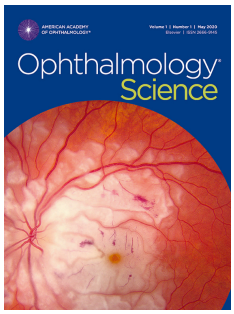
Journal Pre-proof



Quantitative Analysis of Retinal Fluid by a Deep Learning Model in Uveitic Macular Edema

Anthony Wu, BS, Adrian Au, MD, PhD, Justin Hanson, MD, Marcus Yamamoto, BS, Joy Cheng, Simon Lee, MS, Tal Eshkoly Lior, MD, Oren Avram, PhD, SriniVas R. Sadda, MD, Alison B. Coyne, MPH, Nisha R. Acharya, MD, MS, Brian Madow, MD, PhD, Jeffrey N. Chiang, PhD, Edmund Tsui, MD, MS

PII: S2666-9145(26)00179-X

DOI: https://doi.org/10.1016/j.xops.2026.101241

Reference: XOPS 101241

To appear in: *Ophthalmology Science*

Received Date: 20 August 2025

Revised Date: 31 March 2026

Accepted Date: 6 April 2026

Please cite this article as: Wu A., Au A., Hanson J., Yamamoto M., Cheng J., Lee S., Lior T.E., Avram O., Sadda S.R., Coyne A.B., Acharya N.R., Madow B., Chiang J.N. & Tsui E., Quantitative Analysis of Retinal Fluid by a Deep Learning Model in Uveitic Macular Edema, *Ophthalmology Science* (2026), doi: https://doi.org/10.1016/j.xops.2026.101241.

Quantitative Analysis of Retinal Fluid by a Deep Learning Model in Uveitic Macular Edema

Anthony Wu, BS[1,2,*]; Adrian Au, MD, PhD[1,*]; Justin Hanson, MD[1]; Marcus Yamamoto, BS[1,3]; Joy Cheng[2,4]; Simon Lee, MS[2]; Tal Eshkoly Lior, MD[1]; Oren Avram, PhD[2,4,5]; SriniVas R. Sadda, MD[1,6]; Alison B. Coyne, MPH[7]; Nisha R. Acharya, MD, MS[7]; Brian Madow, MD, PhD[8]; Jeffrey N. Chiang, PhD[2,9]; Edmund Tsui, MD, MS[1]

[*] Denotes equal contribution

1. Jules Stein Eye Institute, University of California, Los Angeles, Los Angeles, CA, USA
2. Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA, USA
3. John A. Burns School of Medicine, University of Hawai'i at Mānoa, Honolulu, HI, USA
4. Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
5. Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA, USA
6. Doheny Eye Institute, University of California, Los Angeles, Pasadena, CA, USA
7. Francis I. Proctor Foundation for Research in Ophthalmology, University of California, San Francisco, San Francisco, CA, USA
8. Ira G. Ross Eye Institute, Department of Ophthalmology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo/State University of New York (SUNY), Buffalo, NY, USA
9. Department of Neurosurgery, University of California, Los Angeles, Los Angeles, CA, USA

Corresponding Author: Edmund Tsui (etsui@mednet.ucla.edu)

Abbreviations and Acronyms

AI, artificial intelligence; **AMD**, age-related macular degeneration; **CMT**, central macular thickness; **DRIL**, disorganization of the retinal inner layers; **EZ**, ellipsoid zone; **FAST**, First-line Antimetabolites as Steroid-sparing Treatment; **IRF**, intraretinal fluid; **OCT**, optical coherence tomography; **PED**, pigment epithelial detachment; **RVO**, retinal vein occlusion; **SRF**, subretinal fluid; **UME**, uveitic macular edema

Financial Support

None

Conflict of Interest

S.R.S. has affiliations with Abbvie/Allergan, Alnylam, Alexion, Amgen, Apellis, ARVO, Astellas, Bayer, Biogen, Boerhinger Ingelheim, Carl Zeiss Meditec, Centervue, CharacterBio, Eyepoint, Heidelberg, ONL Therapeutics, iCare, IvericBio, Jannsen, Macula Society, Nanoscope, Nidek, NotalVision, Novartis, Optos, Ocular Tx, Regeneron, Roche, Samsung Bioepis and Topcon. N.R.A. has served as a consultant to Roche. The other authors declare no competing interests.

Keywords

Retinal optical coherence tomography, uveitis, fluid segmentation, macular edema, deep learning

Abstract

Objective

To assess whether AI-derived fluid volume provides prognostic value for visual outcomes in uveitic macular edema (UME) and to compare model performance to central macular thickness (CMT) measurements alone.

Design

Secondary subanalysis of patients with UME in the First-line Antimetabolites as Steroid-Sparing Treatment (FAST) Clinical Trial using a deep learning segmentation model trained on patients with age-related macular degeneration (AMD) and patients with retinal vein occlusion (RVO).

Participants

Patients with UME secondary to noninfectious uveitis from the FAST Uveitis Trial.

Methods

A 2D U-Net model, trained on patients with AMD and RVO using the RETOUCH dataset, was applied to segment intraretinal fluid (IRF) and subretinal fluid (SRF) in Heidelberg Spectralis OCT scans from the FAST Uveitis Trial. Model performance was validated against binary fluid gradings. Linear mixed effects models evaluated fluid resolution during treatment. Likelihood ratio tests and leave-one-subject-out cross-validation assessed whether baseline IRF and SRF volumes provided improved model fit for visual acuity change over 6 months of treatment compared to CMT alone.

Main Outcome Measures

Segmentation accuracy, correlation of fluid volume with visual acuity, and association of visual acuity change with baseline fluid volumes and CMT versus CMT only.

Results

The model achieved Dice scores of 0.61 for IRF and 0.74 for SRF on the RETOUCH dataset. In the FAST dataset, fluid segmentation correlated with binary gradings (biserial correlation: IRF 0.39, SRF 0.63; Mann–Whitney U-test: IRF and SRF $P < 0.01$). Longitudinal fluid modeling showed no significant treatment differences in IRF or SRF resolution. However, baseline IRF volume had a significant interaction with treatment assignment ($P = 0.02$) when modeling visual acuity change, suggesting that baseline IRF was more strongly associated with worse outcomes in the methotrexate group than the mycophenolate mofetil group. The likelihood ratio test confirmed that adding IRF and SRF improved fit over a model using CMT alone ($P < 0.01$) while leave-one-subject-out cross-validation showed a small, non-significant difference in prediction error.

Conclusions

AI-based segmentation of IRF and SRF enabled quantitative fluid measurements in UME and may provide additional prognostic signal for visual acuity change in UME compared to CMT alone, although larger cohorts will be needed to determine the magnitude of generalizable performance gains, if any. These findings support integrating AI-driven fluid analysis in clinical workflows and suggest that future clinical trials should consider stratifying by baseline fluid characteristics using AI fluid analysis.

Introduction

The integration of artificial intelligence (AI) in ophthalmology has revolutionized the analysis of optical coherence tomography (OCT) images, enabling rapid and automated detection of retinal biomarkers. This advancement is largely driven by the availability of large, annotated OCT datasets, such as the Kermany dataset[1] for classification tasks and the RETOUCH dataset[2] for fluid segmentation. These datasets, which are enriched in common retinal diseases such as age-related macular degeneration (AMD) and retinal vein occlusions (RVO), provide a foundation for developing deep learning algorithms capable of biomarker identification and treatment monitoring. While most models are developed and validated on common retinal pathologies, many structural biomarkers that these models identify are not disease-specific, such as ellipsoid zone (EZ) attenuation, intraretinal fluid (IRF), and subretinal fluid (SRF).

For the management of noninfectious uveitis, antimetabolites are frequently used as first-line therapies to achieve corticosteroid-sparing control of chronic inflammation. Among the most commonly prescribed agents are methotrexate and mycophenolate mofetil. The First-line Antimetabolites as Steroid-Sparing Treatment (FAST) Uveitis Trial—a multicenter, observer-masked clinical study—was designed to compare the corticosteroid-sparing effectiveness of these two medications in treating noninfectious uveitis. A key component of this trial involved collecting spectral-domain OCT images of the macula at all study visits to assess for macular edema associated with noninfectious intermediate, posterior, and panuveitis[3].

Despite advancements in AI-based analysis of edema across retinal diseases[4,5], their application in uveitic populations remains limited. In this study, we apply a fluid segmentation algorithm, trained on the RETOUCH dataset, to the FAST Uveitis Trial dataset. Central macular thickness is often

used as a prognostic factor for visual acuity, macular edema, and by extension, inflammatory disease activity[6], but its association with visual acuity vary across disease cohorts[7–9]. By assessing intraretinal fluid and subretinal fluid segmentation volumes from baseline to month 6 (prespecified primary outcome timepoint) in patients treated with methotrexate or mycophenolate mofetil, we aim to demonstrate the clinical utility and external validity of the AI fluid segmentation algorithm in the management of uveitic macular edema.

Methods

The RETOUCH Dataset

The RETOUCH dataset[2] was originally assembled from a challenge created for the 2017 Medical Image Computing and Computer Assisted Intervention (MICCAI) conference, in which OCT volumes were annotated for IRF, SRF, and pigment epithelial detachment (PED) at the pixel level. The slices and ground truth labels are stored in the ITK MetaImage format. The training dataset consists of 70 OCT volumes, with 24 volumes acquired using Zeiss Cirrus OCT (Carl Zeiss Meditec, Dublin, CA) devices, 24 acquired using Heidelberg Spectralis OCT (Heidelberg Engineering, Heidelberg, Germany) devices, and 22 acquired with Topcon OCT (Topcon Corporation, Tokyo, Japan) devices. Heidelberg Spectralis volumes were acquired with 49 B-scans, Zeiss Cirrus with 128 B-scans, and Topcon with 128 B-scans. The challenge submissions were evaluated on a separate evaluation dataset without the segmentation labels provided and thus, we did not use these volumes. 10 (four from Cirrus, three from Spectralis, and three from Topcon) of the 70 volumes were excluded from any training or validation process and were reserved for use in the evaluation of the model. The remaining 60 volumes are split into 5 folds, each fold being

used as a test set in one of the 5 splits. For each split in the model training, the 4 folds used for training are split into 75% for model fitting and 25% for validation. The validation set is used for early stopping[10] and binarizing the segmentation mask, choosing the cutoff for each fluid type that maximizes the Dice score[11]. The Dice score is defined as $Dice = \frac{2 \times intersection\ of\ positive\ pixels}{positive\ ground\ truth\ pixels + positive\ prediction\ pixels}$, performed for each fluid type.

Model Training

A 2D U-Net model[12] was used for this analysis due to its relative simplicity and high performance in data-constrained scenarios, as shown in the popularity of U-Net-based frameworks such as nnU-Net[13]. The model was developed using PyTorch[14] and MONAI[15], trained on a Nvidia A100 for 25 epochs. The model used a combination of the Dice loss[16] and cross entropy as the loss function since Dice is particularly effective in dealing with class imbalance and cross entropy can enforce a precise boundary[17]. The AdamW optimizer[18] was used with fast.ai[19] default parameters for training, with a peak learning rate of 5e-3 found using a range test[20]. The fit-one-cycle policy[21] was applied to the learning rate to ensure proper convergence to a minimum. The weights of the epoch with the lowest validation loss were recorded for each model. The preprocessing step required each B-scan to be padded so that the image was squared and then resized to 256 by 256 pixels. During training, random augmentations were applied to artificially increase the variety of training images and to prevent overfitting. Each of the following augmentations was performed with a 50% chance: horizontal flipping, random affine transformation (e.g., rotation, translate, scale, and shear), and elastic deformation.

Model Evaluation on RETOUCH

The evaluation of the model on the 10 withheld volumes is based on all five models, each trained from a different split. Each pixel was classified as a particular type of fluid according to a majority vote of the five different models. The five models allow us to perform a pixel-wise Fleiss' kappa score for each fluid type, acting as a voting system for the classification of fluid and a measure of internal consistency. The volume of each fluid type generated from the model was also compared to the ground truth volume of each fluid type and the induced R^2 was computed. *(Fig. 1)*

The FAST Dataset

The FAST Uveitis Trial is a National Eye Institute–funded, block-randomized, observer-masked clinical trial which has been previously described by Rathinam et al.[3]. The study was approved by the institutional review boards at all sites, adhered to the Declaration of Helsinki, and obtained written informed consent from all participants. The trial enrolled patients aged 16 years or older with active noninfectious intermediate, posterior, or panuveitis in at least one eye requiring corticosteroid-sparing therapy at baseline. All patients received an initial corticosteroid course (1 mg/kg or 60 mg daily, whichever was lower), tapered per Standardization of Uveitis Nomenclature (SUN) guidelines[22]. Patients were block-randomized to receive oral methotrexate (15 mg for two weeks, then 25 mg weekly) or mycophenolate mofetil (500 mg twice daily for two weeks, then 1.5 g twice daily). Macular edema was not included in the definition of treatment success, which was determined by achieving corticosteroid-sparing control of uveitis, defined as a prednisone dose of $\leq$ 7.5 mg with resolution of anterior chamber, vitreous haze, and retinal or choroidal inflammation. At six months, patients who achieved treatment success continued their assigned therapy, while those with treatment failure switched to the alternative medication for another six months.

Periocular or intraocular steroid injections were permitted only for macular edema within the first 90 days; afterward, they were considered protocol deviations. OCT imaging of the macula was conducted at every visit, with images analyzed at the University of South Florida reading center to assess macular edema presence, morphology, and CMT. Of all the patients evaluated in the clinical trial, patients with macular edema, defined as CMT $\geq$ 315 μm on Spectralis and cystic changes on OCT, were included for analysis. Patients were excluded from CMT analysis if they had a serous detachment associated with a diagnosis of Vogt-Koyanagi-Harada Disease[6].

FAST Dataset External Validation

To ensure accuracy, 20 segmentation masks were qualitatively assessed for IRF, SRF, and PED. PED was excluded in the FAST dataset external validation, as posterior uveitis patients rarely develop macular neovascularization and PED segmentation accuracy was low. *(Fig. 2)* As there were no segmentation labels for the FAST dataset, we were unable to directly evaluate the results of the segmentation model on the dataset at a pixel level. However, the dataset does contain binarized labels for the presence of SRF and IRF for each scan, which we were able to compare against our segmentation volumes using point biserial correlation to measure the relationship between the dichotomous labeling and quantitative variable, and single-tailed Mann–Whitney U-tests[23] to compare volumes in scans graded as fluid present to those graded as fluid absent, as the distribution of fluid are right-tailed.

Fluid Modeling in FAST

To assess the relationship between fluid volume, visual acuity, central macular thickness, and treatment assignment, macular fluid was quantified by summing pixel counts from five central B-

scans (scans 22–26) extracted from 49 B-scan OCT volumes obtained using Heidelberg Spectralis devices, as the majority of the imaging available was in this configuration. In all models, treatment assignment was effect coded as +1 for mycophenolate mofetil and -1 for methotrexate.

Retinal Fluid Changes in Treatment Groups

To assess IRF and SRF changes in the five central scans over time between patients treated with methotrexate and those treated with mycophenolate mofetil, we first fit linear mixed effects models for each fluid type. We modeled the change in fluid compared to baseline up to month 6 (prespecified primary endpoint) to avoid confounding from switching medication as part of the protocol. A total of 28 eyes in 19 patients and 41 eyes in 33 patients in the methotrexate and mycophenolate mofetil groups, respectively, were included. The analysis inherently accommodated unbalanced data under the assumption that data were missing at random, as some visits did not use Heidelberg Spectralis and thus were excluded from our analysis. The modeling included fixed effects for visit month, baseline fluid levels, and two-way interactions with treatment assignment. Patient and eye were included as nested random effects to account for repeated measures. Treating visit month as a factor was deemed unnecessary based on likelihood ratio tests. Restricted cubic splines were fitted for baseline fluid levels and visit month, but deemed unnecessary based on likelihood ratio tests.

The model was estimated using maximum likelihood, and the estimated marginal means at each month were computed using the emmeans R package[24], adjusting for baseline fluid volume. To compare treatment groups at each month, pairwise comparisons of the estimated marginal means were performed and two-sided P-values were calculated using Satterthwaite approximations for degrees of freedom[25]. These P-values assess whether treatment assignment had a statistically

significant effect on fluid levels at each month during treatment after adjusting for baseline differences.

Effect of Baseline Fluid, CMT, and Treatment on Visual Acuity Change

To evaluate whether baseline retinal fluid metrics or CMT influenced visual acuity outcomes and whether their effects differed by treatment assignment, we modeled the change in logMAR visual acuity from baseline to follow-up, up to month 6, to avoid confounding from switching treatment. We used baseline IRF, baseline SRF, baseline CMT, baseline visual acuity, and two-way interactions with treatment assignment as fixed effects, with eye and patient as nested random effects to account for repeated monthly follow-ups. Only eyes with OCT scans and visual acuity recordings at baseline were included, as long as they had at least one visit during the first 6 months of treatment. This resulted in 28 eyes from 19 patients in the methotrexate group and 43 eyes from 34 patients in the mycophenolate mofetil group. In fitting the model, we assumed that data was missing at random and did not attempt to impute missing data points. We again evaluated treating visit month as a factor and restricted cubic splines for each continuous variable, but chose to use the simplest model after evaluating every combination using likelihood ratio tests.

We anticipated collinearity among the OCT measurements given that CMT serves as a surrogate marker for macular fluid. Thus, variable selection is necessary to avoid overfitting and to ensure the interpretability of effect estimates. Stepwise backward variable selection is performed by removing the coefficient with the largest P-value while ensuring that if an interaction term is included, then all of its lower-order terms are included, until no further reduction in AIC is possible. Importantly, visit month and random intercepts for eye nested within patient were retained in all models, regardless of statistical significance, due to their fundamental role in

231 modeling repeated-measures longitudinal data. The quantile-quantile plot of the residuals in the
232 final model is visually inspected to ensure no heteroskedasticity. We then repeated the variable
233 selection process starting with the same variables except for baseline IRF and SRF to determine
234 the best model possible without using fluid metrics.

235 We utilized likelihood ratio tests as in-sample evidence of improved fit for pre-specified nested
236 models. We first compared a full model with baseline visual acuity, treatment assignment, visit
237 month, baseline CMT, IRF, SRF, and a random intercept for eye nested within patient, to a reduced
238 model that removed baseline IRF and SRF volume as variables to assess the incremental value of
239 baseline IRF and SRF volume measurements. To assess the incremental value of CMT, we
240 similarly compared the full model to a model in which baseline CMT was omitted using a
241 likelihood ratio test. Variable selection was not performed for these comparisons because
242 likelihood ratio tests require pre-specified, nested models to quantify incremental improvement in
243 in-sample fit. To evaluate the generalization of the improvement, incremental model performance
244 was assessed with leave-one-subject-out cross-validation using mean absolute error.

245 Results

246 Model Results on RETOUCH

247 Each test set is evaluated on the model created from the split where it acts as a test set, which
248 allows the entire 60 volume subset to be evaluated and bootstrapped 1000 times at the volume
249 level to get 95% confidence intervals (CI) for the mean Dice score. The Dice score of the 60
250 volume subset cross-validation is 0.61 (95% CI: 0.57-0.65) for IRF, 0.74 (95% CI: 0.67-0.79) for
251 SRF, and 0.61 (95% CI: 0.48-0.70) for PED. Our results are similar to those reported in Bogunović

et al.[2], with all three fluid types within a standard deviation of the challenge submissions. It should be noted that our training used less data than challenge submissions, since we withheld a subset of the dataset, whereas submissions were able to train using the entire 70 volume dataset and then evaluated on a subset of the dataset whose segmentation labels are not available outside the challenge. Fluid annotation is inherently a difficult task, with previous publications reporting a human inter-rater mean Dice score of 0.75[4].

Evaluating on the 10 withheld volumes gave us additional confidence in the reliability of the model in segmenting fluid in the context of retinal diseases. For all three fluid compartments, Fleiss' kappa was over 0.6, signifying substantial internal consistency for IRF and SRF, according to the divisions suggested by Landis et al.[26]. The mean Dice score of each fluid ranged from 0.54 to 0.72, which suggests reasonably accurate results. The regression performed on IRF and SRF volume showed good correlation between volume readings in the ground truth and the model output, with an R^2 of 0.86, 0.95, and 0.53 for IRF, SRF, and PED, respectively. *(Table 1)*

Applying Model on FAST

Comparing the fluid segmentation volume in the central five B-scans against the binary grading of fluid resulted in a biserial correlation of 0.39 for IRF and 0.63 for SRF. Single-tailed Mann–Whitney U-tests comparing the volume of fluid in the five central B-scans graded as fluid present to those graded as fluid absent resulted in $p < 0.01$ for both IRF and SRF. Regressing IRF and SRF volumes within the 5 central B-scans on CMT gathered from the same scan resulted in an R^2 of 0.413, with $p < 0.01$ for both IRF and SRF as independent variables, using 515 total scans from the FAST macular edema cohort.

Treatment Effects on Retinal Fluid

For IRF modeling, all main effects and interactions were statistically insignificant based on Satterthwaite's method[25], except for the baseline IRF main effect. Pairwise contrasts of the estimated marginal means at each month had $P > 0.38$ at all months and a P-value of 0.46 at month 6. SRF modeling had similar results, with non-significant main effects and interactions except for the baseline SRF main effect. The pairwise contrasts had $P > 0.81$ at all months and a P-value of 0.82 at month 6. *(Table 2)* Backwards stepwise variable selection eliminated all variables in both models except baseline fluid levels and removed patient level random effects. After adjusting for baseline fluid levels, methotrexate and mycophenolate mofetil had comparable effects on IRF and SRF reduction within the five central B-scans up to month 6. Longitudinal differences in fluid volume while on either treatment is inconclusive.

Significant Interaction Between Baseline IRF and Treatment Assignment

In modeling visual acuity change during treatment, we started with a full model including all baseline measurements and two-way interactions with treatment assignment. As expected, baseline CMT had a correlation of 0.67 with baseline IRF and 0.49 with baseline SRF, demonstrating moderate to strong collinearity. All other pairwise correlations were less than 0.49, indicating weaker associations. The full model prior to variable selection had an AIC of -200 and BIC of -141. Stepwise backward variable selection removed patient level random effects and removed all interaction terms with treatment assignment except for the interaction term with baseline IRF. *(Table 3)* The interaction term shows that for every mm^3 of IRF in the 5 central B-scans, the associated change in logMAR visual acuity is 2.57 (worse vision) in methotrexate and -0.02 (slightly better to no change in vision) for mycophenolate mofetil, with a P-value of 0.02. *(Fig. 3)*

Time was not statistically significant in this model (P = 0.31), though retained due to its structural role in the repeated-measures design. The other main effects of baseline SRF, CMT, and logMAR visual acuity coefficients had P-values of 0.04, 0.01, and < 0.01, respectively. Both the AIC and BIC decreased compared to the full model, with AIC of -202 and BIC of -163. The fixed effects explained 53% of the variance in visual acuity change (marginal R^2), while accounting for the random effect of each eye, the model explains 85% of the variance (conditional R^2). After performing the same backward variable selection process without fluid metrics, the variables retained were baseline CMT, baseline visual acuity, visit month, and random intercepts for eyes nested within patient. The model without fluid metrics prior to variable selection had an AIC of -187 and BIC of -144. After variable selection, the AIC was -194 and BIC was -170, with a marginal R^2 of 42% and conditional R^2 of 85%.

To assess the incremental value of baseline IRF and SRF volume measurements, we first fit a model using baseline CMT, baseline IRF, baseline SRF, and baseline logMAR VA, treatment assignment, time, and random intercepts for eye, nested within patient. We then fit another model that removed baseline IRF and SRF volume. The full model had had an AIC of -199 and BIC of -163, while removing IRF and SRF resulted in an AIC of -192 and BIC of -164. A likelihood ratio test showed that the more complex model including baseline IRF and SRF provided significantly better in-sample fit (P < 0.01). We then fit another model that removed baseline CMT while retaining baseline IRF and SRF, resulting in an AIC of -191 and BIC of -160. Compared to the model without CMT, the full model provided significantly better in-sample fit (P < 0.01). In leave-one-subject-out cross-validation, the paired difference in absolute error between the full model and the model without fluid was not statistically significant, with a mean absolute error of around 2 ETDRS lines (0.22 logMAR) in both models.

Discussion

This study evaluated an AI-driven fluid segmentation model for analyzing OCT images in uveitic macular edema and utilized the segmentation output to compare fluid metrics to CMT in prognosticating visual acuity during treatment. While AI algorithms have been widely used in other retinal diseases[27,28], their application to fluid evaluation in uveitic macular edema remains unexplored. The AI segmentation model, initially trained on a dataset enriched with AMD and RVO cases, demonstrated external validation in a distinct uveitis cohort, supporting its generalizability across different exudative retinal pathologies. Importantly, the pathophysiology of UME is due to inflammatory mechanisms that differ substantially from exudative and ischemic process in AMD and RVO. Our results indicate that despite these mechanistic differences, the model's fluid volume estimates still provide clinical utility in UME, suggesting that cross-disease application of segmentation models may be feasible. Given the limited availability of well-labeled UME datasets, demonstrating that a model trained entirely on non-uveitic diseases can be applied to UME helps address a key barrier to developing AI tools for rare inflammatory conditions. Additionally, our results show that directly quantifying IRF and SRF using the AI segmentation model may offer incremental prognostic signal for vision over CMT alone, particularly when evaluating how treatment modifies these relationships. These findings suggest that AI-based fluid analysis may provide a more reliable metric for assessing disease progression and treatment response in uveitic macular edema.

While CMT has traditionally been regarded as a reliable indicator of fluid regression, it provides a single global measure that cannot distinguish between fluid compartments. This is particularly important because IRF and SRF may have distinct prognostic and pathophysiological implications.

For instance, IRF in AMD has been linked with worse visual acuity outcomes whereas persistent SRF was often associated with better or neutral outcomes[29–31]. However, AMD-specific fluid findings cannot be assumed to hold in UME, as the two arise from fundamentally different processes of fluid accumulation. No analogous compartment-specific fluid analysis has been performed in UME, where most prior work rely exclusively on CMT. A prior study analyzed the quantitative effect of SRF on visual acuity in patients with uveitic cystoid macular edema but did not quantify IRF volume[32]. Against this background, our study demonstrated that baseline IRF and SRF volumes provided significant incremental in-sample prognostic signal for visual acuity change beyond CMT alone in prespecified mixed-effects models, as supported by likelihood ratio testing, while leave-one-subject-out cross-validation showed a small, non-significant improvement in prediction error. Together, these results suggest that while CMT remains a useful parameter, AI-driven fluid segmentation may provide a better assessment at baseline and nuanced assessment of disease progression and visual prognosis in this uveitic macular edema cohort, although larger cohorts will be needed to determine the magnitude of generalizable performance gains.

While Tsui et al.[6] found no difference in CMT between treatment groups, our segmentation of IRF and SRF revealed that baseline IRF may have a differential impact on visual outcomes depending on treatment. Specifically, higher baseline IRF was associated with worse visual acuity in methotrexate-treated eyes, an association not observed in those receiving mycophenolate mofetil. This finding parallels prior observations in AMD, where IRF is a strong negative prognostic marker for vision[33]. This treatment-specific prognostic heterogeneity only emerged when fluid compartment measurements were incorporated into the modeling framework, as the model restricted to CMT alone showed no treatment effect or interaction after model selection, indicating that thickness-based measurements alone were insufficient in capturing this differential prognostic

signal. While both treatments appear similarly effective in resolving fluid anatomically, shown in our analysis of IRF and SRF changes during treatment and previous analysis using CMT in Tsui et al.[6], these results suggest that baseline IRF may influence functional outcomes in a treatment-specific manner. However, as the trial was not designed to evaluate macular edema outcomes and treatment groups were not stratified by baseline fluid characteristics, these exploratory findings may be subject to confounding. Further prospective studies are warranted to investigate whether baseline fluid burden modifies treatment response in uveitic macular edema.

Deep learning frameworks have been developed to identify OCT biomarkers of uveitis with high accuracy, including hard exudates, hyper-reflective foci, intraretinal cysts, and neurosensory detachment[34]. Additional OCT biomarkers have demonstrated prognostic value in uveitic macular edema but have yet to be adopted in this domain. Disorganization of the retinal inner layers (DRIL), intraretinal cyst area, EZ disruption, hyperreflective foci, and CMT have all been strongly associated with visual acuity, whereas external limiting membrane disruption has shown less predictive utility[35]. Pre-treatment EZ integrity and the presence of cystoid spaces or subretinal fluid have been linked to improved therapeutic responses, while the absence of EZ and the presence of DRIL correlate with poorer visual outcomes[36,37]. Machine-learning algorithms have already been implemented to identify biomarkers such as DRIL in diabetic retinopathy[38] and EZ attenuation in retinitis pigmentosa[39], demonstrating their utility in detecting structural changes associated with visual outcomes[40]. However, the limited availability of large, annotated uveitis datasets, along with the heterogeneity and rarity of uveitis, has impeded the widespread adoption of AI in this domain. Addressing these limitations through the development of uveitis-specific datasets and refining AI models for this population could significantly enhance disease monitoring and prognostication for uveitic macular edema. These findings suggest that integrating AI-driven

386 fluid segmentation with other structural biomarkers may provide a more comprehensive
387 assessment of visual prognosis in treatment response.

388 These findings also carry practical implications for ongoing and future UME clinical trials[41–44].
389 These trials currently rely on standard endpoints such as change in best-corrected VA and central
390 subfield thickness, which provide limited insight into fluid compartment-specific burden. Our
391 results suggest that incorporating AI-driven IRF and SRF quantification at baseline for
392 stratification or longitudinally for mechanistic response may improve trial power, reduce outcome
393 variability, and reveal treatment-specific biomarker interactions.

394 A major strength of this study is the use of a clinical trial cohort, ensuring standardized imaging
395 protocols and rigorous data collection. The inclusion of a diverse range of uveitic conditions
396 enhances the generalizability of findings but also introduces variability that may impact the
397 consistency of these associations. Additionally, as the AI model was not specifically trained on a
398 uveitis cohort, its accuracy in this population may be limited; incorporating a manually segmented
399 uveitis-specific dataset could allow for model finetuning, improved performance, and provide a
400 formal benchmark for evaluating how well models trained on non-uveitic diseases perform when
401 applied to UME. Another limitation is the lack of direct comparison between binary gradings by
402 the reading center and segmentation-level data, which may have constrained the validation of AI-
403 based fluid segmentation against standard clinical assessments. Furthermore, our modeling was
404 done on a relatively small cohort and assumed that data was missing at random, which may limit
405 the generalizability of this study. In leave-one-subject-out validation, differences in prediction
406 error between models were small and not statistically significant, suggesting that larger cohorts
407 may be needed to precisely estimate generalizable performance gains. Alternatively, additional
408 fluid metrics may not significantly improve prognostic performance over CMT alone. Lastly,

media opacities (e.g., cataract or vitreous inflammation) may confound the quality of OCT images and subsequently proper segmentation. Addressing these limitations in future studies could further refine the role of AI-driven fluid analysis in uveitic macular edema.

Overall, this study provides evidence supporting AI-driven fluid segmentation as an important addition to traditional CMT-based measurements in uveitic macular edema. By leveraging OCT imaging, AI models can automate fluid quantification with greater precision and speed, offering a more refined assessment of visual response to treatment and identify subtle differences between treatments. Integrating AI-based segmentation into clinical practice and clinical trials has the potential to enhance clinical decision-making, improve patient outcomes, and advance the management of uveitic macular edema.

References

1. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*. 2018;172(5):1122-1131.e9. doi:10.1016/j.cell.2018.02.010

2. Bogunovic H, Venhuizen F, Klimscha S, et al. RETOUCH: The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge. *IEEE Trans Med Imaging*. 2019;38(8):1858-1874. doi:10.1109/TMI.2019.2901398

3. Rathinam SR, Gonzales JA, Thundikandy R, et al. Effect of Corticosteroid-Sparing Treatment With Mycophenolate Mofetil vs Methotrexate on Inflammation in Patients With Uveitis: A Randomized Clinical Trial. *JAMA*. 2019;322(10):936. doi:10.1001/jama.2019.12618

4. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017;8(7):3440. doi:10.1364/BOE.8.003440

5. Wang Z, Zhong Y, Yao M, et al. Automated segmentation of macular edema for the diagnosis of ocular disease using deep learning method. *Sci Rep*. 2021;11(1):13392. doi:10.1038/s41598-021-92458-8

6. Tsui E, Rathinam SR, Gonzales JA, et al. Outcomes of Uveitic Macular Edema in the First-line Antimetabolites as Steroid-Sparing Treatment Uveitis Trial. *Ophthalmology*. 2022;129(6):661-667. doi:10.1016/j.ophtha.2022.02.002

7. Ou WC, Brown DM, Payne JF, Wykoff CC. Relationship Between Visual Acuity and Retinal Thickness During Anti–Vascular Endothelial Growth Factor Therapy for Retinal Diseases. *Am J Ophthalmol*. 2017;180:8-17. doi:10.1016/j.ajo.2017.05.014

8. Hannouche RZ, Ávila MPD, Isaac DLC, Silva RSCE, Rassi AR. Correlation between central subfield thickness, visual acuity and structural changes in diabetic macular edema. *Arq Bras Oftalmol*. 2012;75(3):183-187. doi:10.1590/S0004-27492012000300007

9. Torres-Villaros H, Timoumi R, Fajnkuchen F, Klokner A, Giocanti-Aurégan A. Macular thickness fluctuations and visual acuity outcomes after intravitreal dexamethasone implant for diabetic macular edema. *Retina*. 2024;44(11):1937-1944. doi:10.1097/IAE.0000000000004213

10. Prechelt L. Early Stopping — But When? In: Montavon G, Orr GB, Müller KR, eds. *Neural Networks: Tricks of the Trade*. Vol 7700. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2012:53-67. doi:10.1007/978-3-642-35289-8_5

11. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302. doi:10.2307/1932409

12. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv*. Preprint posted online 2015. doi:10.48550/ARXIV.1505.04597

13. Isensee F, Petersen J, Klein A, et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *arXiv*. Preprint posted online 2018. doi:10.48550/ARXIV.1809.10486

14. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv*. Preprint posted online 2019. doi:10.48550/ARXIV.1912.01703

15. Cardoso MJ, Li W, Brown R, et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv*. Preprint posted online 2022. doi:10.48550/ARXIV.2211.02701

16. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE; 2016:565-571. doi:10.1109/3DV.2016.79

17. Taghanaki SA, Zheng Y, Zhou SK, et al. Combo Loss: Handling Input and Output Imbalance in Multi-Organ Segmentation. *arXiv*. Preprint posted online September 15, 2021. doi:10.48550/arXiv.1805.02798

18. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. *arXiv*. Preprint posted online 2017. doi:10.48550/ARXIV.1711.05101

19. Howard J, Gugger S. fastai: A Layered API for Deep Learning. Published online 2020. doi:10.48550/ARXIV.2002.04688

20. Smith LN. Cyclical Learning Rates for Training Neural Networks. *arXiv*. Preprint posted online April 4, 2017. doi:10.48550/arXiv.1506.01186

21. Smith LN, Topin N. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates. *arXiv*. Preprint posted online 2017. doi:10.48550/ARXIV.1708.07120

22. Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. *Am J Ophthalmol*. 2005;140(3):509-516. doi:10.1016/j.ajo.2005.03.057

23. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann Math Stat*. 1947;18(1):50-60. doi:10.1214/aoms/1177730491

24. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means. Published online 2025. https://rvlenth.github.io/emmeans/

25. Satterthwaite FE. An Approximate Distribution of Estimates of Variance Components. *Biom Bull*. 1946;2(6):110. doi:10.2307/3002019

26. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159. doi:10.2307/2529310

27. Avram O, Durmus B, Rakocz N, et al. Accurate prediction of disease-risk factors from volumetric medical scans by a deep vision model pre-trained with 2D scans. *Nat Biomed Eng*. 2024;9(4):507-520. doi:10.1038/s41551-024-01257-9

28. Rudas A, Chiang JN, Corradetti G, et al. Automated large-scale prediction of exudative AMD progression using machine-read OCT biomarkers. Liu N, ed. *PLOS Digit Health*. 2023;2(2):e0000106. doi:10.1371/journal.pdig.0000106

29. Guymer RH, Markey CM, McAllister IL, et al. Tolerating Subretinal Fluid in Neovascular Age-Related Macular Degeneration Treated with Ranibizumab Using a Treat-and-Extend Regimen. *Ophthalmology*. 2019;126(5):723-734. doi:10.1016/j.ophtha.2018.11.025

30. Chaudhary V, Matonti F, Zarranz-Ventura J, Stewart MW. IMPACT OF FLUID COMPARTMENTS ON FUNCTIONAL OUTCOMES FOR PATIENTS WITH NEOVASCULAR AGE-RELATED MACULAR DEGENERATION: A Systematic Literature Review. *Retina*. 2022;42(4):589-606. doi:10.1097/IAE.0000000000003283

31. Patil NS, Mihalache A, Dhoot AS, Popovic MM, Muni RH, Kertes PJ. Association Between Visual Acuity and Residual Retinal Fluid Following Intravitreal Anti–Vascular Endothelial Growth Factor Treatment for Neovascular Age-Related Macular Degeneration: A Systematic Review and Meta-analysis. *JAMA Ophthalmol*. 2022;140(6):611. doi:10.1001/jamaophthalmol.2022.1357

32. Weldy EW, Patnaik JL, Pecen PE, Palestine AG. Quantitative effect of subretinal fluid and intraretinal edema on visual acuity in uveitic cystoid macular edema. *J Ophthalmic Inflamm Infect*. 2021;11(1):38. doi:10.1186/s12348-021-00266-y

33. Gadiollet E, Kodjikian L, Vasson F, et al. Effect of baseline fluid localization on visual acuity and prognosis in type 1 macular neovascularization treated with anti-VEGF. *Eye*. 2024;38(16):3161-3168. doi:10.1038/s41433-024-03256-1

34. S V A, G DB, Raman R. Automatic Identification and Severity Classification of Retinal Biomarkers in SD-OCT Using Dilated Depthwise Separable Convolution ResNet with SVM Classifier. *Curr Eye Res*. 2024;49(5):513-523. doi:10.1080/02713683.2024.2303713

35. Grewal DS, O'Sullivan ML, Kron M, Jaffe GJ. Association of Disorganization of Retinal Inner Layers With Visual Acuity In Eyes With Uveitic Cystoid Macular Edema. *Am J Ophthalmol*. 2017;177:116-125. doi:10.1016/j.ajo.2017.02.017

36. Ciulla TA, Kapik B, Barakat MR, et al. Optical Coherence Tomography Anatomic and Temporal Biomarkers in Uveitic Macular Edema. *Am J Ophthalmol*. 2022;237:310-324. doi:10.1016/j.ajo.2021.10.024

37. Cicinelli MV, Gerosolima C, Scandale P, et al. Clinical and imaging biomarkers of response to intravitreal dexamethasone implant in eyes with non-infectious uveitic macular oedema. *Eye*. 2024;38(5):910-916. doi:10.1038/s41433-023-02802-7

38. Singh R, Singuri S, Batoki J, et al. Deep Learning Algorithm Detects Presence of Disorganization of Retinal Inner Layers (DRIL)–An Early Imaging Biomarker in Diabetic Retinopathy. *Transl Vis Sci Technol*. 2023;12(7):6. doi:10.1167/tvst.12.7.6

39. Wang YZ, Juroch K, Birch DG. Deep Learning-Assisted Measurements of Photoreceptor Ellipsoid Zone Area and Outer Segment Volume as Biomarkers for Retinitis Pigmentosa. *Bioengineering*. 2023;10(12):1394. doi:10.3390/bioengineering10121394

40. Toto L, Romano A, Pavan M, et al. A deep learning approach to hard exudates detection and disorganization of retinal inner layers identification on OCT images. *Sci Rep*. 2024;14(1):16652. doi:10.1038/s41598-024-63844-9

41. Hoffmann-La Roche. *A Study to Investigate Vamikibart in Participants With Uveitic Macular Edema (Meerkat)*. https://clinicaltrials.gov/study/NCT05642312

42. Hoffmann-La Roche. *Vamikibart in Participants With Uveitic Macular Edema (Sandcat)*. https://clinicaltrials.gov/study/NCT05642325

43. Kodiak Sciences Inc. *A Phase 3 Study to Evaluate the Efficacy and Safety of Intravitreal KSI-101 in Participants With Macular Edema Secondary to Inflammation (MESI) - PEAK (PEAK)*. https://clinicaltrials.gov/study/NCT06990399

44. Kodiak Sciences Inc. *A Phase 3 Study to Evaluate the Efficacy and Safety of Intravitreal KSI-101 in Participants With Macular Edema Secondary to Inflammation (MESI) - PINNACLE (PINNACLE)*. https://www.clinicaltrials.gov/study/NCT06996080

Table and Figures

Fig. 1: Model segmentation output of a B-scan from the withheld portion of the RETOUCH Dataset. Top images are the B-scans obtained from Spectralis OCT. Bottom images are segmentation outputs for the corresponding images overlaid on the B-scan. Red represents areas identified as intraretinal fluid (IRF), green represents areas identified as subretinal fluid (SRF), blue represents areas identified as pigment epithelial detachment (PED).

Fig. 2: Model generated segmentation mask of patients with uveitis and macular edema from the FAST Uveitis Trial. Top images are the B-scans obtained from Spectralis OCT. Bottom images are segmentation outputs for the corresponding images overlaid on the B-scan. Red represents pixels identified as intraretinal fluid (IRF) and green represents pixels identified as subretinal fluid (SRF).

Table 1: Evaluation of U-Net Training Pipeline on 10 Withheld Volumes from RETOUCH

Table 2: Fluid change in subset of 19 methotrexate-treated and 30 mycophenolate mofetil-treated eyes with both baseline and month 6 measurements, showing insignificant differences between treatments at month 6. Some eyes included in the mixed effects modeling had partial follow-up data but were not included in this table due to missing month 6 measurements.

Table 3: Coefficient estimates of model predicting logMAR visual acuity change from baseline. Stepwise backward variable selection was used to select these variables and interactions. Two-way interactions with treatment assignment are considered but only the interaction term with baseline VA is not removed by variable selection. Patient level random effects are also removed. Baseline fluid levels are calculated from the sum of fluid areas in each of the 5 central B-scans, multiplied by the interscan distance. Visit month is coded as months since first follow-up in month 1. Treatment assignment was effect coded with +1 for mycophenolate mofetil and -1 for methotrexate.

Fig. 3: Marginal effect of baseline IRF within the 5 central B-scans on change in logMAR visual acuity during treatment by treatment assignment. The slopes suggest that higher baseline IRF is associated with worse visual acuity outcomes in eyes treated with methotrexate but this association is not observed in eyes treated with mycophenolate mofetil.

Fluid Type	Fleiss' Kappa	Mean Dice Score	R^2 of Volume
IRF	0.75	0.67	0.86
SRF	0.85	0.72	0.94
PED	0.67	0.54	0.53

	Methotrexate (19 eyes)	**Mycophenolate Mofetil (30 eyes)**	**P-value**
Median Baseline IRF (IQR), mm^3	0.019 (0.004 – 0.037)	0.011 (0.003 – 0.032)	
Median Month 6 IRF (IQR), mm^3	0.007 (0.002 – 0.018)	0.004 (0.001 – 0.009)	
Median Change in IRF (IQR), mm^3	-0.001 (-0.023 – 0.004)	-0.001 (-0.013 – 0.001)	**0.46**
Median Baseline SRF (IQR), mm^3	0.006 (0.002 – 0.033)	0.014 (0.001 – 0.053)	
Median Month 6 SRF (IQR), mm^3	0.004 (0.001 – 0.010)	0.006 (0.002 – 0.017)	
Median Change in SRF (IQR), mm^3	-0.001 (-0.023 – 0.003)	-0.011 (-0.036 – 0.001)	**0.82**

Table 2: Fluid change in subset of 19 methotrexate-treated and 30 mycophenolate mofetil-treated eyes with both baseline and month 6 measurements, showing insignificant differences between treatments at month 6. Some eyes included in the mixed effects modeling had partial follow-up data but were not included in this table due to missing month 6 measurements.

	Coefficient Estimate	Standard Error	P-value
Intercept	0.27	0.10	0.01
Treatment Assignment	0.028	0.031	0.37
Baseline IRF (per mm^3)	1.28	0.77	0.10
Baseline SRF (per mm^3)	-1.00	0.48	0.04
Baseline CMT (per mm)	-0.89	0.32	< 0.01
Baseline logMAR VA	-0.30	0.067	< 0.01
Visit Month	-0.005	0.004	0.31
Baseline IRF : Treatment Assignment	-1.30	0.55	0.02

Table 3: Coefficient estimates of model predicting logMAR visual acuity change from baseline. Stepwise backward variable selection was used to select these variables and interactions. Two-way interactions with treatment assignment are considered but only the interaction term with baseline VA is not removed by variable selection. Patient level random effects are also removed. Baseline fluid levels are calculated from the sum of fluid areas in each of the 5 central b-scans, multiplied by the interscan distance. Visit month is coded as months since first follow-up in month 1. Treatment assignment was effect coded with +1 for mycophenolate mofetil and -1 for methotrexate.