Biomedical Signal Processing and Control 54 (2019) 101605

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images

Vineeta Das*, Samarendra Dandapat, Prabin Kumar Bora

Electro Medical and Speech Technology Lab, Indian Institute of Technology Guwahati, India

ARTICLE INFO

Article history:
Received 19 February 2019
Received in revised form 12 June 2019
Accepted 10 July 2019
Available online 9 August 2019

Keywords:
Optical coherence tomography (OCT)
Multi-scale
Convolutional neural network (CNN)
Feature fusion
Classification
Macular pathology

ABSTRACT

Identification of the macular pathologies at an early stage can prevent vision loss. Similarity in the pathological manifestations of common macular disorders like age related macular degeneration (AMD) and diabetic macular edema (DME) can make manual screening fallible. There is a growing interest among researchers for reliable automated detection of macular pathologies using computer methods. Therefore, in this paper we present a novel method for classification of DME and two stages of AMD namely the drusens (early stage) and the choroidal neo vascularization (CNV) (late stage) from healthy optical coherence tomography (OCT) images. The proposed method introduces a multi-scale deep feature fusion (MDFF) based classification approach using convolutional neural network (CNN) for reliable diagnosis. The MDFF captures the inter-scale variations in images to introduce discriminative and complementary information to the classifier. The proposed method is evaluated on an OCT dataset containing 84,484 images with different class distributions. The imbalance in the dataset is handled by introducing the cost sensitive loss function during the learning of the classifier. The proposed method achieves an average sensitivity, specificity and accuracy of 99.6%, 99.87% and 99.6% on the test set. The promising classification results make the proposed method highly suitable for preliminary automated diagnosis of macular pathologies in health care centres and eye clinics.

1. Introduction

Macular pathologies like age related macular degeneration (AMD) and diabetic macular edema (DME) are highly prevalent retinal disorders affecting nearly 8% of the world's population [1,2]. AMD is the most common cause of blindness in the elderly and labelled a 'priority eye disease' by the WHO [3]. DME on the other hand is a complication of diabetes and the primary cause of vision impairment and blindness in patients with diabetes [4]. It can cause substantial vision loss if left untreated for a year or longer [5].

The detection of these diseases at an early stage is important mainly because of the following reasons. (i) These diseases are progressive in nature leading to visual impairments like blurred vision and blank spots at the advanced stages. Success of treatment for these diseases is defined by preservation of remaining vision rather than improvement in vision [6]. (ii) The treatment methods for late AMD and DME include frequent administration of intravitreal injections of anti- *vascular endothelial growth factor* (VEGF) agents and laser photo-coagulation [5]. These treatments are associated with substantial financial cost and frequent visits to the ophthalmologist for follow up. Therefore, the diagnosis of these diseases at an early stage can preserve vision and prevent psychological and financial distress. To this end, we design an automated algorithm for identification of early and late stages of AMD and DME from healthy control subjects.

1.1. Clinical background

AMD at its early stages is characterized by the deposition of extracellular fluid underneath the retinal pigment epithelium (RPE). These deposits, called drusens, accumulate over time leading to the damage of the RPE and subsequent loss of the photo receptor cells [7]. Late stages of AMD are characterized by the growth of abnormal and leaky blood vessels (choroidal neo-vascularization (CNV)) and atrophy of the RPE and the photo receptor cells (geographic atrophy) [8]. In DME, chronic hyperglycemia increases the permeability of the blood vessels and causes increased angiogenesis [9]. This causes breakdown of the blood retinal barrier, swelling of the retina and accumulation of extracellular fluids disrupting the structure of the retinal layers [10].

* Corresponding author.
E-mail addresses: vineetadas@iitg.ernet.in (V. Das), samaren@iitg.ac.in (S. Dandapat).

https://doi.org/10.1016/j.bspc.2019.101605
1746-8094/© 2019 Elsevier Ltd. All rights reserved.

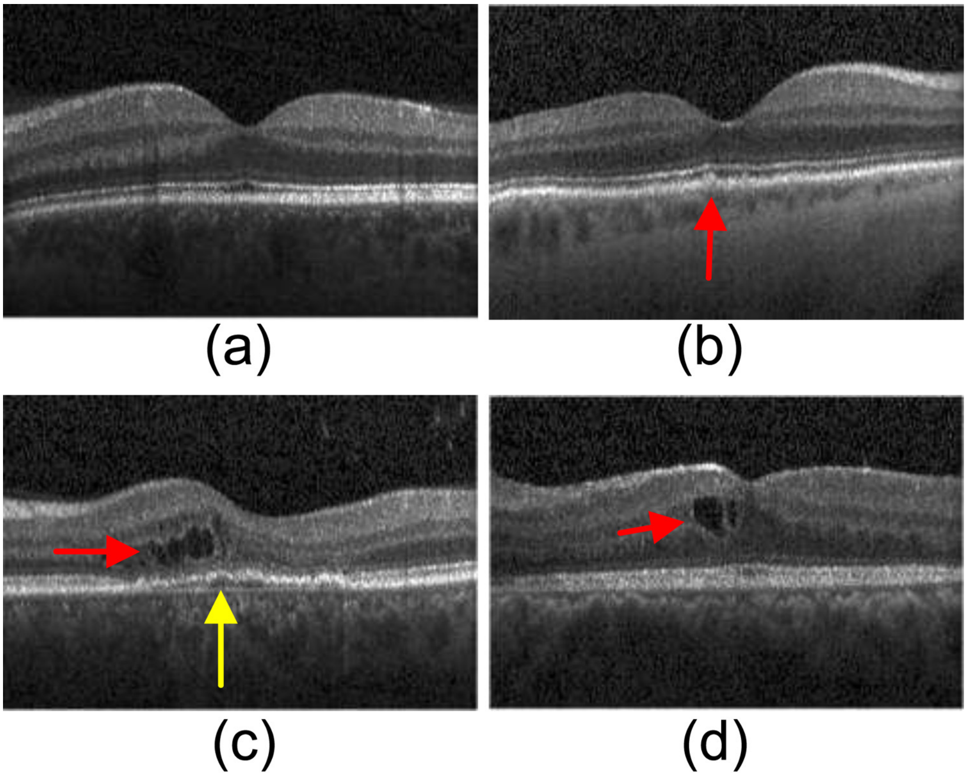

Fig. 1. (a) Normal OCT image, (b-c) Early (drusen) and later stages (CNV) of AMD and (d) DME affected image.

The pathological manifestations of AMD and DME can be visualized using the optical coherence tomography (OCT) images [11]. It provides a high-resolution cross-sectional and volumetric analysis of the tissue micro-structures of the retina [12]. It is highly preferred by the ophthalmologists to assess the health of the retina and identify retinal disorders because of its non-invasive nature and ease of image acquisition [13]. Fig. 1 shows the OCT B-scan for normal and pathological retina. Fig. 1(a) shows the OCT B-scan for a normal retinas. Fig. 1(b) shows an OCT scan at the early stage of AMD containing a drusen (red arrow in the image). Fig. 1(c) shows an OCT image with CNV which occurs at an advanced stage of AMD and is manifested as intra-retinal and sub-retinal fluid (red and yellow arrows respectively) deposits. Fig. 1(d) shows an OCT B-scan with intra-retinal fluid deposits (red arrows in the image) as the manifestation of DME.

Ophthalmologists scrutinize multiple B-scans of the target retinal regions to assess the severity of the pathologies. However, with the increase in the patient population, manual inspection of the OCT images may be time consuming [14]. It can also be seen from Fig. 1 that the pathological manifestations of the DME and CNV are very similar. Also in the early stages of AMD, the drusens appear very sparsely and in small diameters. This makes the drusens at early stage difficult to be distinguished from healthy OCT images. The speckle noises and the similarity of the disease markers of DME and AMD may lead to erroneous diagnosis during manual screening. So, the computer assisted diagnosis of these diseases is a viable solution for accurate mass screening of patients with retinal disorders. In the following section, we discuss the related literature on automated detection of macular pathologies, their bottlenecks for automated diagnosis and the motivation for the present work.

1.2. Related work

Various works have been proposed in the literature for the classification of retinal pathologies from OCT images. Early works on automated classification of retinal pathologies were mainly focussed on designing hand crafted discriminative features for accurate classification. Liu et al. [15] proposed multi-scale local binary pattern (LBP) based features followed by support vector machine (SVM) classifier to classify DME, AMD and macular hole. Farsiu et al. [16] estimated the volume of the total retina, the RPE volume and the RPE drusen complex (RPEDC) as features to classify the AMD and healthy control subjects. Various other hand-crafted features based classification methods have been explored in [17–20].

The limitations of hand crafted features are that the obtained features are highly dependent on the expertise and knowledge of the designers and may not be optimal. The method proposed in [16] requires the accurate segmentation of the the retinal layers for estimating the retinal volume. The low contrast and speckle noises present in the OCT images make the identification of the retinal layers difficult [21]. Also, in presence of anomalies, the accurate delineation of the retinal layers may not be optimal. The performance of these methods show poor generalization across patients and datasets. Therefore, these methods may not be suitable for automated diagnosis.

The deep learning based convolutional neural network (CNN) architectures have shown encouraging results in the classification of the retinal diseases. Karri et al. [22] proposed a transfer learning approach by fine tuning a pre-trained CNN (GoogLeNet) for classification of AMD, DME and healthy control images. Kermany et al. [23] and Li et al. [24] also employed a transfer learning based approach for classification of CNV, DME, drusen and normal OCT images. Transfer learning based methods fine-tune the pre-trained neural networks for specific applications. As the pre-trained networks are learned for natural images, understanding the behaviour of the filter outputs is inscrutable for medical images. Recently, Rasti et al. [25] proposed a multi-scale CNN ensemble classifier with a new cross-correlation based cost function for discriminative and fast learning of image features. The method provides promising classifi-

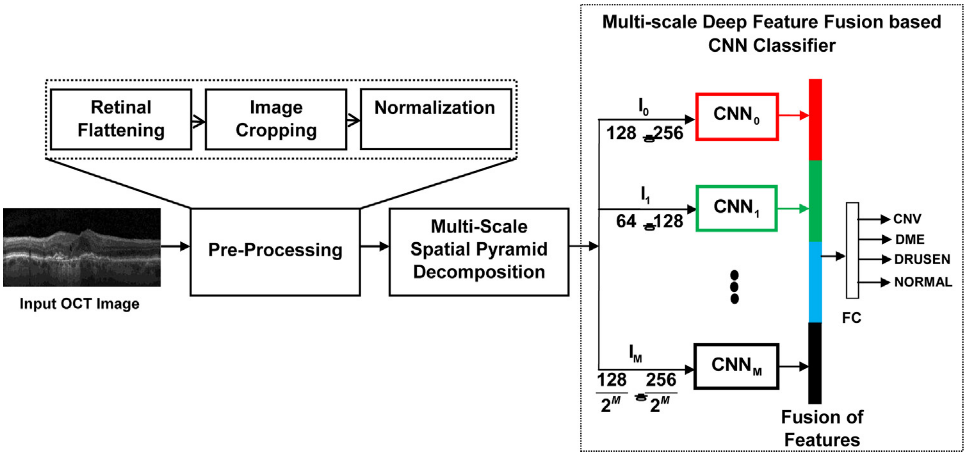

Fig. 2. Block diagram of the proposed method.

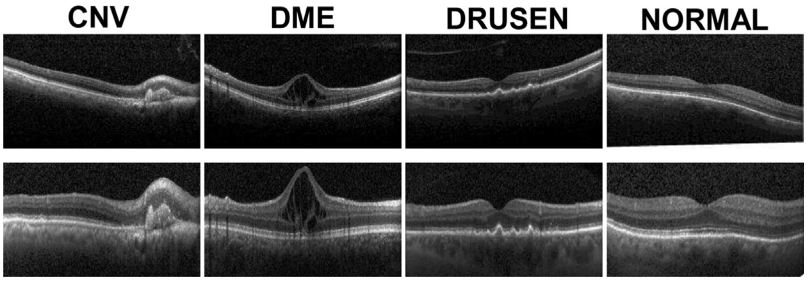

Fig. 3. Original Images (top row) and pre-processed images (bottom row).

cation accuracy. However, its effectiveness in automated diagnosis is limited because the best accuracy is highly dependent on manual setting of hyper-parameters of the loss function.

Therefore, in this work, we propose a novel *multi-scale deep feature fusion (MDFF)* approach using CNN for accurate classification of macular pathologies. The proposed framework is designed for the classification of CNV, DME, drusen and normal OCT images. The method is designed with the intuition that the fusion of features from multiple scales can capture the inter-scale variations introducing complementary information to the classifier. The fusion of features also transforms the fused feature vector to higher dimensions with increased non-linearity thereby assuring better discrimination capability to the classifier. Experimental results on large scale OCT dataset indicate that the accuracy of the proposed method is superior to the state of the art methods. The method does not require any speckle removal step or additional tuning of hyper-parameters (other than the parameters of CNN during training) for obtaining the desired classification accuracy.

The rest of the sections are organized as follows: Section 2 describes the database and the proposed method, Section 3 discusses the results and Section 4 provides the conclusion.

Table 1
Detailed architecture of the proposed MDFF classifier.

Layer	CNN_0	CNN_1	CNN_2	CNN_3
Input	128 × 256	64 × 128	32 × 64	16 × 32
CONV-	CF=3	CF=6	CF=12	CF=24
BN-ReLU-	CFS=7 × 7	CFS=5 × 5	CFS=3 × 3	CFS=3 × 3
POOL	PKS=2 × 2	PKS=2 × 2	PKS=2 × 2	PKS=2 × 2
CONV-	CF=6	CF=12	CF=24	CF=48
BN-ReLU-	CFS=3 × 3	CFS=3 × 3	CFS=3 × 3	CFS=3 × 3
POOL	PKS=2 × 2	PKS=2 × 2	PKS=2 × 2	PKS=2 × 2
CONV-	CF=12	CF=24	CF=48	
BN-ReLU-	CFS=3 × 3	CFS=3 × 3	CFS=3 × 3	-
POOL	PKS=2 × 2	PKS=2 × 2	PKS=2 × 2	
CONV-	CF=24	CF=48		
BN-ReLU-	CFS=3 × 3	CFS=3 × 3	-	-
POOL	PKS=2 × 2	PKS=2 × 2		
CONV-	CF=48			
BN-ReLU-	CFS=3 × 3	-	-	-
POOL	PKS=2 × 2			
FC	15	15	15	15
Dropout	0.5	0.5	0.5	0.5
Feature Fusion				60
Dropout	0.5			
Output	4			

2. Material and methods

In this section, we discuss the details of the database used for evaluation of the proposed method and the pipeline of the method.

2.1. Database description

The proposed method is evaluated on the University of California San Diego (UCSD) database [1] presented by Kermany et al. [23]. The database contains a train and a test set with OCT images from four categories i.e. CNV, DME, drusen and normal. The training set contains a total of 83,484 OCT images with 11,348 DME, 37,205 CNV, 8,616 drusen and 26,315 healthy control images. The test set contains 1000 images (250 images from each category) from 633 patients. The database is collected from patients (males and females) of varied age group and ethnic background.

We have trained our model on 80% of the images from the training set and validated on the remaining 20% of the set. The proposed model has been evaluated on the test set. The training set contains challenging images with many variations in the lesion size

[1] Database collected from: https://data.mendeley.com/datasets/rscbjbr9sj/2.

Table 2
Class distributions for the train, the validation and the test set and weights assigned during training.

Class	Train	Validation	Test	Weights	
				Categorical cross entropy	Weighted categorical cross entropy
CNV	29,764 (44.5%)	7,441	250	1	1
DME	9,079 (13.6%)	2,269	250	1	3.28
DRUSEN	6,893 (10.3%)	1,723	250	1	4.32
NORMAL	21,052 (31.5%)	5,263	250	1	1.41

Table 3
Effects of class weighting of the loss function on the classification performance.

Method	Class	Sensitivity (%)	Specificity (%)	G-mean (%)
Proposed MDFF (categorical cross entropy)	CNV	97.85	97.85	97.85
	DME	91.64	99.27	95.38
	DRUSEN	87.69	98.57	92.97
	NORMAL	96.91	98.14	97.52
Proposed MDFF (class weighted categorical cross entropy)	CNV	95.83	98.92	97.83
	DME	**94.41**	98.82	**96.59**
	DRUSEN	**91.57**	97.99	**94.73**
	NORMAL	96.8	98.42	97.61

and shape. The test set however is very small and contains good quality images with distinctive features. Therefore, we have discussed the results for the validation set along with the test set in the results section.

2.2. The proposed method

The block diagram of the proposed method is shown in Fig. 2. The proposed method is implemented in three steps. First, in the pre-processing step, a graph based curvature removal method [26] is applied to remove the natural curvature of the retina followed by cropping of the regions of interest. In the second step, the pre-processed image is decomposed into various scales to obtain a multi-scale view of the image. In the final step, the deep CNN features extracted from the multi-scale images are fused together to classify the OCT images into any of the four categories. The details of the different stages in the block diagram are discussed as follows:

2.2.1. Pre-processing

As the OCT images in the dataset are of different spatial resolution, the images are resized to 496 × 512 to maintain a uniform field of view. The OCT images of the retina have a natural curvature which varies among patients. To prevent the sensitivity of the classifier to the retinal curvature, retinal flatening is performed. In this work, the image flattening method proposed by Chiu et al. [26] is employed. The flattening process starts with the estimation of the RPE. The RPE is estimated by finding out the location of the brightest pixel in each column of the B-scan as the RPE is the most hyper-reflective layer of the retina. A second order polynomial is fitted on the obtained locations and each column of the image is shifted up or down so as to obtain a flat RPE.

The diagnostic information in the OCT images is confined to the retinal layers. Therefore, each B-scan is cropped horizontally considering 160 pixels above and 90 pixels below the RPE to remove the vitreous and choroid-sclera regions. The OCT B-scans are then resized to 128 × 256 for further processing. To remove the inter-patient variabilities, the B-scans are normalized to have zero mean and unit standard deviation. Fig. 3 shows the original images (top row) and the pre-processed images (bottom row). It can be seen that the curvature of the retina has been removed by the flattening process and some portions of the vitreous and the sclera have been cropped as they may not contribute much to the classification process.

2.2.2. Multi-scale spatial pyramid decomposition (MSSP)

The lesions in DME and AMD have varied shape, orientation and sizes. The drusens at early stages of AMD are very small in size and need to be analysed at a finer scale. The intra-retinal fluid accumulation in DME can be visualized in a coarse scale as their homogeneous texture within the retinal layers stand prominent. It has also been reported in literature that the retinal pathologies in DME exibits key features in multiple scales [25]. Therefore, we introduce the MSSP decomposition to create multi-scale views of the image to capture key multi-scale information for better classification [27]. The MSSP creates an image pyramid by the reduced and Gaussian low pass filtered versions of the image of the previous level. An OCT B-scan I after pre-processing is considered to be at level $j=0$. The multi-scale image at level j given the image at level $(j-1)$ is obtained as follows:

$$I_j(m,n)=\sum_{p=-2}^{2}\sum_{q=-2}^{2}w(p,q)I_{j-1}(2m+p,2n+q) \qquad (1)$$

where I_j is the image obtained at scale j and w is the kernel function. In this study, the separable filter $w(p,q)=w(p)w(q)$ with $w=[1/4-a/2, 1/4, a, 1/4, 1/4-a/2]$ and $a=0.375$ are used for simulation [25].

2.2.3. Classification methodology

The MSSP decomposed images are then fed to the proposed multi-scale feature fusion (MDFF) based classifier. The architecture of the proposed classifier is shown in Fig. 2. The images at different scales ($\mathrm{I_j}, j\in\{0,1,..,M\}$) are fed to different CNNs ($\mathrm{CNN_j}, j\in\{0,1,....,M\}$) for extracting scale specific information. Table 1 presents the detailed network architecture of the proposed multi-scale feature fusion and classification framework. In this work, we have considered and discussed the results for the multi-scale decomposition up to scale 4 ($M=3$). This is because the image resolution becomes too small to capture essential diagnostic information if the scale is increased beyond $M=3$. The CNNs at all the scales have a sequence of CONV-BN-ReLU-POOL [2] architecture followed by a fully connected (FC) layer. The first convolutional layers of the finer scale CNNs ($\mathrm{CNN_0}$ and $\mathrm{CNN_1}$) have a large convolution filter size (CFS) of

[2] CONV-Convolution layer, BN- Batch normalization, ReLU- Rectified linear unit and POOL- Max pooling.

Table 4
Average performance comparison of the baseline methods and the proposed method on the validation set.

Method	Configuration	Sensitivity (%)	Specificity (%)	Accuracy (%)	F1-Score (%)	Precision (%)	Kappa
Feature based	HoG+SVM	75.36	94.54	85.71	83.14	77.64	0.62
	LBP+SVM	48.27	88.26	71.33	64.04	44.17	0.41
Off-the-shelf CNNs	VGG16 [30]	83.51	96.16	89.47	84.53	85.98	0.72
Single scale CNNs	SC-CNN$_0$	91.62	97.86	93.88	91.55	91.48	0.84
	SC-CNN$_1$	91.76	97.93	93.93	91.46	91.46	0.83
	SC-CNN$_2$	87.21	97.23	92.36	88.56	90.33	0.79
	SC-CNN$_3$	79.71	94.84	84.13	78.18	77.06	0.57
Ensemble classifier	Averaging	92.84	98.32	95.26	93.27	93.75	0.87
Proposed method	**MDFF based classification**	**94.65**	**98.54**	**95.5**	**93.77**	**92.98**	**0.88**

Table 5
Average performance comparison of the baseline methods and the proposed method on the test set.

Method	Configuration	Sensitivity (%)	Specificity (%)	Accuracy (%)	F1-Score (%)	Precision (%)	Kappa
Feature based	HoG+SVM	84.8	84.8	94.93	84.36	88.07	0.59
	LBP+SVM	52.2	84.07	52.2	42.5	47.95	0.42
Off-the-shelf CNNs	VGG16 [30]	91.5	97.17	91.5	91.5	92.7	0.77
	Karmany et al. [23]	97.8	97.4	96.6	-	-	-
	Li et al. [24]	97.8	99.4	98.6	-	-	-
Single scale CNNs	SC-CNN$_0$	98.1	99.37	98.1	98.15	98.15	0.95
	SC-CNN$_1$	98.9	99.63	98.9	98.9	98.91	0.97
	SC-CNN$_2$	97.2	99.07	97.2	97.21	97.36	0.92
	SC-CNN$_3$	93.4	97.8	93.4	93.37	93.45	0.82
Ensemble classifier	Averaging	99.3	99.7	99.3	99.3	99.31	0.98
Proposed method	**MDFF based classification**	**99.6**	**99.87**	**99.6**	**99.6**	**99.6**	**0.99**

7 × 7 and 5 × 5 respectively to prevent the speckle noises in the OCT images from penetrating the deeper layers of the network. A CFS of 3 × 3 is used for all the other convolutional layers. BN after each convolution layer ensures fast and stable training [28]. The ReLU layer maintains the sparsity of the convolution kernel [29]. Subsampling of the feature maps is performed using max pooling with a pooling kernel size (PKS) of 2 × 2 to learn position and rotation invariant features.

For all scales, the number of convolution filters (CFs) doubles with the increasing depth. This design is inspired by the VGG16 [30] architecture which is one of the best performing classifiers for object detection. The VGG16 network incorporates a two-fold increment in the number of CFs after each pooling layer. We also introduce a two-fold increment in the number of CFs as we move across scales at a particular convolution layer. This helps the classifier to learn dissimilar features across scales. Fully connected layers with 15 neurons at each scale are fused through concatenation to merge the multi-scale feature information. A dropout layer with a probability of 0.5 is included prior and posterior to the fusion layer to curb over-fitting [31]. At the end, a Softmax output layer with 4 neurons provides the class probabilities for the input images.

The dataset described in section II-A is highly imbalanced with varied class distributions. Table 2 shows the class distribution of the OCT dataset used for evaluation of the proposed method. It can be seen that majority of the images belong to the CNV and the normal classes. The DME and the DRUSEN classes have a handful of images. The imbalance in the datasets can bias the classification result towards the majority class and thereby leading to poor detection of the minority class samples. To tackle this issue, we use a cost sensitive loss function instead of the conventional categorical cross-entropy loss while training the classifier.

Let $\{(\boldsymbol{X}_1, y_1), (\boldsymbol{X}_2, y_2), ..., (\boldsymbol{X}_N, y_N)\}$ denote a set of N training examples with $\boldsymbol{X}_i$ and y_i representing the OCT images and the corresponding class labels respectively. The conventional cross entropy loss (L) is given as

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i(\boldsymbol{X}_i, \theta)) \tag{2}$$

where θ is the network's trainable parameters and $\hat{y}_i(\boldsymbol{X}_i, \theta)$ represents the predicted posterior probability obtained by applying the Softmax function on the network's output layer. Eq. (2) computes the total loss as the average loss incurred by the samples of the training dataset. This framework provides equal weights to the classification errors of all classes. This influences the total loss towards the majority class errors thereby leading to misclassification of the minority class samples.

To mitigate this issue, we use the class weighted categorical cross entropy loss which penalizes the different class errors differently. In this case, the classification errors of the drusen and the DME class (minority class) samples are privileged with a higher weight value compared to the CNV and the normal classes (majority class). The class weighted cross entropy loss function (L_w) is defined as

$$\begin{aligned} L_w(\theta) \;\; = \;\; & -\frac{1}{N}\sum_{i=1}^{N} \Big(w_{cnv} \Pi_{cnv}(\boldsymbol{X}_i) y_i \log(\hat{y}_i(\boldsymbol{X}_i, \theta)) + \\ & w_{dme} \Pi_{dme}(\boldsymbol{X}_i) y_i \log(\hat{y}_i(\boldsymbol{X}_i, \theta)) + \\ & w_{drusen} \Pi_{drusen}(\boldsymbol{X}_i) y_i \log(\hat{y}_i(\boldsymbol{X}_i, \theta)) + \\ & w_{normal} \Pi_{normal}(\boldsymbol{X}_i) y_i \log(\hat{y}_i(\boldsymbol{X}_i, \theta)) \Big) \end{aligned} \tag{3}$$

where w_{cnv}, w_{dme}, w_{drusen} and w_{dme} are the weights assigned to the classification errors of samples of the CNV, DME, drusen and the normal classes. Π_{cnv}, Π_{dme}, Π_{drusen} and Π_{normal} are the indicator functions defining the belongingness of a sample $\boldsymbol{X}_i$ to a particular class. The weight for a particular class 'c' is computed as

$$w_c = \frac{\#\ \text{training samples in the class with highest majority (i.e. CNV class)}}{\#\ \text{training samples in class c}} \tag{4}$$

To alleviate the effects of the imbalance, the class which has less number of samples should be given priority during the loss computation. Therefore, we designed the weights to adapt to the class distribution of the dataset. These weights are inversely proportional to the class distribution. Table 2 shows the weights assigned

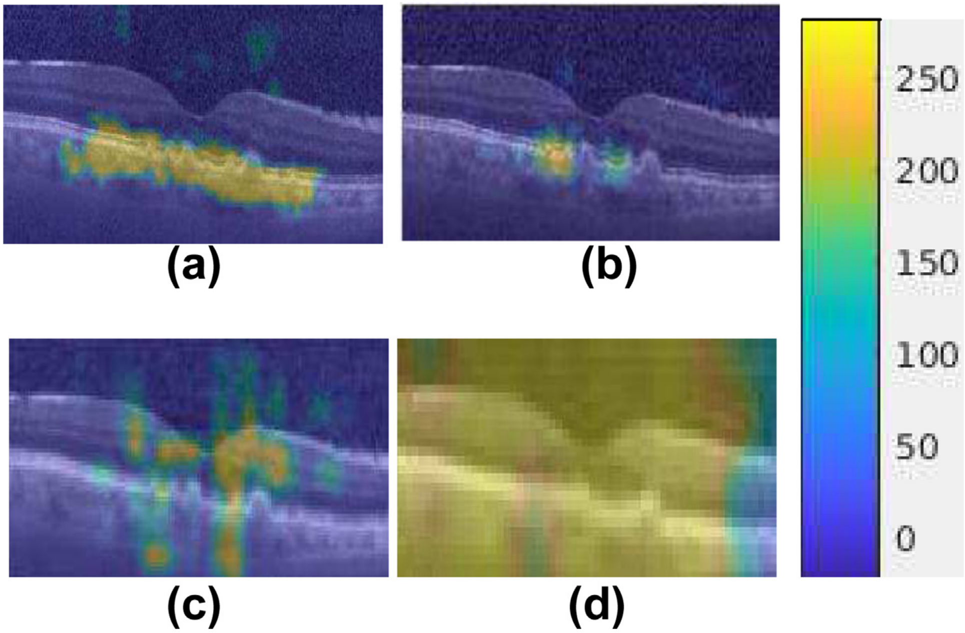

Fig. 4. Occlusion maps: (a) Proposed MDFF classifier and (b-d) Single scale CNN at $j \in \{0, 1, 2\}$ respectively.

to different classes by the categorical and the weighted categorical cross entropy loss functions. As the categorical cross entropy provides equal priority to all samples, its weights for all classes can be considered as one. The weighted categorical cross entropy loss assigns higher weight to the class that has a low sample size. The reason for choosing weights in this manner is to treat one instance of class 'c' as w_c instances of the majority class (CNV). The obtained weights equalize the degree of the impact of the losses from the individual classes towards the total loss. Also this selection of the weights is very simple with minimal computation overhead.

3. Results and discussion

In this section we discuss the various performance measures used to evaluate the proposed method and compare the method with the existing state-of-the-art methods. We show the experimental results of the effect of cost sensitive learning towards the classification performance. The effects of the multi-scale feature fusion towards classification has been discussed. We also show experimental results for the optimum number of scales to consider for fusion during classification.

3.1. Performance measures

Classification performance of the proposed method is evaluated using the performance measures obtained from the 4- classes confusion matrix. The sensitivity, specificity, accuracy, precision, F1-score, G-mean and Cohen's Kappa are used for performance analysis. Sensitivity and specificity measure the effectiveness of the classifier in identifying the positive and negative labels respectively [32]. Accuracy is the percentage of the correct predictions of the classifier. Precision measures the portion of the positive identifications to be correct. F1- score is the harmonic mean of the precision and the sensitivity. It is a single value summerizing the prediction of the positive class labels. Cohen's Kappa is a statistical measure to ensure the agreement of the obtained class labels with the ground truth labels [33]. The Kappa value lies between [-1, 1] with a value close to 1 specifying high agreement and a value close to -1 specifies complete disagreement between the obtained and the ground truth labels.

3.2. Effect of cost sensitive learning towards classification performance

In this section, we study the effects of employing the class weighted categorical cross entropy loss function for classification. The class weighted categorical cross entropy loss function is supposed to improve the detection sensitivity for the minority class samples. We compare the classification results obtained by using the conventional and the class weighted categorical cross entropy for the validation set. We use the sensitivity, specificity and the G-mean [34] as the performance measures. G-mean is a popularly used measure to validate the imbalanced datasets [35]. Table 3 shows the classification results for this experiment. It can be observed that there is an improvement in the sensitivity of drusen class from 87.69% to 91.57% and the DME class from 91.64% to 94.41% while learning the MDFF classifier using the class weighted categorical cross entropy over the conventional categorical cross entropy loss. Therefore, it can be deduced that the class weights used in this work aid in improving the detection rate of the minority class without affecting the detection rate of the majority class much.

3.3. Performance comparison with existing methods

In this section, we compare the performance of the proposed MDFF based CNN classifier with the existing baseline methods. The details of the methods and the parameter settings for experimentation are discussed as follows:

- *Feature based methods:* The Histogram of Oriented Gradients (HoG) [36] and local binary pattern (LBP) [37] histogram are the popularly used gradient and texture based feature descriptors for image classification. These features have also been used for classification of retinal disorders from OCT images [15,18]. In this experiment, the OCT images at scale $j = 0$ are used to extract these features followed by classification using the SVM classifier.

  The HOG features are computed for a cell size of 16 × 16 with nine histogram bins. The HOG feature vector with 3780 dimensions is obtained for an image of size 128 × 256 for block size of 2 × 2 cells with 50% overlap. Similarly, the LBP feature descriptor

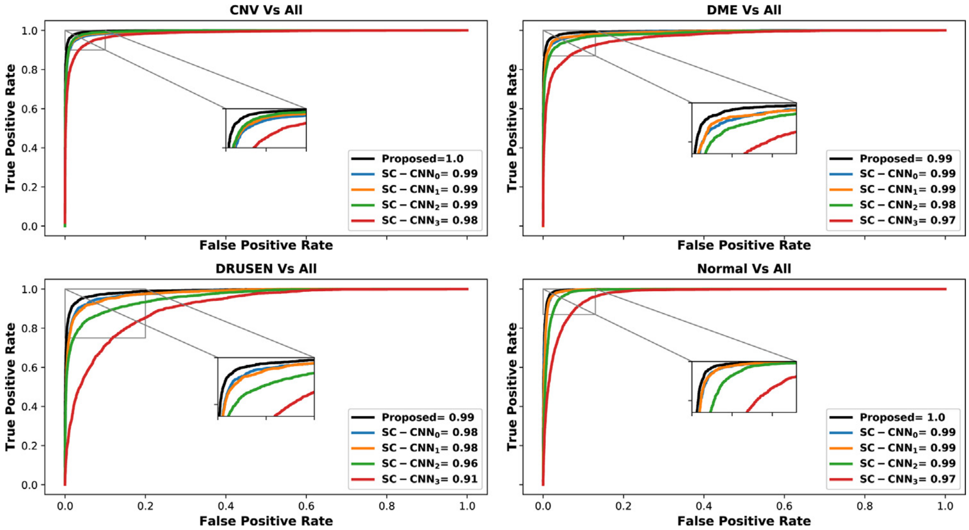

Fig. 5. ROC curves for the single scale CNNs and the proposed MDFF classifier.

is computed in three steps. For each pixel, its eight neighbours are identified along a circle with radius, $r = 1$ and the pixel value is used to threshold each neighbour value, resulting in eight binary numbers. These binary numbers are then concatenated to form an 8-bit-coded integer, which takes a value between [0-255]. A 256− dimensional histogram is obtained from the frequency of each integer occurring over the entire image. In this experiment, we have considered the rotation invariant uniform LBP features (59 dimension) [38] to make the classifier invariant to the the varied orientations of the lesions. These feature descriptors are separately provided to the SVM classifier for obtaining the class predictions.

- *Off-the-shelf CNNs:* The proposed method has been compared with the competitive *off-the-shelf* CNNs like the VGG16 [30]. We compare the proposed method with two recent macular OCT classification methods [23,24] based on fine-tuning existing CNN architectures. The VGG16 network has been successfully used for object recognition from natural images. Few changes that are incorporated to the VGG16 model for fine-tuning it to our problem includes: (a) resizing the input image at scale $j = 0$ to 224×224 and replicating it three times to construct the RGB input and (b) replacing the output layer of the network with 4 neurons for the four categories of macular OCTs. Fine-tuning is performed by fixing the network parameters for all except the output layer. The training process of the output layer is executed on mini-batches of size 32 with RMSprop algorithm [39] for 50 epochs.
- *Single Scale CNN (SC-CNN):* SC-CNN is designed to classify macular pathologies by considering input images from only one scale. For this experiment, we define SC-CNN$_j$ as the SC-CNN which is learned with OCT images at scale j for classification. The network architecture of the SC-CNN$_j$ is same as that of the CNN$_\mathrm{j}$ employed for the MDFF classifier until the first dropout layer (Table 1). An output layer with a Softmax activation function with 4 neurons is used after the dropout layer for classification. Four SC-CNN$_j$ at scales $j \in \{0, 1, 2, 3\}$ are learned for the purpose of comparison. Each of the SC-CNN is learned on mini-batches of size 32 with an adam optimizer [40] for 50 epochs with the class weighted categorical cross-entropy loss function (Eq. (3)).
- *Ensemble Classifier:* To show the effectiveness of the feature level fusion, the performance of the proposed method is compared with the average ensemble classifier. The ensemble classifier performs a score level fusion of the classification results obtained from the SC-CNNs. In this experiment, the class probabilities of the SC-CNNs at four scales $j \in \{0, 1, 2, 3\}$ are averaged to generate the final classification score.

The training process of the proposed MDFF based classifier is executed using the class weighted categorical loss function described in Eq. (3) in mini-batches of 32 using the adam optimizer with an initial learning rate of 10^{-3} with a decay of 10^{-4} per epoch for 50 epochs. The training of the proposed method, *off-the-shelf* CNNs, SC-CNNs and the ensemble classifier is performed on a system with *i*5 processor with 12 GB RAM and Nvidia Tesla K20 graphics processor. The entire code is implemented in Keras library using python.

Table 4 and 5 show the average performance of the baseline methods and the proposed MDFF classifier for the validation and the test set respectively. It is observed from the tables that the feature based methods have a very low performance compared to all the listed methods. This is because these features extract limited information from the gradients and textures of the images. The CNN based methods on the other hand extract a varied range of data specific features like edges, texture, shape, geometry, color etc that help in achieving better classification performance.

The fine-tuned VGG16 attains an average sensitivity, specificity and accuracy of of 91.5%, 97.17% and 91.5% respectively for the test set. An average precision, F1- score and Kappa of 92.7%, 91.5% and 0.77 are obtained. The values of the performance measures in Table 4 and 5 show that the *off-the-shelf* CNNs perform better than the feature based models. The tables also show that these methods cannot outperform the SC-CNNs, the ensemble model and the proposed method. The limited performance of these methods can be attributed to freezing of the initial layers during fine-tuning. As these networks are initially designed for natural images, freezing of the initial layers prevents the network from learning OCT features specific to the macular pathologies. The SC-CNNs show better performance than the feature based and the *off-the-shelf* CNN models. This is because the SC-CNNs have been trained completely on the OCT images and are capable of extracting pathology specific information. It can be seen that the SC-CNN$_0$ and SC-CNN$_1$ perform similarly. SC-CNN$_2$ and SC-CNN$_3$ show a degraded performance. This is because as the scale increases, the resolution of the input image goes down and the finer details in the images become obscure. The SC-CNNs cannot capture multi-scale information.

The results of the average ensemble (score level fusion) classifier shows an improvement in the performance over the SC-CNNs thereby validating that the multi-scale fusion of scores/features

Table 6
Class wise performance of the proposed method and the ensemble classifier.

Method	Class	Sensitivity (%)	Specificity (%)	Accuracy (%)
Ensemble Classifier	CNV	97.14	97.79	97.49
	DME	92.71	98.96	98.81
	DRUSEN	84.11	98.82	97.28
	NORMAL	97.38	97.73	97.62
MDFF Classifier (Proposed)	CNV	96.6	98.73	97.78
	DME	**94.14**	98.97	98.33
	DRUSEN	**90.49**	98.32	97.52
	NORMAL	96.9	89.26	97.85

Table 7
Average performance of the MDFF classifier based on the different combination of scales for the validation set.

Scale	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1-Score (%)	Kappa	#Trainable parameters
CNN_0	91.62	97.86	91.48	93.88	91.55	0.83	17875
CNN_0-CNN_1	91.97	98.08	92.65	94.55	92.29	0.85	40618
CNN_0-CNN_1-CNN_2	94.54	98.58	93.64	95.74	94.08	0.88	62653
CNN_0-CNN_1-CNN_2-CNN_3	94.65	98.54	92.98	95.5	93.77	0.88	82168

have a potential to outperform the SC-CNNs. It can be observed from the Tables 4 and 5 that the average performances of the proposed MDFF and the average ensemble classifier are similar. However, the ensemble model has certain issues. In an ensemble classifier, the CNNs at different scales are learned independent of each other. At the test time, the classification scores from the different CNNs are fused through averaging to obtain the final classification result. The independent learning of the CNNs prevents the ensemble classifier to exploit the complementary across scale information during the learning process. On the other hand, the proposed MDFF method incorporates this complementary information and also transforms the fused feature vector to higher dimensions with increased non-linearity thereby assuring better discrimination capability to the classifier.

Table 6 shows the advantages of the proposed method over the ensemble classifier. It shows the class wise sensitivities, specificities and accuracies of the proposed and the ensemble classifier. We can observe that the there is a significant improvement in the sensitivity from 84.11 % to 90.49% for the drusen class while using the proposed classifier over the ensemble classifier. Drusens occur at early stage of AMD and are sometimes difficult to detect because of their sparse nature. The utilization of the scale specific information during the learning process has aided in the improvement in the sensitivity. This makes the proposed method more generalizable than the ensemble classifier.

The proposed MDFF method fuses the multi-scale information at feature level to capture the inter-scale variabilities in the pathologies which results in the gain in the classification performance. The proposed method outperforms the existing methods for macular pathology classification with a promising average sensitivity, specificity and accuracy of 99.6%, 99.87% and 99.6% for the test set. The encouraging classification accuracy of the proposed method makes it highly suitable for automated diagnosis and pre-screening of macular pathologies.

3.4. Effect of multi-scale deep feature fusion

In this section, we show the effectiveness of the fusion of multi-scale features towards classification. The multi layer non-linear structure of the DNNs make it difficult to interpret the process of arriving at a classification decision [41]. The occlusion sensitivity testing is an indirect way to identify regions of the images contributing most to the network's assignment of predicted labels. The test systematically occludes different areas of the image with a black patch and monitors the classification output. The occluded areas which lead to a drop in the classification performance can be considered as important image structures affecting classification [42]. The test does not generate any overhead during training as it is performed after the training process.

In this experiment, we analyse the occlusion sensitivity of the proposed MDFF classifier to find out if the model truly identifies the location of the pathology to make a decision or just learns the surrounding context. Fig. 4(a) shows the occlusion map for the MDFF classifier using an occluding pixel patch of size 20 × 20 with a stride of 4 for an image containing drusens. The highlighted portions of the image mark the areas which when occluded generate a low classification performance. It can be seen from the figure that the proposed method successfully localizes the lesion before assigning a decision.

We also discuss the results of the occlusion sensitivity test of the SC-CNNs at three scales. The occlusion maps are shown in Fig. 4(b–d) for SC-CNN_0, SC-CNN_1 and SC-CNN_2 respectively. It can be seen that the SC-CNN_0 localizes a portion of the anomaly while missing some of the drusens. The diminishing image resolution with increasing scales truncate the detail structures of the image forcing the classifier to rely on the background regions for decision making as can be seen in Fig. 4(c–d). The occlusion maps show the advantages of using the proposed MDFF based classifier for macular pathology classification over the SC-CNNs. Fig. 5 presents the ROC curves of each class using the one-vs-all scheme. The area under the curve (AUC) values show that the proposed MDFF framework for classification performs better than the SC-CNNs thereby validating the idea of using multi-scale feature fusion for classification.

3.5. Effect of choice of scales

In this section, we find out through experimentation, the optimum number of feature scales to fuse to get better classification performance. We also discuss the the computational overhead incurred with increasing the number of scales. For this experiment, we start with learning a CNN with input at scale $j = 0$ with the architecture same as SC-CNN_0. Each time, we append the CNN at scale $j + 1$ and learn the combined architecture. In this way, we obtain the results shown in Table 7 for the validation set. It can be seen that an improvement in the classification performance is observed when the features of CNN_1 are fused with the features of CNN_0. Similar trends are observed for the fusion of features of CNN_0, CNN_1 and CNN_2. These observations justify that the fusion of multi-scale deep features enhance the classification result.

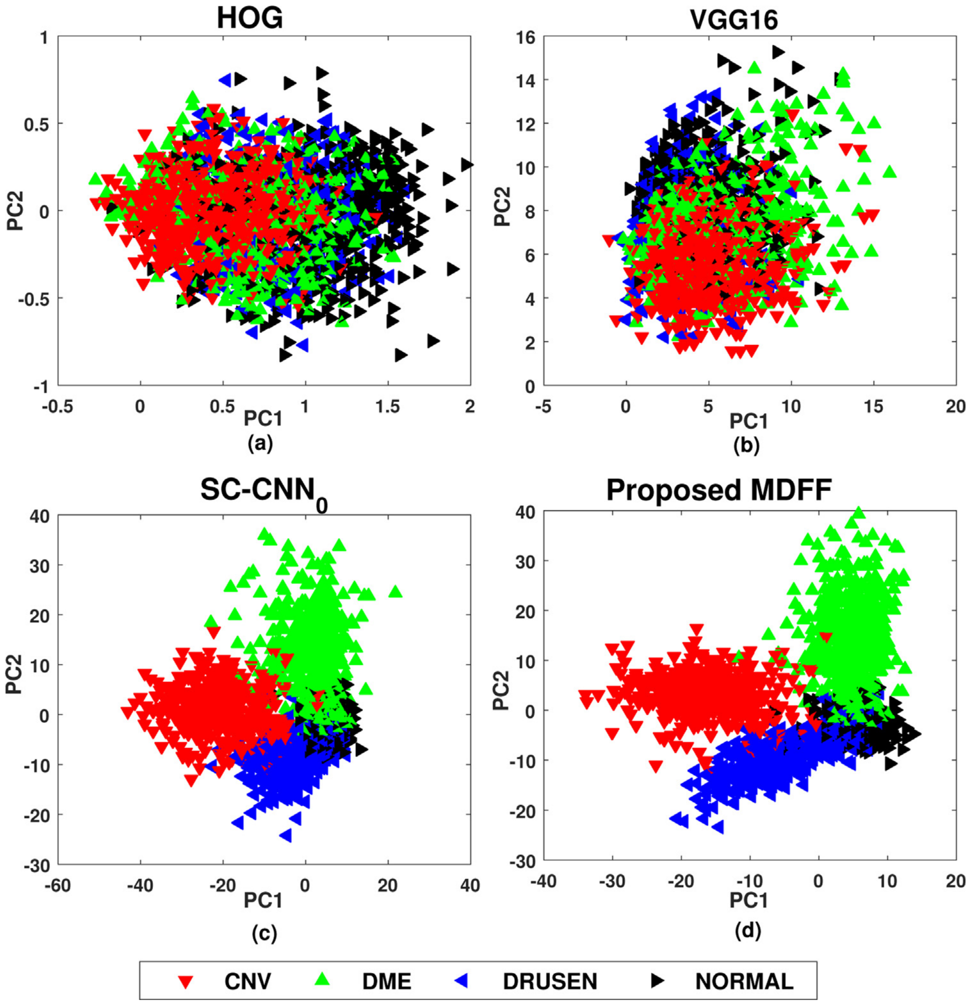

Fig. 6. Visualization of the learned features for the baseline methods and the proposed method.

It is also observed that the number of trainable parameters increases as the features from different scales are appended. Therefore, it is essential to find a trade-off between the classification accuracy and the computational burden. It can be seen from Table 7 that performance improvement of the fusion of four scales over the three scales is minimal. However, the computational load increases from learning 62653 to 82168 parameters. Hence, to reduce the computational burden and still obtain a reliable classification performance, we consider the classification result of the fusion of features at scale $j \in \{0, 1, 2\}$ as final.

3.6. *Visualization of learned features*

In this section, we visualize the learned features for the HOG, VGG16, SC-CNN$_0$ and the proposed MDFF method. The features are obtained for the images in the validation set. For the CNN based approaches, features are extracted from the fully connected layer prior to the output layer. The HOG and the VGG16 network generate feature vectors of 3780 and 4096 dimensions respectively. The SC-CNN$_0$ and the proposed MDFF approach generate 15 and 60 dimensional feature vectors respectively. In order to visualize the high dimensional learned features, we take advantage of the unsupervised principal component analysis (PCA) [43] to reduce the feature dimensions.

Fig. 6 shows the first two principal components for the baseline and the proposed methods by randomly considering 500 samples from each class of the validation set. It can be observed from Fig. 6(a) and (b) that the features of the different classes have high overlap for the HOG and VGG16 network. The classes are well separated for the SC-CNN$_0$ and the proposed method. It can be seen in Fig. 6(c) that the low dimensional representations of the drusen and the normal classes are clustered very close to each other with some overlap. The close clustering can be attributed to the high resemblance of the drusen images to the normal OCT images. The proposed MDFF features better discriminate the drusen and the normal class features as can be seen by the increased separability in Fig. 6(d).

4. Conclusion

In this paper, we proposed a novel automatic method for the detection of retinal pathologies from OCT images. The main contribution of the proposed method was the design and analysis of the multi-scale deep feature fusion architecture towards efficient classification. We also introduced the class weighted categorical cross entropy loss function to handle the class imbalance in the datasets. The advantage of the proposed method is that it does not require manual fine-tuning of parameters for improved accuracy. It also does not rely on any speckle removal step. The results demonstrate that the method outperforms the state-of-the-art macular OCT classification methods. These features make the proposed method convenient for use in eye clinics for preliminary screening of the retina and assisting ophthalmologists in making diagnostic decisions and planning the appropriate treatments.

References

[1] R. Gale, P.H. Scanlon, M. Evans, F. Ghanchi, Y. Yang, G. Silvestri, M. Freeman, A. Maisey, J. Napier, Action on diabetic macular oedema: achieving optimal patient management in treating visual impairment due to diabetic eye disease, Eye 31 (2017) S1.

[2] W.L. Wong, X. Su, X. Li, C.M.G. Cheung, R. Klein, C.-Y. Cheng, T.Y. Wong, Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis, Lancet Global Health 2 (2014) e106–e116.

[3] D.J. Taylor, A.E. Hobby, A.M. Binns, D.P. Crabb, How does age-related macular degeneration affect real-world visual ability and quality of life? a systematic review, BMJ Open 6 (2016) e011504.

[4] JohnR. Gonder, ValeryM. Walker, Martin Barbeau, Bryan H. Nancy, James R. Zaour, Zachau Hartje, Ruihong Li, Costs and quality of life in diabetic macular edema: Canadian Burden of Diabetic Macular Edema Observational Study (C-REALITY), J. Ophthalmol. (2014).

[5] N.M. Bressler, R. Varma, Q.V. Doan, M. Gleeson, M. Danese, J.K. Bower, E. Selvin, C. Dolan, J. Fine, S. Colman, et al., Underuse of the health care system by persons with diabetes mellitus and diabetic macular edema in the united states, JAMA Ophthalmol. 132 (2014) 168–173.

[6] J. Hassell, E. Lamoureux, J. Keeffe, Impact of age related macular degeneration on quality of life, Br. J. Ophthalmol. 90 (2006) 593–596.

[7] R. Zhao, A. Camino, J. Wang, A.M. Hagag, Y. Lu, S.T. Bailey, C.J. Flaxel, T.S. Hwang, D. Huang, D. Li, et al., Automated Drusen detection in dry age-related macular degeneration by multiple-depth, en face optical coherence tomography, Biomed. Opt. Express 8 (2017) 5049–5064.

[8] Y. Kanagasingam, A. Bhuiyan, M.D. Abramoff, R.T. Smith, L. Goldschmidt, T.Y. Wong, Progress on retinal image analysis for age related macular degeneration, Progr. Retinal Eye Res. 38 (2014) 20–42.

[9] N.M. Holekamp, Overview of diabetic macular edema, Am J Manag Care 22 (2016) s284–s291.

[10] J. Lee, B.G. Moon, A.R. Cho, Y.H. Yoon, Optical coherence tomography angiography of dme and its association with anti-vegf treatment response, Ophthalmology 123 (2016) 2368–2375.

[11] S.-D. Ţălu, Optical coherence tomography in the diagnosis and monitoring of retinal diseases, ISRN Biomed. Imaging 2013 (2013).

[12] W. Drexler, J.G. Fujimoto, Optical Coherence Tomography: Technology and Applications, Springer Science & Business Media, 2008.

[13] A. Al-Mujaini, U.K. Wali, S. Azeem, Optical coherence tomography: clinical applications in medical practice, Oman Med. J. 28 (2013) 86.

[14] U. Schmidt-Erfurth, A. Sadeghipour, B.S. Gerendas, S.M. Waldstein, H. Bogunović, Artificial intelligence in retina, Progr. Retinal Eye Res. (2018).

[15] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J.S. Schuman, J.M. Rehg, Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding, Medical Image Anal. 15 (2011) 748–759.

[16] S. Farsiu, S.J. Chiu, R.V. O'Connell, F.A. Folgar, E. Yuan, J.A. Izatt, C.A. Toth, A.-R.E.D.S.A.S.D.O.C.T.S. Group, et al., Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography, Ophthalmology 121 (2014) 162–172.

[17] A. Albarrak, F. Coenen, Y. Zheng, et al., Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction, Proceedings of international conference on medical image, understanding and analysis (2013) 59–64.

[18] P.P. Srinivasan, L.A. Kim, P.S. Mettu, S.W. Cousins, G.M. Comer, J.A. Izatt, S. Farsiu, Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, Biomed. Optics Express 5 (2014) 3568–3577.

[19] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. J. van Grinsven, R. Philipsen, C. Hoyng, T. Theelen, C. I. Sánchez, Automated agerelated macular degeneration classification in oct using unsupervised feature learning, in: Medical Imaging 2015: Computer-Aided Diagnosis, volume 9414, International Society for Optics and Photonics, p. 94141I.

[20] G. Lemaître, M. Rastgoo, J. Massich, C.Y. Cheung, T.Y. Wong, E. Lamoureux, D. Milea, F. Mériaudeau, D. Sidibé, Classification of sd-oct volumes using local binary patterns: experimental validation for dme detection, J. Ophthalmol. 2016 (2016).

[21] J. Cheng, D. Tao, Y. Quan, D.W.K. Wong, G.C.M. Cheung, M. Akiba, J. Liu, Speckle reduction in 3d optical coherence tomography of retina by a-scan reconstruction, IEEE Trans. Med. Imaging 35 (2016) 2270–2279.

[22] S.P.K. Karri, D. Chakraborty, J. Chatterjee, Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration, Biomed. Optics Express 8 (2017) 579–592.

[23] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (2018) 1122–1131.

[24] F. Li, H. Chen, Z. Liu, X. Zhang, Z. Wu, Fully automated detection of retinal disorders by image-based deep learning, Graefe's Arch. Clin. Exp. Ophthalmol. (2019) 1–11.

[25] R. Rasti, H. Rabbani, A. Mehridehnavi, F. Hajizadeh, Macular oct classification using a multi-scale convolutional neural network ensemble, IEEE Trans. Med. Imaging 37 (2018) 1024–1034.

[26] S.J. Chiu, X.T. Li, P. Nicholas, C.A. Toth, J.A. Izatt, S. Farsiu, Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation, Optics Express 18 (2010) 19413–19428.

[27] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, in: Readings in Computer Vision, Elsevier, 1987, pp. 671–679.

[28] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv: 1502.03167 (2015).

[29] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 807-814.

[30] K. Simonyan, A. Zisserman, ery deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

[32] D.G. Altman, J.M. Bland, Diagnostic tests. 1: Sensitivity and specificity, Br. Med. J. 308 (1994) 1552.

[33] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Measure. 20 (1960) 37–46.

[34] Kubat, R. Holte, S. Matwin, Learning when negative examples abound, in: European Conference on Machine Learning, Springer, 146-153.

[35] J. Du, C.-M. Vong, C.-M. Pun, P.-K. Wong, W.-F. Ip, Post-boosting of classification boundary for imbalanced data using geometric mean, Neural Networks 96 (2017) 101–114.

[36] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005, CVPR, IEEE Computer Society Conference on, volume 1, IEEE, 2005, pp. 886–893.

[37] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intel. 24 (2002) 971–987.

[38] T. Ahonen, J. Matas, C. He, M. Pietikäinen, Rotation invariant image description with local binary pattern histogram fourier features, in: Scandinavian Conference on Image Analysis, Springer, 61-70.

[39] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747, 2016.

[40] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

[41] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, K.-R. Müller, Evaluating the visualization of what a deep neural network has learned, IEEE Trans. Neural Netw. Learn. Syst. 28 (2017) 2660–2673.

[42] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European conference on computer vision, Springer, 818-833.

[43] I. Jolliffe, Principal Component Analysis, Springer, 2011.