tvst

https://doi.org/10.1167/tvst.7.6.41

Article

Deep Learning-Based Automated Classification of Multi-Categorical Abnormalities From Optical Coherence Tomography Images

Wei Lu[1], Yan Tong[1], Yue Yu[2], Yiqiao Xing[1], Changzheng Chen[1], and Yin Shen[1]

[1] Eye Center, Renmin Hospital of Wuhan University, Eye Institute of Wuhan University, Wuhan, Hubei, China
[2] Wuhan Haixingtong Technology Limited Company, Wuhan, Hubei, China

Correspondence: Yin Shen, Eye Center, Renmin Hospital of Wuhan University, Eye Institute of Wuhan University, Wuhan, Hubei, China. e-mail: yinshen@whu.edu.cn

Received: 8 June 2018
Accepted: 10 November 2018
Published: 28 December 2018

Keywords: deep learning; optical coherence tomography; artificial intelligence; ResNet

Citation: Lu W, Tong Y, Yu Y, Xing Y, Chen C, Shen Y. Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. Trans Vis Sci Tech. 2018;7(6):41, https://doi.org/10.1167/tvst.7.6.41

Purpose: To develop a new intelligent system based on deep learning for automatically optical coherence tomography (OCT) images categorization.

Methods: A total of 60,407 OCT images were labeled by 17 licensed retinal experts and 25,134 images were included. One hundred one-layer convolutional neural networks (ResNet) were trained for the categorization. We applied 10-fold cross-validation method to train and optimize our algorithms. The area under the receiver operating characteristic curve (AUC), accuracy and kappa value were calculated to evaluate the performance of the intelligent system in categorizing OCT images. We also compared the performance of the system with results obtained by two experts.

Results: The intelligent system achieved an AUC of 0.984 with an accuracy of 0.959 in detecting macular hole, cystoid macular edema, epiretinal membrane, and serous macular detachment. Specifically, the accuracies in discriminating normal images, cystoid macular edema, serous macular detachment, epiretinal membrane, and macular hole were 0.973, 0.848, 0.947, 0.957, and 0.978, respectively. The system had a kappa value of 0.929, while the two physicians' kappa values were 0.882 and 0.889 independently.

Conclusions: This deep learning-based system is able to automatically detect and differentiate various OCT images with excellent accuracy. Moreover, the performance of the system is at a level comparable to or better than that of human experts. This study is a promising step in revolutionizing current disease diagnostic pattern and has the potential to generate a significant clinical impact.

Translational Relevance: This intelligent system has great value in increasing retinal diseases' diagnostic efficiency in clinical circumstances.

Introduction

Deep learning, a burgeoning technology of Artificial Intelligence (AI), has significantly improved the state-of-the-art in image recognition, speech recognition, and navigation.[1,2] The astounding methodology has also been applied into a variety of medical fields in an attempt to enhance management of various healthcare problems. Multiple studies have shown that deep learning algorithms performed at a high level when applied to breast histopathology analysis,[3] skin cancer classification,[4] cardiovascular diseases risk prediction,[5] lung cancer detection,[6] and diabetic retinopathy diagnosis.[7] Gulshan et al.[7] was the first to report the application of deep learning in diagnosing eye diseases. In the last 2 years, a number of deep learning models have been developed for the automated detection of retinal diseases. Diabetic retinopathy, age-related macular degeneration, and glaucoma were the most intensively studied diseases.[8] More strikingly, the first medical device to detect mild or worse diabetic retinopathy by AI (IDx-DR; Technologies Inc., Coralville, IA)[9] has been authorized to market by the US Food and Drug Administration recently. However, the majority of these studies focused mainly on the analysis of fundus photographs. The implementation of auto-

mated diagnosis based on other imaging techniques, such as optical coherence tomography (OCT), remains insufficient.

OCT, a noninvasive, noncontact imaging technique, has become an indispensable tool for the diagnosis of retinal diseases based on its high resolution and convenience in clinical practice.[10–12] OCT is considered the best diagnostic approach to diagnose macular diseases when compared to other imaging techniques, such as ultrasound, fundus photography, and fluorescein angiography. OCT images are useful in facilitating decision-making regarding medical interventions, such as anti-vascular endothelial growth factor (anti-VEGF) injection and vitrectomy surgery. The development of automatic and reproducible OCT classifications should be helpful in supporting clinical work by promoting diagnosis efficiency and improving access to care and professional knowledge, especially in situations where qualified readers are scarce.

Few prior studies have applied deep learning methods to diagnose eye diseases by OCT images.[13–16] ElTanboly et al.[13] developed a deep learning-based computer-aided system to detect diabetic retinopathy from a small sample of data (52 OCT scans), achieving an AUC of 0.98. Kermany et al.[14] reported an accuracy of 96.6%, with a sensitivity of 97.8%, and a specificity of 97.4% in classifying age-related macular degeneration and diabetic macular edema. Schlegl et al.[15] and Lee et al.[16] also proposed deep learning method in detecting cystoid macular edema and achieved an AUC of 0.94 and a cross-validated dice coefficient of 0.911, respectively.

However, the abovementioned studies focused only on a binary classification method to address a "one disease versus normal" task. It is difficult to extend simple binary classifiers into a real clinical setting where visiting patients suffer from various retinal diseases. Multiclass classifiers, which can differentiate a specific abnormality among multi-categorical abnormalities, is more conformed to the clinical circumstances. Nevertheless, the implementation of multiclass classification aimed at identifying diverse retinal diseases through AI still faces challenges.

With the aging of the population, patients suffer from vision-threatening macular diseases continue to increase. Serous macular detachment, cystoid macular edema, macular hole, and epiretinal membrane are treatable macular diseases primarily affecting elderly patients and can lead to a severe visual loss. Treatment of anti-VEGF injection or vitrectomy surgery is generally most effective if carried out earlier. In the present, OCT is the best modality for the detection and treatment decision-making of these four abnormalities in clinical practice.

In order to provide earlier detection as well as earlier intervention of multiple treatable macular diseases, we establish an intelligent system based on deep learning to implement multiclass classification for OCT images. The system has the potential to increase diagnostic efficiency, enable easier access to expert knowledge, facilitate therapeutic decision-making, and decrease overall healthcare costs. This study is the first one to design a multiclass classifier through deep learning to categorize four macular abnormalities.

Methods

Data Set

This study followed the tenets set forth in the Declaration of Helsinki, and approval from the institutional review board of Eye Center, Renmin Hospital of Wuhan University was obtained. A total of 60,407 completely anonymized OCT scans (imaged by Cirrus HD-OCT 4000, Carl Zeiss Meditec, Inc., Dublin, CA, with Macular Cube 512 × 128 protocol and Optic Disc Cube 200 × 200 protocol from February 1, 2012, to October 1, 2014) were exported from the Wuhan University Eye Center. All images were deidentified and encrypted to protect the privacy and security of patients' health information. The data set contains a great diversity of OCT images from all kinds of patients, including males, females, adults, and children. In addition, the data set contains scans from the same patients who underwent the OCT examination in their follow-up study at different time.

Image Labeling and Training Process

The current study invited 17 licensed ophthalmologists, who are specialized in retinal diseases diagnosis, to screen and label 60,407 OCT scans. Images were assigned randomly to 1 of 14 junior retinal experts for the first-round screening and labeling. Each of them reviewed 4314 OCT images. In the second round, three senior retinal experts were invited to confirm (or correct) the labeling results and each of them reviewed 20,135 OCT images. The criterion for inclusion is that the image only contained ONE of the four abnormalities (serous macular detachment, cystoid macular edema, macular hole, and epiretinal membrane), also we include the normal images. The

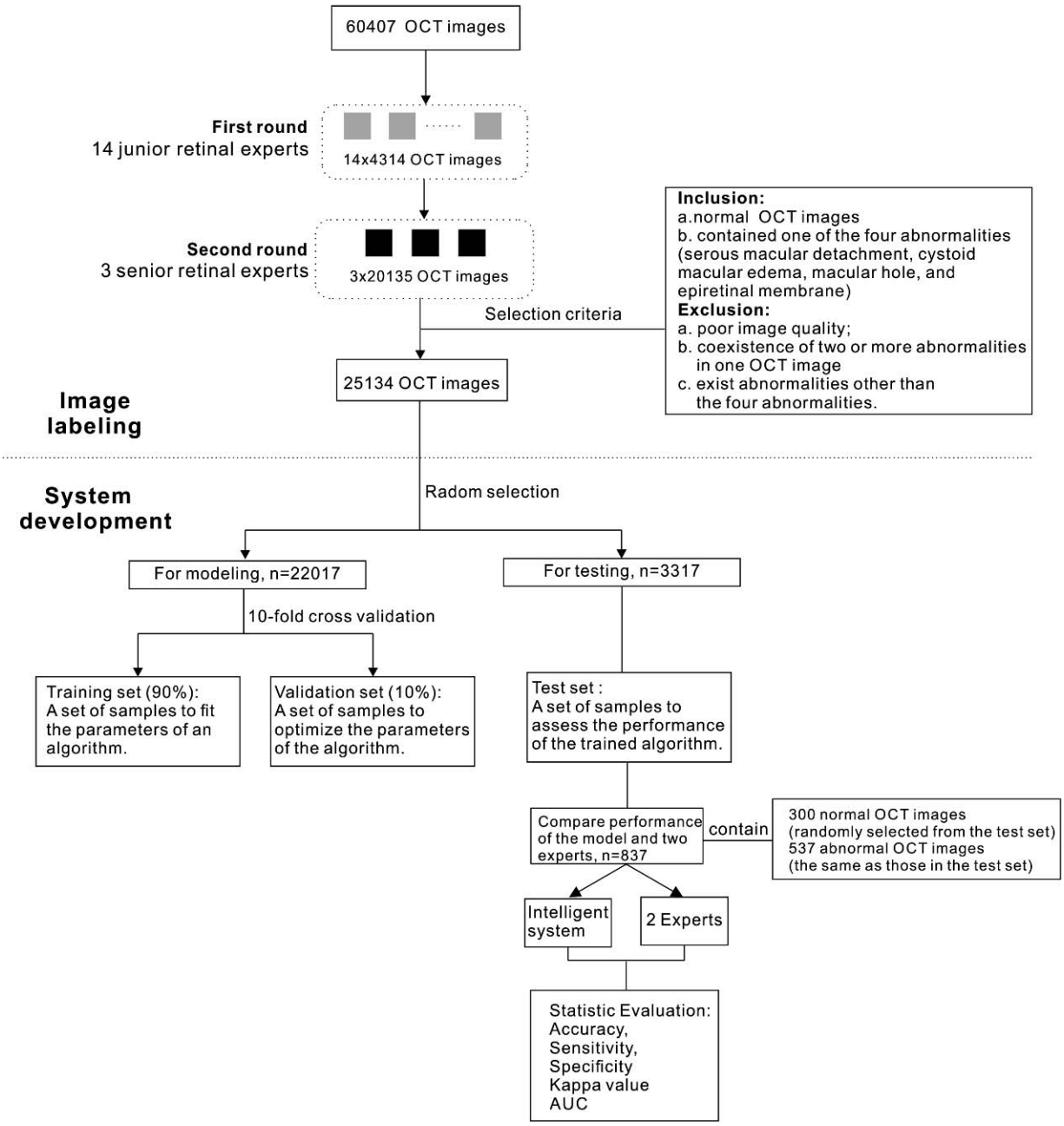

Figure 1. Overall experimental design.

exclusive criteria are (1) poor image quality; (2) coexistence of two or more abnormalities; (3) existence of abnormalities other than the four abnormalities. There were no images excluded based on age, gender, or race. The blind expert panel had no access to deep-learning predictions.

According to the selection criteria, eventually, 25,134 OCT images (including a full range of macular holes, serous macular detachment, cystoid macular edema, and epiretinal membrane) were included in the current study to build intelligent system. Representative OCT images are shown in Figure 2. Image numbers of each category in training set and test set are summarized in Table 1. We selected 3317 images randomly using a simple random sampling method and treated them as a test set. The remaining 22,017 images were used as the training set (Fig. 1). The training set is used to fit the parameters of a model. The test set is used to evaluate the final performance of the trained model. During the training process, we

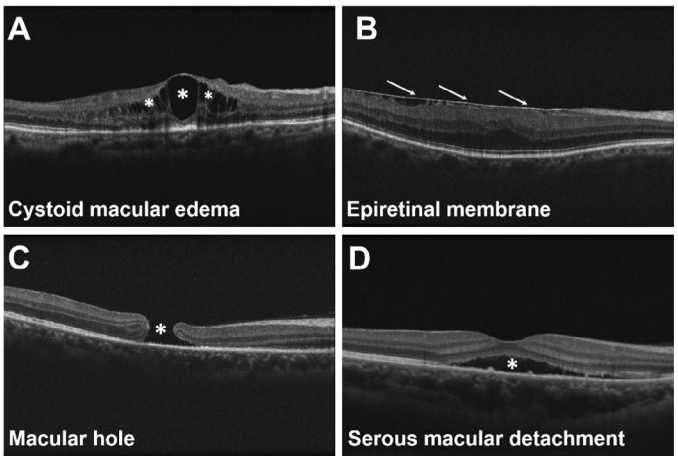

Figure 2. Representative OCT images. The *arrows* and the *asterisks* in A–D indicate the lesion sites. (A) Cystoid macular edema. (B) Epiretinal membrane. (C) Macular hole. (D) Serous macular detachment.

used 10-fold cross-validation method. Cross-validation methods have been widely used to estimate and optimize algorithms.[17] The training data set is randomly and equally divided into 10 subsets. Nine subsets were used to train the model, and the remaining one subset was used to estimate how well the model had been trained and optimized the parameters. This process was repeated 10 times before the algorithms were ready to be tested.

Development of the Intelligent System

In this study, we used the 101-layer deep convolutional neural network (CNN)-ResNet. It was the championship model from the ImageNet Large Scale Visual Recognition Challenge of 2015.[18] The CNN consists of multiple convolutional layers that extract features and transform input images into hierarchical feature maps: from simple features, such as edges and lines, to complicated features, like shapes and colors. It also includes pooling layers (including average pool and max pool) that merge semantically similar features into one to reduce the dimensionality of the extracted features and fully connected layers to combine these features and output a final probability value for the class. The more network layers we use, the more features that algorithm can learn. Recent studies indicate that network depth is beneficial for classification accuracy.[19] However, with the network depth increasing, accuracy gets saturated and then degrades rapidly.[20] The ResNet framework can tackle this problem. In the network, shortcut connections are added for every three convolutional layers along the whole deep network (Fig. 3). The shortcut connections simply perform identity mapping, which does not add any extra parameter or computational complexity. These connections make the deep network easier to optimize, and easier to tackle the problem of gradient vanishing (or gradient explode) during the training process. Hence, ResNet makes it possible to gain higher accuracy from a considerably deeper network than obtained from shallower networks when performing image classification tasks.[18]

An algorithm can apply cumulative knowledge learned from other data sets to a new task by using transfer learning.[21] A large data set is required to train a completely blank deep CNN, which has millions of weights to adjust. Transfer learning is a highly effective technique that has been used increasingly in the application of deep learning. The methodology is to retrain an algorithm that has

Table 1. Numbers of Images in the Training Set and Test Set

Categories	Training Set	Test Set
Normal	15,485	2580
Cystoid macular edema	1220	105
Serous macular detachment	2377	246
Epiretinal membrane	2252	141
Macular hole	683	45
Total	22,017	3117

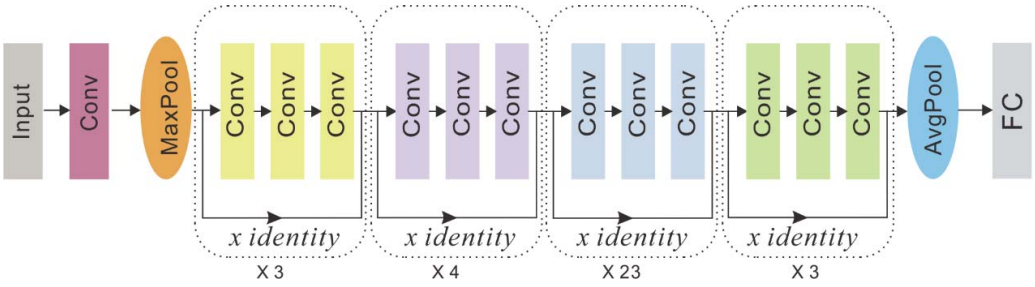

Figure 3. A diagram showing a 101-layer ResNet. Conv, convolutional layer; AvgPool, average pool; FC, fully connected layer.

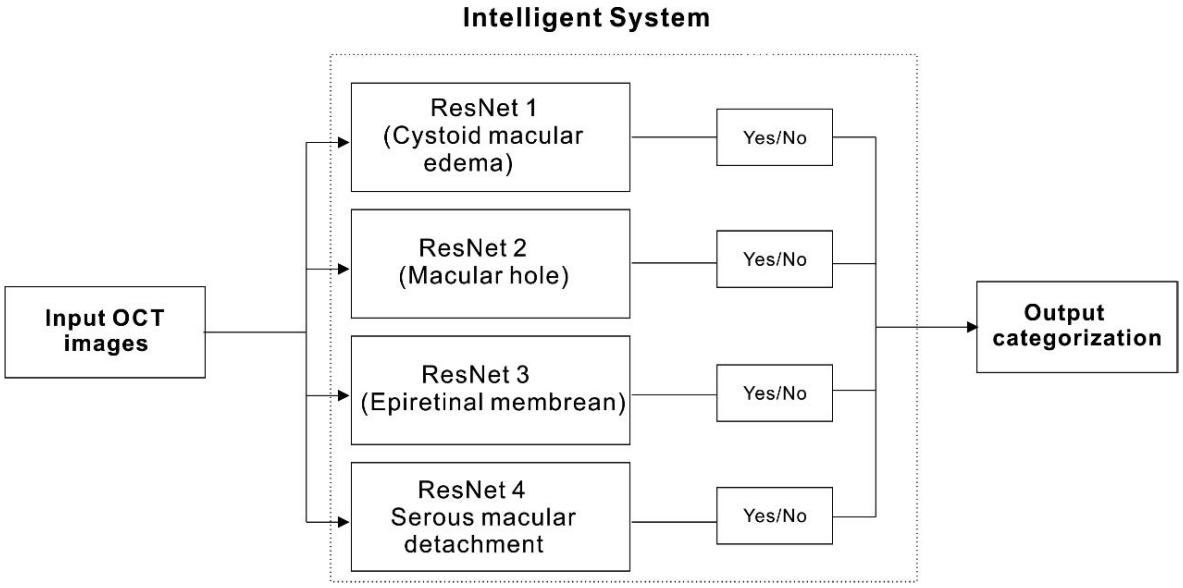

Figure 4. The workflow of our intelligent system.

already been pretrained on millions of general images by a specific data set. Therefore, transfer learning makes it possible to obtain a highly accurate model with a relatively small training data set. The ResNet used in our study was pretrained on ImageNet.[22] Universal features learned from the pretraining were reused for the OCT classification tasks in this study.

We independently trained four binary classifiers to discriminate abnormalities from normal OCT images and combined the four classifiers as a system. When testing the system, each OCT image will go through four rounds of category and then the system will output a final categorization (Fig. 4).

Statistical Evaluation

To evaluate the performance of the intelligent system, accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (95% CIs) were used (Fig. 1). A receiver operating characteristic (ROC) curve was created by plotting the detection probability for each algorithm across a continuum of the threshold. For each threshold, the sensitivity and the false positive rate (1-specificity) were plotted against each other. The AUC can be very useful for the quantitative assessment of a model. The AUCs of effective models range from 0.5 to 1.0; the higher the value of AUC, the better the performance of the model.

A kappa value was also calculated to examine the agreement between the system with the ground truth on the assignment of categories of a categorical variable. Kappa generally ranges from 0 to 1, where larger numbers mean better reliability.[23]

The performance of the system was further assessed by comparing to results obtained by two experts with rich clinical experience (Fig. 1). The data set used for comparison consisted of 300 normal images (randomly selected from normal images in the test set) and 537 abnormal images (the same as those in the test set). The statistical analyses were performed using GraphPad Prism software version 7.0 (La Jolla, CA) and IBM SPSS Statistics 19 (Armonk, NY).

Results

This intelligent system was evaluated in diagnosing cystoid macular edema, epiretinal membrane, serous macular detachment, and macular hole from OCT images. An accuracy of 0.959 with a sensitivity of 0.942 and a specificity of 0.964 was obtained in the multiclass classification (Table 2). A ROC curve was generated to evaluate the system's performance on

Table 2. Performance of the Proposed System and Human Clinicians for Classification of the Four Abnormalities

	Accuracy	Sensitivity	Specificity	Kappa
System	0.952	0.940	0.973	0.929
Expert 1	0.959	0.970	0.940	0.882
Expert 2	0.904	0.931	0.950	0.891

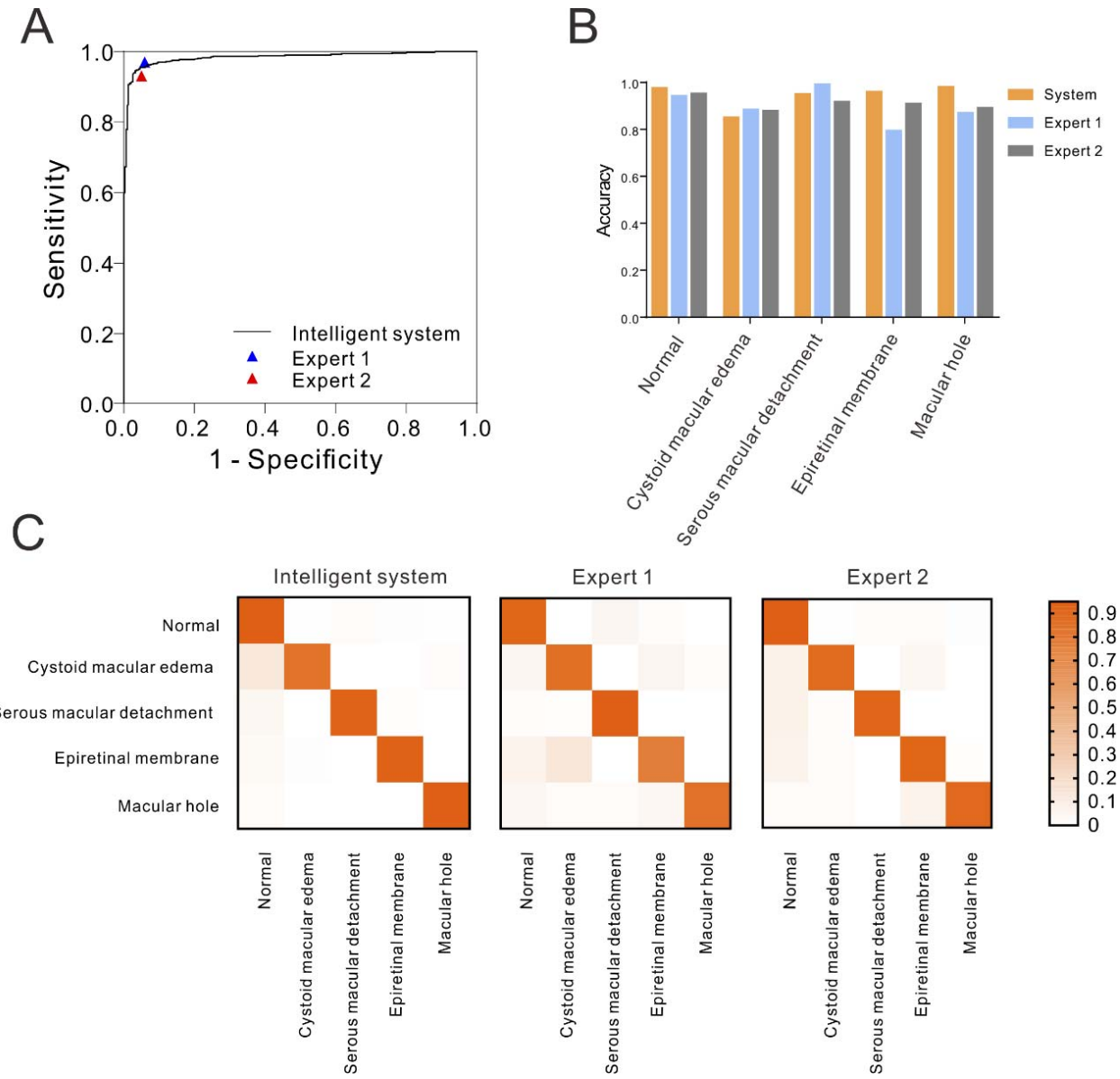

Figure 5. (A) ROC curve for diseases detected by the system, with the operating points of the two experts shown for comparison. (B) A histogram comparing the specific accuracy of each category between the two experts and the system. (C) Three confusion matrixes for the intelligent system and the two experts' predictions, respectively.

discriminating diseases from normal control (Fig. 5A). The AUC was 0.984 (95% CI, 0.976–0.991).

We further compared the performance of the system with the results obtained by two experts. Expert 1 got a sensitivity of 0.97, an accuracy of 0.959, and a specificity of 0.94, while expert 2 got a sensitivity of 0.931, an accuracy of 0.904, and a specificity of 0.95 (Table 2). The sensitivities and specificities of the two experts were plotted on the ROC curve in Figure 5A for comparison. The operating point of expert 1 fell on the ROC curve, while the operating point of expert 2 fell beneath the curve.

Table 3 shows the specific accuracy of each category obtained from the proposed system and the two experts. The results showed that the system can correctly identify the healthy control and the four abnormalities with accuracies of 0.973, 0.848, 0.947, 0.957, and 0.978, respectively (Table 3). The AI system outperformed the human experts in diagnosing normal, epiretinal membrane, and macular hole while the physicians did slightly better for cystoid macular edema. In the categorization of serous retinal detachment, the system's accuracy was 0.032 better than expert 2, with 0.041 poorer than expert 1 (Fig. 5B).

Table 3. The Specific Accuracy of Each Category of the Proposed System

Category	System	Expert 1	Expert 2
Normal	0.973	0.940	0.950
Cystoid macular edema	0.848	0.881	0.876
Serous macular detachment	0.947	0.988	0.915
Epiretinal membrane	0.957	0.794	0.908
Macular hole	0.978	0.867	0.889

Three confusion matrixes shown in Figure 5C reveal the specific assignment of each image of different predictions. The rows represent the samples' true label, and the columns provide the predicted label. Each diagonal element of the heatmap represents the percentage of images classified correctly for the corresponding class. Off diagonal elements show the percentage of misclassified images and how they are misclassified. Misclassification cases and types in the intelligent system were significantly fewer than those from ophthalmologists. The kappa values in Table 2 also described the statistics results, with 0.929 for the system, 0.882 and 0.889 for the two experts, respectively.

The ROC curves of the four-constituent binary classifiers in identifying the corresponding OCT images are shown in Figure 6. The AUC of the four binary classifiers were 0.996 (95% CI, 0.993–0.999), 0.997 (95% CI, 0.994–0.999), 0.998 (95% CI, 0.997–0.999), and 0.999 (95% CI, 0.998–1.000), respectively.

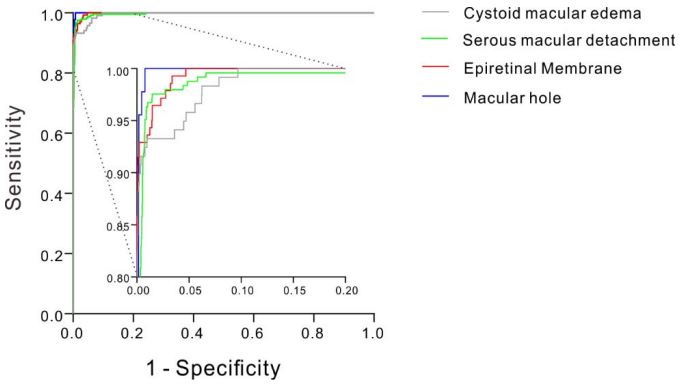

Figure 6. ROC curves for binary classifiers to discriminate cystoid macular edema, macular hole, epiretinal membrane, and serous macular detachment from normal ones.

Discussion

In this study, a deep learning-based system was built to automatically classify four category OCT images. The proposed system yielded a robust accuracy at a level equivalent to or better than that of human experts in identifying the four abnormalities. Furthermore, the predictions of the system had better consistency and higher reliability than experts.

Serous macular detachment, cystoid macular edema, macular hole, and epiretinal membrane have been chosen in this initial study. They were selected based on (1) easy detectable OCT characteristics; (2) large sets of data; (3) common in clinical practice. Automatic detection of these abnormalities would enable early detection as well as early intervention and can effectively reduce the burden on patients and promote their quality of life. We trained four binary classifiers independently, with the images of the specific category as positive samples, and normal macular images as negatives. The four binary classifiers showed excellent performance in identifying corresponding abnormalities from normal images. We then integrated four binary classifiers to build the intelligent system and evaluated its performance in categorization. The robust performance shown in the results demonstrated that our methods were plausible and effective to implement automated categorization of the four abnormalities from OCT images. Furthermore, our system has the ability to recognize multiple abnormalities from complex OCT scans according to the algorithms we trained. We tried a small number of complicated images, including epiretinal membrane with macular edema, macular hole with macular edema, and macular edema with serous retinal detachment. The accuracy was much lower (data not shown), demonstrating the need for further work to improve accuracy in recognizing complicated images.

The most prominent advantage of our study is probably the attempt to address the multiclass OCT image classification. Although many other deep learning approaches involved multiple classes, they mainly focused on the staging of diseases.[21,24–26] Few studies have focused on the automated classification of diseases/abnormalities in ophthalmology. Choi et al.[27] carried out a study on fundus images, applying deep learning to automatically detect multiple retinal diseases with accuracy that ranged from 30.5% to 87.4%. Kermany et al.[14] established a system based on OCT images to automatically detect age-related macular degeneration and diabetic macular edema, achieving an accuracy of 96.6%.

In our study, we developed a multiclass diagnostic system for the automated diagnosis of macular hole,

epiretinal membrane, serous macular detachment, and cystoid macular edema from OCT images. Such a system enables a timely and accurate diagnosis of severe conditions on a tissue map and can discriminate the four abnormalities from each other with accuracy comparable to human experts. It can also facilitate therapeutic decision-making (e.g., close follow-ups, repeated anti-VEGF injection, or prompt surgery) and help to prevent diseases from getting worse. Prospectively, this system has the potential to enhance diagnostic efficiency and improve patient outcomes with better applicability to the clinical circumstances than simplified binary classification models.

Most AI studies based on OCT focused on the image segmentation, which involved complicated feature selection and extraction.[13,16,28–31] Moreover, a minor error in segmentation would lead to misclassification. But for the deep learning technologies used in this study, no OCT segmentation was needed. These substantially deeper networks (101 layers) can acquire richer and more discriminative image characteristics for more accurate recognition than either low-level, hand-crafted features or CNN with shallower architectures.[1,18] In addition, the data set used to train, validate, and test the deep learning algorithms was larger than those of previous studies. Sufficient data is the premise for good performance in deep learning; otherwise, the algorithm may fail to learn the accurate features of each abnormality and therefore compromise accuracy.

Limitations of this study must be considered. First, the OCT images in this study were collected from only one image center. Device settings, camera systems, and population characteristics may affect the OCT images, and so does the system's performance. Additionally, our data set contained scans from the same patients who underwent the OCT examination in their follow-up study at a different time, which might reduce data diversity and affect algorithms' generalization ability. To further validate this intelligent system, data sets from different eye centers and larger patient cohorts will be required in subsequent studies.

Second, to identify complex OCT images is of great importance toward clinical translation. Due to the limit number of images, we only tried a small number of complicated images to test the system's performance in recognizing OCT scans with multiple abnormalities. Lager data sets of complex OCT images are needed to validate and optimize our system and make it an efficient intelligent tool for the clinical circumstances.

Third, deep learning's nature of "black box" makes it unclear how the algorithms analyze patterns and make decisions at the image level. A visualization of the pixels contributing the most to the algorithm's classification is needed in subsequent studies. This could potentially assist real-time clinical validation and future reviews or analysis for both patients and physicians.[14,32,33]

Overall, this study proposed a novel deep learning-based system that can implement automated categorization of four abnormalities from OCT images with robust performance. This is an initial step toward clinical translation. More data sets and abnormalities should be involved to further optimize the system and make it a practical intelligent tool in clinical circumstances. Moreover, while this study offers a promising framework for an automated identification of abnormalities, single OCT images cannot always guarantee the correct diagnosis of a specific retinal disease (e.g., diabetic retinopathy or glaucoma) in clinical practice. To this end, multimodal clinical images, such as images of OCT angiography, visual field testing, and fundus photography, should be included in the AI diagnosis of retinal diseases. This generalized AI system based on multimodal data has the potential to revolutionize current disease diagnostic pattern and generate a significant clinical impact.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (81470628) and International Science & Technology Cooperation Program of China (2017YFE0103400).

Disclosure: **W. Lu**, None; **Y. Tong**, None; **Y. Yu**, None; **Y. Xing**, None; **C. Chen**, None; **Y. Shen**, None

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
2. Banino A, Barry C, Uria B, et al. Vector-based navigation using grid-like representations in artificial agents. *Nature*. 2018;557:429–433.

3. Bejnordi BE, Zuidhof G, Balkenhol M, et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *J Med Imaging*. 2017;4:044504.
4. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
5. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12:e0174944.
6. van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol Physics Tech*. 2017;10:23–32.
7. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402.
8. Rahimy E. Deep learning applications in ophthalmology. *Curr Opin Ophthalmol*. 2018:1.
9. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthamol Vis Sci*. 2016;57:5200.
10. Adhi M, Duker JS. Optical coherence tomography–current and future applications. *Curr Opin Ophthalmol*. 2013;24:213–221.
11. Gabriele ML, Wollstein G, Ishikawa H, et al. Optical coherence tomography: history, current status, and laboratory work. *Invest Ophthamol Vis Sci*. 2011;52:2425–2436.
12. Chen J, Lee L. Clinical applications and new developments of optical coherence tomography: an evidence-based review. *Clin Exp Optom*. 2007; 90:317–335.
13. ElTanboly A, Ismail M, Shalaby A, et al. A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. *Med Phys*. 2017;44:914–923.
14. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018; 172:1122–1131.e9.
15. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology*. 2018;125:549–558.
16. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017;8: 3440.
17. Cherkassky V, Mulier F. Statistical learning theory. *Encyclopedia Sci Learning*. 1998;41:3185.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*; 2016:770–778.
19. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA*; 2015:1–9.
20. He K, Sun J. Convolutional neural networks at constrained time cost. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA*; 2015:5353–5360.
21. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *Jama Ophthalmol*. 2017;135:1170.
22. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211–252.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
24. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125: 1207–1208.
25. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med*. 2017;82:80–86.
26. Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Med Biol Eng Comput*. 2017; 55:1959–1974.
27. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. *PLoS One*. 2017;12: e0187336.
28. Roy AG, Conjeti S, Karri SPK, et al. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed Opt Exp*. 2017;8:3627.

29. Muhammad H, Fuchs TJ, De CN, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma*. 2017:1.
30. Xu X, Lee K, Zhang L, Sonka M, Abramoff MD. Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical OCT data. *IEEE Trans Med Imaging*. 2015;34:1616–1623.
31. Fang L, Cunefare D, Wang C, Guymer RH, Li S, Farsiu S. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Opt Exp*. 2017;8:2732.
32. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124:962–969.
33. Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. *Med Image Anal*. 2017; 39:178.