

AI Grading of Lateral Canthal Lines: Novel Models for Unseen Synthetic Image Generation and Data Augmentation

Ting-Ting Yang ^{1,2}, Ching-Wen Ma ³, Chiu-Hsien Lee^{3,*}, Shi-Xuan Qiu^{3,*}, Cheng-Che E Lan^{1,4,5}

¹Graduate Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; ²Department of Dermatology, Kaohsiung Medical University Gangshan Hospital, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan; ³College of Artificial Intelligence, National Yang Ming Chiao Tung University, Hsinchu, Taiwan; ⁴Department of Dermatology, Kaohsiung Medical University Hospital and College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; ⁵Biomedical Artificial Intelligence Academy, Kaohsiung Medical University, Kaohsiung, Taiwan

*These authors contributed equally to this work

Correspondence: Cheng-Che E Lan, Department of Dermatology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, 100 Shih-Chuan 1st Road, Kaohsiung, Taiwan, Email laneric@cc.kmu.edu.tw

Purpose: Large and balanced datasets are required to train artificial intelligence (AI) algorithms but are often difficult to acquire using clinical dermatologic photographs. We aimed to develop a new diffusion-based generative AI algorithm that generates patient photographs with modifiable details, thereby creating large and balanced datasets for neural network training. The newly developed model was tested using lateral canthal lines, a common patient's reason for seeking treatment.

Patients and Methods: Five hundred and sixty-six photographs of the lateral oblique face of graded by certified dermatologists according to the severity of lateral canthal lines. We developed a zero-shot and few-shot image generation model that adds structured compositional labels as control variables to the diffusion model. This allows us to create synthetic images from original photos while adjusting the severity of lateral canthal lines with only a few examples. The generated images were used to train a convolutional neural network (CNN) with ResNet-34 backbone for classifying the grade of lateral canthal lines.

Results: We successfully generated 10,500 patient images similar to the original photographs with different grades of lateral canthal lines. The accuracy (82% vs 91%) and the area under the receiver operating characteristic curve (0.935 vs 0.981) of the classification CNN remarkably improved after training with the new dataset containing generated images.

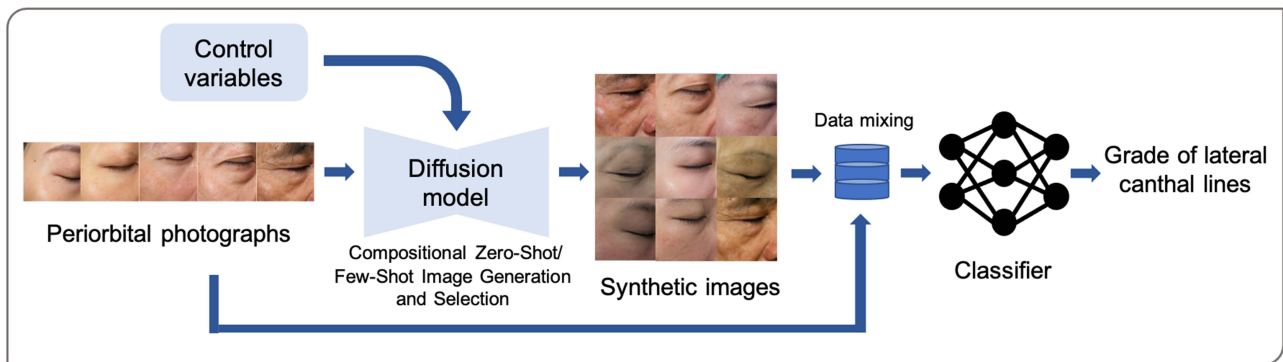
Conclusion: The compositional zero-shot and few-shot generation model is able to generate images similar to original clinical photographs, and the features of the images can be modified to match the needs of the specific task, allowing researchers to create a larger and more balanced dataset to improve neural network training outcomes. This is especially important in dermatology, where large-scale clinical photographs are difficult to acquire for machine learning. The results of this study are limited by low patient diversity and a lack of external validation.

Keywords: artificial intelligence, skin wrinkling, lateral canthal lines, generative AI, data augmentation

Introduction

The application of artificial intelligence (AI) in dermatology has been increasing with promising results.¹⁻⁴ A large labeled training dataset is usually required to develop accurate and reliable AI algorithms for clinical applications since small datasets are prone to overfitting and cannot handle extreme cases.⁵ However, large-scale clinical images are frequently lacking and difficult to acquire. Human labeling of these images is also time and labor-consuming. Moreover, imbalanced or biased data in training datasets also affects the outcome of AI systems. These constraints may limit the development of AI algorithms for dermatologic applications.

Graphical Abstract



To address these issues, traditional data augmentation methods that involve transforming existing data through operations such as rotation, flipping, scaling, translation, and color transformation have been utilized.⁶ However, traditional data augmentation may encounter problems including excessive deformation leading to distortion and is not suitable for all tasks.⁶ Therefore, a more recent approach to data augmentation involves using Generative Adversarial Networks (GANs).⁷ Synthetic images generated from GANs have been applied to medical imaging, including images of skin cancer, onychomycosis, and acne.^{8–11} Recently, Cho et al utilized GAN to generate synthetic data from non-standardized images to train a convolutional neural network (CNN) for classifying melanocytic skin lesions with high accuracy.¹² However, GANs may face challenges such as insufficient generation diversity and difficulty in neural network training.¹³ Diffusion-based image generation models have been developed to address these limitations. They are increasingly applied in medical imaging for tasks such as image segmentation, reconstruction, anomaly detection, and synthetic image generation, among others.^{14,15} Although diffusion-based image generation models offer greater controllability compared to GANs, achieving precise localized control of the generated image is still a challenge. To date, diffusion-based image generation models have not been applied to dermatological clinical images. In this study, we introduce a compositional zero-shot and few-shot generation model that incorporates structured compositional labels into the diffusion framework, thereby enhancing image generation controllability for generating clinical dermatologic photographs.

Prominent lateral canthal lines, or crow's feet, are a common reason for patients seeking aesthetic treatments. The treatment outcome is evaluated using validated grading scales.^{16,17} Standardized grading of lateral canthal lines is important for providing objective severity assessment for treatment planning and outcome monitoring, allowing for improved patient satisfaction and increased reproducibility of research in this area. In this study, we aimed to create a novel generative AI algorithm capable of generating high-quality patient images of the periorbital region with modifiable lateral canthal line severity based on standardized clinical photographs. In this paper, we describe the process of generating patient images with canthal wrinkles using the compositional zero-shot and few-shot generation model.⁴ We demonstrate how the generated images can address the issues of insufficient medical imaging samples or data imbalance in a population of East Asians, thereby improving the overall accuracy of the classifier performance. The generated photographs were used to train a CNN classifier for grading lateral canthal line severity. We also trained CNNs for lateral canthal line grading based on original patient photographs or a dataset containing images generated using traditional data augmentation techniques. The classification accuracies of the above three CNNs were compared.

Materials and Methods

Dataset and Image Preprocessing

Photographs of 121 subjects (6 males, 115 females; mean age 43.53 ± 10.6 years) receiving procedures for aesthetic enhancement, including laser treatments and injectables, were included. All patients are of East Asian descent and have

Fitzpatrick skin type III or IV. Standardized photographs of the left and right lateral oblique face under resting conditions were obtained using the VISIA[®] skin analysis system under standard lighting (7th generation, Canfield Scientific).¹⁸ 283 treatment sessions and 566 (left and right) photographs were used for evaluation. The severity of lateral canthal lines of each image was evaluated by two board-certified dermatologists using a five-point rating scale (0, 1, 2, 3, 4) with a grade 0 indicating no wrinkles and higher grades indicating more severe wrinkles (Figure 1a).¹⁷ In cases where discrepancies occurred, the evaluators discussed the case and made the final decision. All photographs used in this study were preprocessed by cropping, resizing, and normalization. The periorbital region containing complete lateral canthal lines was cropped from the original images. The original image resolution was approximately 4K, and the periorbital region containing complete lateral canthal lines was cropped to 1024*1024 pixels. The cropped images were resized to 128*128 pixels and normalized with a mean of 0.5 and a variance of 0.5 (Figure 1a). This normalization aims to stabilize the model training process and reduce the impact of large differences in input data. This study was approved by the Kaohsiung Medical University Hospital Institutional Review Board (KMUHIRB-E(I)-20190418), and the requirement for patient consent was waived as all images used were deidentified. The study was conducted in compliance with the Declaration of Helsinki and followed the CLEAR Derm consensus guidelines¹⁹ for dermatology AI and the complete checklist is provided in [Supplementary Table 1](#).

Generative Model: The Diffusion Process

The image generation model was based on a diffusion model architecture.²⁰ Diffusion models consist of two processes: forward diffusion and inverse diffusion. Forward diffusion occurs in the form of a Markov chain progressively adding Gaussian noise with variance β_t to the latent variable x_{t-1} to generate a new latent variable x_t . The Forward Diffusion Process can be represented as

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I)$$

This process is akin to continuously adding noise to an image until it becomes completely blurred, resembling a situation where the image is entirely composed of noise, as depicted in [Supplementary Figure 1](#). The inverse diffusion process involves progressively reversing the forward process and is represented by the following equation

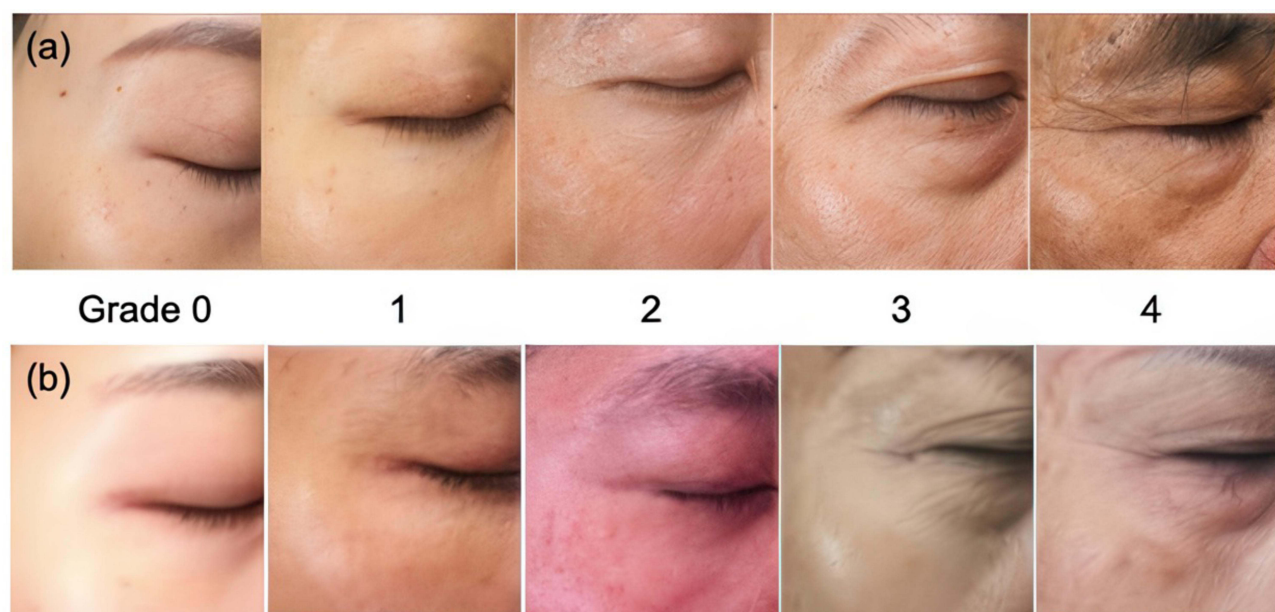


Figure 1 Original dataset of patient photographs with lateral canthal line grading from grade 0 to 4 after cropping the periorbital areas (a) and synthetic images with corresponding severity of lateral canthal lines using the compositional zero-shot and few-shot generation model (b).

$$p_{\theta}(x_{t-1}|x_t) = N(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t))$$

This process entails using a trainable neural network to remove the added noise and restore the original data ([Supplementary Figure 2](#)). The key to the inverse diffusion process lies in how the model is trained to reconstruct clear images from the observed blurry images.

When the model can accurately predict the noise at each moment, we only need to sample once from the original Gaussian noise and iteratively denoise it to reconstruct a complete image. Through the layers of denoising in the inverse diffusion process, we can generate the desired dataset. Due to the control of added variance, each generated photograph can have slight variations.

With many major companies investing significant resources in building and researching generative models, numerous impressive models have emerged, such as OpenAI DALL-E 2,²¹ Google Images,²² or large-scale image generation models like Stable Diffusion.²³ However, applying these large models trained on massive datasets, often in the millions, to our tasks presents certain challenges. Their datasets encompass a wide range of styles, species, and forms, which differ greatly from the images in our task, making it difficult for subsequent classification models to learn key features and even causing counterproductive effects. How to mitigate this problem will be discussed in the following section.

The Compositional Zero-Shot and Few-Shot Generation Model

To make the above-described diffusion model generate specific images, a control variable c must be added that constitutes a conditional generation model. The forward diffusion model thus becomes

$$q(x_t|x_{t-1}, c) = N\left(x_t \middle| c; \sqrt{1 - \beta_t}x_{t-1} \middle| c, \beta_t I\right)$$

and the inverse diffusion process is represented as

$$p_{\theta}(x_{t-1}|x_t, c) = N(x_{t-1}; \mu_{\theta}(x_t, t, c), \Sigma_{\theta}(x_t, t, c))$$

To mitigate the issues concerning existing generative AI models, this study utilizes fully initialized parameters of the generative model in conjunction with a specialized label embedding layer. By training the label embedding layer to induce the model with an understanding of the concepts and corresponding relationships among different compositions, the advantage lies in the likelihood that the distribution of images generated by the model will exhibit a higher correlation with the training data.

Before introducing conditional generative models, it is essential to elucidate the concept of compositionality. Compositionality refers to the cognitive ability of humans to decompose, recombine, and validate compositional concepts. The recombined concepts may be entirely new or reinforce existing ones. For instance, “red apple”, “running lion”, and “beautiful rose” can be decomposed and recombined into “red rose” and “running apple”, where “red rose” is plausible while “running apple” is not. Applied to wrinkle generation, this concept translates to the concepts of images corresponding to wrinkle grades in the dataset, such as the distribution of wrinkle grades from grades 0 to 4. If the generative model can integrate different wrinkle data from the dataset and learn their concepts through neurons, it can subsequently infer and generate a large number of images corresponding to their grades but exhibit slight variations ([Figure 2](#)). This process aids significantly in augmenting the dataset and subsequent wrinkle classification.²⁴ A more detailed explanation of the concept of compositional zero-shot image generation is illustrated in the supplementary file ([Supplementary Figure 3](#)).

The model architecture is depicted in [Figure 3](#). The noise fitting in the backward process is predicted by Unet.²⁵ The training process runs a total of 1000 iterations, with each iteration consisting of 100 noise training and parameter update steps. The conditioning control here is represented by compositional class labels and the model learns the corresponding relationships during training. The hyperparameters of the generation model are summarized in [Supplementary Table 2](#). All generated images were manually reviewed, and unsuitable images were discarded.

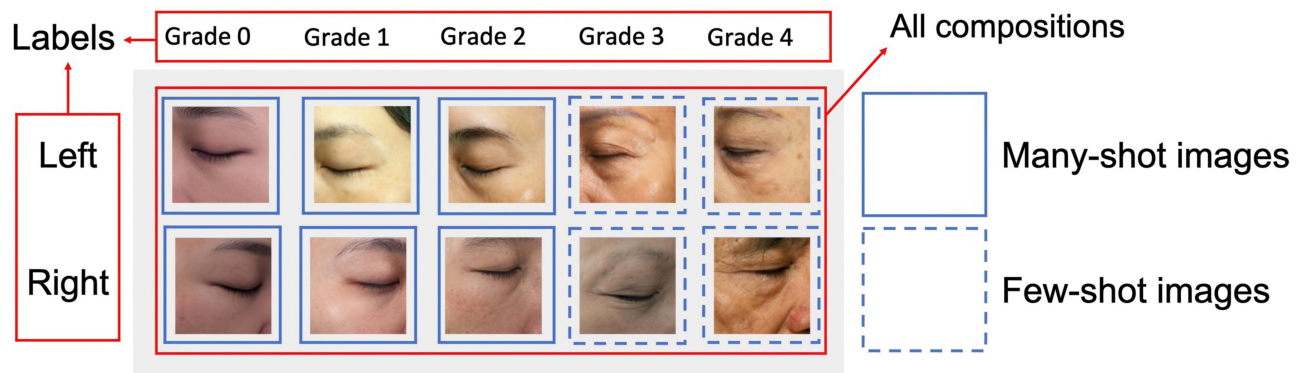


Figure 2 Illustration of applying compositional few-shot image generation to lateral canthal lines.

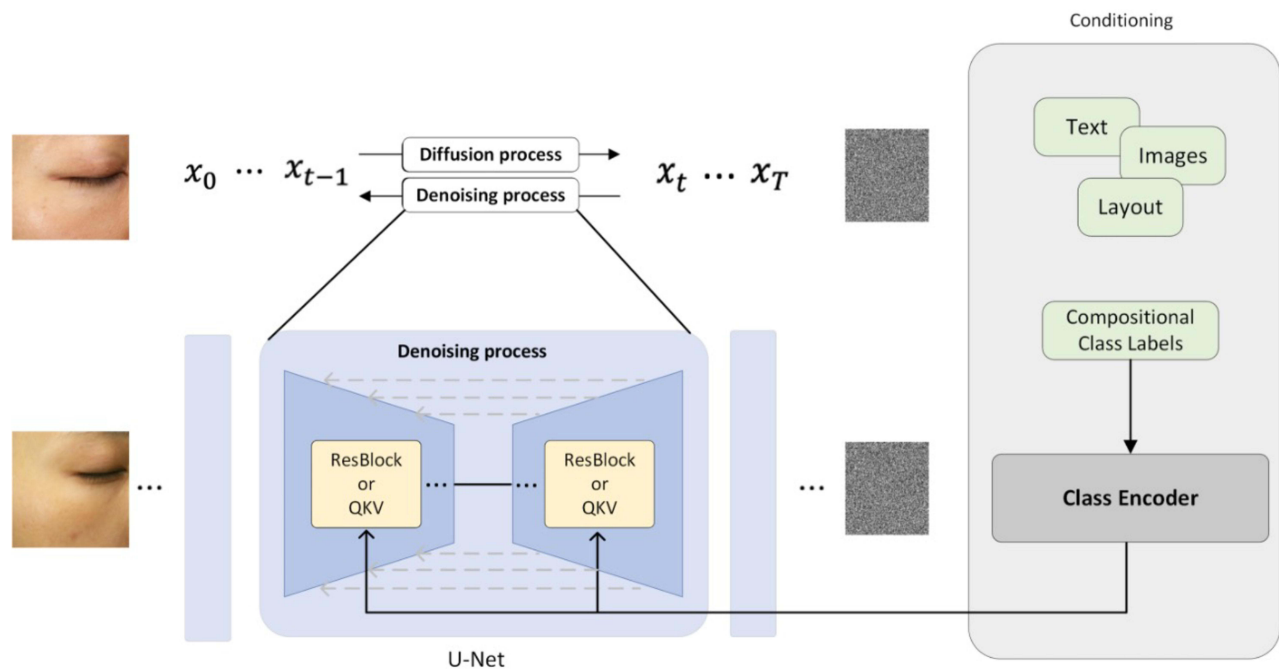


Figure 3 Architecture of the generation model.

Classification Model

A neural network composed of a CNN feature extractor and a classifier was employed to classify the severity of lateral canthal lines. As shown in [Supplementary Figure 4](#), the feature extractor was based on the ResNet-34 network pre-trained on ImageNet.²⁶ The output of the average pooling layer was a 512-dimensional feature vector. The feature vector was further connected to a classifier composed of three fully connected (Fc) layers and three dropout layers. The activation function of the last output layer was a Softmax function, which generates the conditional probability estimates for each class. The classifier gradually reduced the output dimension from 512, 512, to 256. The final output indicates the probability of the grade of lateral canthal lines of the input image and the highest score is chosen as the model’s final prediction result. The neural network in the present study was built and stimulated on PyTorch. The training details of the classification model are summarized in [Supplementary Table 3](#).

Model Training and Validation

To verify the impact of the compositional zero-shot and few-shot generation model, we compared the lateral canthal line classification accuracy, learning curves, and area under the receiver operating characteristic curve of the CNN trained on

three different datasets. The first approach involved using the original dataset and splitting it into training, validation, and test sets in a 64%–16%–20% ratio. The second approach also used the original dataset but incorporates image processing techniques including rotation (-10° to $+10^\circ$), random sharpening, automatic contrast adjustment, and random flipping (10%). The data split remained the same as in the first approach. These operations were performed using PyTorch functions. The third approach, which is the focus of this study, involved using a compositional zero-shot and few-shot generation model to augment the dataset in addition to traditional data augmentation methods described in the second approach. More specifically, 2100 training images were generated for each grade, resulting in 10,500 images generated across 5 grades. Sixty-four percent of the original data and 80% of the generated images form the new training dataset, while 16% of the original data and 20% of the generated images were used for the validation set, and 20% of the original data are used for the test set. Each model underwent 200 epochs, with validation and calculation of validation accuracy performed after each training iteration. Following training, the model with the highest validation accuracy is selected as the final model. The outcomes of the model trained on the three datasets were evaluated in terms of overall classification accuracy.

Results

Generated Images

A total of 2100 generated images were selected for each class, with the aid of human experts, resulting in 10,500 generated images across 5 grades of lateral canthal lines. The new dataset consisting of 80% of original photographs and generated images was more balanced compared to the original dataset as the proportion of each grade is similar in the new dataset ([Supplementary Figure 5](#)). Examples of images generated by using the compositional zero-shot and few-shot generative model are shown in [Figure 1b](#). The overall style of the generated images is similar to the original photographs.

Grading Results: Overall Accuracy Improved After Adding Generated Images to the Dataset

The grading accuracies of the three approaches are summarized in [Table 1](#). In one test that showed the greatest performance difference, the model trained on the original dataset achieved an accuracy of 82%, while the second approach, utilizing traditional data augmentation methods, yielded an accuracy of 85%. The third approach, using the compositional zero-shot and few-shot generation model in addition to the traditional data augmentation methods, achieved an accuracy of 91%. The mean \pm standard deviations of classification accuracies are 0.8346 ± 0.0233 for the original dataset, 0.8582 ± 0.0123 with traditional augmentation, and 0.8964 ± 0.0073 when generated images were included. Clearly, the accuracy of the three methods increases progressively, which aligns with our hypothesis that increasing the amount of data can improve the model's fitting ability, resulting in higher accuracy.

Comparison of Learning Curves

The orange line in [Supplementary Figure 6](#) represents the accuracy of the training set at different time intervals, while the blue line represents the accuracy of the validation set at different time intervals. From left to right of the graph, it can be observed that the model's accuracy improves in the rightmost graph (82% to 91%). The larger gap between the training

Table 1 Performance of Classification Models Trained on the Dataset with and without Image Augmentation

	Original Dataset	Original Dataset Plus Image Augmentation with Traditional Methods	Dataset Augmented with Both Traditional Methods and Synthetic Images
Overall classification accuracy (largest difference)	82%	85%	91%
Overall classification accuracy, mean (SD)	0.835 (0.023)	0.858 (0.012)	0.896 (0.007)
Overall AUC	0.935	0.95	0.981
AUC for grade 4 lateral canthal lines	0.94	0.937	0.997

Abbreviations: AUC, area under the curve; SD, standard deviation.

and testing curves in the three graphs also indicates that the rightmost graph has a denser gap, suggesting that using generated images for data augmentation slightly alleviates overfitting issues. While overfitting did not occur in the middle graph, its accuracy plateaued around 85%.

Receiver Operating Characteristics

The areas under the curves (AUC) of the three approaches are summarized in [Table 1](#). The AUC of the classification model trained on the dataset including the generated images was superior compared to the other two methods (0.981 vs 0.950 and 0.934) ([Figure 4f](#)). Moreover, within the same method, we also conducted receiver operating characteristic (ROC) comparisons across different output categories. It is noted that our proposed method outperforms the other two methods in four out of five grades, with only one grade (grade 1) slightly lower than the traditional data augmentation approach ([Figure 4a–d](#)). This further validates the enhancement of our method in classification effectiveness.

Confusion Matrices

As observed from the confusion matrices ([Supplementary Figure 7](#)), each model generally does not exhibit significant errors, such as misclassifying grade 4 lateral canthal lines as grade 0 or 1. Most errors occur in misclassifications with a difference of one grade. Moreover, models trained on the dataset augmented with the generated images tend to accurately predict the correct answers and confine errors within one grade difference ([Supplementary Figure 7c](#)). On the other hand, models trained on the original dataset or dataset augmented with traditional methods sometimes exhibit misclassifications with a difference of two grades or more ([Supplementary Figure 7a and b](#)). It can also be noted from the plots that using the dataset augmented with generated images shows significant improvements in the classification of grade 4 categories compared to the other two methods. This also validates the notion that augmenting data with the generated images can alleviate classification errors caused by data imbalance.

Discussion

In this study, we developed an AI image generator based on a compositional zero-shot and few-shot generation model, a conditional image generation algorithm. The innovation of this study lies in developing and training a diffusion model specifically tailored to our dataset, ensuring that the generated images are consistent with the original dataset distribution and characteristics, thereby avoiding the creation of incongruous images. Additionally, we employed a novel classification model training method, generating a large volume of synthetic data from a small initial dataset to enhance the training process. This approach was validated by comparing the classification model trained with these synthetic data to models trained without them, demonstrating superior accuracy and stability in the model's performance.

There are increasing applications of deep learning on dermatologic images, but large datasets with high-quality images are difficult to obtain.⁵ Increasing training data improves the performance of CNN classifiers.²⁷ Traditional data augmentation techniques, including image rotation, flipping, cropping, noise injection, and color space transformation, have been utilized to increase data size.⁶ However, important information may be discarded during image transformation.⁶ Recently, increasing studies using a generative adversarial network (GAN) for data augmentation have produced reliable results.¹² Nevertheless, GAN models still face many challenges, including insufficient generation diversity and difficulty in neural network training.¹³

Among various image generation models, diffusion models have gained great popularity recently for their ability to generate high-quality, realistic images.²⁰ Initially, currently off-the-shelf diffusion models (eg, Stable Diffusion) were used to generate images for this project. However, with Stable Diffusion, we were not able to generate images that matches the style of our original photographs and could not control the severity of lateral canthal lines of each generated image to the desired grade. By using the compositional zero-shot and few-shot generation model, images similar to the original photographs with controllable lateral canthal line severity were generated ([Figure 1b](#)). This approach allows researchers to generate realistic images according to the specific requirements of each task. In this study, the generated images were incorporated into the original dataset of clinical photographs for training a classification model to overcome the limitations of a small dataset. The model trained on the dataset with synthetic images demonstrated both superior classification accuracy (91% vs 85% and 82%) and AUC (0.981 vs 0.950 and 0.934) compared to the original dataset and

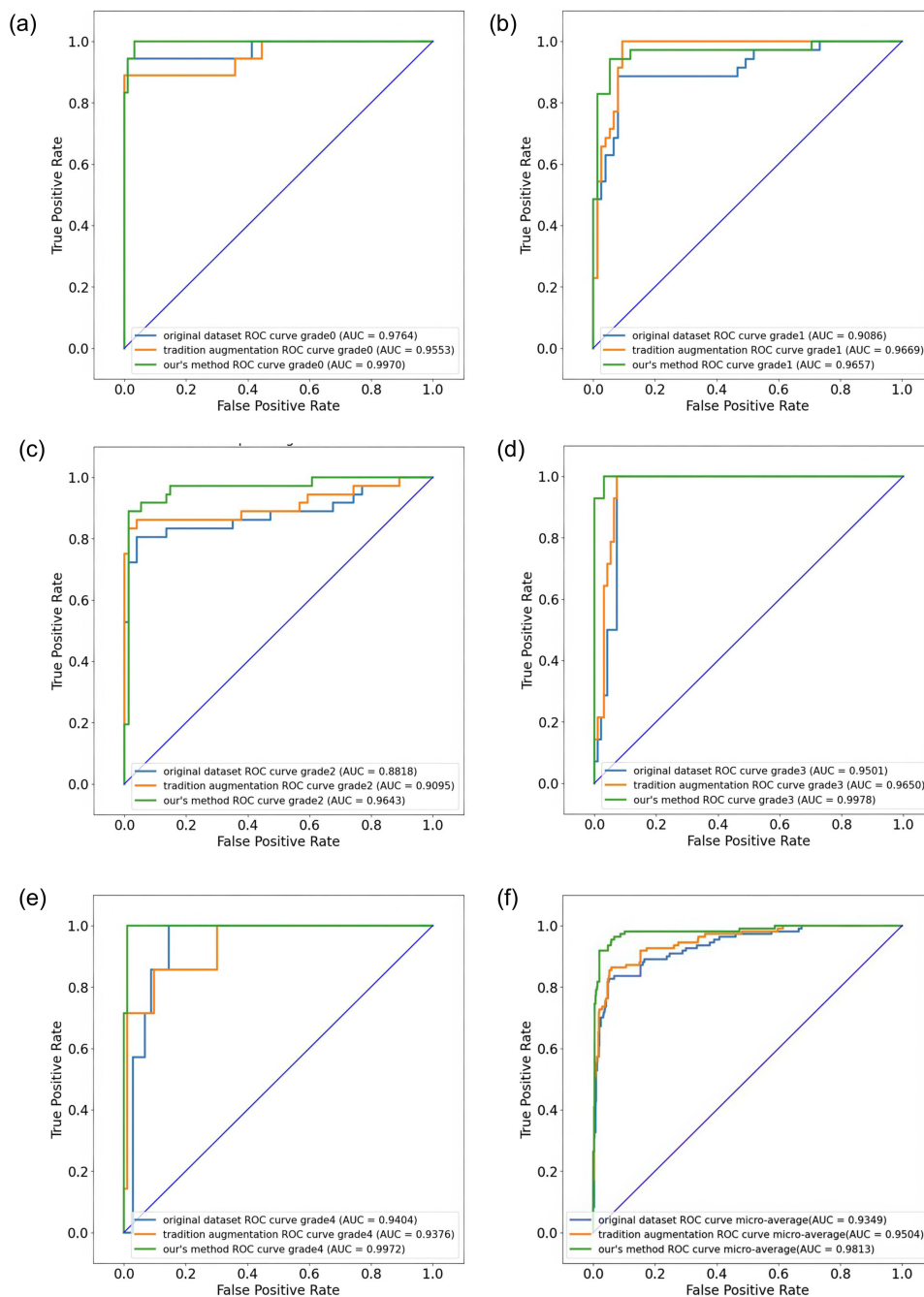


Figure 4 The receiver operating characteristic (ROC) curves for each grade (a–e) of the original dataset (blue line), the traditional data augmentation dataset (orange line), and the dataset with generated images (green line). The area under the curve of the dataset with images generated using our proposed generative model was greater than that of the other two datasets (f).

the dataset augmented using traditional augmentation methods, respectively. This result suggests that the compositional zero-shot and few-shot generation model is an effective approach for data augmentation in AI model training.

In addition to the size of datasets, imbalanced data also affect the outcome of CNN classifiers, resulting in a less accurate classification for significantly underrepresented categories.²⁸ Imbalanced data is commonly encountered in clinical data. In the present study, our original dataset was also imbalanced and contained a higher proportion of grade 1 and grade 2 images and significantly fewer grade 4 photographs (Supplementary Figure 5a). Consequently, the AUCs for grade 4 lateral canthal lines were both below 0.95 for both the classification neural network trained on the original dataset

and the dataset using traditional augmentation methods (Figure 4e). With generated images (Figure 4e), however, the AUC for grade 4 lateral canthal lines was improved to 0.99 after training with a more balanced dataset. This result confirmed that generated images allow for a more balanced dataset and improve classification outcomes.

Another important issue in AI model training is the lack of data representativeness and diversity.²⁹ Lack of training data incorporating underrepresented populations may lead to biased results and reinforcement of social inequalities.²⁹ In dermatology, the underrepresentation of darker skin tones in training datasets contributes to reduced AI model accuracy in patients with these skin types.³⁰ We believe that the compositional zero-shot and few-shot model may also be applied to overcome this problem, as it is able to generate images with limited examples. Incorporating generated images of underrepresented populations for AI model training will be a valuable topic for future research.

This study has several limitations. First, there is limited diversity of the study population as most of the subjects enrolled are females of East Asian descent with Fitzpatrick skin types III–IV. The results should be interpreted with caution when applied to other populations. However, as the compositional zero-shot and few-shot generation model is able to generate images with limited dataset, it may also be used to overcome this issue by incorporating examples of underrepresented populations. More importantly, this study aims to demonstrate that incorporating images generated by the newly developed model can enhance the training of a classification model, and our results support this objective. Secondly, the quality of the generated images was not evaluated by objective metrics. However, all generated images were examined by human experts to ensure only adequate images were included for training. We believe that this mechanism allows us to integrate and leverage knowledge from both human judgment and data-driven processes. Lastly, as this is a single-center study, further external validation will be valuable to confirm the generalizability of our results.

Conclusion

The results of this study suggest that the compositional zero-shot and few-shot generation model can be used to produce clinical images similar to real patients with different characteristics. Incorporating these synthetic images could help build larger and more balanced datasets for AI algorithms training, thereby improving classification outcomes. While future projects focusing on AI applications using dermatologic photographs could potentially benefit from this image generation method, further validation is needed to determine its reliability in a more diverse setting.

Acknowledgments

We sincerely thank Dr Yng Sun for her assistance in grading patient photographs.

Funding

This study was funded by Kaohsiung Medical University Hospital (grant numbers KMUH111-1R60, KMUH-DK(B) 113001-2, and SH11206) and the H&J Global Chair.

Disclosure

Mr Chiu-Hsien Lee reports grants from Kaohsiung Medical University Chung-Ho Memorial Hospital, during the conduct of the study. Mr Shi-Xuan Qiu reports grants from Kaohsiung Medical University Chung-Ho Memorial Hospital, during the conduct of the study. The author(s) report no conflicts of interest in this work.

References

- Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W. Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatol Ther*. 2020;10(3):365–386. doi:10.1007/s13555-020-00372-0
- Jeong HK, Park C, Henao R, Kheterpal M. Deep learning in dermatology: a systematic review of current approaches, outcomes, and limitations. *JID Innovations*. 2023;3(1):100150. doi:10.1016/j.xjidi.2022.100150
- Puri P, Comfere N, Drage LA, et al. Deep learning for dermatologists: part II. Current applications. *J Am Acad Dermatol*. 2022;87(6):1352–1360. doi:10.1016/j.jaad.2020.05.053
- Lai SL, Chen PC, Ma CW. Compositional conditional diffusion model. In: 2024 10th International Conference on Applied System Innovation (ICASI); Kyoto, Japan; IEEE; 2024. 377–379. doi:10.1109/ICASI60819.2024.10547979
- Murphree DH, Puri P, Shamim H, et al. Deep learning for dermatologists: part I. Fundamental concepts. *J Am Acad Dermatol*. 2022;87(6):1343–1351. doi:10.1016/j.jaad.2020.05.056

6. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6(1):60. doi:10.1186/s40537-019-0197-0
7. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM*. 2020;63(11):139–144. doi:10.1145/3422622
8. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. Sakakibara M, ed. *PLoS One*. 2018;13(1):e0191493. doi:10.1371/journal.pone.0191493
9. La Salvia M, Torti E, Leon R, et al. Deep convolutional generative adversarial networks to enhance artificial intelligence in healthcare: a skin cancer application. *Sensors*. 2022;22(16):6145. doi:10.3390/s22166145
10. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal*. 2019;58:101552. doi:10.1016/j.media.2019.101552
11. Zein H, Chantaf S, El-Saleh R, Nait-Ali A. Generative adversarial networks based approach for artificial face dataset generation in acne disease cases. In: *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. IEEE; 2021:1–4. doi:10.1109/BioSMART54244.2021.9677572
12. Cho SI, Navarrete-Dechent C, Daneshjou R, et al. Generation of a melanoma and nevus data set from unstandardized clinical photographs on the internet. *JAMA Dermatol*. 2023;159(11):1223. doi:10.1001/jamadermatol.2023.3521
13. Bau D, Zhu JY, Wulff J, et al. Seeing What a GAN Cannot Generate. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea: IEEE; 2019:4501–4510. doi:10.1109/ICCV.2019.00460
14. Kazerouni A, Aghdam EK, Heidari M, et al. Diffusion models for medical image analysis: a comprehensive survey. *arXiv*. 2022. doi:10.48550/ARXIV.2211.07804
15. Nazir M, Aqeel M, Setti F. Diffusion-based data augmentation for medical image segmentation. *arXiv*. 2025. doi:10.48550/ARXIV.2508.17844
16. Carruthers A, Carruthers J, Hardas B, et al. A validated grading scale for crow's feet. *Dermatologic Surg*. 2008;34:S173–S178. doi:10.1111/j.1524-4725.2008.34367.x
17. Lee SJ, Kim J-I, Yang YJ, Nam JH, Kim W-S. Treatment of periorbital wrinkles with a novel fractional radiofrequency microneedle system in dark-skinned patients. *Dermatologic Surg*. 2015;41(5):615–622. doi:10.1097/DSS.0000000000000216
18. Goldsberry A, Hanke CW, Hanke KE. VISIA system: a possible tool in the cosmetic practice. *J Drugs Dermatol*. 2014;13(11):1312–1314.
19. Daneshjou R, Barata C, Betz-Stablein B, et al. Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR dermatology consensus guidelines from the international skin imaging collaboration artificial intelligence working group. *JAMA Dermatol*. 2022;158(1):90. doi:10.1001/jamadermatol.2021.4915
20. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. *Adv Neural Inf Process Syst*. 2021;34:8780–9434. doi:10.48550/ARXIV.2105.05233
21. Marcus G, Davis E, Aaronson S. A very preliminary analysis of DALL-E 2. *arXiv*. 2022. doi:10.48550/ARXIV.2204.13807
22. Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv*. 2022. doi:10.48550/ARXIV.2205.11487
23. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, Louisiana, USA: IEEE; 2022:10674–10685. doi:10.1109/CVPR52688.2022.01042
24. Xu G, Kordjamshidi P, Chai J. Zero-shot compositional concept learning. In: *Proceedings of the 1st Workshop on Meta Learning and Its Applications to Natural Language Processing*. Bangkok, Thailand (Online): Association for Computational Linguistics; 2021:19–27. doi:10.18653/v1/2021.metanlp-1.3
25. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Vol 9351. Lecture Notes in Computer Science*. Springer International Publishing; 2015:234–241. doi:10.1007/978-3-319-24574-4_28
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, Nevada, USA: IEEE; 2016:770–778. doi:10.1109/CVPR.2016.90
27. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211–252. doi:10.1007/s11263-015-0816-y
28. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*. 2018;106:249–259. doi:10.1016/j.neunet.2018.07.011
29. Shams RA, Zowghi D, Bano M. AI and the quest for diversity and inclusion: a systematic literature review. *AI Ethics*. 2025;5(1):411–438. doi:10.1007/s43681-023-00362-w
30. Joerg L, Kabakova M, Wang JY, et al. AI-generated dermatologic images show deficient skin tone diversity and poor diagnostic accuracy: an experimental study. *Acad Dermatol Venereol*. 2025:jdv.20849. doi:10.1111/jdv.20849

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