Received: 28 February 2024 | Accepted: 18 March 2024

DOI: 10.1111/srt.13690

ORIGINAL ARTICLE

WILEY

The relevance and accuracy of an AI algorithm-based descriptor on 23 facial attributes in a diverse female US population

Zhi Yu[1] | Frederic Flament[2] | Ruowei Jiang[1] | Jeff Houghton[1] | Camille Kroely[3] | Nathalie Cabut[3] | Diala Haykal[4] | Cassidy Sehgal[5] | Nina G Jablonski[6] | Aurelie Jean[7] | Parham Aarabi[1]

[1]Modiface – A L'Oréal Group Company, Toronto, Canada

[2]L'Oréal Research and Innovation, Clichy, France

[3]L'Oréal CDO – Digital Service Factory, Clichy, France

[4]Centre Médical Laser Esthétique, Palaiseau, France

[5]L'Oréal Research and Innovation, Clark, USA

[6]Department of Anthropology, The Pennsylvania State University, University Park, Pennsylvania, USA

[7]Hult International Business School, Cambridge, USA

Correspondence
Flament Frederic, L'Oreal Research and Innovation, 9 Rue Pierre Dreyfus, 92110, Clichy, France.
Email: frederic.flament@rd.loreal.com

Funding information
L'Oreal Research & Innovation

Abstract

Background: The response of AI in situations that mimic real life scenarios is poorly explored in populations of high diversity.

Objective: To assess the accuracy and validate the relevance of an automated, algorithm-based analysis geared toward facial attributes devoted to the adornment routines of women.

Methods: In a cross-sectional study, two diversified groups presenting similar distributions such as age, ancestry, skin phototype, and geographical location was created from the selfie images of 1041 female in a US population. 521 images were analyzed as part of a new training dataset aimed to improve the original algorithm and 520 were aimed to validate the performance of the AI. From a total 23 facial attributes (16 continuous and 7 categorical), all images were analyzed by 24 make-up experts and by the automated descriptor tool.

Results: For all facial attributes, the new and the original automated tool both surpassed the grading of the experts on a diverse population of women. For the 16 continuous attributes, the gradings obtained by the new system strongly correlated with the assessment made by make-up experts ($r \geq 0.80$; $p < 0.0001$) and supported by a low error rate. For the seven categorical attributes, the overall accuracy of the AI-facial descriptor was improved via enrichment of the training dataset. However, some weaker performance in spotting specific facial attributes were noted.

Conclusion: In conclusion, the AI-automatic facial descriptor tool was deemed accurate for analysis of facial attributes for diverse women although some skin complexion, eye color, and hair features required some further finetuning.

KEYWORDS
age, AI-automatic facial descriptor, ancestry, diverse women, facial attributes, skin phototype

1 | INTRODUCTION

Over the past few years, the intensive development and refinement of artificial intelligence (AI) based systems enabled the rise of technologies dedicated to medicine/health fields and especially dermatology.[1–4]

These AI analytical based systems, coupled with novel connected devices allow a high number of people to receive a skin diagnosis or medical advice with particular focus on those with reduced accessibility to practicians.[5,6] With regard to the wide diversity of patients, physicians, or consumers worldwide, it became critical to perform validations in real life conditions to demonstrate the accuracy and relevance of all AI-based analyses made to run on very diverse populations.[7,8] The distinctive attributes of different human population with respect to their ancestry or skin color[9–16] has affected AI-based evaluations.[17–20] These factors are especially crucial in the cosmetic field where facial appearance such as the perception of age and radiance/glow is paramount in screening.[21–25] The ensuing care or adorning procedures comprise a wide variety of cosmetic products or interventions that fulfill individual needs.[26–32]

By modifying or correcting facial traits, make-up application is a key process that often requires the input of make-up professionals. In this context, an original AI-based automated facial descriptor system was developed[33] to assist professionals in assessing and grading facial features in line with their specificities. Twenty-three facial attributes pertaining to face morphology and color were defined by a quorum of make-up artists, and among the latter, priority was given to those more accurately informed of the development of personalized make-up routines.

This automated system was initially developed from 12,000 selfie images that were previously graded by experts on the 23 attributes and covered both gender, age-classes, ancestry, and skin phototype.[33] As a continuation to this initial investigation that helped to identify areas where the grading's accuracy needed improvements, a follow-up study was designed with two goals (i) to improve the accuracy of the original AI-based automatic descriptor and (ii) to validate and assess the relevance of this newer model when applied to subjects from a diverse population. As the design of the most accurate datasets for both training and validation phases is critical to the performance of the resulting algorithm, a transverse study that comprises subjects from a country featuring a diverse, multi-ethnic population, such as the United States of America, was conducted.

Using on-line recruitment, this remote study was designed and conducted to cover four broad categories of ancestry over a large age range (18—80 years old) and six phototypes according to the Fitzpatrick scale. The goal was to improve the original system[33] and explore the relevance and accuracy of this new automated algorithm-based system in analyzing the 23 facial attributes of each subject of the study. The results of this cross-sectional investigation carried out in real-life conditions are the main topic of this study.

2 | MATERIAL AND METHODS

2.1 | Participating subject and make-up experts

In this cross-sectional study conducted online in the USA, 1041 adult American women aged 18—80 years (average: 42.1 years old) were selected, based on an anonymous, short online self-questionnaire reporting age, ancestry and estimated skin tone using the Fitzpatrick phototype classification.[34] Informed written consent was obtained for all subjects participating in this study. Given that no interventions on the human volunteers were carried out and only selfie images were used, an ethics committee approval of the study protocol was not warranted.

In order to avoid blatant discrepancies or errors in this estimation, 40% of these were randomly cross-checked by an expert grader. To cover a highly diverse population, selected participants comprised of four major ancestries (Non-Hispanic Euro-American, African American, Hispanic Euro-American, and East Asian), to include (i) the I–VI Fitzpatrick classification and (ii) five age-classes of 10 years each.

Of a total of 120 possible combinations, some were unlikely such as phototypes I and II in African American ancestry, phototypes V and VI in Non-Hispanic Euro-American, and phototypes I and VI in both East Asian population and Hispanic Euro-American ancestries (see Table 1). The residences of the subjects covered four main regions of the USA, that is, Northeast (25%), Midwest (16%), South (39%), and West (20%).

Inclusion criteria were (i) a facial skin devoid of any cutaneous disorder or lesion, (ii) possessing a smartphone (any brand) with a high-resolution camera (≥5 Megapixels), and (iii) familiar in taking selfie pictures. All subjects were fully informed about the objective of the study and signed an informed consent that guaranteed total confidentiality of their images (blind-coded) and deleted once analyzed. The average age of the 24 make-up experts was 37.9 years old, mostly composed of women (~70%) originating from different regions of the world and of three different ancestries, as shown by Table 2.

2.2 | Data collection

Participants signed an informed consent form and were instructed to take frontal selfie pictures at home with a neutral expression, wearing no skincare, or cosmetics (i.e., bare skin) and unobstructed (without glasses and no filters for example). The time when the picture was taken, and the lighting conditions were at the subject's own discretion. Examples of acceptable selfie pictures to indicate the optimum distance and angle were initially shown to all subjects. All pictures were then sent to a dedicated and secured platform for further processing. Once the study was completed, the pictures were destroyed.

2.3 | AI-based automated facial descriptor

An automated descriptor system of facial features was previously designed[33] to evaluate multiple attributes from a single selfie image

TABLE 1 New dataset: Distribution of the 1041 studied American women according to age-classes, phototypes, locations, and ancestries.

Studied American women characteristics	Sub-groups	Non-Hispanic Euro-American	African American	Hispanic Euro-American	East Asian	Global population
Age-classes	18y–29y	64	62	66	57	249
	30y–39y	65	63	61	56	245
	40y–49y	64	61	55	43	223
	50y–64y	60	56	51	29	196
	65y–80y	60	38	20	10	128
Skin phototype (Fitzpatrick)	I	77	NA	NA	NA	77
	II	80	NA	78	58	216
	III	79	73	74	68	294
	IV	77	77	62	43	259
	V	NA	73	39	26	138
	VI	NA	57	NA	NA	57
Geographical locations	Northeast	106	47	61	48	262
	Midwest	62	45	20	35	162
	South	109	145	113	44	411
	West	36	43	59	68	206
Global population	Ancestries	313	280	253	195	1,041

Note: NA, not applicable.

TABLE 2 Distribution of the 24 make-up experts involved in selfie images' grading according to age-classes, gender, and ancestries/locations.

Dermatologists' characteristics	Sub-groups	Population
Age-classes	**18y–35y**	9
	35y–50y	9
	50y–59y	6
Gender	**Female**	17
	Male	7
Ancestries / Locations	**South America**	8
	East Asia	8
	Africa	8

captured by the frontal camera of a smartphone. The proposed objective behind this study is not only motivated by recent advances[35–37] in the prediction of a facial attribute but takes into account the intrinsic correlation of facial attributes provided by the Convolutional Neural Networks (CNN). Briefly, the issue to address is framed as a supervised classification problem associated with a deep learning framework, using a CNN as network backbone.

The proposed methodology consists of two separate models, each in charge of extracting images of facial signs and classifying the corresponding attributes, divided into sub-groups. One group deals with shape-related attributes and another one deals with color-related attributes. Each model consists of a shared feature extractor (ResNet)[38] for all attributes, followed by sorting heads that will classify each attribute. The shared feature learning then exploits the relationship between tasks to achieve a robust and discriminative featuring. When running the test, given a facial image x, the system builds a vector of confidence scores $y = \mathrm{fh}(x)$, where fh is a neural network parameterized by h. Each component of y corresponds to the possibility of the model to predict x to belong in a specific class. When training the algorithm, multiple samples are defined with human inputs in the form (xi, yi), with xi being the i-th training image and yi the corresponding vector of class annotation, each annotation being a preset of classes chosen for each attribute.

2.4 | Protocol

From November 2020 to March 2021, 1041 American women performed a facial diagnostic using the original AI-based automatic descriptor. Using the same distribution of age, phototype, ancestry and location, the population of 1041 was broken into two groups namely, a training dataset (521 subjects) and a validation dataset (520 subjects). The training dataset was dedicated toward improving the original algorithm. Meanwhile, the validation dataset aimed at exploring the functionality of the AI in a diverse population and assess its accuracy during grading.

Step 1: All 1041 subjects were instructed to take one selfie image at maximum resolution (full face, frontal camera of smartphone) and send it to our secure platform.

Step 2: The 24 trained make-up experts were asked to grade all the 23 facial attributes on the 1041 selfie images within their specific scales.[33] The attributed grade of a continuous facial attribute was therefore an average of grades calculated by all the experts.

As for categorical attributes, the outcome of its evaluation is dependent on a unanimous decision based on common ground truth which may prove to be complex. In such cases, the majority consensus was considered as ground truth for all subjects in the given attribute. When no clear majority was reached among the experts, the classes with the most votes was considered correct. For example, if 3 "round" and 3 "almond" was scored for eye shape, both would be counted as correct. To ensure the robustness and the repeatability of this process, several pictures were presented twice to the experts during the evaluation process. To best standardize and help the experts' grading sessions, in addition to providing scales with illustrative pictures,[33] the same monitors with similar settings were used.

Step 3: Keeping the same distribution in age-classes, phototype, ancestry and locations, a sub-dataset of 521 women (or selfie pictures) was defined (training dataset) to improve the original AI-based automatic descriptor.[33]

Step 4: The other 520 women (or selfie pictures) (validation dataset) were analyzed by both the original[33] and the new automated grading systems. The resulting scores for the 16 continuous and the 7 categorical attributes (Table 3) were sent blind-coded to our secure platform.

Step 5: Automatic scores were analyzed versus experts' average grading in each subject and for each continuous facial attribute or versus agreement sorting for each categorical attribute.

2.5 | Computing methods and statistical analysis

For each continuous facial attribute of a given subject, the Pearson product-moment correlation coefficients between experts' grading and the original or the new AI-based algorithms' predicted scores were determined. Their overall significance was calculated by considering the respective selfie images of the 520 participants of the validation dataset, either as total (all subjects), by ancestry, by phototype or by age-class.

For each continuous facial attribute and each subject, the Pearson's product-moment correlations between the individual grading of each expert and the average grading of all experts were determined. Their overall significance was calculated by considering the 520 participants of the validation dataset.

Several statistical calculations were used to evaluate the performances of the automatic grading system of continuous attributes. The retained parameter was the mean absolute error (MAE), that is, the average of the absolute differences (in grading units) between the AI predicted gradings and those made by all make-up experts. MAE was therefore computed, for continuous attribute, as the average of the absolute differences (in grading units) between individual expert gradings and those made by all experts.

TABLE 3 The 23 facial attributes obtained by the automatic descriptor system and the 24 experts with their associated ranges and nature (16 continuous and 7 categorical).

Face attributes	Range scale	Nature scale
1. Skin tone complexion	6 points	Categorical
2. Undertone	3 points	Categorical
3. Freckles density	4 points	Continuous
4. Face length	6 points	Continuous
5. Jawline shape	6 points	Continuous
6. Eyebrows' shape	4 points	Categorical
7. Eyebrows' thickness	4 points	Continuous
8. Eyelashes length	3 points	Continuous
9. Nose width	3 points	Continuous
10. Eyes color	6 points	Categorical
11. Eyes shape	2 points	Continuous
12. Eyes crease	3 points	Continuous
13. Eyes orientation	3 points	Continuous
14. Distance eyes-nose	3 points	Continuous
15. Eyes depth	3 points	Continuous
16. Dark circles	2 points	Continuous
17. Head hair color	12 points	Categorical
18. Hair form	6 points	Categorical
19. Hair length	6 points	Categorical
20. Forehead height	3 points	Continuous
21. Lips thickness	3 points	Continuous
22. Lips width	3 points	Continuous
23. Cupid's bow size	3 points	Continuous

For categorical attributes, estimation of performance was approached with the percentage of correct prediction for both original and new automatic systems and the ground truth from 24 experts.

3 | RESULTS

3.1 | Continuous facial attributes

Figure 1 illustrates the correlation coefficients for the 16 continuous facial attributes obtained by both new and original automated systems and compared to average experts' assessments, taken as references. Interestingly, the maximum correlation coefficients calculated between every possible pair of graders was consistently lower across all facial attributes compared to those provided between the automatic prediction and the average expert graders. This suggested a high accuracy and stability of the improved AI-based system ($\geq$0.75 for two attributes to $\geq$0.80 for twelve attributes, $p < 0.0001$).

Table 4 illustrates the values of MAE's of all 16 attributes found in either global population or in all age-classes, ancestries and photo-

TABLE 4 Values of the mean absolute errors (MAE) in the gradings of 16 continuous facial attributes obtained by the original and the new automatic descriptor systems against the grading by the 24 experts. Results on all validation dataset, that is, by age-classes, phototypes and ancestries. African American (AA), East Asian (EA), Non-Hispanic Euro-American (NHEA), Hispanic Euro-American (HEA).

Facial attributes	MAE original automatic descriptor						MAE new and inclusive automatic descriptor						MAE make-up experts					
Global panel	$N = 520$						$N = 520$						$N = 520$					
3. Freckles density	0.26						0.20						0.45					
4. Face length	0.15						0.15						0.28					
5. Jawline shape	0.28						0.27						0.57					
7. Eyebrows' thickness	0.22						0.19						0.41					
8. Eyelashes length	0.24						0.21						0.41					
9. Nose width	0.23						0.20						0.45					
11. Eyes shape	0.14						0.13						0.23					
12. Eyes crease	0.16						0.15						0.44					
13. Eyes orientation	0.31						0.28						0.49					
14. Distance eyes-nose	0.22						0.19						0.48					
15. Eyes depth	0.26						0.21						0.50					
16. Dark circles	0.12						0.11						0.35					
20. Forehead height	0.23						0.21						0.48					
21. Lips thickness	0.21						0.17						0.36					
22. Lips width	0.23						0.20						0.46					
23. Cupid's bow size	0.32						0.24						0.43					
Age	18y–29y	30y–39y	40y–49y	50y–64y	65y–80y	NA	18y–29y	30y–39y	40y–49y	50y–64y	65y–80y	NA	18y–29y	30y–39y	40y–49y	50y–64y	65y–80y	NA
3. Freckles density	0.25	0.27	0.28	0.22	0.24	NA	0.19	0.21	0.23	0.18	0.15	NA	0.46	0.48	0.46	0.43	0.44	NA
4. Face length	0.17	0.13	0.13	0.16	0.15	NA	0.16	0.13	0.13	0.15	0.17	NA	0.27	0.28	0.27	0.29	0.33	NA
5. Jawline shape	0.24	0.23	0.30	0.29	0.38	NA	0.27	0.24	0.31	0.27	0.28	NA	0.54	0.55	0.59	0.63	0.54	NA
7. Eyebrows' thickness	0.21	0.18	0.20	0.25	0.28	NA	0.19	0.19	0.18	0.19	0.23	NA	0.44	0.40	0.39	0.44	0.44	NA
8. Eyelashes length	0.23	0.22	0.27	0.28	0.20	NA	0.22	0.20	0.22	0.22	0.15	NA	0.52	0.45	0.42	0.33	0.20	NA

(Continues)

TABLE 4 (Continued)

Age	18y–29y	30y–39y	40y–49y	50y–64y	65y–80y	NA	18y–29y	30y–39y	40y–49y	50y–64y	65y–80y	NA	18y–29y	30y–39y	40y–49y	50y–64y	65y–80y	NA
9. Nose width	0.24	0.22	0.21	0.24	0.23	NA	0.22	0.19	0.20	0.22	0.20	NA	0.46	0.46	0.44	0.43	0.49	NA
11. Eyes shape	0.16	0.16	0.14	0.11	0.12	NA	0.13	0.14	0.13	0.10	0.10	NA	0.26	0.26	0.25	0.20	0.17	NA
12. Eyes crease	0.13	0.15	0.15	0.18	0.21	NA	0.14	0.14	0.13	0.16	0.20	NA	0.36	0.42	0.46	0.49	0.55	NA
13. Eyes orientation	0.29	0.30	0.31	0.34	0.33	NA	0.29	0.30	0.29	0.30	0.21	NA	0.48	0.47	0.51	0.48	0.51	NA
14. Distance eyes-nose	0.22	0.21	0.23	0.21	0.21	NA	0.18	0.18	0.20	0.20	0.21	NA	0.47	0.48	0.50	0.49	0.51	NA
15. Eyes depth	0.26	0.26	0.25	0.26	0.28	NA	0.19	0.20	0.22	0.20	0.23	NA	0.49	0.46	0.48	0.53	0.51	NA
16. Dark circles	0.12	0.12	0.12	0.12	0.14	NA	0.11	0.11	0.12	0.10	0.11	NA	0.35	0.38	0.35	0.30	0.36	NA
20. Forehead height	0.20	0.25	0.21	0.25	0.28	NA	0.23	0.21	0.19	0.21	0.25	NA	0.47	0.48	0.47	0.48	0.54	NA
21. Lips thickness	0.21	0.20	0.21	0.24	0.21	NA	0.15	0.16	0.17	0.22	0.18	NA	0.36	0.38	0.39	0.32	0.31	NA
22. Lips width	0.24	0.22	0.23	0.23	0.22	NA	0.20	0.18	0.20	0.21	0.24	NA	0.44	0.43	0.46	0.48	0.50	NA
23. Cupid's bow size	0.31	0.30	0.34	0.30	0.41	NA	0.24	0.22	0.23	0.25	0.30	NA	0.43	0.42	0.44	0.42	0.41	NA
Photo-types	**I**	**II**	**III**	**IV**	**V**	**VI**	**I**	**II**	**III**	**IV**	**V**	**VI**	**I**	**II**	**III**	**IV**	**V**	**VI**
3. Freckles density	0.33	0.24	0.24	0.26	0.24	0.27	0.20	0.18	0.21	0.21	0.19	0.16	0.44	0.46	0.49	0.46	0.42	0.37
4. Face length	0.17	0.14	0.13	0.15	0.16	0.13	0.14	0.15	0.15	0.15	0.14	0.12	0.35	0.27	0.27	0.27	0.31	0.36
5. Jawline shape	0.27	0.26	0.29	0.30	0.24	0.25	0.28	0.26	0.28	0.28	0.25	0.20	0.60	0.58	0.60	0.54	0.54	0.48
7. Eyebrows' thickness	0.26	0.23	0.21	0.21	0.18	0.25	0.21	0.18	0.20	0.19	0.16	0.24	0.46	0.41	0.40	0.42	0.46	0.45
8. Eyelashes length	0.25	0.23	0.25	0.23	0.25	0.27	0.23	0.21	0.22	0.19	0.20	0.19	0.43	0.45	0.41	0.42	0.36	0.30
9. Nose width	0.29	0.23	0.23	0.22	0.22	0.18	0.21	0.21	0.20	0.20	0.19	0.23	0.48	0.46	0.46	0.44	0.43	0.43
11. Eyes shape	0.18	0.15	0.13	0.14	0.14	0.10	0.16	0.14	0.11	0.12	0.11	0.11	0.17	0.24	0.23	0.28	0.23	0.19
12. Eyes crease	0.16	0.18	0.16	0.14	0.16	0.13	0.16	0.18	0.15	0.13	0.14	0.13	0.51	0.45	0.46	0.43	0.41	0.30
13. Eyes orientation	0.28	0.34	0.30	0.30	0.31	0.26	0.27	0.38	0.26	0.27	0.28	0.26	0.46	0.48	0.47	0.51	0.51	0.52
14. Distance eyes-nose	0.20	0.19	0.22	0.21	0.27	0.26	0.18	0.18	0.20	0.18	0.19	0.21	0.47	0.47	0.49	0.49	0.50	0.57
15. Eyes depth	0.26	0.28	0.26	0.27	0.23	0.23	0.21	0.23	0.18	0.21	0.20	0.20	0.50	0.46	0.51	0.50	0.47	0.54
16. Dark circles	0.17	0.13	0.13	0.11	0.10	0.09	0.13	0.12	0.11	0.11	0.10	0.08	0.44	0.39	0.35	0.31	0.30	0.28

(Continues)

TABLE 4 (Continued)

Photo-types	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
20. Forehead height	0.23	0.26	0.22	0.21	0.23	0.28	0.21	0.22	0.18	0.22	0.23	0.28	0.47	0.46	0.48	0.49	0.54	0.49
21. Lips thickness	0.25	0.22	0.22	0.22	0.19	0.14	0.18	0.16	0.16	0.21	0.15	0.11	0.34	0.36	0.38	0.37	0.32	0.30
22. Lips width	0.24	0.22	0.23	0.24	0.19	0.30	0.18	0.21	0.21	0.19	0.19	0.21	0.48	0.45	0.45	0.44	0.45	0.44
23. Cupid's bow size	0.43	0.38	0.30	0.30	0.27	0.26	0.23	0.26	0.24	0.25	0.22	0.20	0.40	0.43	0.44	0.42	0.39	0.45
Ancestries	**AA**	**EA**	**NHEA**	**HEA**	**NA**	**NA**	**AA**	**EA**	**NHEA**	**HEA**	**NA**	**NA**	**AA**	**EA**	**NHEA**	**HEA**	**NA**	**NA**
3. Freckles density	0.26	0.28	0.20	0.26	NA	NA	0.21	0.20	0.17	0.20	NA	NA	0.45	0.45	0.45	0.45	NA	NA
4. Face length	0.16	0.14	0.15	0.13	NA	NA	0.16	0.13	0.17	0.13	NA	NA	0.28	0.29	0.28	0.29	NA	NA
5. Jawline shape	0.28	0.29	0.26	0.27	NA	NA	0.29	0.28	0.28	0.24	NA	NA	0.57	0.57	0.57	0.57	NA	NA
7. Eyebrows' thickness	0.25	0.22	0.17	0.21	NA	NA	0.21	0.19	0.17	0.19	NA	NA	0.41	0.42	0.41	0.42	NA	NA
8. Eyelashes length	0.25	0.24	0.24	0.24	NA	NA	0.19	0.22	0.22	0.20	NA	NA	0.41	0.41	0.41	0.41	NA	NA
9. Nose width	0.18	0.27	0.24	0.21	NA	NA	0.18	0.21	0.22	0.21	NA	NA	0.45	0.45	0.45	0.45	NA	NA
11. Eyes shape	0.12	0.18	0.14	0.12	NA	NA	0.09	0.16	0.11	0.12	NA	NA	0.23	0.23	0.24	0.23	NA	NA
12. Eyes crease	0.14	0.17	0.18	0.15	NA	NA	0.13	0.16	0.17	0.14	NA	NA	0.44	0.44	0.44	0.44	NA	NA
13. Eyes orientation	0.32	0.32	0.33	0.28	NA	NA	0.29	0.30	0.32	0.24	NA	NA	0.49	0.49	0.49	0.49	NA	NA
14. Distance eyes-nose	0.23	0.18	0.19	0.26	NA	NA	0.20	0.18	0.18	0.21	NA	NA	0.48	0.48	0.48	0.48	NA	NA
15. Eyes depth	0.24	0.30	0.26	0.23	NA	NA	0.18	0.23	0.19	0.20	NA	NA	0.49	0.49	0.49	0.50	NA	NA
16. Dark circles	0.12	0.14	0.13	0.10	NA	NA	0.12	0.12	0.11	0.09	NA	NA	0.35	0.35	0.34	0.34	NA	NA
20. Forehead height	0.22	0.22	0.24	0.25	NA	NA	0.20	0.19	0.24	0.23	NA	NA	0.48	0.48	0.48	0.48	NA	NA
21. Lips thickness	0.21	0.22	0.20	0.22	NA	NA	0.16	0.18	0.16	0.19	NA	NA	0.36	0.36	0.36	0.36	NA	NA
22. Lips width	0.25	0.21	0.22	0.23	NA	NA	0.21	0.19	0.20	0.21	NA	NA	0.45	0.45	0.45	0.45	NA	NA
23. Cupid's bow size	0.30	0.38	0.31	0.29	NA	NA	0.25	0.24	0.25	0.23	NA	NA	0.43	0.43	0.43	0.43	NA	NA

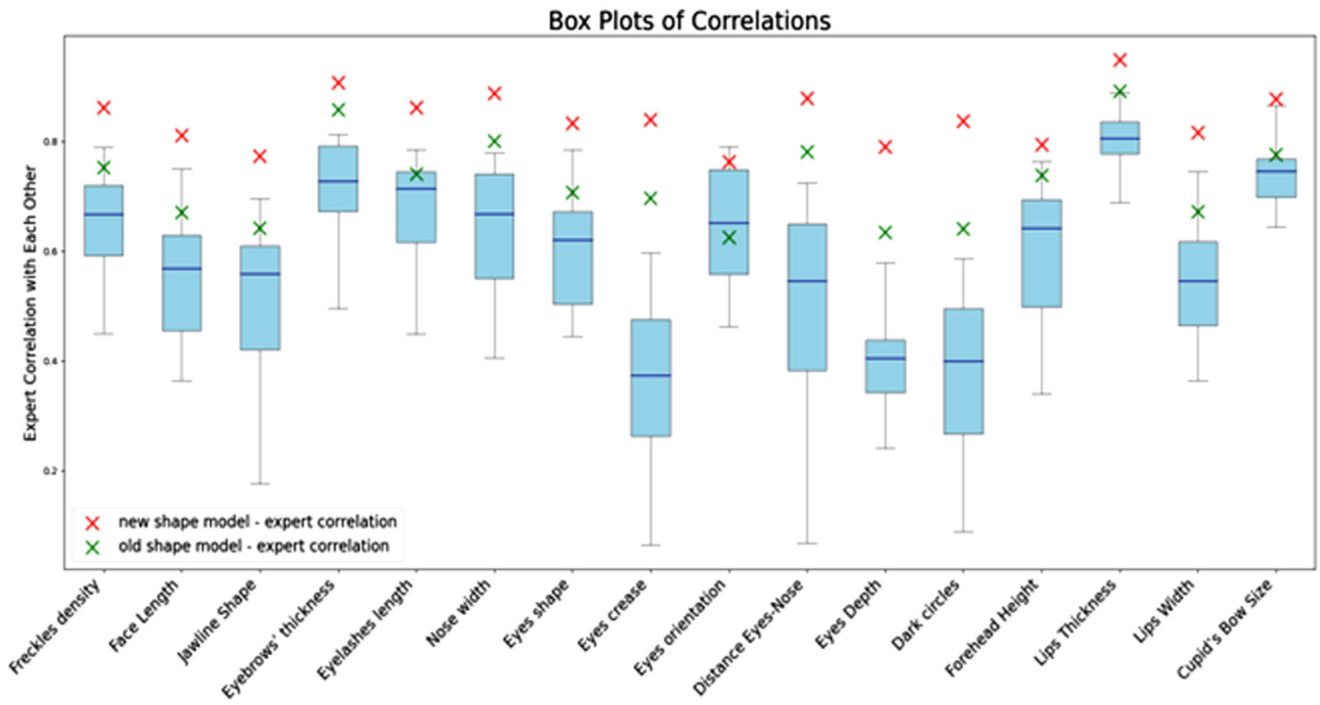

FIGURE 1 Box plots graph showing the correlation coefficients of 16 continuous facial attributes. Red cross represents values of correlation of the new AI-automatic descriptor and average of all experts scores, taken as reference, for each facial attribute and each selfie image. Green cross describes values of correlation for the original AI-based automatic descriptor. Blue square shows the correlation of each individual experts versus average of all experts scores, taken as a reference, and its variation.

types, by the new and the original systems and those obtained by the experts. It becomes clear that the new system presented lower MAE's than the original system and much lower values than those provided by the 24 experts, by about two-fold. The MAE of the automatic system obtained on the 16 continuous criteria can be subdivided into three classes according to thresholds (i.e., <0.3, <0.5, <1.0), expressed in absolute errors in grading units. The new system clearly outperformed the original one with 78.5% versus 73.1% (MAE < 0.3), 94.3% versus 90.8% (MAE < 0.5), and 99.9% versus 99.6% (MAE < 1.0).

Figure 2 illustrates the absolute errors (in grading units) recorded by the two AI-based automatic systems and by each make-up expert for each of the 16 facial attributes. For all attributes, predicted scores from the original and the new systems showed lower error rates compared to the panel of experts. Errors made by the new system were generally of lower values, with the exception of eye crease, jawline shape and face length where the errors were similar in both systems. Overall, errors for the new system were <0.25 grading units whereas those from experts were >0.35 grading units. Interestingly, the maximum error rates were far higher for the experts, suggesting a good stability and reliability of the new AI-based automatic descriptor.

3.2 | Categorical facial attributes

Table 5 shows that the performance of the new system for the seven categorical attributes was better than the original one in all cases, that is, considering the global population, age-classes, phototypes and ancestries. Moreover, the new system performed better than the experts' gradings.

Contrary to continuous facial attributes, the new automated system presented a lower accuracy when observing sub-groups. For the elderly group, after 50 years old, with the exception of the eyebrows' shape, the six other attributes presented a lower prediction capacity. For the different phototypes, some areas of improvement were (i) Phototype I: eyes color, head hair color; (ii) Phototype IV: skin tone complexion, hair length; (iii) Phototype V: skin tone complexion, hair length; and (iv) Phototype VI: hair form, hair length. A difference in accuracy could be observed between ancestries. This led to a lower prediction in eyebrows' shape of East Asian women; eye color in Non-Hispanic Euro-American women; skin tone complexion for African American women. As for Hispanic Euro-American women, possible fields of improvement included skin tone complexion and undertone assessments.

4 | DISCUSSION

Defining and collecting a specific dataset devoted to inclusivity and representative of a large and diverse country such as the USA is critical to develop and demonstrate the new potential of an AI-based algorithms in personalizing medical or cosmetic regimen. A fully remote and online design makes it possible for the dataset to acquire a representative sample size of the US population.

In addition to its intrinsic results, the study aimed at demonstrating the accuracy of an AI-based automated facial descriptor in real-life

TABLE 5 Estimation of the performance (as %) in the gradings of seven categorical facial attributes obtained by the original and the new automatic descriptor systems and the 24 experts. Results on all validation dataset, that is, by age-classes, phototypes and ancestries. African American (AA), East Asian (EA), Non-Hispanic Euro-American (NHEA), Hispanic Euro-American (HEA).

Facial attributes	Original automatic descriptor						New and inclusive automatic descriptor						Make-up experts					
Global Panel	$N = 520$						$N = 520$						$N = 520$					
1. Skin tone complexion	60.7						74.0						56.6					
2. Undertone	52.0						64.5						55.4					
6. Eyebrows' shape	69.8						71.7						55.2					
10. Eyes color	50.5						69.6						63.2					
17. Head hair color	45.1						69.7						61.5					
18. Hair form	53.9						74.8						67.5					
19. Hair length	31.2						72.4						63.9					
Age	**18y–29y**	**30y–39y**	**40y–49y**	**50y–64y**	**65y–80y**	**NA**	**18y–29y**	**30y–39y**	**40y–49y**	**50y–64y**	**65y–80y**	**NA**	**18y–29y**	**30y–39y**	**40y–49y**	**50y–64y**	**65y–80y**	**NA**
1. Skin tone complexion	57.3	61.5	56.8	69.0	59.3	NA	75.8	72.3	71.2	77.0	74.1	NA	56.9	57.4	56.7	57.4	53.5	NA
2. Undertone	52.4	52.3	48.6	34.0	24.1	NA	71.0	67.7	63.1	58.0	57.4	NA	51.2	57.9	57.1	56.1	55.2	NA
6. Eyebrows' shape	73.4	69.2	64.9	61.0	66.7	NA	85.4	87.7	93.7	90.0	87.0	NA	68.3	64.0	61.6	67.7	69.6	NA
10. Eyes color	42.7	46.9	40.5	34.0	35.2	NA	69.3	70.8	73.9	67.0	63.0	NA	65.3	66.5	62.2	59.7	56.9	NA
17. Head hair color	47.6	52.3	46.8	51.0	59.3	NA	70.2	80.0	70.3	62.0	57.4	NA	62.3	63.2	59.4	62.4	52.5	NA
18. Hair form	59.7	53.1	49.5	50.0	59.3	NA	79.8	80.0	69.4	70.0	70.4	NA	66.8	66.5	62.9	65.5	66.7	NA
19. Hair length	34.7	24.6	28.8	30.0	46.3	NA	75.0	70.0	79.3	65.0	72.2	NA	63.6	64.4	65.6	59.2	63.6	NA
Photo-types	**I**	**II**	**III**	**IV**	**V**	**VI**	**I**	**II**	**III**	**IV**	**V**	**VI**	**I**	**II**	**III**	**IV**	**V**	**VI**
1. Skin tone complexion	66.0	79.3	57.6	50.8	50.0	58.3	80.8	81.9	72.2	67.4	64.3	91.7	53.4	59.1	54.0	54.5	61.8	68.1

(Continues)

TABLE 5 (Continued)

Photo-types	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
2. Undertone	38.3	42.2	40.3	47.0	60.7	54.2	66.0	66.4	61.1	58.3	73.2	87.5	59.4	56.2	56.5	54.3	53.1	50.3
6. Eyebrows' shape	61.7	69.8	69.4	67.4	69.6	50.0	85.1	90.5	91.0	87.1	85.7	91.7	68.4	61.0	69.8	65.3	68.7	55.6
10. Eyes color	42.5	56.0	40.3	38.6	28.6	8.3	59.6	65.5	67.4	71.2	82.1	83.3	50.9	59.0	61.7	67.5	71.4	68.7
17. Head hairs color	29.8	37.1	50.0	53.8	75.0	83.3	66.0	75.0	66.0	72.0	66.1	70.9	50.5	56.9	62.7	62.4	65.3	71.5
18. Hair form	53.2	63.8	50.0	50.0	60.7	37.5	72.3	75.0	75.0	77.3	75.0	62.5	69.7	66.9	64.2	63.3	68.1	67.0
19. Hair length	36.2	31.0	34.0	28.8	28.6	25.0	78.7	72.4	78.5	71.2	60.7	58.3	62.1	65.6	63.8	62.2	63.2	59.4
Ancestries	**A A**	**E A**	**NH EA**	**H EA**	**NA**	**NA**	**A A**	**E A**	**NH EA**	**H EA**	**NA**	**NA**	**A A**	**E A**	**NH EA**	**H EA**	**NA**	**NA**
1. Skin tone complexion	48.6	49.5	75.0	65.1	NA	NA	67.9	77.8	80.1	68.0	NA	NA	56.6	56.3	56.3	56.6	NA	NA
2. Undertone	25.0	59.8	35.3	50.8	NA	NA	66.4	65.1	65.4	59.8	NA	NA	55.4	55.0	54.4	55.0	NA	NA
6. Eyebrows' shape	65.0	71.1	64.7	70.6	NA	NA	74.3	63.9	71.8	73.0	NA	NA	65.2	65.2	65.2	65.3	NA	NA
10. Eyes color	65.7	68.0	30.1	45.2	NA	NA	72.9	69.8	57.0	84.5	NA	NA	63.2	62.6	64.6	64.4	NA	NA
17. Head hair color	44.3	61.9	35.9	44.4	NA	NA	62.1	65.9	75.6	76.3	NA	NA	61.5	61.6	62.8	62.1	NA	NA
18. Hair form	45.7	70.1	54.5	50.0	NA	NA	65.0	68.2	83.3	83.5	NA	NA	67.5	68.2	67.7	67.8	NA	NA
19. Hair length	30.0	24.7	35.3	32.5	NA	NA	61.4	73.0	82.7	71.1	NA	NA	63.9	63.5	63.4	63.5	NA	NA

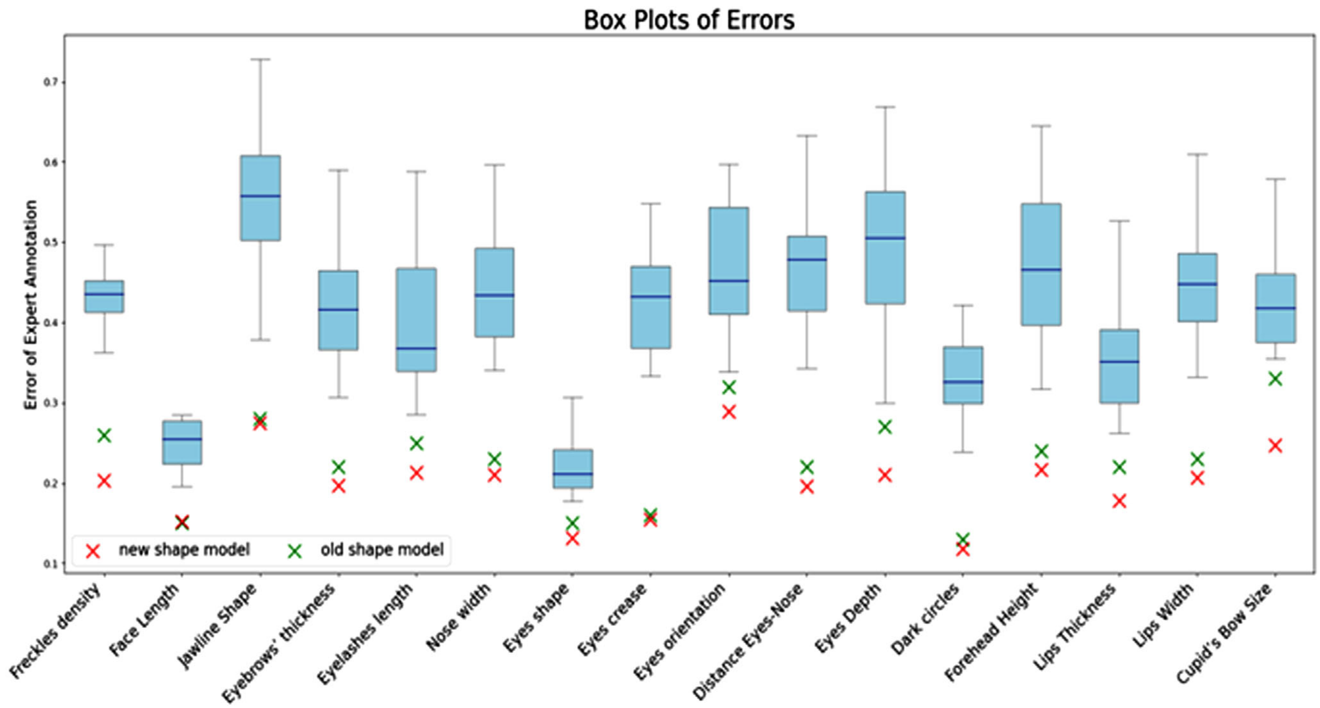

FIGURE 2 Box plots graph showing the error of expert annotations in 16 continuous facial attributes. Red cross represents values of errors of the new AI-automatic descriptor and the average of all experts scores, taken as reference, for each facial attribute and each selfie image. Green cross describes values of errors for the original AI-based automatic descriptor. Blue square shows the error of each individual expert versus all experts scores, taken as a reference, and its variation.

without any specific instructions in shooting selfie pictures, such as lighting conditions, time of the day, location, distance. This emphasizes the drop in a risk to cause algorithm bias, approximations, or errors related to the aforementioned variables. Compared to the original system, the new AI-based automatic descriptor was found to be more accurate for both continuous and categorical attributes and performed twice better than the assigned experts in this study.

The creation and the definition of an inclusive dataset previously used in clinical grading[8] not only tremendously improved the AI-based automated descriptor but also demonstrated its versatility with regard to different age groups, skin phototypes and different ancestries.

Using the new dataset greatly benefited and updated the grading of different continuous attributes. Categorical signs, however, only showed significant improvements in accuracy. It is worthwhile to note that categorical variables were described as such in this investigation due to the existing familiarity of make-up experts with conventional categorical descriptors of traits such as eye color, and so forth. rather than metrical characterization of these traits.

With a well-designed dataset, the improvements of the original system show no usual limitations for continuous signs, as previously discussed,[17–20] with only two attributes (#7,20) with higher MAE in phototype VI and four attributes (#7,15,22,23) in the elder group. In general, a lower recognition in the upper and lower ends of the criterion: darker phototypes in skin color and younger/older groups in the age-class could be attributed to demographic factors that influenced the algorithm leading to a bias in under-represented demographic groups.[17–20]

This highlights the importance of using well distributed datasets across all demographics from the algorithmic evaluation. Clearly, the different training datasets used in the creation on new AI-based automated facial descriptor[8,33,39–41] fulfill these requirements by covering different geographical locations (USA, Brazil, India, China, Korea, Japan, South Africa, and France) with different ages, phototypes, gender and ancestries. To continue to enrich the system, new datasets of selfie images are currently collected in Mexico, Indonesia, or Nigeria.

One limitation in assessing accuracy likely grounds on the use of Fitzpatrick scale, the currently accepted classification for skin phototypes. However, the shortcomings of the scale with respect to the classification of darker skins, including the lower reliability for phototypes V and VI, are widely recognized.[42–44] With new proposition(s) for an improved classification for skin color,[45] or nomad connected tool usable at home,[46] new standards will raise soon, pinpointing the advantage of the present US dataset lies in its capacity to be re-analyzed at the light of new scales to improve performance and results of our automated system.

With combination of this new AI-based automated facial attributes descriptor and previous AI-based automated facial signs grading system[39–41] more than 40 facial features could be assessed from a selfie image with high accuracy and coherence, taking dermatologists and make-up experts' assessments as references.

The validation of the accuracy of the system in all ages, phototypes or ancestries is a pathway to offer to everyone (and anywhere) the most accurate diagnostic or answer to personalized treatment, esthetic procedures, or routine. Indeed, facial appearance can hardly be restricted to a sole criterion as it necessarily integrates a sum of many other facial features, making facial appearance a multi parametric balance. These two systems seem paving the road toward large investigations[47] to better understand patients' or consumers' specificities including their intimate aspirations in personalizing their global facial look,[48] in turn improving their well-being and self-confidence. Such desire is especially the case of "boomers", where changing hair color and facial features with unbalanced facial contrasts[31] mainly due to changes (thinner lips, sparser eyebrows, etc.) in key horizontal traits. Relevant and accurate automatic grading on facial features, applied to videos, becomes critical when assessing dynamic behavior[49] when faces are in motion[50] as well as emotions.

This study investigated the use of an AI-based automated facial descriptor to ascertain its accuracy and validate its functionality toward facial attributes devoted to the makeup and skincare routines of women. It was reported that when 16 continuous facial traits were assessed, the AI based system delivered highly accurate and consistent gradings throughout for all age-classes, ancestries and phototypes. When categorical facial attributes were assessed, the algorithm performed better and exceeded the grading given by the experts. While some further finetuning is recommended for this AI, it is overall accurate and relevant for the analysis of facial attributes across a diverse population of women.

ACKNOWLEDGMENTS

Authors would like to thank Mr Didier Saint-Leger, Mr Marc Toulemonde, Mr Guive Balooch, Mrs Laurence Lebarbanchon, Mr Stephane Ortiz, Mr Stephane Lannuzel, Mrs Chiara Schaad and Mrs Margaret Jonhston-Clarke, not only for their invaluable help in completion of this paper and global study, but also for their strong support and enthusiasm. All costs of the present work were entirely met by L'Oréal Research & Innovation Department.

CONFLICT OF INTEREST STATEMENT

Y.Z., F.F., R.J., J.H., C.K., C.S, and P.A. are employees of L'Oréal group. A.J. has received honoraria as member of the Global Diversity, Equity, and Inclusion Board of L'Oréal. N.G.J. has received honoraria as member of the Scientific Advisory Board and of the Global Diversity, Equity, and Inclusion Board of L'Oréal.

DATA AVAILABILITY STATEMENT

The data presented in this study are available on request from the corresponding author upon reasonable request.

INSTITUTIONAL REVIEW BOARD STATEMENT

This study is a non-interventional consumer study where each adult respondent was a member of an online panel that had previously agreed to be emailed surveys as part of their panel membership.

INFORMED CONSENT STATEMENT

Informed consent was acquired beforehand for the images collected from the participants. The participants waivered their right to the image they sent and were remunerated accordingly.

ORCID

Frederic Flament https://orcid.org/0000-0003-1128-9535
Diala Haykal https://orcid.org/0000-0001-7528-5088

REFERENCES

1. Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W. Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatol Ther (Heidelb)*. 2020;10(3):365-386.
2. Elder A, Ring C, Heitmiller K, Gabriel Z, Saedi N. The role of artificial intelligence in cosmetic dermatology—current, upcoming, and future trends. *J Cosmet Dermatol*. 2021;20(1):48-52.
3. Young AT, Xiong M, Pfau J, Keiser MJ, Wei ML. Artificial intelligence in dermatology: a primer. *J Invest Dermatol*. 2020;140(8):1504-1512.
4. Han SS, Park I, Chang SE, et al. Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. *J Invest Dermatol*. 2020;140(9):1753-1761.
5. Jain A, Way D, Gupta V, et al. Development and assessment of an artificial intelligence-based tool for skin condition diagnosis by primary care physicians and nurse practitioners in tele-dermatology practices. *JAMA Netw Open*. 2021;4(4):e217249.
6. Liu Y, Jain A, Eng C, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med*. 2020;26(6):900-908.
7. Flament F, Jacquet L, Ye C, et al. Artificial Intelligence analysis of over half a million European and Chinese women reveals striking differences in the facial skin ageing process. *J Eur Acad Dermatol Venereol*. 2022;36(7):1136-1142.
8. Flament F, Jiang R, Houghton J, et al. Accuracy and clinical relevance of an automated, algorithm-based analysis of facial signs from selfie images of women in the United States of various ages, ancestries and phototypes: A cross-sectional observational study. *J Eur Acad Dermatol Venereol*. 2023;37(1):176-183. doi:10.1111/jdv.18541
9. Jablonski NG. Skin color and race. *Am J Phys Anthropol*. 2021;175(2):437-447.
10. Jablonski NG. Why humans come in colors. *Anthro Notes*. 2011;32:7-10.
11. Jablonski NG, Chaplin G. Epidermal pigmentation in the human lineage is an adaptation to ultraviolet radiation. *J Hum Evol*. 2013;65(5):671-675
12. Chaplin G, Jablonski NG. Vitamin D and the evolution of human pigmentation. *Am J Phys Anthropol*. 2009;139(4):451-461.
13. Jablonski NG, Chaplin G. The evolution of human skin coloration. *J Hum Evol*. 2000;39:57-106.
14. Del Bino S, Ito S, Sok J, et al. Chemical analysis of constitutive pigmentation of human epidermis reveals constant eumelanin to pheomelanin ratio. *Pigment Cell Melanoma Res*. 2015;28(6):707-717.
15. Saint-Leger D. The colour of the human skin: fruitful research, unsuitable wordings. *Int J Cosmet Sci*. 2015;37(3):259-265.
16. de Rigal J, Mazis ID, Diridollou S, et al. The effect of age on skin color and color heterogeneity in four ethnic groups. *Skin Res Technol*. 2010;16(2):168-178.
17. Klare BF, Burge MJ, Klontz JC, Vorder Bruegge RW, Jain AK. Face recognition performance: role of demographic information. *IEEE Trans Inf Forensics Secur*. 2012;7(6):1789-1801.
18. Drozdowski P, Rathgeb C, Dantcheva A, Damer N, Busch C. Demographic bias in biometrics: a survey on an emerging challenge. *IEEE Trans Technol Soc*. 2020;1(2):89-103.
19. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classsification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR. 2018;81:77-91.
20. Kinyanjui NM, Odonga T, Cintas C, et al. Fairness of classifiers across skin tones in dermatology. In: Martel AL, ed. *Medical Image Computing and Computer Assisted Intervention e MICCAI 2020. Lecture Notes in Com-*

puter Science. Cham: Springer; 2020;12266. doi:10.1007/978-3-030-59725-2_31
21. Porcheron A, Mauger E, Russell R. Aspects of facial contrast decrease with age and are cues for age perception. *PLoS One*. 2013;8(3):e57985.
22. Porcheron A, Mauger E, Soppelsa F, et al. Facial contrast is a cross-cultural cue for perceiving age. *Front Psychol*. 2017;8:1208.
23. Flament F, Abric A, Adam AS. Evaluating the respective weights of some facial signs on perceived ages in differently aged women of five ethnic origins. *J Cosmet Dermatol*. 2021;20(3):842-853.
24. Batres C, Porcheron A, Latreille J, Roche M, Morizot F, Russell R. Cosmetics increase skin evenness: evidence from perceptual and physical measures. *Skin Res Technol*. 2019;25(5):672-676.
25. Flament F, Ye C, Mercurio DG, et al. Evaluating the respective weights of some facial signs on the perceived radiance/glow in differently aged women of six countries. *Skin Res Technol*. 2021;27(6):1116-1127.
26. Shaiek A, Flament F, François G, et al. Morphological criteria of feminine upper eyelashes, quantified by a new semi-automatized image analysis: application to the objective assessment of mascaras. *Skin Res Technol*. 2018;24:135-144.
27. Flament F, Seyrek I, François G, Zhu T, Ye C, Saint-Leger D. Morphometric characteristics of feminine eyebrows: variations with ethnicities and age. *Int J Cosmet Sci*. 2018;41:443-449.
28. Flament F, François G, Seyrek I, Saint-Leger D. Age-related changes to characteristics of the human eyes in women from six different ethnicities. *Skin Res Technol*. 2020;26:520-528.
29. de Rigal J, Des Mazis I, Diridollou S, et al. The effect of age on skin color and color heterogeneity in four ethnic groups. *Skin Res Technol*. 2010;16(2):168-178.
30. Russell R, Porcheron A, Sweda JR, Jones AL, Mauger E, Morizot F. Facial contrast is a cue for perceiving health from the face. *J Exp Psychol Hum Percept Perform*. 2016;42(9):1354-1362.
31. Russell R, Batres C, Courrèges S et al. Differential effects of makeup on perceived age. *Br J Psychol*. 2019;110(1):87-100.
32. Caisey L, Grangeat F, Lemasson A, Talabot J, Voirin A. Skin color and makeup strategies of women from different ethnic groups. *Int J Cosmet Sci*. 2006;28(6):427-437.
33. Flament F, Zhang Y, Yu Z, et al. Developing an artificial intelligence (A.I)-based descriptor of facial appearance that fits with the assessments of makeup experts. *Skin Res Technol*. 2021;27(6):1081-1091.
34. Fitzpatrick TB. The validity and practicability of sun-reactive skin type I through VI. *Arch Dermatol*. 1988;124:869-871.
35. Han H, Jain AK, Wang F, Shan S, Chen X. Heterogeneous face attribute estimation: a deep multi-task learning approach. *IEEE Trans Pattern Anal Mach Intell*. 2017;40(11):2597-2609.
36. Rothe R, Radu T, Van Gool L. Deep expectation of real and apparent age from a single image without facial landmarks. *Int J Comput Vis*. 2018;126(2):144-157.
37. Wang F, Han H, Shan S, Chen X. Deep Multi-Task Learning for Joint Prediction of Heterogeneous Face Attributes. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. 2017;173-179.
38. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016;770-778. doi:10.1109/CVPR.2016.90
39. Jiang R, Kezele I, Levinshtein A, et al. A new procedure, free from human assessment, that automatically grades some facial skin structural signs. Comparison with assessments by experts, using referential atlases of skin ageing. *Int J Cosmet Sci*. 2019;41(1):67-78.
40. Zhang Y, Jiang R, Kezele I, et al. A new procedure, free from human assessment, that automatically grades some facial skin signs in men from selfies pictures. Application to changes induced by a severe aerial chronic urban pollution. *Int J Cosmet Sci*. 2020;42(2):185-197.
41. Flament F, Lee YW, Lee DH, et al. The continuous development of a complete and objective automatic grading system of facial signs from selfie pictures: Asian validation study and application to women of three ethnic origins, differently aged. *Skin Res Technol*. 2021;27(2):183-190.
42. He SY, McCulloch CE, Boscardin WJ, Chren MM, Linos E, Arron ST. Self-reported pigmentary phenotypes and race are significant but incomplete predictors of Fitzpatrick skin phototype in an ethnically diverse population. *JAMA Dermatol*. 2014;71(4):731-737.
43. Ware OR, Dawson JE, Shinohara MM, Taylor SC. Racial limitations of fitzpatrick skin type. *Cutis*. 2020;105(2):77-80.
44. Khosla NN, Grullon K, Rosenblatt AE. Prevention of racialized medicine in pediatric dermatology: a call to re-examine skin tone typing. *Pediatr Dermatol*. 2021;38:167-169.
45. Dadzie OE, Sturm RA, Fajuyigbe D, Petit A, Jablonski NG. The eumelanin human skin colour scale: a proof-of-concept study. *Br J Dermatol*. 2022;187(1):99-104.
46. Gougeon S, Hernandez E, Chevrot N, et al. Evaluation of a new connected portable camera for the analysis of skin microrelief and the assessment of the effect of skin moisturisers. *Skin Res Technol*. 2022;29(1):e13190.
47. Goh CL. Will digitization, big data and artificial intelligence- and deep learning-based algorithm govern the practice of medicine? *J Eur Acad Dermatol Venereol*. 36(7):947 (2022).
48. Batres C, Porcheron A, Courrèges S, Russell R. Professional versus self-applied makeup: do makeup artists add value? *Perception*. 2021;50(8):709-719.
49. Flament F, Bazin R, Piot B. Influence of gravity upon some facial signs. *Int J Cosmet Sci*. 2015;37(3):291-297.
50. Kurosumi M, Mizukoshi K, Hongo M, Kamachi MG. Does age-dynamic movement accelerate facial age impression? Perception of age from facial movement: studies of Japanese women. *PLoS One*. 2021;16(8):e0255570.

How to cite this article: Yu Z, Flament F, Jiang R, Houghton J, Kroely C, Cabut N, et al. The relevance and accuracy of an AI algorithm-based descriptor on 23 facial attributes in a diverse female US population. *Skin Res Technol*. 2024;30:e13690. https://doi.org/10.1111/srt.13690