Journal of the Formosan Medical Association xxx (xxxx) xxx

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

Original Article

Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening

Yi-Ting Hsieh [a,*], Lee-Ming Chuang [b,c], Yi-Der Jiang [b], Tien-Jyun Chang [b], Chung-May Yang [a,c], Chang-Hao Yang [a,c], Li-Wei Chan [a,d], Tzu-Yun Kao [a,e], Ta-Ching Chen [a], Hsuan-Chieh Lin [f], Chin-Han Tsai [g], Mingke Chen [g]

[a] *Department of Ophthalmology, National Taiwan University Hospital, Taipei, Taiwan*
[b] *Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan*
[c] *College of Medicine, National Taiwan University, Taipei, Taiwan*
[d] *Department of Ophthalmology, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei, Taiwan*
[e] *Department of Ophthalmology, Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan*
[f] *Department of Ophthalmology, National Taiwan University Hospital, Hsinchu Branch, Hsinchu, Taiwan*
[g] *Acer Inc., New Taipei, Taiwan*

Received 1 September 2019; received in revised form 9 December 2019; accepted 30 March 2020

KEYWORDS
Artificial intelligence;
Deep learning;
Diabetic retinopathy;
Convolutional neural network;
Retinal fundus photography

Purpose: To develop a deep learning image assessment software VeriSee™ and to validate its accuracy in grading the severity of diabetic retinopathy (DR).
Methods: Diabetic patients who underwent single-field, nonmydriatic, 45-degree color retinal fundus photography at National Taiwan University Hospital between July 2007 and June 2017 were retrospectively recruited. A total of 7524 judgeable color fundus images were collected and were graded for the severity of DR by ophthalmologists. Among these pictures, 5649 along with another 31,612 color fundus images from the EyePACS dataset were used for model training of VeriSee™. The other 1875 images were used for validation and were graded for the severity of DR by VeriSee™, ophthalmologists, and internal physicians. Area under the receiver operating characteristic curve (AUC) for VeriSee™, and the sensitivities and specificities for VeriSee™, ophthalmologists, and internal physicians in diagnosing DR were calculated.
Results: The AUCs for VeriSee™ in diagnosing any DR, referable DR and proliferative diabetic retinopathy (PDR) were 0.955, 0.955 and 0.984, respectively. VeriSee™ had better sensitivities

* Corresponding author. Department of Ophthalmology, National Taiwan University Hospital, 7 Zhongshan S. Rd., Zhongzheng Dist., Taipei 10002, Taiwan. Fax: +886 2 23934420.
E-mail address: ythyth@gmail.com (Y.-T. Hsieh).

https://doi.org/10.1016/j.jfma.2020.03.024

Please cite this article as: Hsieh Y-T et al., Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening, Journal of the Formosan Medical Association, https://doi.org/10.1016/j.jfma.2020.03.024

in diagnosing any DR and PDR (92.2% and 90.9%, respectively) than internal physicians (64.3% and 20.6%, respectively) ($P < 0.001$ for both). VeriSee™ also had better sensitivities in diagnosing any DR and referable DR (92.2% and 89.2%, respectively) than ophthalmologists (86.9% and 71.1%, respectively) ($P < 0.001$ for both), while ophthalmologists had better specificities.

Conclusion: VeriSee™ had good sensitivity and specificity in grading the severity of DR from color fundus images. It may offer clinical assistance to non-ophthalmologists in DR screening with nonmydriatic retinal fundus photography.

Introduction

Diabetic retinopathy (DR) is one of the major microvascular complications in diabetes mellitus. Almost all patients who have type 1 diabetes for 20 years or more develop DR; among them, 20—30% suffer visual loss.[1] During the lifetime, over 60% of patients with type 2 diabetes develop DR, which is also the main cause for legal blindness among the population within 20—74 years of age.[2] The American Diabetes Association[3] suggests that patients of type 1 diabetes should receive fundus examinations for DR within 5 years after diagnosis, and patients of type 2 diabetes should receive DR screening at the time of diagnosis. However, not all diabetic patients receive DR screening regularly. In Taiwan, only 45.5% of patients with diabetes received fundus examinations for DR screening in 2019.[4] This results in delayed diagnosis of DR in many patients, to whom adequate referral or treatment cannot be given in time.

One of the reasons for the poor DR screening rate is that many general practitioners or internal physicians are not confident in diagnosing DR with retinal fundus photography by themselves.[5,6] In recent years, artificial intelligence (AI) with deep learning has been developed for the diagnosis of DR, and previous studies have demonstrated its applicability,[7—15] while most of them used open-access datasets only for the development and validation of their algorithms.[7—10,13—15] Compared to the open-access datasets, the fundus pictures obtained from the nonmydriatic fundus camera in daily practice may contain more pictures with poor quality or artifacts. It has been shown that the validation results were less precise when using the real-world datasets compared to those using the open-access datasets with the same algorithm.[7,16] VeriSee™ (Acer Inc., Taiwan) is a computer-aided diagnosis system developed for the diagnosis of DR. It is developed using convolutional neural network (CNN), the state of the art on image recognition and classification. In this study, we used an open-access dataset as well as a local dataset collected in Taiwan to train the VeriSee™ for diagnosing DR, and to validate its efficacy by comparing the precision rate of diagnosis with those made by internal physicians and ophthalmologists.

Methods and materials

Datasets

Diabetic patients who underwent single-field, 45-degree color fundus photography with a nonmydriatic fundus camera (Canon CR-2, Tokyo, Japan) in the Diabetes Education Center of National Taiwan University Hospital (NTUH) between July 2007 and June 2017 were retrospectively recruited. All patients in this study were of Asian population. The images were deidentified after retrieving the color fundus photos of these patients from the picture archiving and communication system (PACS) of NTUH. Only pictures containing both the optic disc and the central fovea were included. Images that were judged as ungradable by ophthalmologists were excluded. Finally, a total of 7524 gradable fundus pictures were collected. Furthermore, 35,126 color fundus images from the open-access EYEPACS dataset[17] were also used for model training and validation.

Grading for diabetic retinopathy

All fundus pictures from NTUH were graded by two primary graders who were board-certified ophthalmologists in Taiwan with at least one year of fellowship training in the field of retina. The severity of DR was graded as one of the following: no DR, mild nonproliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, or proliferative diabetic retinopathy (PDR) according to the International Clinical Diabetic Retinopathy Disease Severity Scale proposed by the Global Diabetic Retinopathy Project Group.[18] Referable DR was defined as moderate NPDR or worse. If the results were not concordant between the two primary graders, a secondary grader who was a board-certified ophthalmologist with at least 10 years of experience as a retina specialist would make the final diagnosis. The diagnostic results served as the golden standard for this study. As for the fundus pictures retrieved form the EYEPACS, the diagnostic results offered by the EYEPACS were treated as the golden standard.

Image pre-processing

Two different procedures were performed during preprocessing before the images were used for model

Please cite this article as: Hsieh Y-T et al., Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening, Journal of the Formosan Medical Association, https://doi.org/10.1016/j.jfma.2020.03.024

training. Eye tracking was used to find the edge of effective fundus photography and crop it as a uniform size. Gaussian blur was used to normalize its color.[19]

Architecture and training dataset

CNN was used for learning in VeriSee™. The training process can be divided into two steps: (1) pre-training, and (2) fine-tune training. In the pre-training step, 31,612 color fundus photos from the EyePACS dataset were used. We tried several different deep learning models that had been trained on ImageNet to speed up the training and to obtain the base models. Among all the modified Inception-v4 architectures[20] had the best performance for diagnosing any DR and referable DR, while a modified ResNet architecture,[21] which used ResNet with the modified Feature Pyramid Network (FPN) architecture,[22] had the best performance for diagnosing PDR. Therefore, the modified Inception-v4 model was used to train the model for any DR and referable DR, and the modified ResNet model was used to train the model for PDR. In the fine-tune step, 5649 color fundus images from the NTUH dataset were used to train the base models on the same architecture by fine-tune learning. In order to improve the prediction for referable DR, an ensemble of two different Inception-v4 models trained on the same data was used.

Validation

A total of 1875 pictures from the NTUH dataset were used for validation. The diagnoses from the VeriSee™ pre-trained model, the VeriSee™ final model, ophthalmologists, and internal physicians were compared to the gold standards. The diagnoses from the internal physicians were retrieved from chart records, of which the severity of DR was only graded as three classes: no DR, NPDR, and PDR. Therefore, there were no diagnostic results in referable DR for internal physicians.

Statistical analysis

For the VeriSee™ pre-trained model and final model, the receiver operating characteristic (ROC) curves for any DR, PDR, and referable DR were drawn across a range of classification thresholds. The area under the curve (AUC) for each ROC curve was calculated, and a point with the optimal sensitivity & specificity on each ROC curve was chosen. The sensitivity and specificity for the diagnoses of any DR and PDR by internal physicians and ophthalmologists and the diagnosis of referable DR by ophthalmologists were also calculated, and their sensitivities and specificities for the diagnosis of DR were compared to those from the VeriSee™ final model with McNemar tests.

This study followed the tenets of the Declaration of Helsinki. It was approved by the Institutional Review Board of the National Taiwan University Hospital. Because of the retrospective nature, the requirement of informed consent was waived.

Results

The image characteristics and numbers are listed in Table 1. Of the 5649 images from NTUH for training, 2695 had any DR, 931 had referable DR, and 187 had PDR. Of the 1875 images for validation, 654 had any DR, 223 had referable DR, and 33 had PDR. In the NTUH dataset, the intraclass correlation coefficients (ICC) of two ophthalmologists for any DR, referable DR, and PDR were 0.856 (95% CI, 0.842—0.868), 0.736 (95% CI, 0.711—0.759), and 0.919 (95% CI, 0.912—0.926), respectively.

VeriSee™: pre-trained model vs. final model

After pre-trained with the EYEPACS dataset, the pre-trained models had AUCs of 0.812, 0.944 and 0.895, respectively for the diagnosis of any DR, PDR and referable DR when validated with the NTUH dataset. After the fine-tuned training with the NTUH dataset, the AUCs of the final models improved to 0.955, 0.984 and 0.950, respectively for the diagnosis of any DR, PDR and referable DR (Table 2).

Comparisons among VeriSee™, ophthalmologists and internal physicians

Any diabetic retinopathy

For detecting any DR, the algorithm of VeriSee™ achieved an AUC of 0.955 (95% CI, 0.947—0.964) (Fig. 1). At the optimal operating point, the algorithm had the best sensitivity among all (92.2%, 95% CI: 89.4%—94.7%) and the specificity was 89.5% (95% CI: 87.4%—91.6%). The ophthalmologists had the best specificity among all (97.5%, 95% CI: 96.9%—98.1%), and the sensitivity was 86.9% (95% CI: 85.2%—88.8%) in detecting any DR. As for the internal physicians, the sensitivity (64.3%, 95% CI: 60.7%—68.0%) and the specificity (71.9%, 95% CI: 69.3%—74.4%) were both the poorest (Table 3).

Proliferative diabetic retinopathy

For detecting PDR, the algorithm of VeriSee™ achieved an AUC of 0.984 (95% CI: 0.967—0.998) (Fig. 2). At the optimal operating point, the algorithm had a sensitivity of —90.9% (95% CI: 81.1%—100.0%) and a specificity of 99.3% (95% CI:

Table 1 Image characteristics and the severity of diabetic retinopathy of the dataset.

	EYEPACS for training	NTUH for training	NTUH for validation
No of images	31,612	5649	1875
Image quality			
Clear and gradable	NA	2732	585
Not clear but gradable	NA	2917	1290
Severity of DR			
No DR	23,181	2954	1221
Mild NPDR	2188	1764	431
Moderate NPDR	4799	646	167
Severe NPDR	800	98	23
PDR	644	187	33
Any DR	8431	2695	654
Referable DR	6243	931	223

DR: diabetic retinopathy; NPDR: nonproliferative diabetic retinopathy; PDR: proliferative diabetic retinopathy.

Please cite this article as: Hsieh Y-T et al., Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening, Journal of the Formosan Medical Association, https://doi.org/10.1016/j.jfma.2020.03.024

Table 2 Comparisons between the pre-trained model and final model of VeriSee™.

	VeriSee™ pre-trained model	VeriSee™ final model
Any DR		
Sensitivity (95% CI)	64.5% (71.6%–82.6%)	92.2% (89.8%–94.7%)
Specificity (95% CI)	90.4% (88.8%–91.6%)	89.5% (87.4%–91.6%)
AUC (95% CI)	0.812 (0.794–0.831)	0.955 (0.9145–0.964)
PDR		
Sensitivity (95% CI)	90.9% (81.1%–100.0%)	90.9% (81.1%–100.0%)
Specificity (95% CI)	91.0% (89.7%–92.3%)	99.3% (98.9%–99.7%)
AUC (95% CI)	0.944 (0.894–0.982)	0.984 (0.967–0.998)
Referable DR		
Sensitivity (95% CI)	80.3% (60.9%–68.2%)	89.2% (85.2%–93.3%)
Specificity (95% CI)	88.3% (88.8%–92.1%)	90.1% (88.7%–91.6%)
AUC (95% CI)	0.895 (0.871–0.917)	0.950 (0.937–0.961)

98.9%–99.7%). The ophthalmologists had a sensitivity of 89.1% (95% CI: 81.9%–96.2%) and a specificity of 99.8% (95% CI: 99.6%–99.9%) in detecting PDR. There were no significant differences between the VeriSee™ algorithm and ophthalmologists. As for the internal physicians, the sensitivity was much poorer (20.6%, 95% CI: 7.3%–35.2%), although the specificity was good (99.5%, 95% CI: 99.1%–99.8%) (Table 3).

Referable diabetic retinopathy

For detecting referable DR, the algorithm of VeriSee™ achieved an AUC of 0.950 (95% CI: 0.937–0.961) (Fig. 3). At the optimal operating point, the algorithm had a sensitivity of 89.2% (95% CI: 85.2%–93.3%) and a specificity of 90.1% (95% CI: 88.7%–91.6%). As for the ophthalmologists, the sensitivity was poorer (71.1%, 95% CI: 66.8%–75.1%, $P < 0.001$), but the specificity was better (98.9%, 95% CI: 98.5%–99.2%, $P < 0.001$) than the algorithm of VeriSee™ (Table 3).

Discussion

This study demonstrated that the deep learning algorithms of VeriSee™ had a high reliability in diagnosing DR with good sensitivity and specificity. The sensitivity was 89.2% and the specificity was 90.1% for diagnosing referable DR, which were comparable to previous published studies using deep-learning algorithms.[7–15] Furthermore, we also demonstrated that the precision rates of VeriSee™ in diagnosing any DR and PDR were much better than those of internal physicians, which we believe has never been shown in previous studies.

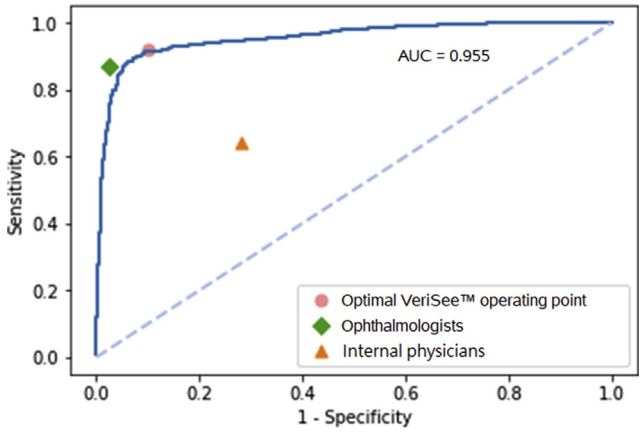

Figure 1 Receiver operating characteristic curves and the areas under the curve of the VeriSee™ pre-trained model and final model for detection of any diabetic retinopathy, in comparison with ophthalmologists' and internal physicians' performances.

Before the launching of deep learning algorithms, conventional automated software has been used for DR detection.[23–28] Previous studies showed that these automated programs were quite good at detecting DR, with sensitivities ranging from 87 to 95.2%; however, the specificities were rather low, ranging from 49.6 to 68.8%.[29] Although a tool with a high sensitivity is more important for screening, a low specificity will result in a high false positive rate in real-world practice, which is insufficient for screening work. The application of CNN has largely improved the ability of artificial intelligence in imaging identification. In the ImageNet Large Scale Visual Recognition Challenge 2015, the winner ResNet reduced the classification error to 3.6%; which was even lower than that of an average human (5.1%).[30] Therefore, the CNN technology is leading a new generation of image recognition. For VeriSee™, both modified ResNet and Inception-v4 architectures were used for different characteristics of the disease entity. The modified ensemble Inception-v4 model was used to predict DR stages, and the modified ResNet was used to detect PDR.

Several previous studies have demonstrated the applicability of deep learning algorithms in diagnosing DR. Abramoff et al.[7] first reported a sensitivity of 96.8% and a specificity of 87.0% in diagnosing referable DR using the device IDx-DR X2.1. Gulshan et al.[8] demonstrated an even better result with a sensitivity of 97.5% and a specificity of 93.4% in detecting referable DR. However, most studies used open-access datasets only for the development and validation of their algorithms.[7–10,13–15] In the present study, we used the open-access datasets for pre-training, and then used the fundus pictures taken with non-mydriatic fundus cameras at NTUH for fine-tune training; another set of pictures taken in NTUH were used for validation. The AUC for diagnosing referable DR with VeriSee™ was 0.950, with a sensitivity of 89.2% and a specificity of

Please cite this article as: Hsieh Y-T et al., Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening, Journal of the Formosan Medical Association, https://doi.org/10.1016/j.jfma.2020.03.024

Table 3 Sensitivity and specificity of diagnosis for diabetic retinopathy by VeriSee™, ophthalmologists and internal physicians.

	VeriSee™	Ophthalmologists	Internal physicians	*P* values by McNemar test	
				VeriSee™ vs. ophthalmologists	VeriSee™ vs. internal physicians
Any DR					
Sensitivity (95% CI)	92.2% (89.8%–94.7%)	86.9% (85.2%–88.8%)	64.3% (60.7%–68.0%)	<0.001	<0.001
Specificity (95% CI)	89.5% (87.4%–91.6%)	97.5% (96.9%–98.1%)	71.9% (69.3%–74.4%)	<0.001	<0.001
Accuracy (95% CI)	90.7% (89.4%–92.3%)	90.6% (84.0%–91.8%)	68.3% (67.2%–71.4%)	0.333	<0.001
PDR					
Sensitivity (95% CI)	90.9% (81.1%–100.0%)	89.1% (81.9%–96.2%)	20.6% (7.3%–35.2%)	1	<0.001
Specificity (95% CI)	99.3% (98.9%–99.7%)	99.8% (99.6%–99.9%)	99.5% (99.1%–99.8%)	0.064	0.68
Accuracy (95% CI)	99.1% (98.7%–99.6%)	99.5% (99.2%–99.8%)	98.1% (97.5%–98.8%)	0.087	<0.001
Referable DR					
Sensitivity (95% CI)	89.2% (85.2%–93.3%)	71.1% (66.8%–75.1%)	–	<0.001	<0.001
Specificity (95% CI)	90.1% (88.7%–91.6%)	98.9% (98.5%–99.2%)	–	<0.001	–
Accuracy (95% CI)	90.0% (88.6%–91.3%)	92.9% (91.9%–94.0%)	–	<0.001	

CI: confidence interval; DR: diabetic retinopathy; PDR: proliferative diabetic retinopathy.

90.1% at an optimal operating point. Such results were comparable to those in previous studies using real-world datasets for validation.[11,12] We believe that such validation results are more applicable to the real-world practice.

After the pre-training process, we used the VeriSee™ algorithms to examine the validation dataset and found several false positive cases that were diagnosed as referable DR or PDR after the pre-training. After the fine-tune training, however, these cases were then correctly diagnosed as no DR or mild NPDR. A certain portion of these cases were found to have artefacts caused by poor photographing or severe vitreous opacity such as asteroid hyalosis. Fig. 4 shows two typical cases with these artefacts. Such artefacts are common in real-world nonmydriatic fundoscopy. Our present study documented that training with real-world fundus images of inconsistent qualities is important for imaging identification with deep leaning algorithms; validation with these fundus images of inconsistent qualities is also important to reveal the situations in clinical practicing.

In the present study, we compared the diagnostic results of the VeriSee™ algorithms to those of ophthalmologists as well as internal physicians. We found that the ophthalmologists had a poorer sensitivity than the VeriSee™ algorithms in detecting referable DR. Such discrepancy between artificial intelligence and ophthalmologists may be attributed to the limitations of human eyes in the case of poor environment and poor image quality. It was noted during the study that, if the ophthalmologists were reading the fundus images in an environment with bright background light or if they have read the images continuously for a long time, small lesions tended to be ignored. Such condition was even more severe if the image quality was poor. This highlights the advantage of artificial intelligence in avoiding the effect of the environment or fatigue. For the cases of PDR, it would be less of a problem for ophthalmologists to recognize neovascularization and vitreous

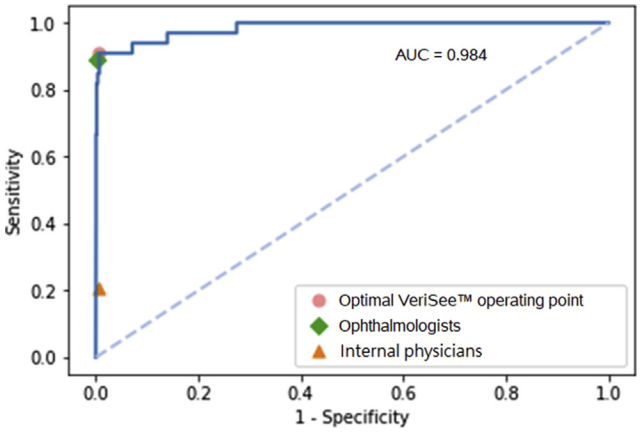

Figure 2 Receiver operating characteristic curves and the areas under the curve of the VeriSee™ pre-trained model and final model for detection of proliferative diabetic retinopathy, in comparison with ophthalmologists' and internal physicians' performances.

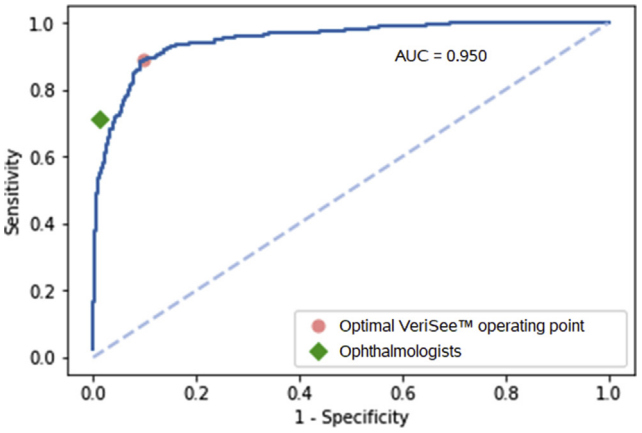

Figure 3 Receiver operating characteristic curves and the areas under the curve of the VeriSee™ pre-trained model and final model for detection of any diabetic retinopathy, in comparison with ophthalmologists' performances.

Please cite this article as: Hsieh Y-T et al., Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening, Journal of the Formosan Medical Association, https://doi.org/10.1016/j.jfma.2020.03.024

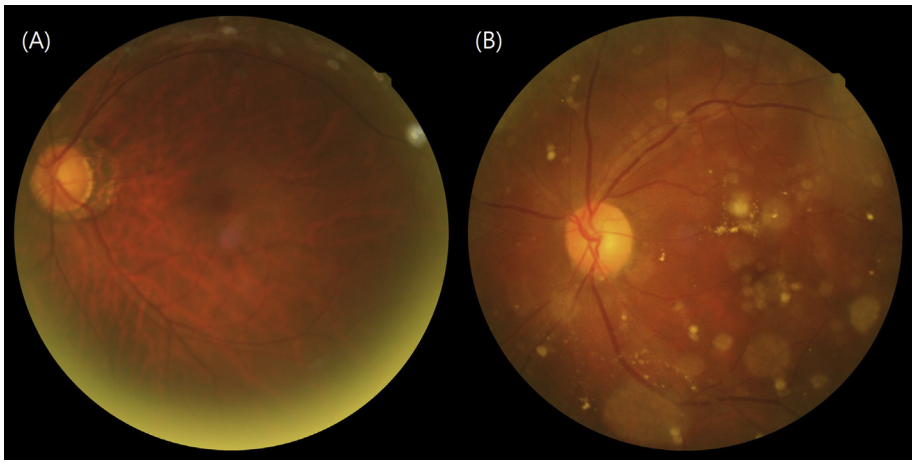

Figure 4 (A) A case of no diabetic retinopathy misdiagnosed as proliferative diabetic retinopathy by VeriSee™ due to artefects caused by poor photographing. (B) A case of no diabetic retinopathy misdiagnosed as proliferative diabetic retinopathy by VeriSee™ due to asteroid hyalosis.

hemorrhage even with poor image quality or under poor circumstance. However, non-ophthalmologists might be less experienced in distinguishing neovascularized vessels from retinal hemorrhage or intraretinal microvascular abnormalities, and could not effectively distinguish vitreous hemorrhage from other media opacity, resulting in a low sensitivity in diagnosing PDR. These observations highlight the role of automated imaging detection system using deep learning algorithms in assisted diagnosis for general practitioner and internal physicians as well as reducing the workload for ophthalmologists.

There are some limitations of this study. First, the datasets used for fine-tune training and validation were all taken with the same style of fundus camera. Further validation with images taken with different styles of fundus camera should be performed. Second, the algorithms in this study were not trained to differentiate diabetic retinopathy from other retinal diseases such as age-related macular degeneration, retinal vein occlusion, etc. On the other hand, cases with these retinal diseases were not excluded from the study. During the validation, most of these cases in fact were diagnosed as referable DR. Although the diagnosis was wrong, these cases indeed should be referred for further evaluation. Third, there was no available data for the diagnosis of referable DR from internal physicians. This also reflects the reality that internal physicians were not familiar with the scoring criteria of referable DR. Nonetheless, this is the first study to compare the diagnoses of DR/PDR made by deep learning algorithms and those by internal physicians, indicating a true value of AI-assisted DR screening in clinical management of diabetes individuals by primary care physicians. Finally, only one fundus picture containing both fovea and optic disc was used for the diagnosis of DR in this study. Compared to seven-field standard fundus photography, single-field fundus photography offers less information; however, it had been proved to be good enough as a screening tool for diabetic retinopathy.[31,32] Single-field fundus photography has also been used for deep learning model development and validation by the Google Inc. team.[8] We chose to use single-field fundus photography in this study because most non-ophthalmologists in Taiwan use it for DR screening in clinical practicing.

In conclusion, this study demonstrated that the algorithms of the automated imaging identification system VeriSee™ had good sensitivity and specificity in diagnosing DR. This is also the first study to demonstrate the superiority of deep learning algorithms in diagnosing DR when compared with internal physicians, implicating that deep learning algorithms may be helpful in assisting non-ophthalmologists in screening DR with fundus photography.

Declaration

Acer Inc. and National Taiwan University own the intellectual property of VeriSee™, and were involved in the design and conduct of the study. Y-T Hsieh, L-W Chan, T-Y Kao, T-C Chen and H-C Lin report receipt of grants from Acer Inc. for retinal fundus photography reading.

Conflict of interest

Yi-Ting Hsieh, Li-Wei Chan, Tzu-Yun Kao, Ta-Ching Chen and Hsuan-Chieh Lin report receipt of grants from Acer Inc. for retinal fundus photography reading. Other authors have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jfma.2020.03.024.

References

1. Klein R, Knudtson MD, Lee KE, Gangnon R, Klein BE. The Wisconsin epidemiologic study of diabetic retinopathy: XXII the twenty-five-year progression of retinopathy in persons with type 1 diabetes. *Ophthalmology* 2008;**115**:1859–68.
2. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet* 2010;**376**:124–36.

3. American Diabetes A. 10. Microvascular complications and foot care: standards of medical care in diabetes-2018. *Diabetes Care* 2018;**41**:S105—18.
4. http://www1.nhi.gov.tw/mqinfo/SearchPro.aspx?Type=DM&List=4.
5. Al Rasheed R, Al Adel F. Diabetic retinopathy: knowledge, awareness and practices of physicians in primary-care centers in Riyadh, Saudi Arabia. *Saudi J Ophthalmol: Off J Saudi Ophthalmol Soc* 2017;**31**:2—6.
6. Delorme C, Boisjoly HM, Baillargeon L, Turcotte P, Bernard PM. Screening for diabetic retinopathy. Do family physicians know the Canadian guidelines? *Can Fam Physician* 1998;**44**:1473—9.
7. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, 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.
8. 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;**316**:2402—10.
9. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;**124**:962—9.
10. 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.
11. Takahashi H, Tampo H, Arai Y, Inoue Y, Kawashima H. Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy. *PLoS One* 2017;**12**:e0179790.
12. 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;**318**:2211—23.
13. Keel S, Wu J, Lee PY, Scheetz J, He M. Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma. *JAMA Ophthalmol* 2019;**137**:288—92.
14. Lam C, Yi D, Guo M, Lindsey T. Automated detection of diabetic retinopathy using deep learning. *AMIA Jt Summits Transl Sci Proc* 2018;**2017**:147—55.
15. Sayres R, Taly A, Rahimy E, Blumer K, Coz D, Hammel N, et al. Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. *Ophthalmology* 2019;**126**:552—64.
16. van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol* 2018;**96**:63—8.
17. Cuadros J, Sim I. EyePACS: an open source clinical communication system for eye care. *Stud Health Technol Inform* 2004;**107**:207—11.
18. Wilkinson CP, Ferris 3rd FL, Klein RE, Lee PP, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;**110**:1677—82.
19. Chen Q, Peng Y, Keenan T, Dharssi S, Agro NE, Wong WT, et al. A multi-task deep learning model for the classification of Age-related Macular Degeneration. *AMIA Jt Summits Transl Sci Proc* 2019;**2019**:505—14.
20. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. learning. In: *AAAI'17 proceedings of the thirty-first AAAI conference on artificial intelligence*; 2017. p. 4278—84.
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 27—30 June 2016*; 2016. p. 770—8.
22. Lin T-Y, Dollár P, Girshick RB, He K, Hariharan B, Belongie SJ. Feature pyramid networks for object detection. In: *CVPR*. IEEE Computer Society; 2017. p. 936—44.
23. Philip S, Fleming AD, Goatman KA, Fonseca S, McNamee P, Scotland GS, et al. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol* 2007;**91**:1512—7.
24. Bouhaimed M, Gibbins R, Owens D. Automated detection of diabetic retinopathy: results of a screening study. *Diabetes Technol Ther* 2008;**10**:142—8.
25. Goatman K, Charnley A, Webster L, Nussey S. Assessment of automated disease detection in diabetic retinopathy screening using two-field photography. *PLoS One* 2011;**6**:e27524.
26. Fleming AD, Goatman KA, Philip S, Williams GJ, Prescott GJ, Scotland GS, et al. The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy. *Br J Ophthalmol* 2010;**94**:706—11.
27. Niemeijer M, Abramoff MD, van Ginneken B. Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Trans Med Imaging* 2009;**28**:775—85.
28. Soto-Pedre E, Navea A, Millan S, Hernaez-Ortega MC, Morales J, Desco MC, et al. Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists' workload. *Acta Ophthalmol* 2015;**93**:e52—6.
29. Norgaard MF, Grauslund J. Automated screening for diabetic retinopathy - a systematic review. *Ophthalmic Res* 2018;**60**:9—17.
30. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;**115**:211—52.
31. Vujosevic S, Benetti E, Massignan F, Pilotto E, Varano M, Cavarzeran F, et al. Screening for diabetic retinopathy: 1 and 3 nonmydriatic 45-degree digital fundus photographs vs 7 standard early treatment diabetic retinopathy study fields. *Am J Ophthalmol* 2009;**148**:111—8.
32. Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR. Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. *Ophthalmology* 2004;**111**:1055—62.

Please cite this article as: Hsieh Y-T et al., Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening, Journal of the Formosan Medical Association, https://doi.org/10.1016/j.jfma.2020.03.024