



OPEN ACCESS

EDITED BY

Xiaogang Wang,
Shanxi Eye Hospital, China

REVIEWED BY

Fei Shi,
Soochow University, China
Gui-shuang Ying,
University of Pennsylvania, United States

*CORRESPONDENCE

Weiwei Dai
✉ daiweiwei@aierchina.com

†These authors have contributed equally to this work

RECEIVED 12 May 2023

ACCEPTED 03 July 2023

PUBLISHED 19 July 2023

CITATION

Hu M, Wu B, Lu D, Xie J, Chen Y, Yang Z and Dai W (2023) Two-step hierarchical neural network for classification of dry age-related macular degeneration using optical coherence tomography images. *Front. Med.* 10:1221453. doi: 10.3389/fmed.2023.1221453

COPYRIGHT

© 2023 Hu, Wu, Lu, Xie, Chen, Yang and Dai. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Two-step hierarchical neural network for classification of dry age-related macular degeneration using optical coherence tomography images

Min Hu^{1†}, Bin Wu^{2†}, Di Lu³, Jing Xie¹, Yiqiang Chen⁴, Zhikuan Yang⁵ and Weiwei Dai^{1,6*}

¹Changsha Aier Eye Hospital, Changsha, China, ²Department of Retina, Shenyang Aier Excellence Eye Hospital, Shenyang, China, ³Department of Retina, Shenyang Aier Optometry Hospital, Shenyang, China, ⁴Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, ⁵Aier Institute of Optometry and Vision Science, Changsha, China, ⁶Anhui Aier Eye Hospital, Anhui Medical University, Hefei, China

Purpose: The aim of this study is to apply deep learning techniques for the development and validation of a system that categorizes various phases of dry age-related macular degeneration (AMD), including nascent geographic atrophy (nGA), through the analysis of optical coherence tomography (OCT) images.

Methods: A total of 3,401 OCT macular images obtained from 338 patients admitted to Shenyang Aier Eye Hospital in 2019–2021 were collected for the development of the classification model. We adopted a convolutional neural network (CNN) model and introduced hierarchical structure along with image enhancement techniques to train a two-step CNN model to detect and classify normal and three phases of dry AMD: atrophy-associated drusen regression, nGA, and geographic atrophy (GA). Five-fold cross-validation was used to evaluate the performance of the multi-label classification model.

Results: Experimental results obtained from five-fold cross-validation with different dry AMD classification models show that the proposed two-step hierarchical model with image enhancement achieves the best classification performance, with a f1-score of 91.32% and a kappa coefficients of 96.09% compared to the state-of-the-art models. The results obtained from the ablation study demonstrate that the proposed method not only improves accuracy across all categories in comparison to a traditional flat CNN model, but also substantially enhances the classification performance of nGA, with an improvement from 66.79 to 81.65%.

Conclusion: This study introduces a novel two-step hierarchical deep learning approach in categorizing dry AMD progression phases, and demonstrates its efficacy. The high classification performance suggests its potential for guiding individualized treatment plans for patients with macular degeneration.

KEYWORDS

optical coherence tomography (OCT), age-related macular degeneration (AMD), nascent geographic atrophy (nGA), convolutional neural network (CNN), deep learning

1. Introduction

Age-related macular degeneration (AMD) (1) is an ocular disease that manifests with a degenerative change in the retina and choroid of the macular region. According to the World Health Organization, ~1.3 billion people globally suffer from varying degrees of vision loss, with AMD as the third leading cause of vision loss among patients (2, 3). AMD predominantly affects individuals over the age of 50, and its prevalence increases with advancing age (4). As the global population continues to age, the number of people affected by AMD is expected to rise, further highlighting the importance of understanding and addressing this debilitating ocular disease (5).

Two distinct categories of AMD are typically distinguished considering both clinical and pathological features: the atrophic or dry form, and the exudative or wet form. The exudative form is defined by the presence of abnormal retinal changes caused by the growth of newly formed vessels within the macula. The dry form, on the other hand, is characterized by a progressive course that culminates in degeneration of the retinal pigment epithelium (RPE), thickening of the Bruch membrane, and photoreceptor loss (6). In its late stages, dry AMD manifests with localized RPE degeneration and photoreceptor loss, known as geographic atrophy (GA), which ultimately leads to progressive and irreversible loss of visual function.

Early detection and intervention of GA can facilitate the timely intervention from clinicians and better management of the disease, leading to improved patient outcomes. The recent approval by the U.S. Food and Drug Administration of a new treatment for GA

accentuates the importance of early detection of this condition (7, 8). The identification of GA is also valuable in understanding the natural progression of AMD and characterizing the disease spectrum. However, GA is typically small in its nascent stage and can be difficult to identify accurately, even for experienced clinicians. This difficulty is exacerbated by the fact that the imaging characteristics of GA are heterogeneous in shape, size, and location, making it challenging to detect and differentiate from other kind of retinal lesions such as drusen. By accurately detecting the earliest signs of GA, clinicians can provide timely and effective interventions, leading to better management of the condition and improved patient outcomes.

A recent study (9) demonstrated that *nascent GA* (nGA) is a strong predictor for the development of GA, providing supportive evidence of the potential value of nGA as a surrogate endpoint in future intervention trials for the early stages of AMD. In OCT images, the presence of nGA is defined as the presence of the subsidence of the inner nuclear layer (INL) and outer plexiform layer (OPL), and/or the presence of a hyporeflective wedge-shaped band (Figure 1).

OCT has emerged as a critical imaging technique for the diagnosis and classification of atrophy, with a consensus established based on the assessment of OCT images (10). The Classification of Atrophy Meetings (CAM) group has promulgated guidelines for the employment of OCT in the diagnosis of dry AMD. According to CAM's consensus, nGA was suggested to be retained as the term to describe incomplete retinal pigment epithelium and outer retinal atrophy (iRORA) in the absence of choroidal neovascularization (CNV) (10). To ensure accurate diagnosis and

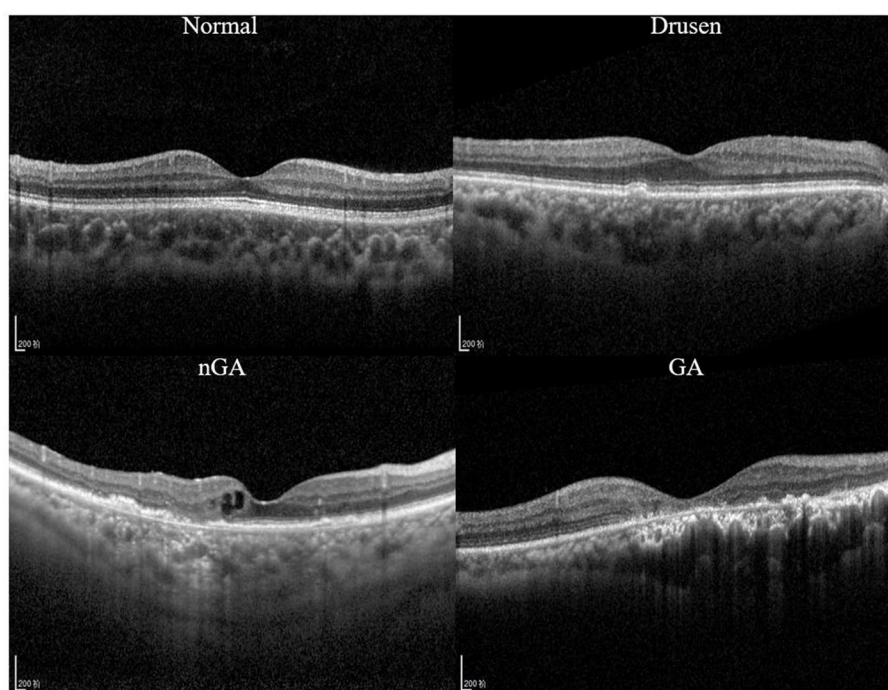


FIGURE 1
Optical coherence tomography (OCT) images of a healthy retina, drusen, nascent geographic atrophy (nGA) and geographic atrophy (GA).

treatment, reproducibility is crucial in identifying iRORA and nGA with OCT.

Advances in OCT technology and the integration of artificial intelligence algorithms will likely improve the accuracy and reliability of iRORA identification as clinicians and reading centers become more familiar with OCT findings (11). Recently numerous studies have reported the application of deep convolutional neural networks (CNNs) in the diagnosis of AMD (12–15). In these studies, promising results have been demonstrated in detecting AMD with medical images such as color fundus photography and OCT. With the aid of deep learning algorithms and medical image analysis, identifying the early signs of GA (i.e., nGA) has become increasingly feasible.

However, no previous studies have explored the feasibility of detecting the early stages of dry AMD. One of the challenges is that currently available datasets for AMD classification tend to focus on the more advanced phases of the disease, and do not contain sufficient examples of early stages. The variability in image quality, diagnosis, and definition of AMD phases across different datasets further complicates the task of detecting early stages of the disease using CNNs.

In seeking to improve the early detection of dry AMD, including the characteristics such as nGA, we developed and validated a novel two-step hierarchical CNN model along with image enhancement for dry AMD classification. The major contributions of this research can be outlined as follows:

- To the best of our knowledge, this is the first investigation to use CNNs to classify the early stages of dry AMD, including nGA.
- To leverage the domain expertise regarding the characteristics of OCT images, including those of nGA, we combine image enhancement and hierarchical classification techniques, effectively highlighting features associated with nGA while preserving the overall integrity of the OCT images.
- The proposed method demonstrates promising classification results and can be utilized as a useful computer-aided diagnostic tool for clinical OCT-based AMD diagnosis.

2. Materials and methods

2.1. Datasets and labeling

OCT images in this study were collected from dry AMD patients admitted to the Shenyang Aier Excellence Eye Hospital (Shenyang, China) in 2019–2021. The device used to obtain the images was a Heidelberg Spectralis HRA + OCT (Heidelberg, Germany), and the scan length was 6 mm × 6 mm. A total of 3,401 qualified OCT images from 338 patients were selected for model development. Each image was classified by a retinal specialist with over 15 years of clinical expertise. Example OCT images of both healthy retinas and dry AMD classes are presented in Figure 1. These images were divided into a training dataset (~80% of the patients) for model development and a validation dataset (~20% of the patients) for validating the models based on the patient's identification number. The partitioning of the dataset into approximate proportions of 80 and 20% is contingent upon the

distribution of patients rather than images. Given that each patient may present an unequal number of OCT images acquired during a single visit, the precise ratios of the division may exhibit minor discrepancies across the five-fold cross validation dataset.

The dataset, as detailed in Table 1, contains 712 images of normal, 1,167 images of drusen, 711 images of nGA, and 811 images of GA. We conducted five-fold cross-validation to evaluate the model performance, which involved partitioning the dataset into five equally sized folds, as described in Table 1. In each iteration of the five-fold cross-validation, four of the five subsets are used for training the model, while the remaining subset is utilized for testing its performance. This process is carried out five times, ensuring that each subset serves as the test set once. The final performance metric is obtained by computing the average of the results from each of the five iterations. This approach allows for a more accurate assessment of the model's generalizability and performance on diverse subsets of the data, reducing the risk of overfitting (16).

2.2. Development of a deep learning classifier

2.2.1. Image preprocessing

Image enhancement was performed to improve the quality and contrast of the OCT images, as they are inevitably susceptible to speckle noise, detection noise, and photon shot noise (17–19). Numerous studies have investigated the impact of image enhancement techniques on OCT images (20–22). The current study applied filtering, exponential enhancement, and linear enhancement to enhance the visibility of important features in the images, such as drusen and atrophic areas, as shown in Figures 2, 3. This process was intended to improve the performance of the CNN models by providing them with clear and informative images.

The outcome of the image enhancement procedure, performed on a representative image, is presented in Figure 3. The original image is presented in Figure 3A. Prior to further processing, an anisotropic diffusion filtering method (23) was employed to eliminate the speckle noise and accentuate the contrast between the distinct retinal layers, as demonstrated in Figure 3B. In the process of anisotropic diffusion, the parameters were set as $\alpha = 0.05$ and $K = 0.3$. Subsequently, exponential enhancement (24) was applied to account for the light attenuation in OCT images and improve the image contrast, leading to the image of Figure 3C. The exponential enhancement was performed with parameters $c = 0.4$ and $\gamma = 1.1$. Following this, a linear enhancement (25) method was implemented to further emphasize the layer by adapting the pixel values within a specific intensity range, thus rendering the layer more apparent and distinguishable from other features present in the image. The linear enhancement was applied with coefficient $m = 20$. The final result of image enhancement is demonstrated in Figure 3D.

2.2.2. Hierarchical classification

Within three phases of dry AMD, drusen-associated atrophy and nGA are both types of age-related macular degeneration that share similar imaging characteristics, making them difficult to

TABLE 1 Details of the training and validation datasets in five-fold cross validation.

	Fold 1		Fold 2		Fold 3		Fold 4		Fold 5	
	Train	Val	Train	Val	Train	Val	Train	Val	Train	Val
No of patients (%)	260 (79.5)	67 (20.5)	260 (79.5)	67 (20.5)	261 (79.8)	66 (20.2)	263 (79.9)	66 (20.1)	266 (80.9)	63 (19.1)
No of images (3,401)	2,720	681	2,806	595	2,587	814	2,742	659	2,749	652
Normal (712)	591	121	621	91	501	211	566	146	569	143
Drusen (1,167)	970	197	875	292	937	230	941	226	945	222
nGA (711)	554	157	592	119	564	147	560	151	574	137
GA (811)	605	206	718	93	585	226	675	136	661	150
Total images (patients)	3,401 (338)									

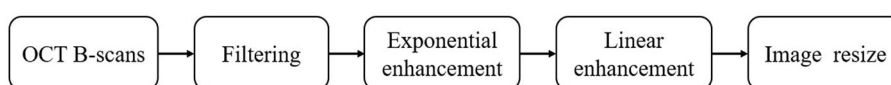


FIGURE 2
Flowchart of image enhancement methods adopted in preprocessing.

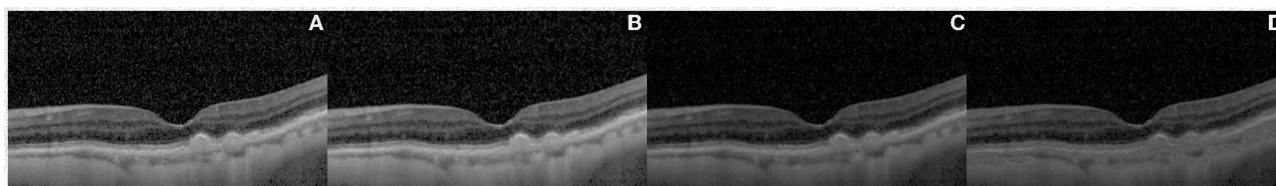


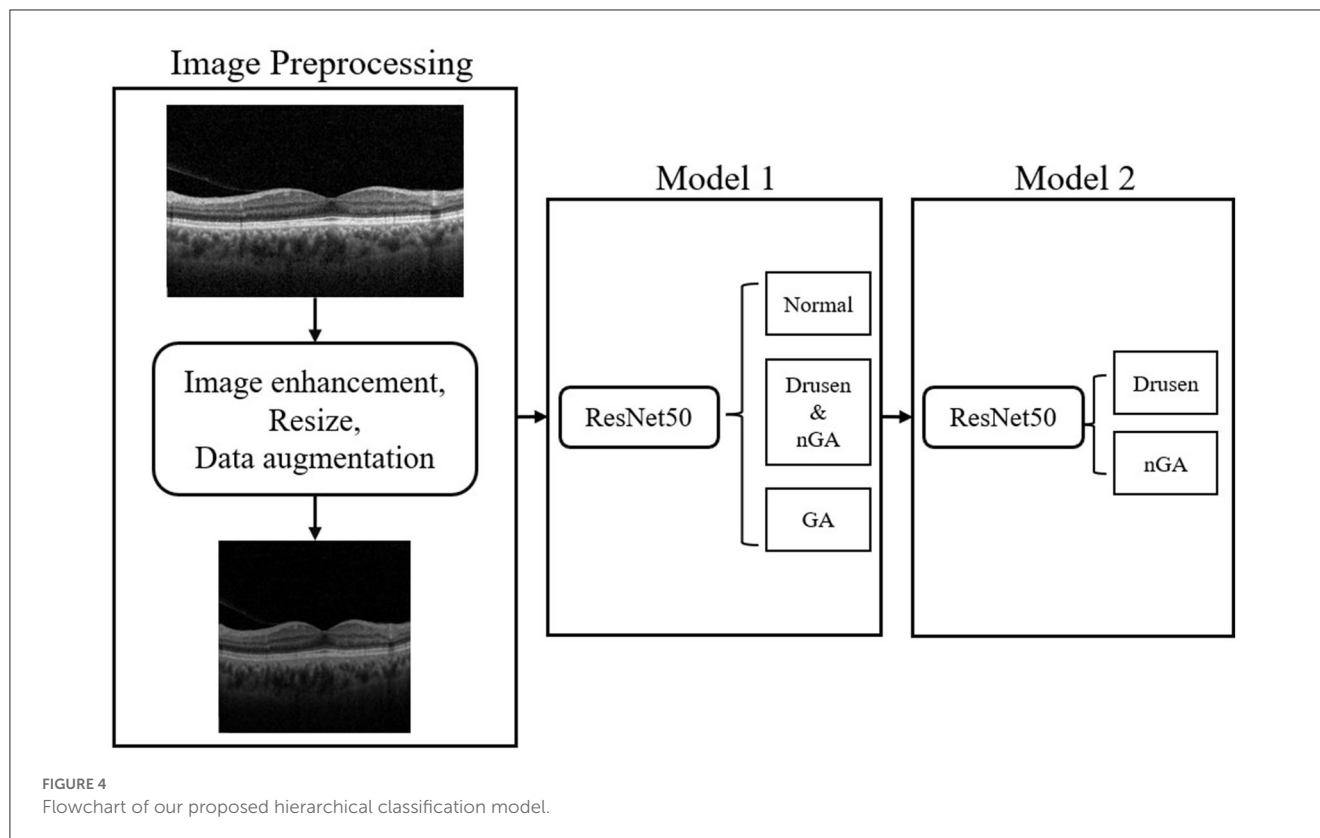
FIGURE 3
Results of the image enhancement procedure on a representative OCT image of retinal layers. The original image is shown in (A), while (B–D) illustrate the effect of different enhancement techniques, namely anisotropic diffusion filtering (B), exponential enhancement (C), and linear enhancement (D), on the image. The final image, shown in panel (D), highlights the retinal layer by adapting the pixel values within a specific intensity range, making it more distinguishable from other features in the image.

differentiate in OCT images. To tackle this problem, hierarchical models were developed to conduct a two-step classification of dry AMD. Hierarchical models perform better than flat models in image classification (26). Recent studies have leveraged the hierarchical organization of object categories to break down classification tasks into multiple stages, leading to successful outcomes (27–29). In this research, we implemented hierarchical classification by training two models, as depicted in the flowchart presented in Figure 4. The initial model generates a general dry AMD classification, distinguishing between normal, early stages of GA (i.e., drusen or nGA), and GA. The subsequent model concentrates on discriminating between drusen and nGA. Upon merging the outputs of the two models, a final classification result consisting of four labels is obtained.

In the hierarchical classification method, the base models serve as the primary feature extractor and the extracted features are then used to classify the images into their respective classes at different levels of the hierarchy. CNNs, which exhibit advantages in many application areas including medical diagnosis (30–33), are selected as the candidates of the base model in the hierarchical

classification. Four CNNs, specifically Normalizer-Free ResNet-50 (34), Xception (35), DenseNet169 (36), and EfficientNetV2 (37), were evaluated as classification models for dry AMD. These models are widely recognized for their remarkable performance in a variety of computer vision and medical image analysis applications, making them suitable for this study. All the models were initially trained on a vast labeled dataset known as ImageNet (38) and were subsequently fine-tuned on OCT images. The final layer of these models was removed, and a new fully connected layer with an output size of four was inserted to represent the four distinct classes (normal, drusen, nGA, and GA) for classification. To minimize the computation cost and select the model with the best performance for further analysis, we conducted a hold-out validation using only the Fold 1 dataset of Table 1, rather than employing five-fold cross-validation, to compare different CNN models.

We trained the deep learning models using PyTorch (39), a commonly used library in the deep learning community. During the training process, we updated the model parameters using the Adam optimizer (learning rate of 0.00002) for every minibatch of four images. The training was stopped after 20 epochs once



the accuracy values no longer increased or started to decrease. All experiments were conducted on a server with Intel Core i5-10600KF, using an NVIDIA GeForce RTX 3090 24GB GPU for training and validation, with 32GB RAM.

Cross-entropy loss curves were plotted against the number of training steps to visualize the model convergence during training. The loss function estimates the discrepancy between the predicted and actual labels and is employed to optimize the model during training. The loss curves enable us to observe how the models learn and converge toward the optimal solution.

Gradient-weighted class activation maps (Grad-CAM) was also employed to generate heatmaps that provide visual explanations of the classification model's predictions by highlighting the areas in retinal images that contributed significantly to the classification of normal, drusen, nGA, and GA cases. Grad-CAM is a visualization technique that employs the gradients of target class scores with respect to feature maps in the final convolutional layer to produce a coarse localization map, highlighting the important regions in the input image for a specific class prediction (40). This visualization technique not only assists in the interpretation of the model's decisions but also aids in identifying potential misclassifications or biases, consequently improving the model's usability and clinical applicability.

2.2.3. Evaluation metrics

To assess the efficacy of the four models, various metrics were employed, including accuracy, sensitivity, specificity, and the f1-score for each class, as well as the macro-f1 and kappa coefficients

for overall classification performance. Sensitivity gauges the ratio of true positive predictions for a given class, while specificity measures the proportion of true negative predictions. The f1-score is the harmonic mean of precision and recall, and offers a holistic measure of the model's accuracy for a particular class. The macro-f1 and kappa coefficients were used to evaluate the models' overall performance. The macro-f1 score is the arithmetic mean of f1-scores for each class, while the kappa coefficient measures the level of agreement between the predicted and actual labels, considering the possibility of chance agreement.

3. Results

3.1. Base model selection

Table 2 presents the results given by the four CNN models evaluated on a single fold of the validation dataset, with each of the four rows corresponding to the classification metrics for individual classes, i.e., normal, drusen, nGA, and GA.

Notably, all four CNN models produce commendable results for the normal, drusen, and GA categories, with all f1 scores exceeding 90%. However, the classification performance for nGA is generally inferior, with the EfficientNetV2, DenseNet169, Xception, and ResNet50NF models giving f1 scores of 82.49%, 82.68%, 80.44%, and 83.58%, respectively. Among the evaluated models, the ResNet50NF model produces the highest overall f1 scores, with 98.91% for normal, 94.60% for drusen, 83.58% for nGA, and 91.21% for GA. The macro-f1 and kappa scores using ResNet50NF are 92.08% and 95.45%, respectively.

TABLE 2 Classification results (percentage) of four CNN models on one fold of validation dataset.

Methods	Classes	Accuracy	Sensitivity	Specificity	F1	Macro-f1	Kappa
EfficientNetV2	Normal	98.66	100.00	98.41	95.79	89.93	94.01
	Drusen	94.96	92.12	97.69	94.72		
	nGA	92.44	89.08	93.28	82.49		
	GA	96.13	81.72	98.80	86.86		
DenseNet169	Normal	99.33	100.00	99.21	97.85	90.35	94.33
	Drusen	95.46	94.18	96.70	95.32		
	nGA	92.61	88.24	93.70	82.68		
	GA	95.80	79.57	98.80	85.55		
Xception	Normal	98.82	100.00	98.61	96.30	88.12	93.88
	Drusen	95.29	91.78	98.68	95.04		
	nGA	91.09	91.60	90.97	80.44		
	GA	94.62	72.04	98.80	80.72		
ResNet50NF	Normal	99.66	100.00	99.60	98.91	92.08	95.45
	Drusen	94.96	90.07	99.67	94.60		
	nGA	92.61	94.12	92.23	83.58		
	GA	97.31	89.25	98.80	91.21		

Each rows gives the classification metrics for individual classes: normal, drusen, nGA and GA. The best results are shown in bold.

Figure 5 illustrates the performance of the four CNN models on the training dataset in terms of cross-entropy loss (A), and on the validation dataset in terms of cross-entropy loss (B), macro-f1 score (C), and kappa coefficient (D) during the training process. Notably, ResNet50NF achieves the highest macro-f1 score and kappa coefficient, while also exhibiting fastest convergence, as evidenced in Figures 5A, B. Given the superior classification performance and fast convergence of ResNet50NF, this CNN was selected as the base model for the subsequent hierarchical classification. The proposed model with image enhancement and hierarchical classification achieves macro-f1 of $91.32 \pm 9.06\%$, and kappa score of $96.09 \pm 4.44\%$.

3.2. Ablation study

We compared the performance of each components of the proposed method by ablation study. The ablation results are presented in Table 3. The accuracy, sensitivity, specificity, and f1 score are listed for the normal, drusen, nGA, and GA classes. The macro-f1 and kappa values are also displayed as metrics of the overall classification performance.

The baseline model achieves a high f1 score on the normal (98.99%) and GA (90.07%) classes, but has a limited ability to classify drusen (85.24%) and nGA (66.79%) classes. Image enhancement slightly improves the classification performance of nGA (from 66.79% to 67.83%) and GA (from 90.07% to 91.11%), but does not improve the performance over the baseline for the normal and drusen classes. By applying hierarchical classification to the baseline model, the classification performance of nGA was improved (from 66.79% to 72.74%, and GA from 90.07% to 93.28%). The proposed method which includes image

enhancement along with hierarchical structure, achieves the highest accuracy for all classes, with a significant improvement on the nGA class (from 66.79% to 81.65%).

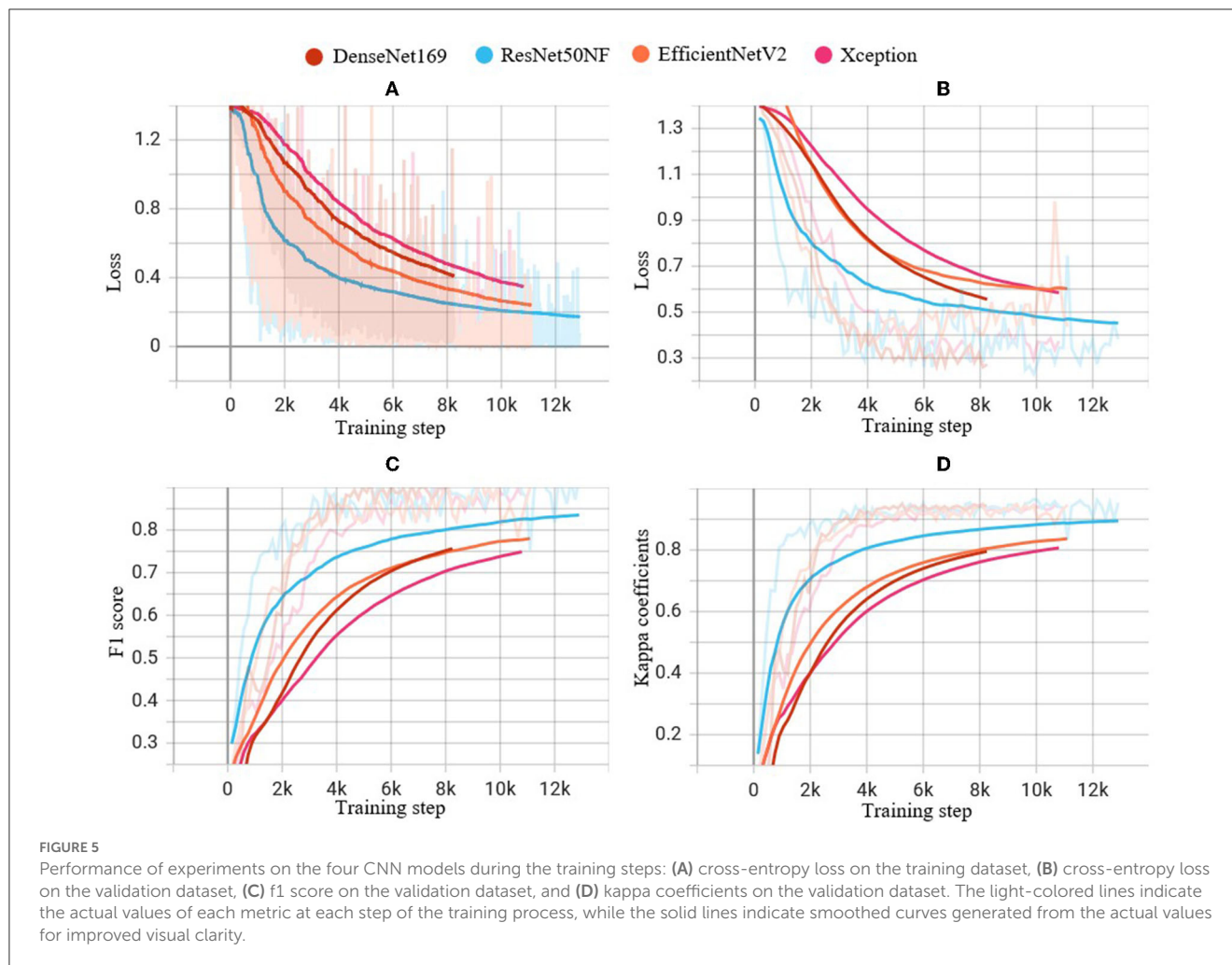
In terms of sensitivity and specificity, the proposed method outperforms the other two models in all classes, indicating that it can better distinguish between true positives and true negatives for each class. The macro-f1 and kappa coefficient for the proposed method, are 91.32% and 96.09% respectively, which are higher than the corresponding values without image enhancement and hierarchical structure. This demonstrates that the proposed model achieves better overall performance.

3.3. Model explainability

The heatmaps generated through Gram-CAM confirm that the proposed model produces an accurate diagnosis by leveraging distinctive features and pertinent regions or lesions within the image. As illustrated in Figure 7, the heatmaps demonstrate that the proposed model diagnoses the pathology based on the correct lesion. In the validation dataset, the model successfully detected pathological changes and identified distinguishing features for the three representative OCT results of the eyes that developed drusen, nGA, and GA. In particular, as depicted in the nGA example, the proposed model correctly highlighted the area where subsidence of OPL and INL were observed (gray mark in Figure 7C).

4. Discussion

The current study aimed to investigate the performance of classifying early stages of dry AMD based on OCT images.



The study proposed a novel hierarchical classification method that combines image enhancement with the ResNet50NF base CNN, which improves the classification performance of nGA, a challenging task for both ophthalmologists and deep learning models.

While most related studies have predominantly focused on using OCT images to classify exudative AMD, particularly those involving CNV, there has been a noticeable lack of research on dry AMD stages such as GA. These investigations (41, 42) have demonstrated promising results in identifying CNV within OCT images. However, when comparing CNV characteristics with those of dry AMD stages such as GA, several key differences in image features can be observed. For example, CNV typically presents with subretinal fluid, hemorrhages, and a distinct network of new blood vessels, whereas GA is characterized by a more uniform thinning of the retinal pigment epithelium and photoreceptor layers, along with the absence of fluid or hemorrhages.

GA is a chronic ocular condition that causes a decline in visual function, leading to difficulties in performing everyday activities such as reading, recognizing faces, and driving, ultimately resulting in a loss of independence. As the disease progresses and the lesions expand, patients often experience a slow and steady decline in visual function. Managing and treating GA can be particularly challenging

because the available treatment options are currently limited. Most current treatment options primarily aim to slow the progression of the disease, rather than reversing the damage that has already occurred. Therefore, early detection, intervention, and ongoing monitoring and management of the disease are crucial for effective disease control.

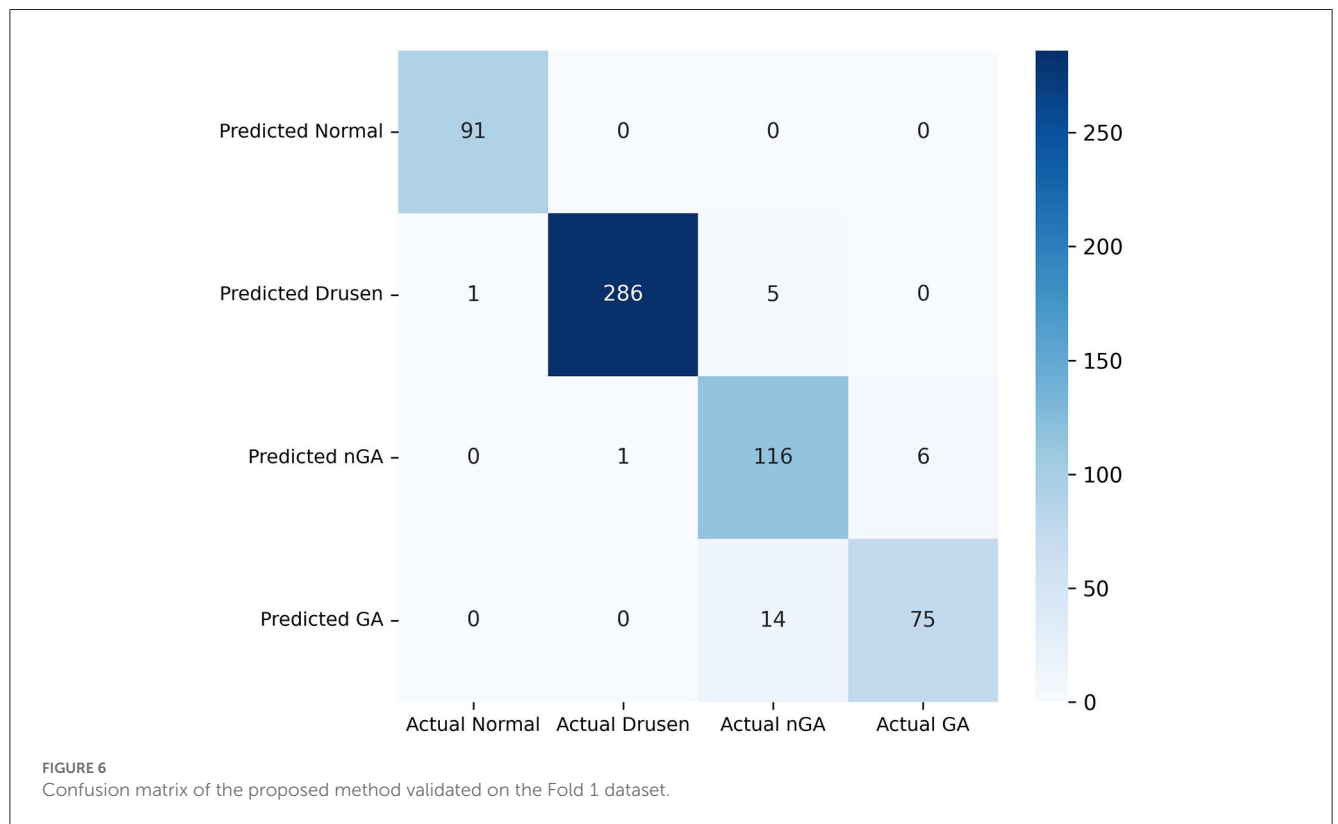
Numerous previous studies have demonstrated that OCT is an effective tool for training CNNs to identify common retinal diseases such as dry AMD. Nonetheless, due to the complicated manifestations of nGA, implementing CNN models that classify the early stages of dry AMD is a significant challenge. The results of our experiments reveal that, among the evaluated CNN models, all models provide satisfactory classification performance for the normal and GA categories. However, nGA classification performance was generally inferior. This emphasizes the challenge of accurate classification of the nGA category and the need for novel approaches to improve the accuracy of classification.

We adapted normalization free ResNet50 as the backbone of the model, which is a variant of ResNet50. This variant omits normalization layers, which traditionally standardize the range of data features (34). The absence of normalization boosts computational efficiency and maintains a wider range of data

TABLE 3 Ablation study results (%) of the baseline model, the model with only image enhancement, the model with only hierarchical classification, and the proposed method.

Methods	Classes	Accuracy	Sensitivity	Specificity	F1	Macro-f1	Kappa
ResNet50NF (baseline)	Normal	98.99 ± 1.44	99.73 ± 0.76	98.78 ± 1.65	97.75 ± 3.02	84.96 ± 3.98	93.01 ± 2.19
	Drusen	90.33 ± 4.41	86.11 ± 11.41	92.05 ± 6.8	85.24 ± 7.97		
	nGA	86.16 ± 4.53	67.82 ± 19.39	91.03 ± 6.0	66.79 ± 11.17		
	GA	95.55 ± 2.95	86.09 ± 6.89	98.34 ± 1.78	90.07 ± 5.42		
Baseline+ Image enhancement	Normal	98.74 ± 0.72	98.27 ± 3.6	98.86 ± 1.06	97.06 ± 1.57	85.17 ± 4.58	93.23 ± 1.67
	Drusen	89.89 ± 4.79	86.18 ± 13.2	91.37 ± 8.05	84.68 ± 8.83		
	nGA	86.78 ± 4.76	68.31 ± 20.94	91.69 ± 6.15	67.83 ± 12.18		
	GA	95.94 ± 3.36	87.78 ± 7.97	98.39 ± 1.96	91.11 ± 6.52		
Baseline+ Hierarchical classification	Normal	99.48 ± 0.87	99.45 ± 1.15	99.54 ± 0.92	99.01 ± 1.76	89.78 ± 4.91	95.46 ± 2.35
	Drusen	93.26 ± 5.98	89.34 ± 13.98	94.75 ± 7.25	88.86 ± 8.77		
	nGA	89.96 ± 5.13	73.24 ± 20.38	93.85 ± 6.53	72.74 ± 12.88		
	GA	96.84 ± 2.97	91.57 ± 7.89	99.07 ± 1.68	93.28 ± 6.49		
Proposed method	Normal	99.68 ± 0.9	99.67 ± 0.92	99.68 ± 0.89	99.12 ± 2.45	91.32 ± 9.06	96.09 ± 4.44
	Drusen	94.12 ± 6.63	90.04 ± 17.64	95.89 ± 5.16	90.3 ± 13.23		
	nGA	91.56 ± 9.26	81.87 ± 13.4	94.0 ± 10.76	81.65 ± 14.65		
	GA	97.12 ± 3.95	93.7 ± 13.29	98.57 ± 1.92	94.22 ± 7.05		

The best results are shown in bold.



features, thus enhancing the learning and performance capabilities in our specific image-based task of classifying dry AMD stages.

The proposed method’s effectiveness can be attributed to its ability to combine image enhancement with hierarchical

classification. An ablation study was conducted to compare the proposed method’s performance with that of the baseline model and the baseline with image enhancement. The results demonstrated that hierarchical classification along with image

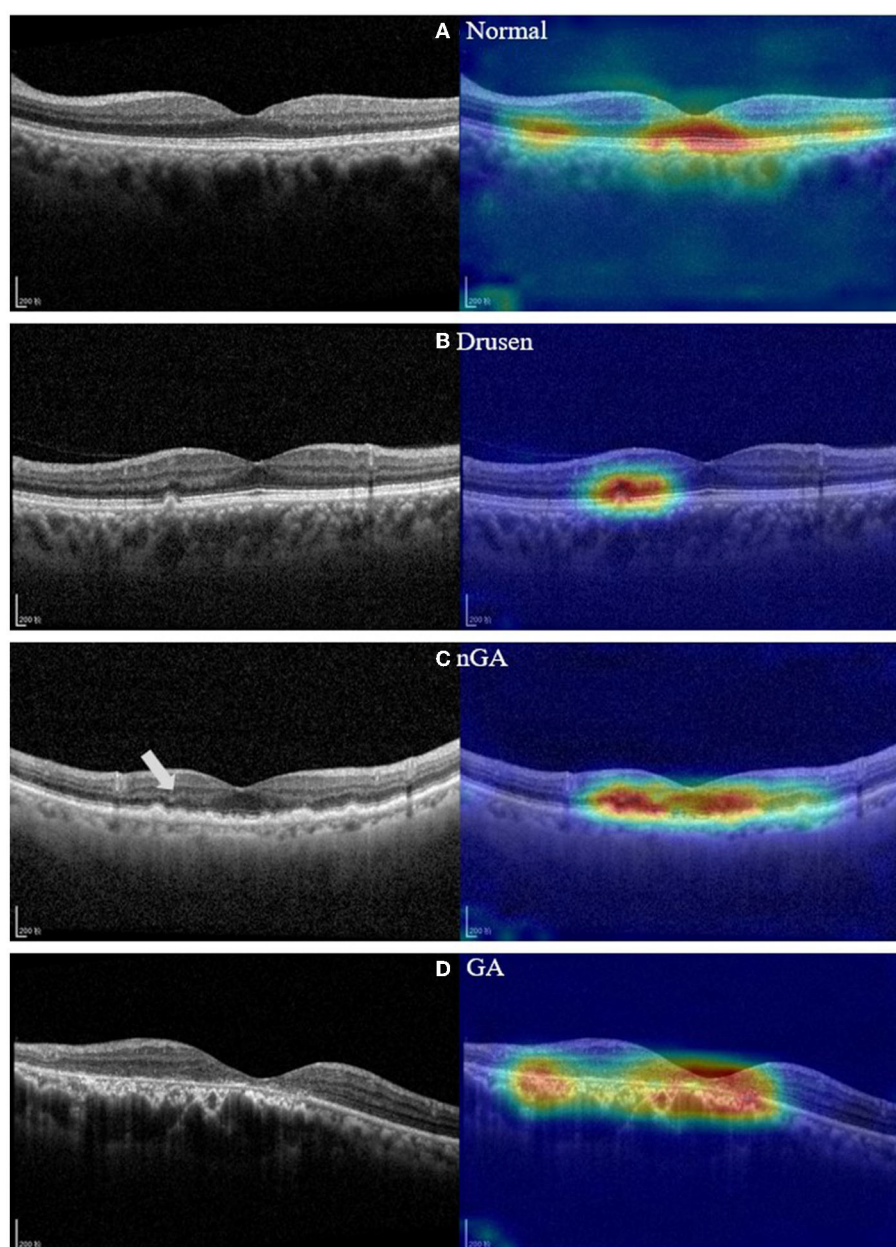


FIGURE 7

OCT images and feature heatmaps demonstrating example cases of normal, drusen, nGA and GA. (A) Example of OCT image of control group; (B) Example of a drusen regression detected by our model; (C) Example of nGA detected by our model. (D) Example of GA detected by our model. On the left side of the nGA example (C), a discernible subsidence of OPL is marked.

enhancement significantly improves the classification performance for all categories, particularly nGA. The proposed method achieved the highest accuracy, sensitivity, specificity, f1 score, macro-f1 score, and kappa value for the three phases of dry AMD (drusen, nGA, and GA). The superiority of the proposed method suggests that it has the potential to be an effective and reliable tool for early detection and monitoring of AMD.

Our model exhibits substantial effectiveness in correctly predicting each phase of dry AMD, with high true positive values across all categories, as observed from the confusion matrix in Figure 6. A key observation is the high sensitivity in the

identification of the drusen and GA phase. However, the matrix also highlights areas for potential improvement in the model, primarily in reducing false positives between the nGA and GA stages.

The image enhancement component improves the image quality, resulting in more accurate feature extraction and better CNN classification performance. The hierarchical classification component utilizes a two-step approach, in which the first CNN model classifies images into three categories, and then the second classification stage distinguishes between nGA and GA. This approach allows for accurate and reliable classification of nGA, which is a challenging category due to its similarity to drusen. These

findings contribute to the development of a reliable and efficient automated diagnostic tool for early detection and monitoring of AMD, ultimately leading to improved patient outcomes. Further validation and testing of the proposed method on larger and more diverse datasets are necessary to confirm its generalizability and robustness.

Our model can precisely identify distinguishing characteristics within OCT images of different dry AMD phases as demonstrated by the heatmaps of Figure 7. In Figure 7B, the heatmap of our model focuses on the disruptions in the overlying retina layers, corresponding to the presence of drusen. This indicates that our model accurately identifies drusen features in this image. In Figure 7C, the model highlighted a relatively larger area in which OPL subsidence can be observed, as marked on the left side of the image. The heatmap of Figure 7D highlights regions where evidence of RPE thinning and depression of the outer retinal layers is present. These imaging features indicate the loss of retinal tissue and disruption of the normal retinal architecture, which are typically observed in GA.

4.1. Limitations

The present study has some limitations that could be addressed in future research. First, the study used the dataset from one center for evaluation. Further multi-center validation on larger and more diverse datasets is necessary to confirm the proposed method's robustness and generalizability. Second, the study evaluated four popular CNN models. It would be interesting to compare the proposed method's performance with state-of-the-art CNNs. Third, we did not perform exhausted fine-tuning on the image enhancement methods, which may have led to limited performance improvement solely through image enhancement. With further fine-tuning of the image enhancement stage, the image noise on the OCT images may be better eliminated, enabling the better identification of image characteristics for nGA. Future research should address these limitations and further validate and optimize the proposed method for practical implementation in clinical settings. Fourth, the phases of dry AMD was graded by a single clinician. While the clinician ensured a consistent evaluation standard across all OCT images, future work could benefit from double grading of OCT images by two or more experienced clinicians with consensus adjudication. Despite this limitation, our study has demonstrated the promising potential of our model for accurate classification of dry AMD stages.

Future work will expand upon this study by incorporating longitudinal OCT data from multiple follow-up visits to examine the progression of dry AMD from drusen to nGA and from nGA to GA. This will allow the investigation of the deep learning model's capability to identify and correctly predict the progression risk of dry AMD.

5. Conclusion

This study proposed and validated a novel two-step hierarchical CNN model with image enhancement for the classification of early-stage dry AMD, including the identification of nGA. The

proposed method combines image enhancement with a CNN-based hierarchical model to improve the classification of early-stage dry AMD using OCT images. The results demonstrate the superior classification performance of the proposed method, particularly for the challenging nGA category. The proposed method's effectiveness could contribute to the development of a reliable and efficient automated diagnostic tool for early detection and monitoring of dry AMD. We believe that the proposed approach provides a valuable computer-assisted diagnostic tool for clinical diagnosis of dry AMD based on OCT. This could facilitate prompt and efficient interventions, leading to better management of the condition and patient outcome, and reducing the burden on ophthalmologists.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the data used in this study are not publicly available due to concerns about patient privacy. However, the code used to analyze the data and generate the results is available. The corresponding author can provide access to the code upon reasonable request. Any queries regarding the availability of materials should be addressed to the corresponding author. Requests to access these datasets should be directed to WD, daiweiwei@aierchina.com.

Ethics statement

This study received ethics approval from the Ethics Committee of Shenyang Aier Excellence Hospital (approval number: 2021KJB002). The participants' confidentiality and anonymity were maintained throughout the study, and the research was conducted in accordance with the principles outlined in the Declaration of Helsinki. The study was conducted in compliance with all applicable ethical standards and regulations.

Author contributions

MH and JX conducted the development of the models and drafted the manuscript. BW and DL contributed to the collection and labeling of the dataset used in this study. YC, ZY, and WD provided critical manuscript revisions and supervised the study. All authors contributed to the study concept and design, acquired and interpreted study data, and read and approved the final manuscript.

Acknowledgments

The authors of this paper wish to express their sincere gratitude to the Science and Innovation Foundation of Hunan Province of China (2020SK50110) and the Science and Innovation Leadership Plan of Hunan Province of China (2021GK4015) for

providing the financial support necessary to conduct the research presented herein. Without their generous funding, this work would not have been possible.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Nowak JZ. Age-related macular degeneration (AMD): pathogenesis and therapy. *Pharmacol Rep.* (2006) 58:353–63.
- Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. *Lancet.* (2018) 392:1147–59. doi: 10.1016/S0140-6736(18)31550-2
- Fleckenstein M, Mitchell P, Freund KB, Sadda S, Holz FG, Brittain C, et al. Age-related macular degeneration. *Nat Rev Dis Primers.* (2021) 7:31. doi: 10.1038/s41572-021-00265-2
- Yang K, Liang YB, Gao LQ, Peng Y, Zhang JJ, Wang JJ, et al. Prevalence of age-related macular degeneration in a rural Chinese population: the Handan Eye Study. *Ophthalmology.* (2011) 118:1395–401. doi: 10.1016/j.ophtha.2010.12.030
- Steinmetz JD, Bourne RRA, Briant PS, Flaxman SR, Braithwaite T, Cicinelli MV, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Global Health.* (2021) 9:e144–60. doi: 10.1016/S2214-109X(20)30489-7
- Ambati J, Ambati BK, Yoo SH, Ianchulev S, Adamis AP. Age-related macular degeneration: etiology, pathogenesis, and therapeutic strategies. *Surv Ophthalmol.* (2003) 48:257–93. doi: 10.1016/S0039-6257(03)00030-4
- Liao DS, Grossi FV, El Mehdi D, Wang H, Nair AA, Schönbach EM, et al. Complement C3 inhibitor pegcetacoplan for geographic atrophy secondary to age-related macular degeneration: a randomized phase 2 trial. *Ophthalmology.* (2020) 127:186–95. doi: 10.1016/j.ophtha.2019.07.011
- U.S. Food and Drug Administration, Center for Drug Evaluation and Research. *Syfovre NDA 217171 Approval Letter, February 17, 2023.* Available online at: https://www.accessdata.fda.gov/drugsatfda_docs/applletter/2023/217171Orig1s000ltr.pdf (accessed February 21, 2023).
- Wu Z, Ayton LN, Luu CD, Guymer RH. Prospective longitudinal evaluation of nascent geographic atrophy in age-related macular degeneration. *Ophthalmol Retina.* (2020) 4:568–75. doi: 10.1016/j.oret.2019.12.011
- Sadda SR, Guymer R, Holz FG, Schmitz-Valckenberg S, Curcio CA, Bird AC, et al. Consensus definition for atrophy associated with age-related macular degeneration on OCT: classification of atrophy report 3. *Ophthalmology.* (2018) 125:537–48. doi: 10.1016/j.ophtha.2017.09.028
- Guymer RH, Rosenfeld PJ, Curcio CA, Holz FG, Sadda SR, Staurenghi G, et al. Incomplete retinal pigment epithelial and outer retinal atrophy in age-related macular degeneration: classification of atrophy meeting report 4. *Ophthalmology.* (2020) 127:394–409. doi: 10.1016/j.ophtha.2019.09.035
- Peng Y, Dharssi S, Chen Q, Keenan TD, Agrón E, Wong TY, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology.* (2019) 126:565–75. doi: 10.1016/j.ophtha.2018.11.015
- Treder M, Lauermann 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. doi: 10.1007/s00417-017-3850-3
- Keenan TD, Grossi FV, Jun G, Thakur N, Khanna A, Dandekar S, et al. A deep learning approach for automated detection of geographic atrophy from color fundus photographs. *Ophthalmology.* (2019) 126:1533–40. doi: 10.1016/j.ophtha.2019.06.005
- Sarici K, Çelik M, Ersan İ, Gul A, Tufan HA, Yıldırım E, et al. Risk classification for progression to subfoveal geographic atrophy in dry age-related macular degeneration using machine learning-enabled outer retinal feature extraction. *Ophthalmic Surg Lasers Imaging Retina.* (2022) 53:31–9. doi: 10.3928/23258160-20211210-01
- Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv.* (2010) 4:40–79. doi: 10.1214/09-SS054

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Bashkansky M, Reintjes J. Statistics and reduction of speckle in optical coherence tomography. *Opt Lett.* (2000) 25:545–7. doi: 10.1364/OL.25.000545
- Vanselow A, Heuke S, Burkhardt M, Riedel M, Dienesowitz M, Zeitner UD, et al. Frequency-domain optical coherence tomography with undetected mid-infrared photons. *Optica.* (2020) 7:1729–36. doi: 10.1364/OPTICA.400128
- Jensen M, Thøgersen J, Hermann N, Larsen H. Noise of supercontinuum sources in spectral domain optical coherence tomography. *JOSA B.* (2019) 36:A154–60. doi: 10.1364/JOSAB.36.00A154
- Halupka KJ, Wong A, Clausi DA. Retinal optical coherence tomography image enhancement via deep learning. *Biomed Opt Express.* (2018) 9:6205–21. doi: 10.1364/BOE.9.006205
- Devalla SK, Agrawal R, Gupta V, Tan JCK, Cheung CY, Wong TY, et al. A deep learning approach to denoise optical coherence tomography images of the optic nerve head. *Sci Rep.* (2019) 9:1–13. doi: 10.1038/s41598-019-51062-7
- Xu X, Li Y, Xie H, Shen X, Liu S, Yan S, et al. Automatic segmentation and measurement of choroid layer in high myopia for OCT imaging using deep learning. *J Digital Imaging.* (2022) 35:1153–63. doi: 10.1007/s10278-021-00571-x
- Weickert J. *Anisotropic Diffusion in Image Processing.* Stuttgart: Teubner (1998).
- Shen J, Castan S. An optimal linear operator for step edge detection. *CVGIP Graph Models Image Process.* (1992) 54:112–33. doi: 10.1016/1049-9652(92)90060-B
- Castellanos P, del Angel PL, Medina V. Deformation of MR images using a local linear transformation. In: *Medical Imaging 2001: Image Processing.* San Diego, CA: SPIE (2001), p. 909–16. doi: 10.1117/12.430963
- Kowsari K, Heidarysafa M, Brown DE, Barnes LE, Alamoody AH. HDLTEX: hierarchical deep learning for text classification. In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA).* Cancun: IEEE (2017), 364–71. doi: 10.1109/ICMLA.2017.0-134
- Seo Y, Shin KS. Hierarchical convolutional neural networks for fashion image classification. *Expert Syst Appl.* (2019) 116:328–39. doi: 10.1016/j.eswa.2018.09.022
- Ranjan N, Rashmi K, Jain A. Hierarchical approach for breast cancer histopathology images classification. In: *International Conference on Medical Imaging with Deep Learning 2018.* Amsterdam (2018).
- Zhu X, Bain M. B-CNN: branch convolutional neural network for hierarchical classification. *arXiv.* [preprint]. doi: 10.48550/arXiv.1709.09890
- Abdel-Hamid O, Mohamed AR, Jiang H, Penn G. Convolutional neural networks for speech recognition. *IEEE/ACM Trans Audio Speech Lang Process.* (2014) 22:1533–45. doi: 10.1109/TASLP.2014.2339736
- Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging.* (2018) 9:611–29. doi: 10.1007/s13244-018-0639-9
- Anwar SM, Majid M, Qayyum A, Awais M, Alnowami MR. Medical image analysis using convolutional neural networks: a review. *J Med Syst.* (2018) 42:1–13. doi: 10.1007/s10916-018-1088-1
- Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging.* (2016) 35:1299–312. doi: 10.1109/TMI.2016.2535302
- Brock A, Donahue J, Simonyan K. High-performance large-scale image recognition without normalization. In: *Proceedings of Machine Learning Research 2021.* (2021). p. 1059–71.
- Chollet F. Xception: deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Honolulu, HI: IEEE (2017), 1251–8. doi: 10.1109/CVPR.2017.195

36. Huang G, Liu Z, Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017), p. 4700–8. doi: 10.1109/CVPR.2017.243
37. Tan M, Le Q. EfficientNetV2: Smaller models and faster training. In: *Proceedings of Machine Learning Research 2021*. (2021). p. 10096–106.
38. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. (2017) 60:84–90. doi: 10.1145/3065386
39. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: *Advances in Neural Information Processing Systems 32*. Red Hook, NY: Curran Associates, Inc. (2019). pp. 8024–35.
40. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE (2017) 618–26. doi: 10.1109/ICCV.2017.74
41. Wang J, Hormel TT, Gao L, Zang P, Guo Y, Wang X, et al. Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning. *Biomed Opt Express*. (2020) 11:927–44. doi: 10.1364/BOE.379977
42. Jin K, Yan Y, Chen M, Wang J, Pan X, Liu X, et al. Multimodal deep learning with feature level fusion for identification of choroidal neovascularization activity in age-related macular degeneration. *Acta Ophthalmol*. (2022) 100:e512–20. doi: 10.1111/aos.14928