Original Article

Conditional generative adversarial network simulation of hypertrophic scar appearance across three treatments

Chengfei Li[1], Shiyi Li[1], Qiang Fu[1], Guiwen Zhou[1], Qian Wu[1], Fanting Meng[1], Xiao Xu[1], Minliang Chen[1]

[1]Department of Burn and Plastic Medicine, the Fourth Medical Center of the Chinese People's Liberation Army General Hospital, Beijing 100048, China

Correspondence to: Prof. Minliang Chen, The Department of Burn and Plastic Medicine, the Fourth Medical Center of the Chinese People's Liberation Army General Hospital, Beijing 100048, China, Email: chenml@sohu.com

Received: 30 July 2026 | Approved: 27 August 2026 | Online: 27 August 2026

Abstract

Aim: To develop and internally evaluate a deep learning system for simulating hypertrophic scar appearance after three treatments from a pretreatment photograph.

Methods: Paired pretreatment and post-treatment photographs from 300 patients at a single center (2020-2025) were divided into training and internal test sets (8:2; 20 test cases per treatment). Separate conditional generative adversarial networks were developed for surgical excision, autologous emulsified fat injection, and micro-plasma radiofrequency (MPR) using a ResNet-style generator, PatchGAN discriminator, adversarial loss, and L1 reconstruction loss. Five attending plastic surgeons assessed 60

generated images. Mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) were calculated against treatment-matched references, and correlations with expert identification were assessed.

Results: Generated images generally preserved scar location and morphology while reflecting treatment-associated changes. Overall, 148 of 300 ratings identified images as model-generated (49.33%; 95% CI, 43.7%-55.0%). For excision, fat injection, and MPR, respectively, MSE was 0.050 ± 0.020, 0.060 ± 0.024, and 0.053 ± 0.024; PSNR was 36.151 ± 3.856, 34.601 ± 3.118, and 37.069 ± 3.215 dB; and SSIM was 0.715 ± 0.039, 0.708 ± 0.026, and 0.736 ± 0.036. Expert identification correlated with MSE (ρ = 0.527, $P < 0.01$) and SSIM (ρ = -0.690, $P < 0.01$), but not PSNR (ρ = -0.170, $P = 0.194$).

Conclusion: The prototype showed preliminary visual realism and quantitative consistency, supporting further validation for pretreatment communication, shared decision-making, and expectation management.

Keywords: Hypertrophic scar, conditional generative adversarial network, image-to-image translation, appearance simulation, clinical communication

INTRODUCTION

Hypertrophic scars are fibroproliferative sequelae of dermal injury, characterized by excessive extracellular matrix deposition, persistent inflammation, increased vascularity, pigmentary alteration, and raised, firm tissue that remains within the boundaries of the original wound. In addition to their visible appearance, these scars may cause pain, pruritus, contracture, functional restriction, and sustained psychosocial burden. Their clinical course is heterogeneous: scar height, color, texture, symptoms, and recurrence risk may differ substantially according to wound characteristics, anatomical tension, maturation, and patient-specific biological factors[1-3]. Consequently, hypertrophic scars are better regarded as a chronic management problem in which appearance, symptoms,

and function must be balanced rather than as a single morphological abnormality.

Management is correspondingly individualized and often multimodal. Available strategies include silicone and pressure therapy, intralesional agents, laser- and energy-based procedures, fat-based regenerative interventions, surgical excision, and selected adjuvant measures[4,5]. These approaches produce different visible trajectories, ranging from structural excision and reconstruction to gradual changes in contour, color, pliability, and surface texture. However, potential outcomes remain difficult to communicate consistently before treatment. Verbal descriptions such as “flatter”, “lighter,” or “softer” cannot fully convey patient-specific variation, whereas photographs of other patients are confounded by scar type, anatomical site, skin tone, maturity, aftercare, and imaging conditions. Because the fulfillment of preoperative expectations is associated with patient-reported outcomes, this communication gap may complicate shared decision-making and expectation management[6,7].

Current scar outcome instruments principally document an existing scar. The Patient and Observer Scar Assessment Scale (POSAS), for example, captures observer-rated features and patient-reported symptoms, while standardized clinical photographs can support reproducible assessment of selected scar characteristics[8-10]. These methods are valuable for measurement and follow-up but do not show how an individual scar might appear after a specific treatment. Artificial intelligence (AI) has recently extended image-based scar analysis to automated screening, classification, and severity estimation[11,12]. Nevertheless, contemporary AI-based scar research remains centered largely on recognizing what a scar is or quantifying its current severity; recent reviews continue to identify limited datasets, inconsistent image acquisition, and insufficient benchmarking as major barriers to clinical translation[13]. A pretreatment visualization tool that generates modality-specific appearances from the same pretreatment photograph would therefore address a distinct and clinically relevant unmet need.

Generative deep learning provides a potential technical route for this task. In paired image-to-image translation, a conditional generative adversarial network (cGAN) learns a mapping from a conditioning image to a target-domain image. Adversarial learning encourages perceptually plausible outputs, whereas reconstruction constraints help retain spatial correspondence with the source image[14,15]. Related cGAN-based systems have been investigated for preoperative outcome simulation in gluteal augmentation and rhinoplasty, demonstrating the feasibility of producing patient-specific visualizations from clinical photographs[16,17]. Nevertheless, a systematic review of generative adversarial networks in plastic surgery identified substantial methodological heterogeneity and emphasized the need for standardized evaluation and clinical validation[18]. Evidence for treatment-specific appearance simulation in hypertrophic scars remains limited.

Accordingly, this study aimed to develop and internally evaluate a treatment-specific cGAN framework for hypertrophic scars. Using paired pretreatment and post-treatment clinical photographs, separate models were trained for surgical excision, autologous emulsified fat injection, and micro-plasma radiofrequency (MPR). We evaluated the visual detectability and quantitative similarity of the generated images and examined associations between clinicians' judgments and objective image-quality metrics. The system was deliberately positioned as a two-dimensional visualization aid for pretreatment communication and expectation management, not as a substitute for physical examination, comprehensive efficacy assessment, or individualized treatment decisions.

METHODS

Clinical study design

This single-center retrospective study was conducted to develop and internally evaluate a deep learning system for simulating the post-treatment appearance of mature hypertrophic scars. Clinical records and photographs were obtained from patients

treated at the Fourth Medical Center of the Chinese PLA General Hospital between 2020 and 2025. Treatment was not assigned for research purposes; patients were grouped according to the scar-directed treatment received during routine clinical care. The study protocol was approved by the institutional ethics committee and was conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent for the use of their clinical data and images in research, and all data were de-identified before analysis. The overall study workflow, including clinical image acquisition, paired dataset construction, image preprocessing, training-set augmentation, model development, and internal evaluation, is summarized in Figure 1.

Patients were eligible if they (1) had a clinical diagnosis of a mature, clinically stable hypertrophic scar; (2) were 18-70 years of age; (3) had received only one of the three prespecified treatments—surgical excision, autologous emulsified fat injection, or MPR—with complete treatment records; and (4) had clear, complete, pairable pretreatment and post-treatment photographs acquired within the prespecified time windows. Exclusion criteria were an uncertain or non-hypertrophic-scar diagnosis; age outside the eligible range; an unstable lesion, acute wound-healing phase, or recent additional scar intervention; two or more scar-directed treatments during the study period; incomplete treatment documentation; a missing photograph at either required time point; major blur, occlusion, glare, overexposure, tattoo interference, or other artifacts that prevented reliable analysis; or a post-treatment photograph outside the permitted follow-up window.

After eligibility screening and pair-level image-quality review, 300 patients were included, each contributing one pretreatment photograph and one corresponding post-treatment photograph. The dataset therefore comprised 300 paired samples (600 photographs) and was balanced across the three treatments: 100 patients underwent surgical excision, 100 received autologous emulsified fat injection, and 100 received MPR.

Pretreatment photographs were preferentially selected from records acquired within 1 day before treatment. Post-treatment photographs were selected at approximately 6 months after completion of the treatment course, with an allowable window of ±2 weeks. The observed post-treatment photograph served as the paired target during supervised model development and as the reference for objective image comparison. A reference image was available only for the treatment actually received and did not constitute a verified counterfactual outcome for either alternative treatment.

For cohort assignment, surgical excision was defined as complete excision of the target scar followed by tension-managed layered suturing and primary closure. Autologous emulsified fat injection was defined as the harvesting, purification, mechanical emulsification, and layered intralesional injection of autologous adipose tissue. MPR was defined as fractional application of radiofrequency-generated plasma energy to create controlled microthermal treatment zones and induce surface remodeling. Patients who received an additional scar-directed modality during the study window were excluded from the corresponding single-treatment group. The operational definitions and treatment protocols are summarized in Table 1.

Table 1. Operational definitions and treatment protocols for the three treatment modalities.

Treatment modality	Operational definition	Key treatment parameters
Surgical excision (n = 100)	Complete excision of the target scar followed by tension-managed layered suturing and primary closure.	Anesthesia: local anesthesia; Suture size: 4-0 Suture technique: tension-reducing suturing; Postoperative scar management: usual care

Autologous emulsified fat injection (n = 100)	Harvesting, purification, mechanical emulsification, and layered intralesional injection of autologous adipose tissue.	Donor tissue: thigh fat; Emulsification method: mechanical emulsification; Number of sessions and interval: 3 times with an interval of 3 months
Micro-plasma radiofrequency (n = 100)	Fractional application of micro-plasma radiofrequency to the scar surface to induce controlled microthermal injury and subsequent remodeling.	Device and handpiece: Alma Lasers; Energy setting: 60 ± 10 W; Number of passes: 2; Number of sessions and interval: 5 times with an interval of 1 months

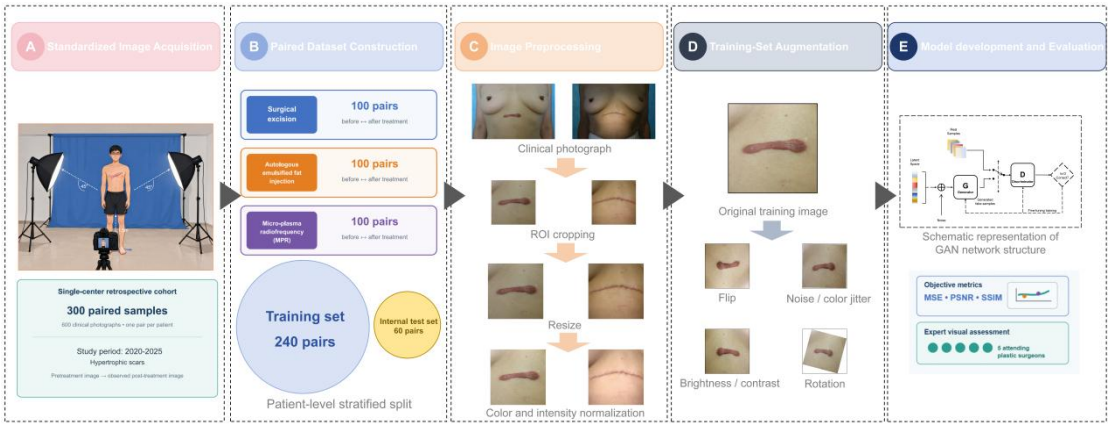

Figure 1. Overall workflow of dataset construction, image preprocessing, treatment-specific cGAN development, and internal evaluation.

Dataset construction and preprocessing

Most photographs were acquired with the same digital camera (Nikon D3400; Nikon Corporation, Tokyo, Japan), and the camera body and lens configuration were generally consistent during routine clinical photography. Two light sources were positioned slightly anterior to the participant on the left and right, creating an incidence angle of approximately 45°. A blue backdrop was used, and the camera-to-participant distance

was approximately 1.5 m. The scar was fully exposed, patient positioning was kept as stable as possible, and the camera height and viewing angle were adjusted to place the lesion near the center of the image. Because the data were retrospective, minor variations in device, distance, angle, and illumination could not be eliminated.

All image pairs underwent usability and consistency review. Images were excluded when hair, tattoos, glare, overexposure, blur, or other artifacts obscured diagnostically relevant scar details and could not be addressed by reasonable cropping. Images with minor localized obstruction or reflection were retained only when the scar remained clearly interpretable.

For each eligible photograph, a region of interest (ROI) centered on the scar was cropped so that the lesion occupied the principal portion of the image while unnecessary background was minimized. The preprocessing workflow included denoising, white-balance correction, and normalization of brightness and color to reduce non-lesion variation related to illumination, color temperature, and exposure. Region-of-interest images were resized to 256 × 256 pixels, converted to red-green-blue format, and normalized to the range -1 to 1 for model input.

Dataset partitioning was performed at the patient level to ensure that the paired photographs from an individual patient were assigned exclusively to either the training set or the internal test set. Cases were stratified by treatment modality and randomly divided at an 8:2 ratio using a fixed random seed of 42. Accordingly, each treatment group contributed 80 paired cases to model training and 20 paired cases to internal testing, resulting in a training set of 240 pairs and an internal test set of 60 pairs.

Data augmentation was applied only to the training set after partitioning. Operations included random horizontal flipping; rotation within ±15°; adjustment of brightness, contrast, and saturation within ±20%; hue perturbation within ±10%; and noise injection

at 5%. Transformations were randomly combined, while a proportion of images remained unaugmented. All augmentation ranges were reviewed to avoid clinically implausible appearances. The internal test set was not augmented or resampled and was not used for parameter adjustment.

Model development

The task was formulated as supervised image-to-image translation, with the pretreatment image as input and the treatment-matched post-treatment photograph from the same patient as the target. A pix2pix-inspired cGAN framework was used. To avoid mixing treatment-specific transformation patterns, three independent cGANs with identical architectures and training settings were trained for surgical excision, autologous emulsified fat injection, and MPR.

The generator used a ResNet-style encoder-residual-block-decoder architecture. An initial 7 × 7 convolution produced 64 feature channels, followed by two downsampling layers that increased the channel depth to 128 and 256. Nine residual blocks operated at 256 channels. Two transposed-convolution layers then restored the spatial resolution while reducing the channels to 128 and 64. Instance normalization and rectified linear unit activation were used throughout, and a final 7 × 7 convolution with tanh activation produced a three-channel 256 × 256 image.

The PatchGAN discriminator received the channel-wise concatenation of the pretreatment image and either the real or generated post-treatment image. Four 4 × 4 convolutional blocks increased the feature channels from 64 to 512. Instance normalization was applied after all convolutional layers except the first, and LeakyReLU used a negative slope of 0.2. The final convolution generated a 1 × 15 × 15 logit map. No sigmoid layer was applied because the adversarial objective used BCEWithLogitsLoss.

The training objective combined conditional adversarial loss with L1 reconstruction loss; the L1 weighting coefficient was 100. The generator and discriminator were optimized alternately using Adam (learning rate, 2 × 10-4; β1, 0.5; β2, 0.999). Each model was trained for 1,000 epochs with a batch size of 8, and automatic mixed-precision autocasting was used. Convolutional and transposed-convolutional weights were initialized from a normal distribution with a mean of 0 and a standard deviation of 0.02. Experiments used Python 3.11 and PyTorch 2.5.0 with CUDA 12.4 on a workstation with an Intel Core i7-14700KF processor and an NVIDIA GeForce RTX 4070 Ti graphics processor. During inference, the network parameters were fixed and only the trained generator for each treatment was retained.

To assess training stability and convergence, Figure 2 presents the generator-loss trajectories of the three treatment-specific models. Generator loss decreased rapidly during early training and then gradually stabilized across all three models.

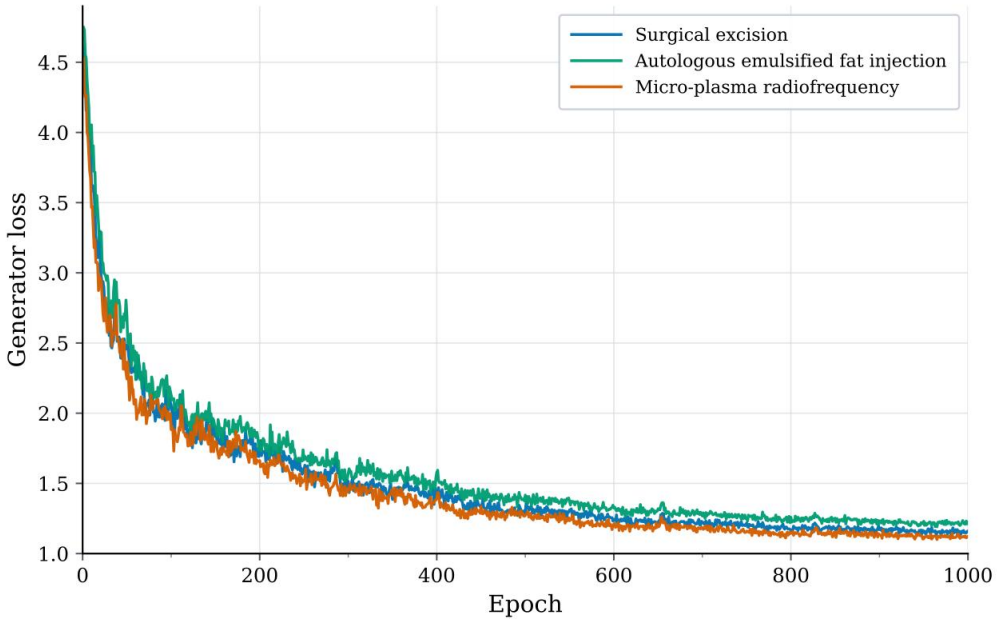

Figure 2. Generator loss curves of the three treatment-specific cGANs over 1,000 training epochs.

Internal testing

The three treatment-specific generators were evaluated on the locked internal test set of 60 cases (20 per treatment). Test images underwent the same deterministic preprocessing as training images but were not augmented. For quantitative evaluation, only the generated image from the model corresponding to the treatment actually received was compared with that patient's post-treatment reference image. Outputs from the other two models were used only for illustrative treatment-scenario visualization.

Five attending plastic surgeons independently assessed the 60 generated images. Images were presented individually in randomized order under standardized display conditions, with consistent resolution, zoom, and background. Reviewers were blinded to patient information, model details, and reference images. Each image was classified as "identified as model-generated" or "not identified as model-generated" based on overall realism, color and texture continuity, scar contour, structural plausibility, and visible artifacts. Because only generated images were presented, this experiment measured the detectability of generation-related features rather than formal discrimination between intermixed real and generated photographs.

First, the RGB pixel-intensity distributions of the observed and generated images were compared. Second, MSE, PSNR, and SSIM were calculated between each generated image and its treatment-matched post-treatment reference. Before metric calculation, 256 × 256-pixel RGB images were rescaled to the range 0 to 1, and the data range for PSNR and SSIM was set to 1.0. MSE quantified pixel-level error, PSNR summarized reconstruction fidelity, and SSIM assessed structural similarity. Metrics were calculated for individual image pairs and summarized within each treatment group as the mean ± standard deviation.

Statistical analysis

Reviewer-specific and pooled identification proportions were reported with Wilson 95%

confidence intervals. Interrater agreement was quantified using Fleiss' κ. At the image level, the identification rate was defined as the proportion of the five reviewers who identified an image as model-generated. Spearman rank correlations assessed associations between this rate and MSE, PSNR, and SSIM. Tests were two-sided, and $P < 0.05$ was considered statistically significant.

RESULTS

Study cohort and dataset composition

A total of 300 patients with mature hypertrophic scars met the eligibility criteria. Each patient contributed one pretreatment photograph and one treatment-matched post-treatment photograph, yielding 300 paired samples (600 photographs). The dataset was balanced across surgical excision, autologous emulsified fat injection, and MPR, with 100 patients in each group. Patient-level stratification allocated 240 pairs to training and 60 to internal testing, with 80 training pairs and 20 test pairs per treatment.

In the overall cohort, the mean age was 30.0 ± 10.8 years, and 190 patients (63.3%) were female. The median scar duration was 18 months (interquartile range, 9-36 months). The chest was the most common anatomical site (102/300, 34.0%), followed by the abdomen (86/300, 28.7%), shoulder and back (69/300, 23.0%), and other sites (43/300, 14.3%). The baseline characteristics of the training and internal test sets are summarized in Table 2.

Table 2. Baseline characteristics of the study cohort.

Characteristic	Overall cohort (*N* = 300)	Training set (*n* = 240)	Internal test set (*n* = 60)	*P* value
Age, years, mean ± SD	30.0 ± 10.8	30.3 ± 10.7	28.9 ± 11.4	0.372
Sex, *n* (%)				0.765
Male	110 (36.7)	87 (36.3)	23 (38.3)	
Female	190 (63.3)	153 (63.8)	37 (61.7)	

Characteristic	Overall cohort (*N* = 300)	Training set (*n* = 240)	Internal test set (*n* = 60)	*P* value
Scar duration, months, median (IQR)	18 (9-36)	18 (9-30)	18 (12-36)	—
Anatomical site, *n* (%)				0.938
Chest	102 (34.0)	83 (34.6)	19 (31.7)	
Shoulder and back	69 (23.0)	55 (22.9)	14 (23.3)	
Abdomen	86 (28.7)	69 (28.8)	17 (28.3)	
Other sites	43 (14.3)	33 (13.8)	10 (16.7)	

Representative simulation results and failure patterns

Representative cases from the internal test set are shown in Figure 3. For each case, the pretreatment image, paired real post-treatment reference, and outputs of the three treatment-specific models are presented side by side. The generated images generally preserved scar location and overall contour. The surgical excision model produced more regular linear morphology and clearer margins. The autologous emulsified fat injection model showed more gradual changes, with smoother edge transitions and reduced local color contrast. The micro-plasma radiofrequency model predominantly generated a flatter surface appearance, reduced color heterogeneity, and more uniform fine texture. The treatment-matched outputs broadly reflected the appearance changes observed in the corresponding reference images, although some patient-specific details were not fully reproduced.

Representative failure patterns are presented in Figure 4. Hair occlusion was associated with incomplete scar delineation and inaccurate local reconstruction. Other observed errors included loss or excessive smoothing of local texture, discontinuous or unstable edge transitions, and localized color-tone deviations from the surrounding skin or paired reference image.

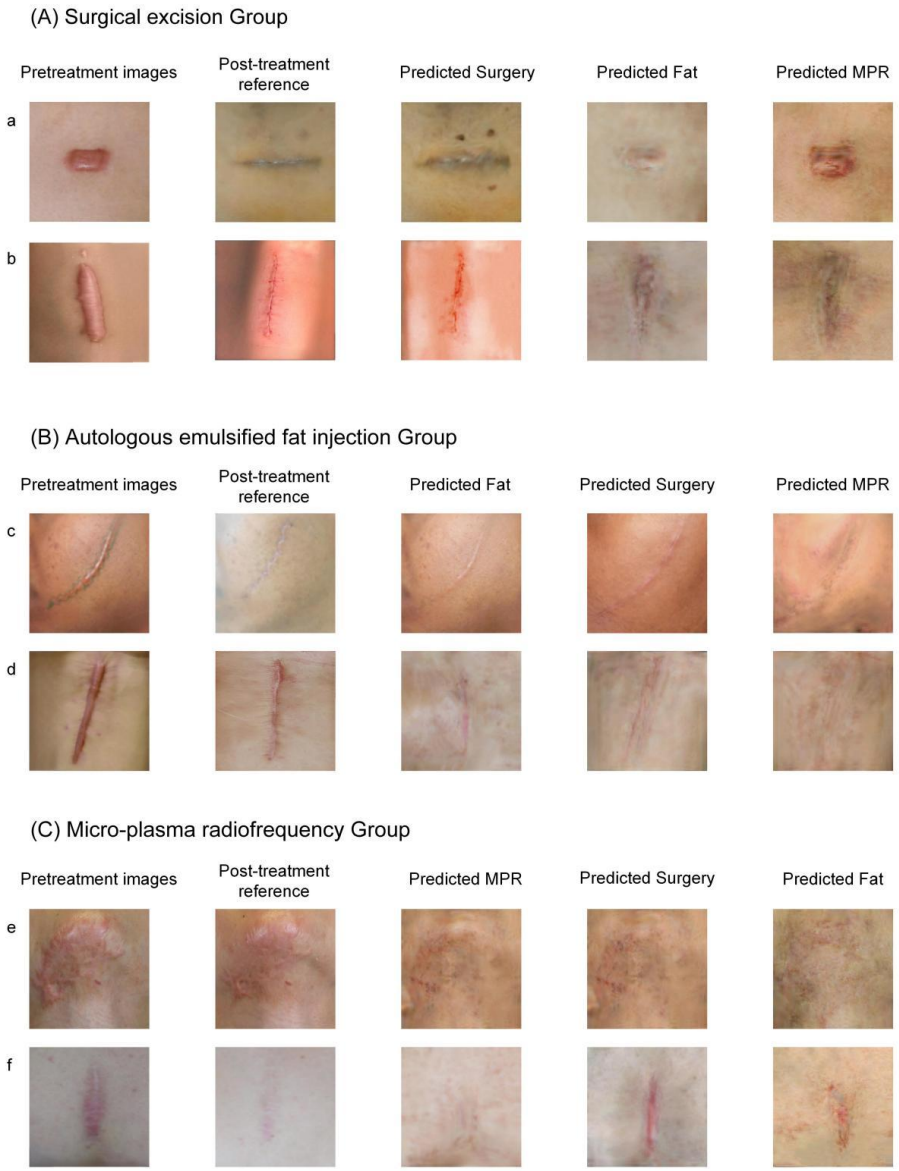

Figure 3. Representative simulation results from the three treatment-specific models. Each row presents the pretreatment image, paired real post-treatment reference, and simulated outputs for surgical excision, autologous emulsified fat injection, and micro-plasma radiofrequency.

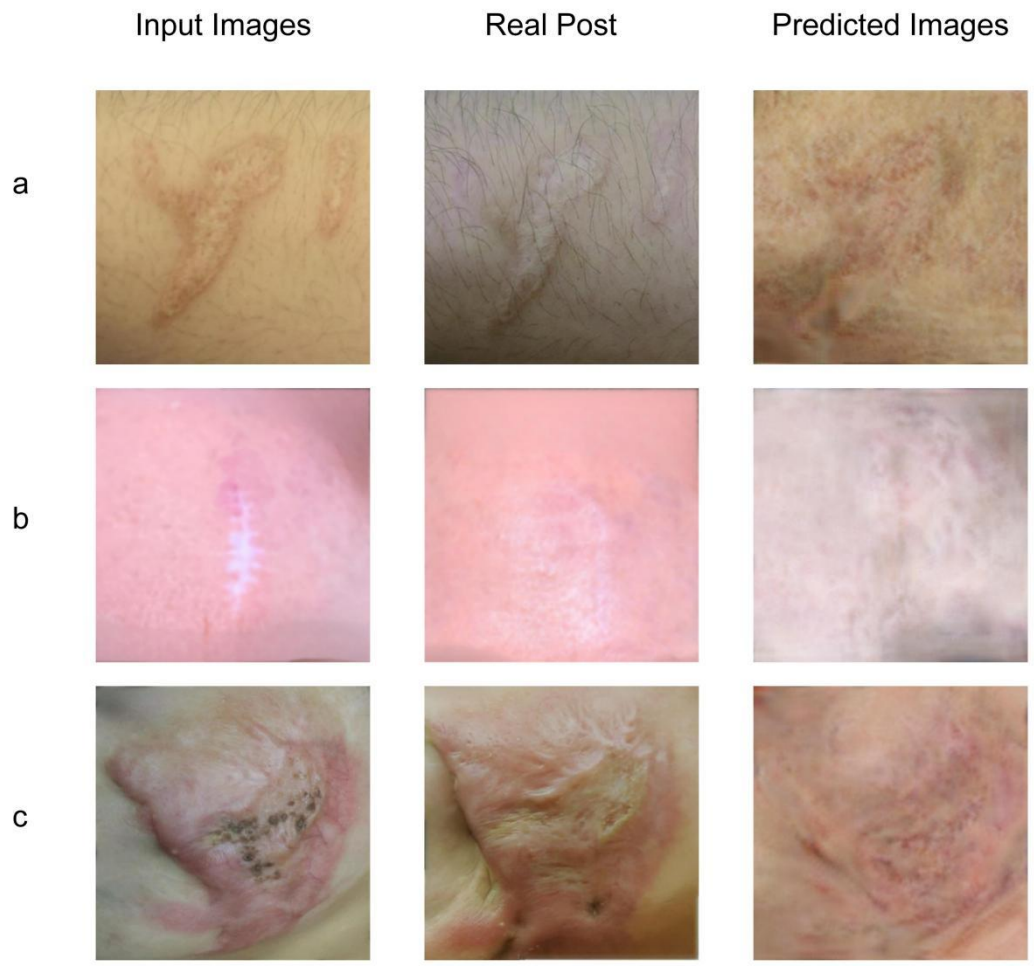

Figure 4. Representative failure patterns: a, hair occlusion; b, strong reflection; c, local texture loss.

Expert visual identification

Five attending plastic surgeons provided 300 independent ratings for the 60 generated images. Overall, 148 of 300 ratings identified the displayed image as model-generated, corresponding to a pooled identification rate of 49.33% (95% confidence interval, 43.7%-55.0%). Reviewer-specific identification rates ranged from 36.67% to 58.33% [Table 3]. Interrater agreement was substantial (Fleiss' $\kappa = 0.693$).

Table 3. Expert identification of model-generated images.

Reviewer	Identified as model-generated, *n/N*	Identification rate, % (Wilson 95% CI)
Reviewer 1	34/60	56.67% (44.10%-68.43%)
Reviewer 2	29/60	48.33% (36.18%-60.69%)
Reviewer 3	35/60	58.33% (45.73%-69.94%)
Reviewer 4	28/60	46.67% (34.63%-59.11%)
Reviewer 5	22/60	36.67% (25.62%-49.32%)
Overall	148/300	49.33% (43.72%-54.96%)

Objective image similarity metrics

The RGB pixel-intensity distributions of the observed post-treatment and cGAN-generated image sets are shown in Figure 5. The distributions were broadly similar across the red, green, and blue channels.

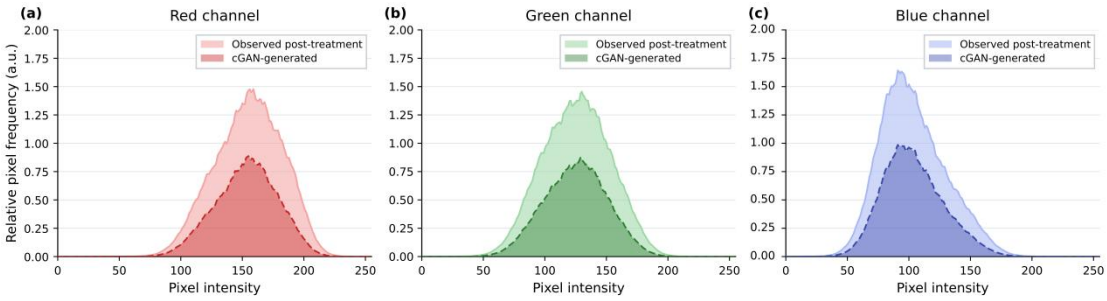

Figure 5. RGB pixel-intensity distributions of observed post-treatment and cGAN-generated images in the internal test set.

The surgical excision group achieved an MSE of 0.050 ± 0.020, a PSNR of 36.151 ± 3.856 dB, and an SSIM of 0.715 ± 0.039. The corresponding values were 0.060 ± 0.024, 34.601 ± 3.118 dB, and 0.708 ± 0.026 for autologous emulsified fat injection, and 0.053 ± 0.024, 37.069 ± 3.215 dB, and 0.736 ± 0.036 for micro-plasma radiofrequency [Table 4]. Descriptively, the micro-plasma radiofrequency group showed the highest mean PSNR and SSIM, whereas the surgical excision group showed the lowest mean MSE. Image-level boxplots demonstrated within-group variability and overlapping

distributions across the three treatment groups [Figure 6].

Table 4. Objective image similarity metrics in the internal test set.

Treatment group	*n*	MSE	PSNR, dB	SSIM
Surgical excision	20	0.050 ± 0.020	36.151 ± 3.856	0.715 ± 0.039
Autologous emulsified fat injection	20	0.060 ± 0.024	34.601 ± 3.118	0.708 ± 0.026
Micro-plasma radiofrequency	20	0.053 ± 0.024	37.069 ± 3.215	0.736 ± 0.036

Data are presented as mean ± standard deviation. MSE, mean squared error; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure.

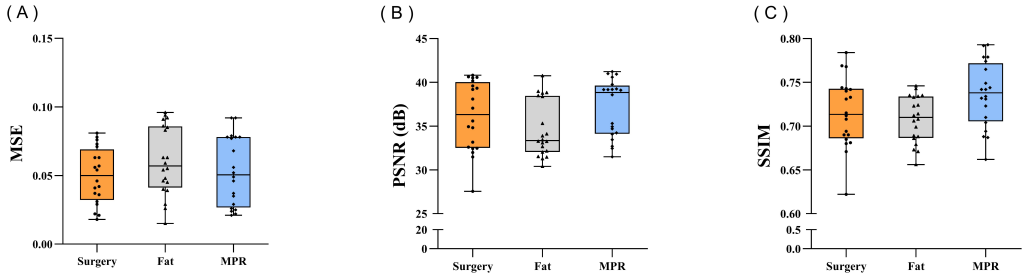

Figure 6. Image-level distributions of objective similarity metrics in the internal test set. Boxplots show the distributions of (A) MSE; (B) PSNR; and (C) SSIM across the three treatment groups. Boxes represent the interquartile range, central lines indicate the median, and individual points represent image pairs (n = 20 per group). MPR, micro-plasma radiofrequency.

Spearman rank correlation analysis

To examine the correspondence between expert visual judgments and objective image similarity, Spearman rank correlations were calculated between the image-level expert identification rate and MSE, PSNR, and SSIM across the 60 generated images. The identification rate was moderately positively correlated with MSE (ρ = 0.527, P < 0.01) and strongly negatively correlated with SSIM (ρ = -0.690, P < 0.01). Its correlation with PSNR was weak and not statistically significant (ρ = -0.170, P = 0.194) [Table 5]. Thus,

images with greater pixel-level error or lower structural similarity tended to be more readily identified as model-generated.

Table 5. Associations between expert identification rate and objective image metrics.

Metric	*n*	Spearman's *ρ*	*P* value	Direction of association	Strength of correlation
MSE	60	0.527	< 0.01	Positive correlation	Moderate correlation
PSNR	60	-0.170	0.194	Negative correlation	Weak correlation
SSIM	60	-0.690	< 0.01	Negative correlation	Strong correlation

DISCUSSION

This study developed and internally evaluated three treatment-specific cGANs for simulating the post-treatment appearance of mature hypertrophic scars from a single pretreatment photograph. The generated images generally preserved lesion location and global morphology while expressing treatment-associated differences in contour, margins, color, and surface texture. Descriptively, surgical excision had the lowest mean MSE, whereas MPR had the highest mean PSNR and SSIM. Expert identification was associated with MSE and SSIM but not PSNR. Together, these findings provide preliminary image-level evidence that treatment-specific appearance simulation is feasible and that expert review and objective similarity metrics capture complementary aspects of output quality.

Previous applications of artificial intelligence to scar images have primarily focused on screening, severity prediction, segmentation, perfusion-based assessment, and automated scar scoring[19-21]. Although these approaches support objective assessment and longitudinal monitoring, their outputs are generally presented as classifications, numerical scores, or segmented regions rather than patient-specific post-treatment visualizations. GAN-based outcome simulation has been investigated in plastic surgery.

The present study extends this line of research to hypertrophic scars by generating modality-specific appearances for three distinct treatment scenarios. The use of separate models also enabled treatment-related visual patterns to be represented without combining potentially conflicting transformations within a single generator.

The selected cGAN architecture was consistent with the paired nature of the present task. Conditional image-to-image translation allows the generator to learn a mapping between a pretreatment photograph and its paired post-treatment reference while retaining spatial correspondence with the input. In the present framework, the ResNet-style generator was intended to preserve global scar anatomy while learning treatment-associated changes, whereas the PatchGAN discriminator emphasized local plausibility in scar margins, pigmentation, and surface texture. The combination of adversarial and L1 reconstruction losses further balanced perceptual realism against correspondence with the patient-specific reference image. Training separate models for the three modalities allowed each generator to learn a more homogeneous treatment-specific target distribution.

Qualitative examination demonstrated distinguishable appearance patterns across the three modalities. Surgical excision outputs predominantly showed more regular linear morphology and clearer boundaries, consistent with a structurally constrained transformation. Autologous emulsified fat injection generated more gradual changes, including smoother edge transitions and reduced local chromatic contrast. Micro-plasma radiofrequency outputs more frequently showed surface smoothing, finer texture, and greater color uniformity. These patterns were broadly consistent with the visible changes represented in the corresponding treatment datasets and with previously reported scar-related applications of fat grafting[22-24]. They should nevertheless be interpreted as learned image-domain patterns rather than direct representations of the biological mechanisms of treatment.

The objective metrics showed modest numerical differences among treatments. Surgical excision had the lowest mean MSE, MPR had the highest mean PSNR and SSIM, and autologous emulsified fat injection had the highest mean MSE and lowest mean PSNR and SSIM. Because no formal between-treatment hypothesis tests were performed, these differences are descriptive and do not establish statistical superiority of one model or clinical treatment over another. A plausible explanation is that surgical excision produces a spatially constrained structural transformation, whereas MPR predominantly affects local texture and color features that may be well captured by patch-based discrimination. By contrast, appearance after fat injection may vary with graft retention, tissue integration, revascularization, and remodeling[4,23,25]. When similar pretreatment appearances correspond to multiple plausible post-treatment outcomes, a deterministic model trained with reconstruction loss may generate an averaged representation, reducing correspondence with any single observed reference. This interpretation concerns image-mapping predictability, not comparative treatment efficacy.

The expert assessment provided a clinically oriented complement to the pixel- and structure-based metrics. Overall, 49.33% of judgments identified the displayed image as model-generated, with reviewer-specific rates ranging from 36.67% to 58.33%. Fleiss' κ was 0.693, indicating substantial consistency among the five evaluators. These findings indicate that detectability varied among outputs: some images did not contain obvious generation-related features, whereas others remained recognizable because of unnatural margins, texture discontinuity, color deviation, or local artifacts.

The image-level correlation analysis further clarified the relationship between clinical visual judgment and objective similarity. The expert identification rate was moderately positively correlated with MSE (ρ = 0.527), (P < 0.01) and strongly negatively correlated with SSIM (ρ = -0.690), (P < 0.01)). Thus, images with greater pixel-level error or lower structural similarity were more likely to be identified as model-generated. In contrast, the correlation with PSNR was weak and not statistically significant (ρ =

-0.170), (P = 0.194). SSIM evaluates luminance, contrast, and structural relationships and was originally designed to better reflect perceptually relevant image information[26]. PSNR primarily summarizes global reconstruction error and may show limited correspondence with human perception when localized structural or textural abnormalities determine visual judgment[27]. The present results therefore support the combined use of expert review and multiple objective metrics rather than reliance on a single numerical measure.

Subjective realism and patient-specific outcome correspondence should also be distinguished. A generated image may appear visually natural because its background, skin tone, and local transitions resemble a clinical photograph, while still differing from the actual post-treatment image in incision direction or scar morphology. Conversely, an output with a less favorable pixel-based metric may retain clinically relevant features such as scar location, overall contour, and the principal direction of appearance change. Consequently, image realism alone does not establish accurate prediction of an individual outcome, and similarity to one observed follow-up image does not encompass the full range of clinically plausible results.

The principal clinical value of this framework lies in its potential to support pretreatment communication. Generating different treatment scenarios from the same pretreatment photograph provides a more individualized visual reference than verbal descriptions or photographs from unrelated patients. Such side-by-side visualization may help clinicians explain differences in expected contour, pigmentation, margins, and texture and may support shared decision-making and expectation management, both of which are important components of surgical care. The outputs should be presented within a clinician-led consultation as possible appearance trends rather than guaranteed outcomes or comparative evidence of treatment efficacy. They are intended to supplement, rather than replace, physical examination and individualized assessment of scar thickness, pliability, symptoms, functional impairment, treatment indications, and

patient preferences.

Further clinical translation should focus on prospectively evaluating whether the system improves patient understanding, expectation concordance, decisional conflict, consultation efficiency, and satisfaction with the decision-making process. Such evaluation should compare model-assisted visualization with conventional verbal counseling and representative clinical photographs and should incorporate patient-reported and workflow-related outcomes. This progression from internal image evaluation to prospective clinical evaluation is consistent with recommended approaches for assessing the real-world value of medical artificial intelligence systems[28,29].

Limitations

Several limitations should be considered. First, this retrospective single-center study included 300 image pairs, with only 60 cases in the internal test set. The relatively standardized acquisition setting and the predominance of trunk scars may have introduced center-, device-, population-, and anatomical site-specific bias; data augmentation cannot substitute for independently collected external data[30]. Moreover, the framework covered only three treatment modalities and relied primarily on two-dimensional photographs. It therefore could not incorporate scar thickness, pliability, symptoms, functional impairment, treatment parameters, or other factors influencing individual responses[31,32]. Minor variations in illumination, viewing angle, exposure, and occlusion also affected simulation quality, emphasizing the need for standardized photography and automated image-quality control[33]. Future studies should conduct multicenter external validation in more diverse populations and integrate clinical variables, three-dimensional imaging, ultrasonography, and uncertainty-aware generation to represent multiple plausible outcomes.

Second, the current evaluation established only preliminary image-level feasibility.

MSE, PSNR, and SSIM measure similarity to a single observed reference image but cannot determine whether a simulation represents the full range of clinically plausible outcomes. Because the expert assessment included only generated images, the 49.33% identification rate should not be interpreted as evidence that generated and real photographs were indistinguishable. In addition, no formal between-treatment comparisons or patient-centered outcomes—such as understanding, decisional conflict, expectation concordance, or satisfaction—were evaluated. Prospective comparative studies are therefore required to determine whether model-assisted visualization provides clinical benefits beyond conventional counseling. Until such evidence is available, the outputs should be regarded as clinician-interpreted illustrations of possible appearance trends rather than deterministic predictions, treatment recommendations, or guarantees of individual outcomes.

CONCLUSION

This study developed and internally evaluated treatment-specific cGANs that generated two-dimensional post-treatment appearance simulations of mature hypertrophic scars from a single pretreatment photograph for surgical excision, autologous emulsified fat injection, and MPR. The outputs generally preserved scar location and global morphology while expressing treatment-associated changes in contour, margins, color, and surface texture. Objective metrics and plastic-surgeon assessments provided complementary evidence of preliminary image-level feasibility, with expert identification more closely associated with SSIM and MSE than with PSNR. These findings support further evaluation of cGAN-based visualization as an adjunct to pretreatment communication, shared decision-making, and expectation management. The outputs should nevertheless be interpreted as possible appearance trends rather than deterministic predictions of individual outcomes.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the study and interpretation of the results: Li CF, Li SY, Chen ML;

Writing of the original draft: Li CF;

Acquisition and curation of paired clinical photographs, image preprocessing, cGAN model development, data analysis and investigation: Li CF;

Provided clinical supervision, administrative coordination, technical guidance, and material support: Li CF, Li SY, Fu Q, Zhou GW, Wu Q, Meng FT, Xu X, Chen ML;

All authors read and approved the final manuscript.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available because they contain sensitive clinical photographs, and unrestricted disclosure could compromise participant privacy. The data may be available from the corresponding author upon reasonable request.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool DeepSeek (version V4-Flash, released 2026-04-24) was used solely for language translation. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

The authors are accountable for all aspects of the work, including ensuring that any questions relating to its accuracy or integrity are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki and subsequent amendments.

Consent for publication

Not applicable.

Copyright

REFERENCES

1. Jeschke MG, Wood FM, Middelkoop E, et al. Scars. *Nat Rev Dis Primers*. 2023;9:64.[DOI: 10.1038/s41572-023-00474-x]

2. Frech FS, Hernandez L, Urbonas R, et al. Hypertrophic scars and keloids: Advances in treatment and review of established therapies. *Am J Clin Dermatol*. 2023;24:225-45.[DOI: 10.1007/s40257-022-00744-6]

3. Ogawa R. The Most Current Algorithms for the Treatment and Prevention of Hypertrophic Scars and Keloids: A 2020 Update of the Algorithms Published 10 Years Ago. *Plastic & Reconstructive Surgery*. 2022;149:79e-94e.[DOI: 10.1097/PRS.0000000000008667]

4. Ahmad N, Anker A, Klein S, et al. Autologous fat grafting-a panacea for scar tissue therapy? *Cells*. 2024;13:1384.[PMID: 39195271 DOI: 10.3390/cells13161384]

5. Baroni A, Verolino P. Plasma Radiofrequency Ablation for Scar Treatment. *JCM*. 2021;11:140.[PMID: 35011879 DOI: 10.3390/jcm11010140]

6. Waljee J, McGlinn EP, Sears ED, et al. Patient expectations and patient-reported outcomes in surgery: A systematic review. *Surgery*. 2014;155:799-808.[PMID: 24787107 DOI: 10.1016/j.surg.2013.12.015]

7. Niburski K, Guadagno E, Mohtashami S, et al. Shared decision making in surgery: A

scoping review of the literature. *Health Expectations*. 2020;23:1241-9.[PMID: 32700367 DOI: 10.1111/hex.13105]

8. Carrière ME, Mokkink LB, Tyack Z, et al. Development of the patient scale of the patient and observer scar assessment scale (POSAS) 3.0: A qualitative study. *Qual Life Res*. 2023;32:583-92.[PMID: 36355319 DOI: 10.1007/s11136-022-03244-6]

9. Kantor J. Reliability and photographic equivalency of the scar cosmesis assessment and rating (SCAR) scale, an outcome measure for postoperative scars. *JAMA Dermatol*. 2017;153:55-60.[PMID: 27806156 DOI: 10.1001/jamadermatol.2016.3757]

10. Carrière ME, Mokkink LB, Pleat J, et al. Development of the observer scales of the patient and observer scar assessment scale: An international delphi study. *Plast Reconstr Surg Glob Open*. 2025;13:e6416.[PMID: 40027471 DOI: 10.1097/GOX.0000000000006416]

11. Ito H, Nakamura Y, Takanari K, et al. Development of a Novel Scar Screening System with Machine Learning. *Plast Reconstr Surg*. 2022;150:465e-72e.[PMID: 35687417 DOI: 10.1097/PRS.0000000000009312]

12. Kim Jemin, Oh I, Lee YN, et al. Predicting the severity of postoperative scars using artificial intelligence based on images and clinical data. *Sci Rep*. 2023;13:13448.[PMID: 37596459 DOI: 10.1038/s41598-023-40395-z]

13. Fuhua Hu, Shao Y, Liu J, et al. Advances in intelligent recognition and diagnosis of skin scar images: concepts, methods, challenges, and future trends. *Front Med*. 2025;12:1667087.[PMID: 40978739 DOI: 10.3389/fmed.2025.1667087]

14. Zhou T, Li Q, Lu H, et al. GAN review: Models and medical image fusion applications. *Information Fusion*. 2023;91:134-48.[DOI: 10.1016/j.inffus.2022.10.017]

15. Isola P, Zhu J-Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (Honolulu, HI, 2017),* p. 5967-76.[DOI:10.1109/CVPR.2017.632]

16. Knoedler L, Odenthal J, Prantl L, et al. Artificial intelligence-enabled simulation of gluteal augmentation: A helpful tool in preoperative outcome simulation? *Journal of*

Plastic, Reconstructive & Aesthetic Surgery. 2023;80:94-101.[PMID: 37001299 DOI: 10.1016/j.bjps.2023.01.039]

17. Knoedler S, Alfertshofer M, Simon S, et al. Turn Your Vision into Reality—AI-Powered Pre-operative Outcome Simulation in Rhinoplasty Surgery. *Aesthetic Plast Surg*. 2024;48:4833-8.[PMID: 38777929 DOI: 10.1007/s00266-024-04043-9]

18. Zargaran A, Sousi S, Glynou SP, et al. A systematic review of generative adversarial networks (GANs) in plastic surgery. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2024;95:377-85.[PMID: 38996662 DOI: 10.1016/j.bjps.2024.04.007]

19. Zhou J, Dai Y, Liu D, et al. Improving deep learning based segmentation of scars using multi-view images. *Biomedical Signal Processing and Control*. 2024;94:106254.[DOI: 10.1016/j.bspc.2024.106254]

20. Li S, Wang H, Xiao Y, et al. A Workflow for Computer-Aided Evaluation of Keloid Based on Laser Speckle Contrast Imaging and Deep Learning. *Journal of Personalized Medicine*. 2022;12:981.[PMID: 35743764 DOI: 10.3390/jpm12060981]

21. Pham TD, Karlsson M, Andersson CM, et al. Automated VSS-based burn scar assessment using combined texture and color features of digital images in error-correcting output coding. *Sci Rep*. 2017;7:16744.[PMID: 29196632 DOI: 10.1038/s41598-017-16914-0]

22. Lee G, Hunter-Smith DJ, Rozen WM. Autologous fat grafting in keloids and hypertrophic scars: A review. *Scars Burn Heal*. 2017;3:2059513117700157.[PMID: 29799555 DOI: 10.1177/2059513117700157]

23. Riyat H, Touil LL, Briggs M, et al. Autologous fat grafting for scars, healing and pain: A review. *Scars Burn Heal*. 2017;3:2059513117728200.[PMID: 29799544 DOI: 10.1177/2059513117728200]

24. Laukka M, Kauhanen S, Hockerstedt A, et al. Tissue-level effects of autologous fat grafting in hypertrophic scars-a case series study. *J Surg Res*. 2025;305:246-57.[PMID: 39937557 DOI: 10.1016/j.jss.2024.11.039]

25. VinayKumar D, Mohsina S, Tripathy S, et al. Histological analysis of the effect of nanofat grafting in scar rejuvenation. *J Cutan Aesthet Surg*. 2022;15:147.[PMID: 35965912 DOI: 10.4103/JCAS.JCAS_106_21]

26. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-12.[PMID: 15376593 DOI: 10.1109/tip.2003.819861]

27. Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment. *Electron Lett*. 2008;44:800-1.[DOI: 10.1049/el:20080522]

28. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195.[PMID: 31665002 DOI: 10.1186/s12916-019-1426-2]

29. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28:924-33.[PMID: 35585198 DOI: 10.1038/s41591-022-01772-9]

30. Khalifa NE, Loey M, Mirjalili S. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif Intell Rev*. 2022;55:2351-77.[DOI: 10.1007/s10462-021-10066-4]

31. Elrefaie AM, Salem RM, Faheem MH. High-resolution ultrasound for keloids and hypertrophic scar assessment. *Lasers Med Sci*. 2020;35:379-85.[PMID: 31240510 DOI: 10.1007/s10103-019-02830-4]

32. Verhaegen PDHM, Van Der Wal MBA, Middelkoop E, et al. Objective Scar Assessment Tools: A Clinimetric Appraisal. *Plastic and Reconstructive Surgery*. 2011;127:1561-70.[PMID: 21460665 DOI: 10.1097/PRS.0b013e31820a641a]

33. Kashetsky N, Mar K, Liu C, et al. Photography in dermatology - a scoping review: Practices, skin of color, patient preferences, and medical-legal considerations. *J Dtsch Dermatol Ges*. 2023;21:1102-7.[PMID: 37515311 DOI: 10.1111/ddg.15129]