

Evaluation of a novel deep learning based screening system for pathologic myopia

Pei-Fang Ren¹, Xu-Yuan Tang¹, Chen-Ying Yu¹, Li-Li Zhu¹, Wei-Hua Yang², Ye Shen¹

¹Department of Ophthalmology, the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, Zhejiang Province, China
²Shenzhen Eye Hospital, Jinan University, Shenzhen 518040, Guangdong Province, China

Correspondence to: Wei-Hua Yang. Shenzhen Eye Hospital, Jinan University, Shenzhen 518040, Guangdong Province, China. benben0606@139.com; Ye Shen. Department of Ophthalmology, the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, Zhejiang Province, China. idrshen@zju.edu.cn

Received: 2023-04-04        Accepted: 2023-05-10

Abstract

- **AIM:** To evaluate the clinical application value of the artificial intelligence assisted pathologic myopia (PM-AI) diagnosis model based on deep learning.
- **METHODS:** A total of 1156 readable color fundus photographs were collected and annotated based on the diagnostic criteria of Meta-pathologic myopia (PM) (2015). The PM-AI system and four eye doctors (retinal specialists 1 and 2, and ophthalmologists 1 and 2) independently evaluated the color fundus photographs to determine whether they were indicative of PM or not and the presence of myopic choroidal neovascularization (mCNV). The performance of identification for PM and mCNV by the PM-AI system and the eye doctors was compared and evaluated via the relevant statistical analysis.
- **RESULTS:** For PM identification, the sensitivity of the PM-AI system was 98.17%, which was comparable to specialist 1 ($P=0.307$), but was higher than specialist 2 and ophthalmologists 1 and 2 ($P<0.001$). The specificity of the PM-AI system was 93.06%, which was lower than specialists 1 and 2, but was higher than ophthalmologists 1 and 2. The PM-AI system showed the $Kappa$ value of 0.624, while the $Kappa$ values of specialists 1, 2 and ophthalmologists 1 and 2 were 0.864, 0.732, 0.304 and 0.238, respectively.
- **CONCLUSION:** In comparison to the senior ophthalmologists, the PM-AI system based on deep learning exhibits excellent performance in PM and mCNV identification. The effectiveness of PM-AI system is an auxiliary diagnosis tool for clinical screening of PM and mCNV.

- **KEYWORDS:** artificial intelligence; deep learning; pathologic myopia; choroidal neovascularization

DOI:10.18240/ijo.2023.09.07

Citation: Ren PF, Tang XY, Yu CY, Zhu LL, Yang WH, Shen Y. Evaluation of a novel deep learning based screening system for pathologic myopia. Int J Ophthalmol. 2023;16(9):1417-1423

INTRODUCTION

Myopia, a most prevalent ocular refractive error status, has caused considerable impact on the public health and socioeconomic well-being burden. Pathologic myopia (PM), a severe form of myopia commonly seen in high myopia worse than -6.00 D, is complicated with the characteristic maculopathy lesions and may result in the great risk of sight-threatening complications. The global incidence of PM was about 1%-4% in the general population and 5%-10% in diagnosed myopia, with a more serious situation in the Eastern Asia nations.

The pathological lesions of PM are often irreversible, due to the lack of effective treatment and poor prognosis. The early identification and timely medical intervention are critically important for alleviating the vision impairment caused by PM. Moreover, as the progression of PM lesions is slow, the patients may neglect the ocular symptoms and thus miss the chance of intervention in the early stage of PM. Therefore, the effective and professional screening for the relevant PM lesions is an urgent mission for public vision function health care, but also a great challenge in view of the large myopic population.
A Meta-analysis for PM (Meta-PM) study group provided a novel simplified and uniform systematic classification system for myopic maculopathy based on color fundus photography in 2015, which provided a practical screening criterion for PM diagnosis and analysis⁹[^9]. According to this criterion, PM was defined as myopic maculopathy of a category equal to or more serious than diffuse choroidal atrophy or with at least one “Plus” lesion [lacquer crack (LC)/myopic choroidal neovascularization (mCNV)/Fuchs spot]. Among the “Plus” lesions, Fuchs spots are considered dry fibrovascular scars of mCNV and barely need treatment and LC, which may be accompanied by chorioretinal hemorrhage and may progress to mCNV or patchy atrophy[^10], is often difficult to be observed on color fundus photographs[^11]. The mCNV is the major cause of visual impairment in PM and may result in patchy choroidal or macular atrophy, leading to irreversible visual loss if without prompt intervention⁹[^9,12]. Recently, the intravitreal anti-vascular endothelial growth factor drugs showed definite therapeutic effects and is the first-line option for mCNV treatment⁹[^12-14]. Therefore, early detection and treatment of mCNV are of paramount importance.

Of note, the large scale screening task for the aforementioned PM lesions in specialist eye clinic or medical care institutes is exhausting if completed by man power alone, especially in the nations with high prevalence of myopia/PM. In addition, primary care general ophthalmologists often lack the requisite expertise to identify PM and its complications, while there is always a serious shortage of ocular fundus specialists.

With the rapid development of artificial intelligence (AI) technology, the AI-assisted medical diagnosis has started to play an important role in the clinical management. With the aid of AI technology, the mission with daunting working volume of PM screening become feasible. Among AI, the deep learning system is a sophisticated subclass that mimics the way of the human brain works and has excellent performance comparable to that of board-certified experts in the respect of massive medical image categorizing[^15-16]. The deep learning related diagnosis systems have been successfully applied in the identification of diabetic retinopathy, glaucoma and other ocular diseases[^17-20].

In previous studies, we developed a series of color fundus photograph based AI models (PM-AI), and finally achieved the following three tasks: 1) automatic identification of PM; 2) automatic classification of myopic maculopathy categories; 3) detection and localization of “Plus” lesions. The algorithms and AI-models achieved the robust performance in the external verification set[^20-23].

The aim of this study is to further evaluate the feasibility and effectiveness of our in-house developed AI models for PM screening and mCNV detection by comparing the performance of PM-AI models with retinal specialists and senior primary ophthalmologists in a real world clinical scenario.

**SUBJECTS AND METHODS**

**Ethical Approval** The study was approved by the Ethics Committee of the First Affiliated Hospital, School of Medicine, Zhejiang University (approval ID: No.2020–693), and adhered to the tenets of the Declaration of Helsinki. Informed consent was waived by the ethics committee due to the retrospective and fully anonymized nature of the data.

**Dataset** A total of 1265 color fundus photographs were collected from the eye center of the First Affiliated Hospital of School of Medicine, Zhejiang University, between January 2021 and December 2021. Among them, 965 color fundus photographs were from the patients with high myopia (spherical equivalent less than -6.00 D or axial length longer than 26.0 mm), and 300 color fundus photographs were from the patients without high myopia. The desktop nonmydriatic retinal cameras and digital retinography systems (Canon, Japan) were used to capture the color fundus photographs, which were maculalutea-centered 45° color fundus photographs.

**Annotation** The quality of the included color fundus photographs was assessed by one ophthalmic technician to exclude ungradable images caused by the factors of underexposure, false focus, or clouding of the refractive media. The criteria to determine a gradable image includes: 1) Images should be the standard 45° field with entire optic nerve head and macula included; 2) Images should have perfect exposure without dark and washed-out areas; 3) Images should with fewer artifacts, avoiding dust spots, arc defects due to eyelid blocking and eyelash images; 4) The image focus should be good for small retinal lesions identification. The gradable images were then assigned to three retinal specialists for independent annotation: PM or non-PM, mCNV or no mCNV. The diagnosis of PM was made according to the Meta-PM (2015) criteria as described above. The unanimous results agreed by all three retinal specialists were used as the reference standard for the subsequent study. A senior retinal specialist was required to arbitrate the results to make the final annotation decision, if any inconsistent results among the three specialist annotators existed.

**Artificial Intelligence Identification** The deep learning system (PM-AI) was used to identify the relevant lesions in the color fundus photographs. This PM-AI system was previously developed based on ResNet18 and the Faster Region-based Convolution Neural Network (R-CNN) algorithm, and was trained using a dataset of 32 010 color fundus photographs[^23]. The PM-AI system automatically analyzed and identified fundus signs in the input images. It then generated the PM screening report including the results of PM or non-PM, the
category of myopic maculopathy, and the presence of Plus lesions, if any. The results of whether these images had PM or mCNV were recorded.

**Human Identification** Two retinal specialists with more than 10y clinical experience from the tertiary hospitals (specialists 1 and 2) and two senior ophthalmologists from the district hospitals (ophthalmologists 1 and 2), who were not taught with the Meta-PM (2015) diagnose criteria or participated in the previous work of the algorithm development, independently annotated the images based on their own clinical expertise. The judgement of PM/non-PM and presence/absence of mCNV were then made by the two specialists and ophthalmologists respectively.

**Statistical Analysis** SPSS Statistics 25.0 (IBM, Armonk, NY) was used for the statistical analyses. The statistical indicators included sensitivity, specificity, and diagnostic test consistency (*Kappa* value).

According to the reference standard, *Kappa* values <0.40 were considered to indicate poor consistency, values between 0.41-0.75 were considered to indicate good consistency, and values between 0.75-1.0 were considered to indicate perfect consistency. The 95% confidence intervals (CIs) were calculated using the modified Wald method. The McNemar test was performed to compare the performances of PM-AI and different doctors, with a *P* value of <0.05 considered statistically significant. The detailed calculation formulae were as follows: Accuracy=(TP+TN)/(TP+FN+FP+TN), Sensitivity=TP/(TP+FN), Specificity=TN/(TN+FP), TP: true positive, FN: false negative, FP: false positive, TN: true negative.

**RESULTS**

A total of 1265 color fundus photographs were collected and 109 unreadable images were excluded. The remaining 1156 images (including 877 with high myopia and 279 without high myopia) were all considered as readable after being uploaded to the PM-AI software. The characteristics of the color fundus photograph dataset were shown in Table 1.

For PM/non-PM identification, sensitivity of the PM-AI system was 98.17%, which was not statistically significantly different from specialist 1 (96.96%), but was higher than that of specialist 2 and ophthalmologists 1 and 2 (*P*<0.05). The specificity of the PM-AI system was 93.06%, which was lower than that of specialists 1 and 2 (both 99.55%), but higher than that of ophthalmologists 1 and 2 (89.59% and 84.16%), and the differences were all statistically significant. Compared with the reference standard, the *Kappa* value of the PM-AI system was 0.904, indicating the perfect consistency (Table 2).

For the identification of mCNV, the PM-AI system showed the sensitivity level of 84.06%. There was no statistically significant difference in sensitivity level between the PM-AI system and specialists 1, 2 or ophthalmologist 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher sensitivity than ophthalmologist 1 (63.77%). The specificity of the PM-AI system was 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificity of the PM-AI system was 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%). The specificities of the PM-AI system were 95.31%, which was lower than specialists 1 and 2 (81.16%, 86.96%, and 92.75%, respectively), while the PM-AI system showed significantly higher specificity than ophthalmologist 1 (63.77%).
C2 was misclassified as C1, or in images with atypical LCs (Figure 1). On the other hand, the false positives for mCNV detection were mainly due to the factors of macular hemorrhages caused by other diseases, small macular atrophy, and photographic reflective artifacts. While the false negatives for mCNV detection were mostly due to occult CNVs, which were relatively small in size or with only trace hemorrhages (Figure 2).

**DISCUSSION**

In previous work, we developed the algorithms to...
automatically identify the PM lesions based on a large image dataset consisting of 37 659 color fundus photographs. ResNet18 was used as the basic architecture, and the Faster R-CNN+FPN was used for the localization algorithm. ResNet has shown excellent performance in the ImageNet competition classification task\cite{24,25}. Meanwhile, the Faster R-CNN is also one of the most advanced object detection networks\cite{26}. The algorithms and AI models developed based on the two advanced neural network architectures achieved precise and efficient detection for PM\cite{27}.

Tan et al\cite{28} introduced an approach of PAMELA to detect PM according to the peripapillary atrophy feature in 2009. With the establishment of the Meta-PM classification system in 2015, the Meta-PM (2015) criterion was broadly adopted for color fundus photograph annotation and machine learning training in the recent years. Based on the Meta-PM, Du et al\cite{29} developed an AI model showing an overall accuracy of 92.08% in identifying PM, and accuracy of 37.07% in detecting mCNV. Another deep learning system developed by a multi-center study showed a sensitivity of 94.2%-98.4%, a specificity of 85.5%-95.9%, and area under the curves (AUCs) of 0.969 or higher for PM identification in the multiple validation dataset\cite{30}.

The dual-stream deep convolutional neural networks (DCNN-DS) model developed by Li et al\cite{31} achieved sensitivities of 93.3% and 91.0%, specificities of 99.6% and 98.7%, and AUCs of 0.998 and 0.994 for detecting PM. Recently, Tang et al\cite{32} developed a co-decision model combined with the ResNet-50 network and DeepLabv3+ network and achieved a sensitivity of 96.67%, a specificity of 99.15%, and an AUC of 99.80% for PM diagnosis. This model can identify different levels of chorioretinal atrophies and LCs, but cannot identify mCNV and Fuchs spots. The authors ascribed the weakness to the small sample size of CNV and Fuchs spots recruited in their dataset.

Besides the good capability in PM identification (sensitivity 0.939, specificity 0.981, AUC 0.995), the PM-AI system previously developed by our team also had excellent performance in the identification of different chorioretinal atrophies and “Plus” lesions, with the sensitivity and specificity of mCNV detection being over 97\%\cite{33}. In this extended study, the performance of the PM-AI system for the identification of PM and mCNV was basically comparable to the level of ocular fundus specialists, and was significantly better than that of senior ophthalmologists. The results from the human-machine comparison test suggested that our PM-AI system is promising in assisting the diagnosis of PM and mCNV in real world settings of either clinical or healthcare setting.

In particular, our PM-AI model performed well in identifying PM, showing better sensitivity than one specialist participant and two ophthalmologist participants, and higher specificity than two ophthalmologist participants, with strong agreement with the reference standard (Kappa value 0.904). However, the performance of mCNV detection, especially the sensitivity, was slightly inferior to the data in our previous work. This might be due to the relatively small dataset in the present study. The mCNV lesions in PM patients are typically small, dark in color (mostly gray-green or gray-black), with little hemorrhage, but often difficult to be detect in a tessellated fundus with diffuse or patchy choroidal atrophy. Nevertheless, our PM-AI model still exhibited the comparable sensitivity of mCNV identification when compared with the specialists, demonstrating its robust performance.

According to our data, similar misclassification might occur by both human and AI in the cases with lesions of classification boundary. For example, “tessellated fundus (C1)” and “diffuse atrophy (C2)” could be misdiagnosed as each other. Besides, some other small and uncertain lesions also can be misclassified. Human’s mistake occurs mainly due to the factors of distraction, fatigue or clinical experience, while AI can make the mistakes, such as misclassifying the hemorrhage of diabetic retinopathy as mCNV and misclassifying the small vitreous opacity artifact as patchy atrophy, etc. The reason can be ascribed to the limited disease contents of training dataset for AI system development. Therefore, to reduce the false diagnosis of AI, the more training contents for AI regarding the disease/lesions category are always the effective way. This is also the part of our ongoing work for AI algorithm development in the future. It is noticeable that ophthalmologist 2 showed the highest sensitivity (92.75%) in identifying mCNV, but the dramatically low specificity (74.61%). The overall diagnostic agreement was also significantly lower than the other three doctors and the PM-AI model. This might indicate the fluctuation of performance of manpower caused by the variation of working status or expertise of human, while the AI system ought to be more stable and superior especially in heavy and repetitive task.

It should be noted that the four human participants are all high standard ophthalmologists with two retinal specialists and two senior ophthalmologists, who are all from the hospitals with advanced medical system and high clinical level. After validation, the PM-AI model outperformed the two senior ophthalmologists in PM diagnosis and mCNV detection. This result highlighted the great potential of PM-AI to assisting the eye clinicians to screen PM and its lesions.

PM is a comprehensive diagnosis based on the grade of choriotreinal atrophies and “Plus” lesions, of which mCNV is one of the most critical macular lesions affecting central vision. Therefore, the present study mainly focused on the diagnosis of PM and only mCNV of three “Plus” lesions, considering
their clinical priority. After all, the larger volume of dataset of mCNV images is always desirable to draw a more convincing conclusion.

Taken together, our in-house developed PM-AI system based on deep learning showed the excellent performance of identifying PM and mCNV that was comparable to the retinal specialists in this human-machine comparison test study. When compared with the senior ophthalmologists, the PM-AI system exhibited superior performance for PM and mCNV identification. In view of the shortage in current medical care resources, the AI-assisted diagnosis system should have wide application prospects in PM screening and many other ocular diseases.

ACKNOWLEDGEMENTS

Conflicts of Interest: Ren PF, None; Tang XY, None; Yu CY, None; Zhu LL, None; Yang WH, None; Shen Y, None.

REFERENCES


21 Wang RY, Zuo G, Li KK, Li WT, Xuan QZ, Han YZ, Yang WH. Systematic bibliometric and visualized analysis of research hotspots and trends on the application of artificial intelligence in diabetic retinopathy. Front Endocrinol (Lausanne) 2022;13:1036426.


