IFSCC 2025 full paper (IFSCC2025-1797)

AI-Driven Multi-Ethnic Study on Facial Dermis Ageing: Insights from in vivo 3D LC-OCT Measurements

Raoul MISSODEY[*] [1], Ali ASSI[1], Colombe LOPEZ[2], Remy LECONGE[3], Sylvie TREUILLET[3], Yves LUCAS[3], Jean-Hubert CAUCHARD[1], Samuel RALAMBONDRAINY[1], Franck BONNIER[1]

[1]LVMH Recherche, Saint-Jean de Braye, [2]DAMAE Medical, Paris, [3]Laboratoire Prisme, Université d'Orléans, Orléans, France

1. Introduction

Line-Field Confocal Optical Coherence Tomography (LC-OCT) has emerged as a high-resolution, in vivo imaging modality for skin microstructures, bridging the gap between macroscopic and invasive methods [1,2]. Providing "optical biopsies" with cellular resolution [2], it enhances insights into skin biology and aging [3]. LC-OCT integrates optical coherence tomography and confocal microscopy to generate 3D skin representations with approximately 1 µm resolution [4], proving valuable for dermatological research and clinical practice. The technology's evolution includes miniaturization [5], dermoscopy coupling [6], and AI-driven image analysis for segmentation and quantification [5,6]**.** Like Reflectance Confocal Microscopy (RCM) and Multiphoton Microscopy (MPM) [7], LC-OCT offers potential for identifying aging biomarkers [8]. Confocal imaging techniques, including RCM and MPM, allow visualization of dermal architecture and age-related modifications [9–11]. LC-OCT provides access to the characterization of the superficial dermis [12], as demonstrated by studies investigating the correlation between skin relief and fibrous network characteristics through qualitative morphological characterization of collagen under mechanical folding applied to the skin surface [13]. Objective tools, particularly those leveraging AI, are increasingly valuable for skin aging assessment, addressing limitations in clinical scoring [14]. Deep learning applied to large databases of face photographs has witnessed a major step forward in recent years to achieve more objective models (e.g., PhotoAgeClock [15], FaceAge [16], etc.) for correlating features of aging with the prediction of age in diverse populations [17], yielding high accuracy with errors as low as 2-3 years for the best approaches [18], and video-based methods are emerging [19]. While large datasets can be challenging to acquire for confocal imaging, automation of image analysis for features like epidermal patterns and dermal fiber organization has been reported for RCM [20], and the importance of deep learning and segmentation of MPM images has been highlighted [21]. Regarding LC-OCT, deep learning has been used to classify collagen quality [22], and a model has been reported to predict chronological age [23]. The present study refines deep learning models for age prediction using a database of LC-OCT images from over 300 women across three ethnic groups. It evaluates model performance based on skin type and pigmentation, and explores the potential for inclusive models.

2. Materials and Methods

2.1 Study population

The reported results come from clinical studies in the US, China, and France, all adhering to the Declaration of Helsinki. Approvals were obtained from the Allendale Investigational Review Board (CS231001), the Shanghai Ethics Committee for Clinical Research (SECCR/2022-77-01), and the Comité de protection des personnes sud méditerranée I (IDRCB: 2021-A00101-40). All participants provided written informed consent. The cohort included 100 Caucasian, 109 Asian, and 116 African-American healthy female volunteers (Fitzpatrick skin phototypes I-VI), evenly distributed across five age groups ([20-30], [31-40], [41-50], [51-60], and [61-70]). Volunteers' chronological age, from their identification, was used for predictive model construction.

2.2. Clinical scoring

On each healthy female volunteer, a trained assessor conducted clinical scoring on facial features to assess skin firmness, density, plumpness, elasticity, and the presence of crow's feet wrinkles. The side of the face evaluated (left or right) was determined randomly. The clinical scoring process involved standardized positioning and lighting conditions to ensure consistency and accuracy.

2.3 LC-OCT 3D imaging

3D LC-OCT images were acquired using the DeepLive™ system (DAMAE MEDICAL, France). Detailed characteristics of DeepLive™ and LC-OCT can be found in the literature [24]. Data acquisition was performed as described previously [23].The resulting dataset comprised more than 2700 3D image stacks, corresponding to 3D stacks collected for each facial area and for each volunteer.

2.4 Age prediction models

The methods used to construct prediction models from Multiple Linear Regression Analysis (MLR) on clinical scores and deep learning applied to LC-OCT images have been described previously by Assi et al. [23]. Briefly, for the deep learning model, a 3D ResNet18 network, adapted from Hara et al. [25] and trained in PyTorch as described in the Caucasian women study [23]], was used. The process involved Dermis Region Of Interest (ROI) selection, preparation of data training inputs (2-fold downscaling, image splitting into tiles, and data augmentation), and ordinal classification for age prediction, consistent with the previously described approach [23]. However, models were trained over 30 epochs with a batch size of 64, using binary cross-entropy loss and the AdamW optimizer. A learning rate of $5e^{-4}$ with a MultiStepLR scheduler decreasing by 0.1 at epochs 10 and 20 was used. Additionally, a five-fold cross-validation was implemented to generalize the model, avoiding the splitting of volunteer data between the training and validation sets.

3. Results and discussions

3.1 Ethnicity-Specific Predictive Models

a. Assessment of Facial Aging through Clinical Scoring

Figure 1A shows the correlation between chronological and predicted age for Caucasian women using a Multiple Linear Regression (MLR) model on clinical scores (r = 0.92, MAE = 4.9 years). Prediction errors ranged from 0.2 to 14.2 years, with 53% of volunteers having errors less than 5 years, 42% between 5 and 10 years, and 5% between 10 and 15 years (Table 1). For Asian women (Figure 1C, Table 1), the correlation was similar (r = 0.93, MAE = 4.1 years), with errors ranging from 0.1 to 14.9 years. A higher percentage (71%) had errors under 5 years, and 23% were between 5 and 10 years. 5% still had errors between 10 and 15 years. For African-American women (Figure 1E, Table 1), the correlation was lower (r = 0.69, MAE = 8.3 years), with errors ranging from 0.01 to 26.6 years. Only 37% had errors under 5 years, while 28% were between 5 and 10 years, 19% between 10 and 15 years, and 16% over 15 years.

Clinical scoring revealed significant age-related changes in skin parameters for both Caucasian and Asian women ($p < 0.01$) (data not shown). Wrinkles increased substantially with age in both groups, while firmness, skin density, plumpness, and elasticity decreased. The magnitude of these changes was generally more pronounced in Caucasian women compared to Asian women. African-American women also showed significant age-related variations in wrinkles, firmness, plumpness, and elasticity, but the extent of these changes was less pronounced than in the other two groups. The smaller range of clinical scores in African-American women limited the ability of Multiple Linear Regression (MLR) to accurately correlate scores with aging. It is known and documented that aging manifests differently in different populations, such as ethnicities or phototypes, with features that appear at a later age or are weaker in darker skin compared to lighter skin [26].

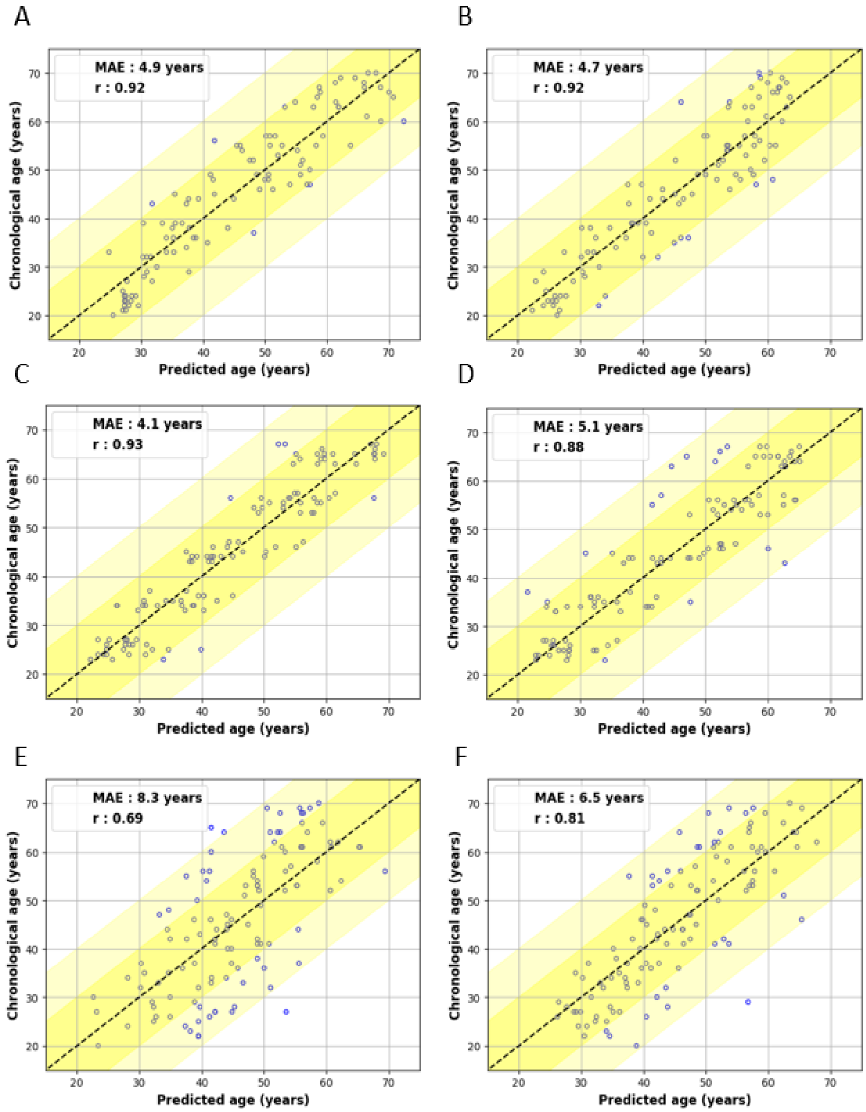

Figure 1. LEFT : Chronological age regressed against predicted age using Multiple Linear Regression applied to clinical scores. RIGHT : Chronological age regressed against predicted age using deep learning on LC-OCT images. A-B: Caucasian panel, C-D: Asian panel and E-F: African-American panel. The black dotted line represents the linear regression x=y. The shaded areas indicate ranges of ±10 years and ±20 years relative to the black line. Mean Absolute Error (MAE) and pearson correlation (r) are also provided for each model.

Table 1: Distribution of absolute errors, expressed in years, between chronological age and age predicted using Multiple Linear Regression (MLR) applied to clinical scores.

Absolute Error range (Years)	**< 1**	**(1-2.5]**	**(2.5-5]**	**(5-10]**	**(10-15]**	**>15**
Caucasian	8%	20%	25%	42%	5%	-
Asian	14%	22%	35%	23%	5%	-
African-American	9%	9%	20%	28%	19%	16%

The differences in the evolution of clinical scores with aging between ethnicities are aligned with the literature, which notably indicates that higher levels of melanin in darker skin provide greater protection against photoaging. This results in fewer visible signs of aging, such as wrinkles and loss of firmness, in individuals with darker skin compared to those with lighter skin [27]. Self-assessment studies on aging [28] indicate that Black women report the least severe facial aging, with many not reporting moderate to severe aging until ages 60-79. Caucasian women reported the most severe aging starting at ages 40-59, while Hispanics and Asians reported it later (ages 50-69). These observations support a connection between skin pigmentation, aging perceptions, and the severity of visible signs.

b. Predicting Individual Age Using LC-OCT 3D Features of the Dermis Fibrous Network

For Caucasian women, deep learning on LC-OCT images (Figure 1B) showed a strong correlation (r = 0.92, MAE = 4.7 years) with prediction errors from 0.0 to 17.9 years. 56% had errors under 5 years, 32% between 5 and 10 years, 11% between 10 and 15 years, and 1% over 15 years (Table 2). This MAE was comparable to the clinical scores model (Figure 1A), though with a broader error distribution. For Asian women (Figure 1D, Table 2), the correlation was slightly lower (r = 0.88, MAE = 5.1 years) with errors from 0.0 to 19.7 years. The error distribution was similar to Caucasians: 54% under 5 years, 32% between 5 and 10 years, and 9% from 10 to 15 years, but with 3.67% exceeding 15 years. For African-American women (Figure 1F, Table 2), the correlation was lower (r = 0.81, MAE = 6.5 years) with errors from 0.0 to 27.8 years. 45% had errors under 5 years, 31% between 5 and 10 years, 17.24% between 10 and 15 years, and 6.90% over 15 years.

Table 2: Distribution of absolute errors, expressed in years, between chronological age and age predicted using Deep Learning analysis applied to LC-OCT 3D images.

Absolute Error range (years)	**< 1**	**(1-2.5]**	**(2.5-5]**	**(5-10]**	**(10-15]**	**>15**
Caucasian	15%	19%	22%	32%	11%	1%
Asian	17%	19%	18%	32%	9%	4%
African-American	14%	10%	21%	31%	17%	7%

3.2 Predictive models multi-ethnicities

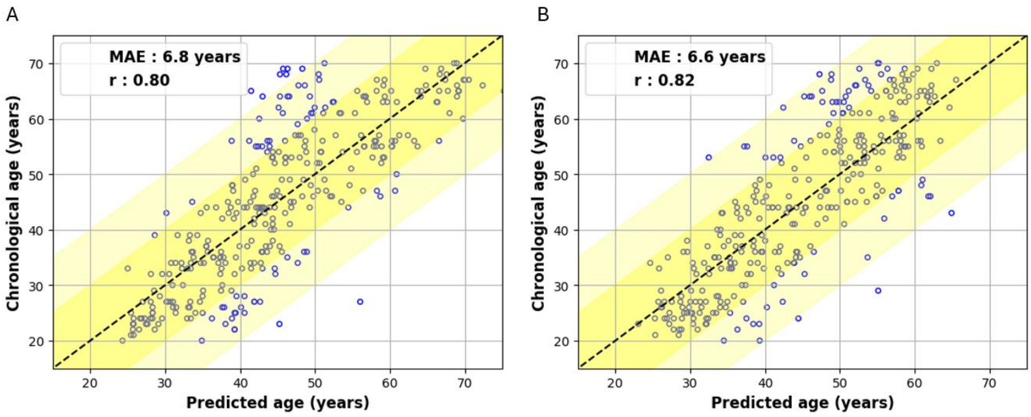

Figure 2. *A : Chronological age regressed against predicted age using Multiple Linear Regression applied to clinical scores. B : Chronological age regressed against predicted age using deep learning on LC-OCT images. The black dotted line represents the linear x=y. The shaded areas indicate ranges of ±10 years and ±20 years relative to the black line. Mean Absolute Error (MAE) and pearson correlation (r) are also provided for each model.*

In Figure 2A, the multiethnic MLR model yielded a MAE of 6.8 years (r = 0.80), which is higher than the individual Caucasian and Asian models (Figure 1A and 1C) but significantly lower than the African-American model (Figure 1E). When extracting errors by ethnicity, it was found that for Caucasian women, mean absolute errors in prediction ranged from 0.0 to 17.2 years, with 58% of volunteers displaying a predicted error of less than 5 years, which is slightly higher than the individual model. However, 27% of volunteers had an error within the range of 5 to 10 years, 11% had an error in the range of 10 to 15 years, and 2% of volunteers had errors greater than 15 years, indicating that the percentage of errors above 10% increased (Table 4). Similarly for the Asian women panel, the minimum and maximum errors were 0.0 and 16.4 years, that is comparable to the individual model. The percentage of volunteers displaying a predicted error of less than 5 years dropped to 61% compared to the individual model (Figure 1C, Table 1). Consequently, 30% of volunteers were observed in the 5 to 10 years range, and 6% of volunteers remained in the range of 10 to 15 years of error compared to the individual model. Additionally, 3% had errors greater than 15 years, suggesting that greater errors were found. For the African-American women panel, the minimum and maximum mean absolute errors in prediction ranged from 0.1 to 29 years, which is comparable to the individual model (Figure 1E). The percentage of volunteers displaying a predicted error of less than 5 years was reduced to 25% in the multiethnic model compared to 38% in the individual model (Table 1). Similarly to other models, a shift to higher errors was found with 32% of volunteers having an error within the range of 5 to 10 years, 24% in the 10 to 15 years range, and 19% of volunteers having errors greater than 15 years.

Table 4: Distribution of absolute errors, expressed in years, between chronological age and age predicted using MLR clinical scoring.

Absolute Error range (years)	**< 1**	**(1-2.5]**	**(2.5-5]**	**(5-10]**	**(10-15]**	**>15**
Caucasian	10%	21%	27%	29%	11%	2%
Asian	12%	16%	33%	30%	6%	3%
African-American	2%	10%	13%	32%	24%	19%

The multiethnic deep learning model (Figure 2B) trained yielded a mean absolute error (MAE) of 6.6 years (r = 0.82), which, consistent with the MLR, falls between the Asian and African-American models (Figure 1). Analyzing the outcome by ethnicity showed that for the Caucasian women panel, the mean absolute errors in prediction ranged from 0.1 to 20.5 years, which is broader compared to the individual model. The percentage of volunteers with a predicted error of less than 5 years was reduced to 42%. Additionally, 37% of volunteers had an error within the range of 5 to 10 years, 16% had a difference ranging from 10 to 15 years, and only 5% had a difference greater than 15 years, corresponding to an increase of about 4-5% for each range (Table 5). For the Asian women panel, there was also a broader range in mean absolute errors in prediction, ranging from 0.0 to 22.0 years. With 53% of volunteers displaying a predicted error of less than 5 years, this is comparable to the individual model (56%, Table 1). However, a decrease to 32% of volunteers within the range of 5 to 10 years was observed, accompanied by a slight increase to 10% in the 10 to 15 years range and a more noticeable increase to 6% with errors greater than 15 years. For the African-American women panel, mean absolute errors in prediction ranged from 0.0 to 26.2 years, which is consistent with the individual model. While 41% of volunteers displayed a predicted error of less than 5 years, which is 4% less than the individual model, the 5 to 10 years range saw a percentage increase by 4% to 35% of volunteers. The percentage of volunteers in the range of 10 to 15 years remained steady at 18% (compared to 17%), and the percentage of volunteers with errors greater than 15 years increased to 10%.

Table 5: Distribution of absolute errors, expressed in years, between chronological age and age predicted using Deep Learning analysis applied to LC-OCT 3D images.

Absolute Error range (years)	**< 1**	**(1-2.5]**	**(2.5-5]**	**(5-10]**	**(10-15]**	**>15**
Caucasian	10%	15%	17%	37%	16%	5%
Asian	12%	17%	24%	32%	10%	6%
African-American	7%	9%	25%	35%	18%	10%

Deep learning algorithms offer high accuracy and automated analysis for quantitative models by capturing complex image patterns and reducing human error. These models automatically extract relevant features from LC-OCT images, learning directly from raw data labeled with volunteer age. This automation is especially valuable in large-scale studies and scalable across diverse populations, including Caucasian, Asian, and African-American groups

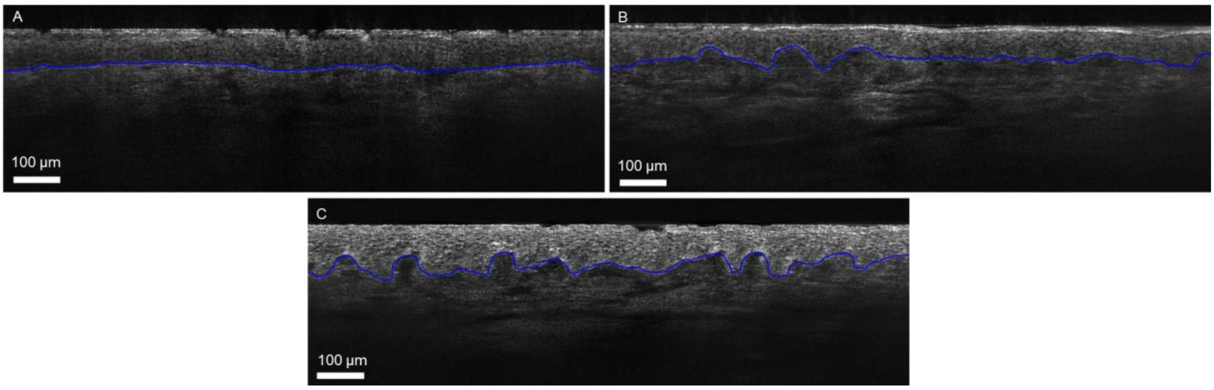

Figure 3: *Examples of 2D transversal reconstructed LC-OCT images representing the dermal-epidermal junction segmentation (DEJ). A: Caucasian woman (45 years old). B: Asian woman (45 years old). C: African-American woman (45 years old). The blue line indicates the position of the DEJ.*

Deep learning models, such as 3D ResNets, require large datasets for optimal training and accuracy, but generalization across different ethnicities can be challenging. A multiethnic model, combining data from Caucasian, Asian, and African-American women, showed reduced age prediction performance compared to ethnicity-specific models. Biological factors, such as variations in skin structure, pigmentation, and aging patterns, can impair generalization. For instance, the dermal-epidermal junction exhibited different degrees of undulation across ethnicities (Figure 3) [8,29]. Technical limitations related to melanin's absorbance of near-infrared light, which affects the LC-OCT signal, also play a role [30]. The African-American model showed a significant increase in MAE, suggesting that melanin's optical properties can compromise training. In contrast, Caucasian and Asian models performed well (MAE = 4.7 years and 5.1 years, respectively). Future improvements could include advancements in GPU technology, optimization of computational resources, and the development of more comprehensive and inclusive databases to refine algorithm training. Confocal imaging should aim to build large, diverse databases to improve the reliability and efficiency of in vivo skin analysis.

5. Conclusion

LC-OCT coupled with deep learning holds promise for developing new tools for in vivo skin characterization. Presently, the training of a 3D ResNet18 has demonstrated promising capabilities in facial age prediction in healthy Caucasian (MAE = 4.7 years) and Asian women (MAE = 5.1 years). While the method could, in principle, be generalized, in practice, the heterogeneity in skin structures suggests that it is best to train specific models per ethnicity to ensure accuracy. The African-American model was associated with higher errors in age prediction, which can be a combined consequence of lesser modifications in the fibrous network due to aging and the attenuated signal due to higher melanin content. Therefore, the application of models to darker skin needs further optimization and investigation for a better understanding to adapt the methodology accordingly.

Bibliography

[1] Chauvel-Picard J et al. 2022;15:e202100236. https://doi.org/10.1002/jbio.202100236.
[2] Monnier J et al. 2020;34:2914–21. https://doi.org/10.1111/jdv.16857.
[3] Latriglia F et al. 2023;13:2268. https://doi.org/10.3390/life13122268.
[4] Pezzini C et al. 2023;32:348–58. https://doi.org/10.1111/exd.14708.
[5] Ogien J et al. 2022;11934:1193409-1193409–6. https://doi.org/10.1117/12.2607344.
[6] Fischman S et al. 2022;12:481. https://doi.org/10.1038/s41598-021-04395-1.
[7] Chen K et al. 2023;29:e13340. https://doi.org/10.1111/srt.13340.
[8] Ali A et al. 2024;30:e13643. https://doi.org/10.1111/srt.13643.
[9] Guida S et al. 2021;11:e2021068. https://doi.org/10.5826/dpc.1103a68.
[10] Koehler MJ et al. 2006;31:2879. https://doi.org/10.1364/ol.31.002879.
[11] Montgomery KL et al. 2024;14:26129. https://doi.org/10.1038/s41598-024-76908-7.
[12] Pedrazzani M et al. 2020;26:398–404. https://doi.org/10.1111/srt.12815.
[13] Ayadh M et al. 2023;6:6. https://doi.org/10.1051/fopen/2023005.
[14] Voegeli R et al. 2021;43:547–60. https://doi.org/10.1111/ics.12727.
[15] Bobrov E et al. 2018;10:3249–59. https://doi.org/10.18632/aging.101629.
[16] Zalay O et al. 2023:2023.09.12.23295132. https://doi.org/10.1101/2023.09.12.23295132.
[17] Ahn I et al. 2024;14:30084. https://doi.org/10.1038/s41598-024-78695-7.
[18] ELKarazle K et al. 2022;6:128. https://doi.org/10.3390/bdcc6040128.
[19] Dibeklioğlu H et al. 2015;24:1928–43. https://doi.org/10.1109/tip.2015.2412377.
[20] Robic J et al. 2021;35:e68–70. https://doi.org/10.1111/jdv.16810.
[21] Pena A-M et al. 2022;12:14863. https://doi.org/10.1038/s41598-022-18657-z.

[22] Breugnot J et al. 2023;29:e13221. https://doi.org/10.1111/srt.13221.
[23] Assi A et al. 2024;14:24113. https://doi.org/10.1038/s41598-024-74370-z.
[24] OGIEN J et al. 2023;158:171–9. https://doi.org/10.23736/s2784-8671.23.07613-2.
[25] Hara K et al. 2018:6546–55. https://doi.org/10.1109/cvpr.2018.00685.
[26] 2007;56:AB33. https://doi.org/10.1016/j.jaad.2006.10.194.
[27] Campiche R et al. 2019;18:614–27. https://doi.org/10.1111/jocd.12806.
[28] Alexis AF et al. 2019;45:1635–48. https://doi.org/10.1097/dss.0000000000002237.
[29] Bonnier F et al. 2023;13:13881. https://doi.org/10.1038/s41598-023-40340-0.
[30] Davis A et al. 2019;10:694. https://doi.org/10.1364/boe.10.000694.