Automatic Discrimination for Screening of the Eyes in a Deep Learning-Based Ensemble Model Using Optical Coherence Tomography Images

Masakazu Hirota (hirota.ortho@med.teikyo-u.ac.jp)
Teikyo University

Shinji Ueno
Nagoya University Graduate School of Medicine

Taiga Inooka
Nagoya University Graduate School of Medicine

Yasuki Ito
Fujita Health University Hospital

Hideo Takeyama
Aichi Health Promotion Foundation

Yuji Inoue
Teikyo University

Emiko Watanabe
Teikyo University

Atsushi Mizota
Teikyo University

Article

Keywords:

Posted Date: March 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1440482/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

The average life expectancy has increased globally, and the risk of visual impairment is expected to increase as well. Therefore, eye checkups have become increasingly important to maintain good vision and quality of life. As the population requiring eye checkups increases, so does the clinical work burden of physicians. Hence, an automatic discrimination algorithm to reduce the clinicians' workload is necessary. The convolutional neural network (CNN), a deep learning algorithm, has recently become one of the chief techniques for automated image recognition and is a helpful tool for identifying ocular diseases. However, the accuracy of a single CNN model ranges within 70%-95%, and it may fail to identify some diseases. In this study, we aimed to compare the diagnostic performances of four CNN models trained with optical coherence tomography (OCT) images, the machine-learning (ML) model trained with data on the retinal and choroidal area, and the ensemble model that integrated the CNNs and ML models using OCT images obtained during eye checkups. Our results show that the ensemble model had a superior diagnostic performance over the CNN and ML models. The ML model, which evaluated diseases using data regarding the temporal peripheral retinal area, improved on the CNN model, which misrecognized the temporal peripheral retinal structures. Our study indicates the strong potential of the ensemble model combining the CNN and ML models in accurately predicting abnormalities during eye checkups.

1. Introduction

The prevalence of visual impairment is higher among older people, and the significant causes include glaucoma, age-related macular degeneration, and diabetic retinopathy.1,2 The global average life expectancy has increased, and the risk of visual impairment is expected to increase accordingly.3 Therefore, eye checkups are essential to maintain good vision and quality of life.

In Japan, eye checkups are performed at a frequency of 16.2% by the local governments,4 and all eye checkups mostly use fundus photography. The use of optical coherence tomography (OCT)5 has become widespread globally, as it is more accurate in detecting retinal disease than fundus photography.6 Therefore, OCT is expected to be introduced into standard eye checkups.

As the population requiring eye checkups increases, concern has grown regarding the corresponding burden on clinicians. Each day, a single clinician must check the findings of hundreds of person-visits,4 thereby requiring an automatic classification algorithm to reduce clinicians' burden. Furthermore, automatic classification is helpful when non-ophthalmologists are involved in health checkups, as due to a shortage of ophthalmologists in some areas.7,8

The convolutional neural network (CNN), a deep learning algorithm, has recently become one of the main techniques in automated image recognition.9 Recent studies have shown that the CNN model can automatically identify retinal disease using OCT images10,11 and has a slightly higher diagnostic performance than the human eye.11,12 These studies suggest that the CNN is a useful tool for identifying
ocular diseases and for reducing the burden on ophthalmologists. However, the accuracy of a single CNN model is about 70–95%, and it may fail to identify some diseases.\textsuperscript{10,11,13}

Ensemble learning, which creates and integrates several different classifiers, is a technique used to improve the accuracy of the CNN model.\textsuperscript{14} An earlier study reported that ensemble learning is useful, even with small datasets.\textsuperscript{15} Ensemble learning is commonly used to divide a single dataset into multiple parts and integrate them at the end. If a discriminator is created using different datasets, a higher accuracy is obtained.

Other studies have shown that retinal and choroidal thicknesses are indicators for retinal diseases.\textsuperscript{16,17} Thus, we hypothesized that integrating a CNN model trained with OCT images with a machine learning (ML) model trained with data values of the retinal area will improve the accuracy of disease estimation. This study aimed to compare the diagnostic performances of the CNN model trained with OCT images, the ML model trained with retinal data values, and the ensemble model integrating the CNN and ML models using OCT images obtained during an eye checkup.

2. Methods

2–1. Data acquisition

Patients underwent a health and eye checkup between April 2017 and December 2019 at Aichi Health Promotion Foundation (Aichi, Japan), Nagoya University Hospital (Aichi, Japan), and Teikyo University Hospital (Tokyo, Japan). We implemented an opt-out method of obtaining patient's informed consent for this study. This investigation adhered to the tenets of the World Medical Association Declaration of Helsinki. The study was approved by the Institutional Review Board of Nagoya University (Approval No. 2017–0283) and Teikyo University (Approval No. 18–161).

2–2. OCT imaging

OCT images from both eyes were obtained using an OCT-HS100 (Canon Co., Ltd., Tokyo, Japan) and RS-3000 Advance (RS-3000; Nidek Co., Ltd., Aichi Japan). OCT-HS100 and RS-3000 have an auto-eye-tracking feature for the posterior direction, auto-alignment, and an auto-focus system. Thus, the OCT-HS100 and RS-3000 provide multiple OCT images and are suitable for eye checkups.

OCT-HS100 has an A-scan rate of 70,000 scans/s with a superluminescent diode with a $\lambda_{\text{max}}$ of 855 nm and creates a cross-sectional image (B-scan). In this study, the B-scan image (OCT image) captured a single shot with a horizontal and vertical angle of view of 9 mm, a resolution of $1,024 \times 1,176$ pixels, and TIFF compression.

RS-3000 has an A-scan rate of 53,000 scans/s with a superluminescent diode with a $\lambda_{\text{max}}$ of 880 nm and creates a B-scan image. In this study, the OCT image captured a single shot with a horizontal and vertical angle of view of 9 mm, a resolution of $1,024 \times 512$ pixels, and JPG compression. The OCT
images from RS-3000 were resized to a resolution of 1,024 × 1,176 pixels and converted to TIFF compression.

2–3. Datasets

A total of 7,703 OCT images were captured over the course of 3 years. All OCT images were double reviewed and labeled by two ophthalmologists at each hospital (S.U. and T.I. labeled the OCT images from OCT-HS100; Y.I and E.W. labeled the OCT images from RS-3000). In case of multiple ocular diseases, the ocular disease with the most abnormalities was diagnosed. Images with findings that the ophthalmologists could not mutually agree on and those that did not lead to a diagnosis were excluded.

Of the OCT images, 655 were classified as abnormal findings, whereas 6,050 were normal. The OCT images of 998 eyes were not used because of difficulties in their interpretation. The number of images in both classifications was then adjusted to match the number of abnormal findings; thus, 655 normal images were extracted randomly. The training and test datasets were randomly divided into 1,210 and 100 images, respectively (with an abnormal-to-normal ratio of 1:1).

3. Experiment 1

In Experiment 1, we compared the diagnostic performances using transfer learning from CNN models pretrained with ResNet-152, DenseNet-201, and EfficientNet-B7.

3.1. Methods

3.1.1. Preprocessing

The original OCT images were center-cropped to 600 × 600 pixels (probability = 50%) and then resized to 512 × 512 pixels. After resizing, data augmentation was applied to the input images as follows: random brightness from 0.8 to 2.0 times, random contrast from 0.8 to 1.5 times, random rotation within 10 degrees, random horizontal and vertical shift within 50 pixels, and random horizontal mirroring (probability = 50%). For margins created by image processing, we used padding with blue (red, green, and blue color information of 0, 0, and 255, respectively) to prevent misrecognition (Supplementary Fig. 1).

3.1.2 Network

In the training phase for transfer learning, supervised learning was used where the network model was given training images. The accuracy of the classification measured as the weights of the deep layers were changed. The weights were changed based on the optimization function. In this study we used the Adam optimizer for all CNN models. The layers of deep neural networks were frozen until just before the output layer in order to use the ImageNet weight parameters. We created the fully connected layer as an output layer. The fully connected layer provided two outputs (abnormal or normal eyes) using the softmax function. We defined an abnormal eye as a predicted value $\geq 0.5$. 
All CNN models trained with 2,000 epochs. The optimizer used an adaptive learning rate; the primary learning rate was 0.02, which was subsequently reduced to 0.5 times at 25%, 50%, 75%, and 90% of the total number of epochs. The training data were divided into three parts and cross-validated.

We used Python 3.8.5 for Windows 10 (Microsoft Co., Ltd., Redmond, WA, USA), with the following libraries: Matplotlib 3.3.2, Numpy 1.18.5, OpenCV 3.3.1, Pandas 1.1.3, Pytorch 1.7.0, Torchvision 0.8.1, Scikit-learn 0.23.2, and Seaborn 0.11.0.

3.1.3. Data visualization

The explanations for the abnormal predictions by the CNN models were visualized using gradient-weight class activation mapping (Grad-CAM). Grad-CAM can generate visual explanations from any CNN-based network without requiring architectural changes or retraining. Grad-CAM images were generated using the feature map in the last convolutional layer.

3.1.4. Ensemble model

The ensemble model used a soft voting algorithm to average the predictions of multiple models.

3.1.5. Statistical analysis

We used receiver-operating characteristic (ROC) curves and calculated the corresponding area under each curve with 1,000 times bootstrap to evaluate the diagnostic performance of the CNN and ensemble models. Then, the area under the ROC curve (AUC) was compared among the models using Scheffé test.

IBM SPSS Statistics version 26 (IBM Corp., Armonk, NY, USA) was used for statistical analysis, and a $P$-value < 0.05 was considered significant.

3.2. Results 1

The abnormal OCT images included 127 of age-related macular degeneration, 3 of asteroid hyalosis, 34 of central serous chorioretinopathy, 60 of drusen, 114 of epiretinal membrane (ERM), 25 of glaucoma, 5 of location error, 17 of macular degeneration, 75 of macular edema, 19 of macular hole, 77 of high myopia, 5 of postoperative complications (ERM or macular edema), 91 of retinitis pigmentosa (RP), 4 of rhegmatogenous retinal detachment, and 5 of vitreomacular traction syndrome (Table 1).
Table 1
Detailed overview of the abnormal OCT images

<table>
<thead>
<tr>
<th>Disease</th>
<th>Training and validation data</th>
<th>Test data</th>
<th>ResNet Failure</th>
<th>DenseNet Failure</th>
<th>EfficientNet Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD</td>
<td>122</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asteroid hyalosis</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSC</td>
<td>33</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drusen</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERM</td>
<td>107</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glaucoma</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location error</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macular degeneration</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macular edema</td>
<td>60</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macular hole</td>
<td>15</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High myopia</td>
<td>72</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-operation</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP</td>
<td>76</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RRD</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VMTS</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>605</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

AMD, age-related macular degeneration; CNN, convolutional neural network; CSC, central serious chorioretinopathy; ERM, epiretinal membrane; ML, machine learning; OCT, optical coherence tomography; RP, retinitis pigmentosa; RRD, rhegmatogenous retinal detachment; VMTS, vitreomacular traction syndrome.

The accuracies of the CNN models with ResNet-152, DenseNet-201, and EfficientNet-B7 and the ensemble model were 95.0%, 95.0%, 96.0%, and 98.0%, respectively (Supplementary Fig. 2). The AUCs of the CNN models with ResNet-152, DenseNet-201, and EfficientNet-B7 and the ensemble model were 0.989 (95% confidence interval [CI], 0.968–1.000), 0.997 (95% CI, 0.986–1.000), 0.997 (95% CI, 0.986–1.000), and 0.998 (95% CI, 0.989–1.000), respectively (Fig. 1). The diagnostic performance was significantly greater in the ensemble model than in the other CNN models ($P<0.001$).

In most cases, the CNN models focused on abnormal images of the retina to predict the disease (Fig. 2). The images that the CNN models misjudged did not have remarkable lesions in the inner retinal layers,
and in such cases, the CNN models focused on the nasal and temporal retinal regions to make decisions (Fig. 3).

These findings suggest that combining the CNN model with a model that detects abnormalities in the peripheral nasal and temporal retinal regions will improve disease prediction accuracy.

4. Experiment 2

In Experiment 2, we examined the ability of the ML model to identify anomalies in the thickness of peripheral nasal and temporal retinal regions, which was the weak point of the CNN models created in Experiment 1. Then, we developed an ensemble model combining the CNN models (ResNet-152, DenseNet-201, and EfficientNet-B7) and the ML model and verified the accuracy of disease estimation.

4.1. Methods

4.1.1. Preprocessing

The original OCT images in the left eye were flipped horizontally. Twenty percent of the total pixel size of all original OCT images (Fig. 4A) was removed from the right and left edges to avoid depression of the optic disc (Resolution: 615 × 1,176 pixels; Fig. 4B). The OCT images were divided into five sections (Resolution: 123 × 1,176 pixels; Fig. 4C): The peripheral temporal retina, temporal perimacular area, central macular area, nasal perimacular area, and peripheral nasal retina as segments 1, 2, 3, 4, and 5, respectively. Each section was binarized using the discriminant analysis method (Fig. 4D). Morphological closing was applied to these segment images to pad the dark area related to the inner retinal layer and choroidal vessels. The sum of the retinal and choroidal areas (Fig. 4E) was then calculated, and the area (in pixels) of each section was exported to an Excel file (Microsoft Co., Ltd.).

4.1.2. ML model

The random forests algorithm was used in the ML model in this study. The random forests algorithm is an ensemble learning method based on bagging. The input data were sampled randomly using bootstrap and divided into multiple groups. Each group trained with decision trees to parallel, which falls to overfitting. We then averaged the prediction values in each group to prevent overfitting. Data on the retinal and choroidal areas were divided into 100 groups. The depth of the decision trees was set to 5.

4.1.3. Ensemble model

The ensemble model used a soft voting algorithm among the CNN models with ResNet-152, DenseNet-201, EfficientNet-B7 and the ML model.

4.1.4. Statistical analysis

We determined the differences in the retinal and choroidal areas between the abnormal and normal images using the Mann–Whitney $U$ test with Bonferroni correction for each segment.
We used ROC curves and calculated the AUC with 1,000 times bootstrap to estimate the diagnostic performance of the ML, ensemble, and ensemble with ML models. The AUC was compared among the models using Scheffé test.

IBM SPSS Statistics version 26 (IBM Corp., Armonk, NY, USA) was used for statistical analysis, and a $P$-value $< 0.05$ was considered significant.

## 4.2 Results

The sum of the retinal and choroidal areas was significantly thicker in the abnormal eyes than that in the normal eyes in segments 3 and 4 ($P < 0.001$) and was significantly thinner in the abnormal eyes than that in the normal eyes in segment 5 ($P < 0.001$; Fig. 5). The accuracies of each segment as evaluated by the ML were 83% for segment 1 (abnormal 47/50, normal 36/50; Supplementary Fig. 3A), 87% for segment 2 (abnormal 48/50, normal 39/50; Supplementary Fig. 3B), 96% for segment 3 (abnormal 50/50, normal 46/50; Supplementary Fig. 3C), 88% for segment 4 (abnormal 47/50, normal 41/50; Supplementary Fig. 3D), and 89% for segment 5 (abnormal 45/50, normal 44/50; Supplementary Fig. 3E).

The ML, ensemble, and ensemble with ML models showed accuracies of 89% (abnormal 45/50, normal 44/50; Supplementary Table 1), 98.0% (abnormal 50/50, normal 48/50), and 99% (abnormal 50/50, normal 49/50), respectively (Supplementary Fig. 4). The diagnostic performances of the ML, ensemble, and ensemble with ML models were 0.962 (95% confidence interval [CI], 0.922–1.000), 0.998 (95% CI, 0.989–1.000), and 0.999 (95% CI, 0.996–1.000), respectively (Fig. 6). The diagnostic performance was significantly greater in the ensemble model with ML than in both ML and ensemble models ($P < 0.001$).

The ML model prediction was unsuccessful when the retinal and choroidal areas were underestimated during image processing (Fig. 7). The ensemble with the ML model analyzed 0.025 image/s.

## 5. Discussion

This study investigated the diagnostic performances of CNN, ML, and ensemble models using OCT images obtained during eye checkups. The CNN models focused on the structural changes in the retina in abnormal eyes, with accuracy from 95–96% (Figs. 1 and 2). Our finding is consistent with earlier studies that reported a classification accuracy of a single CNN model of about 70–95%; this model may also miss some diseases using OCT images.\textsuperscript{10,11,13,24,25} RP was false-negatively predicted, suggesting insufficient training with abnormalities in the retinal pigment epithelium and photoreceptor layer, including interdigitation and ellipsoid zones in the external limiting membrane. In cases with no apparent edema in the inner retinal layer, our CNN models tended to predict an abnormality based on the peripheral temporal and nasal retinal shapes (Fig. 3). Russakoff et al.\textsuperscript{26} described CNN models trained with OCT images of age-related macular degeneration that focused on the temporal and nasal retinas and differentiated between progressors and non-progressors. These findings suggest that the CNN can detect subtle differences in the morphology of the peripheral temporal and nasal retinal regions to differentiate between abnormal and normal eyes or between progressors and non-progressors.
The ML model used the central macular (segment 3) and nasal (segments 4 and 5) areas as bases for determining eye abnormalities, as a significant difference was found between abnormal and normal retinal areas in each segment (Fig. 5). Of the five diseases that were misdiagnosed by the ML model, the retinal and choroidal areas were underestimated because the morphological transformation did not fill in the inner retinal layer with white in ERM (Figs. 7A and 7A') and macular edema (Figs. 7B, 7B', 7C, and 7C'). Furthermore, the OCT images of RP had small retinal and choroidal areas (Figs. 7D, 7D', 7E, and 7E'). Most diseases were correctly classified by the ML model, although cases with an underestimated macular area were misclassified. Therefore, the features of the retinal and choroidal areas extracted by dividing the OCT images into five segments were useful.

The ensemble model had a significant better diagnostic performance than the single CNN models (Fig. 6). The findings show that the OCT images misclassified by each CNN model varied. Conversely, the ensemble model improved the diagnostic performance for retinal disease. However, the risk of misjudging normal features as abnormal remains.

Nonetheless, the ensemble with the ML model significantly improved that risk. The ML model, which evaluated for disease using the nasal retinal area, thereby improving on the CNN models, which misrecognized the nasal peripheral retinal structures (Figs. 3 and 5). Thus, the ensemble model combining the CNN models trained with OCT images and the ML model trained in the retinal area can improve disease prediction during an eye or health checkup in which only OCT images are acquired. Furthermore, the ensemble with the ML model may be useful to clinicians, given its diagnostic accuracy of 0.999 at 0.025 image/s.

The purpose of screening during an eye checkup is to detect abnormalities. The ensemble model used in our study successfully performed screening using already labeled data. However, we did not evaluate whether our ensemble model could be adapted to anonymous screening data. Moreover, we excluded OCT images in which the ophthalmologists had difficulty determining the disease by reading the images alone. The OCT image quality is higher at the clinic than during the eye checkup because the eye checkup often prioritizes speed and minimizes image processing (e.g., the number of images is decreased to enable averaging). The shift in the dataset is one of the problems encountered in applying artificial intelligence from the laboratory level to actual practice. The performance of artificial intelligence depends on the training dataset. Thus, data distribution (e.g., image quality or unknown disease) differs from the training dataset and, consequently, decreases prediction performance. To prevent a dataset shift in this study, we used single shot OCT images captured during eye checkups. However, because we did not divide into segmention of each retinal layer in the OCT images, the ML model included information about the choroid, and we expected data shifts to occur. Therefore, the accuracy of the ensemble model at actual eye checkups will need to be confirmed in a future investigation.

**Conclusion**
The ensemble with the ML model significantly improved diagnostic performance compared with the CNN and ML models because the ML model, which evaluates diseases using the nasal retinal area, improved on the CNN model, which misrecognized the nasal retinal structure. These findings suggest that the ensemble model combining the CNN model trained with OCT images and the ML model trained with data on the retinal and choroidal areas can be used in eye checkups to accurately predict abnormalities.

Declarations

Acknowledgments

Grants-in-Aid supported this work for Early-Career Scientists, Scientific Research (A) and (B), Challenging Exploratory Research, Japan Society for the Promotion of Science [18H04116 (M.H.), 19K09928 (S.Ueno), 19K09989 (S.Ueno), 19K20728 (M.H.), 19K21783 (M.H.), and 20K04271 (M.H.), respectively]; Charitable Trust Fund for Ophthalmic Research in Commemoration of Santen Pharmaceutical's Founder (M.H.); Takeda Science Foundation (M.H.); Nakatani Foundation (M.H.); and Aichi Health Promotion Foundation (S.U.).

Author contributions

M.H., S.U., Y.I., and T.I. conceived the study and designed the experiments. S.U, T.I., Y.I, and E.W. labeled the dataset. M.H. and A.M. analyzed the data. M.H. and A.M. wrote the manuscript. All authors reviewed the manuscript.

Competing interests.

M. Hirota, (P); S. Ueno, (P); T. Inooka, (P); Y. Ito, (P); H. Takeyama, (P); Y. Inoue, None; E. Watanabe, None; A. Mizuta, (P).

Data Availability

The data that support the findings of this study are available on request from the corresponding author, M.H. upon reasonable request. After the request is accepted, access to the data is temporarily permitted.

References


**Figures**
Figure 1

Diagnostic performances of the ResNet-152 (green), DenseNet-201 (purple), EfficientNet-B7 (blue), and ensemble models (red).

The diagnostic performances of the CNN model with ResNet-152, DenseNet-201, and EfficientNet-B7 and the ensemble model were 0.989 (95% CI, 0.968–1.000), 0.997 (95% CI, 0.986–1.000), 0.997 (95% CI, 0.986–1.000), and 0.998 (95% CI, 0.989–1.000), respectively. CNN, convolutional neural network; CI, confidence interval. The diagnostic performance was significantly greater in the ensemble model than in the other CNN models ($P < 0.001$).
Figure 2

Representative visual explanations of the feature map of the CNN model in the corrected OCT image.

The heat maps in (A), (B), and (C) indicate the relative activation intensity of predicting abnormalities in the OCT images. The CNN model made a decision based on the location of the warm color. The CNN model focused on the retinal structural changes in the abnormal eyes. OCT, optical coherence tomography; CNN, convolutional neural network.
Representative visual explanations of the feature map of the CNN model in the misrecognized OCT images.

The heat maps in (A), (B), (C), (D), and (E) indicate the relative activation intensity of predicting abnormalities in OCT images. The CNN model made a decision based on the location of the warm color. The CNN model classified (A) an abnormal eye (retinitis pigmentosa) as normal and (B–E) a normal eye as abnormal. OCT, optical coherence tomography; CNN, convolutional neural network.

Figure 4

Preprocessing for training the ML model.

All original OCT images (A) had 20% of pixels removed from the right and left edges (red line and red arrows) to avoid (B) depression of the optic disk. (C) The OCT images were divided into five sections, each with a resolution of 123 × 1,176 pixels. The peripheral temporal retina, temporal perimacular area, central macular area, nasal perimacular area, and peripheral nasal retina were defined as segments 1, 2, 3, 4, and 5, respectively. (D) Each section was binarized using the discriminant analysis method. (E) Morphological closing was applied to these segment images to pad the dark area related to the inner retinal layer and choroidal vessels. The sum of the retinal and choroidal areas was then calculated. OCT, optical coherence tomography; ML, machine learning; seg, segment.
The peripheral temporal retina, temporal perimacular area, central macular area, nasal perimacular area, and peripheral nasal retina were defined as segments 1, 2, 3, 4, and 5, respectively. The sum of the retinal and choroidal areas was significantly thicker in abnormal eyes than that in normal eyes in segments 3 and 4. Furthermore, the retinal and choroidal areas were significantly thinner in abnormal eyes than those in normal eyes in segment 5.

seg, segment. ***$P < 0.001$. 

Figure 5

Sum of the retinal and choroidal areas in each segment.
Figure 6

Diagnostic performances of the ML (purple), ensemble (blue), and ensemble with ML (red) models.

The diagnostic performances of the ML, ensemble, and ensemble with ML models were 0.962 (95% CI, 0.922–1.000), 0.998 (95% CI, 0.989–1.000), and 0.999 (95% CI, 0.996–1.000), respectively. ML, machine learning; CI, confidence interval. The diagnostic performance was significantly greater in the ensemble model with ML than in both ML and ensemble models ($P < 0.001$).
Figure 7

Original (A–E) and post-processing (A¢–E¢) OCT images misjudged by the ML model.

The ML model misinterpreted epiretinal membrane (A), macular edema (B and C), and retinitis pigmentosa (D and E). In the OCT images of the epiretinal membrane and macular edema, the inner retinal layer was not filled in with white by morphological transformation. The OCT images of retinitis pigmentosa indicate small retinal and choroidal areas. OCT, optical coherence tomography; ML, machine learning.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx