Diabetes Ther (2019) 10:1811–1822
https://doi.org/10.1007/s13300-019-0652-0

ORIGINAL RESEARCH

An Evaluation System of Fundus Photograph-Based Intelligent Diagnostic Technology for Diabetic Retinopathy and Applicability for Research

Wei-Hua Yang · Bo Zheng · Mao-Nian Wu · Shao-Jun Zhu · Fang-Qin Fei · Ming Weng · Xian Zhang · Pei-Rong Lu

Received: April 12, 2019 / Published online: July 9, 2019

ABSTRACT

Introduction: In April 2018, the US Food and Drug Administration (FDA) approved the world's first artificial intelligence (AI) medical device for detecting diabetic retinopathy (DR), the IDx-DR. However, there is a lack of evaluation systems for DR intelligent diagnostic technology.

Methods: Five hundred color fundus photographs of diabetic patients were selected. DR severity varied from grade 0 to 4, with 100 photographs for each grade. Following that, these were diagnosed by both ophthalmologists and the intelligent technology, the results of which were compared by applying the evaluation system. The system includes primary, intermediate, and advanced evaluations, of which the intermediate evaluation incorporated two methods. Main evaluation indicators were sensitivity, specificity, and kappa value.

Results: The AI technology diagnosed 93 photographs with no DR, 107 with mild non-proliferative DR (NPDR), 107 with moderate NPDR, 108 with severe NPDR, and 85 with proliferative DR (PDR). The sensitivity, specificity, and kappa value of the AI diagnoses in the primary evaluation were 98.8%, 88.0%, and 0.89, respectively. According to method 1 of the

Wei-Hua Yang and Bo Zheng contributed equally to this work.

Enhanced Digital Features To view enhanced digital features for this article go to https://doi.org/10.6084/m9.figshare.8256950.

W.-H. Yang · P.-R. Lu (✉)
Department of Ophthalmology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China
e-mail: lupeirong@suda.edu.cn

W.-H. Yang
Department of Ophthalmology, The First People's Hospital of Huzhou, Huzhou, Zhejiang, China

B. Zheng · M.-N. Wu · S.-J. Zhu
The Information Engineering College of Huzhou University, Huzhou, Zhejiang, China

W.-H. Yang · B. Zheng · M.-N. Wu · S.-J. Zhu · F.-Q. Fei
Key Laboratory of Medical Artificial Intelligence, Huzhou University, Huzhou, Zhejiang, China

M. Weng
Department of Ophthalmology, Wuxi Third People's Hospital, Wuxi, Jiangsu, China

X. Zhang
Department of Ophthalmology, Ningbo Medical Center Lihuili Eastern Hospital, Ningbo, Zhejiang, China

F.-Q. Fei
Department of Endocrinology, The First Affiliated Hospital of Huzhou University, Huzhou, Zhejiang, China

intermediate evaluation, the sensitivity of AI diagnosis was 98.0%, specificity 97.0%, and the kappa value 0.95. In method 2 of the intermediate evaluation, the sensitivity of AI diagnosis was 95.5%, the specificity 99.3%, and kappa value 0.95. In the advanced evaluation, the kappa value of the intelligent diagnosis was 0.86.

Conclusions: This article proposes an evaluation system for color fundus photograph-based intelligent diagnostic technology of DR and demonstrates an application of this system in a clinical setting. The results from this evaluation system serve as the basis for the selection of scenarios in which DR intelligent diagnostic technology can be applied.

Keywords: Deep learning; Diabetic retinopathy; Evaluation studies; Ophthalmological diagnostic techniques

INTRODUCTION

Diabetes is a common disease worldwide, with an estimated 422 million adults suffering from it as of 2014 [1, 2]. According to the World Health Organization, this number has quadrupled since 1980 and is projected to rise rapidly [1, 3]. Diabetic retinopathy (DR) is a microvascular complication of diabetes and carries the risk of developing into severely impaired vision or blindness as well as diabetic macular edema. With appropriate laser photocoagulation and timely intraocular injection of vascular endothelial growth factor inhibitors, patients can be spared the potential blindness caused by retinopathy and macular edema. As early stage retinopathy may be asymptomatic, regular eye examinations in diabetic patients are very important for diagnosis of this disease.

Traditionally, diagnosis of DR mainly relies on color fundus photographs obtained by mydriatic or non-mydriatic fundus cameras which are then diagnosed by experienced ophthalmic specialists. Relying entirely on ophthalmologists to diagnose a large number of fundus photographs is very inefficient and makes it difficult to complete a large number of DR screening tasks. Furthermore, a doctor's inexperience and physical or mental exhaustion may lead to diagnostic errors. As there are limited ophthalmologists and even fewer are engaging in the initial screening of DR, the ratio of doctors to patients is extremely low, especially in China.

Artificial intelligence (AI) technology is now available which can be used to obtain preliminary diagnostic results of the disease. Using AI has the advantage of creating time for the ophthalmologists so that they can focus their skills on the further review and confirmation of abnormal results. This reduces the burden on doctors and greatly improves the efficiency of diagnosis and treatment. At the same time, because the machine relies on data rather than experience to make judgments, its diagnostic results are more reliable owing to the minimal influence of subjective factors. Therefore, the application of AI technology to DR diagnosis can improve the diagnostic efficiency of doctors. While these advantages are clear, further validation of the accuracy of AI methods would continue to encourage its use.

Various DR intelligent diagnostic technologies exist, which often adopt the machine learning technology of AI, mainly achieved through deep learning technology [4–9]. Machine learning is a category of algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. Deep learning is a set of algorithms in machine learning that attempt to learn layered models of inputs, commonly neural networks. These technologies are evaluated on the basis of calculated sensitivity, specificity, and kappa values to judge whether the technology meets the standard of diagnosis [10, 11]. The diagnostic results presented by these technologies are based on the international clinical classification of diabetic retinopathy [12]. However, not all diagnostic technologies are able to achieve good results based on these parameters. This therefore limits the application of these devices depending on the level of sophistication required by the medical institution using the device. It is of high importance to establish a standardized and unified evaluation system of DR intelligent diagnostic technologies prior to their

application to clinical practice. This need has developed greater urgency with the FDA approval in April 2018 of the world's first AI medical device for testing DR, the IDx-DR. IDx-DR is an AI diagnostic system that autonomously analyzes images of the retina for signs of diabetic retinopathy. The autonomous AI system, IDx-DR, has two core algorithms, an image quality AI-based algorithm, and the diagnostic algorithm proper. For people with diabetes, autonomous AI systems have the potential to improve earlier detection of DR, and thereby lessen the suffering caused by blindness and visual loss. Therefore, this study proposes an evaluation system for DR intelligent diagnostic technologies and discusses its application value in intelligent diagnostic technologies.

METHODS

Subjects

Five hundred color photographs of diabetic patients' fundi were selected from the Intelligent Ophthalmology Database of Zhejiang Society for Mathematical Medicine in China. The DR severity varied from grades 0 to 4, with 100 photographs selected for each grade. The photographs were then diagnosed by both professional ophthalmologists and the intelligent technology as described in Section "Methods". The fundus photographs were taken by a nonmydriatic fundus color camera, with a maximum of one image per eye. These were macula lutea-centered 45° color fundus photographs requiring high readability with no obvious image blur caused by manual operation. We used 45° color fundus photographs for two reasons: (1) the publicly available fundus photo sets are all 45° photos and (2) most hospitals that are equipped with fundus cameras are only able to get 45° photos. The required selection criteria were as follows:

1. In addition to diabetic retinopathy-related features such as proliferative membrane, preretinal hemorrhage, and vitreous hemorrhage, 90% of the blood vessels in the photograph should be identifiable.
2. The main fundus structures such as optic disc and macula should be in the correct position.
3. No shadows and/or highlighted reflective areas that affect interpretation were within the imaging range.
4. Exposure should be moderate, meaning with no overexposure or underexposure.
5. There should be no staining on the lens, no shielding shadows from eyelids and/or eyelashes, and no motion artifacts.
6. There should be no other errors in the fundus photograph, such as absence of objects in the picture, inclusion of non-fundus areas, etc.

METHODS

Data Anonymization

Five hundred color fundus photographs were selected from the Intelligent Ophthalmology Database of Zhejiang Society for Mathematical Medicine in China. Since it is the photographs of the fundus not the patients themselves that were used in the study, and data anonymization was applied before the study, the Ethics Committee of Huzhou University decided that neither consent from the patients nor approval from them was required.

Clinical Diagnostic Group

The color fundus photograph of the same eye was evaluated by three retina-trained ophthalmologists. Each ophthalmologist was asked to make an independent diagnosis of the DR fundus photographs. Final diagnostic results of the specialists were achieved when the same diagnosis was given three times. And if the diagnoses of the three physicians were the same, the diagnosis would be confirmed. Otherwise we would invite another two physicians to consult and make a final diagnosis. Those with inconsistent diagnoses were evaluated by another set of two senior specialists whose diagnoses would be taken as the final clinical diagnosis.

According to the international classification of diabetic retinopathy [12, 13], the clinical diagnostic group diagnosed and classified the subjects into five grades: no DR, mild non-proliferative DR (NPDR), moderate NPDR, severe NPDR, and proliferative DR (PDR). We refer to the four categories of the above international classification as grade 0–4, respectively. See Table 1 for details of these classifications.

Intelligent Diagnostic Group

Our team developed an intelligent diagnostic system which was based on a deep learning algorithm acquired through transfer learning. The selected training samples were 10,000 color fundus photographs with grades 0–4 DR as diagnosed by specialists. VGGNet was adopted for training [14]. VGGNet is a deep convolutional neural network developed by researchers from Visual Geometry Group and the Google DeepMind Corp. Finally, the international classification of diabetic retinopathy was set as the diagnostic standard [12, 13]. In this study, we used the DR intelligent diagnostic technology developed by our team to perform intelligent diagnosis. We uploaded the 500 color fundus photographs to the AI diagnostic system and generated the intelligent diagnostic reports. These reports were named the intelligent diagnostic group.

As per the international classification of DR, the intelligent diagnostic group [12, 13] also divided the photographs into five grades, namely no DR, mild NPDR, moderate NPDR, severe NPDR, and PDR. See Table 1 for details.

Evaluation System

We then compared the diagnoses made by the intelligent diagnostic technology against those made by the clinical diagnostic group using our proposed evaluation system. The evaluation system consists of primary evaluation, intermediate evaluation, and advanced evaluation. In the primary evaluation, we calculated the consistency rate of the diagnosis of with or without DR between the two groups, where grade 0 in the above international classification standard counts as no DR, while grades 1–4 mean the subject tested has DR. In the intermediate evaluation, the consistency rate of the severity of the diagnosed DR was calculated. In advanced evaluation, the consistency rate for DR grading (grades 0–4) was calculated.

Table 1 International classification of diabetic retinopathy

Diabetic retinopathy	Findings observable on dilated ophthalmoscopy
No apparent DR	No abnormalities
Mild NPDR	Microaneurysms only
Moderate NPDR	More than just microaneurysms, but less than severe non-proliferative DR
Severe NPDR	Any of the following
	Intraretinal hemorrhages ($\geq$ 20 in each quadrant)
	Definite venous beading (in two quadrants)
	Intraretinal microvascular abnormalities (in 1 quadrant)
	No signs of proliferative retinopathy
PDR	Severe non-proliferative DR and one or more of the following
	Neovascularization
	Vitreous/preretinal hemorrhage

IRMA intraretinal microvascular abnormalities, *NPDR* non-proliferative retinopathy, *PDR* proliferative retinopathy

For the intermediate evaluation, we did a comparative experiment of two different evaluation methods. Combining the international classification standard, method 1 classified grades 0 and 1 as mild DR and grades 2, 3, and 4 as severe DR. Accordingly, the sensitivity, specificity, and kappa value of the intelligent diagnostic technology were calculated. Method 2 classified grades 0, 1, and 2 as mild DR and grades 3 and 4 as severe DR. Similarly, the sensitivity, specificity, and kappa values were calculated. We compared the results of these two methods in order to more comprehensively evaluate our system. The sensitivity in the intermediate evaluation was calculated as the ratio of correct diagnoses of severe DR, and specificity was calculated as the ratio of correct diagnoses of mild DR.

Statistical Analysis

The statistical method used was that the of the SPSS 18.0 software package which evaluates diagnostic tests. The results of this were represented in a fourfold table for diagnostic tests. The statistical data was the number of eyes evaluated and the statistical indicators included sensitivity, specificity, and consistency of the diagnostic test (namely the kappa value). A kappa value of 0.61–0.80 was considered to be significantly consistent and one of greater than 0.80 was considered to be highly consistent.

RESULTS

In the study, the intelligent diagnostic technology diagnosed 93 (18.6%) fundus photographs as showing no DR, 107 (21.4%) as mild NPDR, 107 (21.4%) as moderate NPDR, 108 (21.6%) as severe NPDR, and 85 (17.0%) as PDR. The kappa value of the advanced evaluation for the intelligent diagnostic technology was 0.86. Specialist and intelligent diagnostic results using the international classification method are shown in Table 2.

In the primary evaluation, when the intelligent diagnostic technology group was used, 93 patients had no DR, while the number of patients who had DR was 407. The corresponding sensitivity, specificity, and kappa value were 98.8%, 88.0%, and 0.89 (95% CI 0.83–0.94), respectively. The crude agreement of intelligent diagnosis was 96.6%. The crude agreement is defined as the percentage of cases where the intelligent system and the physicians reached the same diagnosis in all the diagnosed cases. The comparison of primary evaluation between the specialist diagnostic results and the intelligent diagnostic results is shown in Table 3.

In the intermediate evaluation, method 1 identified a total of 200 cases with mild DR and 300 cases with severe DR. The sensitivity, specificity, and kappa value were 98.0%, 97.0%, and 0.95 (95% CI 0.92–0.98), respectively. The crude agreement of the intelligent group diagnosis was 97.6%. According to method 2, 307 cases of mild DR were identified and 193 cases of severe DR. The sensitivity, specificity, and kappa value was 95.5%, 99.3%, and 0.95 (95% CI 0.93–0.98), respectively. The crude agreement of the intelligent diagnosis was 97.8%. Comparison between the specialist and intelligent diagnostic results using method 1 of the intermediate evaluation, where grades 0 and 1 are classified as mild DR, is shown in Table 4. Table 5 shows the comparison between the specialist and intelligent diagnostic results in intermediate evaluation method 2, where grades 0, 1, and 2 are classified as mild DR.

In the advanced evaluation, the kappa values was 0.86 (95% CI 0.83–0.89), the quadratic weighted kappa value was 0.97 (95% CI 0.96–0.98). The crude agreement of the intelligent diagnosis was 88.8%. A comparison of the sensitivity, specificity, and kappa values of the three evaluation stages is shown in Table 6.

There exists high consistency between the intelligent diagnosis group and the clinician group in the primary, intermediate, and advanced evaluations.

DISCUSSION

Findings of the diabetic retinopathy study (DRS) group and the early treatment diabetic retinopathy study (ETDRs) group confirm that effective treatment prevents severe vision loss in

Table 2 Comparison of specialist and intelligent diagnostic results using the international classification method

Clinical diagnosis	Intelligent diagnosis					Total
	No DR	Mild NPDR	Moderate NPDR	Severe NPDR	PDR	
No DR	88	12	0	0	0	100
Mild NPDR	5	89	6	0	0	100
Moderate NPDR	0	6	92	2	0	100
Severe NPDR	0	0	8	91	1	100
PDR	0	0	1	15	84	100
Total	93	107	107	108	85	500

Table 3 Comparison between the specialist and intelligent diagnostic results in the primary evaluation

Specialist diagnosis	Intelligent diagnosis		Total
	with DR	No DR	
With DR	395	5	400
No DR	12	88	100
Total	407	93	500

Table 4 Comparison between the specialist and intelligent diagnostic results using method 1 in the intermediate evaluation

Specialist diagnosis	Intelligent diagnosis		Total
	Severe DR	Mild DR	
Severe DR	294	6	300
Mild DR	6	194	200
Total	300	200	500

90% of DR patients and reduces the blindness rate to less than 5% from 50% [15]. Applying AI technology to DR diagnosis allows for quick acquisition of preliminary diagnostic results, which reduces time for diagnosis and treatment, saving time both for the doctors and the patients. This new technology has created a great deal of interest worldwide on the topic of AI technology for DR screening [16–24].

Table 5 Comparison between the specialist and intelligent diagnostic results using method 2 in the intermediate evaluation

Specialist diagnosis	Intelligent diagnosis		Total
	Severe DR	Mild DR	
Severe DR	191	9	200
Mild DR	2	298	300
Total	193	307	500

With the continuous advancement in AI, intelligent diagnosis has developed rapidly. The vast majority of the existing intelligent diagnostic technologies use deep learning which employs a large number of labeled data (images, text, etc.) as training samples to generate a relatively well-developed diagnosis model. The model is then optimized through validation and finally tested by double-blind clinical experiments. If results are sufficiently accurate, the model will be used in clinical situations with continuous optimization during use.

The color fundus photograph-based intelligent diagnostic technology used in this study was primarily developed by our team by using deep learning algorithms. DR intelligent diagnosis involves using machine learning to automatically detect DR in color fundus photographs and to achieve diagnosis of DR. Through transfer learning, we adopted the VGG model to train the data and develop the DR

Table 6 Comparison results of the sensitivity, specificity, and kappa values of the three evaluation stages

Evaluation indicators	Primary evaluation	Intermediate evaluation		Advanced evaluation
		Method 1	Method 2	
Sensitivity	98.8%	98.0%	95.5%	–
Specificity	88.0%	97.0%	99.3%	–
Kappa (95% CI)	0.89 (0.83–0.94)	0.95 (0.92–0.98)	0.95 (0.93–0.98)	0.86 (0.83–0.89)
				0.97* (0.96–0.98)

*Quadratic weighted kappa

diagnostic system. Transfer learning is the act of transferring the knowledge of one field into the learning of another field. The study adopted VGGNet; that is to transfer the framework and parameters of VGG16 model trained by ImageNet data set to our team's design with some parameter adjustment. Intelligent diagnostic reports were then generated by diagnostic analysis of 500 color fundus photographs.

Currently some traditional classification algorithms first pre-process the inputted images to manually extract the characteristics with which the classification model is trained. Traditional classification algorithms are very dependent on these manually extracted characteristics of DR [25–27]. However, the lesion characteristics which are present in DR color fundus photographs are complex. It is difficult to obtain good classification results based entirely on manually extracted characteristics.

In contrast, deep convolutional neural network (DCNN) in deep learning, which was used by our team, can automatically extract features by algorithms to obtain good classification results. Convolutional neural network (CNN) is a feedforward neural network with convolution computation and deep structure, which mainly consists of input layers, convolution layers, pool layers, and fully connected layers. Using a variety of convolution cores and inputted color photographs, the convolution layer performs convolutional operation, which has translation invariance that can support neurons to learn the characteristics with relatively high robustness [28–31]. The typical CNN structure is shown in Fig. 1. It is useful to understand deep learning.

At present, many intelligent diagnostic systems exist for various diseases [21, 32–34]. But no systematic and comprehensive evaluation system exists to evaluate these intelligent diagnostic technologies. This results in difficulties promoting the application of these technologies to medical professionals as they need to be assured that the technology is accurate. In this paper, we put forward an evaluation system of the DR intelligent diagnostic technology, which divided the evaluation into three stages, namely primary, intermediate, and advanced stage evaluations. Additionally, comparative experiments were done to select the best method in the intermediate evaluation.

The different stages in the evaluation system of DR intelligent diagnosis in this study correspond to different levels of need at different types of medical institutions. Therefore, this technology can adapt to the needs of initial DR diagnosis, which would be performed at basic-level hospitals and more sophisticated hierarchical diagnosis and treatment which would be needed at more specialized hospitals. See Table 7 for the detailed information of the three-stage evaluation system. For example, the primary evaluation only distinguishes between patients with DR and patients with no DR, which is simple and intuitive. Even for a non-ophthalmic physician at a community hospital, they would be able to judge whether the result is correct and give reasonable suggestions. If the primary evaluation results are sufficiently accurate, such intelligent diagnostic technologies could even be used in scenarios where there are no professional medical personnel onsite (such as shopping malls, residential

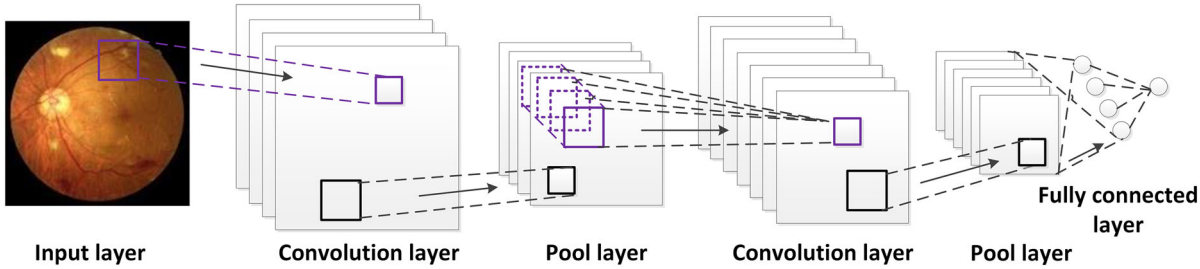

Fig. 1 Typical convolutional neural network (CNN) structure

Table 7 Evaluation system of intelligent diagnostic technology for DR

Stage of evaluation	Objects	Criteria for judgment	Suggested application institutions
Primary	All patients using DR intelligent diagnostic technology	Diagnosed with DR or with no DR	Primary care in community hospitals or health examination institutions
Intermediate	Patients who have been diagnosed with DR	Diagnosed to be mild or severe DR	Ophthalmology departments in county hospitals or in community hospitals
Advanced	All patients using DR intelligent diagnostic technology	Diagnose DR into grades 0–4	Hospitals at the municipal level and above

communities, etc.) as an effective method of initial DR screening. The intermediate evaluation of this diagnostic technology involved evaluating its ability to distinguish between mild and severe grades of DR. A DR intelligent diagnostic technology that meets the standard to be able to perform this type of classification could be used to provide auxiliary diagnostic advice for basic-level medical institutions and medical examination centers that lack ophthalmic professionals. In the case of severe DR patients, referral advice could be provided, achieving hierarchical diagnosis and treatment. Furthermore, advanced evaluation could distinguish among grades 0–4 DR as defined by the international classification method, which could act as an assistant to ophthalmic professionals.

This study confirms that the evaluation system can be relatively well applied to the evaluation of DR intelligent diagnostic technology. This evaluation system fills the gap caused by a lack of evaluation systems for intelligent diagnostic technologies, and could be used as a reference for the evaluation of similar intelligent medical diagnostic technologies.

In this study, specialists were asked to diagnose 500 color photographs of diabetic patients' fundi and classified them according to the grade 0–4 classification system. One hundred photographs of each grade were provided to the intelligent diagnostic system for diagnosis. The results of intelligent diagnosis showed that of the 500 color photographs of diabetic patients' fundi, 93 (18.6%) were with no DR, 107 were with mild non-proliferative DR (NPDR) (21.4%), 107 were with moderate NPDR (21.4%), 108 were with severe NPDR (21.6%), and 84 were with proliferative DR (PDR) (16.8%).

On the basis of the evaluation system proposed in this study, the sensitivity, specificity and kappa value of intelligent diagnosis in the primary evaluation were 98.8%, 88.0% and 0.89, respectively. In intermediate evaluation method 1, the sensitivity, specificity and kappa value of intelligent diagnosis were 98.0%, 97.0% and 0.95, respectively. In intermediate evaluation method 2, the sensitivity, specificity, and kappa value of intelligent diagnosis were 95.5%, 99.3%, and 0.95, respectively. In the advanced evaluation, the kappa value was 0.86

and the quadratic weighted kappa value was 0.97.

In examining the specific results of our evaluation, as shown in Table 6, in the primary and intermediate evaluations, the system's task was to divide the subjects into two categories, which required relatively easy training. Hence, the sensitivity and specificity were relatively high. In contrast, in the advanced evaluation, the system needed to identify and classify the subjects into grades 0–4 DR, which required many more characteristics to be identified, resulting in a slightly worse consistency of test results. The intermediate evaluation divided the subjects into mild DR and severe DR with method 1 defining grades 0 and 1 as mild DR and method 2 defining grades 0–2 as mild DR. The experimental data showed that the sensitivity of method 1 of the intermediate evaluation method was 2.5% higher than that of intermediate evaluation method 2. While the kappa values of methods 1 and 2 are equivalent, the specificity of method 1 is 2.3% lower than that of method 2. Although in accordance with international diagnostic criteria [12], laser treatment and other interventions are required for grade 3 or higher DR, we recommend the use of method 1 in the intermediate evaluation in the consideration of reducing missed diagnoses of patients with severe DR, despite the fact that its diagnosis error rate is higher than that of intermediate evaluation method 2. Our reasoning is that if the intelligent technology mistakenly diagnoses mild DR to be severe DR, these diagnoses will usually require a follow-up examination, and this will not significantly affect their subsequent treatment. Another consideration is that DR is often accompanied by diabetic macular edema (DME) especially in patients with moderate to severe NPDR and PDR [12]. In order to avoid the missed diagnosis of DME, the evaluation method which divides grades 2, 3, and 4 into severe DR is of more practical value.

The IDx-DR, designed by an ophthalmologist at the University of Iowa in the USA, was approved by the FDA in April 2018 for detecting the conditions of moderate or severe DR in adults with diabetes. Its sensitivity and specificity, using an evaluation method similar to intermediate evaluation method 1 as described in this paper, were 87.2% and 90.7%, respectively [21]. Huang et al. constructed an AI deep learning algorithm model to assist in the diagnosis of DR. The sensitivity and specificity of their two-category model (grades 0 and 1 as class 1; grades 2, 3, and 4 as class 2) were 79.5% and 95.3%, respectively (based on the criteria of this study) [35]. The evaluation methods of these studies are similar to that of intermediate evaluation method 1 in our proposed evaluation system. Our evaluation system is therefore suitable for the evaluation of all DR intelligent diagnostic technology in the present and the foreseeable future.

CONCLUSIONS

DR intelligent diagnostic technology based on deep learning designed through transfer learning can achieve high sensitivity and specificity in primary and intermediate evaluations, and is suitable for the initial screening of diabetic patients. The primary, intermediate, and advanced three-stage DR evaluation system proposed in this paper can be applied to different types of hospitals, which fulfills the goal of providing an initial diagnosis of DR at basic-level facilities and allowing for hierarchical diagnosis and treatment. Although both the primary and intermediate evaluation methods divide the subjects into two categories, their evaluation results still have reference value for the DR intelligent diagnostic application scenario selection.

There are some limitations to this study. In this study, the small number of samples may lead to low generalization of the evaluation system. In addition, this study did not consider DME, which is very commonly seen in DR. Furthermore, the photographs used here were rigorously selected to ensure good quality, which could be very different in real-world situations. Hence further research is required to fully realize the clinical application of this evaluation system and AI technology. Along with the continuous advancement of DR intelligent diagnosis technology, the evaluation system of fundus photograph-based intelligent

diagnostic technology for diabetic retinopathy will also gradually improve.

ACKNOWLEDGEMENTS

The authors would like to thank the participants of the studies involved and Prof. Jian Wu from Zhejiang University for his generous help and support of this project.

Funding. This research project received funding from the Zhejiang Basic Public Welfare Research Program (LGF18H120003), the Natural Science Foundation of Zhejiang Province (LQ18F020002), and the Zhejiang Medical and Health Research Project (2018270516). The article processing charges were funded by the authors.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Authorship Contributions. During the study, all authors have contributed significantly. Wei-Hua Yang and Bo Zheng, Fang-Qin Fei analyzed data and drafted the manuscript. Mao-Nian Wu and Shao-Jun Zhu analyzed data and revised the draft. Wei-Hua Yang, Bo Zheng, Mao-Nian Wu, Shao-Jun Zhu and Pei-Rong Lu developed the DR intelligent diagnostic system. Wei-Hua Yang, Ming Weng, Xian Zhang and Pei-Rong Lu participated in the diagnosis as specialists. Bo Zheng, Mao-Nian Wu, Shao-Jun Zhu and Fang-Qin Fei conducted intelligent diagnosis. Pei-Rong Lu designed the experiment and built initial constructs. Pei-Rong Lu and Wei-Hua Yang together proposed the idea and supervised the project.

Disclosures. The authors Wei-Hua Yang, Bo Zheng, Mao-Nian Wu, Shao-Jun Zhu, Fang-Qin Fei, Ming Weng, Xian Zhang, and Pei-Rong Lu have nothing to disclose.

Compliance with Ethics Guidelines. This article does not contain any studies with human participants or animals performed by any of the authors. Since it is the photographs of the fundus not the patients themselves that were used in the study, and data anonymization was applied before the study, the Ethics Committee of Huzhou University decided that neither consent from the patients nor approval from them was required.

Data Availability. All data generated or analyzed during this study are included in this published article.

REFERENCES

1. World Health Organization. Global report on diabetes. http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf?ua=1. Accessed 4 Mar 2019.
2. Tracey ML, Mchugh SM, Fitzgerald AP, Buckley CM, Canavan RJ, Kearney PM. Trends in blindness due to diabetic retinopathy among adults aged 18–-69 years over a decade in Ireland. Diabetes Res Clin Pract. 2016;121:1–8.
3. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract. 2014;103:137–49.
4. Grewal PS, Oloumi F, Rubin U, Tennant MTS. Deep learning in ophthalmology: a review. Can J Ophthalmol. 2018;53:309–13.
5. Raman R, Srinivasan S, Virmani S, Sivaprasad S, Rao C, Rajalakshmi R. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. Eye. 2019;33:97–109.

6. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172:1122–31.

7. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A. Development of a deep learning algorithm for automatic diagnosis of diabetic retinopathy. Stud Health Technol Inform. 2017;245:559–63.

8. Yu FL, Sun J, Li A, Cheng J, Wan C, Liu J. Image quality classification for DR screening using deep learning. In: 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2017, pp 664–667.

9. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunovic H. Artificial intelligence in retina. Prog Retin Eye Res. 2018;67:1–29.

10. Sengupta S, Singh A, Leopold HA, Lakshminarayanan V. Ophthalmic diagnosis and deep learning–a survey. arXiv:1812.07101. 2018.

11. Lou Y, Yang WH, Xu DY, et al. Intelligent Medicine Introduction. Beijing: China Railway Publishing; 2018. p. 216–218.

12. International Council of Ophthalmology. ICO guidelines for diabetic eye care [EB/OL]. http://www.icoph.org/enhancing_eyecare/diabetic_eyecare.html. Accessed 4 Mar 2019.

13. National Technical Guidance Group for Blindness Prevention. Guidelines for the prevention and treatment of diabetic retinopathy in China (for primary medical institutions). Beijing: People's Medical Publishing; 2017. p. 5–6.

14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations, San Diego, CA; 2015.

15. Brown AF, Jiang L, Fong DS, et al. Need for eye care among older adults with diabetes mellitus in fee-for-service and managed Medicare. Arch Ophthalmol. 2005;123:669–75.

16. 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;316:2402–10.

17. Quellec G, Charriere K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. Med Image Anal. 2017;39:178–93.

18. Gargiya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology. 2017;124:962–9.

19. Chandrakumar T, Kathirvel R. Classifying diabetic retinopathy using deep learning architecture. Int J Eng Res Tech. 2016;5:19–24.

20. 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 Ophthalmol Vis Sci. 2016;57:5200–6.

21. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Dig Med. 2018;1:39.

22. Pekala M, Joshi N, Freund DE, Bressler NM, DeBuc DC, Burlina PM. Deep learning based retinal OCT segmentation. arXiv:1801.09749. 2018.

23. Zhang Y, Chung ACS. Deep supervision with additional labels for retinal vessel segmentation task. In: International conference on medical image computing and computer-assisted intervention; September, 2018; Granada, Spain.

24. Liu ZF, Zhang YZ, Liu PZ, et al. Retinal vessel segmentation using densely connected convolution neural network with colorful fundus images. J Med Imaging Health Inform. 2018;8:1300–7.

25. Bu W, Wu X, Chen X, Dai B, Zheng Y. Hierarchical detection of hard exudates in color retinal images. J Soft. 2013;8:2723–32.

26. Al-juboori AM, Bu W, Wu X, Zhao Q. Palm vein verification using Gabor filter. Int J Comput Sci. 2013;10:678–84.

27. Chen X. Automatic detection methods of exudates on diabetic retinal image. Harbin: Harbin Institute of Technology; 2012. p. 32–9.

28. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962;160:106–54.

29. Fukushima K, Miyake S. Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. Competition and cooperation in neural nets. In: Proceedings of the US–Japan joint seminar. February, 1982; Kyoto, Japan. p. 267–85.

30. Sangeethaa SN, Uma Maheswari P. An intelligent model for blood vessel segmentation in diagnosing DR using CNN. J Med Syst. 2018;42:175.

31. Chudzik P, Majumdar S, Calivá F, Al-Diri B, Hunter A. Microaneurysm detection using fully convolutional neural networks. Comput Methods Programs Biomed. 2018;158:185–92.

32. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–8.

33. Orringer DA, Pandian B, Niknafs YS, et al. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. Nat Biomed Eng. 2017;1:0027.

34. Long E, Lin H, Liu Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng. 2017;1:0024.

35. Huang X, Gu S, Ma XY, et al. Artificial intelligence of diabetic retinopathy image recognition used in the real world. Technol Intell Eng. 2018;4:24–30.