Graefe's Archive for Clinical and Experimental Ophthalmology (2019) 257:495–505
https://doi.org/10.1007/s00417-018-04224-8

RETINAL DISORDERS

Fully automated detection of retinal disorders by image-based deep learning

Feng Li[1] · Hua Chen[1] · Zheng Liu[1] · Xuedian Zhang[1] · Zhizheng Wu[2]

Received: 3 August 2018 / Revised: 10 December 2018 / Accepted: 18 December 2018 / Published online: 4 January 2019

Abstract

Purpose With the aging population and the global diabetes epidemic, the prevalence of age-related macular degeneration (AMD) and diabetic macular edema (DME) diseases which are the leading causes of blindness is further increasing. Intravitreal injections with anti-vascular endothelial growth factor (anti-VEGF) medications are the standard of care for their indications. Optical coherence tomography (OCT), as a noninvasive imaging modality, plays a major part in guiding the administration of anti-VEGF therapy by providing detailed cross-sectional scans of the retina pathology. Fully automating OCT image detection can significantly decrease the tedious clinician labor and obtain a faithful pre-diagnosis from the analysis of the structural elements of the retina. Thereby, we explore the use of deep transfer learning method based on the visual geometry group 16 (VGG-16) network for classifying AMD and DME in OCT images accurately and automatically.

Method A total of 207,130 retinal OCT images between 2013 and 2017 were selected from retrospective cohorts of 5319 adult patients from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and the Beijing Tongren Eye Center, with 109,312 images (37,456 with choroidal neovascularization, 11,599 with diabetic macular edema, 8867 with drusen, and 51,390 normal) for the experiment. After images preprocessing, 1000 images (250 images from each category) from 633 patients were selected as validation dataset while the rest images from another 4686 patients were used as training dataset. We used deep transfer learning method to fine-tune the VGG-16 network pre-trained on the ImageNet dataset, and evaluated its performance on the validation dataset. Then, prediction accuracy, sensitivity, specificity, and receiver-operating characteristic (ROC) were calculated.

Results Experimental results proved that the proposed approach had manifested superior performance in retinal OCT images detection, which achieved a prediction accuracy of 98.6%, with a sensitivity of 97.8%, a specificity of 99.4%, and introduced an area under the ROC curve of 100%.

Conclusion Deep transfer learning method based on the VGG-16 network shows significant effectiveness on classification of retinal OCT images with a relatively small dataset, which can provide assistant support for medical decision-making. Moreover, the performance of the proposed approach is comparable to that of human experts with significant clinical experience. Thereby, it will find promising applications in an automatic diagnosis and classification of common retinal diseases.

Keywords Age-related macular degeneration · Diabetic macular edema · Optical coherence tomography · Deep transfer learning · Visual geometry group 16 network

Introduction

Visual impairment represents a major worldwide health concern, which not only affects the normal vision but also carries a significant loss of life quality and a substantial economic burden to both patients and healthcare systems [1, 2]. AMD and DME are the prime causes of irreversible vision loss worldwide [3–8]. As a noninvasive and contactless imaging modality, OCT has emerged as an essential adjunct for the diagnosis and monitoring of patients with AMD and DME

✉ Hua Chen
171655822@qq.com

[1] School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

[2] Department of Precision Mechanical Engineering, Shanghai University, Shanghai 200072, China

[9–11]. As shown in Fig. 1, it can provide a clear cross-sectional representation of the retinal pathology, such as choroidal neovascularization (CNV), DME, DRUSEN, and NORMAL. Traditionally, it is necessary to inspect each cross-section within the OCT volume individually by the clinician, which increases significantly the workload of the ophthalmologist and heavily limits the capacity of acquiring large number of ground truth [12, 13]. Hence, the importance of automated identification for retinal pathologies given retinal OCT images has been recognized.

Various automatic detection technologies using OCT images have been developed for specific retinal pathology identification, either focusing on segmentation or classification. Srinivasan et al. [14] employed multiscale histograms of oriented gradient (HOG) descriptors and the support vector machine (SVM) to classify retinal images within a spectral domain (SD)-OCT volume as normal, AMD and DME. Liu et al. [15] utilized the multi-scale spatial pyramid and local binary pattern (LBP) histograms for encoding texture and shape to detect automatically multiple macular pathologies in ocular OCT images. Alsaih et al. [16] adopted multi pyramids, LBP and HOG descriptors to classify SD-OCT volumes as DME volumes and normal volumes. Although these traditional classification approaches introduced above [14–16] which employed the hand-crafted features for classifiers can obtain promising results, there existed some disadvantages, such as requiring abundant expert knowledge, time-consuming and poor generalization to other domains.

In order to address these disadvantages, machine learning techniques which can learn representations from raw data automatically have been utilized for automated classification of medical images, nowadays. Abràmoff et al. [17] utilized a deep learning enhanced algorithm for automated detection of diabetic retinopathy (DR), which achieved significantly good performance. Schlegl et al. [18] focused their studies on the deep learning based fully automated artificial intelligence approach for detection and quantification of Macular Fluid in macular OCT volume scans, resulting in an excellent accuracy. In addition, Burlina et al. [19] reported image features computed from pre-trained convolutional neural network (CNN) can be effectively transferred to detection of AMD with good preliminary performance. However, these deep learning methods [17–19] either used the raw images to train the CNN from scratch requiring a large amount of training data and weeks to achieve a classification accuracy, or used the feature-based transfer learning method without further optimization and additional fine-tuning resulting in a poor performance of classification.

Considering fine-tuning a VGG-16 network pre-trained on the ImageNet dataset for retinal OCT image classification has not been further explored. Thereby, in this study, we explored the use of deep transfer learning method based on the VGG-16 network to classify OCT images with DME and AMD for providing an accurate and timely diagnosis of key pathology. This method was fully autonomous and no need for engineered features extracted from retinal OCT image dataset for the diagnosis. The proposed approach can be potentially applied to clinics to assist ophthalmologists to make a diagnostic decision, and furthermore extended to a wide range of application in medical imaging techniques to assist clinicians to perform more precise analysis and suitable treatments.

Method

In this section, we conducted an analysis of deep transfer learning method based on the VGG-16 network in diagnosing the most common blinding retinal disorders. The flowchart of the proposed approach for classifying retinal OCT images was shown in Fig. 2. The images were firstly acquired, labeled and graded. Posteriorly, we applied image preprocessing technology such as image normalization to the labeled image dataset. However, we did not perform image denoising for avoiding overfitting and enhancing generalization ability of the classifier. After image preprocessing, the built retinal OCT image dataset was divided into training dataset and validation dataset. Finally, the prediction for the input image was determined by the output from the trained VGG-16 network using deep transfer learning method on retinal OCT images. The proposed approach did not look for specific known features

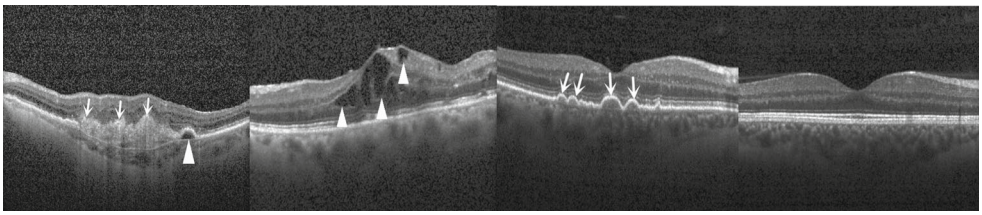

Fig. 1 Layer deformation due to neovascular membrane and retinal fluid in comparison to a normal subject's retina. **a** Choroidal neovascularization (CNV) with neovascular membrane (arrows) and retinal pigment epithelium detachment (white arrowhead). **b** Diabetic macular edema (DME) with retinal-thickening-associated intraretinal fluid (white arrowheads). **c** Multiple drusen (arrows) present in early age-related macular degeneration (AMD). **d** Normal retina

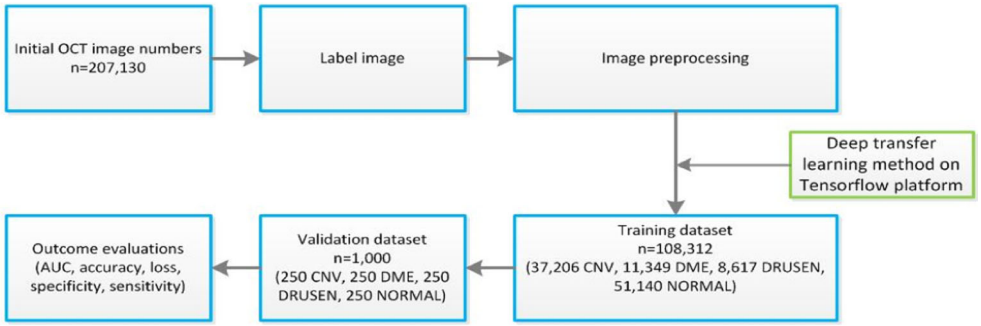

Fig. 2 Workflow diagram of the proposed approach for classifying retinal optical coherence tomography (OCT) images. CNV: choroidal neovascularization; DME: diabetic macular edema; AUC: area under curve of receiver operating characteristic (ROC)

of the input images, but automatically learned the differentiation of several groups composed of a great deal of example images. In medical diagnosis, we classified images with CNV and images with DME as urgent referrals, images with DRUSEN as routine referrals, and NORMAL as observation.

Image acquisition

Retinal images were obtained using the SD-OCT device (Heidelberg Spectralis, Heidelberg Engineering, Germany). The baseline scan protocol consisted of: scan extent = volume scan (15° × 5°); scan sections = B-scans; 7 sections; and OCT automatic real time (ART) averaging, where ART is set to 20, which indicates that 20 SD-OCT images were averaged. This study was performed in compliance with the declaration of Helsinki and the standards of local ethics committee, in which all completely anonymous retinal OCT images (207,130) were selected from retrospective cohorts of adult patients from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center between 2013 and 2017.

Image labeling

A set of retinal OCT images and corresponding labels (CNV, DME, DRUSEN, NORMAL) were needed as training data. Before retraining the VGG-16 network, all images were graded for the presence of CNV, DME, DRUSEN, and NORMAL by two retinal specialists to verify and correct image labels. The grading was performed on full-screen, high-resolution 27-in. monitors. OCT images with severe artifacts causing misalignment and blurring of sections or significant image resolution reductions were removed directly. Only images with a clear consensus annotation between ophthalmologists were taken into the sample and imported into the database. For disagreement in image labels, an experienced retinal specialist performed to arbitrate. Meanwhile, in order to avoid any errors in grading, the resulting evaluation set was further checked by another senior retina expert. The selection and allocation of dataset were displayed in Fig. 2.

Image preprocessing

The approach proposed in this paper needed to preprocess the original images by software. Considering the retinal layers may be tilted, distorted, and shifted vertically resulting from the acquisition process of variant intensity, we needed to normalize and reduce the size of the OCT volumes. Additionally, according to the computational requirements of VGG-16 network, we needed to establish a uniform size of 224 × 224 dimension for all images in the dataset. However, the retina was curved so that simply cropping each B-scan may be ill-suited, which can remove most of the informative data and result in marginal resizing. Therefore, we adopted an effective normalization (flattening) strategy [20], which composed of aligning the individual BM layers, correcting the eye curvature and normalizing variations in volume intensities.

Dataset creation

The retinal OCT dataset for the experiment consisted of 109,312 images from 5319 adult patients, of which 37,456 were affected by CNV, 11,599 represented DME-afflicted patients, 51,390 represented healthy cases, and the others were related to DRUSEN cases. Each image of retinal OCT dataset had passed manually quality review by retinal specialists. For evaluation of the proposed approach, we selected 250 images from each category with CNV, DME, DRUSEN, and NORMAL from the retinal OCT dataset resulting in a total of 1000 images from 633 patients as validation dataset, and the rest images from another 4686 patients were used as training dataset. The built training dataset was used to retrain the weights of the upper levels of VGG-16 network with back propagation, while performance was assessed on the validation dataset. In addition, the same datasets were also used for binary classifiers to compare CNV/DME/DRUSEN from NORMAL so that a breakdown of the model's performance can be determined. Furthermore, in order to compare deep transfer learning performance using limited dataset and using

a large dataset, we trained a "limited model" to distinguish the same four categories. We only extracted 1000 images from each category respectively leading to a total of 4000 images as a limited training dataset, while performance of classification was evaluated using the same validation dataset.

Deep transfer learning method based on the VGG-16 network

VGG-16 network

In this paper, we adopted the macro architecture of VGG-16 network as listed in Table 1. The model was composed of 13 convolution layers, five max-pooling layers, and three fully-connected layers. The convolution operation performed with convolution filters (kernels) to shape feature maps [21, 22] was used for extracting features from the input various images of ImageNet dataset. The max-pooling operation taking the highest value from each kernel reduced the size of the feature maps and ensured a fixed output size. The stride (number of units the filter slides) for convolution and max-pooling was set at 1 and 2 respectively. The fully-connected layer employed a soft-max activation function for the output layer so that the input image can be predicted into 1 of 1000 categories.

The input to VGG-16 network consisted of fixed-size 224 × 224 × 3, where each dimension respectively represented width, height, and channel. The input layer was convolved with 64 3 × 3 × 3 kernels to construct layer 1. Then, a convolution was performed again in layer 1 with 64 3 × 3 × 64 kernels. After which, a 2 × 2 max-pooling operation was performed to generate layer 3. And so on, until the layer 18, in which 7 × 7 × 512 number of neurons were produced. Fully-connected layers were arranged in layers from 19 to 21. The neurons in layer 18 were fully-connected to 4096 neurons in layer 19. Also, layer 19 was fully connected to layer 20 with 4096 neurons. Finally, the layer 20 was fully-connected to the layer 21 (the output layer) with an output of 1000 neurons to represent 1000 categories.

Deep transfer learning method

Considering a relative small retinal OCT image dataset, less training time and high accuracy in this study, we proposed a deep transfer learning method to adapt to our classification by

Table 1 The details of architecture of VGG-16 network

Layers	Type	Number of feature maps	Number of neurons in the layer	The size of the kernel involves to form each feature map	Stride	Number of trainable parameters
0	Input	3	224 × 224 × 3	–	–	0
1	Convolution	64	224 × 224 × 64	3 × 3 × 3	1	1728
2	Convolution	64	224 × 224 × 64	3 × 3 × 64	1	36,864
3	Max-pooling	64	112 × 112 × 64	2 × 2	2	0
4	Convolution	128	112 × 112 × 128	3 × 3 × 64	1	73,728
5	Convolution	128	112 × 112 × 128	3 × 3 × 128	1	147,456
6	Max-pooling	128	56 × 56 × 128	2 × 2	2	0
7	Convolution	256	56 × 56 × 256	3 × 3 × 128	1	294,912
8	Convolution	256	56 × 56 × 256	3 × 3 × 256	1	589,824
9	Convolution	256	56 × 56 × 256	3 × 3 × 256	1	589,824
10	Max-pooling	256	28 × 28 × 256	2 × 2	2	0
11	Convolution	512	28 × 28 × 512	3 × 3 × 256	1	1,179,648
12	Convolution	512	28 × 28 × 512	3 × 3 × 512	1	2,359,296
13	Convolution	512	28 × 28 × 512	3 × 3 × 512	1	2,359,296
14	Max-pooling	512	14 × 14 × 512	2 × 2	2	0
15	Convolution	512	14 × 14 × 512	3 × 3 × 512	1	2,359,296
16	Convolution	512	14 × 14 × 512	3 × 3 × 512	1	2,359,296
17	Convolution	512	14 × 14 × 512	3 × 3 × 512	1	2,359,296
18	Max-pooling	512	7 × 7 × 512	2 × 2	2	0
19	Fully-connected	–	4096	–	–	102,760,448
20	Fully-connected	–	4096	–	–	16,777,216
21	Fully-connected	–	1000	–	–	4096,000
					Total	138,344,128

fine-tuning the pre-trained VGG-16 network on the ImageNet dataset as seen in Fig. 3. Firstly, the convolutional (conv) layers were initialized with loaded pre-trained weights which were initially calculated and stored to reduce redundant processes and speed-up training. Weights of the last three fully connected (fc) layers were randomly initialized to learn the feature space shifts between ImageNet images and retinal OCT images. The last fully connected layer was adjusted to four output classes for CNV, DME, DRUSEN, and NORMAL instead of the 1000 output classes of the ImageNet, and the soft-max activation function was changed to logsoft-max activation function to improve its probabilities and avoid underflow of gradients. Then, we froze the all convolutional layers and the corresponding max-pooling layers as fixed feature extractors. Finally, the newly initialized network took the image bottlenecks which were values of each training and validation images after they had passed through the frozen layers of our model as input and was retrained to detect our specific four classifications on our retinal OCT image dataset. In the retraining process, we attempted to unfreeze and update the pre-trained weights by backpropagation in our retinal OCT images to fine-tune the convolutional layers for avoiding overfitting.

The newly initialized network was trained on an Ubuntu 16.04 operation system with Intel Core i7-2700K 4.6 GHz CPU, 256 Gb RAM, Dual AMD Filepro 512 Gb PCIe-based flash storage, and NVIDIA GTX 1080 8 Gb GPU for training and validation. We utilized SGD optimizer to train layers in batches of 1000 images per step with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 10^{-4}. After 200 epochs, the training was stopped since both accuracy and cross-entropy loss would not be further improved.

Statistical analysis

In this study, all statistical analyses were performed using the Python language with numpy and scikit.learn modules (Anaconda Python, Continuum Analytics). On the training and validation datasets, the classification accuracy (Acc) was calculated. We also calculated classification sensitivity (Sen) and specificity (Spe) for the validation dataset, then ROC curves were plotted by the values of the true positive rate (sensitivity) against the false positive rate (1 – specificity).

The formulas for calculating the Acc, the Sen, and the Spe were defined as:

$$\mathrm{Sen} = \frac{TP}{TP + FN} \tag{1}$$

$$\mathrm{Spe} = \frac{TN}{TN + FP} \tag{2}$$

$$\mathrm{Acc} = \frac{TN + TP}{TN + FP + TP + FN} \tag{3}$$

Where, true positive (TP) referred to the total number of correctly diagnosed urgent referrals; false positive (FP) represented the total number of non-urgent referrals diagnosed as

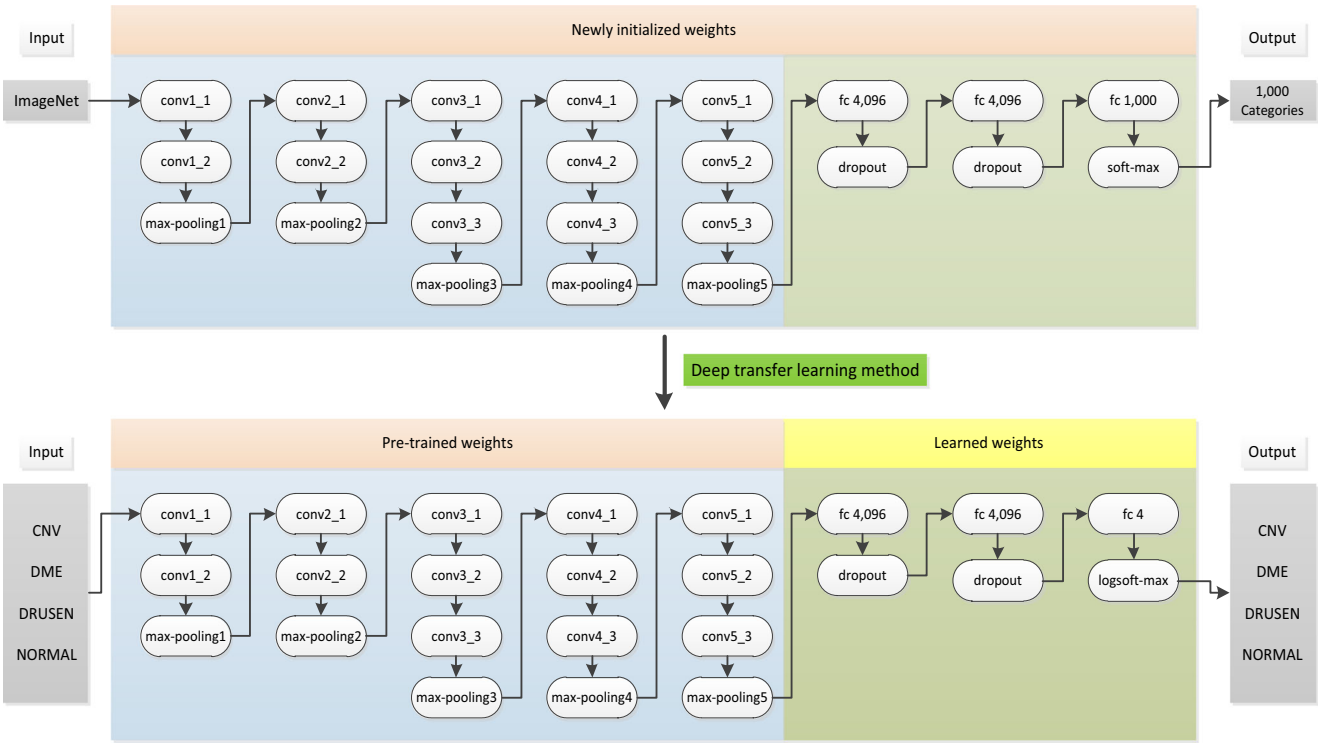

Fig. 3 Schematic of deep transfer learning method. CNV: choroidal neovascularization; DME: diabetic macular edema; conv: convolutional; fc: fully-connected

urgent referrals; false negative (FN) described the total number of urgent referrals diagnosed as non-urgent referrals; and true negative (TN) was the total number of correctly diagnosed non-urgent referrals.

Results

Performance of the model

The performance of the model was evaluated in our built retinal OCT image dataset. When images were identified as urgent referrals (CNV and DME), patients needed to be transferred to an ophthalmologist in time for definitive anti-VEGF treatment. If treatment was delayed, it would almost cause irreversible vision impairment and even lead to blindness. By contrast, for the condition of DRUSEN, there was less urgent to referral to ophthalmologists. The distribution of images over the training and validation was shown in Table 2.

As can be seen from Fig. 4a, in a multi-class comparison between CNV, DME, DRUSEN, and NORMAL, 986 of 1000 images on the validation dataset were accurately identified and the prediction accuracy was up to 98.6%. Meanwhile, it can also be seen in Fig. 4a that 11 images of urgent referrals group (500 images) were incorrectly classified resulting in a sensitivity of 97.8% and 3 images belonging to non-urgent referrals group (500 images) were also incorrectly categorized as urgent referrals leading to a specificity of 99.4%. The ROC curve was generated for quantifying the performance of model to distinguish urgent referrals from DRUSEN and NORMAL. The AUC generated by plotting sensitivity vs 1-specificity reached 100% as shown in Fig. 4b.

After 200 epochs, the training accuracy and cross-entropy loss were plotted in Fig. 5a, and the validation accuracy and cross-entropy loss were plotted in Fig. 5b. In order to facilitate clear visualize trends, plots were normalized with a smoothing factor of 0.6. It can be observed from Fig. 5 that an accuracy of 96.63% on the training dataset and a 98.6% accuracy on the validation dataset were achieved respectively. By comparison with training performance, the validation accuracy and loss showed better performance since images with more noise and lower quality were also included in the training dataset to reduce overfitting and help generalization of the classifier.

In order to further validate performance of model, we also trained the "limited model" for identifying the same four categories using our built limited training dataset including 4000 images consisted of randomly extracted 1000 from each categorization. Compared to results obtained using a large training dataset, a relatively good performance of 96.1% accuracy, 97% sensitivity, and 95.2% specificity was obtained and an AUC of 99.37% was achieved on the same validation dataset (as depicted in Fig. 6). This showed that the proposed approach was robust and capable of detecting automatically and accurately retinal disorders.

For determining a breakdown of the model's performance, we implemented binary classifiers to distinguish CNV/DME/DRUSEN from NORMAL using the same retinal OCT image dataset. The results showed that the model distinguishing CNV from NORMAL achieved a prediction accuracy of 100.0% (Fig. 7a), with a sensitivity of 100.0% and a specificity of 100.0%, while the AUC was 100.0% (Fig. 8a). For discriminating DME from NORMAL, the model yielded 98.8% accuracy (Fig. 7b), 98.8% sensitivity and 98.8% specificity, while an AUC of 99.89% was obtained (Fig. 8b). The model differentiating DRUSEN from NORMAL reached a high level with an accuracy of 99.2% (Fig. 7c), a sensitivity of 98.4% and a specificity of 100%, while the ROC curves had an AUC of 99.99% (Fig.8c).

Discussion

In this paper, we explored a deep transfer learning method based on the VGG-16 network to distinguish AMD from DME using retinal OCT images automatically. The approach avoided an ample amount of example images and higher number of epochs for convergence by fine-tuning the VGG-16 network pre-trained on the ImageNet dataset and achieved comparable performance to that of human experts in diagnosing retinal OCT images [23]. When the model trained with limited training data incorporating 1000 from each category can also yield high performance in accuracy, sensitivity, specificity, and AUC for achieving the correct identification and referral. In addition, the results of binary classification validated a good breakdown of the model's performance. Thereby, the highly effective identification and prediction capabilities

Table 2 Distribution of retinal OCT images over the training and validation datasets

Number of OCT images	Urgent referrals group		Non-urgent referrals group		Total
	CNV	DME	DRUSEN	NORMAL	
Training dataset	37,206	11,349	8617	51,140	108,312
Validation dataset	250	250	250	250	1000

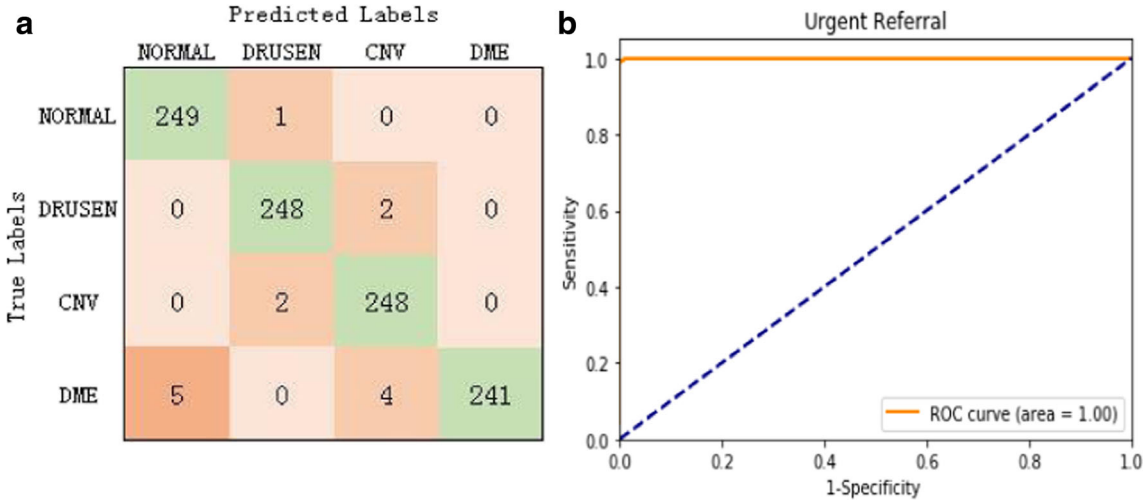

Fig. 4 Multi-class comparison between choroidal neovascularization (CNV), diabetic macular edema (DME), DRUSEN, and NORMAL. **a** The confusion matrix of our model's classification on the validation dataset. **b** Receiver-operating characteristic (ROC) curve reflected the significant performance of the model classification

of the deep transfer learning method based on the VGG-16 network were illustrated, even though using a small training dataset. The proposed approach will find potential application in the pre-diagnosis of other common illnesses like glaucoma, DR, and hypertensive retinopathy.

OCT as a popular technique in the field of ophthalmology can provide images of a cross-sectional visualization with high-resolution of the retinal layers to capture the complex morphology in AMD, and reliably detect DME and other retinal disorders. Especially, it plays an important role in guiding the administration of anti-VEGF therapy. As a general approach, the pre-trained VGG-16 network can be applied to a wide variety of OCT image classifications. There was no need to understand the underlying eye disease process; the results yielded on the training dataset provided by historical treatment decisions determined their performances.

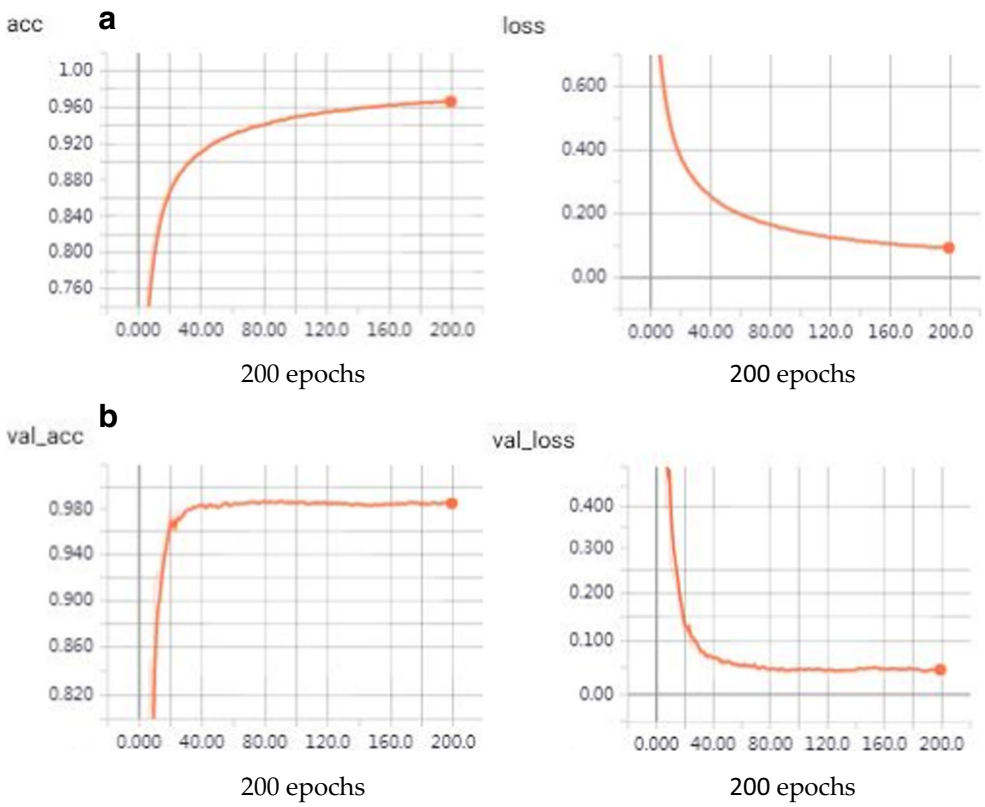

Fig. 5 Plot showing performance on the training and validation datasets using TensorBoard. Acc: accuracy on the training dataset; loss: loss on the training dataset; val_acc: accuracy on the validation dataset; val_loss: loss on the validation dataset

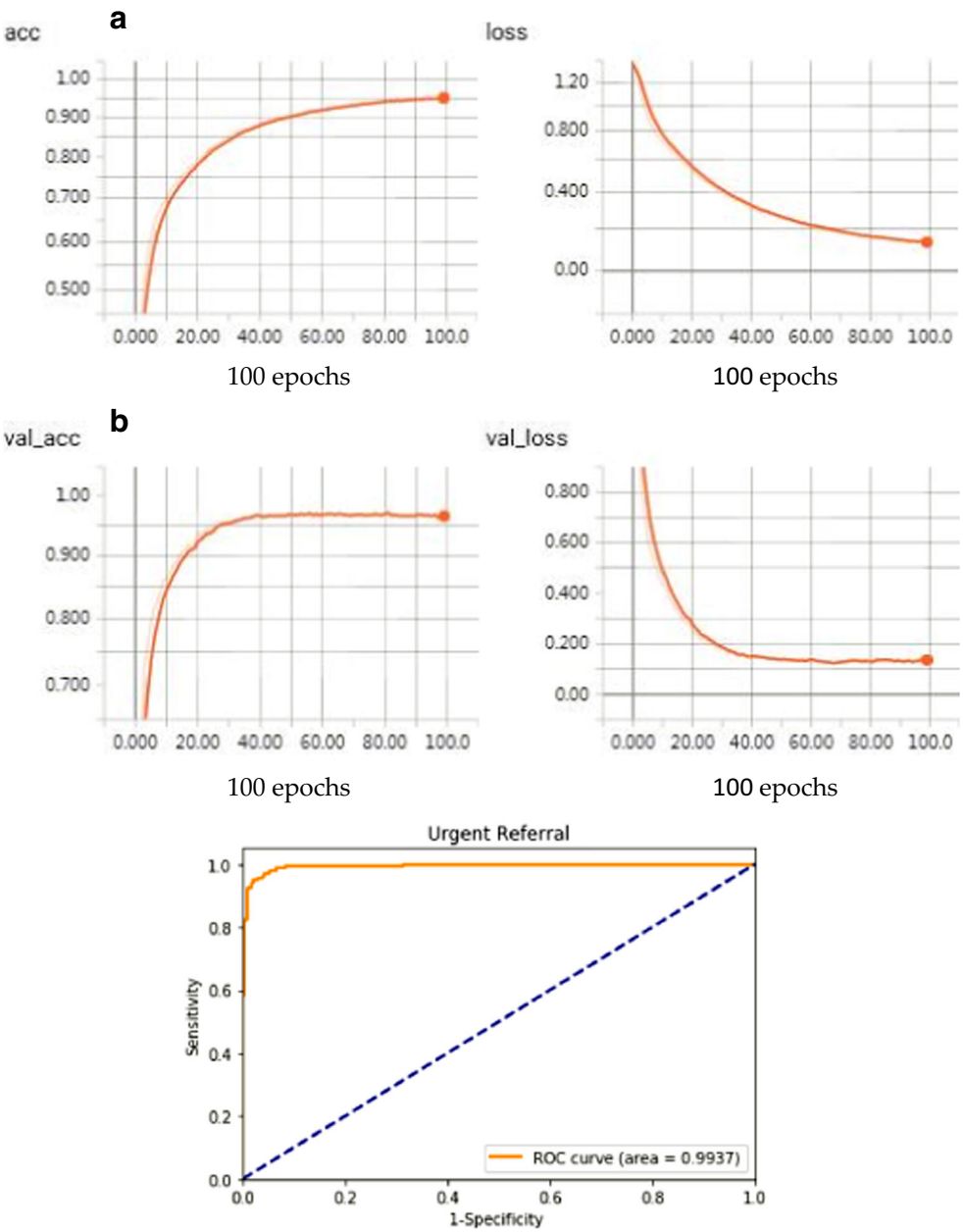

Fig. 6 Plot showing performance on the limited dataset using TensorBoard. Acc: accuracy on the training dataset; loss: loss on the training dataset; val_acc: accuracy on the validation dataset; val_loss: loss on the validation dataset; ROC: receiver operating characteristic

Nevertheless, by comparison with conventional computed explanations of diagnostic tests, the pre-trained VGG-16 network cannot offer transparent interpretations of results presented. Thereby, the output of artificial neural network should not be used to treatment instructions, but be used to offer the clinical support.

Our evaluation clearly illustrated that the deep transfer learning method had a definite advantage over learning from scratch since the performance of blank CNN was subject to selective initialization, which accelerated model training, saved memory, and achieved accurately classification between CNV, DME, DRUSEN, and NORMAL with high accuracy using a relatively small OCT image dataset. The proposed approach can accurately capture patterns within images and hence produce relatively high performance at automated detection of retinal disorders. However, its performance was slightly inferior to that of a blank CNN trained from a random initialization on a large OCT image dataset, since the weights of the blank CNN were directly adjusted and optimized to make accurate predictions of ground truths by minimizing errors in the pre-defined loss function. Unfortunately, the large amount of medical images as the underlying datasets was difficult to collect. Even if collected, the training also needed to adjust the huge parameters so that it would require several weeks to converge to a high accuracy, while a multiclass holdout model trained using deep transfer learning only

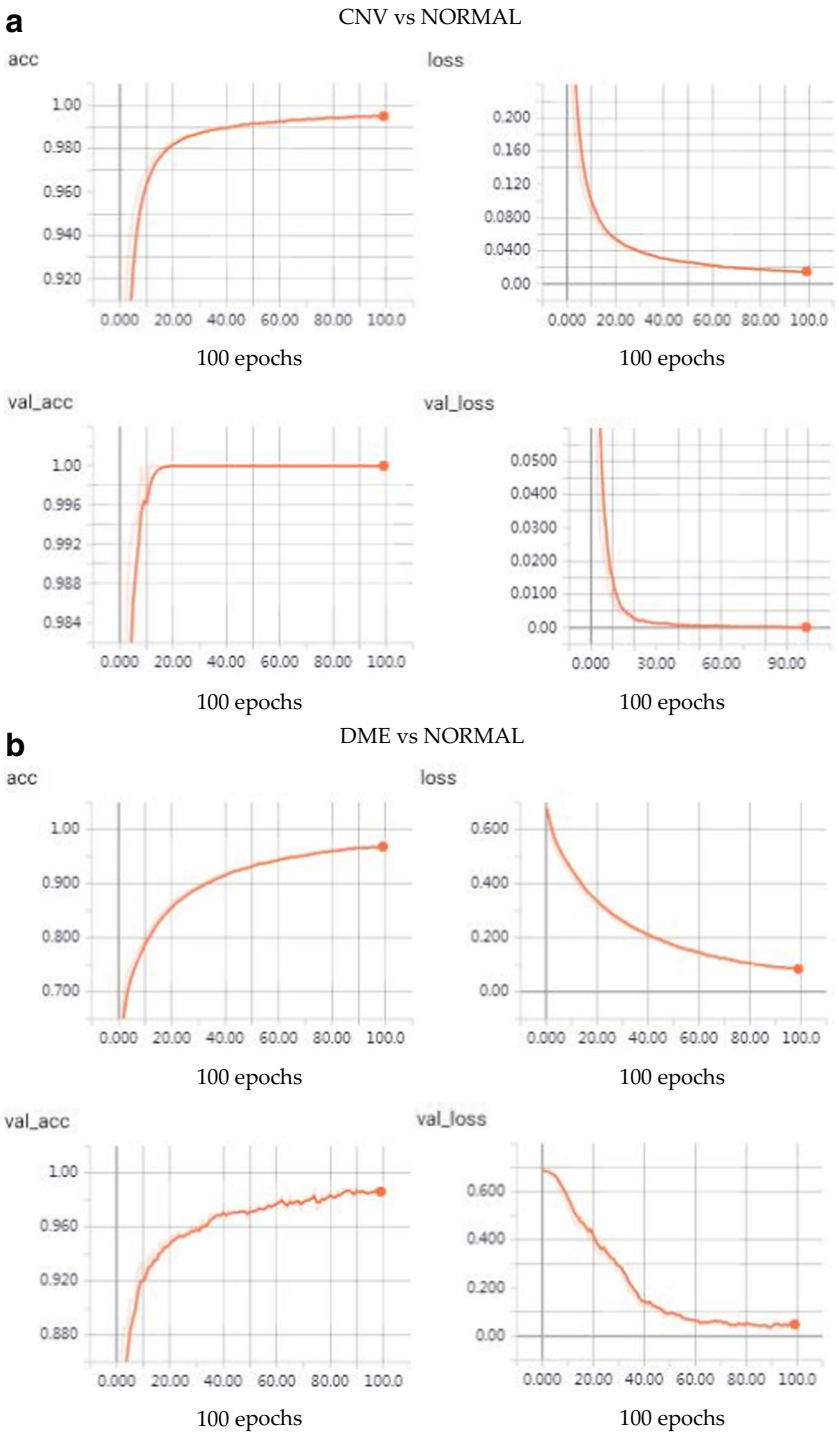

Fig. 7 Plots showing binary performance on the training and validation datasets using TensorBoard. CNV: choroidal neovascularization; DME: diabetic macular edema; Acc: accuracy on the training dataset; loss: loss on the training dataset; val_acc: accuracy on the validation dataset; val_loss: loss on the validation dataset

spent less than 2 h in finishing training and testing on the corresponding datasets. In binary classification, the limited model generated a high accuracy in under 1 h. Consequently, in such conditions, deep transfer learning method based on the VGG-16 network trained on the ImageNet dataset including millions of various medical images can spend less time in generating a high accurate model by retraining layers for retinal OCT images classifications.

Finally, such techniques can be potentially used for a wide range of medical image classification such as chest X-rays and

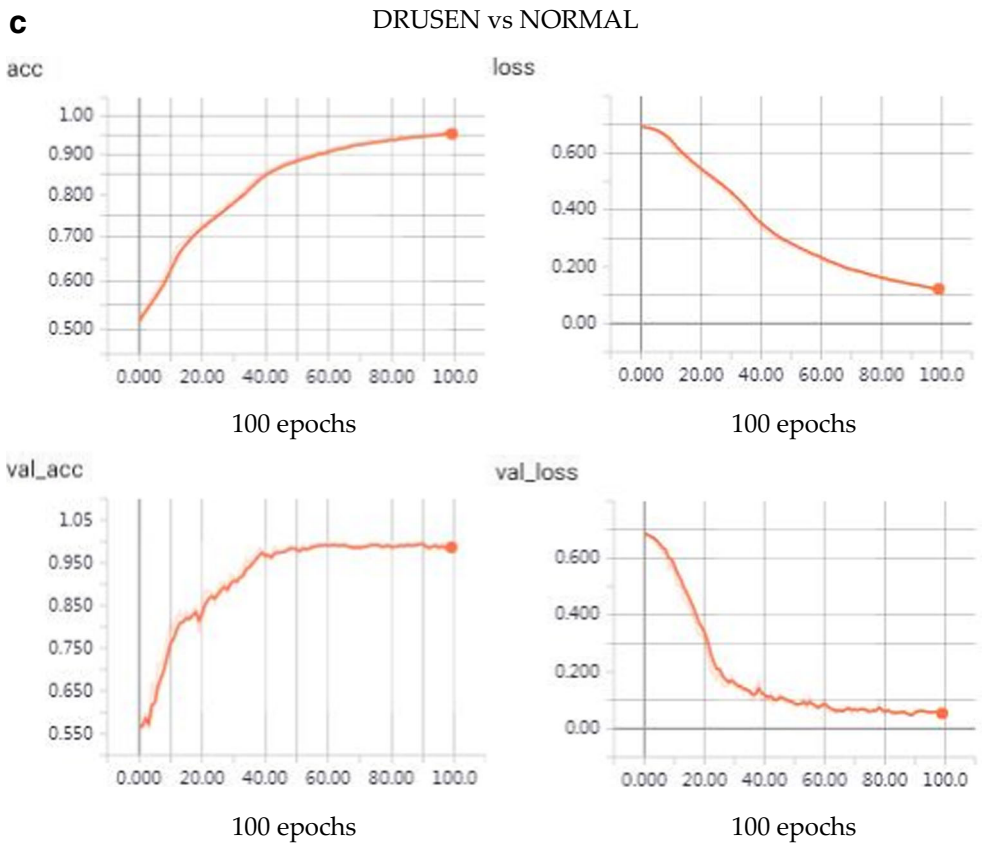

Fig. 7 (continued)

MRI to aid clinicians to make diagnostic decision. The resulting high-accuracy model illustrated that the high generalization capability of this approach can be achieved by learning effectively from increasingly complicated images using a relatively small training dataset. By demonstrating efficacy in detecting retinal disorders, deep transfer learning method based on the VGG-16 network can be extended to a wider range application in biomedical imaging for aiding the clinicians to achieve more precise analysis and suitable treatments.

Conclusion

This paper developed a novel approach to detect retinal disorders accurately using deep transfer learning based on the VGG-16 network. We used a feed-forward approach to fix weights in lower levels of VGG-16 network which had already been optimized in the ImageNet dataset, while retrained the weights of the upper levels of VGG-16 network with back propagation in our built retinal OCT image dataset. Then, the

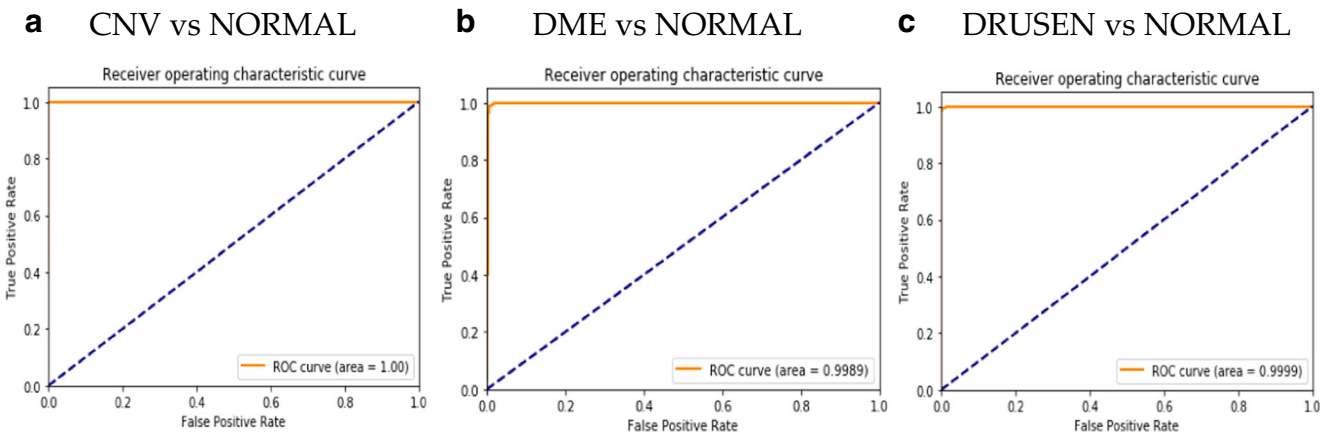

Fig. 8 Receiver operating characteristic (ROC) curves for binary classification. CNV: choroidal neovascularization; DME: diabetic macular edema

performance in distinguishing between AMD and DME was demonstrated on the retinal OCT image dataset, and further compared to that of human experts.

The performance of the proposed approach in diagnosing retinal OCT images manifested superior compared to results obtained by using modern common approaches, and was in accordance with that of ophthalmologist with clinical experience. Furthermore, there was no need for engineered features extracted from retinal OCT image dataset for the diagnosis. Also, the proposed approach was fully autonomous. Therefore, it can be potentially applied to clinics to assist ophthalmologists to make a diagnostic decision.

Given that the prediction of artificial neural network was used as the clinical support, the future work of this study is to translate the proposed approach into software that can be used in medical centers by experts, act as a screening tool and provide a second opinion.

Funding This work was supported by the National Key Research and Development Program of China (2016YFF0101400) and the National Natural Science Foundation of China (51675321).

Compliance with ethical standards

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. For this type of study, formal consent is not required.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Mariotti SP (2012) Global data on visual impairments 2010. http://www.who.int/blindness/GLOBALDATAFINALforweb.pdf . Accessed 6 Jun 2018
2. Bourne RRA, Flaxman SR, Braithwaite T et al (2017) Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health 5:e888–e897. https://doi.org/10.1016/S2214-109X(17)30293-0 Accessed 6 Jun 2018
3. Bourne RRA, Jonas JB, Flaxman SR et al (2014) Vision loss expert Group of the Global Burden of Disease Study, prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990–2010. Br J Ophthalmol 98:629–638. https://doi.org/10.1136/bjophthalmol-2013-304033
4. Pedro RA (2013) Current status in diabetic macular edema treatments. World J Diabetes 4:165–169. https://doi.org/10.4239/wjd.v4.i5.165
5. Koh JEW, Ng EYK, Bhandary SV et al (2018) Automated retinal health diagnosis using pyramid histogram of visual words and fisher vector techniques. Comput Biol Med 92:204–209. https://doi.org/10.1016/j.compbiomed.2017.11.019
6. Ferrara N (2010) Vascular endothelial growth factor and age-related macular degeneration: from basic science to therapy. https://doi.org/10.1038/nm1010-1107 . Accessed 6 Jun 2018
7. Varma R, Bressler NM, Doan QV et al (2014) Prevalence of and risk factors for diabetic macular edema in the United States. Jama Ophthalmol 132:1334–1340. https://doi.org/10.1001/jamaophthalmol.2014.2854
8. Brown DM, Kaiser PK, Michels M et al (2006) Ranibizumab versus verteporfin for neovascular age-related macular degeneration. N Engl J Med 355:1432–1444. https://doi.org/10.1056/NEJMoa062655
9. Keane PA, Patel PJ, Liakopoulos S et al (2012) Evaluation of age-related macular degeneration with optical coherence tomography. Surv Ophthalmol 57:389–414. https://doi.org/10.1016/j.survophthal.2012.01.006
10. Abràmoff MD, Garvin MK, Sonka M (2010) Retinal imaging and image analysis. IEEE Rev Biomed Eng 3:169–208. https://doi.org/10.1109/RBME.2010.2084567
11. Drexler W, Fujimoto JG (2008) State-of-the-art retinal optical coherence tomography. Prog Retin Eye Res 27:45–88. https://doi.org/10.1016/j.preteyeres.2007.07.005
12. Bressler NM (2002) Early detection and treatment of neovascular age-related macular degeneration. http://www.jabfm.org/content/15/2/142.full.pdf . Accessed 6 Jun 2018
13. Ciulla TA, Amador AG, Zinman B (2003) Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies. http://care.diabetesjournals.org/content/diacare/26/9/2653.full.pdf . Accessed 6 Jun 2018
14. Srinivasan PP, Kim LA, Mettu PS et al (2014) Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. Biomedical optics express 5:3568–3577. https://doi.org/10.1364/BOE.5.003568
15. Liu YY, Chen M, Ishikawa H et al (2011) Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. Med Image Anal 15:748–759. https://doi.org/10.1016/j.media.2011.06.005
16. Alsaih K, Lemaître G, Vall JM et al (2016) Classification of sd-oct volumes with multi pyramids, lbp and hog descriptors: application to dme detections. https://doi.org/10.1109/EMBC.2016.7590956 . Accessed 6 Jun 2018
17. Abràmoff MD, Lou Y, Erginay A et al (2016) Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 57:5200–5206. https://doi.org/10.1167/iovs.16-19964
18. Schlegl T, Waldstein SM, Bogunovic H et al (2017) Fully automated detection and quantification of macular fluid in OCT using deep learning. Ophthalmology 125:549–558. https://doi.org/10.1016/j.ophtha.2017.10.031
19. Burlina P, Freund DE, Joshi N et al (2016) Detection of age-related macular degeneration via deep learning. https://doi.org/10.1109/ISBI.2016.7493240 . Accessed 6 Jun 2018
20. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12:629–639. https://doi.org/10.1109/34.56205
21. Lee JG, Jun S, Cho YW et al (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18:570–584. https://doi.org/10.3348/kjr.2017.18.4.570
22. Zhao B, Lu H, Chen S et al (2017) Convolutional neural networks for time series classification. J Syst Eng Electron 28:162–169. https://doi.org/10.21629/JSEE.2017.01.18
23. Kermany DS, Goldbaum M, Cai W et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172:1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010