A Cluster Validity Index-Based Objective Criteria for Aesthetic Evaluation of Periorbital Treatment

Çağlar Cengizler1,*, Ayse Gul Kabakci2,*, Dursun Murat Bozkır3,*, Dilek Sire Eren4,*, Memduha Gülhal Bozkır2,*

1Biomedical Device Technology Program, Vocational School of Health Services, Izmir Democracy University, Izmir, Turkey; 2Department of Anatomy, Faculty of Medicine, Cukurova University, Adana, Turkey; 3Adana Sevgi Eye Center, Adana, Turkey; 4Dr Dilek Eren Medical Aesthetic Clinic, Adana, Turkey

*These authors contributed equally to this work

Correspondence: Ayse Gul Kabakci, Cukurova University Faculty of Medicine Department of Anatomy, Sarıçam, Adana, Turkey, Tel +90 322 338 60 60-3489, Fax +90 322 338 65 72, Email akabakci@cu.edu.tr

Purpose: Dark circles and pigmentation around the eyes are common reasons people see dermatologists. An effective assessment of the severity of infraorbital color and texture differences is critical for determining appropriate treatment. Evaluation of the visual severity of cases is mostly based on visual inspection. Treatment efficiency is often measured using patient questionnaires in many cases. The subjectivity of assessments may lead to a prolonged healing process, misdiagnosis, and excessive use of drugs or chemicals.

Patients and Methods: In this study, a computer-aided objective evaluation approach was proposed for grading periorbital facial rejuvenation. This approach is based on the analysis of numerical features extracted from different facial regions in digital images. Healing was objectively graded by evaluating data clusters formed using the extracted features. Accordingly, an increase in the visual similarity between paired facial regions is accepted as an indicator of healing, which directly affects the form of data clusters. An intracluster validity index was measured to evaluate the clusters as dense and well separated. A total of 144 facial regions were extracted and examined, and the automatically calculated grades were compared with expert evaluations.

Results: The cosmetic effects of the experimental drug were evaluated during the experiments, and expert grades were accepted as the ground truth. The results show that the proposed automated grading approach can evaluate rejuvenation with an accuracy of up to 0.91 accuracy in the upper orbital region.

Conclusion: This study concluded that the proposed data-clustering-based approach is promising and can be functional without any special instruments.

Keywords: colour cosmetics, computer modeling, skin structure, healing

Introduction
Periorbital hyperpigmentation, or lines and wrinkles around the eye area, are generally not a cause of medical concern. However, cosmetic differences in the eye and facial subunits can significantly influence perceptions of a woman’s age and judgments of attractiveness. Factors such as pigmentary disorders and loss of elasticity can lead to cosmetic concerns around the eyes. Consequently, dark circles are among the most common reasons for dermatological consultations. Their etiology is complex and can occur across all age brackets. Assessing the severity of color and texture differences around the eye region is critical for determining appropriate treatment. One study assessed treatment efficacy using questionnaires. Researchers have reported the effects of using non-ablative fractional lasers. In a similar study, the efficacy and tolerability of an eye cream were evaluated using a self-assessment questionnaire completed by the subjects. Researchers in another study evaluated the treatment of periorbital hyperpigmentation. The first treatment step was based on clinical examination to identify the components of periorbital hyperpigmentation. Subjects’ self-assessments similarly evaluated improvement. The appearance of the region around the eye was primarily evaluated based on assessments from subjects and experts. Thus, evaluating a new treatment can be subjective and error-prone,
potentially slowing development. Computer-aided assessment of the relative darkness of the periorbital region is one of the most studied subjects in related literature. For example, an image-processing approach has been proposed to numerically measure the degree of dark circles on the lower eyelid. This image-processing approach is based on splitting an adjusted original color image into an RGB channel image.\textsuperscript{11}

Moreover, in another study, a model for morphing images was proposed. Accordingly, each spectral image was constructed using a morphing procedure where the channels of red, green, and blue are used in conventional color-morphing.\textsuperscript{12} A methodology from a particular study may be highlighted as an example of a digital image analysis-based approach. This study’s approach includes acquiring the reflectance spectra of the color phantoms. The same study produced qualitative maps to reveal local variations in the skin chromophores and blood vessels.\textsuperscript{13}

Previous studies point out that patient satisfaction scores measured with questionnaires can be affected by internal and external factors, and subjectivity in the assessment of facial rejuvenation may lead to a prolonged healing process, misdiagnosis, and excessive use of drugs or chemicals. Accurate quantification and objective assessment of clinical signs are essential for the development of refined methodologies and for enhancing the precision of treatment strategies. Accordingly, a computational approach would be an effective tool for aesthetic medical procedures. In this study, we aimed to reveal the numerical effects of periorbital color change on digital images to present an objective methodology for aesthetic evaluation of cosmetic treatments.

 Accordingly, appearance-characterizing features were extracted from the regions of the periorbital area on the digital images. These numerical features were used to represent each region as a data cluster, and an intra-cluster index was calculated to measure the healing.

**Materials and Methods**

**Dataset**

The dataset consisted of images collected from 12 subjects. All images were captured using a full-face imaging system (Observ 520, InnoFaith Beauty Sciences, Eindhoven, Netherlands). Six photographs were taken of each patient during four consecutive sessions. This study conducted a numerical analysis on cross-polarized skin analysis mode photos from the first and last sessions. Figure 1 displays photographs from the original data. In addition to numerical analysis, experienced dermatologists visually evaluated the images to form an empirical ground truth. The scoring was based on the validated photo-numeric scale for dark circles proposed by O’Mahony et al. They used a 10-point photo-numeric scale for the clinical evaluation. Their scale grades 0 for no dark circles and 9 for extensive, severe dark discoloration.\textsuperscript{14}

Before the photographs were taken, each participant was asked to relax. They were then positioned so that a line from the upper edge of their external ear canal to the lower border of their eye socket was horizontal while standing. The Frankfort horizontal plane was selected as the reference plane to determine the angles related to the face.\textsuperscript{15}

The patient population included that treatment for periorbital facial rejuvenation was applied to female subjects chosen from a 30–55 years age group. All subjects agreed not to change their habits regarding food, physical activity, makeup use, face cosmetics, and cleansing products. Additionally, they were advised to avoid intense UV irradiation of the face, such as UV sessions or sunbaths. Two cosmetic criteria were considered for inclusion. One criterion was mild/moderate cutaneous photoaging according to a photographic reference scale, and the second was the skin phototype, which should be types II and III according to Fitzpatrick’s skin scale.

Each session was completed with three micro-injections into the periorbital area. The treatment sessions were repeated every ten days. The aesthetic results were documented at the end of each session using digital images for further evaluation.

Subjects who underwent plastic surgery, facelift, biomaterial implementation, Botox injection, chemical peeling, or laser treatment 12 months before the study start date were excluded. Also, participants were excluded if they were allergic to the product ingredients, pregnant, breastfeeding, had permanent fillings, did not use proper birth control, or were not in menopause. Moreover, body mass index was also considered for inclusion (variation should not exceed $\pm 1$ during the study period).
Applied Drug

The Sunekos® 200 was used during the experimental sessions. It is an injectable treatment that has been identified as a Class III medical device. It consists of small bottles containing 100 mg of sterile and apyrogenous lyophilized glycine, l-proline, l-leucine, l-lysine HCl, L-Valine, and l-alanine, and sterile vials containing sodium hyaluronate (30 mg in 3 mL of distilled water). The product was manufactured and distributed by Professional Dietetics S.p.A. It was created to correct the photoaging/aging face and body signs (light and medium degrees).\textsuperscript{16}

Data Clustering Approach

This study aimed to investigate the correlation between numerical features and cosmetic healing. It would be possible to accept that one of the most typical indicators of healing is the increasing visual similarity between close facial regions. Therefore, the similarity between the target periorbital regions (green and red masks in Figure 2) and the closest infraorbital area (blue mask in Figure 2) was evaluated. It was assumed that the increasing similarity between these regions indicated the efficiency of the periorbital rejuvenation treatment. Accordingly, several visually characteristic features were extracted from these regions for comparison. Subunits of the face were represented by data clusters formed using these features. The similarity was graded according to the compactness and separability of the data clusters. Accordingly, the evaluation was performed twice using digital images taken after the first and last treatment sessions to assess the effect of the applied drug.

An expert manually marked 12 subunits for each patient; half were segmented from the photograph of the first session, whereas the other half were segmented from the photograph of the last session.

A digital image is a two-dimensional function for a computer that outputs an intensity or color space value where inputs are spatial coordinates.\textsuperscript{17} It is also possible to classify patterns or textures on an image by analyzing pixels or pixel groups according to their locations or features. Moreover, it is possible to characterize regions with numerical features. This study implemented a mesh grid approach, allowing evaluation of the color and texture of subunits with identical

Figure 1 Sample regions extracted from photographs of the first and last sessions from the original data.
windows. Accordingly, the features were extracted from each grid window instead of a single pixel, and each window was an observation of a related cluster.

According to the implemented approach, well-formed and separable subunit clusters indicate that the effect of the applied drug on healing is low because visual likenesses should cause clusters to be close and intertwined. The cluster forms were numerically evaluated using a well-known cluster validity index to present an objective grading mechanism.

Extracted Features
The visual appearance of each segmented area was defined by seven features that characterized the texture and color of the region. It should be noted that energy, homogeneity, contrast, and correlation features were calculated using a gray-level co-occurrence matrix (GLCM). It denotes the frequency of occurrence of a pixel with gray-level value $i$ horizontally adjacent to a pixel with value $j$. All the extracted features are explained below.

Energy
The energy (uniformity) value of an image region measures the localized change. It was measured as the sum of the squared elements in the GLCM. Accordingly, the energy is formulated as follows:

$$\text{Energy} = \sum_{ij} p(i,j)^2$$  \hspace{1cm} (1)
Homogeneity
Homogeneity indicates the existence of strong contrasts within a region.\textsuperscript{18} High homogeneity indicates that the pixel values in the region are close to each other. It was measured by:

\[ Homogeneity = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \]  
\hspace{7cm} (2)

Contrast
This denotes the intensity contrast between any pixel and its neighbor, which is formulated as

\[ Contrast = \sum_{i,j} |i - j|^2 p(i,j) \]  
\hspace{7cm} (3)

Correlation
This defines the correlation between a pixel and its neighbor over the entire image. The formulation of this feature is as follows.

\[ Correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\sigma_i \sigma_j} \]  
\hspace{7cm} (4)

Mean Gray-Level
The average gray level of the sampled window is calculated.

Maximum Value of Hue and Saturation
The sampled window was converted into the hue, saturation, and value (HSV) color space. This space characterizes colors in terms of shade and brightness. Accordingly, the maximum value for each window was measured.

Minimum Saturation
In addition to the maximum value of the hue and saturation, the minimum saturation value of each sampled window was measured.

Different regions of the face and the feature space formed with the two extracted features are shown in Figure 3.

Cluster Evaluation
In this study, the Calinski-Harabasz (CH) index was calculated as an internal cluster validity measure that grades the similarity of facial regions depending on how large inter-cluster distances and proximity of intra-cluster distances.\textsuperscript{19} It was formulated as follows:

\[ CH(k) = \frac{B_c(k)}{(K - 1)} \bigg/ \frac{W_c(k)}{(n - K)} \]  
\hspace{7cm} (5)

Where K represents the number of clusters, and n represents the number of observations. \(B_c\) and \(W_c\) represent the between-group and within-group dispersions, respectively. \(B_c\) was calculated as follows:

\[ B_c = \sum_{k=1}^{K} nk |\overline{C}_k - \overline{x}|^2 \]  
\hspace{7cm} (6)

where nk represents the number of observations in the cluster, \(\overline{C}_k\) and \(\overline{x}\) denotes the centroid of the cluster and the overall centroid, respectively. Finally, \(W_c\) was calculated as follows:

\[ W_c = \sum_{i=1}^{n} |x_{ik} - \overline{C}_k|^2 \]  
\hspace{7cm} (7)
Where \( n_k \), \( x_{ik} \), and \( C_k \) are the number of observations in cluster \( k \), the \( i \)-th observation of cluster \( k \), and the centroid of cluster \( k \), respectively.

This measure evaluates the quality of a clustering solution, which depends on the cluster compactness and the proximity of the intra-cluster distances.

Accordingly, a similar visual appearance corresponds to a lower clustering quality and CH index score.

**Results**

The original dataset includes two photos for each face, taken before and after the initial and final sessions of the drug application. Each photo was manually segmented into six regions: three from the right side and three from the left side of the face. Analyses were conducted based on the features extracted from these segments. In total, 144 regions from 12 individuals were analyzed, and the results were compared to the expert photo-numeric scale grades. It is important to note that the CH score was calculated for the lower and upper orbital subunits. Each subunit was compared to the infraorbital region. The average CH score from the left and right measurements was considered the overall numeric score.

The average CH and expert grades are presented in Table 1, which reveals a potential relationship between the CH scores and empirical expert grades. Furthermore, a binary approach has been used to uncover the correlation between the effects of the applied drug and the CH score. This method is also valuable for assessing the practical relevance and compatibility of the extracted features with the CH score. In this approach, a decrease in the photo-numeric scale grade or CH score at the end of the sessions is indicated by one and is recognized as a cosmetic improvement.

![Figure 3](https://doi.org/10.2147/CCID.S425797)

Cengizler et al. Dovepress

**Figure 3** A sample feature space of different facial regions is plotted. Accordingly, A and C regions are visually dissimilar.
If the score did not decrease in the last session, the treatment was considered ineffective and indicated by a 0. Subjects 1 through 6 were observed to have uniformly positive results across all tests. Subject 7 differed with a negative Empiric and Upper Orbital result. Subject 8 was consistently negative in both human and computer analyses. Subject 9’s computer analysis was mixed, but the Empiric Evaluation was negative. Subject 10 showed negative results across the board. Subject 11 was consistently positive, while Subject 12’s computer analysis indicated a negative Upper Orbital but was positive otherwise.

Photo-numeric scale grades served as the ground truth measures for evaluating the effectiveness of the Sunekos® 200. These grades were also utilized as the ground truth in comparing binary healing results, aiming to determine the correlation between computational and empirical outcomes objectively. Metrics such as F-score, accuracy, sensitivity,
specificity, precision, and recall were employed to gauge the results’ similarity objectively. These metrics are widely accepted for comparing binary decisions. The results are presented in Table 2.

**Discussion**

A numerical grading approach based on digital image analysis was proposed and evaluated to assess healing in the periorbital region objectively. The implemented mechanism measures the variance in cutaneous appearance by numerically comparing adjacent subunits of the face, which means that the reference point for each evaluation was derived from the photographic data of the same face. Additionally, an innovative data-clustering approach was introduced to measure the treatment’s effectiveness numerically, aiming to reduce subjectivity. Many previous studies with similar motivations require specific imaging instruments.

It should be noted that the methods presented in this paper can operate effectively on digital images and may serve as an alternative to manual inspection.

The assumption was that well-formed clusters of color- and texture-characterizing features may be indicators of the visual dissimilarity of subunits around the eye. Accordingly, the CH index value was calculated to compare clusters of facial areas close to each other and was considered a measure of treatment efficiency. Table 2 reveals that in most healing cases, CH values consistently decrease, leading us to conclude that the CH index and manual grades correlate. Additionally, cosmetic improvement was denoted with binary values to compare the calculated results with empirical photo-numeric evaluation. The F-score metrics presented in Table 2 indicate that the calculated values were close to the expert evaluation values, and the proposed validity index was compatible with the numerical grading of treatment efficiency.

A popular approach for quantifying the degree of erythema and pigmentation around the eye region involves the evaluation of the color space. In which, features that characterize not only the color of the region but also its texture were extracted in this study. A number of well-known features were extracted in this study. It should be noted that the introduced mechanism in the study is highly dependent on the significance of the features, which means that the sensitivity of the applied algorithm to cosmetic variance would increase with the significance of extracted features. The grading ability of the applied approach can be increased with additional features that characterize the visual appearance of the periorbital area.

Instead of analyzing the entire surface of the facial region, the area was analyzed with a mesh grid. This approach allowed us to selectively filter the effect of local dents, marks, or irrelevant cosmetic distortions during the overall evaluation. However, the implemented code lacks a preprocessing stage, which would help to suppress possible irrelevant data better on photographs before further analysis.

The original dataset consists of a sufficient amount of photos taken from individuals to demonstrate the functionality of the proposed approach. However, the scope of the dataset could be critical for developing the proposed method for practical applications. A larger dataset, formed with a diverse group of individuals, would enhance robustness.

Furthermore, each image was treated as an independent evaluation source; thus, regions from different images were not compared for healing grades. The textural and color features of subunits around the eye in each photograph were matched against those of a nearby facial infraorbital unit. Consequently, minor photographic variations in digital images from different sessions did not influence the assessment.

The proposed numerical grading method depends on the features of manually segmented regions. Segmentation was performed to determine the contours of subunits for empirical analysis. An automatic segmentation approach would improve the overall functionality by reducing the time and effort required for the proposed evaluation.

<table>
<thead>
<tr>
<th>Table 2 Objective Comparison of Expert Grades and CH Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Lower Orbital</td>
</tr>
<tr>
<td>Upper Orbital</td>
</tr>
</tbody>
</table>

https://doi.org/10.2147/CCID.S425797

Clinical, Cosmetic and Investigational Dermatology 2023:16

2544

Cengizler et al

DovePress

Powered by TCPDF (www.tcpdf.org)
Conclusion
This study is the first attempt to use a cluster validation index to grade periorbital dark circles. The proposed system has been successfully tested in a clinical trial to demonstrate its grading ability. The ground truth was based on visual inspection, and the results showed a strong correlation between manual grading and calculated scores. Therefore, assessing the proposed data-clustering-based approach as effective and not requiring any particular instrument should be possible. In the future, a fully automated grading system based on the proposed approach will be implemented to rapidly and accurately assess the efficiency of cosmetic treatments.

Ethical Statements
This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Written informed consent was provided by the patients for the case details and accompanying images to be published. The protocol was approved by the Non-interventional Ethics Committee of Cukurova University (no.122/25).

Acknowledgments
We thank Professional Dietetics S.p.A. (Via Ciro Menotti 1/A – 20129 Milano) for their important contributions. We would also like to thank the medical aesthetic physicians Dr. E. Gizem Geylani and Dr. Can Mustafa Eren. While preparing this work, generative AI was used to refine the language usage. The authors reviewed and edited the content, taking full responsibility for the publication's content.

Funding
This study funded by Professional Dietetics S.p.A. (Via Ciro Menotti 1/A – 20129 Milano, Italy). Funder website: www.professionaldietetics.com. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure
The authors report no conflicts of interest in this work.

References
2. Fink B, Grammer K, Matts PJ. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. Evolut Hum Behav. 2006;27(6):433–442. doi:10.1016/j.evolhumbehav.2006.08.007


