Journal of Cosmetics, Dermatological Sciences and Applications, 2026, 16(3), 161-175
https://www.scirp.org/journal/jcdsa
ISSN Online: 2161-4512
ISSN Print: 2161-4105

CTK Derma AI: Any Skin Tone-Aware AI Capture Calibration for Objective Quantification of Facial Aging Features: Validation across Fitzpatrick Phototypes I - VI

Shu Li, Labina Shrestha, Mirjalol Tuychiev, Tae Yong Jung, Ryan Wonsuk Choi, Kukizo Miyamoto*

Choicetechkorea Co., Ltd., Yongin-si, Korea
Email: *dr.miyamoto@choicetech.kr

How to cite this paper: Li, S., Shrestha, L., Tuychiev, M., Jung, T.Y., Choi, R.W. and Miyamoto, K. (2026) CTK Derma AI: Any Skin Tone-Aware AI Capture Calibration for Objective Quantification of Facial Aging Features: Validation across Fitzpatrick Phototypes I - VI. *Journal of Cosmetics, Dermatological Sciences and Applications*, **16**, 161-175.
https://doi.org/10.4236/jcdsa.2026.163012

Received: June 16, 2026
Accepted: July 31, 2026
Published: August 3, 2026

Abstract

Background: Image-based assessment of facial aging signs is sensitive to lighting, device-dependent color rendering, and skin phototype, which can compromise reproducibility and fairness across diverse populations. **Objective:** To validate a skin tone-aware AI-based image capture calibration and analysis framework (CTK Derma AI) for the objective quantification of pores, pigmented spots, and wrinkles across Fitzpatrick phototypes I - VI, using expert visual grading as the reference standard. **Methods:** Facial images from 265 participants were captured using a handheld CTK ChoiceDx® platform device under a standardized imaging protocol. Participants were grouped into Fitzpatrick I - II (n = 58), III - IV (n = 61), and V - VI (n = 146). Regions of interest included the periorbital area for wrinkle analysis and the central cheek for pore, spot and skin tone (L-value) analysis by the image analysis, and then graded by the expert graders. CTK Derma AI analysis framework was trained by half of the dataset (n = 132), then validated using the remaining dataset (n = 133). CTK Derma AI pipeline was compared with a conventional non-calibrated image-processing baseline (called Control, hereafter). Scores were normalized to a 0 - 100 scale, with higher scores indicating fewer visible aging signs. Correlations with expert grading were assessed using Pearson's r; differences between dependent correlations were pre-specified for R-based Steiger/Meng-style testing. Repeatability was evaluated across three repeated captures per subject. **Results:** CTK Derma AI showed statistically stronger overall correlations with expert grading than Control for pores, spots, wrinkles and skin tone. The largest improvements were observed in the Fitzpatrick V - VI subgroup. Repeatability was high across all phototypes (CV = 4.5%; provisional ICC (2, 1) = 0.975, 95% CI 0.970 - 0.979). **Conclusions:** CTK Derma AI provides robust, objective, and repeatable quantification of key facial aging features across diverse skin phototypes and outperforms a conventional non-calibrated baseline workflow.

Keywords

Skin Diagnosis, Skin Aging, Wrinkle, Pore, Spots, CTK Derma AI, Skin Phototype

1. Introduction

Visible facial aging is a multidimensional phenotype characterized by structural and pigmentary changes, including wrinkles, pore conspicuity, and hyperpigmented spots. These skin features strongly influence perceived age and are commonly used as endpoints in dermatological and cosmetic evaluation [1]-[8]. However, conventional visual assessment is inherently subjective and prone to intergrader variability. Validated photonumeric scales and structured clinical scoring systems have improved reproducibility, yet image-based quantification remains challenging when capture conditions and skin phototypes (skin tone) vary [5]-[9].

A central limitation of conventional image analysis pipelines is that they are highly sensitive to illumination, exposure, camera-dependent color rendering, and contrast variability. These issues are particularly important in facial skin assessment because the visual prominence of pores, pigmentary variation, and wrinkles differs substantially across skin phototypes [10]-[14]. Consequently, a robust pipeline must not only quantify visible features objectively, but also account for phototype-dependent imaging variability.

To address this challenge, CTK Derma AI was developed as a skin tone-aware capture calibration and image analysis framework designed to standardize image input and improve feature extraction across Fitzpatrick phototypes I - VI. The present study compared CTK Derma AI with a conventional non-calibrated image-processing baseline (Control) using expert visual grading as the reference and further evaluated phototype-stratified performance and repeatability across three repeated captures. Overall study design and flow was shown in **Figure 1** [15]-[19].

2. Materials and Methods

2.1. Study Design and Participants

This observational validation study enrolled N = 265 female adults from age 20 to 59 representing Fitzpatrick phototypes I - VI who met the inclusion criteria of healthy skin, no history of cosmetic procedure and self-declaration of ethnicities, either Caucasians, East Asians, South-East Asians, Middle Asians, Hispanics, or African American. Participants were analyzed in three strata: Fitzpatrick I - II (n = 58), III - IV (n = 61), V - VI (n = 146), following the standard Fitzpatrick scale and segmentation. Mean ages (±SD) were: I - II: 35.4 ± 14.6, III - IV: 37.3 ± 16.2 and V - VI: 36.0 ± 14.1, respectively.

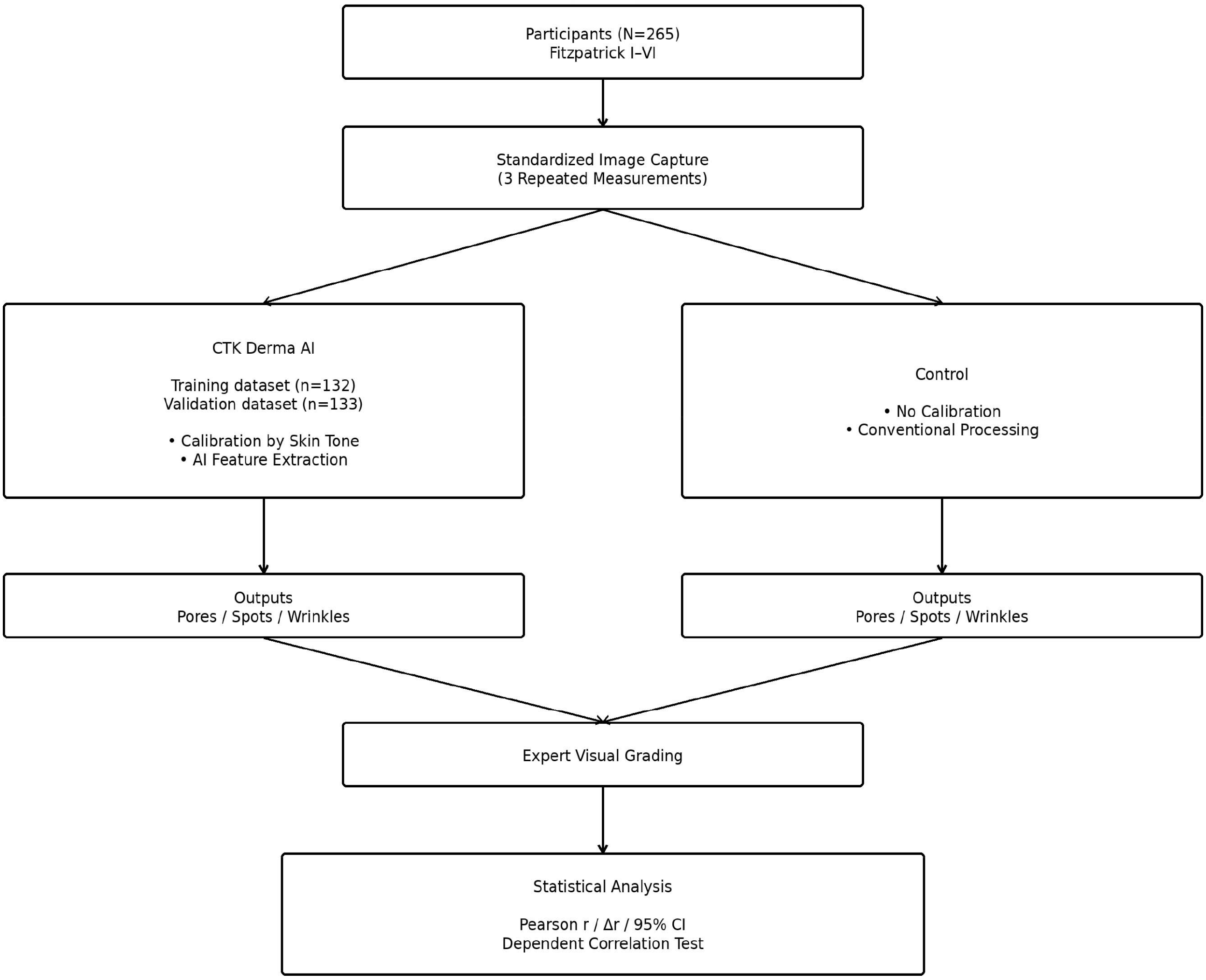
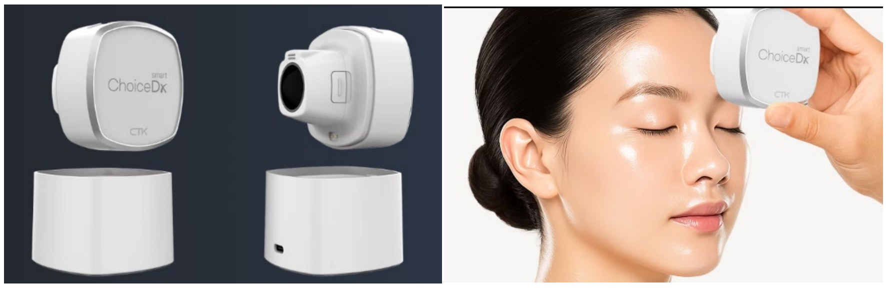


Figure 1. Overview of study design and analytical pipelines (CTK Derma AI vs baseline), evaluation against expert grading, phototype-stratified analysis, and repeatability testing.

The study was conducted in Korea, in accordance with the Declaration of Helsinki and approved by CTK IRB committee, and written informed consent was obtained from all participants.

2.2. Image Acquisition Protocol (Fixed Capture Setting)

All facial skin images were captured using a handheld diagnostic device based on the CTK ChoiceDx platform (e.g., Dx-Pro® portable system; Figure 2), under a standardized and fixed capture protocol, including controlled distance and positioning, closed optical geometry at 6500 K for pore and wrinkle imaging, and spot imaging with cross-polarized filtering, and the consistent CMOS camera settings, to minimize operational variability.



(a) (b)

Figure 2. Handheld CTK ChoiceDx device (e.g., Dx-Pro® portable system) and representative standardized capture setup.

When feasible, a reference target (neutral gray card and/or color calibration chart) was included within the imaging field to enable post hoc normalization of brightness and color characteristics across captures.

Two primary regions of interest (ROI) were defined for quantitative analysis, consistent with established facial aging assessment frameworks. The periorbital region was selected for wrinkle evaluation, reflecting its sensitivity to fine lines and aging-related structural changes, whereas the central cheek region was selected for pore conspicuity, pigmentation (spots) and skin tone (L), as it represents a clinically relevant and relatively homogeneous area for assessing skin texture and discoloration [1] [8]. For each ROI, image-based features were extracted using the CTK Derma AI pipeline or conventional image processing (Control). Specifically: Pore analysis was performed by detecting circular or elliptical microstructures corresponding to follicular openings, followed by quantification of their density, size distribution, and contrast relative to surrounding skin regions. Spot analysis targeted regions of hyperpigmentation, identified through local intensity deviation and chromatic irregularity, with features such as area coverage, edge definition, and color variance contributing to the score. Wrinkle analysis was based on the detection of linear and curvilinear ridge-like structures, incorporating measures of length, depth-associated contrast, orientation consistency, and spatial frequency within the periorbital region. L-value was analyzed at the cheek region for skin tone. Extracted features were integrated into composite scores using algorithm-specific weighting schemes optimized against expert visual grading.

All image analyses were performed in a fully automated manner without manual intervention to eliminate operator bias. For three attributes (pores, spots, wrinkles), analysis outputs were normalized to a continuous 0 - 100 scale, where higher scores represent better skin condition (*i.e.*, lower presence or severity of the corresponding aging signs), and lower scores indicate increased prominence or deterioration. Skin tone analysis output was shown as L-value (higher is lighter skin tone, lower is darker skin tone) This normalization enables cross-subject and cross-attribute comparability and facilitates statistical correlation with expert grading.

These quantitative scores were subsequently compared with expert visual grading using correlation-based metrics and dependent correlation analyses, as described below.

2.3. Expert Visual Grading (Reference Standard)

A panel of dermatologists and/or trained expert graders (N = [3]) scored the following attributes on a color-calibrated monitor:

1) Pores conspicuity.

2) Pigmented spots/uneven pigmentation ("spots").

3) Wrinkles.

4) Skin Tone.

Clinical scale selection and rubric design were informed by validated photoaging and skin aging scoring literature with 0 - 4 grading scale (0: least severe, 1: less severe, 2: moderate, 3: severe, 4: very severe) (photonumeric scales; SCINEXA; related reviews) [5]-[8] [10]. Skin tone expert grading was carried out with 0 - 4 grading scale (0: very light, 1: light, 2: moderate, 3: dark, 4: very dark).

Inter-rater reliability was assessed using intraclass correlation coefficients (ICC) under standard formulations [20] [21].

2.4. Rationale for CIELAB Adoption (Industrial Standard)

Objective color metrics were computed in CIELAB ($L^*a^*b^*$), a device-independent color space standardized in international colorimetry practice and widely used in industry for color specification and difference calculations [22] [23]. This provides a consistent and interpretable foundation for skin tone quantification across devices and illumination when combined with calibration.

2.5. Analytical Pipelines Compared

2.5.1. CTK Derma AI Analysis

CTK Derma AI Analysis comprised:

1) Face/ROI localization.

2) Automated capture-condition calibration (color constancy/correction using reference target when available; learning-based normalization otherwise), motivated by evidence that AI-based color constancy can improve dermatologic image consistency [16].

3) Objective feature extraction for pores, spots, wrinkles, and objective skin tone metrics in calibrated CIELAB space [22] [23].

4) Supervised mapping from calibrated features to predicted attribute scores aligned to expert grading. For the model development, the dataset was randomly divided into a training dataset (n = 132) and an independent validation subset (n = 133). All reported performance metrics by CTK Derma AI were derived from the validation dataset.

5) All image capture calibration and analysis were then carried out by the validated CTK Derma AI framework.

2.5.2. Control: Conventional Feature Detection without Calibration

In short, baseline means conventional non-AI image processing-based feature detection without capture calibration. The control represents conventional rule-based image processing methods widely used prior to AI-based dermatological image analysis. Therefore, the control used the same ROI definition but analyzed images as captured, without calibration; pores/spots/wrinkles were quantified via conventional feature detection and texture/contrast descriptors, and skin tone was estimated using non-calibrated color statistics.

2.6. Test-Retest Repeatability (Three Repeated Captures)

To evaluate measurement repeatability across all skin tones, each participant underwent three repeated captures under the same protocol. Repeatability was quantified using test-retest ICC and coefficient of variation (CV). ICC methodology follows established guidance for rater/repeatability reliability [20] [21].

2.7. Statistical Analysis

The primary endpoint was Pearson's correlation coefficient (r) between algorithm-derived scores and expert visual grading for pores, spots, and wrinkles. Because CTK Derma AI and Control were evaluated on the same participants, differences between correlations were treated as dependent correlations. The pre-specified inferential approach was based on Fisher's r-to-z transformation within a Steiger/Meng-style framework, implemented in R, as recommended for overlapping or dependent correlation comparisons [24]-[26]. Repeatability across the three repeated captures was assessed using the coefficient of variation (CV) and the intraclass correlation coefficient (ICC) based on a two-way random-effects, absolute-agreement model [ICC (2, 1)] [20] [21]. Complementary interpretation followed the principle that correlation strength alone does not equate to full method agreement, as noted by Bland and Altman [27].

3. Results

3.1. Participant Characteristics

A total of N = 265 participants were analyzed across phototype strata: Fitzpatrick I - II (n = 58; mean age 35.4 ± 14.6), Fitzpatrick III - IV (n = 61; mean age 37.3 ± 16.2), and Fitzpatrick V - VI (n = 146; mean age 36.0 ± 14.1) (See **Table 1**).

Table 1. Participant characteristics by phototype stratum.

Participant Demographics	Basesize	Mean ages (±SD)	Sex	Ethnicities
Fitzpatrick I - II	58	35.4 ± 14.6	Female	Caucasians/East Asians
Fitzpatrick III - IV	61	37.3 ± 16.2	Female	East/South East Asians
Fitzpatrick V - VI	146	36.0 ± 14.1	Female	South East Asians/Middle East Asians/Indians/Hispanics/African Americans

3.2. Repeatability across All Skin Tones (Three Repeated Captures)

Across the full cohort and all skin tones, repeatability was high. Each subject underwent three repeated captures under the standardized protocol, and the overall coefficient of variation (CV) was 4.5% across all phototypes, indicating strong measurement stability (see Table 2). Repeatability was additionally summarized with test-retest ICC following established reliability framework after final computation.

Table 2. Repeatability across three repeated captures (template).

Metric	Overall (All phototypes)	Notes
Number of repeated captures per subject	3	Fixed protocol, same capture condition
CV (Coefficient of Variation)	4.5%	Reported across all skin tones
ICC value	[0.975]	
ICC 95% CI	[0.970 - 0.979]	

3.3. Correlation with Expert Visual Grading: CTK Derma AI vs Control (All Participants)

Across all participants (N = 265), CTK Derma AI showed significantly higher correlations with expert grading than control for all evaluated attributes (pores, spots, wrinkles). For each attribute, Pearson's r was computed between algorithm outputs and expert scores, and the difference between dependent correlations (same participants) was tested using Fisher r-to-z transformation within Steiger-type dependent-correlation tests, implemented in R.

Overall results are summarized in Table 3 (Overall: N = 265) and visualized in Figure 3.

Table 3. Correlations vs expert grading and dependent-correlation comparison: Overall (N = 265).

Attribute	r(A) CTK Derma AI vs Expert	r(B) Control vs Expert	$\Delta r = r(A) - r(B)$	p (Δr)*	95% CI (Δr)**
Pores	−0.845	−0.503	0.342	<0.01	0.271 - 0.679
Spots	−0.803	−0.678	0.125	0.03	0.078 - 0.174
Wrinkles	−0.869	−0.639	0.23	0.02	0.177 - 0.289
Skin Tone	−0.882	−0.513	0.269	<0.01	0.243 - 0.631

*p(Δr) computed using Fisher r-to-z dependent-correlation tests (Steiger-type) in R. **95% CI(Δr) computed using dependent-correlation CI procedures consistent with Steiger/Meng frameworks (available via R implementations such as *cocor*).

3.4. Phototype-Stratified Performance (I - II, III - IV, V - VI)

Phototype-stratified analyses showed that CTK Derma AI maintained robust correlations with expert grading across all strata and outperformed baseline for each attribute within each stratum: Fitzpatrick I - II (n = 58), III - IV (n = 61), and V - VI (n = 146). Differences in r between Pipeline A and B were evaluated using the same dependent-correlation Fisher r-to-z/Steiger-type tests in R, with confidence intervals reported via the same framework.

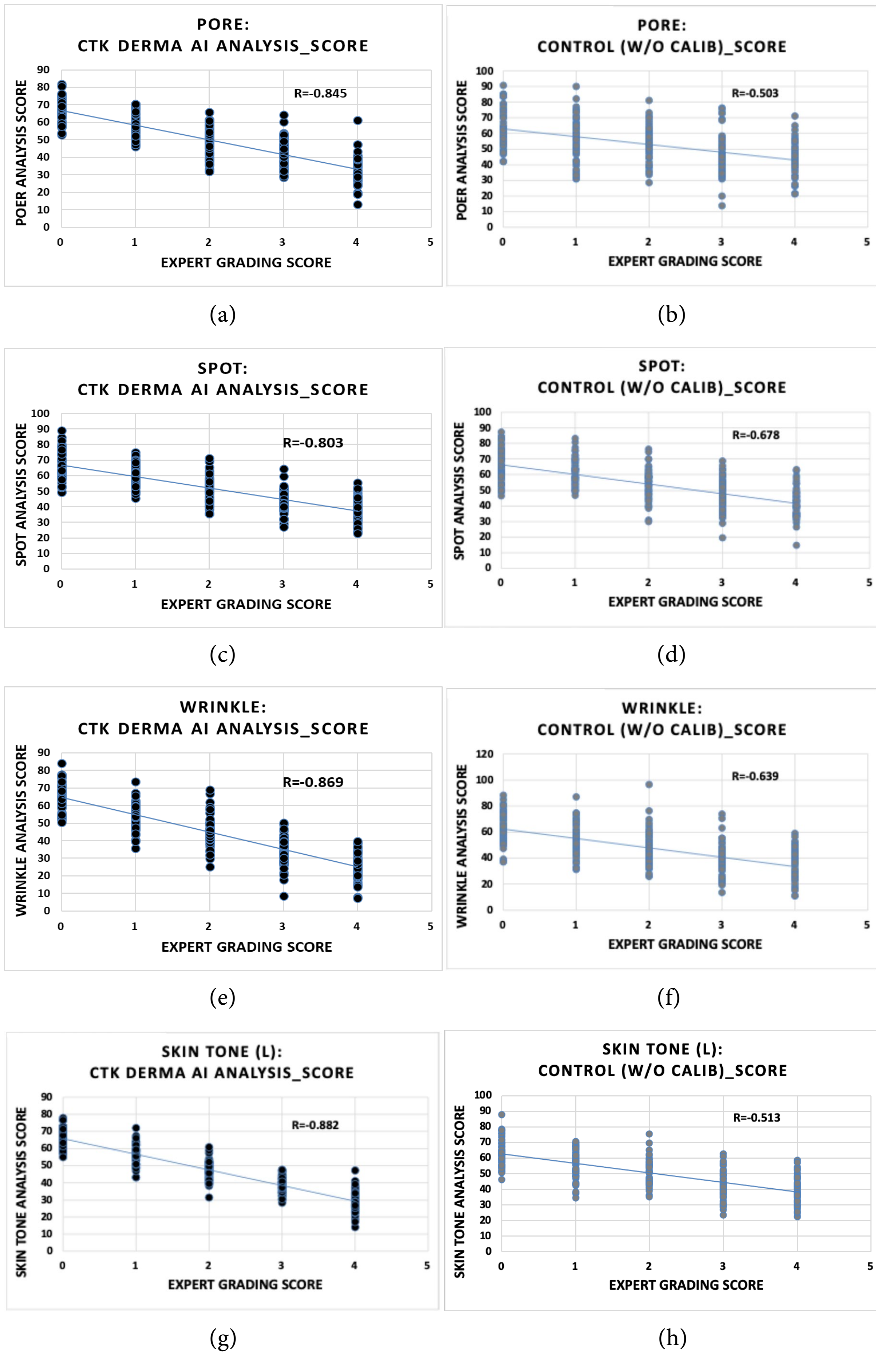


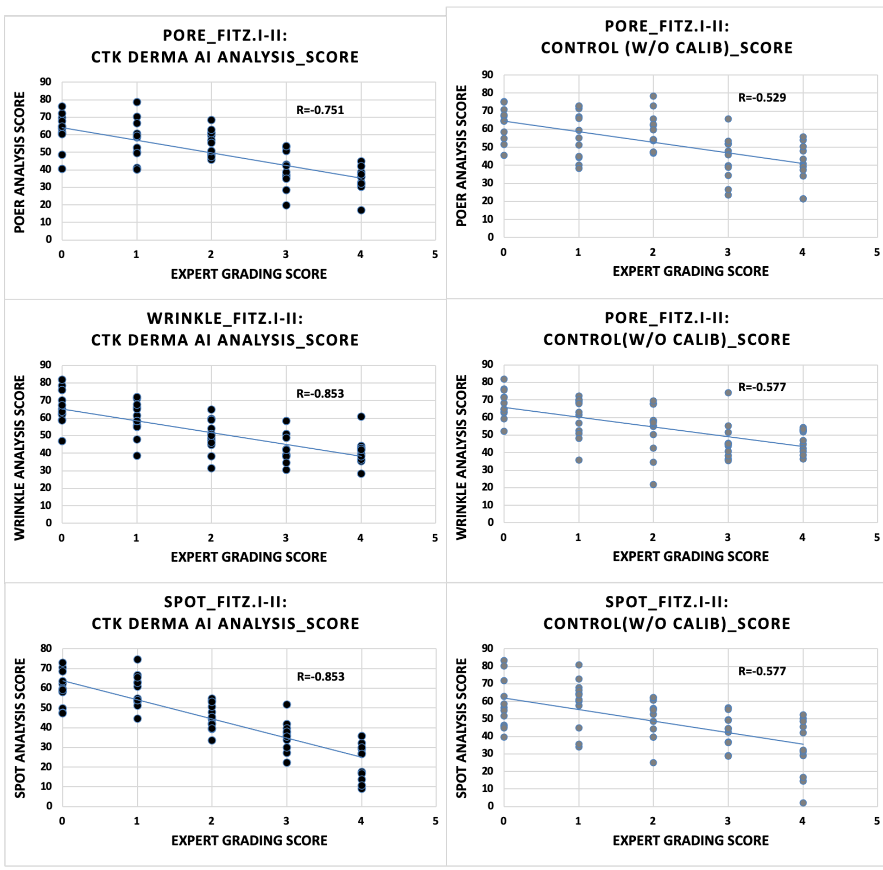
Figure 3. Correlations vs expert grading and dependent-correlation comparison (overall).

Phototype-stratified results are presented in **Table 4** (Stratified) and visualized in **Figure 4**. Example skin images of pore, spot and wrinkle at the original, control analysis (without calibration), and CTK Derma AI analysis were shown in Figure 5.

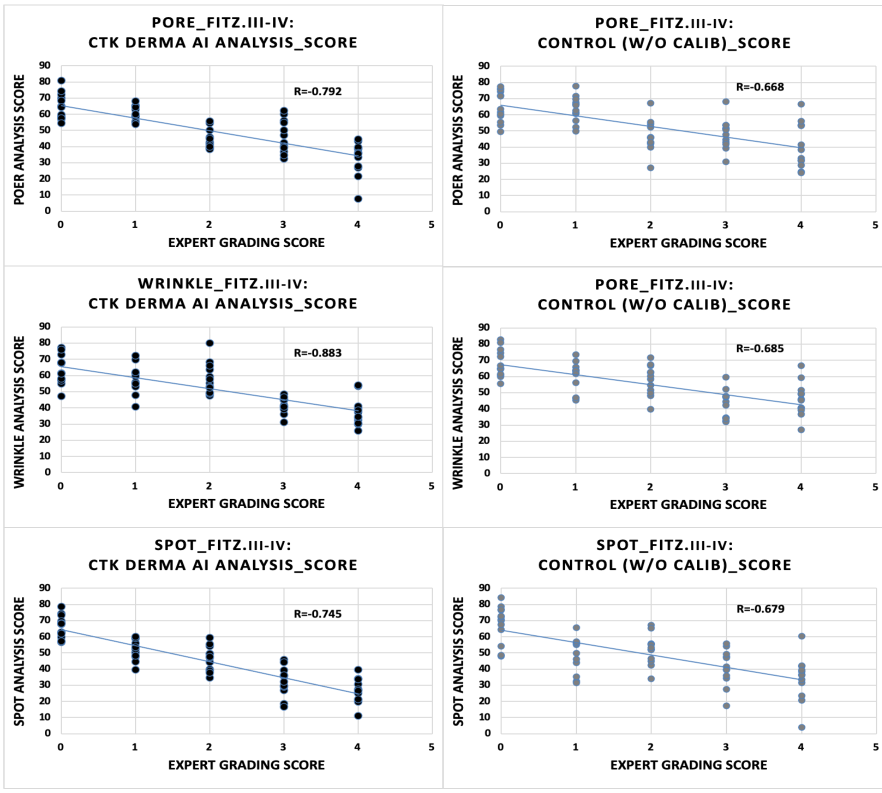
Table 4. Correlations vs expert grading and dependent-correlation comparison: stratified by phototype across Fitzpatrick I - II, III - IV, and V - VI.

Phototype Stratum	n	Attribute	r(A) CTK Derma AI vs Expert	r(B) Control vs Expert	Δr	p (Δr)*	95% CI (Δr)**
Fitzpatrick I - II	58	Pores	−0.721	−0.589	0.132	0.03	0.268 - 0.749
		Spots	−0.733	−0.601	0.132	0.02	0.256 - 0.756
		Wrinkles	−0.853	−0.577	0.276	<0.01	0.427 - 0.748
		Skin Tone	−0.865	−0.612	0.253	<0.01	0.463 - 0.735
Fitzpatrick III - IV	61	Pores	−0.792	−0.668	0.124	<0.01	0.272 - 0.793
		Spots	−0.745	−0.679	0.066	0.03	0.177 - 0.79
		Wrinkles	−0.883	−0.685	0.198	<0.01	0.332 - 0.81
		Skin Tone	−0.896	−0.712	0.184	<0.01	0.297 - 0.758
Fitzpatrick V - VI	146	Pores	−0.708	−0.423	0.285	<0.01	0.405 - 0.671
		Spots	−0.681	−0.366	0.315	<0.01	0.433 - 0.645
		Wrinkles	−0.817	−0.402	0.415	<0.01	0.544 - 0.645
		Skin Tone	−0.879	−0.488	0.391	<0.01	0.583 - 0.697

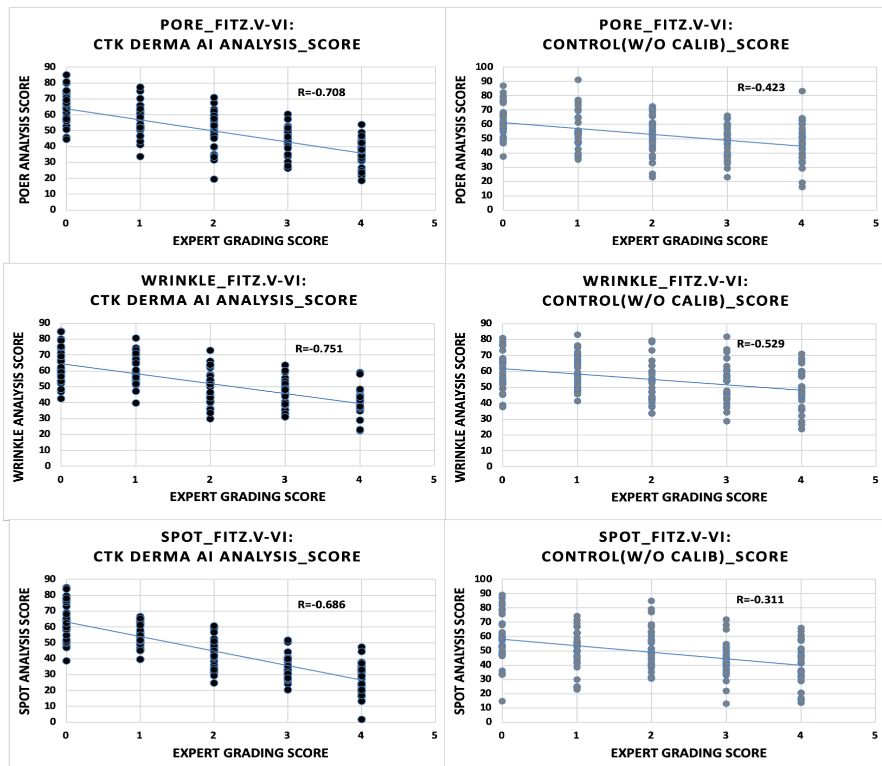
*, ** same as Table 3.



(a)



(b)



(c)

Figure 4. Correlations vs expert grading and dependent-correlation comparison (by Phototype).

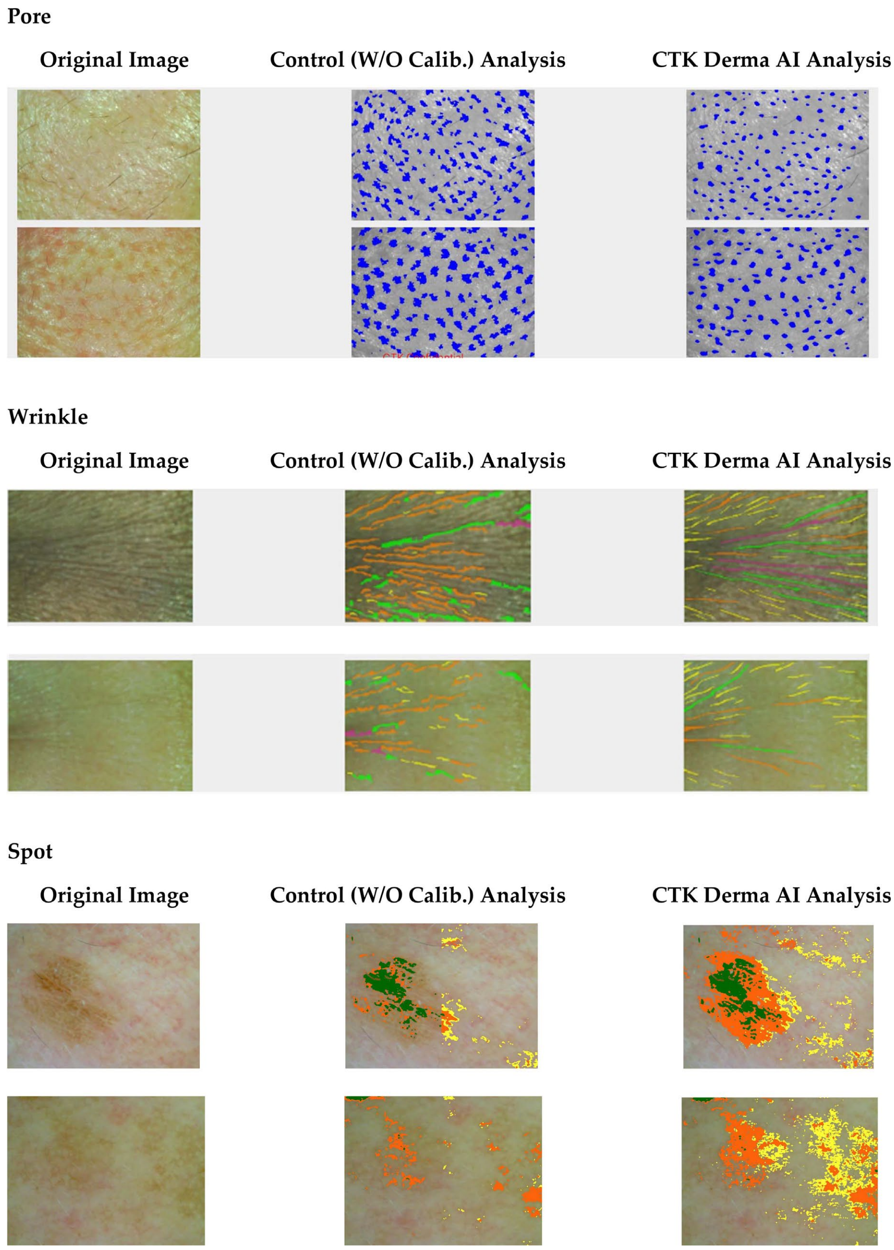


Figure 5. Example skin images of pore, spot and wrinkle at the Original (left), Control analysis (without calibration, middle), and CTK Derma AI analysis (right) for pore, wrinkle and spot.

3.5. Summary of Findings

Across all skin tones and all evaluated attributes, CTK Derma AI achieved consistently superior agreement with expert grading versus the baseline workflow with statistically significant difference, and demonstrated highly repeatability with a provisional ICC (2, 1) of 0.975 (95% CI, 0.970 - 0.979), that was consistent with the low overall coefficient of variation (4.5%) across three repeated captures (Table 2). These findings support CTK Derma AI as a calibrated, phototype-robust image-based measurement framework.

4. Discussion

This study demonstrates that AI-based capture calibration integrated with objective image analysis (CTK Derma AI) yields significantly higher agreement with expert visual grading than a control feature-detection workflow lacking calibration, across pores, spots, and wrinkles together with skin tone (L) in a cohort spanning Fitzpatrick phototypes I - VI (N = 265). The approach builds on the Skin Aging Index concept—translating multi-parameter facial appearance into an interpretable index correlated with visual grading—while explicitly addressing the key barrier of capture-condition variability [1].

The use of CIELAB as the objective color basis is justified by its role as an internationally standardized, device-independent industrial colorimetric system, enabling meaningful comparison of skin tone metrics once calibration reduces device/illumination effects [22] [23]. This is consistent with broader efforts to quantify skin tone objectively and to reduce reliance on coarse or variably assigned scales [9] [11] [12] [14]. The importance of color normalization is further supported by evidence that AI-based color constancy can improve perceived quality and workflow outcomes in dermatologic imaging [16].

For aging attribute validation, prior literature supports that photonumeric scales and validated clinical scoring systems improve repeatability and inter-grader agreement compared to descriptive methods [6] [7]. The skin aging score provides a validated approach to separate intrinsic and extrinsic aging signs [5]. Additionally, independent validation studies show that image analysis methods can correlate moderately to strongly with expert photonumeric grading across wrinkles and pigmentary features when properly validated [8] [19]. These lines of evidence reinforce the necessity of rigorous validation design—exactly what our CTK Derma AI comparison and phototype-stratified analysis provide.

Statistically, we explicitly used tests appropriate for dependent correlations (same participants) rather than treating correlations as independent. This approach follows established methodology for comparing correlated correlations and is implemented in R-based workflows [25] [26]. The emphasis on correct method-comparison statistics echoes the caution that correlation alone is insufficient to claim agreement between measurement methods; complementary agreement analysis (e.g., Bland-Altman) can be added when full numeric results are finalized [27].

Finally, the repeatability assessment with three repeated captures per participant yielded CV = 4.5% across all skin tones, indicating high stability under the standardized protocol. Reliability and ICC frameworks are standard for quantifying repeatability and rater reliability and will be reported fully once ICC values are inserted [20] [21].

Limitations include [e.g., single-site acquisition, limited environmental lighting diversity given fixed protocol], and the focus on 2D imaging for features such as sagging. Future work should incorporate multi-site external validation, expanded lighting/device variability testing, and longitudinal sensitivity to treatment effects.

5. Conclusion

CTK Derma AI, combining AI-based capture calibration with objective multi-attribute analysis, demonstrates significantly improved correlation with expert visual grading compared with conventional feature detection without calibration, across all key attributes (skin tone, pores, spots, wrinkles) and across phototype strata (I - II, III - IV, V - VI). Repeatability across all skin tones was high (CV = 4.5% across three test-retest captures). CTK Derma AI provides a phototype-robust framework suitable for dermatologic and cosmetic science evaluation.

Author Contributions

Shu Li, Labina Shrestha, Mirjalol Tuychiev led the skin image analysis and its AI model development and validation. Tae Yong Jung carried out the device hardware qualification and optimization for accurate skin image calibration and collection. Ryan Wonsuk Choi orchestrated the holistic skin diagnosis system definition and its alignment for the measurement accuracy and sensitivity, and Kukizo Miyamoto led the development and validation data mining for the manuscript proceedings.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Miyamoto, K., Fujii, K., Wanigasekara, S., Ding, B., Munakata, Y., Wang, S., *et al*. (2025) Skin Age Index: Development of a Convenient and Quantitative Indicator of Facial Skin Aging and Its Application to Evaluation of the Efficacy of Three Anti-Aging Skin Care Products Containing GFF (Pitera). *Journal of Cosmetics*, *Dermatological Sciences and Applications*, **15**, 46-61. https://doi.org/10.4236/jcdsa.2025.151004

[2] Nkengne, A., Roure, R., Rossi, A.B. and Bertin, C. (2013) The Skin Aging Index: A New Approach for Documenting Anti-Aging Products or Procedures. *Skin Research and Technology*, **19**, 291-298. https://doi.org/10.1111/srt.12040

[3] Cho, C., Lee, E., Park, G., Cho, E., Kim, N., Shin, J., *et al*. (2022) Evaluation of Facial Skin Age Based on Biophysical Properties *in Vivo*. *Journal of Cosmetic Dermatology*, **21**, 3546-3554. https://doi.org/10.1111/jocd.14653

[4] Robic, J., Nkengne, A., Lua, B.L., Bellanger, A., Hu, R. and Vie, K. (2023) Modeling of Global Skin Aging Indices among Caucasian and Asian Women. *Skin Research and Technology*, **29**, e13222. https://doi.org/10.1111/srt.13222

[5] Vierkötter, A., Ranft, U., Krämer, U., Sugiri, D., Reimann, V. and Krutmann, J. (2009) The SCINEXA: A Novel, Validated Score to Simultaneously Assess and Differentiate between Intrinsic and Extrinsic Skin Ageing. *Journal of Dermatological Science*, **53**, 207-211. https://doi.org/10.1016/j.jdermsci.2008.10.001

[6] Griffiths, C.E.M., Wang, T.S., Hamilton, T.A., *et al*. (1992) A Photonumeric Scale for the Assessment of Cutaneous Photodamage. *Archives of Dermatology*, **128**, 347-351. https://doi.org/10.1001/archderm.1992.01680130061006

[7] Ayer, J., Ahmed, A., Duncan-Parry, E., Beck, P., Griffiths, T.W., Watson, R.E.B., *et*

al. (2018) A Photonumeric Scale for the Assessment of Atrophic Facial Photodamage. *British Journal of Dermatology*, **178**, 1190-1195. https://doi.org/10.1111/bjd.16331

[8] Hamer, M.A., Jacobs, L.C., Lall, J.S., Wollstein, A., Hollestein, L.M., Rae, A.R., *et al*. (2015) Validation of Image Analysis Techniques to Measure Skin Aging Features from Facial Photographs. *Skin Research and Technology*, **21**, 392-402. https://doi.org/10.1111/srt.12205

[9] Harvey, V.M., Alexis, A., Okeke, C.A.V., McKinley-Grant, L., Taylor, S.C., Desai, S.R., *et al*. (2024) Integrating Skin Color Assessments into Clinical Practice and Research: A Review of Current Approaches. *Journal of the American Academy of Dermatology*, **91**, 1189-1198. https://doi.org/10.1016/j.jaad.2024.01.067

[10] Fitzpatrick, T.B. (1988) The Validity and Practicality of Sun-Reactive Skin Types I through VI. *Archives of Dermatology*, **124**, 869-871. https://doi.org/10.1001/archderm.1988.01670060015008

[11] Ulrich, P., Zink, A., Biedermann, T. and Sitaru, S. (2025) Beyond Fitzpatrick: Automated Artificial Intelligence-Based Skin Tone Analysis in Dermatological Patients. *npj Digital Medicine*, **8**, Article No. 378. https://doi.org/10.1038/s41746-025-01770-4

[12] Olsen, D., Nguyen, C., Hall, M., Carletti, M. and Weis, S.E. (2026) The Efficacy of the Fitzpatrick Scale in Clinical Practice. *HCA Healthcare Journal of Medicine*, **7**, Article 2. https://doi.org/10.36518/2689-0216.2207

[13] Eilers, S., Bach, D.Q., Gaber, R., Blatt, H., Guevara, Y., Nitsche, K., *et al*. (2013) Accuracy of Self-Report in Assessing Fitzpatrick Skin Phototypes I through VI. *JAMA Dermatology*, **149**, 1289-1294. https://doi.org/10.1001/jamadermatol.2013.6101

[14] Cu, C.W., Dundas, N.E., Heintz, T., Sheikh, Z.A., Alonso-Bermudez, B., Walker, J., *et al*. (2025) Validity of Two Subjective Skin Tone Scales and Its Implications on Healthcare Model Fairness. *npj Digital Medicine*, **8**, Article No. 595. https://doi.org/10.1038/s41746-025-01975-7

[15] Mbatha, S.K., Booysen, M.J. and Theart, R.P. (2024) Skin Tone Estimation under Diverse Lighting Conditions. *Journal of Imaging*, **10**, Article 109. https://doi.org/10.3390/jimaging10050109

[16] Branciforti, F., Meiburger, K.M., Zavattaro, E., Veronese, F., Tarantino, V., Mazzoletti, V., *et al*. (2023) Impact of Artificial Intelligence-Based Color Constancy on Dermoscopical Assessment of Skin Lesions: A Comparative Study. *Skin Research and Technology*, **29**, e13508. https://doi.org/10.1111/srt.13508

[17] Jung, G., Kim, S. and Yoo, S. (2024) Skin Tone Analysis through Skin Tone Map Generation with Optical Approach and Deep Learning. *Skin Research and Technology*, **30**, e70088. https://doi.org/10.1111/srt.70088

[18] Yoon, H., Kim, S., Lee, J. and Yoo, S. (2023) Deep-Learning-Based Morphological Feature Segmentation for Facial Skin Image Analysis. *Diagnostics*, **13**, Article 1894. https://doi.org/10.3390/diagnostics13111894

[19] Henseler, H., *et al*. (2023) Assessment of the Reproducibility and Accuracy of the Visia® Complexion Analysis Camera System for Objective Skin Analysis of Facial Wrinkles and Skin Age. *GMS Interdisciplinary Plastic and Reconstructive Surgery*, **12**, Doc07.

[20] Shrout, P.E. and Fleiss, J.L. (1979) Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, **86**, 420-428. https://doi.org/10.1037/0033-2909.86.2.420

[21] McGraw, K.O. and Wong, S.P. (1996) Forming Inferences about Some Intraclass Correlation Coefficients. *Psychological Methods*, **1**, 30-46.

https://doi.org/10.1037/1082-989x.1.1.30

[22] International Commission on Illumination (CIE) (2019) Colorimetry—Part 4: CIE 1976 L*a*b* Colour Space (ISO/CIE 11664-4:2019).

[23] Chardon, A., Cretois, I. and Hourseau, C. (1991) Skin Colour Typology and Suntanning Pathways. *International Journal of Cosmetic Science*, **13**, 191-208. https://doi.org/10.1111/j.1467-2494.1991.tb00561.x

[24] Steiger, J.H. (1980) Tests for Comparing Elements of a Correlation Matrix. *Psychological Bulletin*, **87**, 245-251. https://doi.org/10.1037/0033-2909.87.2.245

[25] Meng, X., Rosenthal, R. and Rubin, D.B. (1992) Comparing Correlated Correlation Coefficients. *Psychological Bulletin*, **111**, 172-175. https://doi.org/10.1037/0033-2909.111.1.172

[26] Diedenhofen, B. and Musch, J. (2015) Cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, **10**, e0121945. https://doi.org/10.1371/journal.pone.0121945

[27] Martin Bland, J. and Altman, D. (1986) Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *The Lancet*, **327**, 307-310. https://doi.org/10.1016/s0140-6736(86)90837-8