IFSCC 2025 full paper (IFSCC2025-1419)

Innovative Approaches to Enhance and Deepen Skin Wrinkle Analysis.

Cristina Bonell [1], Haohan Peng [2], Belen Aguirre [1], Mauricio Valerio-Santiago [1], Gemma Mola [1] and Raquel Delgado [1]

[1] Lubrizol Life Science, Lipotec S.A.U., Barcelona, Spain.
[2] Lubrizol, Corporate Technology, Pudong, China.

1. Introduction

Assessing wrinkle parameters is fundamental to dermatological research and to the development of cosmetic products. However, conventional methods for evaluating the effectiveness of cosmetic treatments on the skin, such as visual inspections and manual measurements, present various challenges. Traditional approaches are prone to human error, subjectivity, and inefficiencies, with different evaluators potentially interpreting the same set of wrinkles differently, leading to inconsistent outcomes. Furthermore, the manual evaluation process is both labor-intensive and time-consuming, which can further elevate the risk of inaccuracies. Indeed, in clinical trials, precise and reliable wrinkle assessment is essential, and this is where the integration of computer vision technologies and artificial intelligence (AI) becomes crucial. These advanced technologies provide promising solutions to the limitations of traditional methods by analyzing skin images with precision and consistency that exceeds human capabilities. They can detect subtle changes in wrinkle morphology and distribution, providing more accurate and objective assessments. Additionally, these technologies can process large amounts of data quickly, making the evaluation process more efficient.

The first step to evaluate the *in vivo* efficacy of a cosmetic treatment is to ensure a well-designed clinical trial that carefully considers the panel of volunteers; effective recruitment is key to achieving accurate results. In a study aimed at assessing wrinkle improvement, it is essential to begin with a specific baseline level of wrinkles. A well-established method for this purpose is the ATLAS score classification, however, this is a manual classification method that requires substantial time and relies heavily on a professional evaluator's subjectivity [1]. Using deep learning techniques, we have developed an algorithm capable of classifying wrinkles of crow's feet area according to the ATLAS score. By expanding an internal database with wrinkle images from the crow's feet area of over 1,000 volunteers, representing a wide range of wrinkle severity, we can classify the wrinkles into six distinct ATLAS score categories with an improved model. This wrinkle classification tool allows us to quickly and efficiently recruit a panel of subjects with specific wrinkle grades.

Another important aspect when evaluating the efficacy of a cosmetic treatment in improving wrinkles is the precise analysis. Most products enhance the appearance of wrinkles through

subtle changes in various aspects, significantly improving the overall result. Although there are devices with software capable of analyzing these changes, they do not always provide quick analysis or the parameters of interest for evaluation. Therefore, we have developed two distinct methods to analyze wrinkles from 2D and 3D images.

The first method has been specifically designed based on deep learning technology to automatically detect the crow's feet wrinkles from 2D images and extract data from them. This approach employs neural networks to analyze the images with high precision, identifying and isolating the wrinkles characteristic of the crow's feet area, subsequently providing a comprehensive set of parameters that quantify various features including wrinkle number and length. The second method, based on computer vision capabilities, analyzes 3D images acquired by a commercial device. The device generates high-resolution 3D moving images, which offer a compelling and detailed visual representation of the skin's evolution over time. Our application, specifically designed for this purpose, extracts and provides quantifiable parameters of the wrinkles detected from these images, extending beyond the capabilities of the commercial device's original software. Using both methods significantly enhances the evaluation of wrinkles in quality, precision, and time, providing detailed information about their size and severity, and offering a more comprehensive understanding of their development and progression.

2. Materials and Methods

To develop the wrinkle classification and wrinkle detection models, an internal database of more than 2,000 volunteer's photographs collected from several studies was employed. Algorithms training requires a large number of images to achieve a robust performance. For the analysis tool of detected wrinkles from the 3D images a computer vision technology was applied.

2.1. Automated Detection and Isolation of Crow's Feet Area.

As a first step in the automation process for wrinkle classification and detection, the region of interest (ROI) in the crow's feet area was automatically detected and isolated using an AI-based method.

For this process, 2D RGB facial images captured by a standard camera at a 45-degree angle relative to the facial frontal view were used (Nikon D7200). The algorithm was trained to detect the eye region by comparing an input image to be evaluated with a reference image where the ROI had been manually annotated. Using image registration via template matching, the system slides a small section of the reference image over the input image to find areas of high similarity, generating a correlation matrix where higher values indicate better matches [2].

Initially, the reference image used featured closed eyes, so the model was only effective when the subject's eyes were closed. Therefore, the correct ROI prediction would not be generated for images where volunteers had their eyes open. To address this, we incorporated reference images featuring open eyes in the next phase, enhancing the model's detection capabilities.

2.2. Evaluation Through Wrinkle Grade Classification

A wrinkle classification tool was developed using an initial AI-based method. Wrinkles were graded using the ATLAS scale, which categorizes them into seven different levels, ranging from 0 (no wrinkles) to 6 (severe wrinkles).

We selected a classical artificial neural network for image classification [3,4]. This network processes images through 18 layers, with each layer learning to detect features from simple shapes to complex patterns. To address potential inefficiencies in deep networks, namely the

loss of important information, shortcut connections were incorporated using 18 layers, allowing information to skip layers, thus enhancing learning speed and accuracy.

The model development involved three stages: training, validation, and testing [5,6]. For training, which served as the instruction phase, a total of 919 images were used. For validation stage, aimed at corroborating the training, we included 272 images. Once the model was considered educated, we moved to the test stage, where 159 images were evaluated to perform a final verification of the model's performance. This data split approach follows a typical method in deep learning, with approximately 70% of the data allocated for training, 20% for validation, and 10% for testing. Table 1 shows the distribution of images used for each stage, categorized by the six wrinkle grades. The testing stage results for each category are presented.

Table 1. Data distribution of six categories.

	Categories index	Number of Images
Training	0	179
	1	243
	2	152
	3	165
	4	133
	5	47
Validation	0	51
	1	69
	2	43
	3	58
	4	38
	5	13
Test	0	26
	1	36
	2	23
	3	47
	4	19
	5	8

Activation maps provide visual representations of how the model processes images [7]. As the model analyzes an image, it identifies features such as edges, textures, or patterns. Early layers focus on simple features like lines or colors, while deeper layers capture more complex structures, such as parts of objects. These activation maps show the specific areas of the image that the model focuses on at each stage of its processing.

Figure 1 shows the example of an input image (a) and part of the activation map (b), presenting network layer's impact on the image. Each row of the activation map shows the examples of the filters from the first few convolutional layers of the network. As the depth of the neural network layer increases, features become more prominent, focusing more on the wrinkles, the key feature for classification. This progression illustrates how the model learns to emphasize finer details, like wrinkles, as it processes the image through deeper layers.

Through the examination of these maps by a trained user, we can ensure that the model makes predictions based on the relevant parts of the image.

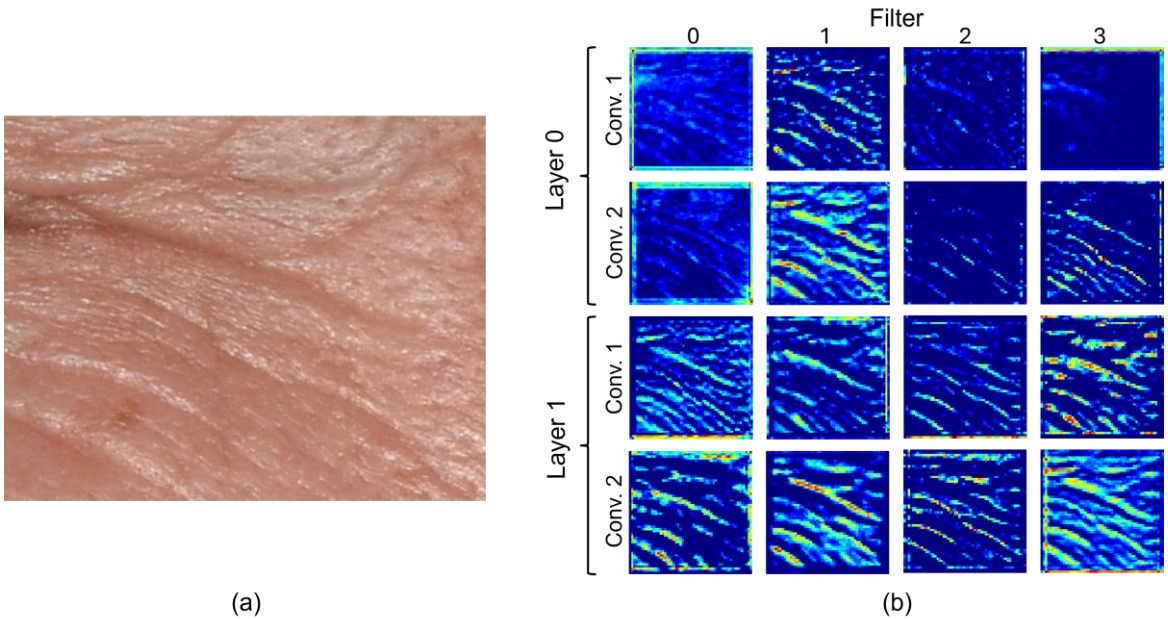


Figure 1. Representation of activation map: (a) Example of input image; (b) A section of the activation map showing layer's impact.

2.3. Detailed and Systematic Wrinkle Detection for Specific Parameters Evaluation.

For training a wrinkle detection method, an initial dataset of images and masks was constructed. Precise mask generation is necessary to indicate to the network where the wrinkles are, but creating these masks can be challenging. Several image processing techniques, such as frequency filtering, were tested to automate mask creation. However, variations in skin brightness and shadows caused by cavities of the face differed among volunteers, making it difficult to find a single process suitable for all images, thereby complicating accurate mask generation.

To ensure precise mask generation, wrinkles were annotated manually using a basic image software. This software allows the user to open images, create a layer, and draw wrinkles on that layer to obtain a wrinkle skeleton. We created 120 layers of images of crow's feet wrinkles, which were exported as masks to include in the dataset. Additionally, 42 images without wrinkles were added.

Once the masks were created, a specialized neural network was chosen to develop the wrinkle detector tool, enhancing accuracy by refining details [3], combined with a powerful image-processing backbone to efficiently recognize patterns [3,4]. This model quickly learns to distinguish different areas within an image with high precision, making it ideal for wrinkle segmentation.

The method was developed by initially augmenting the dataset to improve the training process. Data augmentation was implemented to enhance the quality and diversity of the training set. For the validation and test sets, only horizontal flipping was applied to maintain the realism of the images. The training set underwent various augmentations, including horizontal flips, random color changes, random brightness changes, and random contrast changes. These changes were subtle to avoid unrealistic alterations. After augmentation, the validation and test sets were doubled to 48 and 50 images, respectively, and the training set expanded to 565 images. Figure 2 shows examples of different data augmentations.

Figure 2. Image examples of data augmentation.

2.4. Enhancing Wrinkle Parameter Analysis from 3D Image Assessments

Numerical information from 3D images (LifeViz® Infinity, QuantifiCare) was obtained, improving the understanding of wrinkle features. The 3D images exported from the software's camera show a colormap of wrinkles illustrating their depth (Figure 3). Using the k-means algorithm, an unsupervised clustering algorithm, the colormap can be converted into different categories [8]. This algorithm allows quantizes colored pixels to detect varying depths of wrinkles, allowing for the measurement of the area corresponding to each level and facilitating the extraction of relief information (Figure 4a). Additionally, we created masks showing the wrinkles and their corresponding skeletonization (Figure 4b). This process reduces a binary image to a single-pixel-wide skeleton while preserving structural integrity, enabling accurate measurement of wrinkle parameters by counting pixels within the skeleton.

Furthermore, by generating bounding boxes around each wrinkle, we enabled the identification of forehead and frown wrinkles. By calculating the horizontal and vertical distances of boxes, the direction of wrinkles is determined: forehead wrinkles are identified when the horizontal dimension exceeds the vertical, while frown wrinkles are recognized when the vertical dimension is greater.

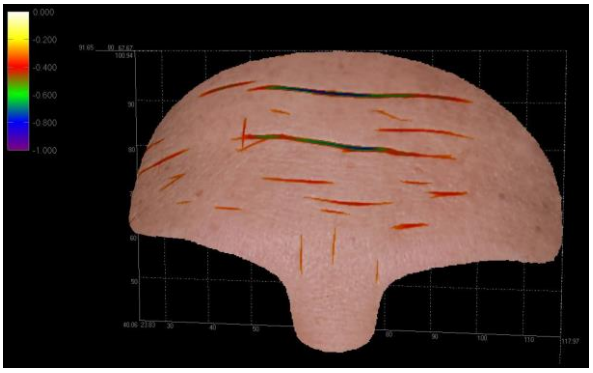
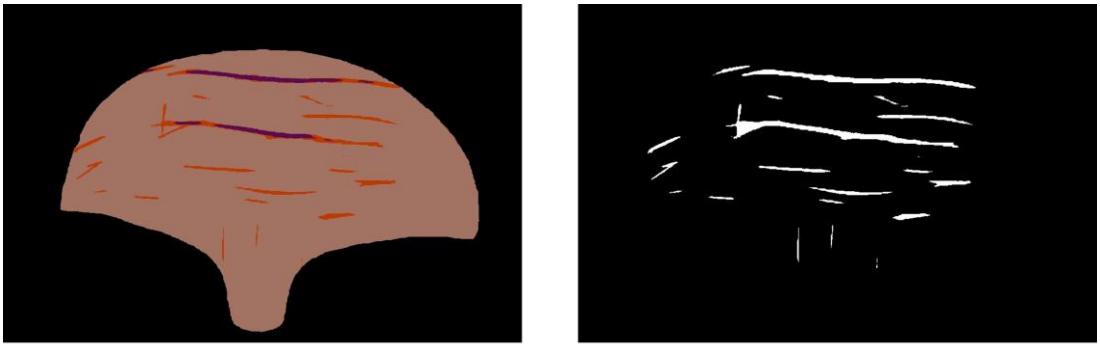


Figure 3. Image from LifeViz® Infinity device.



(a) (b)

Figure 4. (a) Quantification of colors with k-means; (b) Mask generated with wrinkles.

3. Results

3.1. Outcomes of the ROI Detection Method

The images obtained through the ROI detection method were visually assessed. The ROI images focusing on the crow's feet area were accurately cropped, capturing the desired wrinkle region while excluding parts of the eye and eyebrows. Figure 5 presents an example of the algorithm`s output for ROI detection indicating successful matches. The green square highlights the automatically detected eye area, while the yellow square indicates the ROI of the crow's feet region, which will be cropped subsequently for further analysis.

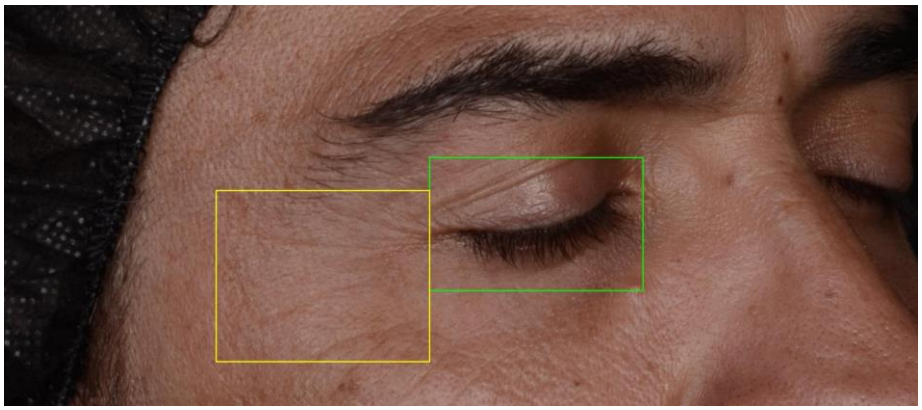


Figure 5. Image example of ROI detection algorithm's output.

3.2. Effective Performance of the Wrinkle Classification Model

The confusion matrix is a tool used to evaluate the performance of a classification model, as it provides a detailed comparison between the model's predictions (Predicted Label) and actual values (True Label) [6]. Figure 6 shows the confusion matrix of the wrinkle classification model, enhanced with a blue heatmap that effectively visualizes the frequency of predictions for each class. The cells along the central diagonal (in dark blue) represent the true positives (TP), while the off-diagonal cells correspond to false negative and positive predictions. Given the percentages shown and an accuracy rate of 68%, we can consider that the model performs well in classifying wrinkles into different grades. This is reasonable considering the challenge of distinguishing between six different grades due to the subtle visual variability among them. These differences can be minimal and difficult to detect, even for human experts.

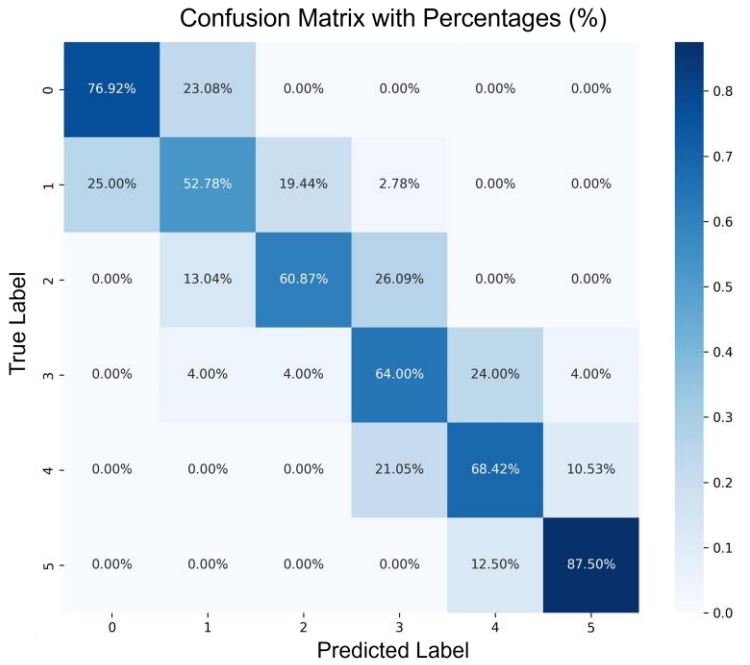


Figure 6. Confusion matrix 6-class classifier based on ATLAS score categories.

3.3. Optimized Efficiency of the Wrinkle Detection Model

The performance of the model was evaluated using the Intersection over Union (IoU) metric [9], which provided crucial parameters to justify the accuracy of the wrinkle detection. IoU measures the ratio of the intersection to the union of the predicted and reference masks, with a value of 1 indicating a perfect match and 0 indicating non-correlation. IoU was measured across the test set images, obtaining an average score of 0.8963, which corresponds to strong performance by the network. Figure 7a, shows an example of an automatically obtained ROI of the crow's feet area, and Figure 7b illustrates the wrinkles skeleton identified by the trained algorithm.

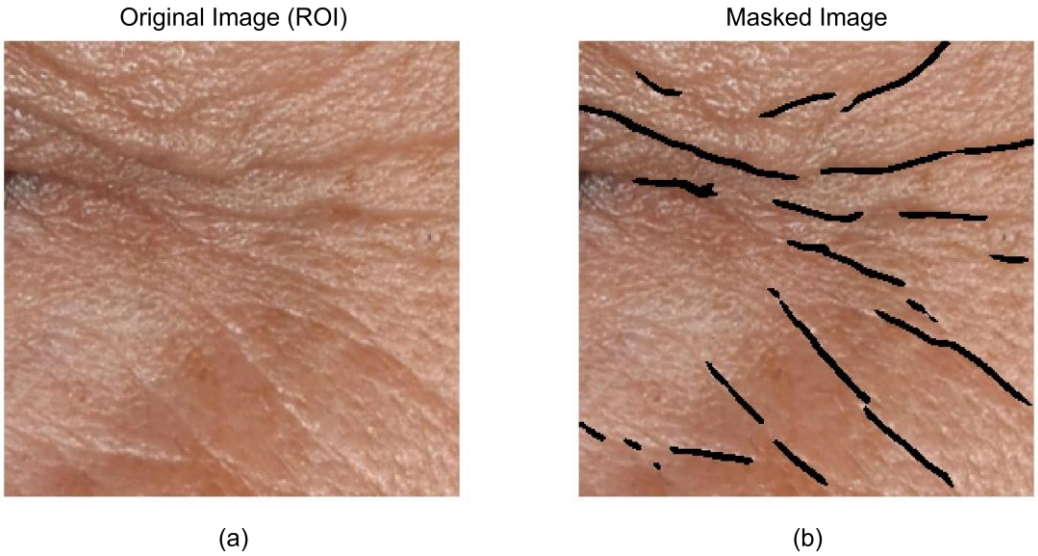


Figure 7. Images of wrinkles prediction process: (a) Image of ROI obtained of crow's feet area; (b) Mask of wrinkle skeleton created by the model.

3.4. Exemplary Results of 3D Image Evaluation

Several parameters were obtained by analyzing the processed images, including wrinkle count, length, total area covered, wrinkle depth (orange and purple areas), and their positions on the forehead and frown regions. Figure 8 shows the bounding boxes surrounding the wrinkles of the mask created in Figure 4b. In this case, wrinkles are identified, and their lengths and positions are illustrated. Table 2 presents the results obtained from the image analysis.

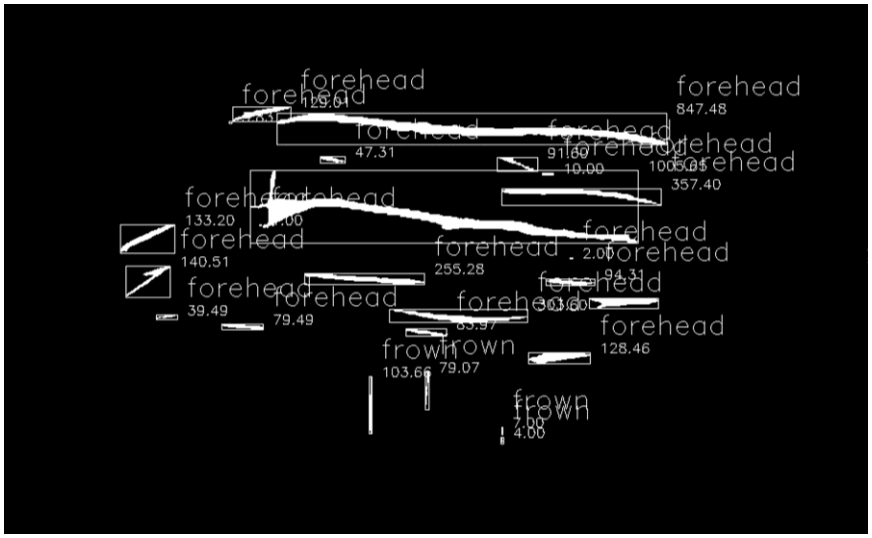


Figure 8. Illustration of wrinkle length measurement and classification of forehead and frown wrinkles.

Table 2. Results of wrinkle parameters obtained from Figure 3 after imaging processing.

Parameter	Value
Total wrinkle number	27
Frown wrinkle number	4
Forehead wrinkle number	23
Mean wrinkle length (px)	140.59
Frown wrinkle length (px)	48.43
Forehead wrinkle length (px)	170.49
Total wrinkle area (%)	5.34
Orange wrinkle area (%)	3.77
Purple wrinkle area (%)	1.57

4. Discussion

The aim of this work was to develop tools that not only reduce the time and effort required for manual image analysis but also enhance the accuracy and precision of the wrinkle assessments. The effectiveness of the presented applications underscores a strong emphasis on accurate wrinkle evaluation, enabling a more comprehensive and reliable analysis that enhances our understanding of the complex nature of wrinkles.

One of the goals was to improve the recruitment of volunteers for clinical trials focusing on crow's feet assessment. By creating the automated wrinkle classification model, the recruitment process for clinical trials becomes more efficient and faster than manual and objective classification. In addition, new subjects are continually added to the volunteer databases, and new images are generated from each study conducted, reflecting changes in their wrinkle grades over the years. The implementation of this model helps maintain updated information

on the volunteers. Moreover, regarding the volunteer databases, correlating information about wrinkles grades and various aspects of the subjects can be done more easily, facilitating the expansion of knowledge about consumer needs and allowing a more targeted approach in future research.

As a next step, the classification could be expanded by adding more images of advanced age subjects to the dataset and retraining the algorithm to include grade 6 wrinkles. This approach could enhance the model's performance and accuracy.

This study also describes the development of models capable of accelerating and/or expanding wrinkles assessments from both 2D and 3D images. By automatically detecting the crow's feet area from 2D images, the model reduces at least five hours of manual effort required for cropping the ROI in a standard clinical trial. Therefore, the algorithm's ability to consistently detect the ROI represents a significant advancement in the evaluation process. Moreover, beyond this advantage, the model is able to detect and analyze crow's feet wrinkles for further quantification, enabling a comprehensive analysis of large number of images with a single click. On the other hand, the latest application presented, powered by k-means algorithm, provides quantifiable parameters from 3D images, offering insights that exceed those delivered by the device's commercial software. In future work, these two applications will be adapted to analyze additional facial wrinkles.

5. Conclusion

In our study, new techniques for analyzing wrinkles from different type of images offering detailed insights with consistent and reliable results that some devices cannot provide by using computer vision and AI. The implementation of these technologies in wrinkle assessment represents a significant advancement in dermatological research and product development. It addresses the challenges of traditional methods, offering more accurate, efficient, and consistent results, which are crucial for evaluating the efficacy of cosmetic or dermatological treatments.

6. References

1. Bazin R, Doublet E. Skin Aging Atlas. Volume 1, Caucasian Type. Editions Med'Com; 2007.
2. Open cv Docs [Internet]. [Cited 2025 March]. Available from: https://docs.opencv.org/4.x/d4/dc6/tutorial_py_template_matching.html.
3. Computer Vision and Pattern Recognition. [Internet]. arXiv; [cited 2025 March]. Available from: https://arxiv.org/list/cs.CV/recent.
4. Kaiming H, Xiangyu Z, Shaoqing, Jian S. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;770-778.
5. Larose D, Larose C, Discovering Knowledge in Data: An Introduction to Data Mining. 2nd ed. Jhon Wiley & Sons, Inc. 2010(134)1-35.
6. Singh P, Singh N, Singh K, Singh A. Machine Learning and the Internet of Medical Things in Healthcare: Diagnosing of disease using machine learning. Academic Press. 2021;89-111.
7. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2921-2929.

8. Dhanachandra N, Manglem K, Chanu YJ. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science. 2015; 54:764-771.
9. Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized intersection over union. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019;658-666.