

A Deep Learning-Based Automated Scoring System for Predicting Eyelid Rejuvenation Outcomes After Monopolar Radiofrequency Treatment

Dong Hye Suh¹, Sang Jun Lee¹, In Yong Kim², Byung Chul Jang³, Hei Sung Kim⁴ 

¹ArumdaunNara Dermatologic Clinic, Seoul, Korea; ²Switz Dermatologic Clinic, Seoul, Korea; ³School of Electronics and Electrical Engineering, Kyungpook National University, Daegu, Korea; ⁴Department of Dermatology, Incheon St. Mary's Hospital, The Catholic University of Korea, Seoul, Korea

Correspondence: Hei Sung Kim, Department of Dermatology, Incheon St. Mary's Hospital, The Catholic University of Korea, 56 Dongsuro, Bupyeong-gu, Incheon, 150-713, Korea, Tel +82-32-280-5105, Email hzlckimhoho@gmail.com

Background: Upper eyelid rejuvenation with monopolar radiofrequency (MRF) is a minimally invasive option for patients with eyelid laxity. However, outcomes vary widely, and conventional evaluation methods rely on subjective photographic assessment and physician judgment, which are prone to observer bias and limited reproducibility. This lack of standardized, objective outcome measures complicates treatment planning and patient counseling.

Objective: To develop and validate a deep learning-based automated scoring system for predicting and assessing clinical outcomes following eyelid MRF treatment.

Methods: A retrospective, multicenter study of 50 patients (47 women, 3 men) treated with eyelid MRF was conducted. Pre- and post-treatment images were used to train a hybrid model combining a convolutional neural network (CNN) and U-Net architecture. The U-Net performed periorbital segmentation, while the CNN generated quantitative improvement scores. Ground truth ratings were provided by five board-certified dermatologists. Model performance was evaluated using root mean square error (RMSE) and mean absolute percentage error (MAPE).

Results: The CNN-U-Net model achieved a RMSE of 0.4 and a MAPE of 0.08, with predicted scores closely aligning with dermatologist evaluations. No significant differences in predictive accuracy were observed across patient age or sex subgroups.

Conclusion: This proof-of-concept study demonstrates the feasibility of an automated deep learning-based scoring system for eyelid MRF outcomes. By providing objective, consistent, and reproducible evaluations, the system has the potential to enhance patient counseling, guide individualized treatment planning, and enable standardized research comparisons across clinics. Larger and more diverse datasets with longer follow-up are needed for further validation.

Keywords: deep learning, automated scoring, eyelid rejuvenation, monopolar radiofrequency, aesthetic dermatology

Introduction

Upper eyelid laxity is a common aesthetic concern that can negatively affect both appearance and self-confidence. While surgical blepharoplasty remains the gold-standard, it carries risks such as under correction, overcorrection, hematoma, infection, and scarring. Consequently, many patients and clinicians turn to non-surgical alternatives such as fractional lasers, chemical peels, and monopolar radiofrequency (MRF).^{1,2} These treatments are appealing because of their minimally invasive nature and short recovery time. However, outcomes following MRF are highly variable, and not all patients achieve satisfactory improvement.³

An additional challenge is the delayed biological response to MRF. Collagen remodeling and neocollagenesis can continue for 3–6 months after treatment, meaning that early assessments may not accurately reflect final outcomes.⁴ This

temporal variability, combined with individual differences in eyelid anatomy, skin thickness, and aging patterns, complicates treatment prediction. Patient dissatisfaction is common when results fail to meet expectations, underscoring the need for reliable predictive and evaluative tools.

Current evaluation methods rely heavily on clinical photography and physician judgment. These subjective assessments are susceptible to observer bias, inter-rater variability, and poor reproducibility, limiting their usefulness in both clinical and research settings. Discrepancies between physician assessments and patient perceptions further complicate treatment planning. Without standardized, quantitative outcome measures, clinicians struggle to provide accurate prognoses, patients may develop unrealistic expectations, and cross-study comparisons remain inconsistent.

Artificial intelligence (AI) and deep learning—particularly convolutional neural networks (CNNs) and U-Net architectures—have demonstrated high performance in medical imaging by enabling automated, objective, and reproducible analysis. Applying such methods to aesthetic dermatology could address critical unmet needs. Automated scoring of eyelid rejuvenation outcomes could directly enhance patient care by providing reliable, data-driven assessments, while improving research comparability and reducing observer bias.

This study aimed to develop and validate a deep learning–based automated scoring system that integrates U-Net segmentation with CNN regression to objectively assess upper eyelid rejuvenation outcomes after MRF. We hypothesized that this system would produce scores closely aligned with expert dermatologist evaluations while offering reproducibility and translational utility in clinical practice.

Method

Study Design and Ethics

This retrospective, multicenter study included 50 patients (47 women, 3 men; age range: 30–80 years; Fitzpatrick skin types III–V) who underwent upper eyelid tightening using MRF (Thermage[®] Inc., Hayward, California, USA).

Given the retrospective nature of the study and the use of de-identified data, the requirement for written informed consent was waived by the Institutional Review Board (IRB) of Incheon St. Mary's Hospital, The Catholic University of Korea (IRB number: OC25RISI0100).

Treatment Protocol

Each patient received a single session targeting the upper eyelid region (from the lash line to the inferior eyebrow). The 450 REP 0.25 cm² eye tip was used. Energy delivery parameters ranged from 0.25–0.45 joules per pulse, with approximately 200 pulses per eyelid, adjusted according to patient tolerance. The protocol ensured uniform coverage of the treatment area.

Clinical Photography

Standardized photographs were taken at baseline and at 8 weeks post-treatment using a high-resolution digital camera under cross-polarized lighting conditions. Patients were positioned at a fixed distance with neutral expression and eyes in primary gaze. All photographs were captured under identical camera settings (aperture, exposure, and ISO). The 8-week interval was chosen to capture early clinical effects, though maximal MRF results typically appear after 3–6 months.

Expert Evaluation

Five board-certified dermatologists independently rated paired pre- and post-treatment images using a 5-point ordinal improvement scale: 1 = No improvement (0%); 2 = Minimal (<25%); 3 = Good (25–50%); 4 = Very good (51–75%); 5 = Excellent (>75%). Evaluation was conducted in a blinded manner without patient information or treatment details. Inter-rater reliability was calculated, and consensus scores were used as the ground truth for model training.

Deep Learning-Based Automated Scoring System

A deep learning model was developed to quantify upper-eyelid rejuvenation following MRF treatment. The system combined periorbital segmentation (U-Net) with quantitative improvement scoring (CNN).⁵ Paired pre- and post-

treatment images were aligned using facial landmarks, cropped to include only the periorbital area, resized to 128×128 pixels, and intensity-normalized to the range [0,1]. To enhance model robustness given the limited dataset, on-the-fly data augmentation was applied during training, including random horizontal flips ($p = 0.5$), rotations ($\pm 15^\circ$), and brightness adjustments ($\pm 20^\circ$).

The augmentations preserved realism and improved robustness to minor variations in illumination and pose. The network architecture comprised two interconnected modules—a U-Net for segmentation and a CNN for regression—jointly trained within a unified framework (Figure 1). The U-Net processed post-treatment image to generate a single-channel periorbital mask. Each encoder block contained two 3×3 convolutional layers with ReLU activation. Each encoder consisted of sequential 3×3 convolutional layers with ReLU activation, followed by 2×2 max-pooling, progressively increasing channel depth from 32 to 128. The bottleneck layer included 128 filters, after which the decoder reconstructed the feature maps through transposed convolutions and incorporated skip connections from corresponding encoder layers to preserve fine textural and boundary details. A final 1×1 convolution layer produced the mask logit output while maintaining the original spatial resolution of 128×128 pixels. This mask localized the region of interest and captured fine-grained features relevant to treatment-induced changes.

The regression module integrated three inputs—the pre-treatment image (3 channels), post-treatment image (3 channels), and U-Net–derived mask (1 channel)—which were concatenated to form a 7-channel input tensor. This configuration allowed the model to evaluate appearance changes while emphasizing features within the mask-localized region. The regressor comprised two convolutional blocks with 3×3 kernels (16 and 32 filters), each followed by ReLU activation and 2×2 max-pooling. The resulting feature maps underwent global average pooling and were passed through a fully connected layer to yield a continuous improvement score ranging from 0 to 5.

The U-Net and CNN were trained jointly within a unified end-to-end framework, enabling the model to learn both spatial localization and quantitative scoring simultaneously. At each training iteration, the U-Net generated a periorbital mask from the post-treatment image, which was concatenated with the pre- and post-treatment images to form a 7-channel input tensor. The regression network then predicted a continuous score representing the degree of eyelid rejuvenation. A weighted combination of segmentation loss and regression loss (mean absolute percentage error, MAPE) was backpropagated to minimize the relative deviation between predicted and true scores across each batch. The dataset of 50 patients was divided into 40 training and 10 test cases. The network was trained for 200 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 4.

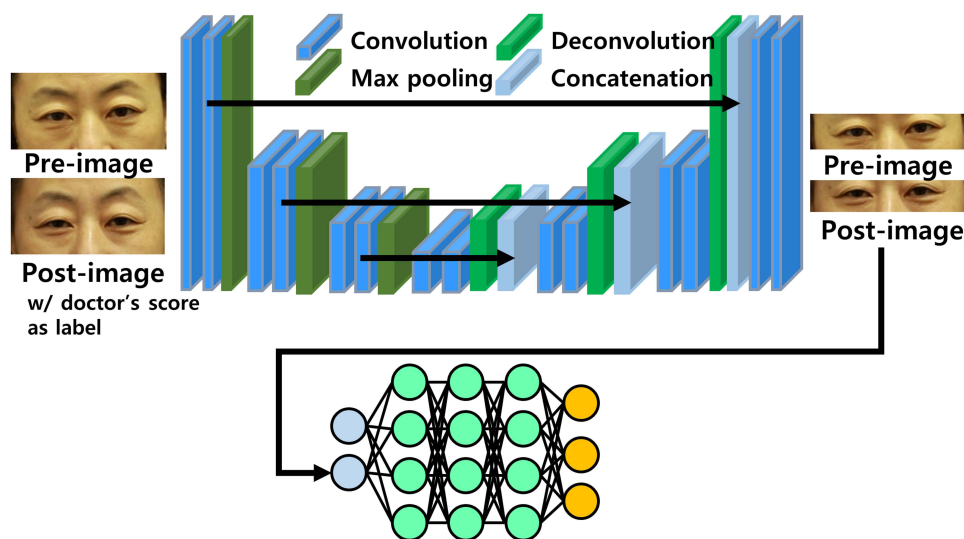


Figure 1 Workflow of the deep learning–based automated scoring system. Pre- and post-treatment eyelid photographs were used as model inputs. Images underwent preprocessing (alignment, cropping, and normalization) before being processed by a U-Net for periorbital segmentation (illustrated by blue, Orange, and green feature blocks). The U-Net output was then concatenated with pre- and post-treatment images and passed to a CNN regression module, which generated a continuous improvement score corresponding to dermatologist assessments.

Evaluation Metrics

Model performance was evaluated using root mean square error (RMSE) and MAPE relative to dermatologist consensus scores. Regression outputs were also visualized across ordinal categories (0–5) using a confusion matrix, demonstrating that predictions closely corresponded to the observed degree of improvement across subjects.

Dataset Limitations

This study represents a proof-of-concept for a deep learning–based automated scoring system designed to predict clinical outcomes after eyelid MRF treatment. The pilot dataset was relatively small ($n = 50$), predominantly female, and limited to Fitzpatrick skin types III–V, which may restrict generalizability. Additionally, the 8-week follow-up period may underestimate final outcomes, as collagen remodeling and neocollagenesis typically continue for 3–6 months post-treatment. Finally, the use of a single train/test split without cross-validation could result in overestimated model performance. Building on these proof-of-concept findings, future work will focus on improving and validating the system by incorporating multi-center data, implementing cross-validation, and extending longitudinal follow-up to ensure reproducibility across diverse skin types, imaging devices, and clinical conditions.

Results

The workflow of the proposed deep learning–based automated scoring system is illustrated in Figure 1, outlining the sequence of image preprocessing, periorbital segmentation using U-Net, and score regression via CNN.^{6,7} From a total of 50 patients, 40 paired pre- and post-treatment images were used for training and 10 for testing. Ground truth ratings were obtained by five board-certified dermatologists.

The model demonstrated strong predictive performance. On the test set, the CNN–U-Net hybrid achieved a RMSE of 0.4 and a MAPE of 0.08, corresponding to an average deviation of only 8% from dermatologist-assigned scores. In eight of ten test cases, the model’s prediction showed exact or near-exact agreement with expert evaluations, while only two cases differed by a single point on the five-point improvement scale.

The confusion matrix shown in Figure 2a demonstrates close alignment between predicted and reference scores, particularly for cases rated “4” or higher, corresponding to very good or excellent clinical improvement. Representative patient photographs with both model-predicted and dermatologist-assigned scores (Figures 2b and c) further illustrate the model’s ability to detect subtle yet clinically meaningful changes in eyelid tightening and wrinkle reduction.

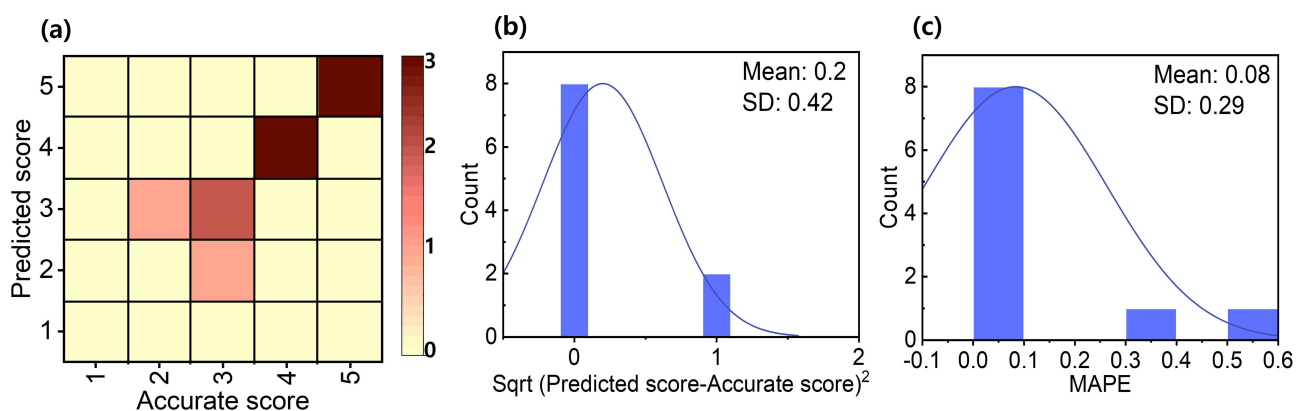


Figure 2 Performance evaluation of the automated scoring system. (a) Confusion matrix comparing dermatologist-assigned ground truth scores with model-predicted values. Strong diagonal alignment indicates high agreement between predictions and expert evaluations. (b) Distribution of root mean square error (RMSE) between predicted and reference scores. The model achieved a mean RMSE of 0.2 (SD = 0.42), demonstrating close correspondence with dermatologist ratings. (c) Distribution of mean absolute percentage error (MAPE). The model achieved a mean MAPE of 0.08 (SD = 0.29), confirming strong predictive robustness and low overall error across the test dataset.

Discussion

This study demonstrates, for the first time, the feasibility of a deep learning–based automated scoring system for evaluating upper eyelid rejuvenation outcomes following MRF treatment. While AI-driven image analysis has been increasingly applied in dermatology, its use in eyelid rejuvenation addresses a notable gap, as traditional assessments have largely relied on subjective photographic evaluation and physician judgment. By integrating U-Net based segmentation with CNN regression, the proposed system provides an objective, reproducible, and observer-independent method that closely aligns with expert dermatologist evaluations.

The clinical implications of this approach are substantial. Automated scoring can enhance patient consultations by delivering consistent visual and quantitative feedback, support personalized treatment planning, and promote research standardization across clinics. In a field often constrained by inter-observer variability and subjective interpretation, such tools represent a meaningful advancement toward reproducibility and transparency in aesthetic outcome assessment.

Despite its promising results, this proof-of-concept study has several limitations. The dataset was relatively small ($n = 50$, including 10 test cases), predominantly female, and limited to Fitzpatrick skin types III–V, which restricts generalizability. The 8-week follow-up period captured only early treatment effects, whereas MRF outcomes typically continue to improve over 3–6 months due to ongoing collagen remodeling. Additionally, the use of a single train/test split without cross-validation may have resulted in an optimistic estimation of model performance.

Future work should focus validating the model with more diverse populations, extended longitudinal follow-up and multi-center datasets. Incorporating additional clinical parameters—such as age, skin thickness, and energy delivery settings—may further enhance predictive accuracy and clinical applicability.

In summary, this study provides preliminary evidence that AI-based image analysis can serve as a standardized and objective tool for assessing non-surgical eyelid rejuvenation outcomes. By bridging the gap between subjective clinical impressions and quantitative measurement, this approach represents a promising step toward improving clinical decision-making, patient communication, and research comparability in aesthetic dermatology.

Conclusion

This proof-of-concept study demonstrates that a CNN–U-Net–based automated scoring system can objectively and consistently evaluate eyelid rejuvenation outcomes following MRF treatment. The model showed strong agreement with dermatologist assessments, supporting its potential as a reproducible, observer-independent, and scalable evaluation tool. Although preliminary and limited by sample size and follow-up duration, these findings mark an important first step toward the integration of AI-driven outcome assessment and standardization in aesthetic dermatology.

Acknowledgment

The abstract of this paper was presented at the Journal of the American Academy of Dermatology [https://www.jaad.org/article/S0190-9622\(24\)01547-0/fulltext](https://www.jaad.org/article/S0190-9622(24)01547-0/fulltext).

Author Contributions

“All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work”.

Funding

“This research was funded by National Research Foundation of Korea (NRF) grant funded by the Korean government, grant number: 2023R1A2C1007759”, “Grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Korea, grant number: RS-2023-KH-136575 & RS-2025-02217860” and “Grant of Translational R&D Project through Institute for Bio-Medical convergence, Incheon St. Mary’s Hospital, The Catholic University of Korea”.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Biesman BS, Baker SS, Carruthers J, Silva HL, Holloman EL. Monopolar radiofrequency treatment of human eyelids: a prospective, multicenter, efficacy trial. *Lasers Surg Med.* 2006;38(10):890–898. doi:10.1002/lsm.20452
2. Suh DH, Hong ES, Kim HJ, Lee SJ, Kim HS. A survey on monopolar radiofrequency treatment: the latest update. *Dermatol Ther.* 2020;33(6):e14284. doi:10.1111/dth.14284
3. Scarano A, Lorusso F, Brucoli M, Lucchina AG, Carinci F, Mortellaro C. Upper eyelid blepharoplasty with voltaic arc dermabrasion. *J Craniofac Surg.* 2018;29(8):2263–2266. doi:10.1097/scs.0000000000004504
4. Elhamaky TR. Small incision upper blepharoplasty in the treatment of upper eyelid solitary nasal pad fat protrusion. *J Cutan Aesthet Surg.* 2023;16(3):210–213. doi:10.4103/jcas.Jcas_33_22
5. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–444. doi:10.1038/nature14539
6. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* 2016;35(5):1285–1298. doi:10.1109/tmi.2016.2528162
7. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* Springer International Publishing; 2015:234–241.

Clinical, Cosmetic and Investigational Dermatology

Publish your work in this journal

Clinical, Cosmetic and Investigational Dermatology is an international, peer-reviewed, open access, online journal that focuses on the latest clinical and experimental research in all aspects of skin disease and cosmetic interventions. This journal is indexed on CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-cosmetic-and-investigational-dermatology-journal>

Dovepress
Taylor & Francis Group