

Automatic classification of retinal three-dimensional optical coherence tomography images using principal component analysis network with composite kernels

Leyuan Fang
Chong Wang
Shutao Li
Jun Yan
Xiangdong Chen
Hossein Rabbani

Automatic classification of retinal three-dimensional optical coherence tomography images using principal component analysis network with composite kernels

Leyuan Fang,^a Chong Wang,^a Shutao Li,^{a,*} Jun Yan,^a Xiangdong Chen,^b and Hossein Rabbani^c

^aHunan University, College of Electrical and Information Engineering, Changsha, China

^bThe First Affiliated Hospital of Hunan University of Chinese Medicine, Department of Ophthalmology, Changsha, China

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

Abstract. We present an automatic method, termed as the principal component analysis network with composite kernel (PCANet-CK), for the classification of three-dimensional (3-D) retinal optical coherence tomography (OCT) images. Specifically, the proposed PCANet-CK method first utilizes the PCANet to automatically learn features from each B-scan of the 3-D retinal OCT images. Then, multiple kernels are separately applied to a set of very important features of the B-scans and these kernels are fused together, which can jointly exploit the correlations among features of the 3-D OCT images. Finally, the fused (composite) kernel is incorporated into an extreme learning machine for the OCT image classification. We tested our proposed algorithm on two real 3-D spectral domain OCT (SD-OCT) datasets (of normal subjects and subjects with the macular edema and age-related macular degeneration), which demonstrated its effectiveness. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JBO.22.11.116011](https://doi.org/10.1117/1.JBO.22.11.116011)]

Keywords: optical coherence tomography; principal component analysis network; composite kernel; retinal disease; image classification.

Paper 170380RR received Jun. 10, 2017; accepted for publication Nov. 8, 2017; published online Nov. 29, 2017.

1 Introduction

Macula is an oval-shaped pigmented area near the center of retina, which is mainly responsible for the central vision. Damages to the macula, such as macular edema (ME) and age-related macular degeneration (AMD), will directly result in the loss of central vision.^{1–3} Clinical diagnosis of ME and AMD relies on the localization of macular structure abnormalities (also called lesions), whose types and numbers are important diagnostic criteria for ophthalmologists. For example, compared with the normal macula (NM) [see an example in Fig. 1(a)], edema and exudates are often related to diabetic retinopathy (one type of ME)⁴ [see an example in Fig. 1(b)], while drusen is a typical lesion often found in AMD eyes⁵ [see an example in Fig. 1(c)]. Therefore, it is essential to investigate the macular lesions for the clinical diagnosis and treatment of ophthalmic disease.

Optical coherence tomography (OCT) can provide *in vivo* three-dimensional (3-D) cross-sectional imaging of human tissue at micrometer resolutions;^{6–8} it has been widely used for a variety of medical imaging applications.^{9–13} High resolution of OCT enables the visualization of multiple retinal cell layers and the capability of volumetric quantitative evaluation of the retinal structures.^{14–16} By employing near-infrared light to image the eye with micron resolution, subtle while valuable pathological structures can be observed, from which many macular and ocular diseases can be identified in their early stages.^{17–20} In the clinical diagnosis, ophthalmologists need to manually identify various macular lesions at each cross section of the OCT volume and then determine the types of disease. Such

manual analysis is time-consuming and demanding for expert graders and often yields subjective results. Consequently, it is urgent to develop an effective computer-assisted OCT image analysis technique.

During the past decades, a multitude of classification methods has been developed for the automatic analysis of OCT images.^{21–28} In general, these OCT image classification approaches mainly consist of the following two key components: feature extraction^{21–27} and classifier design.^{22–26,28} The feature extraction first extracts a set of representative features to describe the original OCT images, and then the classifier determines the type of disease by mapping the extracted features to a category. For the feature extraction of the OCT images, Sugmk et al.²¹ first segmented the retinal pigment epithelium (RPE) layer and then computed the binary features from the layer for the identification of AMD and diabetic macular edema (DME). Liu et al.²² computed the multiscale local binary pattern (LBP) features to encode the texture and shape information of the OCT images. Srinivasan et al.²³ utilized the multiscale histogram of oriented gradient (HOG) features, which are useful for the detection of AMD, DME, and NM, to describe each OCT B-scan. Hassan et al.²⁴ used five features (three thickness profiles and two cyst fluids) based on structure tensors to detect ME and central serous retinopathy. For each specific kind of lesion in the OCT image, the above aforementioned works^{21–24} design the corresponding features and can provide promising classification results. However, the above features (e.g., LBP, HOG, and structure tensor) are designed based on fixed mathematical models, which means that the features designed for one type of lesion may be suboptimal for

*Address all correspondence to: Shutao Li, E-mail: shutao_li@hnu.edu.cn

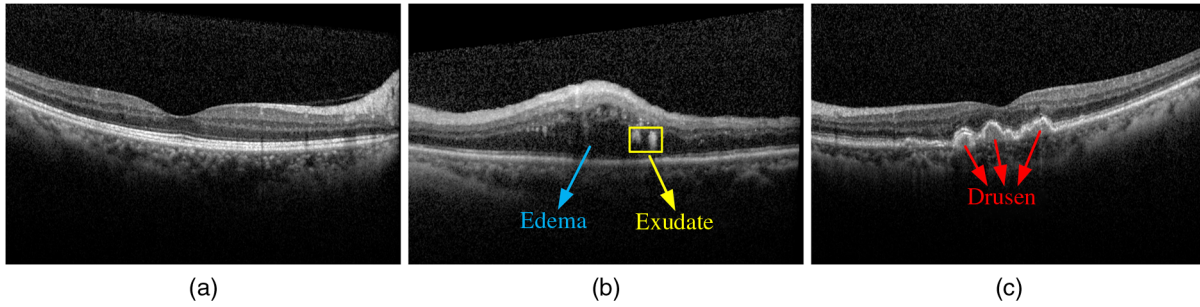


Fig. 1 Examples of (a) OCT image with NM: clear subretinal boundaries; (b) OCT image with ME: swollen retina, edema, and exudates around the macula; and (c) OCT image with AMD: RPE layer interrupted by drusen of various sizes.

representing other kinds of lesions. Since the clinically acquired OCT images usually contain very complex pathological structures, a more desirable strategy is to learn features from the original OCT images. Sun et al.²⁵ used sparse coding and a multiscale dictionary learning method based on scale-invariant feature transform descriptor to extract representative features from the input images for the detection of AMD, DME, and NM. Venhuizen et al.²⁶ introduced an unsupervised feature learning²⁹ approach to distinguish AMD from normal volumes, in which small descriptive image patches are selectively extracted to create patch occurrence histogram features. Very recently, in Ref. 27, a typical deep learning model called the convolutional neural network (CNN)^{30–32} was utilized to extract the very high-level features from each OCTB-scan, and it delivered very satisfactory results. In general, the CNN has many network layers and usually contains at least millions of layer parameters, which thus requires a very huge amount of training datasets and computational cost to train networks.

On the other hand, numerous researchers have attempted to design various classifiers [e.g., support vector machines (SVMs),^{22–25} random forest,²⁶ and sequential minimal optimization (SMO)²⁸] for the OCT image classification task. In Refs. 22–25, the extracted features were considered feature vectors in the SVM classifier, which identifies the presence of NM and each of the pathologies. In Ref. 26, a supervised random forest classifier was trained using the aforementioned feature vectors for category discrimination. Wang et al.²⁸ systematically evaluated these classifiers and demonstrated that the SMO tended to achieve the best performance. However, all the classifiers in the works are only designed for the classification of single B-scan, without considering the correlations among B-scans of the 3-D OCT images.

To address the above issues, we propose a feature learning-based classification algorithm called the principal component analysis network with composite kernel (PCANet-CK) for the automatic diagnosis of AMD, ME, and NM in OCT images. First, a PCANet model³³ is used to automatically extract multiple level features from each B-scan of the 3-D OCT images. Compared with the CNN, the PCANet has a simple network structure and much less parameters, which can be more efficiently trained. Second, a set of the most important features of each volume are carefully selected to construct multiple kernels, which are then combined together to create a composite kernel. Finally, the composite kernel³⁴ is fed into an extreme learning machine (ELM)^{35,36} classifier for the classification of 3-D OCT images. Such a composite kernel-based classification strategy can jointly exploit the correlations among features of

the 3-D OCT images while still reducing the computational cost for classifying the 3-D OCT images.

The rest of this paper is organized as follows. Section 2 briefly reviews the PCANet feature extraction method and ELM classifier. Section 3 introduces the proposed PCANet-CK model. Experimental results on two clinically acquired datasets and related discussions are detailed in Sec. 4. Section 5 concludes this paper and suggests some future works.

2 Review

2.1 PCANet Feature Extraction

As a modified version of the typical deep CNN, the PCANet method³³ adopts a series of PCA convolution filters to extract the features from an input image. A commonly used PCANet model consists of two PCA filter convolution stages and one output stage. Figure 2 shows the typical architecture of the PCANet method, as also described in the following.

In the first convolution stage, given H training images $X_h \in \mathbf{R}^{M \times N}$, $h \in \{1, 2, \dots, H\}$, we convolve each image with L_1 PCA filters in a patch-based pattern to obtain $L_1 \times H$ feature maps $X_h^{1,l}$, $l \in \{1, 2, \dots, L_1\}$. Specifically, we first extract the patches x_h^i (of size $n_1 \times n_2$) for each image X_h and remove the mean for each patch. Then, the extracted patches of each image X_h are vectorized and can be combined into a matrix $P_h = [\bar{P}_h^1, \dots, \bar{P}_h^i, \dots, \bar{P}_h^{M^*N^*}]$, where $\bar{P}_h^i \in \mathbf{R}^{n_1 n_2 \times 1}$ is a vector for the related patch, $M^* = M - \lceil n_1/2 \rceil$, $N^* = N - \lceil n_2/2 \rceil$, and $\lceil n \rceil$ gives the smallest integer greater than or equal to n . The matrices of all the training images are also constructed as a matrix:

$$P = [P_1, P_2, \dots, P_H]. \quad (1)$$

After that, we compute the eigenvectors of PP^T and select the first L_1 principal eigenvectors as the L_1 PCA filters W_l^1 , $l = 1, 2, \dots, L_1$. Finally, L_1 filters are separately applied to each training image X_h :

$$X_h^{1,l} = X_h * W_l^1, \quad l = 1, 2, \dots, L_1, \quad (2)$$

where the operator $*$ denotes the two-dimensional convolution. So, for a given input image X_h , L_1 feature maps $X_h^{1,l}$ can be obtained.

For each feature map $X_h^{1,l}$ from the first stage, the second stage convolves this map with L_2 PCA filters to obtain L_2 feature maps $X_h^{2,l}$, $l \in \{1, 2, \dots, L_2\}$. Similar to the first stage,

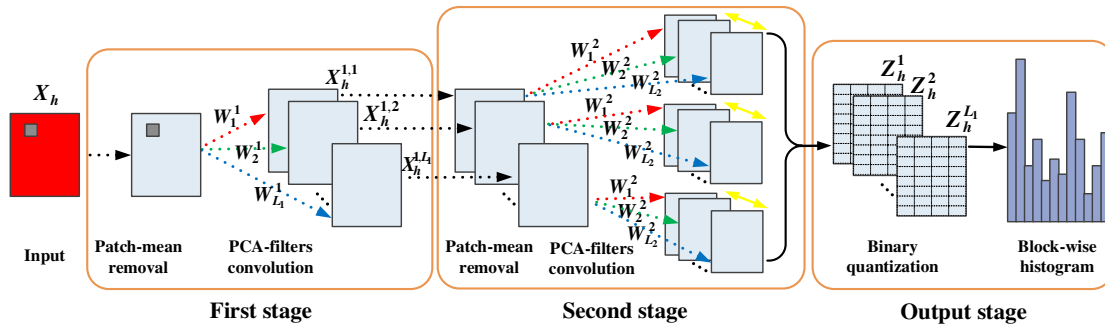


Fig. 2 Illustration of the PCANet architecture with two convolution stages and one output stage.

we also first extract the mean-removed patches from all $L_1 \times H$ maps $X_h^{1,l}$ and use the vector version of patches to construct the matrix U . Then, we compute the eigenvectors of UU^T , and select the first L_2 principal eigenvectors as the L_2 PCA filters W_l^2 , $l = 1, 2, \dots, L_2$. As in Eq. (2), by applying the L_2 PCA filters on each feature map $X_h^{1,l}$ from the first stage, we can obtain $L_1 \times L_2 \times H$ feature maps. Given an input image X_h , the first stage has L_1 feature maps, and the second stage can create $L_1 \times L_2$ feature maps.

In the output stage, we first binarize the $L_1 \times L_2$ feature maps of the second stage using a Heaviside step function and apply the hashing encoding to convert them into L_1 integer-valued images Z_h^l , $l \in \{1, 2, \dots, L_1\}$.³³ Then, each integer-valued image Z_h^l is partitioned into B blocks (of size $n_{B1} \times n_{B2}$), in which the local histogram (overlapping or nonoverlapping) is calculated. After that, all B histograms are concatenated into a vector $\text{Bhist}(Z_h^l)$. Finally, the feature vector of the one input image X_h can be extracted as

$$f_h = [\text{Bhist}(Z_h^1), \text{Bhist}(Z_h^2), \dots, \text{Bhist}(Z_h^{L_1})]^T. \quad (3)$$

The extracted feature vector can be fed into a classifier (e.g., SVM or ELM) for the classification.

2.2 ELM Classifier

ELM is a very efficient supervised learning model, and its objective is to find a decision rule for the classification.^{35,36} To be specific, let $\{x_i, y_i\}_{i=1}^H$ be the training set. $\{x_1, x_2, \dots, x_H\} \in \mathbf{R}^{m \times 1}$ denote the H input training samples and $y_i = [0, \dots, 1, \dots, 0]^T \in \mathbf{R}^{G \times 1}$ represent the corresponding class labels, which are a vector with one for the true label and zero entries for the others. G denotes the number of training classes. In general, the ELM aims to simultaneously minimize the training error and the norm of output weights of the objective function

$$\min_{\beta, \xi} \left\{ \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^H \|\xi_i\|^2 \right\} \\ \text{s.t. } \beta^T \phi(x_i) = y_i - \xi_i, \quad \forall i = 1, 2, \dots, H, \quad (4)$$

where $\phi(\bullet)$ is a feature mapping function determining the hidden layer output. $\beta \in \mathbf{R}^{|\phi(\bullet)| \times G}$ are the output weights, $\xi \in \mathbf{R}^{G \times H}$ is the training error matrix, and $|\phi(\bullet)|$ represents the element number of the vector $\phi(\bullet)$. C is a regularization parameter

balancing the norm of output weights and training errors. Based on the Lagrange multiplier method and the Karush–Kuhn–Tucker optimality conditions,³⁵ the solution of Eq. (4) can be analytically obtained as

$$\hat{\beta} = \Phi^T \left(\frac{I}{C} + \Phi \Phi^T \right)^{-1} \mathbf{Y}^T, \quad (5)$$

where $\Phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_H)]^T \in \mathbf{R}^{H \times |\phi(\bullet)|}$, $\mathbf{Y} = [y_1, y_2, \dots, y_H] \in \mathbf{R}^{G \times H}$, and I is an identity matrix.

As described in Refs. 34 and 37, kernel can further transform the features into a higher dimensional feature space and thus improve the discriminative capacity for the features. Since the mapping $\phi(\bullet)$ in the ELM learning is represented by the inner function, a kernel function K can be defined by

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle. \quad (6)$$

Then, we can establish a liner ELM by the kernel function, without considering the mapping $\phi(\bullet)$ explicitly. The most widely used kernel is the linear kernel, which can be calculated as follows:

$$K(x_i, x_j) = x_i^T x_j. \quad (7)$$

Finally, by incorporating Eq. (6) into Eq. (5), the decision rule of the ELM for any test sample x is determined by

$$f(x) = \hat{\beta}^T \phi(x). \quad (8)$$

3 Proposed PCANet-CK Method

In this paper, we propose a PCANet-CK method, which combines PCANet with composite kernel models for the classification of 3-D OCT images. The PCANet-CK method mainly consists of two parts: (1) apply the PCANet on each B-scan to automatically extract the features and (2) use the composite kernel to exploit the correlations among features of B-scans in each volume for the classification. The outline of the proposed PCANet-CK algorithm is shown in Fig. 3.

3.1 Feature Extraction with PCANet

Since OCT images have large variations in intensity ranges, we first conduct the intensity normalization on each B-scan of the training and testing OCT volumes, which linearly rescales the intensity values to $[0, 1]$. Then, as in Ref. 23, for training and testing OCT images, we adopt the BM3D algorithm³⁸ to remove the noise of each B-scan and use the flattening to reduce

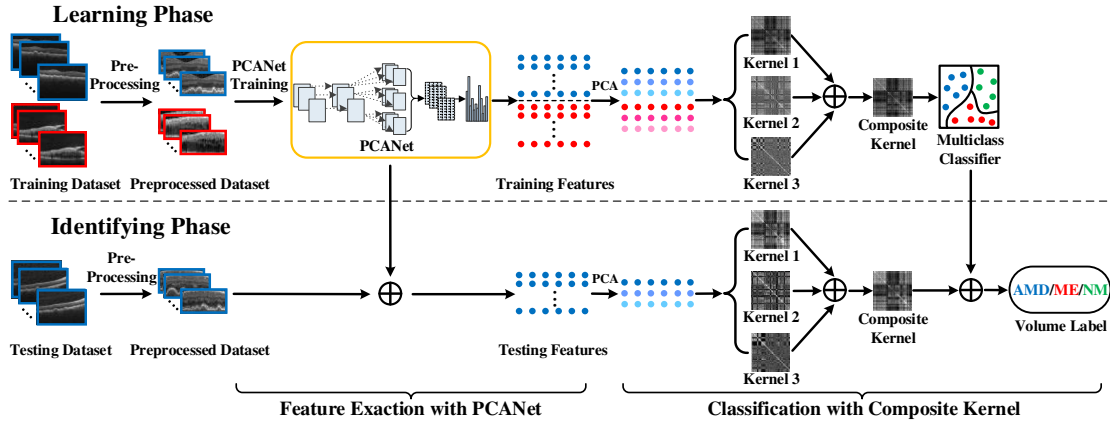


Fig. 3 Outline of the proposed PCANet-CK algorithm.

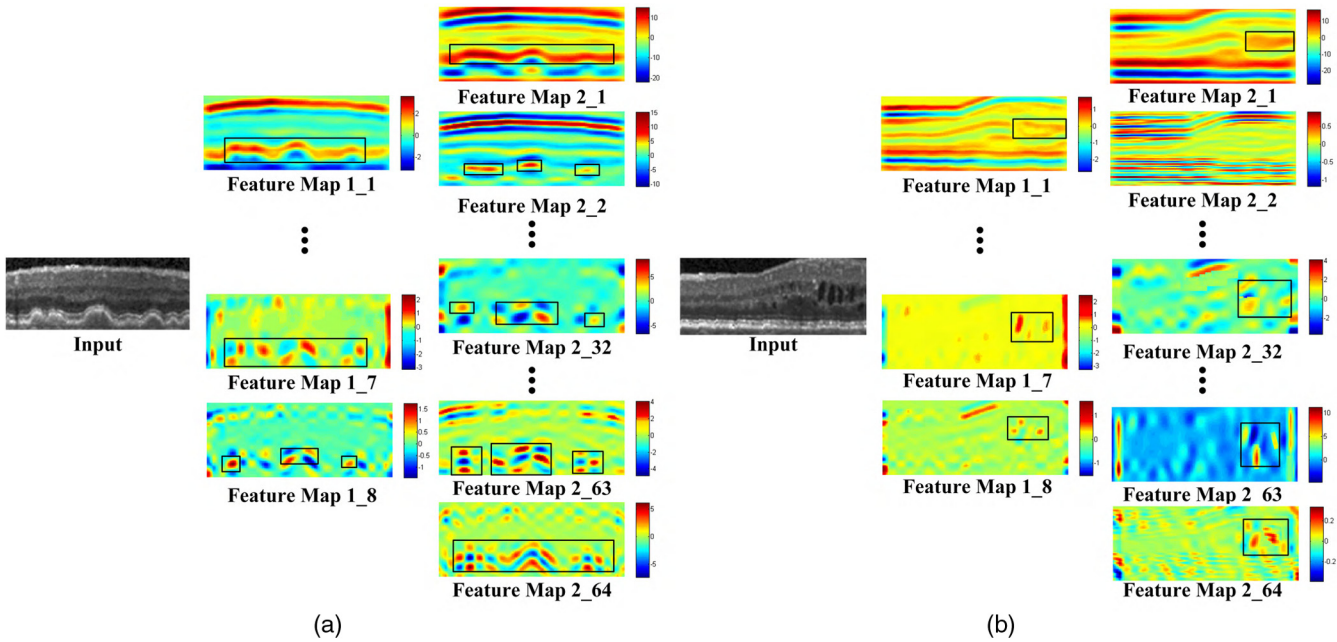


Fig. 4 Two examples of feature maps extracted by PCANet from OCT images on subjects (a) AMD and (b) ME.

the variations for imaged retinas in their angles of inclination, positions, and natural curvatures among B-scans. After the above preprocessing steps, the PCANet feature extraction consists of training and testing phases. Specifically, in the training phase, we first input T 3-D OCT volumes $V_t^{\text{train}} \in \mathbf{R}^{M \times N \times H_t}$, $t = 1, 2, \dots, T$ into the PCANet training model. Each volume contains H_t B-scans $X_h^{\text{train}} \in \mathbf{R}^{M \times N}$, $h = \{1, 2, \dots, H_t\}$. As described in Sec. 2.1, the PCANet will take all B-scans of all training OCT volumes together as input. Then, we decompose the B-scans into patches and use the patches to train L_1 PCA filters W_l^1 , $l = 1, 2, \dots, L_1$ and L_2 PCA filters W_l^2 , $l = 1, 2, \dots, L_2$, in the first and second stages, respectively. Finally, after the binary quantization and calculating the block-wise histogram in the output stage, for each B-scan X_h^{train} , one feature vector f_h^{train} is obtained. For each volume V_t^{train} , its extracted features can be constructed as a matrix $F_t^{\text{train}} = [f_1^{\text{train}}, f_2^{\text{train}}, \dots, f_{H_t}^{\text{train}}]$.

In the testing phase, given an input 3-D OCT volume $V_t^{\text{test}} \in \mathbf{R}^{M \times N \times H_t}$, we apply the PCA filters W_l^1 , $l = 1, 2, \dots, L_1$ and W_l^2 , $l = 1, 2, \dots, L_2$ from the training phase on each B-scan of this OCT volume. Then, after the binary quantization and computation of the block-wise histogram, we obtain one feature matrix $F_t^{\text{test}} = [f_1^{\text{test}}, f_2^{\text{test}}, \dots, f_{H_t}^{\text{test}}]$.

Figure 4 shows two examples of the extracted feature maps from different layers on the AMD and ME OCT images, respectively. As can be observed, for the input AMD image, the drusen regions will have high responses in the different feature maps of different layers [see the zoomed rectangle regions in Fig. 4(a)], while the edema is also very prominent in feature maps extracted from the ME image [see the zoomed rectangle regions in Fig. 4(b)]. Therefore, the learned PCA filters tend to capture meaningful pathology structure information, and the extracted feature maps can be used to achieve an effective classification, even without quantifying the size of the lesions.

3.2 Classification with Composite Kernel

As described above, given an OCT volume, the PCANet can extract one feature vector for each single B-scan in both the training and testing phases. Since the nearby B-scans of the 3-D OCT images are very similar, high correlations should also exist in their extracted features, which can be utilized to enhance the classification.³⁹ One possible way for exploiting the correlations among features within one 3-D OCT volume is to fuse them together. As described in Ref. 40, using the kernel to map the features into the higher dimensional feature space and then combining them as a composite kernel can better utilize the correlations and differences among different features. In addition, the parameters of different features are integrated into the composite kernel function and jointly optimized during the training process,³⁷ which can also effectively combine the information of different features for classification. Specifically, for each feature f_h of a 3-D OCT volume, we use the linear kernel function in Eq. (7) to compute the corresponding kernel K_h . As introduced in Refs. 41 and 42, directly stacking these kernels or a linear combination of them can exploit the correlations among them. Since stacking these kernels will greatly increase the dimension and computational cost, we create the composite kernel by a linearly weighted combination of these kernels

$$K_{\text{comp}} = \sum_{h=1}^H \mu_h K_h, \quad (9)$$

where μ_h is the weight for the kernel K_h . Note that the kernel weights are the same for all the training and testing OCT volumes. In the training phase, for T training volumes, we can create T composite kernels $K_{\text{comp}}^{\text{train},t}$, $t = 1, 2, \dots, T$. Then, we replace the kernel in Eq. (6) with these composite kernels to train an ELM decision rule. In the testing phase, for each input OCT volume, we create one composite kernel $K_{\text{comp}}^{\text{test}}$ and apply the decision rule on this kernel to assign the class label to the volume.

Note that, since PCANet extracts many feature vectors (e.g., more than 30 in our test) from each volume, a number of corresponding weights μ_h , $h = 1, 2, \dots, H$ in Eq. (9) are required to be set, and searching the optimal values for so many weights would be very hard. In addition, different OCT volumes may consist of different numbers H_t of B-scans; thus, the number of kernel weights for different volumes is also varied. To address this issue, before the construction of multiple kernels, we employ the PCA transform⁴³ to reduce the number of features and create the same number of kernel weights for each volume. Specifically, the PCA transform maps the feature matrix $F_t = [f_{1,t}, f_{2,t}, \dots, f_{H_t,t}]$ to a new set of principal components (PCs): $F_t^{\text{PCA}} = [f_{\text{PC1}}, f_{\text{PC2}}, f_{\text{PC3}}]$ using the SVD technique on the covariance matrix of F_t , where several eigenvectors are calculated as the optimal projection axes. In this paper, we only retain the first three PCs, which account for the most information in the feature matrix F_t . After such a feature selection strategy, only three kernels (see Fig. 3) are generated, which are then fused as a composite kernel.

4 Experimental Results

4.1 Datasets

To evaluate the effectiveness of the proposed PCANet-CK method, we tested it on two real OCT datasets acquired from Duke University and First Affiliated Hospital of Hunan

University of Chinese Medicine (HUCM), respectively. The Duke dataset was acquired using the spectral domain (SD)-OCT imaging system (Heidelberg Engineering Inc., Heidelberg, Germany) by Duke University, Harvard University, and the University of Michigan, which was publicly available in Ref. 44. The Duke dataset consists of 45 SD-OCT volumes from 45 patients (15 AMD, 15 DME, and 15 normal), and each volume contains many B-scans (ranging from 31 to 97). The original axial and lateral resolutions of the B-scans are $3.87 \mu\text{m}/\text{pixel}$ and 6 to $12 \mu\text{m}/\text{pixel}$, respectively. More details of the scanning protocols for this dataset can be found in Ref. 23. Three examples from NM, ME, and AMD subjects of the Duke dataset are shown in Fig. 5(a).

The HUCM dataset was captured using the spectral SD-OCT imaging system (Heidelberg Engineering, Heidelberg, Germany) at the First Affiliated Hospital HUCM. The HUCM dataset is composed of 54 SD-OCT volumes (18 AMD, 18 DME, and 18 normal) from 48 patients. All volumes are of size $768 \times 496 \times 31$ voxels, covering $8.8 \times 8.8 \times 2.0 \text{ mm}^3$. Figure 5(b) shows three B-scans from NM, ME, and AMD subjects of the HUCM dataset, respectively. Note that only small pathological structures (e.g., drusen) exist in some B-scans of HUCM dataset, which are very challenging for recognition. The above two datasets used in our study were approved by the local Investigational Review Board and were performed in accordance with the tenets set forth in the Declaration of Helsinki. Written informed consent was obtained before enrolling patients in EUGENDA.

4.2 Experimental Setting

In the proposed PCANet method, the main parameters include the PCA filter size, the block size for local histograms, and the filter numbers L_1 and L_2 . In our experiment, we set the PCA filter size $n_1 = n_2 = 11$, the block size $n_{B1} = n_{B2} = 11$, and the block overlap ratio as 0.1. The effect of the PCA filter size and block size was analyzed in Sec. 4.4. As described in Ref. 33, L_1 and L_2 were set to be 8. In addition, for the composite kernel creation, the kernel weights μ_1 , μ_2 , and μ_3 for the three kernels were set as 0.85, 0.10, and 0.05, respectively. The regularization parameter C in Eq. (5) for the ELM classifier training was set to be 5.

In our experiments, we utilized the cross validation to evaluate the classification performance of the proposed PCANet-CK method. The cross validations were repeated with different random seeds to avoid the dataset splitting bias. The accuracy, specificity, and sensitivity were adopted to evaluate the classification performances, which were defined in a binary classification problem:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (10)$$

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (11)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (12)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

In our three-class classification problem, the sensitivity of a class label is also the prediction accuracy, and the specificity

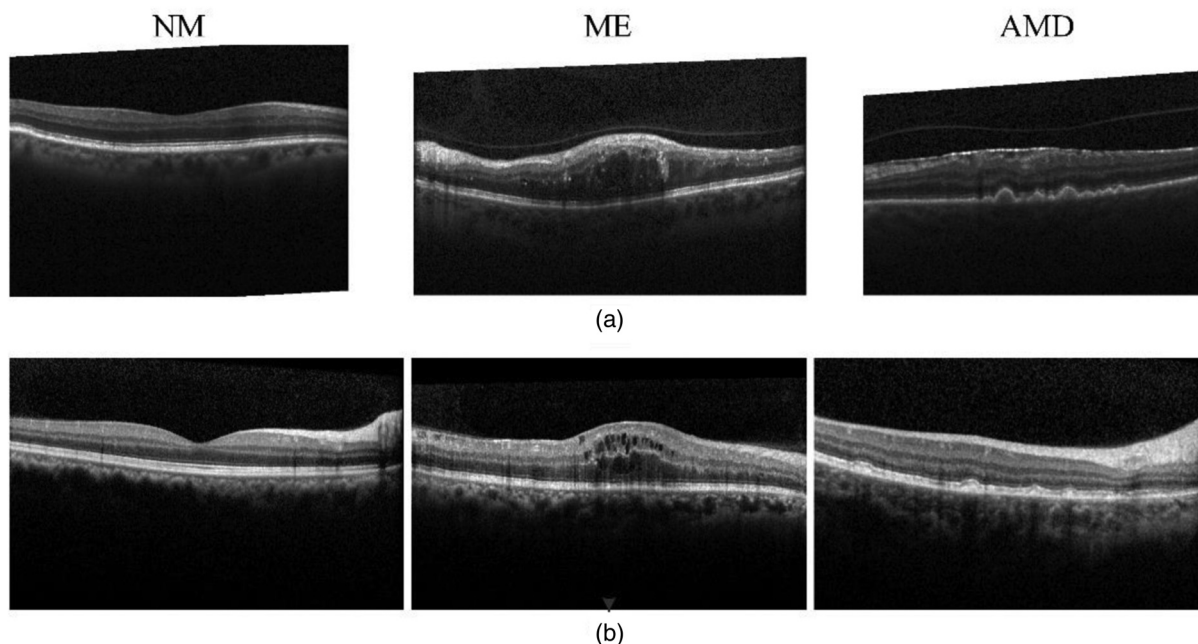


Fig. 5 Examples of SD-OCT images from (a) Duke dataset and (b) HUCM dataset.

is defined in the same way for each class label, where the negative samples are the samples not in the considered class. Therefore, the overall sensitivity (Ov-Se), overall specificity (Ov-Sp), and overall accuracy (Ov-Acc) are averaged over the three-class labels. If the number of samples in each class is equal, the Ov-Se is also defined as the ratio: number of correctly predicted samples/total number of samples. In addition, the mean and standard deviation values of the above metrics are calculated.

4.3 Results Comparisons

The performance of the proposed PCANet-CK method was first compared with the well-known OCT classification method: HOG-SVM.²³ The 15-fold (Duke dataset) and 18-fold (HUCM dataset) cross validations was repeated 10 times with different random seeds. The HOG-SVM method utilizes the HOG descriptor to extract the feature vectors and trains three binary SVMs⁴⁵ for the classification. The leave-three-out cross-validation strategy was applied on the two test datasets; this was achieved by choosing three volumes once from each class as the test dataset and the remaining 42 (in Duke dataset) or 51

(in HUCM dataset) volumes as the training dataset. For different experiments, three different volumes were chosen as test volumes, and 15 or 18 experiments were conducted to cover all the volumes in the two datasets.

The quantitative results on the Duke and HUCM datasets are tabulated in Tables 1 and 2, respectively. As can be observed, in the two test datasets, the proposed PCANet-CK method consistently performs better than the HOG-SVM method in terms of all quantitative metrics. Specifically, in the Duke dataset, the PCANet-CK can accurately classify all the test volumes of the three classes, whereas the Ov-Acc for the HOG-SVM method is about 96.6%. In the HUCM dataset, the Ov-Acc for the PCANet-CK method is 96.9%, whereas the Ov-Acc for the HOG-SVM method is about 90.7%. Moreover, in the classes of AMD and NM on HUCM dataset, the gain of the mean sensitivity of the proposed method over the HOG-SVM method is more than 9%, which demonstrates the effectiveness of the PCANet feature extraction and the composite kernel for exploiting the 3-D information.

Very recently, the deep CNN method was also tested on the Duke dataset.²⁷ The deep CNN method first uses a pretrained CNN model (GoogleNet) and then fine-tunes it on the Duke

Table 1 Classification results (%) on Duke dataset.

Methods	Classes	Sensitivity	Specificity	Accuracy	Ov-Se	Ov-Sp	Ov-Acc
HOG-SVM ²³	AMD	100.0 ± 0.0	96.7 ± 0.0	97.8 ± 0.0	94.9 ± 1.5	97.4 ± 0.8	96.6 ± 1.0
	ME	97.3 ± 3.4	97.0 ± 1.1	97.1 ± 1.5			
	NM	87.3 ± 2.1	98.7 ± 1.7	94.9 ± 1.5			
PCANet-CK	AMD	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	ME	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0			
	NM	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0			

Table 2 Classification results (%) on HUCM dataset.

Methods	Classes	Sensitivity	Specificity	Accuracy	Ov-Se	Ov-Sp	Ov-Acc
HOG-SVM ²³	AMD	82.8 ± 1.8	93.3 ± 1.9	89.8 ± 1.6	86.1 ± 1.6	93.1 ± 0.8	90.7 ± 1.1
	ME	86.1 ± 2.9	99.7 ± 0.9	95.1 ± 1.3			
	NM	89.4 ± 1.8	86.1 ± 0.0	87.2 ± 0.6			
PCANet-CK	AMD	92.2 ± 4.7	96.9 ± 0.9	95.4 ± 1.6	95.4 ± 1.6	97.7 ± 0.8	96.9 ± 1.1
	ME	94.4 ± 0.0	98.3 ± 1.4	97.0 ± 1.0			
	NM	99.4 ± 1.8	97.8 ± 1.8	98.3 ± 1.1			

dataset to extract the features for the identification of AMD, DME, and NM. In the experimental setting of Ref. 27, the cross validation was also utilized by dividing the whole dataset into 15 folds, with each fold containing three volumes (one from each class). However, different from the above validation, each experiment involved eight folds (24 volumes) for training and seven folds (21 volumes) for testing. Folds were sequentially rather than randomly chosen. Here, the proposed PCANet-CK and HOG-SVM methods were tested under the same experimental setting as in Ref. 27. Since only the mean sensitivity was reported in Ref. 27, the value of the deep CNN method was compared with those of the PCANet-CK and HOG-SVM methods, as reported in Table 3. As can be seen, the proposed PCANet-CK method generally performs much better than the HOG-SVM and deep CNN methods for classifying the images of AMD and ME subjects. In addition, for the classification of images from NM subject, the PCANet-CK delivers better performance

Table 3 Classification performances (sensitivities in %) of different methods on Duke dataset.

	HOG-SVM ²³	Deep CNN ²⁷	PCANet-CK
AMD	89	89	94
ME	83	86	94
NM	90	99	98

Note: The best results among different methods are labeled in bold.

than the HOG-SVM, while is very close to the deep CNN. These results show the superiority of the feature learning strategy used in PCANet-CK and deep CNN methods over the hand-crafted HOG feature extraction strategy adopted in the HOG-SVM method.

In our experiments, the proposed PCANet-CK method was implemented on a desktop computer with an Intel (R) Core i7-6700K CPU and 64 GB of RAM under the environment of MATLAB R2016b. The average time for testing one volume requires about 7.2 and 3.1 s on the Duke and HUCM datasets, respectively. Note that, since the training phase of the PCANet-CK method can be an offline process, it does not need to be considered in the testing phase. In addition, our code is not optimized for speed. The processing time is expected to be reduced significantly by more efficient coding coupled with a general purpose graphics processing unit.

4.4 Effect of the PCA Filter Size and Block Size

In this section, the effect of the PCA filter size and the block size on the proposed PCANet model was analyzed on the Duke and HUCM datasets. The leave-three-out cross validation and the Ov-Se were also used here. For the evaluation of the PCA filter size, we varied the PCA filter size ($n_1 = n_2$) in the first two convolution stages from 3 to 19 and kept other parameters (e.g., block size, overlap ratio, and kernel weights) the same as that in Sec. 4.2. The results with different PCA filter sizes are shown in Fig. 6(a). As can be observed, the performance of the proposed PCANet-CK method will generally improve as the PCA filter size increases from 3 to 11. When the PCA

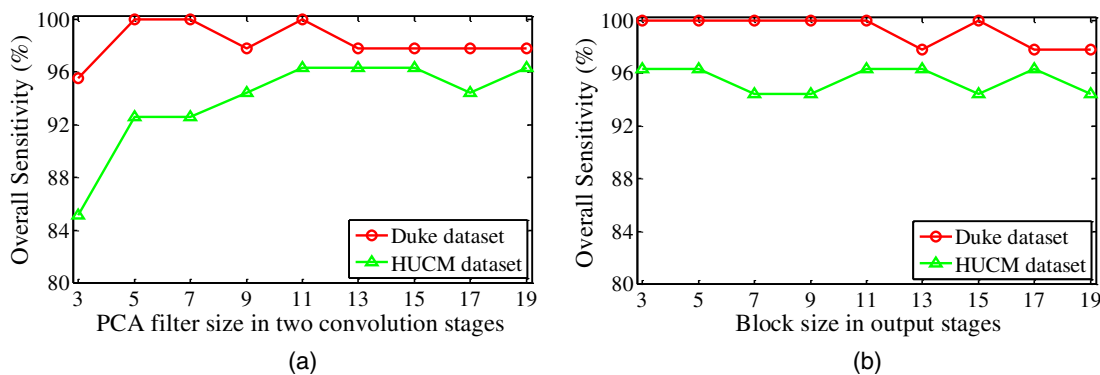
**Fig. 6** Overall sensitivities of the PCANet-CK method on Duke and HUCM datasets for different (a) PCA filter sizes in the first two stages and (b) block sizes in the output stage.

Table 4 Classification results (%) on Duke dataset using single kernel.

Methods	Classes	Sensitivity	Specificity	Accuracy	Ov-Se	Ov-Sp	Ov-Acc
PCANet- K_1	AMD	100.0 ± 0.0	99.0 ± 1.6	99.3 ± 1.1	98.2 ± 1.4	99.1 ± 0.7	98.8 ± 0.9
	ME	98.0 ± 3.2	98.3 ± 1.8	98.2 ± 1.4			
	NM	96.7 ± 3.5	100.0 ± 0.0	98.9 ± 1.2			
PCANet- K_2	AMD	84.0 ± 3.4	92.3 ± 2.3	89.6 ± 2.1	83.8 ± 2.1	91.9 ± 1.1	89.2 ± 1.4
	ME	68.0 ± 2.8	97.3 ± 2.1	87.6 ± 2.2			
	NM	99.3 ± 2.1	86.0 ± 2.1	90.4 ± 1.5			
PCANet- K_3	AMD	52.0 ± 6.9	69.3 ± 7.3	63.6 ± 5.6	43.3 ± 4.2	71.7 ± 2.1	62.2 ± 2.8
	ME	46.7 ± 8.3	67.0 ± 7.6	60.2 ± 5.8			
	NM	31.3 ± 9.5	78.7 ± 5.7	62.9 ± 4.7			
PCANet-CK	AMD	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	ME	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0			
	NM	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0			

Table 5 Classification results (%) on HUCM dataset using single kernel.

Methods	Classes	Sensitivity	Specificity	Accuracy	Ov-Se	Ov-Sp	Ov-Acc
PCANet- K_1	AMD	89.4 ± 4.1	95.0 ± 1.2	93.2 ± 1.3	93.2 ± 1.3	96.6 ± 0.6	95.4 ± 0.8
	ME	93.3 ± 2.3	98.6 ± 1.5	96.9 ± 1.3			
	NM	96.7 ± 2.8	96.1 ± 1.9	96.3 ± 1.2			
PCANet- K_2	AMD	68.9 ± 2.9	86.9 ± 3.9	80.9 ± 2.5	72.4 ± 2.5	86.2 ± 1.3	81.6 ± 1.7
	ME	62.8 ± 7.8	87.8 ± 1.9	79.4 ± 3.0			
	NM	85.6 ± 2.9	83.9 ± 3.7	84.4 ± 2.3			
PCANet- K_3	AMD	36.7 ± 6.9	61.1 ± 7.8	53.0 ± 7.9	34.3 ± 5.9	67.1 ± 2.9	56.2 ± 4.0
	ME	33.3 ± 9.4	70.0 ± 5.4	57.8 ± 6.0			
	NM	32.8 ± 8.4	70.3 ± 6.0	57.8 ± 5.4			
PCANet-CK	AMD	92.2 ± 4.7	96.9 ± 0.9	95.4 ± 1.6	95.4 ± 1.6	97.7 ± 0.8	96.9 ± 1.1
	ME	94.4 ± 0.0	98.3 ± 1.4	97.0 ± 1.0			
	NM	99.4 ± 1.8	97.8 ± 1.8	98.3 ± 1.1			

filter size further increases, our performance will become stable or even worse. Since utilizing a larger PCA filter will create higher computational cost, the PCA filter size is set to 11. For the evaluation of block size, we varied the block size ($n_{B1} = n_{B2}$) in the output stage in a range from 3 to 19. Figure 6(b) shows the classification results with different block sizes on two test datasets. As can be seen, our PCANet-CK method is comparatively stable and achieves the best performance with the block sizes from of 3×3 , 5×5 , and 11×11 on two datasets. In this paper, we set the block size to be 11×11 .

4.5 Comparisons of the Composite Kernel with Single Kernel

We also conducted additional experiments using a single-kernel ELM classifier to validate the superiority of the composite kernel classifier. Specifically, after obtaining three PCs for one volume, we use them to separately create three different kernels. However, instead of fusing these kernels together, only one of them is utilized to train the ELM classifier and test one volume. The method using the first kernel is called the

PCANet-K₁. The method using the second kernel is denoted the PCANet-K₂. The method using the third kernel is termed the PCANet-K₃. The results for the Duke and HUCM datasets are tabulated in Tables 4 and 5. As can be observed, in both the Duke and HUCM datasets, the composite kernel method consistently performs better than the single kernel method in terms of all quantitative metrics, which demonstrates the effectiveness of the composite kernel for exploiting the correlations among adjacent B-scans within one OCT volume. The results also show that the composite kernel ELM can utilize complementary information of different single kernels to further enhance classification.

5 Conclusion and Future Works

In this paper, we present a fully automatic method named the PCANet-CK to identify AMD, ME, and NM using the 3-D retinal SD-OCT images. Instead of adopting hand-crafted features, the proposed PCANet-CK method can automatically learn features from the input OCT images without layer segmentation. In addition, the PCANet-CK utilizes a composite kernel to exploit the strong correlations among features of the 3-D OCT images for classification. Experimental results on two clinically acquired OCT datasets demonstrate the effectiveness of the proposed PCANet-CK method.

In this paper, an automatic retinal OCT image classification algorithm, which can achieve high classification accuracy in identifying AMD, ME, and NM, has been developed. The algorithm may be considered an effective computer-aided diagnosis tool for improving clinically OCT-based ophthalmic disease diagnosis and supporting remote clinical applications.

Note that each kind of disease (e.g., AMD) has large variations on its corresponding lesions (e.g., drusen in different sizes and shapes acquired from patients of different countries and also affected by different illuminations and noise). To better represent and classify the disease, we need to collect more training OCT datasets and then learn a more general model. This is one of our ongoing works, and it is expected to further improve the classification performance.

Since OCT images are 3-D volumetric data and the same pathological structures usually exist in several adjacent cross-sectional slices, one alternative way of feature extraction is to adopt the 3-D-PCANet model with 3-D PCA filters, which can be expected to better capture the main variation of all the 3-D cubes. In addition, our future publication will extend our algorithm to other retinal diseases, such as macular hole, macular telangiectasia, and central serous chorioretinopathy.

Disclosures

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China for Distinguished Young Scholars under Grant Nos. 61325007, the National Natural Science Foundation under Grant Nos. 61771192 and 61471167, the National Natural Science Foundation for Young Scientist of China under Grant No. 61501180, and China Postdoctoral Science Foundation funded Project No. 2017T1100597.

References

1. D. A. Quillen, "Common causes of vision loss in elderly patients," *Am. Fam. Physician* **60**(1), 99–108 (1999).
2. F. E. Hirai et al., "Clinically significant macular edema and survival in type 1 and type 2 diabetes," *Am. J. Ophthalmol.* **145**(4), 700–706 (2008).
3. R. Klein et al., "The five-year incidence and progression of age-related maculopathy: the Beaver Dam eye study," *Ophthalmology* **104**(1), 7–21 (1997).
4. D. A. Antonetti et al., "Diabetic retinopathy," *Diabetes* **55**(9), 2401–2411 (2006).
5. 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**(5), 2732–2744 (2017).
6. D. Huang et al., "Optical coherence tomography," *Science* **254**(5035), 1178–1181 (1991).
7. C. K. Hitzberger et al., "Three-dimensional imaging of the human retina by high-speed optical coherence tomography," *Opt. Express* **11**(21), 2753–2761 (2003).
8. A. M. Zysk et al., "Optical coherence tomography: a review of clinical development from bench to bedside," *J. Biomed. Opt.* **12**(5), 051403 (2007).
9. M. Wojtkowski et al., "In vivo human retinal imaging by Fourier domain optical coherence tomography," *J. Biomed. Opt.* **7**(3), 457–463 (2002).
10. S. S. Gao et al., "Quantification of choroidal neovascularization vessel length using optical coherence tomography angiography," *J. Biomed. Opt.* **21**(7), 076010 (2016).
11. G. J. Tearney, I. K. Jang, and B. E. Bouma, "Optical coherence tomography for imaging the vulnerable plaque," *J. Biomed. Opt.* **11**(2), 021002 (2006).
12. M. Mogensen et al., "OCT imaging of skin cancer and other dermatological diseases," *J. Biophotonics* **2**(6–7), 442–451 (2009).
13. D. C. Adler et al., "Three-dimensional endomicroscopy of the human colon using optical coherence tomography," *Opt. Express* **17**(2), 784–796 (2009).
14. W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," *Prog. Retinal Eye Res.* **27**(1), 45–88 (2008).
15. V. J. Srinivasan et al., "High-definition and 3-dimensional imaging of macular pathologies with high-speed ultrahigh-resolution optical coherence tomography," *Ophthalmology* **113**(11), 2054–2065 (2006).
16. R. J. Zawadzki et al., "Adaptation of a support vector machine algorithm for segmentation and visualization of retinal structures in volumetric optical coherence tomography data sets," *J. Biomed. Opt.* **12**(4), 041206 (2007).
17. C. A. Puliafito et al., "Imaging of macular diseases with optical coherence tomography," *Ophthalmology* **102**(2), 217–229 (1995).
18. L. Fang et al., "Segmentation based sparse reconstruction of optical coherence tomography images," *IEEE Trans. Med. Imaging* **36**(2), 407–421 (2017).
19. M. R. Hee et al., "Quantitative assessment of macular edema with optical coherence tomography," *Arch. Ophthalmol.* **113**(8), 1019–1029 (1995).
20. B. Baumann et al., "Segmentation and quantification of retinal lesions in age-related macular degeneration using polarization-sensitive optical coherence tomography," *J. Biomed. Opt.* **15**(6), 061704 (2010).
21. J. Sugmk, S. Kiattisin, and A. Leelasantham, "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).
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**(5), 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**(10), 3568–3577 (2014).
24. 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**(4), 455–463 (2016).
25. Y. K. Sun, S. Li, and Z. Y. Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," *J. Biomed. Opt.* **22**(1) (2017).

26. F. G. Venhuizen et al., "Automated age-related macular degeneration classification in OCT using unsupervised feature learning," *Proc. SPIE* **9414**, 94141I (2015).
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**(2), 579–592 (2017).
28. 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**(12), 4928–4940 (2016).
29. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013).
30. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
31. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25 (NIPS '12)*, pp. 1097–1105 (2012).
32. P. Prentas et al., "Segmentation of the foveal microvasculature using deep learning networks," *J. Biomed. Opt.* **21**(7), 075008 (2016).
33. T. H. Chan et al., "PCANet: a simple deep learning baseline for image classification?" *IEEE Trans. Image Process.* **24**(12), 5017–5032 (2015).
34. L. Fang et al., "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.* **53**(8), 4186–4201 (2015).
35. G. B. Huang et al., "Extreme learning machine for regression and multi-class classification," *IEEE Trans. Syst., Man, Cybern. B* **42**(2), 513–529 (2012).
36. G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing* **70**(1), 489–501 (2006).
37. L. Fang, H. Zhuo, and S. Li, "Super-resolution of hyperspectral image via superpixel-based sparse representation," *Neurocomputing* **273**, 171–177 (2018).
38. K. Dabov et al., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).
39. L. Fang et al., "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.* **66**(7), 1646–1657 (2017).
40. K.R. Muller et al., "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks* **12**(2), 181–201 (2001).
41. X. Liu et al., "Multiple kernel extreme learning machine," *Neurocomputing* **149**, 253–264 (2015).
42. L. Fang et al., "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.* **53**(12), 6663–6674 (2015).
43. S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.* **2**(1), 37–52 (1987).
44. http://www.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm.
45. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).

Leyuan Fang received his PhD from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2015. From August 2016 to 2017, he was a postdoc researcher at the Department of Biomedical Engineering, Duke University, Durham, USA. Since January 2017, he has been an associate professor at the College of Electrical and Information Engineering, Hunan University. His research interests include sparse representation and deep learning in medical image processing.

Chong Wang received his BS degree from Southern Medical University, Guangzhou, China, in 2017. He is currently working toward the master degree at the Laboratory of Vision and Image Processing, Hunan University, Changsha, China. His research interests focus on deep learning for medical image processing.

Shutao Li received his BS, MS, and PhD degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively. From 2002 to 2003, he was a postdoctoral fellow at the Royal Holloway College, University of London, London, U.K. He is currently a Cheung Kong Scholars Professor at the College of Electrical and Information Engineering, Hunan University. His current research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Jun Yan received his MS degree from Hunan University, Changsha, China, in 2017. His research interests focus on deep learning for optical coherence tomography image processing.

Xiangdong Chen received his MD degree from Hunan University of Chinese Medicine, Changsha, China, in 2017. He is currently a full professor at Hunan University of Chinese Medicine, Changsha, China. His current research interests include analysis and diagnosis of age related macular and glaucoma diseases.

Hossein Rabbani received his MS and PhD degrees in biomedical engineering (bioelectronics) from Amirkabir University of Technology (Tehran Polytechnic). He is currently an associate professor at the Biomedical Engineering Department and also at Medical Image & Signal Processing Research Center of Isfahan University of Medical Sciences. His current research interests include medical image analysis and modeling, statistical (multidimensional) signal processing, sparse transforms, and image/video restoration.