Heliyon 9 (2023) e18773

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Generating OCT B-Scan DME images using optimized Generative Adversarial Networks (GANs)

Aditya Tripathi, Preetham Kumar *, Veena Mayya *, Akshat Tulsani

Department of Information & Communication Technology, Manipal Institute of Technology, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India

ARTICLE INFO

Keywords:
Diabetes mellitus
Diabetic macular edema (DME)
Optical Coherence Tomography (OCT) B-Scan
Generative Adversarial Network (GAN)

ABSTRACT

Diabetic Macular Edema (DME) represents a significant visual impairment among individuals with diabetes, leading to a dramatic reduction in visual acuity and potentially resulting in irreversible vision loss. Optical Coherence Tomography (OCT), a technique that produces high-resolution retinal images, plays a vital role in the clinical assessment of this condition. Physicians typically rely on OCT B-Scan images to evaluate DME severity. However, manual interpretation of these images is susceptible to errors, which can lead to detrimental consequences, such as misdiagnosis and improper treatment strategies. Hence, there is a critical need for more reliable diagnostic methods. This study aims to address this gap by proposing an automated model based on Generative Adversarial Networks (GANs) to generate OCT B-Scan images of DME. The model synthesizes images from patients' baseline OCT B-Scan images, which could potentially enhance the robustness of DME detection systems. We employ five distinct GANs in this study: Deep Convolutional GAN, Conditional GAN, CycleGAN, StyleGAN2, and StyleGAN3, drawing comparisons across their performance. Subsequently, the hyperparameters of the best-performing GAN are fine-tuned using Particle Swarm Optimization (PSO) to produce more realistic OCT images. This comparative analysis not only serves to improve the detection of DME severity using OCT images but also provides insights into the appropriate choice of GANs for the effective generation of realistic OCT images from the baseline OCT datasets.

1. Introduction

Diabetic Macular Edema (DME) is an ocular complication commonly affecting patients diagnosed with diabetes. This condition arises when fluid amasses within the inner and outer plexiform layers of the retina, leading to consequential retinal thickening. This fluid accumulation hinders the eye's ability to discern shapes and attenuates the clarity of distant objects, a capability measured by visual acuity. As DME progresses, the patient's visual acuity deteriorates, making it increasingly challenging to focus. More specifically, DME is a diabetes-related complication that triggers fluid accumulation in the macula - the central part of the retina responsible for sharp and straight-ahead vision. The fovea, located at the center of the macula, is particularly affected by this condition, which leads to gradual vision loss over several months. If not identified and managed in its nascent stages, DME could potentially culminate in permanent visual impairment.

* Corresponding authors.
E-mail addresses: preetham.kumar@manipal.edu (P. Kumar), veena.mayya@manipal.edu (V. Mayya).

https://doi.org/10.1016/j.heliyon.2023.e18773
Received 3 June 2023; Received in revised form 26 July 2023; Accepted 27 July 2023

Available online 2 August 2023

Optical Coherence Tomography (OCT) is a noninvasive technique capable of generating high-resolution images of a patient's retina. It operates by gauging the amount of dim red light reflected off the retina, employing a low-powered laser to produce three-dimensional, color-coded images of the retinal structure. At present, clinicians utilize a range of laboratory and imagery biomarkers derived from OCT images to assess the presence of DME. Laboratory biomarkers typically encompass factors such as obesity, age, gender, and vascular risk conditions. Imagery biomarkers, in contrast, include features such as the Disorganization of Retinal Inner Layers (DRIL), the Ellipsoid Zone (EZ), the Foveal Avascular Zone (FAZ), the count of inner hyperreflective foci, and the height of intraretinal fluid, among others. However, the manual evaluation of DME is subject to potential errors and inconsistencies, which may lead to irreversible visual impairment. A generative model based on the baseline OCT B-Scan (henceforth referred to as OCT) images can assist clinicians in correctly diagnosing and treating DME by showing how the condition progresses in the patient's eyes.

The Generative Adversarial Network (GAN) is a pivotal framework in the realm of computer vision, instrumental in the generation of synthetic, yet high-quality images. A GAN consists of two integral components: a generator network and a discriminator network. The generator network is tasked with producing images of a predefined size, while the discriminator network is designed to classify these images as either artificially generated or real. The input to the generator network is a noise vector, the size of which is determined by a hyperparameter. Depending on this input, the generator generates a fake image. Simultaneously, the discriminator network ingests either a generated or a real image as its input, subsequently outputting a classification of whether the image is real or fake. The discriminator network's loss is calculated using a straightforward binary classification loss function, which categorizes images into one of two classes: real or generated. Conversely, the loss of the generator network is contingent on the discriminator network's output. Notably, GANs diverge from traditional machine learning algorithms as they are generative models, primarily designed to create new synthetic data points. This makes them particularly suitable for unsupervised learning tasks, where labeled data is either scarce or challenging to obtain. GANs are highly valued for their ability to yield high-quality, photorealistic outcomes in the synthesis of images, videos, and music.

In this study, we undertake an investigation and comparison of various GAN-based systems designed to generate predictive OCT images of DME using baseline OCT images. The implementation of such a model holds the potential for significant enhancements in the determination of DME severity in patients. The scope of our analysis encompasses images generated by five distinct GAN variants: Deep Convolutional GAN (DCGAN), Conditional GAN (CGAN), CycleGAN, StyleGAN2, and StyleGAN3.

The remaining sections are organized as follows: Section 2 provides an extensive review of existing DME detection methodologies that employ OCT images, as well as a review of recent developments in GAN technology. Section 3 delineates the preprocessing steps and the GAN architectures utilized for generating DME-afflicted OCT images. Section 4 details the evaluation process for the GAN models tasked with OCT image generation. Section 5 summarizes the experimental study, discusses the DME detection system, its limitations, and outlines future research directions. Finally, Section 6 concludes the study.

2. Literature survey

Liu et al. [1] implemented a GAN to segment retinal layers from OCT imagery. Owing to the limited availability of OCT image datasets for the classification of various eye conditions, GANs have been employed to generate OCT images to train diverse classification networks. These networks are designed to detect conditions such as Age-Related Macular Degeneration (AMD), DME, and Chronic Neovascularization (CNV) [2,3]. Smitha et al. [4] proposed an end-to-end model, based on GANs, designed to segment retinal layers from OCT images and subsequently classify those images as either indicating a certain disease or being normal. Rachmadi et al. [5] utilized a GAN model to predict the evolution of white matter hyperintensities in small vessel disease. Xia et al. [6] forecasted future brain images using a joint distribution of a patient's age and brain morphology. Schlegl et al. [7] constructed a rapid, unsupervised anomaly detection system utilizing a GAN. This system is capable of identifying abnormalities within an image, using only healthy OCT images as a reference.

Suciu et al. [8] presented a comprehensive inventory of the laboratory biomarkers commonly used in determining the severity of DME. These biomarkers encompass obesity, age, gender, and vascular risk factors. In particular, the average body-mass index (BMI) for patients suffering from DME is generally noted to be between 29 and 30 [8,9]. Elevated blood pressure is also associated with a higher severity of DME [8]. Moreover, Klein et al. [10] attempted to link cholesterol levels with DME severity. Their research indicated a positive correlation between the severity of DME and the serum levels of total cholesterol. Conversely, an inverse relationship was identified between DME severity and serum levels of high-density lipoprotein (HDL) cholesterol. On the pharmacotherapy front, Ehlers et al. (2022) investigated the ongoing efficacy of anti-VEGF treatment and intravitreal corticosteroid interventions in mitigating DME [11]. However, Rana et al. [12] underscored the potential drawbacks of relying on laboratory biomarkers, particularly given their invasive nature and inconsistent application in DME evaluation. As an alternative, imaging biomarkers have emerged as a reliable means of assessing DME severity. The advent of OCT, which generates high-definition retinal images, has greatly simplified the identification and reliability of imaging biomarkers.

2.1. Use of imagery biomarkers in grading DME severity

Endo et al. [13] deployed a multivariate analysis to elucidate the relationship between OCT derived features such as DRIL, the count of Hyperreflective Foci (HRF), intraretinal fluid, and visual acuity. Their study discovered that DRIL, EZ disruption, and FAZ circularity significantly influenced visual acuity. The researchers observed that the length of DRIL increased with escalating EZ disruption and vice versa, while the length of DRIL increased as FAZ circularity decreased. This reinforced DRIL as a compelling candidate for a biomarker in DME grading. This importance of DRIL as a critical biomarker in DME was also affirmed by Sun et

al. [14]. Furthermore, Sun et al. [15] revealed that a horizontal extent of DRIL greater than 50% in the 1 mm central foveal area correlated with worse visual acuity (VA) in the eyes.

Nagai et al. [16] demonstrated that both the number and area of cystoids reduced in response to DME treatment. In a similar vein, Yalın et al. [17] reported that (i) there is a 58% likelihood that visual acuity is less than 20/60 in eyes where the horizontal cyst diameter is equal to or greater than 450 μm; (ii) for eyes with a horizontal cyst diameter smaller than 450 μm, the likelihood that visual acuity is equal to or greater than 20/60 is 73%; (iii) eyes with a vertical cyst diameter greater than or equal to 300 μm have a 62% likelihood of VA being less than 20/60; and (iv) eyes with a vertical cyst diameter less than 300 μm have a 69% likelihood that VA is equal to or greater than 20/60.

Besides these biomarkers, there are several other imaging biomarkers in OCT images that help in detecting the severity of DME. Endo et al. [13] showed that DRIL has a positive correlation with EZ, while there is a negative correlation between DRIL and FAZ. Hence, DRIL alone is capable of also indicating if EZ and FAZ have worsened or not. Similarly, Arthi et al. [18] showed hyperreflective foci (HRF) have a positive correlation with subretinal fluid (SRF).

2.2. Generative Adversarial Networks (GANs)

GANs, introduced by Goodfellow et al. [19] in 2014, constitute a pioneering framework for estimating generative models via an adversarial process. This framework incorporates a generator network and a discriminator network. The generator network's primary function is to generate artificial data points mirroring the data distribution, while the discriminator network estimates the likelihood of the generator's output data being sourced from the training dataset rather than synthesized. This paradigm has found extensive application across a broad spectrum of challenges, particularly in the computer vision domain. The potential of GANs extends to numerous applications, such as dataset augmentation [19,20], realistic human image generation [21], cartoon character generation [22], image-to-image translation [23,24], text-to-image translation [25,26], photograph editing [27], and 3D object generation [28]. Chen et al. [29] comprehensively discussed the diverse studies employing GANs to construct medical image augmentation models. Their work encompassed an in-depth examination of the benefits associated with these models, various evaluation metrics, and other pertinent aspects.

Since their inception in 2014, GANs have undergone numerous enhancements to improve the fidelity of generated data and tailor the base model for a wide array of applications. Radford et al. [20] proposed DCGAN, where the generator and discriminator networks are convolutional networks. This structure allowed for the integration of recent advancements in Convolutional Neural Networks (CNNs), such as leveraging different state-of-the-art networks as generators or classifiers to tackle intricate problems. Zhu et al. [24] and Isola et al. [23] built an image-to-image translation model that uses some prior knowledge to output a certain type of desired image. Isola et al. [23] created the Pix2Pix model, which uses image pair mapping to generate one type from an image given another as input. For example, converting line drawings into finished paintings. Zhu et al. [24] on the other hand, the built CycleGAN relaxes the need for image-to-image mapping. Saliman et al. Lin et al. [30] introduced StyleGAN, a variant of GANs capable of generating high-resolution, customizable human portraits. They achieved this by leveraging different noise samples at each upsampling level, thereby offering users greater control in generating customizable images. Li et al. [31] proposed the concept of federated learning with least squared GANs (LSGANs), resulting in a novel federated deep generative learning framework, Fed-LSGAN, for renewable scenario generation. This model efficiently captures the spatial-temporal characteristics of renewable powers, enhances training stability, and improves generation quality, all the while generating high-quality scenarios without compromising generation quality via model parameter transfer. Their simulation results indicate the superior performance of the proposed method over centralized approaches. [32] inferred that GANs are not only useful for generating new data points but also for enhancing other tasks, such as developing a classifier on a semi-labeled dataset.

In this study, the classifier is trained simultaneously to classify the images and detect a data point as fake or real, while a generator synthesizes new data points that can be used to train the classifier even better. DME is a disease that is seen in diabetic patients, and during this disease, the patient's visual acuity reduces sharply. If not diagnosed and treated at an early stage, the disease can even cause the patient to lose his/her visual ability completely. On the basis of the reported baseline OCT, there is a dearth of research predicting the progression of the DME condition. Despite the application of GAN in forecasting the future status of many diseases, a comparable technique in this field is still lacking.

3. Materials and methods

In this study, we conduct an analysis to evaluate the efficacy of five distinct GAN models in their ability to accurately generate OCT images of patients with DME based on their baseline OCT images. The GAN models assessed in this study include DCGAN, CGAN, CycleGAN, StyleGAN2, and StyleGAN2. Fig. 1 delineates the end-to-end methodology for deriving OCT images from the patient's baseline OCT images. The proposed method encompasses two primary stages: the GAN selection phase and the hyperparameter optimization phase.

In the typical workflow of a GAN, a noise vector and baseline OCT image are supplied to the generator (G) of the GAN. This generator subsequently generates a predictive OCT image. This synthesized image, or the original OCT image, is then randomly introduced to the discriminator (D), which determines whether the supplied image is real or fake. Upon identification of the most effective GAN model among those investigated, the model's hyperparameters undergo further refinement using the Particle Swarm Optimization (PSO) algorithm. The PSO algorithm [33–35] is employed to fine-tune the selected GAN model across a specified number of iterations to enhance the realism of the synthesized DME images.

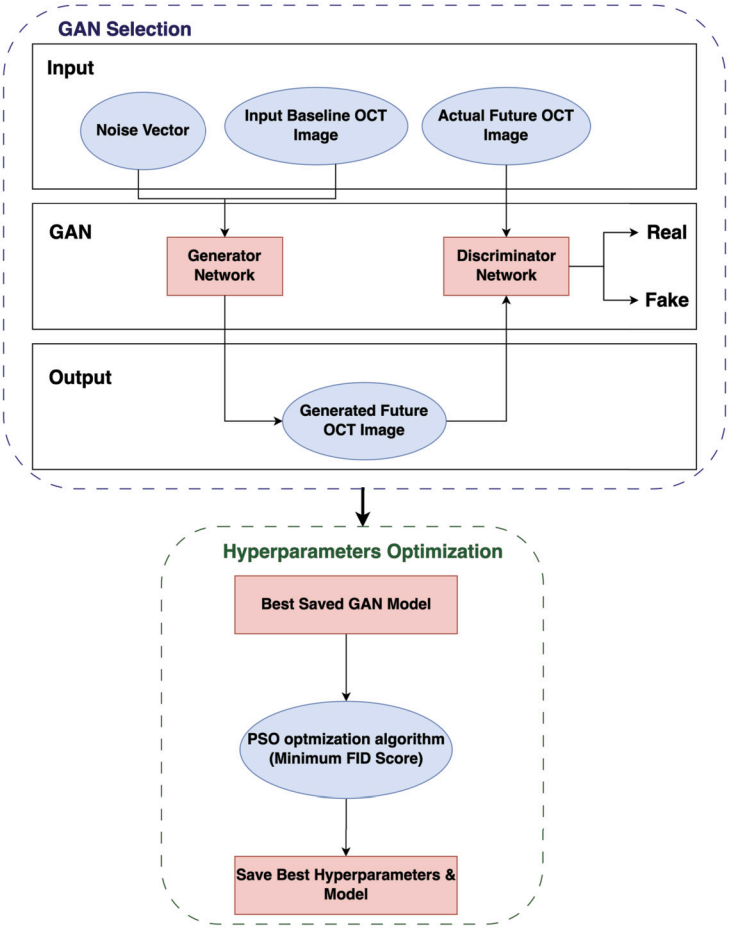

Fig. 1. The general pipeline of proposed methodology.

3.1. Image preprocessing

In our investigation, all images were resized to 496×496 pixels, with pixel values scaled according to a normal distribution to mitigate image noise. We adopted TensorFlow's default transformation mechanism for image normalization. This approach was chosen to effectively eliminate outliers from the images while preserving all essential information. In addition to applying scale and resizing to the input images that were causing the generated image to have a white rectangular area that was irrelevant and not required were also removed.

3.2. GAN

Following the image preprocessing, each image is converted into a vector form. The noise vector is subsequently concatenated with this image vector before being fed into the generator network within the GAN. Our study utilizes five distinct types of GANs. Each of these GANs is trained on the provided dataset, following which their performances are evaluated and compared against one another.

3.3. Deep convolutional GAN (DCGAN)

Radford et al. [20] introduced a GAN with CNN, called DCGAN. Some salient features of DCGAN are as follows:

- The architecture solely relies on convolutional layers, abstaining from the use of pooling layers or fully connected layers.
- Batch normalization is applied after both the hidden and input layers.
- The generator network employs a Rectified Linear Unit (ReLU) activation function for its hidden and input layers, while the hyperbolic tangent (Tanh) function is used in the final output layer.
- The discriminator network utilizes Leaky ReLU activation functions across all its layers.

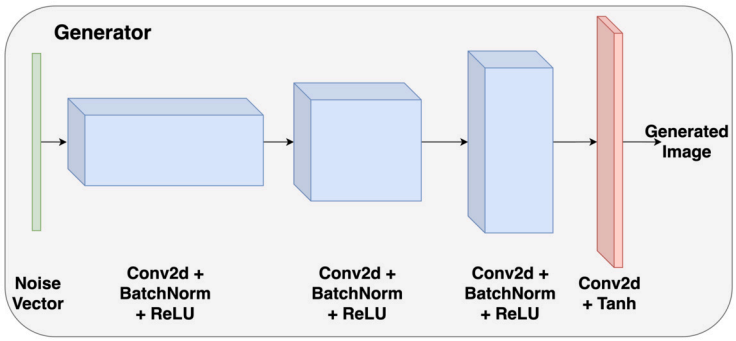

Fig. 2. DC GAN generator architecture.

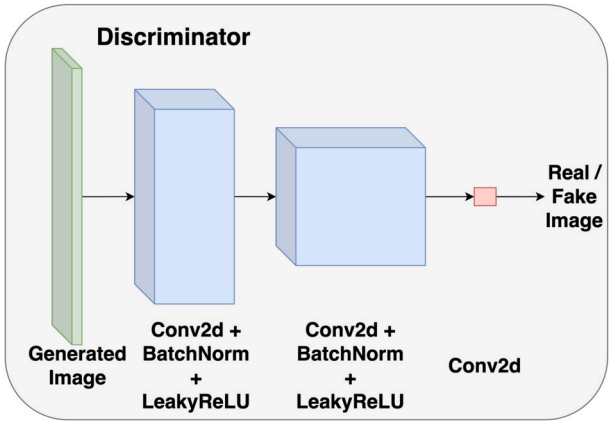

Fig. 3. DC GAN discriminator architecture.

The input to the DCGAN solely consists of a noise vector; the baseline OCT image does not serve as an input to the generator network. In the implementation of DCGAN for this study, the noise vector's size is set at 32. The generator network includes three hidden blocks, with each block constituted by a Conv2D block, followed by batch normalization, and a ReLU activation function. In contrast, the output block contains a Conv2D block paired with a Tanh function. The image generated by the generator network is fed into the discriminator network, which comprises two hidden blocks. Each block includes a Conv2D layer, batch normalization, and a Leaky ReLU activation function. The final block contains only a Conv2D layer. The DCGAN network utilized in this study is visualized in Figs. 2 and 3.

3.4. Conditional GAN

Mirza et al. [36] introduced a CGAN that controls the image generated under certain conditions. This innovation addresses the lack of control over the images generated by the basic generative network. It works by attaching extra information with the noise input to the generative network and attaching the same information again with the discriminator input. A noise vector of size 16 concatenated with a flattened baseline OCT image is given to the generator. The generator network comprises six hidden blocks, each block comprising a Conv2D block followed by batch normalization coupled with ReLU, while the output block comprises a Conv2D block along with Tanh. The image generated is given to the discriminator network, which consists of 3 hidden blocks that comprise a Conv2D layer, batch normalization function, and leaky ReLU, while the final block consists of only a Con2D layer. Figs. 4 and 5 shows the CGAN network used in this study.

3.5. CycleGAN

Zhu et al. [24] developed an unsupervised GAN model known as CycleGAN, which is primarily employed for image transformation tasks. Notably, CycleGAN has the capability to translate a horse image into a zebra image, and vice versa. This model involves the training of two generator and two discriminator networks. In essence, one generator and discriminator pair is trained to convert an image from one class to another (e.g., from horse to zebra), while the other pair is tasked with transforming the image back to its original class (e.g., from zebra back to horse). One unique characteristic of CycleGAN is its generator loss computation, which is derived from the summation of three individual losses:

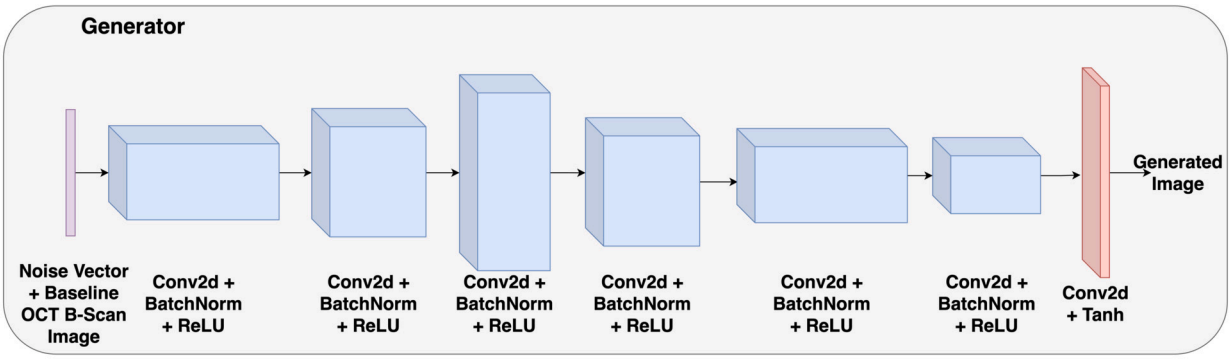

Fig. 4. CGAN generator architecture.

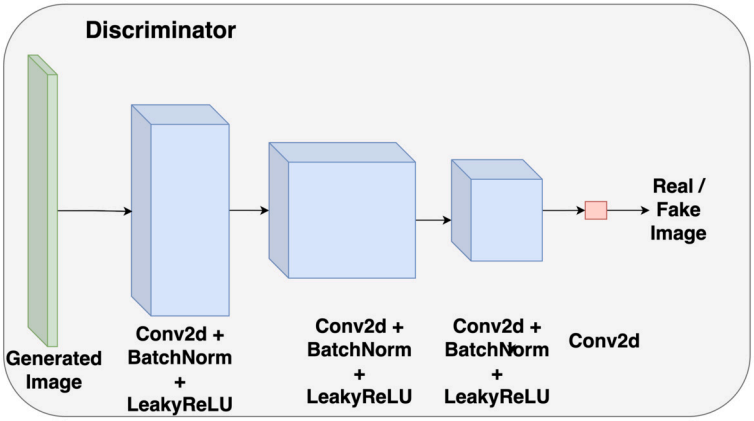

Fig. 5. CGAN discriminator architecture.

- The Mean Squared Error (MSE) loss: This is the adversarial loss calculated from the discriminator network's predictions. This loss is typically found in all GANs.
- L1 cycle loss: This measures the difference between the input image and the image obtained after transforming it to a different class using one generator and then reverting it back to the original class using another generator.
- L1 identity loss: This measures the difference between the input image and the image produced after applying a generator designed to translate an image from another class back to the input image's class. In the case of the horse-to-zebra translation, the identity loss is the L1 loss between the initial horse image and the image generated by applying the zebra-to-horse generator to this image.

Figs. 6 and 7 depict the CycleGAN's generator and discriminator networks, respectively. The baseline OCT image is transformed into a vector and then supplied to the generator network. The generator network comprises 2 downsampling convolution blocks, each including a Conv2D layer, an instance normalization function, and a ReLU activation function. These are followed by 9 residual blocks, each containing 2 convolution blocks, and the output is then concatenated with the residual block's input. Subsequently, two upsampling convolution blocks are used to generate the final image. The discriminator consists of 2 hidden blocks, with batch normalization replaced by instance normalization in each block. The output layer is solely composed of a Conv2D layer.

3.6. StyleGAN

In StyleGAN, the generator function's input deviates from the conventional method of using a random noise vector. Instead, a random point from the latent space is transferred to a mapping network, which operates as an autonomous convolutional network. The vector that the mapping network produces at the end of each convolutional block modulates the resulting feature maps. After the convolutional operation, the noise vector is supplemented with the newly formed feature map from the convolutional layer, and the final feature map is subsequently normalized. This process of modulation and normalization is designated as Adaptive Instance Normalization (AdaIN). The application of AdaIN provides an enhanced level of control over the types of images generated by the generator network, as the output of the mapping function is passed to each convolutional block within the generator network.

Fig. 8 illustrates the StyleGAN2 generator network utilized in our study. StyleGAN2 has integrated modulation and convolution operations (as demonstrated in Fig. 8). Since modulation scales the feature map according to the style vector (the output of the

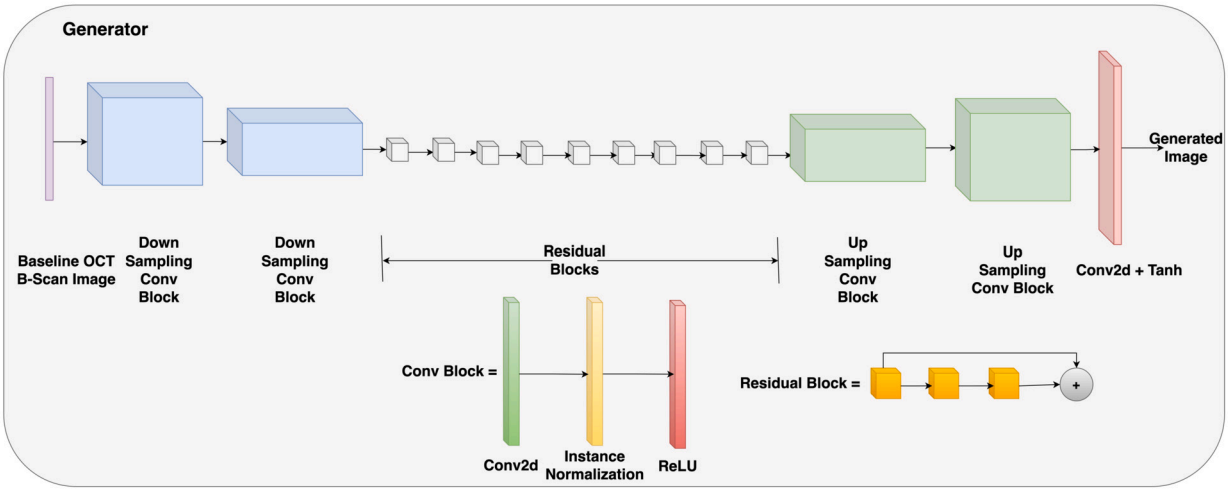

Fig. 6. CycleGAN generator architecture.

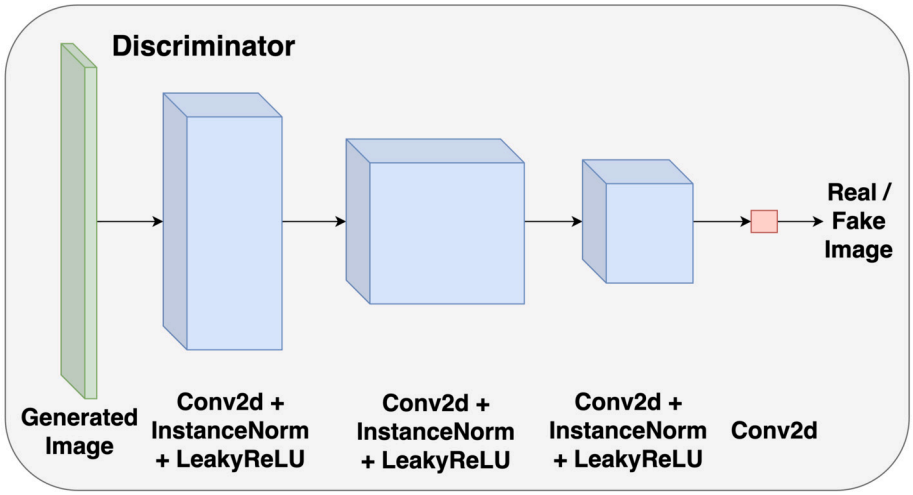

Fig. 7. CycleGAN discriminator architecture.

mapping network), the same result can be achieved by simply scaling the weights used in the convolutional layer. Furthermore, after the convolution operation, the noise vector is added, and weight demodulation is performed instead of normalization. This aims to counteract the effect of the style vector on the new feature map prior to its transmission to the next convolutional block. The inclusion of ResNets in StyleGAN2 has contributed to an improvement in the quality of generated images. The replacement of normalization with demodulation has effectively mitigated the prominent 'water droplet' artefact issue in StyleGAN, as demodulation simply negates the influence of the style vector rather than altering the actual feature map.

StyleGAN3 endeavors to rectify the issue of 'Texture Sticking,' where specific pixels in the image remain stationary during the transition from one image to another, resulting in unnatural appearing image transitions. The root cause of Texture Sticking is identified as positional reference, which is introduced due to factors like per-pixel noise inputs and aliasing. Aliasing is a phenomenon where the output signal from sampling deviates from the original input signal. In response, StyleGAN3 has redesigned the network layers to accommodate continuous feature maps, as aliasing is substantially suppressed in signals existing in continuous spaces compared to signals in discrete spaces. Additionally, it has eradicated per-pixel noise inputs to eliminate positional references.

3.7. Hyperparameters optimization

The role of GAN's hyperparameters is pivotal in the generation of realistic images. We initialize the process by loading the optimal weights of the selected GAN. The Particle Swarm Optimization (PSO) algorithm [33–35] is then utilized to fine-tune the best StyleGAN2 model for a predefined number of iterations. The proposed PSO methodology begins with the random hyperparameters of StyleGAN2, which are represented as PSO particles. These particles are set to navigate randomly towards the optimal hyperparameters. At each stage, we aim to obtain the lowest FID [37] score ever achieved by each hyperparameter and the lowest FID found

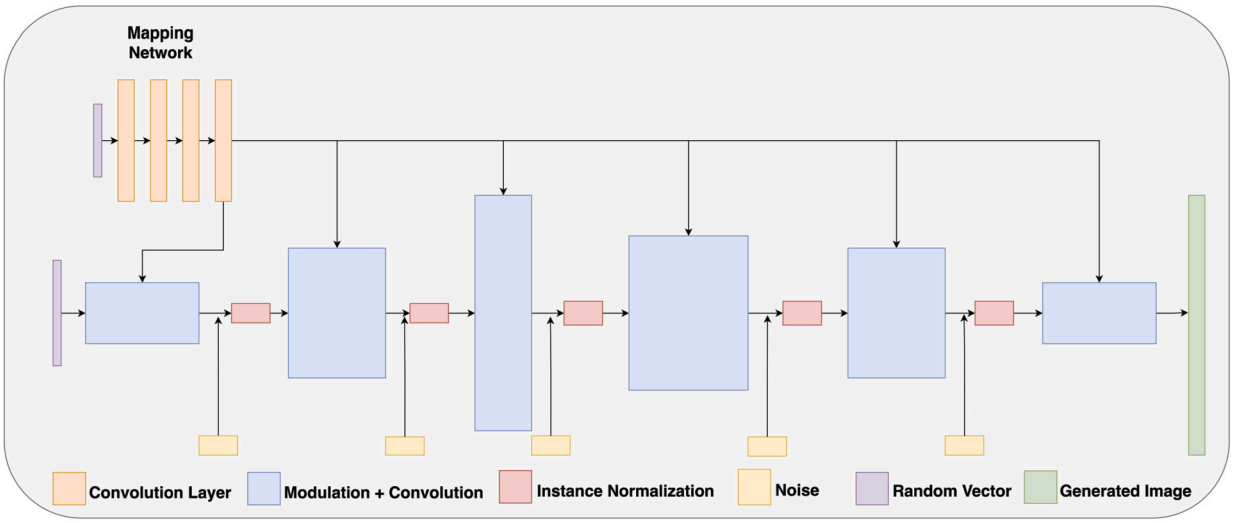

Fig. 8. StyleGAN2 generator architecture.

by the entire particle swarm. This operation is visualized in the flowchart provided in Fig. 9. After a certain number of iterations, the smallest FID score achieved by the particle swarm is considered as the lowest. Consequently, the optimal hyperparameters pertianing to the DME dataset are identified.

The fitness function of the PSO algorithm is evaluated based on the FID score, which is computed between 5000 images generated by the GAN model and the original images. The FID serves to compare the distribution of generated images to that of real images. The identified optimal hyperparameters, along with the model weights, are stored in a JSON file for further use.

4. Results

The subsequent sections detail the datasets and metrics utilized in this study and the results obtained from implementing different GANs using these metrics.

4.1. Dataset

Our experiment utilized the Duke Eye dataset [38] to train and test various GAN models. This dataset comprises spectral-domain OCT images from 45 subjects, which included 15 normal individuals, 15 suffering from DME, and 15 affected by dry age-related macular degeneration. Given that this study focuses on DME, we used the OCT images from the 15 subjects suffering from this condition.

Each subject's data includes 80 to 100 images representing the disease progression timeline through OCT images. The original dataset contained image noise; for example, many images were taken from different angles, resulting in large white and empty areas on one of the four sides. These were removed before feeding the images into the training network.

A total of 15 patients diagnosed with DME were studied. Of these, data from 12 patients was employed to train the model, while the images from the remaining three patients were set aside for testing purposes. Consequently, a comprehensive dataset comprising 2033 images (1115 Normal, 918 DME) was utilized for the training of the GAN models.

4.2. Evaluation metrics

Binary Cross-Entropy Loss (BCE Loss) was used to train DCGAN and CGAN. It can be given as shown in Equation (1).

$$\text{Binary Cross Entropy} = \frac{-1}{N} \times \sum_{i=1}^{N} y_i \times \log(p(y_i)) + (1 - y_i) \times \log(1 - p(y_i)) \tag{1}$$

Where y_i = ground truth (image provided to the discriminator network is real/generated), $p(y_i)$ = discriminator prediction, N = number of samples. L1-loss and Mean Squared Error were used to evaluate the performance of the CycleGAN. They can be given by Equations (2) and (3).

$$\text{L1 Loss} = \frac{1}{N} \times \sum_{i=1}^{N} |y_i - p(y_i)| \tag{2}$$

$$\text{MSE} = \frac{1}{N} \times \sum_{i=1}^{N} (y_i - p(y_i))^2 \tag{3}$$

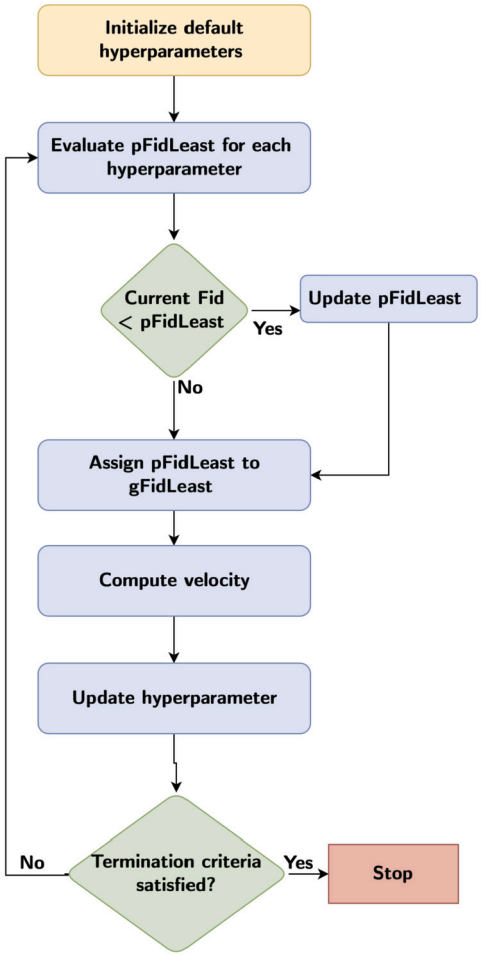

Fig. 9. PSO for hyperparameters optimization.

However, the same evaluation metric is to be used when comparing the performance of five GANs. Hence, L1-loss, and Mean Squared Error were used for training the CycleGAN, but to compare it with other GANs Binary Cross-Entropy loss was calculated for CycleGAN between the generated images and the ground truths.

StyleGAN uses Frechet Inception Distance (FID) as one of the evaluation metric and is given by Equation (4). In this study, FID_N refers the FID score between ground truth images and N generated images. The element-wise means of the inception features for real and generate N images are μ_{real} and $\mu_{generate\_N}$, respectively. The covariance matrices for the real and generated inception feature vectors are σ_real and $\sigma\_generate\_N$, respectively. Tr is used to represent the matrix's trace.

$$\text{FID\_N} = \|\mu_{real} - \mu_{generate\_N}\|^2 - T_r(\sigma\_real + \sigma\_generate\_N - 2\sqrt{\sigma\_real \times \sigma\_generate\_N}) \tag{4}$$

4.3. Comparison of performance of different GANs

In our study, we investigated five GAN models: DCGAN, CGAN, CycleGAN, StyleGAN2, and StyleGAN3. The discriminator and generator losses obtained during the training phase are presented in Figs. 10 (A–E). The decision to halt the training process was made by assessing the visual quality of the generated OCT images, and the number of iterations required varied accordingly.

The result of the comparison with different GAN models is shown in Table 1. Assessing the quality of a GAN model necessitates a broader analytical framework than a sole dependence on the discriminator or generator loss values. Although these loss metrics are significant and warrant consideration, their individual evaluation does not offer a comprehensive understanding of the GAN's overall performance. The discriminator loss within a GAN functions as a quantifier of the discriminator's proficiency in distinguishing between authentic and synthesized samples. An observed decrease in discriminator loss underscores an enhancement in the discriminator's ability to differentiate between these two categories of samples. Contrastingly, the generator loss is an indicator of the generator's capacity, successfully misleading the discriminator. A decrease in the generator loss indicates an increased resemblance of the fabricated samples to their original counterparts.

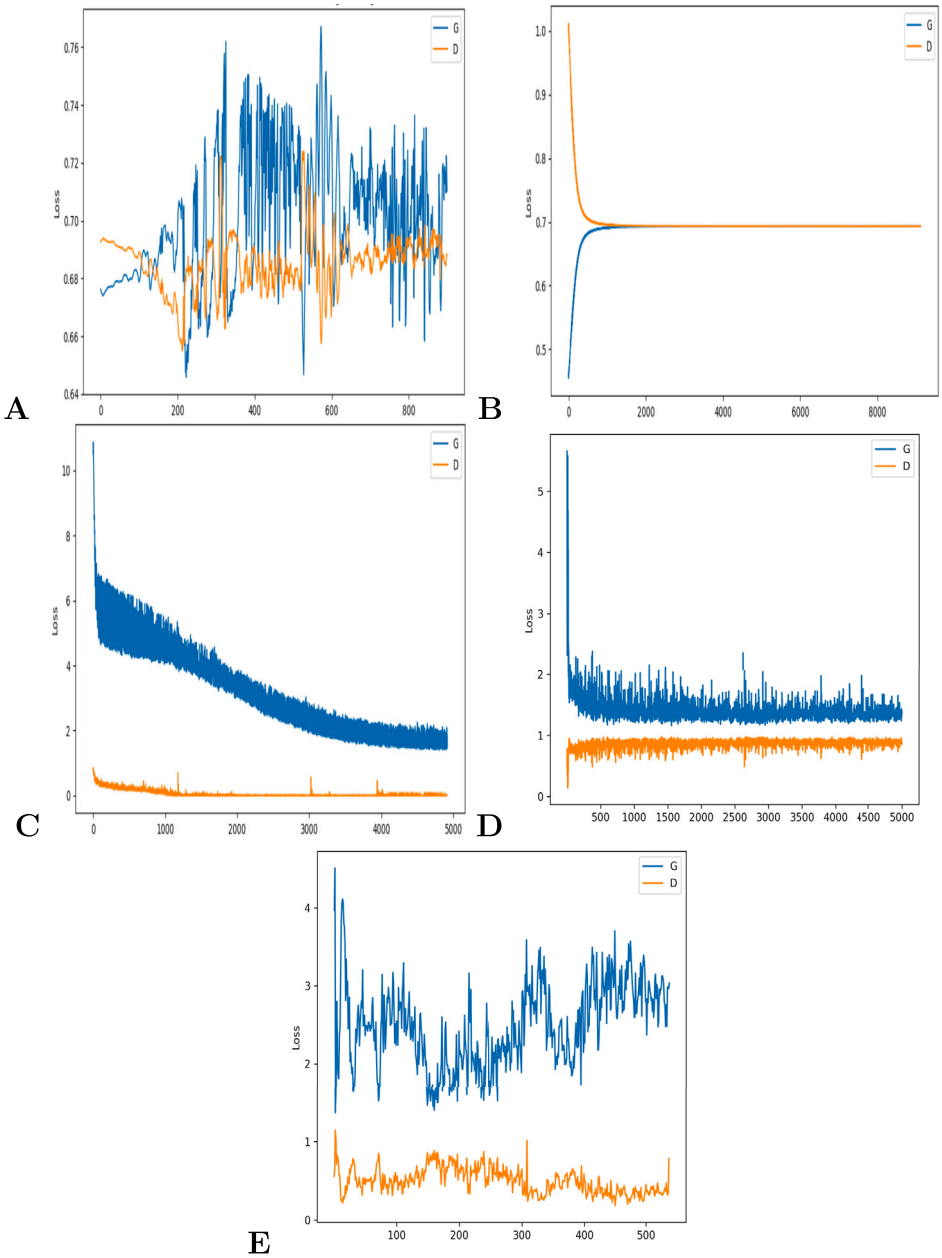

Fig. 10. Generator and discriminator loss during training **A**) DCGAN, **B**) CGAN, **C**) CycleGAN **D**) StyleGAN2, **E**) StyleGAN3.

However, to fully comprehend a GAN model's quality, it is pivotal to consider additional performance metrics alongside a visual inspection of the created samples. FID is a notably significant metric in this context. FID provides a more thorough appraisal of the GAN's performance and the verisimilitude of the generated samples. Upon a detailed evaluation of the outcomes, it becomes clear that, of all the models studied, StyleGAN2 displays the lowest FID score. This finding alludes to the superior capability of StyleGAN2 in generating images with a more realistic quality compared to other models. The sample generated images shown in Fig. 11 (B) confirm StyleGAN2's perceived superiority, and support this inference.

4.4. StyleGAN2 hyperparameters optimization

Specific hyperparameters of StyleGAN2, such as R1 gamma and Adaptive Discriminator Augmentation (ADA) target values, require fine-tuning to suit the custom input data. R1 gamma is a regularization parameter, with the gamma value dictating its weight. The ADA target establishes the augmentation probability. To fine-tune these hyperparameters, we employed the PSO algorithm. Table 2 provides a list of the hyperparameter values, along with the experimental range of values utilized in PSO optimization.

Table 1
Comparison of different GAN models.

GAN	Generator Loss	Discriminator Loss	FID Score
DCGAN	2.670	0.070	37.443
CGAN	0.754	0.673	34.429
CycleGAN	0.640	0.699	25.838
StyleGAN2	1.159	0.143	18.840
StyleGAN3	1.369	0.187	203.399

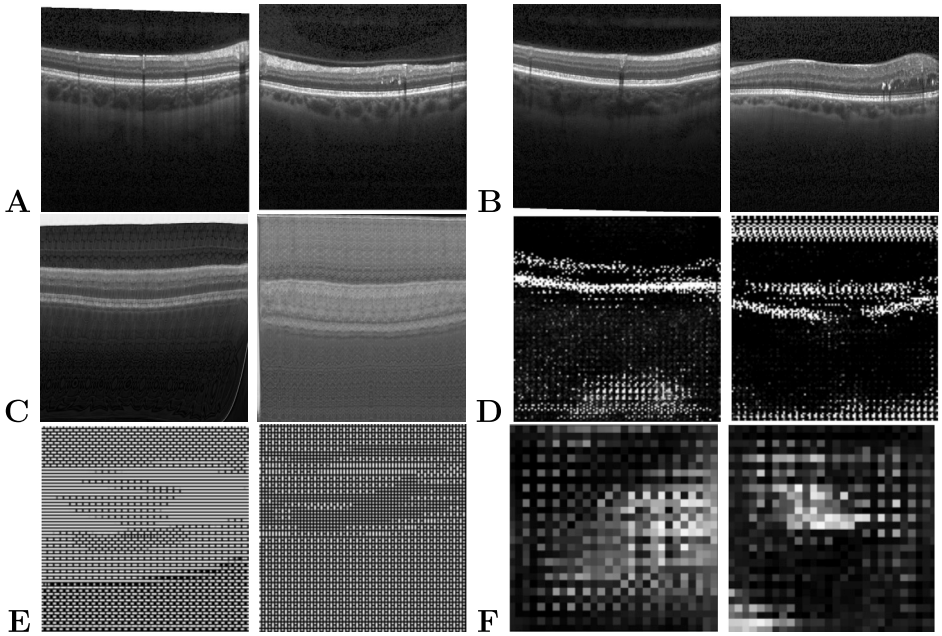

Fig. 11. Results of generated sample images. **A**) Original (Normal, DME), **B**) StyleGAN2 (Normal, DME), **C**) StyleGAN3 (Normal, DME), **D**) CycleGAN (DME, DME), **E**) CGAN (Normal, DME), **F**) DCGAN (DME, DME).

Table 2
Comparison of hyperparameter values.

Hyperparameters	Default value	PSO Experimented Range
R1 gamma	10	5-15
ADA target	0.6	0.5-1

5. Discussion

GANs have been increasingly deployed for detecting the evolution of various diseases in recent times. For instance, Rachmadi et al. [5] utilized a GAN model to project the evolution of White matter hyperintensities in small vessel disease. GANs have also been harnessed to identify abnormal parts of an image using only healthy OCT images [7]. However, this approach has not been employed for generating the DME images. This study aims to generate OCT images of patients diagnosed with DME using GANs, aiding doctors in anticipating the disease's progression, thereby tailoring the patient's treatment regimen accordingly. Additionally, the study provides a comparative analysis of five distinct GAN models: DCGAN, CGAN, CycleGAN, StyleGAN2, and StyleGAN3.

DCGAN [20] does not consider any prior information such as the patient's baseline OCT image, limiting itself to the generation of images resembling a random OCT. This limitation is addressed by the CGAN [36] and CycleGAN [24], with the latter offering a marked improvement in image generation. StyleGANs further enhance performance by modifying the generator architecture, thereby enabling better control over the image generation process. StyleGAN2, which is more efficient in training with a limited number of images, provides superior control over the generated output based on varying conditions. This can be attributed to a distinct mapping network that produces a random vector, which is then fed into every convolutional block along with the feature map generated by the preceding block. This lends significant flexibility to the output. Additionally, the use of ResNets in StyleGAN2's generator network leads to high-quality OCT image production. However, StyleGAN3 does not yield desirable results as it tends to overlook fine details in images. Figs. 11 (A-F) display OCT images generated by the various GAN models: DCGAN, CGAN, CycleGAN, StyleGAN2, and StyleGAN3. StyleGAN3 generates more generalized images, failing to incorporate minor DME biomarkers like HRF in its output.

Table 3
Comparison of hyperparameter and FID5000 scores.

Hyperparameters values	FID5000
(R1 gamma = 10, target = 0.6) - Default	18.8408
(R1 gamma = 6.72852, target = 0.637695) - PSO	13.2612

Table 4
DME detection performance on Kermany dataset [39].

Dataset	Precision	Recall	F1-score	Accuracy
Original	0.62	0.57	0.52	0.57
Without-GAN-500	0.65	0.65	0.65	0.65
With-GAN-500	0.81	0.70	0.68	0.70

StyleGAN2 yields the most accurate results, closely resembling the expected OCT images. The FID scores obtained with the default StyleGAN2 hyperparameters and the hyperparameters selected via PSO are detailed in Table 3. The FID score serves as a measure of the quality of the GAN-generated images, with a lower FID signifying superior quality. The observed decrease in the FID after fine-tuning the hyperparameters using the proposed PSO algorithm indicates an improvement in the realism of the generated DME images.

The effectiveness of the generated DME images was assessed via a DME detection task using the large dataset of labeled OCT Images, one of the most comprehensive publicly accessible datasets in ophthalmology [39]. This dataset includes images from 5,319 patients, including retinas affected by CNV, DME, drusen, and normal retinas. We utilize the second version of this dataset, which provides a balanced class distribution among training (83,484 images), validation (32 images), and testing sets (968 images). A recent study by Emil et al. [40] demonstrated that 92% of the test images correspond to subject-IDs also found in the training set, leading to high accuracy rates in several studies. To avoid this overlap, they proposed an efficient method for fair data splits, achieving an accuracy of 77% for four classes using the LightOCT model [41]. Our study focuses only on DME and normal images from this dataset. We employ the popular, pre-trained ResNet50 [42] model for the DME detection task.

The data was partitioned using the tool provided by Emil et al. [40]. Initially, we trained ResNet50 on 37,667 images (26,331 Normal, 11,336 DME) and tested it on 484 images (242 Normal, 242 DME). The Emil et al. [40] tool determined this split, which will be referred to as 'Original' from here on. Additionally, we reserved 5% of the training data subjects for validation purposes. During the 50-epoch training process, the model providing the highest F1 score on the validation data was saved. We trained ResNet50 on 250 normal and DME images, generated from the highest-performing StyleGAN2 trained on the Duke Eye dataset [38], and the first 500 images from the Kermany dataset [39] training set. This model will be referred to as 'With-GAN-500'. For a fair comparison, we also trained ResNet50 only on the first 500 images from the Kermany dataset [39], referred to as 'Without-GAN-500'. The performance of these experiments is summarized in Table 4, showing that the inclusion of GAN-generated images yielded the best results.

We further analyzed the predominant features learned by ResNet50 for DME detection using Gradient-weighted Class Activation Mapping (Grad-CAM) [43]. Figs. 12 (A–D) display the Grad-CAM visualization for the 'Original', 'With-GAN-500', and 'Without-GAN-500'. The coarse localization map of the last convolution layer, created before AdaptiveAvgPool2d, reflects crucial regions in the input image for DME detection. The generated Grad-CAM is normalized and resized to the original image size, and a mask image generated from the Grad-CAM is superimposed on the original input. The proposed 'With-GAN-500' approach outperformed the 'Original' and 'Without-GAN-500' models in accurately identifying DME for the same test split, thus facilitating explainable predictions. Despite the challenges in discerning subtle DME lesions, the model could capture global distinguishing features between Normal and DME using GAN-generated images, while fine-tuning with minimal images from the required dataset enabled it to learn local features. This significantly reduced the training time and built a more robust, generalized DME detection system that learned accurate DME lesions. The learned features are highlighted in Fig. 12. The 'With-GAN-500' model correctly identified almost all Normal images, barring one test image. Although most DME images were correctly identified, false negatives arose primarily because these types of DME lesions were not seen during training. This could be further reduced by fine-tuning with more such DME images. Despite this, the 'With-GAN-500' model performed more accurately than the model trained with over 30K images. A plausible reason for the performance decrease with 'Original' model may be due to the excessive speckle noise present in some images, leading the model to learn unimportant features, subsequently reducing performance on the test set.

Despite the significant insights this study provides into the role of GAN synthesized images in DME detection, it remains essential to acknowledge the specific limitations embedded within the study. The investigation primarily relies on the pre-trained ResNet50, a commonly employed lightweight model, to illustrate the effectiveness of GAN-synthesized images for DME detection. However, the utilization of more contemporary and potentially advanced classification models in subsequent research could serve to corroborate and augment these initial findings. Furthermore, the study is limited by its reliance on two publicly accessible OCT datasets. This scope restriction may potentially confine the generalizability of the findings to a broader, more diverse array of datasets. To validate the robustness of the classification model, future research could consider incorporating a wider variety of publicly accessible, large-scale OCT datasets. Another constraint of the study pertains to the use of PSO for hyperparameter tuning. Although PSO serves as an effective optimization instrument, it does not consistently assure the acquisition of optimal values. In the present study, the scope of PSO was confined to a predefined range, and it was executed over 100 iterations. The derived hyperparameters were

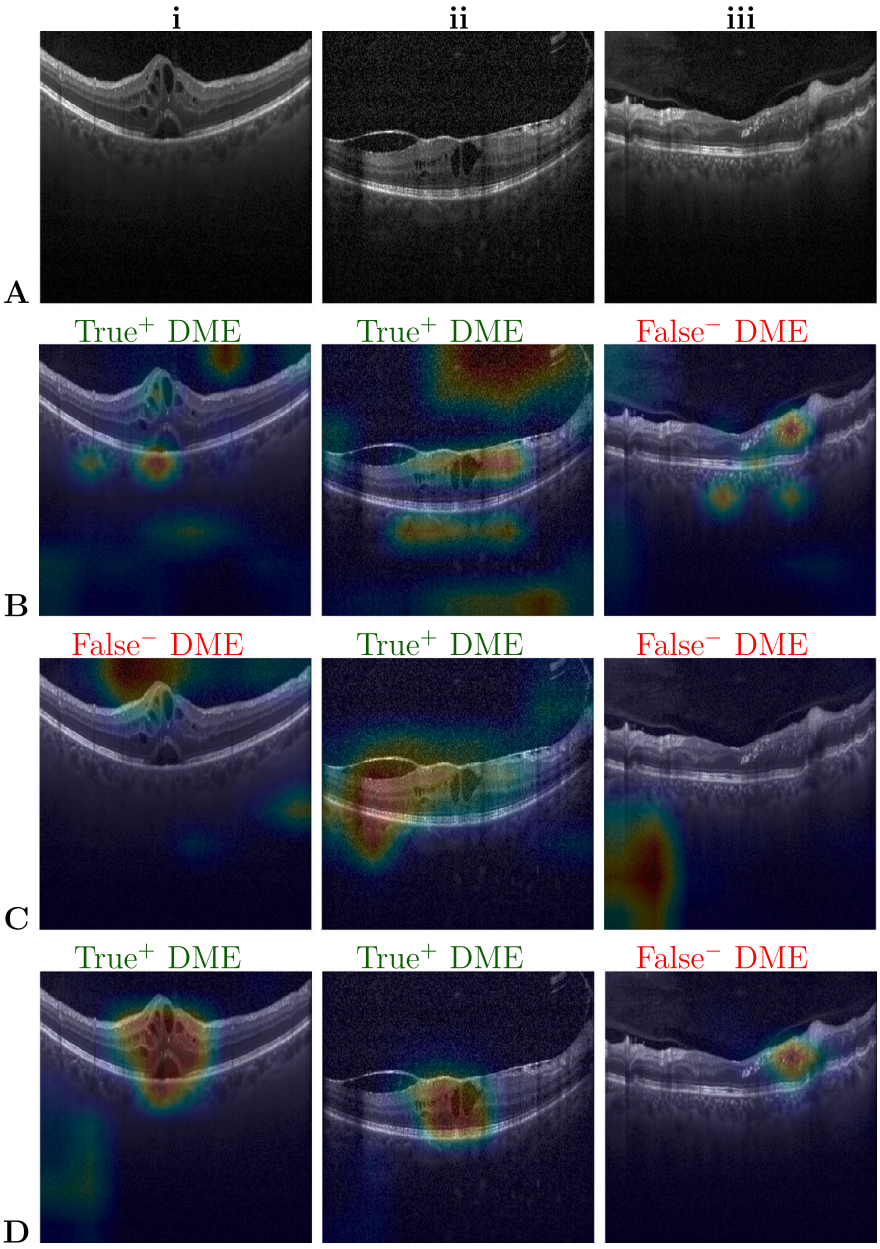

Fig. 12. Visualization of Grad-CAM heatmap on the original input images from test set. The samples are for **A)**Original Images (B) Grad-CAM for 'Original' model for **A** images; **B)**Grad-CAM for 'Without-GAN-500' model for **A** images; and **C)**Grad-CAM for 'With-GAN-500' model for **A** images.

then employed to fine-tune the StyleGAN model across 10 iterations. To the best of our knowledge, this research represents the inaugural exploration of meta-heuristic optimization technique for the fine-tuning of GAN, specifically guided by the FID score. Future research could potentially adopt more exhaustive optimization methods [44,45] and experimentation strategies to determine optimal hyperparameters. While the primary aim of this study was to identify an effective GAN for generating realistic DME images, DME detection was performed solely using the widely adopted ResNet50 model. An area for potential future investigation could involve evaluating the impact on DME detection performance when employing more recent visual transformer [46,47] and attention based models [48]. By acknowledging these limitations, the study paves the way for more comprehensive research in the future.

6. Conclusions

This research introduces a GAN model for generating OCT images of a DME patient, given the patient's baseline OCT image. The investigation commenced with the training of a DCGAN on the Duke Eye dataset. The resulting DCGAN images exhibited high variance from the desired image and lacked any relationship with the baseline OCT image— the model did not incorporate the DME

condition when generating a new image. To address this, a CGAN was employed to generate the OCT images; however, the resulting images were excessively noisy and failed to resemble the target image. Subsequently, CycleGAN was utilized, resulting in improved image quality. The images generated by StyleGAN2 closely resembled actual OCT images for both normal and DME conditions. In contrast, the images generated by StyleGAN3 exhibited more coarse noise patterns. This study includes a comparison of these five different GANs.

Presently, DME severity is determined manually by doctors who interpret various imagery biomarkers present in OCT images. This could be revolutionized by the development of an automated model that inputs the patient's OCT image, extracts distinct biomarkers, and predicts disease severity. Such an approach could standardize the process of detecting severity in DME, minimizing errors and inaccuracies inherent in manual procedures. Although this research juxtaposes different GANs within the medical domain to explore the potential progression of DME in patients, it does not restrict its scope to DME alone. The same methodology could be applied to visualize the progression of other diseases, potentially providing valuable assistance to doctors during treatment planning.

Funding statement

Open access funding provided by Manipal Academy of Higher Education, Manipal.

CRediT authorship contribution statement

Aditya Tripathi: Veena Mayya: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. Preetham kumar: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper. Akshat Tulsani: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are openly available at https://people.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm (Accessed on May 2022) [38].

Acknowledgements

The authors acknowledge the use of high-performance computing resources of the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal, Manipal Academy of Higher Education, Manipal, made available for conducting the research reported in this paper.

References

[1] X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, J. Liu, Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning, IEEE Access 7 (2019) 3046–3061.
[2] V. Das, S. Dandapat, P.K. Bora, A data-efficient approach for automated classification of OCT images using generative adversarial network, IEEE Sens. Lett. 4 (1) (2020).
[3] X. He, L. Fang, H. Rabbani, X. Chen, Z. Liu, Retinal optical coherence tomography image classification with label smoothing generative adversarial network, Neurocomputing 405 (2020) 37–47.
[4] A. Smitha, P. Jidesh, Detection of retinal disorders from OCT images using generative adversarial networks, Multimed. Tools Appl. 81 (20) (2022) 29609–29631.
[5] M.F. Rachmadi, M. del, C. Valdés-Hernández, S. Makin, J.M. Wardlaw, T. Komura, Predicting the Evolution of White Matter Hyperintensities in Brain Mri Using Generative Adversarial Networks and Irregularity Map, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11766, 2019, pp. 146–154.
[6] T. Xia, A. Chartsias, S.A. Tsaftaris, Consistent brain ageing synthesis, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV, Springer-Verlag, Berlin, Heidelberg, 2019, pp. 750–758.
[7] T. Schlegl, P. Seeböck, S.M. Waldstein, G. Langs, U. Schmidt-Erfurth, f-AnoGan: fast unsupervised anomaly detection with generative adversarial networks, Med. Image Anal. 54 (2019) 30–44.
[8] C.-I. Suciu, V.-I. Suciu, A. Cuțaș, S.D. Nicoară, Interleaved optical coherence tomography: clinical and laboratory biomarkers in patients with diabetic macular edema, J. Pers. Med. 12 (5) (2022).
[9] D. Acan, M. Calan, D. Er, T. Arkan, N. Kocak, F. Bayraktar, S. Kaynak, The prevalence and systemic risk factors of diabetic macular edema: a cross-sectional study from Turkey, BMC Ophthalmol. 18 (1) (2018).
[10] B.E.K. Klein, C.E. Myers, K.P. Howard, R. Klein, Serum lipids and proliferative diabetic retinopathy and macular edema in persons with long-term type 1 diabetes mellitus: the Wisconsin epidemiologic study of diabetic retinopathy, JAMA Ophthalmol. (2015).
[11] J.P. Ehlers, S. Yeh, M.G. Maguire, J.R. Smith, P. Mruthyunjaya, N. Jain, L.A. Kim, C.Y. Weng, C.J. Flaxel, S.D. Schoenberger, S.J. Kim, Intravitreal pharmacotherapies for diabetic macular edema: a report by the American Academy of Ophthalmology, Ophthalmology 129 (1) (2022) 88–99.
[12] A. Markan, A. Agarwal, A. Arora, K. Bazgain, V. Rana, V. Gupta, Novel imaging biomarkers in diabetic retinopathy and diabetic macular edema, Ther. Adv. Ophthalmol. 12 (2020) 2515841420950513, https://doi.org/10.1177/2515841420950513.

[13] H. Endo, S. Kase, H. Tanaka, M. Takahashi, S. Katsuta, Y. Suzuki, M. Fujii, S. Ishida, M. Kase, Factors based on optical coherence tomography correlated with vision impairment in diabetic patients, Sci. Rep. 11 (1) (2021).
[14] J.K. Sun, M.M. Lin, J. Lammer, S. Prager, R. Sarangi, P.S. Silva, L.P. Aiello, Disorganization of the retinal inner layers as a predictor of visual acuity in eyes with center-involved diabetic macular edema, JAMA Ophthalmol. 132 (11) (2014) 1309–1316.
[15] J.K. Sun, S.H. Radwan, A.Z. Soliman, J. Lammer, M.M. Lin, S.G. Prager, P.S. Silva, L.B. Aiello, L.P. Aiello, Neural retinal disorganization as a robust marker of visual acuity in current and resolved diabetic macular edema, Diabetes 64 (7) (2015) 2560–2570, https://doi.org/10.2337/db14-0782.
[16] N. Nagai, M. Suzuki, A. Uchida, T. Kurihara, N. Ban, S. Minami, H. Shinoda, K. Tsubota, Y. Ozawa, The area and number of intraretinal cystoid spaces predict the visual outcome after ranibizumab monotherapy in diabetic macular edema, J. Clin. Med. 9 (5) (2020).
[17] G. Yalçın, c. Özdek, F.N.B. Aksakal, Defining cystoid macular degeneration in diabetic macular edema: an OCT-based single-center study [diyabetik maküla Ödeminde kistoid maküler dejenerasyon tanımı: OKT tabanlı tek merkezli Çalışma], Turk. J. Ophthalmol. 49 (6) (2019) 315–322.
[18] M. Arthi, M.D. Sindal, R. Rashmita, Hyperreflective foci as biomarkers for inflammation in diabetic macular edema: retrospective analysis of treatment naïve eyes from south India, Indian J. Ophthalmol. 69 (5) (2021) 1197–1202, https://doi.org/10.4103/ijo.IJO_2627_20.
[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, vol. 27, Curran Associates, Inc., 2014.
[20] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in: 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 2016.
[21] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, in: International Conference on Learning Representations, 2018.
[22] Y. Jin, J. Zhang, M. Li, Y. Tian, H. Zhu, Z. Fang, Towards the automatic anime characters creation with generative adversarial networks, arXiv preprint, arXiv:1708.05509, 2017.
[23] P. Isola, J. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Los Alamitos, CA, USA, 2017, pp. 5967–5976.
[24] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, vol. 2017, October, 2017, pp. 2242–2251.
[25] A. Dash, J.C.B. Gamboa, S. Ahmed, M. Liwicki, M.Z. Afzal, TAC-GAN - text conditioned auxiliary classifier generative adversarial network, arXiv preprint, arXiv:1703.06412, 2017.
[26] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, D. Metaxas, Stackgan: text to photo-realistic image synthesis with stacked generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, vol. 2017, October, 2017, pp. 5908–5916.
[27] G. Perarnau, J. van de Weijer, B. Raducanu, J.M. Álvarez, Invertible conditional GANs for image editing, in: NIPS Workshop on Adversarial Training, 2016.
[28] J. Wu, C. Zhang, T. Xue, W.T. Freeman, J.B. Tenenbaum, Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling, in: Advances in Neural Information Processing Systems, 2016, pp. 82–90.
[29] Y. Chen, X.-H. Yang, Z. Wei, A.A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, Q. Guan, Generative adversarial networks in medical image augmentation: a review, Comput. Biol. Med. 144 (2022).
[30] J. Lin, Y. Xia, T. Qin, Z. Chen, T.-Y. Liu, Conditional image-to-image translation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[31] Y. Li, J. Li, Y. Wang, Privacy-preserving spatiotemporal scenario generation of renewable energies: a federated deep generative learning approach, IEEE Trans. Ind. Inform. 18 (4) (2022) 2310–2320, https://doi.org/10.1109/TII.2021.3098259.
[32] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.
[33] M. Clerc, J. Kennedy, The particle swarm – explosion, stability, and convergence in a multidimensional complex space, IEEE Trans. Evol. Comput. 6 (1) (2002) 58–73.
[34] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks, Vol. 4, 1995, pp. 1942–1948.
[35] J. Kennedy, Particle Swarm Optimization, Springer US, Boston, MA, 2010, pp. 760–766.
[36] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint, arXiv:1411.1784, 2014.
[37] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 6629–6640.
[38] P.P. Srinivasan, L.A. Kim, P.S. Mettu, S.W. Cousins, G.M. Comer, J.A. Izatt, S. Farsiu, Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, Biomed. Opt. Express 5 (10) (2014) 3568–3577.
[39] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M.K. Prasadha, J. Pei, M. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V.A. Huu, C. Wen, E.D. Zhang, C.L. Zhang, O. Li, X. Wang, M.A. Singer, X. Sun, J. Xu, A. Tafreshi, M.A. Lewis, H. Xia, K. Zhang, Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122-1131.e9, https://doi.org/10.1016/j.cell.2018.02.010.
[40] I.E. Tampu, A. Eklund, N. Haj-Hosseini, Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images, Sci. Data 9 (1) (2022), https://doi.org/10.1038/s41597-022-01618-6.
[41] A. Butola, D.K. Prasad, A. Ahmad, V. Dubey, D. Qaiser, A. Srivastava, P. Senthilkumaran, B.S. Ahluwalia, D.S. Mehta, Deep learning architecture "lightoct" for diagnostic decision support using optical coherence tomography images of biological samples, Biomed. Opt. Express 11 (9) (2020) 5017–5031, https://doi.org/10.1364/BOE.395487.
[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016, IEEE Computer Society, 2016, pp. 770–778.
[43] R.R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2019) 336–359.
[44] Y. Liang, F. He, X. Zeng, J. Luo, An improved loop subdivision to coordinate the smoothness and the number of faces via multi-objective optimization, Integr. Comput.-Aided Eng. 29 (1) (2022) 23–41, https://doi.org/10.3233/ICA-210661.
[45] J. Luo, F. He, X. Gao, An enhanced grey wolf optimizer with fusion strategies for identifying the parameters of photovoltaic models, Integr. Comput.-Aided Eng. 30 (1) (2022) 89–104, https://doi.org/10.3233/ICA-220693.
[46] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, 2021, An image is worth 16x16 words: transformers for image recognition at scale.
[47] F. Shamshad, S. Khan, S.W. Zamir, M.H. Khan, M. Hayat, F.S. Khan, H. Fu, Transformers in medical imaging: a survey, Med. Image Anal. 88 (2023) 102802.
[48] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.