

Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier

Reza Rasti
Alireza Mehridehnavi
Hossein Rabbani
Fedra Hajizadeh

Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier

Reza Rasti,^{a,b} Alireza Mehridehnavi,^{a,b,*} Hossein Rabbani,^{a,b} and Fedra Hajizadeh^c

^aIsfahan University of Medical Sciences, School of Advanced Technologies in Medicine, Isfahan Department of Biomedical Engineering, Isfahan, Iran

^bIsfahan University of Medical Sciences, Medical Image and Signal Processing Research Center, Isfahan, Iran

^cNoor Eye Hospital, Noor Ophthalmology Research Center, Tehran, Iran

Abstract. The present research intends to propose a fully automatic algorithm for the classification of three-dimensional (3-D) optical coherence tomography (OCT) scans of patients suffering from abnormal macula from normal candidates. The method proposed does not require any denoising, segmentation, retinal alignment processes to assess the intraretinal layers, as well as abnormalities or lesion structures. To classify abnormal cases from the control group, a two-stage scheme was utilized, which consists of automatic subsystems for adaptive feature learning and diagnostic scoring. In the first stage, a wavelet-based convolutional neural network (CNN) model was introduced and exploited to generate B-scan representative CNN codes in the spatial-frequency domain, and the cumulative features of 3-D volumes were extracted. In the second stage, the presence of abnormalities in 3-D OCTs was scored over the extracted features. Two different retinal SD-OCT datasets are used for evaluation of the algorithm based on the unbiased fivefold cross-validation (CV) approach. The first set constitutes 3-D OCT images of 30 normal subjects and 30 diabetic macular edema (DME) patients captured from the Topcon device. The second publicly available set consists of 45 subjects with a distribution of 15 patients in age-related macular degeneration, DME, and normal classes from the Heidelberg device. With the application of the algorithm on overall OCT volumes and 10 repetitions of the fivefold CV, the proposed scheme obtained an average precision of 99.33% on dataset1 as a two-class classification problem and 98.67% on dataset2 as a three-class classification task. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.23.3.035005]

Keywords: classification; convolutional neural networks; feature learning; macular disease; retinal optical coherence tomography; spatial-frequency information.

Paper 170712R received Nov. 3, 2017; accepted for publication Feb. 27, 2018; published online Mar. 21, 2018.

1 Introduction

Optical coherence tomography (OCT) is a well-known noninvasive imaging technique providing three-dimensional (3-D) images with microscopic resolution (1 to 15 μm).¹ This is the most frequently used imaging technique in ophthalmology since it makes possible the cross-sectional visualization of inner structures. From a clinical point of view, this is a very important ability because it makes possible the early diagnosis of retinal diseases, such as diabetic edema, and the monitoring of the response to treatment.

The retina contains two main regions called macula and optic nerve head. Being responsible for the central vision, the macula is located near the central area of the retina. The main ophthalmic diseases in this area include diabetic macular edema (DME) and age-related macular degeneration (AMD). These pathologies are the major causes of the loss of central vision or even blindness at different ages.^{2,3}

To investigate the macular pathologies in clinical circumstances, ophthalmologists manually explore various abnormalities, such as fluid regions, cystic structures, exudates, and drusens

at each B-scan of the retinal OCT volume. Then, they make a cumulative decision on the type of disease. This tedious routine is a time-consuming and error-prone analysis, so it may yield subjective results especially for elderly stage macular diseases evaluation. Such issues increase the importance of developing computer-aided diagnostic (CAD) systems in retinal OCT. CAD systems can be of great help in providing professional consultations to ophthalmologists in a shorter time. They also enable the remote identification of ocular diseases in public screening programs.⁴

Different computerized algorithms have, therefore, been developed for analysis of the retinal OCT data in the last few years. Some of these algorithms benefit from sophisticated image processing techniques in OCT data analysis field, such as denoising and contrast enhancement,^{5,6} segmentation of retinal layers,⁷⁻¹⁴ segmentation of abnormalities such as regions or cystic structures,¹⁵⁻¹⁹ and also retinal layers alignment^{20,21} in the first steps of the procedures. However, feature extraction and classification techniques²¹⁻²⁸ generally constitute the main subsequent steps of all of these diagnostic algorithms. A brief review of the recent related works is presented as follows.

*Address all correspondence to: Alireza Mehridehnavi, E-mail: mehri@med.mui.ac.ir

Liu et al.²² proposed a multiscale local binary pattern (LBP) feature extraction step and a nonlinear support vector machine (SVM) method for the classification of macular pathologies (i.e., macular edema, macular hole, and AMD). In another study, Srinivasan et al.²³ employed a feature extraction method based on histogram of oriented gradients (HOG) and fed the features to three linear SVM classifiers for the purpose of discrimination between DME, AMD, and normal OCT volumes. The research utilized a preprocessing stage composed of block matching and 3-D-filtering (BM3D) denoising²⁹ and retinal curvature flattening steps. Based on a threshold of 33% of abnormal B-scans for decision-making on a dataset of 45 OCTs, this method achieved a classification rate of 86.67%, 100%, and 100% for normal, DME, and AMD classes, respectively. Hassan et al.²⁶ proposed a feature extraction methodology based on structural tensors. They extracted three thickness profiles and two cyst fluids features for the classification of macular edema, central serous retinopathy, and healthy ones. In Ref. 30, after segmentation of the retinal pigment epithelium (RPE) layer, binary features were computed from the RPE layer to identify AMD and DME pathologies. Koprowski et al.³¹ extracted morphological and textural features of the choroid in OCT images to detect the scarring fibro-vascular tissue, neovascular AMD, and diffuse macular edema. Venhuizen et al.²⁴ proposed a method for unsupervised feature learning³² followed by the bag-of-words approach³³ for discrimination between AMD and normal OCT volumes. The method gained an area under the receiver operating characteristic curve (AUC) of 0.984 in a dataset of 384 retinal OCTs. With the same OCT dataset, an automatic AMD identification method was proposed in Ref. 34 based on convolutional neural networks (CNNs)³⁵ with an AUC of 0.997. For this purpose, the method remapped the OCT volumes to large image mosaics and trained a two-dimensional (2-D) CNN (called RetiNet-C) for the classification of retinal OCTs. Recently, Sun et al.²¹ proposed a macular pathology detection algorithm in OCT images using sparse coding and dictionary learning. After the application of the preprocessing steps consisting of BM3D denoising and retinal curvature correction, the authors performed a dictionary learning technique on shift-invariant feature transform features on partitioned B-scans. Then, they used three two-class linear SVM classifiers for discrimination between normal, DME, and AMD OCT volumes with a classification rate of 93.33%, 100%, and 100%, respectively, on a dataset of 45 OCTs²³ using the majority voting for decision-making. With the same dataset as a part of the study,²⁸ we introduced a multiscale convolutional mixture model to automatically classify the AMD and DME macula from healthy ones. By assessing aligned OCTs and using a diagnostic threshold value of 15% on abnormal B-scans, the method achieved a precision (Pr) rate of 98.33%.

The most recent studies demonstrated that the feature learning from OCT data is a more effective strategy than hand-crafted features in the retinal OCT diagnosis. In this research, adopting the above notion, we propose a fully automated system for identifying different pathologies in retinal OCT volumes, which is termed as the wavelet-based convolutional neural network feature learning with random forests classification (WCNN-RF). With the help of two real retinal OCT datasets captured from different imaging devices, the proposed system tries to address the following issues:

- I. **Minimum preprocessing requirements:** Although preprocessing is performed in many OCT classification methods (such as Refs. 21–23 and 28), since extraction of the retinal boundaries can be challenging, especially for severe abnormal cases, it is desired to detect the ocular diseases without any emphasis on the retinal layer segmentation or curvature correction. Indeed, retinal OCT images are affected by speckle noise and it seems that we need a robust method against the noise corruption. Therefore, CNN-based methods are useful tools, because they are known for their robustness against image noise and distortions. In addition, these models are efficiently shape-, intensity-, and scale-invariant due to their shared-weights architecture.^{36,37}
- II. **Adaptive feature extraction:** The data-driven and task-dependent feature learning procedure that occurs in hidden layers of CNNs is the main prominent advantage of these kinds of intelligent models compared to the hand-crafted feature extraction methods.^{26,30,31} As in Refs. 27, 28, and 34, the proposed system benefits from this ability of CNNs but in the form of a convolutional feature extractor.
- III. **Generalization:** In contrast to Ref. 34, we propose a system with the ability to classify the input OCT volumes with a different number of B-scans (different slicing of OCT data). This method considers the correlation among different B-scans of the input OCT volume in the feature extraction stage. It makes a volumetric diagnosis directly and avoids classic assumptions about the final decision-making ties, such as thresholding or the majority voting.
- IV. **Speed consideration:** The proposed system in this study tries to implement feature learning and classification stages in a fast and efficient mode by considering minimum computational costs for its subsystems.

The rest of the paper is organized as follows: Sec. 2 presents the proposed framework for ocular pathology identification. In this section, having introduced the research datasets, the proposed convolutional model (i.e., WCNN) for retinal OCT image representation and feature learning is described in detail. Section 3 describes the evaluation results. This section includes some baseline studies to evaluate the proposed algorithm. Section 4 presents a comprehensive discussion of the WCNN-RF model and experimental results. Finally, Sec. 5 provides the conclusion of the research and future directions.

2 Materials and Methods

2.1 Optical Coherence Tomography Datasets

Two different SD-OCT datasets were considered for this study. The first one was obtained from the Topcon 1000 device and consists of 30 normal and 30 DME OCTs. Each 3-D-OCT data from this dataset were composed of 128 slices sized 650×512 pixels. The second one is an online available dataset from the Heidelberg device (Heidelberg Engineering Inc., Heidelberg, Germany) that consists of 15 normal, 15 AMD, and 15 DME cases.²³ The OCT data in this dataset include a range from 31 to 97 B-scan slices with the size of 512×496 or 768×496 pixels. In addition to the provided case labels, all B-scans in the two research databases were annotated by an

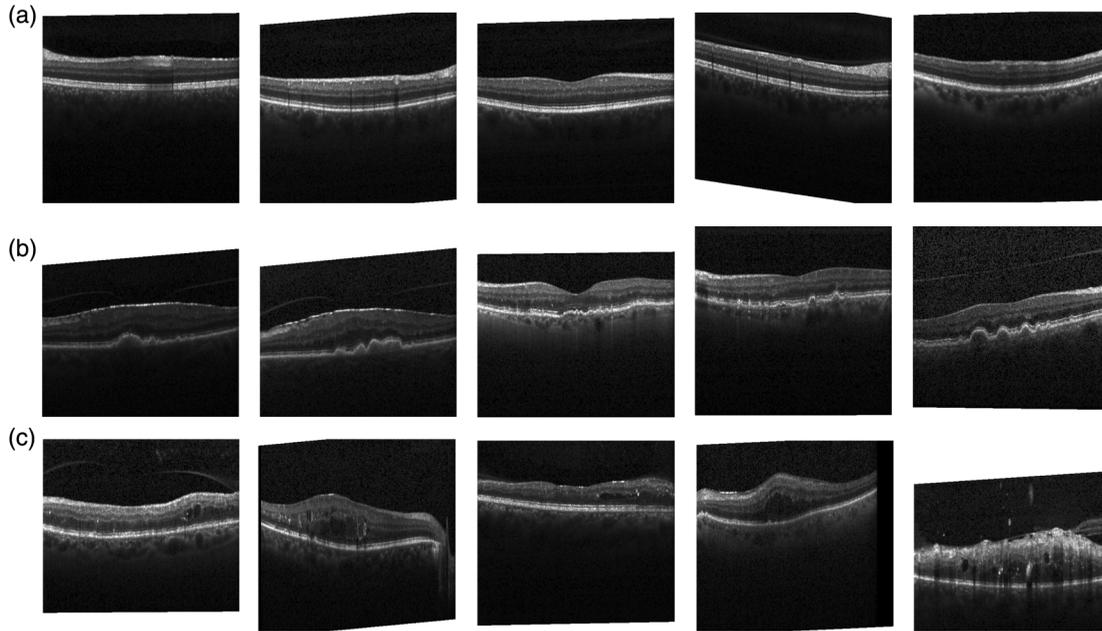


Fig. 1 Sample B-scans: (a) normal, (b) AMD, and (c) DME subjects in dataset2.

expert ophthalmologist experienced in OCT imaging. Figure 1 shows sample B-scans from different volumes of normal, AMD, and DME classes.

2.2 Regular Convolutional Neural Networks

CNN initially proposed by LeCun et al.³⁵ is an image-based neural network model that captures the main spatial information of the input data. Principally, this model is designed and tested for the recognition of 2-D images, such as handwritten digit images. A regular CNN model, generally, consists of three main types of layers:^{38,39} (i) convolutional layers (C-layers), (ii) pooling layers (P-layers), and (iii) fully connected layers (FC-layers). Other CNN layers exist for recently published CNNs such as batch-normalization layers (BN-layers)⁴⁰ and dropout⁴¹ for creating more efficient convolutional networks. In a regular CNN model, layers are arranged in a feedforward structure: stacks of hidden C-P layers (CONV-POOL), some hidden FC-layers, and a final FC-layer called output layer (O-layer). In CNNs, each 2-D layer (C- and P-layers) has several extracted planes which are called as layer's output feature maps (FMs).

2.3 Proposed Approach

In the field of machine vision, a regular CNN performs a hierarchical multiscale modeling of input data for solving problems, which have important features at multiple scales of spatial information. This procedure depends completely on a free run learning process (a time-consuming task of learning thousands of free parameters) to build high-level representations and FMs. Therefore, as the spatial size and complexity of the input data are increased, the efficiency of regular CNNs may be decreased.⁴² Moreover, an important issue in pattern recognition tasks is to analyze different frequency components in data, including high-frequency components such as the edges and corners. So, if we are able to force the CNN to consider different level frequency maps of the input data directly, the computational effort can be reduced by the model to build high-level representations. Moreover, it is possible to have

smaller networks with acceptable and promising performance. One suitable strategy for this aim is to apply the wavelet transform (WT).⁴³ By analysis of the image spatial and frequency characteristics at multiple resolutions, the WT provides a powerful unsupervised representation for image processing. In fact, a combination of different frequency maps information presented by WT subbands causes to attain CNNs with comparable efficiency and lower time complexity.

In this work, we propose a two-stage scheme for the retinal OCT volume classification task which includes: (1) volumetric feature extraction and (2) diagnostic classification. The scheme benefits from the above idea in the feature extraction stage by means of a wavelet-based CNN (WCNN) feature learning subsystem. The WCNN includes a spatial-frequency decomposition layer (SFD-layer) in the first hidden layer of the model and it is exploited as an effective feature learning method for retinal OCT B-scans.

2.3.1 Spatial-frequency decomposition layer

An SFD-layer condenses the input map first by a j -level 2-D orthogonal discrete wavelet transform (DWT). Once the input map is decomposed to different scales, the wavelet coefficient subbands are normalized (with Z-score normalization method⁴⁴) and then convolved by different 2-D kernels of neural weights. After the adding of scale-dependent biases, they are considered as the output FM of the layer. In SFD-layer l , n 'th output FM is calculated as

$$\mathbf{o}_n^l = f_l[\text{DWT}_{j,n}(\mathbf{o}^{l-1}) \otimes \mathbf{w}_n^l + b_n^l], \quad (1)$$

where f_l is the activation function of the layer, \mathbf{o}^{l-1} is the output FM of previous layer, and \mathbf{w}_n^l and b_n^l are the adaptive kernel and bias terms, respectively, associated with n 'th FM in the layer. The choice of DWT type for this layer depends upon the input data and the application. Figure 2 shows a typical 2-D SFD-layer. In the SFD-layer, it is assumed that all subbands provided by the DWT block are of the same size. For a one-level

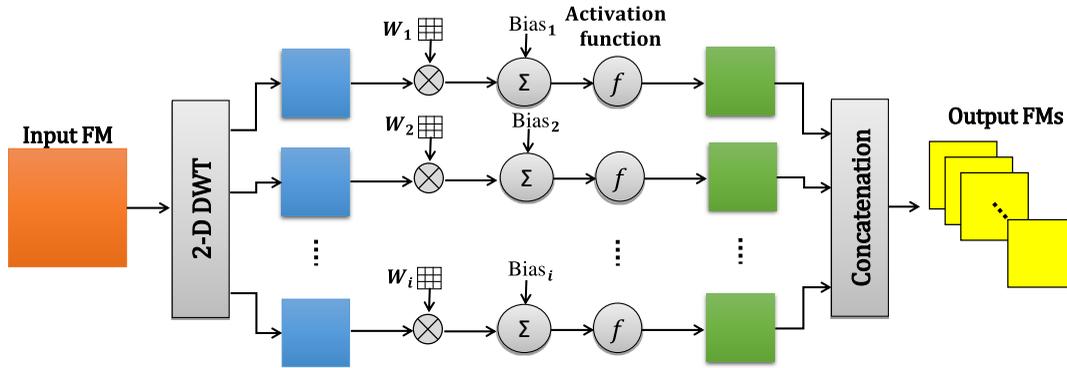


Fig. 2 Schematic diagram of a typical 2-D SFD-layer.

DWT, it needs no further processes. However, for a two, three, and more level DWT, a 2×2 max-pooling filtering is applied for the detailed subbands in the block. This procedure generates output FMs with an identical size in the SFD-layer to feed to the consecutive layers.

By means of the SFD-layer, CNN models benefit from the advantages of different domain multiresolution decomposition both in width and in depth with integrating spatial-frequency information at multiple scales.

2.3.2 Wavelet-based convolutional neural network model for feature learning

In Fig. 3, the proposed WCNN model is demonstrated. The parameters of the model are optimized by training B-scans as the 2-D inputs and the corresponding ground truths. Given a test volume, the output of the last BN layer is considered as the CNN codes for different B-scans in the input volume. In fact, these codes are the learned features at the B-scan level.

In this work, the SFD-layer with the 2-D Daubechies wavelets at one-, two-, and three-level decomposition was used as the first layer of the WCNN. Therefore, the performance effect of the imposed spatial-frequency details of the input data was investigated. The choice of the type of DWT depends upon the input data to be analyzed and the location of the SFD-layer in the WCNN model. Generally, first layers in the recent successful CNN models include some extracted FMs with coarse details. To conform to this attribute, the Daubechies wavelet was found to give more accurate and coarse details for the first hidden layer than other wavelets, such as Haar, biorthogonal, Coiflets, Morlet, and Meyer for retinal OCT image representation.⁴²

Wavelet-based convolutional neural network training algorithm. Training of the WCNN models is based on the batch error backpropagation (BP) method and mean square

error (MSE) objective function. Numerous optimization algorithms can be applied for minimizing the error gradients of different layers in the model.⁴⁵ In this work, for training the WCNN model, the mini-batch Adam method was used as the first-order gradient-based optimization approach.⁴⁶

2.3.3 WCNN-RF structure for retinal optical coherence tomography diagnosis

This section introduces the proposed method for discriminating normal retinal OCT volumes from abnormal macula classes (i.e., DME and/or AMD). The main blocks of the WCNN-RF pipeline are outlined in Fig. 4 and the details are described in the following sections.

Preprocessing. In this block, we generate a volume of interest (VOI) of the input OCT to reduce the time complexity of the whole algorithm by forcing the model to process relevant information. So, for a given OCT volume, the most important regions of different B-scans are cropped, which contain main morphological information of retinal layers. The main steps for this purpose are as follows: first, a preparing process is needed. In the research databases, the B-scans in different subjects and imaging systems have various sizes with possible missing background data.

The missing data are regions with an intensity value of 255. To handle these issues, all B-scans are first resized to 512×496 pixels, and the missing regions are compensated by means of the “imfill” morphological operation⁴⁷ with an intensity value of zero similar to the image background. Second, we perform a cropping step. For this purpose, the middle row position of the maximum intensity values in B-scans of current OCT volume is selected as the central row of the case. Then, for each B-scan, 256-row pixels around the calculated central row are selected as the cropped image (i.e., 135 rows above and 120

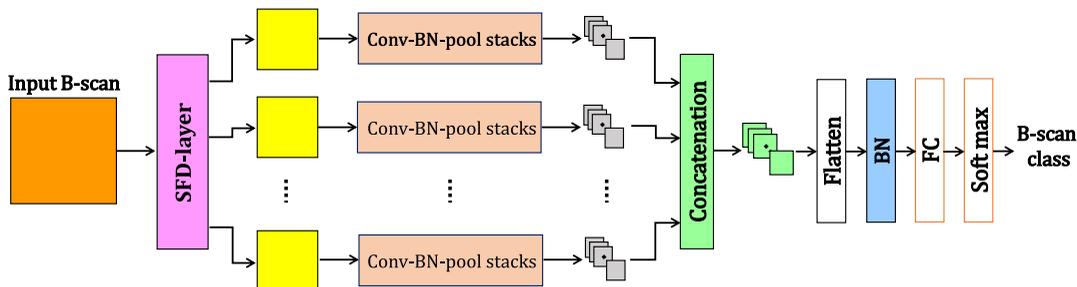


Fig. 3 The proposed WCNN model for B-scan's feature representation.

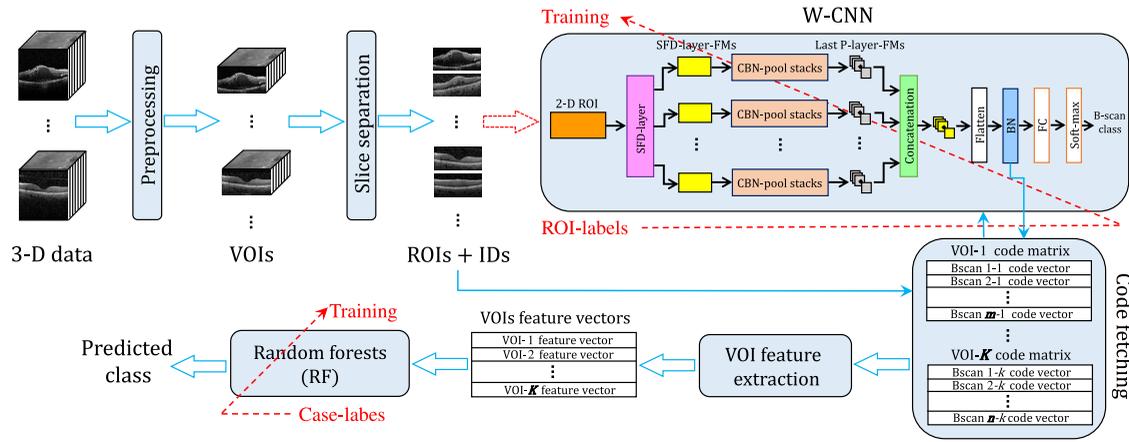


Fig. 4 Overview of the proposed WCNN-RF scheme for classification of retinal OCT volumes. This figure consists of both the training and testing phases. For the testing phase, only the solid arrows are the active paths.

rows below empirically). In some cases with very low or high central row (i.e., severely misaligned data), 256 rows located on top or bottom of the image are selected for cropping purpose. Finally, all of these cropped B-scans are concatenated to generate the VOI of the current OCT data.

Slice separation. The target of this block is to generate the training and testing region of interest (ROI) collections with corresponding ground truths. Also, the case IDs are reserved for all B-scans in the VOIs for diagnostic evaluation purpose at the patient level. In the first step, here, a centered 256×470 pixels bounding box is defined as a field of view (FOV) in a preprocessed B-scan. This FOV is used to generate central ROIs for a given VOI. In the training phase for generalization of the problem and to have an efficient training process, the selected FOVs in training cases are horizontally flipped, translated by $[\pm 10, \pm 20]$ pixels, and/or rotated by $[\pm 3 \text{ deg}, \pm 5 \text{ deg}]$ angles to generate augmented training sets. This augmentation trend increases the number of samples with a factor of 18 in our training process. Furthermore, all the extracted ROIs are resized to 128×256 pixels for subsequent processes. In the testing phase, only the resized central ROIs in a given volume are considered for the evaluation purpose. A sample result of the ROI selection process is demonstrated in Fig. 5 for a Heidelberg B-scan.

Wavelet-based convolutional neural network and code-fetching blocks. In the early phase of learning, the WCNN

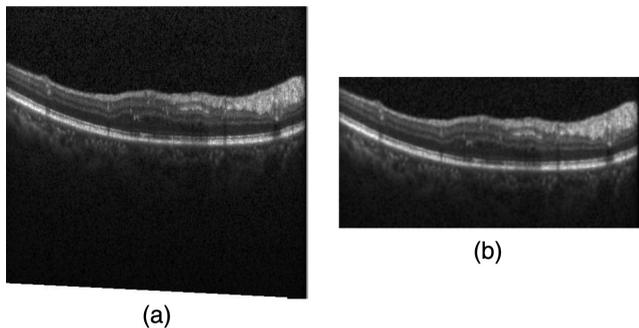


Fig. 5 ROI selection: (a) original DME slice from the Heidelberg OCT dataset with the size of 768×496 pixels and (b) extracted central ROI with 128×256 pixels size.

is trained with augmented training B-scans and the corresponding ground truths. When the training process is completed in WCNN block, the model is used as the CNN code extractor for each B-scan in the volumes. To do that, the output values of the normalization layer in the trained WCNN model are fetched by the code-fetching block (e.g., with a dimension of $1 \times v$). These values are stacked with considering the ID indices to generate a code matrix for each input volume ($\mathbf{X} = [\mathbf{v}_1; \mathbf{v}_2; \dots; \mathbf{v}_M] \in \mathbb{R}^{M \times V}$). In fact, these code matrices are the primary learned features for input volumes. In the testing phase, the above strategy is conducted without any learning consideration for the WCNN block.

Volume of interest feature extraction. In this block, a global feature representation is built for each OCT volume. For this purpose, the code matrix of each retinal OCT (i.e., \mathbf{X} matrix) is mapped to a vector of representative features. As mentioned before, different OCT volumes may consist of different number of B-scans and obtain code matrices with various sizes [e.g., $m \times v$ matrices with various numbers of m (rows) for different cases]. To handle this diversity, the following strategy is applied; in a given code matrix with a size of $m \times v$, mean, standard deviation, and maximum values are extracted from each column (which corresponds to a specific CNN code for different slices) to generate a final $1 \times (3 \times v)$ vector as the final representative features for the given OCT volume.

Random forests classifier. In the proposed framework, a random forests (RF) classifier⁴⁸ is used as the final decision maker, which is exploited at the patient level. After training the RF with volumetric extracted feature vectors and the corresponding case-level ground truths, it will be ready to be used in the testing phase for evaluation purpose.

3 Experimental Design and Results

3.1 Baseline Studies

As the first baseline study, to obtain a criterion for the comparison of the performance of the proposed scheme in the research databases, two recent feature-based methods were considered. These two approaches were a multiscale feature extraction via LBP²² and HOG²³ followed by SVM classification method. As the second study, to evaluate the SFD-layer proficiency in the

proposed WCNN feature learner, a CNN-based framework (hereafter called CNN-RF framework) was considered with topology similar to the proposed scheme and without any SFD-layer. This baseline was compared based on the performance results and also the time complexity of the overall scheme. It should be noted that the baselines were evaluated based on the extracted VOIs described before in Sec. 2.3.3 in preprocessing paragraph.

3.2 Evaluation Setup

3.2.1 Fivefold cross validation

In this study, 10 repetitions of the unbiased fivefold cross-validation (CV) method were applied at the patient level. The generated VOIs, according to Sec. 2.3.3, are used to train and evaluate the diagnostic efficacy of the proposed scheme and the baselines. For evaluation purpose, in each repetition, the Topcon dataset was reshuffled initially and partitioned into 5 case folds of 12 patients (6 normal versus 6 DME cases). By applying the augmentation method, 648 VOIs (i.e., 31,860 ROIs) for training the convolutional models were extracted on average per iteration. Similarly, for the Heidelberg dataset, the extracted ROIs were partitioned randomly 10 times into fivefolds constituted of nine different patients (three cases for each class). According to the augmentation approach, 864 VOIs (i.e., 21,870 ROIs) for training the convolutional models were considered on average per iteration. In addition, the subsequent learning of the RF classifier for the volumetric decision-making

was performed according to the corresponding training labels at the patient level.

3.2.2 Performance measures

Diagnostic performance in this study was computed according to the confusion matrix analysis and the values of accuracy (Acc), Pr, recall (Re), MSE, and also AUC curve.

3.3 WCNN-RF Scheme Characterization

Here, we start with this hypothesis that an efficient algorithm for retinal OCT diagnosis should be high performance at the B-scan level classification to build discriminative features. So, we investigated the proposed WCNN feature learner model by optimizing the SFD-layer in the model and also three different levels of DWT decomposition.

3.3.1 B-scan level analysis of wavelet-based convolutional neural network

To assess the SFD-layer effect on the overall performance of the proposed model, the WCNN structure was investigated by performing the following two different studies and considering the WCNN models in Table 1. According to a grid search on a nested fivefold CV within the training sets, the performed studies were:

- **One-level SFD-layer investigation:** For a specific WCNN model (i.e., WCNN1), different neural kernels

Table 1 WCNN structures detail for the two-class classification problem.

WCNN 1		WCNN 2		WCNN 3	
Layer name	Kernel size	Layer name	Kernel size	Layer name	Kernel size
Model configuration					
SFD1	$4 \times 3 \times 3$	SFD1	$7 \times 3 \times 3$	SFD1	$10 \times 3 \times 3$
CBN2	$4@4 \times 3 \times 3$	CBN2	$7@3 \times 3 \times 3$	CBN2	$10@2 \times 3 \times 3$
P3	$4@4 \times 2 \times 2$	P3	$7@3 \times 2 \times 2$	P3	$10@2 \times 2 \times 2$
CBN4	$4@4 \times 3 \times 3$	CBN4	$7@3 \times 3 \times 3$	CBN4	$10@2 \times 3 \times 3$
P5	$4@4 \times 2 \times 2$	P5	$7@3 \times 2 \times 2$	P5	$10@2 \times 2 \times 2$
CBN6	$4@4 \times 3 \times 3$	CBN6	$7@3 \times 3 \times 3$	Flatten	—
P7	$4@4 \times 2 \times 2$	P7	$7@3 \times 2 \times 2$	BN	—
CBN8	$4@4 \times 3 \times 3$	Flatten	—	O6	$2 \times 1 \times 1$
P9	$4@4 \times 2 \times 2$	BN	—	NTP = 2502	
Flatten	—	O8	$2 \times 1 \times 1$		
BN	—	NTP = 3908			
10	$2 \times 1 \times 1$				
NTP = 4562					

Note: CBN is a unit, which consists of a convolutional layer and a BN layer, NTP indicates the number of trainable parameters, and the sign of @ implies the number of parallel branches in the models.

with the size of 1×1 , 3×3 , 5×5 , and 7×7 were investigated in the SFD-layer.

- **Multilevel SFD-layer evaluation:** Three different structures of WCNN were explored, which include one-level SFD-layer (WCNN1), two-level SFD-layer (WCNN2), and three-level SFD-layer (WCNN3).

For training the WCNN models, considering the Adam optimization method,⁴⁶ the learning rate, β_1 , β_2 , ϵ , decay, and max-epoch were tuned to be 0.001, 0.9, 0.999, 1×10^{-08} , 1×10^{-4} , and 50, respectively. Furthermore, the mini-batch training size of 16, 32, 64, and 128 was explored for all investigated models. Moreover, for SFD-layers, C-layers, P-layers, and the output layer, the activation functions were considered to be “ReLU,”

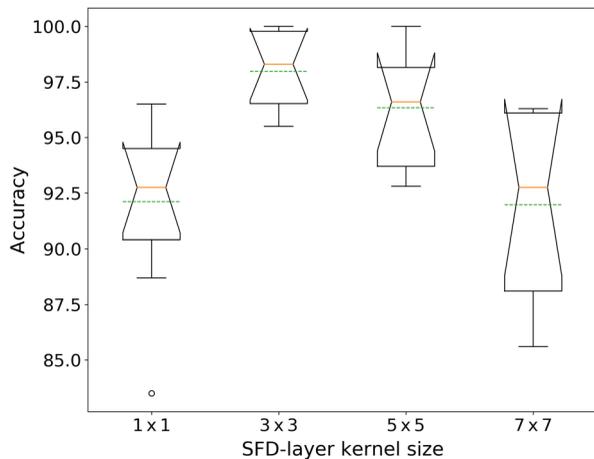


Fig. 6 The effects of the SFD-layer kernel size on the Acc measure for the WCNN1 model at the B-scan level on dataset1. The SFD-layer was considered to include the “ReLU” activation function.

Table 2 Test performance of the WCNN models on the Topcon database at the B-scan level.

Performance	Evaluated models		
	WCNN 1	WCNN 2	WCNN 3
Acc (%)	97.98 ± 1.7	95.33 ± 3.2	92.13 ± 4.5

Note: The best Acc value is indicated in bold.

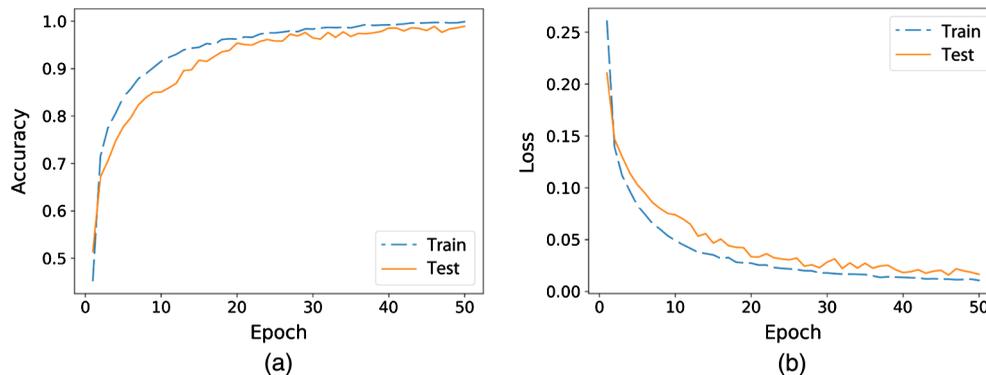


Fig. 7 Comparison of the WCNN1 training and testing phases on Topcon dataset based on the fivefold CV method at B-scan level: (a) Acc curves and (b) MSE curves.

“ReLU,” “Linear,” and “Softmax” functions, respectively. To prevent probable overfitting during the training process, a drop-out factor of 60% was also considered for the flattened layers. The considered WCNN models are introduced in detail in Table 1. Note that for the three-class classification problem (i.e., the Heidelberg data), O-layers had three output neurons.

For this examination, the Topcon dataset was considered and evaluated based on 10 repetitions of the fivefold CV results at the B-scan level. Indeed, the optimum batch size for learning of these models was 32 B-scans. Figure 6 exhibits a comparison among different kernel sizes in the SFD-layer for WCNN1. This study showed that the kernel size of 3×3 pixels was the best nomination for the SFD-layer kernel size in analysis of the retinal OCT B-scans.

Table 2 reports the performance results of the evaluated WCNN models. According to the table, WCNN1 outperforms the other models, so it is the best choice to consider as the CNN code extractor in the overall WCNN-RF framework. To provide more insights on the WCNN1 performance at the B-scan level, Fig. 7 includes average plots of Acc versus iteration and loss versus iteration functions for the train and test folds in the CV5 for the Topcon dataset.

3.3.2 C-scan level analysis of the proposed WCNN-RF framework

Table 3 reports the average performance of the LBP, HOG, CNN-RF baselines, and the proposed WCNN1-RF framework at the patient level based on the fivefold CV. For the CNN-RF framework, we considered a topology similar to the WCNN1 for feature learning step, where the SFD-layer was substituted with a stack of C-P layers. For the two-class classification problem (i.e., Topcon dataset), this baseline framework includes 4562 free parameters, the same as the WCNN1. For the CNN-RF and the WCNN1-RF frameworks, the number of fetched CNN codes for each B-scan was 192 scalar codes, which finally mapped to a $1 \times (3 \times 192)$ feature vector for each input OCT volume in the feature extraction block.

In addition, the RF classifier was explored to have 500, 1000, 2000, and 3000 trees with the max-depth of equal to the number of features ($n = 3 \times 192$). The experimental exploration showed that the RF with 1000 trees outperformed its other configurations.

To assess the generalization ability and robustness of the proposed framework and the settings, we combined the two Topcon and Heidelberg datasets into one. This dataset was evaluated

Table 3 Baseline classification performance on the research databases.

Method	Database	Classification performance			
		Pr (%)	Re (%)	MSE	AUC
Multiscale LBP + RBF SVM ²²	Topcon	95.38 ± 3.8	94.95 ± 3.9	0.077	0.959
	Heidelberg	92.88 ± 4.9	92.27 ± 4.8	0.152	0.942
Multiscale HOG + linear SVM ²³	Topcon	95.71 ± 3.4	95.31 ± 3.5	0.061	0.960
	Heidelberg	94.09 ± 4.6	93.47 ± 4.5	0.122	0.951
CNN-RF framework	Topcon	99.00 ± 1.2	98.67 ± 1.4	0.013	0.990
	Heidelberg	98.17 ± 1.4	97.56 ± 1.9	0.025	0.985
WCNN1-RF framework	Topcon	99.33 ± 0.8	99.11 ± 1.1	0.009	0.993
	Heidelberg	98.67 ± 1.2	98.22 ± 1.7	0.018	0.989

Note: All best values are indicated in bold characters.

by the proposed approach based on 10 repetitions of the CV5, in which the average Pr criterion was computed to be 96.45% ± 2.9.

All convolutional models were implemented in Python 2.7 using the Theano v0.8.2⁴⁹ and Keras v1.2⁵⁰ Toolkits. Training of the networks was executed on an NVIDIA GTX 1080-8GB graphic card, Cuda Toolkit v8.0, and accelerating cuDNN library v5.1. Main codes were run with Corei7 CPU at 3.4 GHz (Intel 6800K: 15M), and 32 GB of RAM. For the time complexity comparison, overall training phase of the WCNN1-RF framework took 10.2 s/VOI on average for both datasets. This time was 11.1 s/VOI for the CNN-RF framework. It should be noted that once the WCNN-RF framework trained it took about 1.4 s to analyze an OCT volume including 128 retinal B-scans.

4 Discussion

In this study, we proposed and evaluated a fully automatic system for the diagnosis of retinal pathologies in 3-D OCTs. The proposed WCNN-RF algorithm did not rely on the routine computerized processes, such as denoising, segmentation of retinal layers, and also retinal curvature correction. This is a significantly important feature when dealing with severe retina diseases where segmentation and alignment of pathological retinas are very challenging tasks.

The proposed system included two learning stages: (i) adaptive feature learning and (ii) classifier learning. In adaptive feature learning stage, the authors introduced a convolutional neural model based on wavelet decomposition in CNNs for benefiting from spatial-frequency information fusion, which included a hidden layer named as the SFD-layer. They also addressed a strategy for feature extraction of 3-D OCTs in the system. In the classifier learning stage, classification of representative and data-driven features of input volumes performed via a RF classifier at the patient level.

The system evaluated on two different datasets and diagnostic problems based on fivefold CV method: (i) the diagnosis of DME and normal cases in a Topcon dataset of 60 subjects with a Pr of 99.33% and (ii) the diagnosis of AMD, DME, and normal cases in a Heidelberg dataset of 45 patients with a Pr of 98.67%.

Experimental results in Table 3 showed that the WCNN1-RF outperformed the considered baseline methods (i.e., LBP-SVM,²² HOG-SVM,²³ and CNN-RF frameworks) in terms of performance measures on both datasets. The results confirm the WCNN1-RF's strength in generating more discriminative features and classification of retinal OCT data. In fact, the SFD-layer imposes the CNNs to have a greater depth for data representation with considering different frequency information. Most likely, when one or more frequency maps (mapped sub-bands) are not closely relevant for discriminative information fusion for a specific class, another one can be efficiently used. This capability allows the WCNN1-RF to have less error than the comparable spatial domain CNN-RF model. In Fig. 8, the middle and output FMs of the SFD-layer in WCNN1 model are depicted for a sample OCT B-scan image, in which the middle FMs are the one-level 2-D Daubechies wavelet subbands.

Although the recent thresholding techniques used in Refs. 23 and 28 are effective trends to design a CAD system in retinal OCT with acceptable sensitivity, they depend entirely on the stages of the diseases in the target database. Ideally, it is expected that an efficient CAD system in retinal OCT be sensitive to the presence of even one abnormal B-scan in OCT volumes. Unlike these methods, which used a threshold of 33% and 15%, respectively, the proposed framework in this paper dealt with this issue automatically by learning a diagnostic role with the RF classifier on extracted OCT features. Compared to Ref. 28, using the diagnostic threshold on abnormal B-scans in the Heidelberg dataset resulted in an average Pr of 98.33%, where our strategy outperformed the method with 0.34% Pr rate without performing the alignment preprocessing for retinal B-scans.

For the evaluation of the robustness and generalization of the proposed WCNN-RF, its diagnostic ability was also evaluated in a more challenging situation with combining the two datasets. For the dataset, there would be more challenges for the analysis and classification, because (i) the number of samples in each class was no longer equal (class imbalance in the dataset), (ii) there was a greater variety of miss-aligned B-scans that included more variations for retinal curvatures, and (iii) there were different levels of noise disruptions in the two basic databases. However, the

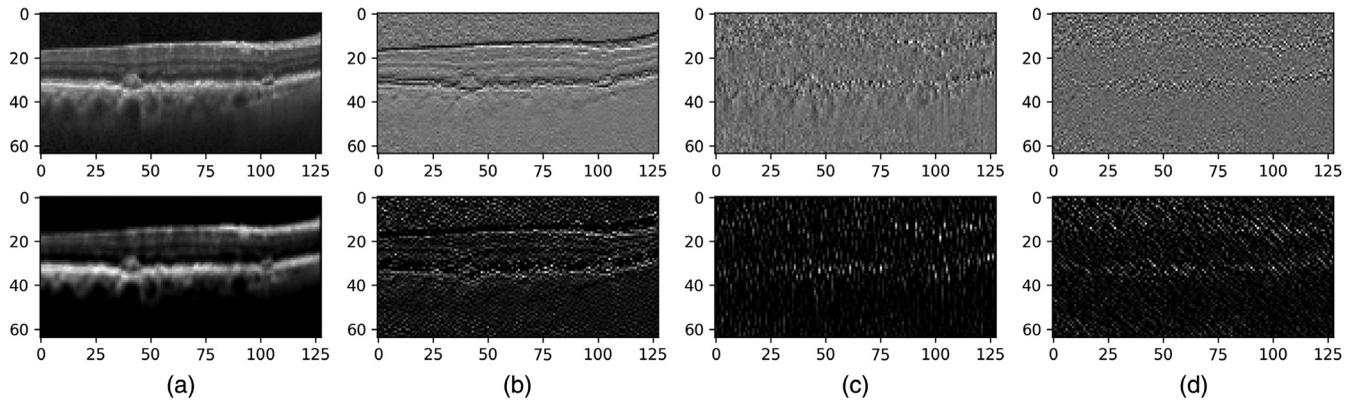


Fig. 8 An example of 2-D FMs in an SFD-Layer for an OCT B-scan image sample: (top row) 2-D Daubechies wavelet subbands and (bottom row) output FMs of the SFD-layer. (a) Low pass approximation image, (b) horizontal detail image, (c) vertical detail image, and (d) diagonal detail image.

proposed algorithm could effectively manage these variations and showed an acceptable diagnostic performance.

In addition, the authors found a reduced time complexity using the WCNN-RF model compared to the equivalent model based on regular CNNs (i.e., CNN-RF). The main reason for this time efficiency is due to the direct application of tunable convolutional kernels on the ROI images in the first hidden layer of the CNN-RF model as well as the error BP process for tuning the kernels in the layer, where the WCNN-RF utilizes pre-defined wavelet kernels instead.

Overall, SFD in the WCNN1 feature learning step provided by the SFD-layer causes the WCNN1-RF framework to have a high potential for fast and discriminative feature extraction. So, the WCNN1-RF has higher performance and lower time complexity than the CNN-RF framework in the classification of retinal 3-D OCT data and presents a robust model for retinal OCT CAD systems.

5 Conclusion

This paper presented an automatic system for diagnosis of AMD and DME patients from healthy subjects in retinal OCT. The presented system consists of a two-stage method for adaptive feature learning and diagnostic scoring. Introducing and exploiting the WCNN model to generate OCT representative features in the spatial-frequency domain, the final diagnosis was made using a RF classifier. Evaluation results on two different SD-OCT datasets showed that by applying the WCNN-RF for spatial-frequency information fusion and automatic mapping from B-scan feature space to OCT level, we can design an efficient and reliable CAD system in retinal 3-D OCT without engaging costly retinal image processing steps (e.g., denoising, segmentation, and alignment processes) and different empirical voting strategies for decision-making. In the future works, we are confident that with the use of a larger database, exploiting of the extended WCNN-RF model, and dealing with the staging problem of macular diseases, the proposed system will gain the potential to support the ophthalmologists in real clinical conditions.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This work was supported in part by the Isfahan University of Medical Sciences, vice-chancellor of Research and Technology under Grant No. 395645.

References

1. J. G. Fujimoto, "Optical coherence tomography for ultrahigh resolution in vivo imaging," *Nat. Biotechnol.* **21**, 1361–1367 (2003).
2. N. M. Bressler, "Age-related macular degeneration is the leading cause of blindness," *J. Am. Med. Assoc.* **291**, 1900–1901 (2004).
3. F. E. Hirai et al., "Clinically significant macular edema and survival in type 1 and type 2 diabetes," *Am. J. Ophthalmol.* **145**, 700–706 (2008).
4. U. Schmidt-Erfurth et al., "Guidelines for the management of neovascular age-related macular degeneration by the European Society of Retina Specialists (EURETINA)," *Br. J. Ophthalmol.* **98**, 1144–1167 (2014).
5. H. Rabbani, M. Sonka, and M. D. Abramoff, "Optical coherence tomography noise reduction using anisotropic local bivariate Gaussian mixture prior in 3D complex wavelet domain," *J. Biomed. Imaging* **2013**, 417491 (2013).
6. Z. Amini and H. Rabbani, "Statistical modeling of retinal optical coherence tomography," *IEEE Trans. Med. Imaging* **35**, 1544–1554 (2016).
7. D. C. DeBuc et al., "Reliability and reproducibility of macular segmentation using a custom-built optical coherence tomography retinal image analysis software," *J. Biomed. Opt.* **14**, 064023 (2009).
8. M. D. Abramoff et al., "Automated segmentation of the cup and rim from spectral domain OCT of the optic nerve head," *Invest. Ophthalmol. Visual Sci.* **50**, 5778–5784 (2009).
9. Q. Yang et al., "Automated layer segmentation of macular OCT images using dual-scale gradient information," *Opt. Express* **18**, 21293–21307 (2010).
10. R. Kafieh et al., "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map," *Med. Image Anal.* **17**, 907–928 (2013).
11. M. S. Miri et al., "A machine-learning graph-based approach for 3D segmentation of Bruch's membrane opening from glaucomatous SD-OCT volumes," *Med. Image Anal.* **39**, 206–217 (2017).
12. L. Fang et al., "Automatic segmentation of nine layer boundaries in OCT images using convolutional neural networks and graph search," *Invest. Ophthalmol. Visual Sci.* **58**, 666 (2017).
13. L. Fang et al., "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express* **8**, 2732–2744 (2017).
14. K. Lee et al., "Multi-layer 3D simultaneous retinal OCT layer segmentation: just-enough interaction for routine clinical use," in *VipIMAGE 2017: Proc. of the VI ECCOMAS Thematic Conf. on Computational Vision and Medical Image Processing*, 18–20, October 2017, Porto, Portugal, J. M. R. S. Tavares and R. M. Natal Jorge, Eds., pp. 862–871, Springer International Publishing, Cham, Switzerland (2018).

15. D. C. Fernandez, "Delineating fluid-filled region boundaries in optical coherence tomography images of the retina," *IEEE Trans. Med. Imaging* **24**, 929–945 (2005).
16. S. J. Chiu et al., "Automatic segmentation of closed-contour features in ophthalmic images using graph theory and dynamic programming," *Biomed. Opt. Express* **3**, 1127–1140 (2012).
17. M. Esmaili et al., "Three-dimensional segmentation of retinal cysts from spectral-domain optical coherence tomography images by the use of three-dimensional curvelet based K-SVD," *J. Med. Signals Sens.* **6**, 166–171 (2016).
18. M. Esmaili, A. M. Dehnavi, and H. Rabbani, "3D curvelet-based segmentation and quantification of drusen in optical coherence tomography images," *J. Electr. Comput. Eng.* **2017**, 4362603 (2017).
19. A. Rashno et al., "Fully-automated segmentation of fluid/cyst regions in optical coherence tomography images with diabetic macular edema using neutrosophic sets and graph algorithms," *IEEE Trans. Biomed. Eng.* **PP(99)** (2017).
20. R. Kafieh et al., "Curvature correction of retinal OCTs using graph-based geometry detection," *Phys. Med. Biol.* **58**, 2925–2938 (2013).
21. Y. Sun, S. Li, and Z. Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," *J. Biomed. Opt.* **22**, 016012 (2017).
22. Y.-Y. Liu et al., "Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding," *Med. Image Anal.* **15**, 748–759 (2011).
23. P. P. Srinivasan et al., "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express* **5**, 3568–3577 (2014).
24. F. G. Venhuizen et al., "Automated age-related macular degeneration classification in OCT using unsupervised feature learning," *Proc. SPIE* **9414**, 94141I (2015).
25. Y. Wang et al., "Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images," *Biomed. Opt. Express* **7**, 4928–4940 (2016).
26. B. Hassan et al., "Structure tensor based automated detection of macular edema and central serous retinopathy using optical coherence tomography images," *J. Opt. Soc. Am. A* **33**, 455–463 (2016).
27. 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," *Biomed. Opt. Express* **8**, 579–592 (2017).
28. R. Rasti et al., "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Med. Imaging* **PP(99)** (2017).
29. K. Dabov et al., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**, 2080–2095 (2007).
30. J. Sugmk, S. Kiattisin, and A. Leelasantitham, "Automated classification between age-related macular degeneration and diabetic macular edema in OCT image using image segmentation," in *7th Biomedical Engineering Int. Conf. (BMEiCON '14)*, pp. 1–4 (2014).
31. R. Koprowski et al., "Automatic analysis of selected choroidal diseases in OCT images of the eye fundus," *Biomed. Eng. Online* **12**, 117 (2013).
32. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2013).
33. U. Avni et al., "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imaging* **30**, 733–746 (2011).
34. S. Apostolopoulos et al., "RetiNet: automatic AMD identification in OCT volumetric data," arXiv:1610.03628 (2016).
35. Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**, 2278–2324 (1998).
36. M. Matsugu et al., "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Networks* **16**, 555–559 (2003).
37. W. Zhang et al., "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," *Appl. Opt.* **29**, 4790–4797 (1990).
38. R. Rasti, M. Teshnehlab, and S. L. Phung, "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks," *Pattern Recognit.* **72**, 381–390 (2017).
39. R. Rasti, M. Teshnehlab, and R. Jafari, "A CAD system for identification and classification of breast cancer tumors in DCE-MR images based on hierarchical convolutional neural networks," *Comput. Intell. Electr. Eng.* **6**, 1–14 (2015).
40. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Int. Conf. on Machine Learning*, pp. 448–456 (2015).
41. N. Srivastava, "Improving neural networks with dropout," Master's Thesis, Vol. 182, University of Toronto, Ontario, Canada (2013).
42. T. Williams and R. Li, "Advanced image classification using wavelets and convolutional neural networks," in *15th IEEE Int. Conf. on Machine Learning and Applications (ICMLA '16)*, pp. 233–239 (2016).
43. C. K. Chui, *An Introduction to Wavelets*, Elsevier, San Diego, California (2016).
44. J. Tang et al., "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Trans. Geosci. Remote Sens.* **53**, 1174–1185 (2015).
45. E. K. Chong and S. H. Zak, *An Introduction to Optimization*, Vol. **76**, John Wiley & Sons, Hoboken, New Jersey (2013).
46. D. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
47. P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer Science & Business Media, New York (2013).
48. L. Breiman, "Random forests," *Mach. Learn.* **45**, 5–32 (2001).
49. T. T. D. Team et al., "Theano: a python framework for fast computation of mathematical expressions," arXiv:1605.02688 (2016).
50. F. Chollet, "Keras," in *GitHub Repository*, GitHub, <https://github.com/keras-team/keras> (2015).

Reza Rasti is a PhD researcher at Isfahan University of Medical Sciences. He received his BSc degree in electronics engineering and his MSc degree in biomedical engineering from Shahid Rajaei University and K. N. Toosi University of Technology, Tehran, Iran, in 2009 and 2012, respectively. His current research interests include machine learning, pattern recognition, medical image and signal analysis, and computer-aided diagnosis systems.

Alireza Mehridehnavi received his BSc degree in electronic engineering from Isfahan University of Technology in 1988. He received his MSc degree in measurement and instrumentation from Indian Institute of Technology Roorkee, India, in 1992 and his PhD in medical engineering from Liverpool University in 1996. He is a full professor in the Biomedical Engineering Department, Isfahan University of Medical Sciences, Isfahan, Iran. His research interests are medical optics, devices and signal, and image processing.

Hossein Rabbani received his BSc degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2000, and his MSc and PhD degrees in bioelectrical engineering from Amirkabir University of Technology, Tehran, Iran, in 2002 and 2008, respectively. He is now a full professor in the Biomedical Engineering Department, Isfahan University of Medical Sciences, Isfahan. His research interests are medical image analysis and modeling, signal processing, sparse transforms, and image restoration.

Fedra Hajizadeh received her MD degree from Tehran University of Medical Sciences, Tehran, Iran, in 1995 and completed the Ophthalmology Residency and Vitreo-Retinal Fellowship both at Farabi Eye Hospital, Tehran University of Medical Sciences in 1999 and 2004, respectively. Since 2008, she has been a consulting surgeon of vitreo-retinal diseases and research scientist at Noor Eye Hospital, Tehran, Iran. Her current research includes retinal optical coherence tomography (OCT), ocular trauma, and retinal fluorescein angiography.