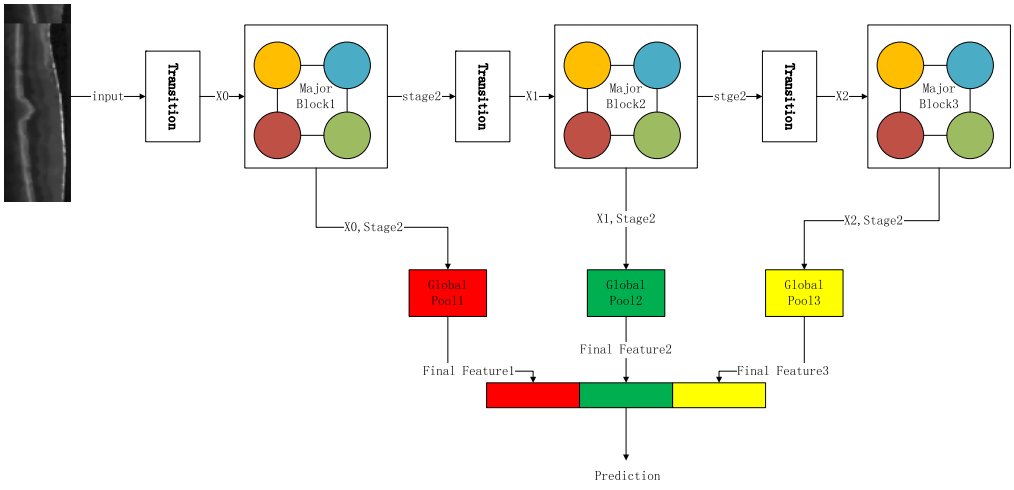
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

On OCT Image Classification via Deep Learning

Volume 11, Number 5, October 2019

Depeng Wang
Liejun Wang



DOI: 10.1109/JPHOT.2019.2934484

On OCT Image Classification via Deep Learning

Depeng Wang [1] **and Liejun Wang** [2]

[1]College of Software Engineering, Xinjiang University, Urumqi 830046, China
[2]College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

DOI:10.1109/JPHOT.2019.2934484

Manuscript received July 21, 2019; accepted August 7, 2019. Date of publication August 12, 2019; date of current version August 27, 2019. This work was supported in part by the National Science Foundation of China under Grant 61771416, in part by the Creative Research Groups of Higher Education of Xinjiang Uygur Autonomous Region under Grant XJEDU2017T002, and in part by the CERNET Innovation Project under Grants NGII20180201, NGII20170325, and NGII20180320. Corresponding author: Liejun Wang (e-mail: wljxju@xju.edu.cn).

Abstract: Computer-aided diagnosis of retinopathy is a research hotspot in the field of medical image classification. Diabetic macular edema (DME) and age-related macular degeneration (AMD) are two common ocular diseases that can result in partial or complete loss of vision. Optical coherence tomography imaging (OCT) is widely applied to the diagnosis of ocular diseases including DME and AMD. In this paper, an automatic method based on deep learning is proposed to detect AME and AMD lesions, in which two publicly available OCT datasets of retina were adopted and a network model with effective feature of reuse feature was applied to solve the problem of small datasets and enhance the adaptation to the difference of different datasets of the approach. Several network models with effective feature of reusable feature were compared and the transfer learning on networks with pre-trained models was realized. CliqueNet achieves better, classification results compared with other network models with a more than 0.98 accuracy and 0.99 of area under the curve (AUC) value finally.

Index Terms: Deep learning, optical coherence tomography, diabetic macular edema, age-related macular degeneration automated diagnosis, computer-aided diagnosis.

1. Introduction

The macular area is an important area of the retina, which is related to fine vision, color vision, and other visual functions. The vision will be negatively affected once lesions occur in the macular area. [1]. Senile macular degeneration, also known as age-related macular degeneration (AMD), is the aging change in the macular area, which occurs mainly in people over 45 years old and the prevalence rate of AMD increases with age, making it one of the primary causes of blindness in the elderly [2]. Diabetic macular edema (DME) is a major complication of diabetes in the eyes, which is the leading cause of blindness in young adults in developed countries [3]. Therefore, early detection is important for the treatment of AMD and DME [4], [5].

OCT is an imaging technique that detects the back reflection of the incoming weak coherent light or several scattered signal in different levels of the biological tissue based on the fundamental of weak coherence light interference, thus the two or three dimentional structural image of the biological tissue can be obtained through scanning [6]. OCT is widely used in the diagnosis of AMD and DME.

Automated detection and diagnosis of retinopathy can assist clinicians in diagnosis and treatment, which is conducive to the early detection and treatment of patients, saves medical resources and helps to reduce the cost of treatment of patients [7]–[9]. In addition, the automatic analysis of medical images through the existing image analysis and processing technologies has become an important research field [10]. There are two problems in retinal OCT images analysis. First, the retinal OCT image public dataset used for research is smaller in number scale compared with the dataset in the fields of traditional image analysis. Second, there are huge differences in the quality and size of retinal OCT the images collected in different tissues (huge differences also exist in the retinal OCT images encountered by doctors in clinical medicine).

In this paper, we propose to use the characteristics of the network structure to reduce the dependence on the size of the dataset and enhance the adaptability to the differences among different datasets. Some network models with effective features of reusable features were selected and trained. For the selected network models, corresponding comparative evaluation was conducted and transfer learning has been achieved for the network with the pre-training model. And the parameters of each network model was tuned to make it achieve better results. In order to verify and compare the advantages and disadvantages of the network model with reusability function with the traditional network model and method, the transfer learning of VGG16 [11], VGG19 [11], Inception-v3 [12] and the classification method using SVM [13] were realized. The models involved in the training evaluation were as follows: CliqueNet [14], DPN92 [15], DenseNet 121 [16], ResNet 50 [17], ResNext101 [18]. The rest of this paper is arranged as follows: In Section 2, the related work was introduced. In Section 3, the approach we propose was illustrated. See Section 4 for experimental results and analysis. Finally, the conclusion of our experiment has been drawn in Section 5.

2. Related Work

With the rise of machine learning, more people are beginning to study the application of machine learning algorithms and various neural networks in medical image analysis.

In 2014, Srinivasan *et al*. [19] applied machine learning to the diagnosis and analysis of AMD and DME diseases based on OCT images. They extracted HOG features of OCT images and used support vector machines (SVM) for classification and recognition. They also created a public OCT dataset from Duke, Harvard, and Michigan University to facilitate future researches. In 2016, Wang *et al*. [20] used the public dataset provided by Srinivasan *et al*. to conduct the classification analysis of OCT images and performed recognition classification by employing the linear configuration pattern (LCP) feature of OCT image and sequential minimal optimization (SMO) algorithm. It is worth mentioning that multi-scale method was used in the extraction of the feature so that the LCP features can be calculated on multiple OCT scales, which makes the local information and the overall information of the OCT image complementary in the extracted features. In 2017, Rasti *et al*. [21] not only studied the classification and recognition of OCT images using the public dataset provided by Srinivasan *et al*., but also established another public OCT dataset from Noor Eye Hospital in Tehran. They used the multi-scale convolution neural network to classify and recognize two OCT datasets and the local and overall information of OCT image were also combined.

In 2017, Karri *et al*. [22] also used the public OCT dataset proposed by Srinivasan *et al*. to classify and recognize OCT images. They fine-tuned the pre-trained convolution neural network (CNN) GoogleLeNet to train and classify OCT images so that the neural network can be trained and good results can be achieved by using limited data. In 2019, Feng *et al*. [23] also used transfer learning to classify OCT images in order to reduce the dependence on dataset size. They fine-tuned and trained the pre-trained VGG16. In 2019, Juan *et al*. [24] conducted a full comparative analysis of the recognition and diagnosis performance of VGG19, GoogLeNet, ResNet50 and DeNet neural networks on glaucoma diseases by color fundus image. Cheng *et al*. [25] proposed a deep hashing algorithm based on ResNet 50 to perform image retrieval and classification tasks. Although satisfying results have been realized in the diagnosis of AMD and DME by traditional machine learning method [19], [20], higher accuracy is hard to be obtained. In [21] the network

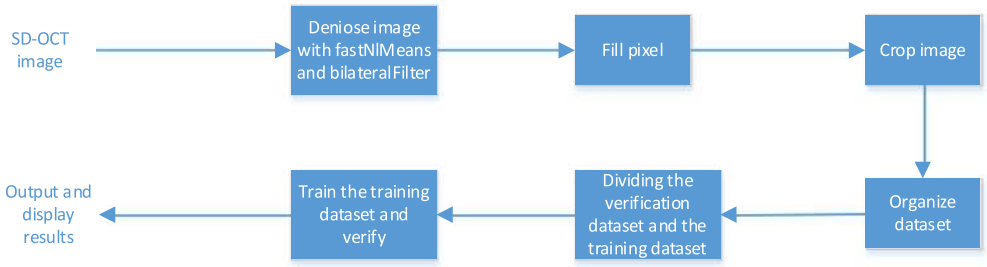


Fig. 1. Overview of the algorithm for classifying OCT images.

structure was improved and the local and overall information was combined with the multi-scale method to extract and retain the effective features as much as possible, thus good results was achieved. In [22] and [23], transfer learning was proposed to reduce the training parameters of neural networks, which reduces the dependence on the size of datasets, but such method has poor adaptation to the differences among different datasets.

Combining the ideas in [19], [20], [21], [23], this study proposes to use the characteristics of the network structure to reduce the dependence on the size of the dataset and enhance the adaptability to the differences among different datasets. Meanwhile, related research has been performed on the diagnosis and analysis of AMD and DME by two published OCT datasets proposed by Srinivasan *et al.* and Rasti *et al.*. We demonstrate the method of using the characteristics of network structure to solve the problem of scarcity of data and combining with the previous experience of transfer learning method for network training, thus good results have been achieved and the adaptability of each network to different datasets has been verified.

3. Material and Methods

In this section, the classification methods and the validation methods were analyzed. First, the whole process of the classification method for retinal OCT images is as follows: (1) FastNIMeans [26] and bilateralFller [27] methods were used for denoising. (2) Pixel filling was performed and finally the image was cropped to complete the preprocessing. Then different datasets were organized and divided into verification set and training set. Finally, the datasets were sent to the neural network model for training and output. All the processes of our methods are shown in Fig. 1.

3.1 Datasets

The datasets used in this study were two public datasets. The dataset 1 was provided by Srinivasan *et al.*, and obtained in Duke University, Harvard University, and the University of Michigan via Spectralis SD-OCT imaging. The dataset 2 was provided by Rasti *et al.*, and obtained at the Noor Eye Hospital in Tehran using Spectralis SD-OCT imaging. There are 3231 pictures in the dataset 1, of which 1101 are DME, 723 are AMD, and 1407 are normal. There are 5,084 pictures in the dataset 2, of which 1185 are DME, 1524 are AMD, and 2375 are normal. However, there are many eyeball images contained in dataset 2, not just retina images, so dataset 2 needs to be properly screened. Since the two public datasets are different in quality and size, they were separated for training and evaluation, which can also verify the adaptability of different neural network models to the difference among datasets differences. An example of two datasets is shown in Fig. 2.

3.2 Image Pretreatment

The same 2 procedures were adopted in the pre-processing of two datasets. But due to the difference in quality between the two datasets, some fine-tuning was made in the process of

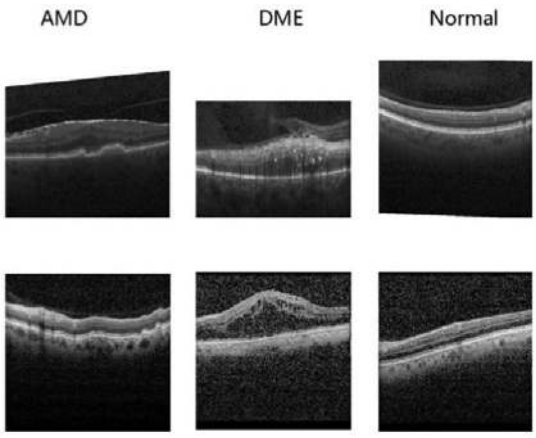


Fig. 2. The first line of the sample image comes from the dataset 1 and the second line comes from dataset 2.

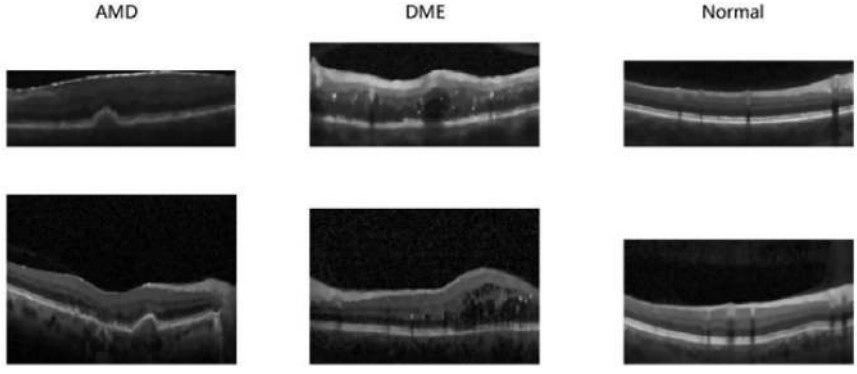


Fig. 3. The first line of the sample image comes from the dataset 1. The second line comes from the dataset 2.

cropping the dataset 2 to make the preprocessing result visually close to the result of the dataset 1. For dataset 1, first, FsatNlMeans method and bilateralfilter method were used for denoising and preventing subsequent steps from being interfered by the image noise. Then pixels were filled in the blanks of each image so that the lower boundary can be distinguished in the cropping. At last, the images were cropped and the upper boundary and lower boundary of the bright part of the retina of each image were detected using the pixel value, and the pixel range from 40 pixels above the upper boundary to 40 pixels below the lower boundary was cut out as our final preprocessing result. Because the image of dataset 2 is relatively dark, thus the difference between the preprocessing of dataset 2 and dataset 1 is that, in the pre-processing of dataset 2, the edge detection method was used to detect the lower boundary, and finally the pixel range from the top of the image to 40 pixels below the lower boundary was cut out. In this way, for dataset 2, the information in the bright part of the retina can be preserved and the loss of the main information caused by excessive cropping of the image can be avoided. The OCT image after the preprocessing is shown in Fig. 3.

3.3 Dataset Organizion and Dividsion

In the division of datasets, 20% of the images were randomly extracted from datasets as the verification set, which is not engaged in training to ensure the reliability of the experimental results,

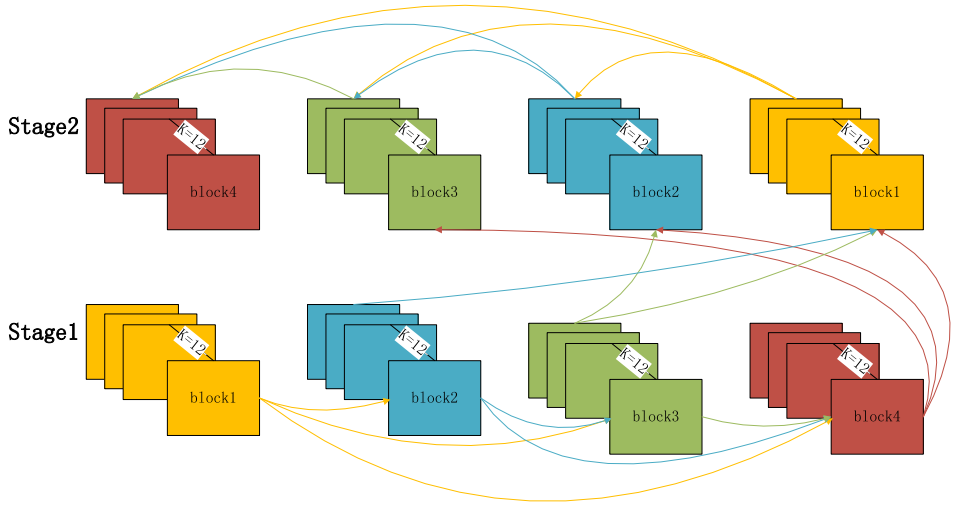


Fig. 4. The output of each block in CliqueNet structure is its block input, among which block 1 means the input block at the beginning of data. k indicates that there are 12 convolution kernels.

and each image was labeled according to the location of the file, where AMD is the class 0, DEM is class 1 and Normal is class 2, and the labels are all in the form of one-hot encoding.

3.4 Training Method

Srinivasan *et al*. [19] used the SVM method to classify AMD, DME, and Normal in dataset 1 to achieve an average accuracy of 95.5%. Karri *et al*. [22] used transfer learning and GoogLeNet for classification based on dataset 1 to achieve an average accuracy of 91.3%. Feng *et al*. [23] used VGG-16 transfer learning method for classifying the dataset they collected and achieved an accuracy rate of 98.6%. Rasti *et al*. [21] used dataset 2 and multi-scale convolutional networks for classification and training to achieve an accuracy of 98.33%. In order to determine whether the hyperparameters we use are appropriate, the accuracy of the two datasets are required to be no less than 91.3% of the average accuracy in the fining tuning of our network models. Since CliqueNet does not have a pre-trained model, so transfer learning was not used. Fine-tuning of the network was conducted according to the features of the OCT dataset and our experimental environment. Three major blocks were adopted, each of which has 4 convolutional blocks, and each convolution block has 12 convolution kernels. The structure of a single major block is shown in Fig. 4.

In the first phase, the forward pass of the feature was carried out (the features extracted by each block was passed into its next block). In the second stage, back pass of the feature was conducted during the forward pass (the feature extracted by each block was passed to the last block). In the third stage, each main block will output the input feature x_i $(i = 0, 1, 2)$ and the features extracted in its stage2 which will be used as the input of the next block, which also plays the role of multi-scale convolution. Finally, the stage2 output features of each block were pooled overall with the input features $x_i (i = 0, 1, 2)$ to generate its final features. The final features of the three main blocks were stitched to generate our final predictive feature, which is called Prediction that generates our final prediction label through a layer of fully connected layers using the softmax activation function. The Prediction generation mode is shown in Fig. 5.

The parameters of CliqueNet were optimized on both datasets and 10 batchsize and Momentum [28] gradient optimization methods were adopted. On dataset 1, a learning rate of 0.1 with a dropout of 0.8 was used, and on dataset 2 a learning rate of 0.01 with a dropout of 0.9 was employed.

ResNet has been fully applied in the study of the diagnosis of glaucoma through color fundus image [24] and good results have been achieved. ResNet is also widely used in traditional image

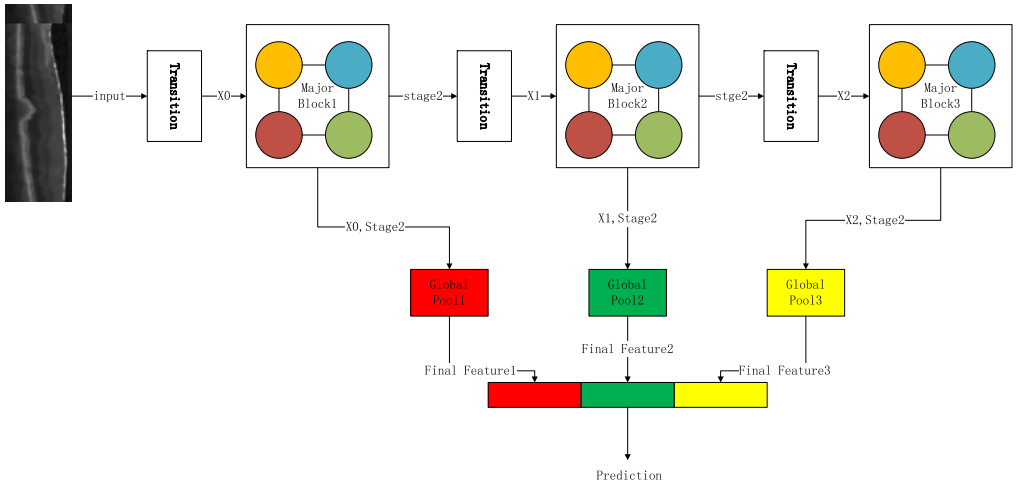


Fig. 5. In the figure, x_i $(i = 0, 1, 2)$ is the input characteristic of each layer, stage2 is the output feature of each block, Final Feature is the final prediction feature of each block, and three Final Features are stitched to generate Prediction. The convolution and pooling operations in the Transition block in the figure play the role of feature extraction.

classification and retrieval tasks [29]. The pre-trained model of ResNet50 selected in [24] was chosen for transfer learning. The neural network model and parameters were fine-tuned as follows: the output channel of the last fully connected layer was adjusted to 3, and make its output the probability values of AMD, DEM and Normal. SGD was used as the gradient optimizer with a learning rate of 0.0001.

ResNext introduces Inception's multi-branch structure, transforming ResNet's single convolution into multi-branch convolution [18]. ResNext101's pre-training model was used for transfer learning. The network model and parameters were fine-tuned as follows: the number of the output neuron of the output layer was adjusted to 3 with SGD as the gradient optimizer and the dropout value of 0.5 and the learning rate of 0.0001.

DenseNet has both strength feature propagation and encourage feature reuse [16]. Because of these two characteristics, the pre-training model of DenseNet121 was selected for the transfer training. The fine-tuning of the network model and parameters of DenseNet121 are the same as those of ResNext101.

ResNet enables feature re-usage while DenseNet enables new features exploration which are both important for learning good representations. Dual Path Network shares common features while maintaining the flexibility to explore new features through dual path architectures [15]. DPN92's pre-training model was employed for training in transfer learning. The fine-tuning of the network model and parameters of DPN92 are the same as those of ResNext101.

3.5 Statistical Analysis Method

Because the missed diagnosis and misdiagnosis events will have a serious impact on patients, so the precision, accuracy and recall rate are important indexes that should be considered in the evaluation. The confusion matrix and ROC curve were drawn with precision, accuracy, and recall rate as the evaluation criteria [30]. The verification images were randomly selected from the verification set. The precision, accuracy, and recall rate are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

TABLE 1
Acc Results Achieved by Different Neural Networks and Methods

Acc	CliqueNet	DPN92	DenseNet121	ResNet50	ResNext101	SVM	VGG16	VGG19	Inception-v3
Dataset 1	0.990	0.996	0.987	0.987	0.973	0.968	0.916	0.982	0.927
Dataset 2	0.986	0.958	0.942	0.952	0.948	0.810	0.863	0.951	0.853

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

True positive (TP) refers to the total number of patients correctly diagnosed; False positive (FP) represents the total number of non-patients wrongly diagnosed as patient; False negative (FN) means the total number of patients diagnosed as non-patients; True Negative (TN) is the total number of non-patients correctly diagnosed.

Accuracy is the ratio of the number of correctly classified samples to the total number of samples for a verification dataset. Precision is the ratio of the number of correctly classified positive samples to all correctly classified samples. Recall ratio is used to indicate the ratio of the number of correctly classified positive samples in the classifier to the total positive samples.

3.6 Experimental Environment

The experimental code is based on python3.6 (based on pytorch1.0.1, keras2.2.4, tensorflow 1.12.0). The Linux system is used with a version of Ubuntu16.04. The GPU used an NVIDIA 2080ti with 11G memory for training. The experimental used NVIDIA Cuda v 9.2, and cuDNN v 7.3.1 accelerated library.

4. Experiment and Results

We performance of CliqueNet, DPN92, DenseNet121, ResNet50, and ResNext101 neural network in two public OCT datasets was evaluated and discussed based on our evaluation methodology. The comparative neural network VGG16, VGG19, Inception-v3, and comparative algorithm SVM were added.

4.1 Comparison With Previous Models

The performance of our network model on two OCT public datasets was evaluated and compared with that of SVM method and Vgg16, Vgg19 and Inception-v3 neural networks (all trained using transfer learning) according to accuracy (Acc) criteria. The processing procedures in the SVM method are the same as in [1]. Firstly, HOG features were extracted and then sent to SVM for training. The results are shown in Table 1.

As shown in Table 1, all of our methods have achieved good results on public OCT dataset 1, but the Acc results of the methods previously used on our public OCT dataset 2 have declined considerably. CliqueNet and DPN92 achieved significantly higher accuracy results than the method previously used, which indicates that these neural network models with effective feature of reusability are more adaptable to the difference among different datasets, and the effective feature of reusable characteristics can also help the neural network to achieve higher accuracy.

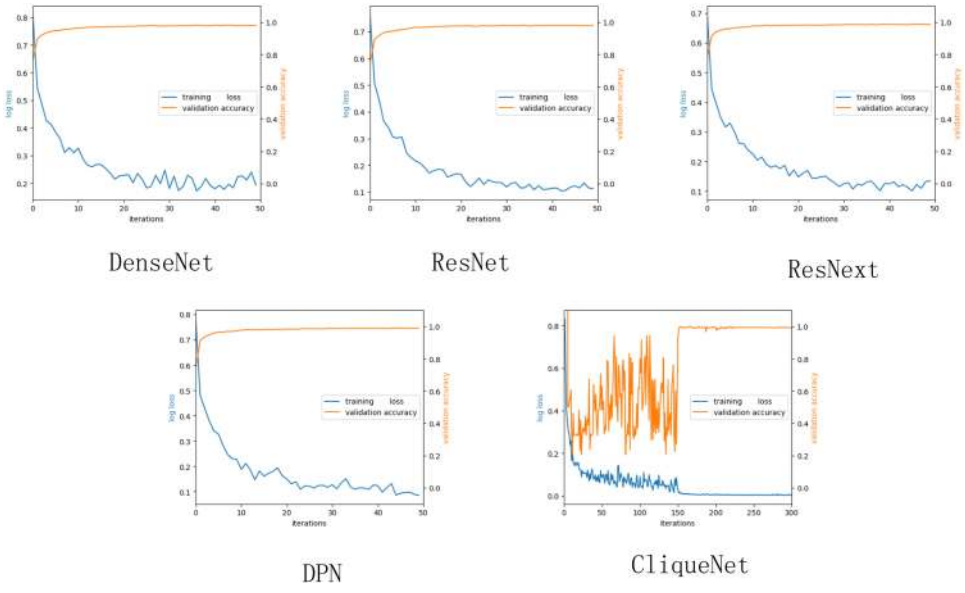


Fig. 6. We used the Matplotlib library to draw our Acc and Loss curves. Acc is the accuracy obtained during the verification, and Loss is the Loss obtained during the training.

4.2 Evaluation of Selected Models

In the actual clinical diagnosis, various evaluation diagnoses are usually included, and the medical images used tend to have different imaging qualities or even different sources. So we used two different public OCT datasets with different imaging quality and sources to train and evaluate our neural network models.

The training loss and accuracy curve of CliqueNet, DPN, DenseNet, ResNet, ResNext neural network on two public OCT datasets are shown in Fig. 6 and Fig. 7. The training is stopped to avoid overfitting of models when Acc no longer improves and the training loss basic converges.

On the dataset 1, DenseNet, ResNet, and DPN have a convergence value after 20 steps of training. ResNext has a convergence value in about 30 steps of training, CliqueNet has a convergence value after about 150 steps of training. CliqueNet's accuracy and training loss curves oscillated before convergence.

On dataset 2, DenseNet, ResNet, and DPN have a convergence loss value at about 30 steps of training. ResNext has a convergence loss value at about 40 steps of training, and CliqueNet has a convergence loss value at about 180 steps of training. In dataset 2, the Acc and loss curves of CliqueNet also greatly oscillated before convergence. The accuracy scores of other network models other than CliqueNet on dataset 2 have all declined, CliqueNet and DPN scored the top two in the accuracy score. Therefore, we conclude that dataset 2 is more complex and more difficult to fit than dataset 1.

It can be seen that the training loss of other network models except for CliqueNet quickly converges on the two datasets and the accuracy no longer improves, but CliqueNet basically converges after 180 steps of training. Before this, CliqueNet's accuracy significantly oscillated. The reasons for this situation are as follows: (1) CliqueNet does not use the pre-training model to train from the beginning, so it is difficult to achieve convergence; (2) CliqueNet uses batch training during training so there will be oscillation; (3) CliqueNet makes multiple reuses of the effective features within a batch and is likely to be overfitting within a batch. Therefore, in order to prevent overfitting in dataset 2, its dropout value was adjusted to 0.9.

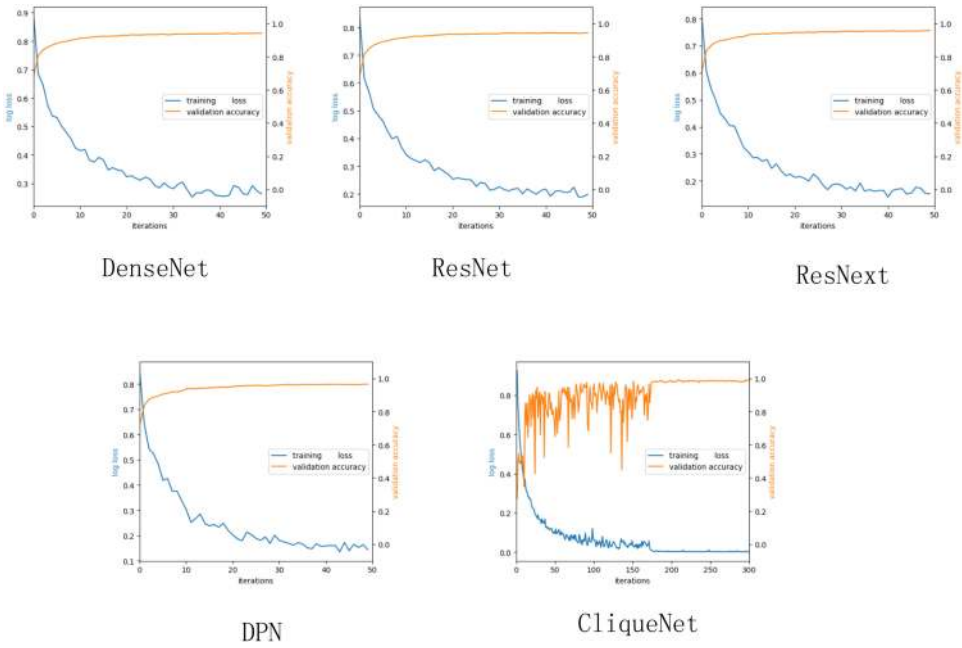


Fig. 7. We used the Matplotlib library to draw our Acc and Loss curves. Acc is the accuracy obtained during the verification, and Loss is the Loss obtained during the training.

TABLE 2
The Results Achieved by Different Neural Networks on Dataset 1

Dataset 1	DenseNet	ResNet	DPN	ResNext	CliqueNet
Acc	0.987	0.987	0.996	0.973	0.990
Precision AMD	1.000	0.992	0.993	0.979	0.956
Precision DME	0.972	0.985	1.000	0.980	1.000
Precision Normal	0.992	0.985	0.996	0.965	1.000
Recall AMD	0.971	0.992	1.000	0.985	1.000
Recall DME	0.990	0.981	0.990	0.943	0.990
Recall Normal	0.992	0.989	1.000	0.989	0.986

After the training, each network model's scores on the randomly selected verification set are shown in Table 2 and Table 3, and their ROC curve [30] is shown in Fig. 8 and Fig. 10 and the confusion matrix is shown in Fig. 9 and Fig. 11.

DenseNet, ResNet, DPN, ResNext, and CliqueNet all achieved good results on the publicly OCT dataset 1, in which CliqueNet and DPN's accuracy, precision, and recall scored the top two, which shows that the network models we selected can fit the characteristics of dataset 1 very well. Although the dataset 1 is small, it still has good results. This confirms that the network model with feature reuse features can fit the features of smaller datasets very well. CliqueNet and DPN make more thorough use of effective features than other network models, so the results are better.

The scores of Acc, precision, and recall of DenseNet, ResNet, and ResNext all greatly declined on the public OCT dataset 2. The DPN network also experienced a large drop in the three ratings of accuracy, precision, and recall, but each score is above 0.9, which demonstrates that dataset 2

TABLE 3
The Results Achieved by Different Neural Networks on Dataset 2

Dataset 2	DenseNet	ResNet	DPN	ResNext	CliqueNet
Acc	0.942	0.952	0.958	0.948	0.986
Precision AMD	0.894	0.927	0.925	0.961	0.979
Precision DME	0.953	0.990	0.995	0.953	1.000
Precision Normal	0.971	0.953	0.963	0.939	0.985
Recall AMD	0.952	0.959	0.965	0.931	1.000
Recall DME	0.907	0.889	0.907	0.907	0.938
Recall Normal	0.954	0.980	0.978	0.980	1.000

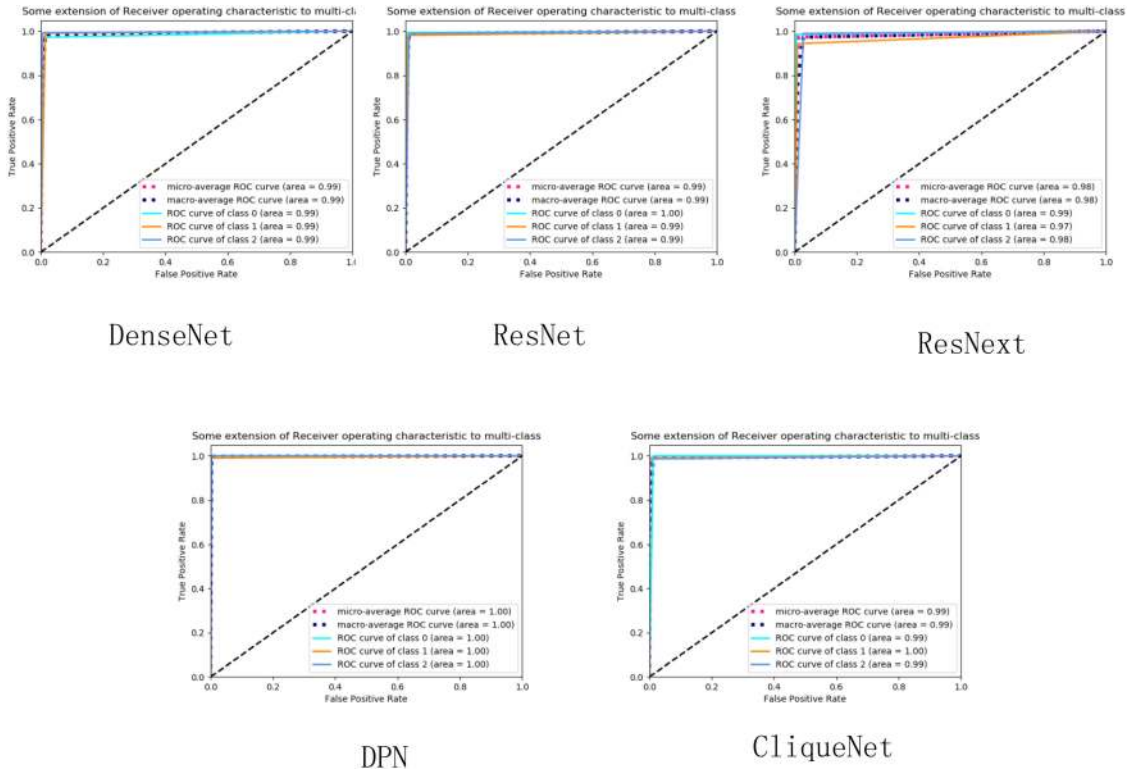


Fig. 8. The micro-average ROC curve is obtained by the micro method in the sklearn.metrics.roc _auc_score function. The macro-average ROC curve is obtained by the macro method in the sklearn.metrics.roc_auc_score function. Class 0, class 1, and class 2 in the figure represent AMD, DME, and Normal, respectively.

is more complex and more difficult to fit than dataset 1. CliqueNet's accuracy, precision, and recall scores are higher than other networks and each score did not decrease significantly. This shows that CliqueNet is highly adaptable to differences among different datasets. CliqueNet make more thorough use of effective features than other network models, so the results are better. This also shows that a network make more thorough for effective feature reuse can be more adaptable to different differences among different datasets.

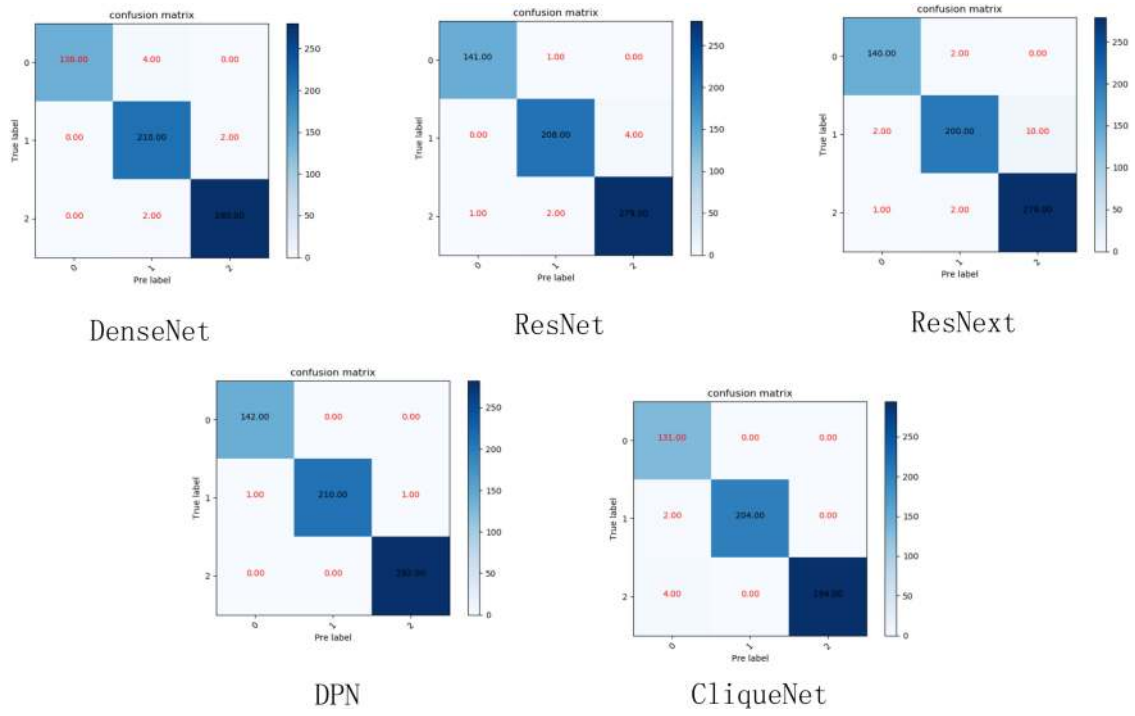


Fig. 9. 0, 1, and 2 in the figure represent AMD, DME, and Normal, respectively.

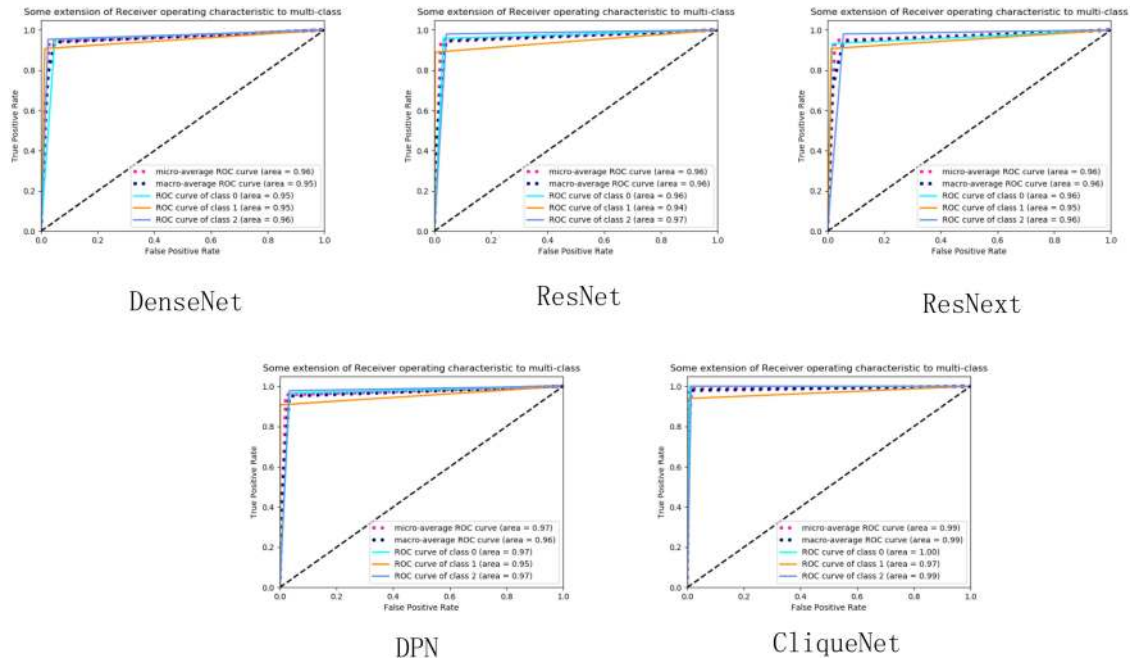


Fig. 10. The micro-average ROC curve is obtained by the micro method in the sklearn.metrics.roc _auc_score function. The macro-average ROC curve is obtained by the macro method in the sklearn.metrics.roc_auc_score function. Class 0, class 1, and class 2 in the figure represent AMD, DME, and Normal, respectively.

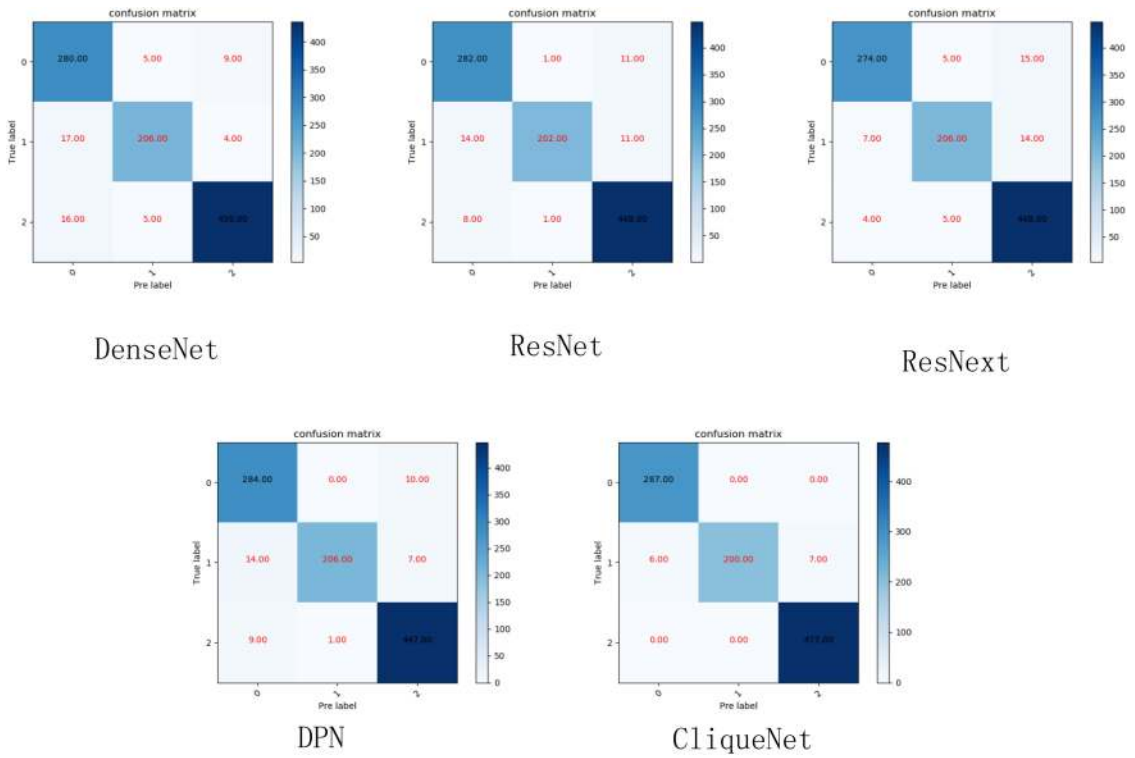


Fig. 11. 0, 1, and 2 in the figure represent AMD, DME, and Normal, respectively.

The AUC values obtained by DenseNet, ResNet, DPN, ResNext, and CliqueNet are all above 0.96 on the public OCT dataset 1 and the average AUC values of three types are all above 0.97, which shows that the network model we trained on dataset 1 can well balance the precision and recall indicators.

The AUC values achieved by DenseNet, ResNet, DPN, ResNext, and CliqueNet on publicly OCT dataset 1 are all above 0.93, and the average AUC values of the three categories are all above 0.96. Although the AUC values of DenseNet, ResNet, DPN fell, yet the values are not much different from those of 0.96 AUC and 0.97 mean AUC value of three types. The AUC of CliqueNet does not decrease, which indicates that the high degree of effective feature reuse can balance the precision and recall indexes well.

Our verification experiment proved that CliqueNet, except for the accuracy score on the first dataset slightly less than DPN, CliqueNet, achieved higher scores than other neural network models. On the recall indicator, CliqueNet achieved a three-category average recall rate of 0.992 and 0.979 on the two OCT datasets. This shows that the use of CliqueNet for the diagnosis of AMD and DEM disease can better prevent missed diagnosis. On the precision indicator, CliqueNet achieved a three-class average precision of 0.985 and 0.988 on the two OCT datasets, which demonstrates that the use of CliqueNet can prevent misdiagnosis of AMD and DEM. In Fig. 8 and Fig. 10, it is shown that CliqueNet achieved a higher AUC value, indicating that the trained model of CliqueNet can better balance the recall and precision indicators. Regarding DPN, although the scores on the dataset2 have declined, the accuracy, recall, precision, and AUC it achieved on the two public OCT datasets are acceptable. And when DPN has a pre-trained model, it takes up fewer resources in training than CliqueNet and is simpler to implement than CliqueNet. Therefore, DPN should also be considered in practical applications.

5. Discussion

In this paper, the application of the neural network models with effective feature reuse features based on the retinal OCT datasets in the diagnosis of AMD, DME was tested and evaluated. The optimization strategies of different neural network models, the effects of different neural network models on final diagnostic recognition and the adaptability of different neural network models to different datasets were studied. Two publicly available OCT datasets were used to evaluate the performance of the neural networks we selected and the impact of differences among datasets. Five neural network structures were tested and evaluated (DenseNet121, ResNet50, DPN92, ResNext101, CliqueNet) and VGG16, VGG19, inception-V3 neural networks, and SVM methods were added to improve experimental comparisons. Accuracy, recall, precision, AUC were used as evaluation criterias to show the ROC curve and confusion matrix of each neural network. After the testing, CliqueNet achieves the best score, with a 0.99/1.00 three-types average AUC, a 0.992 three-types average recall rate, 0.985 three-types average precision, and a 0.990 of accuracy rate on dataset 1 and a 0.99/0.99 of three-types average AUC, a three-types average recall of 0.988, an average precision of 0.979, and an accuracy of 0.986. The effects achieved by CliqueNet on different datasets are similar. Our experiments have proved that the features of feature reuse are beneficial to improve the performance of the neural network models and the adaptability to differences among different datasets. After analysis and comparison, two neural network models CliqueNet and DPN were obtained, which have better effects on different datasets. CliqueNet and DPN can be applied to actual clinical diagnosis. CliqueNet training is complex but result is better. Although DPN is not as effective as CliqueNet on dataset 2, the training method is easy, showing that CliqueNet can better help doctors prevent misdiagnosis and missed diagnosis. We believe that when dealing with medical datasets with less data, some neural network models with effective feature of reuse can be taken into account, but also, data amplification through data enhancement and generative adversarial networks (GAN) can be considered.

References

[1] R. R. A. Bourne *et al.*, "High-income countries and in Eastern and Central Europe: 1990-2010," *Brit. J. Ophthalmology*, vol. 98, no. 5, pp. 629–638, 2014.

[2] D. S. Friedman *et al.*, "Prevalence of age-related macular degeneration in the United States," *Arch Ophthalmol.*, vol. 122, no. 4, pp. 564–572, 2014.

[3] P. Romero-Aroca, "Current status in diabetic macular edema treatments," *World J Diabetes*, vol. 4, pp. 165–169, 2013.

[4] N. M. Bressler, "Early detection and treatment of neovascular age-related macular degeneration," *Amer. Board Family Pract.*, vol. 15, pp. 142–152, 2002.

[5] A. N. Kollias and M. W. Ulbig, "Diabetic retinopathy early diagnosis and effective treatment," *Dtsch. Arztebl. Int.*, vol. 107, pp. 75–84, 2010

[6] D. Huang *et al.*, "Optical coherence tomography," *Science*, vol. 254, pp. 1178–1181, 1991

[7] J. C. Javitt, L. P. Aiello, Y. Chiang, F. L. Ferris, J. K. Canner, and S. Greenfield, "Preventative eye care in people with diabetes is cost-saving to the federal government: implications for health-care reform," *Diabetes Care*, vol. 17, pp. 909–917, 1994.

[8] H. M. J. Kraus, M. Porta, and H. Keen, "Diabetes care and research in Europe: The St Vincent Declaration action programme," *Diabete. Metab.*, vol. 18, pp. 334–377, 1992.

[9] T. J. Hendra and A. J. Sinclairm, "Improving the care of elderly diabetic patients: The final report of the St Vincent Joint Task Force for Diabetes," *Age Ageing*, vol. 26, pp. 3–6, 1997.

[10] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Anal.*, vol. 42, pp. 60–88, 2017.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv*: 1409.1556, Sep. 2014.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 2818–2826.

[13] J. Platt, B. Scholkopf, C. Burges, and A. Smola, "Fast training of support vector machines using sequential minimal optimization," in *Adv. Kernel Methods-Support Vector Learn.*, MIT Press, 1998.

[14] Y. Yang, Z. Zhong, T. Shen, and Z. Lin, "Convolutional neural networks with alternately updated clique," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 2413–2422.

[15] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 4470–4478.

[16] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 2261–2269.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.
[18] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 5987–5995.
[19] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 5, pp. 3568–3577, 2014.
[20] Y. Wang *et al.*, "Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images," *Biomed. Opt. Exp.*, vol. 7, pp. 4928–4940, 2016
[21] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Medical Imag.*, vol. 37, no. 4, pp. 1024–1034, Apr. 2018.
[22] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomed. Opt. Exp.*, vol. 8, pp. 579–592, 2017.
[23] F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, "Fully automated detection of retinal disorders by image-based deep learning," *Graefes Arch. Clin. Exp. Ophthalmol.*, vol. 257, pp. 495—505, 2019.
[24] J. J. Gómez-Valverde *et al.*, "Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning," *Biomed. Opt.*, vol. 10, pp. 892–913, 2019.
[25] S. Cheng, L. Wang, and A. Du, "Histopathological image retrieval based on asymmetric residual hash and DNA coding," *IEEE Access*, vol. 7, pp. 101388–101400, 2019.
[26] H. Oulhaj, A. Amine, M. Rziza, and D. Aboutajdine, "Noise reduction in Medical Images—comparison of noise removal algorithms-," in *Proc. Int. Conf. Multimedia Comput. Syst.*, Tangier, Morocco, 2012, pp. 344–349.
[27] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Medical Imag.*, vol. 37, no. 4, pp. 1024–1034, Apr. 2018.
[28] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. 30th Int. Conf. Mach. Learn.*, Atlanta, GA, USA, 2013, pp. 1139–1147.
[29] S. Cheng, L. Wang, and A. Du, "An adaptive and asymmetric residual hash for fast image retrieval," *IEEE Access*, vol. 7, no. 1, pp. 78942–78953, Jun. 2019.
[30] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, New York, NY, USA, 2006, pp. 233–240.