



Published in final edited form as:

Ophthalmology. 2024 February ; 131(2): 219–226. doi:10.1016/j.ophtha.2023.09.016.

Developing a continuous severity scale for MacTel type 2 using Deep Learning and implications for disease grading

Yue Wu^{1,2}, Catherine Egan^{3,4}, Abraham Olvera-Barrios^{3,4}, Lea Scheppeke^{5,6}, Tunde Peto⁷, Peter Charbel Issa^{8,9}, Tjebo Heeren³, Irene Leung³, Anand Rajesh^{1,2}, Adnan Tufail^{3,4}, Cecilia Lee^{1,2}, Emily Chew¹⁰, Martin Friedlander^{5,6}, Aaron Lee^{1,2}

¹Department of Ophthalmology, University of Washington, Seattle, WA, USA

²The Roger and Angie Karalis Johnson Retina Center, Seattle, WA, USA

³Moorfields Eye Hospital, London, UK

⁴University College London, Institute of Ophthalmology, London, UK

⁵Lowy Medical Research Institute, La Jolla, CA, USA

⁶The Scripps Research Institute, La Jolla, CA, USA

⁷Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

⁸Oxford Eye Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

⁹Nuffield Laboratory of Ophthalmology, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

¹⁰Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD, USA

Abstract

Purpose: Deep learning (DL) models have achieved state-of-the-art medical diagnosis classification accuracy. Current models are limited by discrete diagnosis labels but could yield more information with diagnosis in a continuous scale. We developed a novel continuous severity scaling system for Macular Telangiectasia Type II (MacTel) by combining a DL classification model with Uniform Manifold Approximation and Projection (UMAP).

Design: We used a DL network to learn a feature representation of MacTel severity from discrete severity labels, and applied UMAP to embed this feature representation into 2 dimensions, thereby creating a continuous MacTel severity scale.

Subjects: A total of 2003 OCT volumes were analyzed from 1089 MacTel Project participants.

Corresponding Author: Aaron Y. Lee, MD, MSCI, Associate Professor, C. Dan and Irene Hunter Endowed Professorship, Department of Ophthalmology, University of Washington, 325 Ninth Avenue, Box 359608, Seattle, WA 98104, T: (206) 543-7250, leeay@uw.edu.

Meeting Presentation: Poster presentation at the Association for Research in Vision and Ophthalmology, Annual Meeting New Orleans, LA, May 4, 2023

This article contains additional online-only material. The following should appear online-only: Supplemental Figures 1–7 and Supplemental Video.

Methods: We trained a multi-view DL classifier using multiple B-scans from OCT volumes to learn the discrete 7-step Chew et al. MacTel severity scale. The classifiers' last feature layer was extracted as input for UMAP, which embedded these features into a continuous 2D manifold. The DL classifier was assessed in terms of test accuracy. Rank correlation for the continuous UMAP scale against Chew et al. was calculated. Additionally, the UMAP scale was assessed in the kappa agreement against 5 clinical experts on 100 pairs of patient volumes. For each pair of patient volumes, clinical experts were asked to select the volume with more severe MacTel disease, and compared against the UMAP scale.

Outcome Measures: Classification accuracy for the DL classifier, and kappa agreement vs clinical experts for UMAP.

Results: The multi-view DL classifier achieved top-1 accuracy of 63.3% (186/294) on held-out test OCT volumes. The UMAP metric shows a clear continuous gradation of MacTel severity that has a Spearman Rank Correlation of 0.84 with the Chew et al. scale. Furthermore, the continuous UMAP metric achieved kappa agreements of 0.56–0.63 with 5 clinical experts, which was comparable to inter-observer kappas.

Conclusions: Our UMAP embedding generated a continuous MacTel severity scale, without requiring continuous training labels. This technique can be applied to other diseases, and may lead to more accurate diagnosis, improved understanding of disease progression and key imaging features for pathologies.

Precis

Deep learning can achieve state-of-the-art disease severity classification using optical coherence tomography scans. We demonstrate that features identified by deep learning can be embedded into a continuous scale to allow more refined disease tracking.

Keywords

Deep learning; UMAP; feature embedding; continuous scale; optical coherence tomography; retina; MacTel

Introduction

Macular Telangiectasia Type II (MacTel) is a progressive, bilateral macular disease of unknown etiology that affects the neuroretina and capillary network. The first widely used MacTel severity grading system was created by Gass and Blodi in 1993 by analyzing color fundus photography and fluorescein angiography.¹ With increased usage of optical coherence tomography (OCT) in clinical practice, Chew et al. and the MacTel Research Consortium created a new grading system of MacTel in 2022 that incorporates disease features from new ophthalmic imaging modalities including spectral domain OCT (SD-OCT) and fundus autofluorescence (FAF). This severity classification system has 7 distinct severity classes which correspond to MacTel progression and visual acuity.² Discrete severity scales are used by clinicians to characterize a patient's MacTel disease severity better, discuss progression, and categorize subtypes, but they have limitations. For instance, a discrete scale does not allow for differentiation in severity between patients of the same

class, or a single patient's progression within the same class over time. Additionally, there may be difficulty with inter-rater reliability, with one study finding low inter-rater reliability among expert clinicians between Gass-Blodi classes for patients with less severe MacTel pathology.³ A continuous severity scale derived from imaging data may allow clinicians to have a more granular rating system to better assess and quantify a patient's disease progression. However, developing this continuous scale with human expertise alone is challenging.

Deep learning, specifically using neural networks, has been extremely successful when used to predict disease severity in ophthalmology.⁴⁻⁶ The neural network models are able to learn latent representations of the input data, which are features of the data that are not directly observable, but only appear in the deeper layers of the network. These latent representations are then used by the model to perform the final task. If a model is trained to perform a task like classification of disease severity, the latent representations of input data within the model's network should have some important characteristics that allow distinction between images from different classes of severity.⁷ The latent representations from the network are not constrained to discrete classes of the classification task, and contain information that may be used to generate a continuous scale of disease severity.

In this work, we set out to develop a continuous grading scale for MacTel severity with the goal of creating a more granular grading system. To achieve this, we trained a convolutional neural network to predict MacTel severity based on the most current grading scheme, the Chew et al. 7-class grading scheme, from multiple OCT B-scans of the same patient. Then we used uniform manifold approximation and projection (UMAP) to perform a dimensionality reduction of the latent features from the lower levels of the network to create a continuous severity scale. The new severity scale was evaluated against expert clinical grading to assess for validity.

Method

Data

This study was conducted in accordance with the Declaration of Helsinki. The subjects studied were from the MacTel Project. The MacTel Project is a collaboration of 49 clinical sites in 7 countries. Each participant in the MacTel Project was 18 years of age or older enrolled into the Natural History Study after a diagnosis of MacTel was confirmed on clinical examination at the study sites based upon on stereoscopic color fundus photographs, optical coherence tomography (OCT), fluorescein angiography and fundus autofluorescence images which were graded by the Reading Center at Moorfields Eye Hospital, London, UK. Each participating clinical site obtained approval from their institutional review board (IRB) or independent ethics committee for the protocol and each participant provided written informed consent.

We obtained 2,003 (Heidelberg Spectralis) OCT volumes from 1,089 participants in the MacTel Project. All OCT volumes imaged a 6mm*6mm area centered at the macula. However, the volume rasters varied between 49, 97 and 261 B-scan slices, with 97 the most common. Moreover, volume B-scan resolutions varied between 384px * 496px, 512px

* 496px, 784px * 496px, and 1024px * 496px. Volume scans that did not conform to 49, 97, 261 were rejected, resulting in 78 participants removed.

The 2,003 OCT volumes were matched to the closest clinical measurements within a 6 month window of the scan capture. The clinical measurements, derived from OCT scan features -OCT pigmentation, OCT hyper-reflectivity, presence of inner segment/outer segment (IS/OS) break, and neovascularization - were used to determine the 7 stage MacTel severity score from Chew et al., using their decision tree model reproduced in Supplemental Figure 1. Sample OCTs for the 7 Chew et al. grades are provided in Figure 2.

The data was split 70-15-15 in terms of training, validation and test at the patient level. To standardize B-scan resolutions, all B-scans were center cropped to be 384px * 384px. Vertical center cropping was achieved by detecting the ILM and RPE layers, and centering vertically around these layers. Horizontal cropping was performed by trimming from the center of the B-scan. The demographics of patients for each data partition along with the prevalence of the 7 Chew et al. grades at eye visit level are given in Table 1. We note that 97 participants were included in the study that only had OCT volumes, but no clinical scores and thereby no Chew et al. grades. These OCT volumes were used along with some OCT volumes from the test dataset for clinical validation.

Deep Learning Classifier

We adapted an EfficientNet-B0 to classify the OCT volumes to the 7 stage Chew et al. grades. Specifically, 20 central B-scans from the 97 raster OCT volumes and the 20 anatomically equivalent B-scans from the 49 and 261 OCT volumes, were used as inputs for EfficientNet-B0. The intuition for using multiple scans as input is to incorporate volumetric information that may not be necessarily captured by any individual scan; similar multi-view approaches have been previously used in^{8,9}. This input tensor was collapsed from 5 dimensions Batch(B), View(V), Channel(C), Height(H), Width(W) to into 4 dimensions B*V, C, H, W. EfficientNet-B0 then learnt a mapping between the input and the 7 class target labels using the cross entropy loss.

Once the multiview classifier was trained and model parameters tweaked on the validation dataset, the classifier was frozen. All train, validation and test data were run through the classifier so that the tensor from the final feature layer of the classifier could be extracted for UMAP Embedding. A detailed schematic of the DL training process and feature layer extraction is provided in Supplemental Figure 3.

UMAP Embedding

UMAP¹⁰ was applied to the extracted feature vectors to produce a lower dimensional interpretation of the feature vectors. Specifically, a 2D UMAP embedding, using the umap-learn 0.5.2 package for python 3.6.12 with non-default parameters min-dist=0.01, n_components=2, random_state=42 (full UMAP parameters are provided in the supplement, along with an ablation study using PCA embedding in Supplemental Figure 4), was fitted to the extracted feature vectors without any knowledge of the Chew et al. grades. The resulting 2D UMAP embedding is plotted and then color-coded for Chew et al. grades in Figure 5. A 1D continuous metric was extracted from the 2D UMAP embedding by fitting an Ordinary

Least Squares line to the embedding. Then for any point in the 2D UMAP embedding, its 1D UMAP metric is calculated by projecting it onto the line, and computing the distance of the projected point to the most severe MacTel diagnosis endpoint of this line. Supplemental Figure 6 is added to illustrate the method.

Evaluation

The classifier was evaluated on the test dataset using top-1 accuracy. The 1D continuous UMAP derived metric was compared against Chew et al. grades in terms of Spearman rank correlation. The Spearman rank correlation is used as Chew et al. grades are discrete so that a Pearson correlation is not a good metric of the correspondence between the continuous UMAP metric and the Chew et al. grades.

The consistency of the UMAP metric was validated by sampling. Specifically, the UMAP metric for each Chew et al. stage was compared against that of all stages, by sampling 100 pairs of volumes from each combination of stages. Then the UMAP metric for each volume was calculated and the delta difference between the pair of UMAP metrics was computed.

Five clinical experts were recruited to clinically validate the continuous UMAP metric. Four experts were retinal specialists with 10+ years of sub-specialty retinal experience, and the fifth was a reading center technician with 15+ years of MacTel experience. The five experts were presented with 100 pairs of OCT volumes and asked to choose if volume 1 had more severe MacTel (1), if volume 2 had more severe MacTel (2), or if they could not tell the difference (0). The 100 volume pairs were sampled to be: a) 35 volumes with 1-step difference in Chew et al. severity grade, b) 35 volumes with the same Chew et al. severity grade, and c) 30 longitudinal pairs from the same patient and same eye from different time points (time between volume pair ranged from 2.8 to 4.3 years). Only the 30 longitudinal pairs came from the same patient and same eye, but from different time points. The time between the volumes for the longitudinal data ranged from 2.75 to 4.25 years. The 70 volume pairs with the same, and 1-step difference, severity in the Chew et al. classification were from different patients and not necessarily the same eye as MacTel is diagnosed at the eye level. The inter-observer kappas for MacTel severity for all three setups were computed based on the experts' outputs of (0,1,2). Next, the UMAP metric was used to judge MacTel severity on these 100 volume pairs, outputting (0,1,2) like the clinical experts. Finally, the UMAP metric's kappa was computed against all observers.

Results

The multiview DL classifier achieved top-1 accuracy of 63.3% (186/294) on the test OCT volumes for a 7 class problem. The test dataset confusion table is given in Supplemental Figure 7, and shows mostly diagonal (63.3%) or correct top-1 predictions, with the miscalls coming in adjacent or at worst 2 grades away.

The 2D embedding, in Figure 5, shows a clear continuous gradation of MacTel severity. The 1D continuous UMAP severity metric extracted from this 2D embedding was found to have a Spearman Rank Correlation of 0.84 with Chew et al. grades. An interactive tool was developed to show the UMAP embedding along with 3 central OCT slices, the reference IR,

autofluorescence (FAF) and blue reflectance (BR) images captured at the same time where available. A video capture of the interactive tool is provided in the Supplemental Materials.

More importantly, the continuous UMAP metric achieved kappa agreements of 0.590 to Grader 1 on all 100 volume pairs. Overall, it performed comparably to all graders on all 100 volume pairs, on the 35 one-grade difference pairs, on the 35 same class pairs and the 30 longitudinal pairs (Figure 8). For 1 grade difference pairs, the UMAP scale achieved a kappa of 0.771 with Grader 1, and is comparable to the inter-observer kappas for this setup.

To investigate if there were more granular signals in MacTel than discernible by the 7-stage Chew et al. scale, we presented 5 clinical experts with 35 pairs of OCT volumes that had the same Chew et al. stage (5 pairs of volumes from each of the 7 stages), and asked them to determine, if possible, the OCT volume corresponding to more severe MacTel. The clinical experts achieved substantial kappa agreement between 0.45 and 0.6, which is evidence of more granular MacTel stages that is not captured by the current 7-stage scale, and which suggests the usefulness of finer scales. Our UMAP metric achieved comparable kappa (0.481) to the clinical experts on this task, demonstrating substantial agreement with clinical experts and suggesting that it is a suitable continuous scale that can capture more granular MacTel severity differences.

To understand disagreements between the UMAP scale and the clinical experts, we computed the absolute delta in UMAP metric between the volume pairs and plotted them against severity agreement between clinical expert and UMAP (blue), disagreement (red) and when clinicians was agnostic between the pair (green) in Figure 9, it is clear that most disagreements occur where the UMAP metric's absolute delta are close to zero.

Furthermore, we investigated the UMAP metric's absolute delta by experimental setup in Supplemental Figure 10. The 1 grade different Chew et al. classes have the biggest UMAP metric delta, and are correspondingly easier to grade, with graders agreeing most with the UMAP metric. This is also borne out by the high kappa agreements in Figure 8. When the Chew et al. class is the same for both volumes, the UMAP metric's absolute delta are smaller leading to more disagreement. Finally, the longitudinal setup had the smallest UMAP metric's absolute, leading to most disagreement, and this illustrates why kappa agreement for the longitudinal setup was lower for the other setups in Figure 8.

The UMAP metric consistency experiment results are shown in Supplemental Figure 11. The lower diagonal plots show the delta UMAP metric for each Chew et al. grade versus each other. The leading diagonal is the delta UMAP metric for the same Chew et al. grades, and exhibit a symmetric unimodal distribution around 0. The off diagonals are the less severe stages vs more severe stages, and the delta UMAP metric is shifted more and more negative as the steps between Chew et al. stages increases. This shows the UMAP metric is a consistent continuous metric scale.

Discussion

Our multi-view classifier and UMAP embedding generated a continuous severity scale for MacTel, without requiring continuous training labels. The UMAP severity scale had a high

Spearman rank correlation with the current discrete MacTel grades. More importantly, the UMAP scale achieved high kappa agreements against clinical experts, especially in the same Chew et al. grades and longitudinal setups. This shows that both clinicians and the UMAP severity scale can distinguish more granularly among MacTel cases within the same discrete scale.

Many diseases in medicine have a discrete severity scale with ordinal bins representing corresponding degrees of disease progression. However, these ordinal bins cannot represent continuous scale biology. Li et al attempted to create a continuous severity scale for retinopathy of prematurity and knee osteoarthritis using deep learning models. Their model was successful, but required manual labeling of 100 images in an ordinal scale of increasing severity for both disease domains in order to train their models. Our method, on the other hand, does not require any additional manual labeling to assess severity, but instead used existing clinical labels to generate latent representations that were used to generate the severity scale. Although latent representations have been used to study diseases in medical imaging, they have not been used to create continuous severity scales.⁷

Continuous severity scales may better represent underlying biological variation, and should be able to accurately reflect existing grading scales. In this study, we found that there was a high kappa agreement between graders as well as between the grader and UMAP grading when evaluating 1 grade different Chew et al. pairs. This suggests that the UMAP is reliable for finding the differences between eyes of the existing clinical grading criteria. When comparing kappa agreement among eyes of the same Chew grade, there was less agreement between graders and the UMAP, specifically among eyes that had a delta close to zero in the UMAP space. Among these eyes, there was also lower intergrader agreement for these eyes, so perhaps the disagreement does not reflect the UMAP being incorrect, but rather, that there is little agreement on which is more severe disease when the eyes are within the same Chew class among experts.

Vision outcomes in patients with MacTel may not change over the course of 2 years but there may be underlying pathological changes within that time.¹¹ Ideally, a continuous severity scale would be able to capture longitudinal changes within the same Chew grade that a clinician may be able to detect, but the traditional discrete severity grading scale cannot. Picking up longitudinal changes could allow for improved patient counseling and studying of disease. In our study, we attempted to compare the agreement of the physician raters with each other and with the UMAP scale. The longitudinal setup had relatively lower kappa agreements among the clinical experts and the UMAP scale suggesting that it was the hardest task, and which was borne out by the difference in UMAP scale metrics. Even for this difficult task, inter-observer and UMAP kappa were comparable and on the order of 0.3–0.5, which is substantially above a random chance kappa of 0. Although there was no general consensus on the severity among experts, the agreement of the UMAP with experts suggests that the UMAP metric is accurately assessing longitudinal changes. Therefore, the UMAP metric provides an easy way to provide a continuous scale diagnosis of MacTel, which can differentiate even minor changes in MacTel severity.

This study has some limitations. First, the continuous UMAP scale only applies to subjects with MacTel. It does not differentiate between MacTel and other retinal diseases. Second, the UMAP embedding was extracted from a supervised network, and required training labels. Therefore to replicate this continuous embedding workflow for other diseases or biological systems requires labeled data. Third, the 2D UMAP embedding for MacTel was linear, and allowed a straight-forward extraction of 1D continuous scale. The UMAP was trained using only OCT data. It is not known whether the same result would be obtained by training the UMAP with visual acuity as the only outcome. The Chew et al. grade requires multimodal imaging for the clinician label. It is not known whether a UMAP trained with all imaging modalities would achieve better performance.

Finally, our methodology would have relevance to other diseases with complex pathophysiology and large clinical imaging datasets that are currently constrained by discrete human labels of disease severity. For example, a similar OCT scan analysis can be applied to age-related macular degeneration (AMD)¹². Furthermore, it may apply to non-ophthalmic diseases with abundant routine clinical images, such as the potential to develop a continuous severity scale for COVID-19 from chest x-rays¹³ or chest CT¹⁴, for stroke severity from MRI¹⁵, and for continuous bowel inflammation severity from histopathology¹⁶. Future research will determine if these continuous severity scales, harnessed to linked clinical and multi-omic information and advanced data analysis and prognostic modeling will unlock previously unsuspected biological mechanisms and associations and new diagnostic and therapeutic approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Financial Support:

NIH/NIA R01AG060942 (Cecilia S. Lee); NIH/NEI K23EY029246 (Aaron Y. Lee), NIA/NIH U19AG066567 (Cecilia S. Lee, Aaron Y. Lee), Mexican Council of Science and Technology grant #2018-000009-01EXTF-00573 (AO-B), Lowy Medical Research Institute (Aaron Y. Lee), Research to Prevent Blindness Career Development Award (Aaron Y. Lee); Latham Vision Innovation Award (Cecilia S. Lee, Aaron Y. Lee), and an unrestricted grant from Research to Prevent Blindness (Cecilia S. Lee, Aaron Y. Lee). National Institutes of Health (Bethesda) Intramural research program funding (Emily Y. Chew). The sponsor or funding organization had no role in the design or conduct of this research. National Institute for Health Research Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology (Adnan Tufail, Catherine Egan). The views expressed are those of the authors and not necessarily those of the National Health Service, National Institute for Health Research, Department of Health, or the U.S. Food and Drug Administration. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of Interest:

Aaron Lee reports support from the US Food and Drug Administration, grants from Santen, Regeneron, Carl Zeiss Meditec, and Novartis, and personal fees from Genentech, Roche, and Johnson and Johnson, outside of the submitted work. This article does not reflect the opinions of the Food and Drug Administration. C. Egan, Heidelberg Engineering (personal fees), Novartis Pharmaceuticals (grant) A. Tufail, Allergan (C, R), Bayer (F, C, R), Kanghong (R), Heidelberg Engineering (C, R), Novartis (F, C, R), Roche/Genentech (C, R), Iveric Bio (C, R), Apellis (C, R), Thea (C, R) Tunde Peto OPTOS, Heidelberg, Zeiss, Alimera, Roche, Novartis, Boehringer-Ingelheim: speaker fee paid to myself; Novartis and B-I: grant paid to Institution. The remaining authors have no financial disclosures to report.

References

1. Gass JD, Blodi BA. Idiopathic juxtafoveolar retinal telangiectasis. Update of classification and follow-up study. *Ophthalmology* 1993;100:1536–1546. [PubMed: 8414413]
2. Chew EY, Peto T, Clemons TE, et al. Macular Telangiectasia Type 2: A Classification System Using MultiModal Imaging MacTel Project Report Number 10. *Ophthalmology Science* 2023;3:100261. [PubMed: 36846105]
3. Charbel Issa P, Helb H-M, Rohrschneider K, et al. Microperimetric assessment of patients with type 2 idiopathic macular telangiectasia. *Invest Ophthalmol Vis Sci* 2007;48:3788–3795. [PubMed: 17652753]
4. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016;316:2402–2410. [PubMed: 27898976]
5. Burlina PM, Joshi N, Pekala M, et al. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol* 2017;135:1170–1176. [PubMed: 28973096]
6. Varadarajan AV, Bavishi P, Ruamviboonsuk P, et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. *Nat Commun* 2020;11:130. [PubMed: 31913272]
7. Kang H, Xia L, Yan F, et al. Diagnosis of Coronavirus Disease 2019 (COVID-19) With Structured Latent Multi-View Representation Learning. *IEEE Trans Med Imaging* 2020;39:2606–2614. [PubMed: 32386147]
8. Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3D shape recognition. *arXiv [csCV]* 2015:945–953. Available at: https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Su_Multi-View_Convolutional_Neural_ICCV_2015_paper.html [Accessed February 22, 2023].
9. Elkahky AM, Song Y, He X. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. In: *Proceedings of the 24th International Conference on World Wide Web. WWW '15. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2015:278–288.* Available at: [Accessed February 21, 2023].
10. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [statML]* 2018. Available at: <http://arxiv.org/abs/1802.03426>.
11. Lamoureux EL, Maxwell RM, Marella M, et al. The longitudinal impact of macular telangiectasia (MacTel) type 2 on vision-related quality of life. *Invest Ophthalmol Vis Sci* 2011;52:2520–2524. [PubMed: 21217104]
12. Ferris FL, Davis MD, Clemons TE, et al. A simplified severity scale for age-related macular degeneration: AREDS Report No. 18. *Arch Ophthalmol* 2005;123:1570–1574. [PubMed: 16286620]
13. Danilov VV, Litmanovich D, Proutski A, et al. Automatic scoring of COVID-19 severity in X-ray imaging based on a novel deep learning workflow. *Sci Rep* 2022;12:12791. [PubMed: 35896761]
14. Saeed GA, Gaba W, Shah A, et al. Correlation between Chest CT Severity Scores and the Clinical Parameters of Adult Patients with COVID-19 Pneumonia. *Radiol Res Pract* 2021;2021:6697677. [PubMed: 33505722]
15. Sheth SA, Giancardo L, Colasurdo M, et al. Machine learning and acute stroke imaging. *J Neurointerv Surg* 2023;15:195–199. [PubMed: 35613840]
16. Peters LA, Friedman JR, Stojmirovic A, et al. A temporal classifier predicts histopathology state and parses acute-chronic phasing in inflammatory bowel disease patients. *Commun Biol* 2023;6:95. [PubMed: 36694043]

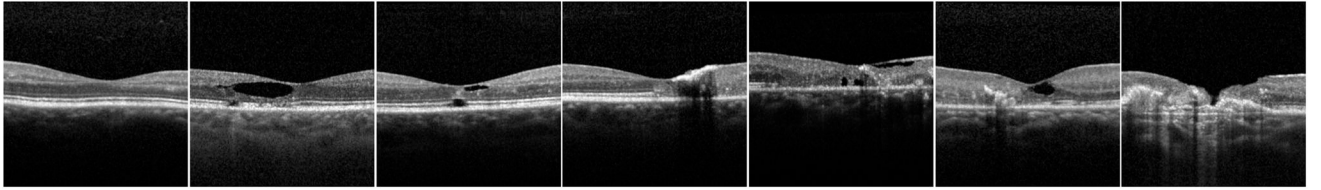


Figure 2.
A sample central OCTs for each of the 7 Chew et al.² MacTel severity stages, order by least severe (leftmost column) to most severe (rightmost column)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

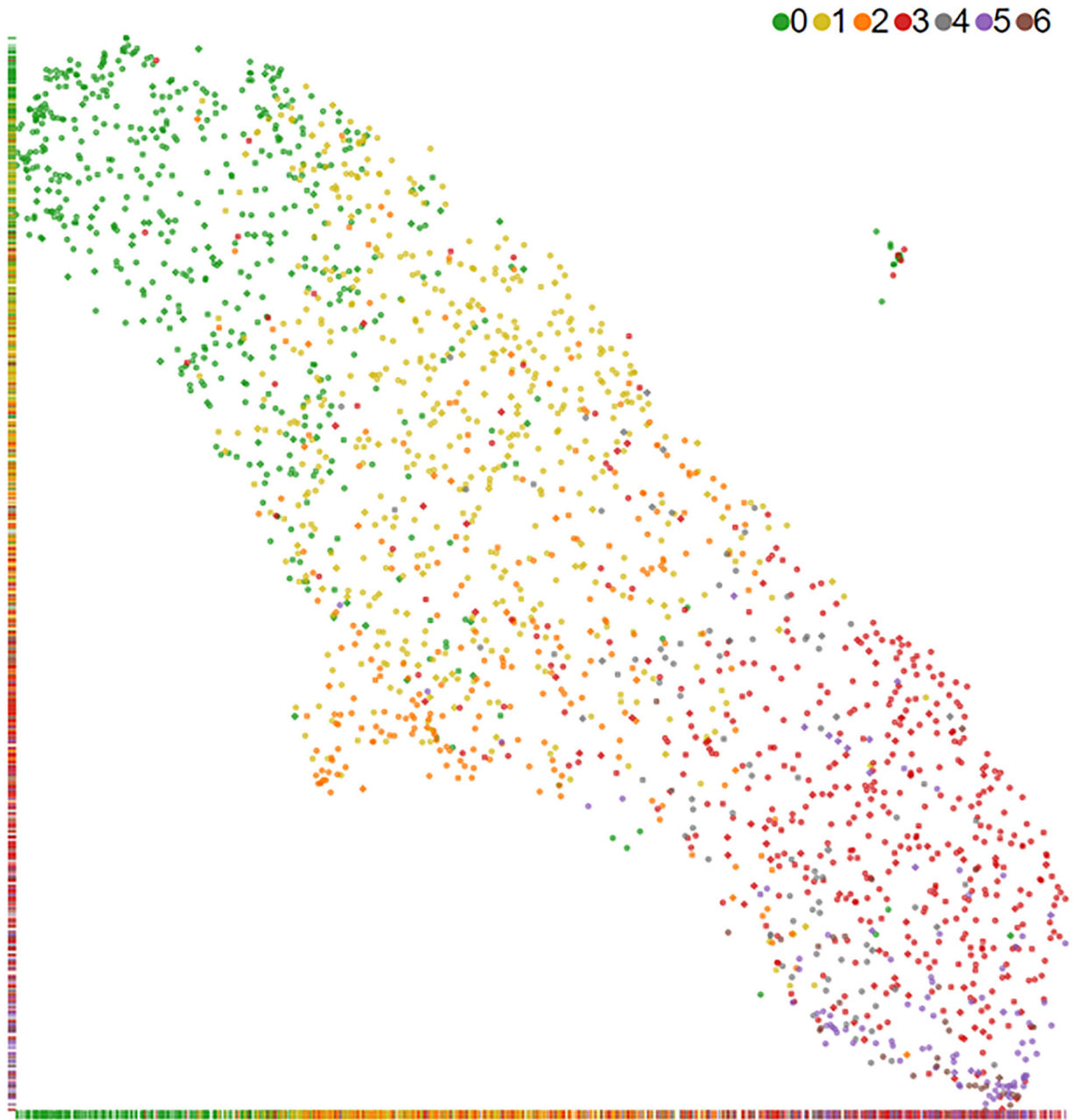


Figure 5. 2D UMAP embedding of MacTel OCT volumes. The Chew et al.² severity grade colors were added after embedding and shows a clear trend of least severe (green) to most severe (brown) from the top left to the bottom right in this UMAP embedding.

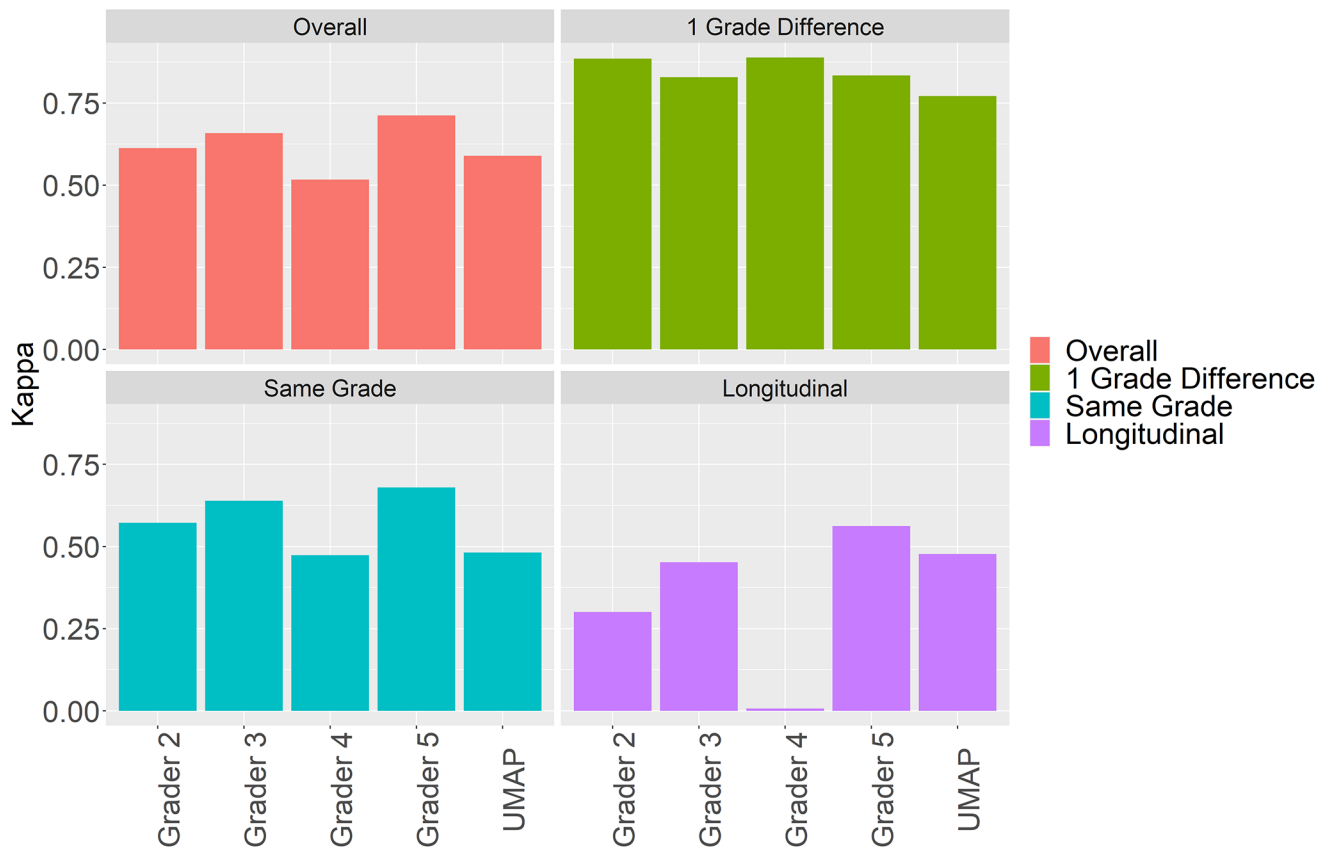


Figure 8.

Kappa agreement to Grader 1 for all 100 volume pairs, 35 pairs that were 1 Grade different on the discrete Chew et al.² scale, 35 same grade pairs, and 30 longitudinal pairs. UMAP achieved comparable kappa to Grader 1 as the other clinical experts.

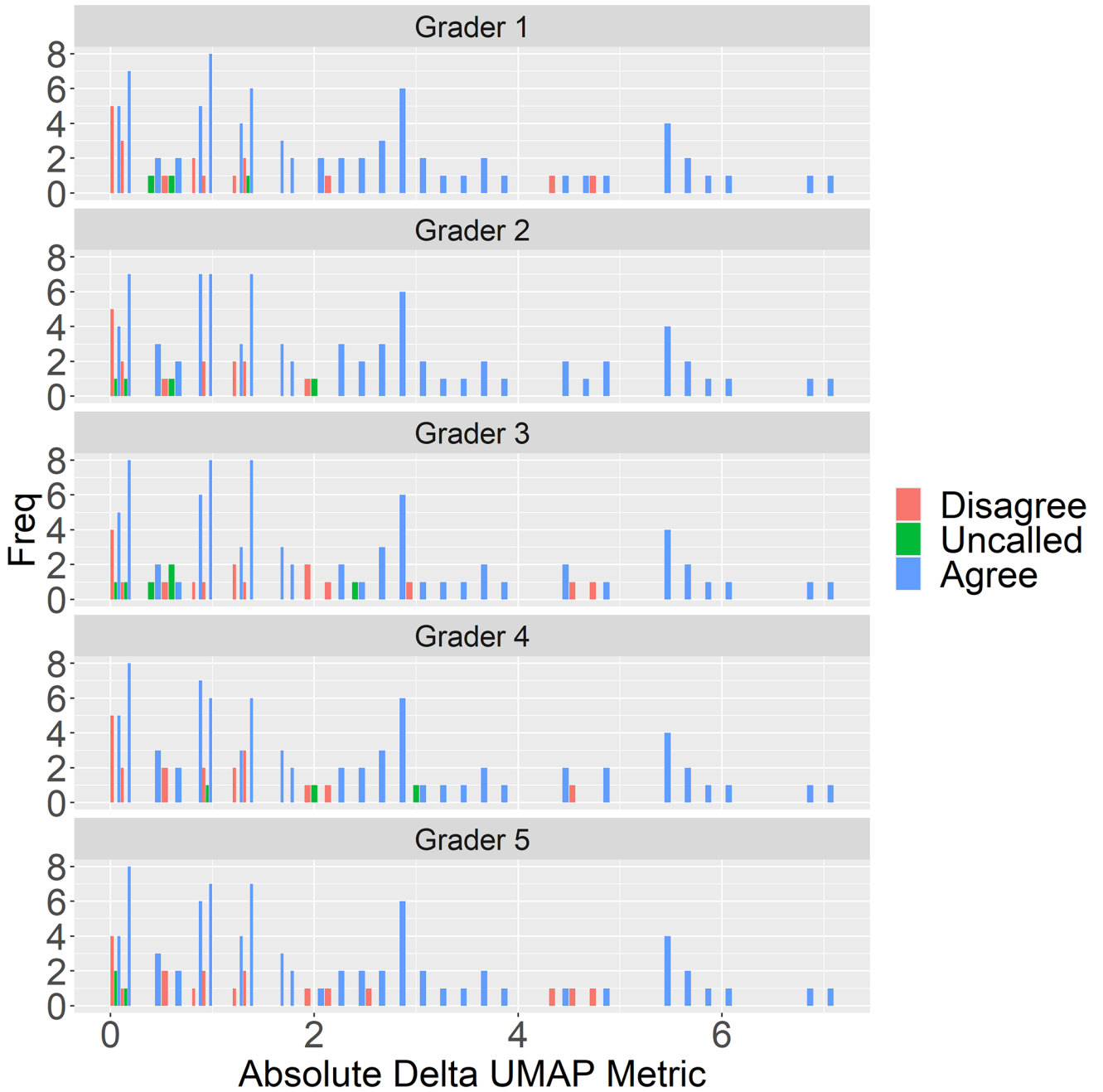


Figure 9. Graders (Grader 1–5) and UMAP metric agreement on 100 volume pairs vs absolute delta in UMAP metric for the volumes. Most disagreement (red) occurs when the absolute delta in UMAP metric is close to zero. Uncalled (green) are cases where graders could not distinguish between MacTel severity in the volume pair.

Table 1.

Demographics breakdown by data partition. Raw numbers are given as well as percentage of partition in parenthesis.

Category	Condition	Split			
		Train	Validation	Test	Unlabelled
Patient	Total	647	133	134	97
Sex	Male	237 (36.63%)	49 (36.84%)	47 (35.07%)	29 (29.90%)
	Female	409 (63.21%)	83 (62.41%)	86 (64.18%)	51 (52.58%)
	Unknown	1 (0.15%)	1 (0.75%)	1 (0.75%)	17 (17.53%)
Age	Mean	61.64	62.46	61.50	62.03
	Std	11.55	11.01	11.01	10.10
	10% Percentile	47.35	46.23	46.99	49.60
	25% Percentile	55.56	56.60	56.08	56.72
	50% Percentile	62.64	63.48	62.84	62.06
	75% Percentile	69.39	70.69	68.74	69.89
Race	90% Percentile	75.10	74.20	74.05	75.00
	White	587 (90.73%)	121 (90.98%)	121 (90.30%)	69 (71.13%)
	Asian	22 (3.40%)	4 (3.01%)	3 (2.24%)	4 (4.12%)
	Black/African	11 (1.70%)	1 (0.75%)	1 (0.75%)	1 (1.03%)
	American Indian	3 (0.46%)	1 (0.75%)	1 (0.75%)	0 (0.00%)
	Pacific Islander	0 (0.00%)	0 (0.00%)	1 (0.75%)	0 (0.00%)
Ethnicity	Other	23 (3.55%)	5 (3.76%)	6 (4.48%)	6 (6.19%)
	Unreported	1 (0.15%)	1 (0.75%)	1 (0.75%)	17 (17.53%)
	Not Hispanic or Latino	610 (94.28%)	129 (96.99%)	125 (93.28%)	77 (79.38%)
	Hispanic or Latino	31 (4.79%)	3 (2.26%)	8 (5.97%)	2 (2.06%)
	Unknown	6 (0.93%)	1 (0.75%)	1 (0.75%)	18 (18.56%)
	Eyes	Unique Eyes	1313	270	275
Total Scans		1431	278	294	975
Chew et al. grades	0	381	80	80	0
	1	382	63	78	0
	2	194	36	20	0
	3	312	74	81	0
	4	66	13	13	0
	5	75	5	19	0
	6	21	7	3	0
	Unknown	0	0	0	975