Artificial Intelligence Research and Development
T. Alsinet et al. (Eds.)

doi:10.3233/FAIA240405

SynthRetina: Revolutionizing Fundus Image Analysis Through Synthetic Data Enhancement

Najwa SHARAF [a,1], Hatem A. RASHWAN [a], Saddam ABDULWAHAB [a], and Domenec PUIG [a]

[a] *DEIM, Universitat Rovira i Virgili, Tarragona, Spain*

Abstract.
In an era characterized by the rapid evolution of data-driven applications, the generation of high-quality synthetic data has become increasingly indispensable. This serves as a crucial element for advancing research, development, and ensuring the responsible management of sensitive information. However, the synthesis of fundus images presents unique challenges due to the intricate and highly detailed structures inherent in retinal images. While Generative Adversarial Networks (GANs) show promise in image synthesis, they often encounter training difficulties and struggle to produce truly realistic images. This paper introduces SynthRetina, an innovative system that harnesses the capabilities of GANs to generate lifelike fundus images. SynthRetina amalgamates a generator network and a discriminator network, facilitating the creation of synthetic fundus images with diverse applications across the medical field. The generator network specializes in transforming input fundus images from one class to another, while the discriminator network rigorously evaluates the authenticity of the generated images. SynthRetina effectively addresses the challenge of limited availability of medical data for research and development, offering a solution that enhances data augmentation and improves the performance of fundus image classification tasks. An evaluation of the SynthRetina architecture using a real fundus image dataset demonstrates its ability to produce a more diverse and realistic collection of fundus images compared to other GAN-based methods.

Keywords. Deep learning, Synthesis Date, Generative Adversarial Networks, Fundus Images, Mixture of GANs.

1. Introduction

The demand for high-quality training data has surged in recent years, primarily propelled by the pervasive adoption of data-driven applications. However, the collection and annotation of datasets essential for machine learning applications, especially in domains like medicine, are often laborious and time-intensive. Synthetic data generation emerges as an appealing alternative, offering promise for research, development, and safeguarding sensitive information, particularly in light of challenges related to patient data re-identification and data availability delays [1].

[1]Corresponding Author: Najwa Sharaf. E-mail: najwaabdoalihasan.sharafaldeen@urv.cat

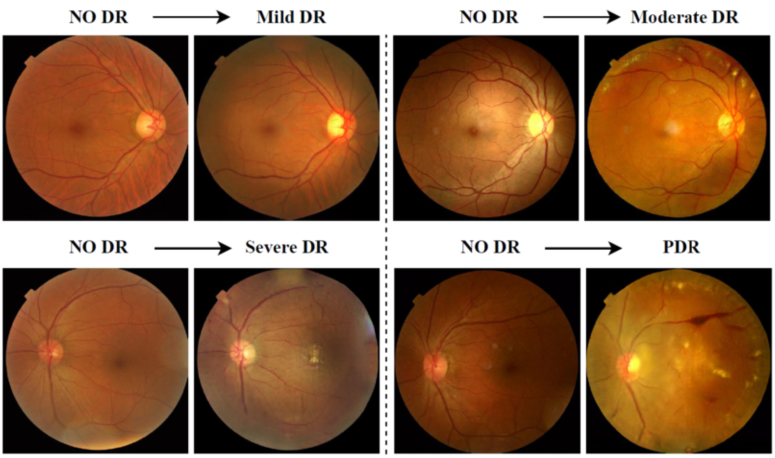

Figure 1. Our algorithm autonomously learns to "translate" images from one collection to another when presented with two image collections, denoted as X and Y. This capability is depicted in the following scenarios: (top-left) translating from the NO DR class to the Mild DR class, (top-right) translating from the NO DR class to the Moderate DR class, (bottom-left) translating from the NO DR class to the Severe DR class, and (bottom-right) translating from the NO DR class to the PDR class. This example effectively showcases our method's ability to generate synthetic fundus images, facilitating seamless transformations across various classes.

However, the collection and annotation of datasets that machine learning approaches require are often labor-intensive and time-consuming, particularly in fields such as medicine, one appealing alternative is rendering synthetic data [2], and given the risks of re-identification of patient data and the delays inherent in making such data more widely available, synthetically generated data is a promising alternative or addition to standard anonymization procedures [3].

Medical imaging is essential for diagnosing and monitoring eye conditions, with fundus images being crucial for detecting and assessing retinal issues. However, collecting well-annotated datasets for training machine learning algorithms is expensive. An alternative is rendering synthetic data [4], which addresses privacy concerns and data availability delays, making it a viable addition to real patient data [5]. Ensuring patient data privacy and security is vital to prevent severe legal and ethical consequences. The inconsistent quality and limited availability of data hinder the advancement of machine learning. The quality of training data significantly impacts the accuracy of machine learning models, and insufficient relevant data restricts their performance and adoption [6].

Deep learning models rely on the data used for their training, and the quality of this data directly affects the accuracy of their decisions and predictions, so a lack of relevant training data limits machine learning models' accuracy and the quality of their output [7]. To address this issue, we propose a system that utilizes the SynthRetina model, which produces class-balanced data samples to mitigate the issue of data scarcity where synthetic data is artificially created to resemble real-world data in its statistical properties, distribution, and structure, and can be used to supplement or replace scarce, expensive, or confidential real-world data in machine learning, computer vision, and data mining applications. Our contributions show that synthetic data improves neural network performance on real medical data. Despite having limitations and challenges, the generation of synthetic data remains a crucial tool for data scientists and researchers and is expected to become even more prevalent with increasing concerns regarding data privacy. The following are the main contributions of this work:

- We introduce SynthRetina, a pioneering system that leverages a combination of Generative Adversarial Networks (GANs) to generate lifelike fundus images with remarkable realism.
- SynthRetina addresses the prevalent challenge of limited availability of medical data for research and development, providing an effective solution that not only enhances data augmentation but also significantly improves the performance of fundus image classification tasks.
- In order to ensure that the synthetic fundus images closely resemble real images, thus maximizing their utility in subsequent classification tasks, we integrate the Structural Similarity Index (SSIM) loss function into the SynthRetina framework.
- Through a comprehensive evaluation conducted on a real fundus image dataset, SynthRetina demonstrates its exceptional capability to generate a diverse array of realistic fundus images, surpassing the performance of existing GAN-based methods in terms of realism and diversity.

Figure 2 and 3 illustrate the proposed SynthRetina model for both training and testing stages. The remainder of the paper is organized as follows: Section 2 summarizes related work in the field. Section 3 presents the methodology for the SynthRetina model. Experimental findings and performance metrics are detailed in Section 4. Finally, Section 5 concludes the work and outlines potential avenues for future research.

2. Related work

The need for high-quality training data has driven researchers to explore innovative methods to tackle the challenges associated with data scarcity, privacy concerns, and the limitations of real-world data collection. In this section, we review the state of the art in data augmentation, Generative Adversarial Networks (GANs), and the broader landscape of synthetic data generation, highlighting their significance in addressing these challenges.

Data augmentation techniques have long been employed in machine learning to enhance the robustness and generalization capabilities of models. These methods involve generating new training samples by applying various transformations to existing data, such as rotations, translations, and mirroring [8,9]. Additionally, in [10], the authors proposed the idea of creating synthetic images from BIM images using a CycleGAN, enabling the transformation of style between different domains and the automated generation of synthetic data.

Recent years have witnessed significant progress in the field of medical image synthesis, particularly in the domain of fundus images. The integration of Generative Adversarial Networks (GANs) into medical image generation has displayed substantial potential, facilitating the creation of synthetic datasets for machine learning model training [11]. While several techniques for data augmentation are available, GAN-based approaches for fundus image synthesis have been relatively scarce. In [12], the authors examined the impact of the quality of synthetic images generated by GANs on the classification performance of models, emphasizing the advantages of GAN-based data augmentation over traditional geometric transformations.

Medical imaging, particularly in ophthalmology, relies heavily on high-quality data for the diagnosis and treatment of various eye conditions. Fundus images, capturing the inner surface of the retina, are instrumental in early detection and monitoring [4]. However, the cost and scarcity of well-annotated fundus image datasets hinder the development and deployment of machine learning solutions in this domain. Synthetic data generation has emerged as a valuable tool to bridge this gap, enabling the training of robust and accurate machine learning models for fundus image analysi [5]. For instance, [13] employed synthesized short-axis CMR images generated using a segmentation-informed conditional GAN to improve the robustness of heart cavity segmentation models.

Our work builds upon the previous research on data augmentation, GANs, and synthetic data. We propose a novel system, SynthRetina, that combines the strengths of these methods to generate high-quality synthetic fundus images. SynthRetina stands out by adopting an innovative strategy that combines multiple GANs, yielding a diverse and high-quality collection of synthetic fundus images.

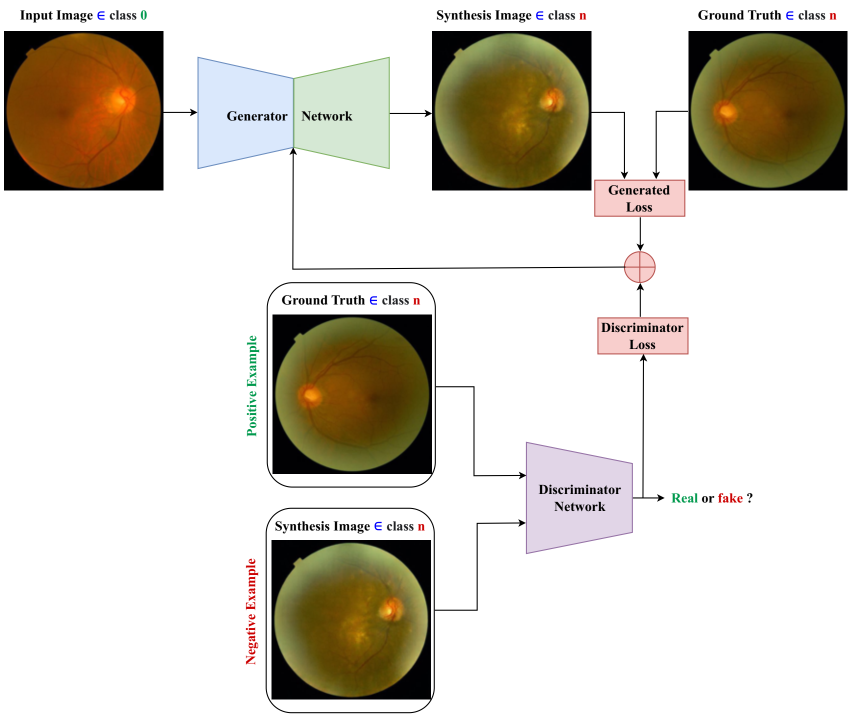

Figure 2. Illustrates the general overview of the proposed SynthRetina model during the training stage. The architecture comprises a Generator network ($\mathcal{N}_G$) responsible for synthesizing retinal fundus images and a Discriminator network ($\mathcal{N}_D$) tasked with discerning between real and synthetic images.

3. Proposed Methodology

In this section, we outline the proposed methodology for addressing the challenge of data scarcity and imbalance in retinal fundus image analysis for Diabetic Retinopathy (DR)

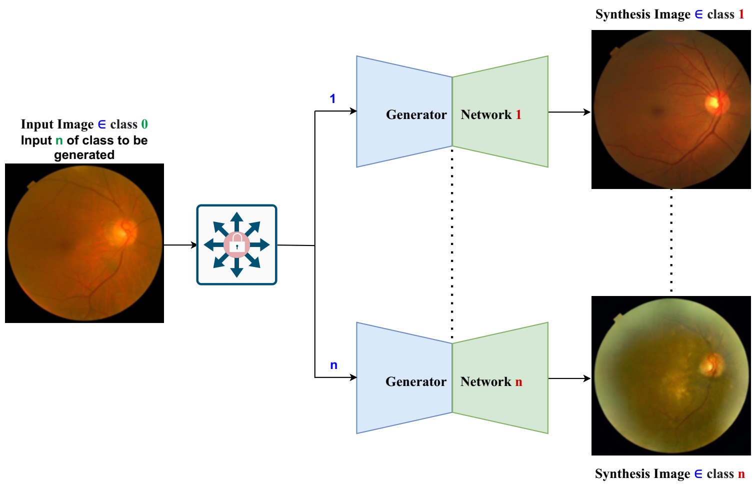

Figure 3. Depicts the general overview of the proposed SynthRetina model during the testing stage. The model consists of a Generator network ($\mathscr{N}_G$), which generate synthetic fundus images based on input from the real fundus images.

grading. We begin by elucidating the problem formulation, wherein we define the objectives and scope of our study. Subsequently, we delve into the architecture of the networks utilized in our proposed approach, highlighting the design choices and considerations. Following this, we provide insights into the training process, including parameter settings and optimization techniques employed. Through this comprehensive overview, we aim to provide a clear understanding of the methodology employed in our study to tackle the challenges inherent in DR grading using retinal fundus images.

3.1. Problem Formulation

To formally define the problem, we consider the task of synthesizing retinal fundus images (X_{synth}) that closely resemble real fundus images (X_{real}). Given a dataset of real fundus images (D_{real}) $\in$ class A, our objective is to train a model that can generate synthetic images (D_{synth}) $\in$ class B.

Mathematically, our goal is to find a mapping function G such that:

$$G : D_{\text{real}} \in A \to D_{\text{synth}} \in B \tag{1}$$

This mapping function G takes the real fundus image $x \in D_{\text{real}}$ from class A as input and produces a synthetic image $X_{\text{synth}} \in D_{\text{synth}}$ from class B.

3.2. Networks Architecture

In our network architecture, we have incorporated two fundamental components: the Generator Network ($\mathscr{N}$G) and the Discriminator Network ($\mathscr{N}$D). These components are pivotal in orchestrating the process of retinal image synthesis, each contributing uniquely to the overall procedure.

3.3. Generator Network

The generator network ($\mathcal{N}_{\mathrm{G}}$) learns the mapping from an input $x \in D_{\text{real}} \subset A$ to the output $x \in D_{\text{synth}} \subset B$. The input to generator network is a fundus image, $x \in$ class A, and it generates a fundus image, $x \in$ class B.

The generator network ($\mathcal{N}_{\mathrm{G}}$) is responsible for synthesizing retinal fundus images from class A to class B. It takes the $x \in D_{\text{real}} \subset A$ as input and generates synthetic images $G(x) \in D_{\text{synth}} \subset B$:

$$\mathcal{N}_{\mathrm{G}}(x \in D_{\text{real}} \subset A) = (D_{\text{synth}} \subset B) \tag{2}$$

To assess the performance of optimizing the training of the network with respect to the structural similarity between the generated image (synthesized image) from a real image and the ground-truth image, we use the Structural Similarity Index (SSIM) as a loss function. SSIM is a comprehensive metric that evaluates the structural similarity between two images. SSIM values closer to 1 indicate a higher similarity between the synthetic and real images.

The SSIM formula is given by:

$$\mathrm{SSIM}(X_{\text{real}}, X_{\text{synth}}) = \frac{(2\mu_{X_{\text{real}}}\mu_{X_{\text{synth}}} + C_1)\cdot(2\sigma_{X_{\text{real}}X_{\text{synth}}} + C_2)}{(\mu^2_{X_{\text{real}}} + \mu^2_{X_{\text{synth}}} + C_1)\cdot(\sigma^2_{X_{\text{real}}} + \sigma^2_{X_{\text{synth}}} + C_2)} \tag{3}$$

where:

- $\mu_{X_{\text{real}}}$ and $\mu_{X_{\text{synth}}}$ are the means of the real and synthesized images, respectively.
- $\sigma_{X_{\text{real}}}$ and $\sigma_{X_{\text{synth}}}$ are the standard deviations of the real and synthesized images, respectively.
- $\sigma_{X_{\text{real}}X_{\text{synth}}}$ is the covariance between the real and synthesized images.
- C_1 and C_2 are constants to stabilize the division by preventing a zero denominator.

These components together assess the structural similarity by comparing luminance, contrast, and structure between the two images.

3.3.1. Discriminator Network

The Discriminator network assumes a pivotal role in the adversarial training framework by discerning between real and synthetic images. It accepts inputs comprising both real images ($x \in D_{\text{real}} \subset B$) and synthetic images generated by the generator network ($X_{\text{synth}} \subset B$), aiming to output a probability score indicative of the authenticity of the input image.

Functioning as a binary classifier, the Discriminator undergoes training to assign high probabilities to real images and low probabilities to synthetic ones. Its architecture typically encompasses convolutional layers followed by fully connected layers, culminating in a single probability score as the final output. The primary objective is to optimize the Discriminator's parameters to minimize the binary cross-entropy loss during the training process.

Mathematically, the Discriminator's output ($\mathcal{N}_{\mathrm{D}}(x)$) for a given image x is defined as follows:

$$\mathcal{N}_{\mathrm{D}}(x) = \begin{cases} \text{High probability} & \text{if } x \in D_{\mathrm{real}} \\ \text{Low probability} & \text{if } x = X_{\mathrm{synth}} \end{cases} \tag{4}$$

Training the Discriminator involves iteratively updating its parameters to accurately differentiate between real and synthetic images. This adversarial interplay, where the Generator and Discriminator engage in oppositional learning, fosters a dynamic equilibrium. Over time, the Generator excels at producing realistic synthetic images, while the Discriminator becomes adept at discerning between real and synthetic data.

3.3.2. Training

In this section, we delve into the training process of the SynthRetina model for synthesizing realistic fundus images. The training procedure involves optimizing the parameters of both the generator and discriminator networks using suitable loss functions. The generator network aims to produce synthetic fundus images that closely resemble real images, while the discriminator network seeks to distinguish between real and synthetic images.

For the Generator network ($\mathcal{N}_{\mathrm{G}}$), the loss function is based on the Structural Similarity Index (SSIM) between the generated images and the real images. This loss encourages the Generator to produce images that closely match the real fundus images:

$$\mathcal{L}_{\mathrm{SSIM}}(X_{\mathrm{real}}, X_{\mathrm{synth}}) = 1 - \mathrm{SSIM}(X_{\mathrm{real}}, X_{\mathrm{synth}}) \tag{5}$$

For the Discriminator network ($\mathcal{N}_{\mathrm{D}}$), the loss function is the binary cross-entropy loss. This loss encourages the Discriminator to accurately classify the images as real or synthetic:

$$\mathcal{L}_{\mathrm{BCE}}(y_{\mathrm{real}}, y_{\mathrm{synth}}) = -\frac{1}{N}\sum_{i=1}^{N}[y_{\mathrm{real}} \cdot \log(D(x_{\mathrm{real}})) + (1 - y_{\mathrm{synth}}) \cdot \log(1 - D(x_{\mathrm{synth}}))] \tag{6}$$

By iteratively updating the networks' parameters based on these loss functions, the SynthRetina model learns to generate high-quality synthetic fundus images that can be used for various medical applications, including Diabetic Retinopathy (DR) grading.

3.4. Parameter settings

We used the ADAM optimizer introduced in [14] to train our model with parameters of $beta_1 = 0.5$, $beta_2 = 0.999$, and an initial learning rate of 0.0002. The optimal combination was with a batch size of 2 and 15 epochs. The PyTorch [15] deep learning framework was used to run all experiments on a 64-bit Core i7-6700, 3.40 GHz CPU with 16 GB of memory, and an NVIDIA GTX 1080 GPU under Ubuntu 16.04. The proposed model's computational cost for the training process is about 2.5 hours per epoch with a 2 batch size. The performance of the online depth map estimation is around 0.028 seconds.

4. Experiments and Results

In this section, we detail the experiments conducted to evaluate the effectiveness of utilizing synthetic images generated by the SynthRetina model in retinal fundus image analysis for Diabetic Retinopathy (DR) grading. We begin by discussing the dataset utilized for training and evaluation, followed by an introduce the evaluation measures used to quantify the performance of the classification model. Finally, we present and discuss the results obtained from our experimental analysis, providing insights into the impact of synthetic images on classification performance.

4.1. Dataset

In our experimental analysis, we employed a dataset obtained from the public DDR dataset [16], which includes 13,673 meticulously curated retinal fundus images used for grading Diabetic Retinopathy (DR). The first step in our process involves resizing all images from 1024 × 1024 to 256× 256. Following this, we utilized the training set to produce synthetic images for training the generator of the MonGANs model. The model's performance is then assessed using the testing set. Synthetic images generated from the testing set are crucial in our thorough evaluation process. These synthetic images are employed to augment the data, aiming to balance the classes and enabling us to assess how well the model performs with and without synthetic data. This approach yields valuable insights into the efficacy of our model to balance the classes.

4.2. Evaluation Measures

This work focuses on creating retinal fundus images, where the influence of synthesized images on classification tasks is pivotal, particularly in scenarios with imbalanced data. In this section, we employ classification evaluation criteria to assess the effectiveness of the generated images. Initially, we train the model using the original images X_{real} and subsequently incorporate augmented images to balance the classes. Evaluation is then performed on the original images X_{real} alongside the augmented generated images X_{synth} after classes balancing. The evaluation metrics encompass Accuracy, Precision, Recall, and F1 Score. These metrics are defined based on the counts of true positive (TP), false positive (FP), true negative (TN), and false negative (FN). these metrics are defined as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{9}$$

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

These classification metrics provide valuable insights into the practical utility of the synthesized images in classification scenarios.

4.3. Results and Discussion

The performance of the SynthRetina model is evaluated through classification metrics. The results are presented and discussed in detail below. In Table 1 provides an overview of the classification evaluation metrics, which consist of Accuracy, Precision, Recall, and F1 Score, for both balanced datasets achieved through traditional methods and synthetic images. A detailed comparative analysis is undertaken to assess the effectiveness of a trained model using original images before balancing, contrasted with the performance of the same model trained on original images augmented with traditional techniques. Moreover, the evaluation extends to the model's performance when trained on original images augmented by images generated through our SynthRetina model.

Table 1. Classification Evaluation Metrics for the test model after training using Balanced Data Utilizing Traditional and Synthetic Images Generated by the SynthRetina Model, Evaluated with Different Metrics Using the VGG16 Model [17].

Techniques	Accuracy	Precision	Recall	F1 Score
Traditional images	0.7721	0.7605	0.7612	0.7641
Synthetic images	**0.7875**	**0.7685**	**0.7864**	**0.7761**

The outcomes depicted in Table 1 reveal a notable advantage for the test model trained on synthetic images compared to the model trained on traditional images. Specifically, the model trained on synthetic images achieves an Accuracy of 78.75%, outperforming the traditional image-based model with an Accuracy of 77.21%. This improvement in accuracy underscores the efficacy of utilizing synthetic images generated by the SynthRetina model. Moreover, a closer examination of the Precision, Recall, and F1 Score metrics further corroborates the superior performance of the synthetic image-based model. The synthetic image-based model exhibits a higher Precision of 76.85%, Recall of 78.64%, and F1 Score of 77.61% compared to the traditional image-based model, which achieved Precision, Recall, and F1 Score of 76.05%, 76.12%, and 76.41% respectively. These findings collectively highlight the effectiveness of the SynthRetina model in generating high-quality synthetic images for training purposes. By augmenting the dataset with synthetic images, the model's ability to generalize to unseen data is significantly improved, leading to enhanced classification accuracy and robustness. Furthermore, the integration of synthetic images into the training data helps alleviate the challenges associated with data imbalance, resulting in more reliable and effective classification outcomes.

5. Conclusion and Future Directions

SynthRetina represents a breakthrough solution for the challenges faced by limited fundus image datasets in the medical realm. By integrating a generator network and a discriminator network, SynthRetina facilitates the generation of synthetic fundus images with multifaceted applications in medicine. The generator network adeptly transforms input fundus images across different classes, while the discriminator network rigorously assesses the authenticity of the generated images. Our evaluation of the SynthRetina architecture using a real fundus image dataset has showcased its capacity to produce a more diverse and realistic assortment of fundus images. The successful development of SynthRetina opens avenues for further research, including optimization for enhanced efficiency and accuracy. Additionally, SynthRetina hold promise beyond fundus images, potentially extending to synthesizing images of other anatomical structures, enriching medical research and healthcare practices.

Acknowledgements: We gratefully acknowledge the financial support provided for this work through the Programa Martí i Franquès grant 2023 PMF-PPF-20.

References

[1] B. van Breugel, T. Kyono, J. Berrevoets, and M. van der Schaar, "Decaf: Generating fair synthetic data using causally-aware generative networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 221–22 233, 2021.

[2] C. Gao, B. D. Killeen, Y. Hu, R. B. Grupp, R. H. Taylor, M. Armand, and M. Unberath, "Synthetic data accelerates the development of generalizable learning-based algorithms for x-ray image analysis," *Nature Machine Intelligence*, vol. 5, no. 3, pp. 294–308, 2023.

[3] J. Yoon, L. N. Drumright, and M. Van Der Schaar, "Anonymization through data synthesis using generative adversarial networks (ads-gan)," *IEEE journal of biomedical and health informatics*, vol. 24, no. 8, pp. 2378–2388, 2020.

[4] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," *arXiv preprint arXiv:1709.01872*, 2017.

[5] J. Dahmen and D. Cook, "Synsys: A synthetic data generation system for healthcare applications," *Sensors*, vol. 19, no. 5, p. 1181, 2019.

[6] A. Desai, C. Freeman, Z. Wang, and I. Beaver, "Timevae: A variational auto-encoder for multivariate time series generation," *arXiv preprint arXiv:2111.08095*, 2021.

[7] J.-H. Wu, T. A. Liu, W.-T. Hsu, J. H.-C. Ho, and C.-C. Lee, "Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis," *Journal of medical Internet research*, vol. 23, no. 7, p. e23863, 2021.

[8] H. Salehinejad, S. Valaee, T. Dowdell, and J. Barfett, "Image augmentation using radial transform for training deep neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3016–3020.

[9] P. Oza, P. Sharma, S. Patel, F. Adedoyin, and A. Bruno, "Image augmentation techniques for mammogram analysis," *Journal of Imaging*, vol. 8, no. 5, p. 141, 2022.

[10] Y. Hong, S. Park, and H. Kim, "Synthetic data generation for indoor scene understanding using bim," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 37. IAARC Publications, 2020, pp. 334–338.

[11] M. Gong, S. Chen, Q. Chen, Y. Zeng, and Y. Zhang, "Generative adversarial networks in medical image processing," *Current Pharmaceutical Design*, vol. 27, no. 15, pp. 1856–1868, 2021.

[12] X. Qin, F. M. Bui, H. H. Nguyen, and Z. Han, "Learning from limited and imbalanced medical images with finer synthetic images from gans," *IEEE Access*, vol. 10, pp. 91 663–91 677, 2022.

[13] Y. Al Khalil, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images," *Medical Image Analysis*, vol. 84, p. 102688, 2023.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration," 2017.

[16] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.