BMJ Open Ophthalmology

Classification of diabetic maculopathy based on optical coherence tomography images using a Vision Transformer model

Liwei Cai [ID],[1] Chi Wen,[2] Jingwen Jiang,[3] Congbi Liang,[1] Hongmei Zheng,[1] Yu Su,[1] Changzheng Chen [ID] [1]

For numbered affiliations see end of article.

Correspondence to
Dr Changzheng Chen;
whuchenchzh@163.com and

Dr Yu Su; sy_daisy1206@163.com

ABSTRACT

Purpose To develop a Vision Transformer model to detect different stages of diabetic maculopathy (DM) based on optical coherence tomography (OCT) images.

Methods After removing images with poor quality, a total of 3319 OCT images were extracted from the Eye Center of the Renmin Hospital of Wuhan University and randomly split the images into training and validation sets in a 7:3 ratio. All macular cross-sectional scan OCT images were collected retrospectively from the eyes of DM patients from 2016 to 2022. One of the OCT stages of DM, including early diabetic macular oedema (DME), advanced DME, severe DME and atrophic maculopathy, was labelled on the collected images, respectively. A deep learning (DL) model based on Vision Transformer was trained to detect four OCT grading of DM.

Results The model proposed in our paper can provide an impressive detection performance. We achieved an accuracy of 82.00%, an F1 score of 83.11%, an area under the receiver operating characteristic curve (AUC) of 0.96. The AUC for the detection of four OCT grading (ie, early DME, advanced DME, severe DME and atrophic maculopathy) was 0.96, 0.95, 0.87 and 0.98, respectively, with an accuracy of 90.87%, 89.96%, 94.42% and 95.13%, respectively, a precision of 88.46%, 80.31%, 89.42% and 87.74%, respectively, a sensitivity of 87.03%, 88.18%, 63.39% and 89.42%, respectively, a specificity of 93.02%, 90.72%, 98.40% and 96.66%, respectively and an F1 score of 87.74%, 84.06%, 88.18% and 88.57%, respectively.

Conclusion Our DL model based on Vision Transformer demonstrated a relatively high accuracy in the detection of OCT grading of DM, which can help with patients in a preliminary screening to identify groups with serious conditions. These patients need a further test for an accurate diagnosis, and a timely treatment to obtain a good visual prognosis. These results emphasised the potential of artificial intelligence in assisting clinicians in developing therapeutic strategies with DM in the future.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Artificial intelligence (AI) has been used to detect different macular diseases based on optical coherence tomography (OCT) images. However, this new grading system for diabetic maculopathy (DM) has not yet been used in deep learning (DL) research, which may be able to predict the treatment outcome and visual prognosis of DM better in the future.

WHAT THIS STUDY ADDS

⇒ Our DL model based on Vision Transformer demonstrated a relatively high accuracy in the detection of OCT grading of DM, which can help with patients in a preliminary screening to identify groups with serious conditions.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our model can help ophthalmologists to develop personalised treatment plans for DM patients. These results emphasise the potential of AI in reducing the necessary time of clinical diagnosis, assisting clinical decision-making and guaranteeing the cure rate in the future.

which is caused by fluid accumulation in the macula due to a breakdown of the blood–retinal barrier.[2 3] DME is the most common cause of visual impairment in people with diabetes, and its global prevalence is expected to increase from nearly 18.83 million in 2020 to 28.61 million in 2045.[4] Patients with DME can be at great risk of irreversible vision loss if not treated promptly.[2] Early diagnosis and timely treatment can effectively protect and restore the vision of DR patients.[5]

Previously, based on the location of retinal thickening and hard exudates, DME was classified into involved central and non-involved central types.[6] According to patterns of DME on optical coherence tomography (OCT) examination, DME can be divided into diffused retinal thickening (DRT), cystoid macular oedema (CME) and serous retinal

INTRODUCTION

Diabetic retinopathy (DR) is one of the most common complications of diabetes.[1] At any time during the progression of DR, patients may develop diabetic macular oedema (DME),

detachment (SRD).[7] All these classifications only focus on the location and its relationship with the fovea of macular thickening, or the overall patterns on OCT, lacking the assessment of macular atrophy and failing to consider the alteration in subtle structure of different manifestations of DME, which cannot meet the needs of treatment.

In recent years, with the widespread application of spectral domain OCT (SD-OCT), more biomarkers have been identified in DME, which are of great significance for the treatment and prognosis.[8–10] The location and size of intraretinal cysts (IRC) correlated with visual acuity at baseline in DME and the cysts of the inner nuclear layer were more sensitive to corticosteroid or anti-VEGF treatment than the outer nuclear layer.[11 12] Visual acuity was closely related to central photoreceptor damage and the percentage of ellipsoid zone (EZ) destruction, and whether the photoreceptor status was restored determined the final visual acuity.[13 14] The greater the range of disorganisation of the inner retinal layers (DRIL) in DME eyes at baseline, the worse the prognosis of vision.[15] In 2019, an international panel of experts attempted to combine more OCT-related morphological features with central subfoveal thickness (CST) to elaborate a new grading system for diabetic maculopathy (DM) applicable to clinical and scientific research.[16] DM includes all the phenotypes of macular involvement in DR irrespective of the presence of macular thickening. According to foveal thickness, the size of the IRC, the EZ and/or external limiting membrane (ELM) status, and the presence of DRIL, DM is classified as early DME, advanced DME, severe DME and atrophic maculopathy. Therefore, this classification may be able to predict the treatment outcome and visual prognosis of DM better in the future.

Patients with DM often need to take regular OCT examinations to record the occurrence and development of the disease. The increasing number of patients with DM makes it a significant burden for clinicians to manually determine the presence or progression of DM on OCT images.[4] Artificial intelligence (AI) that can help with screening may reduce the burden on ophthalmologists. Intelligent systems have been developed for diagnosing and classifying DME based on OCT images.[17 18] In these studies, AI was only used to detect different macular diseases and the overall morphology of DME, and this new grading system has not yet been used in research. In this study, we aimed to build a deep learning (DL)-based training AI system to automatically classify DM images based on the novel classification standard, in order to help ophthalmologists develop personalised treatment plans for patients with DM.

MATERIALS AND METHODS
Image dataset
In this study, all completely anonymous retinal OCT images were selected from retrospective cohorts of adult patients from the Eye Center of the Renmin Hospital of Wuhan University between 2016 and 2022. 4076 OCT images of DM patients centred at the fovea were extracted from an OCT device (Optovue RTVue, Optovue, Fremont, California, USA) for training and validating the DL model. After randomly splitting the dataset into training and validation sets in a 7:3 ratio, the training data and the validation data were completely independent of each other. In the model's training and testing phases, we made no distinction between the patient's left and right eyes. It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Grading of DM
Seven qualitative and quantitative characteristics of DM are considered and scored according to the grading system called TCED-HFV, including foveal thickness (T), intraretinal cyst (C), EZ and/or ELM status (E), presence of DRIL (D), number of hyper-reflective foci (H), subfoveal fluid (F), and vitreoretinal relationship (V).[16] Based on the first four variables, namely T, C, E and D, disease can be classified into four distinct stages, that is, early DME, advanced DME, severe DME and atrophic maculopathy (figure 1). Different stages reflect the severity of the disease.

Image labelling
Before training, all OCT images were graded into four stages by trained graders with increasing expertise for verification and correction of image labels. A trained grader (LC) excluded images with low image quality. These images were taken of improper positioning during image acquisition or scans with strong motion artefacts, causing misalignment and blurring of sections. Then, two retinal specialists (YS and HZ) independently labelled each image, and images with a clear consensus annotation between ophthalmologists were taken into the sample. Images with different grading opinions were adjudicated by a senior retinal expert (CC) with more than 20 years of experience and the final labels were also imported into the database.

Images preprocessing
The preprocessing part was used to enhance the effective area of OCT images, suppress background noise, increase the number of training samples and improve the generalisation ability and robustness of the model. We chose our segmentation method to simply extract the effective region of the OCT images and perform pixel-level enhancement. On the one hand, this was to reduce the noise interference. On the other hand, we wanted to make the model focus more on the effective region of the images.

The background area occupied most of the OCT image and there was a certain amount of noise overall, which may affect model training. Therefore, we hoped to suppress image noise to make it easier for the network to focus on the effective area of the OCT image and converge faster.
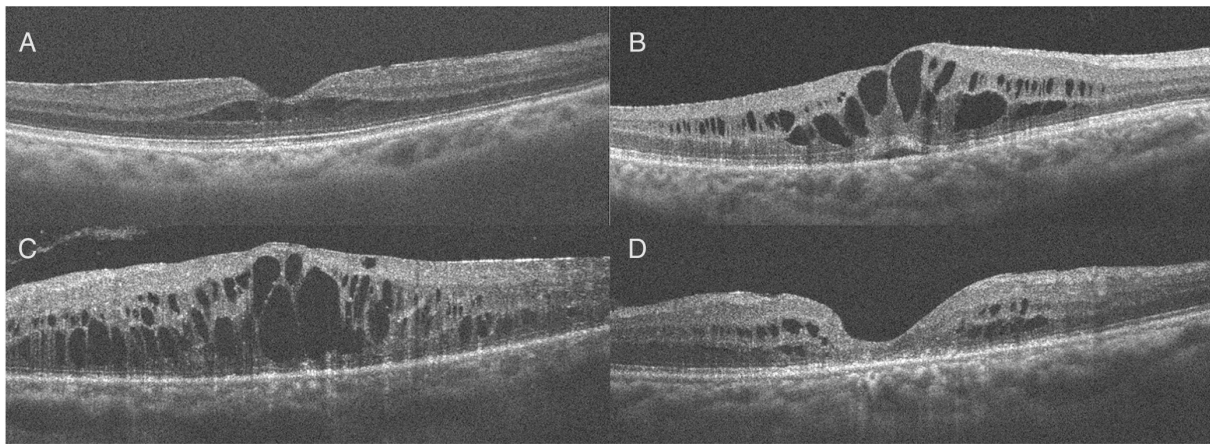
Figure 1 Representative optical coherence tomography images. (A) Early diabetic macular oedema (DME). (B) Advanced DME. (C) Severe DME. (D) Atrophic maculopathy.

To this end, we proposed to use the Otsu method to binarise the image, and fuse the resulting binary image with the original image at a certain ratio. This enhanced the effective area of the image and suppressed background noise.

More specifically, first, according to the interclass variance of the histogram of the OCT image, a binary segmentation was performed to obtain a binary image, denoted as P. The original image was denoted as T. The dot product operation was performed on P and T and add up P and T to get the final result. The formula was as follows:

$$I_{prepro} = (T \cdot P) \cdot \alpha + T \cdot (1 - \alpha)$$

The value of α is between 0 and 1, set manually.

Then, conventional data augmentation methods were used to enhance the preprocessed images. Specific measures will be introduced in the model validation section.

Model training

In the training part, guided by the idea of normalisation, we designed a classifier and loss function to alleviate the problem of network overconfidence and improve its robustness.

Considering that OCT images were prone to noise, we chose the Vision Transformer as the backbone to extract features from OCT images, which was more robust to noise compared with convolutional neural network (CNN), as the Vision Transformer can better mine global information through its self-attention mechanism and had less bias towards local texture features. The features extracted by the backbone were classified through our self-designed classifier. As there was a long-tailed problem in the dataset, we redesigned the classifier and loss function using some normalisation techs. The specific implementation method was as follows:

1. Based on an unbiased linear layer, we calculated the logits using the weights of the classifier and the input feature vectors with L2-Norm.
2. We combined the CrossEntropy Loss and logits with L2-Norm, which will keep the magnitude of logits a constant during training, to create a new loss function.

Concretely, the formula of the logits calculated by classifier that we proposed was as follows:

$$g = K \frac{w^T x}{\|w\| \|x\|}$$

where g is denoted as the logits output of the backbone, w is the weight parameters of the classifier and x stands for the feature input of the classsifier, K is a hyperparameter.

And the new loss function can be defined as:

$$L = -\sum_{i=1}^{n} y_i \log \frac{e^{\frac{g}{\|g\|}}}{\sum_{j=1}^{k} e^{\frac{g_j}{\|g\|}}}$$

Where the temperature parameter τ controls the magnitude of the logits and y_i represents the label.

Model inference

In practice, the distribution of OCT images often varied between training and test data, potentially leading to a decrease in the model's performance. To address this issue, we have introduced an adaptive mechanism that allowed the model to dynamically update its parameters based on test data, thereby enhancing both its performance and generalisation ability. For each test sample, we performed multiple random data augmentation operations, such as rotation, cropping and flipping, to obtain diversified test samples and improve the robustness and generalisation ability of the model.

For each test sample x, a series of random enhancement operations were performed m times to obtain a sample set $X=\{x_1,x_2,…,x_m\}$. Take this set as a batch

and input into the model to obtain the output distribution $p(y|x)$. Here, every single $x \in X$ must predict a label $y \in Y$, but it was important to note that y was not the ground truth. We hoped that the model could maintain relatively stable in the predictions of the same sample under numerous data augmentation operations, which meant an improvement of the model's robustness. To achieve this goal, we took the entropy of the average output distribution of the model as the optimised goal. By minimising the entropy value, the parameters of the model are updated. Then we input the original sample x into the model and got the final result. Specifically, the formula for the optimisation objective was as follows:

$$H = -\sum_{y \in Y} \hat{p}(y|x) \log \hat{p}(y|x)$$

Note that for every test sample x, the model's parameters that update in test time will not be saved.

Experiment

In order to verify the generalisation of our proposed schemes, we conducted experiments on the dataset constructed by ourselves.

First, the preprocessing method we proposed was applied to the OCT images with the $alpha$ value set to 0.2 to obtain enhanced OCT images. Then, some conventional data augmentation methods were applied to these OCT images after they were resized to 224×224. Specifically, includes 'RandomResizedCrop', 'RandomHorizontalFlip', 'RandomVerticalFlip', 'GaussianBlur' and 'Normalise'. The model architecture we proposed using vision transformer will use these images and Adam optimizer to train. The loss function was the one we proposed above, with the temperature parameter set to 1.0. The learning rate was set to 0.001, weight decay to 0.0005, batch size to 128 and the K value of the classifier was set to 8.

Stochastic gradient descent (SGD) was a very common optimiser for model training in machine learning, which meant that it updated the model parameters by using a random subset of the data at each iteration. When it came to test time, we used SGD optimiser with 0.001 learning rate, perform 32 data augmentation operations on each sample, specifically including: 'random rotation', 'histogram equalisation', 'invert pixels', 'colour quantisation', 'shear image along x-axis or y-axis', 'translate image along x-axis or y-axis'.

Model validation

Some metrics have been used to show model performance. The correlation between the true labels and the predicted labels from our model was depicted as a confusion matrix, which was used to calculate the accuracy, precision, sensitivity, specificity and F1 score for image recognition. We also used the area under (AUC) the receiver operating characteristic (ROC) curve to evaluate the accuracy of the model in detecting the four stages. All the metrics mentioned above were calculated using TorchMetrics. We used TorchMetrics to generate the ROC curve for each stage of DM. This was done by taking the model's predictions for all the validation data and the corresponding ground truth labels as input. This approach was also used for other metrics mentioned in the paper. For the generation of the ROC curve of the multiclass classifiers, TorchMetrics employed the One-vs-Rest (OvR) strategy. This strategy treated each class as the positive class and all other classes as the negative class, calculating the ROC curve for each class separately. The cut-off values were determined dynamically by TorchMetrics based on all the validation data, varying for each category, eliminating the need for manual setting.

RESULTS

A total number of 4076 OCT images were collected. After removing images with severe artefacts causing misalignment and blurring of sections or significant image resolution reductions, 3319 OCT images were used in this study, which were 1254 images of early DME, 991 images of advanced DME, 672 images of severe DME and 402 images of atrophic maculopathy. Among these, 70% of images were randomly selected as training dataset to establish our model, 30% of images were selected as validation dataset. The number of images for each stages in training and validation set is shown in table 1. The representative enhanced OCT images are shown in online supplemental file 1.

On the validation dataset, we constructed, we achieved an accuracy of 82.00%, an F1 score of 83.11%, an AUC of 0.96. The confusion matrix of the classification results is shown in figure 2. And accuracy, precision, sensitivity and specificity for every stage are shown in table 2. AUC and F1 score for every stage are shown in online supplemental file 1. The AUC for the detection of early DME was 0.96, with an accuracy of 90.87%, a precision of

Table 1 The distribution of images

DM stages	No of total images	No of training images	No of validation images
Early DME	1254	884	370
Advanced DME	991	695	296
Severe DME	402	290	112
Atrophic maculopathy	672	464	208

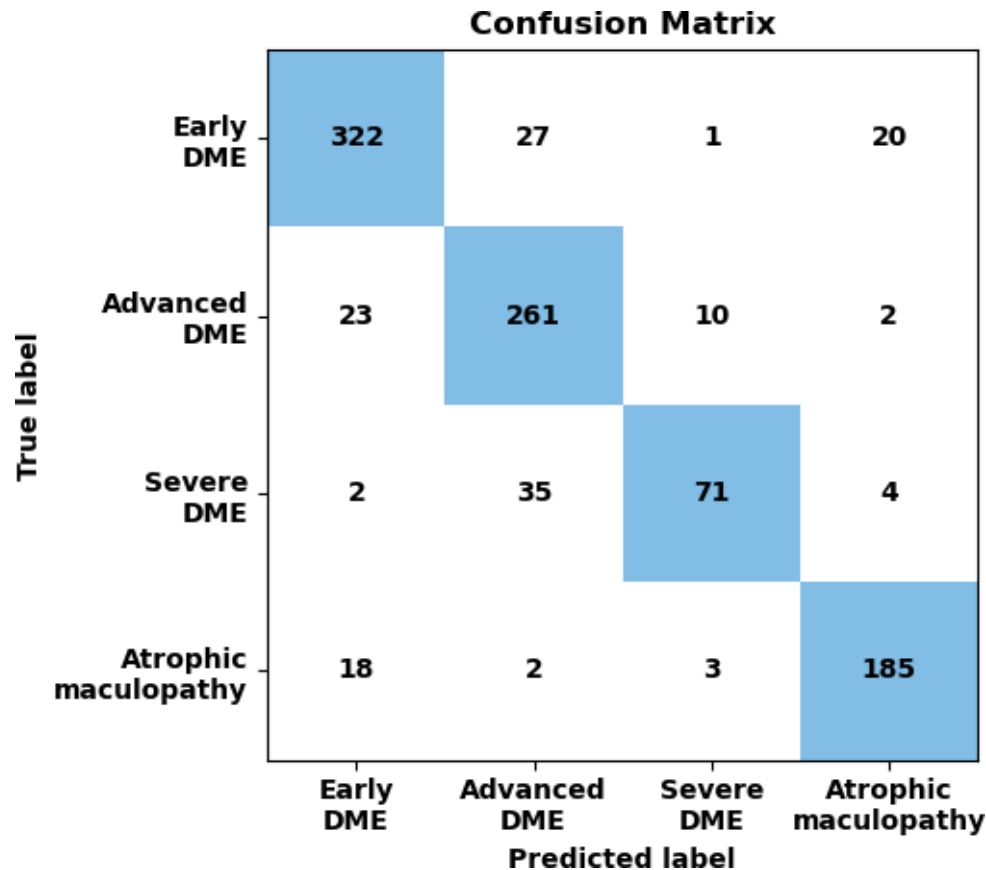DM, diabetic maculopathy; DME, diabetic macular oedema

Figure 2 Confusion matrixes of the model. DM, diabetic maculopathy; DME, diabetic macular oedema.

88.46%, a sensitivity of 87.03%, a specificity of 93.02% and an F1 score of 87.74%. The AUC for the detection of advanced DME was 0.95, with an accuracy of 89.96%, a precision of 80.31%, a sensitivity of 88.18%, a specificity of 90.72% and an F1 score of 84.06%. The AUC for the detection of severe DME was 0.87, with an accuracy of 94.42%, a precision of 89.42%, a sensitivity of 63.39%, a specificity of 98.40% and an F1 score of 88.18%. The AUC for the detection of atrophic maculopathy was 0.98, with an accuracy of 95.13%, a precision of 87.74%, a sensitivity of 89.42%, a specificity of 96.66% and an F1 score of 88.57%. The ROC curve of the classification results is shown in figure 3. The results showed that the

method we proposed can effectively classify patients with DM into different stages.

DISCUSSION

In this study, we developed a DL model based on vision transformer for DM grading based on OCT-related morphological features. We achieved an accuracy of 82.00%, an F1 score of 83.11% and an AUC of 0.96. Our research showed that the accuracy of our model in this novel grading system was promising, which can help with patients in a preliminary screening to identify groups with serious conditions. As this classification may be able to predict the treatment outcome and visual prognosis of DM better in the future, our model can help ophthalmologists to develop personalised treatment plans for patients with DM.

DM at four different stages reflects the severity of the disease. Early DME usually corresponds to a short duration of hyperglycaemic state.[16] So most of the time patients can maintain a good vision if they can take good control of their blood glucose. EZ/ELM state is different between advanced and severe DME. In the former one, EZ/ELM may be damaged but still visible, and the layers of the inner retina are usually recognisable. In the latter one, the internal retinal layers and/or EZ/ELM are mostly destroyed and undetectable. These two groups of patients may have distinct differences in treatment

Table 2 Accuracy, precision, sensitivity and specificity of our model in validation dataset

DM stages	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)
Early DME	90.87	88.46	87.03	93.02
Advanced DME	89.96	80.31	88.18	90.72
Severe DME	94.42	89.42	63.39	98.40
Atrophic maculopathy	95.13	87.74	89.42	96.66

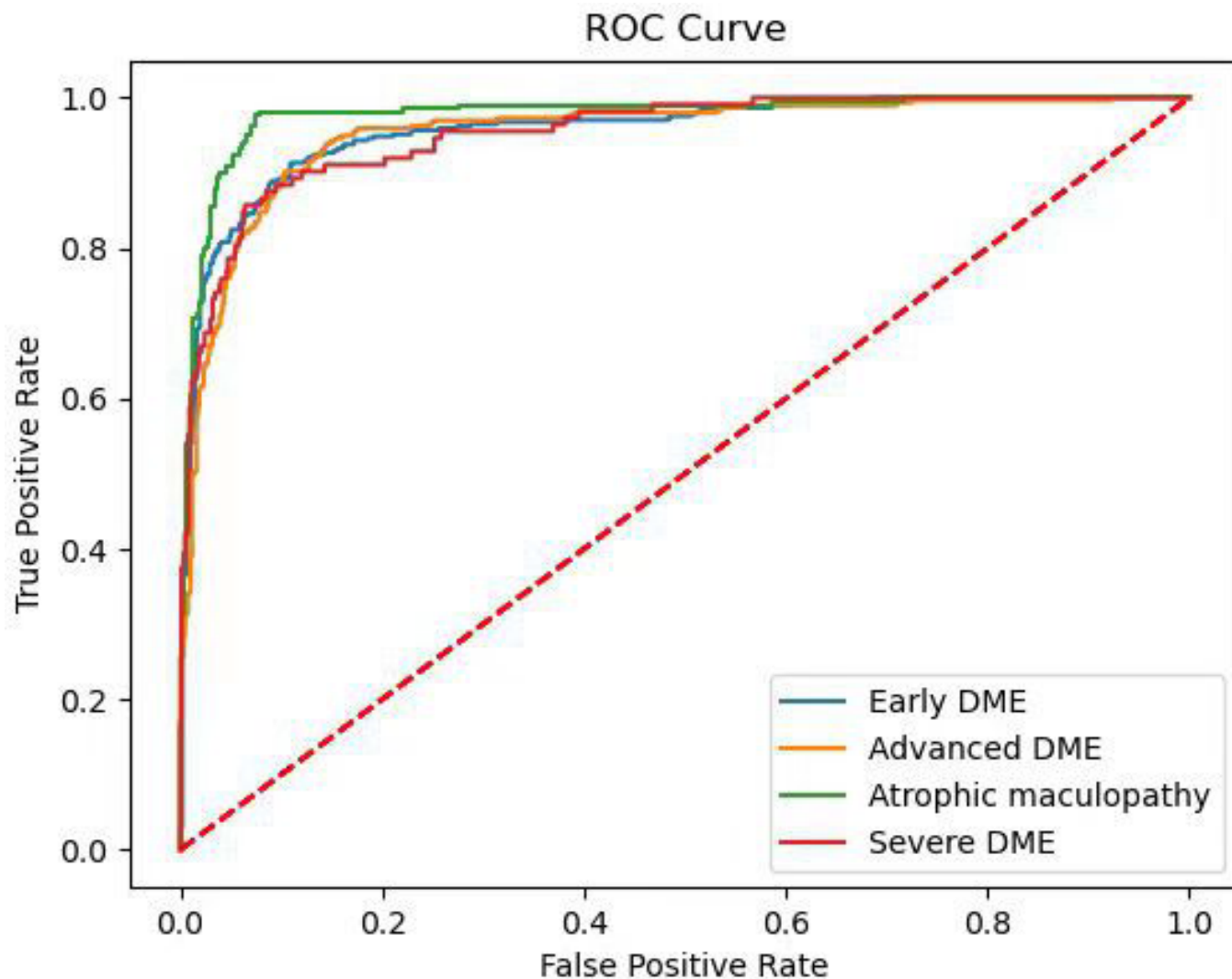DM, diabetic maculopathy; DME, diabetic macular oedema

Figure 3 ROC curve analysis results of our model. DME, diabetic macular oedema; ROC, receiver operating characteristic.

response and visual prognosis and should be distinguished.[16] Patients with advanced DME should be treated promptly. Anti-VEGF treatment may prevent progression of the disease into next stage with ELM and/or EZ being recovered and CST decreasing to normal values. While once the disease progresses into severe DME, it may be difficult in resolution of oedema despite positive treatment, and finally may inevitably develop into atrophy stage. Macular atrophy is characterised by complete EZ/ELM destruction and DRIL, usually as a result of long-term macular oedema, and has a poor visual outcome.[16]

Hence, this novel grading system can assist the ophthalmologists in predicting the prognosis of patients with DM in their clinical work, and personalised therapeutic strategies could be made according to the OCT grading. Especially in the former two stages, taking good control of blood glucose and timely treatment are significant to promote recovery and prevent them from progressing into the more severe stages. For these patients, early screening and long-term follow-up can maintain a better vision outcome. However, detection and grading of DM

currently required expertise and are time-consuming. Thus, it is particularly beneficial and promising to develop an intelligent system for the DM grading based on this new system to assist the clinical decision-making processes in patients.

With the continuous development of DL technology, now we all have more opportunities to achieve automatic diagnosis and classification of diseases. Numerous studies have demonstrated the expert performance of DL technology in detecting DME. For instance, Alqudah[19] proposed a multiclassification model based on SD-OCT for four types of retinal diseases (age-related macular degeneration, choroidal neovascularisation, DME and drusen) as well as normal cases. The proposed CNN architecture with softmax classifier correctly identified 99.17% of DME cases overall. Zhang et al[20] proposed a multiscale DL model, which were divided into two parts: self-enhancement model and disease detection model, with achieving 94.5% accuracy in identifying DME. Meanwhile, they proved that this model provided a better ability to recognise low-quality medical images. Wu et al[21]

trained a DL model using Visual Geometry Group 16 (VGG16) network as the backbone to detect three OCT morphologies of DME, including DRT, CME and SRD. The accuracy was 93.0%, 95.1% and 98.8%, respectively. All the above studies indicated that DL model had good feasibility and application prospects in diagnosing DME. However, there is still a lack of DL model for automatic detection for this OCT-based grading of DM. Meanwhile, it should be noted that most of the above studies were based on CNN. The advantage of CNN is that it can extract image features well, which has been verified by a large number of scholars. However, there is still little research on Visual Transformer, which has better classification capabilities than CNN to solve image classification problems.[22]

In current study, we trained a DL model using Vision Transformer as the backbone to detect this novel grading in OCT images. Vision Transformer proposed in 2020 is a new image classification model, which is considered to be the best image classification model at present, showing better performance than traditional CNN model.[22] Vision Transformer is not dependent on any CNN and is completely based on transformer structure designed with different feature extraction methods from CNN.[23] Research has proved the recognition ability of Vision Transformer for OCT images is stronger than CNN models and traditional machine learning algorithms.[23] In the accuracy comparison of the same test set between Vision Transformer and four CNN models: VGG16, Resnet50, Densenet121 and EfficentNet, Vision Transformer has the highest classification accuracy of 99.69%. Meanwhile, both VGG16 and Vision Transformer are faster than other CNN models in the recognition speed of a single image.[23] Although our result was slightly less impressive than the previous studies using other DL architectures to detect DME based on OCT images and the detection of the OCT patterns. It can be more complicated and challenging than distinguishing DME from other retinal diseases or simply detecting the overall patterns of DME, with less obvious differences in characteristics and subtle lesions between different OCT grading.

To our knowledge, this is the first article to detect the severity of DM according to the novel classification standard based on OCT images by DL and the first article to use Vision Transformer to detect DM. As mentioned above, this classification may be able to predict the treatment outcome and visual prognosis of DM better in the future and help ophthalmologists develop precise treatment plans for patients. And as the Vision Transformer can better mine global information through its self-attention mechanism and has less bias towards local texture features, it is more robust to noise compared with CNN commonly used in past studies. So our model combined with these advantages is very promising for detecting OCT images of DM or other retinal diseases. Our model had a slightly lower performance in predicting severe DME. Possibly because there were fewer images compared with other stages, and patients almost always had poor vision after progressing into this stage, resulting in worsen image quality. However, our model can help with patients in a preliminary screening to identify groups with serious conditions. These patients need a further test for an accurate diagnosis, and a timely treatment to prevent further deterioration in time. Overall, the result achieved by our DL model was promising and encouraging.

Although our model showed great potential, there are still several limitations in the study. First, OCT images only obtained from the Optovue RTVue imaging system in our study. The model needs to be further validated by images from different OCT equipment. Second, We only perform the classification training in this model. In the future, studies can train models to predict treatment outcomes based on this new grading system. Finally, type of data we used only included images from one eye centre. More OCT images from other multicentre trials in the future can be used to improve our model.

In conclusion, our DL model based on Vision Transformer demonstrated a relatively high accuracy in the detection of the different OCT-based stages of DM. This DM grading model can reduce the burden on clinical ophthalmologists and provide a reference in making personalised therapeutic strategies. These results emphasise the potential of AI in reducing the necessary time of clinical diagnosis, assisting clinical decision-making and guaranteeing the cure rate in the future.

Author affiliations
[1]Department of Ophthalmology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China
[2]Wuhan University School of Computer Science, Wuhan, Hubei, China
[3]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou, Guangdong, China

includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Liwei Cai http://orcid.org/0000-0003-1258-232X
Changzheng Chen http://orcid.org/0000-0002-7281-552X

## REFERENCES

1 Wilkinson CP, Ferris FL, Klein RE, *et al*. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–82.
2 Ciulla TA, Amador AG, Zinman B. Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies. *Diabetes Care* 2003;26:2653–64.
3 Chiu SJ, Allingham MJ, Mettu PS, *et al*. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed Opt Express* 2015;6:1172–94.
4 Teo ZL, Tham Y-C, Yu M, *et al*. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology* 2021;128:1580–91.
5 Hariprasad SM, Mieler WF, Grassi M, *et al*. Vision-related quality of life in patients with diabetic macular oedema. *Br J Ophthalmol* 2008;92:89–92.
6 Scott IU, Danis RP, Bressler SB, *et al*. Effect of focal/grid photocoagulation on visual acuity and retinal thickening in eyes with non-center-involved diabetic macular edema. *Retina* 2009;29:613–7.
7 Otani T, Kishi S, Maruyama Y. Patterns of diabetic macular edema with optical coherence tomography. *Am J Ophthalmol* 1999;127:688–93.
8 Reznicek L, Cserhati S, Seidensticker F, *et al*. Functional and morphological changes in diabetic macular edema over the course of anti-vascular endothelial growth factor treatment. *Acta Ophthalmol* 2013;91:e529–36.
9 De S, Saxena S, Kaur A, *et al*. Sequential restoration of external limiting membrane and ellipsoid zone after intravitreal anti-VEGF therapy in diabetic macular oedema. *Eye (Lond)* 2021;35:1490–5.
10 Sun JK, Lin MM, Lammer J, *et al*. Disorganization of the retinal inner layers as a predictor of visual acuity in eyes with center-involved diabetic macular edema. *JAMA Ophthalmol* 2014;132:1309–16.
11 Deák GG, Bolz M, Ritter M, *et al*. A systematic correlation between morphology and functional alterations in diabetic macular edema. *Invest Ophthalmol Vis Sci* 2010;51:6710–4.
12 Karst SG, Lammer J, Mitsch C, *et al*. Detailed analysis of retinal morphology in patients with diabetic macular edema (DME) randomized to ranibizumab or triamcinolone treatment. *Graefes Arch Clin Exp Ophthalmol* 2018;256:49–58.
13 Maheshwary AS, Oster SF, Yuson RMS, *et al*. The association between percent disruption of the photoreceptor inner segment-outer segment junction and visual acuity in diabetic macular edema. *Am J Ophthalmol* 2010;150:63–67.
14 Yanyali A, Bozkurt KT, Macin A, *et al*. Quantitative assessment of photoreceptor layer in eyes with resolved edema after pars plana vitrectomy with internal limiting membrane removal for diabetic macular edema. *Ophthalmologica* 2011;226:57–63.
15 Moein H-R, Novais EA, Rebhun CB, *et al*. Optical coherence tomography angiography to detect macular capillary ischemia in patients with inner retinal changes after resolved diabetic macular edema. *Retina* 2018;38:2277–84.
16 Panozzo G, Cicinelli MV, Augustin AJ, *et al*. An optical coherence tomography-based grading of diabetic maculopathy proposed by an international expert panel: the European school for advanced studies in ophthalmology classification. *Eur J Ophthalmol* 2020;30:8–18.
17 Lu W, Tong Y, Yu Y, *et al*. Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. *Transl Vis Sci Technol* 2018;7:41:41.:.
18 Tsuji T, Hirose Y, Fujimori K, *et al*. Classification of optical coherence tomography images using a capsule network. *BMC Ophthalmol* 2020;20:114.
19 Alqudah AM. AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Med Biol Eng Comput* 2020;58:41–53.
20 Zhang Q, Liu Z, Li J, *et al*. Identifying diabetic macular edema and other retinal diseases by optical coherence tomography image and multiscale deep learning. *Diabetes Metab Syndr Obes* 2020;13:4787–800.
21 Wu Q, Zhang B, Hu Y, *et al*. Detection of morphologic patterns of diabetic macular edema using a deep learning approach based on optical coherence tomography images. *Retina* 2021;41:1110–7.
22 Bazi Y, Bashmal L, Rahhal MMA, *et al*. Vision transformers for remote sensing image classification. *Remote Sensing* 2021;13:516.
23 Jiang Y, Wang L, Wu Q, *et al*. Computer-aided diagnosis of retinopathy based on vision transformer. *J Innov Opt Health Sci* 2022;15.

Supplemental figure
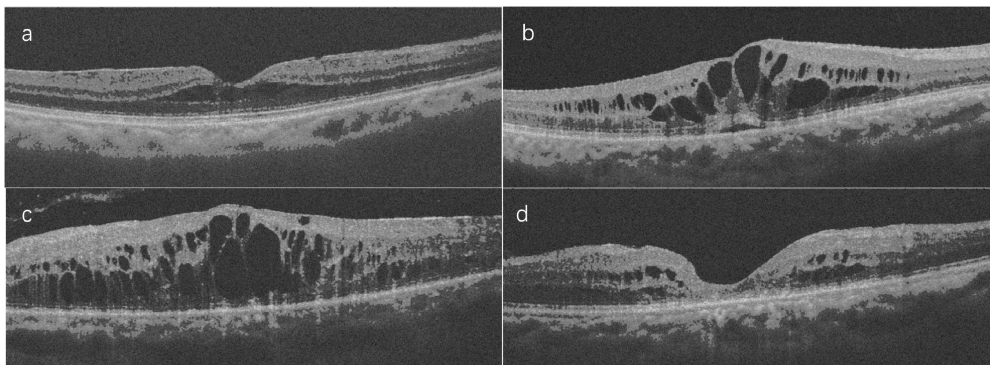
Representative enhanced optical coherence tomography images: a. early diabetic macular edema (DME). b. advanced DME. c. severe DME. d.atrophic maculopathy.

Supplemental table AUC and F1 score of our model in validation dataset		
DM stages	AUC	F1 score (%)
Early DME	0.9607	87.74
Advanced DME	0.9542	84.06
Severe DME	0.8703	88.18
Atrophic maculopathy	0.9770	88.57

DM: diabetic maculopathy; DME: diabetic macular edema; AUC: area under the receiver operating characteristic curve