Artificial Intelligence

Deep Learning-Based Automated Detection of Retinal Breaks and Detachments on Fundus Photography

Merlin Christ[1], Oussama Habra[1], Killian Monnin[2], Kevin Vallotton[1], Raphael Sznitman[2], Sebastian Wolf[1,3], Martin Zinkernagel[1,3], and Pablo Márquez Neila[2]

[1] Department of Ophthalmology, Inselspital, Bern University Hospital, Bern, Switzerland
[2] ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland
[3] Bern Photographic Reading Center, Department of Ophthalmology, Inselspital, Bern University Hospital, Bern, Switzerland

Correspondence: Martin S. Zinkernagel, University Hospital Bern, Bern 3010, Switzerland. e-mail: martin.zinkernagel@insel.ch

Purpose: The purpose of this study was to develop a deep learning algorithm, to detect retinal breaks and retinal detachments on ultra-widefield fundus (UWF) optos images using artificial intelligence (AI).

Methods: Optomap UWF images of the database were annotated to four groups by two retina specialists: (1) retinal breaks without detachment, (2) retinal breaks with retinal detachment, (3) retinal detachment without visible retinal breaks, and (4) a combination of groups 1 to 3. The fundus image data set was split into a training set and an independent test set following an 80% to 20% ratio. Image preprocessing methods were applied. An EfficientNet classification model was trained with the training set and evaluated with the test set.

Results: A total of 2489 UWF images were included into the dataset, resulting in a training set size of 2008 UWF images and a test set size of 481 images. The classification models achieved an area under the receiver operating characteristic curve (AUC) on the testing set of 0.975 regarding lesion detection, an AUC of 0.972 for retinal detachment and an AUC of 0.913 for retinal breaks.

Conclusions: A deep learning system to detect retinal breaks and retinal detachment using UWF images is feasible and has a good specificity. This is relevant for clinical routine as there can be a high rate of missed breaks in clinics. Future clinical studies will be necessary to evaluate the cost-effectiveness of applying such an algorithm as an automated auxiliary tool in a large practices or tertiary referral centers.

Translational Relevance: This study demonstrates the relevance of applying AI in diagnosing peripheral retinal breaks in clinical routine in UWF fundus images.

Introduction

Rhegmatogenous retinal detachment (RRD) is a possible devastating complication of retinal breaks that can lead to permanent vision loss.[1] It mostly affects the older working population around the globe and is currently of growing concern, because an important risk factor for both complicated and uncomplicated retinal breaks is high myopia,[2] which prevalence might double before 2050.[3] Because retinal breaks might remain asymptomatic until a retinal detachment has already taken place,[1] it is of high importance to develop machine-learning based algorithms that can detect them in an automated manner. Such a tool could be of great value, because visual impairment due to RRD can be averted with surgical treatment if recognized early enough.[4]

The preferred method for detecting retinal breaks relies on careful indirect ophthalmoscopy.[1] Fundus photography can assist in finding retinal breaks,[5] but cannot fully replace clinical ophthalmoscopy.[1] Recent

advances in wide field imaging techniques, namely ultra-widefield (UWF) imaging, have increased the sensitivity of those imaging modalities, especially for the peripheral retina.[6] UWF imaging devices enable the acquisition of 200 degrees of the retina in a single image, whereas conventional fundus photography is restricted to the central 30 degrees to 60 degrees. However, the increased sensitivity of the newer imaging modalities comes along with a more challenging, time-consuming analysis for the clinical practitioner, because a standard single UWF fundus image device generates an image series consisting of up to 15 images per eye. Here, machine learning plays an important role thanks to its efficiency and accuracy in processing large data sets.[7] In addition, the use of wide-field imaging to diagnose retinal breaks without the help of machine learning does not appear to have high enough sensitivity and specificity to be implemented as an effective screening tool.[8]

In ophthalmology, image computing using artificial intelligence (AI)-based technology has found its way into clinics due to robust capability in detecting and classifying retinal lesions.[9–11] Recently, Oh et al. even demonstrated the automated detection of retinal breaks on video streams taken from an operating microscope.[12] However, automatic classification of retinal detachment or retinal tears from UWF remains a quite understudied field of research, with most studies having tested their methods on controlled and selected data.[10,12–14] Especially in the setting of a tertiary referral center, where patients are imaged before consultation with a resident doctor or specialist, an accurate automated review of the vast amount of images taken would be desirable to avoid missing retinal breaks. This may also have medico-legal implications, as lesions missed in clinical examination leading to subsequent complications may be found documented on UWF imaging in hindsight.

The aim of this study was to build a deep learning system, based on a pretrained model which reduces the amount of data needed to establish a solid classifier, to collectively detect retinal breaks and detachments, and then correctly classify the images of non-healthy retinas into three subgroups.

Materials and Methods

Data Set

Data Collection

In this study, UWF Optos images (Daytona, Optos PLC, Dunfermline, UK) of retrospectively enrolled patients with a diagnosis of rhegmatogenous retinal detachment with or without visible retinal break and retinal breaks without detachment who have consulted the Department of Ophthalmology at the University Hospital Bern, Bern, Switzerland, were retrieved. Ethical approval for this study was obtained with the study identification number 2019-01588. Written informed consent was waived due to the retrospective and irreversible anonymization of the data. Duplicate acquisitions were excluded. Additionally, UWF images of healthy age- and sex-matched subjects without retinal pathologies, including history of retinal break or rhegmatogenous detachment, were collected. The images were fully anonymized and further processed on a per image basis.

Data Labeling

The anonymized UWF images were included or excluded independently by two retinal specialists. We applied following exclusion criteria: (1) image was acquired less than 3 months after any surgical procedure addressing RRD. (2) Insufficient image quality due to artifacts (e.g. insufficient lighting of the fundus due to intraocular medium opacities, eyelash images) making less than 60% of the peripheral region being assessable. (3) Other distinct retinal pathologies/lesions (such as hemorrhages, cotton-wool lesions, and exudates) or visible manifestations of prior surgical interventions rendering the assessment of retinal lesions specious. Therefore, a UWF image of an eye having undergone a surgical intervention can remain included if the image was acquired more than 3 months after any surgery and the assessment of retinal lesions is still sufficiently possible.

The included images were classified into the following groups: 0 = "No Study Lesion" being the healthy control group, 1 = retinal breaks without detachment, 2 = retinal breaks with retinal detachment, 3 = retinal detachment without visible retinal break, and 4 = a combination of groups 1 to 3. An example of each category is presented in Figure 1. Any classification disagreement was conciliated by a third senior retinal specialist with over 20 years of clinical experience. When no consensus was reached, the image was excluded. Figure 2 displays the image processing workflow. The labeling process of the images generated the ground truth for the deep learning system development.

After reviewing the eligible UWF images, 2489 images remained to develop the deep learning system. The Table shows their distribution per group. The chosen images were randomly split into 2 exclusive sets, one for training and one for testing, in a ratio of approximately 8:2. This was made on a per-patient basis, ensuring that no patient was present in both sets
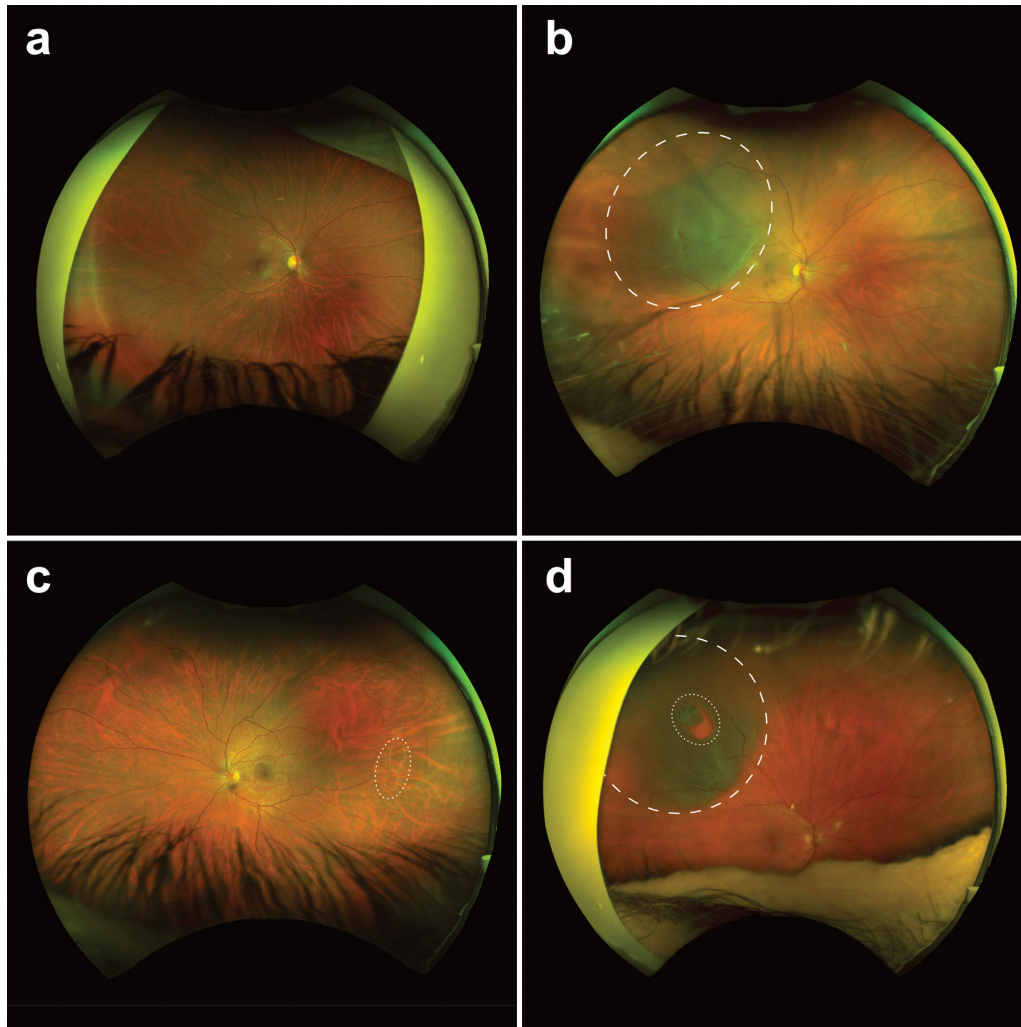
Figure 1. Examples of ultra-widefield fundus images. (**a**) Image with no study lesion. (**b**) Retinal detachment (*dashed circle*). (**c**) Two retinal breaks (*dotted circle*). (**d**) Retinal detachment (*bigger dashed circle*) and retinal break (*smaller dotted circle*).
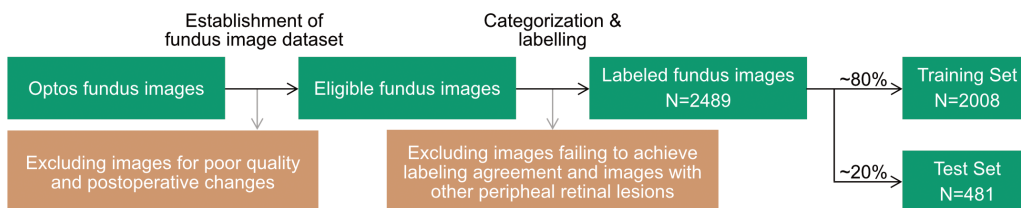

Figure 2. The workflow of establishing the Optos image dataset.

simultaneously. Furthermore, the split was stratified so that the proportions of elements of each group were similar in both the training and testing sets.

Deep-Learning System

Our deep learning system consists of a single convolutional neural network with three binary outputs. The first two outputs produce the log-probability of retinal break and retinal detachment. The third output produces the log-probability of the joint study lesion as a linear combination of the first two log-probabilities. PyTorch 1.10 was used to implement and train the neural network, scikit-learn to compute the evaluation metrics, and Matplotlib for visualization.

Convolutional Neural Network Architecture

We chose an EfficientNet-b0 architecture with wide squeeze-and-excitation layers pretrained on the

Table.   Number of Images Obtained Per Group

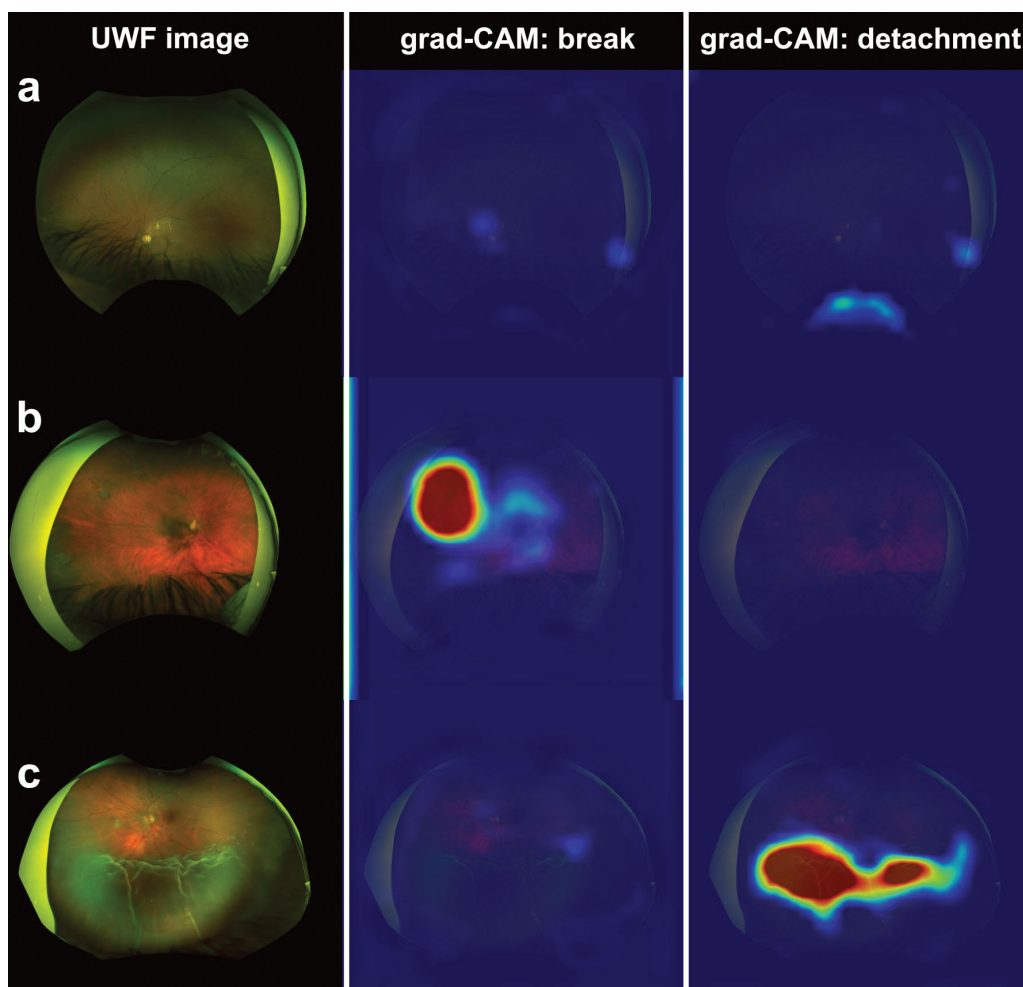| Groups	Images		Groups 1–3	Images	
	N	% of Total		*n*	% of Group 4
0) No study lesion	915	36.8			
4) Groups 1–3	1574	63.2			
			1) Retinal break only	685	43.5
			2) Retinal break with retinal detachment	546	34.7
			3) Retinal detachment only	343	21.8
Total	*n* = 2489	100%	Total group 4	1574	100%

Figure 3.   Examples of UWF images with corresponding grad-CAM overlay. (**a**) Image with no study lesion. (**b**) Image with retinal break. (**c**) Image with retinal detachment.

ImageNet dataset as the backbone of our convolutional neural network (CNN). We appended a linear layer with two outputs to this backbone to produce the log-probabilities for each lesion type. We also stacked an additional linear layer combining these two outputs to predict the joint study lesion.

Training

The network was trained with the training split of 2008 annotated images for 200 epochs. To compensate for the small amount of training data and the high variance of the model, we applied Polyak averaging by keeping two copies of the network: the first

network is updated following a typical training procedure with Adam optimizer with learning rate 10^{-3}, and the second network is updated computing the running average of the weights of the first network. Polyak averaging helped to smooth and stabilize the training. Additionally, we applied online data augmentation as detailed in the following section.

Data Augmentation

At training time, images were resized to 1060 by 1060 pixels and randomly cropped to 1024 by 1024 pixels. We applied random horizontal flipping, random rotation between −5 degrees and 5 degrees, random brightness shift with a factor between 0.8 and 1.6 and random contrast scale in the range of 0.8 to 1.2.

At inference time, we resized test images to 1060 by 1060 pixels and then cropped the central region of 1024 by 1024 pixels. No additional random augmentations were applied.

In all cases, pixel values were normalized to the range [0, 1] before passing the images to the network.

Evaluation Methodology

To assess performance, each output of our system (retinal break, retinal detachment, and study lesion) was evaluated independently using the collection of 481 annotated images of the test set. For each output, we measured the area under the receiver operating characteristic (ROC) curve and the average precision. Additionally, we chose an operating point of the model maximizing the training performance and measured the sensitivity, specificity, precision, and recall of the model on the test data.

We trained 100 models to provide 95% confidence intervals for all considered metrics. The 100 models were obtained by running 10 independent training procedures with different random seeds, and then keeping 10 models from the last 10 epochs of each training sequence.

To understand regions of interest in each image for the model to make its prediction, a gradient-weighted class activation mapping (grad-CAM) function was implemented to visualize which part of the image is detected to make the prediction of an image. Examples are visible in Figure 3.

Results

Performance

The EfficientNet-b0 classification model with three binary outputs was trained with the training set and the final prediction computed for the test set. The perfor-
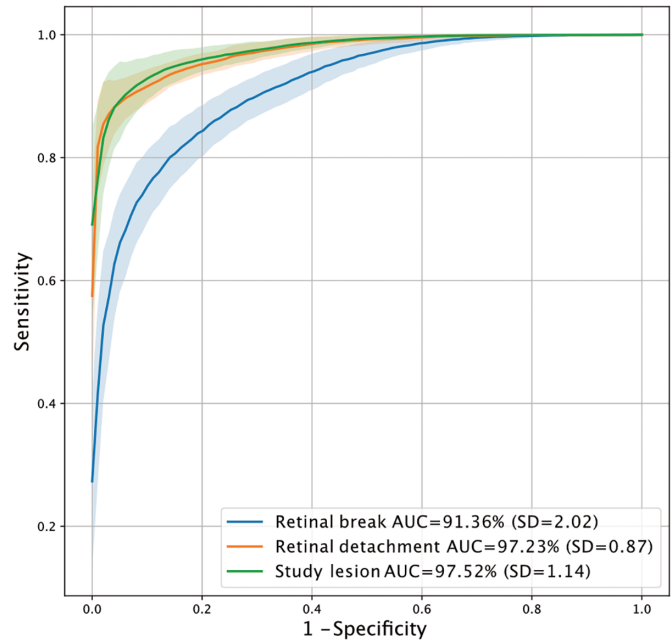

Figure 4. Receiver operating characteristic (ROC) curves for retinal breaks, retinal detachment, and study lesions with area under the curve (AUC) and standard deviation (SD) values.

mance is presented in Figure 4. An area under the curve (AUC) of 0.975 was achieved for lesion recognition, whereas AUCs of 0.972 and 0.913 were obtained for retinal detachment and retinal breaks, respectively.

Error Analysis

To investigate errors made by the deep-learning system, one model with representative performance was chosen. At the operating point for the selected model, we found the following metrics. Study lesion: Sensitivity = 0.919, Specificity = 0.979, Precision = 0.985, Accuracy = 0.943; Detachment: Sensitivity = 0.869, Specificity = 0.954, Precision = 0.916, Accuracy = 0.923; Break: Sensitivity = 0.863, Specificity = 0.828, Precision = 0.808, and Accuracy = 0.844. The performance and confusion matrices of the chosen model is visible in Figure 5. All images misclassified by this model were manually analyzed.

The common characteristics of false-negative cases for the study lesion or no study lesion model ($n = 23$) included visible artifacts of prior laser photocoagulation ($n = 14$, 60.8%) and break localizations in the temporal superior quadrant ($n = 16$, 69.5%). Many of the false-negative cases ($n = 17$, 73.9%) were images only showing a retinal break but no retinal detachment, a minority ($n = 4$, 19%) revealed retinal breaks and detachment and only two ($n = 2$, 9.5%) showed retinal detachment and no retinal break.
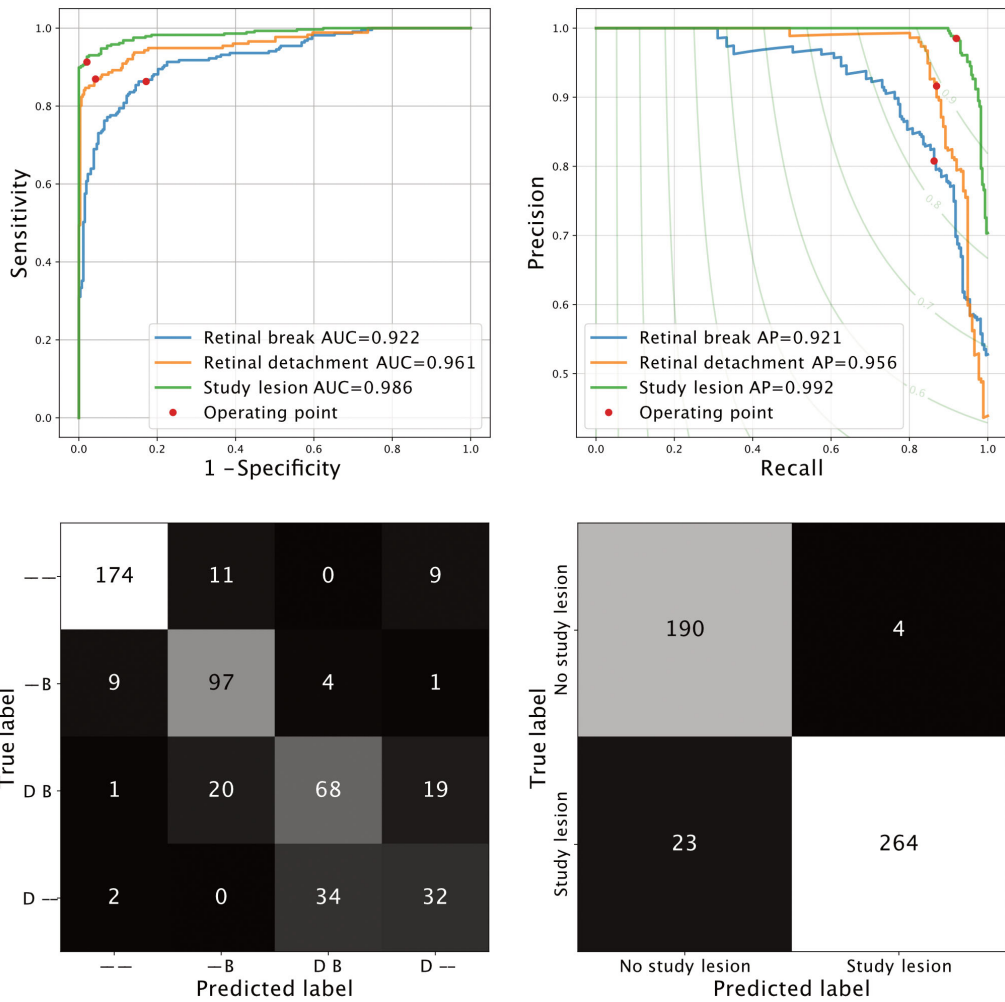
Figure 5.   ROC curves and confusion matrices of the one model chosen for error analysis.

The analysis of false-positive cases for the study lesion or no study lesion model ($n = 4$) revealed that all four images were impacted by obvious eyelash artifacts, whereas one image ($n = 1$) additionally displayed a small reflection artifact and signs of a thin retinal layer.

The false-negative cases in the retinal break subcategory ($n = 30$) consisted of a majority of images ($n = 20$, 66.7%) revealing retinal detachment as well. Signs of prior laser photocoagulation or cryocoagulation were found in seven images ($n = 7$, 23.3%) and one ($n = 1$, 3.3%) image, respectively. In a few images ($n = 10$, 33.3%), eyelash artifacts optically covered part of the retinal break, as displayed in Figure 6.

False-positive cases in the retinal break subcategory ($n = 45$) commonly presented retinal detachment ($n = 33$, 73.3%) and/or lattice degeneration ($n = 27$, 60%). Signs of intravitreous hemorrhage ($n = 15$, 33.3%) or prior laser photocoagulation ($n = 21$, 46.7%) were visible occasionally. During manual analy-

Figure 6.   Image with visible retinal break (inside *dotted circle*) partially covered by eyelash, wrongly classified as "no retinal break" by the algorithm.
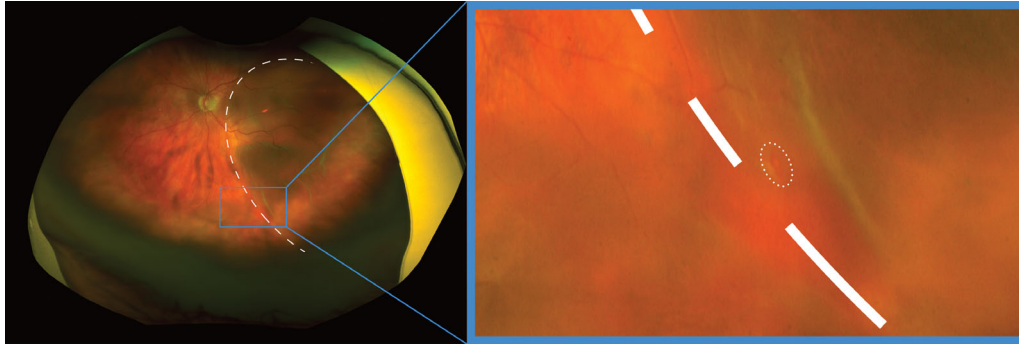
Figure 7. Presumably false-positive retinal break image with retinal detachment (*dashed circle*), which actually shows a small retinal break (*small dotted circle in enlarged section*).

sis of misclassified images, one photo originally labeled by the retinal team as no retinal break and presumably wrongly classified as a retinal break by the classification model, showed a small retinal break in the temporal inferior quadrant, as highlighted in Figure 7.

In the retinal detachment subcategory, false-negative cases ($n = 23$) shared following common characteristics: visible retinal break ($n = 19$, 82.6%), retinal detachment location in the superior half of the retina ($n = 17$, 69.5%), signs of prior laser photocoagulation ($n = 10$, 43.4%), and intravitreous hemorrhage ($n = 4$, 17.4%).

The false-positive cases in the retinal detachment subcategory ($n = 14$) demonstrated an impacted image quality due to eyelash artifacts ($n = 7$, 50%), visible retinal breaks ($n = 5$, 35.7%), reflection artifacts ($n = 5$, 35.7%), and visible signs of a thin retinal layer ($n = 2$, 14.2%).

Discussion

Our results demonstrate that the evaluated algorithm can predict the presence of retinal breaks and retinal detachment with a high degree of accuracy, with an AUC of 91.36% for retinal break and 97.23% for retinal detachment. The relatively lower AUC for a retinal break could be explained by the fact that retinal breaks are usually smaller in size and therefore visually less apparent in comparison to retinal detachment.

Importantly, most (60.8%) of the false-negatives were associated with prior laser photocoagulation, which supports the fact that this algorithm would probably show better performance in primary screening for retinal breaks and detachments prior to therapy, in comparison to being used as an auxiliary tool at a tertiary referral center. However, up to 14% of patients with prior laser photocoagulation develop

new breaks elsewhere[15,16] and, as such, we considered it important to include these patients into our study.

On the contrary, false-positives seem to be partially avoidable with a better triage of the images excluding poor quality images (e.g. Optic medium opacity), which could thus certainly be further improved.

In 2017, Ohsugi et al.[9] demonstrated promising results with a deep learning system for detecting RRD using UWF imaging. In 2019 and 2020, Li et al.[10,14] specifically pointed the potential of a deep learning method in the detection of different retinal pathologies using UWF images. Direct comparison to recent studies relating to detection of retinal breaks or detachments using deep learning is limited due to different inclusion and exclusion criteria concerning baseline requirements or included retinal lesions. Although the system developed in this study showed good performance for retinal breaks (AUC = 0.913) and for retinal detachment (AUC = 0.972), other deep learning systems aiming to detect these conditions achieved even higher performance. Li et al. obtained AUC values of 0.989 for retinal detachment in 2020.[14] Zhang et al. developed a deep learning method, which achieved an AUC of 0.953 and 1.000 for the detection of retinal breaks and retinal detachment, respectively, in tessellated eyes in 2021.[13] Oh et al. achieved an AUC of 0.957 for retinal breaks with their novel object detection based algorithm.[12]

Although the mentioned deep learning systems were trained with UWF images labeled by trained retina specialists, different grading parameters including experience, exclusion criteria, and number of involved graders can further influence the quality of the grading process. Zhang et al. specifically excluded fundus images of eyes with visible signs of previous vitreoretinal surgery or retinal photocoagulation.[13] In our data set, we deliberately included images with signs of previous surgery or retinal photocoagulation,

if the image was acquired at least 3 months after any surgical intervention, with the aim to create a data set that better corresponds to real clinical conditions. Indeed, nearly 6 to 7% of patients having undergone laser photocoagulation of the retina because of a retinal break experience progression to retinal detachment.[17,18] Li et al. and Oh et al. do not further elaborate if such images were included or excluded in their study.[10,12,14] In a clinical setting where technicians are performing UWF imaging as preliminary examination before clinics and clinicians are faced with a large number of images per eye, which they may not be able to examine in detail, this tool may be useful to avoid missing breaks. Even more so, as medico-legal consequences could arise as breaks may be documented on UWF but may have been missed in clinical examination by resident doctors or even retina specialists. The rate of missed retinal breaks in large practices has been reported to be as high as 27%.[19]

The strengths of our study include the robust training data set (>2000 images) and a high threshold for image exclusion (e.g. prior photocoagulation not being an exclusion criterion), with the aim of approaching real clinical data conditions.

Several limitations should be mentioned. First, the final data set used to train the deep learning system consisted of only approximately one third of images with no study lesion. In real-world settings, less than 18% of people are found to have retinal breaks, if not substantially fewer.[2] Whereas the data set used for training and testing in this study has been balanced for machine learning purposes, it does not properly represent the real-world situation in this relation. If this system was to be applied as a sort of screening approach in clinical settings, this may inflict new challenges. Second, the image database used for data collection only contained images from a single University Eye Clinic. Different patient population, devices, or workflows in other institutions could impact the generalizability of our findings. Third, although UWF imaging enables the acquisition of 200 degrees of the retina in a single image, some peripheral regions are still not covered.[8] As our study evaluated the performance regarding the detection of retinal breaks or detachments on UWF images, lesions outside the image acquisition area remain unnoticed by the algorithm. This aspect would have to be assessed if such a system is to be further developed with the idea of a screening tool.

The current gold standard to screen for and detect retinal breaks or retinal detachment is through clinical examination, generally by binocular indirect ophthalmoscopy with or without indentation. The current performance of our system would need to be improved in order to come into consideration for screening purposes in a general population. However, this technique would be useful as an auxiliary tool for ophthalmology referral centers.

In conclusion, the deep learning system developed in this study was able to achieve good performance for identifying retinal detachment and retinal breaks using UWF images. Future clinical studies will be necessary to evaluate the cost-effectiveness of applying this algorithm as an automated approach to detect retinal detachment and retinal breaks in clinical settings.

References

1. Flaxel CJ, Adelman RA, Bailey ST, et al. Posterior vitreous detachment, retinal breaks, and lattice degeneration preferred practice pattern. *Ophthalmology*. 2020;127(1):P146–P181.
2. Kazahaya M. Prophylaxis of retinal detachment. *Semin Ophthalmol*. 1995;10(1):79–86.
3. Holden BA, Fricke TR, Wilson DA, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*. 2016;123(5):1036–1042.
4. Byer NE. Subclinical retinal detachment resulting from asymptomatic retinal breaks: prognosis for progression and regression. *Ophthalmology*. 2001;108(8):1499–1503.
5. Shoughy SS, Arevalo JF, Kozak I. Update on wide- and ultra-widefield retinal imaging. *Indian J Ophthalmol*. 2015;63(7):575–581.
6. Lee J, Sagong M. Ultra-widefield retina imaging: principles of technology and clinical applications. *J Retina*. 2016;1(1):1–10.
7. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–1358, doi:10.1056/NEJMra1814259.
8. Kornberg DL, Klufas MA, Yannuzzi NA, Orlin A, D'Amico DJ, Kiss S. Clinical utility of ultra-widefield imaging with the Optos Optomap compared with indirect ophthalmoscopy in the set-

ting of non-traumatic rhegmatogenous retinal detachment. *Semin Ophthalmol*. 2016;31(5):505–512.

9. Ohsugi H, Tabuchi H, Enno H, Ishitobi N. Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. *Sci Rep*. 2017;7(1):9425.

10. Li Z, Guo C, Nie D, et al. A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images. *Ann Transl Med*. 2019;7(22):618.

11. Gallardo M, Munk MR, Kurmann T, et al. Machine learning can predict anti–VEGF treatment demand in a treat-and-extend regimen for patients with neovascular AMD, DME, and RVO associated macular edema. *Ophthalmol Retina*. 2021;5(7):604–624.

12. Oh R, Oh BL, Lee EK, Park UC, Yu HG, Yoon CK. Detection and localization of retinal breaks in ultrawidefield fundus photography using a yolo V3 architecture-based deep learning model. *Retina*. 2022;42(10):1889–1896.

13. Zhang C, He F, Li B, et al. Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: a pilot study. *Graefes Arch Clin Exp Ophthalmol*. 2021;259(8):2225–2234.

14. Li Z, Guo C, Nie D, et al. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. *Commun Biol*. 2020;3(1):15.

15. Smiddy WE, Flynn HW, Nicholson DH, et al. Results and complications in treated retinal breaks. *Am J Ophthalmol*. 1991;112(6):623–631.

16. Goldberg RE, Boyer DS. Sequential retinal breaks following a spontaneous initial retinal break. *Ophthalmology*. 1981;88(1):10–12.

17. Khan AA, Gupta A, Bennett H. Risk stratifying retinal breaks. *Can J Ophthalmol*. 2013;48(6):546–548.

18. Garoon RB, Smiddy WE, Flynn HW. Treated retinal breaks: clinical course and outcomes. *Graefes Arch Clin Exp Ophthalmol*. 2018;256(6):1053–1057.

19. Takkar B, Azad S, Shashni A, Pujari A, Bhatia I, Azad R. Missed retinal breaks in rhegmatogenous retinal detachment. *Int J Ophthalmol*. 2016;9(11):1629–1633.