Graefe's Archive for Clinical and Experimental Ophthalmology (2022) 260:1663–1673
https://doi.org/10.1007/s00417-021-05503-7

MEDICAL OPHTHALMOLOGY

End-to-end diabetic retinopathy grading based on fundus fluorescein angiography images using deep learning

Zhiyuan Gao[1] · Kai Jin[1] · Yan Yan[1] · Xindi Liu[1] · Yan Shi[2] · Yanni Ge[1] · Xiangji Pan[1] · Yifei Lu[3] · Jian Wu[3] · Yao Wang[1] · Juan Ye[1]

Received: 7 May 2021 / Revised: 11 October 2021 / Accepted: 14 November 2021 / Published online: 23 January 2022

Abstract

Purpose To develop and validate a deep learning system for diabetic retinopathy (DR) grading based on fundus fluorescein angiography (FFA) images.

Methods A total of 11,214 FFA images from 705 patients were collected to form the internal dataset. Three convolutional neural networks, namely VGG16, RestNet50, and DenseNet, were trained using a nine-square grid input, and heat maps were generated. Subsequently, a comparison between human graders and the algorithm was performed. Lastly, the best model was tested on two external datasets (Xian dataset and Ningbo dataset).

Results VGG16 performed the best, with a maximum accuracy of 94.17%, and had an AUC of 0.972, 0.922, and 0.994 for levels 1, 2, and 3, respectively. For Xian dataset, our model reached the accuracy of 82.47% and AUC of 0.910, 0.888, and 0.976 for levels 1, 2, and 3. As for Ningbo dataset, the network performed with the accuracy of 88.89% and AUC of 0.972, 0.756, and 0.945 for levels 1, 2, and 3.

Conclusions A deep learning system for DR staging was trained based on FFA images and evaluated through human–machine comparisons as well as external dataset testing. The proposed system will help clinical practitioners to diagnose and treat DR patients, and lay a foundation for future applications of other ophthalmic or general diseases.

Keywords Diabetic retinopathy · Fundus fluorescein angiography · Deep learning · Grading

Key messages

- Currently, artificial intelligence has already played an important role in a series of studies on automatic diabetic retinopathy (DR) diagnosis, and achieved excellent performances. Past studies were mainly conducted based on color fundus photography (CFP), and few studies have focused on more precise DR diagnosis based on fundus fluorescein angiography (FFA) images.
- FFA is a powerful tool for precisely assessing DR severity. We collected a large internal dataset consisted of 11214 FFA images, and developed a deep learning system for DR grading based on the dataset. We validated the best model on two external datasets (Xian dataset and Ningbo dataset) to show the ability of wide application.
- Considering the difficulties of dealing with the input of FFA images, we came up with a 'nine-square grid input' method in this study. Compared with other input methods, the model using nine-square grid inputs yielded the best results.

Zhiyuan Gao and Kai Jin contributed equally to this work and share the first authorship.

✉ Yao Wang
wangyao@zju.edu.cn

✉ Juan Ye
yejuan@zju.edu.cn

[1] Department of Ophthalmology, The Second Affiliated Hospital of Zhejiang University, College of Medicine, Hangzhou 310009, China

[2] Department of Ophthalmology, Ningbo First Hospital, Ningbo 315010, China

[3] College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

Introduction

Diabetic retinopathy (DR) is the most frequent cause of preventable blindness across the globe [1–3]. It has been estimated that the number of DR patients will increase to more than 200 million by 2040 [2], resulting in a tremendous social burden [4]. However, vision loss mainly occurs in the advanced stages of DR [5], and the most effective treatment for DR can be achieved in the early stages [6]. Therefore, it is necessary to provide patients with precise diagnoses of DR grade.

Clinical DR screening and diagnosis are typically performed based on color fundus photography (CFP) or fundus fluorescein angiography (FFA) images [7–9]. CFP is a simple and effective examination for DR screening because it is rapid, non-invasive, and widely available [10, 11]. However, FFA can detect typical pathological changes such as microaneurysms, non-perfusion regions, and vascular leakage and provides dynamic information regarding retinal blood vessels that CFP cannot identify [1, 7, 12]. FFA is a more powerful tool for precisely assessing DR severity, which directly guides individual treatment plans and plays a crucial role in DR diagnosis.

The traditional procedure for DR diagnosis is time-consuming, and its accuracy depends on the personal experience of experts. Deep learning is a machine learning technique that avoids complex engineering tasks by learning the most predictive features from a labeled dataset of medical images [13, 14]. This technology has already played an important role in a series of studies on automatic DR screening [15–18]. Gulshan's deep learning algorithm for DR detection has achieved excellent performance and can increase the efficiency, reproducibility, and coverage of screening programs [19]. Sayres et al. used a convolutional neural network (CNN) model to assist in DR grading and provided direct evidence for classification using explanatory heat maps [20]. Regardless of those great results, the existing CFP-based DR screening systems showed rather poor sensitivity in the diagnosis of early DR, and the binary diagnosis might not be directly enough to guide further treatment. For another, in many DR-related artificial intelligence studies, the data were heterogeneous, thus restricting their application in real clinical environment. For a comprehensive investigation and better application, the acquisition of multi-center data is essential.

On the other hand, few studies have focused on more precise DR diagnosis based on FFA images [21–23]. Firstly, unlike CFP images, there is a lack of FFA image dataset, because FFA images are relatively hard to obtain. Then, an FFA report always contains nine or more fundus images, and former studies focused on either single FFA image or a continuous input of each image from the report, making the information not fully used or the whole process pretty complex. Recently, Pan et al. compared three CNNs for the multi-label classification of DR lesions and laid a foundation for the automatic analysis of FFA images and DR diagnosis [12].

In this study, we came up with a novel "nine-square grid input" to deal with the input of FFA images and then developed a deep learning algorithm based on FFA images for DR grading. We trained and validated the system on an internal dataset and evaluated its functionality through human–machine comparisons and multi-center dataset testing. Its quick and precise DR grading can potentially aid decision-making for other ophthalmic or general diseases.

Methods

An overview of the study approach and methodology is presented in Fig. 1. This is a sub-analysis of FFA data from a clinical study (A New Technique for Retinal Disease Treatment, ClinicalTrials.gov identifier: NCT04718532).

FFA report acquisition

In this study, we retrospectively collected 11,214 FFA images from 1114 eyes (563 right eyes and 551 left eyes) of 705 diabetes patients (age range of 22 to 85 years, male-to-female ratio of 1.39:1) who received FFA examinations at the Eye Center at the Second Affiliated Hospital of Zhejiang University School of Medicine between August 2016 and July 2020. Images of both early and late phase of different retina fields are included in this dataset. The images are JPEG format, with a resolution of 768*768. Operations were performed by two experienced ophthalmologists using the Heidelberg retina angiograph (Heidelberg Engineering, Heidelberg, Germany) with a 30° field of view. Subjects were not included if their fundus could not be photographed because of a turbid medium or overexposure.

Annotation

The DR severity level of each eye was determined based on a set of images from an FFA examination report. Images were graded according to the international clinical diabetic retinopathy (ICDR) disease severity scale [24], as shown in Table 1. Examples of lesions in FFA images with different DR severity levels are presented in Fig. 2.

Three ophthalmologists were involved in the grading project. They were considered to be experienced and well trained prior to this project. Labels were determined without previous communication, and a ground-truth label was confirmed if there was no divergence for a single eye. Otherwise, another round of discussion and adjudication was

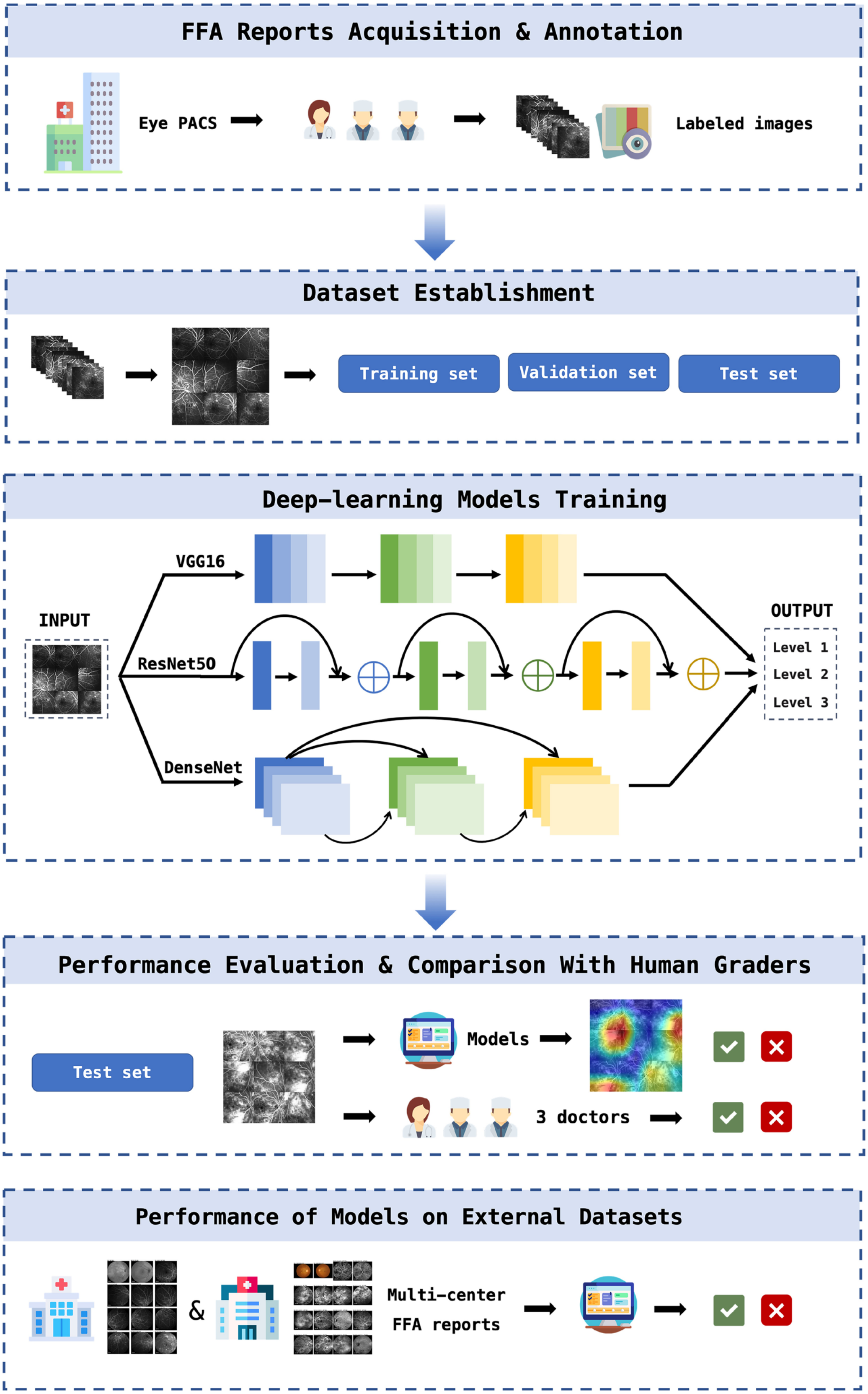

Fig. 1 Illustration of the entire procedure

Table 1 Standard for DR severity levels

Severity level	ICDR severity scale	Findings on FFA images
1	Mild DR and moderate DR	Microaneurysm-like punctate or clustered intense fluorescence; or small sheet capillary leaking fluorescence; or non-perfusion areas ≤ 5 DD
2	Severe DR	4 quadrants with extensive microaneurysm-like fluorescence or retinal hemorrhagic obscured fluorescence and 1 quadrant with severe intra-retinal microvascular abnormalities; or IRMA; or non-perfusion areas ≥ 5 DD
3	Proliferative DR	Retinal neovascularization, vitreous hemorrhage or pre-retinal hemorrhage

ICDR international clinical diabetic retinopathy, *FFA* fundus fluorescein angiography, *DR* diabetic retinopathy, *DD* disc diameter, *IRMA* intra-retinal microvascular abnormalities

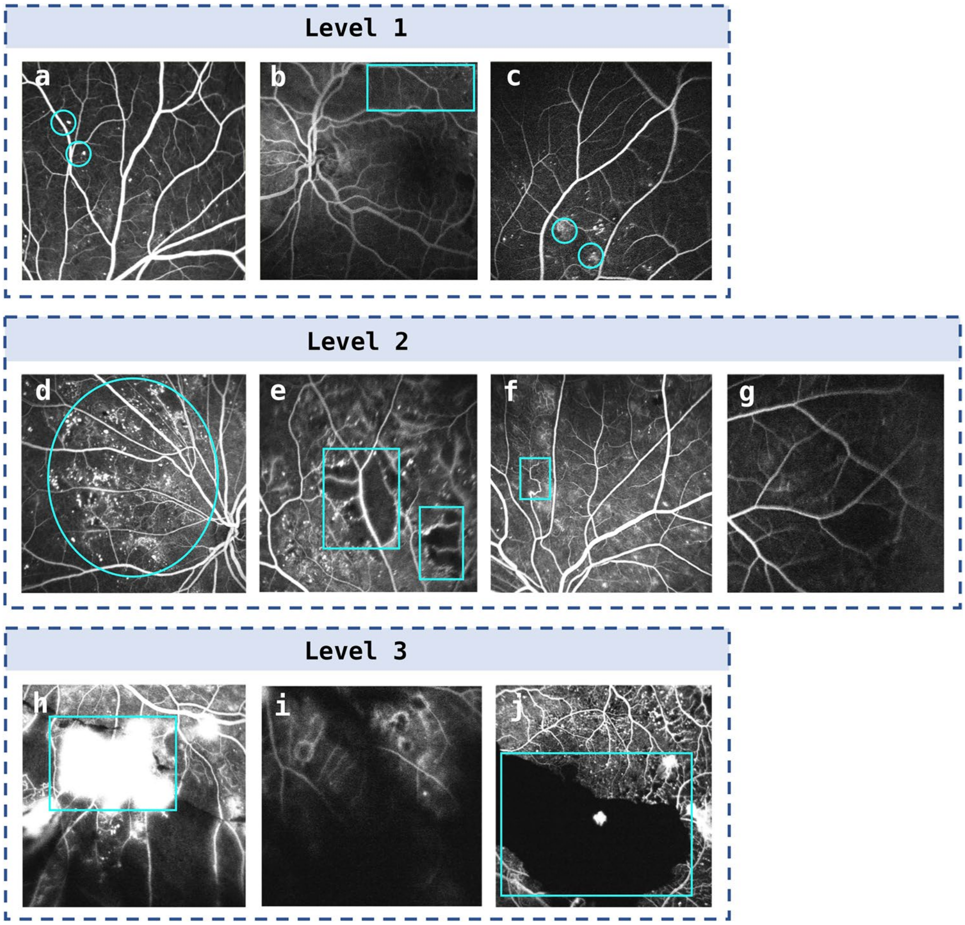

Fig. 2 Examples of FFA images with different DR severity levels. **a** Images with mild microaneurysm-like punctate fluorescence. **b** Images with non-perfusion areas ≤ 5 DD in size. **c** Images with small sheet capillary leaking fluorescence. **d** Images with extensive micro-aneurysm-like fluorescence and intra-retinal microvascular abnormalities (IRMAs). **e** Images with retinal hemorrhagic obscured fluorescence. **f** Images with IRMAs (indicated by red squares). **g** Images with non-perfusion areas ≥ 5 DD in size. **h** Images with fluorescence leakage caused by retinal neovascularization. **i** Images with pre-retinal hemorrhage. **j** Images with vitreous hemorrhage

required, and the final results were decided by the majority rule.

Dataset establishment

Only images considered gradable by human graders were used for further AI analysis. Techniques such as rotating, flipping, or otherwise were not used to increase the number of images. Before training CNNs, we randomly divided the entire dataset into training, validation, and testing sets. The training set was used to train the models, the validation set was used to validate the results of each training cycle (epoch), and the testing set was used to validate the functionality of the models.

Training of deep learning models

Because severity levels were determined based on sets of images, a suitable method was required to handle inputs. Several methods were tested, and a nine-square grid input provided the best results. The specific process for generating a nine-square grid input is to select nine FFA images randomly from one patient's FFA report and then combine them along their edges to form a nine-square grid in chronological order of imaging. Additionally, a histogram equalization technique was applied in this study to enhance the contrast of images prior to training.

Performance evaluation

We tested three mainstream CNNs (VGG16, ResNet50, and DenseNet) to achieve the best possible grading performance. A confusion matrix was generated using five-fold cross-validation to ensure the stability of each model. Following classification, we also developed a visualization model to interpret results.

Comparison to human graders

To demonstrate the capabilities of our models more comprehensively, we compared the performance of the proposed algorithm to that of human graders on the testing set. The performances of the CNNs were assessed according separated accuracies for each level of image, overall accuracy, F1-score, and AUCs. Trained ophthalmologists were then asked to label the images manually, and their performances were evaluated in terms of separated and overall accuracy.

Performance of models on external datasets

There were two external datasets in this study, namely Xian dataset and Ningbo dataset. The Xian dataset collected FFA reports from the Second Affiliated Hospital of Xi'an Jiaotong University (Xibei Hospital), including 308 eyes of 160 DR patients. The image format is JPG, and the resolution is 412*412. The Ningbo dataset consisted of FFA reports from 27 eyes of 14 DR patients, who received examination from the Eye Center at Ningbo First Hospital. The image format is PNG, with a resolution of 518*518. Each report was processed and labeled and then tested by the CNNs. Accuracies for each level of image, overall accuracy, F1-score and AUCs were generated to evaluate the performance.

Results

Internal dataset

Table 2 presents the patient characteristics and distribution of the internal dataset. The numbers of eyes with DR severity levels of 1, 2, and 3 are 180, 395, and 539, respectively. For all three levels, patients were divided into groups with similar age, sex, and right eye proportion distributions. The training set accounted for 60% of the dataset, while the validation set and testing set each accounted for 20%. Each set was well balanced in terms of severity levels.

In this work, the accuracies and F1-scores of the CNN models were evaluated, as shown in Table 3. VGG16 performs the best, with a maximum accuracy of 94.17% and average accuracy of 91.11%, while RestNet50 yields values of 91.03% and 90.22%, respectively, and DenseNet yields values of 91.93% and 90.87%, respectively. The F1-scores for VGG16 are 85.30%, 91.92%, and 98.62% for the three severity levels, while for RestNet50 are 71.64%, 89.03%, and 98.61%, and for DenseNet are 87.18%, 88.00%, and 96.33%.

To further evaluate the performances of the CNNs, receiver operating characteristic (ROC) curves were drawn.

Table 2 Population characteristics and statistical information from the internal dataset

Type	Age (year)	Male sex (%)	OD/OS	Number of eyes	Training	Validation	Test
Level 1	59 ± 10	51.11	1.02	180	108	36	36
Level 2	57 ± 10	57.72	0.98	395	237	79	79
Level 3	52 ± 10	60.67	1.05	539	323	108	108
Total	55 ± 10	58.08	1.02	1114	668	223	223

Table 3 Comparison of CNNs and three ophthalmologists

	Best Acc	Overall Acc	Level 1		Level 2		Level 3	
			Acc	F1-score	Acc	F1-score	Acc	F1-score
VGG16	94.17%	91.11%	80.55%	85.30%	93.67%	91.92%	99.07%	98.62%
ResNet50	91.03%	90.22%	66.67%	71.64%	92.40%	89.03%	98.15%	98.61%
DenseNet	91.93%	90.87%	94.44%	87.18%	83.54%	88.00%	97.22%	96.33%
Doctor 1	92.96%		95.74%	89.10%	89.76%	92.70%	94.38%	93.21%
Doctor 2	88.05%		82.93%	80.01%	87.02%	92.82%	90.50%	93.47%
Doctor 3	91.87%		91.84%	87.38%	93.36%	92.49%	90.79%	93.56%

Acc accuracy

Figure 3 presents the ROC curves and best confusion matrix for each CNN. The areas under the ROC curves (AUCs) of VGG16 are 0.972, 0.922, and 0.994 for levels 1, 2, and 3, respectively. For ResNet50 and DenseNet, the values are 0.948, 0.946, and 0.990 and 0.989, 0.898, and 0.994, respectively. Confusion matrices for the multiclass task were generated for all three networks, as shown in Fig. 3. VGG16 and DenseNet exhibit better stability, while ResNet50 yields poor overall results. The original images and images dealt with visualization models are presented in Fig. 4. In the heat maps, the red areas represent areas of interest to machine, and the blue areas represent areas deemed unimportant by the machine, and the white arrows indicate areas that the doctors considered important.

The results of the human–machine comparisons are presented in Table 3 and Fig. 5. There were 2248 FFA images from 223 eyes in the test set. The overall accuracies of three ophthalmologists were 92.96%, 88.05%, and 91.87%. In Fig. 5, we present a few examples from the testing set and their labels assigned by the networks and

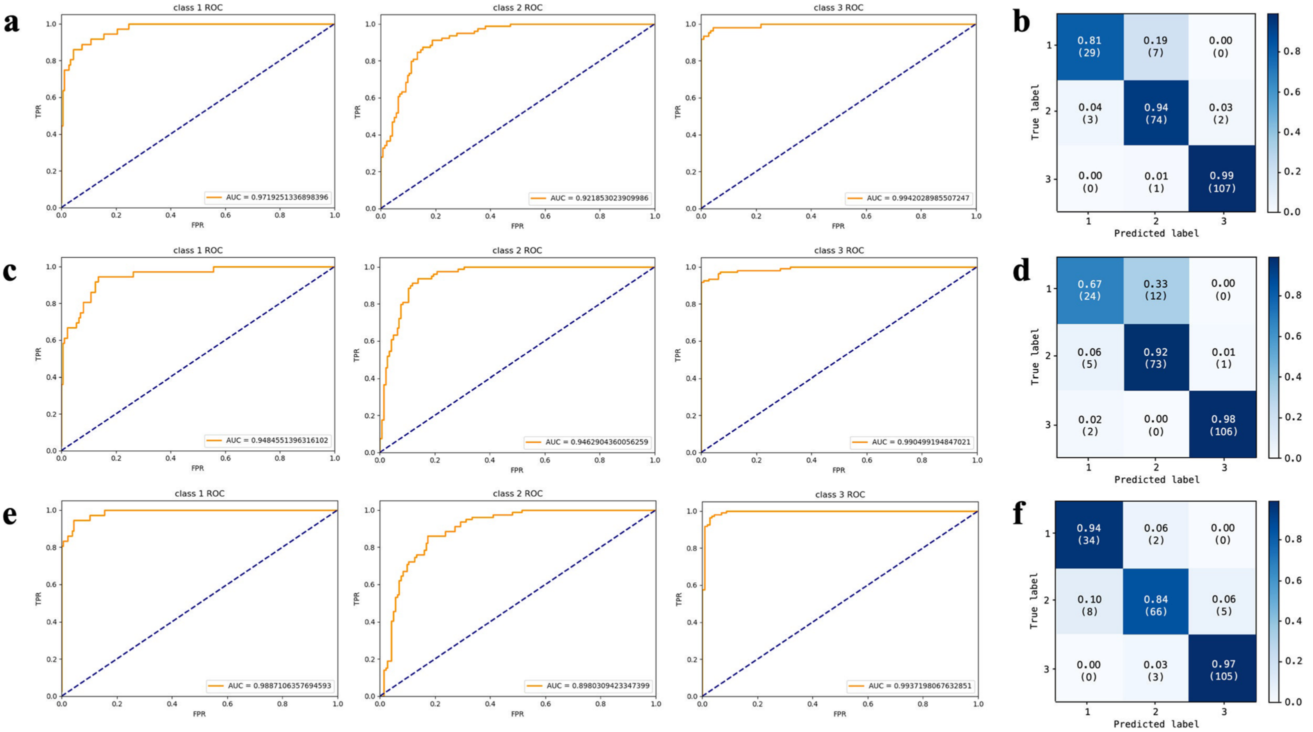

Fig. 3 ROC curve and confusion matrix for the multiclass task for three CNNs. **a** and **b** VGG16. **c** and **d** ResNet50. **e** and **f** DenseNet

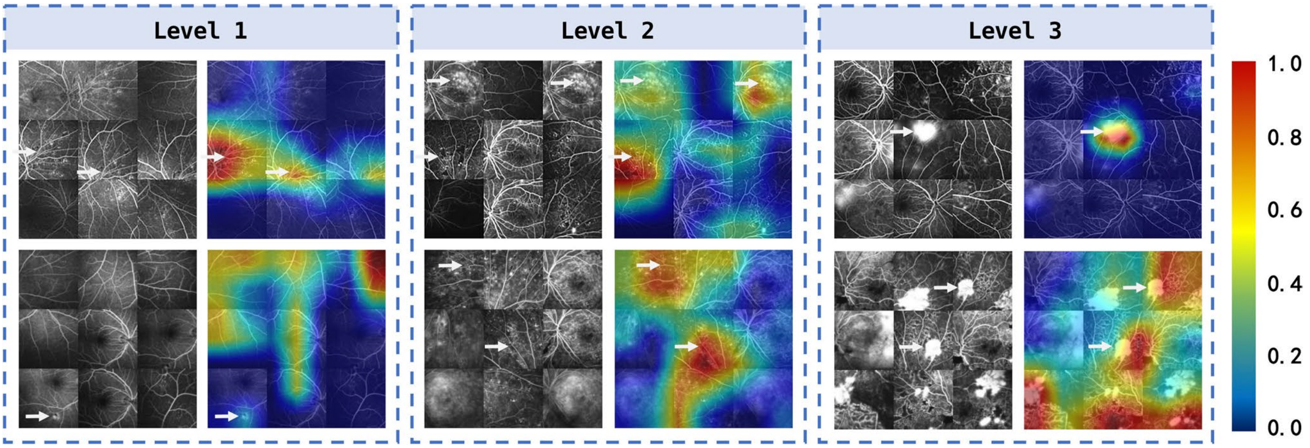

Fig. 4 Heat map of the deep learning model. The white arrows point to crucial areas

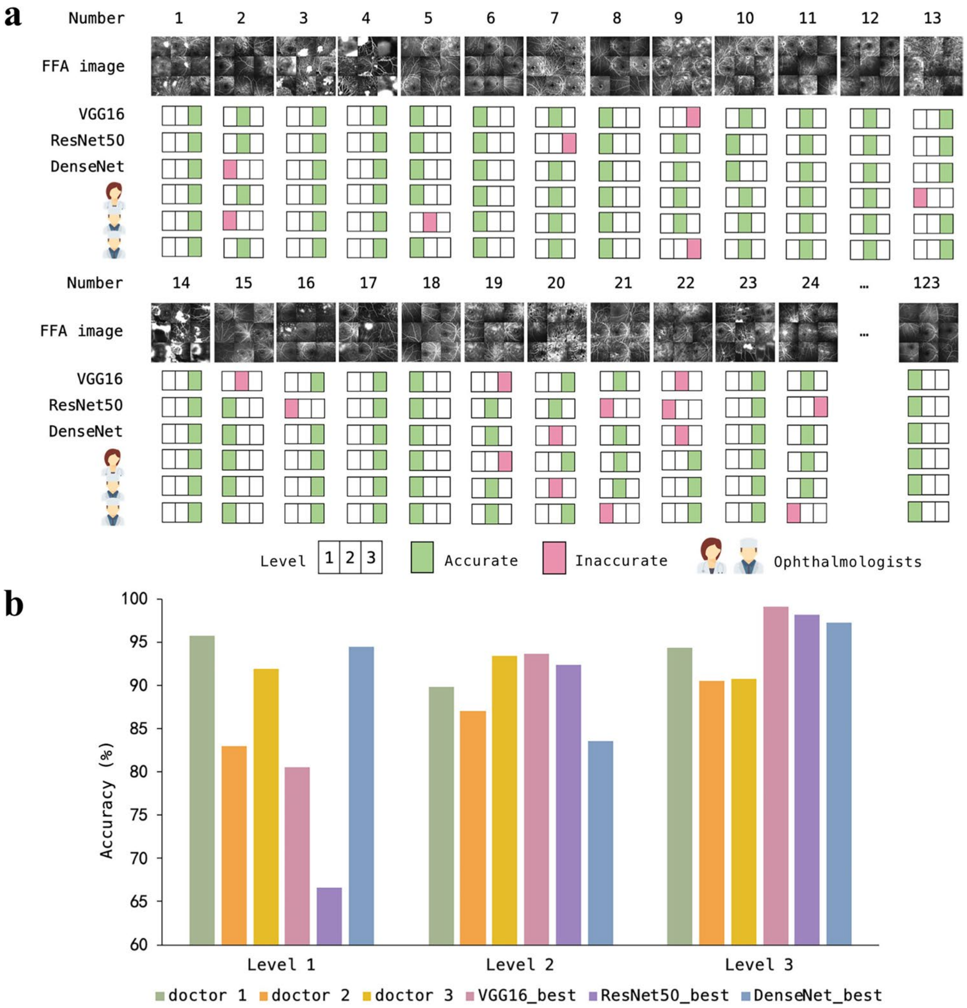

Fig. 5 Comparison between human graders and algorithms. **a** Diagnosis of test set by three CNNs and three ophthalmologists. **b** Performance of three CNNs and three ophthalmologists for each level of image

ophthalmologists. If a label is the same as the ground-truth label, it is shown in green. Otherwise, it is shown in pink. Also, accuracies between human and machine were shown directly using histogram.

External datasets

The patient characteristics and distribution of two external datasets are available in Table 4. The numbers of eyes with DR severity levels of 1, 2, and 3 are 147, 108, and 53 for Xian dataset and 11, 11, and 5 for Ningbo dataset. We applied the best performing network (VGG16) to them. For Xian dataset, our model reached the beat accuracy of 82.47% and the AUC of 0.910, 0.888, and 0.976 for levels 1, 2, and 3. As for Ningbo dataset, the network performed with the best accuracy of 88.89% and the AUC of 0.972, 0.756, and 0.945 for levels 1, 2, and 3. Figure 6 presents the ROC curves and best confusion matrix for the multi-center datasets.

Table 4 Population characteristics and statistical information from the external dataset

Dataset	Type	Age (year)	Male sex (%)	OD/OS	Number of eyes
Xian dataset	Level 1	58 ± 9	66.67	1.01	147
	Level 2	57 ± 10	70.37	0.89	108
	Level 3	53 ± 11	69.81	0.96	53
	Total	56 ± 10	68.51	0.96	308
Ningbo dataset	Level 1	49 ± 15	63.64	1.20	11
	Level 2	62 ± 10	72.73	0.83	11
	Level 3	69 ± 4	60.00	1.50	5
	Total	58 ± 14	66.67	1.08	27

Discussion

In this retrospective multi-center study, we tested and compared three CNN models on an internal dataset to determine the optimal CNN architecture, conducted comparison between human and machine to do evaluation, and applied the best model to external datasets. To the best of our knowledge, this is the first study to develop a deep learning method for DR grading based on a dataset consisting of FFA images. This study fills the gap of FFA images for DR artificial intelligence diagnosis and has the potential to realize better DR classification results than CFP analysis.

Over the past few years, deep learning has become a popular tool for assisting in DR staging. Although few studies have applied artificial intelligence to FFA images, a number of algorithms have been developed based on CFP [25–30]. Gulshan et al. were the first to present a CNN network that was only given an image and associated grade for detecting referable DR (RDR), resulting in an AUC of 0.991 for EyePACS-1 and 0.990 for Messidor-2 [19]. They later studied the performance of their network on a cohort of patients with diabetes in India and found that it could match or exceed the performance of manual grading [31]. Li et al. developed a similar deep learning algorithm and tested its functionality on a large multiethnic dataset, achieving an AUC of 0.955 [32]. Bellemo et al. adopted an ensemble AI model consisting of two CNNs to classify RDR and vision-threatening DR in a poor African population. Their model achieved an AUC of 0.973 [33]. Takahashi et al. proposed a neural network trained on four photographs for grading DR involving a retinal area that is not typically visualized in fundoscopy [34]. The prevalence and bias-adjusted Fleiss' kappa (PABAK)-to-modified-Davis-grading ratio was 0.64 and the PABAK-to-real prognosis-grading ratio was 0.37. Raumviboonsuk et al. presented a method for staging DR on a more granular five-point grade. Their study involved one of the largest clinical validations of a deep learning algorithm [35]. Sayres et al. further improved this concept and investigated visualization models using predicted DR scores and explanatory heat maps, starting a trend toward higher accuracy, confidence, and grading times [20]. All of the studies discussed above have made significant contributions to the computer-aided diagnosis of DR and have promoted the development of deep learning algorithms for medical images. However, CFP images are qualitative, rather than quantitative [11], making them less accurate for staging DR severity. As indicated by the discussion above, most algorithms in previous studies on DR classification have simplified various levels of DR into binary predictions. The not ideal sensitivity in early DR detection and the heterogeneity of data restricted their clinical application.

Compared to CFP, FFA is more invasive and time-consuming. Nevertheless, DR mainly affects retinal blood vessels [36, 37], and FFA images can show more details of retinal blood vessels, including non-perfusion areas and vascular leakage, resulting in greater effectiveness for DR grading. In other words, FFA is the gold standard for DR grading. Additionally, CFP photos are only partial views of the retina, whereas FFA is able to capture peripheral lesions, leading to a more complete assessment of the entire retina. In real-world scenarios, patient management can be different at each level of DR [38]. Therefore, it is crucial to provide a precise diagnosis of DR stages. Up to now, there have been few studies focused on FFA images, which may be explained by the difficulty of data obtaining and processing.

As for dataset acquiring, our work gathered one large internal dataset and collected two external datasets in other

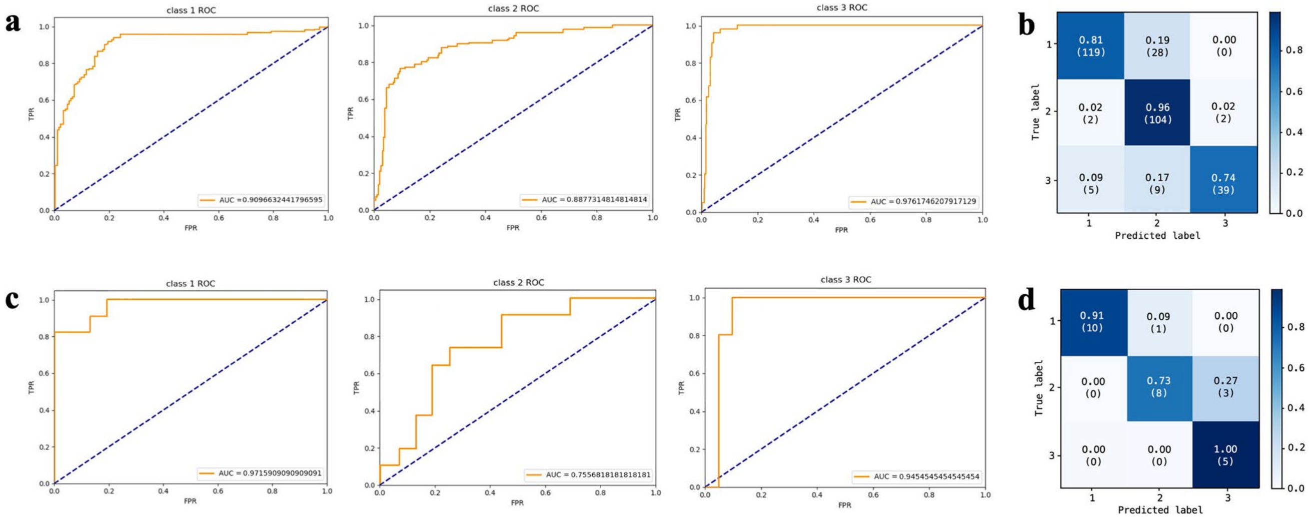

Fig. 6 ROC curve and confusion matrix for the external datasets. **a** and **b** Xian dataset. **c** and **d** Ningbo dataset

centers. The model was developed on the internal one, consisting of 11,214 FFA images from 1114 eyes. Among three mainstream CNNs, VGG16 achieved the best results in terms of accuracy, F1-score, and AUC, which can be indicated from Table 3 and Fig. 3. The level 1 images yield relatively poor results, which can be attributed to an insufficient number of level 1 training samples. Additionally, our standard combines both mild and moderate DR into level 1, which may be ambiguous when lesions lie between the moderate and severe level. This was also an issue during manual labeling. Level 3 images yield the best performance, which is intuitive based on the presence of prominent features such as leakage and vitreous hemorrhage. The heat map in Fig. 4 presents the areas that the proposed algorithm found to be the most important, which are consistent with the white arrows identified by the ophthalmologists. This indicates that the model can aid in making clinical decisions in a logical manner. The human–machine comparison in Fig. 5 showed that CNNs could reach a comparable level to the ophthalmologists. Afterwards, the external datasets were used for testing. The Xian dataset consisted of 308 eyes, while the Ningbo dataset consisted of only 27 eyes. Also, the number and form of FFA reports from each dataset are slightly different due to the devices. However, the results showed that regardless of the size of dataset or the form of FFA reports, our model achieved an acceptable performance, indicating the possibility of wide application.

Considering the difficulties of dealing with the input of FFA images, the nine-square grid inputs used in this study are innovative. Each sample in our dataset contains nine or more images. Therefore, to ensure consistent input sizes, we randomly selected nine images and arranged them into a nine-square grid according to their order in the corresponding inspection report. Additionally, we also considered other input methods such as overlay inputs (for each sample, nine pictures were selected randomly and depth-wise concatenation was performed in ascending order according to the order in the report) and average inputs (for each sample, nine pictures were selected randomly, added pixel by pixel, and average values were calculated to obtain the final input). Experimental comparisons revealed that the model using nine-square grid inputs yielded the best results.

There are several limitations in this study. First, there are no normal FFA images in our dataset because FFA is an invasive fundus examination and a normal image is relatively difficult to obtain. Second, it may be argued that FFA is invasive, and those with mild or moderate DR might have no need to do this test. However, without FFA images, we may find it difficult to say whether these patients are diagnosed with the right DR stage. Additionally, the proposed algorithm may not perform well on images with subtle lesions, which can be attributed to the input method we adopted and the features of deep learning algorithms. Moreover, we only gathered images from patients with DR without any other fundus diseases. This means that our model may not distinguish between individuals with combined diseases, which could lead to incorrect classification. Lastly, images used in this study were obtained from Chinese patients, and the model may not be suitable for other races.

In summary, we proposed a deep learning algorithm for DR staging based on FFA images, compared it to human graders, and analyzed its effects on two external clinical datasets. Our work may significantly improve the current clinical procedure by providing automated real-time evaluations to expedite diagnosis. With appropriate technical help, we can embed the algorithm into the FFA machine, or build a website to automatically read the FFA reports online. Future studies should focus on enhancing classification results, the ability to discover tiny lesions, use of time information in FFA pictures, and incorporating additional modalities such as CFP images.

Acknowledgements We thank the Second Affiliated Hospital of Xi'an Jiaotong University and the Eye Center at Ningbo First Hospital for their contributions to data collection.

Author contribution Z.G., K.J., J.W., and J.Y. conceived and designed the experiments. Y.Y., X.L, Y.S, X.P, and Y.W. collected and processed the data. Y.G. and Y.L. analyzed the results. All authors reviewed the manuscript.

Funding This work was financially supported by the National Key Research and Development Program of China (grant number 2019YFC0118401), Zhejiang Provincial Key Research and Development Plan (grant number 2019C03020), and the Natural Science Foundation of China (grant number 81670888).

Availability of data and material and code availability The datasets presented in this study are available from the corresponding author upon request.

Declarations

Ethics approval Ethical approval for the study was obtained from Ethics Committee of the Second Affiliated Hospital, Zhejiang University School of Medicine (No.Y2020-1027). Informed consent was obtained from the research subjects prior to the study. The research complied with the tenets of the Declaration of Helsinki and the Health Portability and Accessibility Act.

Competing interests The authors declare no competing interests.

References

1. Cheung N, Mitchell P, Wong TY (2010) Diabetic retinopathy. The Lancet 376(9735):124–136. https://doi.org/10.1016/s0140-6736(09)62124-3

2. Yau JWY, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, Chen SJ, Dekker JM, Fletcher A, Grauslund J, Haffner S, Hamman RF, Ikram MK, Kayama T, Klein BEK, Klein R, Krishnaiah S, Mayurasakorn K, O'Hare JP, Orchard TJ, Porta M, Rema M, Roy MS, Sharma T, Shaw J, Taylor H, Tielsch JM, Varma R, Wang JJ, Wang NL, West S, Xu L, Yasuda M, Zhang XZ, Mitchell P, Wong TY, Meta-EYE Study (2012) Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 35(3):556–564. https://doi.org/10.2337/dc11-1909
3. Lu JY, Ma XJ, Zhou J, Zhang L, Mo YF, Ying LW, Lu W, Zhu W, Bao YQ, Vigersky RA, Jia WP (2018) Association of time in range, as assessed by continuous glucose monitoring, with diabetic retinopathy in type 2 diabetes. Diabetes Care 41(11):2370–2376. https://doi.org/10.2337/dc18-1131
4. Lee R, Wong TY, Sabanayagam C (2015) Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. Eye Vis 2:17. https://doi.org/10.1186/s40662-015-0026-2
5. Stitt AW, Curtis TM, Chen M, Medina RJ, McKay GJ, Jenkins A, Gardiner TA, Lyons TJ, Hammes HP, Simo R, Lois N (2016) The progress in understanding and treatment of diabetic retinopathy. Prog Retin Eye Res 51:156–186. https://doi.org/10.1016/j.preteyeres.2015.08.001
6. Li X, Hu X, Yu L, Zhu L, Fu CW, Heng PA (2020) CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. IEEE Trans Med Imaging 39(5):1483–1493. https://doi.org/10.1109/TMI.2019.2951844
7. Cole ED, Novais EA, Louzada RN, Waheed NK (2016) Contemporary retinal imaging techniques in diabetic retinopathy: a review. Clin Exp Ophthalmol 44(4):289–299. https://doi.org/10.1111/ceo.12711
8. Wong TY, Cheung CM, Larsen M, Sharma S, Simo R (2016) Diabetic retinopathy. Nat Rev Dis Primers 2:16012. https://doi.org/10.1038/nrdp.2016.12
9. Porwal P, Pachade S, Kokare M, Deshmukh G, Son J, Bae W, Liu L, Wang J, Liu X, Gao L, Wu T, Xiao J, Wang F, Yin B, Wang Y, Danala G, He L, Choi YH, Lee YC, Jung SH, Li Z, Sui X, Wu J, Li X, Zhou T, Toth J, Baran A, Kori A, Chennamsetty SS, Safwan M, Alex V, Lyu X, Cheng L, Chu Q, Li P, Ji X, Zhang S, Shen Y, Dai L, Saha O, Sathish R, Melo T, Araujo T, Harangi B, Sheng B, Fang R, Sheet D, Hajdu A, Zheng Y, Mendonca AM, Zhang S, Campilho A, Zheng B, Shen D, Giancardo L, Quellec G, Meriaudeau F (2020) IDRiD: diabetic retinopathy - segmentation and grading challenge. Med Image Anal 59:101561. https://doi.org/10.1016/j.media.2019.101561
10. Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR (2004) Single-field fundus photography for diabetic retinopathy screening. Ophthalmology 111(5):1055–1062. https://doi.org/10.1016/j.ophtha.2004.02.004
11. Kwan CC, Fawzi AA (2019) Imaging and biomarkers in diabetic macular edema and diabetic retinopathy. Curr Diabetes Rep 19(10):95. https://doi.org/10.1007/s11892-019-1226-2
12. Pan X, Jin K, Cao J, Liu Z, Wu J, You K, Lu Y, Xu Y, Su Z, Jiang J, Yao K, Ye J (2020) Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning. Graefes Arch Clin Exp Ophthalmol 258(4):779–785. https://doi.org/10.1007/s00417-019-04575-w
13. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539
14. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer Learning. IEEE T Med Imaging 35(5):1285–1298. https://doi.org/10.1109/TMI.2016.2528162
15. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina PM, Chiang MF, Schmetterer L, Pasquale LR, Bressler NM, Webster DR, Abramoff M, Wong TY (2019) Deep learning in ophthalmology: the technical and clinical considerations. Prog Retin Eye Res 72:100759. https://doi.org/10.1016/j.preteyeres.2019.04.003
16. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, van den Driessche G, Lakshminarayanan B, Meyer C, Mackinder F, Bouton S, Ayoub K, Chopra R, King D, Karthikesalingam A, Hughes CO, Raine R, Hughes J, Sim DA, Egan C, Tufail A, Montgomery H, Hassabis D, Rees G, Back T, Khaw PT, Suleyman M, Cornebise J, Keane PA, Ronneberger O (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 24(9):1342–1350. https://doi.org/10.1038/s41591-018-0107-6
17. Verbraak FD, Abramoff MD, Bausch GCF, Klaver C, Nijpels G, Schlingemann RO, van der Heijden AA (2019) Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. Diabetes Care 42(4):651–656. https://doi.org/10.2337/dc18-0148
18. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunovic H (2018) Artificial intelligence in retina. Prog Retin Eye Res 67:1–29. https://doi.org/10.1016/j.preteyeres.2018.07.004
19. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410. https://doi.org/10.1001/jama.2016.17216
20. Sayres R, Taly A, Rahimy E, Katy B, David C, Naama H, Jonathan K, Arunachalam N, Rastegar Zahra Wu, Shawn DX, Scott B, Anthony J, Michael S, Jesse S, Sood Arjun B, Corrado Greg S, Lily P, Webster Dale R (2019) Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. Ophthalmology 126(4):552–564. https://doi.org/10.1016/j.ophtha.2018.11.016
21. Rasta SH, Nikfarjam S, Javadzadeh A (2015) Detection of retinal capillary nonperfusion in fundus fluorescein angiogram of diabetic retinopathy. Bioimpacts 5(4):183–190. https://doi.org/10.15171/bi.2015.27
22. Hemalakshmi GR, Santhi D, Mani VRS, Hemalakshmi GR, Santhi D, Mani V, Geetha A, Prakash NB (2020) Deep residual network based on image priors for single image super resolution in FFA images. CMES 125(1):125–143. https://doi.org/10.32604/cmes.2020.011331
23. Jin K, Pan X, You K, Wu J, Liu Z, Cao J, Lou L, Xu Y, Su Z, Yao K, Ye J (2020) Automatic detection of non-perfusion areas in diabetic macular edema from fundus fluorescein angiography for decision making using deep learning. Sci Rep 10(1):15138. https://doi.org/10.1038/s41598-020-71622-6
24. Wilkinson CP, Ferris FL, Klein RE, Wilkinson CP, Ferris FL, Klein RE, Lee PP, Agardh CD, Davis M, Dills D, Kampik A, Pararajasegaram R, Verdaguer JT (2003) Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 110(9):1677–1682. https://doi.org/10.1016/s0161-6420(03)00475-5
25. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, Peng L, Webster DR (2018) Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology 125(8):1264–1272. https://doi.org/10.1016/j.ophtha.2018.01.034
26. Son J, Shin JY, Kim HD, Jung KH, Park KH, Park SJ (2020) Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. Ophthalmology 127(1):85–94. https://doi.org/10.1016/j.ophtha.2019.05.029
27. Liu TYA (2019) Smartphone-based, artificial intelligence-enabled diabetic retinopathy screening. JAMA Ophthalmol 137(10):1188–1189. https://doi.org/10.1001/jamaophthalmol.2019.2883

28. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M (2016) Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 57(13):5200–5206. https://doi.org/10.1167/iovs.16-19964
29. de la Torre J, Valls A, Puig D (2020) A deep learning interpretable classifier for diabetic retinopathy disease grading. Neurocomputing 396:465–476. https://doi.org/10.1016/j.neucom.2018.07.102
30. Asiri N, Hussain M, Al Adel F, Alzaidi N (2019) Deep learning based computer-aided diagnosis systems for diabetic retinopathy: a survey. Artif Intell Med 99:20. https://doi.org/10.1016/j.artmed.2019.07.009
31. Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, Whitehouse K, Coram M, Corrado G, Ramasamy K, Raman R, Peng L, Webster DR (2019) Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmol 137(9):987–993. https://doi.org/10.1001/jamaophthalmol.2019.2004
32. Li ZX, Keel S, Liu C, He YF, Meng W, Scheetz J, Lee PY, Shaw J, Ting D, Wong TY, Taylor H, Chang R, He MG (2018) An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. Diabetes Care 41(12):2509–2516. https://doi.org/10.2337/dc18-0147
33. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie YC, Yip MYT, Hamzah H, Ho JY, Lee XQ, Hsu WN, Lee ML, Musonda L, Chandran M, Chipalo-Mutati G, Muma M, Tan GSW, Sivaprasad S, Menon G, Wong TY, Ting DSW (2019) Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. Lancet Digit Health 1(1):E35–E44. https://doi.org/10.1016/S2589-7500(19)30004-4
34. Mori K, Takahashi H, Tampo H, Inoue Y, Kawashima H (2017) Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic retinopathy. PLoS ONE 12(6):11. https://doi.org/10.1371/journal.pone.0179790
35. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, Campana BJL, Phene S, Hemarat K, Tadarati M, Silpa-Archa S, Limwattanayingyong J, Rao C, Kuruvilla O, Jung J, Tan J, Orprayoon S, Kangwanwongpaisan C, Sukumalpaiboon R, Luengchaichawang C, Fuangkaew J, Kongsap P, Chualinpha L, Saree S, Kawinpanitan S, Mitvongsa K, Lawanasakol S, Thepchatri C, Wongpichedchai L, Corrado GS, Peng L, Webster DR (2019) Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med 2:25. https://doi.org/10.1038/s41746-019-0099-8
36. Duh EJ, Sun JK, Stitt AW (2017) Diabetic retinopathy: current understanding, mechanisms, and treatment strategies. JCI Insight 2(14):13. https://doi.org/10.1172/jci.insight.93751
37. Antonetti DA, Klein R, Gardner TW (2012) Mechanisms of disease: diabetic retinopathy. New Engl J Med 366(13):1227–1239. https://doi.org/10.1056/NEJMra1005073
38. Ting DSW, Carin L, Abramoff MD (2019) Observations and lessons learned from the artificial intelligence studies for diabetic retinopathy screening. JAMA Ophthalmol 137(9):994–995. https://doi.org/10.1001/jamaophthalmol.2019.1997