



OPEN

# Automatic detection of microaneurysms in optical coherence tomography images of retina using convolutional neural networks and transfer learning

Ramin Almasi<sup>1</sup>, Abbas Vafaei<sup>1</sup>, Elahe Kazeminasab<sup>1</sup> & Hossein Rabbani<sup>2</sup>

Microaneurysms (MAs) are pathognomonic signs that help clinicians to detect diabetic retinopathy (DR) in the early stages. Automatic detection of MA in retinal images is an active area of research due to its application in screening processes for DR which is one of the main reasons of blindness amongst the working-age population. The focus of these works is on the automatic detection of MAs in en face retinal images like fundus color and Fluorescein Angiography (FA). On the other hand, detection of MAs from Optical Coherence Tomography (OCT) images has 2 main advantages: first, OCT is a non-invasive imaging technique that does not require injection, therefore is safer. Secondly, because of the proven application of OCT in detection of Age-Related Macular Degeneration, Diabetic Macular Edema, and normal cases, thanks to detecting MAs in OCT, extensive information is obtained by using this imaging technique. In this research, the concentration is on the diagnosis of MAs using deep learning in the OCT images which represent in-depth structure of retinal layers. To this end, OCT B-scans should be divided into strips and MA patterns should be searched in the resulted strips. Since we need a dataset comprising OCT image strips with suitable labels and such large labelled datasets are not yet available, we have created it. For this purpose, an exact registration method is utilized to align OCT images with FA photographs. Then, with the help of corresponding FA images, OCT image strips are created from OCT B-scans in four labels, namely MA, normal, abnormal, and vessel. Once the dataset of image strips is prepared, a stacked generalization (stacking) ensemble of four fine-tuned, pre-trained convolutional neural networks is trained to classify the strips of OCT images into the mentioned classes. FA images are used once to create OCT strips for training process and they are no longer needed for subsequent steps. Once the stacking ensemble model is obtained, it will be used to classify the OCT strips in the test process. The results demonstrate that the proposed framework classifies overall OCT image strips and OCT strips containing MAs with accuracy scores of 0.982 and 0.987, respectively.

Nowadays, a lot of researches have been conducted on the retinal image analysis area. One of the most important applications of these works is Diabetic retinopathy (DR) screening because this disorder is one of the leading reasons for vision loss in the working-age population, especially in developed countries. Since Microaneurysms (MAs) are early signs of DR, their detection in retinal images aids screening tasks of diagnosis DR. That is why, automatic detection of MAs from retinal images is extremely important. Although several researches have been done in automatic detection of MAs, detection of MAs is still a challenging task due to the variations in MAs appearance in retinal images<sup>1,2</sup>. The focus of these researches is on the en face retinal images like FA and color fundus photograph. The main steps in MA detection include (1) preprocessing (2) detection of MA candidates and (3) classification of candidates. In the first step, preprocessing tackles the problems related to contrast and non-uniform illumination of MA regions. Afterwards, the initial set for MA candidates is detected and, finally some techniques (often machine learning) are applied to eliminate false-positive cases of MA and enhance the

<sup>1</sup>Department of Computer Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran. <sup>2</sup>Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. ✉email: abbas\_vafaei@eng.ui.ac.ir; rabbani.h@ieee.org

Author	Dataset	Image type	Initial candidates method	Classifier for false positive reduction	Results
Habib et al. <sup>1</sup>	New dataset based on the MESSIDOR dataset	Color fundus	Gaussian matched filter	Tree Ensemble classifier	ROC score = 41.5%
Hatanaka et al. <sup>9</sup>	DIARETDB1 and ROC datasets	Color fundus	DCNN (GoogLeNet)	DCNN and three-layer perceptron with 48 features	Sensitivity = 84%
Shan et al. <sup>16</sup>	DIARETDB dataset	Color fundus	SSAE with two hidden layers plus a Softmax classifier	AUC = 96.2%	
				With tenfold CV f-score = 91.3%	
Deepa et al. <sup>4</sup>	DIARETDB1 dataset	Color fundus	CNN		Sensitivity = 97.62% Specificity = 100% Accuracy = 97.75%
Zhang et al. <sup>23</sup>	IDRiD_VOC dataset	Color fundus	Deep neural network with a multi-layer attention mechanism	Refining initial candidates using spatial relationships between MAs and BVs	Sensitivity = 86.8%
Eftekhari et al. <sup>5</sup>	ROC and E-Ophtha-MA datasets	Color fundus	Basic CNN	Thresholded probability map and final CNN	CPM = 0.461
Long et al. <sup>13</sup>	E-Ophtha-MA and DIARETDB1 datasets	Color fundus	BV removal, shape characteristics and connected components analysis	Classification using NB, KNN and SVM	AUC -ROC = 87% on E-Ophtha-MA AUC of ROC = 86% on DIARETDB1
Wu et al. <sup>21</sup>	ROC dataset	Color fundus	MMMMF	Extracting 37 dimensional features and using modified KNN, LLDA, and SVM	Averaged number of false positives per image = 0.286

**Table 1.** Recent works on MA detection from fundus photographs (CPM = Competition measure, CV = cross validation, BV = Blood vessel, ROC dataset = retinopathy online challenge dataset and AUC = area under curve, ROC = receiver operating characteristic, SSAE = stacked sparse auto encoder, NB = Naive Bayesian, KNN = K-nearest neighbor, SVM = support vector machine, MMMF = multi-scale and multi-orientation sum of matched filter, and LLDA = local linear discrimination analysis).

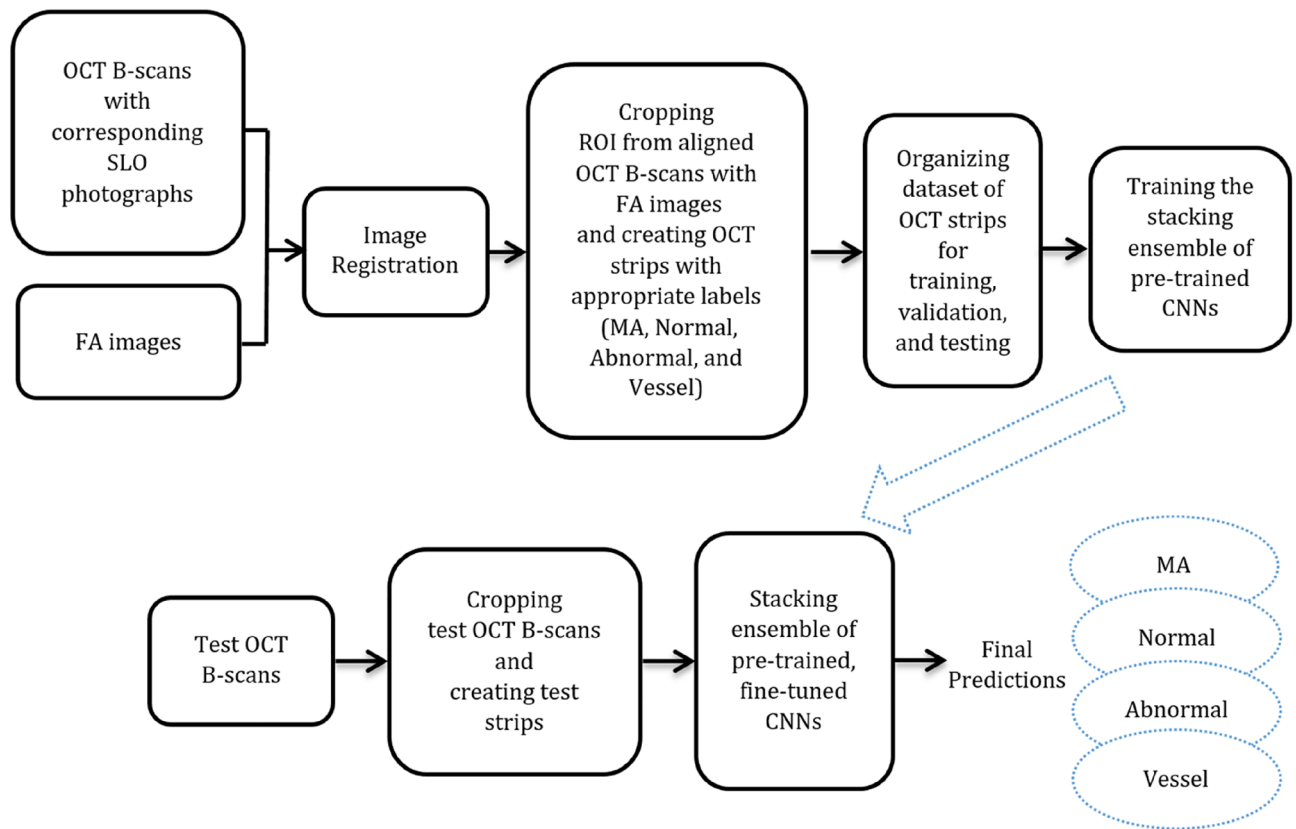
Author	Dataset	Image type	Method	Results
Lemaitre et al. <sup>24</sup>	SERI Dataset of normal, DME-cyst, DME-Exudate	SD-OCT images	LBP features, different mapping strategies and using linear and nonlinear classifiers	Sensitivity = 81.2% Specificity = 93.7%
Rasti et al. <sup>25</sup>	Duke university dataset and private dataset of AMD, DME, and normal	Macular OCT	MCME model containing CNN experts and gating network are fed by specific scales of the input pattern	Precision = 98.86% AUC-ROC = 99.85%
Shih et al. <sup>26</sup>	Dataset of DME, CNV, Drusen, and normal <sup>30</sup>	OCT images	Pre-trained VGG16	Accuracy = 99.48%
Tsuji et al. <sup>27</sup>	Dataset of DME, CNV, Drusen, and normal <sup>30</sup>	OCT images	Capsule network	Accuracy = 99.6%
Kazeminasab et al. <sup>29</sup>	Dataset of MA and normal OCT strips	OCT images	BoF and SURF, MLP classifier	Accuracy = 94.5%

**Table 2.** Some recent works on OCT image classification considering diabetic retinopathy (MCME = multi-scale convolutional mixture of experts, LBP = local binary patterns, BoF = bag of features, SURF = speeded up robust features, and MLP = multilayer perceptron).

accuracy of MA detection. Many algorithms are proposed in<sup>1-23</sup>, each of which brings some improvements into the above-mentioned steps. Table 1 indicates some of the recent works in details.

On the other hand, since OCT images depict the in-depth structure of retinal layers, to the best of our knowledge, optimistically there are few researches that pursue detection of MAs in OCT photographs. As shown in Table 2, most of the researches in OCT classification present the algorithms to classify OCT volumes or OCT B-scans into the AMD, DME, and normal classes<sup>24-27</sup>. AMD is a retinal disease resulting in blurred, blind spots or complete loss of vision in the center of visual field and DME is the most common cause of diabetic vision loss in different societies<sup>28</sup>. The method presented in<sup>29</sup> classifies OCT images into normal and MA categories. For this purpose, features are extracted from the images using Bag of Features (BoF) and SURF descriptor. After that, images are classified into normal and MA classes utilizing a multilayer perceptron network.

Amid the above-mentioned researches, only<sup>29</sup> has focused on the detection of MAs in OCT images. Detection of MAs from OCT images has the following advantages: first, OCT is a non-invasive imaging technique that does not require injection, therefore is safer. Second, as mentioned before, OCT images are normally used for distinguishing between DME, AMD, and normal cases. Detection of MAs from OCT images results in achieving comprehensive information from this single imaging technique. This may reduce the need for using other imaging methods. It is worth mentioning that, there is another retinal imaging technique called Optical Coherence Tomography Angiography (OCTA) which is fast and non-invasive. This method provides volumetric data with the clinical capability of specifically localizing and delineating pathology along with the ability to show both structural and blood flow information in tandem. However this method has some limitations. First, this



**Figure 1.** Overall block diagram of proposed method (SLO = Scanning Laser Ophthalmoscopy).

method is relatively new and not yet very common. Second, it has a relatively small field of view and is not able to show leakage well.

For doing so, we need a dataset comprising OCT image strips with suitable labels including MA, normal, abnormal, and vessel. The abnormal category includes OCT strips that contain objects in the shape of cysts and fluid-associated abnormalities in the retina. Abnormal strips do not have the normal layer structure of a normal OCT strip. To the best of our knowledge, such large labelled datasets are not yet available for OCT images, therefore we have created it. Since MAs and vessels are hard-to-detect objects in OCT images, to create OCT strips with appropriate labels, an accurate image registration method is performed to align OCT images and FA photographs. After that, with the help of corresponding FA images, the OCT strips are created from OCT B-scans in four labels, namely MA, normal, abnormal, and vessel. Once the OCT strips are created and organized as four-class dataset, a stacking ensemble<sup>31</sup> comprising four fine-tuned, pre-trained CNNs is trained to classify the strips of OCT images into the mentioned classes. After the model is obtained, it classifies the strips from the test B-scans. It should be noted that FA images are used once to create OCT strips for training process. After that, the FA images are no longer needed and the stacking ensemble model classifies the OCT strips in the test process independently. To apply our model to the new test images, test B-scans should be divided into strips and the resulted strips should be fed to the stacking ensemble model to be classified into one of the above-mentioned classes. For clinical applications, to find MAs in an OCT volume, first, the B-scans are extracted, then overlapping OCT strips are created (to find all MAs) and after that, the strips are fed to the model for testing. Once the strip label is determined as MA, it can be stated in which B-scan and which strip (known location) the MA is located.

The outline of this paper is organized as follows: “[Methods](#)” section describes the data acquisition, dataset organization, and proposed stacking ensemble method used in this study. In “[Results](#)” section, the evaluation criteria and the results are presented, and “[Discussion](#)” section concludes this study.

## Methods

The overall block diagram of proposed method is presented in Fig. 1 and the details are explained in the next subsections.

**Image registration and OCT strips preparation.** For classification purposes, we need a dataset comprising OCT image strips with suitable labels including MA, normal, abnormal, and vessel. To the best of our knowledge, such labelled datasets are not yet available for OCT images, therefore we had to create it. MAs and vessels are hard-to-detect objects in OCT images, therefore to prepare the dataset of OCT image strips with appropriate labels, a precise registration is performed to align OCT images and FA photographs. To do so, the dataset and the method proposed in<sup>32</sup> is used for accurate registration of OCT and FA images using SLO photo-

graphs as intermediate images. In<sup>32</sup>, dataset includes 36 pairs of FA and SLO images of 21 subjects with diabetic retinopathy, where SLO image pixels are perfectly in correspondence with OCT B-scans. The FA, OCT, and SLO images are captured via Heidelberg Spectralis HRA2/OCT device. Moreover, the FA and SLO images are the same size as  $768 \times 768$  pixels and FA images were captured with two different fields of views (30 and  $55^\circ$ ). In this method, after preprocessing, retinal vessel segmentation is applied to extract blood vessels from the FA and SLO images. Afterwards, a global registration is used based on the Gaussian model for curved surface of retina and for this purpose, first a global rigid transformation is applied to FA vessel-map image using a feature-based method to align it with SLO vessel-map photograph and then the transformed image is globally registered again considering Gaussian model for curved surface of retina to improve the precision of the previous step. Next, a local non-rigid transformation is performed to register two images perfectly.

After that, as shown in Fig. 2, with the help of related FAs, the OCT strips are created from OCT B-scans in four labels including MA, normal, abnormal, and vessel. The FA images are associated only once to create OCT strips for the training process. In the test process, the FA images are no longer needed. In our dataset, MA, normal, abnormal, and vessel classes comprise 87, 100, 72, and 131 strips of OCT images, respectively. In this study, the scale factor of the OCT images in the x direction equals 0.0115. So, in the OCT image, the value of x per pixel is 0.0115 mm. On the other hand, as reported in<sup>33</sup>, the MA has a maximum external diameter of 266  $\mu\text{m}$ . Applying this to our dataset, the maximum external diameter of the MA is calculated as 23.2 pixels, approximately. Therefore, here, the width of the OCT strip is considered to be 31 pixels, which is slightly larger than 23 pixels. The images are cropped in a way that they contain only retinal layers while the other pixels are withdrawn. For this purpose, first the segmentation method presented in<sup>34</sup> is used to detect the Retinal Nerve Fiber Layer (RNFL) and Retinal Pigment Epithelium (RPE) layer, afterwards the OCT B-scan is cropped to include the highest part of NFL and the lowest part of RPE. This process is demonstrated in Supplementary Fig. S1 online. That is why the images have various heights. The process of gathering the dataset of OCT strips and testing B-scans is depicted in Supplementary Figs. S2 and S3 online. It should be noted that when they are used as input for CNNs, the image will be resized to a new dimension of  $150 \times 150 \times 3$ . The dataset is publicly available in <https://misp.mui.ac.ir/en/four-class-dataset-oct-image-strips-png-format-%E2%80%8E-1>.

**Organizing train, validation and test data.** The dataset is organized into train, validation, and test folders each of which includes MA, normal, abnormal, and vessel image folders. Twenty percent of entire dataset is allocated to the test set, and is not utilized in the training process. This is referred to as Hold-out method. To validate each CNN, the Bayesian optimization tuner class of the Keras tuner<sup>35</sup> is used to run the search over the search space. The search space includes the learning rate, momentum, and the number of units in the first dense layer. The number of trials and epochs in the validation process is considered to be 10 and 80, respectively. The tuned hyperparameters for each CNN are listed in Supplementary Table S1 online. Fifteen percent of the entire dataset is allocated to the validation set. Because the dataset of our work is small compared to common deep learning task datasets and this may lead to overfitting, the data augmentation technique is applied. Using this technique, some transformations including rotation, zooming, horizontal flip, re-scaling, and shift are applied to the dataset images in each single training epoch by the image data generator. This helps the model not to memorize images and as a result not to overfit.

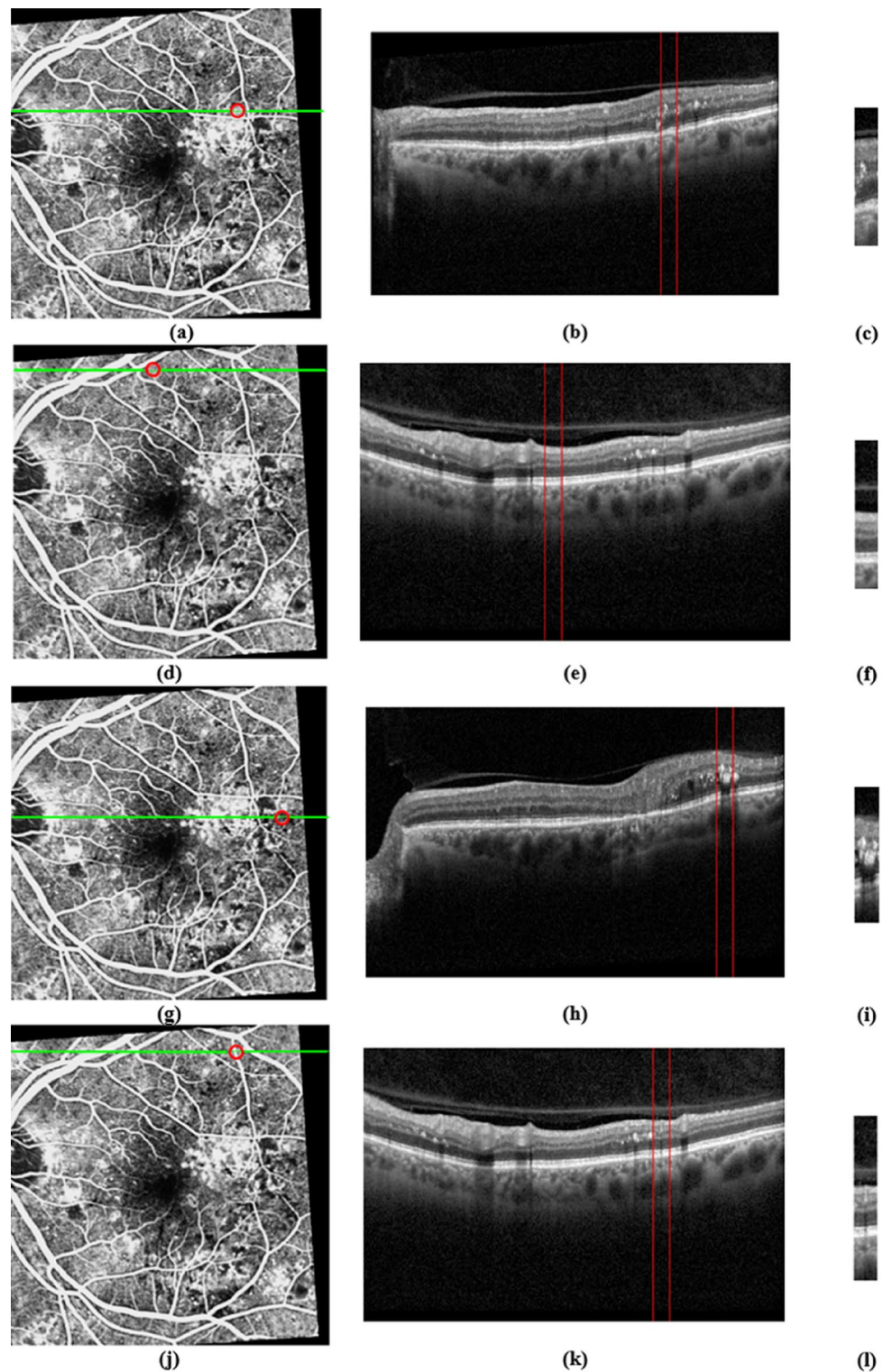
**Stacked generalization ensemble.** The overall structure of the classifier presented in this research is shown in Fig. 3. As can be observed, the stacked generalization (stacking) ensemble of four CNNs pretrained on the ImageNet dataset is utilized. Stacking ensemble includes two levels, namely 0 and 1. The elements and training process of each level is elaborated in the next sections.

*Level 0 of stacking ensemble.* The CNNs used in this stacking ensemble include VGG16, VGG19<sup>36</sup>, Xception<sup>37</sup>, and InceptionV3<sup>38</sup>. These CNNs or so-called base-learners form level 0 of the stacking ensemble. The basic architecture of these base-learners are depicted in Fig. 4. Here, the image size for the CNNs input is  $150 \times 150 \times 3$  and the average pooling is used. To use these networks, the last layer is removed and then, flatten, batch normalization, dense, drop-out and dense layers are added to the end one after the other. The added drop-out layer has the factor of 0.35, and the last added dense layer includes 4 units with Softmax activation function to deal with our 4-class classification problem. The Relu activation function is considered for the first added dense layer, while the number of its units is determined using Keras tuner in validation process. Afterwards, by freezing the previous layers, only the added layers are trained with the data in both train and validation folders for 5 epochs. In this step, the Adam optimizer with the learning rate of 0.0001 is selected for training the newly-added layers in each CNN. Now, the added layers have the initial weights.

After validating each CNN, each network is trained for 100 epochs using whole train and validation data. In this second round of the training process, some of the last layers are trained and the previous layers remain frozen, therefore their weights are not adjusted.

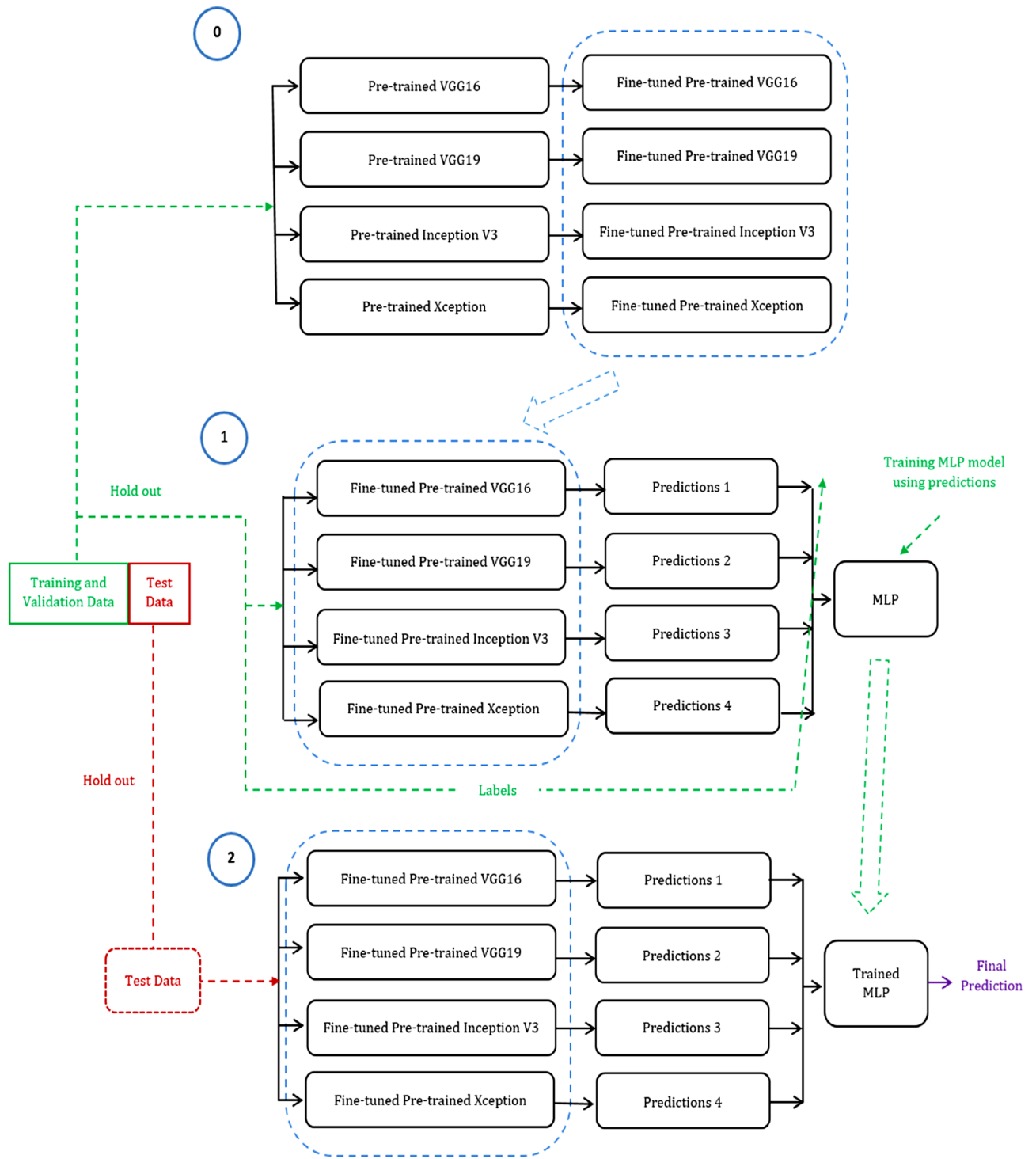
The Stochastic Gradient Descent with Momentum (SGDM) is considered as the optimizer. The learning rate, momentum values, and the number of trainable layers for CNNs are listed in Supplementary Table S1 online. Moreover, the categorical cross-entropy is used as loss function which should be minimized in both training processes.

In the training process, two callbacks are used for early-stopping and saving the model with the highest efficiency. Therefore, if there is no improvement in the performance of the model (minimizing the loss function) for a certain number of epochs (patience parameter), the training process is stopped before the maximum number of epochs is met. In this work, the patience parameter equals 25 and the minimization of the loss function is monitored for early-stopping. Also, since the model obtained from completing the training epochs or stopping



**Figure 2.** The process of creating OCT strips for MA, normal, abnormal, and vessel with the help of corresponding FA. (a) Red circle shows MA in FA. (b) B-scan Corresponds to green line in (a). (c) Cropped ROI from (b). (d–f) Creating strip for normal class. (g–i) Creating strip for abnormal class. (j–l) Creating strip for vessel class (in color).

early (the last trained model) might not necessarily be the best model, saving the model with the highest accuracy and loading it will resolve this issue.



**Figure 3.** Overall structure of the stacked generalization ensemble.

*Level 1 of stacking ensemble.* After training all the base-learners by the training dataset, a meta-learner is presented as level 1 part of stacking ensemble and trained to achieve higher accuracy through combining the results of the trained base-learners. In this study, MLP classifier is used as meta-learner for level 1 of stacking ensemble. This meta-learner is trained on the whole train and validation dataset taking the outputs (predictions) of the base-learners trained in the previous step as input. For this purpose, the predictions from base-learners are stacked and reshaped to be used as input tensors for the MLP model. In fact, the base-learners of the previous step were trained directly by the training dataset, and the MLP model is indirectly trained by the training dataset. To apply obtained model to the new test images, test B-scans should be divided into strips and the resulted strips should be fed to the stacking ensemble model to be classified into one of the mentioned classes. MLP classifier includes 100 hidden layers and maximum iteration parameter is set to be 300.

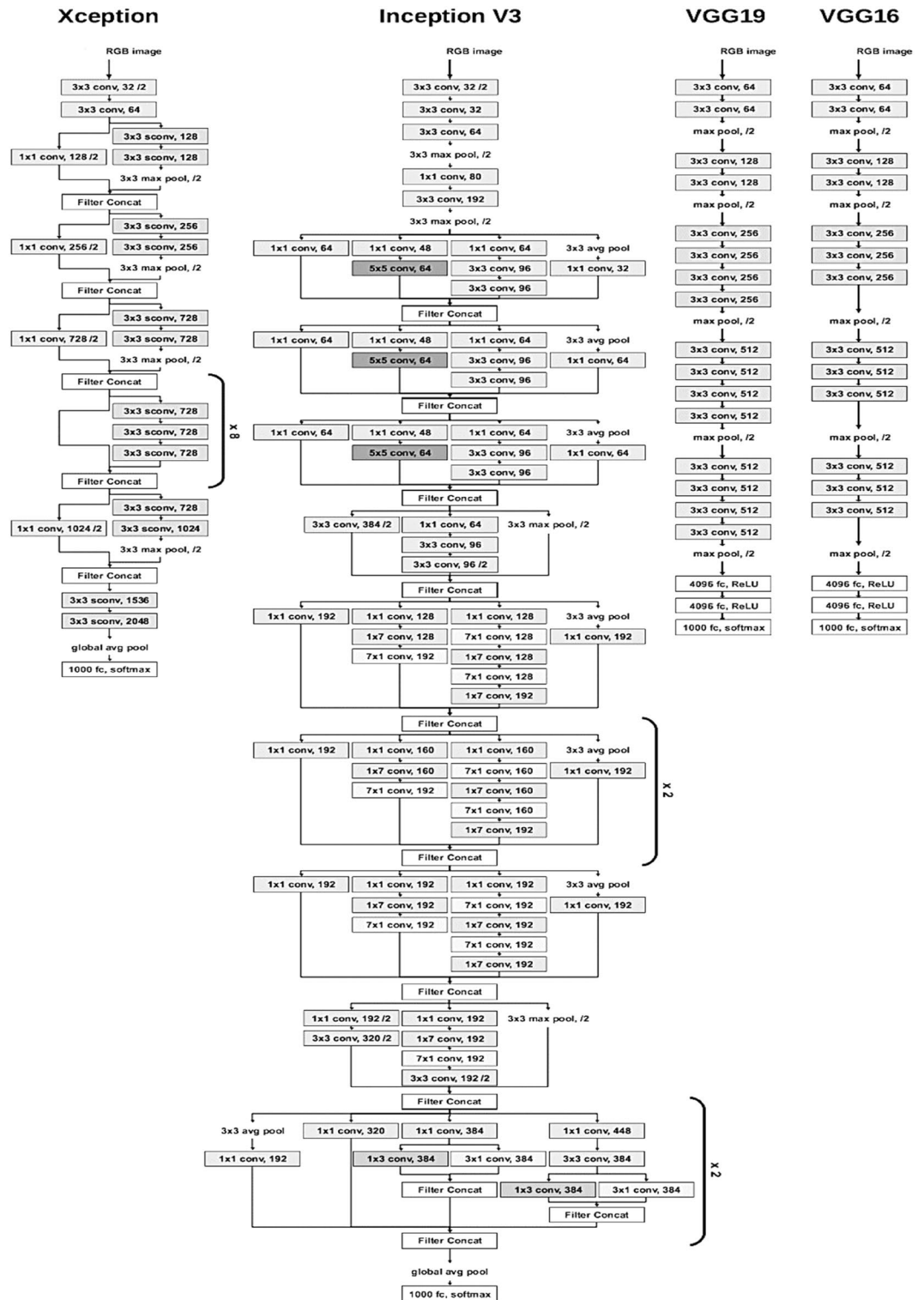


Figure 4. Basic structure of CNNs used in stacking ensemble<sup>39</sup>.

## Results

In this work, to have a complete evaluation of the proposed classification model, a number of criteria have been considered, which are mentioned in the caption of Table 3.

Here, TP (True Positive) for a specific class represents the number of cases that have been correctly classified in that class and TN (True Negative) represents the number of cases that have been correctly classified as not belong to that class. The FP (False Positive) for a specified class also determines the number of items that were

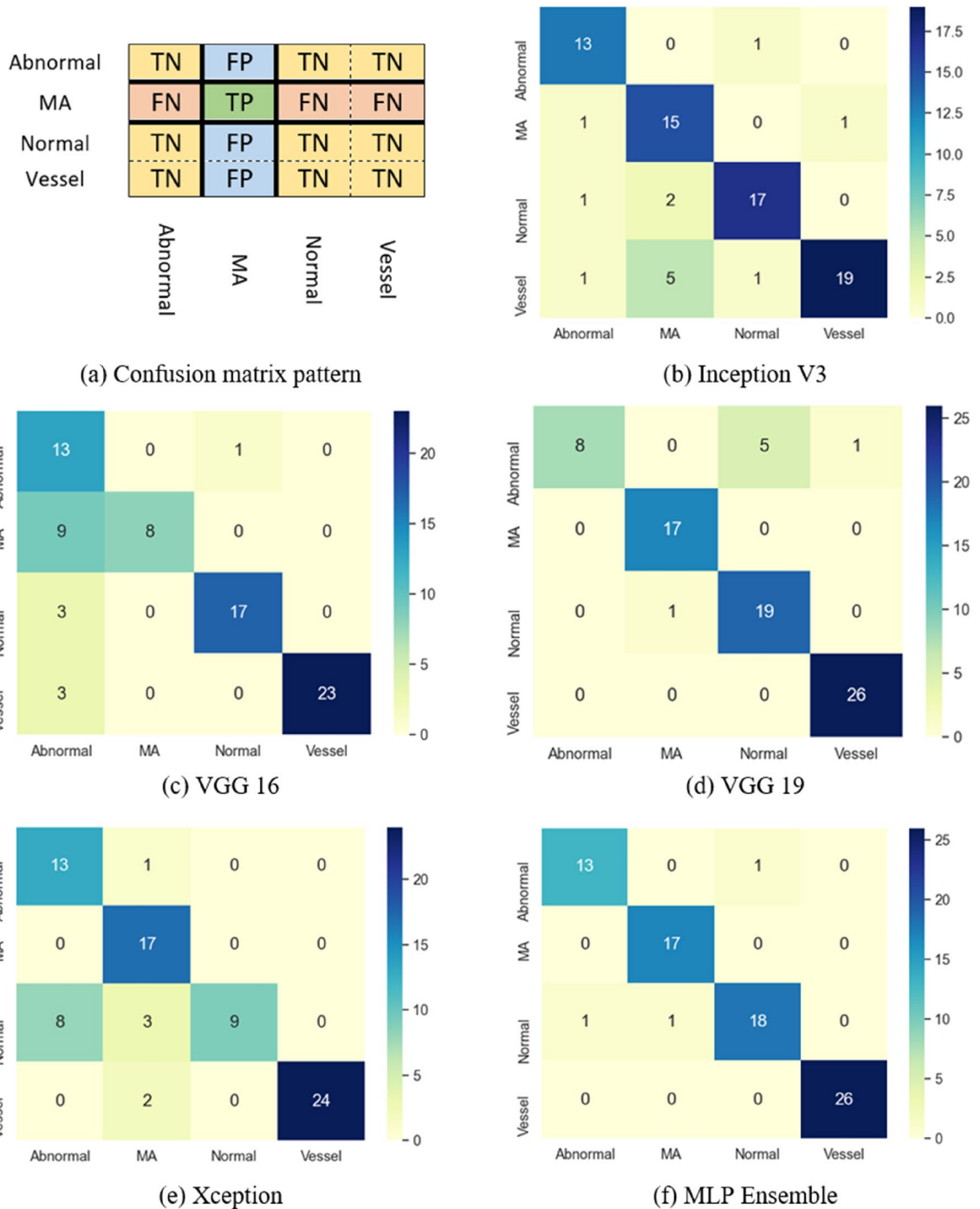
Model per class parameter	Inception V3	VGG16	VGG19	Xception	Method in <sup>29</sup>	MLP ensemble
<b>Accuracy</b>						
Abnormal	0.948	0.792	0.922	0.883	0.931	0.974
MA	0.883	0.883	0.987	0.922	0.962	0.987
Normal	0.935	0.948	0.922	0.857	0.943	0.961
Vessel	0.896	0.961	0.987	0.974	0.946	1
<b>Weighted average accuracy</b>	0.913	0.91	0.958	0.916	0.945	0.982
<b>Precision</b>						
Abnormal	0.813	0.464	1	0.619	0.952	0.928
MA	0.682	1	0.945	0.739	0.964	0.944
Normal	0.895	0.944	0.797	1	0.97	0.947
Vessel	0.95	1	0.963	1	0.964	1
<b>Weighted average precision</b>	0.851	0.888	0.921	0.873	0.963	0.961
<b>Recall (sensitivity)</b>						
Abnormal	0.929	0.929	0.571	0.929	0.86	0.928
MA	0.882	0.471	1	1	0.913	1
Normal	0.85	0.85	0.95	0.45	0.907	0.9
Vessel	0.731	0.885	1	0.923	0.883	1
<b>Weighted average recall</b>	0.831	0.792	0.909	0.818	0.890	0.961
<b>Specificity</b>						
Abnormal	0.952	0.762	1	0.873	0.866	0.984
MA	0.883	1	0.983	0.9	0.941	0.983
Normal	0.964	0.982	0.912	1	0.86	0.982
Vessel	0.980	1	0.980	1	0.887	1
<b>Weighted average specificity</b>	0.95	0.952	0.967	0.818	0.888	0.989
<b>F1-score</b>						
Abnormal	0.867	0.619	0.727	0.743	–	0.928
MA	0.769	0.64	0.971	0.85	–	0.971
Normal	0.872	0.895	0.863	0.621	–	0.923
Vessel	0.826	0.939	0.981	0.96	–	1
<b>Weighted average F1-score</b>	0.833	0.803	0.902	0.808	–	0.961
<b>Weighted average ROC-AUC score</b>	0.981	0.97	0.99	0.989	–	0.998

**Table 3.** Per class and weighted average measures for method in<sup>29</sup>, stand-alone CNNs and stacking ensemble of our method. Accuracy =  $(TP + TN)/(TP + TN + FP + FN)$ , Precision =  $TP/(TP + FP)$ , Recall(sensitivity) =  $TP/(TP + FN)$ , Specificity =  $TN/(TN + FP)$ , F1score =  $(2 \times \text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision}) = (2TP)/(2TP + FP + FN)$ , ROC-AUC score: the closer this criterion is to one, the greater the number of correctly predicted cases. On the other hand, the closer this criterion is to zero, the greater the number of incorrectly predicted cases.

incorrectly predicted in that class, and finally the FN (False Negative) indicates the number of items that were incorrectly predicted in other categories than the class under review.

As can be seen in Table 3, the above-mentioned measures can be expressed per class for stand-alone CNNs and the stacking ensemble. Also, the measures of different classes can be combined to have a single measure for the model. In the second case, the weighted average is used taking a weighted mean of the measures. The weights for each class are the total number of samples of that class. In addition, Table 3 indicates the weighted average measure for stand-alone CNNs and the stacking ensemble. As can be seen in Table 3, the stacking ensemble has the accuracy of 0.987, 0.961, 0.974, and 1 in classifying MA, normal, abnormal, and vessel strips. Also, from this Table, the overall accuracy of stacking ensemble is 0.982. Some misclassified examples can be found in Supplementary Fig. S4 online. According to Tables 3, proposed method outperforms the method presented in<sup>29</sup>. Figure 5b–f indicates confusion matrices resulted from performing classification utilizing different CNNs and stacking ensemble. The experiment is repeated on the dataset whose test data is prepared from the images of cases who have not been included in the training and validation processes. Confusion matrix is a matrix or table in a way that one axis represents true labels and the other one expresses predicted labels. In this matrix, according to the pair of the true and predicted values for each class label, the matrix entries or the table cell values are calculated. The relationship between confusion matrix and TP, TN, FP, and FN is shown in Fig. 5a for MA class. The results show that the ensemble outperforms each stand-alone CNN. The experiments are repeated on the dataset whose test data is prepared from the images of cases that have not been included in the training and validation processes, and the results are shown in Supplementary Fig. S5 and Supplementary Table 2 online.





**Figure 5.** Confusion matrices. (a) The relationship between confusion matrix and TP, TN, FP, and FN for MA class. (b–f) Confusion matrices of different CNNs and stacking ensemble (rows: True labels and cols: Predicted labels) (in color).

### Discussion

In this paper, a method is presented to detect MAs in OCT images using deep convolutional neural networks and transfer learning. For the lack of large labeled datasets which comprise OCT image strips with suitable labels, we have created our dataset. Because MAs and vessels are hard-to-detect objects in OCT images, to create OCT strips with appropriate labels, an accurate image registration method is performed to align OCT images and FA photographs. After that, with the help of corresponding FA images, the OCT strips are created from OCT B-scans in four labels, namely MA, normal, abnormal, and vessel. Once the OCT strips are created and organized as four-class dataset, a stacking ensemble comprising four fine-tuned, pre-trained CNNs is trained to classify the strips of OCT images into the mentioned classes. Once the model is obtained, it can be used to classify the strips

from the test B-scans without the need for FA images. The experimental results show that the proposed method classifies OCT image strips and specially detects OCT strips containing MA in a more precise way.

In the current study, we created and organized a dataset containing a limited number of OCT image strips with specific labels, namely MA, normal, abnormal, and vessel. For future works, increasing the number of samples of the dataset enables creating, using, and training deep convolutional neural networks from scratch. Also, analyzing information from the 3D nature of OCT imaging and especially neighboring B-scans can lead to more accurate classification results. For future work, and with a larger dataset, B-scans or OCT volumes can be taught to the CNNs according to their several labels such as normal, MA, etc. without the need for cropping operations, and test data can also be fed as B-scans or volumes to the network.

### Data availability

The authors declare that the main data supporting the findings of this study are available within the article and its Supporting Information files. Extra data are available from the corresponding author on a reasonable request.

Received: 15 October 2021; Accepted: 8 August 2022

Published online: 17 August 2022

### References

- Habib, M. *et al.* Detection of microaneurysms in retinal images using an ensemble classifier. *Inform. Med. Unlocked* **9**, 44–57 (2017).
- Lazar, I. & Hajdu, A. Retinal microaneurysm detection through local rotating cross-section profile analysis. *IEEE Trans. Med. Imaging* **32**, 400–407 (2012).
- Adal, K. M. *et al.* Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning. *Comput. Methods Prog. Biomed.* **114**, 1–10 (2014).
- Deepa, R. & Narayanan, N. in *IOP Conference Series: Materials Science and Engineering*. 012057 (IOP Publishing).
- Eftekhari, N., Pourreza, H.-R., Masoudi, M., Ghiasi-Shirazi, K. & Saeedi, E. Microaneurysm detection in fundus images using a two-step convolutional neural network. *Biomed. Eng. Online* **18**, 1–16 (2019).
- Fleming, A. D., Philip, S., Goatman, K. A., Olson, J. A. & Sharp, P. F. Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE Trans. Med. Imaging* **25**, 1223–1232 (2006).
- Giancardo, L. *et al.* in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 5939–5942 (IEEE).
- Giancardo, L. *et al.* in *Medical Imaging 2010: Image Processing*. 76230U (International Society for Optics and Photonics).
- Hatanaka, Y. *et al.* in *2018 International Workshop on Advanced Image Technology (IWAIT)*. 1–2 (IEEE).
- Hipwell, J. *et al.* Automated detection of microaneurysms in digital red-free photographs: A diabetic retinopathy screening tool. *Diabet. Med.* **17**, 588–594 (2000).
- Inoue, T., Hatanaka, Y., Okumura, S., Muramatsu, C. & Fujita, H. in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5873–5876 (IEEE).
- Junior, S. B. & Welfer, D. Automatic detection of microaneurysms and hemorrhages in color eye fundus images. *Int. J. Comput. Sci. Inform Technol* **5**, 21 (2013).
- Long, S. *et al.* Microaneurysms detection in color fundus images using machine learning based on directional local contrast. *Biomed. Eng. Online* **19**, 1–23 (2020).
- Qin, L., Ruixiang, L., Shaoguang, M. & Jane, Y. in *3rd International Conference on Multimedia Technology (ICMT-13)*. 1334–1341 (Atlantis Press).
- Sánchez, C. I., Hornero, R., Mayo, A. & García, M. in *Medical Imaging 2009: Computer-Aided Diagnosis*. 72601M (International Society for Optics and Photonics).
- Shan, J. & Li, L. in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. 357–358 (IEEE).
- Sopharak, A., Uyyanonvara, B. & Barman, S. Automatic microaneurysm detection from non-dilated diabetic retinopathy retinal images using mathematical morphology methods. *IAENG Int. J. Comput. Sci.* **38**, 295–301 (2011).
- Spencer, T., Olson, J. A., McHardy, K. C., Sharp, P. F. & Forrester, J. V. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput. Biomed. Res.* **29**, 284–302 (1996).
- Streeter, L. & Cree, M. J. Microaneurysm detection in colour fundus images. *Image Vision Comput. New Zealand*, 280–284 (2003).
- Valverde, C., Garcia, M., Hornero, R. & Lopez-Galvez, M. I. Automated detection of diabetic retinopathy in retinal images. *Indian J. Ophthalmol.* **64**, 26 (2016).
- Wu, J., Xin, J., Hong, L., You, J. & Zheng, N. in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 4322–4325 (IEEE).
- Zhang, B., Wu, X., You, J., Li, Q. & Karray, F. Detection of microaneurysms using multi-scale correlation coefficients. *Pattern Recogn.* **43**, 2237–2248 (2010).
- Zhang, L., Feng, S., Duan, G., Li, Y. & Liu, G. Detection of microaneurysms in fundus images based on an attention mechanism. *Genes* **10**, 817 (2019).
- Lemaître, G. *et al.* Classification of SD-OCT volumes using local binary patterns: Experimental validation for DME detection. *J. Ophthalmol.* **2016**, 1–14 (2016).
- Rasti, R., Rabbani, H., Mehridehnavi, A. & Hajizadeh, F. Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Trans. Med. Imaging* **37**, 1024–1034 (2017).
- Shih, F. Y. & Patel, H. Deep learning classification on optical coherence tomography retina images. *Int. J. Pattern Recognit. Artif. Intell.* **34**, 2052002 (2020).
- Tsuji, T. *et al.* Classification of optical coherence tomography images using a capsule network. *BMC Ophthalmol.* **20**, 1–9 (2020).
- Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 743–800 (2015).
- Kazeminasab, E., Almasi, R., Shoushtarian, B., Golkar, E. & Rabbani, H. Automatic detection of microaneurysms in OCT images using bag of features. *Comput. Math. Methods Med.* <https://doi.org/10.1155/2022/1233068> (2022).
- Kermany, D., Zhang, K. & Goldbaum, M. Labeled optical coherence tomography (oct) and chest x-ray images for classification. *Mendeley Data* <https://doi.org/10.17632/rscbjbr9sj.2> (2018).
- Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
- Almasi, R. *et al.* Registration of fluorescein angiography and optical coherence tomography images of curved retina via scanning laser ophthalmoscopy photographs. *Biomed. Opt. Express* **11**, 3455–3476 (2020).

33. Wang, H. *et al.* Characterization of diabetic microaneurysms by simultaneous fluorescein angiography and spectral-domain optical coherence tomography. *Am. J. Ophthalmol.* **153**, 861–867.e861 (2012).
34. Pangyuteng. *Graph-based segmentation of retinal layers in oct images* <<https://www.mathworks.com/matlabcentral/fileexchange/43518-graph-based-segmentation-of-retinal-layers-in-oct-images>>, MATLAB Central File Exchange (2013).
35. O'Malley, *et al.* KerasTuner, <[https://keras.io/api/keras\\_tuner/tuners/bayesian/](https://keras.io/api/keras_tuner/tuners/bayesian/)> (2019).
36. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* <https://arxiv.org/abs/1409.1556> (2014).
37. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1251–1258.
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.
39. Leonardo, M. M., Carvalho, T. J., Rezende, E., Zucchi, R. & Faria, F. A. Deep feature-based classifiers for fruit fly identification (Diptera: Tephritidae). in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. 41–47 (IEEE).

## Acknowledgements

Authors would like thank Prof. Sina Farsiu from Duke Eye Center for his valuable ideas and comments in starting and development of this study.

## Author contributions

R.A. has developed the stacking ensemble comprising four fine-tuned, pre-trained CNNs and wrote the manuscript. R.A. and E.K. have collected the dataset. In addition, E.K. has prepared Fig. 2 and done the baseline study. H.R. designed the main study. This work has been done under the supervision and guidance of A.V. and H.R. Also, A.V. and H.R. have helped in manuscript edit. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18206-8>.

**Correspondence** and requests for materials should be addressed to A.V. or H.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022