Received: 5 December 2025 | Revised: 27 March 2026 | Accepted: 18 May 2026 | Published online: 23 July 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewJCCE62028693

BON VIEW PUBLISHING

Lesion-Conditioned Fundus Image Synthesis for Diabetic Retinopathy Using Pix2Pix Conditional GANs

Anju Asokan[1] and Manju Bhaskara Panickar Radha[2,*]

[1] *Department of Mathematics, Amrita Vishwa Vidyapeetham-Amritapuri, India*

[2] *Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Amritapuri, India*

Abstract: Deep learning models for diabetic retinopathy (DR) demonstrate superior performance when trained on an extensive, class-balanced fundus dataset. However, widely used collections, such as EyePACS, exhibit significant class imbalances, particularly with the underrepresentation of severe stages. We propose a lesion-conditioned, pathology-aware synthesis framework to augment DR fundus images using a Pix2Pix-based conditional generative adversarial network. Lesion structures are incorporated through segmentation masks for four clinically significant lesion types—microaneurysms, hemorrhages, hard exudates, and soft exudates—derived from pixel-level annotations and/or a trained U-Net segmenter, which serves as conditioning inputs for the generator. To enhance data availability for minority grades, we implemented a minimum-real subsampling baseline and augmented each grade with lesion-fidelity-filtered synthetic images using a mask cycling strategy. We present qualitative results that demonstrate spatial consistency between the conditioned lesion masks and the synthesized appearance, and we include quantitative image-quality descriptors (Fréchet Inception Distance), discussing their limitations in small, low-diversity medical datasets. Finally, we evaluated the downstream utility through a real-only versus real + synthetic robustness study across three random seeds, reporting mean ± std and explicitly noting the small held-out test subset used in this paper. Overall, the proposed framework facilitates controlled, lesion-preserving synthesis for class balancing and offers an auditable augmentation protocol for DR model development.

Keywords: diabetic retinopathy, fundus image synthesis, Pix2Pix, conditional GAN, data augmentation

1. Introduction

Deep learning has made substantial advancements in medical image analysis, enabling automated diagnosis, disease screening, segmentation, and decision support across a diverse array of clinical applications. The continuous evolution of convolutional neural networks and extensive deep architectures has significantly enhanced the performance of tasks such as classification, detection, and pixel-level analysis, frequently surpassing traditional machine learning methods. Nevertheless, despite these technological advancements, the application and generalization of deep learning models in clinical settings remain fundamentally constrained by data-related challenges [1, 2]. A persistent issue identified in the literature is the scarcity of large, well-annotated medical-imaging datasets. Numerous studies have highlighted the lack of sufficiently annotated large-scale datasets as a primary impediment to medical image analysis [3, 4]. This challenge is further compounded by the increasing complexity of modern deep learning models, which often consist of millions to billions of trainable parameters, necessitating vast amounts of labeled data for stable and effective training [3–6]. Assembling such datasets imposes significant demands on clinicians and involves considerable financial and time investments. In addition to data volume, there are additional constraints related to data accessibility and feasibility. Medical image annotation relies on domain experts, such as radiologists, ophthalmologists, and other specialists, who must perform meticulous manual labeling [6]. This process is labor-intensive and is further restricted by privacy regulations and legal considerations that limit the collection, sharing, and reuse of patient data [7, 8]. Consequently, many medical imaging datasets remain relatively small, heterogeneous, and biased toward specific conditions [9–11].

In addition to the challenge of limited data volume, class imbalance poses a significant obstacle to deep learning in medical imaging. In numerous clinical settings, pathological findings are considerably less prevalent than normal findings, resulting in datasets predominantly composed of healthy or nearly normal samples [12, 13]. Given that abnormal cases often represent only a small fraction of real-world data, these skewed class distributions can bias learning algorithms, leading to low sensitivity for minority classes and unstable decision boundaries [14, 15]. Segmentation and detection models encounter similar issues, such as increased false-positive and false-negative rates, unclear lesion

***Corresponding author:** Manju Bhaskara Panickar Radha, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Amritapuri, India. Email: manjubr@am.amrita.edu

boundaries, and diminished performance for rare pathologies. These data-related challenges have prompted a gradual shift from solely model-centric optimization to data-centric strategies aimed at directly enhancing the quality, diversity, and balance of the training sets [16].

The creation of synthetic data has emerged as a promising and increasingly vital approach to address issues such as limited data availability, class imbalance, privacy concerns, and the high costs of annotation. Rather than relying solely on limited real-world data, synthetic augmentation seeks to broaden the training distribution, reduce bias, and enhance generalization without imposing additional burdens on patients or healthcare providers. Among the various generative techniques, generative adversarial networks (GANs) have garnered significant attention for their ability to generate high-quality images that closely resemble real clinical data [3, 4]. In a GAN, a generator network is trained to produce synthetic images, whereas a discriminator network attempts to differentiate between the generated and real samples. This adversarial training process enables the generator to progressively improve its outputs until they are nearly indistinguishable from authentic medical images. This capability renders GANs cost-effective and scalable alternatives for collecting new data and has led to numerous demonstrations of improved classification and segmentation performance across MRI, CT, X-ray, ultrasound, and other modalities [3, 5, 13].

In ophthalmology, particularly in the context of fundus imaging, deep learning applications encounter similar data-related challenges despite the significant clinical importance of retinal diseases. Fundus photographs are crucial for the identification and evaluation of conditions such as diabetic retinopathy (DR), glaucoma, and age-related macular degeneration [12]. However, large-scale fundus datasets often exhibit a pronounced class imbalance, with advanced stages of DR being far less prevalent than normal or mild-stage cases. This imbalance significantly constrains a model's ability to accurately detect vision-threatening diseases, as minority classes provide insufficient examples for effective learning [7, 17, 18]. To address this issue, GAN-based synthesis has been explored in ophthalmic imaging to enhance the minority classes and expand the training distributions. Nonetheless, most current fundus synthesis studies rely on unconditional or weakly conditioned GAN architectures, which, while capable of generating visually realistic retinal images, offer limited control over disease-specific pathological features of the images. Consequently, these methods often fail to ensure clinically accurate lesion representation and consistent severity characteristics, which are crucial for reliable DR grading [19, 20]. Despite the clear benefits of GAN-based medical image synthesis, existing methods have notable limitations. Much of the previous research emphasizes overall image realism, often overlooking the structural attributes of lesions, leading to synthetic images that appear plausible but lack precise pathological control [2, 7, 8]. Unconditional GANs, in particular, offer limited interpretability because they lack a direct mechanism to enforce clinically relevant constraints on lesion location, density, or severity, even though these factors are vital for accurate DR staging. Furthermore, segmentation and synthesis are frequently treated as separate tasks rather than incorporating lesion information directly into the generative process [13]. Evaluation protocols are often restricted to global metrics, such as Fréchet Inception Distance (FID) or qualitative visual assessment, which are useful for assessing perceptual realism but do not ensure lesion-level accuracy or clinical reliability. These limitations underscore the need for pathology-aware generative frameworks that explicitly integrate lesion-level structural information into the synthesis process, particularly for retinal imaging tasks in which diagnosis and grading are inherently lesion-driven [7, 8].

To address these challenges, we propose a pathology-aware, lesion-conditioned Pix2Pix framework for generating DR fundus images. Unlike previous GAN-based DR augmentation methods, which are often unconditional or only conditioned on grade labels, our approach employs multi-lesion, pixel-level segmentation masks, such as microaneurysms, hemorrhages, hard exudates, and soft exudates, as structural conditioning inputs for the generator. This conditioning establishes a direct connection between the lesion masks and generated images, enhancing the lesion-level spatial consistency and providing more clinically relevant control over pathological patterns. Because EyePACS lacks pixel-level lesion annotations, the lesion masks used for training and conditioning are sourced from datasets with pixel-level labels (e.g., IDRiD), and the synthesis framework is evaluated under those annotation constraints; EyePACS is primarily cited as a real-world example of imbalance in DR grading. This study makes four key contributions:

Lesion-conditioned synthesis: We developed a Pix2Pix-basedAQ1conditional GAN that uses multi-label lesion masks as explicitconditioning channels for fundus image synthesis, enabling con-trollable generation that maintains lesion topology whilealigningwith DR severity labels.

Class-balanced dataset protocol: Wepropose a minimum-real subsampling strategy with controlledlesion-guided "topping up" to create a strictly class-balanceddataset, where an equal number of real images is retained pergrade, and only the shortfall is supplemented with syntheticsamples to maintain a consistent real-to-synthetic ratio acrossclasses.

Mask cycling with anatomical constraints: To addressimperfect pairing and minimize mask wastage in large-scaledatasets, we introduced a mask–image independence and cyclingmechanism that reuses lesion masks across compatible fun-dus backgrounds while ensuring anatomical plausibility throughexplicit anatomical exclusion constraints (e.g., optic-disk overlaprejection) and quality control (QC) checks.

Class-wise, lesion-aware evaluation: We conducted a class-wise evaluation of thesynthesized images using per-grade FID along with lesion-maskalignment analyses, collectively assessing visual realism and lesionfidelity beyond global image-level metrics.

The proposed framework advances beyond traditional realism-focused synthesis by addressing lesion-specific conditioning, anatomical accuracy, and clinically significant evaluation, thereby offering a more interpretable and pathology-aware augmentation method for DR imaging. Unlike previous GAN-based DR augmentation approaches, which are either unconditional or conditioned on grade labels, our framework integrates lesion structure conditioning with lesion-fidelity filtering and anatomical validity constraints. This enables controllable pathology synthesis, rather than mere realism generation. In the following sections, we first review the related work on GAN-based medical image synthesis, with a particular focus on fundus imaging and DR augmentation (Section 2). We then detail the proposed lesion-conditioned pipeline, which includes preprocessing, lesion segmentation, dataset subsampling, mask cycling, and Pix2Pix-based synthesis (Section 3). Section 4 presents the qualitative synthesis results and quantitative evaluations, including, where applicable, comparisons of downstream robustness between real-only and real + synthetic training. Finally, Section 5 concludes the study and suggests future research directions.

2. Literature Review

2.1. Data constraints in medical imaging and the rise of synthetic augmentation

Recent advancements in deep learning have significantly enhanced the analysis of medical images, including classification, segmentation, detection, and decision support. Despite these advancements, the performance of models in clinical imaging remains considerably constrained by data-related factors, including limited dataset sizes, varied acquisition protocols, annotation inconsistencies, class imbalances, and privacy and governance limitations. These challenges are particularly problematic for neural networks that require large datasets, as a lack of diversity and imbalanced class distributions can lead to unstable decision boundaries and reduced sensitivity to rare but clinically important conditions. Consequently, research has gradually shifted from solely optimizing models to embracing data-centric strategies that enhance the quality, diversity, and balance of datasets [3, 4].

Among these data-centric approaches, GANs have gained traction as a method for generating realistic medical images that closely mimic real clinical data distributions. The application of GANs for image synthesis has been explored across various modalities, including MRI, CT, X-ray, ultrasound, and histopathology, to support training and evaluation processes in scenarios of data scarcity and imbalance. Numerous surveys and reviews have concluded that GANs can expand training datasets, enrich underrepresented categories, and facilitate privacy-preserving data sharing. However, they also emphasized the importance of evaluating synthetic images using clinically relevant criteria to avoid introducing new biases or artifacts into subsequent learning systems [3, 5, 21].

2.2. GAN-based augmentation across clinical imaging tasks

GAN-based synthesis has been employed in a diverse array of medical imaging tasks. In the segmentation domain, synthetic images contribute to the diversity of organ and lesion boundaries during training, thereby enhancing the robustness of boundary learning and addressing the limitations posed by sparse annotation coverage. In classification tasks, synthetic augmentation mitigates imbalanced class distributions and enhances the performance of minority classes, particularly when rare patterns are underrepresented in real-world datasets. Furthermore, synthetic data are utilized in detection and registration scenarios to expose models to a broader spectrum of anatomical variations. Nonetheless, a prevalent limitation of much of this research is the tendency to treat the generator as a black box, with a primary focus on increasing the data volume. Many studies aim for global realism without imposing specific constraints on pathology-related structures such as lesion morphology, spatial distribution, and patterns consistent with severity. Consequently, although synthetic images may appear plausible, they frequently lack sufficient control at the lesion level, which is crucial for clinical interpretability and reliability.

2.3. Unconditional and weakly conditioned GANs: Realism without lesion control

Numerous GAN-based methodologies for synthesizing medical images predominantly utilize either unconditional models or those with minimal conditioning, such as Deep Convolutional Generative Adversarial Network (DCGAN)-style frameworks, progressively grown GANs, and GANs conditioned on class labels. These approaches can generate visually authentic images and are beneficial for oversampling at the image level. Nevertheless, their conditioning strategies are typically broad, focusing on global labels rather than detailed structural indicators. This limitation is particularly pronounced when clinical decisions hinge on localized and sparse pathological features such as microaneurysm clusters in DR or small nodules in thoracic imaging. In such instances, reliance solely on label conditioning provides limited control over the location, size, shape, and density of lesions. Consequently, although the generated images may exhibit high perceptual realism, they do not necessarily ensure an accurate representation of clinically significant lesion configurations that are essential for diagnosis and grading [22, 23].

2.4. Structure-aware and lesion-aware generative models

In response to the constraints of coarse conditioning, a growing, albeit limited, body of research is exploring structure- and lesion-aware generative synthesis. These methodologies condition the generator not only on class labels but also on structured priors, such as segmentation masks, anatomical labels, landmark configurations, or vessel-tree representations. By incorporating such structural data, the alignment between synthetic images and clinically significant structures is enhanced, facilitating a more accurate depiction of the lesion morphology and spatial arrangement. This approach is particularly pertinent to DR, where severity grading is contingent upon the presence, combination, and distribution of various lesion types, rather than merely the overall appearance. Nonetheless, challenges persist in designing effective multi-lesion conditioning mechanisms, ensuring anatomical plausibility across diverse fundus backgrounds, and evaluating lesion fidelity using clinically relevant protocols instead of solely realism-based metrics [17, 24].

A practical issue frequently overlooked in many structure-conditioned frameworks is the discrepancy between idealized paired training data and actual clinical data. Large datasets may exhibit imperfect pairings between images and masks, varied annotation quality, and inconsistent labeling coverage. These issues necessitate robust strategies for conditioning and QC when lesion masks are reused or transferred across different backgrounds, while maintaining physiological plausibility.

2.5. Evaluation practices and limitations: Beyond FID and visual inspection

Comprehensive reviews of GAN-based medical image synthesis indicate that synthetic data frequently enhance visual realism and may improve segmentation or classification performance in specific contexts. However, these reviews underscore persistent evaluation challenges. Numerous studies predominantly depend on global, image-level metrics, such as the FID or similar perceptual scores. Although these metrics evaluate distributional similarity and perceptual plausibility, they do not directly verify lesion-level accuracy, anatomical validity, or clinical applicability. When reader studies are incorporated, they are often small-scale or informal, and only a limited number of studies systematically assess robustness, fairness, or generalization across diverse acquisition settings and devices. Consequently, there is an increasing demand in the literature for evaluation protocols

that are pathology-aware and connect synthetic image quality to clinically significant outcomes such as lesion fidelity, anatomical constraints, and minority class reliability [3, 5, 21, 25].

Significantly, the broader literature on generative modeling also highlights that learning rare events and semantic conditioning necessitates thorough validation beyond global realism. For example, the GAN-based synthesis of rare events in driving recorder data employs semantic information to represent infrequent but critical conditions more accurately [1]. Similarly, generative design in materials science illustrates that conditioning and validation constraints are essential when creating structures that must adhere to domain-specific realism criteria [2]. These insights further emphasize the need for constraint-aware and structure-conditioned synthesis strategies when applying generative augmentation in high-stakes areas.

2.6. Diabetic retinopathy datasets: Class imbalance and minority-grade reliability

In DR fundus imaging, the issue of class imbalance is well documented: severe stages are significantly less prevalent than normal or mild cases; however, they present the highest clinical risk and determine the urgency of referrals. Both clinical and dataset-focused studies underscore that the less common severe grades are often the most critical for ensuring screening reliability and patient safety, necessitating careful modeling considerations [6, 8]. Several studies have also indicated that class imbalance disproportionately affects sensitivity and decision stability for severe grades, even when the overall accuracy appears satisfactory [15, 22, 26]. In the literature on fundus synthesis and augmentation, challenges in controlling pathology and inadequate lesion-level validation are frequently cited as obstacles to reliable integration into DR training processes [19, 27]. These insights collectively underscore the need for augmentation strategies that are not only realistic but also faithful to lesions and are clinically constrained. Crucially, conditioning on DR lesions requires pixel-level lesion annotations, which are absent in many large-scale grading datasets. For instance, EyePACS is extensively used for DR grading and imbalance analysis but lacks pixel-level lesion masks. Consequently, lesion-conditioned synthesis frameworks typically rely on datasets with pixel-level annotations (e.g., the IDRiD dataset) for mask supervision and evaluation under annotation constraints. This distinction is vital for the reproducibility and interpretation of the outcomes of lesion-conditioned synthesis [21, 28].

2.7. Summary of gaps and motivation for this work

The literature review suggests that GAN-based augmentation can enhance performance metrics in situations where data are limited and imbalanced, primarily by increasing exposure to rare patterns through synthetic samples. However, achieving realism and balanced class counts does not inherently lead to clinically significant improvements. Minority classes may still demonstrate considerable variability in performance, and improvements in overall metrics can obscure ongoing deficiencies in rare but clinically significant categories of patients. Furthermore, only a limited number of prior studies have provided class-specific error analyses that assess the impact of synthetic data on sensitivity, specificity, and decision reliability for minority grades, which is essential for DR severity grading. Several gaps, particularly in DR fundus imaging, emerge from this literature:

1) Limited multi-lesion conditioning for DR synthesis: Current methodologies frequently lack conditioning or are weakly label-conditioned, failing to explicitly incorporate DR-relevant lesion structures (such as microaneurysms, hemorrhages, hard exudates, and soft exudates) as multi-channel conditioning input.
2) Lack of transparent balancing protocols: Numerous augmentation pipelines emphasize increasing sample numbers rather than creating strictly class-balanced datasets with controlled and interpretable real-to-synthetic ratios across severities.
3) Evaluation gaps: Current evaluation practices often prioritize global realism (e.g., FID) without integrating class-wise and lesion-aware assessments that collectively verify the image quality, lesion fidelity, and downstream reliability for minority grades.
4) Insufficient handling of practical dataset constraints: Issues such as imperfect mask–image pairing, varied annotation quality, and the necessity for anatomical validity constraints are frequently inadequately addressed in end-to-end augmentation pipelines.

2.8. Positioning of the proposed approach

This study introduces a lesion-conditioned Pix2Pix framework for synthesizing DR fundus images, addressing existing gaps in the field. The proposed method incorporates multi-lesion segmentation masks as structural conditioning channels and utilizes a minimum-real subsampling strategy to maintain a strictly class-balanced dataset with controlled real-to-synthetic ratios across various DR grades. The synthesis quality was assessed using metrics that were both class-specific and lesion-aware. The subsequent section explores the pipeline, providing a detailed account of preprocessing, lesion segmentation, subsampling, mask cycling under anatomical plausibility constraints, and lesion-guided Pix2Pix synthesis, specifically designed for DR imaging.

3. Research Methodology

This study introduces a lesion-conditioned methodology for generating DR fundus images using a Pix2Pix-based conditional GAN (cGAN). The methodology encompasses (i) standardized preprocessing and intensity normalization, (ii) the creation of multi-lesion segmentation masks using a U-Net architecture, (iii) controlled dataset assembly and definition of data splits, (iv) a mask–image independence and cycling strategy to optimize the utilization of available lesion masks, (v) anatomically constrained conditioning to ensure plausibility, (vi) lesion-conditioned Pix2Pix fundus image synthesis, and (vii) both qualitative and quantitative evaluation of the synthesized image quality. Each component of the process is modular and reproducible, ensuring consistent preprocessing, strong alignment between the lesion masks and fundus images, and traceable dataset construction. EyePACS is cited as a motivating imbalanced dataset, whereas pixel-level conditioning employs IDRiD masks. Specifically, as EyePACS lacks pixel-level lesion annotations, the necessary lesion masks for conditioning were sourced from IDRiD, and the verified experiments in this paper were conducted within these annotation constraints.

3.1. Dataset preprocessing and normalization

Preprocessing is applied to reduce variability in fundus images and to provide standardized inputs to both the

segmentation network and the conditional GAN. Fundus images are resized to fixed spatial resolutions and normalized to mitigate differences arising from acquisition devices, illumination conditions, and contrast variations. A standardized enhancement pipeline is applied (including intensity equalization and field-of-view standardization via circular cropping) to suppress non-informative background regions outside the retinal field.

Let $I \in R^{H \times W \times L}$ denote a raw RGB fundus image. Intensity normalization was performed on a per-image basis as

$$I_{norm} = \frac{I - \mu_1}{\sigma_1} \tag{1}$$

where μ_1 and σ_1 are the mean and standard deviation of the pixel intensities in I, respectively. This normalization reduces intra-class variation and improves numerical stability.

Resolution handling across stages: the segmentation network operates at 512 × 512, while the Pix2Pix synthesis model is trained at 256 × 256. Accordingly, after preprocessing, images and corresponding lesion masks are resized to 256 × 256 for the synthesis stage.

3.2. Lesion mask prediction using U-Net segmentation

To provide explicit structural guidance for synthesis, a U-Net–based segmentation model is trained to predict multi-lesion masks from preprocessed fundus images. Given a normalized fundus image I_{norm}, the segmentation network f_θ learns a mapping

$$f_\theta : I_{norm} \rightarrow M \tag{2}$$

where $M \in \{0,1\}^{H \times W \times L}$ is a multi-channel binary mask and $L = 4$ denotes the number of lesion types. The four channels correspond to clinically relevant DR lesions: microaneurysms (MA), hemorrhages (HE), hard exudates (EX), and soft exudates (SE). Each channel encodes pixel-level presence of a specific lesion type.

Training configuration: The U-Net is trained for 30 epochs at 512 × 512 resolution using the Adam optimizer with learning rate 1×10^{-3} and batch size 4.

Loss function: Due to lesion sparsity (particularly MA and SE), the segmentation model is optimized using a combined loss that balances overlap-driven learning with weighted pixel-wise supervision:

$$L_{seg} = 0.5L_{FT} + 0.5L_{WBCE} \tag{3}$$

where L_{FT} is a focal-Tversky loss and L_{WBCE} is a weighted binary cross-entropy loss with positive-class weighting to mitigate foreground imbalance. The resulting multi-lesion masks serve as explicit structural priors for the subsequent Pix2Pix synthesis step.

3.3. Controlled dataset composition and splits (robustness setup)

To ensure transparent reporting and reproducibility, the dataset is organized into explicitly defined train/validation/test splits. In the verified robustness setup used in this paper, the experimental composition is:

1) Train: $N = 108$ (44 real + 64 synthetic)
2) Validation: $N = 3$ (QC subset; explicitly noted due to strict split separation)
3) Test: $N = 13$

All downstream robustness experiments were repeated across three random seeds: 42, 43, and 44 to assess variability and report performance as mean ± standard deviation.

3.4. Mask–image independence, cycling, and anatomical plausibility constraint

In practical settings, strict one-to-one pairing between a lesion mask and its originating fundus image can be limiting due to incomplete correspondences and varying annotation coverage. To maximize the use of available lesion masks, a mask–image independence strategy is adopted:

For each DR grade c, the set of lesion masks M_c and the set of real fundus images D_c^{real} are treated as separate pools. During training, a lesion mask $m \in M_c$ may be paired with a fundus background $I \in D_c^{real}$ of the same grade to form a pseudo-paired training example (m, I) for the conditional generator. This avoids dependence on filename-level matching while preserving severity compatibility.

Anatomical constraint (optic-disk exclusion): To address anatomical plausibility concerns (e.g., lesions overlapping the optic disk), an explicit exclusion constraint is enforced during conditioning. The optic-disk region is estimated on the real background and dilated by a safety margin to form an exclusion zone. Candidate lesion masks that overlap this exclusion region are rejected via a QC loop prior to synthesis, reducing anatomically implausible lesion placement during mask transfer.

3.4.1. Anatomical Plausibility Constraint (Optic-Disk Exclusion)

3.5. Lesion-conditioned fundus image synthesis using Pix2Pix

Synthetic fundus images were generated using a Pix2Pix-based cGAN consisting of a generator G and a discriminator D. The generator learns to translate a multi-channel lesion mask M into a synthesized fundus image $\hat{I}$,

$$G : M \rightarrow \hat{I}, \tag{4}$$

where M is a multi-lesion mask (MA, HE, EX, SE) and $\hat{I}$ is the generated RGB fundus image. The discriminator D receives paired inputs and learns to distinguish real mask–image pairs (M, I) from synthetic pairs $(M, \hat{I})$, thereby enforcing realism conditioned on the lesion structure.

1) Adversarial + reconstruction objective

Training combines an adversarial loss with an $L1$ reconstruction loss:

$$L(G, D) = \mathrm{L_{cGAN}}(G, D) + \lambda_{L1} L_{L1}(G). \tag{5}$$

In our implementation, the adversarial component is trained in an Least Squares Generative Adversarial Network (LSGAN)-style formulation (mean squared error), which stabilizes training compared to the original log-likelihood GAN loss. The reconstruction loss is defined as:

$$L_{L1}(G) = \mathrm{E}_{M,I}[\,||\, I\text{-}G(M) \,||_1\,]. \tag{6}$$

The reconstruction term encourages the synthesized output to preserve global color/illumination consistency with the paired

target image while respecting the lesion topology encoded in M. The $L1$ term is weighted using:

$$\lambda_{L1} = 100.0. \quad (7)$$

2) Training configuration

Pix2Pix was trained for 100 epochs at 256 × 256 resolution using the Adam optimizer with a learning rate of 2×10^{-4}, $\beta_1 = 0.5$, and a batch size of 8. This lesion-conditioned formulation provides dense spatial conditioning, enabling the generator to synthesize images that reflect lesion distribution and severity patterns more explicitly than unconditional GANs.

3.6. Synthetic sample selection (Top-K, lesion-fidelity filtered)

To reduce low-signal or anatomically inconsistent synthetic outputs, synthetic samples were filtered using a lesion-fidelity selection strategy. Specifically, each candidate synthetic image was first subjected to QC checks, including (i) non-empty lesion presence, (ii) mask–image alignment validation, and (iii) anatomical plausibility constraints (e.g., optic-disk overlap rejection where applicable).

Top-K selection was performed by ranking the remaining candidates using the Lesion-Fidelity Score (LFS) and retaining the highest-scoring samples after QC. In other words, Top-K denotes the subset of synthetic images with the largest LFS values among QC-passing candidates.

In the verified robustness run used in this paper, the final synthetic set retained for downstream training contains $N = 64$ LFS-ranked synthetic images, which were combined with $N = 44$ real images to form the classification training set ($N = 108$ total).

3.7. Evaluation of the synthetic image quality

Synthetic image quality was assessed using both qualitative inspection and quantitative measures.

3.7.1. Qualitative assessment

Lesion masks were superimposed on synthesized fundus images to visually confirm alignment and spatial agreement. Inspection focused on whether lesions appeared at intended mask locations and whether retinal structures (optic disk, macula, vascular tree) remained plausible. Real and synthetic images were compared side by side within grade to assess color distribution, contrast, and background texture.

3.7.2. Quantitative assessment using Fréchet Inception Distance (FID)

The FID is used to quantify similarity between real and synthetic images in a deep feature space. Let (μ_r, Σ_r) and (μ_s, Σ_s) denote the mean and covariance of feature embeddings extracted from real and synthetic images. FID is defined as,

$$\mathrm{FID} = ||\mu_r - \mu_s||^2 + \mathrm{Tr}(\Sigma_r + \Sigma_s - 2(\Sigma_r\Sigma_s)^{1/2}). \quad (8)$$

In addition to any global FID measurement, grade-wise (class-wise) FID was reported where applicable to diagnose synthesis quality across severity levels. Basic dataset statistics (real vs synthetic counts per split) were also reported for transparency and reproducibility.

3.8. Downstream classification robustness evaluation (real-only vs real + synthetic)

To evaluate whether lesion-conditioned synthetic augmentation improves DR grading performance, a controlled classification study was conducted with two arms:

1) Real-only: classifier trained on real images only.
2) Real + synthetic: classifier trained on real images plus lesion-fidelity-filtered Top-K synthetic images.

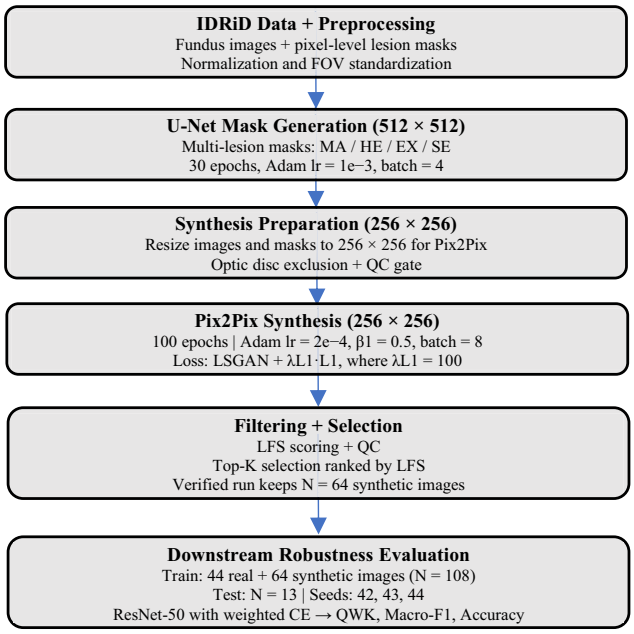
Both arms used the same backbone (ResNet-50) and identical training settings to ensure parity, including weighted cross-entropy loss (inverse class frequency). Experiments were repeated across seeds 42, 43, and 44, and both arms were evaluated on the same held-out test set ($N = 13$) using:

1) Quadratic Weighted Kappa (QWK)
2) Macro F1-score
3) Accuracy

Results are reported as mean ± standard deviation across the three seeds, reflecting variance under the small test-set regime and enabling a robustness-oriented comparison between arms (the overall pipeline is given in Figure 1).

In summary, the proposed methodology introduces a lesion-conditioned Pix2Pix framework tailored for DR fundus synthesis and verified under pixel-level annotation constraints. By combining multi-lesion conditioning, mask cycling, and explicit anatomical plausibility constraints (optic-disk exclusion), the pipeline produces pathology-preserving synthetic images. The study further includes a controlled robustness evaluation comparing real-only versus real + synthetic training under identical settings, providing an evidence-based comparison of downstream utility.

Figure 1
Pix2Pix-based lesion-conditioned fundus image synthesis pipeline



4. Results

This section presents and analyzes the experimental outcomes obtained using the proposed lesion-conditioned Pix2Pix framework for DR fundus image synthesis and augmentation under pixel-level annotation constraints. Although EyePACS is referenced as the motivating large-scale imbalance scenario in DR screening, the verified experiments in this paper are conducted on IDRiD, because IDRiD provides the pixel-level lesion masks required for conditioning. The analysis is structured around five objectives: (i) verifying the reliability of the lesion masks used as conditioning signals, since conditioning quality directly constrains synthesis quality; (ii) validating lesion fidelity and mask–image consistency in synthesized images; (iii) verifying anatomical plausibility through optic-disk exclusion and its measured failure rates; (iv) confirming that synthetic sample selection is driven by Top-K ranking using LFS rather than arbitrary retention; and (v) evaluating downstream utility via a controlled robustness experiment comparing real-only versus real + synthetic training under identical settings on the same held-out test subset.

4.1. Segmentation quality of conditioning masks (U-Net)

Because the Pix2Pix generator is explicitly conditioned on lesion masks, the segmentation stage must be reported and audited rather than treated as a hidden intermediate. We therefore evaluate the U-Net segmentation model that predicts four lesion channels: microaneurysms (MA), hemorrhages (HE), hard exudates (EX), and soft exudates (SE). The per-lesion segmentation performance on the verified IDRiD test partition, consisting of $N = 13$ images, is reported in Table 1. This test regime is strict because retinal lesions are extremely sparse at the pixel level, the labeled test subset is small, and minor annotation or preprocessing issues can disproportionately reduce overlap-based metrics such as Dice and Intersection over Union (IoU). Under these conditions, low Dice and IoU values are expected, especially for MA and SE, where positive lesion pixels can be rare or even absent in several images. As shown in Table 1, hard exudates (EX) achieve the highest segmentation performance among the four lesion categories, with a Dice score of 0.0419 and an IoU of 0.0233. Hemorrhages (HE) achieve a Dice score of 0.0119 and an IoU of 0.0061, reflecting the difficulty of detecting sparse hemorrhagic regions under limited supervision. For microaneurysms (MA) and soft exudates (SE), the Dice and IoU values are approximately zero. This is consistent with the known difficulty of detecting tiny microaneurysms and rare soft exudates, particularly in a small test subset. Importantly, the results in Table 1 are not used to over-claim segmentation performance. Instead, they are reported to make the lesion-conditioned synthesis pipeline transparent and auditable. Since the generated fundus images depend on lesion-mask conditioning, reporting segmentation performance helps clarify the quality and limitations of the conditioning signal used for synthesis.

To keep the synthesis pipeline defensible despite lesion sparsity, the revised workflow includes an integrity audit that blocks common failure modes before synthesis. This audit includes non-empty-mask presence checks to prevent empty-mask propagation and visual overlay verification to confirm that exported masks correspond to the expected lesion topology. This step is critical for reviewer defensibility because, even when segmentation metrics are low, the pipeline remains traceable. Each synthetic sample can be linked back to an explicit conditioning mask that passes the required integrity checks.

4.2. Dataset composition and synthetic image contribution (verified IDRiD run)

In the verified robustness setup, synthetic data are introduced in a controlled and auditable manner to evaluate whether lesion-conditioned augmentation improves DR grading under strict annotation constraints. Because pixel-level lesion masks are required for conditioning, all synthesis and training data in this work are derived from IDRiD, not EyePACS. The training set consists of $N = 44$ real fundus images from the IDRiD segmentation partition combined with $N = 64$ synthetic fundus images generated using Pix2Pix and retained using the Top-K lesion-fidelity selection strategy (Section 4.4), resulting in $N = 108$ training samples. The held-out test subset contains $N = 13$ labeled images (seg_test_with_labels.csv) used consistently across all robustness seeds. This design is intentionally conservative. Rather than claiming large-scale dataset balancing (e.g., "2000 images per class" or "10,000 total"), the work emphasizes a verified, reproducible augmentation protocol where every synthetic image is (i) generated from a traceable conditioning mask, (ii) screened by QC checks, and (iii) retained based on lesion-fidelity ranking. EyePACS remains a motivating imbalance dataset for real-world screening, but it is not used for pixel-level conditioning or for the verified evaluation reported here.

4.3. Lesion fidelity and mask–image consistency in synthesized images

The primary goal of the proposed framework is not merely to generate visually plausible retinal backgrounds but to generate images whose pathological content is semantically aligned with clinically relevant DR lesions. In unconditional GAN settings, fundus realism may appear high while lesion content remains

Table 1
U-Net segmentation performance on verified IDRiD test subset ($N = 13$)

Lesion type	Dice	IoU	Interpretation
EX (hard exudates)	0.0419	0.0233	Low overlap under strict sparse-target regime
HE (hemorrhages)	0.0119	0.0061	Very sparse, low overlap expected
MA (microaneurysms)	0.0000	0.0000	Extremely sparse; tiny lesions often pixel-rare
SE (soft exudates)	0.0000	0.0000	Rare targets; overlap often zero in small subset

uncontrolled. Lesions may be hallucinated, misplaced, distorted, or absent, thereby reducing the clinical value of synthetic data for DR grading. In contrast, the Pix2Pix-based formulation used in this work enforces an explicit mapping from multi-channel lesion masks to synthesized fundus appearance, making the pathological content more controllable and interpretable. To assess this aspect, we qualitatively evaluated lesion fidelity by overlaying the multi-channel lesion masks, including microaneurysms, hemorrhages, hard exudates, and soft exudates, on the corresponding generated outputs. The objective was to inspect whether the synthesized lesions appeared at the intended spatial locations and whether their morphologies were consistent with the provided masks. The qualitative lesion fidelity and mask–image consistency criteria used for this inspection are summarized in Table 2. As shown in Table 2, the inspection focused on four key aspects: spatial correspondence, morphological consistency, background plausibility, and artifact screening. Spatial correspondence verifies whether lesions appear in the regions indicated by the conditioning mask. Morphological consistency checks whether the lesion extent follows the approximate shape of the mask without large drift beyond mask boundaries. Background plausibility ensures that important retinal landmarks, such as the optic disk, macula, and blood vessels, remain visually coherent. Artifact screening verifies that obvious synthesis artifacts do not dominate the generated image.

Across the inspected samples, the synthesized outputs showed lesion placement that followed the conditioning structures. Lesion clusters appeared in regions dictated by the masks rather than emerging as random noise patterns. This is especially important for DR grading, where disease severity depends on lesion type, density, and distribution. By explicitly conditioning on lesion topology, the generator is encouraged to preserve pathological structure, thereby reducing the risk of uncontrolled lesion hallucination typically associated with unconditional synthesis.

Because the verified package is focused on robustness evaluation rather than exhaustive visual statistics, we report this alignment as a qualitative consistency finding supported by mask-overlay inspection. As illustrated in Figure 2 and summarized in Table 2, the generated samples demonstrate visible mask–image consistency under qualitative review. In future extensions, stricter QC criteria may be applied to filter out misregistered or artifact-dominated cases.

4.4. Anatomical plausibility: Optic-disk exclusion and verification

A key reviewer concern in DR synthesis is anatomical plausibility, particularly preventing lesions from appearing on the optic disk, which is physiologically implausible. To address this, we enforce an optic-disk exclusion constraint during conditioning: the optic-disk region is estimated on the real-background image, dilated by a safety margin, and candidate masks that overlap this exclusion region are rejected via a QC loop prior to synthesis. This ensures that the generator is never conditioned on anatomically invalid lesion placement in the optic-disk region.

We verified this constraint using a dual-mode plausibility check on a verification batch (N = 50 per grade). Using the real-background optic-disk mask as the primary metric, the final conditioned masks achieved a 0.00% optic-disk overlap failure rate, confirming that the exclusion mechanism is effective under the verification protocol. In contrast, when using an optic-disk detector on the synthetic images, the failure rate was 92%, indicating that standard heuristic detectors are unreliable on GAN outputs due to artifact-induced landmark drift. This justifies using real-background optic-disk estimation as the valid anatomical check in this pipeline.

4.5. LFS-based Top-K selection of synthetic samples (verified)

To reduce low-signal or anatomically inconsistent synthetic outputs, candidate synthetic images were filtered using a lesion-fidelity selection strategy. Each candidate is first subjected to QC checks (e.g., non-empty lesion presence, mask validity, and anatomical constraints). Top-K selection was performed by ranking candidates using the LFS and retaining the highest-scoring samples after QC. This ensures that the synthetic set used for training is not formed by arbitrary retention but by a reproducible scoring-and-ranking mechanism that prioritizes lesion-salient samples.

In the verified robustness run, the final synthetic set used for training contains N = 64 Top-K (LFS-ranked) synthetic images, combined with N = 44 real images to form the full training set (N = 108). This selection step is central to defensibility: it provides a concrete and auditable rationale for why specific synthetic samples are included in the augmentation arm.

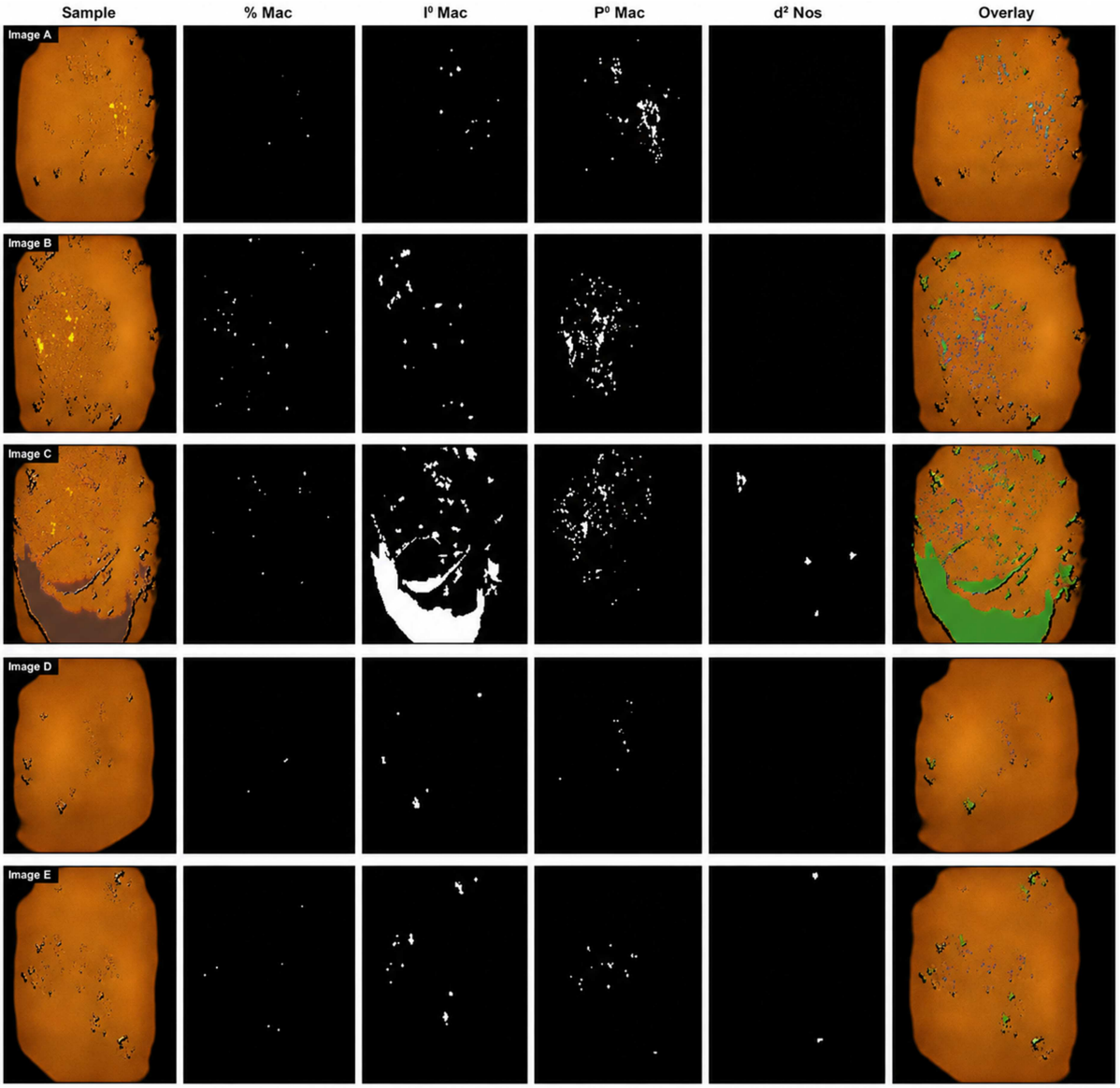
4.6. Downstream robustness: Real-only vs real + synthetic (3-seed study)

To evaluate whether lesion-conditioned synthetic augmentation improves DR grading performance, we conducted a controlled robustness experiment with two arms: (i) real-only, where the classifier is trained on real images only, and (ii) real + synthetic, where the same classifier is trained on real images augmented with the Top-K (LFS-ranked) synthetic images. Both arms use the same backbone (ResNet-50) and identical training settings to ensure parity, including weighted cross-entropy based on inverse class frequency. Experiments were repeated across three random seeds (42, 43, 44) and evaluated on the same held-out

Table 2
Qualitative lesion fidelity and mask–image consistency checklist (reported by inspection)

Check	What is verified	Expected outcome in lesion-conditioned synthesis
Spatial correspondence	Lesions appear where mask indicates	Lesion regions align with mask channels
Morphological consistency	Lesion extent resembles mask shape	No large drift beyond mask boundaries
Background plausibility	Retina landmarks remain coherent	Optic disk/macula/vessels remain plausible
Artifact screening	No obvious synthesis artifacts dominate	Severe artifacts can be excluded via QC

Figure 2
Mask overlays on synthetic outputs (MA/HE/EX/SE channels)



test subset (N = 13) using Quadratic Weighted Kappa (QWK), Macro-F1, and Accuracy.

The results of this robustness evaluation are summarized in Table 3, where performance is reported as mean ± standard deviation across seeds to reflect variance under the small test regime. The augmented arm shows improvement in both QWK and Macro-F1, indicating better ordinal agreement and more balanced class-wise behavior in this constrained evaluation setting. Accuracy remains unchanged on average, which is not unexpected when the test set is very small, and class distribution effects dominate; therefore, we emphasize ordinal and macro-averaged metrics (QWK and Macro-F1) as more informative indicators of robustness than raw accuracy alone.

4.7. Quantitative similarity using Fréchet Inception Distance (FID)

FID is a commonly used metric to quantify similarity between real and synthetic images in a deep feature space by comparing the mean and covariance of feature embeddings extracted

Table 3
Robustness comparison on IDRiD (N = 13 test), mean ± std over 3 seeds

Metric	Real-only	Real + synthetic z (Top-K, LFS)	Δ (Aug – real)
QWK	–0.1835 ± 0.0430	0.0539 ± 0.3401	+0.2374
Macro-F1	0.1763 ± 0.2067	0.1962 ± 0.0983	+0.0199
Accuracy	0.2308 ± 0.2308	0.2308 ± 0.0769	0.0000

from a pretrained network. However, because the verified robustness package in this work is intentionally small, particularly the held-out test subset (N = 13), and because medical-image FID is highly sensitive to preprocessing and limited background diversity, we do not overemphasize absolute FID values as a sole indicator of image quality in this work. Instead, quantitative validation is provided through two complementary, defensible checks: (i) the explicit anatomical plausibility verification (optic-disk exclusion with measured failure rates) and (ii) downstream utility measured through the controlled real-only versus real + synthetic robustness study (mean ± std across seeds). Together, these provide stronger evidence that the synthetic set is not only visually plausible but also carries a diagnostic signal relevant for DR grading.

5. Conclusion

This study presented a lesion-conditioned generative framework for synthesizing DR fundus images using a Pix2Pix-based image-to-image translation model guided by multi-lesion masks. By integrating segmentation-derived structural priors—microaneurysms, hemorrhages, hard exudates, and soft exudates—into the generative process, the framework moves beyond generic realism-oriented GAN augmentation toward pathology-aware synthesis. In contrast to unconditional GANs that rely solely on latent noise and provide limited control over disease-specific content, the proposed approach enforces spatially guided lesion representations, where every synthesized image is explicitly traceable to an input lesion mask, improving the interpretability and controllability of the generated pathology.

Although EyePACS is referenced as the motivating real-world imbalance scenario in DR screening, the verified experiments in this paper were conducted under IDRiD pixel-level annotation constraints, because IDRiD provides the lesion masks required for conditioning. In the verified robustness setup, synthetic augmentation was introduced conservatively, and in an audit-ready manner: the training set consisted of N = 44 real images combined with N = 64 synthetic images generated by the lesion-conditioned Pix2Pix model and retained using a lesion-fidelity selection strategy. Specifically, synthetic candidates were subjected to QC checks, and Top-K selection was performed by ranking candidates using the LFS and retaining the highest-scoring samples after QC, ensuring that the augmentation set is not arbitrary but reproducible and defensible.

Beyond lesion placement fidelity, anatomical plausibility was explicitly enforced and verified using an optic-disk exclusion mechanism. Under the primary verification protocol (real-background optic-disk estimate), the exclusion constraint achieved a 0.00% optic-disk overlap failure rate on the verification batch, while detectors operating on synthetic images exhibited a high failure rate (92%), confirming that ground truth–anchored verification is the valid anatomical check in this setting. Downstream utility was evaluated through a controlled robustness study comparing real-only versus real + synthetic training using identical classifier settings across three random seeds (42/43/44) on a held-out labeled test subset (N = 13). In this constrained evaluation regime, synthetic augmentation improved ordinal agreement and balanced performance (QWK and Macro-F1), while we conservatively report mean ± standard deviation due to the small test size. Overall, this work addresses two key limitations in DR imaging research: the scarcity of lesion-annotated data and the need for augmentation strategies that preserve clinically interpretable pathology. By coupling lesion masks with Pix2Pix-based conditional synthesis, integrating LFS-based Top-K filtering, and enforcing anatomical constraints (optic-disk exclusion), the proposed pipeline provides a modular and computationally practical foundation for pathology-aware augmentation. Future work can scale this framework by expanding the labeled evaluation set, strengthening lesion segmentation under extreme sparsity, and reporting additional quantitative similarity metrics (e.g., FID) under clearly documented sample counts and preprocessing, while preserving the auditability and traceability of mask-conditioned synthesis.

Recommendations

Future research will explore lesion-conditioned diffusion models and GAN–transformer hybrid architectures to further enhance texture realism and structural fidelity under pixel-level conditioning constraints. Additional directions include improving lesion segmentation quality for sparse targets (e.g., microaneurysms and soft exudates), enabling fine-grained control over lesion burden and severity in the conditioning masks, and validating clinical utility through ophthalmologist-in-the-loop assessment. We also plan cross-dataset generalization studies by training or calibrating the synthesis–selection pipeline on IDRiD masks and evaluating downstream robustness on larger imbalance settings (e.g., EyePACS) when compatible conditioning signals or annotations are available.

Acknowledgment

The authors gratefully acknowledge the creators, contributors, and annotators of the IDRiD and EyePACS datasets for providing high-quality retinal fundus images, lesion annotations, and grading information that enabled this study.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The IDRiD datasets that support the findings of this study are openly available at https://idrid.grand-challenge.org/Data/. The EyePACS datasets that support the findings of this study are openly available at https://www.kaggle.com/c/diabetic-retinopathy-detection/data.

Author Contribution Statement

Anju Asokan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Manju Bhaskara Panickar Radha:** Writing – review & editing, Supervision, Project administration.

References

[1] Nakamura, S., Ono, S., & Kawasaki, H. (2021). Flooded road detection from driving recorder: Training deep net for rare event using GANs semantic information. *International Journal of Intelligent Transportation Systems Research*, *19*(1), 1–11. https://doi.org/10.1007/s13177-019-00219-9

[2] Lee, J.-W., Goo, N. H., Park, W. B., Pyo, M., & Sohn, K.-S. (2021). Virtual microstructure design for steels using generative adversarial networks. *Engineering Reports*, *3*(1), e12274. https://doi.org/10.1002/eng2.12274

[3] Islam, S., Aziz, M. T., Nabil, H. R., Jim, J. R., Mridha, M. F., Kabir, M. M., . . . , & Asai, N. (2024). Generative adversarial networks (GANs) in medical imaging: Advancements, applications, and challenges. *IEEE Access*, *12*, 35728–35753. https://doi.org/10.1109/ACCESS.2024.3370848

[4] Heng, Y., Yinghua, M., Khan, F. G., Khan, A., Ali, F., AlZubi, A. A., & Hui, Z. (2025). Survey: Application and analysis of generative adversarial networks in medical images. *Artificial Intelligence Review*, *58*(2), 39. https://doi.org/10.1007/s10462-024-10992-z

[5] Makhlouf, A., Maayah, M., Abughanam, N., & Catal, C. (2023). The use of generative adversarial networks for medical image augmentation. *Neural Computing and Applications*, *35*(34), 24055–24068. https://doi.org/10.1007/s00521-023-09100-z

[6] Qin, X., Bui, F. M., Nguyen, H. H., & Han, Z. (2022). Learning from limited and imbalanced medical images with finer synthetic images from GANs. *IEEE Access*, *10*, 91663–91677. https://doi.org/10.1109/ACCESS.2022.3202560

[7] Krithika alias Anbu Devi, M., & Suganthi, K. (2021). Review of medical image synthesis using GAN techniques. *ITM Web of Conferences*, *37*, 01005. https://doi.org/10.1051/itmconf/20213701005

[8] Osuala, R., Kushibar, K., Garrucho, L., Linardos, A., Szafranowska, Z., Klein, S., . . . , & Lekadir, K. (2023). Data synthesis and adversarial networks: A review and meta-analysis in cancer imaging. *Medical Image Analysis*, *84*, 102704. https://doi.org/10.1016/j.media.2022.102704

[9] Paladugu, P. S., Ong, J., Nelson, N., Kamran, S. A., Waisberg, E., Zaman, N., . . . , & Tavakkoli, A. (2023). Generative adversarial networks in medicine: Important considerations for this emerging innovation in artificial intelligence. *Annals of Biomedical Engineering*, *51*(10), 2130–2142. https://doi.org/10.1007/s10439-023-03304-z

[10] Gao, H., & Ogawara, K. (2020). CGAN-based synthetic medical image augmentation between retinal fundus images and vessel segmented images. In *2020 5th International Conference on Control and Robotics Engineering*, 218–223. https://doi.org/10.1109/ICCRE49379.2020.9096438

[11] Salini, Y., & HariKiran, J. (2022). Deepfakes on retinal images using GAN. *International Journal of Advanced Computer Science and Applications*, *13*(8), 701–708. https://doi.org/10.14569/IJACSA.2022.0130880

[12] You, A., Kim, J. K., Ryu, I. H., & Yoo, T. K. (2022). Application of generative adversarial networks (GAN) for ophthalmology image domains: A survey. *Eye and Vision*, *9*(1), 6. https://doi.org/10.1186/s40662-022-00277-3

[13] S, S. R., J, A., & K, A. M. (2022). A machine learning framework for classification of expert and non-experts radiologists using eye gaze data. In *2022 IEEE 7th International Conference on Recent Advances and Innovations in Engineering*, 314–320. https://doi.org/10.1109/ICRAIE56454.2022.10054277

[14] K, A. R. S., & Shaik, S. S. (2024). Leveraging generative adversarial networks (GANs) for enhancing medical data through augmentation. In *2024 2nd International Conference on Computing and Data Analytics*, 1–12. https://doi.org/10.1109/ICCDA64887.2024.10867363

[15] Karnati, H., & M, S. R. (2024). Enhancing diagnostic accuracy in medical imaging: Integrating GAN-based data augmentation for balanced dataset creation. In *2024 IEEE International Conference on Contemporary Computing and Communications*, 1–7. https://doi.org/10.1109/InC460750.2024.10649058

[16] Shenkut, D., & Bhagavatula, V. (2022). Fundus GAN - GAN-based fundus image synthesis for training retinal image classifiers. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, 2185–2189. https://doi.org/10.1109/EMBC48229.2022.9871771

[17] Chen, Y., Long, J., & Guo, J. (2021). RF-GANs: A method to synthesize retinal fundus images based on generative adversarial network. *Computational Intelligence and Neuroscience*, *2021*(1), 3812865. https://doi.org/10.1155/2021/3812865

[18] Rithani, M., Kumar, R. P., Chatiyode, V., & Anusha, J. S. (2024). Advancing ophthalmic diagnostics: Employing deep learning for precision detection of retinal damage in OCT images. In *2024 International Conference on Data Science and Network Security*, 1–7. https://doi.org/10.1109/ICDSNS62112.2024.10690853

[19] Dash, A., Nayak, D. S. K., & Swarnkar, T. (2025). CiGeN 1.0: A generative AI enabled medical image augmentation pipeline for deep network models. In *2025 International Conference on Ambient Intelligence in Health Care*, 1–6. https://doi.org/10.1109/ICAIHC64101.2025.10956723

[20] Kumar, A., Juliet, S., & Anishin Raj, M. M. (2024). Retinal fundus image synthesis using image segmentation and DCGAN. In *2024 10th International Conference on Communication and Signal Processing, 1089–1094*. https://doi.org/10.1109/ICCSP60870.2024.10544132

[21] Boukhamla, A., Bouziane, M. H., Laib, A., Azizi, N., Rouabhi, R., Merah, A., & Chaib, R. (2023). GANs investigation for multimodal medical data interpretation: Basic architectures and overview. In *2023 International Conference on Control, Automation and Diagnosis*, 1–6. https://doi.org/10.1109/ICCAD57653.2023.10152386

[22] Mohith, V., Raja, K., & Oviya, I. R. (2024). Elevating ocular diagnosis: Harnessing the power of EfficientNet for eye disease classification. In *2024 3rd International Conference on Artificial Intelligence for Internet of Things*, 1–6. https://doi.org/10.1109/AIIoT58432.2024.10574627

[23] Iqbal, T., & Ali, H. (2018). Generative adversarial network for medical images (MI-GAN). *Journal of Medical Systems*, *42*(11), 231. https://doi.org/10.1007/s10916-018-1072-9

[24] Sengupta, S., Athwale, A., Gulati, T., Zelek, J., & Lakshminarayanan, V. (2020). FunSyn-Net: Enhanced residual variational auto-encoder and image-to-image translation network for fundus image synthesis. In *Medical Imaging 2020: Image Processing*, *11313*, 113132M. https://doi.org/10.1117/12.2549869

[25] Akpinar, M. H., Sengur, A., Salvi, M., Seoni, S., Faust, O., Mir, H., . . . , & Molinari, F. (2025). Synthetic data generation via generative adversarial networks in healthcare: A systematic review of image- and signal-based studies. *IEEE Open Journal*

of Engineering in Medicine and Biology, *6*, 183–192. https://doi.org/10.1109/OJEMB.2024.3508472

[26] Middel, L., Palm, C., & Erdt, M. (2019). Synthesis of medical images using GANs. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures: First International Workshop* (pp. 125–134). https://doi.org/10.1007/978-3-030-32689-0_13

[27] Thakur, A., & Thakur, G. K. (2024). Developing GANs for synthetic medical imaging data: Enhancing training and research. *International Journal of Advanced Multidisciplinary Research*, *11*(1), 70–82. https://doi.org/10.22192/ijamr.2024.11.01.009

[28] Qasim, A. B., Ezhov, I., Shit, S., Schoppe, O., Paetzold, J. C., Sekuboyina, A., . . . , & Menze, B. (2020). Red-GAN: Attacking class imbalance via conditioned generation. Yet another medical imaging perspective. In *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, *121*, 655–668.

How to Cite: Asokan, A., & Bhaskara Panickar Radha, M. (2026). Lesion-Conditioned Fundus Image Synthesis for Diabetic Retinopathy Using Pix2Pix Conditional GANs. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62028693