Journal of Computers, Mechanical and Management
Journal homepage: www.jcmm.co.in
e-ISSN: 3009-075X
AN PUBLISHING

Volume 2 Issue 6

Article Number: 230106

Fundus Image Generation using EyeGAN: An Improved Generative Adversarial Network Model

Preeti Kapoor*[1] and Shaveta Arora[1]

[1]Department of Computer Science and Engineering, School of Engineering, The NorthCap University, Gurugram, Haryana, India 122017

Abstract

Deep learning models are widely used in various computer vision fields ranging from classification, segmentation to identification, but these models suffer from the problem of overfitting. Diversifying and balancing the datasets is a solution to the primary problem. Generative Adversarial Networks (GANs) are unsupervised learning image generators which do not require any additional information. GANs generate realistic images and preserve the minute details from the original data. In this paper, a GAN model is proposed for fundus image generation to overcome the problem of labelled data insufficiency faced by researchers in detection and classification of various fundus diseases. The proposed model enriches and balances the studied datasets for improving the eye disease detection systems. EyeGAN is a nine-layered structure based on conditional GAN which generates unbiased, good quality, credible images and outperforms the existing GAN models by achieving the least Fréchet Inception Distance of 226.3. The public fundus datasets MESSIDOR I and MESSIDOR II are expanded by 1600 and 808 synthetic images respectively.

Keywords: Deep Learning; FID; Conditional GAN; Style GAN

1 Introduction

Image generation is the process of creating synthetic images which are realistic in nature [1]. Image augmentation is used in various computer vision problems to resolve the scarcity of original data and to overcome the problem of data imbalance. Various public fundus datasets also suffer from class imbalance and non-labelled data, and there is a lack of good quality images in large datasets. Though deep learning has recently gained popularity and outperformed many current methods in many fields, particularly when it comes to analyzing and reviewing fundus images, these models have poor generalizability and suffer from overfitting. Using an enriched and balanced dataset with labelled data for training can help in overcoming this problem. Well-trained deep learning models prove to provide better results and achieve good accuracy. There are various augmentation methods available for producing images like rotation, flipping, or cropping. With the increase in the use of neural networks in every field, these models are also used for image generation, producing real-like images. GAN models are widely used as augmentation techniques in several areas and also improve the overall results [1]. Lee et al. [2] suggested an any-time-of-the-day camera-based blind spot detection model: GAN + Active BSD system to fight the problem of nighttime sample availability. The cycle GAN was trained using publicly available Nexar data and was used to generate synthetic nighttime images by using the daytime side rectilinear images. Also, these images were used to annotate the synthetic nighttime images. While training the generator, the main aim was to find the mapping between day to night images. The data augmentation using generated samples doubled the samples at all 4 stages of the BSD system. Then all four stages were evaluated with and without data augmentation.

*Corresponding author: preeti19csd006@ncuindia.edu
Received: DD Month 20YY; **Revised:** DD Month 20YY; **Accepted**: DD Month 20YY; **Published:** DD Month 20YY

DOI: 10.57159/gadl.jcmm.2.6.230106.

The accuracy improved after nighttime data augmentation from 26% to 55%, which is two times in comparison to just day images. Liu et al. [3] proposed a GAN model for grape leaf disease images generation to overcome the problems faced in grape disease detection and called it Leaf GAN. The generator was recreated with digressive channel layers and discriminator with the compact connectivity approach and specified example normalization. For further stabilization of the training process, a deep regret analytic gradient penalty algorithm was used to train both generator and discriminator. Then identification was performed with and without using augmented data, also the proposed method LG, LG-RELU, LG-PRELU, was compared with existing DCGAN, WGAN. The proposed method achieved better performance as compared to other GANs and dataset without augmentation, with an efficiency improvement from 95.25% to 98.25% on Xception net. Also, it overcame the overfitting problem. Waheed et al. [4] proposed an auxiliary classifier GAN (ACGAN) and named it CovidGAN for generating artificial images. They then combined it with a CNN model to improve detection. For dimension reduction, Principal Component Analysis was used and confusion matrices were used for result visualization. An improved detection of COVID-19 with an accuracy of 95% was achieved. However, there was no control over the biased image generation since the class labels were not used. This paper focuses on producing good quality images for each class by creating a GAN model which incorporates labels as a condition and avoids biased image generation. The paper demonstrates the use of a GAN model as the augmentation technique [5] to enrich the selected DR Datasets and fight class inequality. A GAN model is presented which is based on conditional GAN for fundus image generation rather than just basic augmentation operations. The proposed method is made lighter than the base model by reducing the convolutional layers, i.e., a nine-layered architecture. Also, EyeGAN was compared with the existing GAN models using the FID score to highlight the better quality of images produced. MESSIDOR I and MESSIDOR II datasets are augmented using EyeGAN to balance the images in all the classes, thus making them enriched and balanced datasets. The major contributions of this paper are as follows:

- A GAN model is proposed which balances the target datasets and produces real-like, high-quality images, surpassing the results of existing models.
- The EyeGAN is compared with conditional GAN and Style GAN using the Fréchet Inception Distance (FID) score.
- Two enriched Diabetic Retinopathy (DR) Datasets are created, enhancing the capabilities for better detection and classification of eye diseases.

The whole work is presented in four sections. Section I is the survey of the various augmentation methods which are used for image generation. Section II explains the methodology, i.e., the structure of the proposed GAN model, the generator loss function, and datasets. Section III shows the comparison and the outcomes. Section IV concludes the research.

2 Data Augmentation

The image datasets are often captured under a certain set of circumstances such as in a different direction, position, scale, or brightness. Data augmentation exploits these characteristics to generate images. Figure 1 shows the various augmentation techniques used in computer vision.

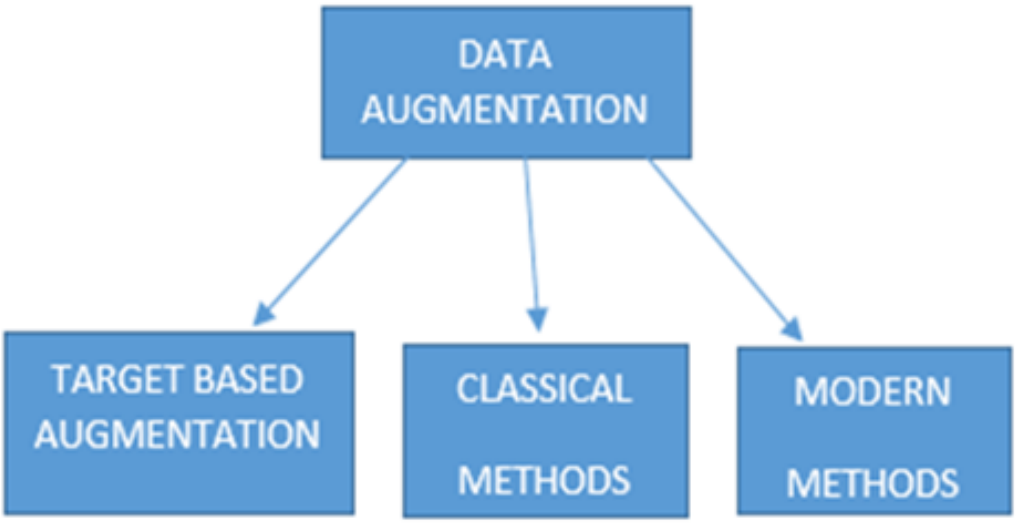

Figure 1: Various Augmentation techniques

2.1 Data augmentation based on target dataset size:

- **Offline augmentation:** This technique helps for small datasets, as the number of images are increased by a large factor. Here the whole dataset is targeted before it is fed to the machine learning models.
- **Online augmentation:** This is used for bigger datasets, where transformations are performed on the mini-batches which are input into the classifier.

2.2 Classical data augmentation methods:

Methods listed in this section are characterized by the ease of execution. These are geometric and space transformations which are relatively simple to use but require more training time and memory. They are not always label-preserving transformations as there is a loss of information during transformations. Table 1 depicts the characteristics of various classical methods.

Table 1: Classical Augmentation Methods

S.No.	Method	Characteristics
1	FLIPPING	Horizontal and vertical flips, easiest method, non-label preserving
2	COLOR SPACE	R,G,B manipulations using color histograms; intensity modification
3	CROPPING	Centre patch cropping, random cropping, non-label preserving
4	ROTATION	Rotation by 1 to 359 degree, depending on the angle of rotation label preservation differs.
5	TRANSLATION	Spatial dimensions are manipulated
6	NOISE INJECTION	Addition of a matrix containing random values selected at random using a Gaussian distribution
7	COLOR SPACE INJECTION	Alteration of color distribution of images, RGB to Grey Scale transformation, color jittering, random color manipulation, edge enhancement and PCA.
8	KERNEL FILTERS	Use to sharpen and blur images using filters like Gaussian blur filters, vertical and horizontal edge filters.
9	MIXING IMAGES	Mixes images by averaging pixels, mixing images by averaging RGB channel values, random cropping and mixing of them to create new images
10	RANDOM ERASING	A n × m patch of an image is randomly selected and masked with either 0, 255, mean pixel values or random values; can be stacked over other techniques; non label preserving.

2.3 Modern techniques: based on deep learning

Deep learning models are extremely effective in converting high-dimensional inputs into low-dimensional representations. They transform pictures into binary classes or nx1 vectors in flattened layers. It is possible to modify neural networks' sequential processing so that transitional depictions can be distinguished from the complete data. In fully connected layers, the lower-dimensional depictions of the image data can be isolated and extracted. Table 2 describes the various deep learning methods which are used for data augmentation in brief.

Table 2: Deep Learning Augmentation Methods[6]

SNO	METHOD	Characteristics
1	Feature space augmentation	Its works using auto encoders to create new inputs, or by isolating a vector representation of a CNN, these are used as input to other machine learning algorithms; but such representations are difficult to interpret.
2	Adversarial training	Two or more networks are used such that their objectives contrast (loss function); strengthen weak spots in machine learning models, may not be useful in problems suffering from overfitting
3	GAN-based Data Augmentation	Two models are used, increases computation speed, enriches datasets, lessen the false positive rate of detection of abnormalities, and solve problems with class imbalance and overfitting.
4	Neural Style Transfer	Manipulates the sequential representation of a CNN models; transferring the style of one image to another; preserves unique content, similar to color space lighting transformations; allows transformations such as lighting variations, texture variations and artistic styles; effort demanding

3 Proposed Model: EyeGAN

3.1 Conditional GAN (CGAN)

CGAN enables the method to depend on external data for enhancing image quality. CGANs control the output of the generator at test time by giving the label for the desired image generation. In CGAN, the generator produces an image for a class using a latent space point and a label as input [7]. Image and class name are sent to the discriminator, and it determines the originality of the input. The class labels help in preventing biased image generation as GAN models tend to generate the images each operation a 4D output is obtained. The output thus obtained is the final generated output. CGANs have proven to achieve good results for image generation [8].

3.2 Proposed GAN architecture

The proposed model works similarly to CGANs, i.e., a condition is applied to the generator. In the proposed model, the class labels are sent to the generator for a particular class of fundus images generation.

3.2.1 Generator architecture

The Generator consists of three blocks, each consisting of 1 convolution layer (Leaky ReLU activated), Up-sampling, and Batch Normalization. As shown in Figure 2, the convolution layer accepts an input in the form $(32, 224, 224, 1)$ where the input is in a batch of 32 images with $(224 \times 224 \times 1)$ dimensions.

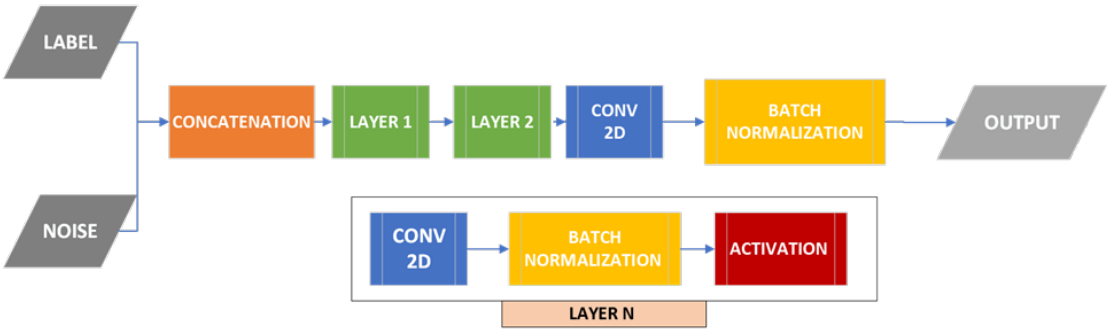

Figure 2: Generator Design of EyeGAN

At the end of each operation, a 4D output is obtained. The output thus obtained is the final generated output. To produce a single $224 \times 224 \times 1$ image, a latent vector of noise and class label are provided as inputs. Each classified input receives the class label as a parameter. A batch normalization layer and an activation layer come after each convolutional layer, except for the final one. The model employs techniques including LeakyReLU activation, a kernel of size $(5, 5)$, a stride of $(2, 2)$, and an activation function for the hyperbolic tangent in the output layer.

3.2.2 Discriminator architecture

Figure 3 shows that in the discriminator model there are two outputs and one image input $(224 \times 224 \times 1)$. The discriminator decides whether the image is real or fake and outputs the class label. Every discriminator block contains a convolutional layer, a batch normalization layer, and an activation layer. There are three such blocks which downsample the input image from $(224 \times 224 \times 196)$ to $(221 \times 221 \times 176)$ and then to $(128 \times 128 \times 160)$. The model employs techniques including LeakyReLU activation, a kernel of size $(5, 5)$, and a stride of $(2, 2)$. An activation function of sigmoid function calculates the authenticity of input, and the other output layer, which outputs the label, uses a softmax function. The complete model architecture is shown in Figure 3.

3.3 Loss Function

Based on the literature survey done, it is observed that cycle GAN is used for image generation where there is a class imbalance [9] and StyleGAN [10] generates high resolution and high diversity images. Therefore, these two are chosen for comparison with the proposed GAN. Also, to find the best loss function for the proposed model; mean squared error (MSE), mean absolute error (MAE), and binary cross-entropy are studied.

3.3.1 Binary cross entropy

The amount of the difference between the predicted probability distribution and the true probability distribution for a binary classification problem is called binary cross-entropy loss. The training data has two classes. The loss is usually

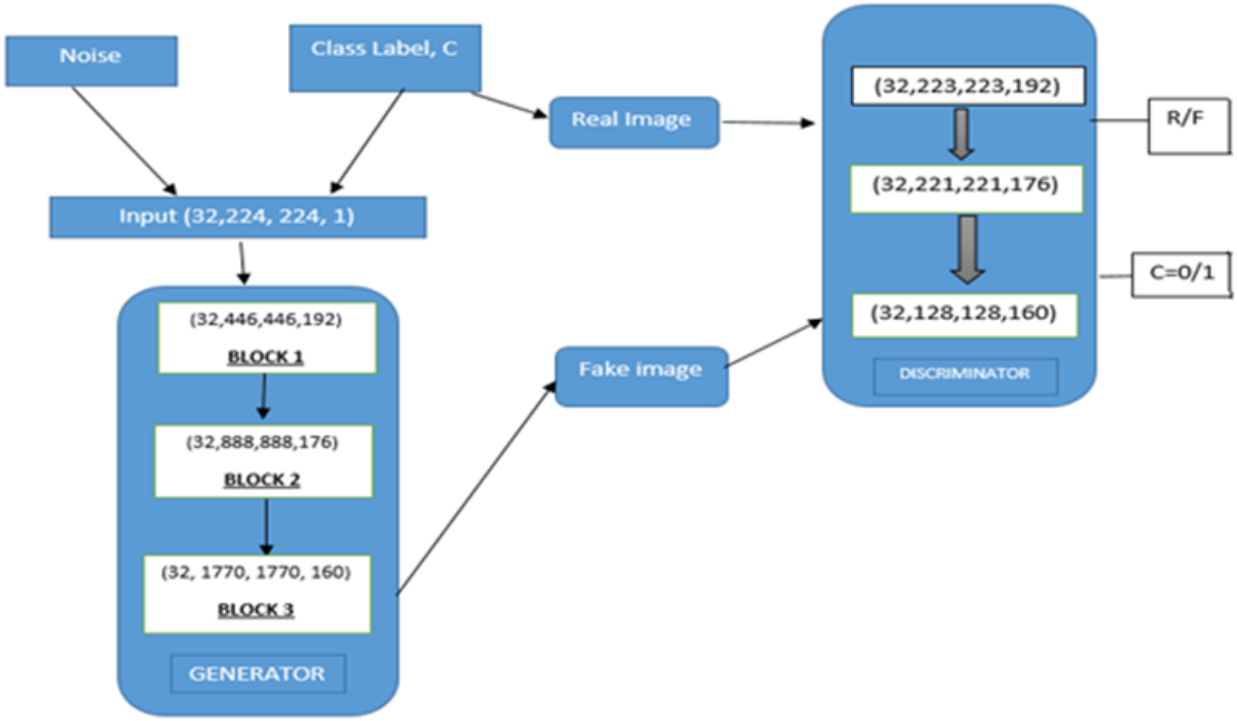

Figure 3: Overall model architecture

employed as the loss function. The binary cross-entropy loss is defined as:

$$\text{logloss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij}) \tag{1}$$

where N is the number of rows and M is the number of columns.

3.3.2 Mean squared error (MSE)

Mean Squared Error (MSE) is a way to quantify the dissimilarity between real values and anticipated values in a regression problem. When a model is trained to produce predictions based on continuous data, this method is frequently employed as the loss function. The MSE is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

where N represents the number of data points on all variables, y is the vector of studied values for the predicted variable, and $\hat{y}$ represents the predicted values.

3.3.3 Mean absolute error (MAE)

Mean Absolute Error (MAE) analyses the average magnitude of the errors in a set of estimates while ignoring the direction. It represents the weighted average of the individual deviations between the actual value and the predictions for the test example. The MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i| \tag{3}$$

Table 3 shows the comparison between the three loss functions values for EyeGAN, Style GAN, and Cycle GAN. Additionally, in Figure 4, it can be seen that the binary loss function gives the least loss value. Therefore, for the proposed method, the binary loss function is used.

3.3.4 Proposed GAN Loss

- **Generator loss**: The objective is to lessen the second term in the discriminator loss equation. The generator loss function for a single created data is expressed as follows:

$$\mathcal{L}(\text{Generator}) = \min\left[\log(1 - D(G(z)))\right] \tag{4}$$

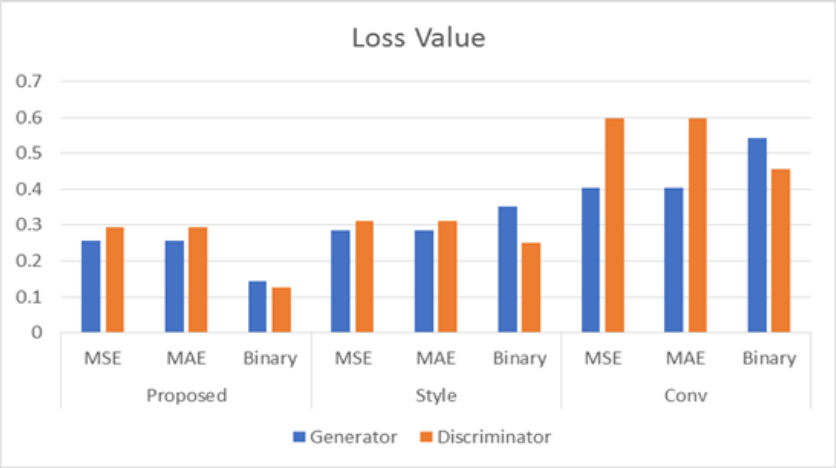

Figure 4: Comparison of Loss Functions

Table 3: Loss Value of Generator and Discriminator

GAN Model	Loss Function	Generator	Discriminator
EyeGAN	MSE	0.2575	0.2941
	MAE	0.2575	0.2941
	Binary	0.1447	0.1248
Style GAN [10]	MSE	0.2858	0.3117
	MAE	0.2858	0.3117
	Binary	0.3511	0.2512
Cycle GAN [9]	MSE	0.4026	0.5964
	MAE	0.4026	0.5964
	Binary	0.5417	0.4568

- **Discriminator loss**: There are two losses for the discriminator: one for a real input and the other for a fake input, which gives the below discriminator loss. Since the model is a conditional GAN, the loss function is represented as shown below:

$$\mathcal{L}(\text{discriminator}) = \max\left[\log(D(X)) + \log(1 - D(G(z)))\right] \tag{5}$$

$$\mathbb{E}_x[\log(D(x|y))] + \mathbb{E}_y[\log(1 - D(G(z|y)))] \tag{6}$$

The modified function has the same fundamental structure with a small modification. With this updated loss function, each part's loss is determined by factoring in the condition, y.

$$\min_G \max_D V(D,G) = \mathbb{E}_x \sim p_{\text{data}}(x)\left[\log D\left(\frac{x}{y}\right)\right] + \mathbb{E}_z \sim p_z\left[\log\left(1 - D\left(G\left(\frac{z}{y}\right)\right)\right)\right] \tag{7}$$

3.4 Database

The Messidor I dataset [11], which contains 1200 fundus color images captured with a 45-degree field of view, is one of the datasets considered. The images vary in size: 1440x960, 2240x1488, or 2304x1536. Fundus images are graded based on the number of microaneurysms, hemorrhages, and the presence of neovascularization, categorized into grades R0, R1, R2, and R3. The Messidor-2 dataset [11], an expansion of the Messidor dataset, consists of diabetic retinopathy examinations, each including two images of the eye fundus with the macula at their center (one per eye). This collection includes 1748 images, also captured with a 45-degree field of view. Neither dataset is balanced across classes. Table V shows the image distribution in the studied datasets. It is evident that class 0 has the maximum number of images, whereas in Messidor I, class 1 and class 2 have very few images, and in Messidor II, class 3 has only 75 images.

4 Experimentation and Results

4.1 4.1 Fréchet Inception Distance (FID)

Various GAN models can be judged based on the graphic quality of synthetic images. For a high-quality dataset, Fréchet Inception Distance (FID) [12] calculates the image quality and diversity, both of which are crucial. Since this is reliable and effective, this score is frequently used to assess the effectiveness of various models. This research compares the GAN model performances using this score. The FID is defined as:

$$FID = d^2((m_r, C_r), (m_g, C_g)) = \|m_r - m_g\|_2^2 + \text{Tr}\left(C_r + C_g - 2(C_r C_g)^{\frac{1}{2}}\right) \tag{8}$$

where d is the FID, (m_r, C_r) represents the real images and (m_g, C_g) represents the generated images, m and C are means and covariance, respectively. The FID score demonstrates the variations between the generated and actual images. A lower score indicates better performance. Table 4 compares the three models in terms of structure, features, and average FID score. The proposed model has fewer layers than the approaches in use, making it faster and lighter. Additionally, a DR class-wise FID score comparison is done in Table 5. The proposed model attained the least FID score for all the classes, promising the best quality image generation.

Table 4: Comparison of GAN Models

Model	Features	Average FID Score
Model 1	Feature Set 1	0.50
Model 2	Feature Set 2	0.60
Model 3	Feature Set 3	0.55

Table 5: DR Class Wise FID Score Comparison

Class	Model 1 FID Score	Model 2 FID Score
Class 0	0.50	0.60
Class 1	0.55	0.65
Class 2	0.60	0.70
Class 3	0.65	0.75

4.1.1 Data Distribution of the Studied Datasets

Figure 5 shows the original images and the respective generative images for each class for the MESSIDOR I dataset. Table 6 shows the data distribution of the studied datasets. This table also illustrates the number of generated images added to overcome the problem of data imbalance.

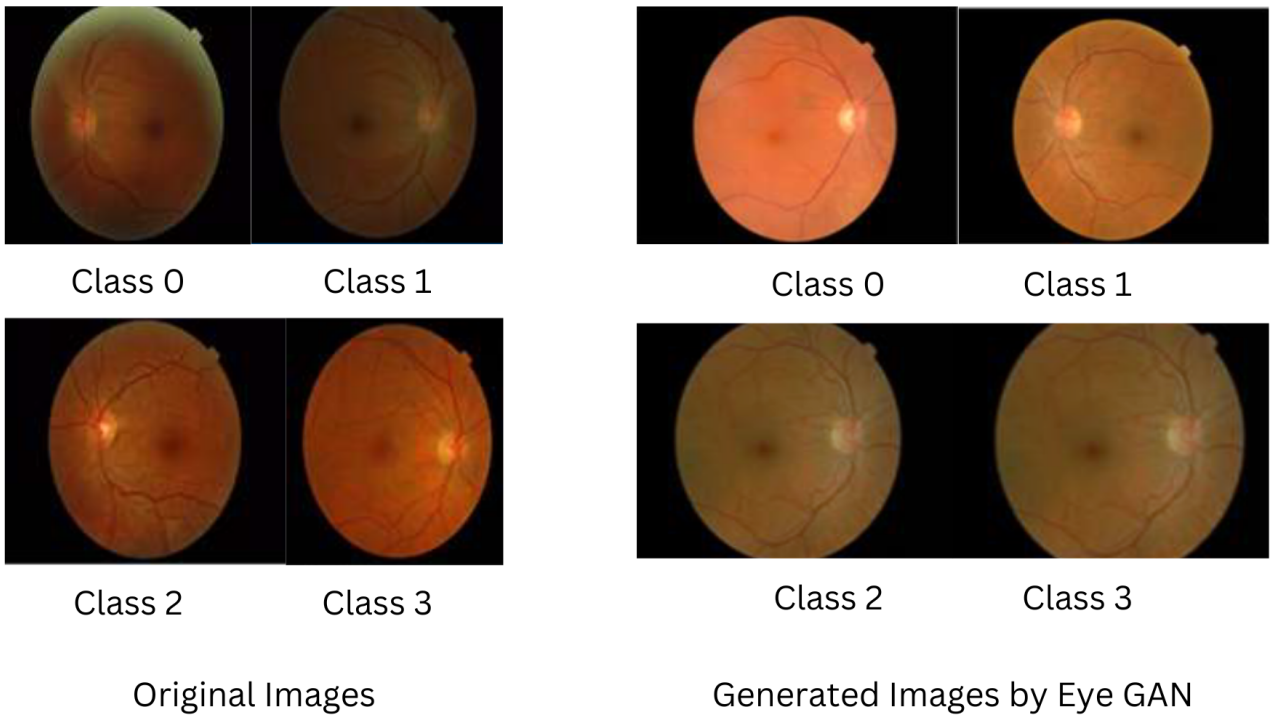

Figure 5: Comparison between original and synthetic images generated

It is observed that:

- The synthetic fundus images were noise-free.
- The EyeGAN produced real-like, good quality images for each class, thus providing an enriched dataset which can be used for improving the results by anyone.
- Both the datasets became class balanced with the application of EyeGAN.

Table 6: Data Distribution of the Studied Datasets

Dataset	Class	Number of Original Images	Images generated by EyeGAN	Total
Messidor I [10]	0	151	349	500
	1	30	470	500
	2	70	430	500
	3	149	351	500
	Total	400	1600	2000
Messidor II [10]	0	1017	0	1017
	1	270	230	500
	2	347	153	500
	3	75	425	500
	Total	1709	808	2517

5 Conclusions

With the need for a balanced and good quality dataset, attaining good accuracy of deep neural networks is conceivable. In the interest of expanding datasets and enhancing the accuracy of deep learning methods in Diabetic Retinopathy (DR) detection, a Conditional GAN-based model named EyeGAN is proposed in this study. The research is implemented on two datasets, MESSIDOR I and MESSIDOR II, which are class imbalanced datasets. By expanding the dataset, synthetic data augmentation increases its variability. EyeGAN attains the least Fréchet Inception Distance (FID) score in comparison with Cycle GAN and Style GAN. Our research demonstrates that the synthesized fundus images have qualities and properties that enrich the datasets with realistic and noise-free images. It can be concluded that the desired results are achieved. The model outperforms the existing GAN models and also balances the datasets studied. This will help in proving a strong system which will perform efficiently and overcome the problem of overfitting faced in fundus diseases detection and classification.

Declaration of Competing Interests

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Funding Declaration

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.

Author Contribution

Preeti Kapoor: Conceptualization, Methodology, Formal Analysis, Investigation, Resources, Data Curation ; **Shaveta Arora**:Writing - Review & Editing, Supervision.

References

[1] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[2] H. Lee, M. Ra, and W.-Y. Kim, "Nighttime data augmentation using gan for improving blind-spot detection," *IEEE Access*, vol. 8, pp. 48049–48059, 2020.

[3] B. Liu, C. Tan, S. Li, J. He, and H. Wang, "A data augmentation method based on generative adversarial networks for grape leaf disease identification," *IEEE Access*, vol. 8, pp. 102188–102198, 2020.

[4] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020.

[5] L. Lan *et al.*, "Generative adversarial networks and its applications in biomedical informatics," *Frontiers in Public Health*, vol. 8, p. 164, 2020.

[6] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[7] Q. Jin, X. Luo, Y. Shi, and K. Kita, "Image generation method based on improved condition gan," in *2019 6th International Conference on Systems and Informatics (ICSAI)*, pp. 1290–1294, IEEE, 2019.

[8] Y. Ikeda, K. Doman, Y. Mekada, and S. Nawano, "Lesion image generation using conditional gan for metastatic liver cancer detection," *Journal of Image and Graphics*, vol. 9, no. 1, 2021.

[9] D. Li, W. Xie, B. Wang, W. Zhong, and H. Wang, "Data augmentation and layered deformable mask r-cnn-based detection of wood defects," *IEEE Access*, vol. 9, pp. 108162–108174, 2021.

[10] X. Cao, H. Wei, P. Wang, C. Zhang, S. Huang, and H. Li, "High quality coal foreign object image generation method based on stylegan-dsad," *Sensors*, vol. 23, no. 1, p. 374, 2022.

[11] P. Kapoor and S. Arora, *Applications of Deep Learning in Diabetic Retinopathy Detection and Classification: A Critical Review*, vol. 91 of *Lecture Notes on Data Engineering and Communications Technologies*, pp. 505–535. Springer Singapore, 2022.

[12] J. Yang, Z. Zhao, H. Zhang, and Y. Shi, "Data augmentation for x-ray prohibited item images using generative adversarial networks," *IEEE Access*, vol. 7, pp. 28894–28902, 2019.