Skin Cancer Image Generation Using WGAN-GP Based on HAM10000 Dataset

Hongye Hao

Bachelor of Engineering in Data Science (Honours), Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor, Malaysia
hongyehao88@gmail.com

Abstract. Skin cancer is a highly prevalent disease, and its early diagnosis relies on the recognition of skin lesion images. However, the acquisition cost of labeled medical image samples is extremely high, which makes it easy for model training to suffer from data imbalance. To address this issue, this paper utilizes the HAM10000 dataset to construct and train a lightweight skin lesion image generation model on the basis of Wasserstein Generative Adversarial Network utilizing Gradient Penalty (WGAN-GP). The study first analyzes and preprocesses the HAM10000 dataset, including image standardization and label encoding. Subsequently, a Convolutional Neural Network (CNN) model and a multi-layer perceptron generator are constructed. By increasing the latent variable dimension to 512, setting different specific rates for the generator and discriminator, and introducing gradient penalty and multi-step discriminator update strategies, the training stability and balance are improved. Through adversarial training, high-quality skin lesion images are generated, and the quality of the generated skin lesion images is verified using the Fréchet Inception Distance (FID) score. Experimental results show that the generation model can synthesize reasonably distributed skin lesion images after 5000 steps of training, providing effective support for sample expansion of subsequent classification models. The research in this paper can provide a new direction for the rapid generation of images in the medical imaging field and lay a foundation for the improvement of clinical diagnostic assistance systems.

Keywords: Skin Lesion Image, WGAN-GP, HAM10000 Dataset, Medical Image Synthesis, Data Augmentation

1 Introduction

Skin cancer has become one of the fastest-growing types of cancer worldwide, and early diagnosis is critical for improving prognosis. Nevertheless, traditional deep learning based diagnostic models in medical imaging face two persistent challenges: (1) limited availability and high acquisition cost of expert-labeled medical data; and (2) severe class imbalance between different disease types. Such constraints significantly hinder model generalization and clinical applicability.

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_66

Goodfellow et al. proposed the Generative Adversarial Network (GAN) in 2014, which method has provided a powerful approach for learning complex data distributions and synthesizing realistic images [1]. Subsequently, Arjovsky et al. proposed the Wasserstein GAN (WGAN), which rephrased the goal of GAN using the earth mover's distance, thereby improving the convergence stability [2]. What's more, a variant of Wasserstein GAN utilizing Gradient Penalty (WGAN-GP) was further proposed by Gulrajani et al. [3]. It involved constrained gradient regularization during training to make the discriminator satisfy Lipschitz continuity, which effectively resolved the training instability and mode collapse problem in traditional GANs, hence greatly promoting the stability and quality of generated images. These advantages have made the GAN-based framework particularly prominent in biomedical imaging applications, such as data augmentation, lesion synthesis, and anomaly detection [4, 5].

In dermatological imaging, the HAM10000 dataset [6] is a large-scale, expert-annotated library of dermoscopy images. These images come from different populations and are collected and stored in various ways. This presents an excellent ground for developing and testing generative frameworks. This study proposed a lightweight model named WGAN-GP, which is specifically designed for the characteristics of medical image data. The model aims to generate various highly realistic images of skin lesions, in order to improve the performance of downstream classification.

The main contributions of this paper are as follows:

1. Describe a light and efficient architecture of WGAN-GP optimized for medical image generation, achieving stable convergence and high visual quality.
2. A structured pre-processing pipeline of HAM10000 is necessary for a well-balanced dataset and normalized inputs for adversarial training.
3. Integrate gradient penalty and multi-step discriminator updates to enforce Lipschitz continuity and maintain generator–discriminator equilibrium.
4. Perform quantitative and qualitative evaluations using Fréchet Inception Distance (FID) [7] to show improved structural realism and distribution consistency in generated samples.

2 Dataset and Methods

2.1 Dataset

The experimental data used in this study are Skin Cancer MNIST: HAM10000 (Human Against Machine with 10,000 training images) [6]. This dataset is one of the most representative and widely used public medical image libraries in the current research on skin lesion image classification. It is annotated by dermatologists and covers seven different types of skin lesions: actinic keratosis and intraepithelial carcinoma, also known as Bowen's disease (AKIEC), basal cell carcinoma (BCC), skin fibroma (DF), melanocytic nevus (NV), melanoma (MEL), benign keratotic lesions (including solar lentigo, seborrheic keratosis and flat lentiginous keratosis, BKL), and vascular lesions (such as hemangioma, vascular keratoma, pyogenic granuloma and hemorrhage, etc.).

Each image is an RGB color image with an original resolution of 600×450 pixels. The total data volume is approximately 10,015 images, and the file size is approximately 5.2 GB. The data comes from the International Skin Imaging Database (ISIC), where over 50 percent of the lesions are confirmed by histopathology (histo), and the diagnosis of the remaining cases is based on either subsequent examinations (follow_up), expert consensus (consensus), or confirmation through in vivo confocal microscopy (confocal). It has high clinical reliability and diversity.

This study automates the download and decompression of the dataset using the Google Colab environment, and realizes authorization and high-speed download through the Kaggle API. And kaggle.json file is used for identity verification and to achieve high-speed access. What's more, for the loading and preprocessing of the HAM10000 skin lesion image dataset, in order to achieve efficient training and avoid repeated data reading, a global caching mechanism is set up to load the processed image data directly during multiple calls. Subsequently, the two folders contained in the downloaded dataset, HAM10000_images_part_1 and HAM10000_images_part_2, are merged into a unified directory, named as /content/ham10000/all_images in this study. If the directory is not empty, the duplicate copy operation will be automatically skipped to save execution time. Then, the official HAM10000_metadata.csv file is used to establish a disease type mapping dictionary, converting the short labels, such as "nv", "mel", "bcc" and so on to the corresponding complete medical names, t thereby improving the interpretability of the samples and the consistency of subsequent classification. Finally, all image files were uniformly converted to the ".jpg" format and resized to 64×64 pixels to balance texture feature preservation and computational efficiency.

During the process of loading the image data, all samples were read as RGB arrays with three channels, and the GAN training specifications required normalization, scaling the pixel values within the range of [0,255] to [-1,1]. Such a transformation will improve not only the numerical stability in generator training but also avoid such phenomena as gradient vanishing or explosion. Then, the train_test_split() was used to perform stratified sampling based on disease categories (dx), dividing the dataset into a training and validation set in an 8 to 2 ratio and balancing the sample distribution across all categories. With the goal of improving the model's generalization, the ImageDataGenerator was employed to carry out the data augmentation of random rotation, including 20°, horizontal and vertical translation, including a range of 0.2 pixels, and horizontal flipping, among others; the validation set only underwent the normalization of pixels, within the range of [0,1]. This finished the processing, caching loading, and data division of 8012 images, providing high-quality input for subsequent GAN generation training and CNN classification validation. This series of pre-processing steps ensured that the GAN model could achieve reasonable feature learning within a short period and laid the data foundation for subsequent generation and discrimination training.

2.2 Methods

Convolutional Neural Networks. Before the formal training of the generative model, a lightweight CNN classifier for multi-class diagnostic recognition in HAM10000 skin

lesion images was developed. The Figure 1 presents the pseudo-code of the model architecture. The architecture of this model consists of two groups of convolutional layers and pooling layers, followed by one fully connected classification layer. It accepts input images of size 64×64×3. The architecture is composed of two convolutional layers. The initial layer employs 32 filters with a 3x3 kernel, and the subsequent layer employs 64 filters, also with a 3x3 kernel. A pooling operation with a 2x2 window performs subsampling after each convolution to reduce feature map dimensions. After convolution and pooling, the resulting feature maps are flattened via a Flatten layer and linked with the 128-dimensional which fully connected layer with ReLU activation. To prevent overfitting, a Dropout rate of 0.3 is used at this stage. The output layer uses the 7-dimensional softmax activation function because there are seven categories of skin lesions in the HAM10000 dataset.

```
Build CNN model:
        Input: (64, 64, 3)
        Conv(32, ReLU) → MaxPool
        Conv(64, ReLU) → MaxPool
        Flatten → Dense(128, ReLU) → Dropout(0.3)
        Dense(7, Softmax)

Compile with Adam optimizer, categorical crossentropy loss
Train CNN for 3 epochs
Plot training and validation accuracy curves
```

Fig. 1. CNN Code (Picture credit: Original)

This model was trained using the Adam optimization algorithm, with the default learning rate set at 0.001, while the loss function and metric to optimize during training are categorical cross-entropy and classification accuracy, respectively. The model was trained during 3 epochs, and each epoch consists of approximately 251 batches with the size of 32.

WGAN-GP Skin Lesion Image Generation Model. Employing the Wasserstein GAN with Gradient Penalty, this study conducts the skin lesion image generation experiments in the second stage.

Unlike traditional GANs that use log-likelihood loss, WGAN-GP measures differences between the real distributions and the generated distributions using the Earth Mover (Wasserstein) distance, which theoretically avoids problems such as mode collapse and gradient vanishing. Its core objective function is defined as:

$$L = \mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] - \mathbb{E}_{x\sim P_r}[D(x)] + \lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2] \quad (1)$$

In this formulation, let P_r indicate the distribution of the real samples and let P_g be the distribution of the generated samples, while the last term represents the gradient penalty. The gradient penalty acts to enforce Lipschitz continuity in the discriminator, a critical factor for stabilizing training.

As illustrated in Figure 2, the Generator accepts a 512-dimensional random noise vector as input. This vector is first projected into feature maps through a fully connected layer. Subsequently, these feature maps are upsampled via transposed convolution (Conv2DTranspose) layers to produce a synthetic skin lesion image with dimensions of 64×64×3. To ensure stable gradient propagation, LeakyReLU activation and BatchNormalization are applied throughout the hidden layers. Finally, a tanh activation function in the output layer scales the generated images to a normalized range of [-1,1].

```
Function build_generator(latent_dim):
        Input: latent vector z ∈ ℝ^(latent_dim)
        Dense(8×8×256)
        → LeakyReLU(0.2)
        → Reshape(8, 8, 256)
        ConvTranspose(128, kernel=4, stride=2, padding="same")
        → BatchNorm(momentum=0.9)
        → LeakyReLU(0.2)
        ConvTranspose(64, kernel=4, stride=2, padding="same")
        → BatchNorm(momentum=0.9)
        → LeakyReLU(0.2)
        ConvTranspose(3, kernel=4, stride=2, padding="same", activation="tanh")
        Output: synthetic image ∈ ℝ^(64×64×3)
End Function
```

Fig. 2. Generator Code (Picture credit: Original)

The Discriminator is a shallow convolutional neural network shown as Figure 3, that extracts high-dimensional features through convolutional and LeakyReLU layers, outputting a Wasserstein distance score for the input image rather than a probability value. The network structure is designed to be lightweight to enhance training stability and computational efficiency.

```
Function build_discriminator(input_shape=(64, 64, 3)):
        Input: Image x ∈ ℝ^(input_shape)
        GaussianNoise(0.05)
        Conv2D(64, kernel=3, stride=2, padding="same")
                → LeakyReLU(0.3)
                → Dropout(0.3)
        Conv2D(128, kernel=3, stride=2, padding="same")
                → BatchNormalization
                → LeakyReLU(0.3)
                → Dropout(0.3)
        Conv2D(256, kernel=3, stride=2, padding="same")
                → BatchNormalization
                → LeakyReLU(0.3)
                → Dropout(0.4)
        Conv2D(512, kernel=3, stride=2, padding="same")
                → BatchNormalization
                → LeakyReLU(0.3)
                → Dropout(0.4)
        Flatten
        Dense(1)
        Output:Scalar score indicating real/fake
End Function
```

Fig. 3. Discriminator Code (Picture credit: Original)

During the training process, an alternating min-max game mechanism is adopted to optimize the Generator and Discriminator. Specifically, the discriminator is trained for D_steps=3 steps, and the generator for G_steps=1 step per training cycle.

Each training cycle comprises the following steps:

(1) Display the real images and the generated images, and then update the parameters of the discriminator to maximize the difference in scores between the real samples and the fake samples.

(2) Freeze the discriminator parameters and minimize the generator loss, making the scores of generated samples as close as possible to those of real samples.

(3) Incorporate the gradient penalty term in each iteration to ensure the Lipschitz constraint.

(4) The model utilizes the Adam optimizer, and the learning rate for the discriminator is set to 0.00003, while that for the generator is set to 0.0001. β_1 is 0.5 and β_2 is 0.9, and the gradient penalty coefficient λ is set to 15.

(5) The total number of training steps is 5000, with generated samples saved and the FID calculated every 500 steps for quality assessment.

3 Experiments and Results

3.1 Experimental Setup

The experiments of this study were conducted on the Google Colab platform, with the running environment being Python 3.10 and TensorFlow 2.19. The main dependent libraries used include Keras, NumPy, Matplotlib, Pandas, Scikit-learn, and the Kaggle API, which is used for online authorization and high-speed download of the dataset.

During the data preprocessing stage, the system loaded a total of 10,015 HAM10000 skin lesion images and implemented automatic normalization, data augmentation (rotation, translation, horizontal flip, etc.) and data division operations through ImageDataGenerator. All images were adjusted to a size of 64×64×3, and normalize the pixel values to the range of [-1, 1] to meet the input requirements of the WGAN-GP model. The training data was divided into a training set and a validation set in 8:2 ratio to ensure the consistency of the distribution of each category.

During the model training stage, each batch contained 32 images. An iterative optimization strategy was adopted, updating the Generator once every 3 times the Discriminator had been trained in an alternating game style, in order to ensure stable and balanced parameter updates with feature confrontation.

In order to verify the model's feasibility, 100 steps of pre-training were firstly conducted to see the convergence trend of the loss and the stability of the generated samples. Later, the training was extended to a full 5,000 steps to ensure that this model can fully learn the features related to the texture, shape and boundary of skin lesions.

During the entire training process, these samples were generated automatically after every 500 steps throughout the training process. Visualizations and FID calculations were performed after every 500 steps to quantify quality and distribution similarity of the generated samples. This achieved a twofold benefit in enhancing both the stability of training and the visual quality without sacrificing too much computational efficiency.

3.2 Results and Analysis

WGAN-GP Skin Lesion Image Generation Model. After normalization and division of the data, the system loaded roughly 10,000 images of skin lesions and sorted them into 7 categories. First, the CNN model was used for multi-class classification training. The training results showed that the model had a stable convergence trend, with the training accuracy rising from 65.0% to 67.1%, and the loss value dropping from 1.13 to 0.92; the accuracy of the validation set increased from 66.9% to 67.0%, with the validation loss decreasing from 0.955 to 0.88. Thus, it can be seen that the model can achieve good convergence performance in a short number of training rounds. Moreover, from the results, the gap between the training accuracy and validation accuracy is small, indicating that the model has not significantly overfitted and has good generalization ability. Additionally, from the performance perspective, this lightweight CNN can achieve a multi-class classification accuracy of approximately 70% in only 3 training epochs, proving its effectiveness and feasibility in extracting skin lesion texture features. The lower loss value reflects the stability of the model in distinguishing the

boundaries of different categories, providing a solid baseline for subsequent WGAN-GP feature generation and adversarial optimization.

Subsequently, a WGAN-GP model based on a convolutional structure was constructed. Through the adversarial game between Generator and Discriminator, skin lesion images were synthesized.

In the early training stage, Step 100, both the Discriminator and the Generator were in an un-converged state. The Wasserstein score output by the Discriminator was significantly different (D(real) equals to 0.9216, D(fake) equals to -12.4929), indicating a significant difference in the distribution between the generated samples and the real samples.

As the training progressed, the Generator gradually learned the texture patterns of the lesion structure, and the Generator loss (G_loss) rapidly decreased from 12.4576 to approximately 3.5, while the Discriminator loss (D_loss) increased from -11.9 to -1.5, indicating that the learning of both sides of the adversarial game tended towards a dynamic balance.

As the training progressed, D_loss gradually increased, while G_loss gradually decreased, and the changes in the two showed a typical adversarial trend:

Table 1. Evolution of Generator and Discriminator Losses During Training.

Training Step	D_loss	G_loss	D(real)	D(fake)	FID
500	-0.0396	1.5883	-1.3659	-1.4351	251.145
1000	-0.5697	-1.7180	2.3643	1.7128	231.130
1500	0.3669	-0.4296	0.1873	0.4840	216.315
2000	-0.5221	1.3948	-1.2072	-1.7641	211.459
2500	-0.1345	0.3643	-0.2810	-0.4343	187.926
3000	-0.4634	0.6822	0.0587	-0.4711	185.298
3500	-0.1523	0.2986	-0.6265	-0.8103	182.828
4000	-0.2830	1.3376	-0.6898	-1.0036	167.079
4500	0.0215	1.9130	-1.6596	-1.7508	176.483
5000	-0.2640	1.0481	-0.6733	-0.9628	168.606

From Table 1, at the initial stage of training (Step 500), the FID value was as high as 251.145, which can be observed, indicating a significant difference between the generated images and the real distribution. After normalization and division of the data, the system loaded roughly 10,000 images of skin lesions and sorted them into 7 categories. First, the CNN model was used for multi-class classification training. The training results showed that the model had a stable convergence trend, with the training accuracy rising from 65.0% to 67.1%, and the loss value dropping from 1.13 to 0.92; the accuracy of the validation set increased from 66.9% to 67.0%, with the validation loss decreasing from 0.955 to 0.88. As training proceeded, the generator further learned the texture and color distribution of skin lesion images, and the FID value continued to drop. And at the 4000th step, it drops to 167.079, which means a large improvement in the perceptual quality of the generated samples of the model. Meanwhile, D_loss and G_loss tended to be stable in the middle stage, from Step 2000 to 3500, and the gap

between D(real) and D(fake) numerically narrowed, which reflected that the discriminator and generator came into a dynamic balance state. The overall trend shows that the model successfully achieved adversarial balance. It can be observed that the training process is undisturbed and stable, with clearly reflected image details, whose texture levels are close to the real structure of skin lesions. This confirms that the gradient penalty mechanism of WGAN-GP can effectively improve the generation quality and enhance the training stability. Overall, the experiment indicates that this improved WGAN-GP model can stably train on the HAM10000 dataset and generate high-quality skin lesion images, providing a feasible solution for medical image data augmentation and imbalance sample supplementation.

Generated Sample Visual Assessment. The WGAN-GP model exhibited stable and representative convergence trends for the five metrics of Discriminator loss, Generator loss, D(real), D(fake), and FID throughout the training process, which could be analyzed in the following distinct stages:

First is the Initial Stage (1–500 Steps): The output images are shown in Figure 4. At this time, the model was in the parameter initialization stage; the generated image is mainly composed of noise and block-like textures. At this point, D_loss was approximately equal to -0.04, G_loss was approximately equal to 1.59, D(real) was approximately equal to -1.36, D(fake) was approximately equal to -1.43, and the FID value was equal to 251.145. These results indicate that the generator has not yet mastered an effective image distribution, while the discriminator is capable of easily distinguishing between real samples and generated samples.

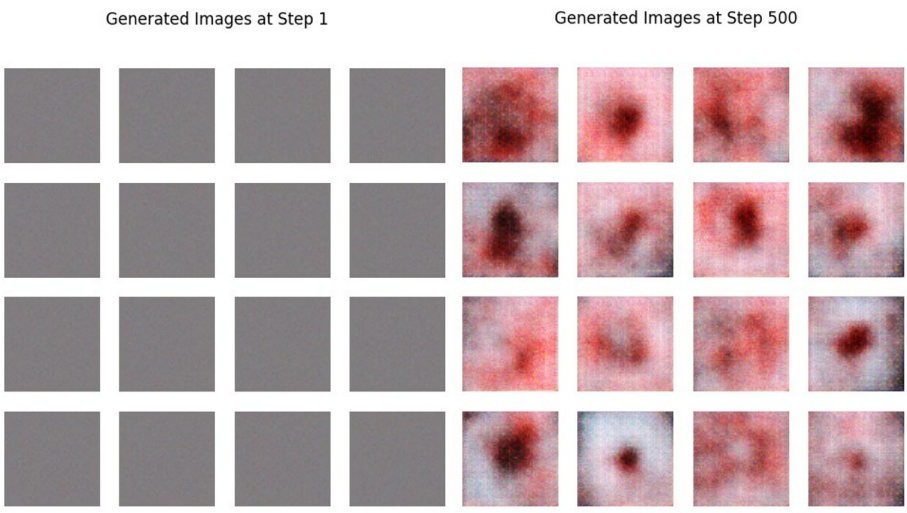

Fig. 4. Generated Images in 1–500 Steps (Picture credit: Original)

Second is the Intermediate Stage (501–3000 Steps): The output images between step 501 and step 3000 are shown in Figure 5, Figure 6 and Figure 7 respectively. The generator began to capture the global structure and color gradations of skin images.

Within the range of Steps 501–3000, D_loss varied from -0.56 to 0.36, while G_loss varied from -1.71 to 1.3, reflecting that the generator and discriminator were in dynamic equilibrium during the adversarial training process.

The difference between D(real) and D(fake) progressively decreased, from approximately 4.07 to 0.53, which indicates that the discriminator is no longer able to easily distinguish the differences between real samples and generated samples.

Meanwhile, the FID score decreased from 231.13 to 185.30, showing that with increases in the generated samples, they were getting closer to the visual distribution of real skin images.

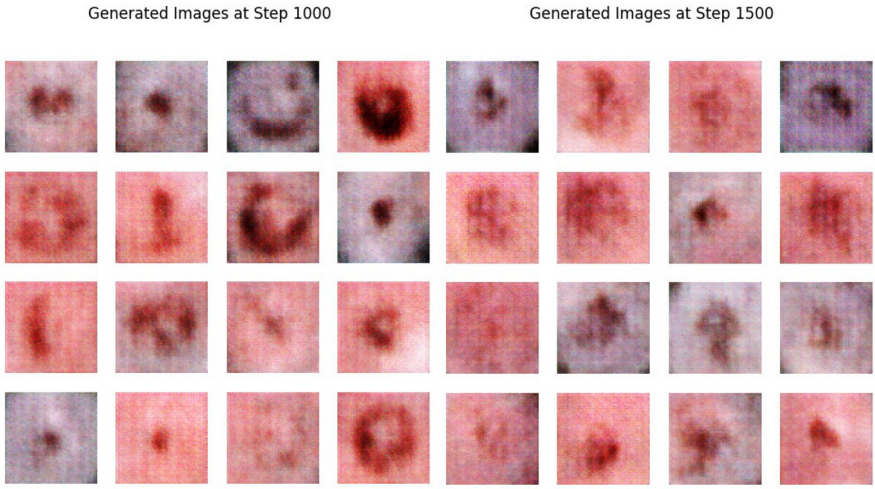

Fig. 5. Generated Images in 501–1500 Steps (Picture credit: Original)

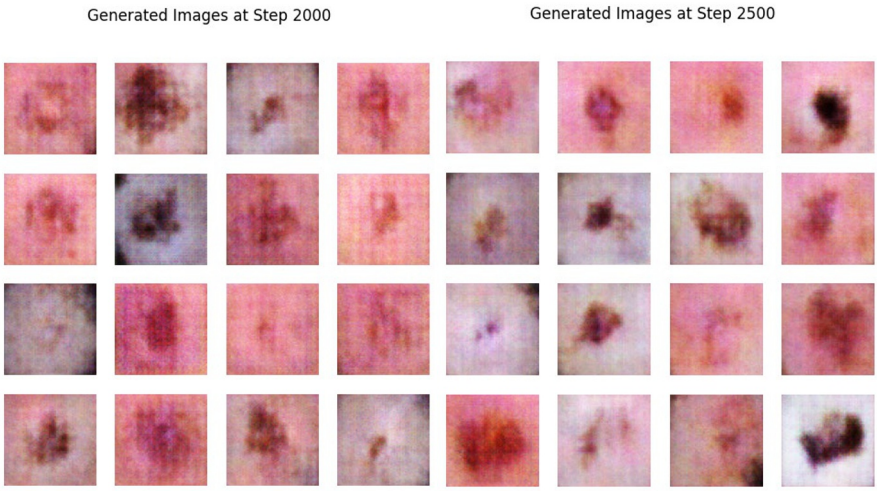

Fig. 6. Generated Images in 1501–2500 Steps (Picture credit: Original)

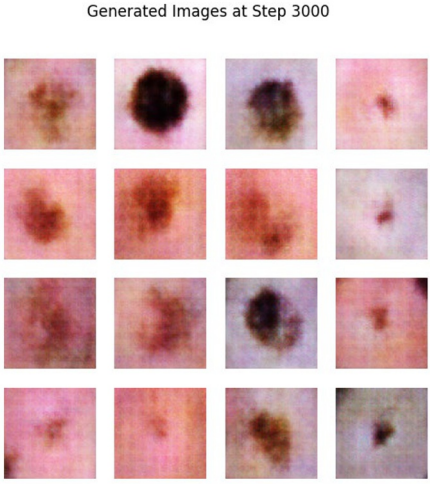

Fig. 7. Generated Images in 2501–3000 Steps (Picture credit: Original)

Third is the Final Stage (3001–5000 Steps): The output images of step 3500, step 4000, step 4500 and step 5000 are shown in Figure 8 and Figure 9. The model entered a stable convergence phase, where only details and textures were further refined.

D_loss stabilized in the range of -0.28 to 0.02, while G_loss has remained between 0.3 and 1.9.

In both D(real) and D(fake), the values converged to near -1.0, indicating a further reduction in the output gap between real samples and generated samples, suggesting that the model reached an adversarial equilibrium.

The FID value reached a minimum of 167.079 at Step 4000, with small fluctuations afterward, reaching stability at 168.606 by Step 5000.

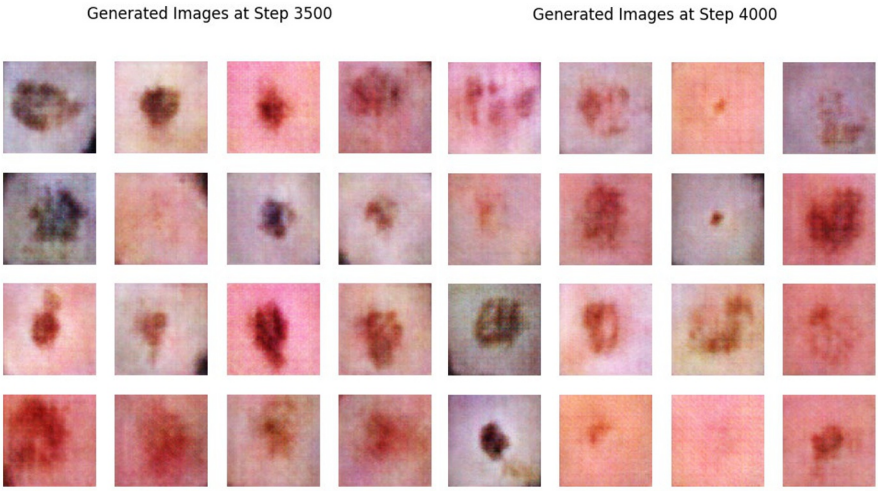

Fig. 8. Generated Images in 3001–4000 Steps (Picture credit: Original)

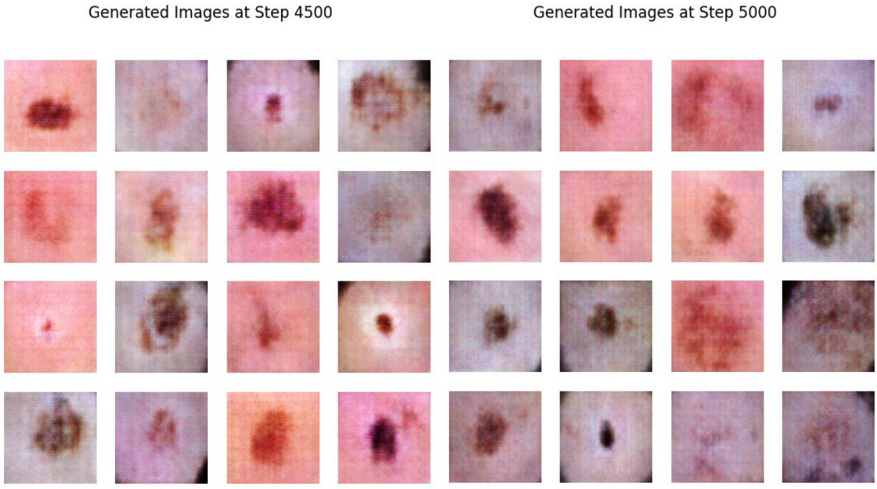

Fig. 9. Generated Images in 4001–5000 Steps (Picture credit: Original)

These observations confirm that the generative quality of the model stabilized in the final phase and enabled the output of skin lesion images with a high degree of textural consistency and structural integrity.

In a nutshell, the WGAN-GP model showed stable convergence trends in 5000 steps of training. The loss variations of the discriminator and generator were mutually

constrained while the continuous decline in the FID score mirrored a steady improvement in terms of the quality of the generated samples. Without the support of a large amount of annotated data, the model succeeded in generating realistic skin lesion images, hence providing high-quality synthetic data support for the subsequent research of medical image enhancement and classification.

Discussion. On the basis of the experimental evidence above, it is clearly evident that the lightweight WGAN-GP model created in this paper has achieved stable training convergence, and has generated high-quality skin lesion images based on the HAM10000 dataset. As compared to the traditional GAN, it was found that WGAN-GP had pronounced advantages regarding training stability, generation quality, and distribution consistency. Experimental results show a steady reduction in the FID metric, with the generated images demonstrating high fidelity to real skin lesions in terms of texture detail, boundary sharpness, and overall distribution consistency. This improvement suggests that the model's outputs can effectively augment clinical datasets and enhance the recognition of rare lesions [8][9].

In 5,000 steps, oscillatory but convergent D_loss and G_loss indicate a balance that is typical of successful adversarial learning [7]; D(real) and D(fake) converge, which means that the texture patterns have been learned well. Additionally, the continuous decline in FID from 251 to 167, as illustrated in Figure 10, confirms a progressively improved alignment in the Inception feature space [4,10]. The inclusion of the gradient penalty effectively mitigated issues such as gradient explosion and mode collapse [3], while multi-step critic updates further contributed to training stability.

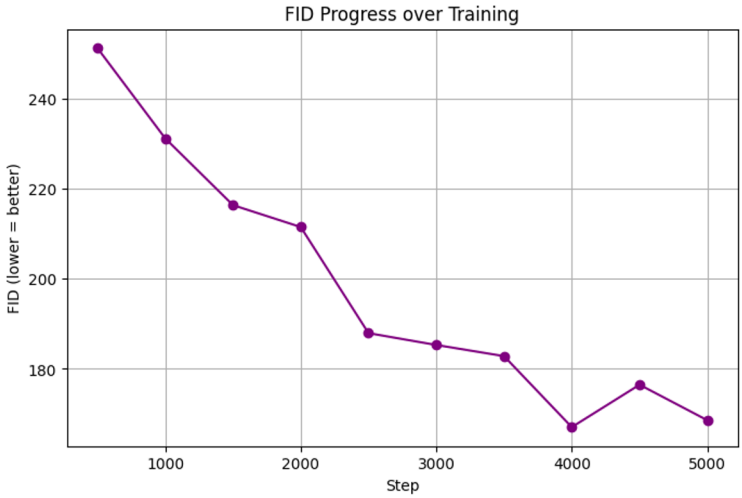

Fig. 10. FID Progress over Training (Picture credit: Original)

The generated images exhibit realistic visual characteristics and can serve as valuable additions to clinical datasets, improving the recognition and classification of

rare lesion types [5,7]. Future work could integrate advanced generative frameworks such as self-attention GANs (SAGAN) [11] or diffusion models [12] to further enhance the fidelity and diversity of synthetic skin lesion images, thereby extending the model's potential applications in medical data augmentation and diagnostic support.

4 Conclusions

This study develops and validates a lightweight WGAN-GP framework for medical image generation based on the HAM10000 skin lesion dataset, targeting the issues of data scarcity and class imbalance. Through systematic data preprocessing, the design of a lightweight convolutional network, and an optimized adversarial training strategy, the model achieved stable convergence and high-quality image synthesis within 5,000 training steps.

The experimental results highlight a consistent reduction in the FID metric, while the generated images maintain strong similarity to real skin lesions in terms of texture detail, boundary clarity, and distributional coherence. These outcomes demonstrate the model's capability to augment clinical datasets and improve the detection of rare lesions. Comparative analyses further confirm that WGAN-GP enhances both the perceptual quality and distributional realism of generated images while maintaining robust training stability, underscoring its suitability for medical image data augmentation.

Additionally, CNN-based classification experiments verified that the generated samples can serve as reliable training priors for subsequent diagnostic models. Overall, the proposed lightweight WGAN-GP framework provides a practical, efficient, and stable solution for medical image augmentation and diagnostic support in dermatological imaging.

Reference

1. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y.: Generative Adversarial Nets. Advances in Neural Information Processing Systems 27 (2014)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein Generative Adversarial Networks. In: International Conference on Machine Learning, pp. 214–223. PMLR (2017)
3. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved Training of Wasserstein GANs. Advances in Neural Information Processing Systems 30 (2017)
4. Yi, X., Walia, E., Babyn, P.: Generative Adversarial Network in Medical Imaging: A Review. Medical Image Analysis 58, 101552 (2019)
5. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-Based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification. Neurocomputing 321, 321–331 (2018)
6. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. Scientific Data 5(1), 1–9 (2018)

7. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Advances in Neural Information Processing Systems 30 (2017)
8. Brock, J., Donahue, C., Simonyan, K.: Large Scale GAN Training for High Fidelity Natural Image Synthesis. In: Proceedings of ICLR (2019)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
10. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
11. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-Attention Generative Adversarial Networks. In: International Conference on Machine Learning, pp. 7354–7363. PMLR (2019)
12. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Yang, M.H.: Diffusion Models: A Comprehensive Survey of Methods and Applications. ACM Computing Surveys 56(4), 1–39 (2023).