



Clinical Image Correction for Extraocular and Periocular Pathologies and Appearance-related Findings Via an AI-based Image-editing Tool

Yapay Zeka Tabanlı Görüntü Düzenleme Aracı ile Ekstraoküler ve Perioküler Patolojiler ve Görünümle İlişkili Bulgularında Klinik Görüntülerinin Düzeltilmesi

¹Buse PULAT¹, ²Beyza GÖKSU¹, ³Hilal HACIOĞLU², ⁴Mehtap ÖZKAHRAMAN KIRIK³, ⁵Furkan KIRIK¹,
⁶Mehmet Hakan ÖZDEMİR¹

¹Bezmialem Vakıf University Faculty of Medicine, Department of Ophthalmology, İstanbul, Türkiye

²Bezmialem Vakıf University Faculty of Medicine, İstanbul, Türkiye

³Op. Dr. Mehtap Özkahraman Clinic, İstanbul, Türkiye

ABSTRACT

Objective: To evaluate the pathology-specific correction performance and photorealistic output quality of the Gemini 2.5 Flash Image model (also known as "Nano Banana"; Google LLC, Mountain View, CA, USA) across multiple periocular and extraocular pathologies and appearance-related findings.

Methods: This retrospective study included 45 standardized clinical photographs representing nine periocular/extraocular pathologies and appearance-related findings. Each image underwent single-step, prompt-based editing using Gemini 2.5 Flash Image. Two independent raters scored images on pathology correction (Q1) and naturalness (Q2) using 5-point Likert scales. Ordinal statistics, Gwet's agreement coefficient (AC), and multilevel cumulative logistic models were applied.

Results: Median scores were high for both Q1 and Q2 across raters. Images rated ≥ 4 ranged from 84-91% for Q1 and reached 100% for Q2. Interrater agreement was excellent (Q1 AC=0.9466; Q2 AC=0.9878). Highest correction performance was observed for blepharoptosis and exotropia, while lower eyelid blepharoplasty need and thyroid eye disease showed comparatively reduced correction.

Öz

Amaç: Bu çalışmanın amacı Gemini 2.5 Flash Image modelinin (Nano Banana; Google LLC, Mountain View, CA, ABD) çoklu perioküler ve ekstraoküler patoloji ve görünümle ilişkili bulguların klinik görüntülerinin düzeltme performansını ve gerçekliğe yakın görüntü üretebilme yeteneğini değerlendirmektir.

Yöntemler: Bu retrospektif çalışma, dokuz perioküler/ekstraoküler patolojiyi ve görünümle ilişkili bulgularını temsil eden 45 standart klinik fotoğrafı içermektedir. Her görüntü, Gemini 2.5 Flash Image kullanılarak tek adımlı, patolojiye özgü metin komutlarıyla düzenlenmiştir. İki bağımsız değerlendirici, görüntüleri patoloji düzeltme (Q1) ve doğallık (Q2) açısından 5 sorulu Likert ölçeği ile skorlamıştır. Ordinal istatistikler, Gwet'in uyum katsayısı (AC) ve çok düzeyli kümülatif lojistik modeller kullanılmıştır.

Bulgular: Her iki değerlendirmede de Q1 ve Q2 için medyan skorlar yüksekti. Q1'de ≥ 4 puan alan görüntülerin oranı %84-91 iken, Q2'de tüm görüntüler %100 olarak değerlendirildi. Gözlemciler arası uyum mükemmeldi (Q1 AC=0,9466; Q2 AC=0,9878). En yüksek düzeltme performansı blefaroptoz ve ekzotropya vakalarında gözlenirken, alt kapak blefaroplasti

Address for Correspondence: Assoc. Prof. Furkan Kırık, Bezmialem Vakıf University Faculty of Medicine, Department of Ophthalmology, İstanbul, Türkiye

E-mail: f.kirik21@gmail.com

ORCID IDs of the authors: B.P.: 0009-0006-4825-9822, B.G.: 0009-0007-2005-2990, H.H.: 0009-0004-9455-8060, M.Ö.K.: 0000-0002-4053-3616, F.K.: 0000-0001-5846-8536, M.H.Ö.: 0000-0002-1719-4265

Cite this article as: Pulat B, Göksu B, Hacıoğlu H, Özkahraman Kırık M, Kırık F, Özdemir MH. Clinical image correction for extraocular and periocular pathologies and appearance-related findings Via an AI-based image-editing tool. Bezmialem Science. [Epub Ahead of Print]

Received: 07.12.2025

Accepted: 11.01.2026

Epub: 26.01.2026



©Copyright 2026 by Bezmialem Vakıf University published by Galenos Publishing House.
Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

Conclusion: Prompt based large language models enabled editing provided clinically meaningful correction with high photorealism across most pathologies, suggesting a practical alternative to dataset dependent conventional image generation approaches.

Keywords: Artificial intelligence, image editing, periocular/extraocular pathology correction, Gemini 2.5 Flash Image, image processing computer-assisted, Nano Banana

Öz

İhtiyacı olan vakalarda ve tiroid okülopatili vakalarda göreceli olarak daha düşük düzeltme elde edildi.

Sonuç: Metin komutlarına dayalı büyük dil modeli tabanlı görüntü düzenleme yaklaşımı, çoğu patolojide yüksek gerçekçilik ile birlikte klinik olarak anlamlı düzeltme sağlamıştır. Bu yöntem, geniş veri setlerine bağımlı geleneksel görüntü üretme modellerine pratik bir alternatif olarak görünmektedir.

Anahtar Kelimeler: Yapay zeka, görüntü düzenleme, perioküler/ekstraoküler patoloji düzeltme, Gemini 2.5 Flash Image, bilgisayar destekli görüntü işleme, Nano Banana

Introduction

The use of generative artificial intelligence (GenAI) for synthetic ocular image generation and editing is increasingly reported in ophthalmology. Most GenAI models, generally based on generative adversarial networks (GANs) and diffusion models, have limited clinical application due to their large-scale training data requirements (1). Visual simulation technologies, which are actively used to predict postoperative appearance, have become an important part of communication with the patient by enabling the prediction of pre- and post-operative appearance, especially in aesthetic surgeries (2-4). However, conventional simulation/editing tools are often time-consuming, costly, and dependent on proprietary software. Additionally, most of these systems fail to produce natural-looking results (5,6). Most GenAI models, generally based on GANs and diffusion models, have limited clinical application due to their large-scale training data requirements. Unlike these traditional approaches, prompt-based editing utilizes the vast pre-trained visual and semantic capabilities of foundation models, thereby overcoming the need for extensive, pathology-specific training datasets (7). New AI models now offer text-to-image (prompt based) and text-guided image editing capabilities; Google Gemini 2.5 Flash Image (also known as "Nano Banana"; Google LLC, Mountain View, CA, USA) is a recent example that can directly edit clinical photographs from textual prompts (8). In the literature, GenAI applications have largely focused on predicting treatment response in intraocular diseases or on single-pathology use cases; also, data on predicting postoperative appearance in periocular procedures are limited. In addition, the fact that the images edited in the previous studies are not evaluated together in terms of both preserving reality and correcting pathology constitutes one of the main limitations in this field (9-13).

This study aims to investigate the correction capacity of the Gemini 2.5 Flash Image model in various periocular/extraocular conditions and appearance-related findings the level of realism of the images produced.

Methods

This single-centered, retrospective study included periocular/extraocular conditions and appearance-related findings; upper eyelid dermatochalasis, the need for lower eyelid blepharoplasty, entropion, ectropion, esotropia, exotropia, brow ptosis, thyroid eye disease, and blepharoptosis. For each condition, images from five patients who presented for treatment were randomly selected, yielding 45 images in total. Inclusion criteria were: presence of one of the listed conditions, availability of a standardized clinical photograph at presentation, and documented consent for image use and processing via AI systems. Photographs were recorded as JPEGs in the standard RGB color space (quality ≥ 90) with file size ≥ 1 megabyte and a short-side dimension ≥ 1024 pixels. Technical exchangeable image file format metadata (e.g., capture device, exposure settings) were preserved to ensure standardization, while all potentially identifiable personal metadata (e.g., geolocation, patient name) were strictly removed prior to analysis. Images were captured to include the face from the hairline to the subnasal and the target finding, against a homogeneous background, centrally framed, under adequate illumination, without digital zoom, and free of obvious artifacts. All images were taken with an iPhone 16 Pro (Apple Inc., Cupertino, CA, USA). Patients with additional congenital or acquired conditions that could confound periocular/extraocular findings, or with photographs not meeting standardization criteria, were excluded. Ethical approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Bezmialem Vakıf University (decision no: 2025/418, date: 22.11.2025); all procedures adhered to the Declaration of Helsinki. Images were uploaded to the Google Gemini 2.5 Flash Image model via the AI studio interface. Access was obtained through a personal account. A separate conversation tab was created for each case, and single-step edits were performed using condition-specific text prompts that requested correction of the target finding while preserving the anatomic and natural appearance of surrounding tissues (full prompt templates are provided in Supplementary Material 1). A single output image was generated for each case. Forty-five images were generated by Gemini 2.5 Flash Image (Figure 1). The generated



Figure 1. Representative examples of AI-assisted editing in clinical photographs. Original images are shown in A, C, and E, and AI-edited outputs in B, D, and F. A-B: Blepharoptosis; C-D: Ectropion; E-F: Simulated upper eyelid blepharoplasty

images were independently assessed, under randomized presentation order, by an experienced ophthalmologist (Rater 1) and an otorhinolaryngologist with expertise in periocular facial aesthetic surgery (Rater 2). For each image, two questions were scored using a 5-point Likert scale: Q1 (correction of pathology; 1=no correction, 5=complete correction) and Q2 (natural-artificial distinguishability; 1=artificial/inappropriate, 5=completely natural).

Statistical Analysis

All analyses were conducted in Stata v19.0 (StataCorp LLC, College Station, TX, USA). Five-point Likert scores were treated as ordinal; descriptive statistics are reported as median (min-max) and category frequencies n (%). Interrater agreement was summarized using Gwet's agreement coefficient (AC) with ordinal weighting and complementary percent agreement, given their lower sensitivity to marginal imbalance and ceiling effects. To evaluate pathology-specific performance and the effect of the two items (Q1: correction; Q2: naturalness), we fit a multilevel cumulative proportional-odds logistic regression including fixed effects for item (Q1/Q2) and random intercepts for image and rater to account for repeated ratings and between-image heterogeneity. For clinical interpretability, adjusted marginal probabilities were derived for "complete correction" [p (score=5)] and "success" [p (score ≥ 4)] by pathology and item, with 95% Confidence intervals (CIs) obtained via the delta method. As a secondary robustness check for the binary endpoint (score ≥ 4), we also fit a Firth-penalized logistic regression. All p -values were two-sided with a significance threshold of $p < 0.05$.

Results

A total of 45 AI-edited images (9 conditions \times 5 cases) were scored by two independent doctors on two items (Q1: correction; Q2: naturalness). Median (min-max) scores were 5 (3-5) for Q1 and 5 (4-5) for Q2 for Rater 1, and 4 (2-5) for Q1 and 5 (4-5) for Q2 for Rater 2. The proportion of images rated ≥ 4 was 84.44% (Q1) and 100% (Q2) for Rater 1, and 91.11% (Q1) and 100% (Q2) for Rater 2 (Table 1). Interrater agreement was very high: for Q1, weighted percent agreement was 97.64% and Gwet's AC was 0.9466; for Q2, 99.03% and 0.9878, respectively.

According to Q1 medians, "need for lower lid blepharoplasty" was selected as the reference category because it had the lowest improvement scores (Rater1/Rater2: 3/4); other pathologies were browlift 5/4, ectropion 5/4, entropion 5/5, esotropia 5/4, ptosis 5/5, thyroid ophthalmopathy 4/4, upper lid blepharoplasty 4/5, exotropia 5/5, respectively. The probability of being in higher score categories compared to the reference was significantly increased for ptosis and exotropia [odds ratio (OR)=327.62; $p=0.001$]; while OR=26.38 ($p=0.013$) was observed as OR=22.25 ($p=0.018$) for browlift and esotropia and OR=22.25 ($p=0.018$) for entropion; a marginal increase was observed in upper eyelid blepharoplasty (OR=9.74; $p=0.065$), but no significant difference was observed in ectropion (OR=4.98; $p=0.182$) and thyroid ophthalmopathy (OR=1.78; $p=0.619$). The item effect favored Q2: compared with Q1, Q2 substantially increased the probability of assigning higher scores (Q2 vs. Q1 OR=34.53; 95% CI: 9.68-123.18; $p < 0.001$), indicating that the same images were rated meaningfully higher when evaluated for naturalness (Table 2).

Table 1. Rater-wise distribution of 5-point Likert scores for pathology correction (Q1) and image naturalness (Q2) in AI-edited periocular/extraocular clinical photographs (n=45 images), reported as median (min-max) and category frequencies.

Rater	Item	n	Median (min-max)	Score=2	Score=3	Score=4	Score=5
Rater 1	Q1 - Correction of pathology	45	5 (3-5)	-	15.56%	33.33%	51.11%
Rater 1	Q2 - Naturalness	45	5 (4-5)	-	-	15.56%	84.44%
Rater 2	Q1 - Correction of pathology	45	4 (2-5)	2.22%	6.67%	42.22%	48.89%
Rater 2	Q2 - Naturalness	45	5 (4-5)	-	-	8.89%	91.11%
All	(Q1+Q2, both raters combined)	180	5 (2-5)	0.56%	5.56%	25.00%	68.89%
Q1: Question 1, Q2: Question 2, AI: Artificial intelligence							

Model-based adjusted probabilities indicated that, under Q1 (correction), the likelihood of “complete correction” ($p=5$) was highest for ptosis and exotropia ($=0.90$), followed by brow ptosis and esotropia ($=0.57$), entropion ($=0.54$), upper-eyelid blepharoplasty/dermatochalasis ($=0.40$), ectropion ($=0.29$), thyroid eye disease ($=0.15$), and lowest for lower-eyelid blepharoplasty need ($=0.10$). Under Q2 (naturalness), these values increased across all pathologies: ptosis/exotropia $=0.996$; brow ptosis/esotropia $=0.956$; entropion $=0.950$; upper eyelid $=0.904$; ectropion $=0.845$; thyroid eye disease $=0.714$; lower eyelid $=0.620$ (Table 3). At the “success” threshold ($p \geq 4$), adjusted probabilities were very high for most conditions: ptosis/exotropia $=0.998$; brow ptosis/esotropia $=0.978$; entropion $=0.975$; upper eyelid $=0.951$; ectropion $=0.920$; with comparatively lower values for lower eyelid $=0.795$ and thyroid eye disease $=0.849$. When restricted to Q1, the ranking was preserved but separations became more pronounced (ptosis/exotropia $=0.996$; brow ptosis/esotropia $=0.955$; entropion/upper eyelid $=0.864$ - 0.905 ; ectropion $=0.848$; thyroid $=0.717$; lower eyelid $=0.625$). For Q2, $p \geq 4$ exceeded 0.96 for all pathologies (Table 4).

Discussion

In this study, the Gemini 2.5 Flash Image model was applied to patient photographs across nine extraocular/periocular conditions using text-based prompts to perform targeted image editing, and its correction performance was evaluated

by two independent physicians. The findings indicate that the model produced clinically meaningful results in several conditions, with the highest performance observed in blepharoptosis and exotropia, and comparatively limited effectiveness in lower eyelid blepharoplasty requirement and thyroid eye disease.

In the current literature, AI-assisted image generation/editing studies predominantly focus on a single pathology or employ GAN/diffusion-based models trained on large datasets. Indeed, a recent study developed a deep learning system to predict postoperative appearance after blepharoptosis surgery using a dataset of 362 patients, reporting patient-satisfying postoperative simulations based on both objective measurements and clinician/patient satisfaction surveys. The authors further emphasized the practical benefits of this approach for managing patient expectations, supporting clinician counseling, and reducing preoperative anxiety (14). In a related study, postoperative appearance after orbital decompression for thyroid eye disease was simulated using a GAN and evaluated on 109 matched pre-postoperative facial image pairs. While the authors highlighted the model’s potential for patient counseling, the images were compiled from Google and synthesized at relatively low resolution (64×64 pixels), resulting in outputs that lacked realism and appeared low quality (15). There are also studies in other fields in terms of generating images with the GAN models. For example, postoperative soft-tissue changes

Table 2. Multilevel cumulative proportional-odds logistic regression results for pathology categories (reference: lower eyelid blepharoplasty) and item effect (Q2 vs. Q1)

Variable (ref: Lower eyelid blepharoplasty: Q1)	OR	95% CI	p-value
Brow ptosis / browlift	26.38	2.02-344.06	0.01
Ectropion	4.98	0.47-52.41	0.18
Entropion	22.25	1.68-293.88	0.02
Esotropia	26.38	2.02-344.06	0.01
Upper eyelid ptosis	327.62	10.81-9.925.03	0.001
Thyroid eye disease	1.78	0.18-17.28	0.62
Upper eyelid blepharoplasty/dermatochalasis	9.74	0.87-108.94	0.06
Exotropia	327.62	10.81-9.925.03	0.001
Item: Q2 vs. Q1	34.53	9.68-123.18	<0.001
Q1: Question 1, Q2: Question 2, ref: Reference, OR: Odds ratio, CI: Confidence interval			

Table 3. Model-based adjusted marginal probabilities of receiving the maximum score [p (score=5), “complete correction/highest rating”] by pathology and rating item (overall, Q1, and Q2)

Pathology	Overall p (=5) (95% CI)	Q1 p (=5) (95% CI)	Q2 p (=5) (95% CI)
Lower eyelid blepharoplasty indicated	0.362 (0.161-0.562)	0.103 (-0.031-0.238)	0.620 (0.335-0.906)
Brow ptosis/browlift	0.765 (0.571-0.959)	0.574 (0.250-0.897)	0.956 (0.882-1.031)
Ectropion	0.566 (0.350-0.781)	0.286 (0.021-0.552)	0.845 (0.658-1.032)
Entropion	0.747 (0.543-0.950)	0.543 (0.209-0.877)	0.950 (0.866-1.033)
Esotropia	0.765 (0.571-0.959)	0.574 (0.250-0.897)	0.956 (0.882-1.031)
Ptosis	0.949 (0.845-1.052)	0.902 (0.706-1.098)	0.996 (0.984-1.008)
Thyroid eye disease	0.434 (0.223-0.646)	0.155 (-0.028-0.338)	0.714 (0.452-0.975)
Upper eyelid blepharoplasty/dermatochalasis	0.650 (0.437-0.862)	0.396 (0.090-0.702)	0.904 (0.768-1.039)
Exotropia	0.949 (0.845-1.052)	0.902 (0.706-1.098)	0.996 (0.984-1.008)

Q1: Question 1, Q2: Question 2, CI: Confidence interval

Table 4. Model-based adjusted marginal probabilities of meeting the success threshold [p (score=4 or 5)] by pathology and rating item (overall, Q1, and Q2)

Pathology	Overall p (≥4) (95% CI)	Q1 p (≥4) (95% CI)	Q2 p (≥4) (95% CI)
Lower eyelid blepharoplasty indicated	0.795 (0.631-0.960)	0.625 (0.342-0.908)	0.966 (0.912-1.021)
Brow ptosis/browlift	0.978 (0.940-1.016)	0.957 (0.885-1.030)	0.998 (0.995-1.002)
Ectropion	0.920 (0.820-1.020)	0.848 (0.662-1.034)	0.992 (0.977-1.008)
Entropion	0.975 (0.932-1.017)	0.951 (0.870-1.032)	0.998 (0.994-1.002)
Esotropia	0.978 (0.940-1.016)	0.957 (0.885-1.030)	0.998 (0.995-1.002)
Ptosis	0.998 (0.992-1.004)	0.996 (0.984-1.007)	1.000 (0.999-1.000)
Thyroid eye disease	0.849 (0.702-0.995)	0.717 (0.456-0.979)	0.980 (0.944-1.016)
Upper eyelid blepharoplasty/dermatochalasis	0.951 (0.880-1.021)	0.905 (0.772-1.039)	0.996 (0.987-1.005)
Exotropia	0.998 (0.992-1.004)	0.996 (0.984-1.007)	1.000 (0.999-1.000)

Q1: Question 1, Q2: Question 2, CI: Confidence interval

after osteotomy have been predicted using approaches that incorporate comprehensive computed tomography and magnetic resonance imaging data from all patients (16,17). Within ophthalmology, GANs have been used to synthesize or predict imaging across several indications, including retinal architecture after epiretinal membrane surgery, macular anatomy following macular hole surgery, and optic coherence tomography appearances after anti-vascular endothelial growth factor therapy (18,19,10,12).

It is noteworthy that GAN models are used in these studies because their methods are based on large data sets and focus on a specific disease entity. In contrast, we demonstrate that with a much smaller dataset, purely text-guided prompting can effectively drive image editing for multiple extra/periocular appearance-related findings. To avoid degrading fidelity, we analyzed images at a minimum of 1024×1024 pixels, which likely contributed to the clinically credible realism observed. This performance drop can be mechanistically linked to the model’s inability to manage volumetric (e.g., fat reduction, hollow formation, maintaining tear trough contour) or multi-planar/3D

spatial changes (e.g., correcting proptosis or eyelid-globe relationship) using a single, 2D text prompt. Clinically, this defines the limit where a prompt-based 2D editing tool fails to simulate the complexities of aesthetic surgery and postoperative appearance. Notably, “naturalness” (Q2) scores were consistently higher than “correction” (Q1), indicating that the model not only attempts the correction but also preserves photorealistic appearance. Given that the model achieves near-perfect photorealism even when the clinical correction is poor, caution should be exercised regarding the appropriate use of this tool. For complex procedures where the model’s clinical success (Q1) is demonstrated to be low (like lower lid), the high realism (Q2) could potentially create unrealistic patient expectations if the output is mistaken for an achievable post-operative result. Therefore, limiting the use of this single-step AI tool to simple, geometry-focused corrections is advised until further validation in volumetric cases is obtained.

Studies in the literature have focused on GANs, and studies using large language models to generate images from

text are limited (20-22). To our knowledge, there are no studies on editing extraocular/periocular pathology images using prompt-based large language models. One of the strengths of our study is that the performance of the model in different pathologies is reported holistically with two separate evaluations (correction/naturalness).

Study Limitations

Limitations include the absence of direct comparison with true postoperative “outcome” images, the single-center design with a limited number of images, reliance on a single AI model (Gemini 2.5 Flash Image), and assessments based on two items rated by two clinicians who were aware of the target condition, which may have introduced potential expectation bias. Future studies should incorporate blinded assessment (e.g., masking the target condition and randomizing mixed pathology sets) to reduce potential bias. We anticipate progress in this field through future studies that perform quantitative comparisons using matched pre-postoperative pairs for each pathology and conduct multi-center investigations with larger datasets that benchmark different pathologies and AI models head-to-head.

Conclusion

This work suggests that prompt-based, large language model-enabled image editing is a promising tool for extra-/periocular conditions, capable of delivering both clinically meaningful correction and photorealistic appearance. Future research with larger, multi-pathology datasets and diverse algorithms will be important to validate and extend these findings.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Bezmialem Vakıf University (decision no: 2025/418, date: 22.11.2025).

Informed Consent: Verbal informed consent was obtained from all patients for the use of their anonymized clinical photographs.

Declaration Regarding the Use of AI and AI-assisted Technologies

During the preparation of this study, the authors utilized Google Gemini 2.5 Flash Image to generate AI-edited clinical photographs from standardized patient images using predefined pathology-specific prompts. The outputs were systematically reviewed by the authors, and their accuracy, clinical plausibility, and consistency with the original image content were verified before inclusion in the analyses. After carefully reviewing and editing the content as necessary, full responsibility for the publication’s scientific and ethical integrity is taken by the authors. The use of this AI tool primarily influenced the generation of edited images used for performance evaluation and the visual dataset supporting the study outcomes.

Footnotes

Authorship Contributions

Surgical and Medical Practices: B.P., B.G., H.H., F.K., M.H., Concept: B.P., B.G., H.H., F.K., M.H., Design: B.P., B.G., H.H., F.K., M.H., Data Collection or Processing: B.P., B.G., H.H., M.Ö.K., F.K., Analysis or Interpretation: B.P., F.K., M.H., Literature Search: B.P., H.H., F.K., Writing: B.P., B.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Sonmez SC, Sevgi M, Antaki F, Huemer J, Keane PA. Generative artificial intelligence in ophthalmology: current innovations, future applications and challenges. *Br J Ophthalmol*. 2024;108:1335-40.
2. Waisberg E, Ong J, Kamran SA, Masalkhi M, Paladugu P, Zaman N, et al. Generative artificial intelligence in ophthalmology. *Surv Ophthalmol*. 2025;70:1-11.
3. Phipps B, Hadoux X, Sheng B, Campbell JP, Liu TYA, Keane PA, et al. AI image generation technology in ophthalmology: use, misuse and future applications. *Prog Retin Eye Res*. 2025;106:101353.
4. Feng X, Xu K, Luo MJ, Chen H, Yang Y, He Q, et al. Latest developments of generative artificial intelligence and applications in ophthalmology. *Asia Pac J Ophthalmol (Phila)*. 2024;13:100090.
5. Ritschl LM, Classen C, Kilbertus P, Eufinger J, Storck K, Fichter AM, et al. Comparison of three-dimensional imaging of the nose using three different 3D-photography systems: an observational study. *Head Face Med*. 2024;20:7.
6. Chang JB, Small KH, Choi M, Karp NS. Three-dimensional surface imaging in plastic surgery: foundation, practical applications, and beyond. *Plast Reconstr Surg*. 2015;135:1295-304.
7. Jin K, Yu T, Ying GS, Ge Z, Li KZ, Zhou Y, et al. A systematic review of vision and vision-language foundation models in ophthalmology. *Adv Ophthalmol Pract Res*. 2025;6:8-19.
8. Fortin A, Vernade G, Kampf K, Reshi A. Introducing Gemini 2.5 Flash Image, our state-of-the-art image model. *Google Developers Blog*. 26 Aug 2025.
9. Moon S, Lee Y, Hwang J, Kim CG, Kim JW, Yoon WT, et al. Prediction of anti-vascular endothelial growth factor agent-specific treatment outcomes in neovascular age-related macular degeneration using a generative adversarial network. *Sci Rep*. 2023;13:5639.
10. Song T, Zang B, Kong C, Zhang X, Luo H, Wei W, et al. Construction of a predictive model for the efficacy of anti-VEGF therapy in macular edema patients based on OCT imaging: a retrospective study. *Front Med (Lausanne)*. 2025;12:1505530.
11. Cao J, You K, Jin K, Lou L, Wang Y, Chen M, et al. Prediction of response to anti-vascular endothelial growth factor treatment in diabetic macular oedema using an optical coherence tomography-based machine learning method. *Acta Ophthalmol*. 2021;99:e19-27.
12. Xu F, Liu S, Xiang Y, Hong J, Wang J, Shao Z, et al. Prediction of the short-term therapeutic effect of anti-VEGF therapy for diabetic macular edema using a generative adversarial network with OCT images. *J Clin Med*. 2022;11:2878.

13. Lee H, Kim N, Kim NH, Chung H, Kim HC. Generative deep learning approach to predict posttreatment optical coherence tomography images of age-related macular degeneration after 12 months. *Retina*. 2025;45:1184-91.
14. Sun Y, Huang X, Zhang Q, Lee SY, Wang Y, Jin K, et al. A fully automatic postoperative appearance prediction system for blepharoptosis surgery with image-based deep learning. *Ophthalmol Sci*. 2022;2:100169.
15. Yoo TK, Choi JY, Kim HK. A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. *Comput Biol Med*. 2020;118:103628.
16. Pan B, Xia JJ, Yuan P, Gateno J, Ip HH, He Q, et al. Incremental kernel ridge regression for the prediction of soft tissue deformations. *Med Image Comput Comput Assist Interv*. 2012;15:99-106.
17. Pan B, Zhang G, Xia JJ, Yuan P, Ip HH, He Q, et al. Prediction of soft tissue deformations after CMF surgery with incremental kernel ridge regression. *Comput Biol Med*. 2016;75:1-9.
18. Kim J, Chin HS. Deep learning-based prediction of retinal structural alterations after ERM surgery. *Sci Rep*. 2023;13:19275.
19. Kwon HJ, Heo J, Park SH, Park SW, Byon I. Accuracy of generative deep learning model for macular anatomy prediction in macular hole surgery. *Sci Rep*. 2024;14:6913.
20. Wu X, Wang L, Chen R, Liu B, Zhang W, Yang X, et al. Generation of fundus fluorescein angiography videos for health care data sharing.
21. Balas M, Micieli JA, Wulc A, Ing EB. Text-to-image artificial intelligence models for preoperative counselling in oculoplastics. *Can J Ophthalmol*. 2024;59:e75-e76.
22. Jong S, Dihan QA, Khodeiry MM, Alzein A, Scelfo C, Elhusseiny AM. Evaluating text-to-image generation in pediatric ophthalmology. *J Pediatr Ophthalmol Strabismus*. 2025:1-7.

Supplementary Material 1. <https://d2v96fxpocvxx.cloudfront.net/c2f7718d-0796-4c18-a6d4-3c339e748b23/content-images/e6a4e027-71e8-463f-8aea-aaae70bf863c.pdf>