Published in Proceedings of the 40th International Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, USA, 17-21 July 2018, which should be cited to refer to this work. DOI: 10.1109/EMBC.2018.8512371

Fusing Results of Several Deep Learning Architectures for Automatic Classification of Normal and Diabetic Macular Edema in Optical Coherence Tomography

Genevieve C. Y. Chan[1], Ravi Kamble[2], Henning Müller[3], Syed A. A. Shah[4], T. B. Tang[1] and Fabrice Mériaudeau*

***Abstract*— Diabetic Macular Edema (DME) is a severe eye disease that can lead to irreversible blindness if it is left untreated. DME diagnosis still relies on manual evaluation from opthalmologists, thus the process is time consuming and diagnosis may be subjective. This paper presents two novel DME detection frameworks: (1) combining features from three pre-trained Convolutional Neural Networks: AlexNet, VggNet and GoogleNet and performing feature space reduction using Principal Component Analysis and (2) a majority voting scheme based on a plurality rule between classifications from AlexNet, VggNet and GoogleNet. Experiments were conducted using Optical Coherence Tomography datasets retrieved from the Singapore Eye Research Institute and the Chinese University Hong Kong. The results are evaluated using a Leave-Two-Patients-Out Cross Validation at the volume level. This method improves DME classification with an accuracy of 93.75%, which is similar to the best algorithms so far on the same data sets.**

I. INTRODUCTION

DME (Diabetic Macular Edema) is an eye disease that is common among diabetes mellitus patients and caused by an abnormal accumulation of extracellular fluids at the macular region. DME only occurs if the patients have Diabetic Retinopathy (DR). DME is one of the leading cause of irreversible blindness among diabetes mellitus patients, if it is not treated [1]. The Optical Coherence Tomography (OCT) imaging technique is widely used to capture 3D cross sectional views of the human eye for the detection of many eye diseases. Compared to fundus photography that captures 2D images of the eye, OCT is able to probe through the retina depth and can image all the retinal layers at a higher resolution. However, since manual diagnosis requires expertise in the field, the assessment can be subjective and time consuming. Thus, it is beneficial to develop an automatic diagnosis system to obtain feedback for physicians.

Over the years, a multitude of DME classification methods have been developed using either classical machine learning methods or deep learning, mostly Convolutional Neural Networks (CNNs). A review was recently published that compares six state of the art machine learning methods for classification of DME and normal patients on OCT images [2]. These models use pre-processing, feature extraction, mapping, feature space reduction and finally classification. The results were evaluated using a majority voting of all B scans in each volume and the performance is measured based on sensitivity (SE), the ability of the system to identify all DME volumes and specificity (SP), the ability of the system to avoid false positives. The highest SE is obtained by [3] with 87.5% while the highest SP is by [4] with 93.8%.

CNNs are a powerful tool for image classification and segmentation thanks to their ability to learn features automatically instead of using handcrafted features. For the ImageNet Large Scale Visual Recognition Competition (ILSVRC), many CNN models designed produced good results. AlexNet (2012) [5], VggNet (2014) [6] and GoogleNet (2014) [7] are explored in this paper. The AlexNet architectures has 8 layers including 3 fully connected layers (FCLs) , VggNet has 16 layers including three 3 FCLs, GoogleNet has 22 layers including 1 FCL and ResNet has 152 layers. Recently, Karri *et al.* [8] fine tuned the GoogLeNet by replacing the last layer of GoogLeNet to classify 3 classes (normal, DME and AMD) in SD-OCT data obtaining an accuracy (ACC) of 96% with a linear SVM classifier. The dataset uses Block Matching 3 Dimension (BM3D) filtering, retinal flattening, cropping and image pyramid reconstruction for image preprocessing. BM3D is a 3D transform domain filtering technique that combines sliding-window transform processing (i.e., a denoising transform that denoises overlapping blocks of the 2D transform domain e.g. Discrete Cosine Transform, DCT) with block-matching (i.e., grouping similar data patches as the image is processed in a sliding manner) [9].

The dataset used in this paper (the SERI dataset) is also used in Chan *et al.* [10], Perdomo *et al.* [11] and Awais *et al.* [12]. This is useful to compare the image classification performance, since the same dataset is used. Chan *et al.* and Awais *et al.* proposed transfer learning for DME detection using a pre-trained CNN. Chan *et al.* used AlexNet and linear SVM classifier while Awais *et al.* used VggNet and KNN and

*This work was supported by the Fundamental Research Grant Scheme (FRGS) grant FRGS/1/2017/TK04/UTP/01/1 Ministry of Education (MOE), Malaysia.

[1]Genevieve C. Y. Chan & T. B. Tang are with Faculty of Electrical and Electronics Engineering, Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi Petronas, Malaysia genevieve.chan94@gmail.com, tongboon.tang@utp.edu.my

[2]Ravi M. Kamble is from Electronics and Telecomm. Dept., SGGSIE& T, Nanded, India kambleravi@sggs.ac.in

[3]Henning Müller is from University of Applied Sciences Western Switzerland (HES-SO) TechnoPole henning.mueller@hevs.ch

[4]Syed A. A. Shah is from COMSATS Institute of Information Technology Abbottabad , Pakistan ayaz@ciit.net.pk

*Fabrice Mériaudeau is Prof. of Dept. of Electrical and Electronic Engg. (CISIR), Universiti Teknologi Petronas, Malaysia fabrice.meriaudeau@utp.edu.my

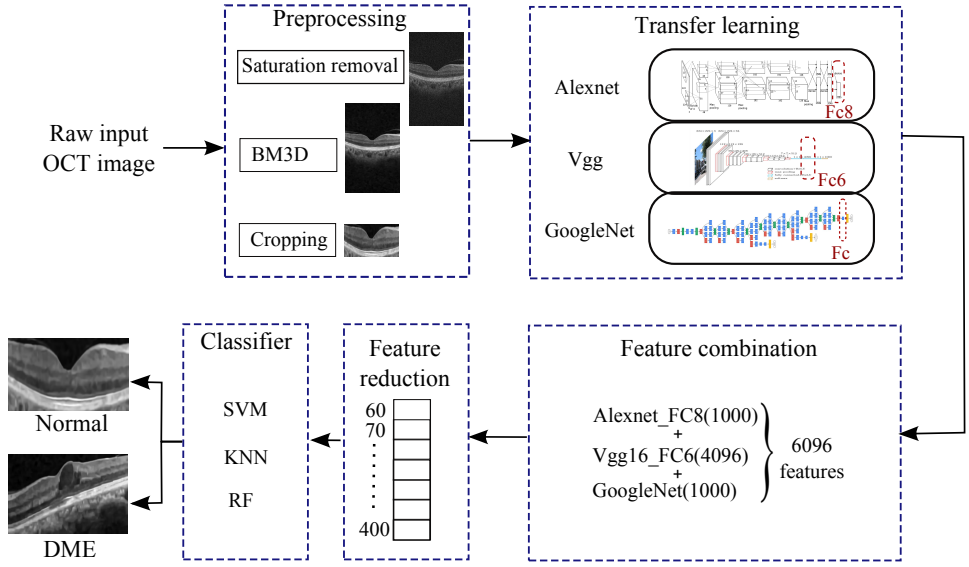

Fig. 1. Overall framework of Feature Combination of AlexNet, VggNet and GoogleNet for Image Classification of DME and Normal of SERI dataset.

a Random Forest classifier. The best performance obtained by Chan *et al.* is 98.85% ACC at the slice level using an 8-fold cross validation while Awais *et al.* obtains 90.6% at the volume level. Although the performance of Chan *et al.* seems better, the training and test sets are not partitioned into volumes, which can create a bias, as slices of the same patient are in training and test data. Not all the OCT slices within a volume have lesions. The decision space in Chan *et al.* and Awais *et al.* is large as AlexNet and VggNet have a dimension of up to 4096 and thus may not generalize well. Perdomo *et al.* propose an end-to-end CNN called OCT-NET. This network consists of 16 layers including 2 FCLs to classify 2 classes i.e. DME and normal. The system is evaluated using a 32−fold Leave One Patient Out Cross Validation (LOPO-CV) and achieves an ACC, SE and SP of 93.75%. This is the baseline model to compare with the results of this paper.

This paper presents DME vs. normal classification based on a decision model of combining AlexNet, VggNet and GoogleNet. Below is the overall contribution of this paper:

1) The proposed method presents a simple framework that combines features extracted from AlexNet, VggNet and GoogleNet, using Principal Component Analysis (PCA) for dimensionality reduction and then classifying DME vs. Normal OCT volumes.
2) A decision model is designed by using the classifications of AlexNet, VggNet and GoogleNet to obtain a result using majority voting based on the plurality rule.

The remainder of this paper is organized as follows: in Section II we present the details of the proposed methods. Section III describes the experiments and Section IV discusses the results. The paper concludes in Section V.

II. Methods

A. Dataset

A set of OCT images and corresponding labels are required as training data. The datasets used in this study were acquired by the Singapore Eye Research Institute (SERI) and the Chinese University of Hong Kong (CUHK), using a Carl Zeiss Meditec, Inc., Dublin, CA (CIRRUS TM) SD-OCT device. The first dataset has 16 volumes of DME cases and 16 volumes of normal cases. The second dataset has 4 volumes of DME cases and 79 volumes of normal cases. Each volume has 128 B-scan slices of 1024×512 pixels. All OCT volumes were read and assessed by trained graders and labelled as normal or DME based on the evaluation of retinal thickening, hard exudates, intraretinal cystoid space formation and subretinal fluid.

Fig. 1 shows the overall framework of the proposed approach. It is composed of pre-processing raw images, transfer learning, a feature combination, feature space reduction and image classifiers for classification of OCT volumes.

B. Image pre-processing of the SD-OCT volumes

The datasets obtained had fuzzy and harsh edges. So, image denoising was used as it had shown to improve image classification performance [10]. First, the saturated pixels (i.e., pixels with intensity value of 255) were removed [8]. Then, the image was filtered with BM3D to produce a smooth image. [13] compared the performance of BM3D with other denoising methods and BM3D had a higher overall effect hence, BM3D was chosen for this study.

Once the images were smoothened, the pixel intensity of the first retinal layer (the Internal Limiting Membrane, ILM) and the last retinal layer (Retinal Pigment Epithelium, RPE) were identified and noted as the borderline to crop. Then, for pixels with values lower than the values of the ILM and RPE layer, were cropped, leaving only the retinal layers with

local intensities that distinguished normal and DME features. Finally, the images were resized based on the requirements of each pre-trained network (i.e., (1) AlexNet is 227×227×3; (2) Vgg-16, Vgg-19 and GoogleNet is 224×224×3) and concatenated thrice to mimic an RGB image.

C. Transfer learning of pre-trained networks

The basic idea of transfer learning is to extract features using the pre-trained weights of each convolutional layer of the network at different depths in the networks after the FCLs for classification. The pre-trained networks used in this study are AlexNet, VggNet and GoogleNet. Since AlexNet and VggNet have 3 FCLs (FC6, FC7 and FC8), features extracted at FC6 and FC7 result each in 4096 features per slice and FC8 obtains 1000 features per slice. GoogleNet has only 1 FCL and 1000 features per image slice are extracted. The volumes were arranged randomly in 16-folds of Leave-Two-Patients-Out Cross Validation (LTPO-CV) (i.e., 1 DME and 1 normal OCT volume).

15 types of classifiers were trained using Classification Learner App on MATLAB 2017-b including Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian and Coarse Gaussian Support Vector Machines (SVMs); fine, medium and coarse trees and fine, medium, coarse, cosine, cubic and weighted K-Nearest Neighbour clasifiers (kNNs). The ACC was computed for each classifier to compare the performance of the classification algorithm and the highest ACC were Cubic SVMs, Fine Trees with 100 splits and Weighted kNN with 10 neighbours, which were used for the remainder of the paper. Experimentation evaluations (i.e. SE, SP and ACC) was similar to the state-of-the-arts methods mentioned in the review, whereby the detailed computations were also explained in [2].

In each fold, the image classification performance was first calculated at the slice level, SE_{Slice} and SP_{Slice}. Then, classification of one OCT volume was based on a quorum rule, so if the SE_{Slice} and SP_{Slice} is at least 50% (i.e., 65 slices per volume), then the OCT volume is predicted as DME (i.e., $SE_{Vol} = 1$) or a normal OCT volume ($SP_{Vol} = 1$), respectively.

D. Decision model of deep learning architectures

After transfer learning, a fusion of deep learning architecture results aims to design a more robust system by combining the classification predictions from AlexNet, VggNet and GoogLeNet. Then, using a plurality rule, an OCT volume is predicted as DME if the following condition is met:

$$SE_{Vol(AlexN)} + SE_{Vol(VggN)} + SE_{Vol(GoogleN)} \geq 2 \quad (1)$$

Similarly, an OCT volume is predicted as normal if the following condition is met:

$$SP_{Vol(AlexN)} + SP_{Vol(VggN)} + SP_{Vol(GoogleN)} \geq 2 \quad (2)$$

E. Feature space reduction

First, features extracted by the FC layers with the highest SE_{Vol} and SP_{Vol} during transfer learning of each pre-trained network were concatenated and PCA is used for dimensionality reduction. Feature space reduction can be useful because features can be correlated and hence redundant. The eigenvectors with the largest eigenvalues were used to reconstruct the feature space in a lower dimension. Although some data can be lost, the most important information should be retained by the remaining eigenvectors. The range in the number of features evaluated was from 60 to 400.

III. Results

Table I shows the image classification performance using pretrained networks. The comparison of SE and SP using majority voting with 3 pre-trained networks and between 3 classifiers helped to select the best FCL for designing a decision model of the deep learning architecture. It can be seen that the decision model constructed by fusing FC8 from AlexNet, FC6 from Vgg16 and GoogleNet FC layers produce a 93.75% ACC. Table II shows a 90.63% ACC after feature combination of the three selected layers with 6096 features and performing majority voting with 3 classifiers. When the feature space is reduced to 140 features, the ACC of the final system using a cubic SVM is 93.75% the same as the baseline method as highlighted in Table III. Table IV highlights the highest performance for each evaluation. The variance column indicates the percentage cumulative sum of the eigenvalues, λ_i up to the selected feature size.

$$Variance = \frac{\sum_1^{Featuresize} \lambda_i}{\sum \lambda_i} \times 100\% \quad (3)$$

This means that the cumulative variance for 140 features is 89%. To test the robustness of the system further, the CUHK dataset is tested on the proposed system and produced a 100% SE, 53.16% SP and 55.42% ACC based on the decision model of the 3 pre-trained networks.

IV. Discussion

The advantage of using LTPO-CV over LOPO-CV is to keep the training data balanced (i.e., 1 DME and 1 normal OCT volume are used for testing). Table IV shows that the feature combination achieved 93.75% for SE, SP and ACC on the SERI dataset. This means that in the 16 folds, it is able to predict the label accurately 15 out of 16 times. The ACC of the proposed method proved that it is simpler but comparable in performance to the baseline model by utilizing readily available CNN models that extract good features instead of designing one from scratch. When the system is tested on the CUHK dataset, the SE and SP show that the proposed method is able to classify all 4 volumes of 'DME' and 42 of 79 volumes of 'normal'cases correctly. The number of OCT volumes for Normal and DME is imbalanced. The drawback of the proposed method is that the weights are pre-trained weights. Therefore, there is a limitation in optimizing the weights to improve classification performance. For future

TABLE I

IMAGE CLASSIFICATION PERFORMANCE OF TRANSFER LEARNING WITH PRETRAINED NETWORKS.

Classifier		**Pre-trained Networks**			**Majority Voting between Networks (%)**		
		AlexNet	**Vgg16/19**	**GoogleNet**	**SP**	**SE**	**ACC**
Cubic SVM		**FC8**	**FC6(16)**	**FC**	87.5	100	**93.75**
Fine Trees		FC8	FC8(19)	FC	81.25	93.75	87.5
Weighted KNN		FC7	FC8(16)	FC	75	93.75	84.38
Majority Voting between Classifiers (%)	**SP**	81.25	81.25	75			
	SE	93.75	93.75	93.75			
	ACC	87.5	87.5	84.38			

TABLE II

MAJORITY VOTING USING SEVERAL CLASSIFIERS FOR COMBINATIONS OF FEATURES.

Majority Voting	SE	SP	ACC
SVM KNN Fine Tree	87.50%	93.75%	90.63%

TABLE III

IMAGE CLASSIFICATION PERFORMANCE AFTER FEATURE SPACE REDUCTION.

Classifier	Feature Size	Variance	SE	SP	ACC
SVM	60	80.48%	75.00%	93.75%	84.38%
	140	**88.79%**	**93.75%**	**93.75%**	**93.75%**
	150	89.36%	87.5%	93.75%	90.63%
	160	89.87%	87.5%	93.75%	90.63%
	210	91.85%	75.0%	93.75%	84.38%
	220	92.16%	75.0%	93.75%	84.38%
KNN	60	80.48%	56.25%	93.75%	75%
	70	82.20%	43.75%	93.75%	68.75%
	110	86.70%	0.00%	100%	50%
	120	87.48%	0.00%	100%	50%
Fine Tree	70	82.20%	87.50%	87.50%	87.50%
	160	89.87%	81.25%	87.50%	84.38%
	170	90.33%	87.50%	87.50%	87.50%
	220	92.16%	81.25%	87.50%	84.38%
	230	92.45%	81.25%	87.50%	84.38%

TABLE IV

COMPARISON OF THE PERFORMANCE BETWEEN THE PROPOSED METHOD AND THE BASELINE METHOD.

Method	SE	SP	ACC
Awais *et al.* [12]	**100%**	81.25%	90.6%
Perdomo *et al.* [11]	93.75%	93.75%	93.75%
Proposed method	93.75%	**93.75%**	**93.75%**

work, fine tuning will be included to optimize the weights to improve performance.

V. CONCLUSION

The proposed method uses information from AlexNet, VggNet and GoogleNet to design a decision model using majority voting of the classification decisions of each model and a system with feature combination from all three networks with PCA as feature space reduction. It showed that this method can classify OCT volumes into DME and normal cases with a 93.75% ACC. The advantage of the proposed method is the implementation of image classification using pre-trained models. This opens up to a simple yet effective method for OCT volume classification.

ACKNOWLEDGMENT

The authors would like to thank FRGS grant FRGS/1/2017/TK04/UTP/01/1 Ministry of Higher Education (MOHE), Malaysia for supporting this research work.

REFERENCES

[1] J. Cunha-Vaz, "Diabetic macular edema," 1998.

[2] J. Massich, M. Rastgoo, G. Lemaître, C. Y. Cheung, T. Y. Wong, D. Sidibé, and F. Mériaudeau, "Classifying dme vs normal sd-oct volumes: A review," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pp. 1297–1302, IEEE, 2016.

[3] G. Lemaître, M. Rastgoo, J. Massich, S. Sankar, F. Mériaudeau, and D. Sidibé, "Classification of sd-oct volumes with lbp: application to dme detection," 2015.

[4] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and J. M. Rehg, "Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding," *Medical image analysis*, vol. 15, no. 5, pp. 748–759, 2011.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[8] S. 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," *Biomedical optics express*, vol. 8, no. 2, pp. 579–592, 2017.

[9] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, *et al.*, "Image denoising with block-matching and 3 d filtering," in *Proceedings of SPIE*, vol. 6064, pp. 606414–606414, 2006.

[10] 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.

[11] O. Perdomo, S. Otalora, 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.

[12] M. Awais, H. Mller, 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, Sept 2017.

[13] D. Fu, H. Tong, S. Zheng, L. Luo, F. Gao, and J. Minar, "Retinal status analysis method based on feature extraction and quantitative grading in oct images," *Biomedical engineering online*, vol. 15, no. 1, p. 87, 2016.