ORIGINAL RESEARCH

Statistical Evaluation of Smartphone-Based Automated Grading System for Ocular Redness Associated with Dry Eye Disease and Implications for Clinical Trials

John D Rodriguez[1], Adam Hamm[2], Ethan Bensinger[1], Samanatha J Kerti[2], Paul J Gomes[3], George W Ousler III[3], Palak Gupta [1], Carlos Gustavo De Moraes[3,4], Mark B Abelson[1,3,5]

[1]Andover Eye Institute, Andover, MA, 01810, USA; [2]SDC, Tempe, AZ, 85288, USA; [3]Ora, LLC, Andover, MA, 01810, USA; [4]Columbia Medical Center, New York, NY, 10032, USA; [5]Harvard Medical School, Boston, MA, 02115, USA

Correspondence: Paul J Gomes, Allergy & Blepharitis, Ora, LLC, 138, Haverhill Street, Suite 102, Andover, MA, 01810, USA, Tel +1-978-685-8900, Fax +1-978-689-0020, Email pgomes@oraclinical.com

Purpose: This study introduces a fully automated approach using deep learning-based segmentation to select the conjunctiva as the region of interest (ROI) for large-scale, multi-site clinical trials. By integrating a precise, objective grading system, we aim to minimize inter- and intra-grader variability due to perceptual biases. We evaluate the impact of adding a "horizontality" parameter to the grading system and assess this method's potential to enhance grading precision, reduce sample size, and improve clinical trial efficiency.

Methods: We analyzed 29,640 images from 450 subjects in a multi-visit, multi-site clinical trial to assess the performance of an automated grading model compared to expert graders. Images were graded on a 0–4 scale, in 0.5 increments. The model utilizes the DeepLabV3 architecture for image segmentation, extracting two key features—horizontality and redness. The algorithm then uses these features to predict eye redness, validated by comparison with expert grader scores.

Results: The bivariate model using both redness and horizontality performed best, with a Mean Absolute Error (MAE) of 0.450 points (SD=0.334) on the redness scale relative to expert scores. Expert graded scores were within one unit of the mean grade in over 85% cases, ensuring consistency and optimal training set for the predictive model. Models incorporating both features outperformed those using only redness, reducing MAE by 5–6%. The optimal generalized model improved predictive accuracy with horizontality such that 93.0% of images were predicted with an absolute error less than one unit difference in grading.

Conclusion: This study demonstrates that fully automating image analysis allows thousands of images to be graded efficiently. The addition of the horizontality parameter enhances model performance, reduces error, and supports its relevance to specific Dry Eye manifestations. This automated method provides a continuous scale and greater sensitivity to treatment effects than standard clinical scales.

Keywords: dry eye hyperemia, deep-learning conjunctiva segmentation, redness, automated grading

Introduction

Conjunctival hyperemia is an important clinical endpoint for the development of therapies to treat inflammatory and infectious diseases of the eye. The current gold standard for the evaluation of redness in clinical trials is based on our validated endpoints using clinical scales developed from ophthalmic practice.[1,2] The development of these scales has relied on extensive clinical observations of the subtle details of characteristic patterns of redness for these various ocular surface diseases such as Dry Eye Disease (DED), ocular conjunctivitis and blepharitis.[1–4] In DED, vasodilation is primarily seen in the inter-palpebral fissure area which is exposed to the environment. This vasodilation is predominantly marked by the presence of fine horizontal blood vessels in this region.[4] These horizontal vessels appear to become

particularly prominent as the severity of the redness condition increases. The Ora CalibraTM Conjunctival Dry Eye Redness Scale or OCDER[5,6] (henceforth referred to as redness scale) was a clinical scale developed to incorporate the parameters of redness intensity, location and prominence of fine horizontal conjunctival vessels (referred to as horizontality or FHCV) observed in DED. This clinical grading scale enables investigators to assess ocular hyperemia on a 0–4 scale, with 0.5-point increments, utilizing both reference images and descriptive criteria for precise evaluation. For over four decades, these scales have been used by clinicians for manual grading extensively in clinical trials, focused on DED. Importantly, patterns of redness vary significantly across different ocular conditions, such as dry eye, allergies, infections, and corneal ulcers, each displaying unique characteristics in vascular presentation, distribution and intensity of redness.[7–10] This variation underscores the necessity for condition-specific grading scales tailored to each disease, enhancing the accuracy and relevance of hyperemia assessment in diverse ocular pathologies.

Although such scales are currently accepted as defining pivotal endpoints in clinical trials by the FDA a number of limitations are inherent in their execution. Previous studies have shown the limits of human perception to clinical grading systems.[11,12] Innate perceptual biases such as primacy have been shown to affect relative scores assessed when grading a series of images. Grader fatigue may also affect grading results.[13] Sensitivity of scales is reduced when using a restricted number of grading levels.[11,14–16] However, human perception is unable to differentiate observations accurately when too many grading levels are used.[16,17] The development of an objective system of grading based on image analysis has the potential for eliminating perceptual bias and inter-grader variability. Eliminating bias is crucial in multi-center clinical trials, especially for Dry Eye treatments, where the subtlety of redness patterns makes consistency across sites essential. The spatial distribution, intensity, and pattern of redness on the conjunctiva can help differentiate between other causes of ocular redness, such as corneal ulcers, lens complications, acute inflammation, and allergies, as each condition presents a unique redness profile and does not face the same consistency challenges.[8,9] Skilled clinicians utilize these variables as diagnostic indicators to refine their assessments.

To capture the desired precision of redness grading, a comprehensive approach to characterize hyperemia based on image analysis has been developed over the years.[5,10,18–24] These investigations have attempted to define characteristic features of the vascular pattern as well as redness hue (chromaticity). These features include vessel width, percent area occupied by vessels, number of vessels and vascular complexity based on fractal dimension as well as many others.[25] In order to identify which features, such as horizontal vascular pattern, are relevant to a given ocular disease, observations from clinical practice as well as psychophysical visual perception must be considered. When comparing clinical scales, a combination of redness and some vascular parameters, such as vessel edge detection or vessel area have been shown to be effective.[10,26]

In an earlier study, we developed and reported an automated computer redness grading system which incorporates both redness and horizontal vascular components as a composite score which is mapped onto the already established the OCDER grading scale.[5,6] These vessel features have been observed to increase with level of dry eye severity in clinical observations by ophthalmologists. In the initial study, a relevant image Region of Interest (ROI) ie, conjunctiva was defined as a manually selected rectangular area centered on the midpoint of the palpebral fissure opening. While this ROI was proportional to the dimensions of each image and varied slightly in size, it effectively represented the area assessed during clinical grading of dry eye redness. This manual assisted approach while helpful, greatly limited high volume grading throughput. In the current study, this process is entirely automated using advanced AI-based segmentation software. Rather than a fixed rectangular ROI, the entire conjunctiva is segmented, enabling a more comprehensive and detailed evaluation of redness parameters across a larger region. This fully automated approach leverages deep learning-based segmentation techniques to define the conjunctiva as the ROI, significantly enhancing the efficiency and scalability of data analysis for multi-site clinical trials.

Additionally, this method offers a more precise and standardized grading system by eliminating potential sources of inter-grader and intra-grader variability inherent in manual assessments, thereby mitigating perceptual biases. This advancement supports more accurate and reproducible evaluations, particularly in large-scale studies. The primary aim of this study is to develop and validate an advanced, fully automated redness grading system for the conjunctiva, leveraging AI-based deep learning segmentation techniques. By automating the process of region-of-interest (ROI)

selection and expanding the assessment area to include the entire conjunctiva rather than a manually selected rectangular region, this study seeks to enhance the precision, comprehensiveness, and scalability of ocular redness evaluations.

Methods

All images were collected in accordance with Institutional Review Board regulations and with the ethical principles that originated with the Declaration of Helsinki. The informed consent and study protocol were approved by a properly constituted IRB (Alpha IRB, San Clemente, CA). Prior to a subject's participation, the study was discussed and subjects wishing to participate gave written informed consent.

Subjects

A total of 450 subjects were recruited as part of a series of multisite study sponsored by Aldeyra Therapeutics Inc. (NCT04971031,[27] NCT05062330,[28] NCT04674358[29]). Clinical resulting in 29,640 images captured. Subjects were recruited over 3 study visits in Phase 2 and 3 trials. Subjects were included if they had a history or desire to use artificial tear substitutes for symptoms of dry eye within the previous 6 months and a patient- or investigator-reported history of DED in both eyes. Participants were excluded if they had any other eye inflammation, infection, or condition that could pose a risk, confound the study results, or interfere with their participation. Additionally, those with a history of laser-assisted in situ keratomileusis (LASIK) or other corneal refractive surgeries, or who used contact lenses or any prohibited medications during the defined washout period of 30 days for medications known to cause ocular drying or 90 days for any prescription treatments for DED, were also excluded. Images of the temporal and nasal conjunctival regions were acquired using the Ora EyeCup™, a smart phone-based AI-supported system enabling patients to capture high-resolution images of their eyes and monitor disease signs and symptoms.[30,31] The system assesses ocular redness, tear film stability, enhancing the quality and efficiency of ophthalmic clinical studies.

Investigator Grading System

Three trained investigators independently graded a total of 20424 photographs each using the clinical grading redness scale described above, developed specifically for dry eye based on manual interactive determination of redness intensity.[5,6] This investigator-based redness scale, calibrated to reference photographs, measures both redness intensity and the presence of horizontal vessels in the bulbar conjunctiva. The scale ranges from 0 to 4, where 0 indicates no redness and 4 indicates severe vasodilation, with increments of 0.5 permitted. In addition, at least one investigator (but not all three) graded an additional 9216 photographs using the same clinical grading system. The same photographs were then graded by computer automated image analysis, which provided two separate grades for redness intensity and horizontal vessels.

Computer Automated Image Analysis

Selection of Region of Interest

The temporal bulbar conjunctiva was segmented from the standard external eye image using an automated algorithm based on the DeepLabV3 architecture,[32] a convolutional neural network (CNN) tailored for semantic image segmentation. DeepLabV3 applies Atrous Spatial Pyramid Pooling (ASPP)[33,34] to capture multi-scale contextual information and improve feature resolution, enabling accurate pixel-wise classification across diverse image regions. Key architectural features, including dilated convolutions, expand the receptive field without sacrificing spatial detail, which is critical for detecting intricate structures in images. Batch normalization and residual connections further enhance model stability and gradient flow during training, ensuring reliable performance in complex segmentation tasks.

DeepLabV3 has been successfully applied to several ophthalmic segmentation challenges, such as retinal layer mapping in optical coherence tomography (OCT) and corneal shape delineation in Scheimpflug images. For instance, García et al demonstrated DeepLabV3's efficacy in segmenting corneal layers in anterior segment OCT, effectively capturing multi-layered ocular structures.[35] Similarly, Yin et al used DeepLabV3 to segment retinal layers and lesions in OCT images for diabetic retinopathy patients, yielding valuable diagnostic insights for clinical monitoring and decision-making.[36]

The segmented image was analyzed for redness hue intensity (Auto-Redness) and vascular pattern (Auto-Horizontality). Redness intensity was measured as the average redness intensity from a single-channel 8-bit image. The vascular pattern was assessed using the blue channel of the image, identifying vessels in regions where the mean intensity exceeded 97%. To calculate horizontality, the first-order Sobel derivatives in the vertical direction were applied to these regions. Specifically, the vertical Sobel derivative of the blue channel image intensity, averaged over each image, defined the horizontality. The vascular pattern feature, chosen based on clinical observations of redness patterns in DED patients, represents the horizontal component of the vessel structure. Details of the calculation of the Auto-Redness and Auto-Horizontality parameters have been described previously.[5] The output of these calculations are two separate continuous values from (0–255) and (0–100) for Auto-Redness and Auto-Horizontality, respectively.

Additional transformations to these outputs were performed to address the non-linearity of the data. These transformations include the Normalization of Redness and Horizontality scores, as well as Z-scores and log transformations, as reported in our previous study.[5,6]

Comparison to Clinical Grades

In order to investigate the relative contributions of the Auto Horizontality and Auto Redness parameters, comparisons with the clinical grades were calculated for Auto-Redness alone as well as a model incorporating both the Auto Horizontality and Auto Redness score.

To make this comparison, two analyses were performed. First, the dataset that contains observations where all 3 graders provided scores was split into an 80–20 Train Test Split, where the model's target is the mean of the 3 expert grader scores, and the model inputs are subset of the transformations of the Auto Horizontality and Auto Redness outputs. Second, the remainder of the observations were added to this dataset and split into an 80–20 Train Test Split, where the model's target is the mean of the 1, 2, or 3 expert grader scores. Multiple linear models were then trained on different combinations of these inputs.

Statistical Models and Methods

A univariate generalized linear model for mean grader score was fit using redness and horizontality. A bivariate generalized linear model for mean grader score was fit using redness and horizontality in the same model. Mathematical forms of the models are below.

$$\text{Univariate Redness Model Mean Grader Score} = \beta_0 + \beta_1 * \text{redness} + \varepsilon$$

$$\text{Univariate Horizontality Model Mean Grader Score} = \beta_0 + \beta_1 * \text{horizontality} + \varepsilon$$

$$\text{Bivariate Mean Grader Score} = \beta_0 + \beta_1 * \text{redness} + \beta_2 * \text{horizontality} + \varepsilon$$

Using training data, estimates were calculated for each of the coefficients in the model. These estimates were then used with the test data to calculate predicted values for the mean grader score. The absolute error for each observation is the absolute value of the difference between the predicted value and the observed value. The mean absolute error (MAE) and standard deviation of absolute error were then calculated for each model for comparison.

Results

We analyzed a total of 29640 images acquired from 450 subjects. The data were collected as part of a phase 2 and 3 trials. Of all the models tested – univariate redness and horizontality and the bivariate model, the best performing linear model was the bivariate redness and horizontality model with all three graders and had a Mean Absolute Error (MAE) of 0.45 pts, with a standard deviation of 0.334, on the redness scale, relative to the mean of the three expert grader scores. The mean absolute errors, along with the standard deviation of the absolute errors, of each model are shown in Table 1.

When comparing a model trained on Auto Redness and Auto Horizontality data to ones trained only with Auto Redness scores, we observed improvement in the model's predictive output. In both models, the addition of horizontality

Table 1 Performance metrics of redness and horizontality models describing mean absolute error and SD against a panel of three graders

Model, Response	MAE (SD)
Univariate Redness, Mean of all 3 graders	0.479 (0.352)
Univariate Horizontality, Mean of all 3 graders	0.512 (0.375)
Bivariate Redness and Horizontality, Mean of all 3 graders	0.450 (0.334)
Univariate Redness, Mean of all available graders	0.481 (0.352)
Univariate Horizontality, Mean of all available graders	0.502 (0.365)
Bivariate Redness and Horizontality, Mean of all available graders	0.457 (0.337)

Table 2 Comparison of Automated Score Prediction to Grader Score Differences

Difference from Predicted Value/Mean Grade	Automated Scores	Mean of Grader Scores
<1 Unit	93.0%	85.8%
≥ 1 Unit	7.0%	14.2%
Mean (SD)	0.450 (0.334)	0.487 (0.362)

to the model reduced the MAE by 5–6% as opposed to using redness alone. The best generalized linear model had an MAE of 0.45 points (SD=0.334) on the redness scale.

Since the mean grader score is commonly used as a ground truth measure, as it reflects the central tendency of the data and serves as a reliable baseline for comparison, we aimed to evaluate the utility of the automated score against this mean grader score. To achieve this, for each image, we calculated the absolute difference between each of the three investigator's individual scores and the mean of these three grader scores. We then categorized these differences, similarly, grouping them by absolute differences of less than one unit and those greater than or equal to one unit.

The comparison of each investigator's scores with the mean of the three graders' scores highlights the strong alignment across assessments. Specifically, 85.8% of the investigator scores were within the clinically relevant difference threshold of one unit from the mean of the grader scores, reflecting a high degree of consistency and supporting the robustness of this dataset for model training. The mean absolute difference of the individual grader score to the mean grader score was 0.487 units with a standard deviation of 0.362. This alignment in scoring illustrates the reliability of the grading approach and ensures that the model is based on data with minimal clinically meaningful discrepancies. The mean absolute difference of the individual grader score to the mean grader score was 0.487 units with a standard deviation of 0.362.

In comparison, the best generalized linear model had an MAE of 0.450 points (SD=0.334) on the redness scale, with 93.0% of images predicted with an absolute error less than one unit, and 7.0% of images predicted with an absolute error greater than or equal to one unit. According to FDA guidelines and historical consensus, a one-unit difference (reported with statistical significance) is considered a clinically significant change in redness grading.[25,59] Table 2 highlights the similarity and potential improvement of using automated scores compared to the mean of the grader scores.

Discussion

This study advances automated grading methods for conjunctival hyperemia by integrating redness and horizontality into a statistical model. Analysis of 29,640 images from 450 subjects in phase 2 and 3 trials showed that the bivariate model

outperformed univariate models, reducing Mean Absolute Error (MAE) by 5–6%. These findings align with prior research showing that models incorporating anatomical context improve accuracy.[17,37]

Previous studies focused on redness intensity but lacked structural parameters like horizontality.[38,39] Specifically, Brea et al[40] developed a machine learning-based tool for grading conjunctival hyperemia, by incorporating a wide variety of features. However, this model is hindered by the large number of features which do not incorporate specific observed characteristics of ocular hyperemia in specific indications such as Dry Eye disease. This may lead to inefficiency and can invariably bias the model toward more common grades and is also dependent on the quality of the inputs.[40] Unlike these approaches, our model incorporates a single feature (horizontality) which has been shown to characterize hyperemia in Dry Eye disease based on clinical observations. Since this parameter is also incorporated into the clinical redness scale, comparison of manual and automated grading is straightforward. In addition, our fully automated model uses a convolutional neural network approach to select region of interest (ROI) and standardize image analysis, minimizing variability and improving consistency. With 85.8% of investigator scores within one unit of the mean grader score and 93.0% of automated predictions achieving similar accuracy, this model demonstrates high reliability, addressing issues like inter-grader variability and fatigue.[25,39]

Implications of the ROI System
The ROI system standardizes grading by isolating relevant anatomical features, addressing variability in framing and lighting. This ensures repeatability and facilitates precise detection of subtle redness changes, critical for evaluating treatment effects in DED. The system's consistency across repeat readings further underscores its reliability.

Broader Implications for Multi-Center Trials
In multi-center trials, inter-grader variability often complicates data interpretation and increases costs. By providing standardized and objective grading, this method can reduce recruitment efforts, trial durations, and costs. Its sensitivity to subtle changes in hyperemia enhances its utility for detecting treatment effects.

Conclusion
The current paper extends the processing of digital images to a fully automated process in which thousands of images can be graded in several minutes. The results here confirm that the incorporation of the horizontality parameter into the statistical model of a large dataset improves the performance of the model and reduces the absolute error relative to manual grading by clinicians. These results support the relevance of this model to specific manifestations of DED associated hyperemia. The automated ROI segmentation method allows for a more robust identification of the conjunctiva and results in a continuous scale and higher sensitivity to the effect of treatment than standard clinical scales.

Although the use of an objective method can provide more sensitivity and precision than manual grading, the most significant aspect is the potential to improve upon the inter-grader variability in multi-center clinical trials. An objective method of analysis based on digital images has the potential of improving the efficiency of multi-center clinical trials by utilizing a method that is not affected by inter-grader variability. It also holds the potential for considerable reduction in time scale, recruitment effort and costs while maximizing repeatability and sensitivity to treatment.

Data Sharing Statement
We appreciate your interest in our research. While our policy generally restricts sharing protected data from sponsored trials, we assess each request on a case-by-case basis. Clarifications regarding our findings and methodology may be provided by the corresponding author, provided they align with our proprietary data and methods policies. For any such inquiries, please contact Paul Gomes (pgomes@oraclinical.com).

Acknowledgments

Disclosure

JDR, EB, PG and MBA are employees at Andover Eye Institute. AH, SK, PJG, GWO and CGDM are consultants of Andover Eye Institute. Aldeyra Therapeutics, Inc. paid Ora, LLC for CRO services. The authors report no other conflicts of interest in this work.

References

1. Abelson MB. Code Red: the Key Features of Hyperemia. https://www.reviewofophthalmology.com/article/code-red-the-key-features-of-hyperemia. Accessed January 21, 2025.
2. Abelson MB, Loeffler O. Conjunctival allergen challenge: models in the investigation of ocular allergy. *Curr Allergy Asthma Rep*. 2003;3 (4):363–368. doi:10.1007/s11882-003-0100-z
3. Abelson MB, Chambers WA, Smith LM. Conjunctival allergen challenge. A clinical approach to studying allergic conjunctivitis. *Arch Ophthalmol*. 1990;108(1):84–88. doi:10.1001/archopht.1990.01070030090035
4. Ousler GW, Gomes PJ, Welch D, Abelson MB. Methodologies for the study of ocular surface disease. *Ocul Surf*. 2005;3(3):143–154. doi:10.1016/s1542-0124(12)70196-9
5. Rodriguez JD, Johnston PR, Ousler GW, Smith LM, Abelson MB. Automated grading system for evaluation of ocular redness associated with dry eye. *Clin Ophthalmol*. 2013;7:1197–1204. doi:10.2147/OPTH.S39703
6. Rodriguez JD, Lane KJ, Ousler III GW, Angjeli E, Smith LM, Abelson MB. Automated Grading System for Evaluation of Superficial Punctate Keratitis Associated With Dry Eye. *Invest Ophthalmol Visual Sci*. 2015;56(4):2340–2347. doi:10.1167/iovs.14-15318
7. How red is a white eye? Clinical grading of normal conjunctival hyperaemia - PubMed. Available from: https://pubmed.ncbi.nlm.nih.gov/16518366/. Accessed January 20, 2025.
8. Baratz KH. Anterior Segment Disease: a Diagnostic Color Atlas. *Archiv Ophthalmol*. 2001;119(6):929–930.
9. FRCSC MBA MD, CM. Code Red: the Key Features of Hyperemia. Available from: https://www.2020mag.com/article/code-red-the-key-features-of-hyperemia. Accessed January 15, 2025.
10. Macchi I, Bunya VY, Massaro-Giordano M, et al. A new scale for the assessment of conjunctival bulbar redness. *Ocular Surf*. 2018;16(4):436–440. doi:10.1016/j.jtos.2018.06.003
11. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Visual Sci*. 1991;32(2):422–432.
12. Sirazitdinova E, Gijs M, Bertens CJF, Berendschot TTJM, Nuijts RMMA, Deserno TM. Validation of Computerized Quantification of Ocular Redness. *Transl Vis Sci Technol*. 2019;8(6):31. doi:10.1167/tvst.8.6.31
13. Bridge P, Fielding A, Rowntree P, Pullar A. Intraobserver Variability: should We Worry? *J Med Imaging Radiat Sci*. 2016;47(3):217–220. doi:10.1016/j.jmir.2016.06.004
14. Amparo F, Yin J, Di Zazzo A, et al. Evaluating Changes in Ocular Redness Using a Novel Automated Method. *Trans Vision Sci Technol*. 2017;6 (4):13. doi:10.1167/tvst.6.4.13
15. Amparo F, Wang H, Emami-Naeini P, Karimian P, Dana R. The Ocular Redness Index: a novel automated method for measuring ocular injection. *Invest Ophthalmol Vis Sci*. 2013;54(7):4821–4826. doi:10.1167/iovs.13-12217
16. Schulze MM, Hutchings N, Simpson TL. Grading Bulbar Redness Using Cross-Calibrated Clinical Grading Scales. *Invest Ophthalmol Visual Sci*. 2011;52(8):5812–5817. doi:10.1167/iovs.10-7006
17. Fieguth P, Simpson T. Automated Measurement of Bulbar Redness. *Invest Ophthalmol Visual Sci*. 2002;43(2):340–347.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
19. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255–268. doi:10.2307/2532051
20. Maldonado MJ, Arnau V, Martínez-Costa R, et al. Reproducibility of digital image analysis for measuring corneal haze after myopic photo-refractive keratectomy. *Am J Ophthalmol*. 1997;123(1):31–41. doi:10.1016/s0002-9394(14)70989-4
21. Masumoto H, Tabuchi H, Yoneda T, et al. Severity Classification of Conjunctival Hyperaemia by Deep Neural Network Ensembles. *J Ophthalmol*. 2019;2019:7820971. doi:10.1155/2019/7820971
22. Murphy PJ, Lau JSC, Sim MML, Woods RL. How red is a white eye? Clinical grading of normal conjunctival hyperaemia. *Eye*. 2007;21 (5):633–638. doi:10.1038/sj.eye.6702295
23. McMonnies CW, Chapman-Davies A. Assessment of conjunctival hyperemia in contact lens wearers. Part I. *Am J Optom Physiol Opt*. 1987;64 (4):246–250. doi:10.1097/00006324-198704000-00003
24. Otero C, García-Porta N, Tabernero J, Pardhan S. Comparison of different smartphone cameras to evaluate conjunctival hyperaemia in normal subjects. *Sci Rep*. 2019;9(1):1339. doi:10.1038/s41598-018-37925-5
25. Papas EB. Key factors in the subjective and objective assessment of conjunctival erythema. *Invest Ophthalmol Vis Sci*. 2000;41(3):687–691.
26. Owen CG, Fitzke FW, Woodward EG. A new computer assisted objective method for quantifying vascular changes of the bulbar conjunctivae. *Ophthalmic Physiol Opt*. 1996;16(5):430–437. doi:10.1046/j.1475-1313.1996.96000373.x
27. Aldeyra Therapeutics, Inc. *A Multi-Center Randomized, Double-Masked, Parallel Design, Vehicle-Controlled Phase 2 Clinical Trial to Assess the Efficacy and Safety of 0.25% Reproxalap Ophthalmic Solution Compared to Vehicle in Subjects With Dry Eye Disease*. 2023. Available from: https://clinicaltrials.gov/study/NCT04971031. Accessed January 18, 2025.
28. Aldeyra Therapeutics, Inc. *The TRANQUILITY 2 Trial: multi-Center Randomized, Double-Masked, Parallel Design, Vehicle-Controlled Phase 3 Clinical Trial to Assess the Efficacy and Safety of 0.25% Reproxalap Ophthalmic Solution Compared to Vehicle in Subjects With Dry Eye Disease*. 2023. Available from: https://clinicaltrials.gov/study/NCT05062330. Accessed January 18, 2025.
29. Aldeyra Therapeutics, Inc. *The TRANQUILITY Trial: multi-Center Randomized, Double-Masked, Parallel Design, Vehicle-Controlled Phase 2/3 Clinical Trial to Assess the Efficacy and Safety of 0.25% Reproxalap Ophthalmic Solution Compared to Vehicle in Subjects With Dry Eye Disease*. Available from: https://clinicaltrials.gov/study/NCT04674358. Accessed January 18, 2025.
30. Sinyak I, Bensinger E, Marquis M, Rodriguez JD, Abelson MB. Reliability of Redness Imaging with the Ora EyeCup Phone. *Invest Ophthalmol Visual Sci*. 2022;63(7):1562–A0287.

 913

31. Marquis M, Abelson MB, Rodriguez JD, Bensinger E, Sinyak I. Patient self acquired photos with the Ora Eyephone compared to self reported eye redness in mobile biocube. *Invest Ophthalmol Visual Sci*. 2022;63(7):1561–A0286.

32. [1706.05587] Rethinking Atrous Convolution for Semantic Image Segmentation. Available from: https://arxiv.org/abs/1706.05587. Accessed January 18, 2025.

33. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv*. 2017. doi:10.48550/arXiv.1706.05587

34. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y editors. *Computer Vision – ECCV 2018*. Springer International Publishing; 2018:833–851. doi:10.1007/978-3-030-01234-2_49.

35. Mathai TS, Lathrop K, Galeotti J. Learning to Segment Corneal Tissue Interfaces in OCT Images. *arXiv*. 2019. doi:10.48550/arXiv.1810.06612

36. DeepRetina. Layer Segmentation of Retina in OCT Images Using Deep Learning - PubMed. Available from: https://pubmed.ncbi.nlm.nih.gov/33329940/. Accessed January 18, 2025.

37. Curti N, Giampieri E, Guaraldi F, et al. A Fully Automated Pipeline for a Robust Conjunctival Hyperemia Estimation. *Appl Sci*. 2021;11(7):2978. doi:10.3390/app11072978

38. Huntjens B, Basi M, Nagra M. Evaluating a new objective grading software for conjunctival hyperaemia. *Contact Lens Anterior Eye*. 2020;43 (2):137–143. doi:10.1016/j.clae.2019.07.003

39. Park IK, Chun YS, Kim KG, Yang HK, Hwang JM. New Clinical Grading Scales and Objective Measurement for Conjunctival Injection. *Invest Ophthalmol Visual Sci*. 2013;54(8):5249–5257. doi:10.1167/iovs.12-10678

40. Sánchez Brea ML, Barreira Rodríguez N, Sánchez Maroño N, Mosquera González A, García-Resúa C, Giráldez Fernández MJ. On the development of conjunctival hyperemia computer-assisted diagnosis tools: influence of feature selection and class imbalance in automatic gradings. *Artif Intell Med*. 2016;71:30–42. doi:10.1016/j.artmed.2016.06.004