Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 16 Issue 1 Year 2026 Pages 1-11
DOI: http://dx.doi.org/10.31940/matrix.v16i1.1-11
URL: http://ojs2.pnb.ac.id/index.php/matrix

Application of conditional lightweight GAN for retinal fundus image synthesis based on diabetic retinopathy severity levels on the IDRiD dataset

Acep Taufik Hidayat [1,3]*, I Made Dendi Maysanjaya [2], I Made Gede Sunarya [1], Made Windu Antara Kesiman [1]

[1] *Graduate Program in Computer Science, Universitas Pendidikan Ganesha, Indonesia*
[2] *Information Systems Study Program, Universitas Pendidikan Ganesha, Indonesia*
[3] *Information Systems Study Program, STMIK Bandung Bali, Indonesia*

*Corresponding Author: *taufiqhidayat50737@gmail.com*

Abstract: Diabetic Retinopathy (DR) is a leading cause of preventable blindness, yet the development of automated diagnostic models using Deep Learning is often hindered by the availability of imbalanced medical datasets. This study aims to address this issue by implementing a Conditional Lightweight Generative Adversarial Network (c-LGAN) architecture to synthesize realistic fundus retinal images corresponding to five DR severity levels from the IDRiD dataset. The c-LGAN model was trained on a balanced dataset, and its performance was quantitatively evaluated using Frechet Inception Distance (FID) and Inception Score (IS) metrics. The results demonstrate that the proposed model is capable of generating high-quality images, evidenced by achieving a best FID score of 121.24 at epoch 100. However, further observation identified significant stability challenges in long-term training, marked by a performance collapse after the model reached its optimal point. This phenomenon was attributed to an overpowering discriminator. This study concludes that c-LGAN is a promising approach for data augmentation but emphasizes the critical importance of periodic metric monitoring and model checkpointing strategies to capture peak performance and overcome training stability issues.

Keywords: Conditional GAN, deep learning, diabetic retinopathy, image synthesis.

History Article: Submitted 16 December 2025 | Revised 19 February 2026 | Accepted 3 March 2026

How to Cite: A. T. Hidayat, I. M. D. Maysanjaya, I. M. G. Sunarya, and M. W. A. Kesiman, "Application of conditional lightweight GAN for retinal fundus image synthesis based on diabetic retinopathy severity levels on the IDRID dataset," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 16, no. 1, pp. 1–11, 2026, doi: 10.31940/matrix.v16i1.1-11.

Introduction

Diabetic Retinopathy (DR) is a microvascular complication of diabetes mellitus and represents a leading cause of visual impairment and blindness among the working-age population worldwide [1], [2]. This disease arises from damage to the fine blood vessels in the retina triggered by chronic hyperglycemic conditions. If not treated at an early stage, such damage can progress to vascular occlusion, fluid leakage, and the growth of abnormal new blood vessels (neovascularization), which may ultimately result in permanent blindness [3], [4]. With the global number of individuals with diabetes projected to continue increasing [5], the development of accurate and efficient early detection methods has become imperative to prevent vision loss [6].

Retinal fundus imaging is considered the gold standard for DR diagnosis due to its ability to provide non-invasive visualization of the internal structures of the eye [7]. To guide diagnosis, the International Clinical Diabetic Retinopathy (ICDR) classification system categorizes DR into five levels of severity: Normal (no abnormalities), Mild Non-Proliferative Diabetic Retinopathy (NPDR), Moderate NPDR, Severe NPDR, and Proliferative Diabetic Retinopathy (PDR), which represents the most critical stage [8].

Along with technological advances, Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), has demonstrated great potential as a fast and accurate method for automated diagnosis [9], [10]. However, the performance of DL models is highly dependent on the

availability of large and diverse training datasets. In many cases, publicly available medical datasets such as the Indian Diabetic Retinopathy Image Dataset (IDRiD) are often insufficient or imbalanced, where the number of images representing severe disease classes is significantly smaller than those of normal conditions [11]. Standard data augmentation techniques, such as rotation and flipping, often fail to address this issue effectively as they merely alter the geometry of existing samples without introducing new pathological feature variations. Consequently, diagnostic models tend to become biased toward the majority class (Normal), resulting in poor sensitivity for critical minority classes (Severe and PDR), which poses a significant risk of missed diagnoses in clinical settings.

To address this challenge, data augmentation through image synthesis has become critically important. Recent advances in the field of Generative Adversarial Networks (GANs), first introduced by Goodfellow et al. [12], have made them a highly promising tool for synthesizing realistic data. GANs consist of two competing neural networks: a generator, which learns to create new data, and a discriminator, which learns to distinguish between real and synthetic data. Previous studies by Zhou et al. [13] introduced DR-GAN to synthesize fine-grained lesions, demonstrating that synthetic data can significantly improve classification accuracy. However, their approach relies on complex architectures that demand substantial computational resources, limiting their applicability in resource-constrained environments often found in developing healthcare systems.

This paper proposes the application of a Conditional Lightweight GAN (c-LGAN) to synthesize retinal fundus images from the IDRiD dataset according to the five specific severity levels. Unlike heavy conventional GANs, the proposed "lightweight" design aims to achieve computational efficiency, enabling model training on hardware with limited specifications [14]. While the synthesis resolution is set to 128 × 128 pixels—a trade-off that sacrifices some fine-grained high-frequency details—this resolution is sufficient to capture global structural pathology and color distribution required to balance class distribution and mitigate model bias. The quality of the synthesized images will be quantitatively evaluated using the Frechet Inception Distance (FID) and Inception Score (IS) metrics. It is expected that the generated images can be utilized for data augmentation to train more robust AI-based classification models and can also serve as supportive diagnostic teaching tools for medical professionals in the future.

Methodology

Experimental Design

The experimental workflow is illustrated in Figure 1. The experiment consists of five main stages: Dataset Collection, Pre-processing, GAN Architecture Selection, Training Process, and Performance Evaluation. This systematic approach ensures that the impact of data augmentation on class balance and image quality can be rigorously assessed.

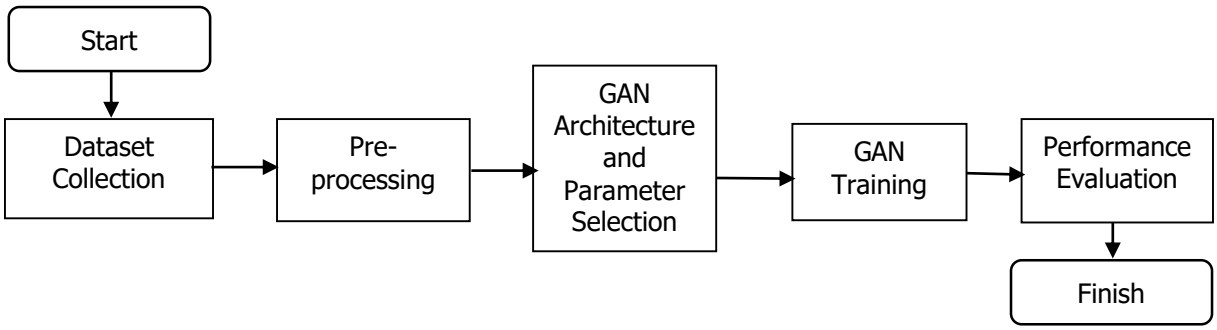


Figure 1. Experimental flow diagram

Dataset Collection

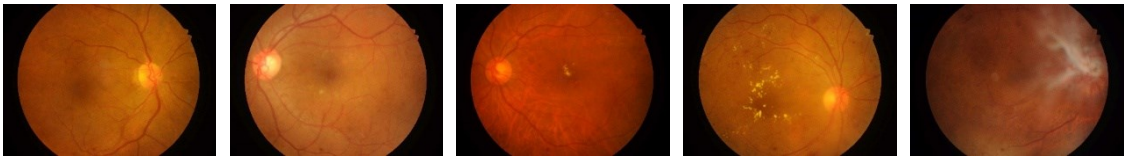
The dataset used in this study is the Indian Diabetic Retinopathy Image Dataset (IDRiD), developed by Porwal et al. [15]. This dataset was specifically designed to support research on the detection and classification of Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME).

The IDRiD dataset is publicly available via the Kaggle platform and provides high-resolution retinal fundus images.

The training subset contains 413 images, each annotated with a DR severity level according to the International Clinical Diabetic Retinopathy (ICDR) scale. As shown in Table 1, the original distribution of the dataset is highly imbalanced, with a significant scarcity of samples in the severe disease classes (Severe NPDR and PDR). This imbalance poses a major challenge for deep learning models, which tend to bias towards the majority class.

Table 1. Distribution of IDRiD dataset before and after balancing

DR Severity Level	Original Count (Images)	Percentage (%)	Target Count After Synthesis
Normal	134	32.4%	500
Mild NPDR	20	4.8%	500
Moderate NPDR	136	32.9%	500
Severe NPDR	74	17.9%	500
Proliferative DR (PDR)	49	11.9%	500
Total	413	100%	2,500



a) Normal *b) Mild* *c) Moderate* *d) Severe* *e) PDR*

Figure 2. Sample IDRiD dataset images based on DR severity levels

Pre-processing

The pre-processing stage was critical to prepare raw data for effective GAN training. The primary objectives were to standardize dimensions, enhance variability, and correct the class imbalance shown in Table 1

Standardization

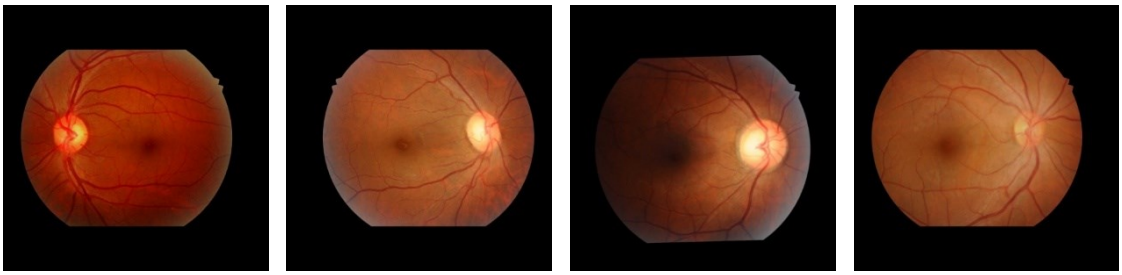
Raw images were first padded to square aspect ratios to prevent distortion and then resized. While the original images are high-resolution, they were resized to 128 × 128 pixels for this study. This resolution was selected as a strategic trade-off to drastically reduce computational load and memory usage, enabling the training of the generative model on standard hardware while still preserving global retinal structures (optic disc, major vessels) required for severity classification.

Augmentation

To enrich the diversity of the training set before synthesis, standard augmentation techniques were applied using the Albumentations library. These included horizontal flipping, brightness/contrast adjustments, color jittering, rotation (+- 10^{o}), and Gaussian blur.

Balancing Strategy

As detailed in Table 1, the final dataset was balanced to contain exactly 500 images per class. This was achieved by combining real augmented images with synthetic images generated by the c-LGAN model, ensuring that the model learns from a uniform distribution of disease severity.



a) Color jitter *b) HFlip* *c) Shift scale rotate* *d) Brightness*

Figure 3. Examples of IDRiD dataset images after pre-processing

GAN Architecture and Parameter Selection

This study employs a Conditional Lightweight GAN (c-LGAN) [16] architecture, adapted from the Lightweight GAN (LGAN) approach proposed by Abd Aziz et al. [14] for retinal fundus image synthesis.

Justification for Lightweight Architecture

Unlike conventional GANs (e.g., DCGAN or StyleGAN2) which rely on deep networks with millions of parameters and require high-end GPU clusters, the c-LGAN is designed for computational efficiency. It utilizes a reduced number of feature maps in the convolutional layers and simpler residual blocks. This design choice makes the model deployable in resource-constrained environments (such as basic medical research facilities in developing countries) and significantly accelerates the training convergence without a massive sacrifice in the semantic quality of the generated medical images.

Generator Structure

The generator is designed to be lightweight yet effective. It takes two inputs, a noise vector z (latent space dimension = 128) and a class embedding label y.

Input Layer

The latent vector and class embedding are concatenated and passed through a dense layer.

Upsampling Blocks

The network consists of a series of transposed convolutional layers (ConvTranspose2d). Instead of deep residual stacks, each block performs efficient upsampling followed by Batch Normalization and ReLU activation.

Output Layer

The final layer uses a Tanh activation function to produce a 3-channel RGB image with values in the range [-1, 1] [17].

Discriminator Structure

The discriminator is a patch-based classifier that also conditions on the class label. It uses Spectral Normalization in its convolutional layers to stabilize training and prevent mode collapse—a common issue in medical image synthesis. The class label is projected and concatenated with the image feature map to ensure the discriminator evaluates both the realism of the image and its correspondence to the target disease severity [18].

Table 2. GAN model training parameters

Parameter	Value	Description
Image size	128 × 128 pixels	Input and output resolution to reduce computation
Latent space dimension	128	Noise vector dimension (generator input)
Optimizer	*Adam*	Used for Generator and Discriminator
Learning rate	0.0002	Learning rate for both networks
β1 (*Adam*)	0.5	First momentum parameter
β2 (*Adam*)	0.999	Second momentum parameter
Maximum epochs	1.000	Total training iterations
Batch size	32	Samples per training iteration
Early Stopping	Yes	Applied to prevent overfitting
Evaluation & checkpointing	Periodic	To monitor and save best model performance

Training parameters were carefully selected to ensure training stability and efficiency. The image size was set to 128 × 128 pixels to reduce computational load, while the latent space dimension was fixed at 128. The Adam optimization algorithm was employed with a learning rate of 0.0002 for both the generator and discriminator, and β1 and β2 values of 0.5 and 0.999, respectively. The model was trained for a maximum of 1,000 epochs with a batch size of 32. In addition, an early stopping mechanism was implemented to terminate training when no significant improvement was observed, thereby preventing overfitting. Periodic evaluation and model checkpointing were also conducted to monitor network performance and stability throughout the training process.

GAN Training

The training process utilized the Adam optimizer (α=0.0002, β1=0.5, β2=0.999) for both networks. The loss function employed was the Binary Cross-Entropy with Logits Loss.

Conditional Training

The model was trained to generate images specific to the five DR grades. This conditional mechanism ensures that the generator learns distinct pathological features (e.g., microaneurysms for Mild DR vs. neovascularization for PDR).

Stability Monitoring

Training was run for a maximum of 1,000 epochs. To address potential instability (e.g., overpowering discriminator), Early Stopping was implemented. If the FID score did not improve for 50 consecutive epochs, training was halted to prevent overfitting and degradation of image quality.

Performance Evaluation

To quantitatively assess the quality and diversity of the synthesized images, two standard metrics were used:

Frechet Inception Distance (FID) [19]

This metric measures the distance between the feature distributions of real and synthetic images. A lower FID indicates that the synthetic images are more similar to real biological data and possess better diversity.

Inception Score (IS) [20]

This metric evaluates the clarity and distinctness of the generated images. A higher IS indicates better image quality.

These metrics were calculated every 10 epochs. The "Best Model" checkpoint was saved not based on the final epoch, but based on the lowest FID score achieved during training, ensuring that the final evaluation uses the most realistic generator.

Results and Discussions

This section presents an in-depth analysis of the experimental results obtained from training the Conditional Lightweight GAN (c-LGAN) model. The discussion focuses on three main aspects: analysis of training progression and stability based on loss data, quantitative evaluation of synthetic image quality using the Frechet Inception Distance (FID) and Inception Score (IS) metrics, and qualitative analysis of the generated images.

Training Progression and Model Stability

The c-LGAN training process was completed over 1,000 epochs with a total computational time of approximately 39.5 hours (142,228 seconds). Analysis of the loss curves (see Figure 4) reveals challenging training dynamics and indicates stability issues.

Loss Dynamics

From the early stages of training, the Discriminator Loss (D_loss) decreased drastically and consistently remained at a very low level, often approaching zero (e.g., 0.0014 at epoch 2). In contrast, the Generator Loss (G_loss) remained high and exhibited significant volatility, fluctuating between 2.0 and peaks above 20.0 (e.g., 20.24 at epoch 921). The condition in which D_loss is extremely low while G_loss is high and unstable is a classic indication of an overpowering discriminator. In this scenario, the discriminator becomes too strong too quickly, providing vanishing gradients that make it difficult for the generator to learn effectively and converge.

Model Stability

This stability issue is further confirmed by the FID evaluation data. Although the model showed improvement up to epoch 100, a drastic performance collapse occurred thereafter. This phenomenon indicates that the model failed to maintain its best performance and entered an unstable training regime known as mode collapse or divergence.

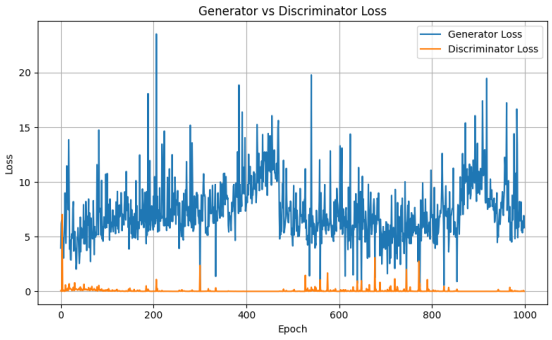


Figure 4. Loss progression curves for the Generator (G) and Discriminator (D) over 1,000 epochs

Synthetic Image Quality

The quality of the synthetic fundus images was quantitatively evaluated every 10 epochs. Key evaluation points highlighting performance trends are summarized in Table 3.

Table 3. Quantitative evaluation results at key epochs

Epoch	FID Score (↓ better)	IS Score (↑ better)	Remarks
10	349.45	1.28 ± 0.04	Early training
40	173.59	1.91 ± 0.18	Significant improvement
70	153.08	1.49 ± 0.19	Approaching peak
90	136.34	1.59 ± 0.18	Near peak
100	121.24	1.56 ± 0.16	Best performance
110	445.85	1.49 ± 0.13	Performance collapse
200	288.68	1.40 ± 0.06	Failed to recover
500	170.16	1.86 ± 0.15	Improved but not recovered
740	156.55	1.40 ± 0.08	Secondary peak (best after collapse)
1000	171.42	1.75 ± 0.27	Final model

The model achieved peak performance at epoch 100, with a minimum FID score of 121.24.

Significance of FID

A lower FID score indicates that the distance between the feature distribution of real and synthetic images is small. This implies that the synthetic images at epoch 100 successfully capture the statistical diversity and global structural features of the real retinal dataset, satisfying the requirement for data augmentation diversity.

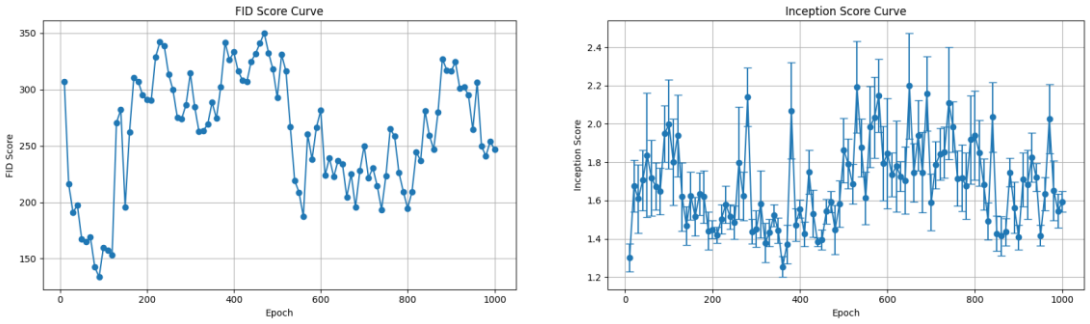
Performance Collapse

However, after this point, a sharp increase in the FID score to 445.85 at epoch 110 signaled a severe quality collapse. Despite partial recovery attempts in later epochs, the model never regained the quality achieved at epoch 100.

Meanwhile, the Inception Score (IS) remained relatively low overall, with the highest value reaching only 1.92 ± 0.20 (at epoch 80).

Interpretation

It is important to note that the IS metric relies on the InceptionV3 network pre-trained on ImageNet (which contains distinct objects like dogs, cars, etc.). Retinal fundus images are visually homogeneous—they all look like orange discs with blood vessels—lacking the distinct object variability found in ImageNet. Consequently, the Inception network does not classify them into "distinct" classes with high confidence, naturally resulting in lower IS values compared to natural image datasets. Therefore, in this specific medical context, the FID score is a more reliable indicator of quality than IS, as FID compares the synthetic distribution directly against the real retinal data distribution.



a) FID curve b) IS curve

Figure 5. Evolution of FID and Inception Score (IS) values during training

Qualitative Analysis

Visual inspection of image samples saved at various training intervals is consistent with the quantitative findings:

Best Images (Epoch 100)

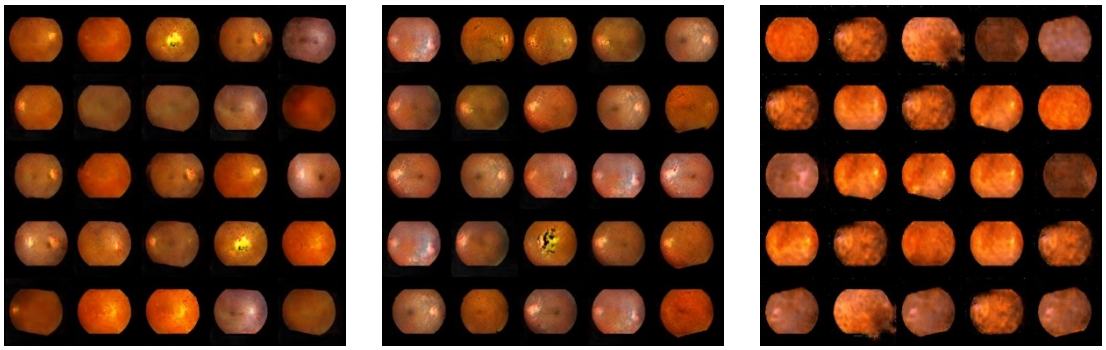
Images generated around epoch 100 exhibit the highest visual quality. Major anatomical structures, such as the optic disc and blood vessels, appear coherent, with natural fundus coloration. The model successfully replicates visual features in accordance with the provided class labels.

Post-Collapse Images (Epoch > 100)

Images generated after epoch 100 show a significant degradation in quality. Numerous visual artifacts, unnatural patterns, and disorganized anatomical structures emerge, corresponding to the deteriorating FID scores.

Final Images (Epoch 1000)

Although better than the complete collapse condition, the visual quality of images at the final epoch does not match that of the images generated at epoch 100.



(a) Samples from epoch 100 (best FID) (b) Samples from epoch 110 (post-collapse) (c) Samples from epoch 1000 (final)

Figure 6. Comparison of synthetic image quality

Discussion

The findings of this study provide two key insights. First, the c-LGAN architecture is capable of generating reasonably realistic retinal fundus images, as evidenced by the minimum FID score of 121.24. This result confirms the potential of GANs as data augmentation tools for imbalanced medical datasets.

Second, and more critically, this study highlights the stability challenges inherent in GAN training. The overpowering discriminator phenomenon (D_loss ≈ 0) emerged as the primary

obstacle, causing the generator to fail to learn stably over extended training periods. The performance collapse after epoch 100 demonstrates that longer training does not necessarily yield better models. This finding underscores the importance of periodic evaluation and model checkpointing during training. Without such strategies, the best-performing model (at epoch 100) would be missed, and the final saved model (at epoch 1,000) would exhibit inferior quality.

This study acknowledges several limitations that need to be addressed in future work:

Resolution

The generated images are limited to 128 x 128 pixels. While sufficient for global structure learning, this resolution restricts the representation of fine retinal micro-lesions (e.g., tiny microaneurysms) crucial for early-stage diagnosis.

Clinical Validation

Currently, the evaluation relies solely on computational metrics (FID/IS). A rigorous clinical validation involving ophthalmologists (e.g., a Visual Turing Test) has not yet been conducted. Expert feedback is essential to verify if the synthetic lesions are medically accurate and not just visually plausible artifacts.

Based on these findings, future research directions may focus on:

Improving Training Stability

Applying advanced regularization techniques to the discriminator, such as Spectral Normalization or Wasserstein GAN with Gradient Penalty (WGAN-GP), to prevent discriminator dominance.

Learning Rate Optimization

Adopting the Two Time-scale Update Rule (TTUR) by employing different learning rates for the generator and discriminator to maintain balanced training dynamics.

High-Resolution Synthesis

Once stability is achieved, scaling up the model to generate images at 512 x 512 pixels to capture fine-grained diabetic retinopathy lesions.

Conclusion

This study successfully implemented and evaluated a Conditional Lightweight GAN (c-LGAN) architecture to synthesize retinal fundus images based on five levels of Diabetic Retinopathy (DR) severity from the IDRiD dataset. The proposed model demonstrated its capability to generate realistic images, as evidenced by the best Frechet Inception Distance (FID) score of 121.24 achieved at epoch 100. These results confirm that generative approaches, particularly c-LGAN, hold substantial potential as data augmentation methods to address data scarcity and class imbalance issues in medical imaging datasets.

Nevertheless, this study also identified significant challenges related to long-term training stability. The occurrence of the overpowering discriminator phenomenon (D_loss ≈ 0) led to a performance collapse after the model reached its optimal point. This finding demonstrates that longer training does not necessarily guarantee better results and underscores the importance of periodic metric monitoring and best-model checkpointing to capture peak performance before instability arises.

Based on the conclusions and limitations identified in this study, several directions for future work can be proposed. Training stability can be improved by mitigating discriminator dominance through the application of more advanced regularization techniques. In particular, incorporating spectral normalization into the discriminator layers or adopting alternative frameworks such as Wasserstein GAN with Gradient Penalty (WGAN-GP) has strong potential to stabilize the adversarial training process. In addition, learning rate optimization represents an important avenue for improvement. Applying different learning rates for the generator and discriminator, a strategy known as the Two Time-scale Update Rule (TTUR), may help maintain a more balanced competitive dynamic between the two networks and prevent one from becoming excessively dominant at an early stage of training. Furthermore, once training stability has been

adequately addressed, future studies should consider increasing the image resolution used during training, for example to 256 × 256 or 512 × 512 pixels. This step is essential for generating synthetic retinal fundus images with richer clinical detail, particularly for the visualization of micro-lesions characteristic of diabetic retinopathy. Finally, clinical validation should be incorporated through qualitative evaluations involving medical experts, such as ophthalmologists, using approaches like the Visual Turing Test. Feedback from domain experts would provide invaluable insight into the clinical realism and potential diagnostic utility of the synthetic images generated by the best-performing model.

This study successfully implemented and evaluated a Conditional Lightweight GAN (c-LGAN) architecture to synthesize retinal fundus images based on five levels of Diabetic Retinopathy (DR) severity from the IDRiD dataset. The proposed model demonstrated its capability to generate realistic images while maintaining computational efficiency, as evidenced by the best Frechet Inception Distance (FID) score of 121.24 achieved at epoch 100. These results confirm that generative approaches, particularly c-LGAN, hold substantial potential as data augmentation methods to address data scarcity and class imbalance issues in medical imaging datasets without requiring high-performance computing infrastructure.

Nevertheless, this study also identified significant challenges related to long-term training stability. The occurrence of the overpowering discriminator phenomenon (D_loss ≈ 0) led to a performance collapse after the model reached its optimal point. This finding demonstrates that longer training does not necessarily guarantee better results and underscores the importance of periodic metric monitoring and best-model checkpointing to capture peak performance before instability arises.

Based on the conclusions and limitations identified in this study, several directions for future work are proposed to enhance both technical robustness and clinical applicability. Future iterations should mitigate discriminator dominance through advanced regularization techniques. Specifically, incorporating Spectral Normalization or adopting the Wasserstein GAN with Gradient Penalty (WGAN-GP) framework has strong potential to stabilize the adversarial training process. Additionally, adopting the Two Time-scale Update Rule (TTUR)—applying different learning rates for the generator and discriminator—may help maintain a balanced competitive dynamic.

To address the limitation of fine-grained detail in the current 128 x 128 output, future studies should aim to increase the synthesis resolution to 256 x 256 or 512 x 512 pixels. This step is essential for capturing micro-lesions (e.g., microaneurysms) characteristic of early-stage diabetic retinopathy, ensuring the data is not just visually similar but diagnostically valuable.

A critical limitation of the current study is the reliance on computational metrics (FID/IS). Future work must incorporate clinical validation through qualitative evaluations involving ophthalmologists, such as a Visual Turing Test. Feedback from domain experts is indispensable to verify the clinical realism and potential diagnostic utility of the synthetic images, ensuring they do not contain misleading artifacts that could confuse diagnostic models.

References

[1] T. E. Tan and T. Y. Wong, "Diabetic retinopathy: Looking forward to 2030," *Front Endocrinol (Lausanne)*, vol. 13, pp. 1–8, Jan. 2023, doi: 10.3389/fendo.2022.1077669.

[2] W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, "Diabetic retinopathy fundus image classification and lesions localization system using deep learning," *Sensors*, vol. 21, no. 11, pp. 1–22, Jun. 2021, doi: 10.3390/s21113704.

[3] A. D. Bhatwadekar, A. Shughoury, A. Belamkar, and T. A. Ciulla, "Genetics of diabetic retinopathy, a leading cause of irreversible blindness in the industrialized world," *Genes (Basel)*, vol. 12, no. 8, pp. 1–17, Aug. 2021, doi: 10.3390/genes12081200.

[4] D. S. Fong, L. P. Aiello, F. L. Ferris, and R. Klein, "Diabetic Retinopathy," *Diabetes Care*, vol. 27, no. 10, pp. 2540–2553, Oct. 2004, doi: 10.2337/diacare.27.10.2540.

[5] L. Wu, "Classification of diabetic retinopathy and diabetic macular edema," *World J Diabetes*, vol. 4, no. 6, pp. 290–294, Dec. 2013, doi: 10.4239/wjd.v4.i6.290.

[6] P. Vashist, S. Singh, N. Gupta, and R. Saxena, "Role of early screening for diabetic retinopathy in patients with diabetes mellitus: An overview," *Indian Journal of Community Medicine*, vol. 36, no. 4, pp. 247–252, Oct. 2011, doi: 10.4103/0970-0218.91324.

[7] I. Welch Allyn and Ynjiun Paul Wang, "FUNDUS IMAGING SYSTEM," Aug. 13, 2019

[8] Z. Yang, T. E. Tan, Y. Shao, T. Y. Wong, and X. Li, "Classification of diabetic retinopathy: Past, present and future," *Front Endocrinol (Lausanne)*, vol. 13, pp. 1–18, Dec. 2022, doi: 10.3389/fendo.2022.1079217.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[10] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, pp. 1–26, May 2023, doi: 10.3390/computers12050091.

[11] P. Porwal *et al.*, "IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge," *Med Image Anal*, vol. 59, no. 1, pp. 1–83, Sep. 2019, doi: 10.1016/j.media.2019.101561.

[12] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," Jun. 10, 2014, *arXiv*. [Online]. Available: http://arxiv.org/abs/1406.2661

[13] Y. Zhou, B. Wang, X. He, S. Cui, and L. Shao, "DR-GAN: Conditional Generative Adversarial Network for Fine-Grained Lesion Synthesis on Diabetic Retinopathy Images," *IEEE J Biomed Health Inform*, vol. 24, no. 1, pp. 56–66, Dec. 2022, doi: 10.1109/JBHI.2020.3045475.

[14] N. Abd Aziz, M. Azman Hanif Sulaiman, A. Zabidi, I. Mohd Yassin, M. Syahirul Amin Megat Ali, and Z. Ismael Rizman, "Lightweight Generative Adversarial Network Fundus Image Synthesis," *INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION*, vol. 6, no. 1, pp. 270–277, Mar. 2022, [Online]. Available: www.joiv.org/index.php/joiv

[15] P. Porwal *et al.*, "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research," *Data (Basel)*, vol. 25, no. 3, pp. 1–8, Jun. 2018, doi: 10.21227/H25W98.

[16] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," Nov. 2014, [Online]. Available: http://arxiv.org/abs/1411.1784

[17] P. Zhang *et al.*, "Fundus Image Generation and Classification of Diabetic Retinopathy Based on Convolutional Neural Network," *Electronics (Switzerland)*, vol. 13, no. 18, Sep. 2024, doi: 10.3390/electronics13183603.

[18] T. Miyato and M. Koyama, "cGANs with Projection Discriminator," Feb. 2018, [Online]. Available: http://arxiv.org/abs/1802.05637

[19] D. A. Chan and S. P. Sithungu, "Evaluating the Suitability of Inception Score and Fréchet Inception Distance as Metrics for Quality and Diversity in Image Generation," in *CIIS 2024 - 2024 the 7th International Conference on Computational Intelligence and Intelligent Systems*, Association for Computing Machinery, Inc, Feb. 2025, pp. 79–85. doi: 10.1145/3708778.3708790.

[20] A. Brock, J. Donahue, and K. Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," Sep. 2018, [Online]. Available: http://arxiv.org/abs/1809.11096

© 2026 by the author; licensee Matrix: Jurnal Manajemen Teknologi dan Informatika. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).