Graefe's Archive for Clinical and Experimental Ophthalmology (2022) 260:1329–1335
https://doi.org/10.1007/s00417-021-05475-8

OCULOPLASTICS AND ORBIT

# Developing an iOS application that uses machine learning for the automated diagnosis of blepharoptosis

**Hitoshi Tabuchi[1,2] · Daisuke Nagasato[1,2] · Hiroki Masumoto[1,2] · Mao Tanabe[2] · Naofumi Ishitobi[1] · Hiroki Ochi[3] · Yoshie Shimizu[2] · Yoshiaki Kiuchi[4]**

Received: 13 September 2021 / Revised: 15 October 2021 / Accepted: 21 October 2021 / Published online: 4 November 2021


## Abstract

**Purpose** To assess the performance of artificial intelligence in the automated classification of images taken with a tablet device of patients with blepharoptosis and subjects with normal eyelid.

**Methods** This is a prospective and observational study. A total of 1276 eyelid images (624 images from 347 blepharoptosis cases and 652 images from 367 normal controls) from 606 participants were analyzed. In order to obtain a sufficient number of images for analysis, 1 to 4 eyelid images were obtained from each participant. We developed a model by fully retraining the pre-trained MobileNetV2 convolutional neural network. Subsequently, we verified whether the automatic diagnosis of blepharoptosis was possible using the images. In addition, we visualized how the model captured the features of the test data with Score-CAM. $k$-fold cross-validation ($k = 5$) was adopted for splitting the training and validation. Sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic curve for detecting blepharoptosis were examined.

**Results** We found the model had a sensitivity of 83.0% (95% confidence interval [CI], 79.8–85.9) and a specificity of 82.5% (95% CI, 79.4–85.4). The accuracy of the validation data was 82.8%, and the AUC was 0.900 (95% CI, 0.882–0.917).

**Conclusion** Artificial intelligence was able to classify with high accuracy images of blepharoptosis and normal eyelids taken using a tablet device. Thus, the diagnosis of blepharoptosis with a tablet device is possible at a high level of accuracy.

**Trial registration** Date of registration: 2021–06-25.
Trial registration number: UMIN000044660.
Registration site: https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000051004

**Keywords** Artificial intelligence · Automatic diagnosis · Blepharoptosis · Convolutional neural network · Tablet device

✉ Daisuke Nagasato
d.nagasato@tsukazaki-eye.net

[1] Department of Technology and Design Thinking for Medicine, Hiroshima University, Hiroshima, Japan

[2] Department of Ophthalmology, Saneikai Tsukazaki Hospital, 68-1 Waku, Aboshi-ku, Himeji City, Hyogo 671-1227, Japan

[3] Department of Medicine, Hiroshima University, Hiroshima, Japan

[4] Department of Ophthalmology and Visual Sciences, Hiroshima University, Hiroshima, Japan

**Key messages**

**What was known before**

- Supervised machine learning systems known as neural networks have been gaining attention in recent years owing to the mass availability of imaging data in the medical domain.
- Artificial intelligence can diagnose blepharoptosis from clinical photographs taken with a regular digital camera.

**What this study adds**

- The diagnosis of blepharoptosis with a tablet device is possible at a high level of accuracy.
- Our application may be useful for screening tests and lead to early diagnosis and treatment by a doctor who can determine the need for medical examination or hospital referral.

## Introduction

Blepharoptosis is an eye condition characterized by the drooping of the upper eyelid to eventually cover the eye, causing visual field impairment and visual dysfunction. This can result in visual, functional, and cosmetic problems for patients. Blepharoptosis can occur in individuals of all ages and can result from a variety of causes. Generally, blepharoptosis is further classified as either congenital or acquired, depending on the age at which symptoms appear. Among the acquired forms, age-related blepharoptosis is the most frequent. However, even in young individuals, blepharoptosis may develop secondary to trauma, eye surgery, or long-term contact lens wear [1]. Blepharoptosis is mainly diagnosed using six clinical measurements: palpebral fissure height, marginal reflex distance-1, upper eyelid crease, ocular surface area, eyebrow position, and a levator function test [2–4].

Supervised machine learning systems, known as neural networks [5, 6], have been gaining attention in recent years due to the mass availability of imaging data in the medical domain. In ophthalmology, the use of deep neural networks (DNNs) has been validated in reports on the prediction of cardiovascular risk factors [7], detection of diabetic retinopathy [8] from retinal fundus photographs, prediction of age and brachial-ankle pulse-wave velocity using ultra-wide-field pseudo-color images [9], and detection of glaucoma, age-related macular degeneration, and retinal detachment [10–13]. The advantage of using deep learning systems for diagnosis and judgment lies in their adaptability. By using convolutional layers, rendering a diagnosis is possible even with a slight noise in the images used [14–16]. Hung et al. [5] reported that artificial intelligence (AI) can diagnose blepharoptosis from clinical photographs taken with a regular digital camera. However, to date, no research has identified the role of machine learning for use in classifying blepharoptosis with a tablet device.

This study aimed to investigate the ability of AI to distinguish between images of blepharoptosis and images of normal eyelids from facial photographs taken with an iPad Mini 5.

## Materials and methods

### Study design

This study was performed following the tenets of the Declaration of Helsinki, and the study was approved by the institutional review board of Saneikai Tsukazaki Hospital. For the model training and the validation, we used images of eyelids taken with an iPad mini 5 at Saneikai Tsukazaki Hospital between October 18, 2017, and August 8, 2018. We obtained written consent from all patients. This study has been registered with the University Hospital Medical Network clinical trials registry under the title "Developing an iOS application that uses machine learning for the automated diagnosis of blepharoptosis, UMIN000044660."

We obtained 1519 eyelid images from 681 adult subjects. In order to obtain a sufficient number of images for analysis, 1 to 4 images were obtained from each participant. When using multiple images from a participant, images taken on different days were used. An ophthalmologist reviewed the images and excluded those with severe eyelid swelling, congenital blepharoptosis, traumatic blepharoptosis, severe facial nerve palsy, and myopathies. These images were classified according to the specialist's certainty factor into the following five categories: 0, definitely no blepharoptosis; 1,

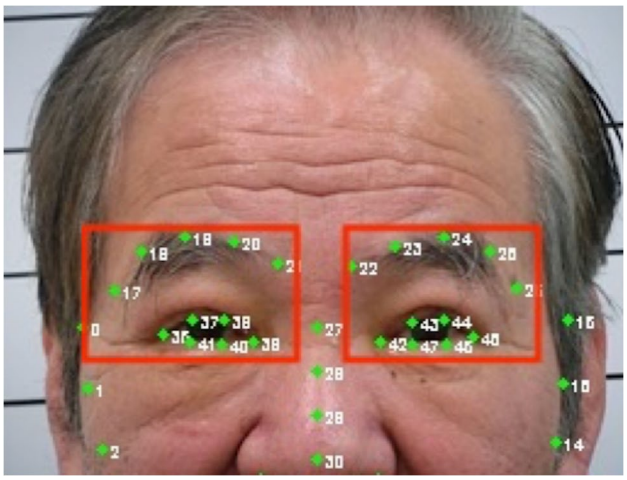

**Fig. 1** Method of cropping the eyelid area from a facial image. Using OpenCV, coordinates (green points with number) on the images of patients' faces were marked and the eyelid area (inside the red line, two locations in this subject) was cropped

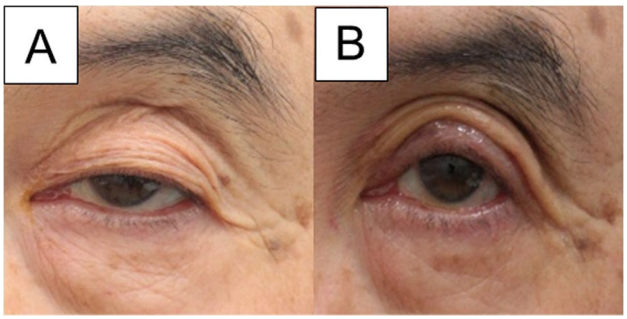


**Fig. 2** Representative eyelid images of **a** blepharoptosis and **b** normal eyelid

unlikely blepharoptosis; 2, unable to determine; 3, probable blepharoptosis; 4, definite blepharoptosis. Then, to increase the accuracy of the model, categories 3 and 4 were reclassified into the "blepharoptosis" group, whereas categories 0 and 1 were reclassified into the "normal eyelid" group. Furthermore, the images in category 2 were excluded. After the reclassification, 1276 eyelid images from 714 subjects (blepharoptosis group: 624 images from 347 subjects; normal eyelid group: 652 images from 367 subjects) were analyzed.

## Image acquisition and library

All facial photographs were taken with an iPad Mini 5. The eyelid area was created using an image processing library. We used the dlib machine learning image processing library to create images of the eyelid area before setting coordinates on the images of patients' faces and cropping the eyelid area using OpenCV (Fig. 1). The method for cropping the eyelid area is shown in the supplementary file (Online Resource: Python code). All images for this study were converted to 224 × 224 pixels in advance and read in 8-bit RGB color with three-channel tensors. The input was normalized to a range of 0–1 by dividing it by 255.

We adopted $k$-fold cross-validation ($k = 5$) [17, 18] for splitting the training and validation data because the number of images was too small for a single training/validation split. Representative blepharoptosis and normal eyelid images used for the deep learning model and training are shown in Fig. 2.

## The neural network

MobileNetV2 is a relatively simple network that is composed of a bottleneck depth-separable convolution with residuals. Its architecture has been described previously

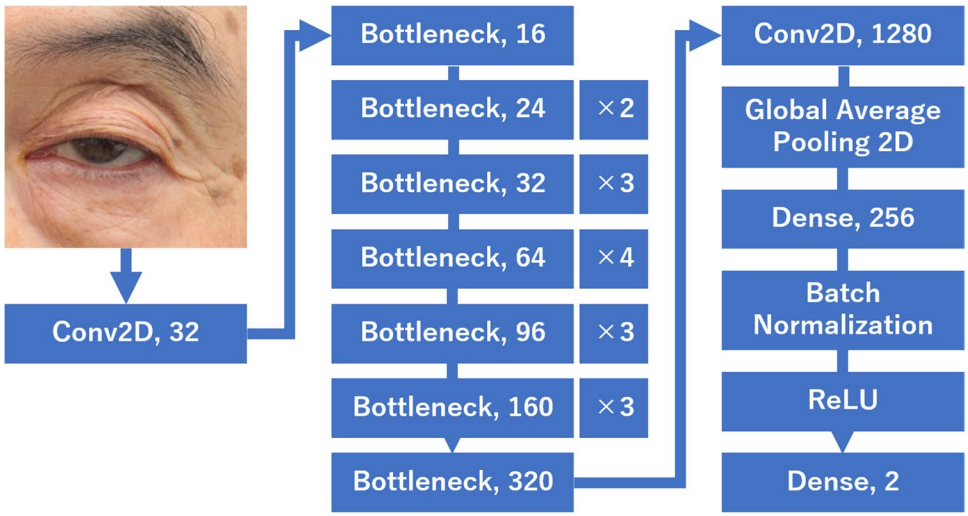


**Fig. 3** Overview of the MobileNetV2 model

[19]. Because it is memory efficient, MobileNetV2 allows for high-speed inference, even on mobile applications.

The global average pooling layer was placed at the end of each block. After flattening the three-dimensional matrix, two layers of the fully connected layer were arranged and classified into two classes using a softmax function. The fully connected layer was used to remove spatial information from extracted features and to distinguish the target from other feature vectors statistically [20]. The model is shown in Fig. 3.

The fully retrained method used in this study relied on parameters that were learned with different images. Strength of this approach is that even small amounts of data could be learned efficiently in a short time [21]. We used ImageNet for the network to learn the initial values. Keras (https://keras.io/en/), which runs Python's TensorFlow backend (https://www.tensorflow.org/), was used to construct and validate the model.

## Heat map

We produced a heat map to use the Score-weighted Class Activation Mapping (Score-CAM) method to visualize how the convolutional neural network model captured the features from the test data [22].

## Outcomes

The area under the curve (AUC), sensitivity, and specificity were determined for blepharoptosis images using the DNN model described above for the validation data.

Images judged to exceed a threshold were defined as positive for blepharoptosis, and a receiver operating characteristic (ROC) curve was generated.

The DNN model was created to return the probability of blepharoptosis. For sensitivity and specificity, we considered results over 83.6% to be positive for blepharoptosis. This threshold was calculated using the Youden index for the ROC curve. The ROC curve was generated using Python scikit-learn (http://scikit-learn.org/stable/tutorial/index.html).

## Statistical analysis

Patients' backgrounds were compared based on age using Student's *t*-test. Fisher's exact test was performed to compare male–female ratios. In all cases, $p < 0.05$ was considered statistically significant. All statistical analyses were performed using Python SciPy (https://www.scipy.org/) and Python Statsmodels (http://www.statsmodels.org/).

For AUC, the 95% confidential interval (CI) was obtained using the following formula, as described previously [23]:

$$95\%\ \mathrm{CI} = A \pm 1.96SE(A),$$

where $A$ means AUC, and $SE(A)$ means the standard error of the AUC.

$SE(A)$ was determined using the formula below [24]:

$$SE(A) = \sqrt{\frac{A(1-A) + (Np-1)\left(Q1 - A^2\right) + (Nn-1)(Q2 - A^2)}{Np \cdot Nn}},$$

where $Np$ is the number of blepharoptosis images; $Nn$ is the number of normal images; $Q1$ is the probability two randomly chosen abnormal images will both be ranked with greater suspicion than a randomly chosen normal image; and $Q2$ is the probability one randomly chosen abnormal image will be ranked with greater suspicion than two randomly chosen normal images.

$Q1$ and $Q2$ were derived using the following formulas [24]:

$$Q1 = \frac{A}{2-A}, Q2 = \frac{2A^2}{1+A}$$

For sensitivity and specificity, a 95% confidential interval (CI) was obtained using the Clopper–Pearson method [25]:

$$\begin{aligned}\text{Clopper} - \text{Pearson CI}(k,n) &= \frac{k}{(n-k+1)F_{0.025}(2(n-k+1),2k)+k} \\ &\sim \frac{(k+1)F_{0.025}(2(k+1),2(n-k))}{(k+1)F_{0.025}(2(k+1),2(n-k))+n-k},\end{aligned}$$

where $F_{0.025}(a,b)$ is the 0.025 quantile from an F-distribution with $(a, b)$ degrees of freedom; $k$ is the number of successes; and $n$ is the number of trials.

The CIs of AUC, sensitivity, and specificity were calculated using SciPy.

**Table 1** Background characteristics of study subjects

| | Total | Blepharoptosis | Normal eyelid | *p* value |
|---|---|---|---|---|
| Number of images | 1276 | 624 | 652 | |
| Number of subjects | 714 | 347 | 367 | |
| Age | 71.5 ± 9.3 (22–100) | 73.2 ± 8.9 (33–100) | 70.0 ± 9.6 (22–92) | $p < 0.001$ (Student's *t*-test) |
| Sex, female | 707 (55.4%) | 333 (53.4%) | 374 (57.4%) | $p = 0.72$ (Fisher's exact test) |
| Eye, left | 640 (50.2%) | 315 (50.5%) | 325 (49.8%) | $p = 1.00$ (Fisher's exact test) |

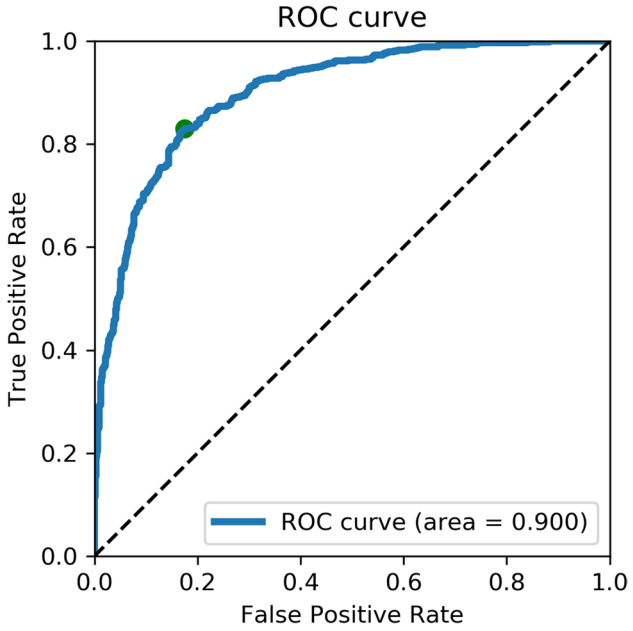


**Fig. 4** ROC curve for the validation data

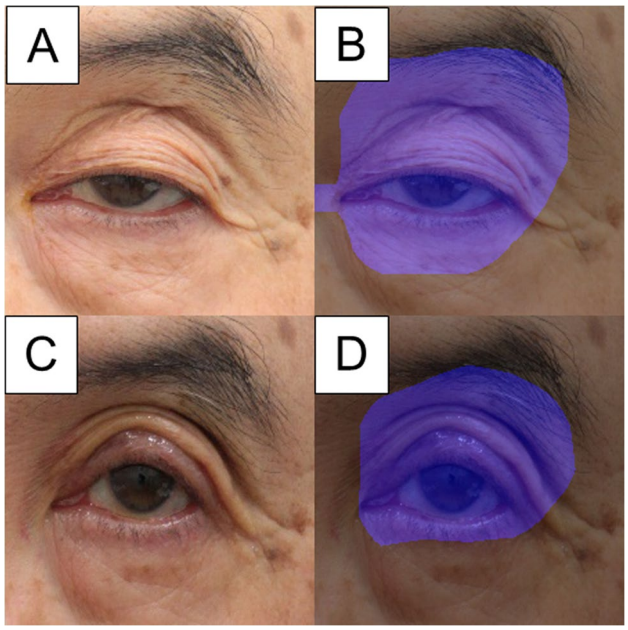


**Fig. 5** Representative eye images of blepharoptosis and normal eyelid and their respective heat maps. **a** Blepharoptosis image. **b** Heat map of blepharoptosis image. **c** Normal eyelid image. **d** Heat map of normal eyelid image

## Results

### Patient characteristics

The patients' backgrounds are described in Table 1. No significant differences were observed in sex, and in the left and right eyes between the "blepharoptosis" and the "normal eyelid" groups ($p = 0.72$ and $p = 1.00$, respectively; Fisher's exact test). However, significant differences were found in age ($p < 0.001$; Student's $t$-test).

### Evaluation of model performance

We found the model had a sensitivity of 83.0% (95% CI, 79.8–85.9) and a specificity of 82.5% (95% CI, 79.4–85.4). The accuracy of the validation data was 82.8% and the AUC was 0.900 (95% CI, 0.882–0.917). The ROC curves for the validation data are shown in Fig. 4.

### Heat map

Figure 5 shows a heat map of a representative blepharoptosis image. We found that the convolutional neural network correctly analyzed eyelids and the dropping of the corners of the eye, as are evaluated during a doctor's diagnosis.

## Discussion

Our AI model was able to identify blepharoptosis from normal eyelids with high accuracy using images of the eyelid areas from facial images. A general point of concern in the use of AI is that AI and ophthalmologists differ in the lesions they focus on when classifying images into blepharoptosis and normal eyelids. However, as shown in the heat maps in our study, our AI model correctly focused on eyelids and the dropping of the corners of the eye, which is consistent with the areas on which ophthalmologists focus when diagnosing blepharoptosis. Therefore, these results indicate that the DNN correctly identified the blepharoptosis lesions in eyelid images and distinguished them from normal eyelid images. However, comparing the diagnostic performance between ophthalmologists and AI is challenging. Liu et al. [23] conducted a systematic review and meta-analysis to examine the diagnostic accuracy of deep learning algorithms for medical imaging and compared it with that of healthcare professionals. They found an average sensitivity of 87.0% (95% CI, 83.0–90.2) and an average specificity of 92.5% (95% CI, 85.1–96.4) for AI, whereas the healthcare professionals had a sensitivity of 86.4% (95% CI, 79.9–91.0) and a specificity of 90.5% (95% CI, 80.6–95.7). These

results suggest that the diagnostic performance of AI was equivalent to that of healthcare professionals. However, Liu et al. [23] pointed that high-quality papers comparing performance between AI and healthcare professionals are limited. Thus, an established comparison method does not currently exist. Therefore, the diagnostic performance of AI can be improved by increasing the number of images for training and validation, as well as by improving AI algorithms.

If blepharoptosis can be shown through our application, a user may be able to refer to objective findings. When subjects photograph their face using a mobile phone, the camera angle may not provide accurate images of the eyelid opening state compared with the diagnostic accuracy of diagnosing blepharoptosis in hospitals; hence, the diagnostic accuracy of diagnosing blepharoptosis using mobile phones is expected to decrease. However, our application, which is readily available, may be useful for screening tests and lead to early diagnosis and treatment by a doctor who can determine the need for medical examination or hospital referral.

Several limitations should be noted when interpreting the results of this study. First, to evaluate the performance of AI for the detection of blepharoptosis, the use of high-resolution images of blepharoptosis and normal eyelid is needed. Second, we did not generalize the findings to other populations or types of ocular diseases. Diagnosis by neural network is a black box [26], but the performance exceeds the capability of human beings. In future verification studies, we need to generalize to other types of ocular diseases or populations, and we may need to consider a method of creating a neural network for calculating the marginal reflex distance and determining blepharoptosis from the target area by neural network segmentation. However, the next concern would be the construction of such a neural network in a manner that coincides with the human methods of identification and subsequent diagnosis. Third, the device was still not able to identity or falsely mark around 15%. Finally, the device has been so far tested only in Asian Eyelids, where ptosis may have several different characteristics as compared to Caucasian eyelids.

AI can be used to diagnose with high accuracy images of blepharoptosis of adult subjects taken with an iPad Mini 5. In the future, we aim to expand the usability of our model by further iteration of the methodology to more closely mimic the diagnostic methods used by humans.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00417-021-05475-8.

**Acknowledgements** Yusuke Endo and the orthoptists of Tsukazaki Hospital contributed to the collation of the data. We would like to thank Enago (www.enago.jp) for the English language review.

**Author contribution** All authors contributed to the study conception and design. Material preparation was performed by Yoshie Shimizu. Data collection was performed by Naofumi Ishitobi and Hiroki Ochi. Analyses were performed by Hiroki Masumoto and Mao Tanabe. The first draft of the manuscript was written by Hitoshi Tabuchi and Daisuke Nagasato and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Data availability** The data that support the findings of this study are available from the corresponding author (DN) upon reasonable request.

**Code availability** Not applicable.

## Declarations

**Ethics approval** All images and data obtained and procedures performed in this study were approved by the Institutional Review Board of Tsukazaki Hospital and adhere to the Declaration of Helsinki and its later amendments or comparable ethical standards.

**Consent to participate** Written informed consent was obtained from all participants included in the study.

**Consent for publication** The authors affirm that human research participants provided informed consent for publication of the images in Fig. 1.

**Conflict of interest** The authors declare no competing interests.

## References

1. Koka K, Patel BC (2021) Ptosis correction. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK539828/. Accessed 21 July 2021
2. Yadegari S (2016) Approach to a patient with blepharoptosis. Neurol Sci 37:1589–1596. https://doi.org/10.1007/s10072-016-2633-7
3. Kokubo K, Katori N, Hayashi K, Sugawara J, Fujii A, Maegawa J (2017) Evaluation of the eyebrow position after levator resection. J Plast Reconstr Aesthet Surg 70:85–90. https://doi.org/10.1016/j.bjps.2016.09.025
4. Zheng X, Kakizaki H, Goto T, Shiraishi A (2016) Digital analysis of eyelid features and eyebrow position following $CO_2$ laser-assisted blepharoptosis surgery. Plast Reconstr Surg Glob Open 4:e1063. https://doi.org/10.1097/GOX.0000000000001063
5. Hung JY, Perera C, Chen KW et al (2021) A deep learning approach to identify blepharoptosis by convolutional neural networks. Int J Med Inform 148:104402. https://doi.org/10.1016/j.ijmedinf.2021.104402
6. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539
7. Poplin R, Varadarajan AV, Blumer K et al (2018) Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2:158–164. https://doi.org/10.1038/s41551-018-0195-0
8. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402–2410. https://doi.org/10.1001/jama.2016.17216
9. Nagasato D, Tabuchi H, Masumoto H et al (2020) Prediction of age and brachial-ankle pulse-wave velocity using ultra-wide-field

pseudo-color images by deep learning. Sci Rep 10:19369. https://doi.org/10.1038/s41598-020-76513-4
10. Ohsugi H, Tabuchi H, Enno H, Ishitobi N (2017) Accuracy of deep learning, a machine-learning technology, using ultra–wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Sci Rep 7:9425. https://doi.org/10.1038/s41598-017-09891-x
11. Matsuba S, Tabuchi H, Ohsugi H et al (2019) Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. Int Ophthalmol 39:1269–1275. https://doi.org/10.1007/s10792-018-0940-0
12. Masumoto H, Tabuchi H, Nakakura S, Ishitobi N, Miki M, Enno H (2018) Deep-learning classifier with an ultrawide-field scanning laser ophthalmoscope detects glaucoma visual field severity. J Glaucoma 27:647–652. https://doi.org/10.1097/IJG.0000000000000988
13. Sonobe T, Tabuchi H, Ohsugi H et al (2019) Comparison between support vector machine and deep learning, machine-learning technologies for detecting epiretinal membrane using 3D-OCT. Int Ophthalmol 39:1871–1877. https://doi.org/10.1007/s10792-018-1016-x
14. Deng J, Dong, W, Socher R, Li L, Kai L, Li F-F (2009) ImageNet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp 248–255.https://doi.org/10.1109/CVPR.2009.5206848
15. Russakovsky O, Deng J, Su H et al (2015) ImageNet large scale visual recognition challenge. Int J Comp Vision 115:211–252. https://doi.org/10.1007/s11263-015-0816-y
16. Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z (2015) Deeply-supervised nets. In: Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS). San Diego, CA, USA: Journal of Machine Learning Research Workshop and Conference Proceedings, pp 562–570
17. Mosteller F, Tukey JW (1968) Data analysis, including statistics. In: Lindzey G, Aronson E, eds. Handbook of social psychology: Vol. 2. Research methods. Addison-Wesley, Reading, pp 80–203
18. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc Int Joint Conf AI 2:1137–1145
19. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C (2018) MobileNetV2: inverted residuals and linear bottlenecks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 4510–4520. arXiv:1801.04381, last revised 21 Mar 2019
20. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 1:1097–1105
21. Agrawal P, Girshick R, Malik J (2014) Analyzing the performance of multilayer neural networks for object recognition. In: Proc Lecture Notes in Computer Science, pp 329–344
22. Wang H, Wang Z, Du M et al. (2020) Score-CAM: score-weighted visual explanations for convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 24–25. arXiv:1910.01279, last revised 13 Apr 2020
23. Liu X, Faes L, Kale AU et al (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 1:e271-297. https://doi.org/10.1016/S2589-7500(19)30123-2
24. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36. https://doi.org/10.1148/radiology.143.1.7063747
25. Clopper CJ, Pearson ES (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404–413. https://doi.org/10.2307/2331986
26. Zhang Z, Beck MW, Winkler DA et al (2018) Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Ann Transl Med 6:216. https://doi.org/10.21037/atm.2018.05.32