



Automated classification of dry eye type analyzing interference fringe color images of tear film using machine learning techniques

Katsumi Yabusaki¹, Reiko Arita^{2,3}, Takanori Yamauchi⁴

¹Tokyo New Drug Research Laboratories, Kowa Company Ltd., Tokyo, Japan; ²Itoh Clinic, Saitama, Japan; ³Lid and Meibomian Gland Working Group, Tokyo, Japan; ⁴Chofu Factory, Kowa Company Ltd., Tokyo, Japan

Keywords: artificial intelligence, diagnostic support, dry eye disease, machine learning, type classification

1. Background and purpose

The unstable balance in secretions of lipids and aqueous fluid to tear film is a significant cause of dry eye disease (DED).^{1,2} Arita et al. demonstrated a simple but very effective method that classifies dry eye types to the aqueous deficient dry eye (ADDE) and the evaporative dry eye (EDE) by focusing on the dry eye type-unique appearances of interference fringe colors and patterns of tear films.³ We thought this simple classification is very helpful for diagnoses and treatments. However, diagnostic bias by unskilled observers remains an issue to be solved.³

The artificial intelligence (AI)-based support for diagnosis is one of the hottest topics in the field of ophthalmology research. We expected that the AI-based model would reduce bias in DED-type diagnoses. Many studies have been reported targeting retinal diseases like age-related macular degeneration and/or diabetic retinopathy. Most of the works established AI-based predicting models using images taken by fundus cameras and/or optical coherence tomography (OCT)

Correspondence: Katsumi Yabusaki, Kowa Company Ltd., 2-17-43 Noguchicho, Higashimurayama, Tokyo 189-0022, Japan.
E-mail: yabusaki@kowa.co.jp

devices to capture disease-related structural disorders.⁴⁻⁷ In contrast, the interference fringes dynamically change the colors and patterns spatiotemporally. To the best of our knowledge, there is no AI-based model studied for distinguishing ADDE and EDE using interference fringe images. However, an AI-based study classifying the condition of the tear lipid layer by analyzing the textures of interference fringes compared to the device-unique grades has been reported.⁸ This suggested the possibility of using the unstructured characteristics, such as colors and/or complexities of interference fringes, as the numerical image features when building AI-based prediction models. In this study, we first examined several types of image characteristics extracted from the colors and patterns of fringes to obtain effective image features for the DED-type classification. We then evaluated whether the AI-based models would have sufficient abilities for this type of prediction by comparing their diagnoses with those made by an ophthalmologist skilled in this classification (the founder of this type classification).³

2. Methods

The interference fringe of tear film is generated by the phase difference between reflected lights from the surface of the lipid layer and the border formed between lipid and water layers (Fig. 1). It seems the colors and appearances of interference fringes are primarily determined only by the lipid conditions; however, we consider that the balance of secreted amounts of lipid and water component in tear film is key to keep the appearance of the interference fringe at normal type. Therefore, the interference fringe would show a colorful and complex appearance (ADDE type) when the balance is off due to conditions such as excess secretion of the lipid component and/or shortage of the water component, since the excess amount of lipid might be pooled at the lower side of the corneal surface due to the shortage in upward movement of the tear film. The pooled lipids would induce an uneven thickness of lipid layer, and this could cause the colorful and complex appearance of the interference fringe.

Interference fringe images were taken with the Kowa DR-1 α tear interferometer (Kowa Company Ltd., Tokyo, Japan) at 5 seconds after blinking (Fig. 1A and B). We used these fringe images to build the linear support vector machine (SVM)-based machine learning (ML) models. We converted the unstructured properties (characteristics), such as fringe colors, color saturations, color diversities, and/or textures of fringes, to numerical values through several image processing techniques. These numerical values were normalized to range from 0 to 1 to define the image features used for this study. Examining image characteristics in a multifaceted manner, we finally chose a minimum necessary of 11 image features to build the models. A detailed definition of each image feature is shown in Table 1 and examples explaining how to extract the numerical values from interference fringe images are shown in

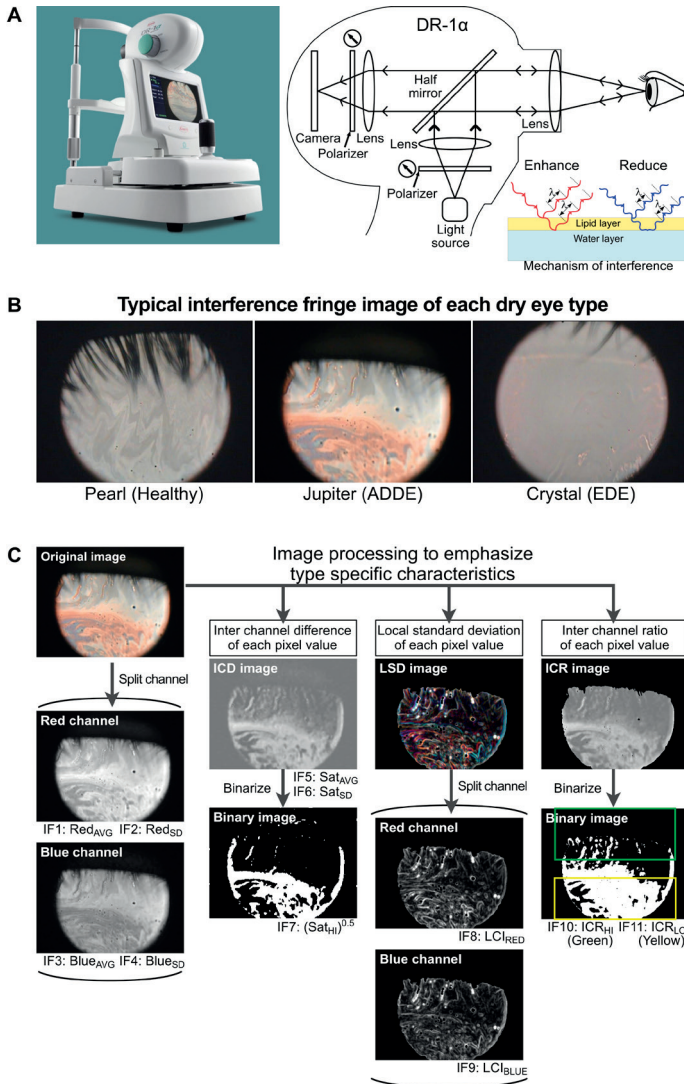


Fig. 1A. Typical interference fringe color images captured by the Kowa DR-1 α . (A) Interference fringes are generated on the surface of tear film under special illumination of the device. The colors of the fringes represent the thickness of the lipid layer, which covers the water layer of tear film. (B) Interference fringe images show dry-eye-specific colors and appearances. (C) Eleven image features (IFs) optimized to classify three dry-eye types were prepared by using several image processing techniques. IFs can be divided to three categories by the functions: 1. IFs represent color properties of IFCl (IF1 ~ IF6); 2. IFs show the existences of breakup of tear film (IF8 and IF9); and 3. IFs show the size and location of interference fringes (IF7, IF10, and IF11).

Table 1. Definitions of image features

ID	Name	Definitions of image features
IF1	RedAVG	Averaged pixel value of red channel of target area
IF2	RedSD	Standard deviation of red channel of target area
IF3	BlueAVG	Averaged pixel value of blue channel of target area
IF4	BlueSD	Standard deviation of blue channel of target area
IF5	SatAVG	Averaged pixel value of saturation of target area
IF6	SatSD	Standard deviation of saturation of target area
IF7	(SatHI)0.5	Square root of pixel number where saturation shows high level
IF8	LCIRED	Local complexity of pixel value of red channel of SD image
IF9	LCIBLUE	Local complexity of pixel value of blue channel of SD image
IF10	ICRHI	Averaged pixel value of upper area of inter-channel ratio (R/B) image
IF11	ICRLO	Averaged pixel value of lower area of inter-channel ratio (R/B) image

Figure 1C and Figure 2A. Image features were averages and/or standard deviations and/or region of interest (ROI) areas calculated from pixel values of color channels, and can be largely sorted into three classes:

1. class-1 represents color properties of interference fringes obtained as averaged pixel values of color channels (including IF1, IF3, and IF5);
2. class-2 indicates complexities of fringe appearance and/or colors determined by the variations of pixel colors for complexities of whole cornea areas or local cornea area (including IF2, IF4, IF6, IF8, and IF9); and
3. class-3 shows localizing areas of fringes and breakups of tear film in the corneal surface (including IF7, IF10, and IF11).

We collected 138 images (46 images were contained in each type) to build ML models. The models were trained using training images (including 31 randomly chosen images from each type, for a total of 93 images) and type predictions were examined using 45 residual images containing 15 images for each type, with averaged values and standard deviations calculated from 10-time-running using randomly chosen data from the data set. DED types predicted by the ML models were compared to ones diagnosed by a skilled ophthalmologist to obtain F-scores and Kappa coefficients. We also ranked image features by the significance in efficacies of type classification from the F-scores under independent use.

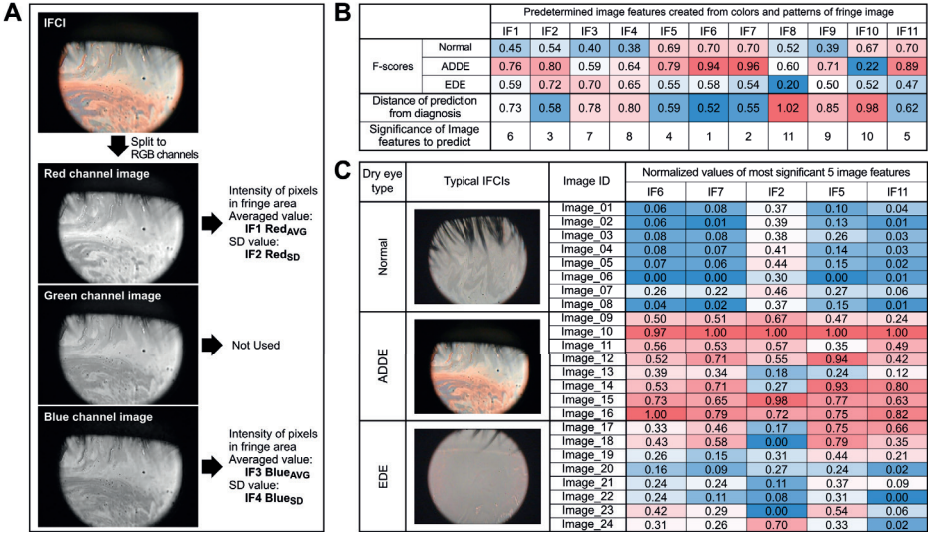


Fig 2. Definition and ability in independent use of each image feature. (A) An example of image processing to extract image features (IF1 – IF4) from the split channels of interference fringe image. Averaged pixel values (IF1: Red_{AVG} and IF3: $Blue_{AVG}$) and standard deviation values (IF2: Red_{SD} and IF4: $Blue_{SD}$) were calculated by only targeting the areas of interference fringe limited by thresholding values applied to pixel brightness. (B) Ability of each image feature to classify DED types under individual use. The abilities were indicated as F-scores and colored as heat map (blue for lower and red for higher values). Image features were ranked by the integrated ability of DED-type prediction (defined by Equation 1 as shown in Results section). (C) DED-type dependent trends were visualized by image feature values shown as heat map.

3. Results

First, we examined the ability of every single image feature to predict DED type. As shown in Figure 2B, the image feature defined as $Blue_{SD}$ (IF4) had the lowest contribution for normal-type prediction, while ICR_{LO} (IF11) had the highest contribution. Likewise, the lowest and highest image features for ADDE-type prediction were ICR_{HI} (IF10) and $(Sat_{HI})^{0.5}$ (IF7), respectively, while LCl_{RED} (IF8) and Red_{SD} (IF2) had the lowest and highest features for EDE-type prediction, respectively. Distance of prediction from diagnosis (DP), indicating combined error in the predicting performance of each image feature (lower value had lower error), was given by each F-score value (FS_{NORMAL} , FS_{ADDE} , and FS_{EDE}) as below:

$$DP = \frac{(1 - FS_{NORMAL}) + (1 - FS_{ADDE}) + (1 - FS_{EDE})}{3} \quad (1)$$

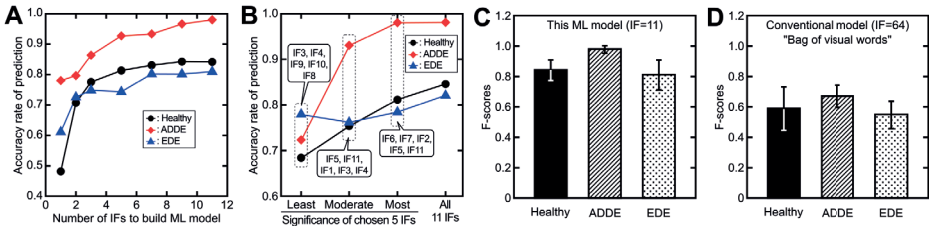


Fig 3. Accuracy rates of our ML model to predict DED type. (A) Dependencies of accuracy rate of DED-type prediction on the increase in applied number of image features. Image features were added to the queue with an order of [IF1, IF3, IF5, IF7, IF8, IF10, IF2, IF4, IF6, IF9, IF11]. (B) Different efficacies observed under the different sets of the five selected image features chosen by the rank in the ability of prediction (determined in Fig 2B). (C) Accuracy rates of DED-type prediction evaluated by our ML model using 11 image features. (D) Accuracy rates of DED-type prediction evaluated by orthodox ML model (bag of visual words) using 64 automatically determined image features.

DP values showed that Sat_{SD} (IF6) was the most significant image feature for type-predicting ability, while LCI_{RED} (IF8) was the least significant one. We found DED-type-specific trends in image feature values when using the five most significant image features determined by the DP values (IF6, IF7, IF2, IF5, and IF11 ranked under independent use, Fig. 2C). Normal type showed very low image feature values in almost all interference fringe images except IF2. ADDE type very clearly indicated very high values in all image features and EDE type showed moderate image feature values except for IF2, which showed very low values. In general, normal-type images tended to show clear fringe patterns with very lower color saturations; ADDE-type images showed complicated fringes containing many colors with higher saturations. In contrast, EDE-type images had a uniform and slightly colored surface with no fringes just after blinking. In many cases of ADDE and EDE patients, fringe images were accompanied by breakups of tear film. In the areas where tear film broke up, local roughness increased as a complicated texture (IF8 and IF9). These trends seemed to be the strong driving forces that would bring accurate type prediction by our ML models.

Next, we determined the minimum number of image features necessary to create accurate ML models (Fig. 3A). All three types (normal, ADDE, and EDE) showed an increase in accuracy rates depending on the number of image features added to the models; the accuracy rate curves were followed by plateaus at nine image features. In addition, predictions using all 11 image features showed higher accuracy rates than ones assessed by the most significant 5 image features (Fig. 3B). This indicated that 11 was the sufficient minimum necessary number of image features for this case.

Our ML model built using 11 image features showed effectively high F-scores (normal, 0.845 ± 0.067 ; ADDE, 0.981 ± 0.023 ; EDE, 0.815 ± 0.095) and high inter-rater agreement value (kappa coefficient = 0.820) with the predetermined DED-types

diagnosed by the skilled ophthalmologist, whereas the orthodox method using “bag of visual words” technique (we used 64 automatically determined image features by the algorithm to predict the types) showed very low F-scores (normal, 0.587 ± 0.146 ; ADDE, 0.669 ± 0.073 ; EDE, 0.549 ± 0.088) and agreement (kappa coefficient = 0.367). These results strongly indicate that the ML model built with a small but highly capable number of image features generated through minute investigation is able to bring effective diagnostic support, even though the image features have been created from non-structural image characteristics.

4. Conclusions and future perspectives

In this study, we originally developed a ML-based predictive model for DED diagnosis. In a previous study, it was reported that inter-rater agreement showed wide variation between observers in a range from 0.57 to 0.94.3 Even though we were not able to examine the same data set used in the report, our ML model demonstrated almost perfect inter-rater agreement with those diagnosed by the skilled ophthalmologist. The results indicate that our predictive model would help reduce diagnostic biases by providing observers with additional accurate diagnostic support, which is one of main purposes of this work. Nevertheless, accuracy rates for normal- and EDE-type predictions are lower than the accuracy rate for ADDE. We believe this is caused by the lower distinction between normal and EDE types, since they showed similar appearances in fringe colors. Our next step will be to further examine other image features that may enhance model performance in distinguishing between normal and EDE types, hence confirming the ability of this model through clinical evaluations using images collected from many more patients.

References

1. Craig JP, Nichols KK, Akpek EK, et al. TFOS DEWS II Definition and Classification Report. *Ocul Surf.* 2017;15: 276-283.
2. Arita R, Morishige N, Koh S, et al. Increased tear fluid production as a compensatory response to Meibomian gland loss: a multicenter cross-sectional study. *Ophthalmology.* 2015;122: 925-933.
3. Arita R, Morishige N, Fujii T, et al. Tear interferometric patterns reflect clinical tear dynamics in dry eye patients. *IOVS.* 2016;57: 3928-3934.
4. Liu YY, Chen M, Ishikawa H, Wollstein G, Schuman JS, Rehg JM. Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Med Image Anal.* 2011;15: 748-759.
5. Schmidt-Erfurth U, Klmscha S, Waldstein SM, Bogunović H. A view of the current and future role of optical coherence tomography in the management of age-related macular degeneration. *Eye.* 2017;31: 26-44.
6. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, et al. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the “i-ROP” system and image features associated with expert. *TVST.* 2015;4: Article 5.

7. Jaya T, Dheeba J, Singh NA. Detection of hard exudates in colour fundus images using fuzzy support vector machine-based expert system. *J Digit Imaging*. 2015;28: 761-768.
8. Remeseiro B, Bolon-Canedo V, Peteiro-Barral D, et al. A methodology for improving tear film lipid layer classification. *IEEE J Biomed Health Inform*. 2014;18: 1485-1493.