

# Automated diagnoses of age-related macular degeneration and polypoidal choroidal vasculopathy using bi-modal deep convolutional neural networks

Zhiyan Xu <sup>1,2</sup>, Weisen Wang,<sup>3</sup> Jingyuan Yang,<sup>1,2</sup> Jianchun Zhao,<sup>4</sup> Dayong Ding,<sup>4</sup> Feng He <sup>1,2</sup>, Di Chen,<sup>1,2</sup> Zhikun Yang,<sup>1,2</sup> Xirong Li,<sup>5</sup> Weihong Yu,<sup>1,2</sup> Youxin Chen<sup>1,2</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bjophthalmol-2020-315817>).

<sup>1</sup>Department of Ophthalmology, Peking Union Medical College Hospital, Dongcheng District, Beijing, China

<sup>2</sup>Key Laboratory of Ocular Fundus Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>3</sup>AI & Media Computing Lab, School of Information, Renmin University of China, Beijing, China

<sup>4</sup>Vistel AI Lab, Visionary Intelligence Ltd, Beijing, China

<sup>5</sup>Department for state-of-the-art ophthalmology AI research & development, Key Lab of DEKE, Renmin University of China, Beijing, China

## Correspondence to

Professor Youxin Chen, Department of Ophthalmology, Peking Union Medical College Hospital, Beijing 100730, China; [chenyouxinpumch@163.com](mailto:chenyouxinpumch@163.com)

Received 4 January 2020

Revised 31 March 2020

Accepted 14 May 2020

Published Online First

4 June 2020



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Xu Z, Wang W, Yang J, et al. *Br J Ophthalmol* 2021;**105**:561–566.

## ABSTRACT

**Aims** To investigate the efficacy of a bi-modality deep convolutional neural network (DCNN) framework to categorise age-related macular degeneration (AMD) and polypoidal choroidal vasculopathy (PCV) from colour fundus images and optical coherence tomography (OCT) images.

**Methods** A retrospective cross-sectional study was proposed of patients with AMD or PCV who came to Peking Union Medical College Hospital. Diagnoses of all patients were confirmed by two retinal experts based on diagnostic gold standard for AMD and PCV. Patients with concurrent retinal vascular diseases were excluded. Colour fundus images and spectral domain OCT images were taken from dilated eyes of patients and healthy controls, and anonymised. All images were pre-labelled into normal, dry or wet AMD or PCV. ResNet-50 models were used as the backbone and alternate machine learning models including random forest classifiers were constructed for further comparison. For human-machine comparison, the same testing data set was diagnosed by three retinal experts independently. All images from the same participant were presented only within a single partition subset.

**Results** On a test set of 143 fundus and OCT image pairs from 80 eyes (20 eyes per-group), the bi-modal DCNN demonstrated the best performance, with accuracy 87.4%, sensitivity 88.8% and specificity 95.6%, and a perfect agreement with diagnostic gold standard (Cohen's  $\kappa$  0.828), exceeds slightly over the best expert (Human1, Cohen's  $\kappa$  0.810). For recognising PCV, the model outperformed the best expert as well.

**Conclusion** A bi-modal DCNN for automated classification of AMD and PCV is accurate and promising in the realm of public health.

## INTRODUCTION

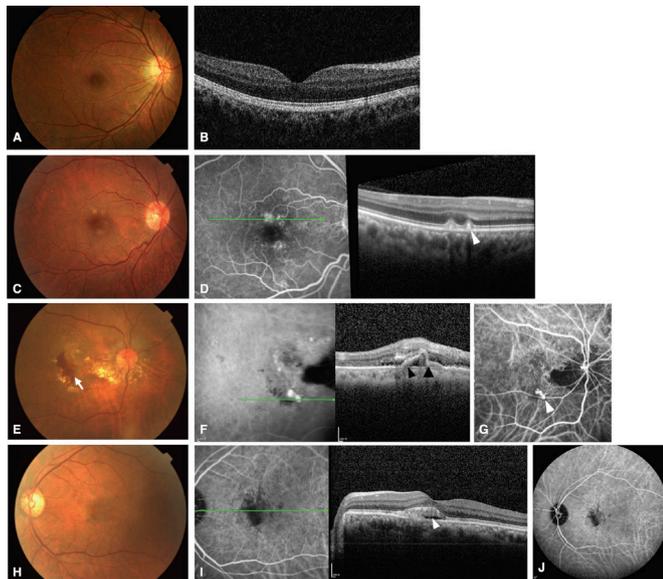
Age-related macular degeneration (AMD) is a disease affecting macular area of retina and causing severe visual impairment.<sup>1</sup> It is characterised by drusen and basal laminal deposits between the retinal pigment epithelium (RPE) and Bruch's membrane in the early stage.<sup>2</sup> Patients in early AMD are usually asymptomatic or presented with mild central vision distortion. Late AMD includes neovascular form and atrophic form. Choroidal neovascularisation (CNV) is further classified into three subtypes, the central vision loss can deteriorate rapidly.<sup>1</sup> Atrophic AMD progresses more stable and is less urgent.<sup>3</sup> Clinical diagnosis of AMD is

based on auxiliary examinations, as colour fundus photographs, spectral-domain optical coherence tomography (SD-OCT) and OCT angiography with higher resolution.<sup>1</sup>

Polypoidal choroidal vasculopathy (PCV) is a subtype of Type 1 CNV, the neovascularisation grows below the RPE.<sup>4</sup> PCV is characterised by recurrent exudation, haemorrhages and aneurysmal dilations.<sup>5</sup> The prevalence of PCV among patients with AMD is much higher in African and Asian populations (22% to 62%) than in European populations (8% to 13%).<sup>6</sup> In both Japanese Study Group Guidelines and EVEREST criteria for PCV diagnosis, colour fundus photographs and indocyanine green angiography (ICGA) are essential for lesion detection.<sup>7</sup> ICG is a dye having different absorption and reflection spectrum from RPE, which facilitates better visualisation of choroid and relating haemorrhage.<sup>8</sup> It has fast first-pass effect in liver and is metabolised through kidney.<sup>8</sup> ICGA is invasive, severe adverse effects as urticaria and anaphylactic reactions in patients have been reported.<sup>9</sup> This study aimed to develop an artificial intelligence system to assist the non-invasive diagnosis of PCV when ICGA is not applicable.

Currently available clinical trials highly recommend initial combination therapy of photodynamic therapy and anti-vascular endothelial growth factor (anti-VEGF) therapy in some patients of PCV can significantly prolong the gap period for retreatment.<sup>10</sup> Nevertheless, the effective strategy for neovascular AMD, anti-VEGF therapy, is less strongly proved to be efficient in PCV.<sup>4</sup> Given the fact that the different subtypes of AMDs can affect clinician's decision for therapy, it is necessary to make an accurate diagnosis and differentiation between PCV and other neovascular AMDs.

Machine learning has been widely used in the area of medical imaging and computer-aided diagnosis (CAD).<sup>11</sup> Its subclassification, deep learning, has revolutionised the learning process from feature input as object segmentation and feature extraction to image input in the use of pixels. This can maintain all information on pixels as well as avoid errors by wrong feature extraction.<sup>12</sup> The use of deep convolutional neural networks (DCNN) in analysing AMD has been applied by scientists widely.<sup>13 14</sup> Burlina, *et al* first applied DCNN to achieve automated grading of AMD into two-class classification (the disease-free/early stage and the referable intermediate/advanced stage) using colour



**Figure 1** Example images of four multi-classification. Example fundus images from healthy controls (A,B), dry AMD (C,D), PCV (E,G) and wet-AMD (H,J). Example of colour photograph (A) and sections of SD-OCT cross the macula (B) of healthy controls. C,D show the colour fundus and OCT images of dry AMD, the left half of OCT image depicts the horizontal section crossing the disease loci, the right side depicts the drusen (white arrowhead) on OCT B-scan. C shows retinal haemorrhage (white arrow) and hard exudation (blue arrow) around the macula on colour fundus image of PCV. F shows specific double layer sign (black arrowhead), notch of the retinal pigmental epithelium detachment of PCV (black arrow) on OCT as demonstrated in previous studies.<sup>18</sup> G depicts the example of ICGA images of PCV, the specific hyperfluorescence nodular 'polyp' sign (white arrowhead) can be seen. H depicts the colour fundus of wet-AMD. Subretinal exudation (white arrowhead) can be seen on this section crossing macula fovea of SD-OCT (I). No fluorescence leakage can be detected on ICGA (J). AMD, age-related macular degeneration; ICGA, indocyanine green angiography; OCT, optical coherence tomography; PCV, polypoidal choroidal vasculopathy; SD-OCT, spectral-domain optical coherence tomography.

fundus images,<sup>15</sup> then he optimised the classification system to first using four-step classification then fusing into the final two classes.<sup>16</sup> Besides the application in AMD grading, detection of specific lesions and subtypes of AMD was studied.<sup>17–20</sup> Such as DCNN for automated detection of exudative AMD from healthy controls using cross-section SD-OCT images,<sup>19</sup> the detection of geographical atrophy from both healthy and other retinal diseases groups in fundus autofluorescence (FAF) images,<sup>20</sup> and the recognition of foveal centre based on pixel-wise classification in OCT scans from AMD patients at different severity stages.<sup>18</sup> Endeavour to technical refinement broadly expand the ability of artificial intelligence (AI), as demonstrated in previous studies for successful retinal layer segmentation,<sup>21</sup> detection and segmentation of pigment epithelium detachment, subretinal fluid and intraretinal fluid in OCT B-scans of AMD and other retinal diseases.<sup>22</sup> Further approaches of DCNN have been applied in decision-making<sup>23</sup> and evaluation of anti-VEGF treatment indication for AMD.<sup>24</sup> All studies involved only one examination method, either colour fundus photographs or SD-OCT.<sup>15 25–28</sup> Besides, PCV is not identified in all deep learning networks for classification of AMD. The new study hereby combined these two examination tools in the aid of diagnoses of PCV and AMD.

## METHODS

Given a colour fundus image and an OCT image from a specific eye, the proposed deep learning model aims to recognise what types of AMD is present in the eye. As such, the model takes simultaneously the fundus image and the OCT image as a bi-modal input. The output of the model has four ways, wet AMD (excluding PCV), dry AMD (atrophic AMD and early stage AMD), PCV and normal. The primary goal of this study is to analyse and compare the performance of DCNNs with different input. The input can be fundus images alone, SD-OCT images alone or their combination. The secondary goal is to compare the performance between models and ophthalmology specialists for PCV categorisation.

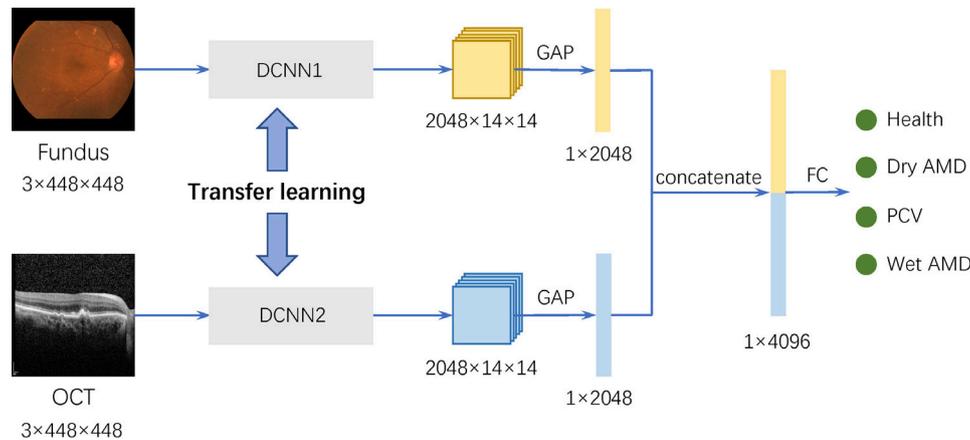
This was a retrospective cross-sectional study of patients from outpatient clinic of the Peking Union Medical College Hospital (PUMCH) from January 2013 to July 2018. The inclusion criteria were diagnosis of AMD or PCV, and images of colour fundus photograph, SD-OCT and ICGA were available. Exclusion criteria were concurrent retinal vascular diseases including diabetic retinopathy, glaucoma and posterior staphylomas, which may confound the classification. Normal eyes images for control were collected from the Medical Examination Center of PUMCH, including both fundus photographs and SD-OCT images.

All patients were first evaluated by dilated fundus examinations according to the Wisconsin grading consortium (online supplementary table 1).<sup>29</sup> Early AMD and late atrophic AMD were grouped as dry AMD. In late exudative AMD, PCV was confirmed based on the EVEREST diagnosis criteria (online supplementary table 1), with focal hyperfluorescent lesions before 6 min on ICGA and at least one of the four evidences from ICGA or orange subretinal nodule on colour photography, or associated massive submacular haemorrhage.<sup>7</sup> While the remaining was classified as wet AMD. Example images per classification were shown in figure 1. All diagnoses were confirmed by two ophthalmology specialists independently, with disagreement further diagnosed by a senior specialist. Those predefined groups serve as the golden standard for evaluating machine learning algorithms.

Colour fundus images from dilated pupils were acquired with Topcon (Topcon Corporation, Tokyo, Japan), at a 45° covering the whole macula and optic disc. One colour-fundus image per eye was downloaded. One or more horizontal cross-section B-Scan images of the diagnosis significant lesion sites from macula region were manually selected and obtained with Topcon and Heidelberg (SPECTRALIS, Heidelberg Engineering, Heidelberg, Germany). All images were downloaded in a standard JPEG format according to the manufacturer's instructions, they were preprocessed to remove any diagnosis information that might reveal personal information and artefacts, such as cross-section lines and scales, which may interfere with model training.

DCNNs have been the most successful approach to medical image categorisation.<sup>15 26 30</sup> A typical DCNN architecture consists of two building blocks, a convolutional block and a task block. The convolutional block, mainly consisting of convolutional layers and pooling layers, is used to extract multilevel discriminative features from a raw input. In the task block, features are fed and converted through multiple fully connected layers into wanted output for a specific task such as classification and regression.

A bi-modal DCNN (Model DCNN-Combo) is generated. It is simultaneously fed with a fundus image and an OCT image (both with a size of 3×448×448, detailed resizing strategy in



**Figure 2** Schematic of DCNN-Combo. The model receives a fundus image (3×448×448) and an OCT image (3×448×448), then outputs a vector with four values representing the probability of healthy, dry AMD, PCV and neovascular AMD. AMD, age-related macular degeneration; DCNN, deepconvolutional neural network; OCT, optical coherence tomography; PCV, polypoidal choroidal vasculopathy.

online supplementary figure 1) and accordingly gives a diagnostic result. **Figure 2** depicts the schematic diagram of Model DCNN-Combo, there are a fundus branch and an OCT branch to extract a 2048-dim feature vector from each modal in parallel. They are concatenated to form a combined 4096-dim vector. The combined feature is fed into a fully connected layer followed by a softmax layer. The output of the softmax layer is a 4-dim output vector, corresponding the probabilistic scores of being one of the four categories. Detailed parameters are shown in online supplementary table 2. For comparison, we train two uni-modal DCNNs from fundus (DCNN-F) and OCT (DCNN-O), respectively. The uni-modal models' architecture is based on ResNet-50.

Deep transfer learning was implemented for better performance. First, weights from a ResNet-50 pretrained on ImageNet are transferred to two uni-modal models that take fundus images and OCT images as input, respectively. These uni-modal models are then fine-tuned on our fundus and OCT data sets, separately. The whole uni-modal models are updated during the fine-tuning procedure. Second, weights of the uni-modal models are transferred to the corresponding branches in DCNN-Combo. We focus on training the task block of DCNN-Combo, with its convolutional blocks fixed.

To evaluate the impact of the task block, we substitute a random forest classifier for the softmax layer in the bi-modal DCNN, designated as RF-Combo. In addition, we re-implemented the bi-modal method (the baseline model RF-Fixed) described in Yoo *et al*,<sup>30</sup> where a VGGNet pretrained on ImageNet is used for feature extraction and a random forest classifier for classification. For a fair comparison, the VGGNet was replaced with the pretrained ResNet50, which is well recognised to be better for visual recognition.

For DCNNs training, we use cross-entropy, a standard loss function for multiclass classification. The loss is minimised by stochastic gradient descent (momentum=0.9, weight decay=1e-4). The model that obtains the best F1 score on the validation set is selected. Our DCNNs models were implemented in the PyTorch (V.1.0.0) framework.

According to the selection criteria mentioned above, a total of 1099 eyes (195 healthy, 107 dry AMD, 301 PCV, 496 wet AMD) with fundus images (online supplementary table 3) and 821 eyes (195 healthy, 62 dry AMD, 197 PCV, 367 wet AMD) with OCT images (online supplementary table 4) were enrolled in this study. A fundus and an OCT scan from the same eye are

paired and labelled by diagnosis. The performance metrics is calculated based on the prediction of image pairs rather than eyes. As mentioned before, some eyes have more than one OCT images, 20 eyes were randomly selected from each class, and were constituted into 143 image pairs for test set. The same strategy was applied to validation set, obtaining 137 pairs. The rest was partitioned into training set (85% fundus, 81% OCT data). All images of a given subject participant were involved only within one single partition subset.<sup>31</sup>

To compare the classification performance of models with human, the same test set was assigned to three retinal specialists, who independently categorised each paired image into one of four classifications. Performance from both specialists and our AI system were evaluated with comparisons to the predefined gold standard classifications.

Our performance metrics included accuracy, defined by the rate of true classification cases among the total number of cases tested. Sensitivity and specificity were measured by dividing the number of true positive classifications and true negative classifications, respectively, by the total number of test cases.<sup>26</sup> Similar to sensitivity and specificity, positive predicted value (PPV) and negative predicted value (NPV) replaced the denominator with the number of all test positives and all test negatives. Cohen's kappa coefficient ( $\kappa$ ) was used to measure inter-rater agreement for categorical items between models and gold-standard.<sup>32</sup> F1 score is another measure of the test's accuracy, which is the harmonic average of precision (ie, PPV) and recall (ie, sensitivity) in binary classification. Here in multi-classification system, we used weighted average of the F1 scores of each class to calculate the overall F1 score.<sup>33</sup> Different classifiers trade-off between specificity and sensitivity. Therefore, receiver operating characteristic (ROC) curve was implemented to evaluate performance levels among different models in distinguishing PCV from other classifications. For human-machine comparison, the performance level of individual retinal specialist was demonstrated as an operating point on the ROC curve plot.<sup>15 26</sup>

## RESULTS

The experiments used two single-modality models (first group) as DCNN-F and DCNN-O. Multi-modality deep learning models (second group) include DCNN-Combo, RF-Combo and RF-Fixed. The overall multi-classification performance levels for those algorithms and three retinal experts (third group) were

**Table 1** The overall multi-classification performance levels for machine models and retinal specialists\*

Method/model (human)	Group 1		Group 2			Group 3		
	DCNN-O	DCNN-F	DCNN-Combo	RF-Combo	RF-Fixed	Human1	Human2	Human3
Accuracy	83.2	75.0	87.4	80.4	61.5	86.0	85.3	81.1
Sensitivity	86.0	75.0	88.8	83.2	66.2	88.1	86.5	81.6
Specificity	94.1	91.7	95.6	93.2	86.3	95.2	94.8	93.4
PPV	87.0	79.9	90.3	87.0	67.1	89.1	86.2	85.0
NPV	94.1	91.9	95.6	94.1	86.4	95.2	94.9	93.4
$\kappa$ score	0.771	0.667	0.828	0.734	0.472	0.810	0.799	0.741
F1 score	0.836	0.754	0.878	0.812	0.616	0.864	0.852	0.816

\*All values present percentages except for  $\kappa$  and F1 score.

DCNN-O, deep convolutional neural network, algorithm O; DCNN-F, deep convolutional neural network, algorithm F; DCNN-Combo, deep convolutional neural network, algorithm Combo; RF-Combo, classifier random forest, combining convolutional layers; RF-Fixed, fixed weights from training in ImageNet; NPV, negative predicted value; PPV, positive predicted value.

presented in table 1. Performance levels of categorising PCV from the best models/expert in each group were further analysed in table 2 (see the confusion matrix in online supplementary figure 2).

As seen in table 1, DCNN-Combo model presented the most promising performance among these three groups, with the highest accuracy (87.4%), sensitivity (88.8%), specificity (95.6%) and perfect agreement with diagnostic gold standard ( $\kappa$ , 0.828), exceeding slightly over the best retinal expert (Human1,  $\kappa=0.810$ ). Its final F1 score was 0.878. The RF-Fixed model showed the lowest performance level of accuracy (61.5%) and moderate agreement with gold standard ( $\kappa=0.472$ ). The training loss and performance of validation data set of two uni-modal models and the bi-modal DCNN were demonstrated in online supplementary figure 3 and online supplementary table 5. The bi-modal DCNN converges fast with the pretrained parameters from uni-modal models.

In single-modality models group, DCNN-O (accuracy 83.2%,  $\kappa$  score 0.771) exhibited outstanding efficacy over DCNN-F (accuracy 75%,  $\kappa$  score 0.667) (table 1). Sensitivity of DCNN-O is more than 10 per cent higher than that of DCNN-F as well (86.0% and 75.0%, respectively). In multimodality algorithms group, the performance levels of DCNN-Combo surpassed another two models RF-Combo and RF-Fixed (accuracy=87.4%, 80.4% and 61.5%,  $\kappa=0.828$ , 0.734 and 0.472, respectively).

Moderate differences in performance levels were noted among three retinal specialists. The accuracy ranged from 81.1% to 86.0%, with an average accuracy of 84.1%. Substantial to near perfect agreement were met in all experts ( $\kappa=0.810$ , 0.799 and 0.741)(table 1).

**Table 2** Performance levels of categorising PCV among other cases for the best model/expert in each group\*

Method/model (human)	DCNN-Combo	Human1	DCNN-O
Sensitivity	89.4	78.7	70.2
Specificity	94.8	96.9	92.7
PPV	89.4	92.5	82.5
NPV	94.8	90.3	86.4
F1 score	0.894	0.850	0.759
$\kappa$ score	0.841	0.786	0.669

\*All values present percentages except for  $\kappa$  and F1 score.

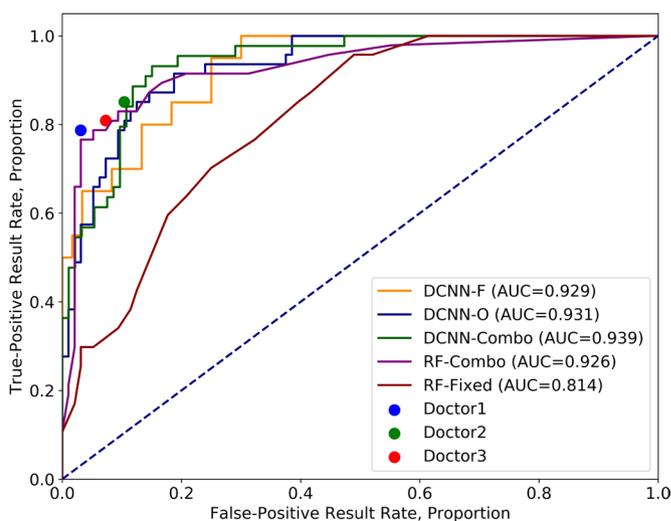
DCNN-O, deep convolutional neural network, algorithm O; DCNN-Combo, deep convolutional neural network, algorithm Combo; NPV, negative predicted value; PCV, polypoidal choroidal vasculopathy; PPV, positive predicted value.

Performance of categorising PCV in all machine models were presented as ROC curves and areas under the curve (AUC) (figure 3), operating points of three retinal specialists were marked as well. In aggregate, the performance results were comparable between our deep learning models and retinal specialists. The AUC of DCNN-Combo was 93.9%, followed by DCNN-O (AUC 93.1%) and RF-Combo (AUC 92.6%).

Table 2 showed metrics depicting the efficacy of the best model/expert in each group categorising PCV among other classifications. Our machine algorithm DCNN-Combo ( $\kappa=0.841$ ) achieved a near perfect agreement with the gold-standard. While the best human expert achieved moderate inter-agreement ( $\kappa=0.786$ ).

**DISCUSSION**

In this study, we described a multimodality deep learning machine for the multi-classifications of AMD. This model was to our knowledge the first deep learning convolutional neural



**Figure 3** Receiver operating characteristic curves for PCV detection among other cases for all algorithms with retinal specialists' performance for comparison. AUC, areas under the curve; DCNN-Combo, deep convolutional neural network, algorithm Combo; DCNN-F, deep convolutional neural network, algorithm F; DCNN-O, deep convolutional neural network, algorithm O; PCV, polypoidal choroidal vasculopathy; RF-Combo, classifier random forest, combining convolutional layers; RF-Fixed, fixed weights from training in ImageNet.

network employing paired images from colour fundus and SD-OCT devices. Early work in multimodality models focussed on other deep learning approaches based on restricted Boltzmann machine and deep belief network instead of CNN.<sup>30</sup> By applying transfer learning algorithm, we acquired updated weights of the convolutional layers efficiently from each single-modality DCNN model to construct the multimodality model. This AI system achieved the highest performance among all models and gained higher agreement with gold-standard compared with human experts. Moreover, our experiments comparing performance levels between Random Forest and Full Connection layer classifier further confirmed the superiority and necessity of every layer in DCNN framework.

Similar to strategies implemented in one previous study of AMD classification from Burlina *et al*,<sup>25</sup> we first performed the aforementioned four-group classification, then fused the other three groups except PCV group, to compare different performance levels of models categorising PCV. Furthermore, to compare the performance levels of models with human, the same test set was assigned to retinal specialists as well, similar to previous AI model studies.<sup>15 16</sup> In order to perform the human-machine comparison under the same test condition, no ICGA images of the testing data set were provided to retinal specialists either. This monitored the realistic situation where the invasive ICGA device may be unavailable, though retinal specialists' judgement would be affected with lack of evidence of ICGA. Nevertheless, results of human-machine comparison could provide a significant assessment for the potential future clinical value of our models. Only the performance level of DCNN-Combo could be comparable with retinal specialists. This proved that although OCT imaging can provide more information of retina in three-dimensional space, it still can't be used alone or take the place of Colour Fundus in diagnosing PCV.<sup>10</sup> This is to date the first multimodality AI system designed for differentiating PCV and AMD. It revealed enormous potentials in clinical translations. The high safety and non-invasiveness enable it to be widely applied in community medical institutions.

The principal limitation of our study is limited data, a common situation in single-centre research. Nevertheless, we collected over 2000 images in total, which was a relatively large quantity involving PCV, and achieved a near perfect agreement with diagnostic gold standard. Based on a previous study Zhang, *et al* published in *Cell* 2018, transfer learning based model can achieve similar performance levels comparing a large data set (n=207 130, accuracy=96.6%) with limited data set (n=4000, accuracy=93.4%).<sup>26</sup> Another limitation is the class imbalance between dry AMD group and other three groups, though inevitable due to the unnecessary to further take OCT images from clinical perspective. Nevertheless, due to the prominent image feature of dry AMD, our final test result also proved perfect performance levels of categorising dry AMD. A third limitation is different SD-OCT devices between the healthy and diseased, but all images from these devices were preprocessed to the same file format and size before training. In the future, we will focus on collecting more data from different examination devices and different retinal diseases.

This study proposed a novel bi-modality deep learning convolutional neural network framework for classification of PCV and AMD. The high accuracy of the algorithm achieved by training of more than 2000 standardised clinical images further proved that transfer learning approach and AI system could effectively learn from limited repository of medical data. Moreover, this work could accelerate the development of multimodality in AI system and more efficient screening systems for different subtypes of

AMD in community medical institutions, promoting the wide application of clinical translation and medical engineering in the realm of public health.

**Contributors** ZX, WY and YC designed the experiments. JY, ZX, WY and YC collected and labelled samples. WW, JZ, DD and XL created and trained the AI models. ZY, DC and FH performed the human-machine comparison test. ZX, WW and XL contributed to the interpretation of the results. ZX and WW wrote the manuscript. WY, XL and YC guided revision of the manuscript.

**Funding** This work was supported by The Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences grant number 2018PT32029, CAMS Initiative for Innovative Medicine grant number 2018-I2M-AI-001, Pharmaceutical collaborative innovation research project of Beijing Science and Technology Commission grant number Z191100007719002, National Natural Science Foundation of China grant number 61672523, Beijing Natural Science Foundation Haidian original innovation joint fund grant number 19L2062 and Beijing Natural Science Foundation grant number 4202033.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. The data sets generated during and/or analysed during the current study are not publicly available due to the issue of copyright but are available from the corresponding author on reasonable request.

#### ORCID iDs

Zhiyan Xu <http://orcid.org/0000-0002-3674-8198>

Feng He <http://orcid.org/0000-0001-7152-3034>

#### REFERENCES

- Coleman HR, Chan C-C, Ferris FL, *et al*. Age-Related macular degeneration. *Lancet* 2008;372:1835–45.
- Green WR. Histopathology of age-related macular degeneration. *Mol Vis* 1999;5:27.
- Joachim N, Mitchell P, Kifley A, *et al*. Incidence and progression of geographic atrophy: observations from a population-based cohort. *Ophthalmology* 2013;120:2042–50.
- Wong CW, Yanagi Y, Lee W-K, *et al*. Age-Related macular degeneration and polypoidal choroidal vasculopathy in Asians. *Prog Retin Eye Res* 2016;53:107–39.
- Kleiner RC, Brucker AJ, Johnston RL. The posterior uveal bleeding syndrome. *Retina* 1990;10:9–17.
- Wong WL, Su X, Li X, *et al*. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health* 2014;2:e106–16.
- Tan CS, Ngo WK, Chen JP, *et al*. Everest study report 2: imaging and grading protocol, and baseline characteristics of a randomised controlled trial of polypoidal choroidal vasculopathy. *Br J Ophthalmol* 2015;99:624–8.
- Owens SL. Indocyanine green angiography. *Br J Ophthalmol* 1996;80:263–6.
- Garski TR, Staller BJ, Hepner G, *et al*. Adverse reactions after administration of indocyanine green. *JAMA* 1978;240:635b–635.
- Cheung CMG, Lai TYY, Ruamviboonsuk P, *et al*. Polypoidal choroidal vasculopathy: definition, pathogenesis, diagnosis, and management. *Ophthalmology* 2018;125:708–24.
- El-Baz A, Gimel'farb G, Suzuki K. Machine learning applications in medical image analysis. *Comput Math Methods Med* 2017;2017:1–2.
- Suzuki K. Pixel-Based machine learning in medical imaging. *Int J Biomed Imaging* 2012;2012:1–18.
- Ting DSW, Cheung CY-L, Lim G, *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.
- Ting DSW, Pasquale LR, Peng L, *et al*. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167–75.
- Burlina PM, Joshi N, Pekala M, *et al*. Automated grading of age-related macular degeneration from color fundus images using deep Convolutional neural networks. *JAMA Ophthalmol* 2017;135:1170–6.
- Burlina P, Joshi N, Pacheco KD, *et al*. Utility of deep learning methods for Referability classification of age-related macular degeneration. *JAMA Ophthalmol* 2018;136:1305–7.
- Yang H-L, Kim JJ, Kim JH, *et al*. Weakly supervised lesion localization for age-related macular degeneration detection using optical coherence tomography images. *PLoS One* 2019;14:e0215076.
- Liefers B, Venhuizen FG, Schreur V, *et al*. Automatic detection of the foveal center in optical coherence tomography. *Biomed Opt Express* 2017;8:5160–78.
- Treder M, Lauerermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. *Graefes Arch Clin Exp Ophthalmol* 2018;256:259–65.

- 20 Treder M, Laueremann JL, Eter N. Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier. *Graefes Arch Clin Exp Ophthalmol* 2018;256:2053–60.
- 21 Shah A, Zhou L, Abrámov MD, et al. Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images. *Biomed Opt Express* 2018;9:4509–26.
- 22 Rashno A, Koozekanani DD, Parhi KK. Oct fluid segmentation using graph shortest path and Convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc* 2018;2018:3426–9.
- 23 Hwang D-K, Hsu C-C, Chang K-J, et al. Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics* 2019;9:232–45.
- 24 Prahs P, Radeck V, Mayer C, et al. Oct-Based deep learning algorithm for the evaluation of treatment indication with anti-vascular endothelial growth factor medications. *Graefes Arch Clin Exp Ophthalmol* 2018;256:91–8.
- 25 Burlina PM, Joshi N, Pacheco KD, et al. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. *JAMA Ophthalmol* 2018;136:1359–66.
- 26 Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018;172:1122–31.
- 27 Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology* 2018;125:1410–20.
- 28 Khalid S, Akram MU, Hassan T, et al. Fully automated robust system to detect retinal edema, central serous chorioretinopathy, and age related macular degeneration from optical coherence tomography images. *Biomed Res Int* 2017;2017:1–15.
- 29 Klein R, Meuer SM, Myers CE, et al. Harmonizing the classification of age-related macular degeneration in the three-continent AMD Consortium. *Ophthalmic Epidemiol* 2014;21:14–23.
- 30 Yoo TK, Choi JY, Seo JG, et al. The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment. *Med Biol Eng Comput* 2019;57:677–87.
- 31 Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision & Pattern Recognition*, 2016.
- 32 Ben-David A. Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Syst Appl* 2008;34:825–32.
- 33 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.