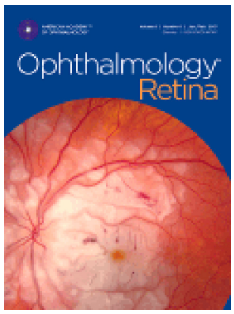
Journal Pre-proof



Deep-learning algorithm for the diagnosis and prediction of hydroxychloroquine retinopathy: An International, multi-institutional study

Peter Woodward-Court, MBBS, MRes, Jeffry Hogg, MBBS, Ph.D., Terry Lee, M.D., Priyal Taribagil, M.D., Cindy S. Zhao, M.D., Vanessa Otti, William R. Tucker, M.D., Michael Allingham, M.D., Ph.D., Oleg Alekseev, M.D., Ph.D., Siegfried K. Wagner, M.D. Ph.D., David Myung, M.D., Ph.D., Loh-Shan Leung, M.D., Eleonora M. Lad, M.D., Ph.D., Hani Hasan, Ch.M., James Talks, MBBchir, Daniel C. Alexander, Ph.D, Pearse A. Keane, M.D., Eliot R. Dow, M.D., Ph.D.

PII: S2468-6530(25)00274-X

DOI: https://doi.org/10.1016/j.oret.2025.06.003

Reference: ORET 2025

To appear in: *Ophthalmology Retina*

Received Date: 14 January 2025

Revised Date: 4 June 2025

Accepted Date: 5 June 2025

Please cite this article as: Woodward-Court P., Hogg J., Lee T., Taribagil P., Zhao C.S., Otti V., Tucker W.R., Allingham M., Alekseev O., Wagner S.K., Myung D., Leung L.-S., Lad E.M., Hasan H., Talks J., Alexander D.C., Keane P.A. & Dow E.R., Deep-learning algorithm for the diagnosis and prediction of hydroxychloroquine retinopathy: An International, multi-institutional study, *Ophthalmology Retina* (2025), doi: https://doi.org/10.1016/j.oret.2025.06.003.

Deep-learning algorithm for the diagnosis and prediction of hydroxychloroquine retinopathy: An International, multi-institutional study

Peter Woodward-Court, MBBS, MRes.[1*], Jeffry Hogg, MBBS, Ph.D..[1,2], Terry Lee, M.D.[3], Priyal Taribagil, M.D.[1], Cindy S. Zhao, M.D.[4], Vanessa Otti[1], William R. Tucker, M.D.[1], Michael Allingham, M.D., Ph.D.[3], Oleg Alekseev, M.D., Ph.D.[3], Siegfried K. Wagner, M.D. Ph.D.[1], David Myung, M.D., Ph.D. [4,5], Loh-Shan Leung, M.D.[4,5], Eleonora M. Lad, M.D., Ph.D.[3], Hani Hasan, Ch.M.[2], James Talks, MBBchir[2], Daniel C. Alexander, Ph.D[6]., Pearse A. Keane, M.D.[1], Eliot R. Dow, M.D., Ph.D.[4,7]

[1] NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust, UCL Institute of Ophthalmology, London, UK

[2] Department of Ophthalmology, Newcastle-upon-Tyne Hospital NHS Foundation Trusts, Newcastle-upon-Tyne, UK

[3] Department of Ophthalmology, Duke University Medical Center, 2351 Erwin Rd, Durham, North Carolina 27705

[4] Byers Eye Institute at Stanford, Department of Ophthalmology, Stanford University School of Medicine, Palo Alto, California 94303

[5] Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA, 94304

[6] Centre for Medical Image Computing, Department of Computer Science, UCL, London, UK

[7] Retinal Consultants Medical Group, Sacramento, CA, USA, 95825

[*] Corresponding author:

Peter Woodward-Court

Institute of Health Informatics, University College London

222 Euston Road

NW1 2DA, United Kingdom

peter.woodward-court@ucl.ac.uk

Keywords: hydroxychloroquine, retinopathy, HCQ, machine learning

Abbreviations and Acronyms:

AUROC area under the receiver-operator characteristic

AUPRC area under the precision-recall curve

CNN convolutional neural network

HCQ hydroxychloroquine

LRS likelihood of retinopathy score

mfERG multifocal electroretinography

NPV negative predictive value

PPV positive predictive value

PACS picture archiving and communication system

SaMD software as a medical device

SD-OCT spectral domain optical coherence tomography

VF visual field

The author(s) have made the following disclosure(s):

Financial Disclosures:

P.W.C - none

E.R.D - none

T.L - none

C.S.Z. - none

P.K has acted as a consultant for Retina Consultants of America, Topcon, Roche, Boehringer-Ingleheim, and Bitfount and is an equity owner in Big Picture Medical. He has received speaker fees from Zeiss, Novartis, Gyroscope, Boehringer-Ingleheim, Apellis, Roche, Abbvie, Topcon, and Hakim Group. He has received travel support from Bayer, Topcon, and Roche. He has attended advisory boards for Topcon, Bayer, Boehringer-Ingleheim, RetinAI, and Novartis.

O.A. - none

M.F.M. - none

P.T. - none

V.O. - none

E.L. - none

Acknowledgments:

PK is supported by a UK Research & Innovation Future Leaders Fellowship (MR/T019050/1).

SW is supported by a Medical Research Council Clinical Research Training Fellowship (MR/T000953/1).

DCA is supported by the NIHR UCLH Biomedical Research Centre, Wellcome Trust award 221915, and EPSRC EP/Y017803/1.

OA is supported by a K08 Career Development Award from NEI/NIH (K08EY033857).

TL is supported by the VRSF Research Award.

Short title: Deep learning algorithm for the prediction of hydroxychloroquine retinopathy

Abstract

Purpose We sought to develop a deep-learning algorithm - HCQuery - to detect the presence of hydroxychloroquine retinopathy and predict its future occurrence from spectral-domain optical coherence tomography (SD-OCT) images.

Design We trained and validated a deep-learning algorithm using retrospective SD-OCT images from patients taking hydroxychloroquine.

Participants The study involved a retrospective, non-consecutive collection of 409 patients (171 positive for hydroxychloroquine retinopathy, 238 negative for retinopathy) and 8251 SD-OCT b-scans (1988 volumes) from five independent international clinical locations.

Methods Imaging macular volumes from two different SD-OCT devices (Heidelberg Spectralis, Zeiss Cirrus) at two clinical sites were used to train and validate a convolutional neural network (EfficientNet-b4) to produce a Likelihood of Retinopathy Score (LRS) for each SD-OCT b-scan. LRS scores were processed across SD-OCT volumes for an eye- and patient-level binary decision output of the presence or absence of retinopathy. The adjudicated consensus of up to three independent retina specialists using patient clinical data and multimodal testing served as the reference standard for hydroxychloroquine retinopathy. The algorithm was tested on four withheld test sets, one internal (Data Set 1) and three external (Data Sets 3, 4, and 5). The test sets were obtained in two countries (United States, United Kingdom) and represented two SD-OCT devices each with diverse acquisition parameters.

Main Outcome Measures The algorithm was assessed with sensitivity, specificity, accuracy, negative predictive value (NPV), positive predictive value (PPV), area under the receiver-operator characteristic (AUROC), and area under the precision-recall curve (AUPRC) for the detection of hydroxychloroquine retinopathy either at the time of clinical diagnosis or up to 18 months in advance of clinical diagnosis.

Results The algorithm demonstrated discriminated hydroxychloroquine retinopathy at the time of clinical diagnosis as well as in advance of clinical diagnosis (Mean 220.8 days prior to clinical diagnosis; Accuracy: 0.987 (95% CI: 0.962-1.00), Sensitivity: 1.00 (95% CI: 0.833-1.00), Specificity: 0.983 (95% CI: 0.952-1.00), PPV: 0.944 (95% CI: 0.836-1.00), NPV: 1.00 (95% CI: 0.937-1.00)). For eyes that developed retinopathy, it was identified as positive by the algorithm on average 2.74 years in advance of the clinical diagnosis.

Conclusions We report a deep learning algorithm that can detect hydroxychloroquine retinopathy at all stages of disease as well as predict retinopathy years in advance of clinical diagnosis.

Financial Disclosure(s): Authors with financial interests or relationships to disclose are listed.

Introduction

Hydroxychloroquine sulfate (HCQ) is an anti-inflammatory medication used to treat systemic lupus erythematosus, rheumatoid arthritis, Sjögren syndrome, and other autoimmune conditions. The drug's minimal cost, high efficacy, and proven survival benefit has led to its widespread use.[1-5] However, hydroxychloroquine retinopathy is a well-characterized drug side effect, affecting up to 7.5% of users.[3] Drug build up in retinal pigment epithelium (RPE) is believed to lead to decompensation in metabolic functions, resulting in secondary damage to the photoreceptors.[6] In people of European descent, the retinopathy generally begins in the parafovea and progressively expands in a ring-like distribution around the fovea characteristically described as a bull's eye when RPE damage makes it visible by ophthalmoscopy. In people of Asian descent, the most common pattern of retinopathy is pericentral occurring near the vascular arcades.[7] The consequences of these macular changes include loss of paracentral visual field, and once the RPE is damaged, a loss of visual acuity. The only available treatment for hydroxychloroquine retinopathy is cessation of the medication.[8]

Current guidelines recommend annual screening after five years of hydroxychloroquine use with spectral-domain optical coherence tomography (SD-OCT) and subjective visual field (VF) testing[9,11] to prevent severe vision loss.[4,9,10] Typical findings show loss of the parafoveal ellipsoid zone as well as functional vision loss distributed in a parafoveal or pericentral ring.

Several shortcomings arise from current approaches to retinopathy screening. First, the established SD-OCT biomarker of retinopathy, ellipsoid zone loss, may not represent the earliest sign in the disease process.[12] Even when hydroxychloroquine is stopped at the time of

ellipsoid zone loss, retinal thinning often progresses after cessation.[13,14] Moreover, recent work has shown that other retinal changes begin several years in advance of ellipsoid zone loss.[6]

Second, a multimodal approach to screening also imposes limitations. VF testing for retinopathy screening is time consuming and frequently unreliable, particularly in patients with physical limitations from connective tissue diseases like rheumatoid arthritis.[15] Tests such as FAF or multifocal ERG have poor consistency of results, high cost and low availability,[9] so that a multimodal testing approach can be burdensome and poorly cost effective. Reflecting this, the most recent screening recommendations from the UK Royal College of Ophthalmology omitted recommending a VF examination unless there is an abnormal FAF or SD-OCT study.[4,11] Finally, due to the low frequency of positive patients and the subtle signs of early retinopathy, clinicians may make diagnostic errors while trying to identify the proverbial needle-in-a-haystack.

An improved system for screening should possess several features. First, it should detect retinopathy prior to the onset of ellipsoid zone loss. Second, it should do so for both the parafoveal and pericentral phenotypes of hydroxychloroquine retinopathy. Third, it should limit the burden on patients and the healthcare system by employing a single diagnostic modality that is widely available. Finally, it should democratize this difficult diagnosis to clinicians throughout all levels of specialization.

Here we present the results of a deep-learning algorithm for the detection of hydroxychloroquine retinopathy from an SD-OCT study. The algorithm demonstrates the ability to detect retinopathy at all stages of disease, including prior to ellipsoid zone loss, and, in many cases, years in advance of the clinical diagnosis. The algorithm is fully automated and renders a diagnosis without human annotation or other manual inputs. The algorithm works on both Zeiss and Heidelberg imaging devices with diverse image-acquisition parameters using a robust, modular

approach to retinopathy prediction using as little as a single b-scan. The algorithm demonstrated high sensitivity and positive predictive value (PPV) with face validity for both retinopathy detection and prediction on external data sets obtained from multiple institutions across two countries.

Methods

Ethics and institutional governance approvals

This study was reviewed and approved by Institutional Review Boards at Stanford University, Duke University, Newcastle-upon-Tyne Hospital, and Moorfields Eye Hospital. Patient consent for inclusion of data was waived for this retrospective analysis which did not alter standard patient-care procedures. Patient data was de-identified. The protocol followed tenets of human research as presented in the Declaration of Helsinki.

Patient Imaging Data

Data was obtained during routine patient care between 2015-2023 from five independent clinical locations (Table 1). SD-OCT data was acquired on two devices, Heidelberg Spectralis and Zeiss Cirrus using macula-centered protocols with wide variation in acquisition parameters including section number, step size, section orientation, and section width. In total, model training and validation was performed with 264 patients (104 retinopathy positive, 160 retinopathy negative) and 4900 SD-OCT b-scans (2450 positive, 2450 negative) from 1702 imaging volumes across two clinical locations. All b-scans with adequate imaging acquisition quality were included. Model testing was performed on 145 patients (67 positive, 78 negative) and 3351 SD-OCT b-scans (1566 positive, 1785 negative), which were fully excluded on the patient level from the training and validation data. The testing data originated from four clinical locations, three that were independent from the training and validation data and one that was a withheld, internal test set.

Data was obtained from both male and female patients as well as diverse self-reported racial or ethnic groups (Table 1). Three patients (6 eyes) in the test set, all of self-reported East Asian

race, demonstrated pericentral pattern retinopathy. Among eyes in the test set 18.5% (53/286 eyes) had one or more macular comorbidities, which were included for analysis. SD-OCT volumes were excluded from the training or test sets if the acquisition quality was insufficient to distinguish retinal layers (1.4% of cases) or if retinal comorbidities were considered severe (a single case).

<< TABLE 1 LOCATION >>

Table 1 - Characteristics of data sets involved in the training and testing of the algorithm.

Clinical Data and Clinical Taxonomy

Each discovered case yielded by the method described above was reviewed by a member of the research team (C.Z., E.D., J.H., P.T., V.O. T.L., S.W., W.T.) who determined whether the patient had undergone SD-OCT imaging that was available for research use through institutional picture archiving and communication system (PACS) systems and whether they had done so while using hydroxychloroquine for four or more years.

The gold standard for hydroxychloroquine retinopathy was determined by an adjudicated consensus of up to three independent retina specialists (E.D., L.L, M.A., O.A., P.A.K., H.H.). The labeling system involved detailed review of all available SD-OCT, FAF, VF, and mfERG studies as well as clinical data including years of medication use, daily medication dose, cumulative medication dose, kidney function (e.g. serum creatinine, glomerular filtration rate), use of tamoxifen, and weight. Reviewers were masked to the clinical diagnosis in the medical record. If the reviewer's diagnosis of retinopathy differed from the clinical diagnosis documented in the medical record, then the case was adjudicated by a third retinal specialist. The majority

label was used as the final label for hydroxychloroquine retinopathy. Explicit criteria for retinopathy were determined by the individual retina specialist, but in general, reviewers examined imaging data for parafoveal ellipsoid zone loss, outer nuclear layer thinning, generalized or progressive retinal thinning, or attenuation or blurring of the interdigitation zone on SD-OCT; parafoveal hyperautofluorescence or bull's eye hypoautofluorescence in fundus autofluorescence images; bilateral parafoveal visual field deficits on reliable testing; and bilateral decreased amplitudes and delayed implicit times in a parafoveal distribution on mfERG. A clinical diagnosis of hydroxychloroquine retinopathy in the medical record was defined as the retina specialist recommending cessation of the medication due to retinal toxicity.

AI algorithm development

We trained a convolutional neural network (CNN), EfficientNet-b4, to take as input SD-OCT b-scans and output a likelihood of retinopathy score (LRS) for each given image.[16] The b4 subtype was chosen to balance model size with computational requirements. For model development, data was randomly partitioned on a patient basis into training, validation, and test sets on an 80:10:10 split with even numbers of hydroxychloroquine retinopathy positive and negative cases in the training and validation sets. Test sets contained a single scan from both eyes of a patient, if available, at the furthest time point between 0 to 18 months prior to clinical diagnosis (Mean 220.8 days prior to clinical diagnosis). Training sets included multiple time points from the same patient wherever available. For patients without hydroxychloroquine retinopathy, SD-OCT images were used in which there was a subsequent encounter at least 18 months in the future without retinopathy.

We undertook model training with four NVIDIA T4 GPUs. The CNN was pre-trained on ImageNet-1k. We utilized Cross Entropy Loss with an Adam Optimiser set with a learning rate of 0.0001 with a scheduler (with a factor of 0.5 and a patience of 2) to prevent overfitting. We also utilized early stopping. At run time, the model ran over twenty epochs.

The model and associated code have been published on GitHub and are available here: https://github.com/peterwoodward-court/HCQuery (accessed 4th June 2025)

Statistical analysis

To improve uniformity and robustness across different SD-OCT acquisition protocols, a mean of LRS from b-scans within the central 1.5 millimeters of each volume was calculated. A volume was classified as positive for retinopathy if the volume mean was equal to or greater than 0.50. If an SD-OCT volume from either the patient's right or left eye was positive, the patient was classified as having a positive screening result for retinopathy (Supplemental Figure 1).

Two-tailed 95% confidence intervals for accuracy, specificity, and positive predictive value were calculated by Wald interval method using z-score of 1.96; confidence intervals for sensitivity and negative predictive value were calculated by Wilson binomial proportion confidence interval. Clinical data was analyzed by two-tailed Student's t-test for continuous variables and Chi-squared test for categorical variables.

Differences in algorithm performance for self-reported gender and race were performed at the eye-level due to the greater number of errors for single eyes than patients thus permitting more opportunities to detect any disparities.

Although there is no definitive severity staging system for hydroxychloroquine retinopathy, staging was based on previous work in which mild retinopathy involved parafoveal or pericentral outer retinal thinning with minimal, focal ellipsoid zone loss of less than 250 micrometers on SD-OCT; moderate retinopathy involved ellipsoid zone loss of greater than 250 micrometers but without a complete ring; and severe retinopathy involved a complete bullseye of ellipsoid zone loss with or without RPE hyperplasia.[3] An additional category was added of early disease that involved an intact ellipsoid zone in the setting of subtle structural changes on SD-OCT including outer nuclear layer thinning, interdigitation zone blurring, or variegated ellipsoid zone appearance.

Results

Positive retinopathy cases were associated with older age (Positive: 65.6 years (SD 12.8), Negative: 59.3 years (SD 12.8), P-value: 0.006), years of medication use (Positive: 15.6 years (SD 8.4), Negative: 12.8 years (SD 7.3), P-value: 0.046), and lifetime cumulative dose (Positive: 1776.7 grams (SD 1199.7), Negative: 1298.2 grams (SD 930.4), P-value: 0.020, Supplemental Table 1). 25.4% patients had early retinopathy (17); 31.3% had mild retinopathy (21); 22.3% had moderate retinopathy (15); and 20.9% had severe retinopathy (14). 25.4% of patients (17) converted to retinopathy during surveillance.

The algorithm discriminated hydroxychloroquine retinopathy with the following performance on the external test sets: Accuracy: 0.954 (95% CI: 0.916-0.992), Sensitivity: 1.00 (95% CI: 0.931-1.00), Specificity: 0.910 (95% CI: 0.834-0.986), PPV: 0.912 (95% CI: 0.839-0.985), NPV: 1.00 (95% CI: 0.931-1.00), AUROC: 0.99, AUPRC: 0.99) (Figure 1). Performance metrics were similar for the internal test set. The AUROC and AUPRC were similar across each of the four independent data sets, and performance was similar when evaluated at both the eye and patient level (Figure 1).The algorithm identified retinopathy across multiple stages of the condition, multiple image-acquisition protocols, and in the setting of image-acquisition artifacts (Figure 2). Three patients (6 eyes) of patients with pericentral pattern retinopathy, all of whom self-reported East Asian race, were correctly identified as being positive for retinopathy. There was no significant difference in algorithm performance by self-reported race (chi-square statistic: 0.7425, p-value: .388846), gender (chi-square statistic: 0.0054, p-value: .941671), or among the 18.5% (53/286 eyes) of eyes with one or more macular comorbidities (chi-square statistic: 2.5417, p-value: .110873).

<<FIGURE 1 LOCATION>>

<<FIGURE 2 LOCATION>>

In accordance with the bilateral, symmetric presentation of hydroxychloroquine retinopathy, the LRS between right and left eyes demonstrated a coefficient of determination of $R^2 = 0.913$ (Supplemental Figure 2). Additionally, the standard deviation of retinopathy likelihood scores across a single SD-OCT volume was low (Mean LRS for positive cases: 0.836 (SD: 0.071), Mean LRS for negative cases: 0.145 (SD: 0.067)), and the algorithm generally detected retinopathy in b-scans from positive eyes even when there was no discernible ellipsoid zone loss in that particular b-scan. However, LRSs tended to decrease with distance from the fovea (Supplemental Figure 3).

Among four eyes with false positive results, each possessed abnormalities of the parafoveal ellipsoid zone including shadowing by vitreous opacities and irregularity from age-related macular degeneration (Supplemental Figure 4). One eye with a false negative result showed strong focal ellipsoid zone loss in the inferior parafovea, which was detected as positive by the algorithm, along with a normal-appearing central and superior parafovea, which were given a low LRS (Supplemental Figure 5). These errors suggest that the algorithm may be attending to the outer retina as a major factor in generating its output.

For prediction of hydroxychloroquine retinopathy from a single SD-OCT volume up to 18 months prior to the patient receiving a clinical diagnosis of retinopathy (Mean 220.8 days prior to clinical diagnosis), the algorithm demonstrated the following performance: Accuracy: 0.987 (95% CI: 0.962-1.00), Sensitivity: 1.00 (95% CI: 0.833-1.00), Specificity: 0.983 (95% CI: 0.952-1.00), PPV: 0.944 (95% CI: 0.836-1.00), NPV: 1.00 (95% CI: 0.937-1.00), (Supplemental Figure 6). Retinopathy-positive eyes showed an upward trend in the LRS as it approached the date of clinical diagnosis with the average eye being identified as positive 2.74 years prior to a clinical

diagnosis (Figure 3, Figure 4). The performance of the model was upheld even when presented with the pericentral retinopathy phenotype an East Asian patient (Figure 4).

<<FIGURE 3 LOCATION>>

<<FIGURE 4 LOCATION>>

Discussion

Automated detection of hydroxychloroquine retinopathy may offer a means to efficiently screen patients for the condition that matches or exceeds expert-level detection. Here we report a deep-learning algorithm for hydroxychloroquine retinopathy screening that can diagnose retinopathy at all stages of the disease as well as predict future retinopathy.

The work builds on previous research that applies machine learning to hydroxychloroquine retinopathy. Tharindu used annotated SD-OCT images from 168 eyes of patients taking hydroxychloroquine to train a deep learning model to identify ellipsoid zone loss, which had a precision of 0.90, recall 0.88, and F1 score 0.89, comparable to human experts.[17] A separate research group used a random forest classifier to detect hydroxychloroquine retinopathy in a data set that included 38 positive patients. The model used inputs of clinical data and SD-OCT biomarkers like retinal layer thicknesses generated by a semi-automated, human-corrected segmentation pipeline.[18] A third group used deep-learning to analyze hand-selected regions of color fundus photographs achieving 97% accuracy in a data set that included 25 patients with hydroxychloroquine retinopathy.[19]

The work in this report offers several advances from prior studies. First, it was developed on a large data set that included 171 patients positive for hydroxychloroquine retinopathy. This meant the model could learn from a wide distribution of examples, including a diversity of race, ethnicity, gender, and age. Second, to guard against overfitting and other deficits in model generalization, the algorithm was validated on three external data sets obtained at different clinical locations across two countries and two SD-OCT devices with a wide variety of acquisition parameters. In addition, case evaluation was performed by a large group of retina

specialists at multiple institutions. Furthermore, the algorithm's analysis could be fully automated after the export of a raw SD-OCT study without the need for image annotation, human evaluation, or input of tabular data. Across these diverse data sets, performance of the model was excellent, missing no true positive cases of retinopathy while maintaining a high positive predictive value - essential characteristics of a useful screening test. Most importantly, in addition to the diagnosis of hydroxychloroquine retinopathy, the algorithm could predict the onset of retinopathy several years in advance of the clinical diagnosis.

As a foundation for these prediction results, Melles and Marmor demonstrated that hydroxychloroquine can be stably tolerated by many patients for years until there is an abrupt decompensation towards retinopathy manifested as rapid retinal thinning.[6] This decompensation can be recognized several years before conventional SD-OCT signs of toxicity through serial measurements of retinal thickness. Interestingly, the authors discuss the implications of clinical decision making following the detection of toxicity that precedes deficits in visual acuity or visual field, or even frank structural changes on SD-OCT. As they state, "*should we recommend discontinuation of hydroxychloroquine, the use of a lower dose, or simply more frequent follow-up? Is the goal of screening to prevent central visual loss or to prevent any structural damage to the retina?*".[6]

A better ability to determine how quickly retinopathy may progress and which patients will continue to progress after cessation of hydroxychloroquine is needed to answer this question. However, an initial approach to incorporating a predictive algorithm into screening workflows could simply involve catching retinopathy at the time of its earliest human-recognizable structural changes on SD-OCT. If the algorithm gave a positive result in a patient without apparent structural abnormalities, the clinician could obtain VF testing and fundus

autofluorescence to seek additional supportive evidence of retinopathy. If no additional abnormalities were seen on ancillary testing, the SD-OCT imaging and algorithmic analysis could then be repeated serially every 4-6 months monitoring for progressive increase in the LRS and early signs of retinopathy.[9]

In our work we have not delineated which of the most salient image features the model uses to make its assessment. Several false positive and false negative cases suggest that the model is attending to outer retinal irregularities. However, LRS outside of the areas of ellipsoid zone abnormality are also elevated in many cases suggesting that other cues may be present. Additional manual analyses of the large hydroxychloroquine data set that has been assembled may further validate the outer retinal abnormalities or other human-interpretable biomarkers.[20,21] Saliency maps and other explainability approaches in machine learning could be employed, however, there are limitations to these methods.[22]

Nevertheless, the model did demonstrate face validity in several respects. Hydroxychloroquine retinopathy is a bilateral, predominantly symmetric condition, which was reflected in the high correlation of LRS between the right and left eyes involved in the study. Additionally, it is a slowly progressive condition, which is represented in the steady increase in likelihood scores over the years preceding clinical diagnosis of retinopathy. The time frame during which the LRS rise are also in accordance with previous work demonstrating the onset of anatomical changes in eyes with retinopathy in many cases 4-6 years in advance of ellipsoid zone loss.[6]

Despite extensive validation, our work was performed as a retrospective analysis and thus needs prospective validation. The rarity of the disease would require large patient cohorts for such a clinical study, a significant hindrance to prospective validation. Furthermore, validation of the prediction of future clinical retinopathy requires years of patient surveillance. Although,

enrichment strategies could be used in prospective validation studies (i.e. years of hydroxychloroquine use), and the low marginal cost of algorithm deployment could allow its testing across many clinical sites in parallel, translating the algorithm into real-world clinical remains challenging due to the time and money required for regulatory approval for Software as a Medical Device (SaMD) in the United States and many other countries. Further work may include an economic evaluation of the algorithm in the eye clinic to justify any further financial investment.

References

1. Erkan D, Yazici Y, Harrison MJ, Paget SA. Physician treatment preferences in rheumatoid arthritis of differing disease severity and activity: the impact of cost on first line therapy. Arthritis Care & Research. 2002 Jun 15;47(3):285-90.

2. Alarcón GS, McGwin G, Bertoli AM, Fessler BJ, Calvo-Alén J, Bastian HM, Vilá LM, Reveille JD. Effect of hydroxychloroquine on the survival of patients with systemic lupus erythematosus: data from LUMINA, a multiethnic US cohort (LUMINA L). Annals of the rheumatic diseases. 2007 Sep 1;66(9):1168-72.

3. Melles RB, Marmor MF. The risk of toxic retinopathy in patients on long-term hydroxychloroquine therapy. JAMA ophthalmology. 2014 Dec 1;132(12):1453-60.

4. Lotery A, Yusuf I, Foot B. Hydroxychloroquine and chloroquine retinopathy: recommendations on monitoring. London: The Royal College of Ophthalmologists, 2020.

5. Pearce FA, Grainge MJ, King AJ, Lanyon PC. O39 Implementing screening for hydroxychloroquine ocular toxicity: how big is the problem? Epidemiology of hydroxychloroquine prescriptions in the UK Clinical Practice Research Datalink. Rheumatology. 2019 Apr 1;58(Supplement_3):kez105-037

6. Melles RB, Marmor MF. Rapid macular thinning is an early indicator of hydroxychloroquine retinal toxicity. Ophthalmology. 2022 Sep 1;129(9):1004-13.

7. Melles RB, Marmor MF. Pericentral retinopathy and racial differences in hydroxychloroquine toxicity. Ophthalmology. 2015 Jan 1;122(1):110-6.

8. Marmor MF, Hu J. Effect of disease stage on progression of hydroxychloroquine retinopathy. JAMA ophthalmology. 2014 Sep 1;132(9):1105-12.

9. Marmor MF, Kellner U, Lai TY, Melles RB, Mieler WF. Recommendations on screening for chloroquine and hydroxychloroquine retinopathy (2016 revision). Ophthalmology. 2016 Jun 1;123(6):1386-94.

10. Rosenbaum JT, Costenbader KH, Desmarais J, Ginzler EM, Fett N, Goodman SM. ACR, AAD, RDS, and AAO 2020 joint statement on hydroxychloroquine use with respect to retinal toxicity. Arthritis Rheumatol. 2021 Feb 9;73(6):0-3.

11. Yusuf IH, Foot B, Lotery AJ. The Royal College of Ophthalmologists recommendations on monitoring for hydroxychloroquine and chloroquine users in the United Kingdom (2020 revision): executive summary. Eye. 2021 Jun;35(6):1532-7.

12. Jayakar G, De Silva T, Cukras CA. Visual Field Sensitivity Prediction Using Optical Coherence Tomography Analysis in Hydroxychloroquine Toxicity. Investigative Ophthalmology & Visual Science. 2022 Jan 3;63(1):15-.

13. Lally DR, Heier JS, Baumal C, Witkin AJ, Maler S, Shah CP, Reichel E, Waheed NK, Bussel I, Rogers A, Duker JS. Expanded spectral domain-OCT findings in the early detection of hydroxychloroquine retinopathy and changes following drug cessation.

International journal of retina and vitreous. 2016 Dec;2:1-1.

14. Allahdina AM, Chen KG, Alvarez JA, Wong WT, Chew EY, Cukras CA. Longitudinal changes in eyes with hydroxychloroquine retinal toxicity. Retina. 2019 Mar 1;39(3):473-84

15. Marmor MF, Melles RB. Disparity between visual fields and optical coherence tomography in hydroxychloroquine retinopathy. Ophthalmology. 2014 Jun 1;121(6):1257-62.

16. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. InInternational conference on machine learning 2019 May 24 (pp. 6105-6114). PMLR.

17. De Silva T, Jayakar G, Grisso P, Hotaling N, Chew EY, Cukras CA. Deep learning-based automatic detection of ellipsoid zone loss in spectral-domain OCT for hydroxychloroquine retinal toxicity screening. Ophthalmology Science. 2021 Dec 1;1(4):100060.

18. Kalra G, Talcott KE, Kaiser S, Ugwuegbu O, Hu M, Srivastava SK, Ehlers JP. Machine learning–based automated detection of hydroxychloroquine toxicity and prediction of future toxicity using higher-order OCT biomarkers. Ophthalmology Retina. 2022 Dec 1;6(12):1241-52.

19. Fan WS, Nguyen HT, Wang CY, Liang SW, Tsao YM, Lin FC, Wang HC. Detection of Hydroxychloroquine Retinopathy via Hyperspectral and Deep Learning through

Ophthalmoscope Images. Diagnostics. 2023 Jul 14;13(14):2373.

20. Ugwuegbu O, Uchida A, Singh RP, Beven L, Hu M, Kaiser S, Srivastava SK, Ehlers JP. Quantitative assessment of outer retinal layers and ellipsoid zone mapping in hydroxychloroquine retinopathy. British Journal of Ophthalmology. 2019 Jan 1;103(1):3-7.

21. Membreno RF, De Silva T, Agrón E, Keenan TD, Cukras CA. Quantitative analysis of optical coherence tomography imaging in patients with different severities of hydroxychloroquine toxicity. British Journal of Ophthalmology. 2023 Jun 1;107(6):849-55.

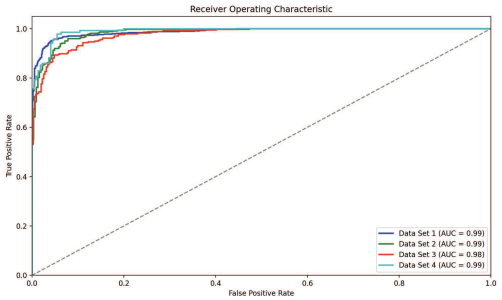
22. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019 May;1:206–215.

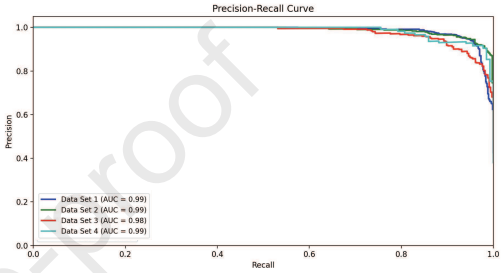
23. Marmor MF, Melles RB. Disparity between visual fields and optical coherence tomography in hydroxychloroquine retinopathy. Ophthalmology. 2014 Jun 1;121(6):1257-62.

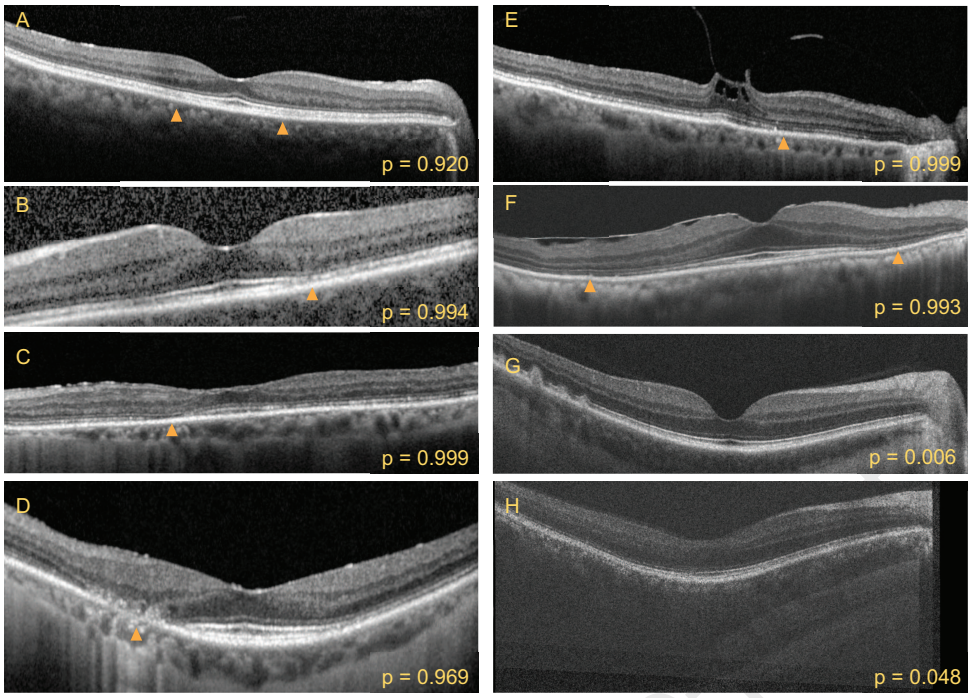
Characteristics of training and test set data	Data set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5	Overall
Role in model development	Training, internal validation	Training	External validation	External validation	External validation	
Location of acquisition	Byers Eye Institute, Stanford, CA, USA	Duke Eye Center, Durham, NC, USA	South Durham, NC, USA	Moorfields Eye Hospital, London, UK	Newcastle-upon-Tyne Hospital, UK	
OCT Device	Zeiss Cirrus	Heidelberg Spectralis	Heidelberg Spectralis	Heidelberg Spectralis	Heidelberg Spectralis	
OCT Protocol	5-line HD, 1-line HD. Variable angle of imaging from 0-7 degrees	61-line, 25 x 30 degree raster	61-line, 25 x 30 degree raster	Multiple line counts (19-61), multiple scan dimensions (20, 25, 30 degrees)	Multiple line counts (19-61), multiple scan dimensions (20, 25, 30 degrees)	
Model training and validation (toxicity positive)						
Number of patients	153 (62)	111 (42)	0 (0)	0 (0)	0 (0)	264 (104)
Number of eyes	302 (121)	252 (84)	0 (0)	0 (0)	0 (0)	554 (205)
Number of b scans	2450 (1225)	2450 (1225)	0 (0)	0 (0)	0 (0)	4900 (2450)
Model testing (toxicity positive)						
Number of patients	37 (15)	0 (0)	22 (13)	49 (22)	37 (17)	145 (67)
Number of eyes	73 (29)	0 (0)	43 (25)	97 (43)	73 (34)	286 (131)
Number of b scans	355 (135)	0 (0)	731 (425)	1579 (671)	686 (335)	3351 (1566)
Age, years (SD)	59.5 (14.7)	*No patients in test set*	63.7 (14.1)	63.2 (11.7)	58.4 (13.1)	61.6 (13.5)
Female sex (n)	91.8% (34)		90.9% (20)	81.6% (40)	97.4% (37)	87.0% (94)
Self-reported ethnicity (n)						
White	54.1% (20)		72.7% (16)	16.3% (8)	75.7% (28)	49.6% (72)
Black	5.4% (2)		27.3% (6)	14.2% (7)	0% (0)	10.3% (15)
Asian	27.0% (10)		0% (0)	20.4% (10)	5.4% (2)	15.2% (22)
Other/Unknown	13.5% (5)		0% (0)	49.0% (24)	18.9% (7)	24.8% (36)

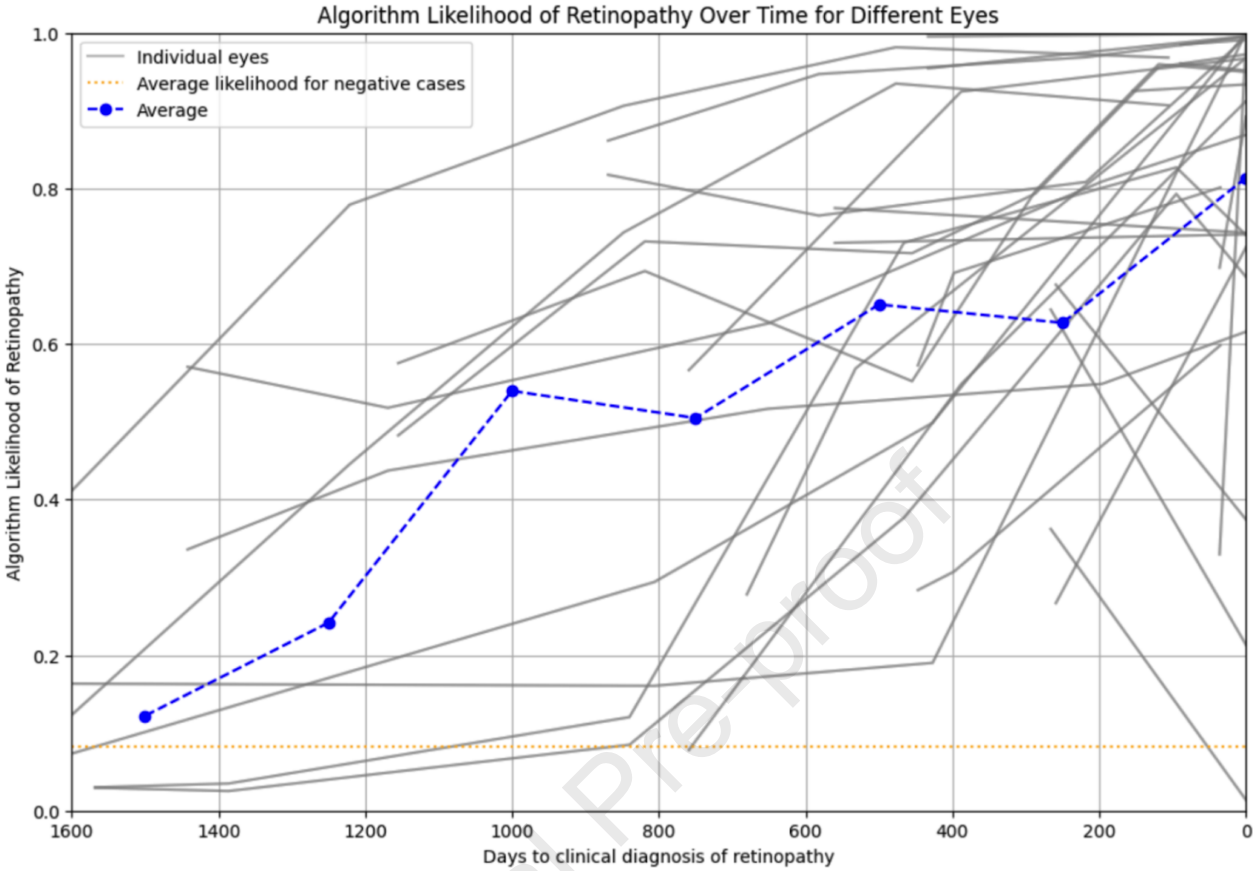
Algorithm	Eye Ground Truth +	Eye Ground Truth -	Patient Ground Truth +	Patient Ground Truth -
+	100	10	52	5
-	2	101	0	51

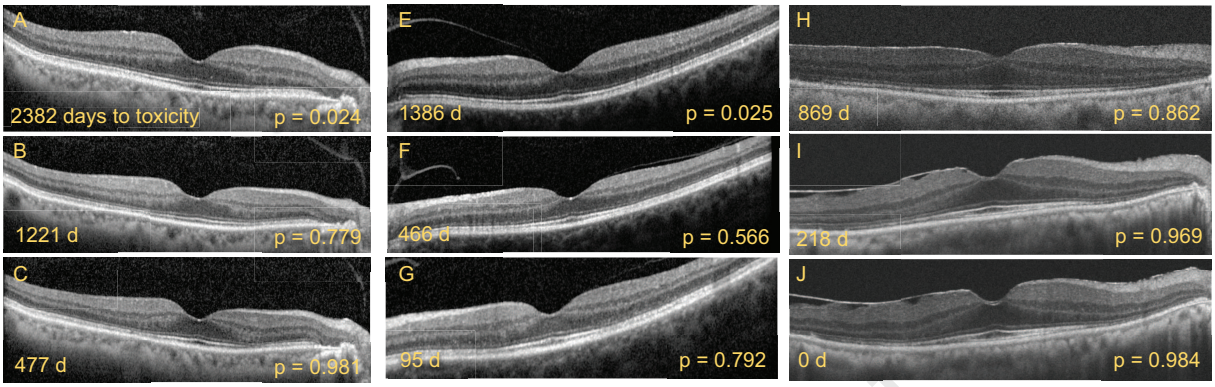
Eye	Patient
Sens = 0.980	Sens = 1.000
Spec = 0.910	Spec = 0.911
Accu = 0.943	Accu = 0.954
PPV = 0.910	PPV = 0.912
NPV = 0.981	NPV = 1.000











A deep learning algorithm, validated on data from multiple institutions, can detect hydroxychloroquine retinopathy with high accuracy from spectral-domain optical coherence tomography images at all stages of disease and in advance of clinical diagnosis.