# scientific data

DATA DESCRIPTOR

Check for updates

## **OPEN** The International Skin Spectra Archive (ISSA): a multicultural human skin phenotype and colour spectra collection

Yan Lu<sup>1</sup>, Kaida Xiao <sup>1</sup><sup>M</sup>, Michael Pointer<sup>1</sup>, Ruili He<sup>1</sup>, Sicong Zhou<sup>1</sup>, Ahmed Nasseraldin<sup>2</sup>, Suchitra Sueeprasan<sup>3</sup>, Cheng Gao<sup>4</sup>, Changjun Li<sup>4</sup>, Ali Sohaib<sup>5</sup>, Yuanyuan He<sup>6</sup>, Yoko Mizokami<sup>7</sup>, Mengmeng Wang<sup>8</sup>, Esther Perales Romero<sup>9</sup>, Faraedon Zardawi<sup>10</sup>, Lesley Gill<sup>11</sup>, Ningfang Liao<sup>12</sup> & Sophie Wuerger<sup>13</sup>

This paper presents the International Skin Spectra Archive (ISSA), a multicultural human skin phenotype dataset, containing 15,256 records of both spectral and colorimetric data derived from 2,113 subjects. These measurements, collected between 2012 and 2024, come from eleven different datasets gathered by international laboratories across eight countries, all adhering to a uniform measurement protocol to ensure data consistency. The ISSA dataset addresses the inherent challenges in measuring human skin colour due to its complex structure and covers a wide variability in skin characteristics such as geography, ethnicity, age, gender, and body location. Providing a broad spectrum of human skin data, the ISSA dataset will advance our understanding of skin colour variations and their biological, cultural, and environmental influences. It will also serve as a crucial resource for scientific research and technological development across various fields where diverse and precise spectral and colour data of real human skin are essential.

### Background & Summary

Measuring the spectral properties of human skin and its variations has been fundamental to research in vision and evolution. As a multi-layered organ, human skin derives its colour mainly from the absorption characteristics of melanin in the epidermis and haemoglobin in the blood<sup>1</sup>. Melanin has adapted to balance protection from ultraviolet radiation with vitamin D synthesis, exhibiting colours from tan to white<sup>2</sup>. Haemoglobin, the red chromophore, introduces the distinctive 'W' shape in the skin reflectance spectrum<sup>3</sup>. This critical feature is hypothesised to influence the development of primate trichromacy for enhancing skin discrimination<sup>4</sup>. Skin colour is a highly variable human trait, with its evolution and genetic bases being complex and far from well-understood. Recent genetic studies have underscored the need for objective measurement data on diverse skin phenotypes to help understand this culturally and biologically important trait<sup>5</sup>.

The spectral composition of human skin determines its colour appearance perceived under various viewing conditions. Thus, reflectance measurements are crucial for assessing the real colour of human skin and reproducing its appearance across multiple fields, particularly in merging facial technologies. For instance, our familiarity with skin colour necessitates precise rendering and reproduction in digital imaging such as computer graphics, photography, cinematography, virtual reality (VR) and augmented reality (AR)<sup>6-9</sup>. Advances in 3D printing technology have highlighted the importance of accurate skin colour reproduction for creating skin

<sup>1</sup>School of Design, University of Leeds, Leeds, UK. <sup>2</sup>School of Design, Dar Al-Hekma University, Jeddah, Saudi Arabia. <sup>3</sup>Faculty of Science, Chulalongkorn University, Bangkok, Thailand. <sup>4</sup>School of Computing, University of Science and Technology Liaoning, Anshan, China. <sup>5</sup>Nordson Corp, Dunstable, UK. <sup>6</sup>School of Engineering, Utsunomiya University, Utsunomiya, Japan. <sup>7</sup>Graduate School of Informatics, Chiba University, Chiba, Japan. <sup>8</sup>School of Digital Technology and Innovation Design, Jiangnan University, Wuxi, China. <sup>9</sup>Department of Optics, University of Alicante, Alicante, Spain. <sup>10</sup>School of Dentistry, University of Sulaimani, Sulaimani, Iraq. <sup>11</sup>School of Healthcare Science, Manchester Metropolitan University, Manchester, UK. <sup>12</sup>National Key Lab of Colour Science and Engineering, Beijing Institute of Technology, Beijing, China. <sup>13</sup>School of Psychology, University of Liverpool, Liverpool, UK. <sup>Se</sup>e-mail: k.xiao1@leeds.ac.uk

prostheses<sup>10,11</sup>. In medical contexts, precise spectral data is vital for analysing skin chromophores to detect and diagnose various skin diseases and other health issues<sup>12-15</sup>. Such data also supports skin assessment, skincare product development, and preference identification in the cosmetics, personal healthcare, and medical beauty industries<sup>16–19</sup>. Furthermore, more inclusive skin datasets are necessary to enhance face recognition and biometric authentication technologies, particularly in data-driven artificial intelligence (AI) algorithms<sup>20–23</sup>.

While skin spectra data provide vital references for the aforementioned research and practical applications, achieving accurate measurements is challenging due to the uneven, multi-layered structure of the skin and the broad variability it exhibits. Measurement techniques and protocols, such as instrument type, aperture size, measurement distance, and pressure applied during measurement, can introduce considerable variation in results<sup>24</sup>. Additionally, the appearance of human skin varies widely affected by ethnicity, gender, age, body location, and even factors like diet, and health<sup>19,25</sup>. A comprehensive, comparable skin spectra dataset should account for the essential variations and adhere to a uniform protocol, with well-documented notes on measurement techniques and detailed subject information.

Previous efforts on collecting spectral datasets have focused on daylight and various natural objects, such as plants, with limited resources dedicated to human skin<sup>26,27</sup>. In 2003, the International Organisation for Standardization (ISO) released the standard '*Graphic technology - Standard object colour spectra database for colour reproduction evaluation (SOCS)*', which included 8,213 human skin reflectance from six datasets: SHISEIDO, KAO, OOKA, KAWASAKI, OULU and SUN<sup>28</sup>. These reflectance data were primarily gathered from Japanese and Caucasian participants using different instrument types and measurement geometry, complicating comparisons between different datasets. Over the past two decades, several other skin datasets have been collected to meet varying research and industry needs<sup>29–33</sup>. However, none adequately represent demographic variations due to limited sample sizes or are easily integrated due to differing measurement methods. Furthermore, most of these datasets are not freely accessible to the public.

In 2013, the International Commission on Illumination (CIE) initiated the Technical Committee TC 1–92, Measurement of Human Skin Colour, to investigate the measurement uncertainty and recommend protocols for good measurement practice. Guided by this committee, we collected eleven high-quality skin spectra datasets between 2012 and 2024 at thirteen global laboratories across eight countries: the UK, Spain, China, Japan, Pakistan, Thailand, Iraq, and Saudi Arabia. This effort led to the establishment of a new spectra dataset for human skin - the International Skin Spectra Archive (ISSA). The purpose of the dataset was to provide a publicly accessible resource of diverse human skin spectra data, supporting both multidisciplinary research and practical applications. To overcome the limitations of previous datasets and enhance measurement accuracy and consistency, all data within the ISSA were collected using spectrophotometers following a standardised measurement protocol, allowing objective comparisons across datasets. The variability using different measurement parameters such as measurement aperture size and pressure and the measurement repeatability was evaluated<sup>24</sup> (see Technical Validation). Figure 1 provides a visual overview of the measurement instruments, the measuring process, the body locations, and the spectral reflectance across all skin measurements in the dataset.

The International Skin Spectra Archive (ISSA), comprises spectral data from 15,256 skin samples collected from 2,113 subjects. Each sample is meticulously documented in a single dataset record. These records detail the record number, origin, subject number, and skin type, including the subject's ethnicity, gender, age, and body location. Additionally, they provide comprehensive instrument information such as type, inclusion of specular components, wavelength range, and interval. Alongside spectral data, the dataset also includes CIE colorimetric data for each sample. These measurements are based on the CIE 1931 standard colorimetric observer and the CIE standard illuminant D65 to simulate skin colour in daylight conditions. This illuminant provides a consistent and universally comparable lighting scenario that closely resembles average daylight, ensuring that our colorimetric assessments are both reliable and replicable across all samples.

The spectral data from the ISSA offers extensive resources and references crucial for various applications, such as reflectance reconstruction, medical diagnosis, and training datasets for algorithms. Users can access specific groups of data tailored to particular needs, such as data of certain ethnic groups, genders, age groups, and body locations. Figures 2, 3 give examples to illustrate the varied human skin spectral reflectance across nationality or ethnicity and gender. These examples demonstrate the variability and specific patterns observable in the spectral data, emphasising the importance of considering these factors in skin colour analysis and research. The dataset also includes colorimetric data, enabling the direct assessment of skin colour appearance under standard D65 daylight illumination and CIE standard observer conditions. Additionally, colour appearance under different viewing conditions can be simulated using spectral data and any self-defined light sources and observer's colour matching functions. The diversity and unique pattern highlighted in Fig. 1d illustrate the ISSA's comprehensive coverage, providing valuable insights for research in skin physiology, genetics, and anthropology, and supporting innovation across a range of scientific and technological disciplines.

#### Methods

**Data collection.** Skin colour measurements were systematically obtained from thirteen global locations across eight countries as detailed in Table 1. The collection encompasses diverse ethnic groups across eleven datasets:

- Dataset 1: Comprises data from Caucasian subjects measured at the University of Sheffield and Manchester Metropolitan University in the United Kingdom.
- Dataset 2: Contains measurements of Chinese skin, conducted at institutions including the Beijing Institute of Technology, Minzu University of China, and Beijing Institute of Graphic Communication.



**Fig. 1** An overview of measurement tools and processes for skin spectra data collection (**a**) Measurement instruments displayed: on the left, the CM 700 d spectrophotometer; on the right, the CM 2600 d spectrophotometer. (**b**) The measurement process illustrated with a spectrophotometer positioned to measure the forehead of a subject. (**c**) Display of 12 standard body locations used for skin measurements. (**d**) Spectral reflectance across all skin measurements (N = 15,256) included in the dataset. Each dotted line represents the spectral reflectance from a specific body location of an individual subject, with the colour of each line providing a visual approximation of the skin colour; the solid red line indicates the average spectral reflectance of human skin.

- .....
- Dataset 3: Includes measurements from Middle Eastern (Iraqi) individuals, undertaken at the University of Sulaimani in Iraq.
- Dataset 4: Southeast Asian (Thai) data gathered from Chulalongkorn University in Thailand.
- Dataset 5: South Asian (Pakistani) measurements taken at COMSATS University Islamabad in Pakistan.
- Dataset 6: Comprised of Caucasian samples, measured at University of Alicante in Spain.
- Dataset 7: A mixed collection involving Chinese, Caucasian, and South Asian subjects, conducted at University of Leeds in the United Kingdom.
- Dataset 8: African samples acquired at University of Leeds in the United Kingdom.
- Dataset 9: Japanese skin measurements obtained from Chiba University in Japan.
- Dataset 10: Chinese samples measured at the University of Science and Technology Liaoning in China.
- Dataset 11: Arabian skin measurements taken at Dar Al-Hekma University in Saudi Arabia.

Ethical approval for this study was obtained from the following committees: the University Research Ethics Committee (School of Clinical Dentistry), The University of Sheffield (Reference Number: 9); the IPHS Research Ethics Committee, University of Liverpool (Reference Number: IPHS-1314-278); the PVAC & Arts joint Faculty Research Ethics Committee, University of Leeds (Reference Number: PVAR 13-057) and the Faculty of Arts, Humanities and Cultures Research Ethics Committee, University of Leeds (Reference Number: PVAR 13-057) and the Faculty of Arts, Humanities and Cultures Research Ethics Committee, University of Leeds (Reference Number: LTDESN-067). Additional datasets were collected under the ratification of host institutions, which accepted the ethical approval from the University of Sheffield, Liverpool or Leeds without requiring additional approvals. For Dataset 4 (Chulalongkorn University, Thailand), the study complied with the university's policy in 2014, which exempted research that did not link data to individual participants. All other datasets were approved by their respective institutional ethics committees, including the Ethics Committee of the University of Science & Technology Liaoning and the Research Ethics Committee of the Scientific Research Centre at Dar Al-Hekma University. All participants across all collection sites provided written informed consent, which included permission to share anonymised data for research purposes.

Note that subjects under age 18 were excluded from the data collection; only healthy skin without visible abnormalities was included in the dataset. Despite its comprehensive scope, the current dataset does not fully represent all segments of the global population. For example, it lacks measurements from Africans living in



**Fig. 2** Spectral reflectance differences of human skin for the eight ethnic groups: CA (Caucasian), CN (Chinese), SA (South Asian - Pakistani), AF (African), IQ (Middle Eastern - Iraqi), TH (Southeast Asian - Thai), JP (Japanese), AB (Middle Eastern - Arabian). Each dotted line represents the spectral reflectance from a specific body location of an individual subject, with the colour of each line providing a visual approximation of the skin colour; the solid red line indicates the average spectral reflectance.



**Fig. 3** Spectral reflectance differences of female and male skin for the eight ethnic groups: CA (Caucasian), CN (Chinese), SA (South Asian - Pakistani), AF (African), IQ (Middle Eastern - Iraqi), TH (Southeast Asian - Thai), JP (Japanese), AB (Middle Eastern - Arabian). The solid lines represent the average spectral reflectance for female (blue line) and male (orange line), and the shaded areas denote  $\pm 2$  standard error.

Dataset	Year(s)	Location	Nationality or ethnicity Group(s)	Method(s)	Subjects	Body Locations
1	2012-2014	UK	Caucasian	SP	172	9
2	2013	China	Chinese	SP	201	9
3	2013	Iraq	Middle Eastern (Iraqi)	SP	149	9
4	2014	Thailand	Southeast Asian (Thai)	SP	710	6
5	2015	Pakistan	South Asian (Pakistani)	SP	117	10
6	2016	Spain	Caucasian	SP	27	3
7	2017-2018	UK	Chinese, Caucasian, South Asian	SP	249 <sup>a</sup>	4
8	2019-2023	UK	African	SP	147	4-10
9	2023-2024	Japan	Japanese	SP	118	9
10	2024	China	Chinese	SP	116	7
11	2024	Saudi Arabia	Middle Eastern (Arabian)	SP	107	9

Table 1. Description of the eleven sample datasets included in the ISSA: SP – Spectrophotometer. <sup>a</sup>including155 Caucasian, 53 Chinese, and 41 South Asian.

.....

Africa and individuals of mixed ethnicity. We acknowledge these limitations and intend to collect more data to update this dataset further, aiming to address these gaps and enhance its inclusivity and representativeness.

Principles of skin colour measurement. For over 70 years, CIE colorimetry has been successfully applied for the objective measurement of colour<sup>34</sup>. The principles of skin colour measurement in this study adhere strictly to the CIE colourimetry standards. The CIE system of colorimetry is widely used for the calculation of appropriate colour-related parameters, for example CIELAB coordinates, from spectral reflectance measurements of surfaces using spectrophotometers with the assumption of a CIE standard observer, and under standard illuminant (e.g., D65, which represents average daylight with a correlated colour temperature of 6504 K)<sup>35</sup>. This standard illuminant is crucial as it provides a consistent basis for colour evaluation, mimicking natural light to ensure that colour measurements are comparable irrespective of the actual lighting conditions under which they are taken. Both skin colour appearance and skin colour differences can be further predicted in the CIELAB uniform colour space. Although CIELAB coordinates are extensively used in industry to assess the 'colour' of a sample, they are typically computed under the standard daylight illuminant and, once calculated it is not easy to re-calculate values for a second, different illuminant<sup>36</sup>. To accommodate applications that require assessments under multiple lighting conditions, the ISSA provides spectral data which can be used to calculate appropriate colorimetric coordinates under any illuminant as long as it is defined by a relative spectral power distribution, either tabulated, as in the case of CIE standard illuminants or measured with an appropriate instrument. This flexibility is crucial not only for accurately simulating skin appearance under different lighting and measuring colour discrimination thresholds but also for extracting detailed information about skin characteristics and chromophores, which is important for various technological and medical applications.

**Measurement instruments.** Portable spectrophotometers (SP) were employed for all skin spectral reflectance measurements. An SPM measures the spectral reflectance factor, defined as the ratio of the radiant flux reflected by a sample to that of a perfect diffuser, normally a calibrated white ceramic tile. Each instrument incorporates a diffused light source, ensuring that measurements remain consistent across conditions. This method is independent of the spectral power distribution and does not account for fluorescence effects. The SPM's aperture directly contacts the sample, eliminating ambient light interference, as demonstrated in Fig. 1b where a portable spectrophotometer is used on a subject's forehead. Also, the instrument uses a pulsed xenon lamp with a UV cut-off filter such that the measurement time is extremely short at approximately 1 second. This should serve to minimise the possible effects of movement during the measurement period bearing in mind that the instrument is being held at the required location on the subject's body. The instrument is also able to include or exclude the specular component of the reflected light. In this dataset, the specular component of the light reflected from the sample was included in the measurements.

In this study, Datasets 1, 2, 3, 6, and 10 were measured with a Konica Minolta CM-2600d (Tokyo, Japan). Datasets 5, 7, 8, 9, and 11 were measured with a Konica Minolta CM-700d (Tokyo, Japan). Dataset 4 was measured with an X-Rite SP62 (Michigan, US). All the devices adhered to the CIE di:8° geometry standards with diffuse illumination and an 8-degree detection angle<sup>34</sup>. Measurements were taken in the specular component included (SCI) mode with an aperture size of either 3 mm or 8 mm. Most skin reflectance data span the visible spectral range from 360 nm to 740 nm, although some datasets are confined to a narrower range of 400 nm to 700 nm, all recorded at 10 nm intervals.

**Measurement protocol.** The nationality or ethnicity, gender, age, and body location of the subject are important and depend on the requirements of the specific study that needs the measurement of human skin colour. Thus, this information for each participant was first determined through self-evaluation via a questionnaire. In regions with homogeneous populations (e.g., China, Thailand, Japan), all participants belonged to the same ethnic group. In regions with mixed populations (e.g., the UK), participants were provided with a questionnaire that included options for existing ethnicities, mixed ethnicity, and "other" (self-defined). This information was recorded using consistent coding schemes in the dataset (see Data Records).

To ensure the integrity and consistency of all skin spectral reflectance measurements in this study, several standardised conditions were rigorously maintained. First, it was essential for the skin of all subjects to be clean, unabraded, and free from any cosmetics, lotions, or medical products that could affect the measurement outcomes. Each subject was prepared accordingly prior to data collection to meet this standard. Additionally, the measurement instruments (portable spectrophotometer - SP) were calibrated according to manufacturer guidelines before each session. Lighting conditions for measurements were carefully controlled; all measurements were conducted under diffuse lighting conditions to avoid discrepancies associated with collimated light sources. This was facilitated by lighting systems integrated within the measurement instruments (SP). Such controlled environments guaranteed that the spectral data collected was accurate and consistent across all subjects and datasets.

During the measurement, the portable spectrophotometer was brought to the subject, ensuring the sample area had no blemishes (e.g. hair, freckles, etc.) and had not been subjected to recent pressure (e.g. to promote blood flow). Areas with visible hair (e.g., the chin of male participants with beards) were avoided during measurements to ensure that only clear skin areas were included in the dataset. When using a spectrophotometer, care was taken to ensure the instrument was gently in contact with the area of skin to be measured, to prevent extraneous light from reaching the detector (see Fig. 1b). It was also necessary to be careful that no excessive pressure was applied to the skin when contacting the device, which might lead to a change in its colour due to the promotion of blood flow beneath the surface. Measurement parameters mentioned in the last section, such as the geometry of illumination and the inclusion of the specular component, must be checked before measurement. Measurements were taken at 4–10 different body locations depending on the site, with the forehead, cheek, and back of the hand covered by all sites. For consistency, measurements were taken from one randomly selected side (left or right) of each subject. Laterality was generally not recorded, except for Datasets 7 and 8, which included bilateral cheek measurements.

After completing the measurements, we reviewed the data and excluded any measurements with zero reflectance (e.g., due to instrument error) or any participant data with colour difference greater than  $15 \Delta E_{ab}^{*}$  between positions. All measurement data were then carefully recorded and reported in accordance with the predefined coding schemes. This protocol not only supports the reliability of our measurements but also enhances the comparability of our data across various locations and time periods.

**Computing colorimetric values.** Based on the CIE colourimetry, the skin spectral data in our study are transformed into CIE colorimetric data using the CIE standard colorimetric observer and CIE standard illuminant D65, the latter simulating the skin colour under daylight conditions. These derived colorimetric data include tristimulus values: *X*, *Y*, *Z*; x, y chromaticity coordinates: *x*, *y*; CIELAB parameters:  $L^*$ ,  $a^*$ ,  $b^*$ ,  $C^*_{ab}$ ,  $h_{ab}$ ; u', v' chromaticity coordinates: *u'*, *v'*, respectively<sup>35</sup>.

The tristimulus values *X*, *Y*, *Z* are obtained:

$$\begin{split} \mathbf{X} &= k \sum_{\lambda} S(\lambda) \mathbf{R}(\lambda) \overline{\mathbf{x}}(\lambda) \Delta \lambda \\ \mathbf{Y} &= k \sum_{\lambda} S(\lambda) \mathbf{R}(\lambda) \overline{\mathbf{y}}(\lambda) \Delta \lambda \\ \mathbf{Z} &= k \sum_{\lambda} S(\lambda) \mathbf{R}(\lambda) \overline{\mathbf{z}}(\lambda) \Delta \lambda \end{split}$$

where  $S(\lambda)$  is the spectral power distribution (SPD) of the CIE standard illumination D65;  $R(\lambda)$  is the spectral reflectance of the skin sample as a function of wavelengths;  $\overline{x}(\lambda)$ ,  $\overline{y}(\lambda)$ , and  $\overline{z}(\lambda)$  are colour-matching functions of the CIE standard observer;  $\lambda$  is the wavelength (in the unit of nm); k is a scaling constant to normalise the tristimulus values.

CIE chromaticity coordinates, x, y, z, are defined as the equations below:

$$x = \frac{X}{X + Y + Z}; Y = \frac{Y}{X + Y + Z}; Z = \frac{Z}{X + Y + Z}$$

where x + y + z = 1.

The coordinates of CIELAB are transformed from the XYZ tristimulus values:

$$L^{*} = 116f(Y/Yn) - 16$$
$$a^{*} = 500[f(X/Xn) - f(Y/Yn)]$$
$$b^{*} = 200[f(Y/Yn) - f(Z/Zn)]$$

$$f(w) = \begin{cases} (w)^{1/3} \text{ for } w > 0.008856\\ 7.787(w) + 16/116 \text{ for } w \le 0.008856 \end{cases}$$

			ORIGIN						
Data		1	UK		7 UK2				
А	Record number	2	China		8 UK3				
В	Origin of data	3	Middle East - Iraq		9 Japan				
C	Subject number	4	Southeast Asia - Thailand		10 China2				
Skin Type		5	Southern Asia - Pakistan	11 Middle East - Saudi Aribia					
D Ethnicity		6 Spain							
F	Gender		ETHNICITY	GENDER		<b>BODY LOCATION</b>			
E	Age group	CA	Caucasian	F	Female	1	Back of hand		
	Age group	CN	I Chinese	М	Male	2	Cheek		
G	Body location	SA	South Asian (Pakistani)	AG	E GROUP	3	Cheek bone		
Instrument		AF	African	1	< 20	4	Chin		
н	Instrument type	IQ	Middle Eastern (Iraqi)	2	20-29	5	Ear lobe		
1	SPIN/SPEX	TH	Southeast Asian (Thai)	3	30-39	6	Forehead		
J	Start wavelength	JP	Japanese	4	40-49	7	Inner arm		
К	End wavelength	AB	Middle Eastern (Arabian)	5	50-59	8	Neck		
L	Wavelength interval		INSTRUMENT	6	> 60	9	Nose tip		
N – BD	Spectral data	SF	Spectrophotometer			10	Outer arm		
BF – BQ	Colorimetry		SPIN			11	Palm		
		sc	Specular included			12	Ring finger		

Fig. 4 Overview of data structure and coding scheme.

where X, Y, and Z and Xn, Yn, and Zn represent the tristimulus values of the object colour and reference white, respectively. And the perceived attributes of chroma  $C_{ab}^*$  and hue  $h_{ab}$  can be predicted in CIELAB colour space:

$$C_{ab}^* = [(a^*)^2 + (b^*)^2]^{1/2}$$

$$\mathbf{h}_{ab} = \tan^{-1}(b^*/a^*)$$

The u', v' chromaticity coordinates are transformed from the XYZ tristimulus values:

$$u' = 4X/(X + 15Y + 3Z)$$
  
 $v' = 9Y/(X + 15Y + 3Z)$ 

This computing process is seamlessly integrated into the dataset file, ensuring that the colorimetric values are directly accessible and systematically calculated to reflect accurate and consistent colour assessments under standardised daylight conditions.

#### **Data Records**

The International Skin Spectra Archive (ISSA) has been deposited at *figshare*<sup>37</sup>, specifically structured for ease of use in coding applications. The file is organised into two spreadsheets: one coding scheme and one datasheet. Figure 4 shows the overview of the data structure and coding schemes. Within the datasheet, data are arranged into columns labelled A to BQ, each representing specific attributes of the recorded measurements. A is the unique identifier for each record; B is the origin of data, linked to the detailed origin table in the coding scheme spreadsheet (also see the top right column in Fig. 4); C is the unique identifier for each subject; D is the ethnicity, linked to the ethnicity key in the coding scheme (also see the middle right column in Fig. 4); E is the gender; F is the age group linked to the age group categories; G is body location, linked to the body location descriptions; H is the instrument type; I is the specular components mode; J and K are the start and end wavelength measured in nanometres (nm); L is the wavelength interval of measurement in nm. The spectral data are given in columns N (360 nm) to BD (780 nm). All records include at least the spectral range of 400–700 nm, which covers the full visible spectrum. Some datasets extend beyond this range, capturing additional data from 360–740 nm. CIE colorimetric data, calculated using the CIE standard colorimetric observer and CIE standard illuminant D65, are listed in columns BF to BQ.

#### **Technical Validation**

The dynamic nature of skin colour and its variability due to many factors as mentioned make consistent measurement difficult, especially in comparison to flat uniform samples. To address this, our methodology involves not only a stringent protocol that ensures consistent measurements across different laboratories and instruments but also a thorough assessment of the variability of skin measurements, focusing on the short-term repeatability and the inter-instrument agreement. The short-term repeatability is evaluated by repeating the measurement of the same body location, using the same instrument in quick succession, ensuring consistent pressure on the skin. Inter-instrument agreement is assessed by measuring the same body location using different settings on the same instrument. The CIELAB colour difference is used to quantify the variability between the measurements<sup>34</sup>. Our results indicate that the short-term repeatability for the SP maintains a variation within  $0.40 \Delta E_{ab}^*$  across different pressure and aperture sizes. The mean colour differences between the different values of measurement pressure were 0.88 and 1.82 for the medium and small aperture respectively. The mean colour differences between the two measurement apertures were 2.48 and 3.04 for the two different pressures, respectively. The greater effect of field size on the variability, compared with that of pressure, implies that the variability in the colour distribution of the area being measured is greater than the variability in the measurement caused by the pressure: thus reasonable pressure on the skin does not significantly affect the measured colour values. Those variances on live human skin are higher than those observed with uniform flat skin-colour samples, such as the PANTONE SkinTone<sup>™</sup> Guide. A detailed investigation of the variability in skin colour measurements, as explored in a prior study<sup>24</sup>, underpins these findings and contributes to our understanding of the complexities involved in skin colour measurement. The validation results have provided useful guidance for the definition of measurement protocol and the establishment of the current skin spectral dataset.

#### **Usage Notes**

The dataset is designed for easy integration into coding environments, such as R Studio or MATLAB. The spreadsheet is formatted to include the code necessary for calculating the CIE colorimetric data using the CIE standard colorimetric observer and CIE standard illuminant D65. Additionally, users have the flexibility to use spectral reflectance data to compute colorimetric values based on any self-defined illuminants and observer's colour matching functions, facilitating versatile applications across different research and practical contexts.

#### Code availability

No custom code was used to generate the data described in this study.

Received: 21 January 2025; Accepted: 17 March 2025; Published online: 23 March 2025

#### References

- 1. Igarashi, T., Nishino, K. & Nayar, S. K. The Appearance of Human Skin: A Survey. Found. Trends<sup>®</sup> Comput. Graph. Vis. 3, 1–95 (2007).
- Jablonski, N. G. & Chaplin, G. Human skin pigmentation as an adaptation to UV radiation. Proc. Natl. Acad. Sci. 107, 8962–8968 (2010).
- 3. Angelopoulou, E. The Reflectance Spectrum of Human Skin Technical Report MS-CIS-99-29. 20.500.14332/7527 (1999).
- 4. Tan, K. W. & Stephen, I. D. Colour detection thresholds in faces and colour patches. *Perception* 42, 733-741 (2013).
- 5. Quillen, E. E. et al. Shades of complexity: New perspectives on the evolution and genetic architecture of human skin. Am. J. Phys. Anthropol. 168, 4–26 (2019).
- Imai, F. H., Tsumura, N., Haneishi, H. & Miyake, Y. Principal component analysis of skin color and its application to colorimetric color reproduction on CRT display and hardcopy. J. Imaging Sci. Technol. 40, 422–430 (1996).
- 7. Bonnier, N. Attributes of image quality for color prints. J. Electron. Imaging 19, 011016 (2010).
- 8. Xiao, K. et al. Improved method for skin reflectance reconstruction from camera images. Opt. Express 24, 14934–14950 (2016).
- 9. Liu, Y. et al. Comparing VR- and AR-Based Try-On Systems Using Personalized Avatars. Electronics 9, 1814 (2020).
- 10. Xiao, K., Zardawi, F., van Noort, R. & Yates, J. M. Color reproduction for advanced manufacture of soft tissue prostheses. J. Dent. 41, e15–e23 (2013).
- Xiao, K., Zardawi, F., Van Noort, R. & Yates, J. M. Developing a 3D colour image reproduction system for additive manufacturing of facial prostheses. Int. J. Adv. Manuf. Technol. 70, 2043–2049 (2014).
- 12. Zonios, G. *et al.* Melanin absorption spectroscopy: new method for noninvasive skin investigation and melanoma detection. *J. Biomed. Opt.* **13**, 014017 (2008).
- Do Hyun Chung & Sapiro, G. Segmenting skin lesions with partial-differential-equations-based image processing algorithms. IEEE Trans. Med. Imaging 19, 763–767 (2000).
- 14. Hirose, M., Akaho, R., Maita, C., Sugawara, M. & Tsumura, N. Optimization of spectral sensitivities of mosaic five-band camera for estimating chromophore densities from skin images including shading and surface reflections. *Opt. Rev.* 23, 544–548 (2016).
- De, D. R., Park, S. E., Shih, T., Shi, V. Y. & Hsiao, J. L. Representation of Skin Color, Disease Severity, and Anatomic Location in Hidradenitis Suppurativa Images. Ski. J. Cutan. Med. 8, 1700–1704 (2024).
- Lu, Y. et al. Skin coloration is a culturally-specific cue for attractiveness, healthiness, and youthfulness in observers of Chinese and western European descent. PLoS One 16, e0259276 (2021).
- 17. Lu, Y. et al. Different colour predictions of facial preference by Caucasian and Chinese observers. Sci. Rep. 12, 12194 (2022).
- Lu, Y., Xiao, K., Pointer, M. & Lin, Y. Predicting Facial Attractiveness from Colour Cues: A New Analytic Framework. Sensors 24, 391 (2024).
- 19. Kikuchi, K., Masuda, Y., Yamashita, T., Kawai, E. & Hirao, T. Image analysis of skin color heterogeneity focusing on skin chromophores and the age-related changes in facial skin. *Ski. Res. Technol.* **21**, 175–183 (2015).
- Bilal, S., Akmeliawati, R., Salami, M. J. E. & Shafie, A. A. Dynamic approach for real-time skin detection. J. Real-Time Image Process. 10, 371–385 (2015).
- Brancati, N., Frucci, M., De Pietro, G. & Gallo, L. Dynamic clustering for skin detection in YCbCr colour space. Pattern Recogn Inf. Process 49–53 (2016).
- 22. Jones, M. J. & Rehg, J. M. Statistical Color Models with Application to Skin Detection. *Int. J. Comput. Vis.* 46, 81–96 (2002).
- 23. Zhang, W., Zhao, X., Morvan, J.-M. & Chen, L. Improving Shadow Suppression for Illumination Robust Face Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 41, 611–624 (2019).
- 24. Wang, M. et al. An investigation into the variability of skin colour measurements. Color Res. Appl. 43, 458–470 (2018).
- Appleton, K. M. et al. The value of facial attractiveness for encouraging fruit and vegetable consumption: analyses from a randomized controlled trial. BMC Public Health 18, 298 (2018).
- Morimoto, T., Zhang, C., Fukuda, K. & Uchikawa, K. Spectral measurement of daylights and surface properties of natural objects in Japan. Opt. Express 30, 3183 (2022).
- Arnold, S. E. J., Faruq, S., Savolainen, V., Mcowan, P. W. & Chittka, L. FReD: The Floral Reflectance Database A Web Portal for Analyses of Flower Colour. 5, 1–9 (2010).
- ISO. ISO/TR 16066:2003 Graphic technology Standard object colour spectra database for colour reproduction evaluation (SOCS). 46 (2003).
- Sun, Q. & Fairchild, M. D. Statistical characterization of face spectral reflectances and its application to human portraiture spectral estimation. J. Imaging Sci. Technol. 46, 498–506 (2002).
- 30. De Rigal, J. et al. The effect of age on skin color and color heterogeneity in four ethnic groups. Ski. Res. Technol. 16, 168–178 (2010).
- De Rigal, J., Abella, M., Giron, F., Caisey, L. & Lefebvre, M. A. Development and validation of a new Skin Color Chart <sup>™</sup>. Ski. Res. Technol. 13, 101–109 (2007).

- 32. Cooksey, C. C., Allen, D. W. & Tsai, B. K. Reference Data Set of Human Skin Reflectance. J. Res. Natl. Inst. Stand. Technol. 122, 26 (2017).
- Washington, D. C. Reference Data Set and Variability Study for Human Skin Reflectance of The 29th Cie Session. https://doi. org/10.25039/x46.2019 (2019).
- 34. CIE. CIE 015:2018 Colorimetry. https://doi.org/10.25039/TR.015.2018 (CIE Publication, 2018).
- 35. Hunt, R. W. G. & Pointer, M. R. Measuring Colour. https://doi.org/10.1002/9781119975595 (Wiley, 2011).
- Luo, M. R. Encyclopedia of Color Science and Technology. Encyclopedia of Color Science and Technology. https://doi.org/10.1007/978-1-4419-8071-7 (Springer New York, 2016).
- Yan, L. *et al.* The International Skin Spectra Archive (ISSA): a multicultural human skin phenotype and colour spectra collection. *figshare* https://doi.org/10.6084/m9.figshare.28228571.v4 (2025).

#### **Author contributions**

Designed research: Y.L., K.X., M.P. Data Collection: Y.L., K.X., R.H., S.Z., S.S., C.G., A.S., Y.H., M.W., E.P.R., F.Z. Data preparation and anonymisation: Y.L., S.Z., A.N., C.G., C.L., Y.H., Y.M., E.P.R., F.Z. Data validation: K.X., S.W. Analysed data: Y.L., K.X., M.P. Wrote the paper: Y.L., K.X., M.P., S.W. All the authors contributed substantially to the critical revision and approved the final version of the paper.

#### **Competing interests**

The authors declare no competing interests.

#### **Additional information**

Correspondence and requests for materials should be addressed to K.X.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025