

# A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images

Ahmed ElTanboly\*

*Department of Mathematical Engineering, Mansoura University, Mansoura 35516, Egypt  
Department of Bioengineering, University of Louisville, Louisville, KY 40292, USA*

Marwa Ismail,\* Ahmed Shalaby,\* Andy Switala, and Ayman El-Baz<sup>a)</sup>

*Department of Bioengineering, University of Louisville, Louisville, KY 40292, USA*

Shlomit Schaal

*Department of Ophthalmology and Visual Sciences, School of Medicine, University of Louisville, Louisville, KY 40202, USA*

Georgy Gimel'farb

*Intelligent Vision Systems Laboratory, Department of Computer Science, University of Auckland, Auckland 1142, New Zealand*

Magdi El-Azab

*Department of Mathematics and Physical Engineering, Mansoura University, Mansoura 35516, Egypt*

(Received 29 July 2016; revised 9 September 2016; accepted for publication 15 November 2016; published 16 March 2017)

**Purpose:** Detection (diagnosis) of diabetic retinopathy (DR) in optical coherence tomography (OCT) images for patients with type 2 diabetes, but almost clinically normal retina appearances.

**Methods:** The proposed computer-aided diagnostic (CAD) system detects the DR in three steps: (a) localizing and segmenting 12 distinct retinal layers on the OCT image; (b) deriving features of the segmented layers, and (c) learning most discriminative features and classifying each subject as normal or diabetic. To localise and segment the retinal layers, signals (intensities) of the OCT image are described with a joint Markov-Gibbs random field (MGRF) model of intensities and shape descriptors. Each segmented layer is characterized with cumulative probability distribution functions (CDF) of its locally extracted features, such as reflectivity, curvature, and thickness. A multistage deep fusion classification network (DFCN) with a stack of non-negativity-constrained autoencoders (NCAE) is trained to select the most discriminative retinal layers' features and use their CDFs for detecting the DR. A training atlas was built using the OCT scans for 12 normal subjects and their maps of layers hand-drawn by retina experts.

**Results:** Preliminary experiments on 52 clinical OCT scans (26 normal and 26 with early-stage DR, balanced between 40–79 yr old males and females; 40 training and 12 test subjects) gave the DR detection accuracy, sensitivity, and specificity of 92%; 83%, and 100%, respectively. The 100% accuracy, sensitivity, and specificity have been obtained in the leave-one-out cross-validation test for all the 52 subjects.

**Conclusion:** Both the quantitative and visual assessments confirmed the high accuracy of the proposed computer-assisted diagnostic system for early DR detection using the OCT retinal images. © 2016 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.12071]

**Key words:** diabetic retinopathy (DR), joint image-region-map model, Markov–Gibbs random field (MGRF), non-negativity-constrained autoencoder (NCAE), optical coherence tomography (OCT)

## 1. INTRODUCTION

Most of today's ophthalmologists identify types of diseases by visual observation and interpretation. Because accurate diagnosis is a must for successful timely treatment, the identification can be notably improved by employing disease-specific computer-aided diagnostic (CAD) systems based on optical coherence tomography (OCT). OCT has become a powerful imaging modality for non-invasive diagnosis of various retinal abnormalities, such as glaucoma, macular degeneration, and diabetic macular edema. However, detecting early retinal changes in type 2 diabetes to give patients a chance to delay further complications is absent so far.<sup>1</sup> Despite up to

21% and 61% of type 2 diabetics have retinopathy of some kind at the time of diagnosis and during the first two decades of the disease, respectively.<sup>2</sup> The main challenge in diagnosing diabetic retinopathy (DR) before complications arise is that the patient is not aware of the disease until the retinal changes have progressed to a considerable level and treatment tends to be less effective. Automated early DR detection, which could help ophthalmologists to more efficiently investigate and treat the disease and make it less severe, was the main goal of this paper. The second goal was to decrease errors in segmenting retinal layers on the OCT images.

Most of CAD systems for early DR detection use fundus images of the patient. Mizutani et al.<sup>3</sup> explored a trainable

automated detection of micro-aneurysms, manifesting, by supposition, early DR on these images. Experiments showed the sensitivity of 65% at 27 false positives per image. Jaafar *et al.*<sup>4</sup> detected hard and soft exudates as the earliest DR indicators. However, their diagnostic success depended on whether those candidates exist. Filtering, morphological processing, and thresholding of the fundus images were used by Pachiyappan *et al.*<sup>5</sup> to detect macular abnormalities caused by DR. Tan *et al.*<sup>6</sup> extracted retinal vasculature automatically to get a network of blood vessels. Knowing their shape, together with connections and salient points, helps the user in editing segmentation results. Fundus imaging is more common in the present DR-detecting CAD systems because of using the same concepts as traditional indirect ophthalmoscopy to form a wide view of the retina and adequately present systemic diseases. However, one of its crucial drawbacks is the acquisition of only 2D images with no account of depth. Unlike the fundus imaging, which is expensive due to required specialized highly trained technicians and allows for only qualitative evaluations, OCT is more cost-effective, supports quantitative measurements, and can evaluate changes with no human bias. Detecting subtle changes in retinal thickness, as diabetes progresses, makes OCT a powerful tool for diagnosing diabetic macular edema. Early detection of DR lesions can be based on the OCT biomarkers, such as retinal volume and total thickness, together with micro aneurysms. Thinning retinal thickness and decreasing optical reflectivity are significant biomarkers for detecting DR changes with OCT.

Segmenting retinal layers from the OCT images, which is a key component of our CAD system, is briefly overviewed below. Rossant *et al.*<sup>7</sup> automatically segmented eight layers using active contours, *k*-means clustering, Markov random field models, and Kalman filtering to account approximately for similarities between the layers. Their approach performed well, but failed on blurred images. A semi-automated extraction of nine layers from the OCT images by Yaz *et al.*<sup>8</sup> used the Chan and Vese's energy-minimizing active contour without an edge model along with shape priors. The segmentation, however, had to be manually initialized by the user and was tested on neither normal nor pathological human retinas. Kafieh *et al.*<sup>9</sup> used graph-based diffusion maps to segment intraretinal layers in the OCT scans for normal controls and glaucoma patients, and Ehnes *et al.*<sup>10</sup> developed a graph-based segmentation of up to 11 layers, yet it worked only with the high-contrast images. An automated choroid segmentation by Chen *et al.*<sup>11</sup> considers properties of choroidal vasculature for enhanced depth imaging OCT (EDI-OCT), in particular, the large vascular of the Haller's layer neighbor with the choroid-sclera junction. Satisfactory, i.e., similar to manual results for normal eyes and pathologies (DR and age-related ones) were obtained, due to good visibility and easy detection of the choroidal vasculature. Nam *et al.*<sup>12</sup> detected lumen contours and stent strut candidates for fundus images by using high-resolution intravascular OCT to visualize the arterial walls microstructure and combining the OCT intensity images with their first- and second-order axial derivatives.

It should be noted that most of the known methods for segmenting the OCT layers are sufficiently accurate only on images with high signal-to-noise ratios (SNR), but fail if the SNR is low; are tested on normal, but not abnormal (disease) cases, and stratify only up to eight layers. More layers could be segmented only on the high-contrast images of normal retinas,<sup>10</sup> so that to be robust and successful the segmentation either focuses on specific retinal parts or requires high-resolution images.<sup>6,11,12</sup> All the limitations led to missed diagnoses of certain abnormalities if more than one segmented layer should be combined, like e.g., in age-related macular degeneration (AMD) cases when the retinal pigment epithelium (RPE) and interdigitation zone (IZ) should be extracted as two separate layers for the better diagnostics.<sup>13</sup>

The introduced below innovative CAD system for automated early DR detection overcomes most of the aforementioned problems by segmenting 12 distinct retinal layers from an input OCT image and extracting three quantitative discriminative features, which can distinguish between the normal and DR cases. To perform well on the realistic OCT images, including the low-SNR ones for both the normal cases and different diseases, our segmentation employs a novel probabilistic model of the input OCT images and output region maps of segmented layers. The model combines image intensities, a prior knowledge of shapes of regions, and descriptors of local spatial dependencies between region labels. To the best of our knowledge, the proposed CAD system is the first to use and segment the OCT scans into 12 distinct layers for detecting the early DR.

## 2. MATERIALS AND METHODS

Figure 1 summarizes the main sequential steps of our automated non-invasive CAD system in processing and analyzing an input OCT image for early DR detection: (a) segmenting the image of a test subject into 12 distinct retinal layers; (b) extracting three global discriminative features (curvature, reflectivity, and thickness) of statistically significant layers; and (c) detecting the normal or DR state of the test subject by using a pre-trained deep fusion classification network (DFCN). That the outer retinal layers were found significant agrees with multiple electroretinogram (ERG) studies,<sup>14</sup> showing the selected features can measure a DR-related photoreceptor damage preceding anatomical vascular changes. Sections 2.A–2.C will detail this CAD system.

### 2.A. OCT layers segmentation

Let  $\mathbf{g} = \{g_{x,y} : (x,y) \in \mathbb{R}; g_{x,y} \in \mathbf{Q}\}$  and  $\mathbf{m} = \{m_{x,y} : (x,y) \in \mathbb{R}; m_{x,y} \in \mathbf{L}\}$  denote a grayscale image and an associated region map on a finite arithmetic lattice  $\mathbb{R}$  and with values from finite sets  $\mathbf{Q}$  and  $\mathbf{L}$  of integer intensities and region labels, respectively. An input OCT image,  $\mathbf{g}$ , co-aligned to a given training database, and its region map,  $\mathbf{m}$ , are described with a joint probability model,  $P(\mathbf{g}, \mathbf{m}) = P(\mathbf{g}|\mathbf{m})P(\mathbf{m})$

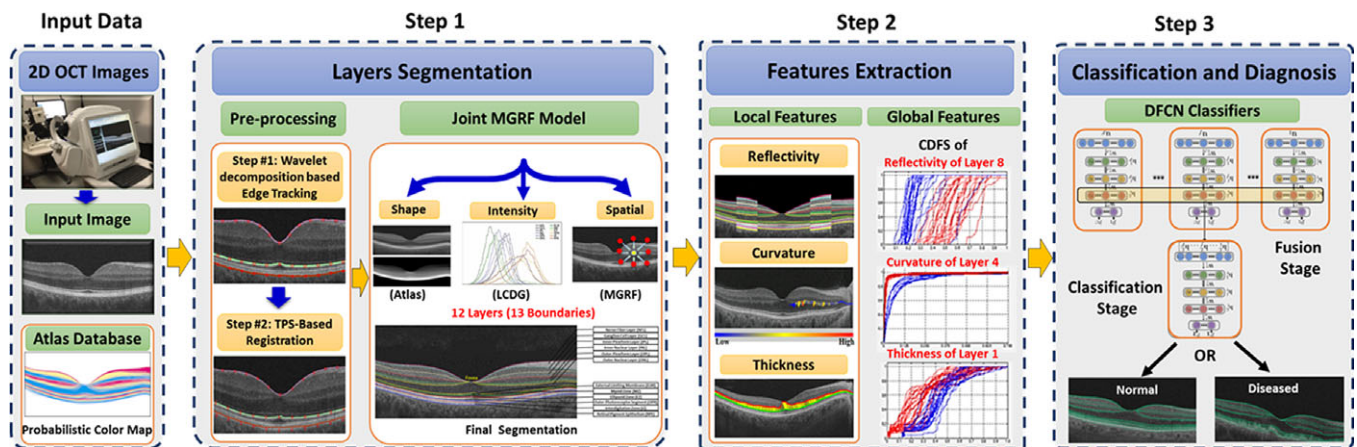


FIG. 1. Basic steps of the proposed diabetic retinopathy (DR) detection. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

including a conditional image probability,  $P(\mathbf{g}|\mathbf{m})$ , given a map, and an **unconditional map probability**,  $P(\mathbf{m}) = P_{\text{sp}}(\mathbf{m})P_{\text{v}}(\mathbf{m})$ . The latter combines an adaptive shape prior,  $P_{\text{sp}}(\mathbf{m})$  and a **second-order MGRF** model,  $P_{\text{v}}(\mathbf{m})$ , of spatially homogeneous maps  $\mathbf{m}$  with Gibbs potentials  $\mathbf{V}$ .

### 2.A.1. Adaptive shape prior,

$$P_{\text{sp}}(\mathbf{m}) = \prod_{(x,y) \in \mathbb{R}} P_{\text{sp}:x,y}(m_{x,y})$$

The shape **prior** is constructed using several training OCT scans (**six male and six female images** in our experiments below), selected to capture biological variability of the whole dataset. Their “ground truth” region maps were delineated under supervision of retina specialists. **A manually segmented retina with 12 distinct layers** — NFL, GCL, IPL, INL, OPL, ONL, ELM, MZ, EZ, OPS, IZ, and RPL — is exemplified in Fig. 2.

Using one good scan (no tilt, central fovea location) as a reference, all the other training images are co-registered by **thin-plate spline** (TPS) deformations.<sup>15</sup> The same deformations are applied to each respective ground truth map. The

pixel-wise labels across a stack of the co-aligned maps lead to the prior shape probabilities,  $P_{\text{sp}:x,y}(k)$ ;  $k \in \mathbb{L}$ , of the typical retina.

An image to be segmented is **first aligned to the shape database** by a new technique integrating the TPS with multi-resolution edge tracking that **identifies control points to initialize the alignment**. First, the “à trous” algorithm<sup>16</sup> decomposes each scan by undecimated wavelet transform. In a three-band appearance of the retina, two hyperreflective bands are separated by a hyporeflective band, corresponding roughly to the layers from ONL to MZ in Fig. 2. Contours following the gradient maxima of this wavelet component provide initial estimates of the vitreous/NFL, MZ/EZ, and RPE/choroid boundaries.

The **foveal peak** is identified as the closest point between the vitreous/NFL and MZ/EZ contours, and control points are located on these boundaries at the foveal peak and at uniform intervals from it. **The optimized TPS aligns the input image to the shape database using these control points**, so that the shape prior can be used to segment the aligned image.

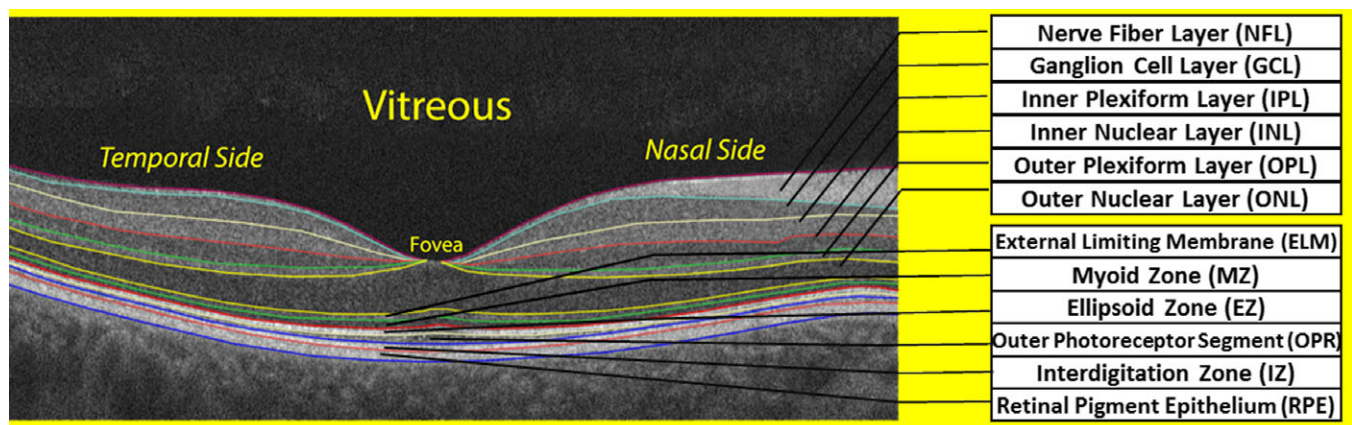


FIG. 2. Twelve distinct retinal layers in a typical optical coherence tomography (OCT) scan for a normal subject. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



### 2.A.2. First-order conditional intensity model,

$$P(\mathbf{g}|\mathbf{m}) = \prod_{(x,y) \in \mathbb{R}} P(\mathbf{g}_{x,y}|\mathbf{m}_{x,y})$$

To account for visual appearance of the input image, its histogram of the pixel-wise intensities (gray levels) is collected and normed to form the empirical marginal probability distribution of the intensities. The latter is closely approximated with a linear combination of sign-alternate discrete Gaussians (LCDG) and separated into individual components (also LCDGs) for different regions, associated each with a dominant mode. This model and its expectation–maximization-based learning are detailed in Ref. (17).

### 2.A.3. Second-order MGRF model of spatial dependences, $P_V(\mathbf{m})$

For better spatial homogeneity of segmentation, and to more accurately account for noise, a second-order Markov–Gibbs random field (MGRF) model of spatial interactions between region labels is combined with the above shape prior and intensity model.<sup>18</sup> For simplicity and by symmetry considerations, we use a popular Potts model of homogeneous region maps, which accounts for pairwise equality or inequality of region labels in the nearest eight-neighborhood of each pixel:

$$P_V(\mathbf{m}) \propto \exp \left( \sum_{(x,y) \in \mathbb{R}} \sum_{(\xi,\eta) \in \mathbf{v}_s} \mathbf{V}(l_{x,y}, l_{x+\xi, y+\eta}) \right)$$

where coordinate offsets  $\mathbf{v}_s = \{(1,0), (-1,1), (0,1), (1,1)\}$  define non-repetitive neighboring pairs of pixels in the lattice  $\mathbb{R}$  and bi-valued Gibbs potentials  $\mathbf{V} = (V(k, k') : k, k' \in \mathbb{L})$  are as follows:  $V(k, k') = \gamma \geq 0$  if  $k = k'$  and  $-\gamma$  otherwise. The potentials are approximated analytically from the empirical probability of equal label pairs in the training region maps.

Algorithm 1 summarizes the segmentation stage of the proposed computer-aided diagnostic (CAD) system.

**Algorithm 1** Segmentation of the optical coherence tomography (OCT) layers.

1. Input a test image and the atlas, or dataset of co-aligned training images and maps.
2. Align the test image to the atlas:
  - (a) Apply the “à trous” wavelet decomposition algorithm to the input image and locate the vitreous/NFL, IPL/EL, and RPE/choroid boundaries.
  - (b) Identify the foveal peak using the three located boundaries.
  - (c) Place fiducial points on each detected boundary, starting at the foveal peak and moving outward at equidistant intervals.
  - (d) Align the input image to the atlas using the control points identified.
3. Initial Segmentation:
  - (a) For each aligned input pixel,  $(x, y) \in \mathbb{R}$ , search within a certain tolerance interval for the closest atlas intensity over an initial 2D window centered at that pixel.
  - (b) Increase, if necessary, the window size and repeat the search until the closest by intensity atlas pixel is found.
  - (c) If the maximum window size is reached, but no atlas pixel is found, increase the intensity tolerance and go back to Step 3a.

Algorithm 1. Continued.

**Algorithm 1** Segmentation of the optical coherence tomography (OCT) layers.

- (d) Take from the atlas the prior probabilities,  $P_{\text{sp},x,y}(k); k \in \mathbb{L}$ , of retinal layers for the input pixel.
4. Final Segmentation:
  - (a) Approximate the marginal intensity distribution of the test image with an LCDG with four dominant modes (Section 2.A.2).
  - (b) Form an initial region map using the estimated LCDGs and the shape prior.
  - (c) Compute analytically the approximate Gibbs potentials for the MGRF model of the initial map (Section 2.A.3).
  - (d) Form the final segmentation map using the joint probability model detailed in Section 2.A.

## 2.B. Feature extraction

Three features quantifying “reflectivity”, “curvature”, and “thickness” of retinal layers are extracted from each segmented OCT scan. The reflectivity is obtained from two regions per scan, comprising the thickest portions of the retina on the nasal and temporal sides of the foveal peak (Fig. 3). A normalized reflectivity scale (NRS) is defined to take a value of 0 arbitrary units (AU) in the vitreous and 1000 AU in the RPE. The average intensity within a segment is calculated using Huber’s M-estimates, which are robust to possible outlying values, such as very bright pixels in the innermost segment that belong to the internal limiting membrane, rather than the NFL.

The curvature of a retinal layer combines all Menger curvature values calculated for each point across the layer (Fig. 3) after local weighted polynomial smoothing of the surface. The thickness of a retinal layer is calculated from streamlines of the Laplace’s equation [19] between the two surfaces of the layer (Fig. 3).

For each subject, these three features of the extracted retina layers are described as a whole with a cumulative distribution function (CDF). The CDF for each layer contains 100 bins for the reflectivity and thickness and 1000 bins for the curvature. The CDFs are considered global discriminatory characteristics, being able to distinguish between the normal and DR cases. One of the contributions of the proposed system is using the CDF as a global input feature for a deep fusion classification network. The CDFs for a training set of the OCT images are used for deep learning of a multistage classifier with a stack of non-negative-constrained autoencoders (SNCAE). Figure 4 exemplifies the CDFs of the curvature of the 4th layer (INL), reflectivity of the 8th layer (MZ), and thickness of the 1st layer (NFL). This figure shows their significant potential, together with other layers, in statistical analysis and separation between the normal and DR cases. Section 3 will present more details.

It is worth noting that more conventional classification techniques, employing directly the pixel-wise curvature, reflectivity, and thickness as discriminative features,

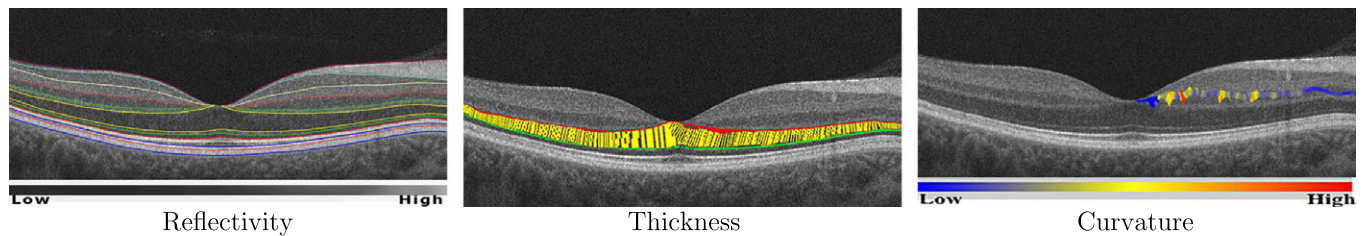


FIG. 3. Features for classification (the bottom gray and color bars encode the reflectivity and curvature values, respectively; the thickness is evaluated from streamlines (yellow) between the upper (red) and bottom (green) borders of a layer). [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

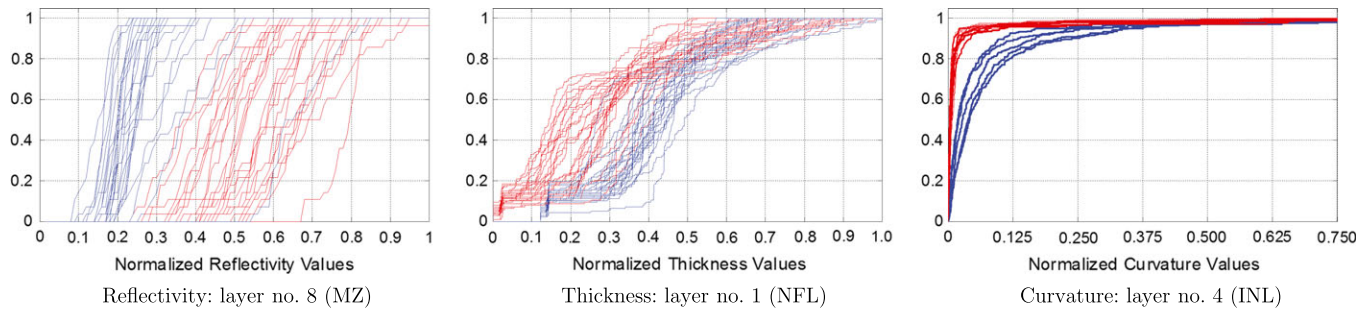


FIG. 4. CDFs of normalized feature values for all the 26 normal (red) and 26 DR (blue) cases. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

encounter serious difficulties. Due to various sizes for each subject, the individual feature collections have to be unified by either truncation of large volumes or zero padding of small ones. Both ways may decrease the classification accuracy of the classification. In addition, large data sizes lead to considerable time expenditures for training and classification. Contrastingly, our classifier exploits only the CDFs of fixed sizes to describe the entire raw data measured for each retinal layer. The fixed data size helps to overcome the above difficulties.

## 2.C. Deep fusion classification network (DFCN)

Deep learning network with a stack of non-negativity-constrained autoencoders (SNCAE) is employed in our CAD

system to discriminate between normal and DR subjects after segmenting the 12 retinal layers and extracting their global discriminatory features — the curvature, reflectivity, and thickness CDFs. Based on a training set of images for normal and DR subjects, the network learns to fuse the extracted CDFs into low-dimensional characteristic descriptors and classify other OCT images.

### 2.C.1 Structure of SNCAE

An individual autoencoder (AE) in Fig. 5 is specified by a set of column vectors of weights,  $\mathbf{W} = \{\mathbf{W}_{e:1}, \dots, \mathbf{W}_{e:n}; \mathbf{W}_{d:1}, \dots, \mathbf{W}_{d:s}\}$ , for encoding (e) and decoding (d) layers. The encoding layer converts an input  $n$ -dimensional column

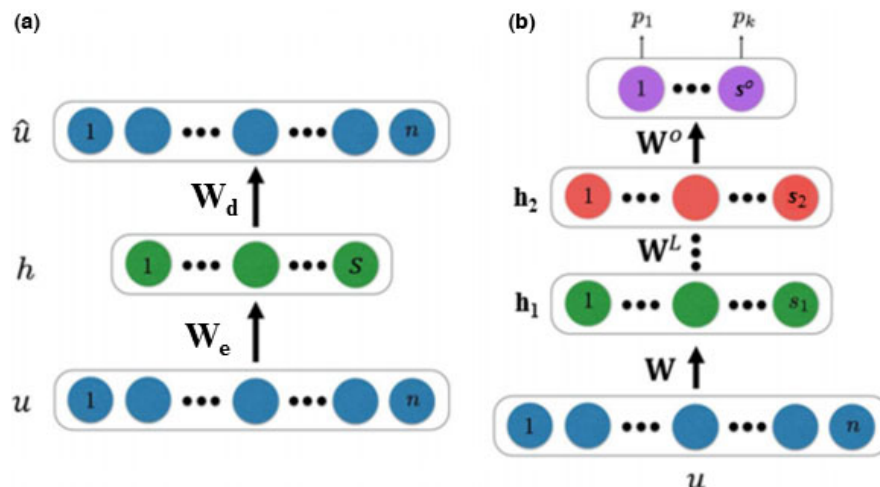


FIG. 5. Structures of an individual AE (a) with encoding,  $\mathbf{W}_e$ , and decoding,  $\mathbf{W}_d$ , weights and of an SNCAE (b) with two NCAE layers and an output softmax layer. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

vector  $\mathbf{u} = [u_1, \dots, u_n]^\top$  of scalar signals in the range  $[0,1]$  into an  $s$ -dimensional,  $s \ll n$ , column vector  $\mathbf{h} = [h_1, \dots, h_s]^\top$  of scalar codes, called features or activations, in the same range. The decoding layer converts the obtained vector  $\mathbf{h}$  of  $s$  codes back to the vector of  $n$  decoded signals,  $\hat{\mathbf{u}} = [\hat{u}_1, \dots, \hat{u}_n]^\top$ , in the same range. Both the encoding and decoding are done by uniform nonlinear transformations of the weighted linear combinations of signals:

$$h_j = \sigma(\mathbf{W}_{e,j}^\top \mathbf{u}) \equiv \sigma\left(\sum_{i=1}^n w_{e,j,i} u_i\right); \quad j = 1, \dots, s;$$

$$\hat{u}_i = \sigma(\mathbf{W}_{d,i}^\top \mathbf{h}) \equiv \sigma\left(\sum_{j=1}^s w_{d,i,j} h_j\right); \quad i = 1, \dots, n$$

where  $\sigma(\dots)$  is a certain sigmoid, i.e., a differentiable monotone scalar function taking values in the range  $[0,1]$ ;  $\top$  indicates the vector transposition, and  $\hat{u}_{w,i}$  is the decoded signal  $u_i$ . **Unsupervised pre-training of the AE minimizes the total reconstruction error**, i.e., the total quadratic ( $\ell_2$ ) norm of deviations between each training input vector,  $\mathbf{u}_k$ ;  $k = 1, \dots, K$ , and the same-dimensional vector,  $\hat{\mathbf{u}}_{w,k}$  decoded from its activation vector,  $\mathbf{h}_{w,k}$ :

$$J_{\text{AE}}(\mathbf{W}) = \frac{1}{2K} \sum_{k=1}^K \|\hat{\mathbf{u}}_{w,k} - \mathbf{u}_k\|^2 \quad (1)$$

To decrease the **number of negative weights and enforce sparsity of coding**, a non-negativity-constrained AE (NCAE) is built by appending the reconstruction error of Eq. (1) with two weighted terms penalizing, respectively, the negative weights and non-zero codes:

$$J_{\text{NCAE}}(\mathbf{W}) = J_{\text{AE}}(\mathbf{W}) + a \sum_{j=1}^s \sum_{i=1}^n f(w_{j,i}) + b J_{\text{KL}}(\mathbf{h}_w; \gamma) \quad (2)$$

Here, the factors  $a \geq 0$  and  $b \geq 0$  determine relative contributions of the non-negativity and sparsity constraints to the overall loss;  $f(w_i) = (\min\{0, w_i\})^2$ ;  $i = 1, \dots, n$ , and  $J_{\text{KL}}(\mathbf{h}_w; \gamma)$  is the Kullback–Leibler (KL) divergence of the codes (activations)  $\mathbf{h}_w$  from a fixed small positive average value,  $\gamma$ , near 0:

$$J_{\text{KL}}(\mathbf{h}_w, \gamma) = \sum_{j=1}^s h_{w,j} \log\left(\frac{h_{w,j}}{\gamma}\right) + (1 - h_{w,j}) \log\left(\frac{1 - h_{w,j}}{1 - \gamma}\right) \quad (3)$$

The proposed SNCAE in Fig. 5 is built by stacking **two NCAE layers with an output softmax layer**. Each NCAE is **pre-trained separately in the unsupervised mode** by using the activation vector of a lower layer as the input to the upper layer. The data compression at the NCAE layers was selected empirically after experimental comparisons of several variants. In the chosen structure, the bottom NCAE compresses the input vector to  $s_1 = 50$  first-level activators. The latter are compressed by the next NCAE to  $s_2 = 5$  second-level

activators, which are decreased in turn by the output softmax layer to the goal  $s^* = 2$  values.

**Separate pre-training of the first and second layers by minimizing the loss of Eq. (2) decreases the total reconstruction error, as well as increases sparsity of the activations and numbers of the non-negative weights.** The activations of the second NCAE layer,  $\mathbf{h}^{[2]} = \sigma(\mathbf{W}_{e,[2]}^\top \mathbf{h}^{[1]})$  in Fig. 5, are inputs of the softmax classification layer, computing probabilities of the normal ( $c = 1$ ) and DR ( $c = 2$ ) decisions:

$$p(c; \mathbf{W}_{\text{soft}:c}) = \frac{\exp(\mathbf{W}_{\text{soft}:c}^\top \mathbf{h}^{[2]})}{\exp(\mathbf{W}_{\text{soft}:1}^\top \mathbf{h}^{[2]}) + \exp(\mathbf{W}_{\text{soft}:2}^\top \mathbf{h}^{[2]})};$$

$$\sum_{c=1}^2 p(c; \mathbf{W}_{\text{soft}:c}) = 1$$

Its separate pre-training maximizes the mean log-likelihood  $L_{\text{soft}}(\mathbf{W}_{\text{soft}})$  of the known training classes, appended with the weighted term penalizing the negative weights:

$$L_{\text{soft}}(\mathbf{W}_{\text{soft}}) = \frac{1}{K} \sum_{k=1}^K \log p(c_k; \mathbf{W}_{\text{soft}}) + \vartheta \sum_{c=1}^2 \sum_{j=1}^{s_2} w_{\text{soft}:j:c} \quad (4)$$

Then the SNCAE is fine-tuned on the training data by traditional error back-propagation through the network and **penalties on only the negative weights** of the softmax layer.

## 2.C.2 Our deep fusion classifier

The classification is performed by a two-stage network of the SNCAE. The first stage consists of the SNCAE with the output softmax regression layer for each input feature (one SNCAE for each feature per significant layer, as shown in Fig. 6). Each SNCAE compresses its input data,  $\mathbf{u}_f$ , in order to capture most prominent variations and is built separately by greedy unsupervised pre-training.<sup>20</sup> The softmax output layer facilitates the subsequent supervised back-propagation-based fine-tuning of the entire classifier by maximizing the penalized log-likelihood of the training data of known classes. Each NCAE<sup>21</sup> yields both more reasonable data codes (features) after unsupervised pre-training and better classification accuracy after supervised refinement. The SNCAE forming the  $f$ -th input feature converts an input  $n$ -dimensional column vector  $\mathbf{u}_f = [u_{1:f}, \dots, u_{n:f}]^\top$ , into a column  $s$ -dimensional vector  $\mathbf{h}_f = [h_{1:f}, \dots, h_{s:f}]^\top$  of the codes (features).

**After each separately pre-trained and fine-tuned first-stage SNCAE formed its own feature vector,  $\mathbf{u}_f$ ;  $f = 1, \dots, F$ , from an input image, the second-stage SNCAE obtains concatenated top-most output codes,  $\mathbf{h}_f^{[F]}$ , of each SNCAE;  $f = 1, \dots, F$ .** The concatenated codes allow for capturing the discriminative information within each feature,  $\mathbf{u}_f$ , by training the second-stage SNCAE on the training data with the known class labels. The same layer-wise pre-training and fine-tuning of this SNCAE are used to accurately classify the fused feature vectors  $[\mathbf{h}_f^{[1]}, \dots, \mathbf{h}_f^{[L]}]$  on the basis of the final decision probabilities (Fig. 6):



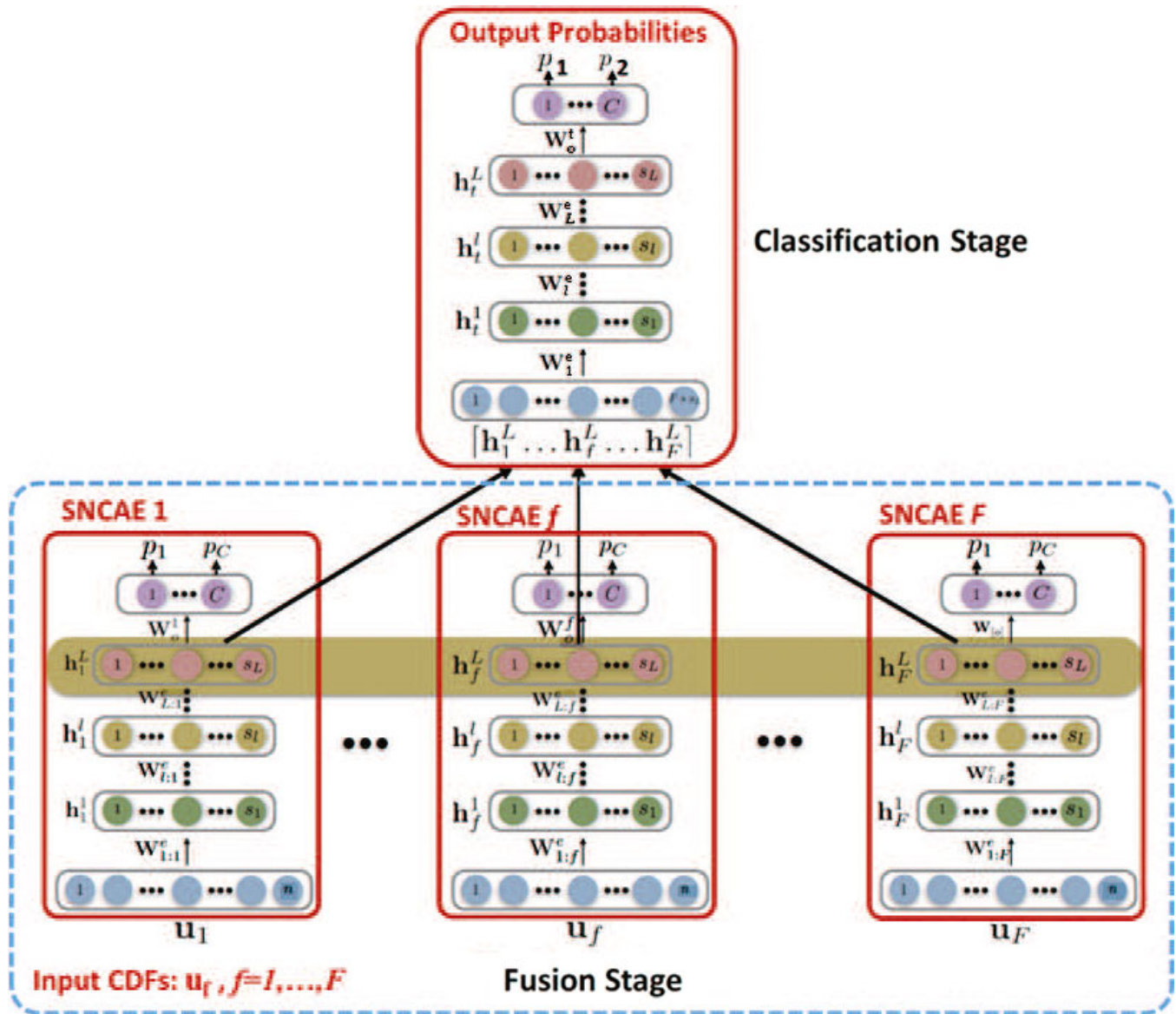


FIG. 6. DFCN structure. [Colour figure can be viewed at wileyonlinelibrary.com]

$$p_i(c; \mathbf{W}_{o:c}^t) = \frac{\exp\left((\mathbf{W}_{o:c}^t)^\top \mathbf{h}_i^{[L]}\right)}{\exp\left((\mathbf{W}_{o:1}^t)^\top \mathbf{h}_i^{[L]}\right) + \exp\left((\mathbf{W}_{o:c}^t)^\top \mathbf{h}_i^{[L]}\right)} \quad (5)$$

The proposed classifier was called the deep fusion classification network (DFCN), Fig. 6, as the decisions are based on the codes  $\mathbf{h}_i^{[L]}$  extracted and fused by a deep structure of layers, starting from the lowest retinal features,  $\mathbf{u}_f$ ,  $f = 1, \dots, F$ . The CDFs for each feature are fed to the AEs resulting in the 64 AEs for each scan.

### 3. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed CAD system was tested and validated on the spectral-domain OCT scans (Zeiss Cirrus HD-OCT 5000)

for 52 subjects (26 normal and 26 abnormal) aged 40–79 yr. Subjects with high myopia ( $\leq -6.0$  diopters), and tilted OCT were excluded. The ground truth was created with the aid of retina specialists by manual delineation of the retina layers.

The average execution time of segmenting and diagnosing a given OCT image 2-D of size  $1024 \times 1024$  pixels is 4 min. This estimated time is based on the implementation of the system on a single core of a 3.3 GHz Intel<sup>®</sup> Xeon processor with 48 GB of memory using MATLAB<sup>®</sup> prototype implementation.

Figure 7 compares our segmentation of 12 distinct retinal layers with the iterative algorithm by Chiu et al.,<sup>22</sup> converting an image into a nearest eight-neighborhood graph, such that weights between any pixel and its neighbours depend on only eight-directional intensity gradients, and using dynamic programming to find the minimum-weight paths to accurately segment eight retinal layer boundaries. In addition to the

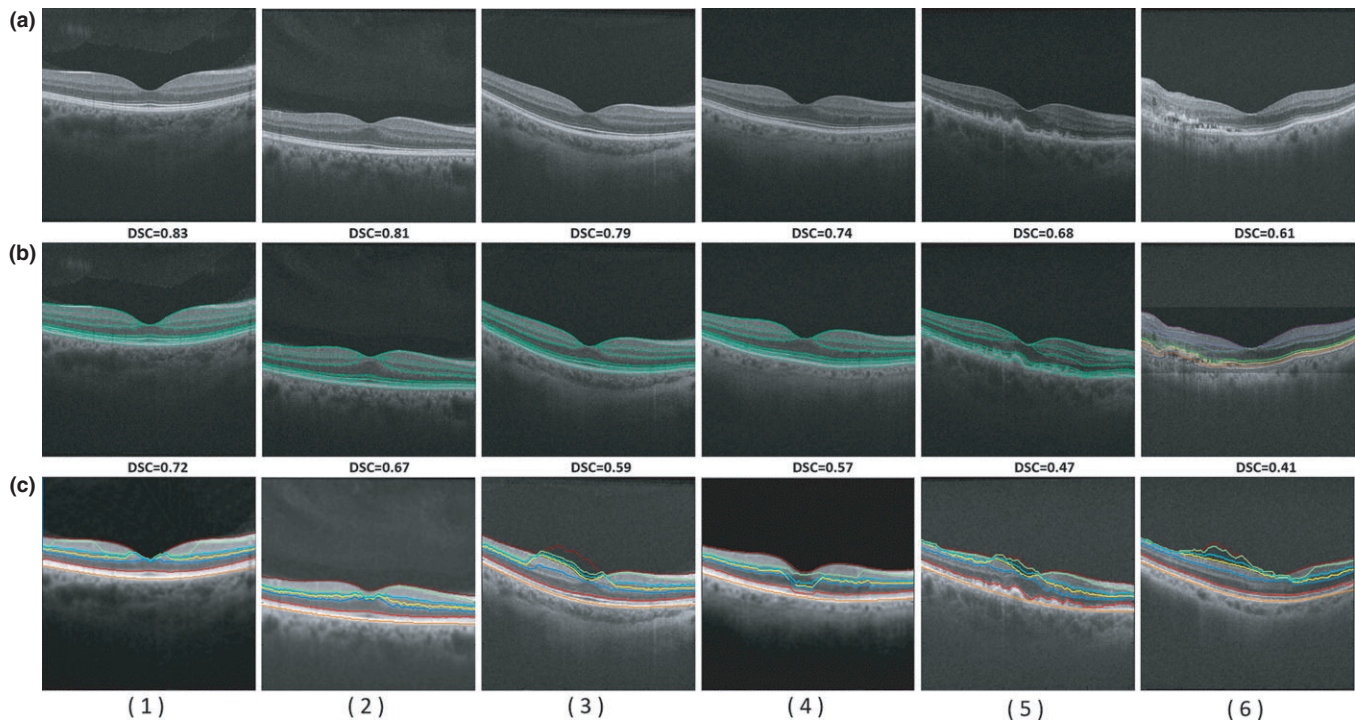


FIG. 7. Original OCT images (a) for normal (1–3), diabetic (4), and AMD (5–6) cases and our segmentation (b) vs. the algorithm in<sup>22</sup>(c). DSC scores above each result illustrate better accuracy of our approach (see also Table I), including noisy and low-contrast cases (4–6). [Colour figure can be viewed at wileyonlinelibrary.com]

visualized region maps, the robustness and accuracy of our approach were evaluated by measuring the agreement coefficient (AC) and Dice similarity coefficient, (DSC)<sup>23</sup> as well as the average deviation distance (AD)<sup>24</sup> to the ground truth. To be compared with the method in,<sup>22</sup> our segmentation was modified first by merging different layers together for obtaining the same number of layers before their statistical analysis. The mean boundary error, averaged across all the 13 boundaries, was 6.9µm from the ground truth. The inner (vitreous) boundary of the retina was placed most accurately, with the 2.8µm mean error. The worst performance of 12µm was on the outer (choroid) boundary. Contrastingly only the RPE/choroid boundary was reliably detected by the segmentation method in Ref. (22).

Table I summarizes quantitative comparisons of both the methods versus the ground truth for all the 52 subjects. According to the paired *t*-test, the found differences between the AC, DSC, and AD metrics for both the methods are statistically significant (the *P*-values <0.05). These results demonstrate the promise of the proposed segmentation in application to the OCT scans.

To evaluate discriminative capabilities of the chosen features (curvature, reflectivity, and thickness) of individual retinal layers, Table II presents statistics of their scatter over the available two groups of 26 normal and 26 diabetic subjects. It is worthy to note that the curvature values show *highly skewed* distribution as shown in Fig. 8 and thickness values have also, moderate skewed distribution. The unpaired *t*-tests suggest that differences between the INL curvature, MZ reflectivity, and NFL thickness over both the groups are

statistically significant (the *P*-values <0.05). For now, these results are encouraging enough to explore the discriminative power of only these layer-wise features.

A trainable DFCN used in our CAD system decides whether the test subject is normal or diabetic, based on the CDFs of the most discriminative features (the INL curvature, MZ reflectivity, and NFL thickness). A pilot quantitative comparison of the DFCN with four conventional classifiers (K-Star, K-Nearest Neighbor, Random Tree, and Random Forest) from the well-known statistical Waikato Environment for Knowledge Analysis (Weka) workbench<sup>25</sup> is summarized in Table III. As recommended by the NIH Biostatistics Group for a small number of datasets, we perform a *leave-one-out cross-validation test with all the 52 OCT images* (26 normal and 26 DR). The DFCN classifier achieves the 100% accuracy with the area under the receiver operating curve (AUC) of 0.98 for all the 52 datasets.

TABLE I. Comparative accuracy (mean ± standard deviation) in terms of the Dice similarity coefficient (DSC), the agreement coefficient (AC), and the average deviation distance (AD).

	Evaluation Metric		
	DSC	AC, %	AD, µm
Our segmentation	<b>0.76 ± 0.16</b>	<b>73 ± 4.5</b>	<b>6.9 ± 2.8</b>
Segmentation method in <sup>22</sup>	0.41 ± 0.26	2.3 ± 9.7	15 ± 8.6
<i>P</i> -value	<0.0001	<0.0001	<0.004

Bold indicates the better results from our approach compared with the other approach.



TABLE II. Feature scatter over the 12 retinal layers (mean±standard deviation).

Retina 12-layers	Reflectivity		Curvature, mm <sup>-1</sup>		Thickness, μm	
	Normal	Diabetic	Normal	Diabetic	Normal	Diabetic
NFL	667 ± 147	665 ± 165	0.71 ± 1.72	0.74 ± 2.03	<b>19.1 ± 13.0</b>	<b>21 ± 12.2</b>
GCL	584 ± 103	590 ± 111	0.89 ± 3.7	1.02 ± 4.63	34.7 ± 18.2	32.5 ± 17.9
IPL	638 ± 96.8	631 ± 117	0.86 ± 3.7	1.2 ± 4.7	34.0 ± 15.3	33.6 ± 16.2
INL	479 ± 95.8	464 ± 101	<b>1.2 ± 4.9</b>	<b>2.2 ± 6.2</b>	32.0 ± 15.5	27.9 ± 15.1
OPL	546 ± 93.7	514 ± 108	1.2 ± 4.9	1.8 ± 5.6	27.2 ± 14.3	30.6 ± 13.9
ONL	303 ± 74.3	292 ± 108	0.84 ± 3.1	1.7 ± 5.0	67.1 ± 23.6	61.6 ± 22.6
ELM	396 ± 135	465 ± 236	0.13 ± 0.25	0.10 ± 0.16	13.8 ± 3.7	14.7 ± 3.0
MZ	<b>414 ± 160</b>	<b>512 ± 309</b>	0.11 ± 0.20	0.10 ± 0.17	15.5 ± 4.3	14.0 ± 4.0
EZ	996 ± 189	1038 ± 281	0.12 ± 0.20	0.11 ± 0.20	14.6 ± 3.7	15.0 ± 3.1
OPR	823 ± 210	883 ± 271	0.13 ± 0.25	0.06 ± 0.01	15.0 ± 5.5	15.0 ± 5.5
IZ	1100 ± 180	1200 ± 190	0.08 ± 0.14	0.06 ± 0.09	14.3 ± 3.8	15.6 ± 3.1
RPE	1046 ± 98.0	1041 ± 135	0.08 ± 0.17	0.06 ± 0.11	23.5 ± 6.4	23.3 ± 6.8

Bold indicates the values of the chosen layers for the next step.

In the second experiment, the datasets were split arbitrarily into training (40 subjects) and testing (12 subjects) sets. The training set was used to adjust the DFCN parameters by a

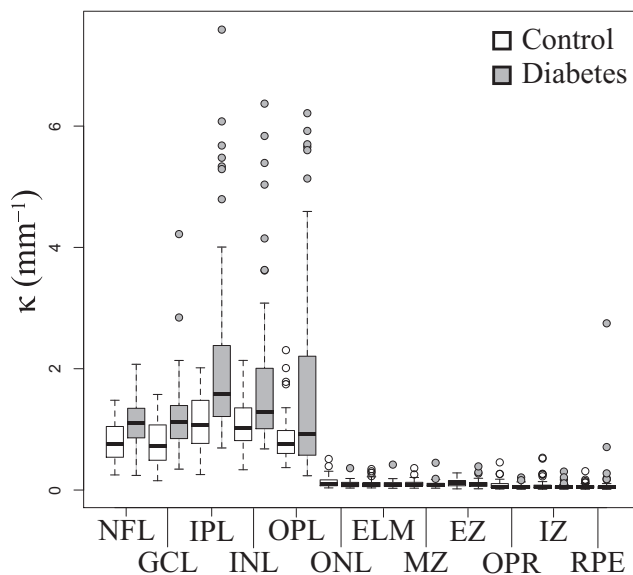


FIG. 8. A box plot of the curvature data for each layer, comparing control (healthy) ones vs. diabetes.

TABLE III. Accuracy, sensitivity, specificity, and AUC (area under the ROC curve) for our CAD system vs. four Weka classifiers.<sup>25</sup>

Classifier	Accuracy	Sensitivity	Specificity	AUC
DFCN (proposed)	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>0.98</b>
K-Star (K*)	95%	95%	95%	0.94
K-Nearest-Neighbor (K-NN)	90%	90%	90%	0.90
Random forest	85%	85%	85%	0.87
Random tree	72%	72%	72%	0.71

Bold indicates the better results from our approach compared with the other approach.

fourfold cross-validation, each fold containing 10 subjects. Threefolds were used for training and onefold for validation, this operation was repeated several times by changing the validation fold each time to evaluate the accuracy resulting in average accuracy of 95%. Then, after the DFCN was trained on the whole training set (40 subjects), the diagnostic accuracy was evaluated on the testing set (12 subjects). The resulting accuracy, sensitivity, and specificity were 92%, 84%, and 100%, respectively. The whole results of this experiment in Table IV confirm that the proposed DFCN outperformed again the four other classifiers.

#### 4. CONCLUSIONS

The proposed computer-aided diagnostic (CAD) system for detecting diabetic retinopathy (DR) on optical coherence tomography (OCT) images of the retina combines novel and efficient segmentation and classification techniques. After segmenting 12 distinct retinal layers, three quantitative features (pixel-wise curvature, reflectivity, and thickness) are

TABLE IV. Comparing our deep fusion classification network (DFCN) with four conventional classifiers: the average fourfold cross-validation accuracy for the training set (40 subjects) and the accuracy, sensitivity, and specificity for the testing set (12 subjects).

Classifier	Training accuracy (fourfold cross-validation)	Testing		
		Accuracy	Sensitivity	Specificity
DFCN (proposed)	<b>95%</b>	<b>92%</b>	<b>83%</b>	<b>100%</b>
K-Star (K*)	93%	89%	89%	89%
K-Nearest-Neighbor (K-NN)	91%	84%	84%	83%
Random forest	85%	82%	82%	82%
Random tree	83%	81%	81%	81%

Bold indicates the better results from our approach compared with the other approach.

extracted each over an own layer showing statistically significant differences of values between normal and diabetic subjects. Cumulative distribution functions of these features are analyzed by a two-stage trainable deep fusion classification network (DFCN) with stacks of non-negativity-constrained autoencoders (SNCAE) to determine whether the subject has the DR or not. Promising initial experimental results suggest that in the future, this CAD system could amplify or even replace current technologies of early DR detection.

## CONFLICTS OF INTEREST

The authors declare there are no relevant conflicts of interest.

\*Equivalent contribution.

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: aselba01@louisville.edu; Website: <https://louisville.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/elbazlab>.

## REFERENCES

- Karumanchi DK, Gaillard ER, Dillon J. Early diagnosis of diabetes through the eye. *Photochem Photobiol.* 2015;91:1497–1504.
- Exalto LG, Biessels GJ, Karter AJ, Huang ES, Quesenberry CP Jr, Whitmer RA. Severe diabetic retinal disease and dementia risk in type 2 diabetes. *J. Alzheimer's Dis.* 2014;42:S109–S117.
- Mizutani A, Muramatsu C, Hatanaka Y, Suemori S, Hara T, Fujita H. Automated microaneurysm detection method based on double ring filter in retinal fundus images. *Proc of SPIE.* 2009;7260:72601N.
- Jaafar HF, Nandi AK, Al-Nuaimy W. Automated detection of exudates in retinal images using a split-and-merge algorithm. In: *2010 18th European Signal Processing Conference.* Aalborg, Denmark: IEEE; 2010:1622–1626.
- Pachiyappan A, Das UN, Murthy TV, Tatavarti R. Automated diagnosis of diabetic retinopathy and glaucoma using fundus and OCT images. *Lipids Health Dis.* 2012;11:73.
- Tan JH, Acharya UR, Chua KC, Cheng C, Laude A. Automated extraction of retinal vasculature. *Med Phys.* 2016;43:2311–2322.
- Rossant F, Ghorbel I, Bloch I, Paques M, Tick S. Automated segmentation of retinal layers in OCT imaging and derived ophthalmic measures. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009 (ISBI'09).* Boston, MA: IEEE; 2009:1370–1373.
- Yazdanpanah A, Hamarneh G, Smith BR, Sarunic MV. Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *IEEE Trans Med Imag* 2011;30:484–496.
- Kafieh R, Rabbani H, Abramoff MD, Sonka M. Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. *Med Image Anal.* 2013;17:907–928.
- Ehnes A, Wenner Y, Friedburg C, et al. Optical coherence tomography (OCT) device independent intraretinal layer segmentation. *Transl Vis Sci Technol.* 2014;3:1.
- Chen Q, Niu S, Yuan S, Fan W, Liu Q. Choroidal vasculature characteristics based choroid segmentation for enhanced depth imaging optical coherence tomography images. *Med Phys.* 2016;43:1649–1661.
- Nam HS, Kim C-S, Lee JJ, Song JW, Kim JW, Yoo H. Automated detection of vessel lumen and stent struts in intravascular optical coherence tomography to evaluate stent apposition and neointimal coverage. *Med Phys.* 2016;43:1662–1675.
- Kiernan DF, Zelkha R, Hariprasad SM, Lim JJ, Blair MP, Mieler WF. En face spectral-domain optical coherence tomography outer retinal analysis and relation to visual acuity. *Retina* 2012;32:1077–1086.
- Tzekov R. Full-field ERG in diabetic retinopathy: a screening tool? *Graefes Arch Clin Exp Ophthalmol.* 2015;253:987–988.
- Lim J, Yang M-H. A direct method for modeling non-rigid motion with thin plate spline. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005 (CVPR 2005).* Vol. 1, San Diego, CA: IEEE; 2005:1196–1202.
- Lega E, Scholl H, Alimi J-M, Bijaoui A, Bury P. A parallel algorithm for structure detection based on wavelet and segmentation analysis. *Parallel Comput.* 1995;21:265–285.
- El-Baz A, Elnakib A, Khalifa F, et al. Precise segmentation of 3-D magnetic resonance angiography. *IEEE Trans Biomed Eng.* 2012;59:2019–2029.
- Alansary A, Ismail M, Soliman A, et al. Infant brain extraction in T1-weighted MR images using BET and refinement using LCDG and MGRF models. *Biomed Health Inform* 2015;20:925–935.
- Khalifa F, Beache GM, Gimel'farb G, Giridharan GA, El-Baz A. Accurate automatic analysis of cardiac cine images. *IEEE Trans Biomed Eng.* 2012;59:445–455.
- Bengio Y, Lamblin P, Popovici D, Larochelle H, et al. Greedy layer-wise training of deep networks. *Adv Neural Inf Process Syst.* 2007;19:153.
- Hosseini-Asl E, Zurada JM, Nasraoui O. Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints. *IEEE Trans Neural Netw Learn Syst.* 2016; 27:2486–2498. [Online]. Available: <http://ieeexplore.ieee.org/document/7310882>
- Chiu SJ, Li XT, Nicholas P, Toth CA, Izatt JA, Farsiu S. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Optics Express.* 2010;18:428.
- Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945;26:297–302.
- Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol.* 2008;61:29–48.
- Witten IH, Frank E, Hall MA. *Data Mining: Practical Machine Learning Tools and Techniques.* Burlington: Kaufmann; 2011.

Copyright of Medical Physics is the property of American Association of Physicists in Medicine and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.