Unsupervised Learning Based on Meibography Enables Subtyping of Dry Eye Disease and Reveals Ocular Surface Features

Siyan Li,[1,2] Yiyi Wang,[1] Chunyu Yu,[3] Qiyuan Li,[1] Pingjun Chang,[1,2] Dandan Wang,[1,2] Zhangliang Li,[1,2] Yinying Zhao,[1,2] Hongfang Zhang,[1,2] Ning Tang,[2] Weichen Guan,[1] Yana Fu,[1,2] and Yun-e Zhao[1,2]

[1]National Clinical Research Center for Ocular Diseases, Eye Hospital, Wenzhou Medical University, Wenzhou, China
[2]Eye Hospital of Wenzhou Medical University at Hangzhou, Hangzhou, China
[3]Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou, China

Correspondence: Yana Fu, Eye Hospital of Wenzhou Medical University at Hangzhou, 618 East Fengqi Road, Hangzhou, Zhejiang 310000, China;
fuyana@eye.ac.cn.
Yun-e Zhao, Eye Hospital of Wenzhou Medical University at Hangzhou, 618 East Fengqi Road, Hangzhou, Zhejiang 310000, China;
zye@eye.ac.cn.

SL, YW and CY contributed equally to this work and should be considered as first authors.

PURPOSE. This study aimed to establish an image-based classification that can reveal the clinical characteristics of patients with dry eye using unsupervised learning methods.

METHODS. In this study, we analyzed 82,236 meibography images from 20,559 subjects. Using the SimCLR neural network, the images were categorized. Data for each patient were averaged and subjected to mini-batch k-means clustering, and validated through consensus clustering. Statistical metrics determined optimal category numbers. Using a UNet model, images were segmented to identify meibomian gland (MG) areas. Clinical features were assessed, including tear breakup time (BUT), tear meniscus height (TMH), and gland atrophy. A thorough ocular surface evaluation was conducted on 280 cooperative patients.

RESULTS. SimCLR neural network achieved clustering patients with dry eye into six image-based subtypes. Patients in different subtypes harbored significantly different noninvasive BUT, significantly correlated with TMH. Subtypes 1 and 5 had the most severe MG atrophy. Subtype 2 had the highest corneal fluorescent staining (CFS). Subtype 4 had the lowest TMH, whereas subtype 5 had the highest. Subtypes 3 and 6 had the largest MG areas, and the upper MG areas of a person's bilateral eyes were highly correlated. Image-based subtypes are related to meibum quality, CFS, and morphological characteristics of MG.

CONCLUSIONS. In this study, we developed an unsupervised neural network model to cluster patients with dry eye into image-based subtypes using meibography images. We annotated these subtypes with functional and morphological clinical characteristics.

Keywords: dry eye disease, meibomian gland dysfunction, unsupervised learning, image clustering

Dry eye disease (DED) is an ocular surface disorder characterized by disruption of tear film stability and affects millions of people worldwide.[1] The latest international consensus of dry eye experts classifies DED as aqueous deficiency dry eye (ADDE) due to insufficient tear secretion and excessive evaporation due to abnormal physiology of the meibomian glands (MGs).[2] The MG is a key factor that should be paid attention to in every patient with dry eye. Meibomian gland dysfunction (MGD) is the most common cause of DED.[3] According to the current estimation of DED, patients with mixed types of DED account for over 25% of all.[4] These individuals also exhibit higher expenditures in comparison to those afflicted solely with MGD or ADDE.[5] However, the current step-wise treatment recommendations for DED faces a challenge in predicting relative benefits of specific management options due to the lack of a better differentiation of patients by DED subtype.[6]

The role of MG's structural changes is becoming more and more important in the diagnosis and treatment of DED. Because meibography enables the morphology of MGs to be visualized, several studies have demonstrated methods for obtaining image parameters and grading MGD.[7–9] The area of dropout (atrophy) of the MGs was the first recognized image-based measure of MGD and is considered one of its most critical and closely related parameters. The meiboscore[10] and Meiboscale[11] are assigned by clinicians through subjective quantification of the area of MG loss. Besides, the Dry Eye Assessment and Management (DREAM) Study[12] has addressed a list of MG morphological parameters including "distorted," "tortuous," "hooked," "abnormal gap," "overlapping," "fluffy areas," "tadpoling," "thinned,"

"thickened," "ghost," "no extension to lid margin," "shortened," and "dropout (atrophy)." Many studies attempted to interpret clinical correlation information from captured MG morphological parameters with the intention of further improving the meibography evaluation system.[13,14] Nevertheless, the correlation of a single parameter with the clinical examination is not comprehensive, and the research results are not very accessible for clinical application situations. Therefore, there is currently no recognized grading standard that can include the multiple morphological parameters mentioned above.

Automated algorithmic systems to detect MG morphological abnormalities based on segmentations of meibography images are currently an area of interest for many researchers.[15–17] Although segmentation-based machine learning achieves a fast production of the manual labeling parameters, there was no attempt to explore the association of information from morphology images with clinical features of dry eye. In this regard, exploring a meibography grading system that can extract integrated image information as well as characterize the clinical functional differences may provide a direct method in the clinical decision making.

The unsupervised learning method is independent of data annotation and can be used to cluster image datasets into different classes,[18] which can be effective to discover new subtypes of clinically relevant features in computed tomography scans and pathological images.[19–21] Previously, one study has applied unsupervised learning approach for the automated assessment of the severity of MG atrophy in meibography images, and demonstrated a high accuracy of 80.9% for the meiboscore grading.[22] However, they did not directly relate image information to assessments of patients with dry eye, including tear secretion function, ocular surface clinical features, and self-conscious symptoms. In the present study, we attempted to establish an unsupervised neural network model based on meibography images to cluster patients with dry eye into image-based subtypes. The main aim was to annotate the different subtypes of patients with dry eye with functional and morphological clinical characteristics based on the meibography. The image-based subtypes bring a new direction to the classification of DED.

PATIENTS AND METHODS

Study Design and Patients

This cross-sectional, observational study included 20,559 subjects with subjective symptoms (burning, stinging, itching, tearing, gritty feeling, foreign-body sensation, and fluctuating visual acuity) and evidence of tear film instability (fluorescein tear breakup time [FBUT] $\leq$ 5 seconds or noninvasive tear breakup time [NIBUT] < 10 seconds or the Schirmer I test (without anesthesia) $\leq$ 5 mm/5 minutes). All subjects were investigated at the Dry Eye Center of the Eye Hospital of Wenzhou Medical University, Hangzhou, for dry eye management from January 2, 2021, to March 31, 2022. The present study was approved by the Ethics Committee of the Eye Hospital of Wenzhou Medical University (2022-020-K-17) and complied with the Declaration of Helsinki. MG images obtained from 20,279 subjects were included in our study; an ethics committee waived the requirement of informed consent for the inclusion of these data. Additionally, 280 subjects provided written informed consent to undergo detailed evaluation of the ocular surface for the study.

Clinical Assessments and Meibography

Infrared images of the MGs in 20,559 patients' bilateral upper and lower eyelids and measurements of tear meniscus height (TMH) and NIBUT were obtained using a Keratograph 5M (Oculus, Wetzlar, Germany). An infrared light that does not cause tearing was focused on the center of each patient's lower lids; the patients were asked to gently blink once, after which the infrared light was quickly replaced with white light to complete the shot. The TMH was measured three times per eye by one experienced examiner using the Keratograph 5M measurement tool, and the average value was recorded. The Keratograph 5M recorded a video of the tear film changes while the patient's eyes were continuously open, and the first NIBUT and average NIBUT were automatically calculated.

For the 280 patients who were able to cooperate in the detailed evaluation, the Ocular Surface Disease Index (OSDI) questionnaire was administered at the first visit. The purpose of the questionnaire was explained by a physician (author H. Z.) specializing in DED. Additionally, regular staff members were present throughout the process and helped to explain each question to ensure that the patients' answers to each question were based on a full understanding. The OSDI contains 12 items, and each answer is scored on a 5-point scale, representing the frequency of the situation. Patients were classified into four categories according to the OSDI score ranges established by valid clinical trials: normal (scores 0–12), mild dry eye (scores 13–22), moderate dry eye (scores 23–32), and severe dry eye (scores 33–100).[23]

Lid margin findings were evaluated for the upper and lower eyelids with the use of a slit-lamp microscope. Corneal fluorescent staining (CFS) was scored from 0 to 12.[24] The FBUT was measured three times consecutively after the instillation of fluorescein, and the mean value was used. For each patient, we pressed the central eight MGs in the central area of the lower eyelid and averaged the meibum quality scores of the opened glands[25] (0 = clear fluid; 1 = cloudy fluid; 2 = cloudy fluid containing particulate matter; and 3 = opaque, toothpaste-like meibum); we then took the average score of the left and right eyes. The expressibility of 5 MGs in the central area of the lower eyelid was tested by applying firm pressure and scoring the presence or absence of meibum secretion on a scale from 0 to 3: 0 = all glands expressible; 1 = 3 to 4 glands expressible; 2 = 1 to 2 glands expressible; and 3 = no glands expressible.[26] Meiboscore and MGs' morphological characteristics were scored by two DED experts (authors Y. F. and S. L.) using the previously defined methods.[27,28] Each patient's four eyelids were scored separately, and the scores of the two experts were averaged to obtain the final meiboscore. Each patient's four eyelids were assessed for main MGs' morphological characteristics[12,29] separately including dropout/atrophy, thinned, thickened, tortuous, overlapping, uneven, and normomorph, with a score of 1 given for each characteristic. The total score for each person is obtained by summing the scores of four eyelids, and the morphological score for each subtype is determined by averaging the scores of specific abnormalities.

Creating Meibography Image Clusters Using Unsupervised Clustering Models

We used Lightly (https://github.com/lightly-ai/lightly), a self-supervised learning framework, to train a SimCLR network with 82,236 images from 20,559 patients. The

model was trained for a total of 500 epochs using the Stochastic Gradient Descent (SGD) optimizer, configured with an initial learning rate of 0.1, a momentum parameter of 0.9, and a weight decay parameter of 5e-4. Our SimCLR implementation uses a projection head with an input dimension of 512, corresponding to the feature dimension of the ResNet-34 backbone, a hidden dimension of 512, and an output dimension of 160. The loss function utilized for training was the Normalized Temperature-Scaled Cross-Entropy Loss (NTXentLoss), and a CosineAnnealingLR scheduler was applied for learning rate decay. The training batch size was set to 256. We used the ResNet-34 architecture as the backbone for feature extraction. Data augmentation techniques applied during training included vertical flip (with a probability of 0.5), random rotation (with a probability of 0.5), horizontal flip (with a probability of 0.5), and conversion of all images to grayscale. The model was trained on a rack-mounted server equipped with an NVIDIA RTX 3090 graphics processing unit (GPU). The network could embed any image into a vector reflecting its features. We encoded four images per patient into four vectors using the SimCLR network and used their mean vector as the final embedding vector. We clustered these vectors with mini-batch k-means algorithm and consensus clustering. We calculated Bayesian information criterion (BIC) and Davies–Bouldin score (DB) to determine the number of subtypes. We removed subtypes with fewer than 50 patients and classified patients into 6 image-based subtypes, and we visualized the embedding vectors using principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), linear discriminant analysis (LDA), and uniform manifold approximation and projection (UMAP) in Supplementary Figure S1.

MG Dropout of the Identified Meibography Image-Based Subtypes

We built two semantic segmentation models to segment MGs and eyelids in meibography images. We used Keras framework to implement Attention U-net model by Oktay et al.[30] For MG segmentation, we used images and masks from the public MGD-1K dataset by Saha et al.[31]

We defined the upper tarsal area vertically from the lid margin line to the tarsal fold edge, and the lower tarsal area horizontally from the lacrimal punctum to the lateral canthus. Y. Wang annotated the images and S. Li verified the images. Four hundred fifty-six images were used to build the E model. We trained another Attention U-net model to segment tarsal plate in meibography images. The 2 models were both trained for a total of 100 epochs using the Nadam optimizer with a learning rate of 1e-4. The input dimension for the model was set to 256 × 256 pixels with 3 color channels. For the loss function, we utilized a custom loss function based on the Dice coefficient, specifically the dice loss, which measures the overlap between the true and predicted segmentation masks. The batch size used during training was set to eight. Additionally, two callbacks were used during training: the ReduceLROnPlateau callback, which reduces the learning rate by a factor of 0.1 whenever the validation loss does not improve for 4 consecutive epochs, and the EarlyStopping callback, which terminates the training process if the validation loss does not improve for 10 consecutive epochs, while restoring the best weights observed during training.

We calculated the MG and the tarsal area in the upper and lower eyelids in different subtypes from segmentations; pixels with >50% probability of being foreground in each model were the MG or tarsal areas. To compare the MG areas between patients and reduce the bias to the tarsal area, we calculated the normalized MG area in the upper or lower eyelid, which is the ratio of the MG area to the corresponding tarsal area; the tarsal area was the sum of the elementwise product of the segmented tarsal plate; and the MG area was the sum of the segmented MGs. We used these models to calculate the normalized MG area for all 82,236 images.

Statistical Analysis

All quantitative variables were analyses as continuous variables in this study. We used 1-way ANOVA to compare clinical features in different subtypes; the Shapiro-Wilk test was used for data normality. For non-normal features, we used the Kruskal-Wallis test. We measured the correlation of clinical features with Spearman's rank correlation coefficient and tested significance with permutation test. We did all analyses with the SciPy package. Missing data were ignored in between-group comparisons for each variable.

RESULTS

The present study included 20,559 patients with DED; the mean age of the patients was 34.0 ± 21.9 years. Figure 1 shows a flow chart of this study. In particular, after constructing a dataset of the patients' meibography images, we built a SimCLR network to embed these images into image vectors, and four vectors from the same patient were averaged to obtain a unique vector for this patient. We implemented a consensus clustering strategy to cluster these patient vectors into 2 to 16 clusters. Figure 2A shows that clustering patients into 8 clusters achieved the best performance. The clustering results are visualized in Figure 2B, which shows 6 major clusters and 2 clusters harboring fewer than 50 patients each. These 2 clusters were removed for downstream analysis, and patients were ultimately classified into 6 image-based subtypes, which contained at least 1882 patients each. For each subtype, we visualized 12 images of 3 patients, and several subtypes showed distinct image features, as shown in Figure 2C. Morphological characteristics of meibomian glands in 280 patients are summarized in the Table, which show that subtype 1 and subtype 5 has the largest atrophy area of the MGs with uneven glands, subtypes 2 and 3 has the most tortuous glands, subtype 3 has the most MGs with normal morphology, subtype 4 has the most overlapping glands, and subtype 6 has the most thickened glands.

Figures 3A and 3B presents that the neural network can efficiently identify the MGs in the images from the MGD-1K dataset, and we visualized the segment results in Supplementary Figure S1. Figure 3C illustrates the differences in the normalized gland area between the different image-based subtypes and shows that subtype 3 and subtype 6 had the largest gland area. Figure 3D compares the correlation of this parameter between the upper and lower eyelids and between the left and right eyes; interestingly, whereas the normalized MG areas of the patients' bilateral upper eyelids were highly correlated, the correlation between the upper and lower eyelids of a single eye was weaker.

A detailed evaluation of the ocular surface, including OSDI, fluorescein BUT, CFS, meibum expressibility, meibum
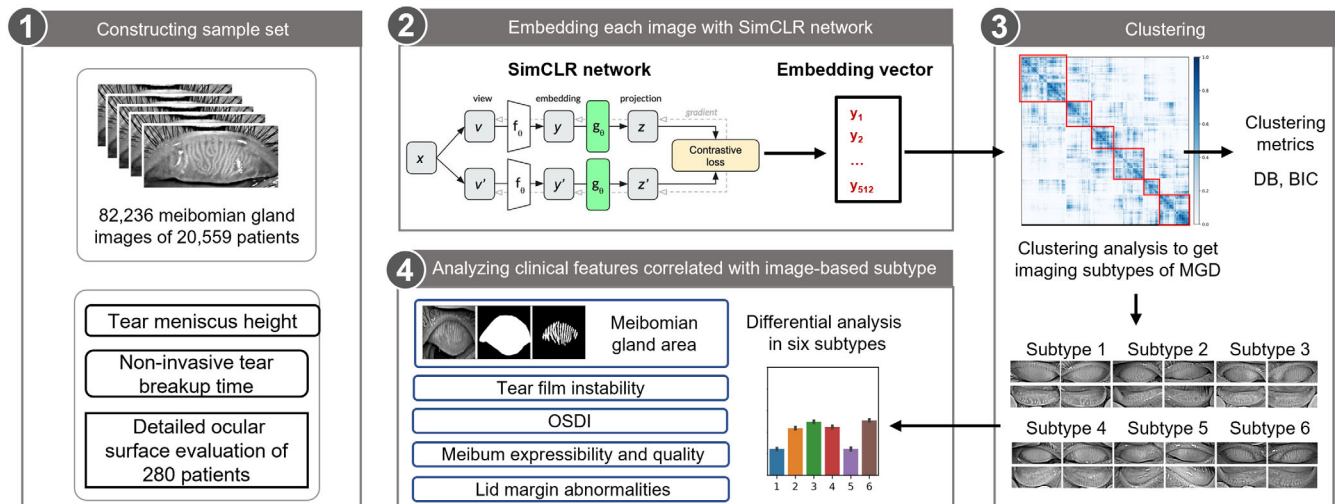
FIGURE 1. Schematic of the analysis pipeline. Meibography images of patients with dry eye and related clinical information were collected from 2021 to 2022. Unsupervised SimCLR network was trained using meibography images, and the images were embedded into vectors using the network. Average vector of four images of one patient were calculated and considered as the image vector of the patient. Then, consensus clustering was applied to group patients into six major image subtypes, and the differences of the clinical features among image subtypes were analyzed.
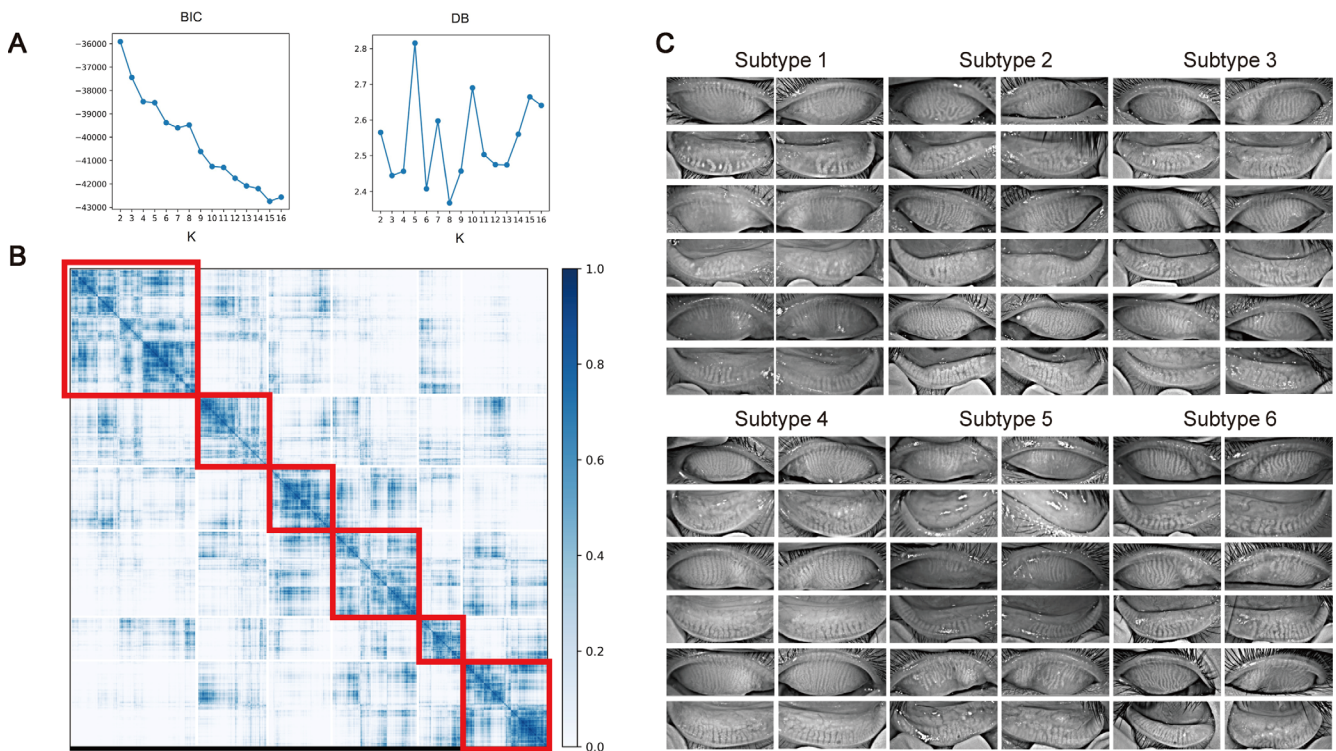
FIGURE 2. Consensus clustering of dry eye patients using meibography images. (**A**) Clustering metrics of different number of clusters (k = 2, 3, …, 16). Davies-Bouldin score (DB) and Bayesian Information Criterion (BIC) of different number of clusters were calculated, and k = 8 achieves lowest DB and less BIC. (**B**) Consensus matrix of k = 8. Six major clusters (> 50 patients) and two minor clusters (< 50 patients) were obtained. (**C**) Examples of clustering results of six major clusters (subtypes). For each cluster, 12 images of 3 patients are shown.

quality, and lid margin abnormalities, was performed in 280 cooperative patients. Missing data were ignored in between-group comparisons for each variable, including 13 patients without OSDI, 22 patients without fluorescein BUT, 19 patients without CFS, and 13 patients without meibum expressibility. We compared these features in different imaging-based subtypes and found a significant difference in fluorescein BUT, CFS, meibum expressibility, meibum quality, and lid margin abnormalities. As witnessed in Figures 4E to 4H and the Table, these results suggest that image-based subtypes are related to the abovementioned clinical features. Figures 4A and 4B showed that patients in different subtypes harbored significantly different tear meniscus heights, for example, patients in subtype 4 had

Investigative Ophthalmology & Visual Science

TABLE. Detailed Evaluation of the Ocular Surface of 6 Image-Based Subtypes in 280 Patients With Dry Eye

| Clinical Features	Subtype 1 ($n = 42$)	Subtype 2 ($n = 56$)	Subtype 3 ($n = 46$)	Subtype 4 ($n = 26$)	Subtype 5 ($n = 39$)	Subtype 6 ($n = 71$)
	Mean (95% CI) or Values N (%)					
CFS (0–12)	1.65 (0.775, 2.525)	2.151 (1.243, 3.059)	1.085 (0.445, 1.726)	0.455 (−0.006, 0.915)	1.875 (0.967, 2.783)	0.53 (0.242, 0.818)
TBUT, s	2.105 (1.709, 2.502)	2.278 (1.961, 2.596)	2.805 (2.374, 3.236)	3.705 (2.87, 4.539)	2.405 (1.839, 2.972)	3.106 (2.709, 3.503)
TMH, mm	0.217 (0.232, 0.247)	0.193 (0.213, 0.233)	0.217 (0.234, 0.251)	0.226 (0.247, 0.268)	0.195 (0.213, 0.231)	0.216 (0.23, 0.243)
Meiboscore, 0–3	1.298 (1.046, 1.549)	0.705 (0.598, 0.812)	0.527 (0.422, 0.632)	0.538 (0.385, 0.692)	1.487 (1.242, 1.732)	0.711 (0.602, 0.821)
Meibum expressibility, 0–3	1.7 (1.452, 1.948)	1.352 (1.164, 1.54)	1.345 (1.079, 1.611)	1.37 (1.154, 1.585)	1.763 (1.502, 2.024)	1.29 (1.134, 1.446)
Meibum quality scores, 0–3	1.164 (0.96, 1.367)	1.302 (1.116, 1.488)	1.013 (0.85, 1.176)	0.797 (0.533, 1.061)	1.549 (1.317, 1.781)	1.261 (1.112, 1.409)
OSDI, 0–100	41.31 (36.093, 46.527)	47.36 (41.952, 52.767)	42.945 (36.903, 48.986)	40.823 (33.372, 48.274)	45.893 (38.946, 52.841)	37.109 (33.131, 41.087)
Lid margin abnormalities						
Congestion	5 (11.9)	6 (10.7)	6 (13.0)	3 (11.5)	10 (25.6)	10 (14.1)
Vascularity	13 (31.0)	13 (23.2)	13 (28.3)	3 (11.5)	21 (53.8)	17 (23.9)
Plugging	29 (69.0)	27 (48.2)	23 (50.0)	11 (42.3)	27 (69.2)	36 (50.7)
Thickening and irregularity	18 (42.9)	10 (17.9)	5 (10.9)	2 (7.7)	24 (61.5)	17 (23.9)
Normal	7 (16.7)	14 (25.0)	12 (26.1)	7 (26.9)	4 (10.3)	9 (12.7)
Max line location forward	13 (31.0)	15 (26.8)	9 (19.6)	4 (15.4)	16 (41.0)	19 (26.8)
Foaming	3 (7.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Scaly	8 (19.0)	3 (5.4)	4 (8.7)	3 (11.5)	6 (15.4)	2 (2.8)
Cuff deposits	11 (26.2)	5 (8.9)	7 (15.2)	4 (15.4)	11 (28.2)	11 (15.5)
Morphological characteristics of MGs						
Normomorph	0.925 (0.553, 1.297)	1.196 (0.96, 1.433)	1.413 (1.103, 1.723)	1.731 (1.326, 2.136)	0.667 (0.375, 0.958)	1.225 (1.005, 1.445)
Dropout/atrophy	2.5 (1.993, 3.007)	1.946 (1.665, 2.228)	1.478 (1.131, 1.825)	1.423 (0.936, 1.91)	3.028 (2.644, 3.411)	1.648 (1.354, 1.942)
Thinned	0.85 (0.564, 1.136)	0.625 (0.372, 0.878)	0.37 (0.159, 0.581)	0.308 (0.033, 0.582)	1.139 (0.717, 1.56)	0.324 (0.165, 0.483)
Thickened	0.025 (−0.026, 0.076)	0.054 (−0.007, 0.114)	0.109 (−0.02, 0.237)	0.077 (−0.082, 0.235)	0.111 (−0.024, 0.246)	0.127 (0.029, 0.224)
Tortuous	0.65 (0.347, 0.953)	1.179 (0.885, 1.472)	1.239 (0.913, 1.566)	1.0 (0.677, 1.323)	0.75 (0.403, 1.097)	1.085 (0.845, 1.324)
Overlapping	0.675 (0.34, 1.01)	0.768 (0.479, 1.057)	1.109 (0.841, 1.376)	0.846 (0.509, 1.183)	0.528 (0.242, 0.814)	0.732 (0.513, 0.951)
Uneven	1.425 (0.934, 1.916)	1.036 (0.693, 1.378)	0.326 (0.148, 0.504)	0.5 (0.117, 0.883)	1.639 (1.095, 2.182)	0.873 (0.633, 1.113)

CI, confidence interval.
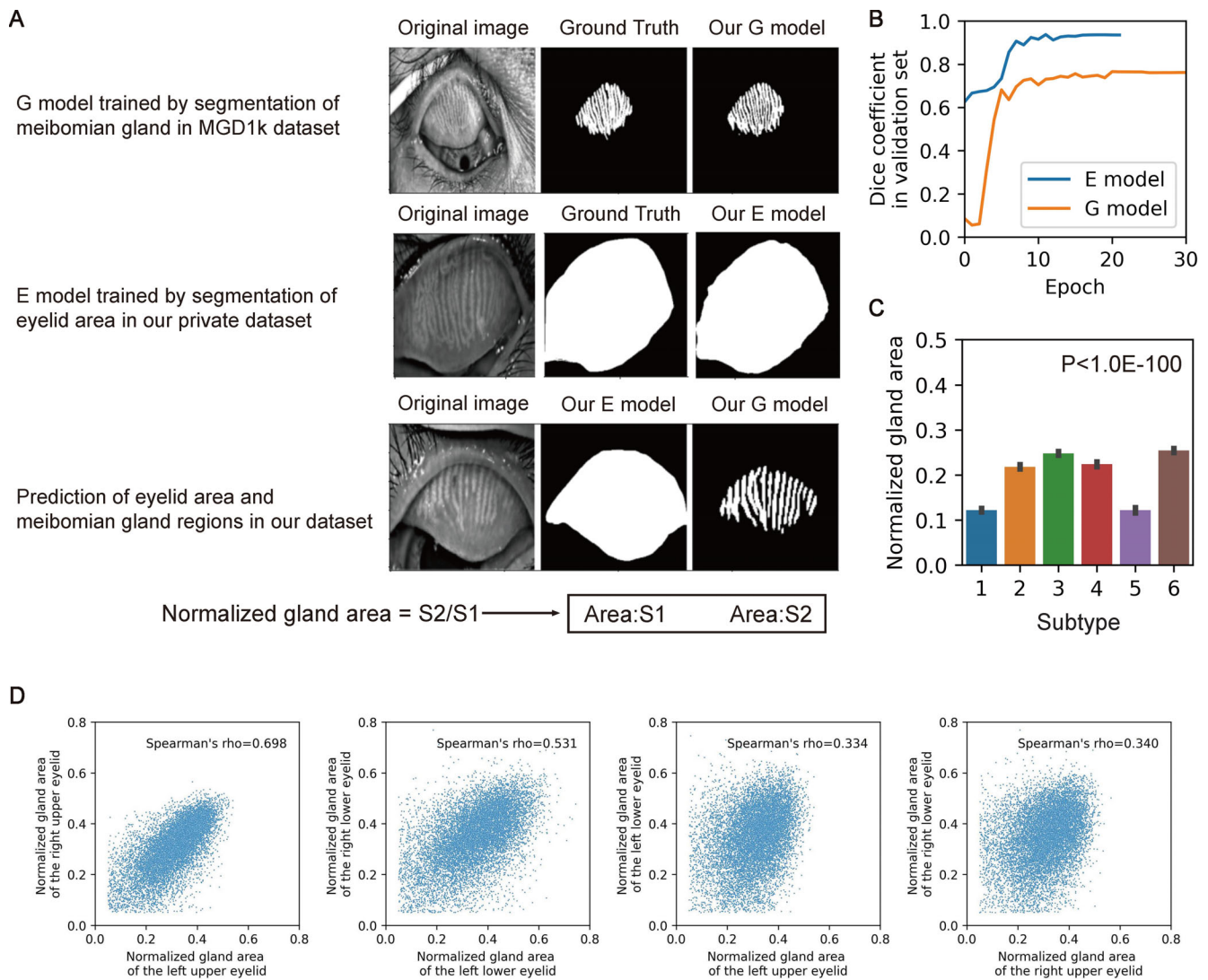
FIGURE 3. Area of meibomian gland regions differ among image subtypes. (**A**) Illustration of deep learning models segment eyelid area (E model) and meibomian gland regions (G model). The ratio of outputs of two models is defined as the normalized gland area of the meibomian gland region. (**B**) Dice coefficient of E model and G model in training processes with an early-stop patience. (**C**) Normalized gland area of 6 subtypes was compared using 1-way ANOVA analysis, and F-test was applied to calculate the *P* values. (**D**) Normalized gland area of different eyelids in patients with dry eye. Spearman's rank correlation coefficient was calculated and shown.

the lowest tear meniscus height, and patients in subtype 5 had the highest tear meniscus height. Data from the left and right eyes showed consistent results. Figure 4C shows that patients in different subtypes also harbored significantly different NIBUT values. As observed in Figure 4D, the NIBUT was significantly correlated with TMH. We analyzed the OSDI of patients in different subtypes to explore the relationship between imaging and subjective symptoms of dry eye in Figure 4I, and the results suggest that image-based subtypes do not differentiate between symptoms.

DISCUSSION

Meibography has become widely used to assess the morphologic changes in the MG for the diagnosis of DED, most previous studies have focused on using artificial intelligence (AI) algorithms for rapid recognition and statistics of already labeled features.[7] Labeling is usually time consuming and

results in an increasing number of parameters,[12] not all of which are necessary. As the DREAM study has introduced a long list of MG morphology metrics, most of them have yet to be fully vetted as DED diagnostic tools.[12,28] The interpretation of the meaning of a single image parameter is not easily applied in actual practice. In addition, the current grading system[32] is dependent on the dropout of the MGs and does not incorporate any of the above image parameters. The staged management algorithm for DED is intricate due to its multifactorial etiology. Therefore, a quickly and directly grading system holds significant relevance in the clinical evaluation and treatment of patients. Unsupervised learning uses a variety of information without being defined that may contain features relevant to clinical examination results. Studies on computed tomography scans and pathological images have provided with new subtypes of clinically relevant features to facilitate clinical decision making, suggesting that image information can provide new insights
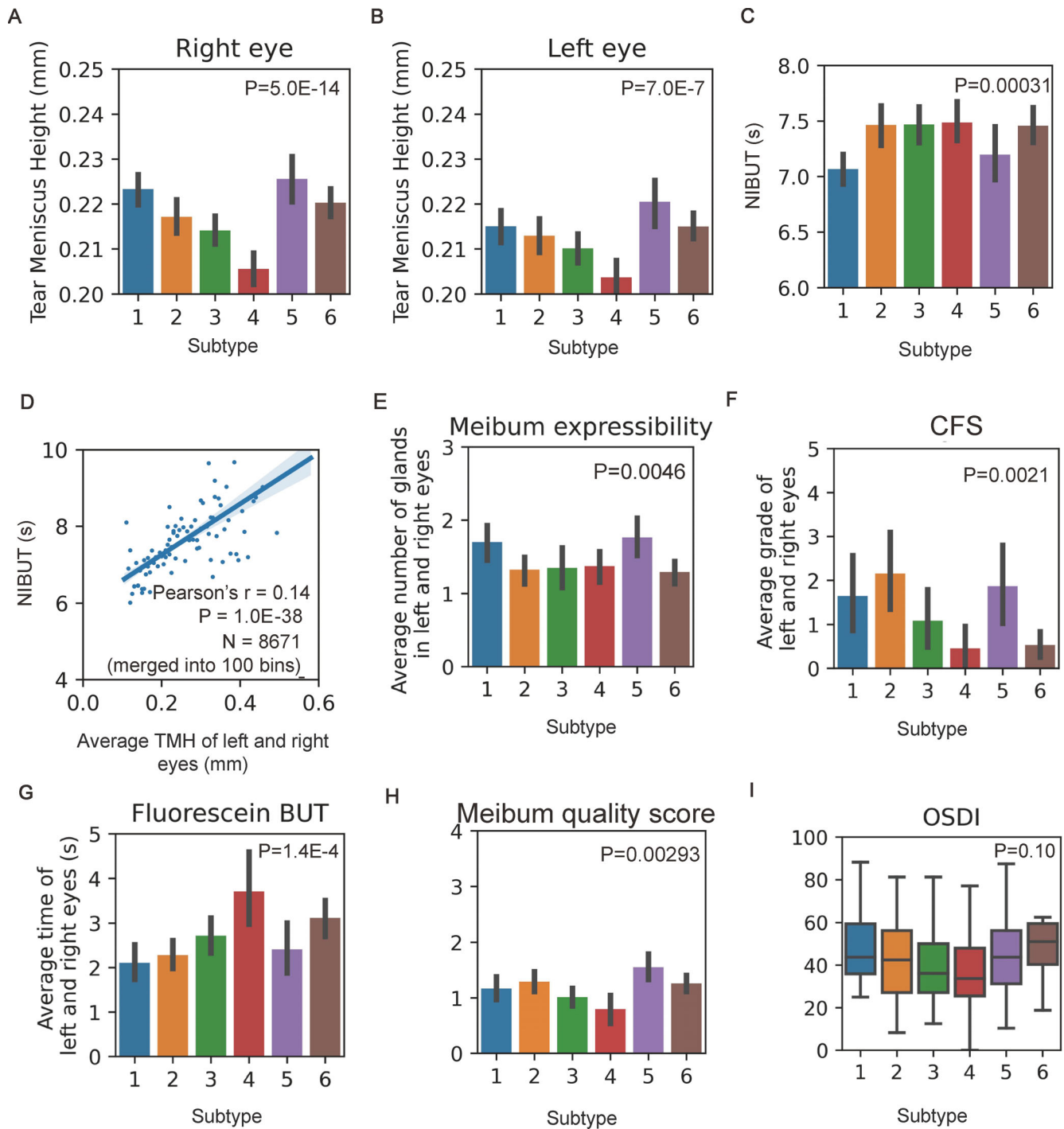
FIGURE 4. Image-based subtypes are corelated with clinical features. (**A, B**) Tear meniscus height of the left and right eyes of 6 subtypes were compared using 1-way ANOVA analysis, and F-test was applied to calculate the *P* values. (**C**) Noninvasive tear breakup time of 6 subtypes was compared using 1-way ANOVA analysis, and F-test was applied to calculate the *P* values. (**D**) Noninvasive tear breakup time is correlated with tear meniscus height. Patients are merged into 100 bins, and the related 2-sided *P* value was shown. Clinical features, including meibum expressibility (**E**), CFS (**F**), fluorescein tear break-up time (**G**), meibum quality score (**H**), and OSDI (**I**) of 6 image-based subtypes was compared using 1-way ANOVA analysis, and F-test was applied to calculate the *P* values.

into disease. However, an unsupervised learning approach was not used to discover new imaging-based subtypes that can characterize the functional differences of DED.

The current study demonstrated the feasibility of using unsupervised learning to discover new imaging-based subtypes of DED. We annotated the six subtypes of patients with different clinical characteristics using tear stability evidence. The different image subtypes were correlated with tear break-up time and tear meniscus height (see Fig. 4). Previous works suggest that tear breakup time is an important examination to assess tear quality for patients with DED and is correlated with tear meniscus height,[33,34] which is

consistent with our results. In addition, we noted that the image-based subtypes could be related to meibum expressibility and cerebrospinal fluid (CFS; see Fig. 4). The International Workshop on MGD indicated that meibum expressibility is useful for the diagnosis and staging of MGD-related dry eye.[32] MGD is divided into two major categories based on meibomian gland secretion: low delivery (containing the hyposecretory and obstructive subtypes) and high delivery (the hypersecretory subtype). Meibum quality and expressibility are used to evaluate MG function. However, some patients have characteristics of both the hypersecretory and obstructive subtypes. Some researchers have suggested that the classification system of MGD should be modified to include patients with various types of characteristics.[13] The results of the present study suggest that our new MGD classification system based on the clusters classified by unsupervised learning may be a good approach to establish the correspondence between image-based features and function.

The latest international consensus on the classification of patients with DED is based on the concept of tear film–oriented diagnosis. There are three basic types of dry eye, defined by their causes: increased evaporation, aqueous deficiency, and decreased wettability.[35] However, it is difficult to strictly distinguish between the aqueous deficiency type of dry eye and the increased evaporation type, and a hybrid form has been proposed.[36] The Asia Dry Eye Society (ADES) has proposed a simple classification of dry eye based on fluorescein breakup patterns.[37] A previous work[38] has demonstrated that the lid margin abnormality score correlates with the meiboscore. One work[39] explored different types of lid margin abnormalities in detail and found that an irregular lid margin was associated with the meiboscore, but vascular engorgement, plugging, and anterior placement of the Marx line did not significantly affect the meiboscore. In this study, six image-based subtypes were correlated with lid margin vascularity, thickness, irregularity, foam, and scaliness. There are no other previous studies correlating meibography parameters with lid margin abnormalities. Some experts have also suggested reconsidering the importance of the frequency of symptoms in classifying DED.[40] Previous studies demonstrated that the OSDI showed good sensitivity and specificity in differentiating between healthy subjects and patients with DED.[41,42] In addition, the OSDI was able to effectively differentiate between mild to moderate and severe DED.[43] In our study, the OSDI scores did not correlate with image-based subtypes (see Fig. 4). This is consistent with previous studies that also failed to find strong correlations between traditional objective clinical measures of dry eye and patient symptoms.[44,45] However, this study also has limitations. We used only 280 detailed ocular surface assessment data points to evaluate this new classification obtained from AI. Another drawback is the exclusive use of the OSDI to assess patient symptoms, without considering other questionnaires like SPEED, potentially influencing the outcomes. These 280 cases of patients were relatively young and their clinical features were unable to cover the population of patients with DED of all ages. Presently, we are also in the process of amassing therapeutic efficacy data for varied subtypes of patients, including the cohort of 280 subjects discussed in this work. Moving forward, we intend to investigate the therapeutic outcomes across distinct subtypes to elucidate a paradigm of precise diagnosis and tailored treatment for DED.

In conclusion, this study describes a SimCLR neural network model for clustering meibography results into six image-based subtypes. These six different subtypes are correlated with BUT, TMH, MG dropout, meibum expressibility, meibum quality, and CFS. However, the integration of AI in DED diagnosis faces challenges related to trust, data security, and usability. Clinicians may harbor skepticism toward AI systems that lack interpretability. Thus, both explainable and secure AI models are crucial for gaining clinical trust and widespread use. Usability is also essential, especially when clinicians seek a simple "readout" for diagnosis.[46,47] Addressing these multifaceted challenges is vital for the successful application of AI in the DED diagnosis and management.

References

1. Stapleton F, Alves M, Bunya VY, et al. TFOS DEWS II epidemiology report. *Ocul Surf*. 2017;15:334–365.
2. Craig JP, Nichols KK, Akpek EK, et al. TFOS DEWS II definition and classification report. *Ocul Surf*. 2017;15:276–283.
3. Bron AJ, de Paiva CS, Chauhan SK, et al. TFOS DEWS II pathophysiology report. *Ocul Surf*. 2017;15:438–510.
4. Bhatt K, Singh S, Singh K, Kumar S, Dwivedi K. Prevalence of dry eye, its categorization (Dry Eye Workshop II), and pathological correlation: a tertiary care study. *Indian J Ophthalmol*. 2023;71:1454–1458.
5. Yang W, Luo Y, Wu S, et al. Estimated annual economic burden of dry eye disease based on a multi-center analysis in China: a retrospective study. *Front Med*. 2021;8:771352.
6. Jones L, Downie LE, Korb D, et al. TFOS DEWS II management and therapy report. *Ocul Surf*. 2017;15:575–628.
7. Swiderska K, Read ML, Blackie CA, Maldonado-Codina C, Morgan PB. Latest developments in meibography: a review. *Ocul Surf*. 2022;25:119–128.
8. Ngo W, Srinivasan S, Jones L. Historical overview of imaging the meibomian glands. *J Optom*. 2013;6:1–8.
9. Wise RJ, Sobel RK, Allen RC. Meibography: a review of techniques and technologies. *Saudi J Ophthalmol*. 2012;26:349–356.
10. Arita R, Itoh K, Inoue K, Amano S. Noncontact infrared meibography to document age-related changes of the meibomian glands in a normal population. *Ophthalmology*. 2008;115:911–915.
11. Pult H, Riede-Pult BH. Non-contact meibography: keep it simple but effective. *Cont Lens Anterior Eye*. 2012;35:77–80.
12. Daniel E, Maguire MG, Pistilli M, et al. Grading and baseline characteristics of meibomian glands in meibography images and their clinical associations in the Dry Eye Assessment and Management (DREAM) study. *Ocul Surf*. 2019;17:491–501.
13. Kim HM, Eom Y, Song JS. The relationship between morphology and function of the meibomian glands. *Eye Contact Lens*. 2018;44:1–5.
14. Xiao J, Adil MY, Olafsson J, et al. Diagnostic test efficacy of meibomian gland morphology and function. *Sci Rep*. 2019;9:17345.

15. Xiao P, Luo Z, Deng Y, Wang G, Yuan J. An automated and multiparametric algorithm for objective analysis of meibography images. *Quant Imaging Med Surg*. 2021;11:1586–1599.

16. Setu MAK, Horstmann J, Schmidt S, Stern ME, Steven P. Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography. *Sci Rep*. 2021;11:7649.

17. Wang J, Li S, Yeh TN, et al. Quantifying meibomian gland morphology using artificial intelligence. *Optom Vis Sci*. 2021;98:1094–1103.

18. Niu C, Shan H, Wang G. SPICE: semantic pseudo-labeling for image clustering. *IEEE Trans Image Process*. 2022;31:7264–7278.

19. Campredon A, Battistella E, Martin C, et al. Using chest CT scan and unsupervised machine learning for predicting and evaluating response to lumacaftor-ivacaftor in people with cystic fibrosis. *Eur Respir J*. 2021;59:2101344.

20. Li C, Hu Z, Chen H, et al. Cervical histopathology image clustering using graph based unsupervised learning. *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)*. New York, NY: Springer; 2020:141–152.

21. Muhammad H, Sigel CS, Campanella G, et al. Unsupervised subtyping of cholangiocarcinoma using a deep clustering convolutional autoencoder. In: Shen D, Liu T, Peters TM, et al. (eds), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Cham, Switzerland: Springer International Publishing; 2019:604–612.

22. Yeh CH, Yu SX, Lin MC. Meibography phenotyping and classification from unsupervised discriminative feature learning. *Transl Vis Sci Technol*. 2021;10:4.

23. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the ocular surface disease index. *Arch Ophthalmol*. 2000;118:615–621.

24. Song X, Zhao P, Wang G, Zhao X. The effects of estrogen and androgen on tear secretion and matrix metalloproteinase-2 expression in lacrimal glands of ovariectomized rats. *Invest Ophthalmol Vis Sci*. 2014;55:745–751.

25. Bron AJ, Benjamin L, Snibson GR. Meibomian gland disease. Classification and grading of lid changes. *Eye (Lond)*. 1991;5(Pt 4):395–411.

26. Pflugfelder SC, Tseng SC, Sanabria O, et al. Evaluation of subjective assessments and objective diagnostic tests for diagnosing tear-film disorders known to cause ocular irritation. *Cornea*. 1998;17:38–56.

27. Arita R, Itoh K, Inoue K, Amano S. Noncontact infrared meibography to document age-related changes of the meibomian glands in a normal population. *Ophthalmology*. 2008;115:911–915.

28. von Ahrentschildt A, Hanenberg L, Robich ML, et al. Morphological characteristics of meibomian glands and their influence on dry eye disease in contact lens wearers. *Ocul Surf*. 2022;24:93–99.

29. Yu X, Fu Y, Lian H, Wang D, Zhang Z, Dai Q. Uneven meibomian gland dropout in patients with meibomian gland dysfunction and demodex infestation. *J Clin Med*. 2022;11:5085.

30. Oktay O, Schlemper J, Folgoc L, et al. Attention U-Net: learning where to look for the pancreas. arXiv Preprint. 2018. https://doi.org/10.48550/arXiv.1804.03999.

31. Saha RK, Chowdhury AMM, Na KS, et al. Automated quantification of meibomian gland dropout in infrared meibography using deep learning. *Ocul Surf*. 2022;26:283–294.

32. Nichols KK, Foulks GN, Bron AJ, et al. The international workshop on meibomian gland dysfunction: executive summary. *Invest Ophthalmol Vis Sci*. 2011;52:1922–1929.

33. Tsubota K. Short tear film breakup time-type dry eye. *Invest Ophthalmol Vis Sci*. 2018;59:Des64–Des70.

34. Wang J, Palakuru JR, Aquavella JV. Correlations among upper and lower tear menisci, noninvasive tear break-up time, and the Schirmer test. *Am J Ophthalmol*. 2008;145:795–800.

35. Tsubota K, Yokoi N, Watanabe H, et al. A new perspective on dry eye classification: proposal by the Asia Dry Eye Society. *Eye Contact Lens*. 2020;46(Suppl 1):S2–S13.

36. Bron AJ, Tiffany JM. The contribution of meibomian disease to dry eye. *Ocul Surf*. 2004;2:149–165.

37. Yokoi N, Georgiev GA, Kato H, et al. Classification of fluorescein breakup patterns: a novel method of differential diagnosis for dry eye. *Am J Ophthalmol*. 2017;180:72–85.

38. Arita R, Itoh K, Maeda S, et al. Proposed diagnostic criteria for obstructive meibomian gland dysfunction. *Ophthalmology*. 2009;116:2058–2063.e2051.

39. Ha M, Kim JS, Hong SY, et al. Relationship between eyelid margin irregularity and meibomian gland dropout. *Ocul Surf*. 2021;19:31–37.

40. Rolando M, Zierhut M, Barabino S. Should we reconsider the classification of patients with dry eye disease? *Ocul Immunol Inflamm*. 2021;29:521–523.

41. Wang MTM, Xue AL, Craig JP. Comparative evaluation of 5 validated symptom questionnaires as screening instruments for dry eye disease. *JAMA Ophthalmol*. 2019;137:228–229.

42. Ozcura F, Aydin S, Helvaci MR. Ocular surface disease index for the diagnosis of dry eye syndrome. *Ocul Immunol Inflamm*. 2007;15:389–393.

43. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the ocular surface disease index. *Arch Ophthalmol*. 2000;118:615–621.

44. Barabino S, Labetoulle M, Rolando M, Messmer EM. Understanding symptoms and quality of life in patients with dry eye syndrome. *Ocul Surf*. 2016;14:365–376.

45. Yazdani M, Chen X, Tashbayev B, et al. Evaluation of the ocular surface disease index questionnaire as a discriminative test for clinical findings in dry eye disease patients. *Curr Eye Res*. 2019;44:941–947.

46. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–2223.

47. Aldhoayan MD, Alghamdi H, Khayat A, Rajendram R. A machine learning model for predicting the risk of readmission in community-acquired pneumonia. *Cureus*. 2022;14:e29791.

Investigative Ophthalmology & Visual Science