How to Cite:
Nisha, S. S., Meeral, M. N., & Sathik, M. M. (2022). JIEB-linknet: An improved linknet with joint input encoder block for segmentation of retinal layers and fluid accumulation in OCT images. *International Journal of Health Sciences*, *6*(S1), 2072–2096. https://doi.org/10.53730/ijhs.v6nS1.5009

JIEB-LinkNet: An Improved LinkNet with Joint Input Encoder Block for Segmentation of Retinal Layers and Fluid Accumulation in OCT Images

S. Shajun Nisha
Research Supervisor, Assistant Professor & Head, PG & Research Department of Computer Science, Sadakathullah Appa College, Rahmath Nagar, Tirunelveli, India

M. Nagoor Meeral
Ph.D. Research Scholar, PG & Research Department of Computer Science, Sadakathullah Appa College, Rahmath Nagar, Tirunelveli, India. Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, India

M. Mohamed Sathik
Principal, Sadakathullah Appa College, Rahmath Nagar, Tirunelveli, India

***Abstract*---**Age-related Macular Degeneration, Diabetic Retinopathy, Edema as well as Glaucoma are considered as leading ophthalmology diseases result in permanent blindness unless diagnosed earlier. Vision Impairment may be associated with factors such as hereditary, lifestyle and age. Structural changes and fluid formation in the retina helps to investigate the prognosis. Optical coherence Tomography helps to visualize the retinal microstructures in a non-invasive manner. Identification of disease progression is time demanding when these OCT biomarkers are segmented by human experts. Recent findings show that, Semantic segmentation in deep learning is effective for class imbalance problems. This research work proposes an improved LinkNet architecture to delineate retinal layers and fluids concerning OCT images. The primary intention concerning the proposed segmentation network is to categorize 7 retinal layers and fluid as distinct classes. Two Joint input encoder blocks are included in the architecture, which accept multiple inputs from prior layers in order to preserve high resolution characteristics for exact segmentation. In addition, the hybrid loss functions were utilized for improving the prediction accuracy. This model was assessed over the

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
***Corresponding author**: Meeral, M. N.; Email: nmeeral1989@gmail.com*
Manuscript submitted: 09 Nov 2021, Manuscript revised: 27 Feb 2022, Accepted for publication: 18 March 2022

public Duke dataset and the outputs demonstrate that this model attains a dice coefficient of 0.9.

***Keywords*---**deep learning, fluid segmentation, OCT, retinal layer, semantic segmentation.

Introduction

Medical Imaging technologies are beneficial to the clinicians in acquiring quick and deeper insights about human organs through image interpretation. In the modern decades, it is revamping the healthcare sector by introducing non-invasive procedures for several disease diagnoses. Optical coherence tomography (OCT) was typically employed in ophthalmology for examining infections in eyes like Glaucoma, Diabetic Retinopathy etc., also requires no human involvement. With the low coherence interferometer, the light waves enter deep over optical tissues for micrometric distances. Being so, it can capture the cross-sectional portion of the retina at a high acquisition speed [1]. This region comprises ten different layers which can be seen as hyper-reflectivity or hypo-reflectivity in OCT. In case of retinal disorders, these cellular layers exhibit a number of anomalies like yellow deposits for AMD, hyper reflective tissue for Choroid Neo-vascularization, hard exudates for Diabetic Retinopathy, decrease in retinal thickness for Geographic Atrophy, presence of cystoid spaces for Macular Edema [2]. Manual annotation of the retinal layers by ophthalmologists is extremely laborious and time-consuming process to diagnose optical diseases [3],[4].

Hence, a myriad of research works are proposed in the literature which falls under two categories 1. Classical computer vision algorithms and 2. Artificial Intelligence based techniques. The first category involves a) Edge detection algorithms- identifies the edges considering discontinuities in local features like grey level, colour, texture etc., b) Region growing algorithms-groups the pixel regions carrying similar properties like intensity, color etc., c) Clustering techniques- A homogeneous clusters are formed by segmenting the pixel with identical features, d) active contour approach- A deformable spline model, controlled by energy and constraint forces to locate contours in the image and e) graph based segmentation- It models an image as a weighted, undirected graph. The graph is divided into pixel clusters by taking all the similar pixels to account as nodes and their connections between them as edges [4-10].

The second category involves Machine Learning (ML) and Deep Learning (DL) algorithms. ML constructs a mathematical model and makes predictions automatically by learning the given input data. For each labelled training dataset, it extracts image features and applies ML algorithms like (neural network, random forest classifier, K-means clustering, support vector machines etc.) for image segmentation. There are some inadequacies associated with employing the above methods for image segmentation like i) Requires human interference ii) designing handcrafted features iii) parameter generation and updating iv) increasing computational cost for large volume of data etc., [11],[12].

Towards solving these issues, deep learning acts as a great problem solver to automate the segmentation of retinal layers efficiently using multiple processing layers. A plethora of research works based on Convolutional Neural Network like AlexNet [13], VGGNet [14], ResNet [15], architectures have been explored in the recent years. Despite their remarkable success in image classification, it experiences information loss in semantic segmentation. The maxpooling operation generates fuzzy edges thereby directly impact the result consistency. Thus, the classic CNN limitations are surmounted by its improved variants like FCN [16], U-Net [17], LinkNet [18], DeepLab [19] model and yields precise semantic masks with limited training data. Semantic segmentation or segmentation based on pixels, labels all the individual pixels of input images to their associated label categories. This paper introduces an encoder-decoder architecture which incorporates inter skip connections between the downsampling layers. These connections allow the encoder to work with multiple inputs. The significant features in the initial layers are given as input to the following layers for future reusability. Moreover, it reduces the count of convolution layers which greatly reduces training time and learning parameters.

The following are the contributions of proposed work:

- This research explores the connection between semantic segmentation and medical segmentation. While comparing the existing U-Net architectures, the proposed model requires lesser training parameters with high segmentation accuracy. Thus, it is an efficient network for retinal layer classification as well as classification of fluid regions concerning OCT images.
- A novel JIEB-LinkNet on the basis of encoding-decoding framework was suggested. This network employs joint input encoder blocks which accept multiple inputs from prior layers which greatly reduce spatial information loss.
- This model was evaluated over the public DUKE dataset, as well as it scores average Dice co-efficient value of 0.9 and 90% accuracy. These findings reveal that the suggested technique considerably improved segmentation results when assessed with well-evaluated existing algorithms.

Motivation and justification

Semantic segmentation could be considered as a dense prediction strategy that groups similar image pixels to a particular class. In the recent years, Deep learning models are robust in pixel wise segmentation of medical images. Motivated by this, the proposed research work aims to construct a deep learning model, JIEB-LinkNet for segmenting the OCT images semantically.

From the recent research works, it is observed that the encoder decoder architecture is efficient in medical applications. It is capable of learning the feature maps without deformations and loss of image structures. Hence, the proposed model implements the contracting path (encoding section) and expansive path (decoding section) to segment 7 layers of retina as well as fluid accumulation. The performance of JIEB-LinkNet architecture is justified by comparing the conventional methods using the dice co-efficient.

Proposed framework – outline

The overall schematic illustration of this suggested approach is depicted in Fig.1

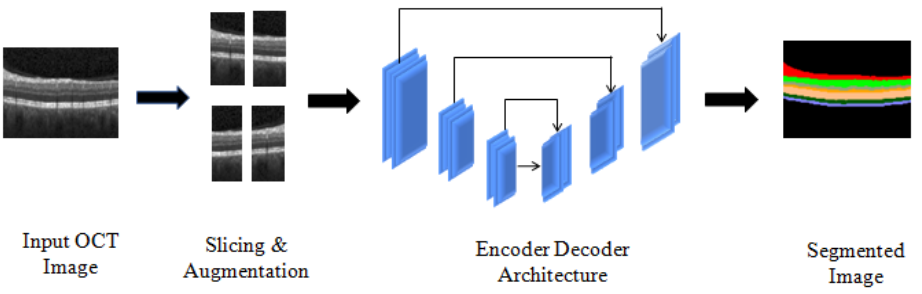


Figure 1. Outline of the research work

Organization of the research

The paper has been organized as given below. Section 2 discusses the analysis of existing literatures, Section 3 elucidates the method and features, Section 4 reveals experimental results and Section 5 concludes this paper.

Literature Review

This section explains two categories of research works 1) Semantic Segmentation 2) Semantic models for Retinal Layer Segmentation.

Semantic segmentation

The Early research works exclusively focused on patch classification in deep learning based image segmentation which has low computational efficiency. Later then, Shelhamer et al [16] formulated the FCN trained on pixel-basis for dense problems. It adapts three networks namely ALexNet, VGGNet and GoogleNet. It replaces the fully connected layer into convolutions to yield output maps from arbitrary sized inputs. The upsampling operation in deconvolution helps to prevent pixel loss which results in high resolution output.

In 2015, Ronneberger et al [17] formulated the U-Net framework based on FCN.U-Net is an encoding-decoding model. A fine segmentation mask is obtained by means of concatenating high resolution features. This is achieved by a skip connection between the contracting path and expansive output, imparts better localization accuracy with limited training time. Chen et al introduced DeepLabv2 with three major contributions. Dilated convolution to improve resolution, Atrous Spatial Pyramid Pooling (ASPP) allows the network to extract multiscale features and finally adopting fully connected CRF (conditional random field) significantly improves boundary localization. Later then, DeepLabv3 is proposed with depth-wise separable convolution [19],[20].

Badrinarayanan et al [21] presented SegNet, a modified version of encoder-decoder architecture. The encoder works on the same principle as VGG16

network. To reconstruct the input image, the max-pool indices value from the contracting path is fed into the decoder during upsampling operations. The foremost objective of this contribution is to minimise memory and computational time. Zhao et al [22] formulated the Pyramid Scene Parsing Networks (PSPNet) endeavouring in preserving contextual information. It encompasses a pyramid pooling module specially designed for extracting the sub-region representations at different scales. These pooled features are finally concatenated to generate final feature map bearing both global and local context information. Chaurasia et al [18] formulated the LinkNet framework on the basis of ResNet18 with one modified encoder operation. Each encoder input is concatenated with the corresponding decoder output to prevent spatial information loss. The network uses limited training parameters and less computational time, making the system more efficient.

Semantic models for retinal layer segmentation

Fang et al [23] implemented the CNN-GS network by integrating Convolutional Neural Network as well as graph searching methods. The probabilistic feature maps from CNN are fed into graph search for post processing to delineate retinal boundaries. Roy et al [24] proposed a ReLayNet model deploying encoder and decoder. The retinal layers and fluid accumulation are segmented using the contextual information obtained from the contracting path. Liu et al [25] proposed the semi-supervised segmentation network with adverse learning named SGNet. Two different layers like global convolutional and boundary refinement layers have been employed for reducing learning parameters as well as for improving the accuracy respectively. Anoop et al [26] implemented the DelNet architecture on the basis of ensemble learning. It encloses 4 basic models subsequently executing the Dilated ReLayNet and prediction blocks. Here, stack generalization merges these feature maps for generating segmented retinal layers.

Wang et al [27] formulated the boundary aware U-Net (BAU-Net) that employs Multi-scale input for improving spatial data. This comprises of one edge aware (EA) element for extracting the edges, U-structure feature enhanced (UFE) element for acquiring the global features and finally one canny edge fusion (CEF) element for merging the outputs and further utilized for better segmentation. Li et al [28] implemented graph convolutional network (GCN) to segment retinal layers as well as optical discs of peripapillary OCT image. Establishing the reasoning block based on graphs helps in extracting ocular global features.

Kugelman et al [29] developed the recurrent neural network on the basis of graph search for delineating the retinal layers from normal and AMD images and it yields better accuracy. Devalla et al [30] implemented the dilated-residual UNet model (DRUNET) for segmenting the ONH tissues. Residual learning and dilated convolutions are enforced for capturing the local and global features efficiently. Liu et al [31] developed a deep learning model with two distinct networks like segmentation and confidence network. The first one is the modified U-Net highly responsible to regulate the topological corrections and the latter one deal with complex regions to obtain accurate segmentation results.

Methodology

Problem statement

The objective of this proposed JIEB-LinkNet was concerned in generating one segmentation mask, where all the pixels x_{ij} of input OCT images I(m,n) is assigned a class label {0,1,2…7} which represents 7 retinal layers and fluid.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \{0,1,2,3,\ldots.7\} \quad where\ x_{ij} \in I(m,n) \tag{1}$$

Encoder-decoder architecture

A typical encoder-decoder architecture contains down sampling (encoder) and upsampling (decoder) towers. The encoder performs convolution followed by batch Normalization and activation functions to extract high level feature maps while reducing the spatial size. Maxpooling operation is performed to downsample the images. On the other side, the upsampling tower restores the feature mappings to its original size using transpose convolution. The skip connection from the encoder ensures fine accurate results. The decoder concatenates both the extracted features and the missing spatial features from the contractor path. Several variants of encoder decoder architectures like LinkNet, ReLayNet are implemented for segmenting medical images. However it requires high training parameters and in case of noise degraded images, it results in low accuracy.

Proposed network architecture

The proposed research work termed JIEB-LinkNet is designed for OCT image segmentation and the architecture is illustrated in Fig 2. It is based on LinkNet in which Semantic segmentation using encoder-decoder architecture is employed in order to attain precise location of retinal layers. The architecture consists of three encoder and three decoder blocks. Both of them repeatedly perform convolution, batch Normalizations as well as ReLu activations. These modules in the proposed framework could be seen from Table 1. Unlike LinkNet, the suggested network introduces Joint Input Encoder Block (JIEB) which accepts multiple inputs. The skip connections between the downsampling layers allow the feature maps generated in the initial layers are propagated as input to the subsequent encoders. As a result, it significantly encourages the reuse of features in order to extract fine-grained information.

The contracting path consists of five layers.

- Convolution Layer: The 256x256 input image is first applied to a convolution layer possessing 64 filters sized 3x3 with stride=2. It results in down sampling the input image to a half scale i.e. 128x128. The operation can be defined on the basis of equation given below,

$$Y_1^{out} = (X * K_1)(i,j) = \sum_m \sum_n X(m,n)K_1(i-m, j-n) \tag{2}$$

Y_1^{out} is layer 1 output, K_1 is Kernel and X(m,n) is the Input Image.

- Maxpooling: This operation is useful to reduce the input dimension. Maxpooling operation is performed with filter size 3x3 and stride 2. It further lowers the input dimension to ¼ scale i.e. 64x64 with deep semantic extracts. The result of the Maxpooling operation is fed into Encoder section. Equation 3 describes the max pooling operation with output Y_2^{out}.

$$Y_2^{out}(i,j) = \max(Y_1^{out}(i-k, j-l)) \quad (3)$$

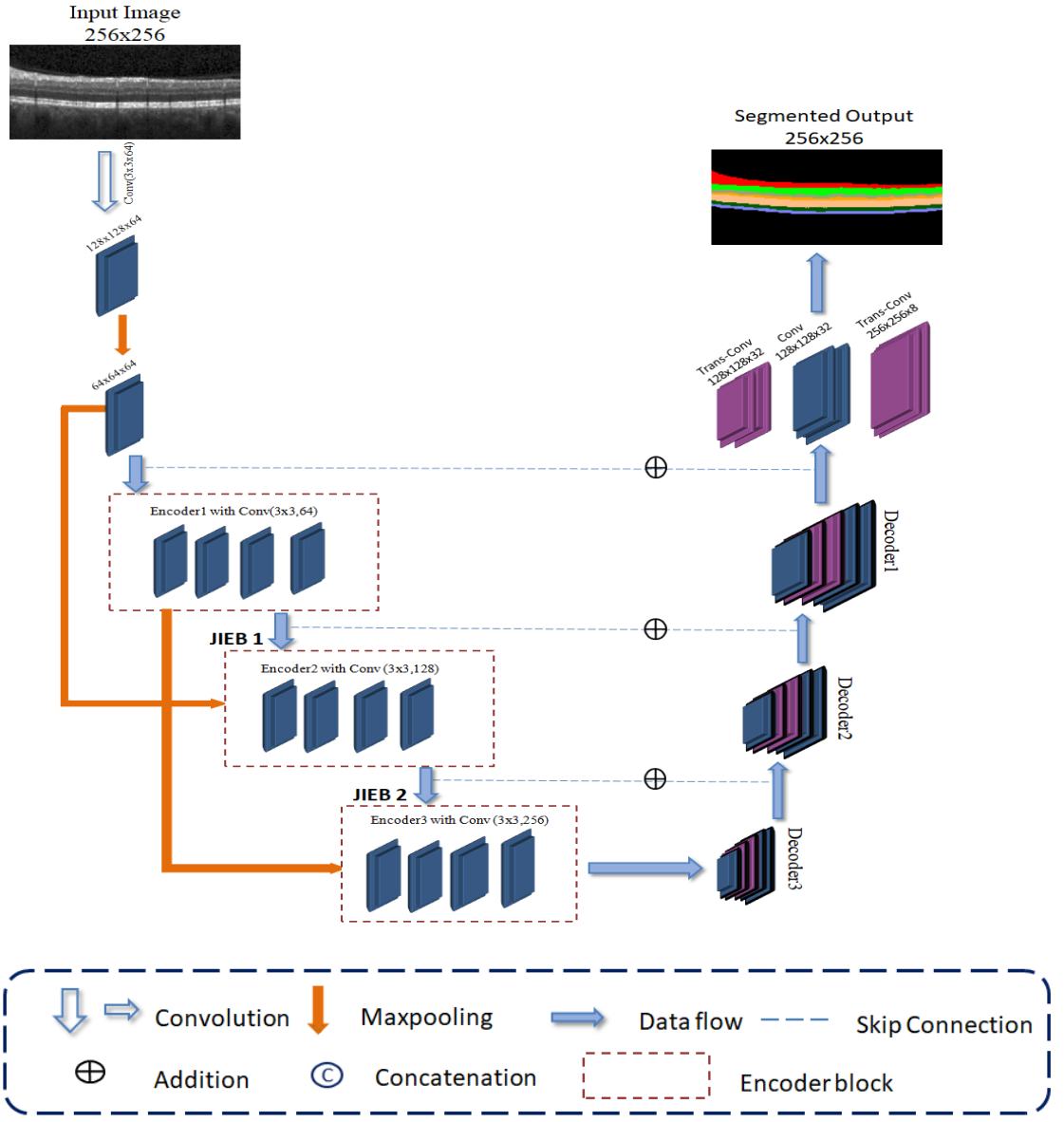


Figure 2. Proposed network architecture

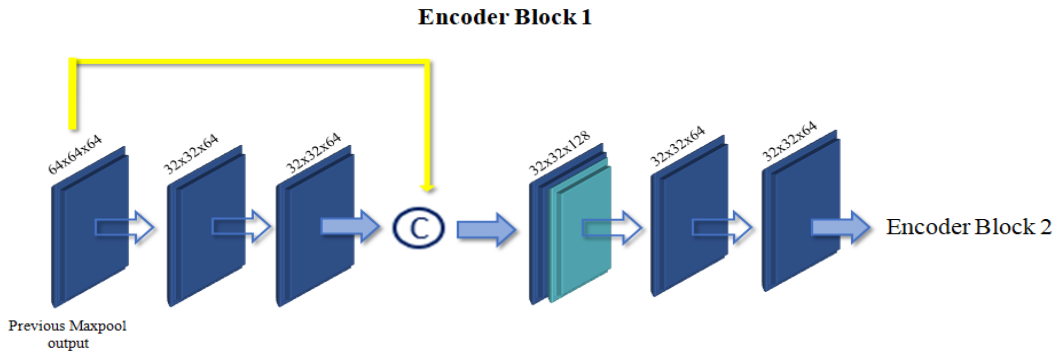


Figure 3(a). Operation in encoder block 1

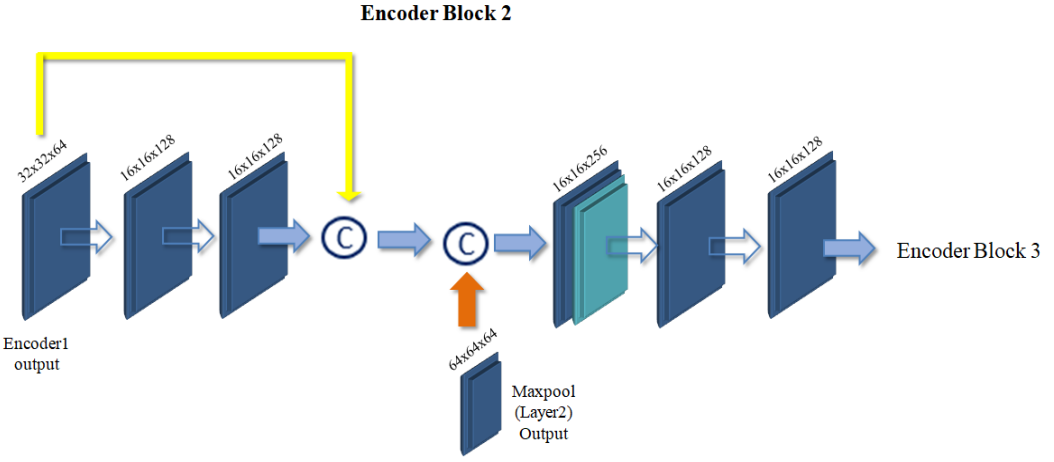


Figure 3(b). Operation in encoder block 2

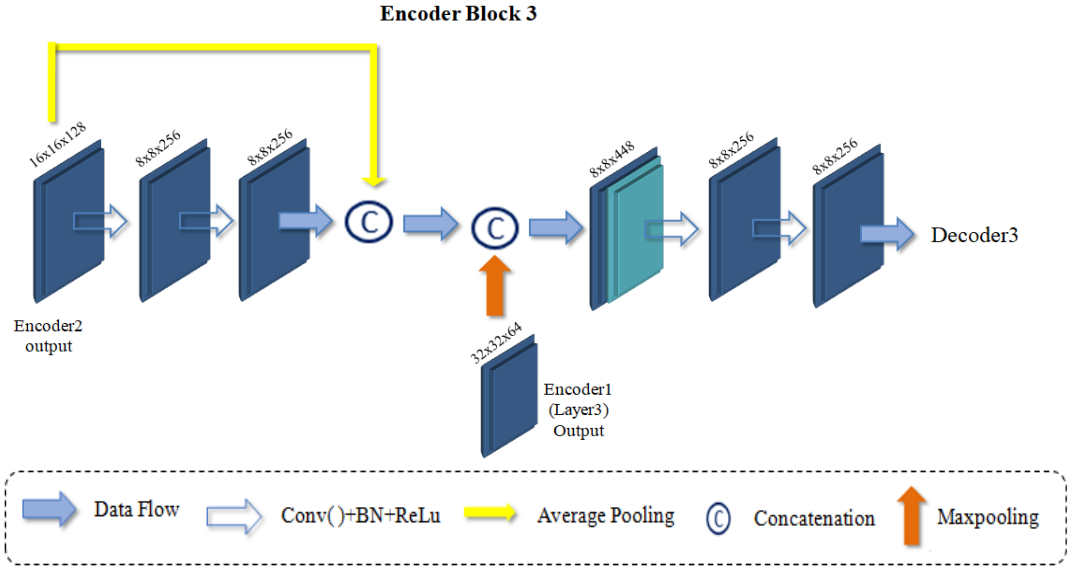


Figure 3(c). Operation in encoder block 3

- Encoder: Totally, three encoders E1, E2, E3 are incorporated in the architecture. Each encoder includes consecutive standard convolution layers which receive input from previous layer. In this context, E2 and E3 are different from E1. E2 and E3 act as a Joint Input Encoder Block (JIEB) as it receives two inputs: one from adjacent layer and another from maxpool value of previous second layer output. Thus E1 receives input from the maxpool layer. Whereas E2 accepts input from both E1 and Maxpool (Maxpool). Finally E3 collects input from E2 and Maxpool (E1). This skip connection with maxpool operation helps for reusing the features to avoid contextual information loss. The operation of E1, E2 and E3 were displayed in Fig.3 (a) (b) and (c).

Table 1
Proposed network modules

Layer	Modules	Filter Count	Output feature map size	Parameters
Layer 1	Conv_1	64	128x128x64	896
Layer 2	Max_pool	-	64x64x64	-
Encoder block 1	Conv_2	64	32x32x64	37,184
	Conv_3	64	32x32x64	37,184
	Avg_pool	-	32x32x64	-
	Concatenation	-	32x32x128	-
Encoder Block 2	Conv_4	64	32x32x64	74,048
	Conv_5	64	32x32x64	37,184
	Conv_6			
	Conv_7	128	16x16x128	72,368
	Avg_pool	128	16x16x128	1,48,096
	Concatenation	-	16x16x64	-
		-	16x16x192	-
	Max_pool	-	16x16x64	-
	Concatenation	-	16x16x256	-
Encoder Block 3	Conv_8	128	16x16x128	2,95,552
	Conv_9	128	16x16x128	1,48,096
	Conv_10			
	Conv_11	256	8x8x256	2,96,192
	Avg_pool	256	8x8x256	5,91,104
	Concatenation	-	8x8x128	-
		-	8x8x384	-
	Max_pool	-	8x8x64	-
	Concatenation	-	8x8x448	-
Decoder block 1	Conv_12	256	8x8x256	1,033,472
	Conv_13	256	8x8x256	5,91,104
	Conv_14	32	8x8x32	8352
	Conv_trans1	-	16x16x32	9376
	Conv_15	256	16x16x256	9472
	Conv_16	256	16x16x256	33,024
	Drop_out	-	16x16x256	-

	Addition	-	16x16x256	-
	Conv_17	16	16x16x16	4176
	Conv_trans2	-	32x32x16	2384
Decoder block 2	Conv_18	128	32x32x128	2688
	Conv_19	128	32x32x128	8320
	Drop_out	-	32x32x128	-
	Addition	-	32x32x128	-
	Conv_173	16	32x32x16	2128
	Conv_trans2	-	64x64x16	2320
Decoder block 3	Conv_18	64	64x64x64	1344
	Drop_out	-	64x64x64	-
	Addition	-	64x64x64	-
Pixel	Conv_trans4	32	128x128x32	18,592
Classification	Conv_19	32	128x128x32	9376
Layer	Conv_trans5	8	65536x8	1032

Encoder block 1

It accepts input from maxpool layers. 3x3 Convolution with 64 filters and stride 2 is performed in E1 which downsamples the input size to 1/8. The input is given to the two convolutional layers. A short skip connection from the encoder 1 input concatenates with these convolution results. This shortcut link helps to preserve local features and the average pooling is used for matching dimensions. The output is passed into the two subsequent convolution layers (3x3, 64, 2). Y_3^{out} is the output of Encoder 1 and the operation is shown in equation 4.

$$Y_3^1(i,j) = Y_2^{out}(i,j) * K_2(i,j)$$

$$Y_3^2(i,j) = Y_3^1(i,j) * K_2(i,j)$$

$$Y_3^3(i,j) = Y_3^2(i,j) \oplus avg(Y_2^{out}(i,j))$$

$$Y_3^4(i,j) = Y_3^3(i,j) * K_2(i,j)$$

$$\boldsymbol{Y_3^{out}(i,j) = Y_3^4(i,j) * K_2(i,j)} \quad (4)$$

Encoder block 2

It receives two inputs. 3x3 Convolution with 128 filters and stride 2 are performed in E2 reducing the input dimension to 1/16. Once the input data is obtained from E1, it is processed in two convolution layers. Then the result is concatenated with second input (skip connection from E1 output). Then it again gets concatenated with the third input: maxpool value of Maxpool layer. This combined output will be fed into the two convolution layers (3x3, 128, 2) and (3x3, 128). Y_4^{out} is the output of Encoder 2 and the operation is shown in equation.

$$Y_4^1(i,j) = Y_3^{out}(i,j) * K_3(i,j)$$

$$Y_4^2(i,j) = Y_4^1(i,j) * K_3(i,j)$$

$$Y_4^3(i,j) = Y_4^2(i,j) \oplus avg(Y_3^{out}(i,j))$$

$$Y_4^4(i,j) = Y_4^3(i,j) \oplus Max(Y_2^{out}(i,j))$$

$$Y_4^5(i,j) = Y_4^4(i,j) * K_3(i,j)$$

$$\boldsymbol{Y_4^{out}(i,j) = Y_4^5(i,j) * K_3(i,j)} \quad (5)$$

Encoder block 3

In this JIEB E3, 3x3 Convolution with 256 filters and stride 2 are performed to downscale the input image to $1/32$. The convolution layer accept E2 output and it is concatenated with two inputs (one from E2 output and other from Maxpool (E1) output) followed by two convolution layers (3x3, 256, 2) and (3x3, 256). Y_5^{out} is the output of Encoder 3 and the operation is shown in equation 6.

$$Y_5^1(i,j) = Y_4^{out}(i,j) * K_4(i,j)$$

$$Y_5^2(i,j) = Y_5^1(i,j) * K_4(i,j)$$

$$Y_5^3(i,j) = Y_5^2(i,j) \oplus avg(Y_4^{out}(i,j))$$

$$Y_5^4(i,j) = Y_5^3(i,j) \oplus Max(Y_3^{out}(i,j))$$

$$Y_5^5(i,j) = Y_5^4(i,j) * K_4(i,j)$$

$$\boldsymbol{Y_5^{out}(i,j) = Y_5^5(i,j) * K_4(i,j)} \quad (6)$$

Decoder

The decoder section is similar as in LinkNet model. The decoder block is responsible for upsampling the feature maps. There are three decoders in which each comprises three operations, two convolutions and one Transpose convolution. To perform the transpose convolutional operation, the singular value done from lower-resolution feature maps were selected and multiplied with every weights of the filter. Then, it projects the weight values to feature maps at output. The long skip connections from the encoder block gets combined with the decoder block and further proceeds for upsampling operation to produce output similar to input size. The operation of Decoder is shown in Fig.4.

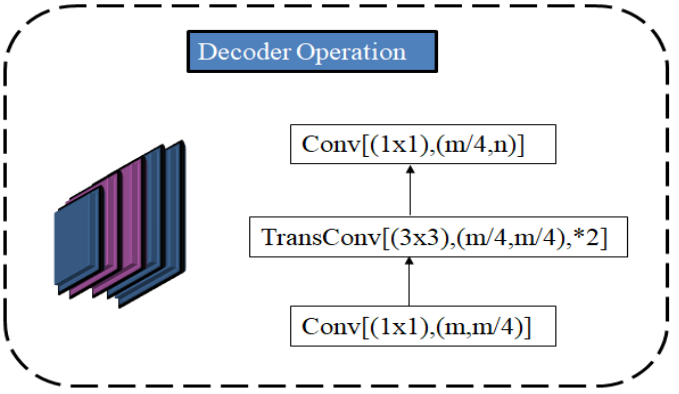


Figure 4. Decoder operation

Pixel classification layer

This layer helps to predict probability map of appropriate classes. The softmax classifier will generate the mask of corresponding output labels of each class.

Loss function

While training the data, Cost function for each class is measured among the predicted maps as well as with the corresponding ground truth values. The proposed network is optimized by combining dice losses as well as categorical cross entropy losses as recommended in [24], [32].

$$Loss_{seg} = Loss_{Dice} + Loss_{cce}$$

Dice Loss

It calculates the similarity amid predicted images as well as original samples. Better the dice coefficient is, higher will be the efficiency of training model. Thus, the dice value of 1 indicates the accurate segmentation. Dice Loss for class C with probability map M and ground truth Y can be expressed as.

$$Loss_{Dice} = 1 - \frac{2\sum_{i=1}^{n}\sum_{j=1}^{C} G_{i,j}M_{i,j} + \emptyset}{\sum_{i=1}^{n}\sum_{j=1}^{C}(G_{i,j} + M_{i,j}) + \emptyset}$$

Here, n represents Number of Pixels and ∅ is the smoothing factor.Dice coefficient is adept for data imbalance problems. However, it offers minimal optimization in some cases and delayed convergence.

Categorical cross entropy loss

This loss function is exploited for classification problems which has one hot encoding output (1's and 0's) for different class. It is used to calculate the dissimilarity between two different probability distributions. The equation of categorical cross entropy is defined as,

$$Loss_{cce} = -\sum_{i=1}^{N} y_i . \log(\hat{y}_i)$$

Here, N represents output classes, $y_i$corresponds to true distributions, $\widehat{y_i}$ represents target probability distributions.

Network training

Dataset generation

The performances of this proposed JIEB-LinkNet have been assessed over DUKE public dataset. It is obtained using Spectralis imaging (Heidelberg Engineering, Heidelberg, Germany). The dataset consists of 10 DME pathological subjects in different qualities. Each subject encloses 11 B scans and totally 110 images are available with image size 512x740. The retinal layers and the fluid region in the images are annotated by two manual experts. For Network training, 1-5 subjects are used and 6-10 subjects are applied for testing [33].

Pre-processing

Before training the data, mxn input OCT B-scan is divided into K non-overlapping portions. To avoid over fitting and to improve model performance with limited data, the OCT slices are involved for augmentations like horizontal flip to change the orientation and random cropping. These changes in the images help to increase the training dataset size to a required factor.

Experimental setup

The experiment is conducted in system with configuration of Intel (R) UHD Graphics 630, NVIDIA GeForce GTX 1050 and 8 GB RAM. The training image input size is reshaped to 256x256 for semantic segmentation. While training the model, the batch size has been fixed as 32 as well as the maximum of 200 epochs is required. Initial learning rate 0.001 is used with RMS prop optimizer [18]. The hybrid loss function along with weight decay of 0.0001 is employed to adjust the optimizer. It takes 6 hours 25 minutes to complete the training process with 3.47M parameters.

Experimental Results

Performance metrics

The proposed algorithm concerning segmentations of OCT images were assessed based on Dice Coefficients, Accuracy, Precision, Recall, F-score as well as Intersection over Union (IoU) as employed in [34]. The findings were assessed over suitable ground truth interpreted by the researchers previously. Dice coefficient is the statistical calculation of similarity between true and predicted images.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

Accuracy estimates the percentage of pixels correctly classified to the appropriate class.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Number\ of\ Pixels}$$

Precision is calculated by the ratio between correctly predicted true value and total actual positives.

$$Precision = \frac{True\ positive}{True\ positive + False\ Positive}$$

Recall or sensitivity calculates the true positive predictions. The model is best if recall value is high.

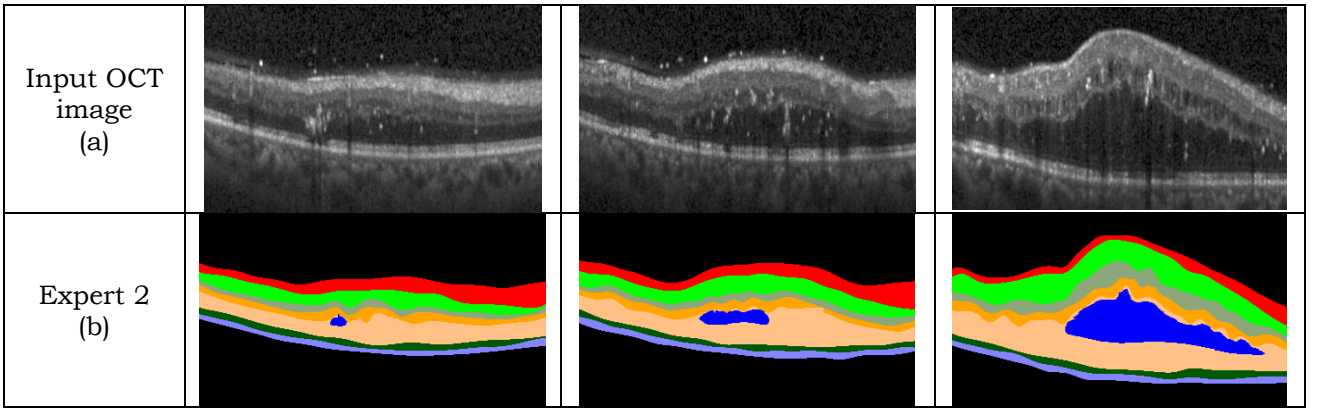
$$Recall = \frac{True\ positive}{True\ positive + False\ Negative}$$

F score simply measure the harmonic average value between precision and recall. If these values are high then F score also provides high value.

$$\text{F score} = 2x\frac{\text{Precision x recall}}{\text{Precision} + \text{recall}}$$

IoU is used to calculate the overlap between ground truth (X) and segmented image (Y). The greater IoU value indicates the better segmentation.

$$IoU = \frac{X \cap Y}{X \cup Y}$$



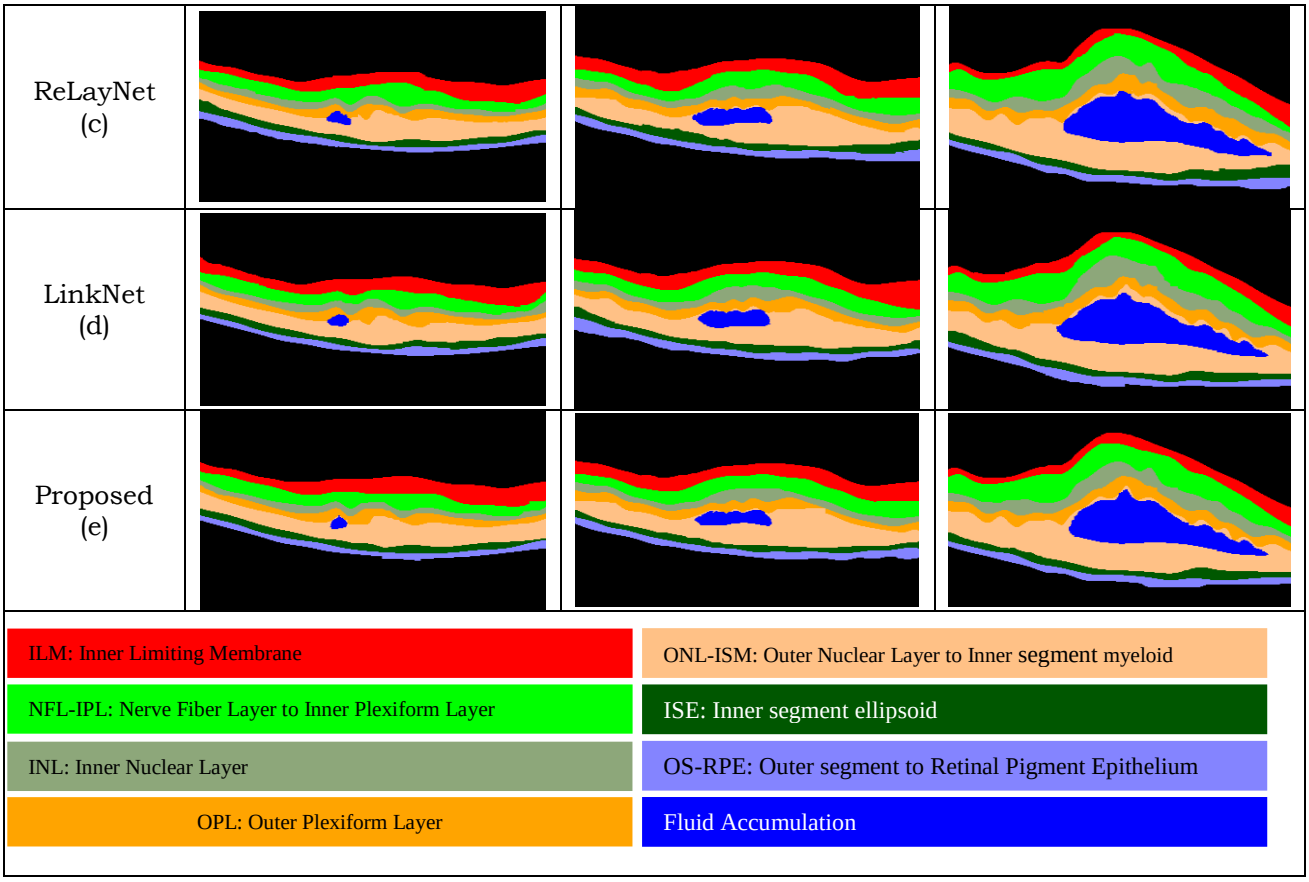


Figure 5. Qualitative assessment of the proposed model with existing techniques and expert 2 annotated images

Result and Discussion

This suggested technique was qualitatively assessed over ReLayNet and LinkNet models in Fig.5. Input OCT Images were presented in Fig.5(a). The expert 2 annotation from duke images were exposed in Fig 5(b). The segmented maps of ReLayNet, LinkNet and proposed method was illustrated in respective Fig 5(c) (d) and (e). From the illustration, it is observed that the proposed model segments the retinal layers and fluid region better than the other methods. There are slight misclassifications of boundaries in other networks than the proposed methodology.

To measure the quantitative analysis of proposed model, the performance metrics like Dice coefficient, Accuracy, Precision, Recall, F score and IoU are taken. Table 2 displays the assessment of this proposed strategy over the traditional methods namely (1) Expert 2 [24], (2) CM-GDP [35], (3) CM-KR [33], (4) CM-LSE [36], (5) CM-Unet [17], (6) CM-FCN [16], (7) ReLayNet [24], (8) LinkNet [18] in terms of dice coefficient for different classes. The corresponding graphical representation is illustrated in Fig.6. It is identified that this suggested methodology displays improved outcomes when associated to other existing methodologies. LinkNet

shows second better result than others. For NFL-IPL, the proposed model and ReLayNet shows similar results. The ONL-ISM layer is predicted similarly by expert 2 and CM-LSE. In OSE-RPE layer, ReLayNet, LinkNet and proposed method shows same outcome. When comparing with Expert 2 annotation, the proposed methods result in high dice coefficient for 7 predicted maps except ONL-ISM layer. ReLayNet, LinkNet competes finely with proposed model. Hence, these methods are compared with proposed one for metrics like Accuracy, Precision, Recall, F score and IoU. Table 3 illustrates comparison of ReLayNet, LinkNet and Proposed methods with respect to accuracy, which reveals that this suggested method shows better accuracy results for each class. Table 4, 5, 6 and 7 shows the predictions of proposed network in terms of Precision, Recall, F score, and IoU concerning eight classes. The overall segmentation performances of the comparison methods were displayed in Table 8 and it reveals that this proposed methodology achieves Precision 89.6%, Recall 94.5%, F score 91.5%, IoU of 0.75 and dice 0.9. Figs.7 & 8 represent the corresponding graphical representation in comparing the overall performances concerning this proposed method. The training as well as validation accuracies of LinkNet and proposed methodology were displayed in Fig.9 (a) & (b). Likewise, training loss as well as validation loss was displayed in Fig.9(c) & 9(d). Considering all these observations, it could be understood that this proposed architecture performs superior concerning OCT retinal layer and Fluid segmentation.

Table 2
Assessment of proposed technique over prevailing techniques with respect to dice coefficient

	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
Expert 2 [24]	0.86	0.9	0.79	0.74	0.94	0.86	0.82	0.58
CM-GDP [35]	0.77	0.77	0.65	0.67	0.86	0.87	0.82	NA
CM-KR [33]	0.85	0.89	0.75	0.74	0.93	0.87	0.82	0.53
CM-LSE [36]	0.87	0.9	0.8	0.77	0.94	0.88	0.86	NA
CM-Unet [17]	0.86	0.91	0.83	0.81	0.91	0.9	0.83	0.67
CM-FCN [16]	0.81	0.84	0.72	0.71	0.88	0.89	0.86	0.28
ReLayNet [24]	0.9	0.94	0.87	0.84	0.93	0.92	0.9	0.77
LinkNet [18]	0.91	0.91	0.87	0.83	0.9	0.91	0.9	0.8
Proposed JIEB-LinkNet	0.94	0.94	0.9	0.85	0.94	0.94	0.9	0.82

Table 3
Assessment of proposed technique over prevailing techniques with respect to accuracy

		ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
Accuracy (%)		0	1	2	3	4	5	6	7
	ReLayNet	94	95	92	92	93	92	93	92
	LinkNet	96	96	93	93	95	95	94	93
	Proposed JIEB-LinkNet	98	98	96	95	97	98	96	95

Table 4
Assessment of proposed technique over prevailing techniques with respect to precision

Precision (%)	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
	0	1	2	3	4	5	6	7
ReLayNet	92	86	90	83	80	91	86	84
LinkNet	93	87	92	84	81	93	88	86
Proposed JIEB-LinkNet	95	89	93	86	83	95	90	88

Table 5
Assessment of proposed technique over prevailing techniques with respect to recall

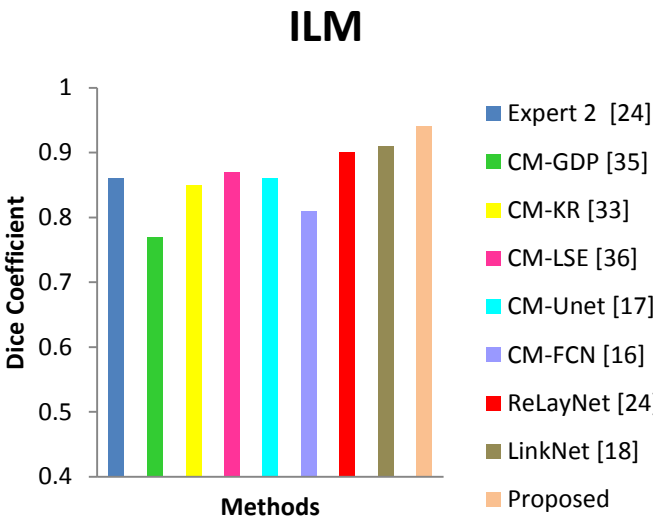
Recall (%)	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
	0	1	2	3	4	5	6	7
ReLayNet	91	91	91	89	85	93	93	91
LinkNet	93	92	93	90	87	94	94	93
Proposed JIEB-LinkNet	96	94	96	92	89	97	96	95

Table 6
Assessment of proposed technique over prevailing techniques with respect to F score

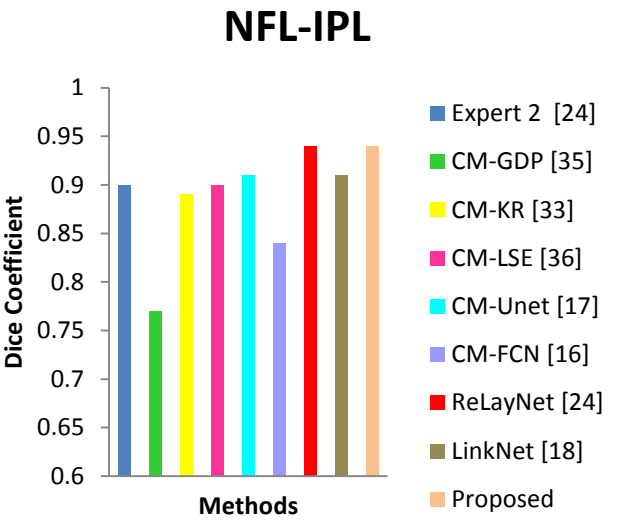
F score (%)	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
	0	1	2	3	4	5	6	7
ReLayNet	91	87	89	83	81	91	87	87
LinkNet	93	89	91	85	83	93	90	88
Proposed JIEB-LinkNet	95	91	94	88	85	96	92	91

Table 7
Assessment of proposed technique over prevailing techniques with respect to IoU

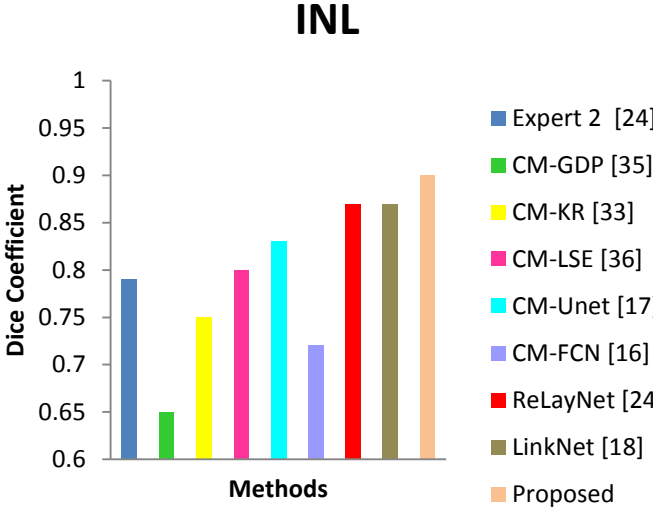
IoU	ILM	NFL-IPL	INL	OPL	ONL-ISM	ISE	OS-RPE	Fluid
	0	1	2	3	4	5	6	7
ReLayNet	0.89	0.69	0.77	0.61	0.55	0.84	0.73	0.69
LinkNet	0.90	0.70	0.78	0.63	0.57	0.85	0.74	0.70
Proposed JIEB-LinkNet	0.93	0.71	0.81	0.65	0.58	0.87	0.75	0.71


ILM
Dice Coefficient
1
0.9
0.8
0.7
0.6
0.5
0.4
Methods
Expert 2 [24]
CM-GDP [35]
CM-KR [33]
CM-LSE [36]
CM-Unet [17]
CM-FCN [16]
ReLayNet [24
LinkNet [18]
Proposed

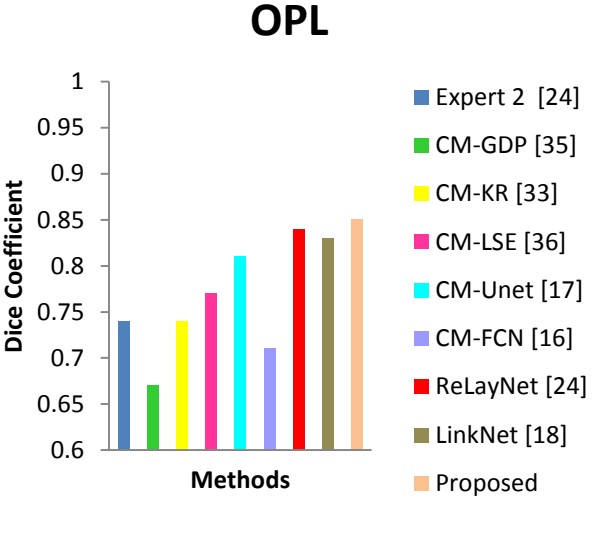
(a)


NFL-IPL
Dice Coefficient
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
Methods
Expert 2 [24]
CM-GDP [35]
CM-KR [33]
CM-LSE [36]
CM-Unet [17]
CM-FCN [16]
ReLayNet [24]
LinkNet [18]
Proposed

(b)


INL
Dice Coefficient
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
Methods
Expert 2 [24]
CM-GDP [35]
CM-KR [33]
CM-LSE [36]
CM-Unet [17
CM-FCN [16]
ReLayNet [24
LinkNet [18]
Proposed

(c)


OPL
Dice Coefficient
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
Methods
Expert 2 [24]
CM-GDP [35]
CM-KR [33]
CM-LSE [36]
CM-Unet [17]
CM-FCN [16]
ReLayNet [24]
LinkNet [18]
Proposed

(d)

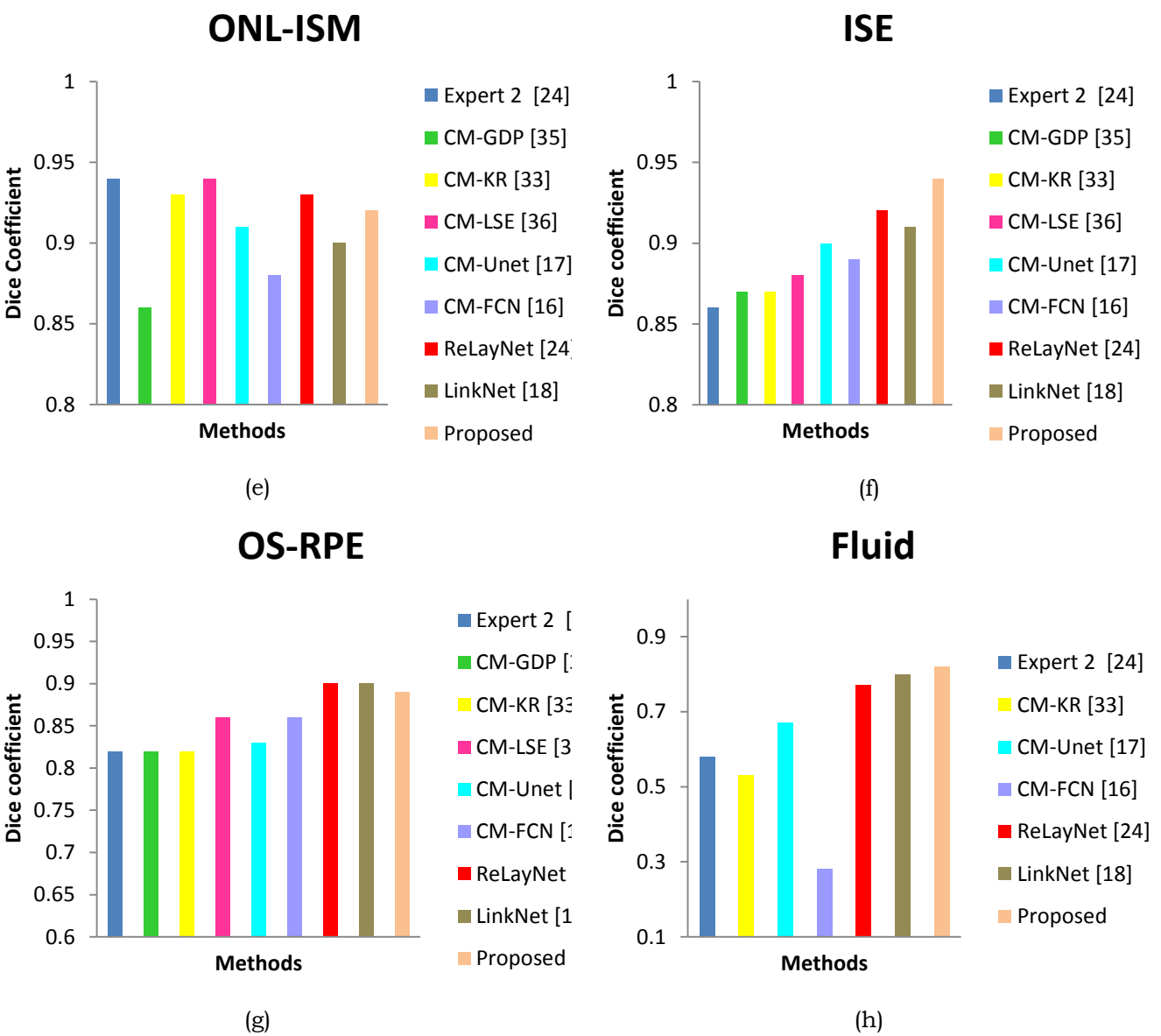


Figure 6. Graphical Representation of comparison of different methods like Expert 2 [24], CM-GDP [35], CM-KR [33], CM-LSE [36], CM-Unet [17], CM-FCN [16], ReLayNet [24], LinkNet [18] and Proposed in terms of Dice Coefficient for Retinal Layers (a) ILM (b) NFL-IPL (c) INL (d) OPL (e) ONL-ISM (f) ISE (g) OS-RPE and (h) Fluid

Table 8
Overall Segmentation performance of relaynet, linknet and Proposed methods

	Accuracy (%)	Recall (%)	Precision (%)	Fscore (%)	IoU	Dice
ReLayNet	92.8	90.5	86.5	87	0.72125	0.88375
LinkNet	94.3	92	88	89	0.73375	0.8845

Proposed JIEB-LinkNet	96	94.5	89.6	91.5	0.750498	0.9

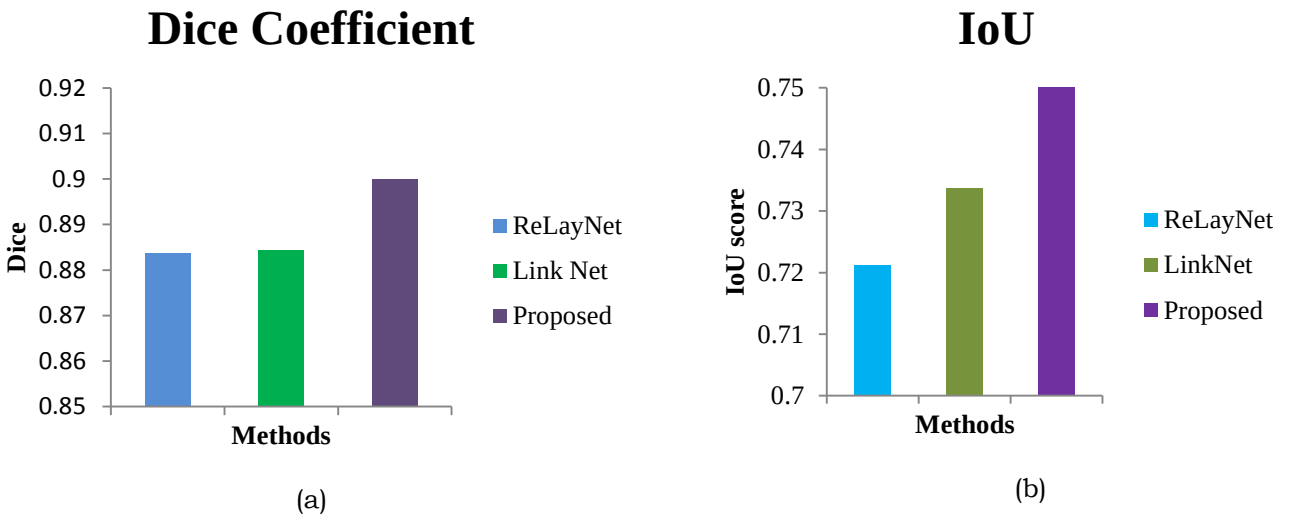


Figure 7. Graphical representation of dice coefficient and iou over relaynet, linknet and proposed method

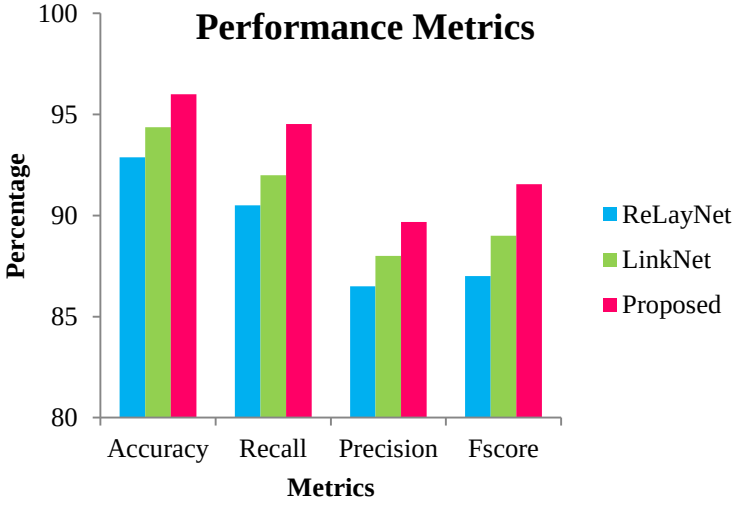


Figure 8. Graphical Representation of accuracy, recall, precision and f score over relaynet, linknet and proposed method

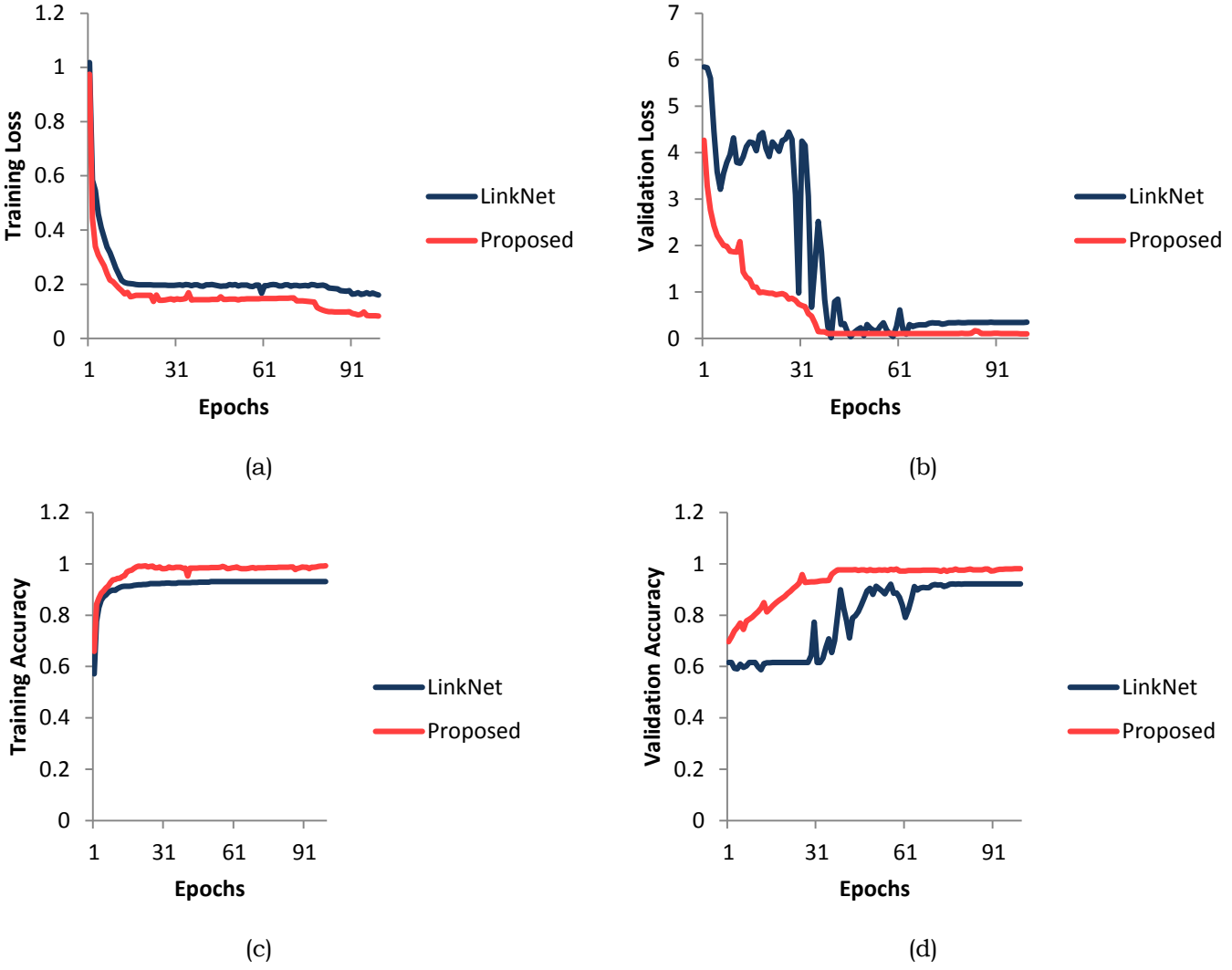


Figure 9. Assessment of (a) Training Losses, (b) Validation Losses, (c) Training Accuracy, (d) Validation Accuracy for LinkNet as well as the Proposed Method

Conclusions

This research work proposes a JIEB-LinkNet for segmenting 7 retinal layers and also fluid masses from OCT B-scan. The training and testing is carried on DUKE dataset with OCT scans of 10 patients which is available publicly. The proposed work introduces Joint Input Encoder Blocks and highlights the advantages of skip connections between the encoder blocks for better feature preservation and improvement in training accuracy. Reducing the count of Convolutional layers results this model in learning with only 3.47M training parameters. This model is optimized using dice and categorical cross entropy losses. The performance of this novel JIEB-LinkNet model was assessed with conventional well-evaluated

techniques and it achieves dice coefficient of 0.9 and 96% accuracy, 89.6% precision, 94.5% recall, 91.5% F score and 0.75 of IoU. The proposed work attains better results in quantitative as well as qualitative manner. In the future work, the proposed network shall be implemented for classification of OCT images for identifying retinal diseases.

References

1. Kramoreva, L. I., & Rozhko, Y. I. (2010). Optical coherence tomography (Review). Journal of Applied Spectroscopy, 77(4), 449–467. doi:10.1007/s10812-010-9354-0.
2. Bhende M, Shetty S, Parthasarathy MK, Ramya S. (2018). Optical coherence tomography: A guide to interpretation of common macular diseases. Indian J Ophthalmol. 66(1):20-35. doi:10.4103/ijo.IJO_902_17.
3. Kafieh, R., Rabbani, H., Abramoff, M. D., & Sonka, M. (2013). Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. Medical Image Analysis, 17(8), 907–928. doi:10.1016/j.media.2013.05.006.
4. Chiu, S. J., Li, X. T., Nicholas, P., Toth, C. A., Izatt, J. A., & Farsiu, S. (2010). Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. Optics Express, 18(18), 19413. doi:10.1364/oe.18.019413.
5. Luo, S., Yang, J., Gao, Q., Zhou, S., & Zhan, C. A. (2017). The Edge Detectors Suitable for Retinal OCT Image Segmentation. Journal of Healthcare Engineering, 2017, 1–13. doi:10.1155/2017/3978410.
6. Abhishek, A. M., Berendschot, T. T. J. M., Rao, S. V., & Dabir, S. (2014). Segmentation and analysis of retinal layers (ILM & RPE) in Optical Coherence Tomography images with Edema. 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES). doi:10.1109/iecbes.2014.7047486.
7. Shijian Lu, Jiang Liu, Joo Hwee Lim, Carol Cheung, & Tien Yin Wong. (2010). Automated layer segmentation of optical coherence tomography images. 2010 5th IEEE Conference on Industrial Electronics and Applications. doi:10.1109/iciea.2010.5515505.
8. Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. International Journal of Computer Vision, 1(4), 321–331. doi:10.1007/bf00133570.
9. Dodo, B.; Li, Y.; Eltayef, K. and Liu, X. (2018). Graph-Cut Segmentation of Retinal Layers from OCT Images. In Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOIMAGING, ISBN 978-989-758-278-3; ISSN 2184-4305, pages 35-42. DOI: 10.5220/0006580600350042.
10. Mayer, M. A., Tornow, R. P., Hornegger, J., & Kruse, F. E. (2008). Fuzzy C-means clustering for retinal layer segmentation on high resolution OCT images. In 19th Biosignal Conf.
11. Dodo, B. I., Li, Y., Kaba, D., & Liu, X. (2019). Retinal Layer Segmentation in Optical Coherence Tomography Images. IEEE Access, 7, 152388–152398. doi:10.1109/access.2019.2947761.
12. Antony, B. J., Abràmoff, M. D., Harper, M. M., Jeong, W., Sohn, E. H., Kwon, Y. H., Garvin, M. K. (2013). A combined machine-learning and graph-

based framework for the segmentation of retinal surfaces in SD-OCT volumes. Biomedical Optics Express, 4(12), 2712. doi:10.1364/boe.4.002712.
13. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. doi:10.1145/3065386.
14. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409.1556, 2014.
15. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.90.
16. Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4), 640–651. doi:10.1109/tpami.2016.2572683.
17. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 234–241. doi:10.1007/978-3-319-24574-4_28.
18. Chaurasia, A., & Culurciello, E. (2017). LinkNet: Exploiting encoder representations for efficient semantic segmentation. 2017 IEEE Visual Communications and Image Processing (VCIP). doi:10.1109/vcip.2017.8305148.
19. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. doi:10.1109/tpami.2017.2699184.
20. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Lecture Notes in Computer Science, 833–851. doi:10.1007/978-3-030-01234-2_49.
21. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. doi:10.1109/tpami.2016.2644615.
22. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.660.
23. Fang, L., Cunefare, D., Wang, C., Guymer, R. H., Li, S., & Farsiu, S. (2017). Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. Biomedical Optics Express, 8(5), 2732. doi:10.1364/boe.8.002732.
24. Roy, A. G., Conjeti, S., Karri, S. P. K., Sheet, D., Katouzian, A., Wachinger, C., & Navab, N. (2017). ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. Biomedical Optics Express, 8(8), 3627. doi:10.1364/boe.8.003627.
25. Liu, X., Cao, J., Fu, T., Pan, Z., Hu, W., Zhang, K., & Liu, J. (2018). Semi-supervised Automatic Segmentation of Layer and fluid region in Retinal Optical Coherence Tomography Images Using Adversarial Learning. IEEE Access, 1–1. doi:10.1109/access.2018.2889321.

26. Anoop, B. N., Pavan, R., Girish, G. N., Kothari, A., & Rajan, J. (2020). Stack generalized deep ensemble learning for retinal layer segmentation in Optical Coherence Tomography images. Biocybernetics and Biomedical Engineering. doi:10.1016/j.bbe.2020.07.010.
27. Wang, B., Wei, W., Qiu, S., Wang, S., Li, D., & He, H. (2021). Boundary Aware U-Net for Retinal Layers Segmentation in Optical Coherence Tomography Images. IEEE Journal of Biomedical and Health Informatics, 25(8), 3029–3040. doi:10.1109/jbhi.2021.3066208.
28. Jiaxuan Li, Peiyao Jin, Jianfeng Zhu, Haidong Zou, Xun Xu, Min Tang, Minwen Zhou, Yu Gan, Jiangnan He, Yuye Ling, and Yikai Su (2021). Multi-scale GCN-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary OCT images. Biomed. Opt. Express 12, 2204-2220.
29. Kugelman, J., Alonso-Caneiro, D., Read, S. A., Vincent, S. J., & Collins, M. J. (2018). Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search. Biomedical Optics Express, 9(11), 5759. doi:10.1364/boe.9.005759.
30. Devalla, S. K., Renukanand, P. K., Sreedhar, B. K., Subramanian, G., Zhang, L., Perera, S., … Girard, M. J. A. (2018). DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. Biomedical Optics Express, 9(7), 3244. doi:10.1364/boe.9.003244.
31. Liu, X., Cao, J., Wang, S., Zhang, Y., & Wang, M. (2021). Confidence-Guided Topology-Preserving Layer Segmentation for Optical Coherence Tomography Images With Focus-Column Module. IEEE Transactions on Instrumentation and Measurement, 70, 1–12. doi:10.1109/tim.2020.3047430.
32. Reddy TGP, Ashritha KS, Prajwala T, Girish G, Kothari AR, Koolagudi SG, et al. (2020) Retinal-layer segmentation using dilated convolutions. Proceedings of 3rd International Conference on Computer Vision and Image Processing. p. 279–92.
33. Chiu, S. J., Allingham, M. J., Mettu, P. S., Cousins, S. W., Izatt, J. A., & Farsiu, S. (2015). Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. Biomedical Optics Express, 6(4), 1172. doi:10.1364/boe.6.001172.
34. Hassan, B., Qin, S., Hassan, T., Ahmed, R., & Werghi, N. (2021). Joint Segmentation and Quantification of Chorioretinal Biomarkers in Optical Coherence Tomography Scans: A Deep Learning Approach. IEEE Transactions on Instrumentation and Measurement, 70, 1–17. doi:10.1109/tim.2021.3077988.
35. Chiu, S. J., Li, X. T., Nicholas, P., Toth, C. A., Izatt, J. A., & Farsiu, S. (2010). Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. Optics Express, 18(18), 19413. doi:10.1364/oe.18.019413.
36. Karri, S. P. K., Chakraborthi, D., & Chatterjee, J. (2016). Learning layer-specific edges for segmenting retinal layers with large deformations. Biomedical Optics Express, 7(7), 2888. doi:10.1364/boe.7.002888.
37. Kumar, S. (2022). Strategic management of carbon footprint using carbon collectible non-fungible tokens (NFTS) on blockchain. Academy of Strategic Management Journal, 21(S3), 1-10

38. Kumar, S. (2021). Review of geothermal energy as an alternate energy source for Bitcoin mining. Journal of Economics and Economic Education Research, 23(1), 1-12
39. Ritika Malik, Aarushi Kataria and Naveen Nandal, Analysis of Digital Wallets for Sustainability: A Comparative Analysis between Retailers and Customers, International Journal of Management, 11(7), 2020, pp. 358-370.
40. Aarushi, Naveen Nandal, Parul Agrawal. AN EXPLORATORY RESEARCH IN PRODUCT INNOVATION IN AUTOMOBILE SECTOR. JCR. 2020; 7(2): 522-529. doi:10.31838/jcr.07.02.98