2020 12th International Conference on Communication Software and Networks

Attention-Aware Convolutional Neural Network for Age-Related Macular Degeneration Classification

Shanshan Li, Zhi Quan
School of Information Engineering
Zhengzhou University
Zhengzhou, China
e-mail: lisan@gs.zzu.edu.cn, iezquan@zzu.edu.cn

***Abstract*—Though age-related macular degeneration (AMD) poses an important personal and public health burden, studies on AMD is hampered by different approaches to classify AMD. In this paper, we propose convolutional neural networks (CNN) based models for fundus retinal images that classify four types of AMD automatically. We use deep residual network (ResNet50) to extract high-dimensional features and be trained end-to-end to classify AMD. Furthermore, we apply attention mechanism to deep residual network (Atten-ResNet) which enables to further select features adaptively. Experimental results show that comparing to HOG-SVM and Visual Geometry Group (VGG), the ResNet50 based method could achieve 17.2% and 12.1% overall classification accuracy improvement. The Atten-ResNet based method has more 0.4% accuracy improvement than ResNet50 based method.**

Keywords-ResNet; Attention Mechanism; OCT; AMD; Classification

I. INTRODUCTION

Age-related macular degeneration (AMD) refers to a degenerative disease that is the main cause of blindness in the elderly [1]. Regular medical examination of the macular area is the key to effectively avoiding vision loss. At present, macular degeneration is mainly diagnosed by professional ophthalmologists. Such manual analysis is time-consuming and demanding for doctors, which is also prone to yield subjective results. As a non-contact and non-invasive imaging technology, high resolution optical coherence tomography (OCT) imaging technique enables the sensitive detection of multiple retinal cell layers and quantitative assessment of these macular lesions within retina [2]. An automatic and reliable computer assisted OCT image analysis techniques are required for efficient diagnosis of AMD.

In the past decade, researchers have proposed a variety of automatic classification techniques for macular lesions based on OCT images. Traditional classification methods are used to detect macular lesions. It mainly uses multi-scale spatial pyramid (SP) [3], local binary pattern (LBP) histogram [3], multi-scale oriented gradient histogram (HOG) [4], principal component analysis (PCA) [5] as feature extractors, and use support vector machine (SVM), Bayesian as a classifiers.

In recent years, Deep learning [6] [7] has achieved great success in the fields of computer vision and pattern recognition. Convolutional neural network (CNN) [8] is a typical deep learning model that can automatically learn features from big data. Lee [9] used the VGG16 model to implement the AMD classification. Rasti [10] extracted OCT image features based on wavelet transform CNN model. Karri [11] used transfer learning methods GoogleLeNet, a pre-trained model from ImageNet, to perform migration optimization to achieve AMD classification, and Feng [12] realized the AMD classification by using the pre-trained model VGG16.

The attention mechanism in image processing is similar to the visual attention mechanism of the human eyes. It selectively focuses on key areas in the image to obtain the required detailed information and optimizes network performance. Soft attention is a kind of attention mechanism commonly used in CNN. It represents the probability of selecting certain parts from the input images, obtaining attention weight, and uses attention weight to realize feature selection. Chen [13] have applied soft attention to the end-to-end model and achieved good performance.

Reference [14] [15] show that deep CNN has a strong ability to automatically extract low, middle, and high dimensional features. Deep residual network (ResNet) is a typical deep CNN model. Given deep neural networks have not been applied to the field of AMD classification, we propose to use the ResNet as the AMD classification model. Furthermore, considering the effectiveness of the attention mechanism in image processing, we use an attention-based deep residual network (Atten-ResNet) for AMD classification. The proposed classification models automatically extract high-dimensional features and select feature from the OCT images to achieve AMD classification.

II. RELATED WORK

A. Residual Network

In this work, the residual network [14] is applied to the automatic classification of macular degeneration. The basic unit of the residual network is the residual block. The residual block structure is shown in Fig. 1, x denoting the inputs to the first layers. Multiple nonlinear layers can asymptotically approximate residual functions F(x). F(x)+x represents the output. It realizes the construction of deep network through the idea of identity mapping. The residual network structure is composed of convolutional layer, pooling layer and fully connected layer. We use ResNet50 as the basic classification model. The main architecture of the

978-1-7281-9815-6/20/$31.00 ©2020 IEEE

ResNet50 is composed of four residual layers which consist 16 bottleneck residual blocks. The final classification layer consists of the global average pooling layer and the multi-class softmax function layer.

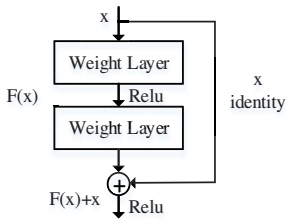


Figure 1. Residual block structure

B. *Residual Attention Network*

The residual attention network (Atten-ResNet) [16] consists of convolutional layer, pooling layer, residual unit and attention module. The network depth reaches multiple layers, and the residual unit adopts the bottleneck design method. The attention module is the core of the residual attention network. The attention module structure is shown in Fig. 2. Each attention module is composed of two parts: feature processing and feature selection.

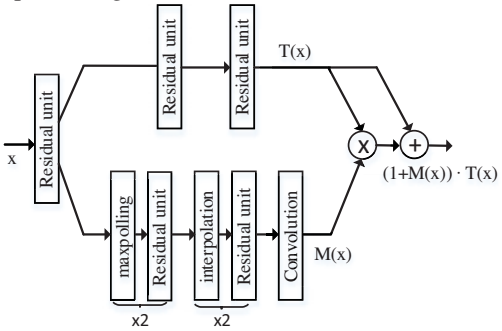


Figure 2. Attention module structure

In Fig. 2, x is the input of the attention module. Feature processing branch output T(x) with input x. The feature selection branch uses bottom-up top-down structure to learn same size mask M(x) that softly weight output features T (x), as control gates for neurons of feature processing branch. The output of Attention Module is (1+M(x))·T(x).

In the feature selection branch, max pooling are performed to increase the receptive field rapidly after Residual Units. After reaching the lowest resolution, the global information is then expanded by a symmetrical top down architecture to guide input features in each position. Linear interpolation up sample the output after some Residual Units. Then output M(x) after convolution layers. During backward propagation, the feature selection branch acts as a gradient update filter to prevent the erroneous gradient from the noise label from updating the features of the feature processing branch. In the Attention Module, each feature processing branch has its own feature selection branch to learn attention that is specialized for its features, realized adaptively select features from the attention module.

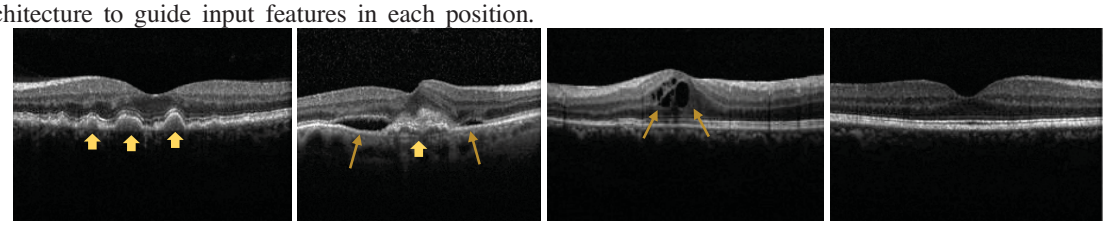
C. *Datasets*

The retinal fundus images used in this work were obtained from the website of https://data.mendeley.com/datasets/rscbjbr9sj/2 [17]. The dataset contains 84484 OCT 2D slice images (B-scans) of the fundus retina, which includes 26565 healthy normal fundus images, 11598 DME images, and 37456 CNV images, 8866 drusen images. DME and CNV are the main factors to result in blindness. The OCT images of the-se four fundus conditions are shown in Fig. 3.

We choose 5581 CNV, 1676 DME, 1264 DRUSEN and 3911 NORMAL, a total of 12432 images as the validation dataset. The remaining 72052 images are divided as the training dataset. The training dataset is used to train the classification model, and the validation dataset is used to evaluate the model performance.

III. PROPOSED METHODS

The overview algorithm of the proposed classification methods in this paper is shown in Fig. 4. Preprocessing block perform the operations (such as screening, cropping, flipping and rotating) for the sent OCT images. The screening operation is to remove the OCT images that do not have classification feature. The crop, flip and rotate operations perform data augmentation, enhance the generalization ability of the classifier, and avoid over fitting. In the classification network block, we use two types of deep neural networks to implement AMD classification. By using the automatic feature extraction capability of deep neural networks, ResNet50 is used as the first classifier. In addition, Atten-ResNet classification model which is an attention-based deep residual network architecture is used to implement AMD classification as the second classifier. The attention mechanism enhances the effective features and suppresses the invalid features, therefore the network can pay more attention to the nuanced parts in the OCT images. In the output category block, the input data is predicted as corresponding category labels and output.



(a) Drusen (b) CNV (c) DME (d) Normal

Figure 3. Four types of OCT images of AMD fundus retina (a) Multiple drusen in early AMD. (b) choroid neovascularization (CNV) with neovascularization membrane and associated sub retinal fluid (c) diabetic macular edema (DME) with retinal thickening associated retina internal fluid. (d) Normal

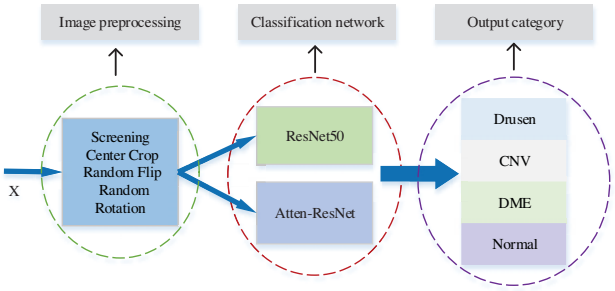


Figure 4. Overview of the algorithm for the proposed AMD classification

IV. EXPERIMENTAL VERIFICATION

A. Evaluation Metrics

Classification performance is evaluated according to the following metrics: the overall accuracy (OA) and overall precision (OP), precision (PR), sensitivity (SE). These metric is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$OA = \frac{correctly\ classified\ samples}{total\ number\ of\ samples} \tag{3}$$

$$OP = \frac{correctly\ positive\ samples}{total\ number\ of\ positive\ samples} \tag{4}$$

where, TP (True Positive) represents positive samples which are identified as positive samples; FP (False Positive) represents negative samples which are identified as positive samples; FN (False Negative) represents positive samples which are identified as negative samples; TN (True Negative) represents Negative samples which are identified as negative samples.

B. Parameter Settings

The parameter settings are as follows. The network is optimized by using stochastic gradient descent algorithm (SGD) with a momentum of 0.9 with batch of 64 images per step and an initial learning rate of 0.1. The weight decay factor is set to 0.001. At each iteration, the accuracy and loss of the models was recorded, and at each iterations, the performance of the models was evaluated using an independent validation set. The maximum number of iterations is chosen to 90 epochs. During training of both ResNet50 and Atten-ResNet, The original OCT images are resized to 224 × 224 as input. To improve robustness of the network, the data augmentation strategies are performed on each B-scan of the OCT datasets by center crop, random flipping and random rotation.

The proposed methods are implemented using the PyTorch framework with NVIDIA Cuda v9.0 and cuDNN v10.0 accelerated library, and is coded in Python. All experiments are performed under an Ubuntu 16.04 operating system on a machine with GPU NVIDIA GeForce 1080 Ti, and 16GB of RAM.

C. Results and Analysis

Table I compares four methods of HOG-SVM, VGG16, ResNet50, and Atten-ResNet for AMD classification. It can be seen that the VGG16, ResNet50 and Atten-ResNet these deep learning-based methods have better classification performance than HOG-SVM method. Deep neural network ResNet50 and Atten-ResNet have a great performance improvement compared to shallow neural network VGG16.

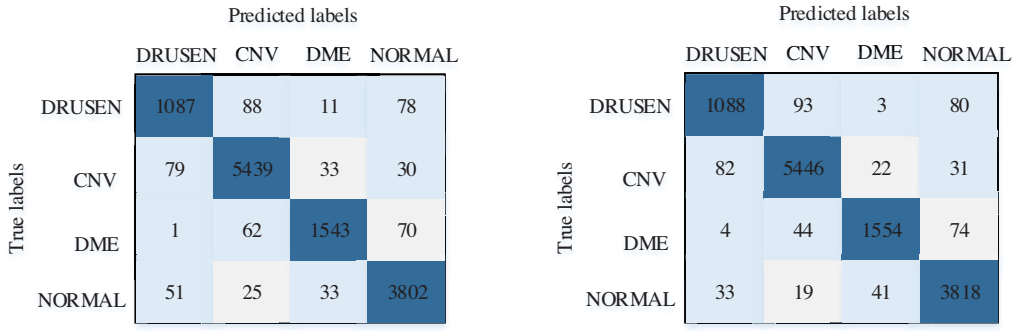
As can be observed, the ResNet50 achieves much improvement in terms of all quantitative metrics compared with the HOG-SVM and VGG network. Specifically, ResNet50 achieved 95.3% in overall classification accuracy, higher than HOG-SVM with 17.2%, and VGG with 12.1%. For each individual class of the Drusen, CNV, DME and Normal, the gains (on the precision) of the ResNet50 over VGG16 are about 34.7%, 3.7%, 20.6% and 12.2%, respectively, which demonstrates the effectiveness of depth of the network for OCT classification. Moreover, the ResNet50 also outperforms the HOG-SVM method with the improvement of about 17.2% in OA, and significant improvement of PR, SE and OP can also be observed. This demonstrates that the ResNet50 can exploit the discriminative high-level features to achieve high automatic classification performance compared with traditional classification method.

In Table I, the proposed Atten-ResNet method achieved 95.7% in OA, which outperforms the HOG method, the VGG16 method and the ResNet50 method with the improvement of about 17.6%, 12.5% and 0.4%, respectively. In addition, compare with the classification evaluation indicator OP, the Atten-ResNet method achieved 95.2%, far superior to the HOG-SVM methods and the VGG16 methods. In comparison with the ResNet50 method, Atten-ResNet leads with a gain of 1.5%. The above experimental results demonstrates the effectiveness of utilizing feature selection. by the attention mechanism to guide CNN for the classification of OCT images. For each individual class of the Drusen, CNV, DME, the gains (on the precision) of the Atten-ResNet over its counterpart without attention mechanism ResNet50 are about 0.9%, 0.3%, 0.4%, respectively. Atten-ResNet adds an adaptive feature selection mechanism on the basis of deep neural network, which makes the model focus more on the subtle features from the high-level lesion feature and improve the classification performance.

As can be seen from Fig. 5(a), the ResNet50 in a multi-class comparison between DRUSEN, CNV, DME, and NORMAL, 11872 of 12432 images on the validation dataset were accurately identified and the prediction accuracy was up to 95.3%. The training and validation accuracy and loss curve of ResNet50 on public OCT dataset are shown in Fig. 6(a). In order to clearly see the changing trend of the graph, we perform normalize to the graph. After 90 iterations, it can be observed from Fig. 6(a), the accuracy on the training set reaches99.6%, and the accuracy on validation reaches 96.8%.

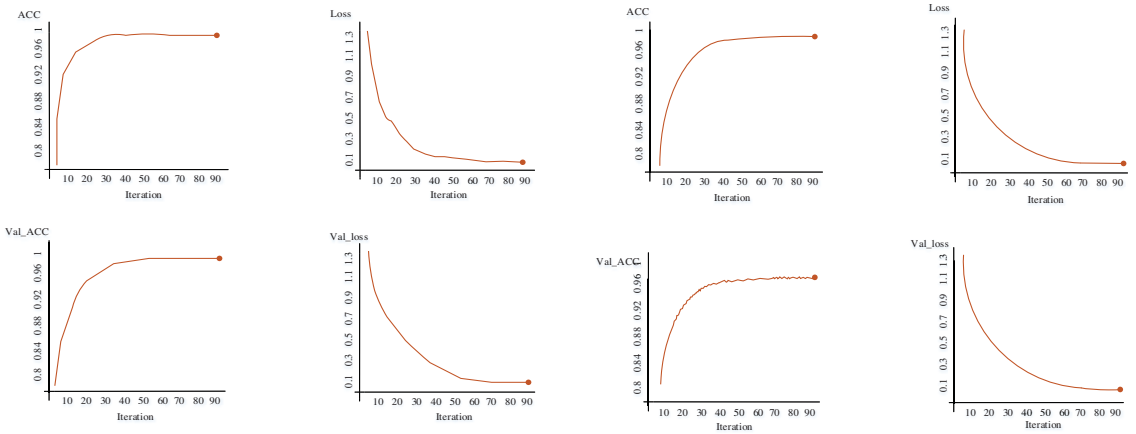
TABLE I. COMPARISON OF CLASSIFICATION PERFORMANCE BETWEEN PROPOSED METHODS AND STATE OF THE ART CLASSIFICATION METHODS

Methods	Classes	PR (%)	SE (%)	OA (%)	OP (%)
HOG-SVM[4]	Drusen	52.6	29.5	78.1	71.8
	CNV	82.0	87.6		
	DME	74.6	53.8		
	Normal	78.1	90.4		
VGG16[15]	Drusen	54.5	54.7	83.2	76.4
	CNV	93.2	86.6		
	DME	74.6	70.9		
	Normal	83.3	92.6		
ResNet50	Drusen	89.2	86.0	95.3	93.7
	CNV	96.9	97.0		
	DME	95.2	92.1		
	Normal	95.5	97.2		
Atten-ResNet	Drusen	90.1	86.1	95.7	95.2
	CNV	97.2	97.1		
	DME	95.6	93.1		
	Normal	94.9	97.6		



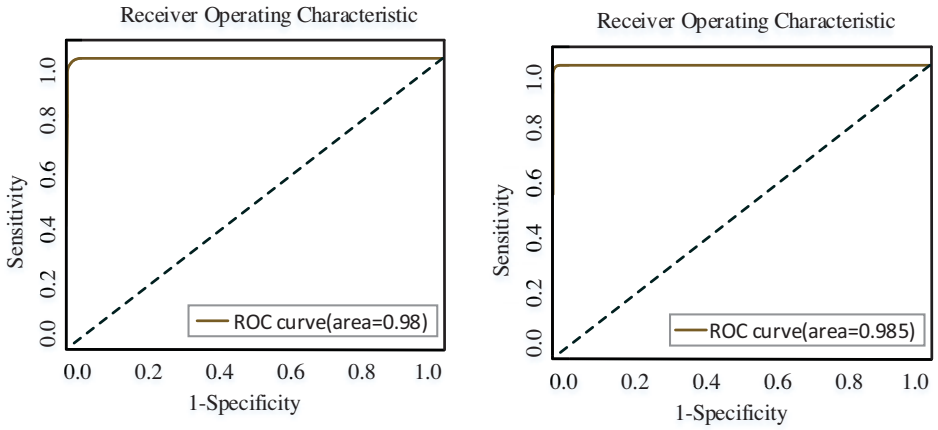
(a) Confusion matrix of ResNet50 (b) Confusion matrix of Atten-ResNet

Figure 5. Confusion matrix of proposed methods classification



(a) The curve of the accuracy and loss function of the ResNet50 (b) The curve of accuracy and loss function of the Atten-ResNet

Figure 6. The curve of accuracy and loss function of the proposed methods in the training and validation stages



(a) ROC curve of the ResNet50 (b) ROC curve of the Atten-ResNet

Figure 7. Receiver-operating characteristic (ROC) curve reflected the classification performance of the proposed methodst

The ROC curve was generated for quantifying the performance of model to distinguish urgent situation CNV, DME from observation situation DRUSEN and NORMAL. The AUC generated by plotting sensitivity vs 1-specificity reached 98% as shown in Fig. 7(a).

Fig. 5(b) shows the confusion matrix by using Atten-ResNet. The prediction accuracy of AMD classification of this method reaches 95.7%. The training and validation accuracy and loss curve are normalized and shown in Fig. 6(b). After 90 iterations, it can be observed that the accuracy on the training set reaches 99.7%, and the accuracy on validation reaches 97.4%. The training and validation losses of Atten-ResNet start with a rapid descent and then reach the convergence soon, which demonstrates that the Atten-ResNet method can significantly accelerate the network training process. Fig. 7(b) shows the ROC curve for quantifying the performance of model to distinguish urgent situation from observation situation. AUC of 98.5% was achieved from the ROC curve. This showed that the proposed Atten-ResNet was robust and capable of detecting automatically and accurately retinal disorders.

The average training time and test time of Atten-ResNet for each OCT image are 0.095 seconds and 0.012 seconds, respectively. For ResNet50 network, the training and testing time of each OCT image is about 0.087 and 0.011 seconds, respectively. Although the time complexity of Atten-ResNet is higher than the ResNet50, the Atten-ResNet can still meet the requirement of the clinical diagnosis for the eye diseases and compared to ResNet50, Atten-ResNet achieves a 0.4% performance improvement (using OA as indicator).

Two methods proposed above are suitable for the classification of age-related macular degeneration. In the future, the proposed methods can be further applied to other complex retinal diseases, such as diabetic retinal disease, glaucoma, cataract, etc.

V. CONCLUSION

We apply two deep convolutional neural network: ResNet50 and Atten-ResNet to achieve automatic classification of Age-relating macular degeneration. Compared with the previous networks, The ResNet50 can automatically extract high-dimensional features from the fundus retinal OCT images. The Atten-ResNet introduces the attention mechanism on the basis of the residual basic unit to achieve the selection of high-dimensional feature, which can pay more attention to the nuanced parts in the OCT image to achieve more efficient and accurate OCT classification. The experimental result show that the proposed two methods have accuracy improvement by comparing to traditional classification method HOG-SVM and shallow neural network VGG16. In addition, guided by the attention information, The Atten-ResNet is showed a better classification performance than ResNet50 method. Furthermore, the methods proposed in this paper realizes end-to-end automatic AMD classification. There is no need for engineered features extracted from retinal OCT image dataset for the diagnosis. The methods in this work can be potentially applied to clinics to assist ophthalmologists to make a diagnostic decision, which reduce the burden on clinicians and increase the accuracy of clinical diagnosis.

ACKNOWLEDGMENT

This work is partially supported by the China Scholarship Council.

REFERENCES

[1] D. C. Neely, K. J. Bray, C. E. Huisingh, M. E. Clark, G. J. McGwin and C.Owsley, “Prevalence of Undiagnosed Age-related Macular Degeneration in Primary Eye Care,” JAMA Ophthalmol, vol. 135, no. 6, Dec. 2017, pp. 570-575, doi: 10.1001/jamaophthalmol.2017.0830.

[2] C. A. Puliafito，M. R. Hee, C. P. Lin, E. Reichel, J. S. Schuman, J. S. Duker, et al., “Imaging of Macular Diseases with Optical Coherence Tomography,” Ophthalmology, vol. 102, no. 2, Dec. 1995, pp. 217–229, doi: 10.1016/S0161-6420(95)31032-9

[3] Y. Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J.S. Schuman and 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,” Med. Image Anal, vol. 15, Dec. 2011, pp.748–759, doi: 10.1016/j.media.2011.06.005.

[4] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, et al,“Fully Automated Detection of Diabetic Macular Edema and Dry Age-related Macular Degeneration from Optical

Coherence Tomography Images," Biomedical optics express, vol.5, Dec. 2014, pp: 3568-3577, doi: 10.1364/BOE.5.003568.

[5] K. Alsaih, G. Lemaitre, M. Rastgoo, J. Massich, D. Sidibé, and F. Meriaudeau, "Machine Learning Techniques for Diabetic Macular Edema (DME) Classification on SD-OCT Images," Biomedical Engineering Online, vol. 16, Dec. 2017, doi: 10.1186/s12938-017-0352-9.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, Dec. 2006, pp. 504-507, doi: 10.1126/science.1127647.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, Dec. 2015, pp. 436-444, doi: 10.1038/nature14539.

[8] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in Proc. Adv. Neural Inform. Process. Syst, vol. 25, Dec. 2012, pp. 1097-1105, doi: 10.1145/3065386.

[9] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep Learning is Effective for Classifying Normal versus Age-related Macular Degeneration OCT images,"Ophthalmology Retina, vol. 1, no. 4, Dec. 2017, pp. 322–327, doi: 10.1016/j.oret.2016.12.009.

[10] R. Rasti, A. Mehridehnavi, H. Rabbani, and F. Hajizadeh, "Automatic Diagnosis of Abnormal Macula in Retinal Optical Coherence Tomography Images using Wavelet-based Convolutional Neural Network Features and Random Forests Classifier," J. Biomed. Opt., vol. 23, no. 3, Dec. 2018, pp. 35-45, doi: 10.1117/1.JBO.23.3.035005.

[11] 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, Dec. 2017, pp. 579–592, doi: 10.1016/j.oret.2016.12.009.

[12] 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, Dec. 2019, pp. 495—505, doi: 10.1007/s00417-018-04224-8.

[13] L. C. Chen, Y. Yang, J. Wang, W. Xu, A. L. Yuille, "Attention to Scale: Scale-aware Semantic Image Segmentation," 2016 IEEE Conference on Computer Vision and Pattern Recognition, Dec. 2016, pp. 3640-3649, doi: 10.1109/CVPR.2016.396.

[14] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Computer Society, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, Dec. 2016, pp: 770-778, doi: 10.1109/CVPR.2016.90.

[15] R. Eldan and O. Shamir, "The Power of Depth for Feedforward Neural Networks," Journal of Machine Learning Research, Dec. 2016, pp:1-34.

[16] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang et al., "Residual Attention Network for Image Classification," 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Dec. 2017, doi: 10.1109/CVPR.2017.683.

[17] D. S. Kermany , M. Goldbaum , W. J. Cai, C. S. Valentim, H. E. Liang, S. L. Baxter et al., "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," Cell , Volume 172, Dec. 2018, pp: 1122-1131, doi: 10.1016/j.cell.2018.02.010 .