Automatic Estimation of Objective Cyclodeviation in Fundus Image Using Machine Learning

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

Teikyo University

Megumi Fukushima

Teikyo University

Kakeru Sasaki

Teikyo University

Kanako Kato

Teikyo University

Chie Usui

Teikyo University

Yoshinobu Mizuno

Teikyo University

Takao Hayashi

Teikyo University

Atsushi Mizota

Teikyo University

Research Article

Keywords:

Posted Date: February 23rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1371470/v1>

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

[Read Full License](#)

Abstract

Calculating the angle of objective cyclodeviation has been based on the difference between the fovea and center of the optic nerve head (ONH) in fundus images. An examiner defines the location of the fovea and center of ONH manually; thus, there is an influence of bias, and the analysis time per fundus image is long. Therefore, in this study, we developed a machine learning-based model for automatic cyclodeviation estimation using fundus images. Our model is divided into two steps: detecting macular and ONH regions using the single shot multibox detector (SSD) and estimating cyclodeviation using gradient boosting on decision trees of CatBoost. The SSD detected the macular and ONH with a 75% average precision of 99.7% and 99.4%. Then, CatBoost estimated the objective manual cyclodeviation with a mean absolute error of 0.804. These findings suggest that the machine learning-based algorithm estimate cyclodeviation with the same accuracy as a human from fundus images.

Introduction

Cyclodeviation is defined as the rotation of an eyeball along the optic axis. The directions of cyclodeviation in both eyeballs differ; one rotates with its roll toward the nose, which is defined as incyclodeviation, and the other rotates with its roll toward the temples, which is defined as excyclodeviation.

A few subjective and objective methods have been established for evaluating cyclodeviation. The double Maddox rod and/or Lancaster red–green tests are standard techniques for evaluating subjective cyclodeviation.^{1–3} However, patients with cyclotropia rarely complain of cyclodiplopia⁴ because the fusional amplitude of cyclodeviation is approximately 9°. ^{5,6} Thus, the detectability of cyclodeviation is higher with objective measurements than with the subjective measurements.

The objective evaluation of cyclodeviation employs fundus images. Bixenman and von Noorden established a measuring method for cyclodeviation.⁷ They calculated the angle of cyclodeviation from the difference between the fovea and center of the optic nerve head (ONH). Recently, Simiera et al. developed web-based software to assess objective cyclodeviation.^{8,9} Methods for calculating objective cyclodeviation from fundus images and software to assist in the analysis are valuable for quantitative evaluation of cyclotropia. However, the conventional methods and software define the fovea and center of ONH by an examiner visually checking the fundus images. The manual analysis is affected by bias among examiners. Because the analysis time per fundus image is quite long, it is unsuitable for processing vast data.

Recently, machine learning (ML) has achieved excellent image processing results. ML-based object detection methods can predict the location and types of objects in one image.^{10–12} Further, the processing speed is faster than simpler conventional methods that use raster scans per image.¹³ Therefore, we considered using ML-based object detection to detect the fovea and center of ONH and then calculate objective cyclodeviation.

This study aimed to investigate whether an ML-based method for automatic objective cyclodeviation estimation comprises manual objective cyclodeviation using fundus images.

Methods

1. Data acquisition

Healthy students who belong to the Department of Orthoptics, Faculty of Medical Technology, Teikyo University, participated in this retrospective study between April 2015 and March 2020. Complete ophthalmological examinations were performed for the subjects, including assessing the ONH and macula using optical coherence tomography (OCT-2000, Topcon Corp., Tokyo, Japan), intraocular pressure using a noncontact tonometer (CT-800, Topcon Corp., Tokyo, Japan), visual field using the Humphrey perimeter (Carl Zeiss Meditec), near-point of convergence, stereoscopic acuity at 40 cm (Original Stereo Fly Stereotest; Stereo Optical Co., Inc., Chicago, USA), heterophoria via an alternating cover test for near (33 cm) and distant (5.0 m) assessments, and best-corrected visual acuity using the decimal visual acuity chart at Teikyo University. We applied an *opt-out* method to obtain subject *consent*. This investigation adhered to the World Medical Association Declaration of Helsinki tenets. The Institutional Review Board approved the study of Teikyo University (approval no. 20–021–2).

2. Fundus photography

Color fundus images were captured with a fundus camera (VX-10i using a charge-coupled device of KD-211C, Kowa Co., Ltd., Tokyo, Japan). The examiner used either internal or external fixation. The subjects fixated on the tip of the internal fixation bar in the camera with the eye photographed for internal fixation. For external fixation, the subjects were asked to look at a fixation target before the nonexamined eye, and the examiner moved the fixation target to ensure the examined eye was always at or near the primary position.

All fundus color images were taken with a pupillary diameter of ≥ 6.0 mm after mydriatic and cycloplegic eye drops (Midrin-^P ophthalmic solution, Santen Pharmaceutical Co., Ltd., Osaka, Japan), flashlight brightness of 6.3, 9.5, or 12.5 mW, angle of view of 50°, resolution of 4288 × 2848 pixels, and jpeg compression.

A total of 985 fundus images were captured over two years (Fig. 1). Then, two examiners (M.H. and M.F.) double-checked the image quality. Finally, 573 fundus images were registered.

3. Manual analysis for objective cyclodeviation

In this study, we used Cyclocheck^R (<http://cyclocheck.com/>) to calculate the objective cyclodeviation (hereinafter called “cyclodeviation”) manually.^{8,9} Cyclocheck^R is a web-based application. First, the user selected the eye for calculating cyclodeviation, checking the left and right eye. Second, the user drags and drops the fundus photograph. Third, the user clicks the foveal center and then clicks the top and bottom

of ONH. Cyclocheck^R calculates the cyclodeviation from the horizontal line through the geometric center of ONH and the other line joining the fovea and center of ONH.

M.H. and M.F. double-checked cyclodeviation using Cyclocheck^R. In this study, exocyclodeviation and incyclodeviation were defined as the angles between the fovea, and the center of the ONH was $> 0^\circ$ and $< 0^\circ$, respectively. Among 573 fundus images, including 552 and 21 for exocyclodeviation and incyclodeviation, respectively.

In the analysis, 115 fundus images were considered (the same image as the test data shown below), and the time required to analyze one fundus image was measured using a stopwatch.

4. Preliminary experiment

As a preliminary experiment, we confirmed the estimation performances using a convolutional neural network (CNN) model based on EfficientNet-B7.¹⁴ The fundus images were randomly divided into 458 (80%) and 115 (20%) for training and testing, respectively. The training data were resized to 512×512 pixels. Afterward, data augmentation was performed for the input images: random brightness from 0.8 to 2.0 times, random contrast from 0.8 to 1.5 times, random horizontal and vertical shift within 50 pixels, and random horizontal mirroring (probability = 50%).

In the training phase for transfer learning, supervised learning was used, and the accuracy of estimation was measured as the weights of deep layers were changed. The weights were changed on the basis of an optimization function. In this study, we used the Adam optimizer for the CNN model.¹⁵ The layers of deep neural networks were frozen until immediately before the output layer to use ImageNet weight parameters. We introduced a fully connected layer as an output layer. Using the identity function, the fully connected layer provided cyclodeviation.

The CNN model was trained for 100 epochs. The optimizer used an adaptive learning rate; the primary learning rate was 0.02, which subsequently reduced to 0.5 times at 25%, 50%, 75%, and 90% of the total epochs. The training data were divided into five and cross-validated.

We used the software Python 3.8.5 for Windows 10 (Microsoft Co., Ltd., Redmond, WA, USA), with the following libraries: Matplotlib 3.3.4, Numpy 1.19.2, OpenCV 4.0.1, Pandas 1.2.3, Pytorch 1.10.1, Torchvision 0.11.2, Scikit-learn 0.24.1, and Seaborn 0.10.1.

The performance of the CNN model evaluated the mean absolute error (MAE) between prediction and manual values, which was 6.750 (Fig. 2).

5. Model for macular and ONH detection

5.1 Network

The object detection model employed the single shot multibox detector (SSD).^{11,16} SSD analyzed the input image using the base CNN for extracting the feature map and branch from the base CNN into two other CNNs to predict the object category and location. The SSD output surrounds the range of detected objects within a rectangle called a bounding box and displays the name of the object category and predicted value. We used a trained visual geometry group-16 (VGG16) as the base CNN.^{17,18} Object categories were target and nontarget (background).

Among 573 fundus images, 228 (40%) images were extracted randomly. The area around macular and ONH was annotated using Labellmg (Tzutalin). Then, the dataset was randomly divided into training data of five groups, and the training data were cross-validated.

The fundus images were resized to 300 × 300 pixels; then, data augmentation was applied to the target images. In the training phase, we set the following parameters: 100 epochs, a batch size of 32, and an Adam optimizer with a learning rate of 0.001. The validation was performed every ten epochs.

5.2 Performance of SSD

The SSD model was evaluated by the average precision of the remaining 345 fundus images. The intersection of union (IoU) was calculated by dividing the area of overlap between the predicted and groundtruth bounding boxes, including names (macular and ONH) and locations, as manually defined by M.H, as well as the area of the union of both bounding boxes. We defined an IoU of $\geq 75\%$ as “correct.” (AP_{75}). Subsequently, the results of the test dataset were classified as follows—true positive (TP): the predicted bounding box has been covered with a groundtruth bounding box ($\text{IoU} \geq 75\%$); false positive (FP): the predicted bounding box has been covered with a groundtruth bounding box ($\text{IoU} < 75\%$ and $\text{IoU} \neq 0$); false negative (FN): the predicted bounding box has not been covered with a groundtruth bounding box ($\text{IoU} = 0$); true negative: the predicted and groundtruth bounding boxes did not exist. Precision was defined as the percentage at which the IoU can be predicted correctly with an accuracy of $\geq 75\%$, given by $TP/(TP + FP)$. Recall was defined as the percentage at which the ratio of the bounding box at apposition can be accurately predicted if the IoU is $\geq 75\%$, given by $TP/(TP + FN)$. The AP_{75} is given by integrating Precision and Recall.

5.3 Calculating fovea and center of ONH

The SSD involves drawing an object's location within a bounding box (Fig. 3A). The bounding box was computed from two coordinates of $(X_{\min}, Y_{\min})$ and $(X_{\max}, Y_{\max})$. The center of the object coordinates was determined by $[(X_{\max} + X_{\min})/2, (Y_{\max} + Y_{\min})/2]$. Thus, the fovea (F_X, F_Y) and center of ONH (ONH_X, ONH_Y) location were defined as the centers of the bounding box (Fig. 3B), and a program output the locations to an Excel file (Microsoft Co., Ltd.) in the inference phase.

First, the difference between the fovea and center of ONH [$\text{abs}(F_X - ONH_X), (F_Y - ONH_Y)$] were calculated (Fig. 3B). Then, the temporary cyclodeviation was calculated following as $\arctan (F_Y - ONH_Y) / \text{abs}(F_X -$

ONH_x). Horizontal differences were calculated in absolute values. Meanwhile, absolute values of vertical differences were not taken because plus and minus signs indicate excyclodeviation and incyclodeviation, respectively.

Three hundred and forty-five fundus images not used for training and validation of SSD were calculated for the temporary cyclodeviation.

6. Estimating cyclodeviation

For estimating cyclodeviation, we used XGBoost,¹⁹ LightGBM,²⁰ and CatBoost,²¹ which are ML-based models, and an algorithm for gradient boosting on decision trees (GBDT). The GBDT uses a decision tree (DT) as a detector, learns the prediction error of the previous DT as the prediction of the subsequent DT, and finally makes inferences by adding the first prediction result and the prediction error of each DT weight.

The tree growing algorithm differs among XGBoost, LightBGM, and CatBoost. XGBoost grows a branch to a specified depth and computes with all data. LightGBM grows the branches to a specified depth, although the developed tree is unbalanced. Random sampling is used for branches with few errors. CatBoost grows the tree to the specified depth; the structure of the CatBoost tree is in bilateral symmetry, with branching equations at each depth applying the same value. CatBoost calculates the probability according to each branch's error for each data point.

Among 345 temporary cyclodeviation values, 230 (66.7%) were randomly extracted for training and validation data, and the remaining 115 (33.3%) were used as test data. Manual analysis data obtained from Cyclocheck^R was used as the supervised data. The optimal parameters in GBDTs were detected by grid-search using five-fold cross-validation.

Ensemble learning, which creates and integrates several different classifiers, improves a model's accuracy.²² An earlier study reported that ensemble learning is valuable, even if the dataset is small.²³ An ensemble model was created using a soft voting algorithm that averages the predictions of the three ML-based models.

The performance of XGBoost, LightGBM, CatBoost, and ensemble model was evaluated using the MAE between prediction and manual values. The automatic cyclodeviation was defined as the model with the lowest MAE.

7. Statistical analysis

Cyclodeviations between the temporary or automatic cyclodeviation and manual cyclodeviation were analyzed by performing a Bland–Altman analysis.^{24,25} The fixed and proportional biases between the two cyclodeviations were analyzed using the paired t-test and simple linear regression analysis.

The difference in the time required for analyzing cyclodeviation between automatic and manual cyclodeviation was obtained via paired t-test.

IBM SPSS Statistics version 26 (IBM Corp., Armonk, NY, USA) was used to determine the significance of the differences, and a P -value of <0.05 was considered significant.

Results

1. Performance of ML-based models

For the SSD model, the AP_{75} for macular and ONH was 99.7% and 99.4%, respectively.

The MAEs in XGBoost, LightGBM, CatBoost, and the ensemble model were 0.887° , 0.899° , 0.804° , and 0.824° , respectively (Fig. 4). Thus, CatBoost was considered the automatic cyclodeviation.

2. Comparison Between Temporary/automatic Cyclodeviation And Manual Cyclodeviation

The temporary cyclodeviation was significantly greater ($7.63^\circ \pm 3.48^\circ$) than the manual ones ($6.36^\circ \pm 3.58^\circ$) ($P < 0.001$) (Fig. 5A). The proportional bias did not differ between temporary and manual cyclodeviations ($R^2 = 0.008$, $P = 0.35$) (Fig. 5B). The mean value of the differences between the temporary and manual cyclodeviations was $1.271^\circ \pm 1.032^\circ$. The 95% limits of agreement ranged from -0.751° to 3.294° .

The fixed ($P = 0.73$) and proportional ($R^2 = 0.034$, $P = 0.051$) biases between automatic ($6.34^\circ \pm 3.43^\circ$) and manual cyclodeviations did not differ (Fig. 6). The mean value of the differences between the automatic and manual cyclodeviation was $-0.036^\circ \pm 1.097^\circ$. The 95% limits of agreement ranged from -2.185° to 2.114° .

The time required for analyzing cyclodeviation was significantly shorter in automatic cyclodeviation (0.443 ± 0.115 s) than the time required in manual cyclodeviation (15.798 ± 1.668 s) ($P < 0.001$).

Discussion

In this study, we developed an ML-based model for automatic cyclodeviation estimation. The automatic cyclodeviation calculated through the SSD and CatBoost did not have any systematic error compared with conventional manual cyclodeviation. Our findings suggest that combining ML—which is outstanding for image recognition—and GBDT—which is outstanding for numerical prediction—can estimate cyclodeviation with the same accuracy as a human from fundus images.

The single task using a CNN did not estimate cyclodeviation from fundus images (Fig. 2). Earlier studies reported that CNN models were highly accurate in classifying retinal disease from fundus images.^{26–28} Some earlier studies visualized the focus of CNN using the gradient-weighted class activation mapping (Grad-CAM).^{29–31} The CNNs classify images by finding differences from controls. However, the angle between the fovea and the center of the ONH from 4° to 9° exists in healthy individuals.⁹ CNN could not learn the features of cyclodeviation from fundus images because the control data was nothing.

In supervised learning using CNNs, the efficiency is improved by having a human specify the training target prior. Because cyclodeviation can be calculated based on the central positions of the fovea and ONH, we used the SSD to learn the macula and ONH preferentially. The SSD detected macular and ONH regions with high accuracy. However, the temporary cyclodeviation defined as the difference between the macular and ONH region centers was significantly greater than the manual cyclodeviation (Fig. 5). By contrast, the proposal bias was not found between temporary and manual cyclodeviations. These findings suggest that the overestimated cyclodeviation (fixation bias) occurred by the compression of fundus images from 4288 × 2848 to 300 × 300 pixels in the image processing. Cyclocheck has been based on two lines between the foveal center and the top or bottom of ONH.⁸ ONH coordinates center, the user-manually defined line connecting the top and bottom. This method can approximate the center of ONH.

The automatic cyclodeviation using ML models corresponded with manual cyclodeviation (Figs. 4 and 6), suggesting that the SSD alone overestimates cyclodeviation, whereas combining GBDTs provides an accurate prediction. Although the GBDTs in XGBoost, LightGBM, and CatBoost have almost the same performance, CatBoost was the best model in this study. An ensemble model with soft voting has often been used to improve a model's accuracy. However, the ensemble model was not the best in this study because the MAE of XGBoost and LightGBM were higher than that of CatBoost. In particular, the accuracy of incyclodeviation was low. We consider that the sample size was too small to identify incyclodeviation using the GBDTs. We acquired 573 fundus images. However, the subjects with incyclodeviation were only 21. This finding is consistent with the earlier studies that reported that the population is greater in excyclodeviation than incyclodeviation.^{3,7,32} Therefore, we consider that it may better to apply our model only to the excyclodeviation until enough subjects with incyclodeviation are accumulated. Our model can detect between excyclodeviation and incyclodeviation from the plus and minus signs calculated from the difference between the fovea and ONH.

The most significant advantage of automated analysis is that the time required for analyzing a single fundus image dramatically reduces. The SSD and CatBoost models can automatically analyze the amount of cyclodeviation around a clock in the fundus images stored in the working directory. The fundus images that fail to be detected as macular or ONH by SSD will be detected by relearning the model. Therefore, the accuracy is expected to improve as the sample size increases. Most ophthalmic hospitals have a fundus camera. We believe that our model is helpful for daily clinical practice.

In addition to the insufficient sample size of subjects with incyclodeviation, our method has other limitations. Our model's estimated objective cyclodeviation agreed with the manual cyclodeviation using Cyclocheck. Subjective cyclodeviation is used in strabismological surgery for cyclotropia to prevent diplopia after surgery, not objective cyclodeviation. This retrospective study established the basic algorithm for automatically estimating objective cyclodeviation from fundus images. In future studies, we will update the model to estimate subjective cyclodeviation from fundus images.

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 (19K20728 [M.H.], 18H04116 [M.H.], 20K04271 [M.H.], and 19K21783 [Y.I. and 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 for Advancement of Measuring Technology in Biomedical Engineering (M.H.); Japan Medical Herb Association Research Grant Program (M.H.).

Author contributions

M.H. conceived the project and designed the experiments. M.H. produced the apparatus. M.H. and M.F. performed experiments. M.H., M.F., K.S., K.K., C.U., Y.M., T.H., and A.M. analyzed the data. M.H., T.H., and A.M. wrote the manuscript. All authors reviewed the manuscript.

Competing interests.

All authors have no competing interests.

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

1. von Noorden, G. K. Clinical and theoretical aspects of cyclotropia. *J. Pediatr. Ophthalmol. Strabismus* **21**, 126–132 (1984).
2. Liebermann, L., Leske, D. A., Hatt, S. R. & Holmes, J. M. Test-retest variability of cyclodeviations measured using the double Maddox rod test. *Journal of American Association for Pediatric Ophthalmology and Strabismus* **22**, 146–148.e141, doi:10.1016/j.jaapos.2017.09.010 (2018).
3. Woo, S. J., Seo, J. M. & Hwang, J. M. Clinical characteristics of cyclodeviation. *Eye* **19**, 873–878, doi:10.1038/sj.eye.6701675 (2005).

4. von Noorden, G. K. & Campos, E. C. *Binocular Vision and Ocular Motility: Theory and Management of Strabismus*. 6 edn, 389 (Mosby, 2002).
5. Sen, D. K., Singh, B. & Mathur, G. P. Torsional fusional vergences and assessment of cyclodeviation by synoptophore method. *Br. J. Ophthalmol.* **64**, 354–357, doi:10.1136/bjo.64.5.354 (1980).
6. Baskaran, A. A., Britto, T., Rajkumar, S., Thomas, P. A. & Jesudasan, C. A. N. Comparison of torsional amplitudes between emmetropes and myopes using after-image slides. *Indian J. Ophthalmol.* **67**, 655–658, doi:10.4103/ijo.IJO_1663_18 (2019).
7. Bixenman, W. W. & von Noorden, G. K. Apparent foveal displacement in normal subjects and in cyclotropia. *Ophthalmology* **89**, 58–62, doi:10.1016/s0161-6420(82)34862-9 (1982).
8. Simiera, J. & Loba, P. Cyclocheck: a new web-based software for the assessment of objective cyclodeviation. *Journal of American Association for Pediatric Ophthalmology and Strabismus* **21**, 305–308, doi:10.1016/j.jaapos.2017.02.009 (2017).
9. Simiera, J., Ordon, A. J. & Loba, P. Objective cyclodeviation measurement in normal subjects by means of Cyclocheck (R) application. *Eur. J. Ophthalmol.* **31**, 704–708, doi:10.1177/1120672120905312 (2021).
10. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv e-prints*, arXiv:1506.02640 (2015).
<<https://ui.adsabs.harvard.edu/abs/2015arXiv150602640R>>.
11. Liu, W. *et al.* SSD: Single Shot MultiBox Detector. *arXiv e-prints*, arXiv:1512.02325 (2015).
<<https://ui.adsabs.harvard.edu/abs/2015arXiv151202325L>>.
12. Ren, S. Q., He, K. M., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, 1137–1149, doi:10.1109/Tpami.2016.2577031 (2017).
13. Cho, H., Rybski, P. E., Bar-Hillel, A. & Zhang, W. Real-time pedestrian detection with deformable part models. 2012 *IEEE Intelligent Vehicles Symposium* **2012**, 1035–1042, doi:10.1109/IVS.2012.6232264. (2012).
14. Tan, M. & Quoc. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv pre-print server*, doi:arxiv:1905.11946 (2020).
15. Diederik & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv pre-print server*, doi:arxiv:1412.6980 (2017).
16. Hirota, M., Hayashi, T., Watanabe, E., Inoue, Y. & Mizota, A. Automatic Recording of the Target Location During Smooth Pursuit Eye Movement Testing Using Video-Oculography and Deep Learning-Based Object Detection. *Translational Vision Science & Technology* **10**, 1, doi:10.1167/tvst.10.6.1 (2021).
17. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *arXiv pre-print server*, arXiv:1512.03385 (2015).
18. Szegedy, C., Vanhoucke, V., Loffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv*, arXiv:1512.00567 (2015).

19. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. arXiv pre-print server, doi:10.1145/2939672.2939785 (2016).
20. Ke, G. *et al.* LightGBM: a highly efficient gradient boosting decision tree. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 3149–3157, doi:10.5555/3294996.3295074 (2017).
21. Prokhorenkova, L., Gusev, G., Vorobev, A., Anna & Gulin, A. CatBoost: unbiased boosting with categorical features. arXiv pre-print server, doi:arxiv:1706.09516 (2019).
22. Rokach, L. Ensemble-based classifiers. *Artificial Intelligence Review* **33**, 1–39, doi:10.1007/s10462-009-9124-7 (2010).
23. Phung, V. H. & Rhee, E. J. A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. *Applied Sciences* **9**, 4500, doi:10.3390/app9214500 (2019).
24. Altman, D. G. & Bland, J. M. Measurement in Medicine - the Analysis of Method Comparison Studies. *Statistician* **32**, 307–317 (1983).
25. Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **8**, 135–160 (1999).
26. Li, B. *et al.* Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography. *Br. J. Ophthalmol.*, doi:10.1136/bjophthalmol-2020-316290 (2021).
27. Tang, M. C. S., Teoh, S. S., Ibrahim, H. & Embong, Z. Neovascularization Detection and Localization in Fundus Images Using Deep Learning. *Sensors-Basel* **21**, 5327, doi:10.3390/s21165327 (2021).
28. Kang, E. Y.-C. *et al.* Deep Learning–Based Detection of Early Renal Function Impairment Using Retinal Fundus Images: Model Development and Validation. *JMIR Medical Informatics* **8**, e23472, doi:10.2196/23472 (2020).
29. Selvaraju, R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. arXiv pre-print server, arXiv:1610.02391v02393 (2017).
30. Hirota, M. *et al.* Effect of color information on the diagnostic performance of glaucoma in deep learning using few fundus images. *Int. Ophthalmol.* **40**, 3013–3022, doi:10.1007/s10792-020-01485-3 (2020).
31. Kim, K. M. *et al.* Development of a Fundus Image-Based Deep Learning Diagnostic Tool for Various Retinal Diseases. *Journal of Personalized Medicine* **11**, 321, doi:10.3390/jpm11050321 (2021).
32. Miyata, M. *et al.* Age-related change and sex difference over 60s in disc-fovea angle in Japanese population: the Nagahama Study. *Acta Ophthalmol. (Copenh.)* **96**, E840-E845, doi:10.1111/aos.13642 (2018).

Figures

Figure 1

Diagrams showing an overview for estimating cyclodeviation automatically.

Excyc.: excyclodeviation; Incyc.: incyclodeviation; ONH: optic nerve head; SSD: single shot multibox detector; Train.: training; Val.: validation; AP₇₅: 75% average precision; GBDT, gradient boosting on decision trees.

Figure 2

Preliminary experimental results

The gray dashed line indicates the equation line. The mean absolute error of the CNN was 6.750° . A single CNN model could hardly estimate cyclodeviation.

CNN: convolutional neural network.

Figure 3

Calculating for temporary cyclodeviation using bounding box with SSD.

A) The cobalt blue and sky blue squares indicate the bounding boxes that SSD detected. SSD detected the fundus image's ONH (cobalt blue) and macular (sky blue).

B) The image's top-left corner is the origin (0.0, 0.0), and the bottom-right corner is the maximum value. Thus, the bounding box coordinates take the minimum and maximum values at the top-left and bottom-right, respectively. The difference between the fovea and center of ONH was defined by $[abs(F_x - ONH_x), (F_y - ONH_y)]$. Then, the temporary cyclodeviation was calculated following using $\arctan (F_y - ONH_y) / abs(F_x - ONH_x)$.

SSD: single-shot multibox detector; ONH: optic nerve head; F: fovea; abs: absolute value.

Figure 4

Performance of machine learning models.

The purple, cobalt blue, red, and sky blue dots indicate XGBoost, LightGBM, CatBoost, and an ensemble model. The gray dashed line indicates the equation line. Catboost has the lowest MAE.

MAE: mean absolute error.

Figure 5

Comparison between temporary and manual cyclodeviations.

A) The red and cobalt blue crosses indicate the mean value. The cyclodeviation was significantly greater in temporary analysis than in manual analysis.

B) The red solid and dashed lines indicate the mean and 95% limits of agreement. The mean value of the differences between the temporary and manual cyclodeviation was $1.271^\circ \pm 1.032^\circ$. The 95% limits of agreement ranged from -0.751° to 3.294° .

***: $P < 0.001$. Paired t-test

Figure 6

Comparison between automatic and manual cyclodeviations.

A) The red and cobalt blue crosses indicate the mean value. The cyclodeviation was not significantly different between the automatic and manual analyses.

B) The red solid and dashed lines indicate the mean and 95% limits of agreement. The mean value of the differences between the temporary and manual cyclodeviation was $-0.036^\circ \pm 1.097^\circ$. The 95% limits of agreement ranged from -2.185° to 2.114° .