand, bypasses manual calibration and improves accuracy. Arteries and veins offer critical insights into DR, hypertension, and age-related macular degeneration [7], [8]. For example, DR is associated with increased arteriolar tortuosity and venular widening [8]. Precise segmentation of these structures is crucial for diagnosis and management. OCTA surpasses traditional fundus photography by visualizing retinal blood flow without dye injection and offering high-resolution microvasculature images [9]. The FAZ is a non-vascularized retinal area that is a significant disease biomarker. However, unclear vessel boundaries pose challenges in FAZ segmentation [10]. Clinically, accurate FAZ segmentation helps diagnose and manage retinal diseases [11].

Deep learning applications in OCTA imaging address the problems of manual calibration and boundary inaccuracies, allowing for robust, automated segmentation. Although there has been tremendous progress in deep learning, accurately segmenting capillaries, arteries, veins, and the FAZ remains a significant research gap. Existing deep learning models often struggle with the intricate and overlapping structures of retinal vasculature, leading to poor differentiation between arteries and veins. The other issues are the undefined and blurry boundaries of the FAZ, especially in diseased states, which degrade the segmentation accuracy. Moreover, high variability in OCTA image quality and noise further worsens these problems and reduces robustness. Current methods cannot address these complexities comprehensively, necessitating the development of novel approaches tailored to overcome these limitations. The proposed approach utilizes annotated datasets to train a model capable of segmenting capillaries, arteries,

---

The associate editor coordinating the review of this manuscript and approving it for publication was Samuel Ortega (*Corresponding author: Varun P. Gopi*).

N. P. Raja, S. Sarvesan, A. Thomas, and Varun P. Gopi are with the Department of Electronics and Communication Engineering, National Institute of Tiruchirappalli, Tiruchirappalli, Tamil Nadu 620015, India (e-mails: 408122004@nitt.edu, 108121124@nitt.edu, varun@nitt.edu, and anju@nitt.edu).

veins, and FAZ regions. Our segmentation process leverages the Dense-PMSFNet architecture, designed for precisely segmenting retinal vasculature and FAZ in OCTA images. The DenseNet encoder and the MultiScale Pyramidal Fusion Module (MSPFM) help to efficiently extract multi-scale features across different scales, capturing fine details and larger structures. This approach improves segmentation accuracy and robustness, making it highly effective for early detection and monitoring of DR. It provides a scalable solution with significant clinical and research applications in ophthalmology.

The rest of the paper is organized as follows: Section II reviews related works on OCTA image segmentation and deep learning methodologies and the research gap. Section III details the proposed methodology, including network architecture and fusion modules. Section IV presents the experiments and analysis with details such as dataset description, experimental settings, and evaluation metrics. Section V presents the results along with qualitative and quantitative analysis of the segmentation results, showcasing the effectiveness of the proposed approach compared to state-of-the-art methods. Section VI includes a detailed discussion of the ablation study and the advantages of the proposed model, the comparison of the computational complexities of different models with the proposed model, and the limitations and areas for development. Finally, Section VII concludes how the proposed model helps to achieve the specific objectives and future research works.

II. RELATED WORKS

This section briefly reviews existing research on segmentation tasks using OCTA images in medical imaging, focusing on retinal vasculature segmentation. Deep learning techniques have been widely employed for this purpose. To address the challenge of labeling OCTA and fundus images, semisupervised learning has been used to extract meaningful areas from image patches [12]. Recent studies introduced scribble-like annotation approaches for retinal structure segmentation in 2D OCTA images and other medical imaging tasks [13]. OCTAve, a weakly supervised framework, combines scribble annotations with expert-labeled, unpaired ground truth data during training. Self-Supervised Deep Supervision (SSDS) further enhances model performance without additional data [14]. Segmentation tasks like FAZ and vessel segmentation have utilized deep neural models for their robustness and superior performance over conventional methods [15], [16], [17]. CS2-Net [18] integrating spatial and channel attention modules within a U-Net framework, achieved improved accuracy compared to standard segmentation algorithms like U-Net [19], U-Net++ [20], and other similar networks. Several studies have utilized multi-scale feature fusion for improved segmentation. FPN-MSPF-Net [21] integrates dual predictions and attention-based multi-scale fusion for refined results. MS-Net [22] combines a Multi-Scale Attention Module with a Stacked Feature Pyramid Module to dynamically adjust receptive fields for better focus. MCFSA-Net [23] emphasizes tiny retinal vessels using attention-based fusion and spatial activation, achieving high accuracy across datasets. PCAM-PRDC-Net [24] incorporates a pyramid channel attention module and optimized convolution blocks to enhance sensitivity for fine vessel segmentation.

Capillaries appear as a dense mesh in OCTA images. An Image Magnification Network refines its cross-sectional view by enhancing topological clarity and suppressing noise and artifacts in the FAZ region using a FAZ mask. Label Adversarial learning further improves vessel visibility and segmentation accuracy [25]. For retinal artery and vein segmentation, traditional methods, such as graph-theoretic algorithms, analyze vascular architecture, classifying vessel segments by crossing points and assigning artery or vein labels [26]. Clustering is also used for reconstructing retinal topology [27]. Deep learning approaches, including cascaded neural networks, enable simultaneous segmentation of arteries and veins from dual-modal fundus images [28]. OCTA surpasses fundus images in artery-vein segmentation, with algorithms incorporating OCTA and color fundus images for enhanced classification [29], as well as techniques using enface OCT images or infrared optical coherence tomography oximetry [30]. Recent methods align OCTA and fundus images via topological message-passing to classify arterioles and venules using clustering techniques [31]. FAZ segmentation has evolved from traditional methods, such as level-set approaches [32] and two-stage image processing [33], which struggled with feature extraction. Modern deep learning frameworks address these limitations with architectures incorporating boundary regression, distance-aware learning, and encoder-decoder models for boundary alignment. Despite advancements, challenges remain, such as blurred FAZ boundaries due to unclear retinal vessel edges [10]. Various approaches have been explored to address the challenges in OCTA segmentation, including high data dependency, poor generalization, and overfitting. Datasets such as OCTA-500 [15] and methods like the 3D-to-2D Image Projection Network (IPN) and its iterations [34] have demonstrated improvements in FAZ segmentation accuracy. Despite these advancements, segmentation performance and robustness remain critical bottlenecks in clinical applications. This work proposes a novel framework that integrates a DenseNet Encoder with MSPFM to overcome these limitations. Our method achieves superior segmentation accuracy and robustness by enhancing feature extraction and allowing dynamic fusion across scales. The proposed approach aims to improve clinical and research applications in ophthalmology. The novel **contributions** of this paper are as follows:

- Dense-PMSFNet: Novel architecture for accurate retinal vasculature and FAZ segmentation in OCTA images.
- Enhances the diagnostic accuracy of DR, which helps in early detection and monitoring.
- Introduces the DenseNet encoder with MSPFM for multi-scale feature extraction.
- Provides a scalable, robust solution for clinical use in medical imaging advancements.

III. METHODOLOGY

This section overviews our proposed model architecture, detailing the DenseNet encoder, the MSPFM module, the decoder, and the deep fusion module. Fig. 1 illustrates the

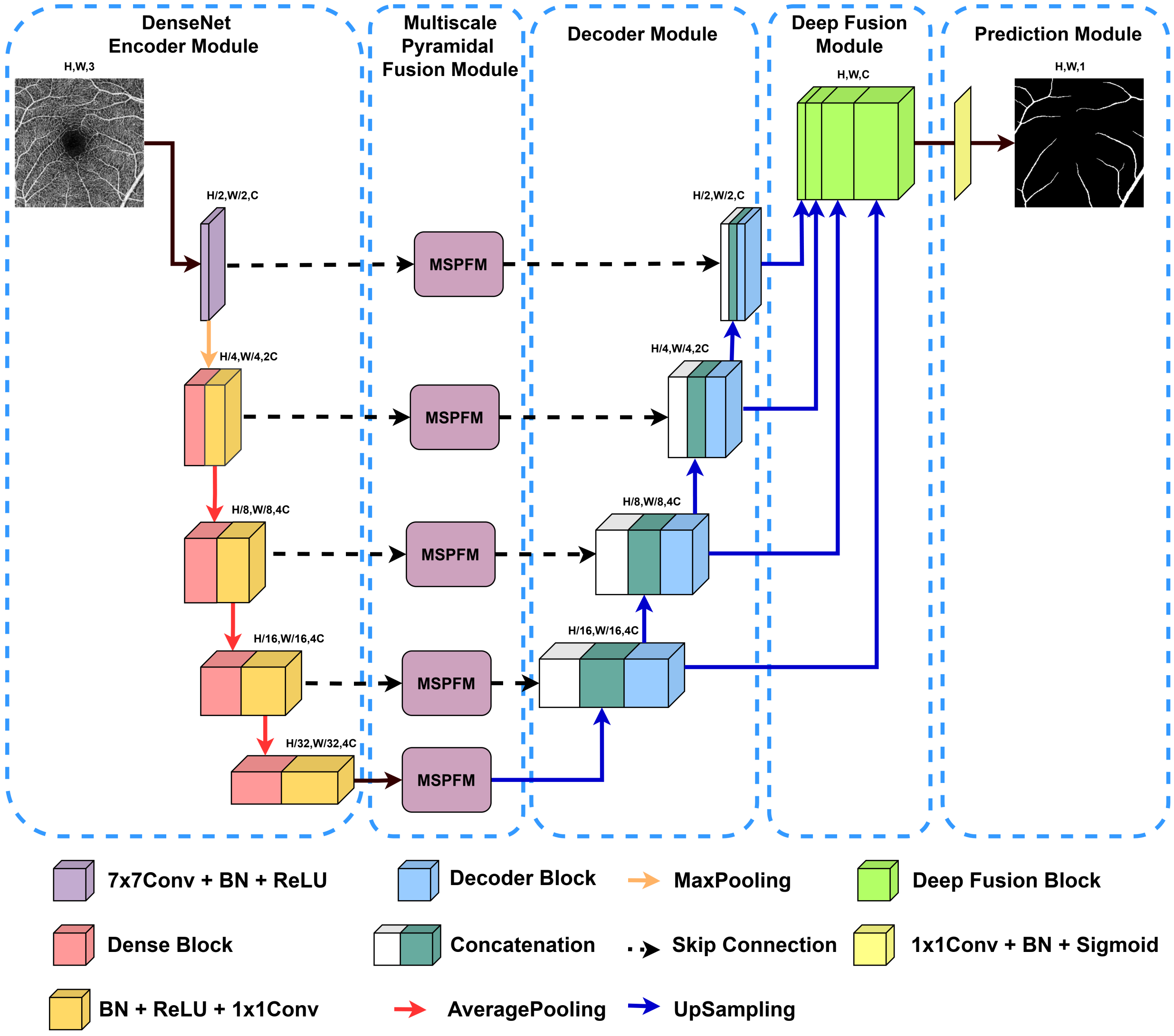


Fig. 1. The proposed model-Dense-PMSFNet Framework consists of a DenseNet121 encoder, a Multi-Scale Pyramidal Fusion Module (MSPFM) for feature extraction, and a decoder with scSE modules to improve feature representation and segmentation performance.

proposed model, featuring a DenseNet121 encoder for accurate medical image segmentation [35]. The MSPFM Module integrates multi-scale outputs from each encoder layer, enhancing feature extraction by combining local spatial details with broader patterns. These features are passed to the decoder, where MSPFM outputs are concatenated at corresponding layers. The decoder uses convolutional blocks with spatial and channel Squeeze-and-Excitation (scSE) [36] blocks for feature recalibration, emphasizing relevant features while suppressing irrelevant ones. Upsampling layers refine segmentation boundaries, and a deep fusion module with auxiliary segmentation heads at the last three decoder levels extracts hierarchical features, improving segmentation accuracy.

A. DenseNet Encoder

The proposed encoder architecture consists of a 7×7 convolutional receptive layer at the initial level ($k = 0$) and four dense blocks (f_{dense}^k where $k = 1, 2, 3, 4$) in the subsequent levels. We use an ImageNet [37] pre-trained DenseNet-121 [35] as the encoder, and the layers $f_{dense}^k, k = 2, 3, 4$ are finetuned on the OCTA 500 dataset. The architecture of DenseNet-121 comprises a 7×7 convolutional receptive field followed by a pooling module and dense blocks, each containing multiple layers connected densely. To maximize feature reuse and reduce the number of parameters, each layer takes input from all previous layers and transmits its output to every subsequent layer.

The operation within a dense block can be described as follows:

$$y_k = G_k([y_0, y_1, \ldots, y_{k-1}]) \tag{1}$$

where y_k represents the output of the k-th layer, G_k is a composite function of operations such as Batch Normalization (BN), Rectified Linear Unit (ReLU), and Convolution (Conv), and $[y_0, y_1, \ldots, y_{k-1}]$ denotes the concatenation of the feature maps generated in previous layers. Combining the strength of pre-training and fine-tuning of the densenet blocks, the encoder extracts local features effectively, leveraging DenseNet's hierarchical feature extraction capabilities.

B. Multi-Scale Pyramidal Fusion Module

Fig. 2 represents MSPFM used at each layer. MSPFM enhances feature extraction from DenseNet, capturing diverse and informative feature information. The module involves several parallel processes and fusion operations to combine the features effectively. As shown in Fig. 2, for the DenseNet feature f^k_{dense}, the pyramidal pooling module [38] is applied with average pooling (AvgPool) at scales 2×2, 4×4 and 6×6. Each pooled feature map is then passed through a 3×3 convolution (Conv) and upsampled (Up) to match the input feature map size. The Up features are then concatenated. The operations are defined as follows:

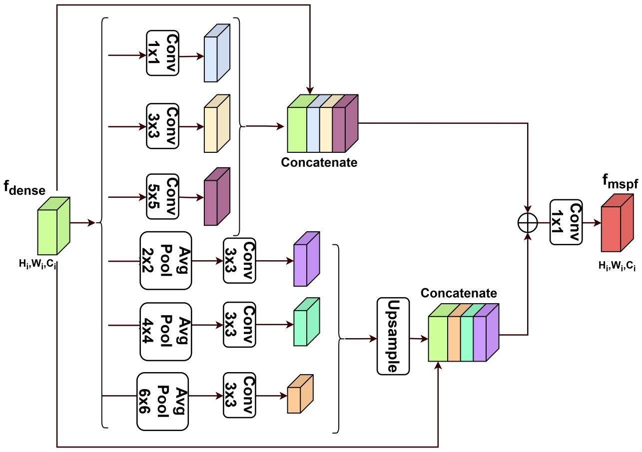


Fig. 2. Multi-Scale Pyramidal Fusion Module (MSPFM) – This module enhances feature extraction by integrating multi-scale contextual information using pooling, convolution, and upsampling operations

$$f_{pool_i} = Up_{2i \times 2i}\left(Conv_{3\times3}\left(AvgPool_{2i\times2i}(f^k_{dense})\right)\right), \quad i = 1, 2, 3 \tag{2}$$

here, f_{pool_i} represents the upsampled features obtained after applying average pooling at scale $2i \times 2i$, followed by a 3×3 convolution and upsampling.

In parallel, the DenseNet features (f^k_{dense}) are also processed through 1×1, 3×3 and 5×5 convolutions. These convolved features are then concatenated:

$$f_{conv} = Concat(Conv_{1\times1}(f^k_{dense}), Conv_{3\times3}(f^k_{dense}), Conv_{5\times5}(f^k_{dense})) \tag{3}$$

The concatenated convolution features f_{conv} are then added to the concatenated outputs of the pyramidal pooling module, producing $\hat{f}^k_{dense}$, as shown below:

$$\hat{f}^k_{dense} = f_{conv} + Concat(f_{pool_1}, f_{pool_2}, f_{pool_3}) \tag{4}$$

here, $\hat{f}^k_{dense}$ represents the fused features, combining local contextual features (f_{conv}) with multi-scale pooled features (f_{pool_i}).

Finally, the multi-scale pyramidal fusion features (f^k_{mspf}) are obtained by applying a 1×1 convolution to the concatenated outputs of $\hat{f}^k_{dense}$:

$$f^k_{mspf} = Conv_{1\times1}\left(Concat(\hat{f}^k_{dense})\right) \tag{5}$$

C. Decoder

Fig. 3 represents the decoder block used in the model. The decoder upsamples MSPFM output to regain spatial resolution, concatenates it with MSPFM features, and processes them through two 3×3 convolutional layers, each augmented with the scSE module [36]. This adaptively recalibrates features, enhancing spatial and channel-wise information integration and enabling the network to focus on the most informative features.

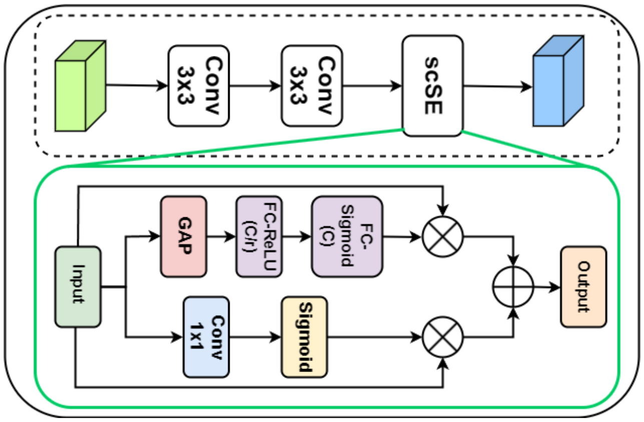


Fig. 3. Decoder Block – The decoder refines segmentation by upsampling MSPFM outputs and recalibrating spatial and channel-wise information through scSE blocks

$$s_c = \sigma(W_2 \cdot \delta(W_1 \cdot GAP(Y)))Y_{cSE} = s_c \cdot Y \tag{6}$$

$$s_s = \sigma\left(Conv_{1\times1}(Y)\right) Y_{sSE} = s_s \cdot Y \tag{7}$$

$$Y_{scSE} = Y_{cSE} + Y_{sSE} \tag{8}$$

The input feature map is denoted by $Y \in \mathbb{R}^{H\times W\times C}$, the Global Average Pooling (GAP) operation computes the spatial average for each channel, reducing Y to a vector of size $\mathbb{R}^{1\times1\times C}$. The fully connected layers' weights are represented by $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $W_2 \in \mathbb{R}^{C\times\frac{C}{r}}$, the reduction ratio is denoted by r, the ReLU activation function by δ, and the sigmoid activation function by σ. After the scSE block, the output features are combined with the upsampled features from the other layers of the decoder, forming the deep fusion module. These combined features are then concatenated and passed through the final stage of the decoder, which involves a 1×1 convolutional layer and a sigmoid activation function, producing the segmentation output with pixel-wise probabilities. Mathematically, this can be represented as:

$$f_{deepfusion} = \text{Concat}\left(\text{Up}_{2i\times2i} f^k_{decoder})\right), \text{where } k = 0, 1, 2, 3 \text{ and } i = 1, 2, 3, 4. \tag{9}$$

$$f_{\text{final}} = \sigma\left(\text{Conv}_{1\times1}\left(\left(f_{\text{deepfusion}}\right)\right)\right), \tag{10}$$

where f_{final} represents the final segmentation output and σ denotes the sigmoid activation function.

IV. EXPERIMENTS AND ANALYSIS

In this section, we introduce the input dataset utilized for training, provide additional experimental details, and present visual and quantitative comparisons of our proposed network.

A. Dataset Description

The proposed technique is validated using the publicly available OCTA-500 dataset [15], which includes 500 subjects classified into two main subsets: OCTA 6mm and OCTA 3mm, based on their Field of View (FOV). The OCTA 6mm subset comprises 300 subjects (No.10001-10300) with grayscale images of resolution 400×400, while the OCTA 3mm subset includes 200 subjects (No. 10301-10500) with grayscale images of resolution 304×304. All images in the dataset are stored in the .bmp format, and for each image in the OCTA-500 3mm and 6mm datasets, there are corresponding ground truth segmentation masks separately for vein, artery, capillary, and FAZ. Acquired with a 70 kHz spectral-domain OCT system (RTVue-XR, Optovue, CA), the dataset covers diseases like retinal vein occlusion, choroidal neovascularization, age-related macular degeneration, and DR.

B. Experimental Settings

The proposed model was implemented using TensorFlow's Keras backend and trained on an NVIDIA GeForce RTX 3090 24 GB GPU with 128 GB RAM. Experiments used the OCTA-500 database: 240/10/50 images for training/validation/testing from the 6mm dataset, and 140/10/50 from the 3mm dataset, sequentially split following the same approach described in [15]:

- **OCTA_6mm**: The training set consists of 240 images from NO. 10001 to NO. 10240, the validation set includes 10 images from NO. 10241 to NO. 10250, and the test set includes 50 images from NO. 10251 to NO. 10300.
- **OCTA_3mm**: The training set consists of 140 images from NO. 10301 to NO. 10440, the validation set includes 10 images from NO. 10441 to NO. 10450, and the test set includes 50 images from NO. 10451 to NO. 10500.

The training set is used to train the network models, the validation set is used to select the best model, and the test set is used for evaluation. In all experiments, the same images are used across different segmentation tasks (capillary, artery, vein, and FAZ) for each respective subset. The original size of the images is used without resizing to ensure the preservation of image resolution and microvasculature details. Data augmentation included horizontal flipping, shearing (up to 0.15), and 20° rotation was applied to expand the training set three times, increasing the 3mm dataset from 140 to 420 images and the 6mm dataset from 240 to 720 images. Horizontal flipping addresses spatial biases by mirroring images and their masks, helping the model learn orientation-invariant features relevant to microvasculature patterns. Shearing introduces minor distortions to mimic the irregularities of shapes and misalignments in clinical settings while keeping images and masks aligned. A rotation of 20° improves the variability of microvasculature orientation and helps improve the model's feature detection under varied conditions. The Adam optimizer was employed with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$, and a learning rate of 10^{-3}. An equally weighted combination of Dice loss and Focal loss was used as a loss function. The training was conducted with a batch size of 2 for a maximum of 400 epochs, with an early stopping criterion applied with patience of 20 epochs.

C. Evaluation Metrics

The performance of our network is assessed using five metrics: Dice coefficient (Dice), Intersection Over Union (IoU), Accuracy (Acc), Specificity (SP), and Sensitivity (SE). The Dice [39] measures the similarity between segmentation and ground truth using

$$Dice = \frac{2 \cdot |S \cap G|}{|S| + |G|}, \tag{11}$$

while IoU evaluates overlap using

$$IoU = \frac{|S \cap G|}{|S \cup G|}. \tag{12}$$

where S and G denote the segmentation result and ground truth, respectively. Both metrics range from 0 to 1, with higher values indicating better performance. The positives refer to the pixels that belong to the target classes, such as retinal blood vessels and the FAZ, while the negatives refer to the background pixels (non-target areas). The sensitivity (recall) refers to the ability of the model to identify positive instances correctly. At the same time, the specificity signifies its capacity to predict the true negatives for each class. Accuracy measures the overall performance of the model by indicating the proportion of correctly classified instances.

V. RESULTS

A. Quantitative Analysis

To evaluate the proposed network comprehensively, we compared its performance against twelve state-of-the-art models: U-Net [19], Attention U-Net [42], UNet++ [40], UNet 3+ [41], CS-Net [43], AV-Net [30], TransUNet [44], and UNETR [45], MS-Net [22], PCAM-PRDC-Net [24], MCFSA-Net [23], FPN-MSPF-Net [21]. For publicly available studies, experiments were reproduced with similar parameters. Segmentation performance for capillaries, arteries, veins, and the FAZ was evaluated using the test dataset consisting of 50 images each from the OCTA 6mm (NO. 10251 to NO. 10300) and OCTA 3mm (NO. 10451 to NO. 10500) datasets. Metrics such as Dice, IoU, Acc, SP, and SE were computed. Table I presents the comparison results of capillary, artery, vein, and FAZ segmentation with other state-of-the-art methods.

Capillary Segmentation: The proposed Dense-PMSFNet achieved the highest Dice of 87.51% (OCTA 6mm) and 90.88% (OCTA 3mm), with IoU values of 77.79% and 82.62%, respectively, surpassing all other models indicating its ability in capturing fine vascular structures.

Artery Segmentation: The proposed method outperformed other models with a Dice of 88.16% for OCTA 6mm and 90.21% for OCTA 3mm. IoU values were 78.84% and 82.16%

TABLE I
COMPARISON RESULTS OF CAPILLARY, ARTERY, VEIN, AND FAZ SEGMENTATION WITH OTHER STATE-OF-ART METHODS

Results of Capillary segmentation										
	6 mm					3 mm				
Publications	Dice	IoU	Acc	SE	SP	Dice	IoU	Acc	SE	SP
U-Net [19]	84.61	73.33	88.08	84.89	90.05	86.98	76.96	88.35	90.64	86.79
UNet++ [40]	87.34	77.52	90.08	89.17	90.61	88.45	79.29	90.49	89.84	90.97
UNet 3+ [41]	87.11	77.16	89.96	88.41	90.87	87.55	77.85	89.61	90.14	89.32
Attention U-Net [42]	87.47	77.59	90.29	89.10	90.97	89.65	81.24	91.44	91.52	91.41
CS-Net [43]	87.16	76.90	**90.63**	89.27	90.14	90.04	81.57	91.67	90.99	90.74
AV-Net [30]	87.44	77.68	90.32	89.04	90.86	89.53	81.05	91.34	91.50	91.17
TransUNet [44]	83.31	71.39	86.98	83.58	89.88	86.51	76.22	88.01	89.92	87.78
UNETR [45]	84.63	73.35	88.11	84.87	90.08	87.22	77.33	88.46	90.75	88.89
MS-Net [22]	85.95	75.36	89.48	89.33	89.57	88.20	78.89	90.13	89.38	90.66
PCAM-PRDC-Net [24]	86.12	75.63	89.35	88.56	99.14	88.11	78.75	90.42	**92.90**	88.58
MCFSA-Net [23]	86.37	76.01	89.53	90.25	89.09	88.84	79.92	90.55	92.19	89.33
FPN-MSPF-Net [21]	85.78	74.21	89.62	85.79	90.02	86.72	76.77	88.29	89.34	87.91
Proposed work	87.51	**77.79**	90.43	**89.52**	**91.74**	**90.88**	**82.62**	**91.84**	91.79	**91.93**
Results of Artery Segmentation										
U-Net [19]	85.64	74.89	98.74	85.10	99.38	87.07	77.11	99.04	86.96	99.50
UNet++ [40]	84.52	73.19	98.65	83.11	99.37	87.45	77.71	99.07	87.97	99.49
UNet 3+ [41]	83.88	72.24	98.58	83.07	99.31	86.75	76.60	99.03	85.71	99.54
Attention U-Net [42]	86.49	76.20	98.83	83.84	99.54	88.57	79.50	99.15	88.50	99.56
CS-Net [43]	79.89	67.22	98.33	78.54	99.24	82.12	69.98	98.73	79.76	99.46
AV-Net [30]	82.72	70.53	98.49	81.31	99.30	85.44	74.58	98.95	83.18	99.56
TransUNet [44]	84.67	73.41	98.51	83.27	99.27	85.52	74.70	98.89	83.21	99.51
UNETR [45]	85.78	75.10	98.57	83.25	99.41	86.71	76.53	98.98	85.64	99.49
MS-Net [22]	85.84	75.19	98.82	83.61	99.53	88.51	79.40	99.19	89.44	99.55
PCAM-PRDC-Net [24]	86.20	75.74	98.84	85.02	99.48	85.21	74.23	99.16	89.75	99.52
MCFSA-Net [23]	85.18	74.19	98.71	84.19	99.39	86.01	75.46	99.02	88.74	99.41
FPN-MSPF-Net [21]	84.81	74.63	98.69	83.19	99.41	85.11	74.34	98.92	87.58	99.42
Proposed work	**88.16**	**78.84**	**98.95**	**87.84**	**99.56**	**90.21**	**82.16**	**99.27**	**90.70**	**99.72**
Results of Vein Segmentation										
U-Net [19]	85.10	74.07	98.54	86.75	99.14	84.82	73.65	99.12	83.52	99.59
UNet++ [40]	84.24	73.79	98.57	84.14	99.29	86.35	75.98	99.21	84.22	99.66
UNet 3+ [41]	83.48	71.65	98.43	83.12	99.19	85.35	74.44	99.14	84.07	99.60
Attention U-Net [42]	86.43	76.11	98.72	85.08	**99.41**	85.68	74.95	99.16	84.77	99.59
CS-Net [43]	80.48	67.78	98.03	82.44	98.84	80.91	68.21	98.93	78.31	99.56
AV-Net [30]	83.57	71.78	98.47	81.36	99.33	70.62	54.58	98.24	71.88	99.04
TransUNet [44]	83.78	72.08	98.45	83.44	99.16	84.27	72.81	99.12	83.97	99.53
UNETR [45]	84.21	72.72	98.53	84.02	99.21	85.21	74.23	99.17	84.01	99.56
MS-Net [22]	84.99	73.92	98.67	86.05	99.28	87.48	77.75	99.26	**90.74**	99.52
PCAM-PRDC-Net [24]	85.34	74.45	98.65	**88.56**	99.14	87.03	77.05	99.27	88.47	99.60
MCFSA-Net [23]	84.61	73.35	98.58	84.87	99.24	84.06	72.50	99.09	86.67	99.47
FPN-MSPF-Net [21]	84.28	72.77	98.58	84.57	99.21	84.32	72.87	99.09	84.63	99.58
Proposed work	**87.38**	**77.58**	**98.83**	87.51	99.37	**89.23**	**80.55**	**99.38**	88.12	**99.71**
Results of FAZ Segmentation										
U-Net [19]	88.57	79.49	99.74	87.58	99.88	96.14	92.57	99.67	96.07	99.82
UNet++ [40]	81.31	68.51	99.60	77.79	99.85	96.31	92.88	99.63	96.37	99.77
UNet 3+ [41]	87.57	77.89	99.73	83.05	99.93	96.33	92.93	99.69	96.17	99.84
Attention U-Net [42]	88.67	79.65	99.74	87.20	99.90	97.43	94.98	99.79	96.43	99.93
CS-Net [43]	86.16	78.03	99.66	85.83	99.87	95.48	91.74	99.65	95.33	99.84
AV-Net [30]	86.13	75.64	99.68	87.83	99.82	95.75	91.85	99.60	95.48	99.79
TransUNet [44]	88.69	79.67	99.71	87.63	99.87	95.56	91.49	99.61	95.37	99.81
UNETR [45]	89.71	81.34	99.78	88.74	99.89	96.44	93.12	99.65	96.36	99.84
MS-Net [22]	90.25	82.24	99.81	86.79	99.96	96.93	94.05	99.77	95.72	99.95
PCAM-PRDC-Net [24]	84.09	72.59	99.67	86.69	99.82	94.36	89.32	99.61	93.36	99.88
MCFSA-Net [23]	90.87	83.32	99.82	89.02	99.94	95.57	91.51	99.69	93.99	99.94
FPN-MSPF-Net [21]	88.12	79.17	99.65	87.51	99.92	95.22	91.30	99.62	95.24	99.87
Proposed work	**94.84**	**90.18**	**99.87**	**94.67**	**99.94**	**97.92**	**95.92**	**99.83**	**97.30**	**99.94**

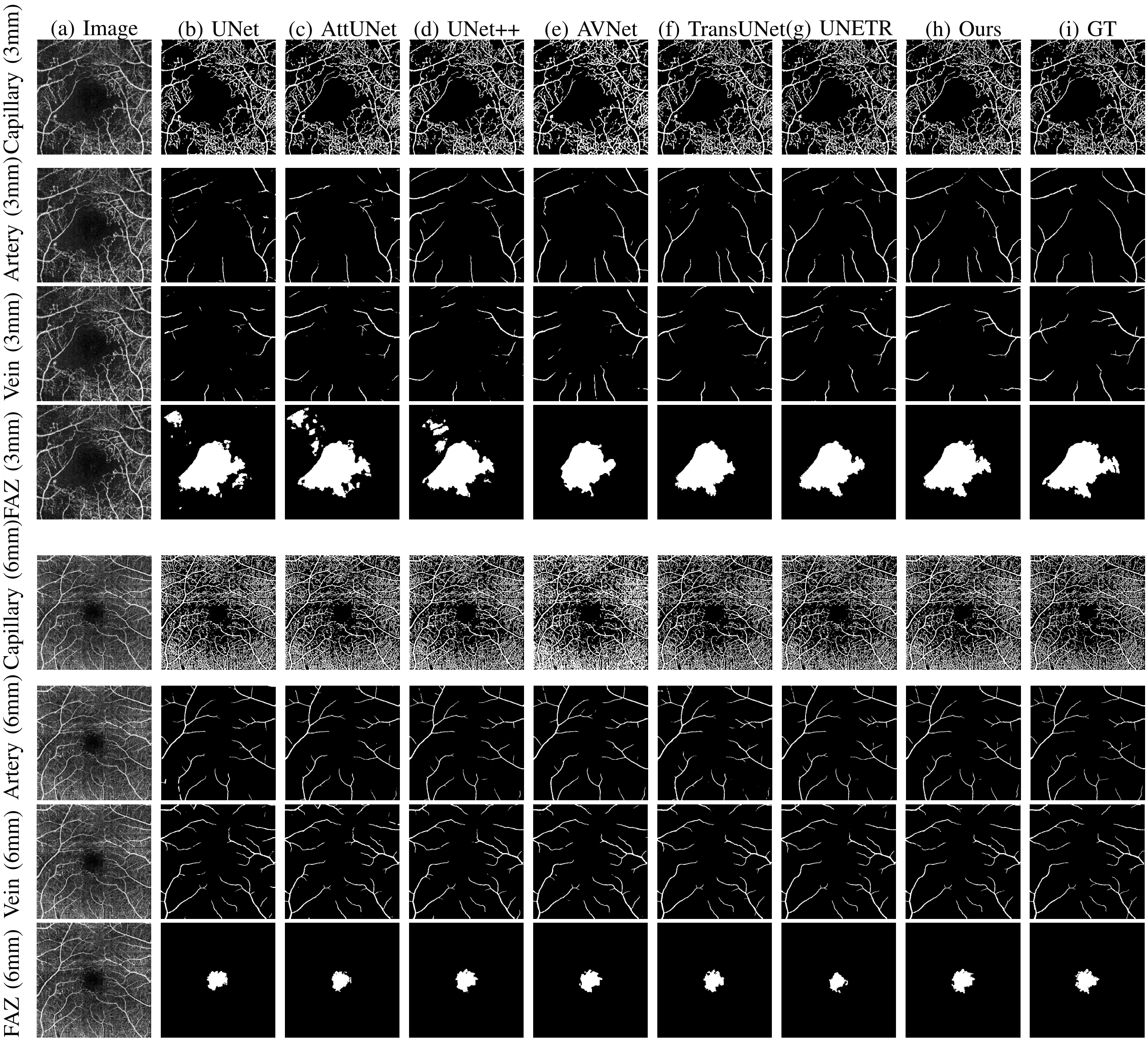


Fig. 4. Segmentation Performance Comparison – This figure shows the segmentation results of Dense-PMSFNet alongside other models for capillaries, arteries, veins, and FAZ in 3mm and 6mm OCTA images. Ground truth labels are also included for comparison.

for OCTA 6mm and OCTA 3mm, respectively, demonstrating the model's ability to delineate arterial structures with precision.

Vein Segmentation: The Dense-PMSFNet achieved Dice of 87.38% for OCTA 6mm and 89.23% for OCTA 3mm. The IoU values were 77.58% for OCTA 6mm and 80.55% for OCTA 3mm. While there was a slight trade-off in SE and SP for vein segmentation on the OCTA 6mm and 3mm datasets compared to Attention U-Net [42], MS-Net [22], PCAM-PRDC-Net [24] overall performance remained superior, especially in terms of Dice, IoU and accuracy crucial for detecting subtle venous changes associated with DR.

FAZ Segmentation: The proposed model achieved outstanding Dice of 94.84% (OCTA 6mm) and 97.92% (OCTA 3mm), with IoU values of 90.18% and 95.92%. While PCAM-PRDC-Net [24] demonstrated slightly better SE, the proposed model excels in overall performance. These results are particularly significant for diagnosing DR, as changes in the FAZ's size and shape indicate disease progression.

Dense-PMSFNet consistently outperformed these models on key metrics, including the Dice and IoU, demonstrating superior accuracy in segmenting fine details and preserving anatomical structures, which is essential for microvasculature segmentation.

B. Qualitative Analysis

Segmentation results from one of the 50 test images, each from the test datasets (Image No. 10317 from the 3mm dataset and Image No. 10086 from the 6mm dataset) are presented in

TABLE II
ABLATION STUDY OF CAPILLARY, ARTERY, VEIN, AND FAZ SEGMENTATION

	6 mm					3 mm				
Methods	Dice	IoU	Accuracy	SE	SP	Dice	IoU	Acc	SE	SP
Results of ablation study on capillary segmentation										
DenseNet121 (Baseline)	86.68	76.50	90.15	89.48	90.54	89.38	80.80	91.09	91.14	91.04
DenseNet121 + MSPFM	86.84	76.74	90.28	89.47	90.62	89.77	81.43	91.22	91.42	91.05
DenseNet121 MSPFM + Deep Fusion	86.91	76.85	90.31	89.49	90.64	89.82	81.52	91.25	91.49	91.04
DenseNet121 + MSPFM + scSE	87.39	77.60	90.39	89.50	90.69	89.92	81.68	91.23	91.67	90.98
Proposed work	**87.51**	**77.79**	**90.43**	**89.52**	**90.74**	**90.88**	**82.62**	**91.84**	**91.79**	**91.93**
Results of ablation study on artery segmentation										
DenseNet121 (Baseline)	87.13	77.20	98.86	86.48	99.44	88.95	80.10	99.20	90.42	99.53
DenseNet121 + MSPFM	87.27	77.11	98.85	87.57	99.37	89.53	81.04	99.25	90.66	99.61
DenseNet121 + MSPFM + Deep Fusion	87.32	77.49	98.87	87.71	99.39	89.65	81.24	99.24	**90.75**	99.56
DenseNet121 + MSPFM + scSE	87.55	77.85	98.89	87.68	99.48	89.87	81.61	99.26	90.49	99.71
Proposed work	**88.16**	**78.84**	**98.95**	**87.84**	**99.56**	**90.21**	**82.16**	**99.27**	90.70	**99.72**
Results of ablation study on vein segmentation										
DenseNet121 (Baseline)	86.50	76.22	98.73	87.39	99.29	87.01	77.00	98.99	86.87	99.51
DenseNet121 + MSPFM	86.82	76.71	98.77	87.44	99.33	87.59	77.91	99.09	86.98	99.57
DenseNet121 + MSPFM + Deep Fusion	86.88	76.80	98.74	87.45	99.29	88.06	78.66	99.17	87.37	99.61
DenseNet121 + MSPFM + scSE	86.86	76.77	98.69	87.43	99.31	89.02	80.22	99.37	87.98	99.71
Proposed work	**87.38**	**77.58**	**98.83**	**87.51**	**99.37**	**89.23**	**80.55**	**99.38**	**88.12**	**99.71**
Results of ablation study on FAZ segmentation										
DenseNet121 (Baseline)	91.57	84.62	99.58	91.34	99.90	96.98	94.13	99.73	96.83	99.91
DenseNet121 + MSPFM	93.86	88.43	99.79	93.52	99.92	97.57	95.25	99.79	97.12	99.93
DenseNet121 + MSPFM + Deep Fusion	94.31	89.23	99.84	94.12	99.94	97.71	95.52	99.81	97.25	99.93
DenseNet121 + MSPFM + scSE	94.59	89.73	99.86	94.43	99.93	97.79	95.67	99.80	97.27	99.94
Proposed work	**94.84**	**90.18**	**99.87**	**94.67**	**99.95**	**97.92**	**95.92**	**99.83**	**97.30**	**99.94**

Fig. 4. These results compare Dense-PMSFNet with U-Net, Attention U-Net, UNet++, AV-Net, TransUNet, and UNETR. Our method excels in capturing fine anatomical details, especially in FAZ segmentation, closely resembling the ground truth and outperforming other models. Accurate FAZ segmentation is pivotal for diagnosing DR, as changes in its size and shape are key indicators of disease progression. Additionally, Dense-PMSFNet demonstrates superior segmentation of arteries, veins, and capillaries, critical for identifying microvascular complications like arterial narrowing and venous dilation in DR. Precise vascular delineation aids in understanding disease-related changes, enabling targeted treatment and improved disease management. The results reflect Dense-PMSFNet's potential to enhance clinical diagnosis and treatment planning, contributing significantly to ophthalmology.

VI. DISCUSSIONS

A. Ablation Study

An ablation study was conducted to evaluate the contributions of individual components of Dense-PMSFNet. Using DenseNet121 as the baseline, we assessed the impact of integrating the MSPFM, scSE Decoder, and Deep Fusion. Table II presents the ablation study results for capillary, artery, vein, and FAZ segmentation. DenseNet121 alone achieved moderate accuracy, but integrating MSPFM significantly improved segmentation performance across capillary, artery, vein, and FAZ cases. The combination of DenseNet121, MSPFM, scSE, and Deep Fusion yielded the best results, achieving Dice of 94.84% (OCTA 6mm) and 97.92% (OCTA 3mm) and IoU of 90.18% and 95.92% for FAZ segmentation. While DenseNet121 + MSPFM occasionally outperformed the proposed model in specific scenarios, the proposed architecture demonstrated superior overall effectiveness. This trend extended to other segmentation tasks, emphasizing the architecture's capability to improve retinal vasculature and FAZ segmentation. The consistent results across OCTA 6mm and 3mm subsets underscore the architecture's capability to improve retinal vasculature and FAZ segmentation, validating the robustness of our approach and its potential for advancing retinal image analysis.

B. Advantage of the Proposed Model

The Dense-PMSFNet model has several advantages over other state-of-the-art methods. It can precisely segment fine vascular structures, such as capillaries, arteries, veins, and the FAZ, showing better anatomical detail preservation. This is critical for microvasculature segmentation, where small errors may greatly affect clinical outcomes. Our proposed method incorporates several key components that set it apart from existing approaches:

- The DenseNet Encoder allows efficient feature extraction through dense connectivity, optimizing gradient flow and feature utilization.
- MSPFM effectively captures multi-scale spatial features to achieve a highly accurate segmentation of fine anatomical structures such as capillaries and the FAZ.
- The Deep Fusion module ensures robust contextual understanding by fusing multi-resolution features and further enhancing the delineation of complex vascular patterns.

The model's exceptional performance in metrics such as Dice and IoU across all tasks demonstrates its potential to improve clinical diagnosis and treatment planning for DR. The quantitative analysis, presented in Table I, underscores the exceptional performance of the proposed model. The model achieved the highest Dice and IoU values across all segmentation tasks when evaluated against twelve state-of-the-art models, including U-Net [19], Attention U-Net [42], and TransUNet [44]. These results, achieved on both the OCTA 6mm and OCTA 3mm datasets, emphasize its effectiveness in segmenting capillaries, arteries, veins, and the FAZ—structures critical for retinal image analysis.

C. Limitations and Areas for Development

Despite the strengths of Dense-PMSFNet, several limitations must be addressed to enhance its applicability in real-world scenarios. The computational complexity of the model, combined with an average inference time of 116 ms per image of size 400×400, remains a challenge, particularly for deployment in resource-constrained environments. Future efforts should aim to optimize the architecture for better computational efficiency with minimal compromise on accuracy. This could include model pruning, quantization, or other approaches that reduce resource requirements. The quality and size of the dataset significantly influence the model's performance. Imbalanced or insufficient data may hinder the generalization capability, and variability in imaging conditions such as resolution, noise levels, and modality differences can also affect segmentation accuracy. To address these issues, future work could explore advanced data augmentation techniques, such as geometric transformations, synthetic data generation, or domain-specific augmentations, that will increase the model's robustness and generalizability. Domain adaptation techniques could be incorporated to allow the model to adapt to different datasets with varying imaging characteristics and improve its applicability across a broader range of scenarios. In addition, further evaluation of larger, more diverse datasets will be essential to validate the model's generalization capabilities. Introducing Dense-PMSFNet into real-time diagnostics is another research direction that would make it more practical. The interpretability of the model can be enhanced through explainable AI techniques, which would facilitate its clinical adoption. By providing transparency in decision-making, explainable AI could help clinicians trust the model's predictions and help to integrate it into healthcare workflows, leading to improved disease diagnosis and treatment planning outcomes.

VII. CONCLUSION AND FUTURE WORKS

This paper presents the Dense-PMSFNet, an advanced framework for segmenting OCTA images, which aids in DR diagnosis. It incorporates a DenseNet Encoder with a MSPFM for robust feature extraction and dynamic fusion, addressing data dependency, generalization, and overfitting challenges. A Deep Fusion module further enhances segmentation by optimizing information integration across multiple scale decoder outputs. Extensive experiments and ablation studies validate its superior performance over state-of-the-art models on four tasks (capillary, artery, vein, and FAZ segmentation) on OCTA images with superior Dice and IoU scores. By automating OCTA image segmentation, Dense-PMSFNet reduces manual workload, accelerates diagnoses, and supports precise treatment planning. Dense-PMSFNet demonstrates significant potential for integration into clinical workflows, advancing early DR interventions, and improving patient outcomes. Future research can optimize architectures, improve domain generalization, enhance decision-making interpretability, and develop data augmentation strategies for better performance with limited data.

DECLARATIONS

Ethical approval: This article does not contain any studies with human participants or animals.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could influence the results reported in this paper.

DATA AVAILABILITY STATEMENT

The study's datasets are publicly available and can be accessed at the mentioned URL: https://ieee-dataport.org/open-access/octa-500. The source of this dataset has been referenced in our manuscript.

REFERENCES

[1] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: A survey," *IEEE Access*, vol. 10, pp. 28 642–28 655, 2022.

[2] S. Lin, P. Ramulu, E. L. Lamoureux, and C. Sabanayagam, "Addressing risk factors, screening, and preventative treatment for diabetic retinopathy in developing countries: a review," *Clinical & experimental ophthalmology*, vol. 44, no. 4, pp. 300–320, 2016.

[3] R. F. Spaide, J. G. Fujimoto, N. K. Waheed, S. R. Sadda, and G. Staurenghi, "Optical coherence tomography angiography," *Progress in retinal and eye research*, vol. 64, pp. 1–55, 2018.

[4] Z. Sun, D. Yang, Z. Tang, D. S. Ng, and C. Y. Cheung, "Optical coherence tomography angiography in diabetic retinopathy: an updated review," *Eye*, vol. 35, no. 1, pp. 149–161, 2021.

[5] M. M. Abdelsalam, "Effective blood vessels reconstruction methodology for early detection and classification of diabetic retinopathy using octa images by artificial neural network," *Informatics in Medicine Unlocked*, vol. 20, p. 100390, 2020.

[6] M. K. Ikram, Y. T. Ong, C. Y. Cheung, and T. Y. Wong, "Retinal vascular caliber measurements: clinical significance, current knowledge and future perspectives," *Ophthalmologica*, vol. 229, no. 3, pp. 125–136, 2013.

[7] C. Y. Cheung, M. K. Ikram, R. Klein, and T. Y. Wong, "The clinical implications of recent studies on the structure and function of the retinal microvasculature in diabetes," *Diabetologia*, vol. 58, pp. 871–885, 2015.

[8] C. Y.-l. Cheung, C. Sabanayagam, A. K.-p. Law, N. Kumari, D. S.-w. Ting, G. Tan, P. Mitchell, C. Y. Cheng, and T. Y. Wong, "Retinal vascular geometry and 6 year incidence and progression of diabetic retinopathy," *Diabetologia*, vol. 60, pp. 1770–1781, 2017.

[9] Y. Jia, S. T. Bailey, T. S. Hwang, S. M. McClintic, S. S. Gao, M. E. Pennesi, C. J. Flaxel, A. K. Lauer, D. J. Wilson, J. Hornegger *et al.*, "Quantitative optical coherence tomography angiography of vascular abnormalities in the living human eye," *Proceedings of the National Academy of Sciences*, vol. 112, no. 18, pp. E2395–E2402, 2015.

[10] M. Guo, M. Zhao, A. M. Cheong, F. Corvi, X. Chen, S. Chen, Y. Zhou, and A. K. Lam, "Can deep learning improve the automatic segmentation of deep foveal avascular zone in optical coherence tomography angiography?" *Biomedical Signal Processing and Control*, vol. 66, p. 102456, 2021.

[11] Y. Liu, L. Zuo, Y. He, S. Han, J. Lei, J. L. Prince, and A. Carass, "Octa segmentation with limited training data using disentangled representation learning," in *Deep Learning for Medical Image Analysis*. Elsevier, 2024, pp. 451–469.

[12] Y. Xu, X. Xu, L. Jin, S. Gao, R. S. M. Goh, D. S. Ting, and Y. Liu, "Partially-supervised learning for vessel segmentation in ocular images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021*. Springer, 2021, pp. 271–281.

[13] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "Scribblesup: Scribble-supervised convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3159–3167.

[14] A. Chinkamol, V. Kanjaras, P. Sawangjai, Y. Zhao, T. Sudhawiyangkul, C. Chantrapornchai, C. Guan, and T. Wilaiprasitporn, "Octave: 2d en face optical coherence tomography angiography vessel segmentation in weakly-supervised learning with locality augmentation," *IEEE Transactions on Biomedical Engineering*, 2022.

[15] M. Li, K. Huang, Q. Xu, J. Yang, Y. Zhang, Z. Ji, K. Xie, S. Yuan, Q. Liu, and Q. Chen, "Octa-500: a retinal dataset for optical coherence tomography angiography study," *Medical Image Analysis*, vol. 93, p. 103092, 2024.

[16] R. Mirshahi, P. Anvari, H. Riazi-Esfahani, M. Sardarinia, M. Naseripour, and K. G. Falavarjani, "Foveal avascular zone segmentation in optical coherence tomography angiography images using a deep learning approach," *Scientific reports*, vol. 11, no. 1, p. 1031, 2021.

[17] T. T. Hormel, T. S. Hwang, S. T. Bailey, D. J. Wilson, D. Huang, and Y. Jia, "Artificial intelligence in oct angiography," *Progress in Retinal and Eye Research*, vol. 85, p. 100965, 2021.

[18] L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi *et al.*, "Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging," *Medical image analysis*, vol. 67, p. 101874, 2021.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015*. Springer, 2015, pp. 234–241.

[20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

[21] T. Van Quyen and M. Y. Kim, "Feature pyramid network with multi-scale prediction fusion for real-time semantic segmentation," *Neurocomputing*, vol. 519, pp. 104–113, 2023.

[22] B. Zhang, Y. Wang, C. Ding, Z. Deng, L. Li, Z. Qin, Z. Ding, L. Bian, and C. Yang, "Multi-scale feature pyramid fusion network for medical image segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 2, pp. 353–365, 2023.

[23] R. Li, Z. Li, H. Fan, S. Teng, and X. Cao, "Mcfsa-net: A multi-scale channel fusion and spatial activation network for retinal vessel segmentation," *Journal of Biophotonics*, vol. 16, no. 4, p. e202200295, 2023.

[24] H. Zhang, W. Fang, and J. Li, "A microvascular segmentation network based on pyramidal attention mechanism," *Sensors*, vol. 24, no. 12, p. 4014, 2024.

[25] K. Huang, N. Su, X. Ma, M. Li, J. Yang, S. Yuan, Y. Liu, and Q. Chen, "Choroidal vessel segmentation in sd-oct with 3d shape-aware adversarial networks," *Biomedical Signal Processing and Control*, vol. 84, p. 104982, 2023.

[26] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1073–1083, 2013.

[27] Y. Zhao, J. Xie, H. Zhang, Y. Zheng, Y. Zhao, H. Qi, Y. Zhao, P. Su, J. Liu, and Y. Liu, "Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering," *IEEE transactions on medical imaging*, vol. 39, no. 2, pp. 341–356, 2019.

[28] S. Zhang, R. Zheng, Y. Luo, X. Wang, J. Mao, C. J. Roberts, and M. Sun, "Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network," *IEEE Access*, vol. 7, pp. 57 561–57 573, 2019.

[29] M. Alam, D. Toslak, J. I. Lim, and X. Yao, "Color fundus image guided artery-vein differentiation in optical coherence tomography angiography," *Investigative ophthalmology & visual science*, vol. 59, no. 12, pp. 4953–4962, 2018.

[30] M. Alam, D. Le, T. Son, J. I. Lim, and X. Yao, "Av-net: deep learning for fully automated artery-vein classification in optical coherence tomography angiography," *Biomedical optics express*, vol. 11, no. 9, pp. 5249–5257, 2020.

[31] J. Xie, Y. Liu, Y. Zheng, P. Su, Y. Hu, J. Yang, J. Liu, and Y. Zhao, "Classification of retinal vessels into artery-vein in oct angiography guided by fundus images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*. Springer, 2020, pp. 117–127.

[32] Y. Zheng, J. S. Gandhi, A. N. Stangos, C. Campa, D. M. Broadbent, and S. P. Harding, "Automated segmentation of foveal avascular zone in fundus fluorescein angiography," *Investigative ophthalmology & visual science*, vol. 51, no. 7, pp. 3653–3659, 2010.

[33] M. Díaz, J. Novo, P. Cutrín, F. Gómez-Ulla, M. G. Penedo, and M. Ortega, "Automatic segmentation of the foveal avascular zone in ophthalmological oct-a images," *PloS one*, vol. 14, no. 2, p. e0212364, 2019.

[34] M. Li, Y. Chen, Z. Ji, K. Xie, S. Yuan, Q. Chen, and S. Li, "Image projection network: 3d to 2d image segmentation in octa images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3343–3354, 2020.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[36] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 540–549, 2019.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[38] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[39] J. Bertels, T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*. Springer, 2019, pp. 92–100.

[40] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[41] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2020, pp. 1055–1059.

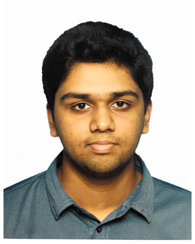
[42] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[43] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "Cs-net: Channel and spatial attention network for curvilinear structure segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*. Springer, 2019, pp. 721–730.

[44] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[45] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 574–584.

Nisan Pranavah Raja is currently pursuing his Ph.D. in the Department of Electronics and Communication Engineering at the National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. His research interests include artificial intelligence, biomedical image processing, and deep learning.



Srivatsan Sarvesan is currently pursuing his B.Tech degree in the Department of Electronics and Communication Engineering at the National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. His research interests include artificial intelligence, biomedical image processing, and deep learning.

Dr. Anju Thomas is a post-doctoral fellow in the Department of Electronics and Communication Engineering at the National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. Her research interests include artificial intelligence, signal processing, and medical image processing.

Dr. Varun P. Gopi is an Associate Professor in the Department of Electronics and Communication Engineering at the National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. His areas of expertise include artificial intelligence, signal processing, medical image processing, circuits and systems, compressed sensing, and hardware for deep learning.