

# Automated Diabetic Macular Edema (DME) Analysis Using Fine Tuning with Inception-Resnet-v2 on OCT Images

Ravi M. Kamble<sup>1</sup>, Genevieve C. Y. Chan<sup>2</sup>, Oscar Perdomo<sup>3</sup>, Fabio A. González<sup>3</sup>, Manesh Kokare<sup>1</sup>, Henning Müller<sup>4</sup> and Fabrice Mériaudeau\*

**Abstract**—Accurate detection of diabetic macular edema (DME) is an important task in optical coherence tomography (OCT) images of the eye. A relatively simple and practical approach is proposed in this paper. A pre-trained convolutional neural network (CNN) is fine tuned for a classification of DME versus normal cases. The fine-tuned Inception-Resnet-v2 CNN model can effectively identify pathologies in comparison to classical learning. Experiments were carried out on the publicly available data set of the Singapore Eye Research Institute (SERI). The developed model was also compared to other fine tuned models, such as Resnet-50 and Inception-v3. The proposed method achieved 100% classification accuracy with the Inception-Resnet-v2 model using a leave-one-out cross-validation strategy. For robustness, the model trained on the SERI dataset was tested on another dataset provided by the Chinese University HongKong (CUHK), also with 100% accuracy. The proposed method is a potentially impactful tool for accurately detecting DME vs. normal cases.

## I. INTRODUCTION

### A. Motivation

Diabetic macular edema (DME) and diabetic retinopathy (DR) are the most common serious retinal disorders. Both cause an irreversible vision impairment [1]. Accurate identification of DME and DR with progression heavily depends on advanced retinal imaging techniques. Optical coherence tomography (OCT) imaging is extensively used to view and capture small changes in the retina [2]. This three-dimensional imaging technique is now considered as standard in the ophthalmology for examination and assessment of the response to retinal treatment. Clinically, when lesions occur in the macular region, the internal limiting membrane (ILM) and the retinal pigment epithelium (RPE) undergo morphological changes [3]. Accurate identification

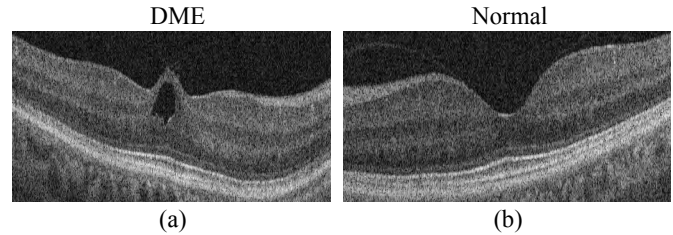


Fig. 1. Sample image from the SERI (Singapore Eye Research Institute) database (a) a DME case and (b) a normal case of an OCT slice.

of these abnormalities as observed in OCT imaging can help to improve diagnosis of retinal diseases like DME and DR.

Many researchers have developed methods for automatic detection of DME [2]. Srinivasan et al. [4] proposed a classification approach to distinguish DME, age-related macular degeneration (AMD), and normal OCT volumes. They used pre-processing using Block Matching 3D filtering (BM3D) and extracted Histogram of Oriented Gradients (HOG) features with an SVM classifier. Lemaitre et al. [5] used LBP (Local Binary Patterns) features to capture the texture of OCT images with dictionary learning using a BoVW (Bag of Visual Words) models. Furthermore, they used a Random forest (RF) classifier for LBP feature descriptors and achieved a sensitivity (SE) of 87.5% and a specificity (SP) of 75%. Awais et al. [6] developed a deep learning approach using features from a pre-trained VGG network with several classifiers for DME volume classification. Perdomo et al. [7] provide an important point of reference by developing the OCT-NET end-to-end deep learning model. The method achieves an accuracy of up to 93.75 %. Other advantages are that the OCT-NET model has a simple architecture with only 12 layers, so is fast to train.

Variants of CNN architectures have been developed over the past years, such as Googlenet [8], Resnet-50 [9], Inception-v3 [10] and Inception-resnet-v2 [11]. The inception architecture with residual connections performs more efficiently by making the network deeper and wider. The recent Inception-Resnet-v2 was chosen for our proposed method as it has given the best performance on ImageNet in the 2015 ImageNet Large Scale Visual Recognition Competition (ILSVRC) challenge [11]. The use of a CNN model pre-trained on natural images and fine-tuned on medical images is widely accepted [12]. Fine tuning has made it feasible to use recent networks for our DME pathology classification. Karri et al. [13] presented a deep learning approach that

\*This work was supported by Universiti Research Internal Fund (URIF) grant, Universiti Teknologi PETRONAS.

<sup>1</sup>Ravi Kamble and Manesh Kokare is from Electronics and Telecomm. Dept., SGGSIE& T, Nanded, India [kambleravi@sggs.ac.in](mailto:kambleravi@sggs.ac.in), [mbkokare@sggs.ac.in](mailto:mbkokare@sggs.ac.in)

<sup>2</sup>Genevieve C. Y. Chan is from Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi Petronas, Malaysia [genevieve.chan94@gmail.com](mailto:genevieve.chan94@gmail.com)

<sup>3</sup>Oscar Perdomo and Fabio A. González is from MindLab Research Group, Universidad Nacional de Colombia [ojperdomoc@unal.edu.co](mailto:ojperdomoc@unal.edu.co), [fagonzalez@unal.edu.co](mailto:fagonzalez@unal.edu.co)

<sup>4</sup>Henning Müller is from University of Applied Sciences Western Switzerland (HES-SO) TechnoPole [henning.mueller@hevs.ch](mailto:henning.mueller@hevs.ch)

\*Fabrice Mériaudeau is Prof. of Dept. of Electrical and Electronic Engg., Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi Petronas, Malaysia [fabrice.meriaudeau@utp.edu.my](mailto:fabrice.meriaudeau@utp.edu.my)

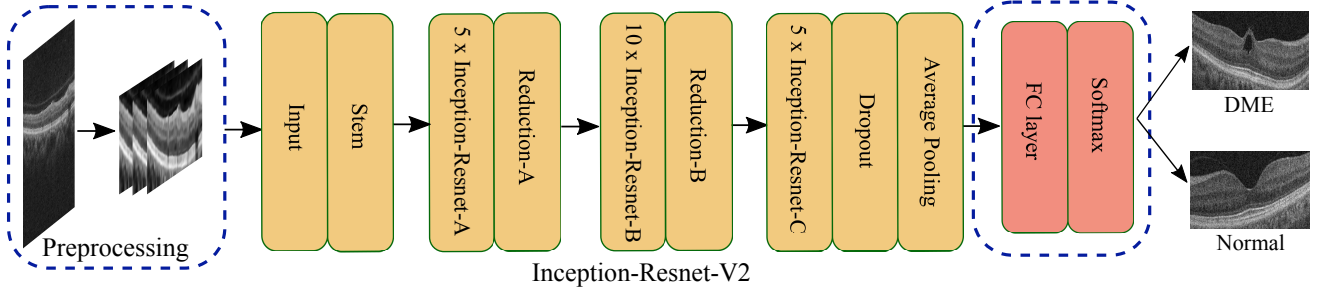


Fig. 2. The overall scheme for fine tuning the Inception-Resnet-v2 for DME vs. normal OCT volume classification.

fine-tunes a pre-trained GoogLeNet on an OCT dataset with DME, dry AMD and normal cases. There are also several recent techniques that applied deep learning approaches for DME pathology detection [14]- [15].

Many of the published methods still suffer from limited performance but good performance is required in real world applications.

### B. Summary of the Contributions

In this paper, we propose an efficient DME vs. normal classification method based on fine tuning recent CNNs. The overall contributions of this paper are:

- 1) We present a simple but effective fine tuning method using the recent Inception-Resnet-v2 CNN model for DME vs. normal classification.
- 2) The proposed method was also trained on the SERI dataset and tested on the CUHK dataset, which shows the network robustness by classifying normal vs. DME.
- 3) We compare to the performance of other networks, such as Inception-v3 and ResNet-50 with fine tuning, also with good results.

The remainder of the paper is organized as follows: in section II we provide the proposed method with fine tuning for the DME classification. The experimental results and discussion are provided in Section III. In Section IV, critical conclusion and discussion.

## II. METHODS

### A. Datasets used

The proposed method uses the SERI and CUHK SD-OCT image datasets. The SERI dataset was acquired by the Singapore Eye Research Institute and consists of 32 OCT volumes (16 DME and 16 normal cases). Each volume contains 128 contiguous 512x1024 pixel B-Scan images. The evaluation was performed first with a k-fold cross-validation (leave one patient out) with 32 independent folds on the SERI dataset and then tested on the CUHK dataset. The CUHK dataset contains DME (4) and normal (79) OCT volumes. The overall framework of the proposed DME disease classification using OCT images is shown in Fig. 2. The methods include preprocessing, the Inception-ResNet-v2 CNN model and fine tuning at the end.

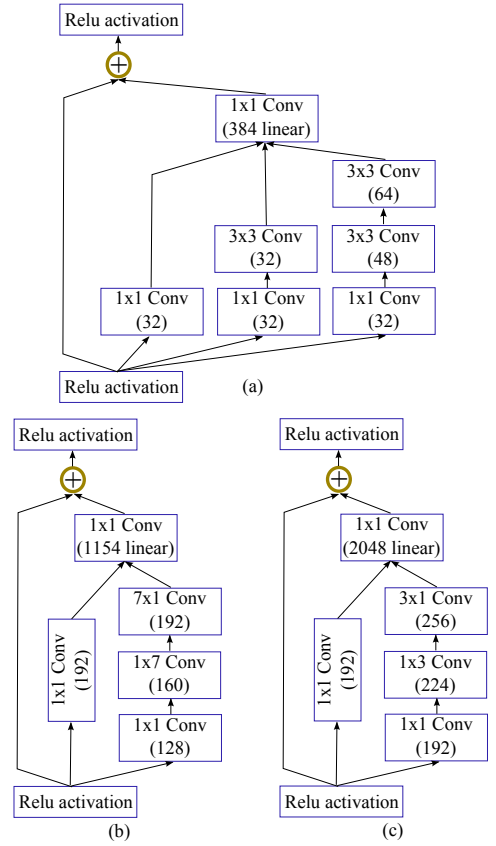


Fig. 3. The schema for interior grid modules of the Inception-ResNet-v2 network [11]. (a) Inception-Resnet-A (b) Inception-Resnet-B and (c) Inception-Resnet-C block of the network shown in Fig. 2.

### B. Preprocessing of OCT data

A typical retinal OCT image has unclear edges with low signal-to-noise ratio, and hence de-noising can be useful as pre-processing step. For noise reduction we use a BM3D filter, which produces a smooth OCT image [16]. BM3D is a 3D filtering technique in the transform domain using the Discrete Fourier Transform (DFT) by combining a sliding-window with block-matching. The proposed approach segments the Internal Limiting Membrane (ILM) and Retinal Pigment Epithelium (RPE) OCT layer of the denoised image for cropping and resizing it to 224x224 pixels [17]. Once preprocessed, the images are fed into several CNNs.

### C. Inception-ResNet-v2

Inception-ResNet-v2 is a combination of two recent networks, residual connections [9] and a recent version of Inception architecture [10]. The Inception models are famous for their multi-branch architectures. They have a set of filters (1x1, 3x3, 5x5, etc.) that are merged with concatenation in each branch. The split-transform-merge architecture of the inception module is observed as a powerful representational ability in its dense layers [10]. The residual model is famous for training very deep architectures. The hybrid Inception-ResNet-v2 network uses residual connections with good efficiency [11]. The stem configuration for the Inception-ResNet-v2 network is not shown here. The interior module of the network includes Inception-Resnet-A, Inception-Resnet-B and Inception-Resnet-C blocks shown in Fig. 3 [11]. The original network is stable by scaling down the residuals before adding them to the previous layer.

### D. Fine tuning for DME vs. normal cases

The proposed method includes fine-tuning an Inception-Resnet-v2 pre-trained model to identify DME in OCT images. The parameters in the last fully-connected layer can adapt better to the DME data. The recent Inception-Resnet-v2 pretrained CNN model is used to initialize the first several layers for the fine-tuned network [11]. The last three layers namely global average pooling, fully connected dense layer and a softmax layer were replaced to tune DME vs. normal. The fully connected dense layer has 3074 parameters for tuning. The softmax layer produces a probability distribution over 2 classes: DME vs. normal. The weights of all fully connected layer are randomly initialized. Then, the model is fine-tuned using an RMSprop optimizer with a momentum of 0.8 and learning rate of 0.0001. RMSprop utilizes the magnitude of recent gradients to normalize the gradients [18]. The experiments are carried out for 10 epochs with a batch size of 16 images. The proposed method was implemented entirely in python using a TensorFlow backend on a GeForce GTX 1080ti card from NVIDIA on linux with an Intel Xeon 2.1GHzx16 CPU with 32GB DDR2 RAM. It takes 8 hours for training with the above system configuration without parallel computing for 10 epochs.

## III. EXPERIMENTAL RESULTS

All the experiments are evaluated in terms of sensitivity ( $Se$ ), specificity ( $Sp$ ) and finally accuracy using Eq.(1) and Eq.(2) by considering  $Tp$ ,  $Tn$ ,  $Fp$  and  $Fn$  obtained from correct prediction. The ( $Se$ ) and ( $Sp$ ) shows the performance of the method with respect to both the DME and normal classes. The SERI dataset is arranged into 32 independent folds using k-fold cross validation, where the network was trained on 31 volumes in the training set and tested on one volume using a leave-one-patient-out strategy. We consider the entire volume as DME if 65 or more OCT images were classified as DME cases by a quorum rule [7]. The fine tuned Inception-Resnet-v2 gives 100% accuracy, sensitive and specificity on the SERI dataset on the volume level. The performance of other networks such as Resnet-50 and

TABLE I  
ACCURACY USING SEVERAL CNN MODELS ON THE SERI DATABASE OF 32 (16 DME AND 16 NORMAL) VOLUMES

CNN	Sensitivity	Specificity	Accuracy
Resnet-50 [9]	93.75%	56.25%	75%
Inception-v3 [10]	100%	87.5%	93.75%
<b>Inception-Resnet-v2 [11]</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

TABLE II  
COMPARISON WITH PREVIOUS METHODS ON THE SERI DATABASE.

Method	Sensitivity	Specificity	Accuracy
Awas et.al [6]	100%	81.25%	90.6%
Perdomo et.al [7]	93.75%	93.75%	93.75%
<b>Proposed</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

Inception-v3 is shown in Table I. The best model on the SERI dataset was then evaluated on the CUHK dataset for testing. The obtained accuracy of DME vs. normal is 100% as well.

$$Se = \frac{Tp}{Tp + Fn} \quad , \quad Sp = \frac{Tn}{Tn + Fp} \quad (1)$$

$$Acc = \frac{Tp + Tn}{Tp + Fn + Tn + Fp} \quad (2)$$

The comparative performance of the method is shown in Table II. The network performs well in terms of sensitivity, specificity and accuracy. Also, the plots of accuracy and loss on the slice level average of 32 OCT volumes on the SERI dataset using several models is shown in Fig. 4. The average accuracy on a slice level for Resnet-50, Inception-v3 and Inception-Resnet-v2 is 86.94%, 99.21% and 99.46% on SERI dataset.

## IV. DISCUSSION

We observed that the Resnet-50 model is performing poor in terms of specificity as can be seen in Table I. Inception-v3 performs well compared to Resnet-50. The accuracy of Inception-v3 reached the baseline method of 93.75% [7]. The model built on the SERI dataset was also tested on a different dataset named CUHK and achieved a performance of 100% highlighting the good generalization of the model. The limitation of fine tuning is that changing the filter configuration of any layer requires to train that respective layer from scratch. Furthermore, experiments need to be performed on larger datasets, as currently both datasets are of small size. In the future, the system can also be extended to other pathologies.

## V. CONCLUSION

In this paper, an OCT based DME pathology classification is proposed using fine tuning of Inception-Resnet-v2. It was evaluated experimentally that the model trained on general images can be fine-tuned to classify SD-OCT volume data. The fine tuned model achieved 100% accuracy on the SERI dataset at the volume level. This method also has the ability

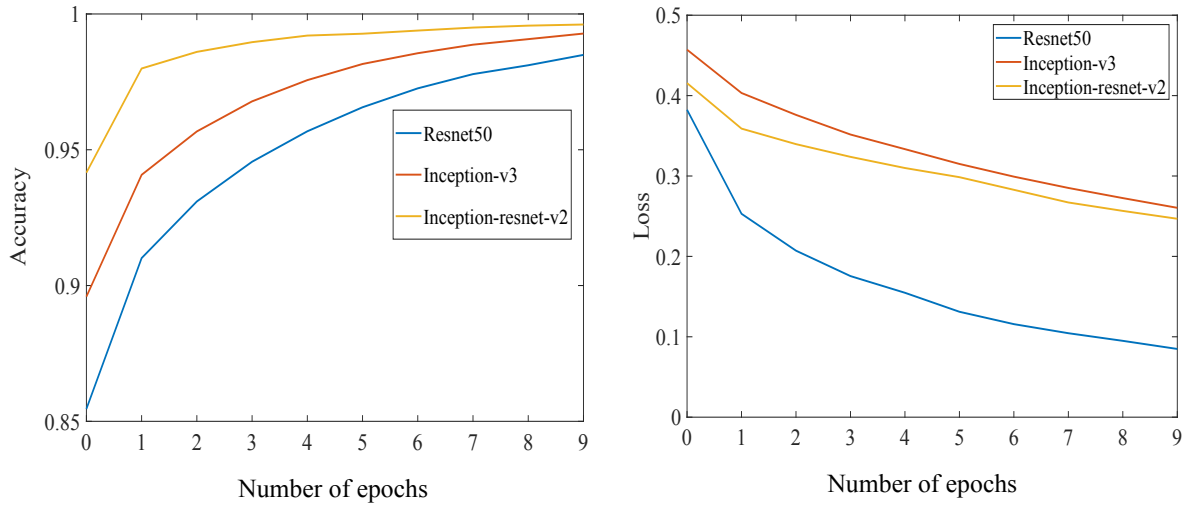


Fig. 4. Results over 32 OCT volumes using a leave one out scheme with several CNNs on the SERI dataset. Left: accuracy and right: loss

to identify DME on the CUHK dataset using the network trained on the SERI dataset. The advantage of the proposed method is the use of recent pre-trained models with limited data for DME classification.

#### ACKNOWLEDGMENT

The authors acknowledge the funding for this research by the Fundamental Research Grant Scheme (FRGS) grant FRGS/1/2017/TK04/UTP/01/1 Ministry of Education (MOE), and Universiti Research Internal Fund (URIF) grant, Universiti Teknologi PETRONAS Malaysia.

#### REFERENCES

- [1] "International diabetes federation. *IDF Diabetes Atlas, 8th edn.* brussels, belgium: International diabetes federation," Online, 2017, <http://www.diabetesatlas.org>.
- [2] J. Massich, M. Rastgoo, G. Lematre, C. Y. Cheung, T. Y. Wong, D. Sidib, and F. Meriaudeau, "Classifying dme vs normal sd-oct volumes: A review," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 1297–1302, Dec 2016.
- [3] Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. de Sisternes, and D. L. Rubin, "Automated drusen segmentation and quantification in sd-oct images," *Medical Image Analysis*, vol. 17, no. 8, pp. 1058 – 1072, 2013.
- [4] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farsiu, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express*, vol. 5, pp. 3568–3577, Oct 2014.
- [5] G. Lemaître, M. Rastgoo, J. Massich, C. Y. Cheung, T. Y. Wong, E. Lamoureux, D. Milea, F. Meriaudeau, and D. Sidibé, "Classification of sd-oct volumes using local binary patterns: experimental validation for dme detection," *Journal of ophthalmology*, vol. 2016, 2016.
- [6] M. Awais, H. Muller, T. B. Tang, and F. Meriaudeau, "Classification of sd-oct images using a deep learning approach," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 489–492, 2017.
- [7] O. Perdomo, S. Otorola, F. A. Gonzalez, H. Muller, and F. Meriaudeau, "Oct-net: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes," in *2018 IEEE International Symposium on Biomedical Imaging (ISBI)*, Accepted, April 2018.
- [8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, June 2015.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, June 2016.
- [10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, June 2016.
- [11] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*, 2017.
- [12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017.
- [13] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomed. Opt. Express*, vol. 8, pp. 579–592, Feb 2017.
- [14] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular oct classification using a multi-scale convolutional neural network ensemble," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1–1, 2017.
- [15] T. Schlegl, S. M. Waldstein, H. Bogunovic, F. Endstraer, 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*, pp. –, 2017.
- [16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, vol. 6064, p. 606414, International Society for Optics and Photonics, 2006.
- [17] G. C. Y. Chan, A. Muhammad, S. A. A. Shah, T. B. Tang, C. K. Lu, and F. Meriaudeau, "Transfer learning for diabetic macular edema (dme) detection on optical coherence tomography (oct) images," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 493–496, Sept 2017.
- [18] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.