

Original Paper

# An Interpretable and Expandable Deep Learning Diagnostic System for Multiple Ocular Diseases: Qualitative Study

Kai Zhang<sup>1,2</sup>, MS; Xiyang Liu<sup>1,3,4</sup>, PhD; Fan Liu<sup>3</sup>, BSc; Lin He<sup>1</sup>, BSc; Lei Zhang<sup>1</sup>, BSc; Yahan Yang<sup>2</sup>, BSc; Wangting Li<sup>2</sup>, BSc; Shuai Wang<sup>3</sup>, BSc; Lin Liu<sup>1</sup>, BSc; Zhenzhen Liu<sup>2</sup>, MD, PhD; Xiaohang Wu<sup>2</sup>, MD, MPH; Haotian Lin<sup>2</sup>, MD, PhD

<sup>1</sup>School of Computer Science and Technology, Xidian University, Xi'an, China

<sup>2</sup>State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

<sup>3</sup>School of Software, Xidian University, Xi'an, China

<sup>4</sup>Institute of Software Engineering, Xidian University, Xi'an, China

**Corresponding Author:**

Haotian Lin, MD, PhD

State Key Laboratory of Ophthalmology

Zhongshan Ophthalmic Center

Sun Yat-sen University

54 Xian Lie South Road

Guangzhou,

China

Phone: 86 (020)87330

Email: [haot.lin@hotmail.com](mailto:haot.lin@hotmail.com)

## Abstract

**Background:** Although artificial intelligence performs promisingly in medicine, few automatic disease diagnosis platforms can clearly explain why a specific medical decision is made.

**Objective:** We aimed to devise and develop an interpretable and expandable diagnosis framework for automatically diagnosing multiple ocular diseases and providing treatment recommendations for the particular illness of a specific patient.

**Methods:** As the diagnosis of ocular diseases highly depends on observing medical images, we chose ophthalmic images as research material. All medical images were labeled to 4 types of diseases or normal (total 5 classes); each image was decomposed into different parts according to the anatomical knowledge and then annotated. This process yields the positions and primary information on different anatomical parts and foci observed in medical images, thereby bridging the gap between medical image and diagnostic process. Next, we applied images and the information produced during the annotation process to implement an interpretable and expandable automatic diagnostic framework with deep learning.

**Results:** This diagnosis framework comprises 4 stages. The first stage identifies the type of disease (identification accuracy, 93%). The second stage localizes the anatomical parts and foci of the eye (localization accuracy: images under natural light without fluorescein sodium eye drops, 82%; images under cobalt blue light or natural light with fluorescein sodium eye drops, 90%). The third stage carefully classifies the specific condition of each anatomical part or focus with the result from the second stage (average accuracy for multiple classification problems, 79%-98%). The last stage provides treatment advice according to medical experience and artificial intelligence, which is merely involved with pterygium (accuracy, >95%). Based on this, we developed a telemedical system that can show detailed reasons for a particular diagnosis to doctors and patients to help doctors with medical decision making. This system can carefully analyze medical images and provide treatment advices according to the analysis results and consultation between a doctor and a patient.

**Conclusions:** The interpretable and expandable medical artificial intelligence platform was successfully built; this system can identify the disease, distinguish different anatomical parts and foci, discern the diagnostic information relevant to the diagnosis of diseases, and provide treatment suggestions. During this process, the whole diagnostic flow becomes clear and understandable to both doctors and their patients. Moreover, other diseases can be seamlessly integrated into this system without any influence on existing modules or diseases. Furthermore, this framework can assist in the clinical training of junior doctors. Owing to the rare high-grade medical resource, it is impossible that everyone receives high-quality professional diagnosis and treatment service.

This framework can not only be applied in hospitals with insufficient medical resources to decrease the pressure on experienced doctors but also deployed in remote areas to help doctors diagnose common ocular diseases.

(*J Med Internet Res* 2018;20(11):e11144) doi: [10.2196/11144](https://doi.org/10.2196/11144)

## KEYWORDS

deep learning; object localization; multiple ocular diseases; interpretable and expandable diagnosis framework; making medical decisions

## Introduction

Although there have been many artificial intelligence-based automatic diagnostic platforms, the diagnostic results produced by such computer systems cannot be easily understood. Artificial intelligence that obtains diagnostic results from the computational perspective cannot provide the reason that is depicted as clinical practice for a given diagnosis. Some researchers have attempted to make the conclusion obtained from artificial intelligence methods explainable, such as Raccuglia et al used a decision tree to understand the classification result from the support vector machine [1]. Hazlett et al used a deep belief network, a reverse trackable neural network, to find diagnostic evidence of autism [2]. Zhou et al used the output of the last full-connected layer of the convolution neural network to infer which part of an image causes the final classification result, which also provides the evidence of classification [3]. In addition, Zeiler et al used occlusion test to study which parts of images produce a given classification result [4]. These studies made great achievements in explainable artificial intelligence, but readily explainable automatic diagnostic systems are still rare. The primary cause is that these explainable methods did not explain their result according to human thought patterns. Therefore, this research aims to make additional progress based on previous studies.

There are many existing works about the automatic diagnosis of different types of diseases with medical imaging, but all these works are isolated; those cannot regard all diseases shown in a specific format of medical images with a unified perspective, which is common in natural image processing and practical medical scenes. On the other hand, once all diseases are regarded as unified, the extensibility for integrating other types of medical imaging or disease will be easy. The diagnosis of ophthalmic diseases is highly dependent on observing medical images, so this work selected ophthalmic images that represent multiple ocular diseases as material and treated them with a consistent view. Of note, the unified automatic diagnostic procedure is the simulation of the work flow of doctors. An explainable artificial intelligence-based automatic diagnosis platform offers many advantages. First, it can increase the confidence in the diagnostic results. Second, it assists doctors to perfect the diagnosing thinking. Third, it helps medical students deepen the medical knowledge. Finally, it can clear a path toward diagnosing higher numbers of diseases from a unified perspective.

Besides, doctors can diagnose diseases by observing medical images, but doctors from many specialties and subspecialties cannot tackle all diseases. If a patient suffers from more than one type of disease, the system can tackle these diseases simultaneously. This work plans to integrate the experience of

doctors from many subspecialties to construct an omnipotent ophthalmologist.

Thus, to create an explainable automatic diagnostic system with artificial intelligence, we simulated the workflow of doctors to help artificial intelligence follow the patterns of human thought. This research aims to apply artificial intelligence techniques to fully simulate the diagnostic process of doctors so that reasons for a given diagnosis can be illustrated directly to doctors and patients.

In this research, we designed an interpretable and expandable framework for multiple ocular diseases. There are 4 stages in this diagnostic framework: primary classification of disease, detection of each anatomical parts and foci, judging the conditions of anatomical parts, and foci and providing treatment recommendations. The accuracies of all stages surpass 93%, 82%-87%, 79%-98%, and 95%, respectively. Not only is this system an interpretable diagnostic tool for doctors and patients but it also facilitates the accumulation of medical knowledge for medical students. Moreover, this system can be enriched to cover more ophthalmic diseases or more diseases of other specialties to provide more services as the workflow of doctors. Telemedicine [5] can combine medical experts and patients with considerable low cost. This research develops an interpretable and expandable telemedical artificial intelligence diagnostic system, which can also effectively improve the undesirable condition that medical resource with high quality is not adequate and the distribution of it is not even. Finally, the health level of people all over the world and the medical condition of underdeveloped countries can be improved with the help of a computer network.

## Methods

### Data Preparation

Data are important for data-driven research [6]. The dataset is examined by all members of our team. Besides, we developed some programs to facilitate the examination of data. All images were collected in the Sun Yat-sen University Zhongshan Ophthalmic Center, which is the leading ophthalmic hospital in China [7]. In order to simulate the experience and diagnostic process of doctors, all images were segmented into several parts according to anatomical knowledge or diagnostic experiences and, then, were annotated. Next, multiple attributes of all parts were classified as the actual states of these parts (including foci). All the relevant aspects of the data (images, coordinates of each part, and the attribute information) were used to train an artificial intelligence system. This data preparation process can not only help simulate the diagnostic process of doctors but also facilitate many follow-up studies such as medical image segmentation,

clinical experience mining, and integration of refined diagnosing of multiple diseases.

We collected 1513 images that can be classified into 5 classes (normal, pterygium [9], keratitis [10], subconjunctival hemorrhage [11], and cataract [12]). Figure 1 lists the number of images of each class. Furthermore, the examples of objects to be detected in images are shown in Figure 2; for fundus images (the last row), the localized objects include an artery (blue), vein (green), the macula (black), the optic disc (light purple), hard exudate (yellow), and so on. For other types of images, the objects to be localized include the eyelid (red), eyelash (green), keratitis focus (yellow), cornea and iris zone with keratitis (pink), the pupil zone (blue), conjunctiva and sclera zone with hyperemia (orange), the conjunctiva and sclera zone with edema (light blue), the conjunctiva and sclera zone with hemorrhage (brown), the pupil zone with cataracts (white), the slit arc of the cornea (black), cornea and iris zone (dark green), the conjunctiva and sclera zone (purple), pterygium (gray), the slit arc of keratitis focus (dark red), and the slit arc of the iris (light brown). Table 1 lists the detailed diagnostic attributes to be classified, and each diagnostic information corresponds to a classification problem. The diagnostic information in Table 1 is corresponding to stage 3 (see Methods). This information is essential and fundamental for diagnosing and providing treatment advice and will be determined in stage 3 of the interpretable artificial intelligence system (see Methods). All information (object annotation and diagnostic information) was double-blind marked by the annotation team, which consisted of 5 experienced ophthalmic doctors and 20 medical students. The annotation of fundus images was completed; however, the experiments on fundus images were not finished. Because of the intrinsic characteristics

of the fundus image, the output of the annotation method for fundus image is suitable for semantic segmentation.

### Methodology

The framework consists of 4 functional stages as follows: (1) judging the class of disease, preliminary diagnosis that is completed with original image without any processing; (2) detecting each part of image, localization of anatomical parts, and foci that are used to discern different parts with different appearance so that more careful checking can be guaranteed; (3) classifying the attributes of each part, severity and illness assessment, which is closely connected to the second stage, is used to determine the condition of the illness; and (4) providing treatment advice according to the results from the first, second, and third stages, except for the treatment advice of a pterygium is from artificial intelligence, whereas treatment advice of other diseases is from experiences of doctors. First, the disease is primarily identified during stage 1. Second, all anatomical parts and foci are localized during stage 2, and important parts (cornea and iris zone with keratitis and pterygium) are segmented for the analysis in stage 3. Then, the attributes of all anatomical parts and foci are determined during stage 3. Then, the treatment advice is provided in stage 4. The whole process imitates the diagnostic procedure of doctors so that the reasons for a given diagnosis can be tracked and used to construct an evidence-based diagnostic report. Finally, treatment advice can be provided according to the full workflow presented above. Figure 3 shows the flowchart of this system. The analysis of fundus images is coming soon and will be easily integrated into this system quickly as the same idea with existing images. The first, second, and third function is fully based on artificial intelligence, which is trained with dataset; the fourth function is dependent on both artificial intelligence and the experience of doctors.

Figure 1. Information of image dataset.

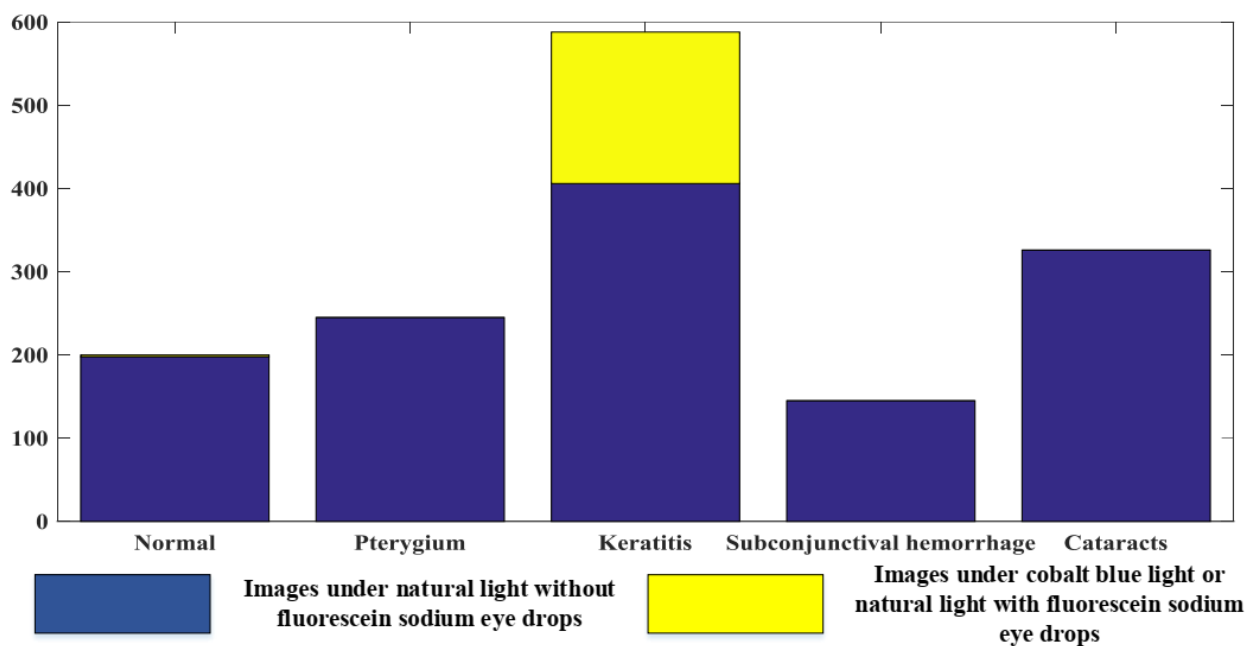


Figure 2. Examples of each object in terms of each type of disease or normal eye.

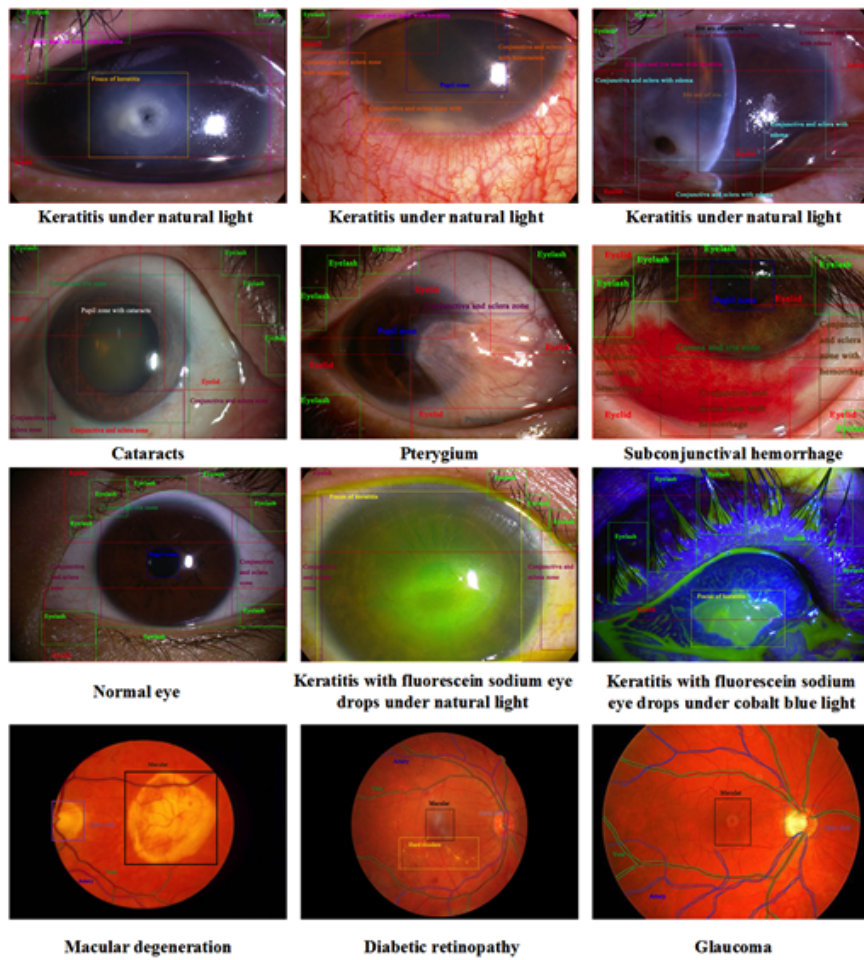


Table 1. Detailed diagnostic information regarding the dataset.

Disease	Diagnostic information (Number of classification problems)	Values of diagnostic information	Type of image
Pterygium	<ol style="list-style-type: none"> <li>Whether the body of the pterygium is hypertrophied</li> <li>Whether pseudo pterygium is present</li> <li>Whether the head of the pterygium is uplifted</li> <li>Whether the head and body of the pterygium is hyperemic</li> <li>Whether the pterygium is in the progressive period</li> </ol>	Yes or no	Images under natural light without fluorescein sodium eye drops
Keratitis	<ol style="list-style-type: none"> <li>Turbidity degree of the cornea</li> <li>Stage of keratitis</li> <li>Corneal neovascularization</li> <li>Edge of foci is clear</li> <li>The condition of illness based on dyeing</li> </ol>	<ol style="list-style-type: none"> <li>Pupil zone is invaded by turbidity or not</li> <li>Infiltration stage and ulcer stage, perforation stage, or convalescence</li> <li>Yes or no</li> <li>No dyeing and dot staining, sheet dyeing, or dyeing with coloboma</li> </ol>	Images under cobalt blue light or natural light with fluorescein sodium eye drops [8]

Machine learning, especially deep learning technique represented by the convolutional neural network (CNN), is becoming the effective computer vision tool for automatically diagnosing diseases using biomedical images. It has been widely

applied in the medical image classification and automatic diagnosis of disease, such as the diagnosis of attention deficit hyperactivity disorder with functional magnetic resonance imaging [13]; gradation of brain tumor [14], breast cancer [15],

and lung cancer [16]; and diagnosis of skin disease [17], kidney disease [18], and ophthalmic diseases [19-23]. In this research, inception\_v4 [24] and residual network (Resnet) [25] (101 layers) were used to carry out stage 1 and stages 3 and 4, respectively. While stage 1 (inception\_v4) can give a general diagnostic conclusion, stages 3 (Resnet) and 4 (Resnet) can provide further information about diseases and treatment recommendations. In this research, cost-sensitive CNN was adopted because the imbalanced classification is common in this research. Inception\_v4 is a wider and deeper CNN that is suitable for careful classification (the difference between all classes is easily neglected sometimes). Resnet is a type of thin CNN, the architecture of which is full of cross-layer connections.

The objective function is transformed to fit the residual function so that the performance of Resnet is improved considerably. In addition, Resnet is suitable for rough classification (the difference between all classes does not need to be carefully analyzed). In addition, we chose Resnet with 101 layers whose volume is adequate for the classification problems in this research. Stage 1 is a 5-classes classification, with some classes being very similar in color and shape; thus, inception\_v4 is chosen in stage 1. As other classification problems are limited in one specific disease, Resnet is selected in stages 3 and 4. Furthermore, the chain rule of derivatives based on the stochastic gradient descent algorithm [26] was used to minimize the loss function.

**Figure 3.** Architecture of the overall framework for interpretable diagnosis of multiple ocular diseases. AI: artificial intelligence.



Faster-region based convolutional neural network (RCNN), an effective and efficiency approach, was adopted to localize the anatomical parts and foci (Stage 2). Faster-RCNN [27] is developed on the basis of RCNN [28] and Fast-RCNN [29], which originally applied superpixel segmentation algorithm to produce proposal regions, whereas Faster-RCNN uses an anchor mechanism to generate region proposals quickly and then adopts 2-stage training to obtain the transformations of bounding box repressor and classifier. The first stage of Faster-RCNN is region proposal network, which is responsible for generating region proposals. Then, whether the proposals are objects or not are judged, and the coordinates of each object are primary regressed. The second stage is judging the class of each object and eventually regressing the coordinate of each object, which is the same as RCNN and Fast-RCNN. In this research, pretrained ZF (Zeiler and Fergus [4]) network was exploited to save training time.

### Experimental Settings

This system was implemented with convolutional architecture for fast feature embedding [30] (Berkeley Vision and Learning Center deep learning framework) and Tensorflow [31]; all models were trained in parallel on four NVIDIA TITAN X GPUs. For the classification problem, indicators applied to evaluate the performance are as follows:

$$Accuracy = \frac{\sum_{i=1}^k TP_i}{N}$$

$$Precision_i = TP_i / (TP_i + FP_i) \quad [TP_i + FP_i]$$

$$Sensitivity_i TPR, Recall_i = TP_i / (TP_i + FN_i) \quad [TP_i + FN_i]$$

$$FNR_i \text{ false-negative rate} = FN_i / (TP_i + FN_i) \quad [TP_i + FN_i]$$

$$Specificity_i = TN_i / (TN_i + FP_i) \quad [TN_i + FP_i]$$

$$FPR_i \text{ (false-positive rate)} = FP_i / (TN_i + FP_i) \quad [TN_i + FP_i]$$

where  $N$  is the total number of samples;  $P_i$  indicates the number of correctly classified samples of  $i$  th class;  $k$  is the number of classes in specific classification problem;  $TP_i$  denotes the number of samples that are correctly classified as  $i$  th class;  $FP_i$  is the number of samples that are wrongly recognized as  $i$  th class;  $FN_i$  denotes the number of samples that are classified as  $j$  th class,  $j \in [1, c] / i$ ;  $TN_i$  is the number of samples recognized as negative  $j$  th class,  $j \in [1, c] / i$ . All the above performance indicators can be computed with a confusion matrix. In addition, the receiver operating characteristics (ROC) curve, which indicates how many samples of  $i$  th class are recognized conditioned on a specific number of  $j$  th class ( $j \in [1, c] / i$ ), are classified as  $i$  th class, PR (precision recall) curve, which illustrates how many samples of  $j$  th class are recognized as samples of  $i$  th class conditioned on a specific number of  $j$  th class ( $j \in [1, c] / i$ ), are classified as  $i$  th class and area under the ROC curve (AUC), which means the area of the zone under the

ROC curve was also adopted to assess the performance [32]. The indicators (precision, sensitivity, specificity, ROC curve with AUC, and PR curve) were only used to evaluate the performance of binary classification problems. Furthermore, accuracy and confusion matrix were used to evaluate the performance of multiclass classification problems.

For object localization problem, the interpolated average precision is always used to evaluate the performance [33]. The interpolated average precision is computed with the PR curve using the equation presented below:



In the equation,  $p(\eta)$  is the measured precision at specific recall  $\eta$ . In this research, 4-fold cross-validation was used to evaluate the performance of this system firmly for all classification problems and localization problems. The application of the cost-sensitive CNN is dependent on the distribution of the dataset in specific classification problems. Except for the classification problems 1, 6, and 8, other classification problems in stages 3 and 4 were completed with the cost-sensitive CNN.

## Results

### Performance of Stages 1 and 2

All stages and the whole work flow of this system were completed with acceptable performance. The 4 stages in the framework were separately trained and validated, and all relevant results in stages 1 and 2 are shown in Figures 4 and 5. The rows and columns of all heat maps stand for ground truth labels and predicted labels, respectively. Figure 4 shows the heat map of stage 1; the accuracy reaches 92%. Figure 5 shows the detection performance of Faster-RCNN in recognizing anatomical parts and foci; the mean value of average precision over all classes surpasses 82% and 90% for images under natural light without fluorescein sodium eye drops, and images under cobalt blue light or natural light with fluorescein sodium eye drops, respectively. The left image in Figure 5 is the performance for localizing objects in images without fluorescein sodium eye drops during stage 2, where I-VX represent the cornea and iris zone with keratitis, the focus of keratitis, the conjunctiva and sclera zone, the slit arc of the cornea, the slit arc of keratitis focus, the eyelid, the slit arc of the iris, the conjunctiva and sclera zone with hyperemia, the conjunctiva and sclera zone with edema, cornea and iris zone, pterygium, eyelash, pupil zone, the conjunctiva and sclera zone with hemorrhage, and the pupil zone with cataracts, respectively. The right image in Figure 5 presents the performance for localizing the objects in images with fluorescein sodium eye drops during stage 2, where I-VII represent the cornea and iris zone with keratitis, the focus of keratitis, the slit arc of the cornea, the slit arc of keratitis focus, the slit arc of the iris, the eyelid, and the eyelash, respectively. The statistical results of stage 2 are shown in Multimedia Appendix 1.

**Figure 4.** Performance of stage 1.

**Figure 5.** Performance of stage 2. AP: average precision.



### Performance of Stages 3 and 4

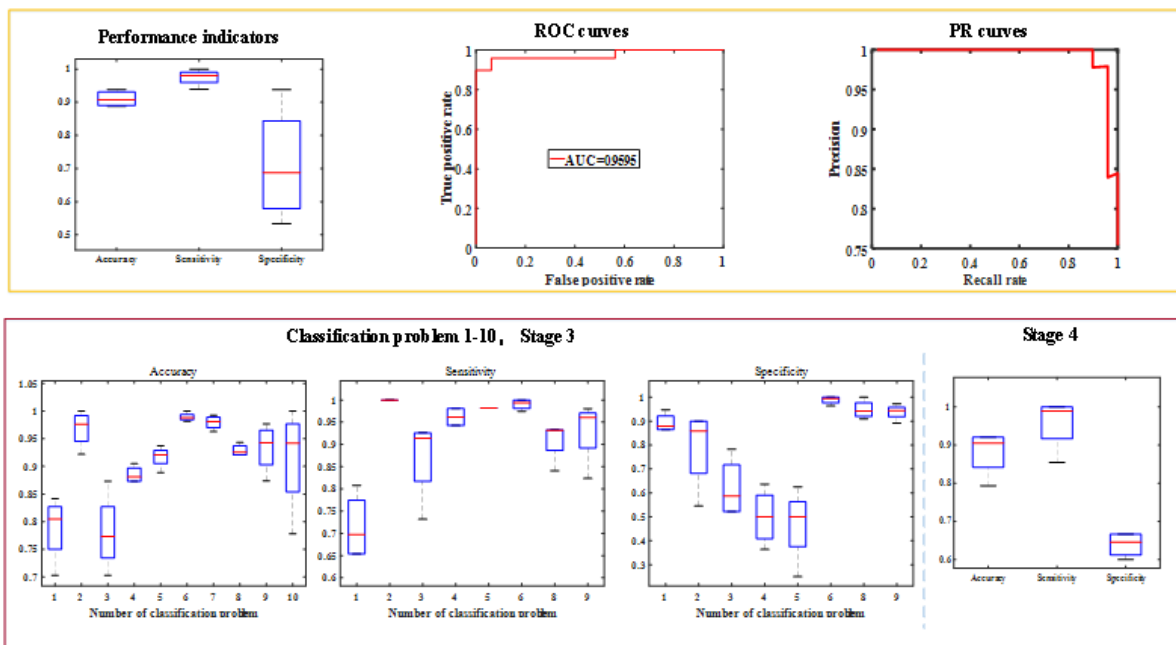
Stage 3 was decomposed into 10 classification problems, and the relevant results are shown in Figure 6, including the boxplots for the accuracy, specificity and sensitivity, ROC curve with the AUC, PR curve for all binary classification problems, and the heat maps with accuracy for all multiclass classification problems. Figure 6 also shows the classification performance of stage 4, which includes boxplot for the accuracy, sensitivity and specificity, ROC curve with the AUC value and PR curve. The only one classification problem addressed by stage 4 is whether a patient who suffers from pterygium needs surgery. In stage 2, the detection rate of some objects is low because Faster-RCNN cannot effectively detect some small objects. We will overcome this issue by adjusting the parameters of Faster-RCNN. In spite of this, stage 3 will not be affected by this drawback because the detection rate of the cornea and iris zone with keratitis and pterygium (the relevant anatomical parts and foci), which is involved with stage 3, is considerably high. In addition, the detection performance of the pupil zone, which is related to vision is also satisfactory. In stage 3, the specificity

of classification problems 1, 3, 4, and 5 is slightly low; the application scene of this system is hospitals where doctors pay more attention to sensitivity than specificity. The result of all classification problems is satisfactory and acceptable. Furthermore, the performance of classification problems 1, 3, 4, and 5 can be improved with more samples under the circumstance of Web-based learning. The statistical results of stages 3 and 4 are shown in Multimedia Appendix 1.

### Performance of Stage 3 and 4 with Original Images

To study which anatomical parts are essential for automatic diagnostic, stages 3 and 4 were repeated with original medical images without processing; all parameters were same as the original parameters used in stages 3 and 4. The relevant results are shown in Figure 7. The classification performance close to that of the classification with anatomical parts and foci. In other words, the important parts, the cornea and iris zone with keratitis and pterygium, are essential for automatic diagnosis. The statistical results of stages 3 and 4 with original images are shown in Multimedia Appendix 1.

Figure 6. Performance of stage 3 and 4. PR: precision recall; ROC: receiver operating characteristics.



**Figure 7.** Performance of stage 3 and 4 with original images. PR: precision recall; ROC: receiver operating characteristics; AUC: area under the curve.



### Web-Based Automatic Diagnostic System

We applied Django framework [34] to develop a telemedical decision-making and automatic diagnosing system to facilitate doctors and patients; this system can analyze inputted medical images, show the diagnostic result as the working process of doctors, and provide treatment advice by producing an examination report. In addition, this telemedical system can finely analyze medical images and provide treatment advice with a diagnostic report (a PDF file) that includes treatment suggestion according to the analysis result and the consultation between a doctor and a patient. The format of the diagnostic report is shown in [Multimedia Appendix 1](#). All diagnostic information can be shown to a doctor and a patient by storing into a database. Administrators and doctors can handle all information and contact patients conveniently. Furthermore, this system can be deployed in multiple hospitals and medical

centers to screen common diseases and collect more medical data, which can be used to improve the diagnosis performance. The website is available in [Multimedia Appendix 1](#).

### Discussion

In this study, we constructed an explainable artificial intelligence system for the automatic diagnosis of multiple ophthalmic diseases. This system carefully mimics the work flow of doctors so that reasons for specific diagnosis can be explained to doctors and patients with high performance. Besides, this system accelerates the application of telemedicine with the assistance of computer network and helps develop the health level and medical condition. Moreover, this system can be easily expanded to cover more diseases as long as the diagnostic processes of other diseases are simulated seamlessly. In addition, this system can help medical students to understand diagnosis and diseases.

In the future, considerable progress can be made in this field. In this research, we did not consider a multilabel classification for those patients with multiple diseases. In the future, multiple-label classification can be adopted to make this system closer to real clinical circumstances. Moreover, because the bound box is not suitable for some anatomical parts, semantic segmentation can be applied in this system for segmenting medical images more accurately.

### Acknowledgments

This study was funded by the National Key Research and Development Program (2018YFC0116500); the National Science Foundation of China (#91546101, #61472311, #11401454, #61502371, and #81770967), National Defense Basic Research Project of China (jcky2016110c006), the Guangdong Provincial Natural Science Foundation (#YQ2015006, #2014A030306030, #2014TQ01R573, and #2013B020400003), the Natural Science Foundation of Guangzhou City (#2014J2200060), The Guangdong Provincial Natural Science Foundation for Distinguished Young Scholars of China (2014A030306030), the Science and Technology Planning Projects of Guangdong Province (2017B030314025), the Key Research Plan for the National Natural Science Foundation of China in Cultivation Project (#91546101), the Ministry of Science and Technology of China Grants (2015CB964600), and the Fundamental Research Funds for the Central Universities (#16ykjc28). We gratefully thank the volunteers of AINIST (medical artificial intelligence alliance of Zhongshan School of Medicine, Sun Yat-sen University).

### Authors' Contributions

XL and HL designed the research; KZ conducted the study; WL, ZL, and XW collected the data and prepared the relevant information; KZ, FL, LH, LZ, LL, and SW were responsible for coding; LH and LZ developed the Web-based system; KZ analyzed and completed the experimental results; and KZ, WL, HL, and XL cowrote the manuscript. HL critically revised the manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Relevant material.

[\[DOCX File, 42KB-Multimedia Appendix 1\]](#)

### References

1. Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* 2016 May 05;533(7601):73-76. [doi: [10.1038/nature17439](https://doi.org/10.1038/nature17439)] [Medline: [27147027](https://pubmed.ncbi.nlm.nih.gov/27147027/)]
2. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, IBIS Network, Clinical Sites, Data Coordinating Center, Image Processing Core, Statistical Analysis. Early brain development in infants at high risk for autism spectrum disorder. *Nature* 2017 Dec 15;542(7641):348-351 [FREE Full text] [doi: [10.1038/nature21369](https://doi.org/10.1038/nature21369)] [Medline: [28202961](https://pubmed.ncbi.nlm.nih.gov/28202961/)]
3. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. 2016 Jun 26 Presented at: International Conference on Computer Vision and Pattern Recognition; June 26th - July 1st, 2016; Las Vegas p. 2015-2019.
4. Zeiler M, Fergus R. Visualizing and Understanding Convolutional Networks. 2014 Sep 6 Presented at: European Conference on Computer Vision; September 6th-12th, 2014; Zurich.
5. Perednia DA, Allen A. Telemedicine technology and clinical applications. *JAMA* 1995 Feb 08;273(6):483-488. [Medline: [7837367](https://pubmed.ncbi.nlm.nih.gov/7837367/)]
6. Vibhu A, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis. *Journal of Medical Internet Research* 2016 Sep 21;18(9):A.
7. Lin H, Long E, Chen W, Liu Y. Documenting rare disease data in China. *Science* 2015 Sep 04;349(6252):1064. [doi: [10.1126/science.349.6252.1064-b](https://doi.org/10.1126/science.349.6252.1064-b)] [Medline: [26339020](https://pubmed.ncbi.nlm.nih.gov/26339020/)]
8. Amparo F, Wang H, Yin J, Marmalidou A, Dana R. Evaluating Corneal Fluorescein Staining Using a Novel Automated Method. *Investigative Ophthalmology & Visual Science* 2017 Jul 15;58(6):BIO168-BIO173.
9. Hirst LW. Treatment of pterygium. *Aust N Z J Ophthalmol* 1998 Nov;26(4):269-270. [Medline: [9843252](https://pubmed.ncbi.nlm.nih.gov/9843252/)]
10. Prajna N, Krishnan T, Rajaraman R, Patel S, Shah R, Srinivasan M, Mycotic Ulcer Treatment Trial Group. Predictors of Corneal Perforation or Need for Therapeutic Keratoplasty in Severe Fungal Keratitis: A Secondary Analysis of the Mycotic Ulcer Treatment Trial II. *JAMA Ophthalmol* 2017 Sep 01;135(9):987-991 [FREE Full text] [doi: [10.1001/jamaophthalmol.2017.2914](https://doi.org/10.1001/jamaophthalmol.2017.2914)] [Medline: [28817744](https://pubmed.ncbi.nlm.nih.gov/28817744/)]
11. Welch JF, Dickie AK. Red Alert: diagnosis and management of the acute red eye. *J R Nav Med Serv* 2014;100(1):42-46. [Medline: [24881426](https://pubmed.ncbi.nlm.nih.gov/24881426/)]

12. Lin H, Ouyang H, Zhu J, Huang S, Liu Z, Chen S, et al. Lens regeneration using endogenous stem cells with gain of visual function. *Nature* 2016 Mar 17;531(7594):323-328 [FREE Full text] [doi: [10.1038/nature17181](https://doi.org/10.1038/nature17181)] [Medline: [26958831](https://pubmed.ncbi.nlm.nih.gov/26958831/)]
13. Riaz A, Asad M, Alonso E, Slabaugh G. Fusion of fMRI and non-imaging data for ADHD classification. *Comput Med Imaging Graph* 2018 Apr;65:115-128. [doi: [10.1016/j.compmedimag.2017.10.002](https://doi.org/10.1016/j.compmedimag.2017.10.002)] [Medline: [29137838](https://pubmed.ncbi.nlm.nih.gov/29137838/)]
14. Mohan G, Subashini MM. MRI based medical image analysis survey on brain tumor grade classification. *Biomedical Signal Processing & Control* 2018;39(1):139-161.
15. Golden JA. Deep Learning Algorithms for Detection of Lymph Node Metastases From Breast Cancer: Helping Artificial Intelligence Be Seen. *JAMA* 2017 Dec 12;318(22):2184-2186. [doi: [10.1001/jama.2017.14580](https://doi.org/10.1001/jama.2017.14580)] [Medline: [29234791](https://pubmed.ncbi.nlm.nih.gov/29234791/)]
16. Yu K, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016 Dec 16;7:12474 [FREE Full text] [doi: [10.1038/ncomms12474](https://doi.org/10.1038/ncomms12474)] [Medline: [27527408](https://pubmed.ncbi.nlm.nih.gov/27527408/)]
17. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 Dec 02;542(7639):115-118. [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)] [Medline: [28117445](https://pubmed.ncbi.nlm.nih.gov/28117445/)]
18. Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM, et al. Association of Pathological Fibrosis With Renal Survival Using Deep Neural Networks. *Kidney Int Rep* 2018 Mar;3(2):464-475 [FREE Full text] [doi: [10.1016/j.ekir.2017.11.002](https://doi.org/10.1016/j.ekir.2017.11.002)] [Medline: [29725651](https://pubmed.ncbi.nlm.nih.gov/29725651/)]
19. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nature Biomedical Engineering* 2017 Jan 30;1(2):0024.
20. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology* 2017 Dec;124(7):962-969. [doi: [10.1016/j.ophtha.2017.02.008](https://doi.org/10.1016/j.ophtha.2017.02.008)] [Medline: [28359545](https://pubmed.ncbi.nlm.nih.gov/28359545/)]
21. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016 Dec 13;316(22):2402-2410. [doi: [10.1001/jama.2016.17216](https://doi.org/10.1001/jama.2016.17216)] [Medline: [27898976](https://pubmed.ncbi.nlm.nih.gov/27898976/)]
22. Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin KW, et al. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Med Image Anal* 2012 Jan;16(1):216-226. [doi: [10.1016/j.media.2011.07.004](https://doi.org/10.1016/j.media.2011.07.004)] [Medline: [21865074](https://pubmed.ncbi.nlm.nih.gov/21865074/)]
23. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: [10.1001/jama.2017.18152](https://doi.org/10.1001/jama.2017.18152)] [Medline: [29234807](https://pubmed.ncbi.nlm.nih.gov/29234807/)]
24. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. 2017 Feb 4 Presented at: AAAI Conference on Artificial Intelligence; February 4th- 9th, 2017; San Francisco, USA.
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 Jun 26 Presented at: International Conference on Computer Vision and Pattern Recognition; June 26th - July 1st, 2016; Las Vegas, USA.
26. Ketkar N. Parallelized stochastic gradient descent. 2010 Presented at: Advances in neural information processing systems; December 6TH -11th, 2010; Vancouver Canada.
27. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. : International Conference on Neural Information Processing Systems; 2015 Presented at: Advances in Neural Information Processing Systems; December 7th -12th, 2015; Montreal Canada.
28. Girshick R, Donahue J, Darrell T, Malik J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans Pattern Anal Mach Intell* 2016 Jan;38(1):142-158. [doi: [10.1109/TPAMI.2015.2437384](https://doi.org/10.1109/TPAMI.2015.2437384)] [Medline: [26656583](https://pubmed.ncbi.nlm.nih.gov/26656583/)]
29. Girshick R. Fast R-CNN. 2015 Dec 13 Presented at: IEEE International Conference on Computer Vision; December 13th - 16th, 2015; Santiago, Chile.
30. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. In: Caffe: Convolutional Architecture for Fast Feature Embedding. Caffe: Acm International Conference on Multimedia; 2014 Nov 3 Presented at: Acm International Conference on Multimedia; November 3rd - 7th, 2014; Orlando, Florida, USA.
31. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. eprint arXiv.0. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems URL:<https://arxiv.org/abs/1603.04467> [accessed 2018-09-24] [WebCite Cache ID 72fPz3sU]
32. Shi C, Pun CM. Superpixel-based 3D deep neural networks for hyperspectral image classification. *Pattern Recognition* 2018 Feb;74:600-616. [doi: [10.1016/j.patcog.2017.09.007](https://doi.org/10.1016/j.patcog.2017.09.007)]
33. Everingham M, Zisserman A, Williams C, Gool L, Allan M, Bishop C, et al. The 2005 PASCAL Visual Object Classes Challenge. In: The 2005 PASCAL Visual Object Classes Challenge. Berlin, Heidelberg: Springer; 2006:117-176.
34. Holovaty A, Kaplan-Moss J. The Definitive Guide to Django: Web development done right. In: The Definitive Guide to Django: Web development done right. California, American: Apress; 2009.

## Abbreviations

**AUC:** area under receiver operating characteristics curve  
**CNN:** Convolutional Neural Network  
**PR:** precision recall  
**RCNN:** region based convolutional neural network  
**Resnet:** residual network  
**ROC:** receiver operating characteristics

*Edited by G Eysenbach; submitted 26.05.18; peer-reviewed by M Banf, L Zhang, Q Pan; comments to author 12.07.18; revised version received 31.07.18; accepted 02.08.18; published 14.11.18*

*Please cite as:*

Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, Li W, Wang S, Liu L, Liu Z, Wu X, Lin H

*An Interpretable and Expandable Deep Learning Diagnostic System for Multiple Ocular Diseases: Qualitative Study*

*J Med Internet Res 2018;20(11):e11144*

URL: <http://www.jmir.org/2018/11/e11144/>

doi: [10.2196/11144](https://doi.org/10.2196/11144)

PMID: [30429111](https://pubmed.ncbi.nlm.nih.gov/30429111/)

©Kai Zhang, Xiyang Liu, Fan Liu, Lin He, Lei Zhang, Yahan Yang, Wangting Li, Shuai Wang, Lin Liu, Zhenzhen Liu, Xiaohang Wu, Haotian Lin. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.