AMERICAN ACADEMY OF OPHTHALMOLOGY®

Artificial Intelligence-Based Disease Activity Monitoring to Personalized Neovascular Age-Related Macular Degeneration Treatment: A Feasibility Study

Zufar Mulyukov, PhD,[1] *Pearse A. Keane, FRCOphth, MD,*[2,3] *Jayashree Sahni, FRCOphth, MD,*[1] *Sandra Liakopoulos, MD,*[4,5] *Katja Hatz, MD,*[6,7] *Daniel Shu Wei Ting, MD, PhD,*[8] *Roberto Gallego-Pinazo, MD, PhD,*[9] *Tariq Aslam, PhD, DM(Oxon),*[10] *Chui Ming Gemmy Cheung, FRCOphth, MD,*[11,12] *Gabriella De Salvo, FRCOphth, MD,*[13] *Oudy Semoun, MD,*[14,15] *Gábor Márk Somfai, MD, PhD,*[16,17,18] *Andreas Stahl, MD,*[19] *Brandon J. Lujan, MD,*[20] *Daniel Lorand, MSc*[1]

Purpose: To evaluate the performance of a disease activity (DA) model developed to detect DA in participants with neovascular age-related macular degeneration (nAMD).

Design: Post hoc analysis.

Participants: Patient dataset from the phase III HAWK and HARRIER (H&H) studies.

Methods: An artificial intelligence (AI)-based DA model was developed to generate a DA score based on measurements of OCT images and other parameters collected from H&H study participants. Disease activity assessments were classified into 3 categories based on the extent of agreement between the DA model's scores and the H&H investigators' decisions: agreement ("easy"), disagreement ("noisy"), and close to the decision boundary ("difficult"). Then, a panel of 10 international retina specialists ("panelists") reviewed a sample of DA assessments of these 3 categories that contributed to the training of the final DA model. A panelists' majority vote on the reviewed cases was used to evaluate the accuracy, sensitivity, and specificity of the DA model.

Main Outcome Measures: The DA model's performance in detecting DA compared with the DA assessments made by the investigators and panelists' majority vote.

Results: A total of 4472 OCT DA assessments were used to develop the model; of these, panelists reviewed 425, categorized as "easy" (17.2%), "noisy" (20.5%), and "difficult" (62.4%). False-positive and false negative rates of the DA model's assessments decreased after changing the assessment in some cases reviewed by the panelists and retraining the DA model. Overall, the DA model achieved 80% accuracy. For "easy" cases, the DA model reached 96% accuracy and performed as well as the investigators (96% accuracy) and panelists (90% accuracy). For "noisy" cases, the DA model performed similarly to panelists and outperformed the investigators (84%, 86%, and 16% accuracies, respectively). The DA model also outperformed the investigators for "difficult" cases (74% and 53% accuracies, respectively) but underperformed the panelists (86% accuracy) owing to lower specificity. Subretinal and intraretinal fluids were the main clinical parameters driving the DA assessments made by the panelists.

Conclusions: These results demonstrate the potential of using an AI-based DA model to optimize treatment decisions in the clinical setting and in detecting and monitoring DA in patients with nAMD.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2024;4:100565*

Supplemental material available at www.ophthalmologyscience.org.

Neovascular age-related macular degeneration (nAMD) is a long-term condition and a common cause of severe and irreversible blindness due to the growth of choroidal neovascularization and associated exudation of blood and serous fluid in the macula.[1–3]

VEGF is a potent angiogenic factor and a key driver of nAMD.[2,4] Treating nAMD with VEGF inhibitors can substantially reduce the degree of vision loss compared with the previous standard of care (e.g., laser photocoagulation and photodynamic therapy).[3,5] Still, anti-VEGF therapy is

https://doi.org/10.1016/j.xops.2024.100565
ISSN 2666-9145/24

associated with a substantial treatment burden owing to several factors, including frequent repeated intravitreal injections and clinic monitoring visits.[3,6] Numerous clinical setting studies of anti-VEGF therapy in nAMD have reported visual outcomes inferior to those reported in randomized clinical trials.[3,7–9] Treatment protocols that provide treatment only when disease activity (DA) is present ("pro re nata") or whereby patients are treated and monitored in extended intervals while observing absence of DA ("treat and extend") have been used to lessen the treatment burden for patients and clinics.[10,11]

Fluid exudation in nAMD is a key target for anti-VEGF therapy. OCT has become the recommended imaging modality for initial assessment and monitoring of treatment response to anti-VEGF agents in the qualitative assessment of DA, based on the presence of markers of fluid exudation such as subretinal fluid (SRF), intraretinal fluid (IRF) (within cystoid spaces or diffuse retinal thickening), increase in pigment epithelial detachments, and hemorrhage.[2,12–14] Recent clinical trials have defined DA assessment based on qualitative exudative OCT features, such as central subfield thickness (CSFT), and best-corrected visual acuity (BCVA) to plan retreatment or extend the treatment interval.[15–18]

The evaluation of OCT images may depend on the expertise/experience of the treating physician and, therefore, could be subject to considerable interphysician variability in OCT interpretation.[19,20] This was observed in the Comparison of Age-related Macular Degeneration Treatments trial, where the grading of macular fluid by a reading center was compared with an ophthalmologist's treatment decision based on OCT-guided macular fluid identification.[20] It was observed that the ophthalmologist's treatment decision and reading center fluid determination from OCT images disagreed in 27.9% of patient visits on presence or absence of macular fluid.[20] These observations highlight that failure to detect DA correctly may lead to undertreatment with a subsequent decline in visual acuity.[19]

Artificial intelligence (AI) approaches based on deep learning (DL) methods, with their capacity to detect and quantify exudative fluid using OCT automatically, have great potential as a diagnostic tool in the complex nAMD treatment landscape.[14,21] Several groups have begun to develop algorithms for predicting treatment response and optimizing treatment regimens in patients with nAMD.[22–24] However, validating and implementing AI in clinical decision-making remains challenging.[21]

A tool designed to guide and support the clinician in considering a particular dosing regimen could result in more effective treatment and optimized visual outcomes. Here, we have developed and evaluated the performance of an AI-based DA model capable of assessing DA in nAMD at each clinic visit to optimize treatment.

Methods

This is a post hoc analysis of the phase III HAWK and HARRIER (H&H) double-masked clinical trials (ClinicalTrials.gov identifiers: NCT02307682 and NCT02434328, respectively). These 2 similarly designed studies were conducted to compare the efficacy and safety of intravitreal brolucizumab, a single-chain antibody fragment that inhibits VEGF-A, with aflibercept in treatment-naive patients with nAMD.[15,25] Figure S1 (available at www.ophthalmologyscience.org) illustrates the study design of the H&H studies. In the brolucizumab arms of the H&H studies, maintenance treatment began with a 12-week treatment interval after 3 loading doses. The interval could be shortened to an 8-week interval (every 8 weeks) based on DA assessments at predefined visits. Eyes of patients treated with brolucizumab demonstrated greater fluid resolution from week 48 to week 96 compared with aflibercept. The H&H studies were conducted in accordance with principles of the Declaration of Helsinki, International Conference on Harmonization E6 Good Clinical Practice Consolidated Guideline, and other regulations as applicable. Before starting the studies, an independent ethics committee/institutional review board approved the study protocols and written informed consent was obtained from all participants.

This post hoc analysis aimed to (1) develop a model capable of assessing DA in participants with nAMD undergoing anti-VEGF therapy and (2) evaluate the model's performance against the DA assessments made by the H&H investigators and an independent panel of retina specialists.

DA Model—Development

The DA model was developed using visual acuity data and morphologic features extracted from OCT images (Figure S2; available at www.ophthalmologyscience.org). A subset of data from the DA assessments performed by the H&H investigators were used for model training. The first DA assessment in the H&H studies occurred at week 16, followed by the assessments at predefined later visits. Once switched to the every 8 weeks regimen, patients remained on this treatment interval until the end of the study. Therefore, this post hoc analysis only includes data from DA assessments conducted before and at the point of switch to an every 8 weeks regimen. Morphologic features, such as fluid volumes, were extracted from spectral-domain (SD)-OCT images acquired with Spectralis SD-OCT (Heidelberg Engineering) with volumes consisting of 49 high-speed B-scans (512 A-scans per B-scan). This additional restriction narrowed down the training data to only Spectralis SD-OCT images, representing the majority (67%) of the OCT images from the H&H studies.

Firstly, a DL algorithm was applied to raw OCT images to measure anatomic features such as IRF, SRF, and pigment epithelial detachment volumes, thickness of retinal layers, and central subfield thickness.[26,27] These data were combined with the corresponding nonimaging data, such as BCVA, time since the last injection, and patient characteristics and medical history. A tree-based machine learning algorithm was applied to the combined data from visits before and at DA assessment to build a classification model of DA. The model used the gradient-boosted decision trees approach implemented by the XGBoost package version 1.4 that generates a prediction model from an ensemble of weakly predictive decision trees. The model

was trained using cross-validation on the entire dataset of DA assessments. The most parsimonious classification model used 11 measurements of DA: IRF, SRF, and pigment epithelial detachment volumes in the scanned 6 × 6 mm area, central subfield thickness, BCVA at the time of DA assessment, changes of these features from the previous visit when treatment was received, and the time since the last injection. OCT images taken at DA assessment visits and previous injection visit were used to predict DA. The main output of the DA model was a continuous metric (DA score), representing the probability of a retina specialist deciding whether DA is present (see section titled *DA Model—Independent Case Review for Performance Evaluation*).

A sample of the DA assessments, which served as the basis for training the models, underwent reevaluation by an independent panel of retina specialists. This subset of reevaluated DA assessments spanned different DA categories defined by the level of agreement between the model predictions and the DA assessment conducted by the H&H investigators. This sampling of DA assessment cases aimed to investigate interpanelist agreement across these DA categories, derive consensus-based DA assessments to correct possible errors made by H&H investigators, and evaluate the model's performance within these categories. Following the completion of the reevaluation process for selected DA cases, the final DA model underwent retraining using updated DA assessments in the training data, where corrections were made to some of the assessments based on the majority vote from the panel. Cases were categorized into 3 types—"easy," "noisy," and "difficult"—using model-derived DA scores and the extent of agreement with investigators' decisions.

"Easy" cases were defined as cases where there was no contradiction between the DA assessments made by the investigators and those made by the model predictions: cases where the model predicted a low DA score when investigators did not indicate DA, or a high DA score when DA was indicated by investigators. For "easy" cases, an agreement was expected between the investigators and the majority-voted decision generated by the panelists. "Easy" cases were not expected to influence the model performance evaluation metrics as training labels would not be changed, and consequently, only a small number were selected for review by the panelists for confirmation purposes.

"Noisy" cases were defined as cases where there was contradiction between the DA assessments made by the investigators and those predicted by the model: cases where the investigators indicated high DA and the model predicted a low DA score and vice versa. "Noisy" cases could prevent optimal model training and impact performance evaluation (e.g., due to true positives being counted as false positives); therefore, a majority of these cases were included in the review process by panelists.

"Difficult" cases were defined as cases where the DA scores predicted by the model varied most between cross-validation runs on different subsets of selected data. As DA score variability reflects heterogeneity between DA assessments by the investigators, higher level of disagreement between panelists was expected for "difficult" cases. We defined the top quintile of cases with most variable DA scores as "difficult." "Difficult" cases were expected to be the most informative from which to evaluate the variability of DA assessments between panelists; thus, the majority of cases selected for review were from this category.

DA Model—Independent Case Review for Performance Evaluation

An independent panel of 10 international retina specialists (S.L., K.H., R.G.P., T.A., C.M.G.C., G.D.S., O.S., G.M.S., A.S., B.J.L.) (panelists) with special interest in retinal imaging reviewed a subset of DA assessments from the H&H studies. The panelists did not have access to the outcome of the DA assessments determined by the investigators and reassessed DA in a similar treatment-masked setting implemented in the H&H studies. Panelists did have access to the participant information available to the investigators: participant-level data (demographics, medical history, concomitant medications, and adverse events and serious adverse events); OCT scans of the study eye (raw OCT images [standard density scans]; and other ocular features [longitudinal BCVA, intraocular pressure, presence/absence of retinal hemorrhage, and retinal fibrosis and atrophy]). An interactive web application was used to provide the panel with the data related to the cases (Shiny R package; Rstudio).[28] Access to the OCT images was given through the image analysis platform RetinAI Discovery (RetinAI Medical AG) (Figure S3, available at www.ophthalmologyscience.org), a data management platform that enables storing, viewing, and processing of patient images and data.[29] Panelists did not have access to all features of RetinAI Discovery.

Additionally, an assessment form was provided within the interactive web application to capture panelists' input. This form was carefully designed to adhere to the decision-making in the H&H studies (Figure S4, available at www.ophthalmologyscience.org). Disease activity was deemed "present" if the panelist indicated that treatment was needed, and "absent" if the panelist indicated no treatment was required. The panelist was also asked to note their confidence level in the treatment decision they had made (low, medium, or high), the importance of features impacting the treatment decision (low, medium, or high), and the feature characteristics, alongside any other factors affecting the decisions.

Disease activity scores generated by the model were calibrated to represent the probability of DA presence using DA assessment data generated during the independent case review. A logistic regression model was used to fit the original model DA scores to the panelists' DA assessments. The resulting logistic transformation was applied to all DA scores initially generated by the model, including cases that were not reviewed. After this transformation, DA scores of 0.1, 0.5, and 0.9 represent 10%, 50%, and 90% probability of having DA detected by a retina specialist, respectively.

The Krippendorff's alpha reliability estimate was used to determine the inter-rater agreement (agreement among the panelists).[30]

DA Model—Performance Evaluation Metrics

The model's performance was evaluated by comparing its ability to determine the DA against the majority-voted

decisions of reviewed cases. The main metrics of performance evaluation were cross-validated accuracy, sensitivity, and specificity. A majority vote was used to summarize the panelists' DA assessments for each case, and ties were resolved using the investigators' assessments.

Results

Dataset

A total of 4472 OCT DA assessment image scans acquired longitudinally from patients across all treatment arms of the H&H studies were used to develop and train the DA model (Figure S2, available at www.ophthalmologyscience.org). Baseline characteristics of the H&H patients included in the DA model (n = 389 [HARRIER], n = 594 [HAWK]) are detailed in Table S1 (available at www.ophthalmologyscience.org). As noted in the Methods, only SD-OCT image scans acquired on a Spectralis SD-OCT device were used. These DA assessments were classified as "difficult" (20.0%), "easy" (76.8%), or "noisy" (3.2%) (Table 2). Of these, 425 DA assessments were selected for independent review by the panelists by random sampling within each category. The DA assessments selected for review purposely represented a small percentage of all "easy" cases (2.1%) and a high percentage of all "noisy" cases (60.8%). "Difficult" cases accounted for less than one-third of all "difficult" cases but represented the majority of cases reviewed (62.4%).

Table 2. Distribution of DA Assessments Reviewed and Available From the H&H Studies According to DA Category

DA Category	All DA Assessments, n (%)	Reviewed DA Assessments, n (%)	Percent of All Assessments Reviewed, %
Difficult	895 (20.0)	265 (62.4)	29.6
Easy	3434 (76.8)	73 (17.2)	2.1
Noisy	143 (3.2)	87 (20.5)	60.8
Total	4472 (100)	425 (100)	9.5

DA = disease activity; H&H = HAWK and HARRIER.
Percentages may not total 100% because of rounding.

Model Performance

The cross-validated contingency table evaluating the DA model's performance (based on the 4472 DA assessments) showed a decrease in false-positive and false negative rates, respectively, from 13.5% and 9.6% before review to 10.3% and 5.9% after review (Table 3). The area under the curve was 0.958 before review and 0.981 after review.

The performances (accuracy, sensitivity, and specificity based on the 425 reviewed DA assessments) of the DA model, investigators, and each panelist were compared against the panelists' majority vote for "easy," "noisy," and "difficult" DA categories (Fig 5). Overall, the DA model achieved an accuracy of 80%, a sensitivity of 83%, and a specificity of 76%. For "easy" cases, the model performed as well as the investigators and individual panelists (accuracy of 96%, 96%, and 90%, respectively). The model outperformed the investigators for both "noisy" (accuracy of 84% and 16%, respectively) and "difficult" cases (accuracy of 74% and 53%, respectively). The model performed similarly to panelists for "noisy" cases (accuracy of 84% and mean accuracy of 86%, respectively) but underperformed the panelists for "difficult" cases (accuracy of 74% and mean accuracy of 86%, respectively) due to lower specificity (44%).

Analysis of the DA Model Scores

The distribution of the DA model scores was plotted against the investigators' assessment decisions and the majority vote of the panelists (Fig 6). For "easy" cases, the scores assigned by the DA model were in close agreement with the investigators' decisions and the majority vote of the panelists; the DA model clearly distinguished between the cases where the presence or absence of DA was assigned by investigators or the majority vote of panelists. For "difficult" cases, the distribution of DA model scores was broader than for "easy" cases and did not distinguish between the assessments made by the investigators. In contrast, the DA model was better aligned with the assessments made by the majority vote of panelists. For "noisy" cases, the investigators and the majority vote of panelists largely disagreed in their DA assessments; for these cases, the DA model largely agreed with the majority vote of panelists.

Interpanelist Variability of Reviewed Cases

Investigators' and panelists' majority vote agreed with regard to their DA assessments for most "easy" cases (96%) (Table 4). Conversely, the proportion of DA assessments for which the panelists reversed the investigators' decisions was about half for "difficult" cases (47%) and even higher for "noisy" cases (84%). In 67% of cases, panelists agreed unanimously in their assessments. The highest agreement was for "noisy" and "easy" cases (71% in both categories) followed by "difficult" (64%) (Table 5). Ties between panelists only occurred in 19 reviewed cases (4.5%) (data not shown).

Table 3. DA Model's Performance (Contingency Table) Based on All DA Assessments Used for DA Model Training (i.e., before and at the Switch to the q8w Dosing Regimen)

	Model	Physician "Yes"	Physician "No"	Total
Prereview	Yes, n (%)	1276 (90.4)	412 (13.5)	1688 (37.7)
	No, n (%)	135 (9.6)	2649 (86.5)	2784 (62.3)
	Total, n	1411	3061	4472
Postreview	Yes, n (%)	1378 (94.1)	310 (10.3)	1688 (37.7)
	No, n (%)	86 (5.9)	2698 (89.7)	2784 (62.3)
	Total, n	1464	3008	4472

DA = disease activity; q8w = every 8 wks.
Physician = HAWK and HARRIER investigators (nonreviewed cases) or the majority vote of panelists (reviewed cases).

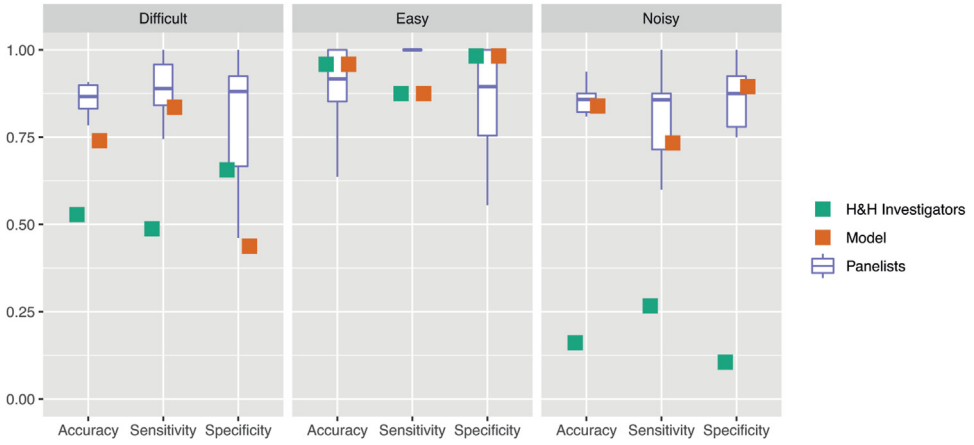

Figure 5. Accuracy, sensitivity, and specificity achieved by the H&H investigators, DA model, and individual panelists from the panel against the majority vote of panelists. A total of 425 DA assessments were evaluated. The boxes show interquartile ranges. The horizontal line across each box denotes the median, and vertical lines extending above and below each box indicate the minimum and maximum values. DA = disease activity; H&H = HAWK and HARRIER.

Clinical Parameters Driving the Panelists' DA Assessments

Analysis of the distribution of the responses provided by the panelists on the DA assessment review forms found that IRF and SRF were the main clinical parameters leading to positive DA assessments. SRF and IRF would be highly important in 86% and 78% of DA assessments, respectively (Table 6).

In addition, analysis of the distribution of the panelists' responses on the dynamics or persistence of the factors leading to positive DA assessments found that new or increased retinal fluid had a greater impact on determining positive DA assessments than static retinal fluid. For example, considering the status of IRF, in 91% of cases, the increase of IRF would be regarded as a driving factor for positive DA assessment (Table 7).

Discussion

In this study, we developed and evaluated the performance of a classification model based on features extracted from OCTs using DL methods in detecting DA in participants with nAMD treated with anti-VEGF therapies. To achieve diverse sampling in the cases reviewed by the panelists, the cases selected were within 3 categories: "easy," "noisy," and "difficult." Sampling of "easy" cases was needed to demonstrate full or close agreement between panelists. Only a limited number of "easy" cases were reviewed. "Noisy" cases could have contaminated the model evaluation results by giving false positives or false negatives; a majority of these cases were reviewed. Sampling of "difficult" cases, which were close to the DA model's decision boundary, allowed to relate the variability in DA assessments between panelists to that between investigators.

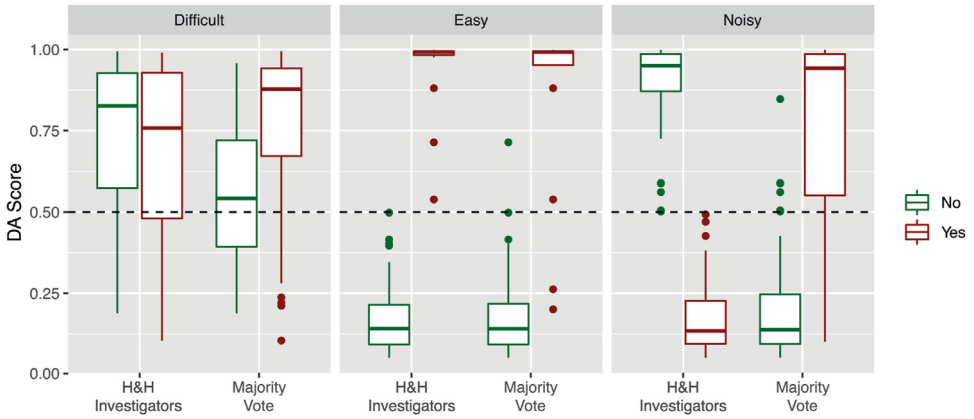

Figure 6. Disease activity score distributions according to the DA assessments from the H&H investigators or the majority vote of the panel of the retina specialists (a total of 425 DA assessments were evaluated). Disease activity score of 0.50 corresponds to the model's positive and negative DA assessment cut-off point. The boxes show interquartile ranges. The horizontal line across each box denotes the median, and vertical lines extending above and below each box indicate the minimum and maximum values. Dots above and below the boxes are outliers. DA = disease activity; H&H = HAWK and HARRIER.

Table 4. Investigators' DA Assessments versus DA Assessments by the Panelists (Majority Vote) According to DA Category

DA Category	"No" to "Yes," n (%)	Unchanged, n (%)	"Yes" to "No," n (%)
Difficult	103 (38.9)	140 (52.8)	22 (8.3)
Easy	2 (2.7)	70 (95.9)	1 (1.4)
Noisy	22 (25.3)	14 (16.1)	51 (58.6)
All	127 (29.9)	224 (52.7)	74 (17.1)

DA = disease activity.
"No" and "Yes" refer to the absence and presence of DA, respectively, i.e., "No" to "Yes" refers to cases where the investigators assigned absence of DA but changed to presence of DA by panelists, and "Yes" to "No" refers to cases where the investigators assigned presence of DA but changed to absence of DA by panelists.

Overall, the false-positive and false negative rates (based on the 4472 DA assessments and after incorporating the review by the panelists) were low: 10.3% and 5.9%, respectively. Based on the reviewed cases, when comparing against the majority vote of panelists, the model outperformed the investigators in correctly assessing DA in "noisy" and "difficult" cases, while the DA model closely agreed with the investigators' DA assessments and the panelists' majority vote for "easy" cases. For "noisy" cases, the DA model largely agreed with the majority panelist vote but disagreed with the investigators, indicating that these were indeed "noisy" cases for which the investigators may have incorrectly assessed DA.

The performance of the DA model noticeably improved after updating the DA assessments following the review by the panelists, even though only a small fraction of DA assessments from the H&H studies were reviewed. The decrease in false-positive and false-negative rates following review was driven by the panelists' majority vote, which resolved a large fraction of disagreements between the DA model and investigators that took place before the review.

The study relied on the assumption that the panelists assessed DA in a way that was consistent with DA assessments made by the investigators. To minimize the risk of systematic differences in DA assessments between the panelists and investigators, the panelists were trained on key aspects of the H&H studies and were presented with data closely matching the data available to the investigators. Additionally, the assessment form was carefully designed to reflect the decision-making in the H&H studies.

The extent of agreement among panelists for the "difficult" cases was markedly lower than for "easy" and "noisy" cases. The "difficult" cases were classified as such owing to variability of DA score during cross-validation caused by variability in DA assessment of similar cases by investigators. This disparity between the investigators in DA assessments for "difficult" cases, and the evidence of some disagreement among the panelists themselves, illustrate the subjective nature of DA assessment and that it may be dependent on the physician's experience and expertise.[31,32] Clinical setting OCT case studies have also shown that OCT findings may be misinterpreted.[33] This subjectivity of the human grader reinforces the utility of an objective, consistent, human-independent DA model trained on data from expert retina specialists.

In this study, a DL-based algorithm was used to quantify retinal fluids in OCT images. Automated quantification of fluids achieves high consistency with manual expert assessments.[34] However, prospective studies are needed to evaluate DL algorithms and their role in guiding treatment decisions in clinical settings.

Interestingly, the presence or absence of IRF only, SRF only, or both was considered particularly important by the panelists, with BCVA or subretinal pigment epithelium fluid having minimal effects on their decisions (Table 6). These findings, alongside the disparity in DA assessment between the investigators and the panelists for "noisy"

Table 5. Agreement between Panelists by DA Categories Presented as Percent of Unanimous DA Assessments and as Krippendorff's Alpha

DA Category	n/N (%)	Alpha (95% CI)
Difficult	170/265 (64.2)	0.43 (0.35–0.49)
Easy	52/73 (71.2)	0.53 (0.36–0.70)
Noisy	62/87 (71.3)	0.55 (0.41–0.66)
Total	284/425 (66.8)	0.54 (0.49–0.59)

CI = confidence interval; DA = disease activity; n = number of unanimously assessed cases; N = number of cases.

Table 6. Distribution of the Panelists' Responses on the Importance of Clinical Parameters Driving Positive DA Assessments

	Importance				
Parameter	*None*	*Low*	*Medium*	*High*	*N*
BCVA	77%	6%	12%	5%	1035
CSFT	40%	17%	23%	20%	1031
IRF	9%	1%	11%	78%	319
SRF	4%	1%	9%	86%	615
Sub-RPE fluid	73%	8%	10%	9%	278

BCVA = best-corrected visual acuity; CSFT = central subfield thickness; DA = disease activity; H&H = HAWK and HARRIER; IRF = intraretinal fluid; N = number of panelists; PED = pigment epithelium detachment; RPE = retinal pigment epithelium; SRF = subretinal fluid.
"High," "Medium," "Low," and "None" indicate the importance of a particular parameter in panelists' decisions. Clinical data regarding CSFT values and the presence/absence of IRF, SRF, and sub-RPE fluid were assessed in the reading centers involved in the H&H studies. PED presence was not assessed.

Table 7. Distribution of Panel Responses on Parameters Driving Positive DA Assessments

Parameter	Characteristic	Percent
BCVA	Decreased	99%
	Low	1%
CSFT	Increased	96%
	High	4%
IRF	New/increased	91%
	Persistent	9%
PED	New/increased	80%
	Persistent	20%
SRF	New/increased	87%
	Persistent	13%
Sub-RPE fluid	New/increased	84%
	Persistent	16%

BCVA = best-corrected visual acuity; CSFT = central subfield thickness; DA = disease activity; IRF = intraretinal fluid; PED = pigment epithelium detachment; RPE = retinal pigment epithelium; SRF = subretinal fluid. Data show the characteristic (low, decreased, persistent, and new/increased) given by panelist to the parameters in the case assessment form (Fig S3).

cases, highlight the need for a universally agreed definition of DA and a consistent OCT interpretation on those studies in which the investigators assess OCT of their participants.

Different studies have demonstrated the potential utility of DL-based algorithms in diagnosing age-related macular degeneration.[26,35] To our knowledge, this is the only model that, by generating a DA score, emulates the consensus on whether to treat or not by experienced retina specialists. While further refinements, improvements, and external validation in a prospective setting would be needed, these initial results highlight the potential of AI to support and improve the quality of DA evaluations and treatment decisions in patients with nAMD. Figure S7 (available at www.ophthalmologyscience.org) shows a mock-up example of the output provided by the DA model. Clinicians can use the DA score, the segmented OCT images, and changes in fluid volumes to guide treatment decisions. Once fully validated, such a DA model could also be used as a clinical decision support tool in clinical trial settings to test the efficacy of different therapies in nAMD. Such a model could even be used as a teaching or training tool for future specialists in nAMD, as well as a diagnostic tool to assess treatment needs in treatment-naive patients.

In conclusion, the results presented here highlight the potential of the DA model to assess DA in patients with nAMD. The DA model could potentially reduce the time and resources needed for monitoring patients with nAMD and improve interpretation consistency, thus providing a powerful tool to optimize treatment decisions at point of care so that each patient is retreated to achieve and maintain the best visual outcome with the lowest possible treatment burden. Digital analysis tools, such as the one reported here, are likely to become essential in the clinical management of patients with nAMD and could be integrated into the existing workflow of retina practices in the near future.

Footnotes and Disclosures

Originally received: October 20, 2023.
Final revision: May 31, 2024.
Accepted: June 11, 2024.
Available online: June 17, 2024. Manuscript no. XOPS-D-23-00249.

[1] Novartis Pharmaceuticals AG, Basel, Switzerland.

[2] NIHR Moorfields Biomedical Research Centre, London, United Kingdom.

[3] UCL Institute of Ophthalmology, London, United Kingdom.

[4] Cologne Image Reading Center and Laboratory, Department of Ophthalmology, Faculty of Medicine and University Hospital Cologne, Cologne, Germany.

[5] Department of Ophthalmology, Goethe University, Frankfurt, Germany.

[6] Vista Augenklinik Binningen, Binningen, Switzerland.

[7] Faculty of Medicine, University of Basel, Basel, Switzerland.

[8] Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore.

[9] Oftalvist Clinic, Valencia, Spain.

[10] School of Pharmacy and Optometry, University of Manchester, Manchester Royal Eye Hospital, Manchester, United Kingdom.

[11] Duke-NUS Academic Clinical Programme, University of Singapore, Singapore, Republic of Singapore.

[12] Singapore National Eye Centre, Singapore, Republic of Singapore.

[13] Ophthalmology Department, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom.

[14] Centre Hospitalier Intercommunal de Créteil, Créteil, France.

[15] Institut d'Ophtalmologie du Panthéon, Paris, France.

[16] Stadtspital Zürich, Department of Ophthalmology, Zürich, Switzerland.

[17] Spross Research Institute, Zürich, Switzerland.

[18] Department of Ophthalmology, Semmelweis University, Budapest, Hungary.

[19] Department of Ophthalmology, University Hospital Greifswald, Greifswald, Germany.

[20] OHSU Casey Eye Institute, Portland, Oregon.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

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

Z.M.: Financial support present manuscript — Novartis (employee); Patent — pending; Stock options — Novartis.

P.A.K.: Financial support present manuscript — Novartis; Consultant— Apellis, AbbVie, Adecco, Boehringer Ingelheim, Novartis, Roche; Payment or honoraria — AbbVie, Alimera, Apellis, Bayer, Boehringer Ingelheim, Gyroscope, Novartis, Thea; Travel expenses— Bayer, Roche; Patents — Google US10198832B2 (issued), Google US20220301152A1 (pending); Data Safety Monitoring Board or Advisory Board participation — AbbVie, Boehringer Ingelheim, Novartis, RetinAI, Roche; Stock — Big Picture Medical; Stock options — Bitfount.

J.S.: Financial support present manuscript — Novartis (employee); Patents — Novartis; Stock options — Novartis (employee); Others — Novartis.

S.L.: Financial support present manuscript — Novartis; Grants or contracts — Novartis Pharma GmbH; Consultant — AbbVie, Apellis, Bayer, Biogen,

Novartis; Payment or honoraria — AbbVie, Apellis, Bayer, Biogen, Heidelberg Engineering, Novartis, Zeiss.

K.H.: Financial support present manuscript — Novartis; Grants or contracts — Allergan/AbbVie, Bayer, Novartis, Roche; Payment and honoraria — Allergan/AbbVie, Bayer, Heidelberg Engineering, Novartis, Roche, Zeiss; Travel expenses — Bayer, Roche; Data Safety Monitoring Board or Advisory Board participation — Bayer, Novartis, Roche.

D.S.W.T: Financial support present manuscript and grants or contracts — Funding body, National Medical Research Council (Singapore), Duke-NUS Medical School, Agency for Science, Technology and Research; Patents planned, issued, or pending — holds a few patents of deep learning systems for retinal diseases.

R.G-P.: Financial support present manuscript and — Novartis; Consultant — Apellis, Boehringer Ingelheim, Carl Zeiss, Heidelberg Engineering, Novartis, ORA Clinical, Roche; Payment or honoraria — Allergan/AbbVie, Bloss Group, Heidelberg Engineering, Novartis, Roche.

T.A.: Financial support present manuscript — Novartis; Grants or contracts — Allergan, Bayer Pharmaceuticals, Canon, NIHR, Topcon, Roche; Consultant — Allergan, Bayer Pharmaceuticals, Canon, Topcon, Roche; Payment or honoraria — Allergan, Bayer Pharmaceuticals, Canon, Topcon, Roche; Travel expenses — Allergan, Bayer Pharmaceuticals, Canon, Topcon, Roche; Leadership or fiduciary role — Bayer, Macular Society, Fight for Sight.

C.M.G.C.: Financial support present manuscript — Novartis; Grants or contracts — Bayer, Boehringer Ingelheim, Roche; Consultant — Bayer, Boehringer Ingelheim, Roche; Payment or honoraria — Allergan, Bayer, Roche, Topcon, Zeiss; Travel expenses — Bayer, Roche; Data Safety Monitoring Board or Advisory Board participation — Boehringer Ingelheim; Stock or stock options — Avirmax.

G.D.S.: Consultant — Boehringer Ingelheim, Roche; Payment or honoraria — AbbVie, Bayer, Heidelberg Engineering, Novartis, Roche; Travel expenses — AbbVie, Bayer, Novartis, Roche; Data Safety Monitoring Board or Advisory Board participation — AbbVie, Apellis, Bayer, Novartis, Teva, Roche.

O.S.: Financial support present manuscript — Novartis; Grants or contracts — Bayer; Consultant, payment or honoraria, payment for expert testimony, Data Safety Monitoring Board or Advisory Board participation — Bayer, Novartis.

G.M.S.: Payment or honoraria — Allergan, Apellis, Bayer, Roche, Zeiss; Data Safety Monitoring Board or Advisory Board participation — Apellis, Bayer, Novartis, Roche, Zeiss.

A.S.: Financial support present manuscript — Novartis; Grants or contracts — Bayer, Novartis; Consultant — Apellis, Bayer, Novartis, Roche; Payment or honoraria — Bayer, Novartis, Roche; Travel expenses — Apellis, Bayer, Novartis, Roche; Data Safety Monitoring Board or Advisory Board participation — ROPROP study; Leadership or fiduciary role — German Retina Society, Retina.net, German Ophthalmological Society; Stock options — SemaThera Inc.; Receipt of equipment, materials, drugs, medical writing, gifts or other services — Bayer, Novartis.

B.J.L.: Financial support present manuscript — Novartis (payments made to Lujan Imaging LLC); Consultant (payments made to Lujan Imaging LLC) — Allergan, Genentech/Roche, Kodiak, Lineage, NGM, Novartis, RegenxBio, Ribomic; Patents planned, issued, or pending — Directional Optical Coherence Tomography (no royalties collected); Stock options — Translational Imaging Innovations.

D.L.: Financial support present manuscript, patents planned, issued, or pending, stock or stock options, other financial or nonfinancial interests — Novartis (employee).

This study was funded by Novartis Pharma AG, Basel, Switzerland. Ana María Rodríguez de Ledesma (Bedrock Healthcare Communications, Fleet, UK) provided medical writing support to the authors, funded by Novartis Pharma AG (Basel, Switzerland).

Daniel Shu Wei Ting, MD, PhD, an Editor of this journal, was recused from the peer-review process of this article and had no access to information regarding its peer-review.

HUMAN SUBJECTS: Human subjects data were included in this study. The H and H studies were conducted in accordance with principles of the Declaration of Helsinki, International Conference on Harmonization E6 Good Clinical Practice Consolidated Guideline, and other regulations as applicable. Before starting the studies, an independent ethics committee/institutional review board approved the study protocols and written informed consent was obtained from all participants.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Mulyukov, Keane, Sahni, Liakopoulos, Hatz, Ting, Pinazo, Lorand

Data collection: Mulyukov, Sahni, Liakopoulos, Hatz, Pinazo, Aslam, Cheung, Salvo, Semoun, Somfai, Stahl, Lujan, Lorand

Analysis and interpretation: Mulyukov, Keane, Sahni, Liakopoulos, Hatz, Pinazo, Lorand

Obtained funding: N/A. Study was performed as part of regular employment duties at Novartis Pharma AG. No additional funding was provided.

Overall responsibility: Mulyukov, Keane, Sahni, Liakopoulos, Hatz, Ting, Pinazo, Aslam, Cheung, Salvo, Semoun, Somfai, Stahl, Lujan, Lorand

Preliminary analysis of this study was presented at the Association for Research in Vision and Ophthalmology (ARVO) Annual Meeting, in June 2022.

Abbreviations and Acronyms:
AI = artificial intelligence; **BCVA** = best-corrected visual acuity; **DA** = disease activity; **DL** = deep learning; **H&H** = HAWK and HARRIER; **IRF** = intraretinal fluid; **nAMD** = neovascular age-related macular degeneration; **SD-OCT** = spectral-domain OCT; **SRF** = subretinal fluid.

Keywords:
Artificial intelligence, Deep learning, Disease activity, Neovascular age-related macular degeneration, Optical coherence tomography.

Correspondence:
Daniel Lorand, MSc, Novartis Pharma AG, WSJ-188, Novartis Campus, CH-4056 Basel, Switzerland. E-mail: daniel.lorand@novartis.com.

References

1. Gale RP, Mahmood S, Devonport H, et al. Action on neovascular age-related macular degeneration (nAMD): recommendations for management and service provision in the UK hospital eye service. *Eye*. 2019;33(Suppl 1):1–21.
2. Hobbs SD, Pierce K. *Wet age-related macular degeneration (wet AMD) [Internet]*. StatPearls; 2022. Available at: http://www.ncbi.nlm.nih.gov/books/NBK572147/. Accessed February 27, 2024.
3. Hussain RM, Shaukat BA, Ciulla LM, et al. Vascular endothelial growth factor antagonists: promising players in the treatment of neovascular age-related macular degeneration. *Drug Des Devel Ther*. 2021;15:2653–2665.
4. Duffy AM, Bouchier-Hayes DJ, Harmey JH. Vascular endothelial growth factor (VEGF) and its role in non-endothelial cells: autocrine signalling by VEGF. Madame Curie Bioscience Database

[Internet]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK6482/; 2013. Accessed February 28, 2024.
5. Daien V, Finger RP, Talks JS, et al. Evolution of treatment paradigms in neovascular age-related macular degeneration: a review of real-world evidence. *Br J Ophthalmol*. 2021;105:1475−1479.
6. Lanzetta P, Loewenstein A. Fundamental principles of an anti-VEGF treatment regimen: optimal application of intravitreal anti−vascular endothelial growth factor therapy of macular diseases. *Graefes Arch Clin Exp Ophthalmol*. 2017;255:1259−1273.
7. Arevalo JF, Lasave AF, Wu L, et al. Intravitreal bevacizumab for choroidal neovascularization in age-related macular degeneration: 5-year results of the Pan-American Collaborative Retina Study Group. *Retina*. 2016;36:859−867.
8. Ciulla TA, Hussain RM, Pollack JS, Williams DF. Visual acuity Outcomes and anti-vascular endothelial growth factor therapy intensity in neovascular age-related macular degeneration patients: a real-world analysis of 49 485 eyes. *Ophthalmol Retina*. 2020;4:19−30.
9. Holz FG, Tadayoni R, Beatty S, et al. Multi-country real-life experience of anti-vascular endothelial growth factor therapy for wet age-related macular degeneration. *Br J Ophthalmol*. 2015;99:220−226.
10. Busbee BG, Ho AC, Brown DM, et al. Twelve-month efficacy and safety of 0.5 mg or 2.0 mg ranibizumab in patients with subfoveal neovascular age-related macular degeneration. *Ophthalmology*. 2013;120:1046−1056.
11. Kertes PJ, Galic IJ, Greve M, et al. Canadian treat-and-extend analysis trial with ranibizumab in patients with neovascular age-related macular disease: one-year results of the randomized canadian Treat-and-Extend analysis Trial with Ranibizumab Study. *Ophthalmology*. 2019;126:841−848.
12. Flaxel CJ, Adelman RA, Bailey ST, et al. Age-related macular degeneration preferred practice pattern®. *Ophthalmology*. 2020;127:P1−P65.
13. Wachtlin J, Spital G, Schmitz-Valckenberg S, et al. Use of imaging modalities in real life: impact on visual acuity outcomes of ranibizumab treatment for neovascular age-related macular degeneration in Germany. *J Ophthalmol*. 2020;2020:8024258.
14. Schmidt-Erfurth U, Reiter GS, Riedl S, et al. AI-based monitoring of retinal fluid in disease activity and under therapy. *Prog Retin Eye Res*. 2022;86:100972.
15. Dugel PU, Singh RP, Koh A, et al. HAWK and HARRIER: ninety-six-week outcomes from the phase 3 trials of brolucizumab for neovascular age-related macular degeneration. *Ophthalmology*. 2021;128:89−99.
16. Holekamp NM, Campochiaro PA, Chang MA, et al. Archway randomized phase 3 trial of the Port Delivery System with ranibizumab for neovascular age-related macular degeneration. *Ophthalmology*. 2022;129:295−307.
17. Kodjikian L, Parravano M, Clemens A, et al. Fluid as a critical biomarker in neovascular age-related macular degeneration management: literature review and consensus recommendations. *Eye*. 2021;35:2119−2135.
18. Patel PJ, Villavicencio P, Hanumunthadu D. Systematic review of neovascular age-related macular degeneration disease activity criteria use to shorten, maintain or extend treatment intervals with anti-VEGF in clinical trials: implications for clinical practice. *Ophthalmol Ther*. 2023;12:2323−2346.
19. Liakopoulos S, Spital G, Brinkmann CK, et al. ORCA study: real-world versus reading centre assessment of disease activity of neovascular age-related macular degeneration (nAMD). *Br J Ophthalmol*. 2020;104:1573−1578.
20. Toth CA, Decroos FC, Ying G-S, et al. Identification of fluid on optical coherence tomography by treating ophthalmologists versus a reading center in the Comparison of Age-Related Macular Degeneration Treatments Trials (CATT). *Retina*. 2015;35:1303−1314.
21. Ferrara D, Newton EM, Lee AY. Artificial intelligence-based predictions in neovascular age-related macular degeneration. *Curr Opin Ophthalmol*. 2021;32:389−396.
22. Feng D, Chen X, Zhou Z, et al. A preliminary study of predicting effectiveness of anti-VEGF injection using OCT images based on deep learning. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020:5428−5431.
23. Fu DJ, Faes L, Wagner SK, et al. Predicting incremental and future visual change in neovascular age-related macular degeneration using deep learning. *Ophthalmol Retina*. 2021;5:1074−1084.
24. Romo-Bucheli D, Erfurth US, Bogunovic H. End-to-end deep learning model for predicting treatment requirements in neovascular AMD from longitudinal retinal OCT imaging. *IEEE J Biomed Health Inform*. 2020;24:3456−3465.
25. Dugel PU, Koh A, Ogura Y, et al. HAWK and HARRIER: phase 3, multicenter, randomized, double-masked trials of brolucizumab for neovascular age-related macular degeneration. *Ophthalmology*. 2020;127:72−84.
26. Mantel I, Mosinska A, Bergin C, et al. Automated quantification of pathological fluids in neovascular age-related macular eegeneration, and its repeatability using deep learning. *Transl Vis Sci Technol*. 2021;10:17.
27. Apostolopoulos S, De Zanet S, Ciller C, et al. Pathological OCT retinal layer segmentation using branch residual U-shape networks. Available at: http://arxiv.org/abs/1707.04931; 2017. Accessed February 27, 2024.
28. R Shiny. Available at: https://shiny.rstudio.com/; 2024. Accessed February 27, 2024.
29. Available at:. *RetinAI Discovery®*; 2024. https://www.retinai.com/research. Accessed February 27, 2024.
30. Krippendorff K. Estimating the reliability, systematic error and random error of interval data. *Educ Psychol Meas*. 1970;30:61−70.
31. Bemme S, Heins A, Lauermann P, et al. Reliability of subjective assessment of SD-OCT pathologies by multiple raters in retinal vein occlusion. *Ophthalmol Sci*. 2021;1:100031.
32. Fu D, Tong H, Zheng S, et al. Retinal status analysis method based on feature extraction and quantitative grading in OCT images. *Biomed Eng Online*. 2016;15:87.
33. Bracha P, Ciulla TA. 'Real world' OCT: subtle findings, critical implications. Available at: https://ophthalmologymanagement.com/issues/2017/june/8216real-world8217-oct-subtle-findings-critical-implications/; 2017. Accessed February 27, 2024.
34. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology*. 2018;125:549−558.
35. Schmidt-Erfurth U, Vogl W-D, Jampol LM, Bogunović H. Application of automated quantification of fluid volumes to anti-VEGF therapy of neovascular age-related macular degeneration. *Ophthalmology*. 2020;127:1211−1219.