Application of Generative Adversarial Network for Data Augmentation in Diabetic Retinopathy

Anuj Kumar
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
anuj.qtx@gmail.com

Rimjhim Padam Singh
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
ps_rimjhim@blr.amrita.edu

***Abstract*—Image generation through generative adversarial network (GAN) has been a well-studied problem. It has been specifically implemented in medical domain for a number of use cases but training GAN has its own challenges. Application of GAN on generation of eye fundus images for the disease diabetic retinopathy (DR) has not been explored in depth. This paper shows a novel approach to generate low resolution eye fundus images for DR using GAN and validate it though a CNN classifier using samplers and schedulers module. The proposed approach beats the current benchmark in terms of generated image quality as well as classification accuracy. The low-quality generated images of 224*224 pixels manages to retain the key aspects of DR which aids in identifying the DR grade during classification predominantly from generated images only.**

Keywords—GAN, DCGAN, data augmentation, diabetic retinopathy, medical imaging, low resolution GAN, MESSIDOR

I. INTRODUCTION

GAN or Generative adversarial network is a popular technique for artificial image generation. Diabetic Retinopathy (DR) is a lethal disease of eye which is caused due to diabetes and may lead to blindness. When the lesions in DR overlays the central area of the eye, significant vision loss could occur. This progression of DR is referred to Diabetic Macular Edema or DME in the medical domain. Deep-learning based algorithms have been extensively used in eye disease treatment for tasks such as diabetic retinopathy classification and grading and also macular degeneration (AMD) diagnosis due to aging amongst patients of different age groups and ethnicity. Deep Learning methods require huge amount of data (images) for model training and this dataset is also required to be manually labelled by human experts before it can be used. DR is classified into five categories based on the seriousness of the condition and often the data of rarer grades are not available in abundance, so model training for classification or grading becomes difficult. Many works done in this area tried to implement a data augmentation strategy, but it did not attain the expected level of success due to huge effort involved in annotation and validation with human experts. Legal concerns about patient privacy have also been a limiting factor. Such development and road blocks led to the idea of creating synthetic retinal pictures that are very close to the real data. GANs are an effective framework for creating synthetic datasets in a variety of domains and thus it has been used towards retinal fundus image generation as well. GANs have been getting a lot of attention in ophthalmology [1] due to its versatility and high-quality image generation. Other well known augmentation methodologies such as transposing, rotation, translation. cropping, and flipping etc. could alleviate the issue in cases where at least a few hundred samples are available, but such methods lack the variety in sample creation and thus restrict the model performance. GAN on the other hand can generate new pictures of retina or eye fundus from a set of real images which are not just copy of original images but has its own characteristics.

Earlier, Diabetic Retinopathy (DR) were studied using machine learning methods where relevant features were extracted depending upon domain knowledge and image processing steps and, then labelled with the help of human experts followed by training the machine learning classification models. The performance of such traditional methods often depends on the feature extraction methods. It required huge manual effort and chances of error always used to be high. Then, deep learning technology became successful and it was extensively implemented in medical image analysis where low level image features such as edge detection, texture analysis and segmentation masks were studied. Such methods could learn the high-level meta features from images naturally from image pixels and labels.

Previously, machine learning ensemble models were also trained to classify diabetic retinopathy after extracting pixel level features and converting the same to tabular data [2]. Texture-based features like LBP have recently been used by some researchers for DR detection [3]. However, deep learning models are known to have significant dependency on a huge amount of data and labels, which is often difficult to obtain in the medical domain. Such lack of good amount of labelled data makes DL methods hard to apply in such use cases. On the contrary, hospitals and clinics can produce a large number of un-labelled data which may act as an important source of information. Therefore, it makes more sense to employ semi-supervised learning such as GAN for DR classification task. As for semi-supervised learning, GAN based models works very well in a various kind of applications for medical image synthesis and also achieves state of the art performance. The efficient technique DCGAN implemented in this paper, expands the idea of GAN by concatenating several layers of convolution operations which upscale the images for low resolution pictures. For quantitative validation and authentication of GAN generated images, first a classification model is trained to grade DR images using real data and then gradually real data are removed and replaced with fake images and classification models are again trained. The two models are compared to show how fake images affect the classification accuracy

II. LITERATURE SURVEY

Synthetic generation of accurate fundus pictures used to be a difficult job in pre-deep learning era. Earlier, such problems were solved using state of the art mathematical models of ocular anatomy [1] but now due to the advancements in computing capacity and artificial intelligence, machine learning can now be applied on an architecture of neural networks with very deep layers resulting into unprecedented applications. GANs are useful frameworks for image generation in modern days [4].

As described in above section, Reddy, G. T et al. (2020) [2] used manual feature extraction from DR images and then trained multiple ML based methods for the prediction as well as detection of DR. They proposition was a model ensemble the combines the existing machine learning approaches such as Random Forest classifier, XGBoost, K-Nearest Neighbor classifier, Decision Tree Classifier, Boosting methods like Adaboost classifier, and Logistic Regression classifier. This ensembled model improved the classification accuracy but low sample for rarer class remained a challenge.

Chetoui, M. et al. (2018) [3] proposed an approach that extracts texture features for classification of DR. They extract features like Local Ternary Pattern i.e. third order image features and Local Energy-based Shape Histogram from the area of digital image processing and they reported a result for 99% precision and recall using Support vector machine with degree 3 polynomial kernel. But they used only a small sample of images for feature extraction, feature selection and model building and training.

Goodfellow et al. (2020) [4] introduced GAN, an unsupervised deep learning method for image generation, consisting of two blocks: a discriminator and a generator. The generator learns from random noise to obtain the data distribution similar to original data and produces a very realistic image from it. After that, the discriminative model tries to differentiate between real and fake pictures by computing the probability that a sample generated from random noise is coming from the original distribution of data or from the generator. Such a simple and intuitive idea found a strong application and GAN became the de facto method for image generation until recently.

Wang, S et al. (2020) [5] proposed a multiple channel generative adversarial network (M-GAN) with semi supervision to classify and grade DR images. The multiple channel model proposed by them generated many images each corresponding to a specific DR feature. They use high resolution 1024*1024 fundus images and detect the those features which cause lesion with the help of three channels in GAN. They demonstrate such models can grades DR images effectively on the public datasets Messidor. However, the algorithm is slow and prone to overfitting and also mode collapse was observed for some scenarios.

Bhatia, K et al. (2016) [6] provides a detailed overview of machine learning and neural network classification methods on DR dataset and thus acts a good starting point. They do not specifically describe about GAN and its implementation for DR, but summarizes the body of work before GAN in this direction.

Priya, R et al (2013) [7] implements three Bayesian models namely Probabilistic neural network, Support vector machine and Bayesian multi-level classification to diagnose diabetic retinopathy and their performances are compared. They show that their algorithms work well in imbalanced data setting as well and features that they extract play a significant role in classification.

Roychowdhury, S et al. (2013) [8] develops a novel idea of combining two-step hierarchical and classification algorithm where the false positives are computed and rejected in the first phase and then in subsequent step, they classify the bright, high intensity spots on eye fundus or lesions as hard exudates - familiar terms in medical domain for DR diagnosis. They used SVM for classification.

Zhou, Y. et al. (2020) [9] proposed DR-GAN short for diabetic retinopathy generative adversarial network to generate eye fundus images of higher resolution which could be perturbed with any type of grading levels or lesion information. They used this method to generate largescale artificial data samples which then were used for augmentation to train and classify lesion segmentation model. Their method used the generator to condition on the structural pattern of lesion masks, and controlled the image generation by grading vectors which are sampled from the high dimensional latent space. They also developed a multi-scale spatial as well as channel wise attention layer so that the generated images have good generalization ability to take into account even the small details of disease condition. They used the joint adversarial loss to optimize the whole network.

Niu, Y et al. (2021) [10] implemented an explainable AI framework for DR grading. They use the domain learning first to define a novel pathological descriptor using activated and non-activated perceptron's of the DR discriminator from GAN to formulate channel wise information of fundus lesions present. After that, the symptoms grabbed by the descriptor are visualized by creating another network called Patho-GAN. They iteratively manipulate these descriptors one at a time, and thus control the x, y co-ordinates such as position, location and also quantity as well as types of lesions generation. They also demonstrate that the generated images show the same symptoms of diabetic retinopathy as original ones. Their solution is also faster compared to previous work done.

Xiao, Q et al. (2019) [11] approached the problem from image segmentation perspective and developed an edge detector method called HEDNet to implement the segmentation and classification approaches. They created pixel by pixel masking of DR grades by first using HEDNet and fused it with Conditional Generative Adversarial Network (cGAN). They used a custom loss metric that combines the adversarial loss and segmentation loss. They showed when the adversarial loss is added into segmentation loss, the lesion segmentation performance improved over the baseline by a good margin.

J.-Y. Zhuet al. (2017) [12] showed image-to-image translation using cycleGAN in which case one domain to other images can be translated and provided some critical hints which can be utilized for DR grading and generation process too.

A. Brock et al. (2019) [13] provided an approach to ensemble together some of best conditional image generation techniques for training class-conditional images. It helps generation of high-resolution as well as high-fidelity images. Based on the conditional GAN frameworks they proposed a retina generator which synthesizes realistic images of high-resolution.

Chong, M. J et al. (2020) [14] provides a comprehensive argument to evaluate the metric to measure the quality of generated images. They suggested two different evaluation metrics for generative models, first the FID score or Fréchet

Inception Distance score and second the IS score or Inception Score. They take into account the bias term to measure the quality of synthetic images. Bias is defined as expectation or mean FID score computed for a finite number of sample set which may or may not be equal to the ground truth score. The paper argues that the comparisons of generated images using FID or IS are not reliable in many situations under bias.

Shmelkov, K et al. (2018) [15] provides a very comprehensive overview on how to measure the quality of generated images using GAN. Before such metrics, generated images were validated by visual inspection. But such traditional methods are often insufficient and in adequate. They introduced two measures- GAN-train and GAN-test, which approximate the metric recall for diversity of image generation and precision for quality of the image respectively. A number of recent GAN models are evaluated based on these two measures and shows a clear difference in performance.

III. Dataset And Pre-Processing

A. *Dataset Exploration*

A well-known public dataset for diabetic retinopathy MESSIDOR [16] is used for this study. This database contains about 1200 images of the retinal fundus. A few samples of images is shown in Fig. 1. For image collection, researchers used a specific camera mounted on top of a non-mydriatic retinograph at 45 degrees angle. The images were captured in three different resolutions of 1440*960 pixel, 2240*1488 pixel and 2304*1536 pixels. Each image was labeled by medical experts and they noted the detailed diagnosis corresponding to each image and weather it contains DR or not. The experts also provided a retinopathy grade wherever detected. Out of 1200 images, 540 images were labelled as normal, next 153 images as mild DR, 247 as moderate DR, and rest 260 as severe as described by E. Decencière et al (2014) [16]. The distribution of classes and images are shown in Table I.

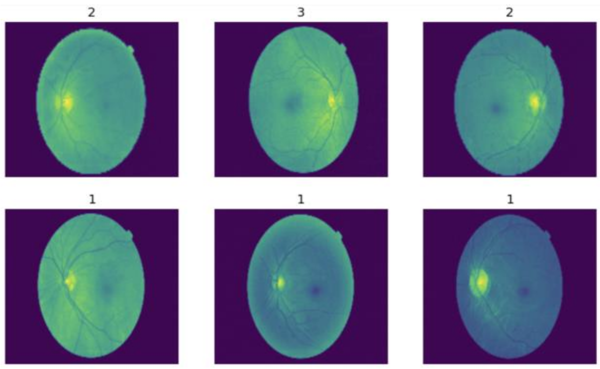

Fig. 1. Sample images from MESSIDOR dataset.

TABLE I. Class distribution of dataset based on DR grade

Grade	Severity	No. of images
R0	Normal	540
R1	Mild	153
R2	Moderate	247
R3	Severe	260

Each image from the dataset is categorized among four lesion grades namely R0, R1, R2, and R3, as mentioned in Table I. ‘R0’ category is for normal images with no lesion fundus. ‘R1’ and ‘R2’ is for the mild and severe variety of DR but a non-proliferative retina and lastly ‘R3’ is for the proliferative retina. R0, R1, R2, and R3 grades individually account for 45.5%, 12.75%, 20.58%, and 21.67% of the total data set as shown below in Table II.

TABLE II. Classification of image types

Class labels	Descriptions	No. of images
DR	R0/R1/R2/R3	546/153/247/254
Normal/Abnormal	R0/R1, R2, R3	546/654
Non -referable / referable	R0, R1/R2, R3	699/501

B. *Pre-processing*

The original images have resolution of 1440*960 which are resized to 224*224 and pixel intensity is normalized between (0,1). Images are resized so that model becomes computationally efficient. We show that reducing the size does not impact the detection in any adverse way rather it makes the model faster and light-weight. A train-test split of 70:30 is performed for all cases.

IV. Models

A. *Solution Architecture*

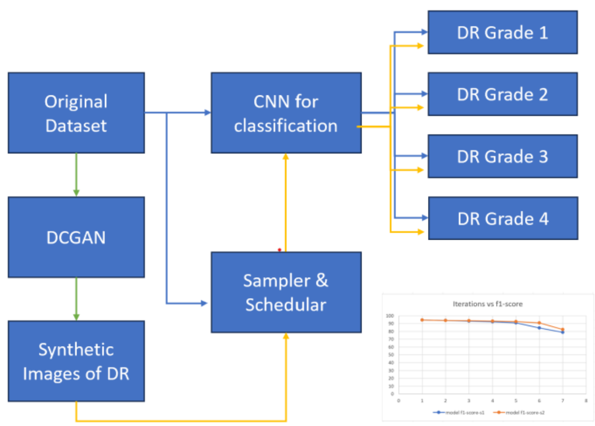

Fig. 2. Solution architecture for DR detection using GAN generated eye fundus images.

Fig. 2 describes the solution architecture for diabetic retinopathy classification on synthetic eye fundus images. First, a CNN model is trained for eye fundus classification using real images from MESSIDOR dataset. Accuracy and F1-Score obtained using this model is used as a benchmark for comparisons in subsequent step. Next, a deep conditional GAN network is trained to generate artificial samples similar to real images from MESSIDOR dataset. To validate the accuracy and quality of generated images, these images are mixed with real images in an iterative process and the CNN classification model is re-trained using same hyper-parameters. The accuracy and F1-score of second iterations of CNN are compared with first model instance. The iterative process of mixing of real and generated images is controlled using sampler and schedular block. The details on these modules are discussed below.

B. *GAN Architecture*

Fig. 3 describes the GAN architecture for DR eye fundus image generation. Generator is a 128 sized noise vector which learns to generate the eye fundus images carrying characteristics of diabetic retinopathy from the original images supplied though a discriminator block.

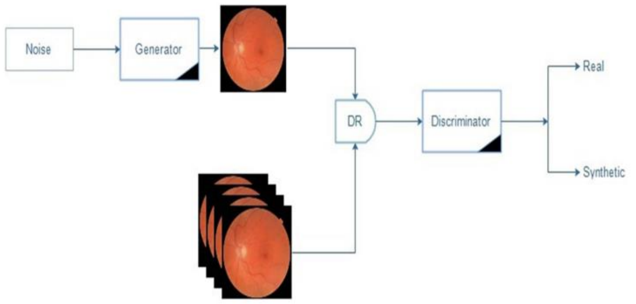

Fig. 3. Block diagram of GAN architecture.

The role of discriminator is to learn the meta features of these images for classification and detect any generated images from generator as fake or artificial. The role of generator is to beat the discriminator in detection. In comparison to uncontrolled GAN, DCGAN creates a deep convolutional layer for better feature understanding in case of discriminator and learning of such features in case of generator. The labels of original images are also passed so that the architecture learns to generate images specific to each class. The model parameters of GAN network, discriminator and generator as provided in Table III.

TABLE III. MODEL PARAMETERS OF GAN

Parameter Name	Value
Discriminator learning rate	0.00025
Generator learning rate	0.00015
Epoch	200
Optimizer	Adam
Loss function	Categorical cross-entropy
Latent dimensions	128
Batch size	32
Number of layers in Generator	14
Number of layers in Discriminator	12
No. of parameters in Generator	2171085
No. of parameters in Discriminator	519105

The model architecture of generator and discriminator is described in Fig. 4 and 5 respectively. The discriminator implements a down sampling convolution network to reduce the original image size from 224 * 224 succcessively to a fully connected layer of 4 neurons for multi-class classification. On the other hand, the generator network starts with a random noise of 128 dimensions and implements a transposed convultional network to convert this noise vector into a 224* 224 image. The final layer of generator outputs an image of size 224*224. The GAN network is iteratively trained with different settings of learning rates, epochs and batch size. The most performant configuration is described in Table III. The network performs the best when the learning rate of generator is kept slightly smaller than that of the discriminator. The model starts converging after about 150 epochs.

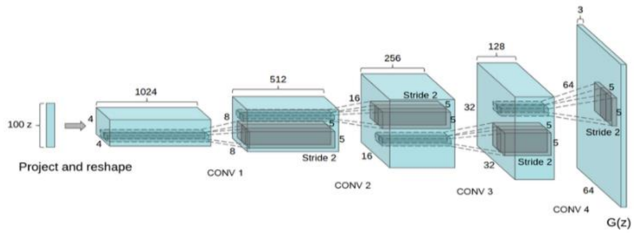

Fig. 4. Generator model architecture for DR eye fundus image generation.

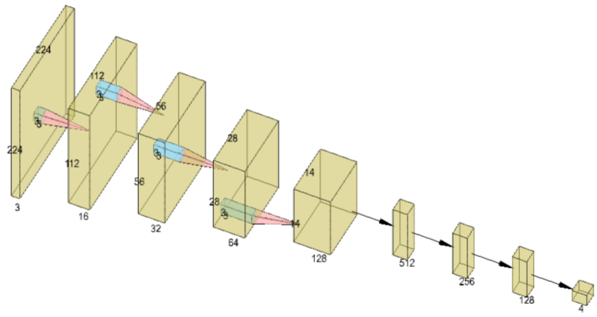

Fig. 5. Discriminator model architecture for DR eye fundus image generation.

C. *CNN Classification Model*

CNN or Convolutional neural network model is designed and trained for eye fundus image classification into DR grading. The model architecture is described in Fig. 6. It implements 11-layer architecture with convolution and max-pooling layers. The total count of model parameters in the network is 26081092. The model is run for 50 epochs, batch size is 32 and optimizer learning rate is 0.001, with Adam optimizer and categorical cross-entropy loss as loss function.

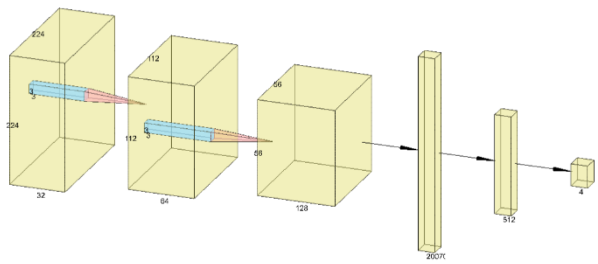

Fig. 6. CNN model architecture.

D. *Training Procedure*

This section describes the detailed training process for the complete solution. CNN and GAN architecture along with the parameters have been described above. The most important aspect which brings the novelty in this work is sampler and schedular modules which have been described in this section. Sampler and schedular is a mechanism to iteratively add more and more synthetic eye fundus images, remove original image and train CNN classifier as shown in Table IV. Initially 10% original images are removed and, in its place, 10% synthetic images are added for the same class and then the percentage of generated images are increased to 20%, 50%, 70% and 90%. In one iteration, the entire 100% images used are synthetic and model performance is measured on the same. Two types of sampling methods are used: uniform and proportional. In case of uniform sampling, ratio of each of the four class is kept same i.e. for 10% scenario, out of 120 synthetic images, 30 images are from each of the four class. Whereas in case of proportional sampling, the ratio of classes is kept same as original training dataset classes.

E. Evaluation Metrics

Precision, recall, accuracy and F1-score are used to measure the CNN classification model performance [17, 18]. The generated images thorough GAN is measured using FID score. F1-score is given preference over accuracy due to imbalance in original training dataset. Therefore, most results are described in terms of F1-score in results section.

TABLE IV. SAMPLER AND SCHEDULAR CONFIGURATION

Iteration	gen. image %	Sampling type	# fake images	# real images
1	0	Uniform	0	1200
2	10	Uniform	120	1080
3	20	Uniform	240	960
4	50	Uniform	600	600
5	70	Uniform	840	360
6	90	Uniform	1080	120
7	100	Uniform	1200	0
8	10	Proportional	120	1080
9	20	Proportional	240	960
10	50	Proportional	600	600
11	70	Proportional	840	360
12	90	Proportional	1080	120
13	100	Proportional	1200	0

V. RESULTS AND ANALYSIS

Generator and discriminator loss for GAN training is shown in Fig. 7. Multiple GAN trainings are done with epoch 50, 100, 150 and 200. Best result is obtained with least generator loss, least discriminator loss and low FID is observed for 200 epochs run. Fig. 7 shows the loss curve for 200 epochs run. Initially, generator loss is quite high and slowly it tries to converge after a few oscillations in the epoch range 25-50. The loss starts converging after epoch 150 and the gap between generator loss and discriminator loss becomes minimum at around epoch 100 and 180. The FID score is shown in Table V and sample of generated images are shown in Fig. 8. Sample synthetic images corresponding to same class labels as shown in Fig. 1 are shown in Fig. 8 for comparison purpose. Due to small image size of 224*224, the generated images are slightly pixeled. The GAN architecture that was described above is much smaller compared to current GAN models available but still the generated images do a very good job doing classification of DR grades as recorded in Table VI. The algorithm performance drops slightly only when the percentage of generated images are more than 50% compared to the original images during classification.

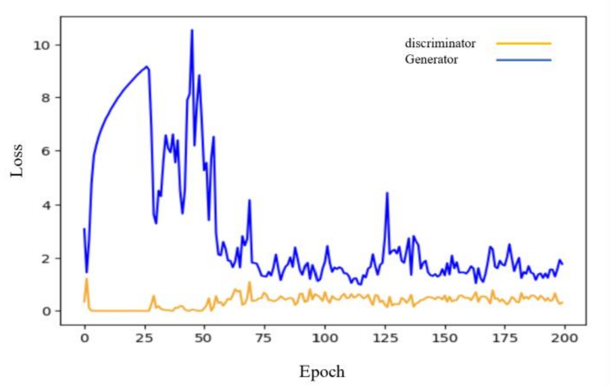

Fig. 7. Discriminator and Generator loss for eye fundus generation.

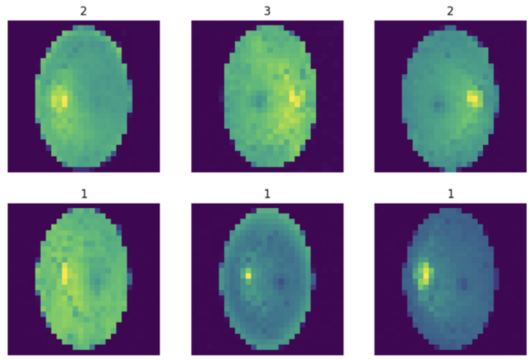

Fig. 8. Sample of generated images from DCGAN on MESSIDOR dataset.

TABLE V. FID SCORE FOR GAN GENERATED EYE FUNDUS IMAGES

Epochs	FID
50	9.126
100	5.854
150	3.546
200	1.854

The Fig. 9 shows the f1-score for iteration 1 to 6 corresponding to 10% to 100% synthetic image mixing up with real images for DR grade classification for two sampling strategy. Proportional sampling (model f1-score-s2) shown with orange color performs slightly better compared to uniform sampling (shown in blue). The performance of both sampler drops slightly when percentage of synthetic images increase more than 50% (iteration 6: 70 % synthetic and iteration 7: 90% synthetic) as visible in the fig. 9. Further, the classification accuracy after mixing generated images to real images is compared with the benchmark result obtained from other researchers and the same is shown in Table VII.

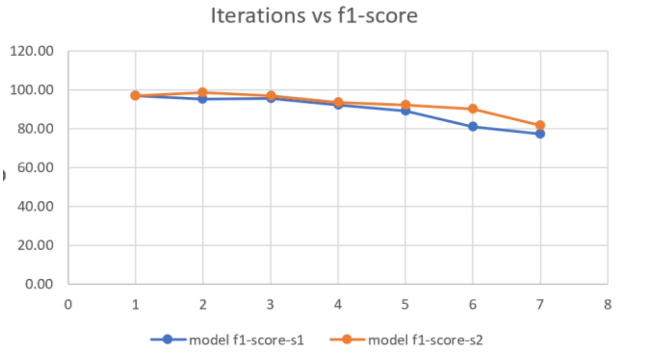

Fig. 9. F1-score comparison for uniform and proportional sampling strategy for schedular.

TABLE VI. CLASSIFICATION SCORE ON GENERATED EYE FUNDUS IMAGES

gen. image %	Sampling type	# fake images	# real images	F1 Score (%)	Acc. (%)	Pre. (%)	recall (%)
0	uniform	0	1200	96.85	96.28	98.4	99.45
10	uniform	120	1080	95.08	95.86	97.3	99.33
20	uniform	240	960	95.58	94.91	98.9	98.11
50	uniform	600	600	92.33	93.44	98.4	96.33
70	uniform	840	360	89.06	92..9 9	95.2	97.22
90	uniform	1080	120	81.10	89.21	94.2	92.32
100	uniform	1200	0	77.40	84.46	90.2	93.45
10	pro.	120	1080	98.51	97.34	99.2	99.8
20	pro.	240	960	97.03	96.44	98.6	99.4
50	pro.	600	600	93.54	95.6	97.5	98.1
70	pro.	840	360	92.25	94.3	96.5	98.2
90	pro.	1080	120	90.28	91.2	95.8	97.5
100	pro.	1200	0	81.64	86.5	90.2	96.8

Our model performs much better compared to other result reported on MESSIDOR dataset using GAN based techniques. Deepfake paper [19] has slightly higher accuracy compared to our result but its F1-Score is relatively very poor. Deepfake paper used up to 70% generated images during classification and hence, we have also reported out result on 70% mixing ratio for a fair comparison as explained in Table VI.

TABLE VII. COMPARISON OF CLASSIFICATION ACCURACY WITH BENCHMARK

Model	Accuracy	F1-Score
VSG-GAN [19]	67%	79%
DR-LL GAN [20]	94.2%	Not reported
Deepfakes [21]	98.19%	60.1%
Our Model	**94.3%**	**92.25%**

VI. CONCLUSION

Image generation through GAN has been solved extensively in literature and yet GAN is notoriously difficult to train for many scenarios. Generating artificial images for diabetic retinopathy using GAN is one such case. The implementation described in this paper attempts to solve this problem and does a very good job at it. We show that the classification performance with synthetic eye fundus images is as good as with original images. The model can be further improved for cases with more than 50% synthetic images if image resolution is increased to a higher resolution.

REFERENCES

[1] Abbood, S. H., Abdull Hamed, H. N., Mohd Rahim, M. S., Alaidi, A. H. M., & Salim ALRikabi, H. T. (2022). DR-LL Gan: Diabetic Retinopathy Lesions Synthesis using Generative Adversarial Network. International Journal of Online & Biomedical Engineering, 18(3).

[2] Reddy, G. T., Bhattacharya, S., Ramakrishnan, S. S., Chowdhary, C. L., Hakak, S., Kaluri, R., & Reddy, M. P. K. (2020, February). An ensemble based machine learning model for diabetic retinopathy classification. In *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)* (pp. 1-6). IEEE.

[3] Chetoui, M., Akhloufi, M. A., & Kardouchi, M. (2018, May). Diabetic retinopathy detection using machine learning and texture features. In 2018 IEEE Canadian conference on electrical & computer engineering (CCECE) (pp. 1-4). IEEE.

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[5] Wang, S., Wang, X., Hu, Y., Shen, Y., Yang, Z., Gan, M., & Lei, B. (2020). Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision. IEEE Transactions on Automation Science and Engineering, 18(2), 574-585.

[6] Bhatia, K., Arora, S., & Tomar, R. (2016, October). Diagnosis of diabetic retinopathy using machine learning classification algorithm. In 2016 2nd international conference on next generation computing technologies (NGCT) (pp. 347-351). IEEE.

[7] Priya, R., & Aruna, P. (2013). Diagnosis of diabetic retinopathy using machine learning techniques. ICTACT Journal on soft computing, 3(4), 563-575.

[8] Roychowdhury, S., Koozekanani, D. D., & Parhi, K. K. (2013). DREAM: diabetic retinopathy analysis using machine learning. IEEE journal of biomedical and health informatics, 18(5), 1717-1728.

[9] Zhou, Y., Wang, B., He, X., Cui, S., & Shao, L. (2020). DR-GAN: conditional generative adversarial network for fine-grained lesion synthesis on diabetic retinopathy images. IEEE Journal of Biomedical and Health Informatics, 26(1), 56-66.

[10] Niu, Y., Gu, L., Zhao, Y., & Lu, F. (2021). Explainable diabetic retinopathy detection and retinal image generation. IEEE journal of biomedical and health informatics, 26(1), 44-55.

[11] Xiao, Q., Zou, J., Yang, M., Gaudio, A., Kitani, K., Smailagic, A., ... & Xu, M. (2019, August). Improving lesion segmentation for diabetic retinopathy using adversarial learning. In International Conference on Image Analysis and Recognition (pp. 333-344). Cham: Springer International Publishing.

[12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2223–2232.

[13] A. Brock, J. Donahue, and K. Simonyan, “Large scale GAN training for high fidelity natural image synthesis,” in Proc. Int. Conf. Learn. Representations, 2019.

[14] Chong, M. J., & Forsyth, D. (2020). Effectively unbiased fid and inception score and where to find them. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 6070-6079).

[15] Shmelkov, K., Schmid, C., & Alahari, K. (2018). How good is my GAN?. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 213-229).

[16] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J-C. Klein. "Feedback on a publicly distributed database: the Messidor database," Image Analysis & Stereology, v. 33, n. 3, p. 231-234, 2014.

[17] Singh RP, Sharma P. “A light-weight change detection method using ycbcr-based texture consensus model”. *International Journal of Pattern Recognition and Artificial Intelligence*. 2020 Aug 6;34(09):2050023.

[18] Vasista, T. Sai, Rimjhim Padam Singh, and Priyanka Kumar. "Advancing Healthcare: Unleashing the Potential of Cryptography, Blockchain, and Machine Learning." In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2023.

[19] Liu, J., Xu, S., He, P., Wu, S., Luo, X., Deng, Y., & Huang, H. (2024). VSG-GAN: A High Fidelity Image Synthesis Method with Semantic Manipulation in Retinal Fundus Image. *Biophysical Journal*.

[20] Abbood, S. H., Abdull Hamed, H. N., Mohd Rahim, M. S., Alaidi, A. H. M., & Salim ALRikabi, H. T. (2022). DR-LL Gan: Diabetic Retinopathy Lesions Synthesis using Generative Adversarial Network. *International Journal of Online & Biomedical Engineering*, *18*(3).

[21] Salini, Y., & HariKiran, J. (2022). Deepfakes on retinal images using GAN. *International Journal of Advanced Computer Science and Applications*, *13*(8).