

Received January 7, 2020, accepted February 9, 2020, date of current version February 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974158

# Accurate Detection of Non-Proliferative Diabetic Retinopathy in Optical Coherence Tomography Images Using Convolutional Neural Networks

MOHAMMED GHAZAL<sup>1</sup>, (Senior Member, IEEE), SAMR SAMIR ALI<sup>1</sup>, (Member, IEEE),  
ALI H. MAHMOUD<sup>2</sup>, (Member, IEEE), AHMED M. SHALABY<sup>2</sup>, (Member, IEEE),  
AND AYMAN EL-BAZ<sup>2</sup>, (Senior Member, IEEE)

<sup>1</sup>Electrical and Computer Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates

<sup>2</sup>Bioengineering Department, University of Louisville, Louisville, KY 40208, USA

Corresponding author: Mohammed Ghazal (mohammed.ghazal@adu.ac.ae)

This work was supported in part by the Abu Dhabi Award for Research Excellence (AARE) 2019.

**ABSTRACT** Diabetic retinopathy (DR) is a disease that forms as a complication of diabetes. It is particularly dangerous since it often goes unnoticed and can lead to blindness if not detected early. Despite the clear importance and urgency of such an illness, there is no precise system for the early detection of DR so far. Fortunately, such system could be achieved using deep learning including convolutional neural networks (CNNs), which gained momentum in the field of medical imaging due to its capability of being effectively integrated into various systems in a manner that significantly improves the performance. This paper proposes a computer aided diagnostic (CAD) system for the early detection of non-proliferative DR (NPDR) using CNNs. The proposed system is developed for the optical coherence tomography (OCT) imaging modality. Throughout this paper, all aspects of deployment of the proposed system are studied starting from the preprocessing stage required to extract input retina patches to train the CNN without resizing the image, to the use of transfer learning principals and how to effectively combine features in order to optimize performance. This is done through investigating several scenarios for the system setup and then selecting the best one, which from the results revealed to be a two pre-trained CNNs based system, in which one of these CNNs is independently fed by nasal retina patches and the other one by temporal retina patches. The proposed transfer learning based CAD system achieves a promising accuracy of 94%.

**INDEX TERMS** Convolutional neural network (CNN), diabetic retinopathy (DR), optical coherence tomography (OCT).

## I. INTRODUCTION

Ophthalmologists today are capable of leveraging computer assisted diagnostic (CAD) systems to inform their opinions, in contrast to the traditional methods of visual interpretation and observation. CAD systems are still a new technology in the field of medicine, and there is a continuous flux of interest in the development of such systems due to their capability of improving the medical services provided to the community in terms of accuracy and reliability in the diagnosis of diseases. Meanwhile, machine learning is paving the way for breakthroughs in the different areas of medical imaging such as in classification [1], segmentation [2], disease

detection [3], and image registration [4]. The application of deep learning [5]–[8] a subset of machine learning algorithms, has made tremendous impact in the area of medical image processing research [9], [10]. Deep learning is the leading machine learning paradigm in computer vision and image processing domains, particularly as regards convolutional neural networks (CNNs) [11]. CNN are especially powerful in solving problems that are computationally difficult or with a high error rate such as medical image recognition with outstanding performance results [12]. In this context, we determined to use CNNs for the early detection of one of the most serious ophthalmological concerns, which is diabetic retinopathy (DR).

Blindness resulting from diabetes, which is becoming an increasingly alarming issue, is a consequence of the

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi<sup>1</sup>.

associated eye disease: DR. Such disease which develops as a complication of diabetes, particularly type II [13], [14], occurs specifically from the chronic high levels of sugar in the blood associated with swelling and damage of the tiny retinal blood vessels in the eye [15]–[17]. This leads to distortion of the vision followed by scarring of the retina in advanced stages, and finally consequent blindness [16]. It is worth mentioning that DR is one of the leading causes of blindness in adults [18], [19]. This problem is further exacerbated by the fact that 75% of the diabetic patients are not aware of the eye complications that they may be experiencing [13]. To prevent such a ramification, it is paramount to diagnose DR as soon as possible. Early detection and intervention can slow down the process and halt it completely [20], [21], which in turn protects the vision of the patient [22]. However, despite the significance of this matter and the notable rise in prevalence of diabetes, a precise procedure to detect early retinal changes for DR prevention is absent [17], [21].

DR can be broadly classified as proliferative (PDR) or non-proliferative (NPDR) [10]. NPDR is characterized by the presence of damaged blood vessels in the retina in addition to fluid leakage, which results in the retina swelling and wetness. In the case of PDR, multiple regions of the retina are affected by the appearance of new abnormal blood vessels, which makes it a severe and advanced DR stage. The work presented in this paper is limited only to NPDR.

One of the ophthalmic imaging modalities used for early DR detection is funduscopy [23], [24]. This modality is an area of active research for the development of CAD systems, because while funduscopy is well understood in terms of its imaging principles, interpretation requires a highly trained ophthalmologist; hence it is expensive [25]. Related work on fundus imaging for early detection of DR includes a trainable system for automated micro-aneurysm detection [26], performing with 65% sensitivity and averaging 27 false positives per image by supposition testing. Another system presented in [27] is able to detect hard and soft exudates based on combining fine and coarse segmentation, however it sometimes does not discriminate between exudates and non-exudate regions if their features are similar. On the other hand, in [28], Pachiyappan *et al.* use a combination of filtering, morphological processing, and thresholding for DR macular abnormalities detection, while in [29], automatic extraction of retinal vasculature was performed in order to obtain the blood vessels network. Similar feature-based algorithms for fundus image interpretation have also been reviewed in [30]. Recently, CNNs have also been used for exudate detection in fundus images from diabetic patients [31]. For the optical coherence tomography angiography (OCTA), Eladawi *et al.* [32] proposed a CAD system for the early detection of DR based on vascular segmentation of various layers of the retina using spatial statistical modeling.

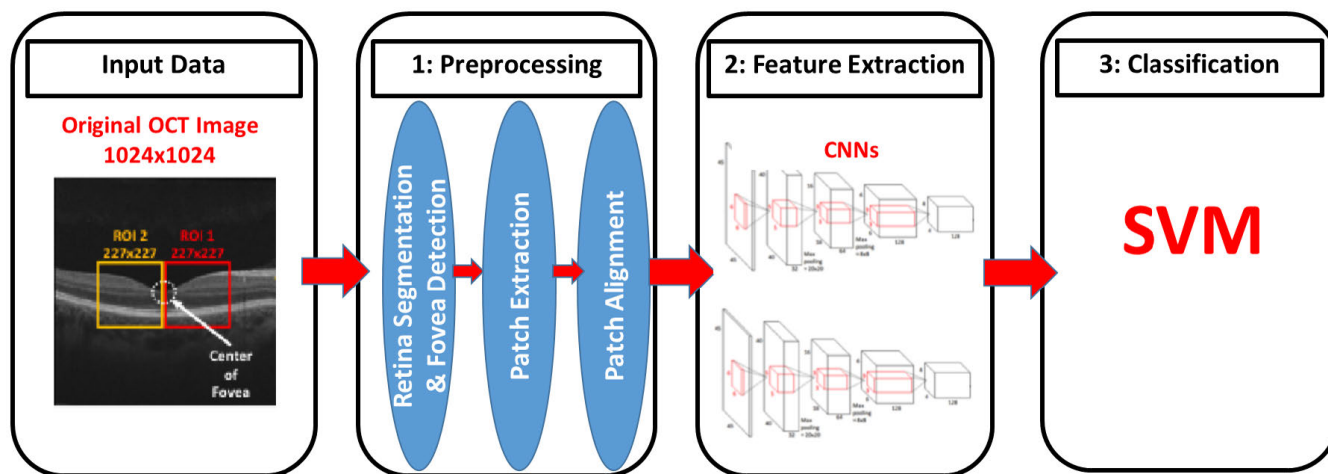
Nonetheless, another medical imaging modality that may be employed in the early detection of DR is optical coherence tomography (OCT) [33], which is useful because it facilitates

retinal morphology evaluation to microscopic resolution [34]. In turn, various retinal abnormalities including glaucoma, macular degeneration, and diabetic macular edema may be diagnosed in a non-invasive manner. Compared to fundus imaging, OCT is more favorable because it supports quantitative evaluations as it capable of capturing depth, in addition to its lower cost, and its ability to allow human bias free monitoring of changes [35]. However, OCT is relatively unexplored in comparison to fundus images in terms of early detection of DR. Related to this, Roychowdhury *et al.* [36] presented an OCT automatic system for localizing cysts with diabetic macular edema (DME) using 6 layers of the retina. Correia *et al.* [37] used Monte Carlo simulation to find the changes in the OCT images of DME patients from the cellular perspective. Trabelsi *et al.* [38] presented a technique to detect cystoids macular edema from OCT images.

This paper investigates the early detection of DR in OCT images, which is principally performed using CNNs. The proposed system has an optimal CNN architecture for early detection of DR through the exploration of various CNN configurations and parameters. In contrast to conventional feature extraction methods, using CNNs effectively classifies normal and DR images without the need for features that are extracted manually. The system starts with roughly segmenting the 12 layers of the retina and localizing the fovea using unsupervised learning [39]. Patches are then extracted from both the nasal and temporal sides of the fovea. These patches which are extracted from different subjects are aligned in both x and y directions, where the x-direction alignment is dependent on the location of the fovea, while the y-direction alignment is performed based on the position of a certain retina layer. Different retina layers have been investigated to be used for the y-direction alignment. The aligned nasal patches and the aligned temporal patches are then used for training CNNs, from which features are extracted and then fused using support vector machine (SVM) that is responsible for giving the final classification (DR vs normal). Throughout the work, different scenarios for training the CNNs have been tested, among which transfer learning is used. The following items are studied: 1) the effect of transfer learning on improving the performance of the proposed CAD system, given the scarcity of the data; 2) the effect of fusing CNNs retrained with different datasets on the overall system performance; 3) the depth of CNN layers required to extract features to train the final classifier used for data fusion; and 4) the OCT layer required to be segmented in order to be used for the y-coordinate axis alignment of the extracted patches for optimum results. The rest of this paper is organized in three sections as follows: a section that presents the materials and methods, followed by a section that discusses the experimental results, and finally the conclusion is given in the last section.

## II. MATERIALS AND METHODS

A simplified block diagram of the proposed CAD system for early detection of DR in OCT images using CNNs is



**FIGURE 1.** Simplified block diagram of the proposed CAD system for early detection of DR in OCT images using CNNs.

shown in Fig. 1. The proposed system is composed of (1) a preprocessing stage which includes rough segmentation of retina layers, and fovea detection, in addition to patch extraction and alignment; (2) a CNN-based feature extraction stage; and (3) a classification stage. The details of these stages as well as the used dataset and the used validation technique are given below.

#### A. OCT DATASET

Patients with and without DR were enrolled at the Kentucky Lions Eye Center at University of Louisville between June 2015 and December 2015 (University of Louisville IRB protocol 18.0010). Informed consent (or assent) was provided by each participant. Exclusion criteria included history of retinal pathology, including diabetes-related, and severe myopia, defined as refractive error  $\leq -6.0$  diopters. In all, 52 subjects were enrolled, 26 of whom had DR, ranging in age from 40 to 79 years.

Data used for training and testing of the CAD system were obtained using a clinical OCT scanner, Cirrus HD-OCT 5000 (Carl Zeiss Meditec, Dublin, California). B-scans were obtained over a 21-line raster across the macula of both eyes. For each eye, a single B-scan, passing through the fovea, was selected for analysis. Images were  $1024 \times 1024$  pixels, 8 bit grayscale, capturing an optical slice 2 mm deep and 9 mm from side to side (nasal-temporal).

#### B. PREPROCESSING

The preprocessing stage is illustrated in Fig. 2, where we extract the input patches to be fed to the CNN as appropriate. This starts with rough segmentation of the retina proper from the rest of the image, and identification of the fovea. The results of the prior two processes are then used for the positioning and extraction of the appropriate patches as per the schema of the proposed system. The segmentation of the original OCT scan into twelve different layers is performed by the application of an unsupervised parametric mixture model

and Markov Gibbs Random Fields [39]. This is inspired from the appearance of the retina in an OCT B-scan, which reveals approximately 12 bands of greater or lesser reflectivity as shown in Fig. 3. Histological studies have correlated these bands with the layers of the retina, proceeding from the vitreous body to the choroid: 1. nerve fiber layer (NFL), 2. ganglion cell layer (GCL), 3. inner plexiform layer (IPL), 4. inner nuclear layer (INL), 5. outer plexiform layer (OPL), 6. outer nuclear layer (ONL), 7. external limiting membrane (ELM), 8. myoid zone (MZ), 9. ellipsoid zone (EZ), 10. outer segments of the photoreceptors (OPR), 11. interdigitation zone (IZ), and 12. the retinal pigment epithelium (RPE).

As shown in the OCT image in Fig. 3, the retina has 12 layers, with the fovea in the middle of the image. The thickness of these 12 layers are not constant all over the image. So we can find that at the fovea, the vitreous body is nearly adjacent to the ONL (layer 6), while layers 1–5 almost vanish. Also, layers 1–5 are thickest in the foveal rim, which is surrounding the fovea. This structure is common among retinas of different subjects, so it could be used to roughly detect the fovea location and guide the patch extraction procedure for consistent representation of the retina across different OCT images. Patches were extracted from both sides of the fovea (the temporal and nasal sides) and oriented to align with the retinal layers. Consequently, the amount of extracted background (vitreous or choroid) is minimized, and feature extraction should be independent of any peculiarities (e.g. slight tilt or off-center) of a given OCT scan.

The fovea coordinate localization starts off with applying a median filter in order to remove any impulsive noise. This is then followed by the “à trous” algorithm [40] that decomposes each scan into scale-space components of coarser and finer detail by undecimated wavelet transform as per the aforementioned details. Edge detection in scale space allows for easy identification of high contrast boundaries in the OCT image: vitreous-NFL, MZ-EZ, and RPE-choroid. Contours are first detected as local gradient maxima in the appropriate wavelet component, then smoothed using adaptive

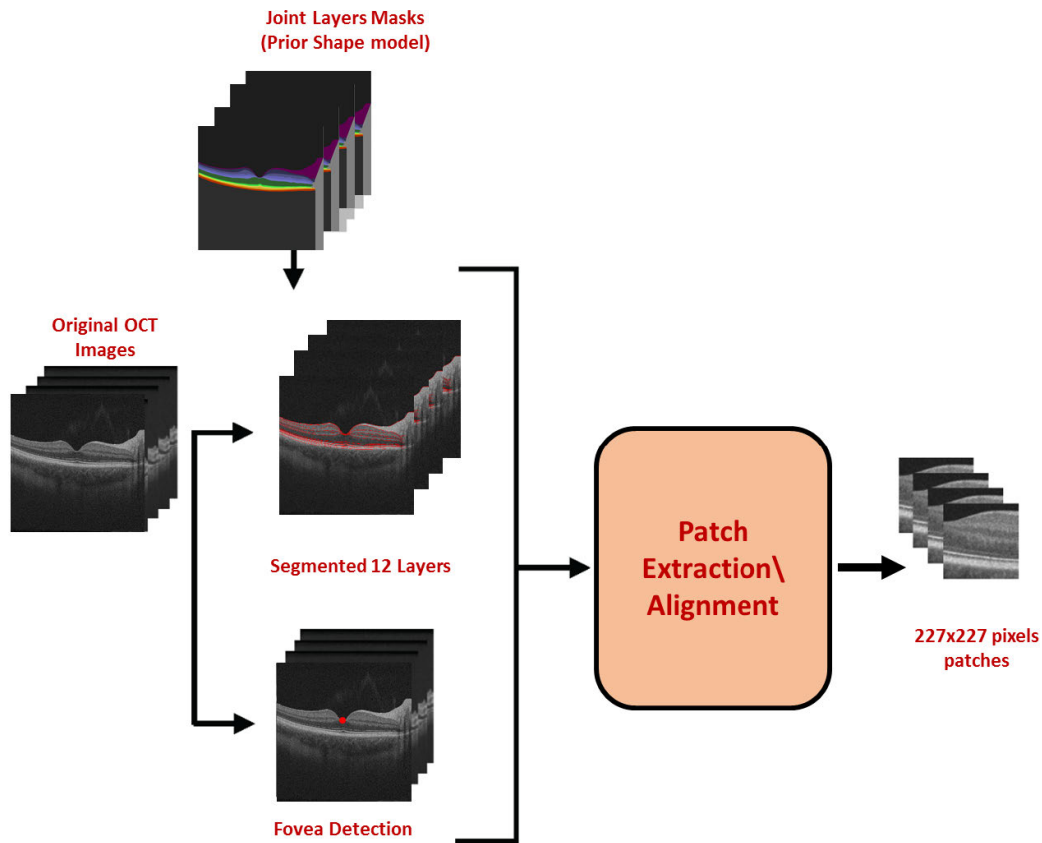


FIGURE 2. Preprocessing stage of the proposed CAD system.

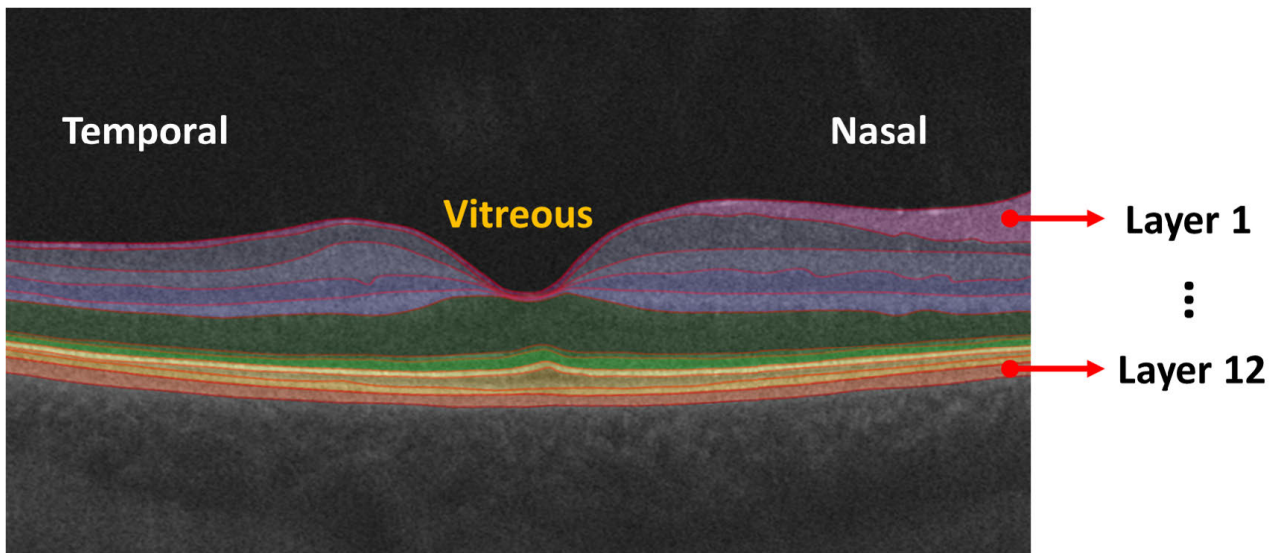


FIGURE 3. OCT retinal layers.

spline smoothing. Considerations of typical retina structure, above, lead to identification of the fovea with the point on the vitreous-NFL boundary at minimum distance from the MZ-EZ boundary. When computing these distances, it is important to correct for the non-square pixel aspect ratio of

typical OCT scanners. The preprocessing algorithm is based on the work presented in [39].

Upon the detection of the fovea, the required patches' locations along the x-coordinate axis are computed appropriately, i.e. with the origin at the fovea. These calculated

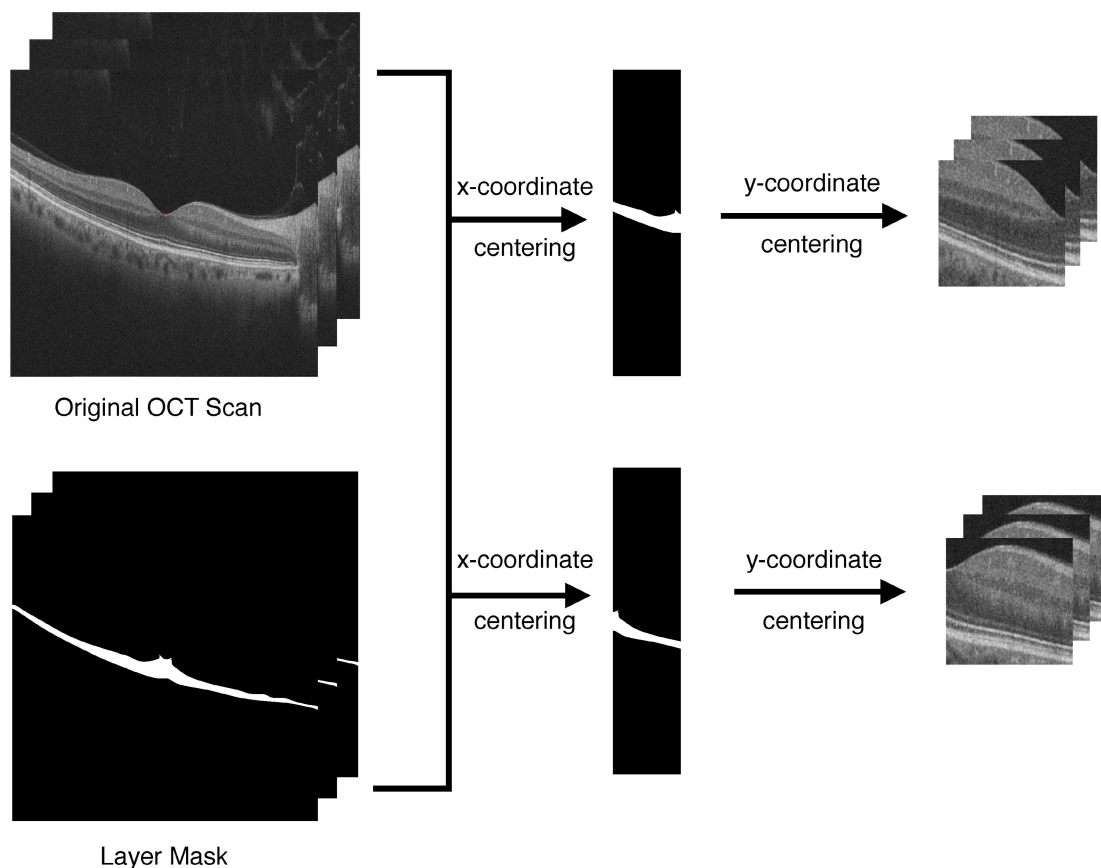


FIGURE 4. Patch extraction framework.

points are used for extracting the corresponding vertical slice from the segmentation mask of the layer that we would be centering the patch extraction at for the y-coordinate axis. The sum of the values of the pixels of these extracted slices along the x-axis is then considered to take into account any orientation or skewness display of the OCT scan. It acts as an efficient measure where if it is greater than zero then the corresponding row contains a significant part of the layer. The resultant of the final algorithm of the preprocessing stage as shown in Fig. 4 is a representative patch for each of the temporal and nasal sides, i.e. extracted image from the original OCT, that is in a matching size to the input layer of the CNN. Similar to the nasal and temporal patch extraction procedure discussed above, the preprocessing module could be tuned to generate center patches with the fovea in the middle as well as nasal and temporal patches that are distal from the fovea. These patches will be used in addition to the nasal and temporal patches to investigate several scenarios for the system setup, in order to select the best one.

The preprocessing stages inherently compensates for the mismatch between the dimensional size of the original OCT scans that are the input image data and that of the input layer of pre-trained CNNs, which will be used for feature extraction as discussed below. The preprocessing stage also eliminates

unimportant information; hence, improving the speed and efficiency of the system.

### C. FEATURE EXTRACTION AND CLASSIFICATION

Generally, the patches provided at the output of the preprocessing stage are fed into a CNN/CNNs from which features are extracted from certain layers. The involved CNN could be pre-trained i.e. based on transfer learning or not. Throughout the work, the AlexNet CNN shown in Fig. 5 is used with an input patch of size  $227 \times 227 \times 3$ . The CNNs with no transfer learning are randomly initialized.

In order to find out the optimum parameters for the proposed system, various scenarios were investigated throughout this work to determine: 1) whether the application of transfer learning improves the accuracy of the algorithm rather than just training the network from scratch, taking into consideration the fact that the dataset size is relatively small; 2) the patches to be used for best performance (distal nasal patches, nasal patches, central patches, temporal patches and/or distal temporal patches); 3) the OCT layer to be used for the y-direction alignment; and 4) The CNN layer to be used for feature extraction. A generic framework that illustrates various scenarios is given in Fig. 6. The figure has 7 CNNs, where the features extracted from each CNNi is stored in a

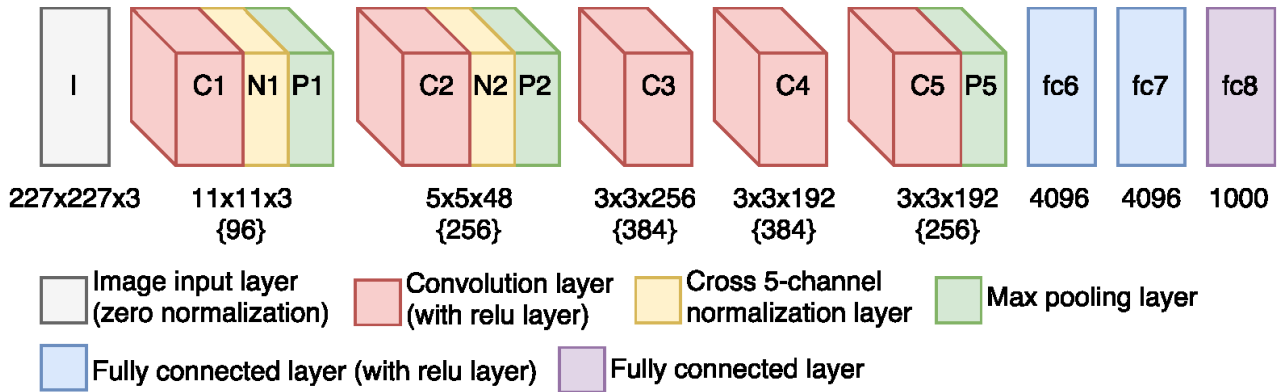


FIGURE 5. AlexNet CNN model architecture.

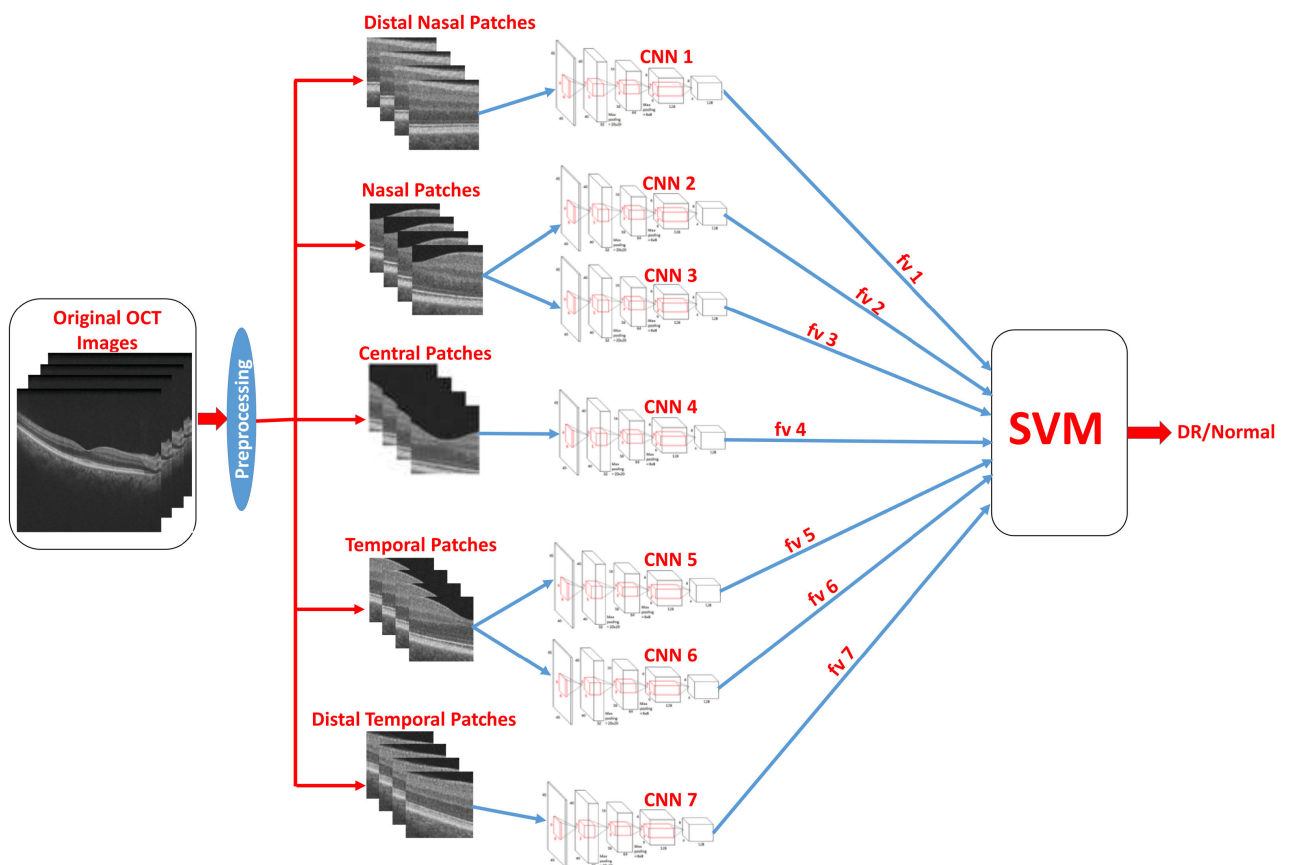


FIGURE 6. Feature extraction and classification generic framework used to investigate several scenarios. Patches provided by the preprocessing stage are fed into Convolutional Neural Networks (CNNs), from which features are extracted and then fed into a Support Vector Machine (SVM) to perform classification. The investigated CNNs are 1) CNN1: Distal Nasal patch pre-trained CNN, 2) CNN2: Nasal patch pre-trained CNN, 3) CNN3: Nasal patch CNN (no transfer learning), 4) CNN4: Central patch pre-trained CNN, 5) CNN5: Temporal patch pre-trained CNN, 6) CNN6: Temporal patch CNN (no transfer learning) and 7) CNN7: Distal Temporal patch pre-trained CNN. fv1, fv2, fv3, fv4, fv5, fv6 and fv7 are the extracted feature vectors. Within a specific scenario, the 7 CNN branches shown in the figure are not used altogether, but 1 or more branches are selected, however all branches are included in the figure for illustration purpose only.

corresponding feature vector  $fv_i$ . Each scenario will investigate the use of 1 or more CNNs out of these 7 CNNs but all the branches shown in the figure will not be used at the same time. They are included in a single figure for illustration purpose only. The extracted feature vector(s) is/are fed into a linear

SVM, a machine learning architecture used in classification problems [41], with a fast stochastic gradient descent solver for the final classification. If more than 1 CNN is involved in the scenario, the corresponding feature vectors are fused at the beginning of the SVM module. This is carried out by

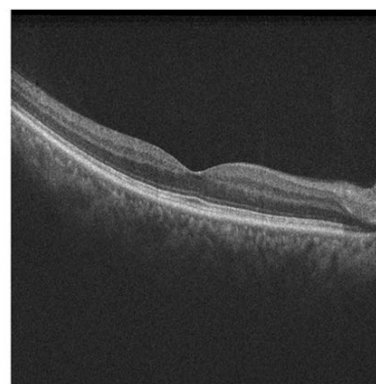
concatenating the features that are extracted by default from the bottleneck stages of the CNNs. The CNNs involved in various scenarios are 1) CNN1: Distal Nasal patch pre-trained CNN, 2) CNN2: Nasal patch pre-trained CNN, 3) CNN3: Nasal patch CNN (no transfer learning), 4) CNN4: Central patch pre-trained CNN, 5) CNN5: Temporal patch pre-trained CNN, 6) CNN6: Temporal patch CNN (no transfer learning) and 7) CNN7: Distal Temporal patch pre-trained CNN.

Scenario 1 and Scenario 2 are concerned only with the nasal patches and are used to investigate whether transfer learning enhances the performance or not. In Scenario 1, CNN2 is used alone, while in Scenario 2, CNN3 is used alone. The training of CNN2 and CNN3 is done using the the normal/DR dataset described above after being preprocessed. The pre-training of the CNN to be used for transfer learning was performed on a subset of the large-scale ImageNet database [42], containing 1.2 million real-life images and 1000 object categories.

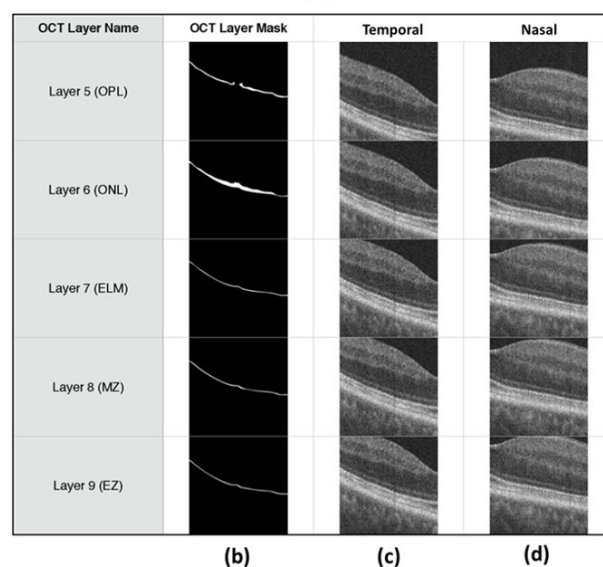
The pre-training is used only as far as layer *pool5* (Fig. 5) for the activations of the hidden layers, while subsequent layers of the AlexNet are treated identically for the comparison between the use and non-use of transfer learning. Similarly, in Scenario 3 and Scenario 4, patches that are extracted from the temporal side of the fovea are likewise input to two AlexNet CNNs, one of which was pre-trained (CNN5 in Scenario 3), and the other was not (CNN6 in Scenario 4), for the purpose of carrying out the same comparison. Note that each grayscale patch is input to all three different channels of the AlexNet, which was designed to operate on color images. Scenario 5 investigates using only the pre-trained CNN4 which is fed by central patches with the fovea in the center.

It was also tested whether improved accuracy would result from fusion of various input data, as in Scenario 6 and Scenario 7, where fusion of features extracted from pre-trained CNNs is involved. It is worth mentioning that our experiment is hypothesized to show that transfer learning does improve the results particularly with small datasets. In Scenario 6, both nasal and temporal patches are used with the pre-trained CNN2 and CNN5 respectively. Scenario 5 and Scenario 6 incorporate information from both the nasal and temporal sides of the fovea. However, information from both sides are used independently in Scenario 6 as patches from each side is fed into an independent CNN, while in Scenario 5, the information from both sides are used with a single CNN (CNN4). Scenario 7 investigates using 4 type of patches to train 4 pre-trained CNNs independently and then feed the corresponding extracted features into the SVM module to perform classification. The patches are distal nasal patches, nasal patches, temporal patches and distal nasal patches that correspond to CNN1, CNN2, CNN5 and CNN7 respectively. The distal nasal and temporal patches are extracted farther away from the fovea compared to the nasal and temporal patches respectively.

For the scenario with the best performance, an investigation is carried out in order to find out the optimum parameters



(a)



(b)

(c)

(d)

**FIGURE 7.** DR sample patches centered with different OCT layers. (a): Original OCT scan from a subject with DR. (b): OCT scan corresponding joint layers mask for segmentation. (c) and (d): Patches extracted to the temporal and nasal side of the foveal pit respectively from a subject with DR centered across the y-coordinate axis with different OCT layers.

for the best performance regarding the layer at which we carry out transfer learning. This is done by varying the CNN layer used for extracting the features. The following layers are used for comparison of the performance results: (i) one layer above the bottleneck features which is relu6 layer, {fc6}; (ii) the layer with the bottleneck features which is pool5 layer, {P5}; (iii) one layer before the bottleneck features which is relu5 layer, {C5}; and (iv) two layers before the bottleneck features which is relu4 layer, {C4}. In this investigation, both the results in terms of accuracy and the computation expense for an overall performance evaluation are taken into consideration.

Furthermore, for each of the previous scenarios, the default preprocessing pipeline for each of the OCT scans includes fovea detection. This is required to extract patches at standardized locations relative to the center of the fovea. This preprocessing step uses the ONL for y-direction alignment. However, it is noteworthy to mention that different experiments were carried out for the scenario with the best

performance, where all other parameters are fixed except for OCT layer used for the y-direction alignment, in order to find the layer that achieves the optimum results and hence further improves the algorithm. The layers that were considered in this sub-experiment are: (i) layer 5 or the OPL; (ii) layer 6 or the ONL; (iii) layer 7 or the ELM; (iv) layer 8 or the MZ; and (v) layer 9 or the EZ. Samples of such extracted patches are shown in Fig. 7. Scenario 8 (not included in Fig. 7) explores using resized whole retina images to directly train a pre-trained CNN, as opposed to using only patches from the retina in Scenarios 1 through 7.

#### D. VALIDATION

Five-fold cross validation is used for the evaluation of each of the investigations of the CNN CAD system for early detection of DR. This is a particular case of  $k$ -fold cross validation, where  $k = 5$  training runs are performed, each time leaving out a fraction  $\frac{1}{k}$  of the data for subsequent testing. In this way every available observation (OCT scan) is used for both training and (exactly once) for testing. This is in contrast to the traditional hold-out method where a fixed proportion of the data are set aside from the beginning of the experiment to be used only for validation. Cross validation has the advantage of making the most of a limited amount of data, but is known to produce biased estimates of system performance.

The performance metrics used are accuracy ( $\alpha$ ), error rate ( $\beta$ ), specificity ( $\chi$ ), precision (PPV), and recall (TPR). If TP is the number of correctly classified DR cases in a particular run of cross validation, TN is the number of correctly classified normal cases, and  $P = TP + FP$  and  $N = TN + FN$  are the total number of DR and normal cases in the test data, then these metrics are defined as:

$$\alpha = \frac{TP + TN}{P + N} \quad (1)$$

$$\beta = 1 - \alpha \quad (2)$$

$$\chi = \frac{TN}{FP + TN} \quad (3)$$

$$PPV = \frac{TP}{TP + FP} \quad (4)$$

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

### III. EXPERIMENTAL RESULTS AND DISCUSSION

Shown in Table 1 are the comparative results of the investigations carried out for various scenarios. In addition to the aforementioned performance metrics, the respective standard deviations across the folds are noted. The standard deviation of the error metric corresponds to that of the accuracy as they are complementary measures. First, we investigate the effect of transfer learning application on the proposed system by comparing Scenario 1 with Scenario 2 and Scenario 3 with Scenario 4. The input to the CNNs in the case of Scenario

1 and Scenario 2 is patches extracted only from the nasal side of the fovea. The CNN used in Scenario 1 was pre-trained while that used in Scenario 2 was not. Moreover, the input to the CNNs in the case of Scenario 3 and Scenario 4 is patches extracted only from the temporal side of the fovea. The CNN used in Scenario 3 was pre-trained while that used in Scenario 4 was not. Both comparisons show a clear increase in performance across all of the measures when using transfer learning. Second, we compared the performance results in the case of training the proposed system with a single patch with the fovea centered (Scenario 5) versus training the system with independent nasal and temporal patches (Scenario 6 and Scenario 7). Scenario 6 uses two independent patch types for training, which are nasal patches and temporal patches, while Scenario 7 uses two additional independent patch types, which are distal nasal patches and temporal nasal patches. It was found that using independent patch training to combine two or four patches results in the highest performance metrics across the board. However, it can also be observed that the training time taken for the four patches scenario is almost four times that of the two patches scenario, with no improvement in any of the metrics. As such, only the two-patch approach is accounted for in the next investigation of finding the optimum CNN layer to extract the features from. It can be derived from the summary in the table that the bottleneck features represent the best choice in terms of balance between the run time required, i.e. computational complexity, and the accuracy. Finally, in choosing the OCT layer to align the extraction of the input patches at, we find that layer 6 is the best given that it reaches the highest performance metrics. Although other layers achieve similar levels of accuracy, error rate, specificity, precision, and recall, they all require more time to train. As such, the final design choices for the proposed system is the CNN retrained with two patches, one on each side of the fovea, with the ONL centered vertically within each patch. The features are extracted from *pool5* CNN layer and transfer learning is applied in the deployment of the proposed system.

Additional testing was carried out in order to confirm that the CNN architecture is not biased due to color space, for the ImageNet dataset is RGB while our dataset is intrinsically grayscale. Hence, we retrained the network with 200 grayscale images and reapplied our investigation. The result was conclusive that the network is independent of the color space as the results obtained exactly matched that of directly using the ImageNet pre-trained CNN.

Significant degradation of performance resulted upon downsampling, as shown in Table 1 (Scenario 8). After training with downsampled images, accuracy was at most 78% (71% specificity, 78% precision, and 100% recall). Every metric, except for the recall, was lower compared to that of the CNN trained at full resolution. Based on this, resampling of input should be avoided, and extraction of patches as per the proposed methodology is recommended.

Moreover, the confusion matrix of the chosen CNN for the proposed early DR detection CAD system is shown in Fig. 8.



TABLE 1. Summary of proposed system findings.

	Accuracy	Accuracy Std	Error Rate	Specificity	Specificity Std	Precision	Precision Std	Recall	Recall Std	Testing Time (s)
Scenario 1: Nasal Alone (Transfer learning) (pool5 - layer6) - 32 batches	92.00%	0.08370	8.00%	89.33%	0.09830	93.78%	0.05700	97.78%	0.05640	0.15
Scenario 2: Nasal Alone (No transfer learning) (pool5 - layer6) - 32 batches	54.00%	0.05480	46.00%	42.86%	0.04120	66.67%	0.20410	84.29%	0.20450	0.48
Scenario 3: Temporal Alone (Transfer learning) (pool5 - layer6) - 32 batches	66.00%	0.08940	34.00%	66.72%	0.18840	73.00%	0.16430	91.00%	0.12450	0.15
Scenario 4: Temporal Alone (No transfer learning) (pool5 - layer6) - 32 batches	50.00%	0.00000	50.00%	30.00%	0.00000	76.79%	0.22520	66.79%	0.20660	0.47
Scenario 5: Center (Transfer learning) (pool5 - layer6) - 32 batches	80.00%	0.10950	20.00%	80.29%	0.13470	87.33%	0.08330	91.28%	0.15350	0.43
Scenario 6: Two Patches (Transfer learning) (pool5 - layer6) - 32 batches	<b>94.00%</b>	<b>0.05480</b>	<b>6.00%</b>	<b>88.00%</b>	<b>0.10950</b>	<b>90.00%</b>	<b>0.09130</b>	<b>100.00%</b>	<b>0.00000</b>	<b>0.42</b>
Scenario 7: Four Patches (Transfer learning) (pool5 - layer6) - 32 batches	94.00%	0.05480	6.00%	88.00%	0.10950	90.00%	0.09130	100.00%	0.00000	0.42
Scenario 6: Two Patches (Transfer learning) (relu6 - layer6) - 32 batches	84.00%	0.05480	16.00%	72.00%	0.17890	79.52%	0.12250	96.00%	0.08940	0.42
Scenario 6: Two Patches (Transfer learning) (relu5 - layer6) - 32 batches	94.00%	0.05480	6.00%	88.00%	0.10950	90.00%	0.09130	100.00%	0.00000	0.42
Scenario 6: Two Patches (Transfer learning) (relu4 - layer6) - 32 batches	92.00%	0.04470	8.00%	84.00%	0.08940	86.67%	0.07450	100.00%	0.00000	0.40
Scenario 6: Two Patches (Transfer learning) (pool5 - layer5) - 32 batches	88.00%	0.04470	12.00%	88.00%	0.10950	90.00%	0.09130	88.00%	0.17890	0.43
Scenario 6: Two Patches (Transfer learning) (pool5 - layer7) - 32 batches	94.00%	0.05480	6.00%	88.00%	0.10950	90.00%	0.09130	100.00%	0.00000	0.42
Scenario 6: Two Patches (Transfer learning) (pool5 - layer8) - 32 batches	92.00%	0.04470	8.00%	84.00%	0.08940	86.67%	0.07450	100.00%	0.00000	0.40
Scenario 6: Two Patches (Transfer learning) (pool5 - layer9) - 32 batches	94.00%	0.05480	6.00%	88.00%	0.10950	90.00%	0.09130	100.00%	0.00000	0.42
Scenario 8: Resized Image (Transfer learning) (pool5) - 32 batches	78.00%	0.13040	22.00%	71.23%	0.12400	78.00%	0.13040	100.00%	0.00000	0.40

	Diabetic Retinopathy	Normal	
Diabetic Retinopathy	True Positive (TP) = (26) 100%	False Negative (FN) = (0) 0%	Sensitivity or Recall = 100%
Normal	False Positive (FP) = (3) 12%	True Negative (TN) = (23) 88%	Specificity = 88%
	Precision = 90%	Negative Predictive Value = 88%	<b>Accuracy = 94%</b>

FIGURE 8. Confusion matrix for proposed computer aided design system for early DR detection with CNNs.

TABLE 2. Comparison of our proposed system with four other machine learning classifiers.

Classifier	Accuracy	Recall	Specificity
CNN+SVM (proposed)	94%	100%	88%
K-Star(*)	89%	89%	89%
K-Nearest Neighbor (kNN)	84%	84%	83%
Random Forest	82%	82%	82%
Random Tree	81%	81%	81%

Finally, a comparison of our proposed technique against other machine learning techniques shows its superiority as can be observed in Table 2. It is worth mentioning that we used the same methodology described in [35] to implement the comparison framework using Matlab-ready K-Star, K-Nearest Neighbor (kNN), Random Forest, and Random tree classifiers.

IV. CONCLUSION

Early intervention is essential to delay or prevent complications of DR, including blindness. As such, in this paper, a novel CAD system for early detection of DR-related

changes in OCT images using CNNs was presented. The system was developed for use with patients with almost clinically normal retina appearances.

Upon investigation of the various scenarios for the proposed CAD system, the optimal conditions for its deployment were found. Foremost, transfer learning should be used to achieve high accuracy given the scarcity of the data. Second, best results are seen when combining the output features of two independently trained CNNs, which operate on both sides of the fovea. Features extracted at the pool5 layers of these CNN provided for the highest accuracy with the least computational complexity. Finally, in order to reach highest accuracy, which our results found to be 94 %, the patches extracted for training and testing should be aligned along the y-axis using the patch extraction algorithm presented with the segmented OCT layer number 6, or the ONL. This paper recommends that further research is directed towards this relatively uncharted topic, especially with OCT images, for the results were observed to be high even given the scarcity of the data and the relative complexity of the problem.

## REFERENCES

- [1] D. V. Pazinato, B. V. Stein, W. R. de Almeida, R. de O. Werneck, P. R. M. Junior, O. A. B. Penatti, R. D. S. Torres, F. H. Menezes, and A. Rocha, "Pixel-level tissue classification for ultrasound images," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 1, pp. 256–267, Jan. 2016.
- [2] D. Marín, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Trans. Med. Imag.*, vol. 30, no. 1, pp. 146–158, Jan. 2011.
- [3] P. Mesejo, D. Pizarro, A. Abergel, O. Rouquette, S. Beorchia, L. Poincloux, and A. Bartoli, "Computer-aided classification of gastrointestinal lesions in regular colonoscopy," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2051–2063, Sep. 2016.
- [4] G. Wu, D. Shen, and M. Sabuncu, *Machine Learning and Medical Imaging*, 1st ed. Cambridge, MA, USA: Academic, 2016.
- [5] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
- [6] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.
- [7] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 215–232, Nov. 2016.
- [8] J. Han, X. Ji, X. Hu, D. Zhu, K. Li, X. Jiang, G. Cui, L. Guo, and T. Liu, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.
- [9] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006.
- [10] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [12] S. Suzuki, X. Zhang, N. Homma, K. Ichiji, N. Sugita, Y. Kawasumi, T. Ishibashi, and M. Yoshizawa, "Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis," in *Proc. 55th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, Sep. 2016, pp. 1382–1386.
- [13] M. E. H. D. Experts. (2016). *Serious Eye Complications Set to Rise as Diabetes Reaches Record Levels in the Middle East Region, say Experts at Moorfields eye Hospital Dubai*. [Online]. Available: <https://www.moorfields.ae/dubai/serious-eye-complications-set-to-rise-as-diabetes-reaches-record-levels-in-the-middle-east-region-say-experts-at-moorfields-eye-hospital-dubai/>
- [14] L. G. Exalto, G. J. Biessels, A. J. Karter, E. S. Huang, C. P. Quesenberry, and R. A. Whitmer, "Severe diabetic retinal disease and dementia risk in type 2 diabetes," *J. Alzheimer's Disease*, vol. 42, no. s3, pp. S109–S117, Sep. 2014.
- [15] O. Faust, R. Acharya, U. E. Y. K. Ng, K.-H. Ng, and J. S. Suri, "Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review," *J. Med. Syst.*, vol. 36, no. 1, pp. 145–157, Feb. 2012, doi: 10.1007/s10916-010-9454-7.
- [16] Native English Institute. (2016). *Facts About Diabetic eye Disease* National Eye Institute. [Online]. Available: <https://nei.nih.gov/health/diabetic/retinopathy>
- [17] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, and J. Grauslund, "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes Care*, vol. 35, no. 3, pp. 556–564, 2012. [Online]. Available: <http://care.diabetesjournals.org/content/35/3/556>
- [18] F. Al-Maskari and M. El-Sadig, "Prevalence of diabetic retinopathy in the United Arab Emirates: A cross-sectional survey," *BMC Ophthalmol.*, vol. 7, no. 1, p. 11, Dec. 2007, doi: 10.1186/1471-2415-7-11.
- [19] American Academy of Ophthalmology. (2013). *Diabetic Retinopathy—Middle East/North Africa*. [Online]. Available: <https://www.aao.org/topic-detail/diabetic-retinopathy-middle-east-north-africa>
- [20] U. M. Akram and S. A. Khan, "Automated detection of dark and bright lesions in retinal images for early detection of diabetic retinopathy," *J. Med. Syst.*, vol. 36, no. 5, pp. 3151–3162, Oct. 2012, doi: 10.1007/s10916-011-9802-2.
- [21] D. K. Karumanchi, E. R. Gaillard, and J. Dillon, "Early diagnosis of diabetes through the eye," *Photochem. Photobiol.*, vol. 91, no. 6, pp. 1497–1504, Nov. 2015.
- [22] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," *Pattern Recognit.*, vol. 46, no. 1, pp. 107–116, Jan. 2013.
- [23] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017.
- [24] N. Gurudath, M. Celenk, and H. B. Riley, "Machine learning identification of diabetic retinopathy from fundus images," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2014, pp. 1–7.
- [25] K. C. Jordan, M. Menolotto, N. M. Bolster, I. A. Livingstone, and M. E. Giardini, "A review of feature-based retinal image analysis," *Expert Rev. Ophthalmol.*, vol. 12, no. 3, pp. 207–220, May 2017, doi: 10.1080/17469899.2017.1307105.
- [26] Y. Hatanaka, T. Inoue, S. Okumura, C. Muramatsu, and H. Fujita, "Automated microaneurysm detection method based on double-ring filter and feature analysis in retinal fundus images," in *Proc. 25th IEEE Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2012, pp. 1–4.
- [27] H. F. Jaafar, A. K. Nandi, and W. Al-Nuaimy, "Automated detection of exudates in retinal images using a split-and-merge algorithm," in *Proc. 18th Eur. Signal Process. Conf.*, Aug. 2010, pp. 1622–1626.
- [28] A. Pachiyappan, U. N. Das, T. V. Murthy, and R. Tatavarti, "Automated diagnosis of diabetic retinopathy and glaucoma using fundus and OCT images," *Lipids Health DisLipids Health Disease*, vol. 11, no. 1, p. 73, 2012.
- [29] J. H. Tan, U. R. Acharya, K. C. Chua, C. Cheng, and A. Laude, "Automated extraction of retinal vasculature," *Med. Phys.*, vol. 43, no. 5, pp. 2311–2322, Apr. 2016.
- [30] M. S. B. M. R. and H. S. Sheshadri, "Feature extraction from the fundus images for the diagnosis of Diabetic Retinopathy," in *Proc. Int. Conf. Emerging Res. Electron., Comput. Sci. Technol. (ICERECT)*, Dec. 2015, pp. 240–245.
- [31] O. Perdomo, J. Arevalo, and F. A. González, "Convolutional network to detect exudates in eye fundus images of diabetic subjects," *Proc. SPIE*, vol. 10160, Jan. 2017, Art. no. 101600T, doi: 10.1117/12.2256939.
- [32] N. Eladawi, M. Elmogy, F. Khalifa, M. Ghazal, N. Ghazi, A. Abouelfetouh, A. Riad, H. Sandhu, S. Schaal, and A. El-Baz, "Early diabetic retinopathy diagnosis based on local retinal blood vessel analysis in optical coherence tomography angiography (OCTA) images," *Med. Phys.*, vol. 45, no. 10, pp. 4582–4599, Oct. 2018.
- [33] J. C. BuAbbud, M. M. Al-latayfeh, and J. K. Sun, "Optical coherence tomography imaging for diabetic retinopathy and macular edema," *Current Diabetes Rep.*, vol. 10, no. 4, pp. 264–269, Aug. 2010, doi: 10.1007/s11892-010-0129-z.
- [34] B. L. Sikorski, G. Malukiewicz, J. Stafiej, H. Lesiewska-Junk, and D. Raczynska, "The diagnostic function of OCT in diabetic maculopathy," *Med. Inf.*, vol. 2013, Nov. 2013, Art. no. 434560.
- [35] A. ElTanboly, M. Ismail, A. Shalaby, A. Switala, A. El-Baz, S. Schaal, G. Gimel'farb, and M. El-Azab, "A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images," *Med. Phys.*, vol. 44, no. 3, pp. 914–923, Mar. 2017, doi: 10.1002/mp.12071.
- [36] S. Roychowdhury, D. D. Koozekanani, S. Radwan, and K. K. Parhi, "Automated localization of cysts in diabetic macular edema using optical coherence tomography images," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 1426–1429.
- [37] A. Correia, L. Pinto, A. Araújo, M. Morgado, S. Barbeiro, F. Caramelo, P. Serranho, P. Menezes, and R. Bernardes, "Monte Carlo simulation of diabetic macular edema changes on optical coherence tomography," *Investigative Ophthalmol. Vis. Sci.*, vol. 55, no. 13, p. 4807, 2014.
- [38] H. Trabelsi, I. Zghal, and N. Slokom, "Segmentation of cyctoids macular edema in optical coherence tomography," in *Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Mar. 2016, pp. 303–306.
- [39] A. A. Farag, A. S. El-Baz, and G. Gimel'farb, "Precise segmentation of multimodal images," *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 952–968, Apr. 2006.
- [40] E. Lega, H. Scholl, J.-M. Alimi, A. Bijaoui, and P. Bury, "A parallel algorithm for structure detection based on wavelet and segmentation analysis," *Parallel Comput.*, vol. 21, no. 2, pp. 265–285, Feb. 1995.

- [41] X. Zhang, J. Tian, D. Xiang, X. Li, and K. Deng, "Interactive liver tumor segmentation from ct scans using support vector classification with watershed," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 6005–6008.
- [42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.



**MOHAMMED GHAZAL** (Senior Member, IEEE) received the B.Sc. degree in computer engineering from the American University of Sharjah, and the M.Sc. and Ph.D. degrees in electrical and computer engineering (ECE) from Concordia University, Montréal, QC, Canada, in 2006 and 2010, respectively. He led the B.Sc. programs in ECE to achieve ABET and WASC accreditation. He also leads ADU E-Learning initiative. He is currently an Associate Professor of ECE and the Chair of the Department with Abu Dhabi University (ADU). He has authored more than 50 publications in technical journals and international conferences, including Elsevier's Renewable and Sustainable Energy Reviews, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS. His current research interests are biomedical signal processing, mobile computing, and embedded systems. He is a member of the ACM and BMES. He was awarded the Alexander Graham Bell Fellowship by NSERC and the QFRNT by the Government of Quebec. He was awarded the President's Cup and the Ministry of Education Shield for Creative Thinking from the American University of Sharjah. During his time at ADU, he received multiple awards, including the Distinguished Faculty Award in 2016 and 2014, and the University Teaching Award in 2013. Under his leadership, the department has grown in the number of students, laboratories, faculty, and programs and is home to two bachelors, two minors, and two Master's programs serving more than 300 students. Under his supervision, ECE students have won major national and international awards and secured top places in competitions, including a Special Innovation Award of 1,000,000 from HH Sheikh Muhammad Bin Rashid Al Maktoum as part of the Best m-Government Service Award Competition during the Government Summit.



**SAMR SAMIR ALI** (Member, IEEE) received the B.Sc. degree (*Summa Cum Laude*) in computer engineering from Abu Dhabi University, United Arab Emirates, in 2015. She is currently pursuing the Ph.D. degree in electrical and computer engineering with Concordia University. She was also a Teaching and a Research Assistant with Abu Dhabi University, during her Master's study in electrical and computer engineering. She is also a Research Assistant with Concordia University. Her current research interests include machine learning, medical image processing, pattern recognition, and computer vision. She has served as the President of the IEEE student branch in Abu Dhabi University. She has led several teams towards winning top places in competitions, such as a Special Innovation Award from HH Sheikh Muhammad Bin Rashid Al Maktoum as a part of the Best mGovernment Service Award Competition during the 2015 Government Summit for the Al Najdah Mobile Application, the United Nations Economic and Social Commission for Western Asia best Digital Arabic Content Application Award for Iqraa Educational Platform, the Ministry of Interior Award, and the Mobile Application Contest.



**ALI H. MAHMOUD** (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Alexandria University, Egypt, in 2005 and 2009, respectively, and the Ph.D. degree in electrical engineering from the University of Louisville, Louisville, KY, USA, in 2014. He is currently working as a Postdoctoral Research Associate with the Bioengineering Department, University of Louisville. His research interests include computer vision and image processing, object detection and tracking, medical imaging, and facial biometrics.



**AHMED M. SHALABY** (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Alexandria University, Egypt, in 2003 and 2009, respectively, and the Ph.D. degree in electrical engineering from the University of Louisville, Louisville, KY, USA, in 2014. He is currently working as a Postdoctoral Research Associate with the Bioengineering Department, University of Louisville. His research interests include computer vision and image processing, object detection and tracking, medical imaging, and facial biometrics.



**AYMAN EL-BAZ** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering, in 1997 and 2001, respectively, and the Ph.D. degree in electrical engineering from the University of Louisville, Louisville, KY, USA, in 2006. He is currently a Distinguished Professor with University Scholar and the Chair of the Bioengineering Department, University of Louisville. He has 17 years of hands-on experience in the fields of bio-imaging modeling and non-invasive computer-assisted diagnosis systems. He has authored or coauthored more than 500 technical articles (145 journals, 27 books, 78 book chapters, 225 refereed-conference papers, 172 abstracts, and 30 U.S. patents, and Disclosures). In 2009, he was named a Coulter Fellow for his contributions to the field of biomedical translational research.

...