IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 December 2022, accepted 5 February 2023, date of publication 14 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244922

RESEARCH ARTICLE

A Deep Ensemble Learning-Based CNN Architecture for Multiclass Retinal Fluid Segmentation in OCT Images

MOHAMMAD RAHIL[1], B. N. ANOOP[1], (Member, IEEE), G. N. GIRISH[2], ABHISHEK R. KOTHARI[3], SHASHIDHAR G. KOOLAGUDI[1], AND JENY RAJAN[1]

[1]Department of Computer Science and Engineering, National Institute of Technology Karnataka, Surathkal 575025, India
[2]Department of Computer Science and Engineering, Indian Institute of Information Technology Sri City, Chittoor 517646, India
[3]Pink City Eye and Retina Center, Jaipur 302015, India

Corresponding author: B. N. Anoop (anoopcem@gmail.com)

This work was supported by the Science and Engineering Research Board, Department of Science and Technology, India, under Grant EMR/2016/002677.

ABSTRACT Retinal Fluids (fluid collections) develop because of the accumulation of fluid in the retina, which may be caused by several retinal disorders, and can lead to loss of vision. Optical coherence tomography (OCT) provides non-invasive cross-sectional images of the retina and enables the visualization of different retinal abnormalities. The identification and segmentation of retinal cysts from OCT scans is gaining immense attention since the manual analysis of OCT data is time consuming and requires an experienced ophthalmologist. Identification and categorization of the retinal cysts aids in establishing the pathophysiology of various retinal diseases, such as macular edema, diabetic macular edema, and age-related macular degeneration. Hence, an automatic algorithm for the segmentation and detection of retinal cysts would be of great value to the ophthalmologists. In this study, we have proposed a convolutional neural network-based deep ensemble architecture that can segment the three different types of retinal cysts from the retinal OCT images. The quantitative and qualitative performance of the model was evaluated using the publicly available RETOUCH challenge dataset. The proposed model outperformed the state-of-the-art methods, with an overall improvement of 1.8%.

INDEX TERMS Optical coherence tomography, retinal cysts, intra retinal fluid, sub retinal fluid, pigment epithelial detachment, ensemble-approach, deep learning, medical image segmentation.

I. INTRODUCTION

The human eye is a complex sensory organ that enables us to see and interpret the scenes around us. The eye receives the reflected light from the different objects in a scene and converts them into electrochemical signals that are further processed by the visual cortex of the human brain. The retina is a thin layered tissue in the eye, which is responsible for converting the optical signals to electrochemical signals that are fed to the visual cortex via the optic fiber nerves [1]. Retinal abnormalities can cause visual impairment due to several pathologies. One such pathology is the presence of fluid-filled regions in the retina, which are called retinal cysts [2], [3], [4], [5]. These cysts are formed due to underlying conditions such as macular edema, diabetic retinopathy, age-related macular degeneration (AMD), central serous chorioretinopathy, and retinal vein occlusion. The pathophysiological categorization of the retinal cysts can aid in the diagnosis of the aforementioned ocular disorders. The retinal cysts can be classified into three types based on their location in the retina:

- Intra-retinal fluid (IRF): It is present between the nerve fiber layer and the outer plexiform layer of the retina [4], [5], [6], [7]. The presence of IRF has a pathological

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

significance in the diagnosis of diabetic macular edema and diabetic retinopathy.

- Sub-retinal fluid (SRF): It appears between the photoreceptor layer and the retinal pigment epithelium (RPE) [7], [8], [9], [10], [11]. SRF is a collection of lipid-rich exudate (serous liquid) that enters the area from the choroid through the damaged pigment epithelium due to inflammation or tumor. SRF appears owing to the breakdown of the normal anatomical course of action of the retina and its supporting tissues. The condition has significance in the diagnosis of AMD and central serous chorioretinopathy. Features of the SRF fluid are the moving of this liquid with postural changes and the smooth arc-like appearance of the detached retina lacking layering or fixed folds.
- Pigment epithelial detachment (PED): It occurs due to the detachment or elevation of the RPE layer from the Bruch's membrane. A PED can be found in certain choroidal diseases, just as in some fundamental conditions [9], [12]. The two roots of PED are ARMD and its variations, such as polypoidal choroidal vasculopathy and retinal angiomatous expansion, and central serous chorioretinopathy. In any case, regardless of whether less as often as possible, PEDs can be found in a few foundational issues of provocative, irresistible, neoplastic, and iatrogenic nature.

Optical coherence tomography (OCT) is a non-invasive technique used to image the retina [13], [14]. OCT employs low coherence light waves to produce different cross-sectional views of the retina using the interferometry technique. OCT scans acquired are 3D images [15], commonly called volumes or C-scans. Each OCT volume consists of B-scans which are 2D images taken at the different cross-sectional locations. OCT aids in imaging different pathologies (like cysts) of the retina by providing cross-sectional views. Commonly called volumes or C-scans. Each OCT volume consists of B-scans that are 2D images taken at various cross-sectional locations. OCT aids in imaging different pathologies (such as cysts) of the retina by providing cross-sectional views.

However, identifying the retinal cysts and categorizing them based on OCT scans requires immense expertise and is a time-consuming task. Hence, in this work, we have proposed an automatic convolutional neural network (CNN)-based model to segment the three different retinal cysts (IRF, SRF, and PED).

Prior to the RETOUCH challenge [16], we could not find a single work addressing the segmentation of all types of retinal cysts, which is mainly due to the unavailability of benchmark dataset. This challenge focused on creating a benchmark dataset by considering all the three types of retinal cysts, and the scans were acquired using three different OCT instruments. Eight global teams participated in this endeavor, namely Helios [17], MABIC [18], NJUST [19], RetinAI [20], RMIT [21], SFU [22], UCF [23], and UMN [24]. Bogunovic et al. [16] have reported the outcome of this challenge, and almost all the teams used CNN-based solutions for retinal cyst segmentation. The team SFU won the challenge.

A patch-wise approach for image processing has been proposed [25]. This approach involves breaking the original images into small patches or images of small dimensions. It has been proven in the past that training a model with small patches gives better results in segmentation than training it with a huge complex image. Once the images are broken into small dimensions, the patch models are able to learn complex functions easily. It means that in patch-wise learning, at a time the model is learning a part of the image only and not the whole image. The other factors that are involved while discussing the patch-wise learning approach is the amount of overlap each patch has over the others and its dimensions. Both these factors play key roles in deciding the final Dice score of the model. Both act as hyperparameters of the model, if they are set correctly, and provide the best results. An ensemble-based segmentation approach was proposed in [26]. The optimal model selection was a drawback of the method proposed in [26]. Also, in [27] a multi-resolution based CNN architecture called RF Net is proposed for the segmentation of retinal cysts from OCT images. Other approaches include the dual attention based CNN [28], annotation efficient joint segmentation [29], and dense atrous convolution & special pyramidal pooling based deep joint segmentation [30].

In this study, we also considered some of the recently published CNN models in the field of medical image segmentation, such as Double UNET [31], Bidirectional UNET [32], KI-UNET [33], DC-UNET [34], attention UNET [35], and Multi-guided UNET [36]. Double UNET [31] is a kind of extension of simple UNET and has two UNET present in it. It has two encoders and two decoders. For the first encoder, we employed a VGG19 architecture, while the second decoder was similar to the UNET encoder. Double UNET performed exceptionally well on the CVC-ClinicalDB [37] and ETIS-Laraib datasets and yielded state-of-the-art-results. We also trained Bidirectional UNET [32], which is a combination of UNET, convolution LSTM, and dense convolution. Apart from UNET, it possesses the advantages of LSTM and dense layer. During our data analysis, we found the presence of cysts of different scales. Since UNET provided only high-level information, we might have missed out on low-level information, such as small cyst regions. Hence, we also tried KI-UNET [33]. which is a combination of UNET and Reverse UNET. We used DC-UNET [34], too, which is based on the concept of multi-resolution layers that help in the segmentation of cysts of different shapes and scales. Furthermore, we attempted to add attention modules to UNET [35] based on the belief that they might be helpful in increasing the Dice score. Moreover, using UNET, we tested the patched approaches [25] and the multi-guided attention modules in multi-guided architecture [36].

The key contributions in this paper are as follows-

- Proposed an one-to-one fluid segmentation architecture with a base CNN for the multi-class retinal cyst segmentation.
- Performed 28 different experiments with the existing CNN based segmentation architectures for finalizing the optimal fluid segmentation network.
- Experimentally showed that, an ensemble approach is an efficient technique for the retinal cyst segmentation.

The rest of the paper is structured as follows, Section II describes the proposed method which includes the data set, pre-processing, and model architecture. Section III discusses the results and comparative analysis with existing methods. Finally, conclusions and remarks are drawn in section IV.

II. METHODOLOGY

The proposed CNN model is an ensemble learning-based approach and contains three different base models for the three kinds of cysts, followed by a predictor block. Ensemble learning is a machine learning technique adopted for better prediction accuracy for multiclass segmentation. Here, by training, multiple deep learning models and outcomes are generated with the help of majority voting. We built three separate models for the segmentation of the retinal cysts—IRF, SRF, and PED. With rigorous experiments, we found that a single model is not sufficient to obtain a good performance. The pictorial representation of the proposed segmentation pipeline is given in figure 1. Each model was trained separately with its corresponding data. The training of the models was performed using the same input OCT images and by changing the corresponding ground truth image. In the following sections, we have elaborated on the different models used for the proposed ensemble-based segmentation architecture.

A. PROPOSED IRF MODEL ARCHITECTURE

For the segmentation of the IRF, we used the well-known UNET model with a combination of the relative layer distance [16]. Furthermore, we employed the data augmentation technique for IRF segmentation. The results revealed that data augmentation does not contribute much to the accuracy of the IRF segmentation. In data augmentation, we performed some basic augmentations such as horizontal flip, vertical flip, rotation, and zooming.

Relative distance is an algorithm that has been used to obtain additional information on the retinal layers to enhance the segmentation of the retinal cysts. This algorithm calculates the distance of each pixel in an image with regard to the internal limiting membrane (ILM) and the RPE. ILM is the topmost retinal layer, and RPE is the bottommost layer. The use of this relative layer distance map as a second channel aided the proper segmentation of the retinal cysts. Based on literature analysis, it was found that the location of the cyst depends on the retinal layer in which it is present. For a pixel (m, n), the relative distance can be calculated using the following formula (Equation 1):

$$f(m,n) = \frac{n - M1(m)}{M1(m) - M2(m)} \tag{1}$$

where f(m, n) is a specific pixel in a B-scan, m and n are the coordinates of that pixel, M1(m) is the y coordinate of ILM, and M2(m) is the y coordinate of RPE. Using the above formula, the relative layer information map was created for every B-scan. Figure 2 represents the architecture utilized for our IRF model. In figure 2, it can be seen that two channels are fed as the input to the model and one channel is obtained as the output, which is the IRF segmented image. The UNET is an encoder (E-Block) and decoder (D-Block) architecture. The depth of UNET is 4. Each block in the encoder contains two simultaneous layers of convolution, with a filter size of 3 × 3. To accelerate the learning process and to reduce the overfitting of the data, data normalization was performed after two layers of convolution. Subsequently, downsampling of the image was done with the help of the max pool layer with a pool size of 2 × 2. The number of neurons was increased in power by 2 in each block, such as 64, 128, 256, and 512, and the bottom layer (B-Block) had 1024 neurons in it.

To preserve the spatial information, we utilized skip connections. To restore the images to their original dimensions, we used the transpose convolution in the decoder blocks. The number of neurons in each decoder block was reduced to 512, 256, 128, and 64, which is exactly the opposite to that of the encoder block. Finally, we had a 1 × 1 convolution layer with one neuron in it and the sigmoid as the activation function. We employed the Binary Cross Entropy as a loss function in our model. The learning rate of our model was 3e–4. We ran this model for 200 epochs and saved the best weights.

TABLE 1. Layers information for the three different models.

Models	IRF	SRF	PED
Relative info	✓	✓	
data augmentation			✓
E-Block(16X16)		✓	
E-Block(32X32)		✓	
E-Block(64X64)	✓	✓	✓
E-Block(128X128)	✓	B-block	✓
E-Block(256X256)	✓		✓
E-Block(512X512)	✓		✓
E-Block(1024X1024)	B-block		B-block
D-Block(512X512)	✓		✓
D-Block(256X256)	✓		✓
D-Block(128X128)	✓	✓	✓
D-Block(64X64)	✓	✓	✓
D-Block(32X32)		✓	
D-Block(16X16)		✓	

B. PROPOSED SRF MODEL ARCHITECTURE

In the extended UNET, we used the structured dropout blocks. One of the characteristics of SRF is its larger volume when

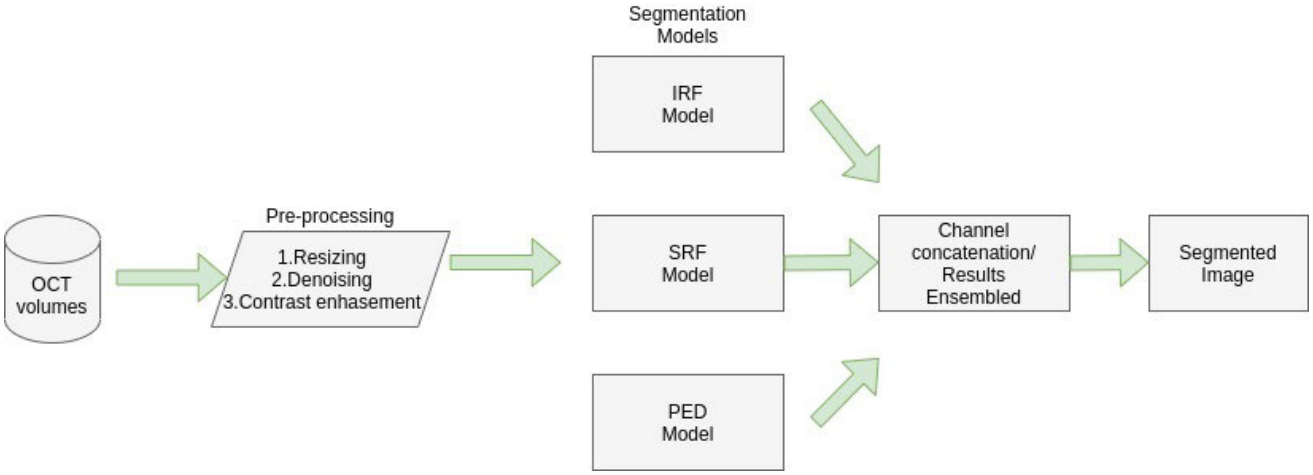

FIGURE 1. Segmentation Pipeline.

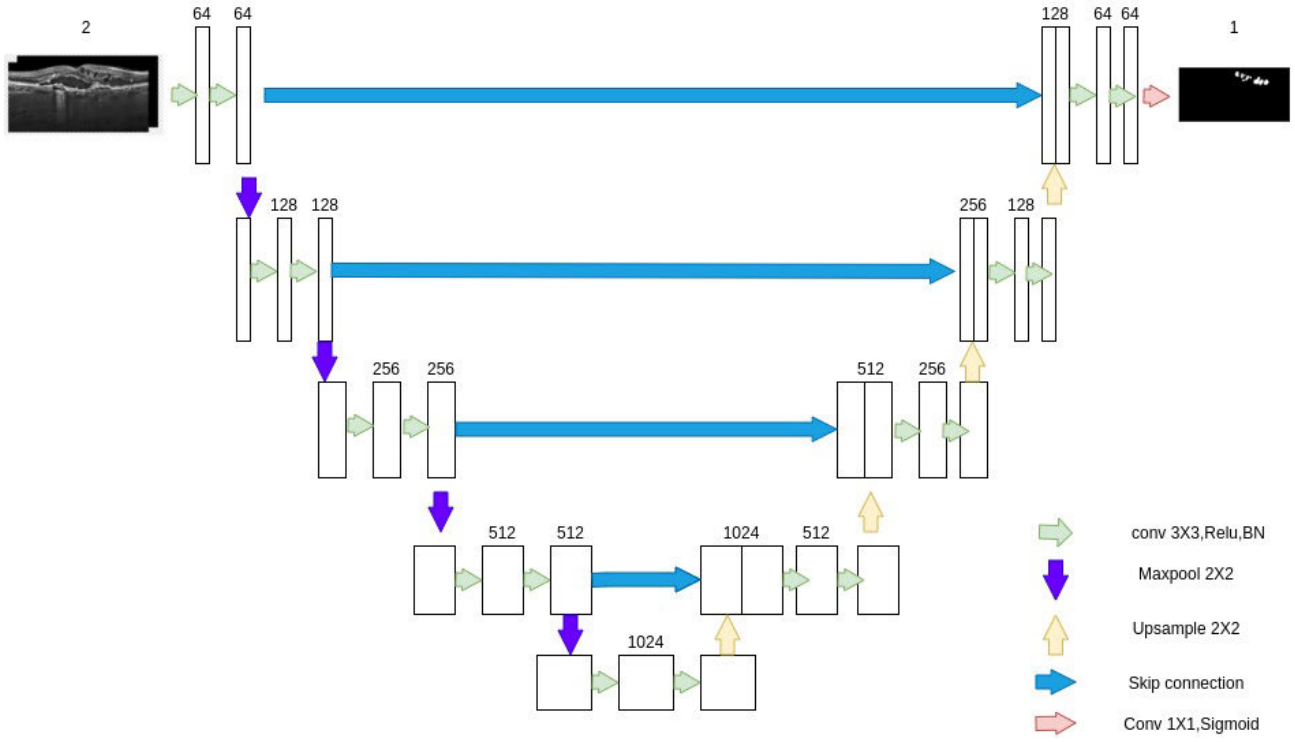

FIGURE 2. Base architecture of the proposed model.

compared with IRF and PED. SRFs are mostly present in the central retinal layers. Hence, we need a less complex model to segment SRF. To avoid model overfitting, we tried various data augmentation techniques; however, the results were inappropriate. To reduce the complexity, we constructed a model with fewer layers and fewer model parameters. Additionally, we included drop blocks to avoid overfitting. The main difference between dropout and drop blocks is that the former removes the random units from the layers while the latter removes the contiguous areas of the feature maps from the layers.

In this model, every convolution layer is followed by a Dropblock layer, a batch normalization layer to regularize the model, and an ReLU activation layer. Subsequently, maxpool layer is applied in the encoding path to downsample the image with a pool size of 2×2. In the decoding path, transpose convolution layer is applied with a filter size of 2×2 to upsample the image. The skip connections aid in concate-

TABLE 2. Dataset brief description.

DATASET	CIRRUS	SPECTRALIS	TOPCORN
No of B-scans	3072	1176	2688
No of volumes	24	24	22
No of B-scans/volume	128	49	128
Dimensions of B-scans	512 X 1024	512 X 496	512 X 885

nating the feature maps from the encoding blocks to their corresponding decoding blocks.

Finally, we obtained a convolution layer of filter size 1×1 with a sigmoid function, yielding a segmentation map of SRF. Similar to the IRF model, we used a relative distance algorithm for the second channel to provide the input. The loss function used in this model was binary cross entropy. We trained the model for 200 epochs and empirically set the learning rate as 3e–4.

C. PROPOSED PED MODEL ARCHITECTURE

The PED model is similar to the model proposed for IRF segmentation but with differences. First, we removed the relative distance channel and used only the OCT images for training. PED is not a retinal layer but a detachment of the retina [38], [39], [40], [41]. Hence, the relative layer information did not contribute much to predicting PED. The second difference is that we used data augmentation in the case of PED. During the course of our experiments, we found that data augmentation helped in PED segmentation. The rest of the aspects were the same as those in the IRF model. Table 1 provides architectural level information on the three different models used in this study concerning the base model.

D. PREDICTION MODEL

The inputs were given in the form of OCT images, and a three-channel output was obtained in which each channel represented the respective segmented cysts. Since we had three different models, we had three different output channels, one from each model. We finally concatenated all three outputs to obtain a single output with three channels in it. The dimensions of this output were the same as those of any other output produced by a single model.

E. DATA SET

We used the RETOUCH challenge dataset in our experiments. This challenge included 112 OCT volumes, of which 70 were given for training and 42 were used for testing. Only 70 OCT volumes were present in the public domain with labels. Hence, we used only these in our experiments and divided them into training, validation, and testing volumes. Out of these 70 OCT volumes, 24 were acquired using a Zeiss Cirrus OCT scanner, 24 using a Heidelberg Spectralis OCT scanner, and 22 using a Topcon OCT scanner. Different vendors have different dimensions and numbers of B-scans in each volume. Table 2 provides a detailed description of the dataset.

TABLE 3. Vendor-wise analysis of OCT B-scans.

Vendor	IRF	Non-IRF	SRF	Non-SRF	PED	Non-PED
Cirrus	523	1269	293	1499	2916	1496
Spectralis	329	337	118	568	135	551
Topcorn	447	1217	783	3377	822	3320

The volume-wise data do not provide insights into the dataset because a volume may contain B-scans ranging from 49 to 128. What is important is the number of B-scans in these volumes that are positive or negative for IRF, SRF, and PED. Table 3 gives the framewise analysis for different vendors. From the table, it is evident that there is a huge gap between the positive number of frames and the negative number of frames for all the three cysts.

We have split those 70 volumes of the data into train, test and validation sets. In medical image processing, 70 volumes is an extensive dataset considering the manual annotation labour.

F. PREPROCESSING

The preprocessing stage focuses on removing the irrelevent and confusing features from the image. Noise is one of the negative features that affects the performance of the segmentation network. Our preprocessing stage included three parts. First, we cropped the images and resized them into 256×512. Second, we denoised the images and used an Unbiased Fast Non Local Means filter to denoise the scans. Third, we performed contrast enhancement and employed CLAHE to distinguish between the fluid and non-fluid regions. For visual understanding, we took one raw frame from the Zeiss Cirrus vendor. The preprocessing stage is demonstrated in Figure 3.

G. MODEL PERFORMANCE ANALYSIS

The performance of the model was evaluated with the help of well-known quality evaluation metrics such as Precision, Recall, and Dice score [42], [43]. The mathematical relations for calculating Precision, Recall, and Dice are provided in equations 2, 3 and 4 respectively.

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$dice = 2 * \frac{| predicted \cap groundtruth |}{| predicted \cup groundtruth |} \tag{4}$$

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section provides details of the experiments conducted to evaluate the proposed method and the strategies used to validate its overall robustness and adaptability. Out of the 70 training OCT volumes available in the RETOUCH dataset, 41 were used for training, 6 for validation, and 23 for testing. Hence, the training set consists of 4142 B-scans (49704 Patches), a test set having 610 B-scans (7320 Patches), and

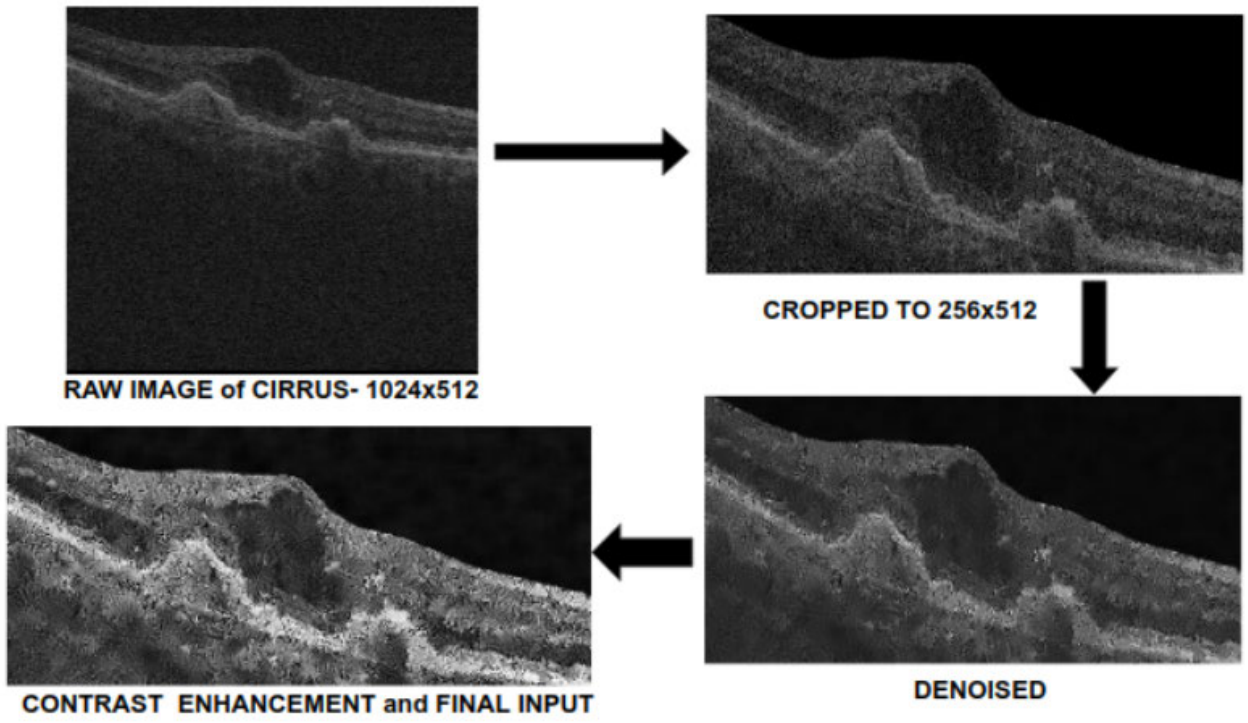

FIGURE 3. Preprocessing stages with one sample.

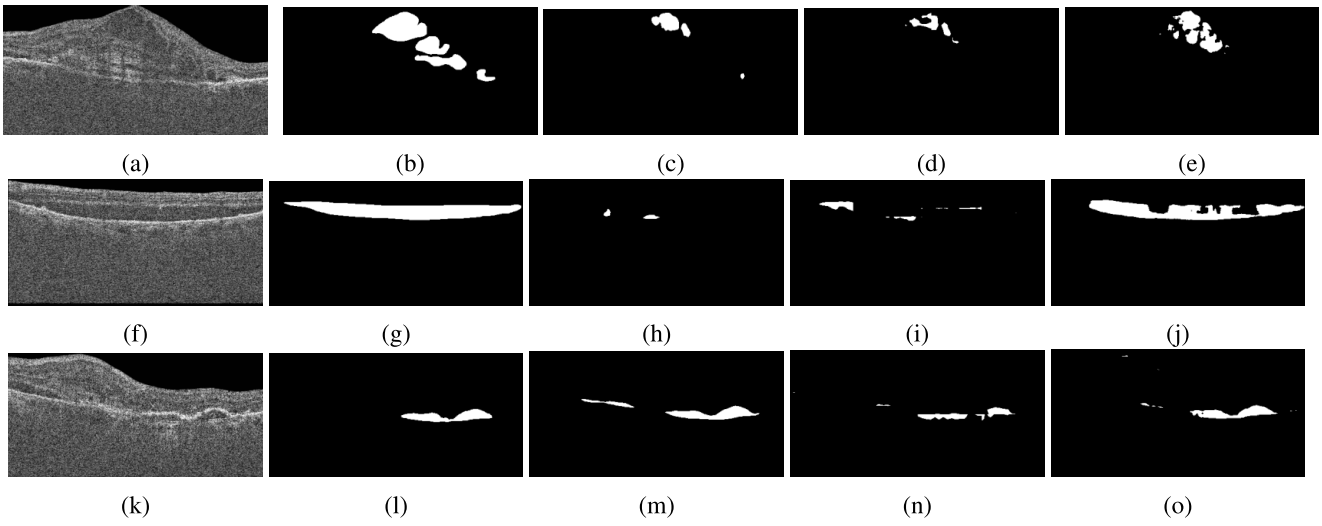

FIGURE 4. The qualitative evaluation over cirrus data: (a) IRF input, (b) IRF ground truth, (c) SFU IRF result, (d) Deeplab IRF result, (e) Ensemble IRF result, (f) SRF input, (g) SRF ground truth, (h) SFU SRF result, (i) Deeplab SRF result, (j) Ensemble SRF result, (k) PED input, (l) PED ground truth, (m) SFU PED result, (n) Deeplab PED result, (o) Ensemble PED result.

the remaining 2312 B scans (27744 Patches) have been employed for testing the model. Also, we have considered the overlapped patches to increase the dataset further. Table 4 provides an overview of the performed experiments and their results. We considered 27 different segmentation networks. The thorough evaluation and its inferences helped us in modeling a novel approach for retinal image segmentation. To prove the generalizability of the proposed model, we created different data splits for training, validation, and testing. All models were tested with and without a relative layer. Table 4 depicts the use of a relative layer in the model. The very first model tested was the SFU [16] and the results were taken as the benchmark. The second model was FCN (fully convolutional neural network), which is the same as

TABLE 4. Models results on RETOUCH dataset.

S.No	Models	IRF (Dice)	SRF (Dice)	PED (Dice)	Average (Dice)
1	UNET+rel (sfu)	0.71	0.70	0.78	0.74
2	FCN	0.73	0.68	0.74	0.74
3	Deeplab (patched)	0.58	0.66	0.73	0.64
4	Double Unet (cascaded unet)	0.70	0.67	0.74	0.70
5	Bidirectional Unet (LSTM and Dense)	0.73	0.73	0.74	0.74
6	Bidirectional Unet+rel	0.70	0.72	0.78	0.74
7	Bidirectional Unet (only LSTM)	0.70	0.71	0.75	0.70
8	SD-Unet	0.71	0.74	0.74	0.72
9	SA-Unet (with spatial layer)	0.66	0.71	0.73	0.69
10	KI-Unet (Unet & Reverse unet)	0.63	0.58	0.53	0.60
11	DC-Unet	0.70	0.70	0.76	0.70
12	DC-Unet+rel	0.72	0.71	0.78	0.74
13	Unet+aug (flip,rotate,zoom)	0.70	0.71	0.76	0.72
14	Unet+aug (flip,rotate,zoom)+rel	0.71	0.70	0.77	0.73
15	Double-Unet+rel	0.67	0.65	0.70	0.69
16	Multi-guided	0.69	0.70	0.75	0.73
17	Multi-guided+rel	0.67	0.72	0.73	0.70
18	Ensemble-sfu	0.72	0.74	0.78	0.74
19	Attention-Unet3	0.68	0.69	0.75	0.69
20	Attention-Unet4	0.71	0.59	0.70	0.73
21	Attention-Unet5	0.68	0.61	0.70	0.70
22	Attention-Unet4+rel	0.68	0.68	0.76	0.70
23	Attention-Unet5+rel	0.67	0.68	0.73	0.69
24	Patched SFU (128,128,0.72)	0.68	0.70	0.76	0.71
25	Patched SFU+rel (128,128,0.65)	0.66	0.68	0.75	0.68
26	sfu-attention	0.67	0.67	0.75	0.69
27	sfu-dense	0.70	0.72	0.77	0.71
28	Proposed model	**0.728**	**0.75**	**0.79**	**0.76**

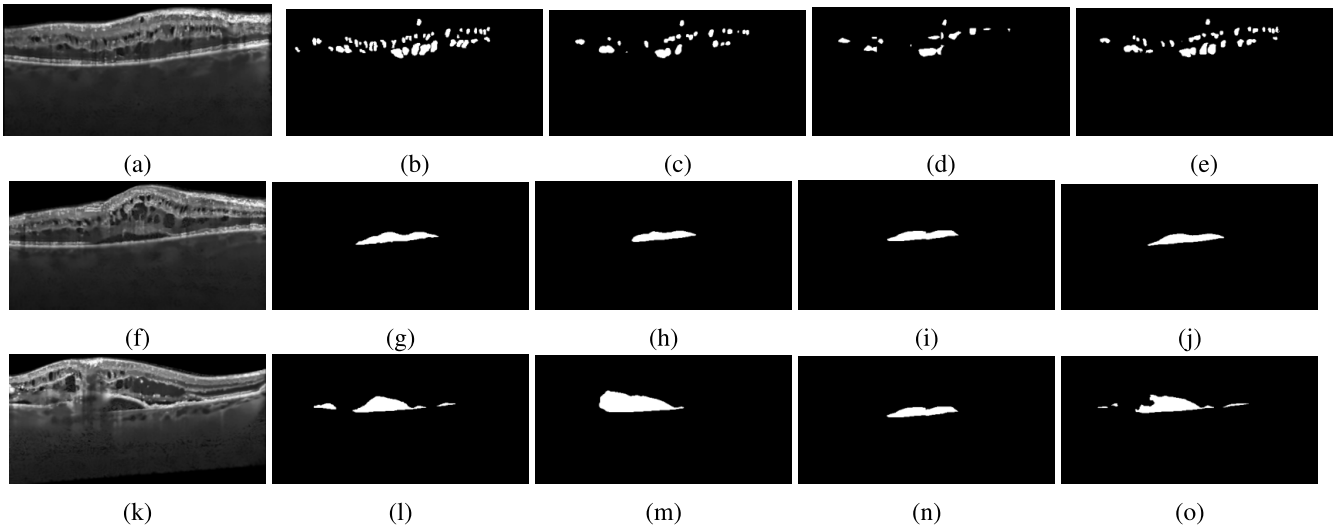

FIGURE 5. The qualitative evaluation over spectralis data: (a) IRF input, (b) IRF ground truth, (c) SFU IRF result, (d) Deeplab IRF result, (e) Ensemble IRF result, (f) SRF input, (g) SRF ground truth, (h) SFU SRF result, (i) Deeplab SRF result, (j) Ensemble SRF result, (k) PED input, (l) PED ground truth, (m) SFU PED result, (n) Deeplab PED result, (o) Ensemble PED result.

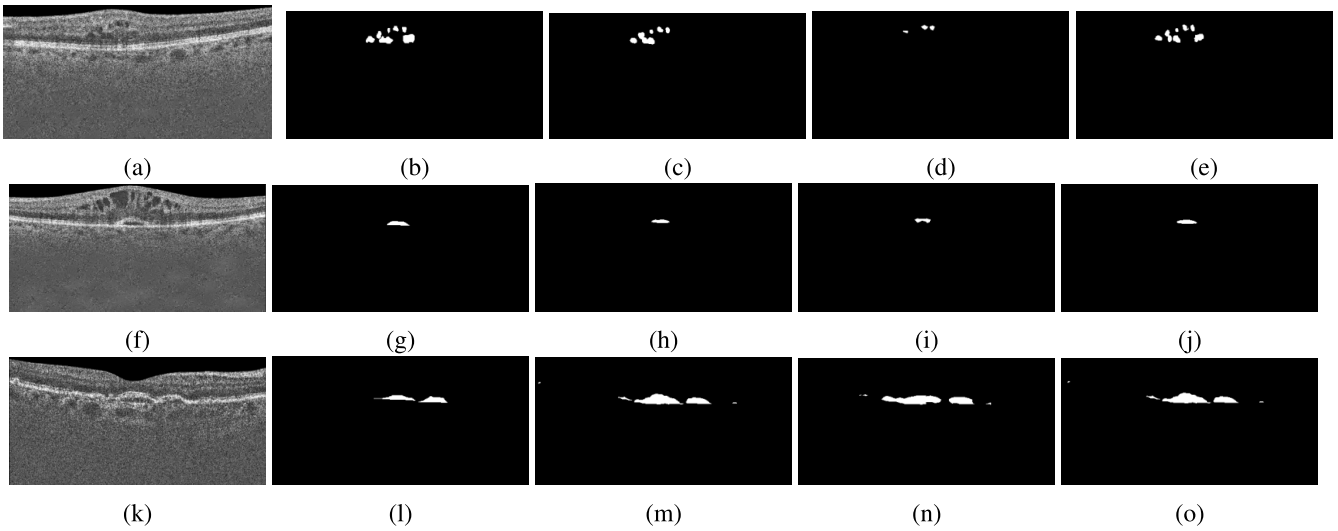

FIGURE 6. The qualitative evaluation over Topcorn data: (a) IRF input, (b) IRF ground truth, (c) SFU IRF result, (d) Deeplab IRF result, (e) Ensemble IRF result, (f) SRF input, (g) SRF ground truth, (h) SFU SRF result, (i) Deeplab SRF result, (j) Ensemble SRF result, (k) PED input, (l) PED ground truth, (m) SFU PED result, (n) Deeplab PED result, (o) Ensemble PED result.

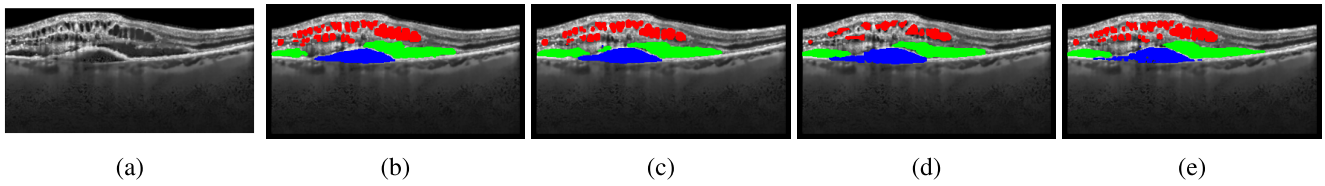

FIGURE 7. Overall qualitative evaluation over spectralis data: (a) input, (b)ground truth, (c) SFU result, (d) Deeplab result, (e) Ensemble result.

the first model except that its depth is less and it does not use relative distance as the second channel. The third is a patch-wise DeepLab [25]. Our experiments also found that the patch-wise approach did not work well with the cysts dataset. The fourth model was double UNET [31] and the fifth model was bidirectional UNET [32]. Bidirectional UNET showed better performance in IRF and SRF when compared with the benchmark model, but the overall average remained the same as that of the benchmark. In the sixth experiment, we employed bidirectional UNET along with relative information, which provided the same results as the benchmark. In the seventh experiment, we utilized a bidirectional model without dense layers. The eighth experiment made use of structured Dropout UNET (SDUNET) explained in section II-B which yielded the best results for SRF. In the ninth experiment, we used SD-UNET with a spatial attention block at the bottom of the UNET, which is known as Spatial Attention UNET (SA-UNET). However, the performance of SA-UNET was inferior to that of SDUNET. In the 10^{th} experiment, we used KI-UNET [33]. which performed poorly on the RETOUCH dataset. In the 11^{th} and 12^{th} experiments, we used we have used DC-UNET [34], which when used along a relative layer yielded the same average results as the benchmark model. In the 13^{th} and 14^{th} experiments, we used data augmentation with UNET and found

TABLE 5. Independent models trained for IRF.

Models	Dice Score(IRF)	Precision	Recall
UNET-IRF+rel	0.728	0.82	0.67
UNET-IRF(no rel layer)	0.70	0.87	0.63
UNET-IRF(no rel layer)+augmentation	0.728	0.81	0.65
UNET-IRF+augmentation	0.722	0.82	0.67
UNET-IRF(depth-1)	0.728	0.83	0.67
Bidirectional Unet-IRF (LSTM and Dense)	0.71	0.84	0.65
attention-Unet4-IRF	0.30	0.26	0.26
FCN-IRF	0.65	0.60	0.53
FCN-IRF+rel	0.68	0.67	0.51

that overall augmentation was not helpful in the RETOUCH dataset. In the 15^{th} experiment, we used double UNET with relative layer. In the 16^{th} and 17^{th} experiments, we used a multi-guided attention network, which has been previously explained [36]. In the 18^{th} experiment, we attempted to ensemble the SFU UNET model by taking the same SFU model and training it individually over three different cysts and finally combining the results. However, this approach also did not improve the results. From the 19^{th} to the 23^{rd} experiments, we tried different depth of attention UNET [35] models, We attempted to tune the attention model, but this approach also did not provide acceptable results. We tested the patched network over SFU, but all these approaches failed to give a higher Dice score than the benchmark result.

TABLE 6. Independent models trained for SRF.

Models	Dice Score(SRF)	Precision	Recall
Sfu-SRF	0.74	0.81	0.69
Sfu-SRF(depth-1)	0.69	0.80	0.65
Sfu-SRF(no rel layer)	0.71	0.74	0.71
Sfu-SRF(no rel layer)+augmentation	0.70	0.81	0.64
Sfu-SRF+augmentation	0.69	0.81	0.64
SD-unet	0.74	0.78	0.72
SD-unet+rel	0.75	0.80	0.72

TABLE 7. Independent models trained for PED.

Models	Dice Score(PED)	Precision	Recall
Sfu-PED	0.77	0.71	0.86
UNET-PED(no rel layer)	0.77	0.73	0.84
UNET-PED(no rel layer)+augmentation	0.79	0.77	0.83
UNET-PED+rel+augmentation	0.77	0.73	0.85
Bidirectional-PED+rel	0.76	0.69	0.85
sfu-PED(depth-1)	0.75	0.74	0.79

In experiment 24^{th}, we tried patch-wise SFU architecture with an overlapping of 0.72 %, while in the 25^{th} experiment we checked the same model with an overlapping of 0.65 %. In the 26^{th} experiment, we applied the attention technique to the SFU model, but it did not surpass the benchmark result. Furthermore, we applied a dense network to the bottom layer of the SFU network in the 27^{th} experiment. Nevertheless, the benchmark could not be surpassed. Finally, experiment 28 yielded favorable results for our proposed ensemble approach. The method gave better results than the benchmark method over split one. We performed multiple experiments in which we trained the model over a single cyst and selected the best model that suited a specific cyst. Table 5 provides the list of experiments performed for IRF cyst only.

From Table 5, it is evident that basic UNET along with relative layer information yielded the best results for IRF. Data augmentation did not help in the case of IRF. Hence, in our ensemble approach, we employed UNET along with relative layer information as our IRF model. If trained independently, it gave 1% more Dice score than the SFU benchmark result.

Table 6 shows the experiments performed for SRF segmentation. In these experiments, the models were trained with SRF cysts only. Spatial Dropout UNET (SD-UNET) performed best in the case of SRF and provided 1% better results than the individual trained SFU network and 5% better results than the single SFU model. Hence, we chose SD-UNET along with relative layer information as our SRF model in the ensemble approach.

Table 7 presents the results of experiments performed for PED cyst independently. It can be seen that data augmentation worked for PED. Relative layer information was not quite useful for the segmentation of PED. Hence, we used basic UNET along with data augmentation for the segmentation of PED in our ensemble approach. The UNET along with data augmentation resulted in 2% improvement over the SFU model trained individually over the PED dataset.

After choosing different models for the different cysts, we compared the obtained results with the benchmark results. For the comparison, we used three different splits of data

TABLE 8. Comparison of our Proposed Ensemble approach and Benchmark SFU model.

Models	IRF	SRF	PED	Multiclass(Average)	Precision	Recall
SFU (split 1)	0.71	0.70	0.78	0.74	0.79	0.73
SFU (split 2)	0.55	0.72	0.64	0.68	0.80	0.62
SFU (split 3)	0.70	0.74	0.58	0.68	0.67	0.71
Proposed-model (split 1)	0.72	0.75	0.79	0.766	0.80	0.75
Proposed-model(split 2)	0.58	0.70	0.65	0.70	0.79	0.66
Proposed-model(split 3)	0.71	0.74	0.52	0.74	0.74	0.68

to ensure that our approach surpassed the state-of-the-art method. Table 8 gives the comparison of our ensemble model with the state-of-the-art model over three different splits.

After calculating the average multiclass Dice score for the three different splits, we got a 70% Dice score for the SFU model, while our model yielded a score of 71.86%, which is an improvement of 1.8% over the benchmark. Hence, it is clear that our ensemble approach surpassed the benchmark by 1.8%.

The qualitative evaluation of the proposed model over different vendors (Cirrus, Spectralis and Topcon) is depicted in Figures figure 4, figure 5, figure 6 and figure 7. From the figures, it is clear that the proposed method yielded better retinal cyst segmentation with fewer false positives.

IV. CONCLUSION

In conclusion, this study has proposed an automated method for the segmentation and detection of three different retinal cysts from OCT images using a deep ensemble learning-based technique. To create the ensemble-based architecture, we used three base models, which are extended versions of UNET architecture, and a predictor block that combined the results of all three models. The experimental results on the RETOUCH dataset indicated that the proposed architecture improved the segmentation and detection accuracies when compared with the stand-alone SFU model, which is the state-of-the-art method, by 1.8%. During our rigorous experiments, we also found that data augmentation does not always help in case of retinal cyst segmentation and detection although it helps in the identification of PED cysts.

REFERENCES

[1] B. Anoop, G. Girish, P. Sudeep, and J. Rajan, "Despeckling algorithms for optical coherence tomography images: A review," in *Advanced Classification Techniques for Healthcare Analysis*. Hershey, PA, USA: IGI Global, 2019, pp. 286–310.

[2] S. R. Irvine, "A newly defined vitreous syndrome following cataract surgery: Interpreted according to recent concepts of the structure of the vitreous, the seventh Francis I. Proctor lecture," *Amer. J. Ophthalmol.*, vol. 36, no. 5, pp. 601–619, 1953.

[3] G. N. Girish, A. R. Kothari, and J. Rajan, "Automated segmentation of intra-retinal cysts from optical coherence tomography scans using marker controlled watershed transform," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1292–1295.

[4] G. N. Girish, B. Thakur, S. R. Chowdhury, A. R. Kothari, and J. Rajan, "Segmentation of intra-retinal cysts from optical coherence tomography images using a fully convolutional neural network model," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 296–304, Jan. 2019.

[5] G. N. Girish, B. Saikumar, S. Roychowdhury, A. R. Kothari, and J. Rajan, "Depthwise separable convolutional neural network model for intra-retinal cyst segmentation," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2027–2031.

[6] F. Venhuizen, M. J. Van Grinsven, B. Van Ginneken, C. C. Hoyng, T. Theelen, and C. I. Sanchez, "Fully automated segmentation of intraretinal cysts in 3D optical coherence tomography," *Investigative Ophthalmol. Vis. Sci.*, vol. 57, no. 12, p. 5949, 2016.

[7] J. Wang, M. Zhang, A. D. Pechauer, L. Liu, T. S. Hwang, D. J. Wilson, D. Li, and Y. Jia, "Automated volumetric segmentation of retinal fluid on optical coherence tomography," *Biomed. Opt. Exp.*, vol. 7, no. 4, pp. 1577–1589, Apr. 2016.

[8] J. Novosel, K. A. Vermeer, J. H. De Jong, Z. Wang, and L. J. Van Vliet, "Joint segmentation of retinal layers and focal lesions in 3-D OCT data of topologically disrupted retinas," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1276–1286, Jun. 2017.

[9] M. Wu, W. Fan, Q. Chen, Z. Du, X. Li, S. Yuan, and H. Park, "Three-dimensional continuous max flow optimization-based serous retinal detachment segmentation in SD-OCT for central serous chorioretinopathy," *Biomed. Opt. Exp.*, vol. 8, no. 9, pp. 4257–4274, Sep. 2017.

[10] T. Schlegl, S. M. Waldstein, H. Bogunovic, F. Endstraßer, A. Sadeghipour, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, and U. Schmidt-Erfurth, "Fully automated detection and quantification of macular fluid in OCT using deep learning," *Ophthalmology*, vol. 125, no. 4, pp. 549–558, 2017.

[11] A. Montuoro, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, and H. Bogunović, "Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context," *Biomed. Opt. Exp.*, vol. 8, no. 3, pp. 1874–1888, 2017.

[12] Z. Sun, H. Chen, F. Shi, L. Wang, W. Zhu, D. Xiang, C. Yan, L. Li, and X. Chen, "An automated framework for 3D serous pigment epithelium detachment segmentation in SD-OCT images," *Sci. Rep.*, vol. 6, no. 1, p. 21739, Feb. 2016.

[13] B. N. Anoop, K. S. Kalmady, A. Udathu, V. Siddharth, G. N. Girish, A. R. Kothari, and J. Rajan, "A cascaded convolutional neural network architecture for despeckling OCT images," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102463.

[14] S. N. Menon, V. V. Reddy, A. Yeshwanth, B. Anoop, and J. Rajan, "A novel deep learning approach for the removal of speckle noise from optical coherence tomography images using gated convolution–deconvolution structure," in *Proc. 3rd Int. Conf. Comput. Vis. Image Process.* Cham, Switzerland: Springer, 2020, pp. 115–126.

[15] R. Kafieh, H. Rabbani, and I. Selesnick, "Three dimensional data-driven multi scale atomic representation of optical coherence tomography," *IEEE Trans. Med. Imag.*, vol. 34, no. 5, pp. 1042–1062, May 2015.

[16] H. Bogunovic, F. Venhuizen, S. Riedl, S. Apostolopoulos, A. Bab-Hadiashar, M. F. Beg, L. Bekalo, Q. Chen, C. Ciller, K. Gopinath, A. K. Gostar, K. Jeon, Z. Ji, S. Kang, D. Koozekanani, D. Lu, D. Morley, K. Parhi, and U. Schmidt-Erfurth, "RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1858–1874, Aug. 2019.

[17] S. Yadav, K. Gopinath, and J. Sivaswamy, "A generalized motion pattern and FCN based approach for retinal fluid detection and segmentation," 2017, *arXiv:1712.01073*.

[18] S. H. Kang, H. S. Park, J. Jang, and K. Jeon, "Deep neural networks for the detection and segmentation of the retinal fluid in OCT images," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, 2017, pp. 9–14.

[19] Q. Chen, Z. Ji, T. Wang, Y. Tand, C. Yu, O. I. Paul, and L. B. Sappa, "Automatic segmentation of fluid-associated abnormalities and pigment epithelial detachment in retinal SD-OCT images," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, 2017, pp. 15–21.

[20] S. Apostolopoulos, C. Ciller, R. Sznitman, and S. De Zanet, "Simultaneous classification and segmentation of cysts in retinal OCT," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, 2017, pp. 22–29.

[21] R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar, "Retinal fluid segmentation and classification in OCT images using adversarial loss based CNN," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, 2017, pp. 30–37.

[22] D. Lu, M. Heisler, S. Lee, G. Ding, M. V. Sarunic, and M. F. Beg, "Retinal fluid segmentation and detection in optical coherence tomography images using fully convolutional neural network," 2017, *arXiv:1710.04778*.

[23] D. Morley, H. Foroosh, S. Shaikh, and U. Bagci, "Simultaneous detection and quantification of retinal fluid with deep learning," 2017, *arXiv:1708.05464*.

[24] A. Rashno, D. D. Koozekanani, and K. K. Parhi, "Detection and segmentation of various types of fluids with graph shortest path and deep learning approaches," in *Proc. MICCAI Retinal OCT Fluid Challenge (RETOUCH)*, 2017, pp. 54–62.

[25] K. Alsaih, M. Z. Yusoff, T. B. Tang, I. Faye, and F. Mériaudeau, "Deep learning architectures analysis for age-related macular degeneration segmentation on optical coherence tomography scans," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105566.

[26] K. Alsaih, M. Z. Yusoff, I. Faye, T. B. Tang, and F. Meriaudeau, "Retinal fluid segmentation using ensembled 2-dimensionally and 2.5-dimensionally deep learning networks," *IEEE Access*, vol. 8, pp. 152452–152464, 2020.

[27] B. Hassan, S. Qin, R. Ahmed, T. Hassan, A. H. Taguri, S. Hashmi, and N. Werghi, "Deep learning based joint segmentation and characterization of multi-class retinal fluid lesions on OCT scans for clinical use in anti-VEGF therapy," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104727.

[28] W. Liu, Y. Sun, and Q. Ji, "MDAN-UNet: Multi-scale and dual attention enhanced nested U-Net architecture for segmentation of optical coherence tomography images," *Algorithms*, vol. 13, no. 3, p. 60, Mar. 2020.

[29] B. Hassan, S. Qin, T. Hassan, R. Ahmed, and N. Werghi, "Joint segmentation and quantification of chorioretinal biomarkers in optical coherence tomography scans: A deep learning approach," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–17, 2021.

[30] B. Hassan, S. Qin, and R. Ahmed, "SEADNet: Deep learning driven segmentation and extraction of macular fluids in 3D retinal OCT scans," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2020, pp. 1–6.

[31] D. Jha, M. Riegler, D. Johansen, P. Halvorsen, and H. Johansen, "DoubleU-Net: A deep convolutional neural network for medical image segmentation," in *Proc. IEEE 33rd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2020, pp. 558–564.

[32] R. Azad, M. Asadi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, Oct. 2019, pp. 406–415.

[33] J. M. J. Valanarasu, V. Sindagi, I. Hacihaliloglu, and V. Patel, "KIU-Net: Towards accurate segmentation of biomedical images using over-complete representations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Jun. 2020, pp. 363–373.

[34] A. Lou, S. Guan, and M. Loew, "DC-UNet: Rethinking the U-Net architecture with dual channel efficient CNN for medical images segmentation," in *Proc. SPIE*, vol. 11596, pp. 758–768, Feb. 2021.

[35] O. Oktay, J. Schlemper, L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," Apr. 2018, *arXiv:1804.03999*.

[36] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 1, pp. 121–130, Jan. 2021.

[37] J. Bernal, N. Tajbakhsh, F. Sanchez, B. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, K. Pogorelov, S. Choi, Q. Debard, L. Maier-Hein, S. Speidel, D. Stoyanov, P. Brandao, H. Córdova, C. Sánchez-Montes, and A. Histace, "Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1231–1249, Jun. 2017.

[38] D. Thomas and G. Duguid, "Optical coherence tomography—A review of the principles and contemporary uses in retinal investigation," *Eye*, vol. 18, no. 6, pp. 561–570, Jun. 2004.

[39] A. S. Wenick and D. E. Barañano, "Evaluation and management of pediatric rhegmatogenous retinal detachment," *Saudi J. Ophthalmol.*, vol. 26, no. 3, pp. 255–263, Jul. 2012.

[40] L. K. Chang and D. Sarraf, "Tears of the retinal pigment epithelium: An old problem in a new era," *Retina*, vol. 27, no. 5, pp. 523–534, 2007.

[41] S. Mrejen, D. Sarraf, S. K. Mukkamala, and K. B. Freund, "Multimodal imaging of pigment epithelial detachment: A guide to evaluation," *Retina*, vol. 33, no. 9, pp. 1735–1762, 2013.

[42] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[43] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Int. J. Mach. Learn. Technol.*, pp. 37–63, 2011. [Online]. Available: https://arxiv.org/abs/2010.16061

MOHAMMAD RAHIL received the B.E. degree from the MGM's Jawaharlal Nehru Engineering College (JNEC), Aurangabad, Maharashtra, India, in 2018, and the M.Tech. degree from National Institute of Technology Karnataka (NITK), Surathkal, Mangalore, India, in 2021, all in computer science and engineering (CSE).

B. N. ANOOP (Member, IEEE) received the Ph.D. degree from the National Institute of Technology Karnataka (NITK), India, in 2021, and the M.Tech. degree from the National Institute of Technology Calicut, India. His research interests include image processing, deep learning, and medical image analysis.

G. N. GIRISH received the Ph.D. degree in computer science and engineering from the National Institute of Technology Karnataka, Surathkal, India. He was a Postdoctoral Research Fellow at the Tearney Laboratory, Wellman Centre for Photomedicine, Massachusetts General Hospital and Harvard Medical School, Harvard University, USA. He is currently working as an Assistant Professor with the Indian Institute of Information Technology Sri City, Chittoor, India. His research interests include image processing, machine learning, deep learning, optical coherence tomography, and medical image analysis.

ABHISHEK R. KOTHARI completed his postgraduate ophthalmology residency at Sawai Man Singh Medical College, Jaipur, India, in 2006. Currently, he is working as Senior Consultant Vitreoretinal Surgeon with the Pink City Eye and Retina Center, Jaipur, India. He authored several chapters and more than 35 research papers in reputed international journals. His research interests include endophthalmitis, diabetic retinopathy, ophthalmic imaging, and medical economics. He received a fellowship in vitreoretinal surgery from Sankara Nethralaya, Chennai, India. He is the Chief Editor of a textbook on *Vitreoretinal Surgery*.

SHASHIDHAR G. KOOLAGUDI received the Ph.D. degree from IIIT Kharagpur, India, in 2011. Currently, he is working as an Associate Professor with the Department of Computer Science and Engineering, National Institute of Technology Karnataka (NITK), India. He has published more than 140 research papers in reputed international journals and conference proceedings. His research interests include image processing, speech processing, and deep learning.

JENY RAJAN received the Ph.D. degree from the Vision Laboratory, University of Antwerp, Belgium, in 2012. Currently, he is working as an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology Karnataka (NITK), India. He has published more than 75 research papers in reputed international journals and conference proceedings. His research interests include image processing, deep learning, and medical image analysis.