IFSCC 2025 full paper (IFSCC2025-163)

Prediction of facial translucency scores and facial skin image generation using neural network models

Kensuke Yotsumoto[1, 2], Aika Kuramoto[2], Mizuho Kokubo[1], Yuki Shibaike[1], Hiroko Kawanobe[1], Akinobu Hayashi[1] and Makoto Hasegawa[2]

[1] Research and Development Department, ALBION CO., LTD., 2-24-11, Higashi-nihombashi, Chuo-ku Tokyo 103-0004, Japan; [2] Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo 120-8551, Japan

1. Introduction

Counseling consumers about skin conditions can lead to more accurate cosmetic proposals and higher customer satisfaction. To this end, many commercial stores have introduced equipment for skin evaluation, including VISIA (Canfield Scientific), which can assess “spots,” “wrinkles,” “pores,” “uneven skin tone,” “dark spots,” “porphyrins,” “texture,” “melanin index,” and “hemoglobin index” [1], and other systems, which can evaluate “skin hydration” and “oil content” [2, 3]. In addition to such directly measurable parameters, more nuanced concepts such as “facial translucency,” which is considered important by many Japanese women, have also been used in skin counseling.

Facial translucency is defined as the degree to which light is reflected from the skin surface and within the skin. It is primarily related to texture conditions, thickening of the stratum corneum, skin hydration, transepidermal water loss, parakeratosis, and color factors from melanin and hemoglobin indices [4–7]. Although these factors interact in complex ways, their relative contributions to facial translucency remain unclear.

Recent studies evaluated facial translucency using equipment and multiple regression analysis [8, 9], which helped clarify relationships between variables using a specific formula [10]. Such techniques enable identifying factors such as “skin texture and roughness,” “uneven skin

tone," "b* value," and "pigmented spots" to determine facial translucency; however, its single-layer linear nature restricts its ability to manage complex evaluations [11].
Deep learning, which is a subset of artificial intelligence (AI) based on artificial neural networks, presents a multilayer architecture for nonlinear fitting [12]. Deep learning is widely applied in image recognition, text summarization, and drug discovery [13] and leverages convolutional, pooling, and dense layers for feature extraction, dimension reduction, and feature combination, respectively [14]. However, deep learning requires curated datasets to achieve high accuracy. In this study, we designed a deep-learning model to estimate facial translucency scores using convolutional neural networks [15].
A generative adversarial network (GAN) is a machine-learning framework that features a generator and discriminator [16]. CycleGAN, which can learn from unpaired datasets, is designed for image-to-image translation [17]. We used cycleGAN to create predicted images with improved or deteriorated "facial translucency" based on real facial images.
In this study, we prepared a dataset for deep learning and developed a deep-learning model for both facial translucency scoring and generative skin imaging based on that dataset. Accordingly, we proposed neural-network-based deep learning as a more effective method for evaluating facial translucency and suggesting potential modifications in predicted facial skin images.

2. Materials and Methods

2.1 Participants and Evaluation of Facial Translucency

A total of 100 Japanese women participated in the study (20s and younger: 24, 30s: 26, 40s: 22, and 50s and older: 28). In addition, a total of 40 experienced individuals from our beauty department (expert evaluators) scored the facial translucency of participants using their professional judgment.

2.2 Image Collection and Dataset Preparation

Facial images were captured under standard lighting using VISIA Evolution software and displayed on a personal computer (HP EliteBook 860 G10 Notebook) with an eye-tracking system (Tobii Nano 60 Hz). All images were rated on a visual analog scale for facial translucency using an expert evaluator.

2.3 Data Handling and Statistical Analysis

Shapiro–Wilk test results indicated non-normal data, and therefore, a robust Z-score standardization was used [18]. Median translucency scores were calculated for each facial image, and a percentile-based clinical scale was developed based on these medians.

2.4 Eye-Tracking and Region of Interest Analysis

An eye-tracking system was used to monitor the line of sight of evaluators during scoring [19]. The regions of interest (ROIs) included the forehead, nose, cheek, and mouth, and the focus was quantified as the duration of fixation (DoF). The ratio of DoF for each ROI was computed as

$$\text{Ratio of DoF} = (\text{DoF of each ROI})/(\text{Tota DoF}) \times 100.$$

The median values of the ratio of DoF were used for the analysis because of non-normality.

2.5 Dataset Preparation for Evaluating the Correlation of Multiple Regression Analysis

A dataset with 30,000 facial images (300 per participant) cropped to 256 × 256 pixels was used in this study. A total of 2,448 images were randomly selected for color feature extraction using MATLAB (MathWorks). The color features included saturation, hue, hue angle, brightness, a*, b*, SD(a*), and SD(b*); subsequently, a multiple regression analysis was performed using the extracted features.

2.6 Deep Learning for Facial Translucency Scoring: Convolutional Neural Network

A dataset of 2,448 images (256 × 256 pixels) with associated translucency scores was used to train a facial translucency deep-learning model based on pretrained convolutional networks from Keras. The performance was evaluated by validation loss on a separate set of 2,448 images that lacked translucency scores. Predicted versus true values were plotted, and the final predicted translucency score for each participant was the median of their 300 cropped images.

2.7 Image Generation for Assessing the Developed Model

Heatmap images of facial translucency were generated for a participant who used skincare and quasi-drug products for four years and eight months. Subsequently, the number of tiles for each translucency grade was calculated.

2.8 Deep Learning for Generative AI: CycleGAN

A dataset graded using a clinical scale was categorized as grade 1 (2,100 images), grades 2 and 3 (8,100 images each), grade 4 (6,300 images), and grade 5 (5,400 images). CycleGAN was trained pairwise across these grades for producing images with predicted improvements or deteriorations in facial translucency.

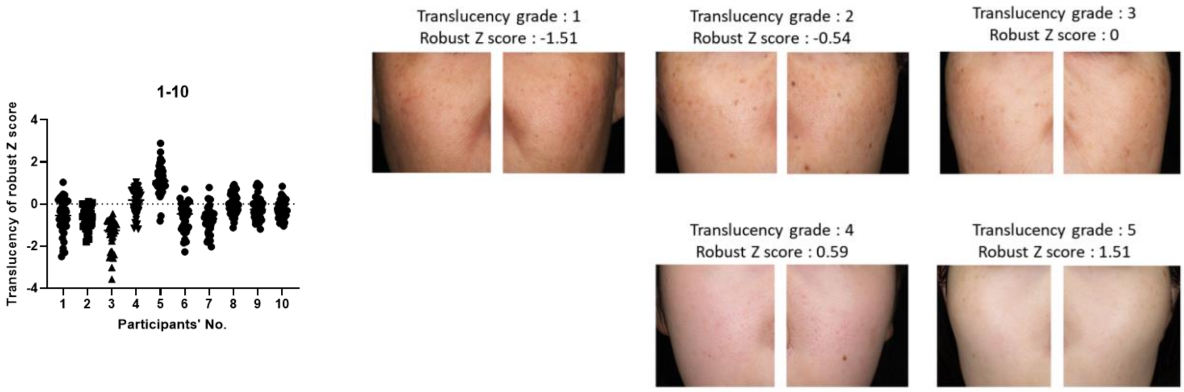
2.9 Ethics

Each experiment was performed in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Tokyo Denki University. The personal information of the participants was strictly protected by this code of ethics, and the participants were allowed to voluntarily participate or discontinue from the study.

3. Results

3.1 Dataset for Facial Translucency Score Determination

A robust and reliable dataset for facial translucency scoring was establied by collecting evaluations from 40 experienced individuals, which helped mitigate subjective bias. Figure 1 shows the collected facial translucency scores and developed clinical translucency grade using percentile rankings. The clinical grade aids in visually interpreting a facial translucency deep-learning model.

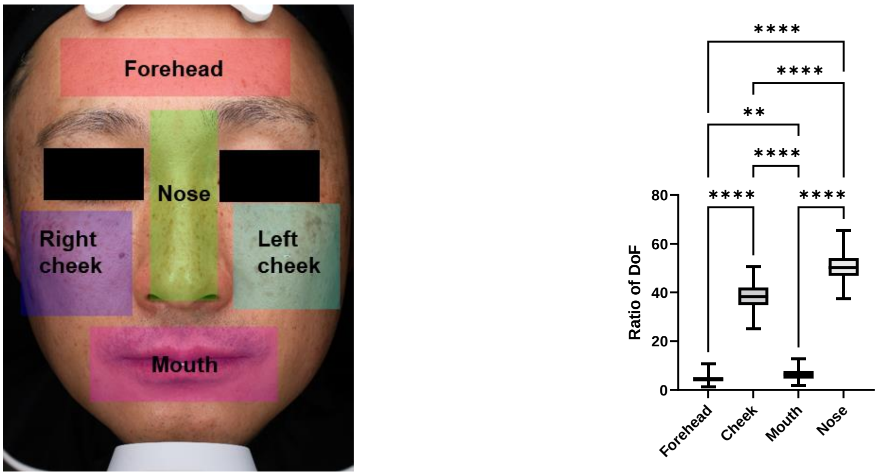


(**a**) (**b**)

Figure 1. Clinical grade of facial translucency based on the median of robust Z scores. (a) Facial translucency scores for participants No. 1–10 and (b) developed clinical translucency grade.

3.2 Eye Position with Facial Translucency Assessment

The eye-tracking system recorded gaze positions of expert evaluators, which were evaluated for the ROI (Figure 2a). A comparison of DoF ratios revealed that forehead and mouth areas had lower fixation ratios, whereas the cheeks and nose had higher fixation ratios (Figure 2b). This implied that the evaluators spent more time focusing on the cheeks and nose when assessing facial translucency scores.

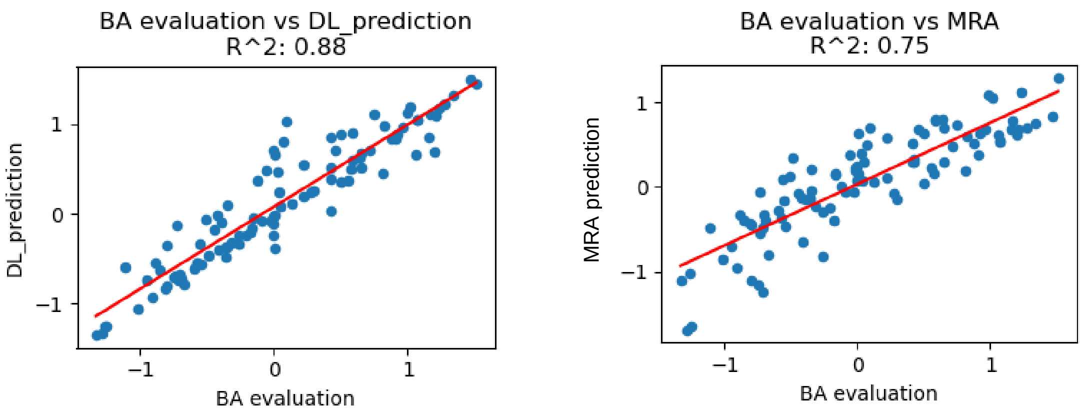


(**a**) (**b**)

Figure 2. Cheek and nose focusing with a translucency evaluation on the eye-tracking system. (a) ROI setting for the face of each participant and (b) analysis of the duration of fixation (DoF) of each part of the facial area. (** $p < 0.01$, **** $p < 0.0001$, Dunn's test).

3.3 Development of a Deep-Learning Model for Facial Translucency Evaluation

We selected a pretrained model from the Keras Applications library as the foundation for our model. Among candidate models, ResNet 50 with fine tuning demonstrated the lowest minimum validation loss (0.338), indicating its suitability for this task. We compared its predictive capabilities with those of multiple regression analysis for validating the performance of our model (Figure 3). Our deep-learning model outperformed both traditional approaches, achieving an R^2 value of 0.88, indicating its superior ability to predict facial translucency scores.



(**a**) (**b**)

Figure 3. Determination of the coefficient of the facial translucency model using multiple regression analysis. Scatter plot and R^2 of the (a) facial translucency model and (b) multiple regression analysis.

3.4 Application of the Facial Translucency Model in Clinical Practice

In a long-term clinical study extending over four years and eight months, we used our developed facial translucency model to observe longitudinal changes in the facial translucency of a single participant. Heatmap analyses (Figures 4a–d) revealed an increase in the high-translucency region, which suggested that consistently applied skincare routines could enhance facial translucency. Further, we integrated the findings of the model with a clinical grading scale for translucency. The histogram illustrating the transition from Grade 3 to Grade 4 (Figures 4e and 4f) provided further evidence to improve facial translucency over time.

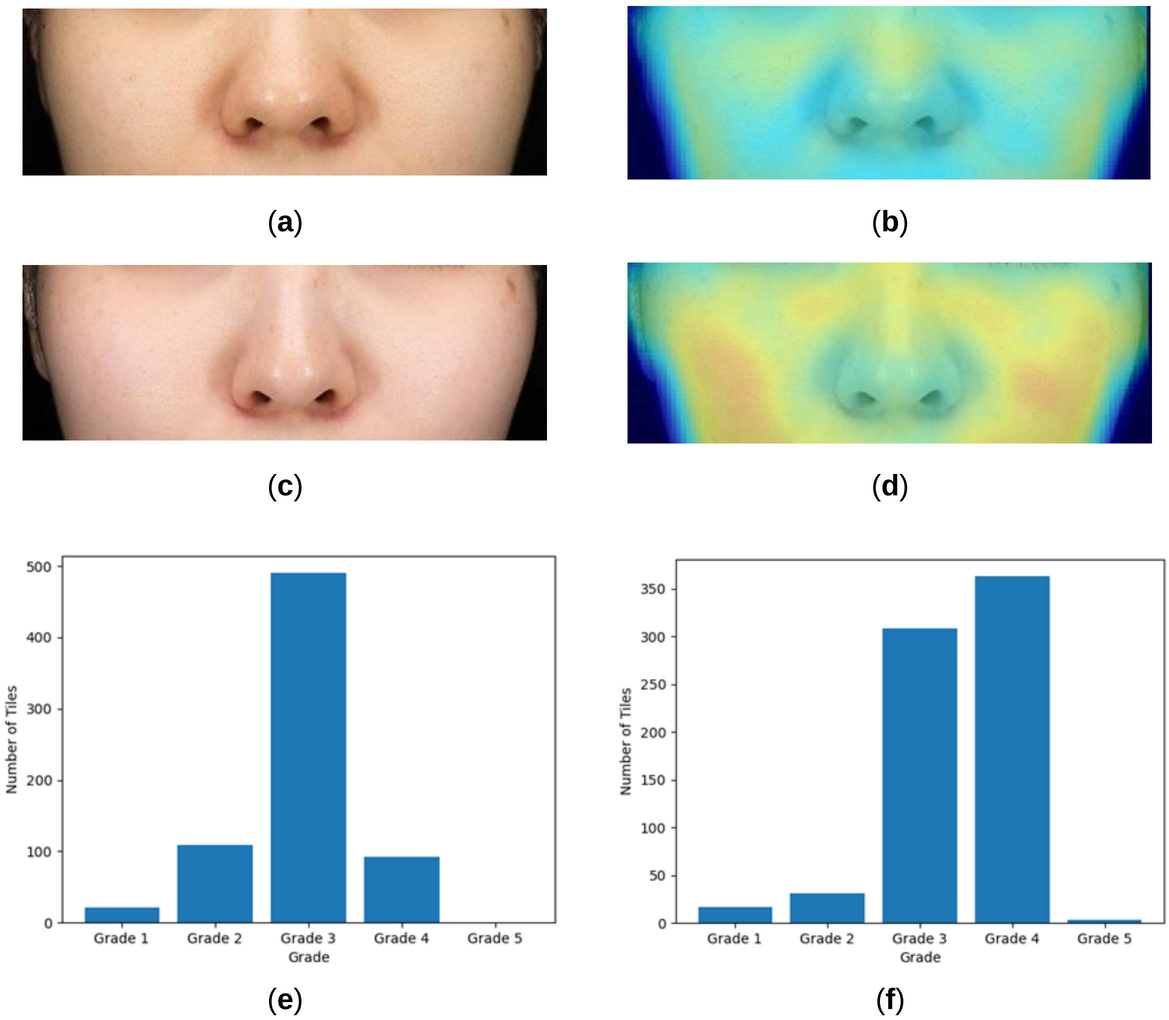


Figure 4. Application of the developed model in clinical studies. Standard light images captured using VISIA Evolution at the (a) baseline and (c) after four years and eight months. Heatmap images generated by the developed model at the (b) baseline and (d) after four years and eight months. Histograms of the clinical grades of the facial area at the (e) baseline and (f) after four years and eight months.

3.5 AI from Real to Generative Images

CycleGAN was used to generate predictive images representing various degrees of facial translucency. Figure 5 shows that the model successfully produced images reflecting enhanced translucency by transforming the images from Grade 1 (Figure 5a) to Grade 5 (Figure 5c) and from Grade 5 (Figure 5b) to Grade 1 (Figure 5d). The ability of this model to preserve image fidelity was demonstrated by the reconstructed output (Figure 5e and 5f), which closely resembled the original image. The high degree of similarity between the original and reconstructed images highlighted the effectiveness of the proposed image prediction methodology.

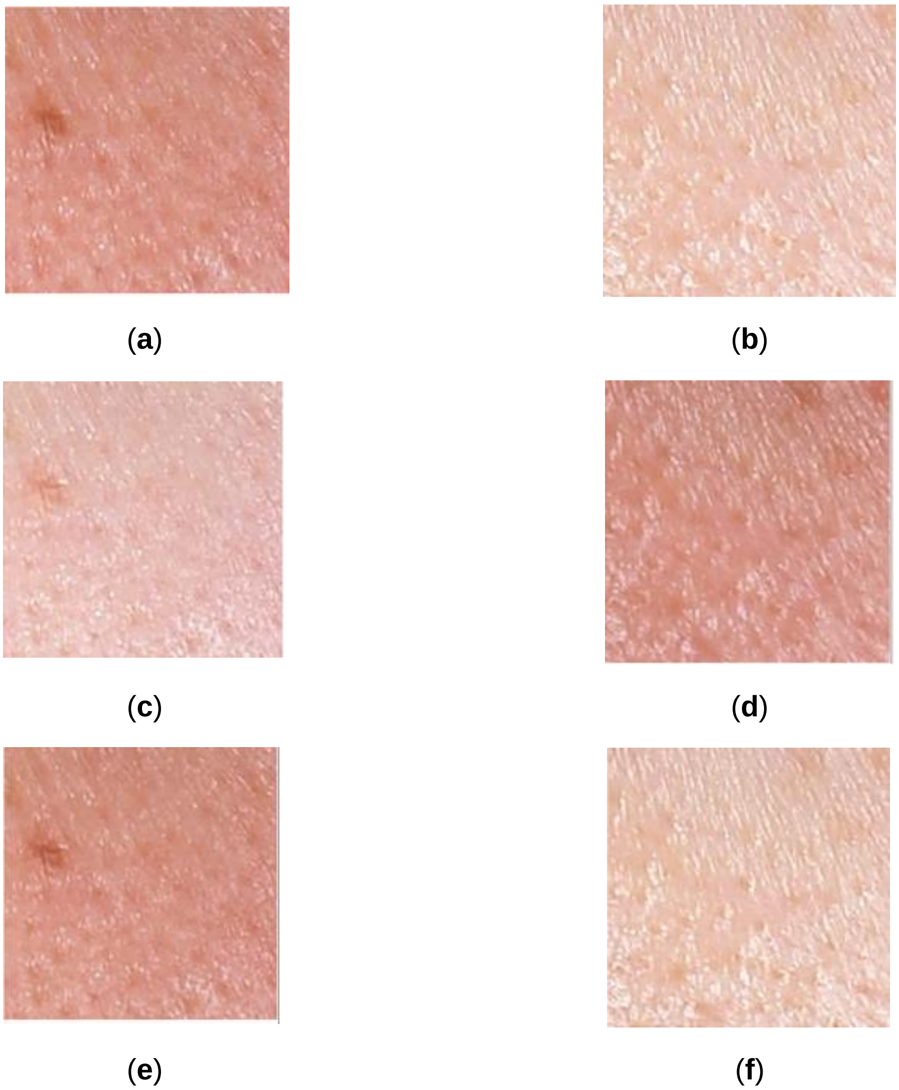


Figure 5. CycleGAN for converting skin images based on clinical grade. Real image cropped around cheek Grade 1(a) and Grade 5 (b), fake image generated from a real image to Grade 5 (c) and Grade 1 (d), and fake image generated from a fake image to Grade 1 (e) and Grade 5 (f).

4. Discussion

4.1 Interpretability of the Enhancement of the Deep-Learning Model for Facial Translucency Evaluation

This study presented a novel deep neural network (DNN) model to evaluate facial translucency. Although deep-learning models demonstrate strong performance, they lack interpretability, making it difficult to understand the effect of individual input parameters on the output. To this end, we examined the correlation between the output of our model and clinically graded facial translucency (Figure 1). Clinical grading is an established evaluation system that offers insights into translucency variations and guides initial visual assessments. These findings enhance the interpretability of the translucency score of the model and support the informed recommendations.

We first considered multiple regression for translucency evaluation; however, it was outperformed by deep learning (Figure 3). Two factors can explain this difference:

(1) Multiple regression is prone to biases arising from subjective selection and transformation of input parameters, whereas the multilayered architecture of the deep-learning model helps mitigate these biases.

(2) Compared to the linear approach of multiple regression, deep learning uses a nonlinear fitting algorithm, which further contributes to a higher coefficient of determination.

These results highlight the value of developing robust facial translucency models and emphasize the need for methods that can clarify decision-making processes.

4.2 Image Recognition using the Developed Model

Although deep-learning models excel in image recognition and computation, their inherent complexity acts as a barrier to understanding their decision-making process. To address this issue, clinical grading is incorporated to enhance the interpretability of the developed model (Figure 1). Clinical grading serves as an intuitive measure for bridging the gap between the output of the model and human perception. Clinical evaluators and retailers that routinely assess facial appearances offer practical insights into model performance. Although these analyses cannot entirely clarify intricate workings of the model, they highlight the features and principles guiding its decisions, helping us better understand its capabilities and limitations.

4.3 Applications of CycleGAN for Generating Realistic and Predictive Facial Skin Images

The CycleGAN model demonstrated significant potential to generate realistic and predictive images of the facial skin, which offers valuable applications in both clinical research and consumer-facing platforms. In clinical studies evaluating the efficacy of cosmetic products, this technology can be employed for visualizing changes in facial translucency scores initially quantified using the developed model. This approach provides a powerful tool to communicate complex scientific data to consumers in an engaging and easily interpretable manner, which can potentially enhance purchase satisfaction by providing a personalized and visually driven understanding of product benefits.

4.4 Limitations of the Study

This study has several limitations, including a relatively small sample size and a restricted demographic composition. Although a coefficient of determination of $R^2 = 0.88$ was obtained, a dataset based on only 100 participants cannot guarantee higher accuracy for the deep-learning model. Further, all participants in this study were Japanese women, which may have confined the applicability of the model to East Asian populations. To address these issues, future work should incorporate a broader range of facial variations by utilizing diverse real-image datasets such as UTKFace and generating artificial facial images through methods such as StyleGAN2 [20].

5. Conclusion

We developed a high-accuracy model for evaluating facial translucency and generating skin images using deep learning. However, this model is subject to several limitations related to sample size and domain of applicability. Despite these constraints, the proposed facial translucency evaluation method is a powerful tool to assess skin care and makeup products in clinical studies, offering both speed and robustness. In addition, this approach has the potential to be applied in retail settings, enabling consumers to evaluate their own facial skin.

References

1. A. Goldsberry, C. W. Hanke, and K. E. Hanke. “VISIA system: A possible tool in the cosmetic practice.” Journal of Drugs in Dermatology: JDD 13.11 (2014): 1312-1314.
2. E. Berardesca and European Group for Efficacy Measurements on Cosmetics and Other Topical Products (EEMCO). “EEMCO guidance for the assessment of stratum corneum hydration: electrical methods.” Skin Research and Technology 3.2 (1997): 126-132.
3. G.E. Piérard, et al. “EEMCO guidance for the in vivo assessment of skin greasiness.” Skin Pharmacology and Physiology 13.6 (2000): 372-389.
4. T. Kuwahara, “Measurement of skin translucency,” Japanese Journal of Optics 39.11 (2010): 524-528.
5. T. Seiya, et al. “The structure of consciousness and the physiological characteristics of the skin regarding skin transparency comparison between the young and middle-aged and elderly women.” Journal of Society of Cosmetic Chemists of Japan 38.2 (2004):115-124.

6. Z. Gajinov, et al. “Optical properties of the human skin.” Serbian Journal of Dermatology and Venereology 2.4 (2010): 131-136.
7. A.M. Nurani, et al. “Development of a method for evaluating skin dullness: A mathematical model explaining dullness by the color, optical properties, and microtopography of the skin.” Skin Research and Technology 29.7 (2023): e13407.
8. R. Nakamura, et al. “Development of mathematical model of skin translucency.” Journal of Society of Cosmetic Chemists of Japan 54.4 (2020): 315-22.
9. Q. Marin, et al. “Analysis of correlation and construction of a predictive model of skin transparency using parameters from digital images of the face.” Skin Research and Technology 28.4 (2022): 582-595.
10. D.C. Montgomery, E.A. Peck, and G.G. Vining. Introduction to linear regression analysis. John Wiley & Sons, 2021.
11. S. Ray, “A quick review of machine learning algorithms.” 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). 2019.
12. A.M. Turing, Computing machinery and intelligence. Springer Netherlands, 2009.
13. Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning.” Nature 521.7553 (2015): 436-444.
14. J. Schmidhuber, “Deep learning in neural networks: An overview.” Neural Networks 61 (2015): 85-117.
15. A. Krizhevsky, I. Sutskever, and G.E. Hinton. “Imagenet classification with deep convolutional neural networks.” Advances in Neural Information Processing Systems 25 (2012).
16. A. Creswell, et al. “Generative adversarial networks: An overview.” IEEE Signal Processing Magazine 35.1 (2018): 53-65.
17. J.Y. Zhu, et al. “Unpaired image-to-image translation using cycle-consistent adversarial networks.” Proceedings of the IEEE International Conference on Computer Vision. (2017).
18. B.M. Konopka, et al. “Exploratory data analysis of a clinical study group: Development of a procedure for exploring multidimensional data.” PloS One 13.8 (2018): e0201950.
19. A. Poole and L.J. Ball. “Eye tracking in HCI and usability research.” Encyclopedia of Human Computer Interaction. IGI global, 2006. 211-219.
20. A.H. Bermano, et al. “State-of-the-art in the architecture, methods and applications of StyleGAN.” Computer Graphics Forum. 41.2. (2022): 591-611.