



# Deep Learning Based Fully Automated Decision Making for Intravitreal Anti-VEGF Therapy

Simran Barnwal<sup>1</sup>(✉), Vineeta Das<sup>2</sup>, and Prabin Kumar Bora<sup>2</sup>

<sup>1</sup> Department of Physics, Indian Institute of Technology Guwahati, Guwahati, India  
simranbarnwal.iitg@gmail.com

<sup>2</sup> Department of Electronics and Electrical Engineering,  
Indian Institute of Technology Guwahati, Guwahati, India

**Abstract.** Intraocular anti-vascular endothelial growth factor (VEGF) therapy is the most significant treatment for vascular and exudative diseases of the retina. The highly detailed views of the retina provided by optical coherence tomography (OCT) scans play a significant role in the proper administration of anti-VEGF therapy and treatment monitoring. With increasing cases of visual impairment worldwide, computer-aided diagnosis of retinal pathologies is the need of the hour. Recent research on OCT-based automatic retinal disease detection has focused on using the state-of-the-art deep convolutional neural network (CNN) architectures due to their impressive performance in image classification tasks. However, these architectures are large in size and take significant time during testing, thus limiting their deployment to machines with ample memory and computation power. This paper proposes a novel deep learning based OCT image classifier, utilizing a small CNN architecture named as SimpleNet. It provides better classification accuracy with 800x fewer parameters, 350x less memory requirement, and is 50x faster during testing compared to state-of-the-art deep CNNs. Unlike other papers focusing on the prediction of specific diseases, we focus on broadly classifying OCT images into needing anti-VEGF therapy, needing simple routine care or normal healthy retinas.

**Keywords:** Deep learning · Optical coherence tomography · Anti-VEGF therapy

## 1 Introduction

The macula, located in the central part of the retina, is the primary sensory region associated with vision. Retinal pathologies like choroidal neovascularization (CNV), diabetic retinopathy (DR), age-related macular degeneration (AMD) and diabetic macular edema (DME) affect macular health. These diseases deteriorate normal vision and are also the primary causes of irreversible

vision impairment across the world [14]. Intraocular administration of anti-vascular endothelial growth factor (anti-VEGF) has become the standard treatment modality for these vision-threatening diseases [2]. Retinal optical coherence tomography (OCT) imaging provides highly detailed views of the morphology of the retina and is being widely used by retinal specialists in the decision making and treatment monitoring process of anti-VEGF indications [1]. A 3D-OCT volume may contain as many as 1600 B-scans [4] and individually interpreting each scan is a time-consuming task for ophthalmologists. Moreover, the interpretation of retinal OCT scans varies heavily among human experts [3]. Automatic computer-aided diagnosis of OCT scans, employing deep convolutional neural networks (CNNs) have the potential to revolutionize retinal disease diagnosis by performing accurate and rapid classifications on large amounts of OCT data [9].

Recent researches have used deep learning on OCT images for retinal pathology identification based on segmentation or classification methods. Nugroho *et al.* [11] showed that deep neural networks like Resnet50 and DenseNet-169 outperformed handcrafted features based HOG and LBP methods for classification of OCT images. Kermany *et al.* [6] utilized transfer learning, based on InceptionV3 architecture, on OCT scans to classify AMD and DME, to guide the administration of anti-VEGF therapy. Rasti *et al.* [13] proposed a multi-scale deep convolutional neural network consisting of multiple feature-learning branches, each having different input image sizes, to identify dry-AMD and DME from normal retinas. Prahs *et al.* [12] focused on predicting anti-VEGF treatment indication from OCT B-scans, by developing a deep CNN based on GoogLeNet Inception models. The authors classified OCT images into two classes, “injection” and “no injection” groups. Li *et al.* [10] integrated handcrafted features with features extracted from VGG, DenseNet and Xception networks for OCT-based retinal disease classification.

However, these studies [6, 10–12] made use of deep CNNs with millions of parameters to extract features from OCT scans, which not only limited their deployment to machines with large memories but also entailed long testing times for each B-scan. Moreover, most studies [6, 10, 11, 13] focused on specific retinal pathology classification without generalizations to other retinal diseases. From an ophthalmologist point of view, it is essential to classify whether anti-VEGF treatment is required or routine care would suffice so that proper strategies and treatment plans can be designed for the patients. Therefore, in this study, we propose a fully automated classification algorithm, utilizing a CNN with fewer parameters for macular OCT B-scans concerning the requirement of anti-VEGF therapy. In this research, OCT scans are classified into “urgent referral”, “routine referral” or “normal” categories. Urgent referral corresponds to OCT scans of patients requiring immediate anti-VEGF therapy to prevent irreversible blindness, whereas routine referral corresponds to mild forms of pathologies which do not require anti-VEGF medication but do require routine care. Normal corresponds to patients with healthy eyes. Not only do we predict whether anti-VEGF injections are needed or not, but we also predict whether a patient has other mild forms of eye disease requiring routine care. The main contribution is that a

novel CNN with 800x less number of parameters, requiring 10x less memory and 50x less testing time per scan is proposed, which provides better classification performance compared to the well known deep CNNs.

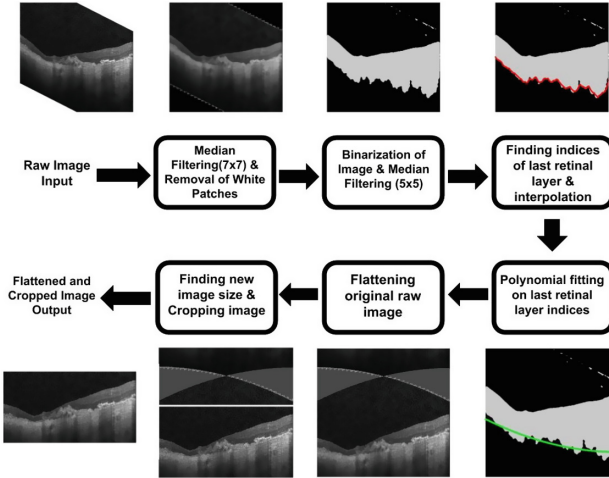


Fig. 1. Data preprocessing: retinal curvature flattening and image cropping

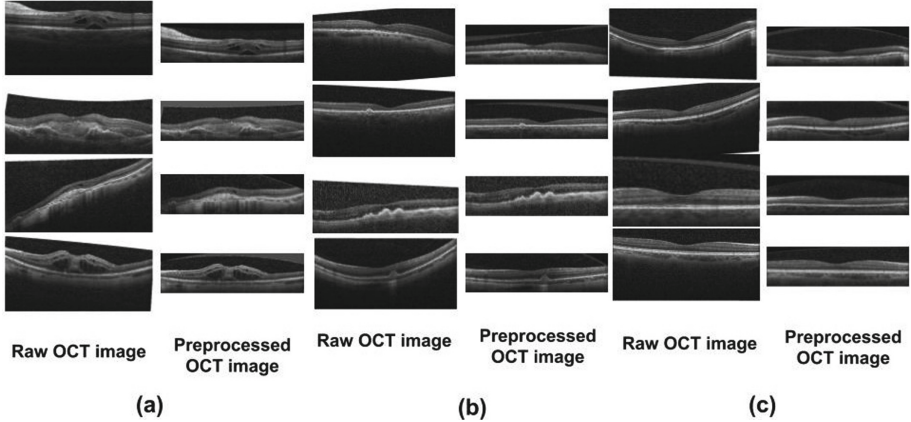
## 2 The Proposed Method

### 2.1 Data Preprocessing

Due to various biological structures and acquisition disturbances, retinal layers in raw OCT scans may be shifted or oriented randomly, consequently causing high variability in their locations and axes [13]. Retinal layers in OCT scans also appear naturally curved due to standard OCT image acquisition practices and display [8] which varies between patients and imaging techniques [16]. Retinal curvature flattening and image cropping gives uniformity to images before automated classification and is widely used in literature. To this end, the retinal flattening algorithm proposed by Srinivasan *et al.* [16] is adopted with some modifications.

Figure 1 shows the steps for preprocessing the retinal OCT images. Median filtering with a mask of size  $7 \times 7$  is initially applied to remove the speckle noises present in the grayscale OCT images. The high-intensity white patches resulting from random orientation in raw OCT scans were removed by replacing pixels with values above 250 with zero. The image is then binarized to obtain image pixels with intensities above the mean intensity of the image, thus, highlighting the retinal layers. The hyper-reflective retinal pigment epithelium (RPE) layer is estimated by locating the last occurrences of high intensities in the binarized

image. Interpolation and a  $5 \times 5$  median filter are applied to remove outliers from the estimated RPE layer. A second-order polynomial fit is obtained on the estimated RPE, and each column in the original raw image is shifted up or down to obtain a flat RPE. The diagnostic information in the OCT images is contained in the retinal layers. Therefore, each B-scan is cropped horizontally, considering 220 pixels above and 20 pixels below the RPE to remove the vitreous and choroid-sclera regions. Finally, each image is resized to  $60 \times 200$  pixels. Figure 2 shows raw OCT scans in various classes before and after preprocessing.



**Fig. 2.** Raw OCT images showing random orientation, retinal curvature, and variability in location. Corresponding preprocessed images after retinal curvature flattening, cropping, and resizing (a) Urgent referral category (b) Routine referral category (c) Normal category

## 2.2 Proposed Architecture for Classification: SimpleNet

In this paper, a small CNN architecture with 30,793 parameters is used as the classifier. The proposed SimpleNet consists of four convolutional layers, two batch normalization layers, two max-pooling layers, and a final fully connected output layer consisting of three nodes with softmax activation function. The convolutional layers extract features from the preprocessed OCT images using convolution operation performed with convolution filters (kernels). The mathematical formulation of the convolution operation across different layers is given as follows:

$$X_j^l = \sum_i \sigma(X_i^{l-1} * w_{i,j}^l + b_j^l) \quad (1)$$

where  $*$  represents the convolution operation,  $X_j^{l-1}$  and  $X_j^l$  represents the feature maps at the convolutional layer  $l - 1$  and  $l$  respectively,  $w_{i,j}^l$  and  $b_j^l$  are

the trainable weights and biases.  $\sigma$  represents the rectified linear unit (ReLU) activation function given as

$$f(x) = \max(0, x) \quad (2)$$

The dropout layer after the convolution layer prevents the CNN from overfitting on the data while the batch-normalization layer accelerates and stabilizes the training of the CNN [18]. The max-pooling operation takes the highest value from each kernel and reduces the size of feature maps. The fully-connected layer employs a soft-max activation function for the output layer to predict input images into three categories, i.e., urgent referral, routine referral, or normal categories. The details of the SimpleNet model is summarized in Table 1.

**Table 1.** Details of the proposed SimpleNet architecture

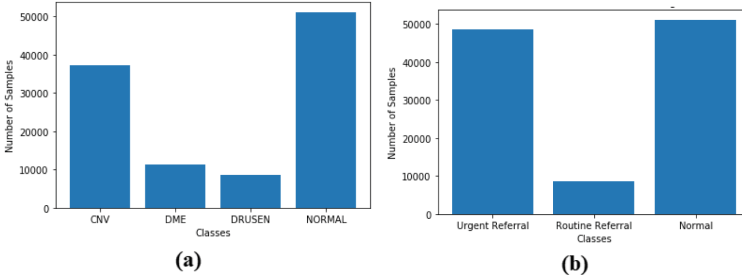
Layers (type)	Number of feature maps	Number of neurons in layer	Size of kernel for each feature map	Stride	Padding	Number of parameters
Input	1	$60 \times 200 \times 1$				
Convolution + ReLU	20	$56 \times 196 \times 20$	$5 \times 5 \times 1$	1	0	520
Convolution + ReLU	20	$52 \times 192 \times 20$	$5 \times 5 \times 20$	1	0	10020
Dropout (10%)		$52 \times 192 \times 20$				
Batch Normalization		$52 \times 192 \times 20$				80
Max-pooling		$13 \times 48 \times 20$	$4 \times 4$	1		
Convolution + ReLU	40	$11 \times 46 \times 40$	$3 \times 3$	1	0	7240
Convolution + ReLU	30	$9 \times 44 \times 30$	$3 \times 3$	1	0	10830
Dropout (20%)		$9 \times 44 \times 30$				
Batch Normalization		$9 \times 44 \times 30$				120
Max-pooling		$2 \times 11 \times 30$	$4 \times 4$	1		
Flatten		660				
Fully-connected		3				1983
Total =						30793

The training process uses Adam as the optimizer with a learning rate of 0.001, while the “categorical cross-entropy” cost function is used to train the output layer. The training was performed in batches of 128 images per step. Given the imbalance between classes in the training set (Fig. 3(b)), different weights are assigned to loss functions of each class based on the number of images it contains.

## 3 Experimental Studies and Results

### 3.1 Database

The publicly available OCT database proposed by Kermany et al. [6] was used in our study. The large database containing 109,309 validated OCT images from 5319 adult subjects, were split into train and test sets of independent patients.



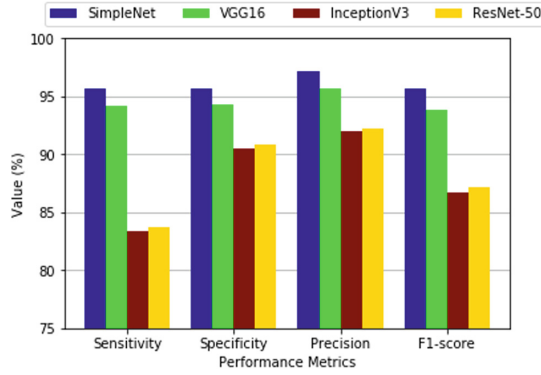
**Fig. 3.** (a) Original training data distribution per class (b) New distribution of training instances per class

The OCT images in the dataset have four categories; CNV, DME, DRUSEN, and NORMAL. The data distribution of the four classes in the training set is shown in Fig. 3(a). The test set contained 250 images from each category. CNV and DME require urgent care in the form of anti-VEGF medication, whereas drusen, which are lipid deposits under the retina, requires routine medications. Thus, we combined CNV and DME into a new class called “urgent referral,” and DRUSEN was labeled as a “routine referral.” The new training set distribution is shown in Fig. 3(b).

### 3.2 Performance Comparison

We compared the proposed method with VGG16 [15], ResNet50 [5] and InceptionV3 [17] based classification. We used the transfer learning technique where we removed the fully connected layers, froze the convolutional layers, and fed the features extracted from it to a fully connected layer with 3 nodes and softmax as the activation function. Batch-normalization layers were, however, trained again for this specific classification problem. 5-fold cross validation was used to present an accurate estimate of the efficacy of each model on unseen data. Each of these models was trained with Adam as the optimizer and categorical cross-entropy as the cost function for the output layer. Class weights of 0.71, 4.19, and 0.74 were given to loss functions of normal, routine referral and urgent referral classes respectively to overcome the imbalance in training set [7]. Table 2 presents the performance metrics for SimpleNet, VGG16, Resnet50, and InceptionV3 for 5-fold cross-validation on the preprocessed dataset. It can be seen that SimpleNet heavily outperforms other architectures in terms of space and testing times while achieving better classification performance (Fig. 4).

Table 3 shows the class wise performance of the proposed SimpleNet architecture. The best SimpleNet model achieves a prediction accuracy of 97.7% with a sensitivity of 100%, 95% and 97.8% for the normal, routine referral and urgent referral classes respectively.



**Fig. 4.** Graph showing performance metrics of the proposed SimpleNet model against transfer learning based classification methodology according to 5-fold cross-validation on preprocessed data

**Table 2.** Details and performance metrics of the proposed SimpleNet model against transfer learning based classification methodology based on 5-fold cross-validation. Testing was done on GPU with 5 GB GPU memory and card model Nvidia Tesla K20c. The card was fitted on a Dell Workstation of Model R7610 with 32 GB memory and 600 GB SAS HDD.

Classification method	Number of parameters	Number of trainable parameters	Average testing time per image (s)	Size on disk (MB)	[Sensitivity, Specificity, Precision, F1-score] (%)
<b>SimpleNet</b>	<b>30,793</b>	<b>30693</b>	<b>0.0464</b>	<b>0.15</b>	<b>[95.67, 95.7, 97.12, 95.63]</b>
VGG16	14,789,955	75,267	2.1444	56.4	[94.15, 94.25, 95.71, 93.89]
InceptionV3	21,956,387	153,603	0.9027	84.4	[83.37, 90.46, 92.00, 86.74]
ResNet-50	23,888,771	354,179	1.4566	73.2	[83.68, 90.86, 92.28, 87.12]

**Table 3.** Performance metrics of the best SimpleNet model for each class on the test set of 1000 preprocessed images

Categories	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
Normal	100.0	99.07	97.28	98.62
Routine referral	95.20	98.67	95.97	95.58
<b>Urgent referral</b>	<b>97.80</b>	<b>98.80</b>	<b>98.79</b>	<b>98.29</b>

## 4 Conclusions

In this paper, a novel deep CNN called SimpleNet was proposed, which performed classification with an accuracy of 97.7%, an average sensitivity of 97.67% and an average specificity of 98.84%. It was shown that our proposed model was highly efficient in terms of storage space and testing times compared to the state-of-the-art CNNs, performing 50 times faster testing and consuming 350 times less space, with better classification performance. The fast and accurate features of the proposed method make it highly suitable for use in eye clinics and remote health care centers.

## References

1. Cheung, N., Wong, I.Y., Wong, T.Y.: Ocular anti-VEGF therapy for diabetic retinopathy: overview of clinical efficacy and evolving applications. *Diabetes Care* **37**(4), 900–905 (2014)
2. Cheung, N., Wong, T.Y.: Diabetic retinopathy and systemic vascular complications. *Prog. Retinal Eye Res.* **27**(2), 161–176 (2008)
3. Ferrara, N.: Vascular endothelial growth factor and age-related macular degeneration: from basic science to therapy. *Nat. Med.* **16**(10), 1107 (2010)
4. Gabriele, M.L., et al.: Three dimensional optical coherence tomography imaging: advantages and advances. *Prog. Retinal Eye Res.* **29**(6), 556–579 (2010)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
6. Kermany, D.S., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**(5), 1122–1131 (2018)
7. King, G., Zeng, L.: Logistic regression in rare events data. *Polit. Anal.* **9**(2), 137–163 (2001)
8. Kuo, A.N., et al.: Correction of ocular shape in retinal optical coherence tomography and effect on current clinical measures. *Am. J. Ophthalmol.* **156**(2), 304–311 (2013)
9. Lee, C.S., Baughman, D.M., Lee, A.Y.: Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol. Retina* **1**(4), 322–327 (2017)
10. Li, X., Shen, L., Shen, M., Qiu, C.S.: Integrating handcrafted and deep features for optical coherence tomography based retinal disease classification. *IEEE Access* **7**, 33771–33777 (2019)
11. Nugroho, K.A.: A comparison of handcrafted and deep neural network feature extraction for classifying optical coherence tomography (OCT) images. In: *2018 2nd International Conference on Informatics and Computational Sciences (ICI-CoS)*, pp. 1–6. IEEE (2018)
12. Prahs, P., et al.: OCT-based deep learning algorithm for the evaluation of treatment indication with anti-vascular endothelial growth factor medications. *Graefe's Arch. Clin. Exp. Ophthalmol.* **256**(1), 91–98 (2018)
13. Rasti, R., Rabbani, H., Mehridehnavi, A., Hajizadeh, F.: Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Trans. Med. Imaging* **37**(4), 1024–1034 (2017)



14. Resnikoff, S., et al.: Global data on visual impairment in the year 2002. *Bull. World Health Organ.* **82**, 844–851 (2004)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
16. Srinivasan, P.P., et al.: Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* **5**(10), 3568–3577 (2014)
17. Szegedy, C., et al.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015)
18. Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.: Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**(4), 611–629 (2018)