The Conception of Fundus Multi-Disease Dataset (FMDD) using Multi-Spectral Generative Adversarial Networks

Karthika Gidijala
Department of Computer Science and Engineering, School of Technology, GITAM Deemed to be University, Hyderabad, Telangana, India
kgidijal@gitam.edu (corresponding author)

Vijaya Kumar Sagenela
Department of Computer Science and Engineering, School of Technology, GITAM Deemed to be University, Hyderabad, Telangana, India
vsagenel@gitam.edu

Received: 8 December 2024 | Revised: 8 January 2025 | Accepted: 19 January 2025

 DOI: https://doi.org/10.48084/etasr.9874

ABSTRACT

The World Health Organization (WHO) reports that 2.2 billion people are affected by visual impairment. Early detection and diagnosis of ocular pathologies can help predict visual impairment. Over the past twenty years, many fundus image datasets have become publicly available due to technological advances. These datasets have primarily focused on the detection of common ocular pathologies such as age-related macular degeneration, diabetic retinopathy, and glaucoma, and recent research in fundus diseases has highlighted the importance of detecting multiple fundus diseases. Existing datasets such as ARIA and RFMiD mainly contain images of the most common pathologies and very few images related to rare pathologies. The existing public datasets have problems in multi-disease classification, such as less data under some under-represented diseases, low-quality photos, and class imbalance among several classes. The main objective of our research is to construct a Fundus Multi-Disease Dataset (FMDD) with 20 courses of ocular diseases from publicly available datasets and with the application of Multi-Spectral Generative Adversarial Networks (MSGANs). The resulting dataset is balanced for all image classes.

Keywords-Fundus Multi-Disease Dataset (FMDD); ocular pathologies; generative adversarial networks; image augmentation; visual impairment

I. INTRODUCTION

Deep neural networks have been widely used to detect Diabetic Retinopathy (DR), a leading cause of visual impairment. Techniques such as principal component analysis and support vector machines have shown promise, whereas AlexNet-based deep convolutional neural networks excel at extracting features from retinal fundus images [1]. Research has focused primarily on common eye diseases such as DR, glaucoma, and age-related macular degeneration. However, existing systems often overlook rare diseases due to unbalanced datasets like RFMiD and ARIA [2-8]. These datasets prioritize prevalent diseases, with limited representation of rare pathologies, poor image quality, and class imbalance, all of which affect model performance. Generative Adversarial Networks (GANs) have been pivotal in medical imaging for tasks like class imbalance, image enhancement, and synthetic image generation. Techniques such as CycleGAN, StyleGAN, and Progressive Growing GANs are widely used. CycleGAN excels in image-to-image translation (e.g., CT to MRI), StyleGAN generates high-resolution synthetic images for rare pathologies, and Progressive Growing GANs improve image resolution and stability, thereby increasing the quality of the dataset for deep learning models [9]. To address the limitations of currently available datasets, this study introduces a Fundus Multi-Disease Dataset (FMDD) balanced with synthetic images generated by GANs. It organizes the findings into a literature review, disease details, augmentation algorithms, and results. Advances in deep learning have improved detection accuracy (~96.5%) for diseases such as DR, but challenges remain in identifying coexisting pathologies. Previous research highlights high sensitivity and F1 scores for common diseases, but neglects rare pathologies. This study addresses issues such as class imbalance, lesion localization, and system trustworthiness, and emphasizes preprocessing and data augmentation to create robust, interpretable models for comprehensive ocular disease detection.

II. DATASET DESCRIPTION

The identification of multi-disease detection using retinal fundus images will be performed when the dataset contains high-quality images and class balance among multiple classes. To conduct our research on fundus multi-disease detection, we construct a dataset by collecting publicly available datasets with high-quality images, such as RFMiD, DRIVE, STARE, CHASE_DB1, and ARIA [3-8]. From these datasets, we collected about 4000 images from 39 classes. Some of the classes contain very few images, so we have converted these 39 classes into 4 classes of fundus diseases. For each of these 4 classes, we apply data augmentation and Multi-Spectral Generative Adversarial Networks (MSGANs) to maintain class balance among them. A brief description of these 4 classes of diseases with their significant characteristics is given below:

1. DR: DR is a type of eye disease that affects a person's vision. Anyone with type 1, type 2, or gestational diabetes can have DR. Over time, high blood sugar levels can damage the retina, but this can be resolved by lowering the diabetes. If not treated in the early stages, it can lead to blindness. The fundus images with DR consist of a retinal dot, hard exudates, and blot hemorrhages, as shown in Figure 1.
2. Branch Retinal Vein Occlusion (BRVO): BRVO causes blockage of small veins in the retina. The blockage is caused by narrowing of the blood vessels due to blood clots or pressure from a thickened artery crossing the vein. This blockage causes blood and fluid to leak onto the retina and swelling of the macula, which can lead to vision loss or blurred vision. The fundus images with BRVO consist of soft exudates, retinal edema, and dot hemorrhages, as shown in Figure 2.
3. Central Retinal Vein Occlusion (CRVO): In CRVO, the central retinal vein, which carries blood from the eye back to the heart, becomes blocked. Slowed blood flow, changes in the vein wall, and changes in the blood can all contribute to the formation of a blood clot. The blockage causes the walls of the vein to leak blood fluid into the retina. When this fluid accumulates in the macula, vision becomes blurred. The fundus images with CRVO consist of flame-shaped hemorrhages, as shown in Figure 3.
4. Age-related Macular Degeneration (AMD): AMD is an age-related process that affects the retina. If we think of our eye as a camera, the retina is like the film in camera. It is the back part of the eye that captures images and sends them back to the brain for processing. The part affected by AMD is the macula, the part of our retina that we use for central vision. The fundus image consists of the fovea, geographic atrophy, and the macular region with multiple drusen, as shown in Figure 4.

III. PROPOSED METHODS FOR FUNDUS DATA AUGMENTATION

The collection of fundus images from various publicly available datasets is organized into 4 classes. The main problems are the high level of class imbalance and the low number of samples for the underrepresented labels, which leads to poor robustness, lack of generalization, and reduced confidence in the predictions. To overcome these problems, we propose data augmentation using GANs and image augmentor.

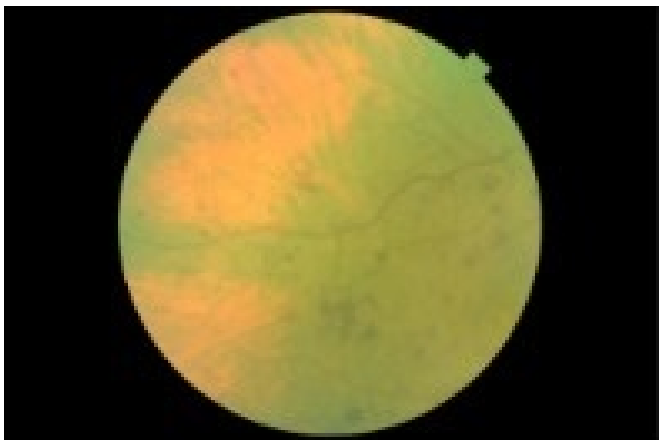


Fig. 1. Fundus image showing characteristics of DR, including retinal dots, hard exudates, and blot hemorrhages.

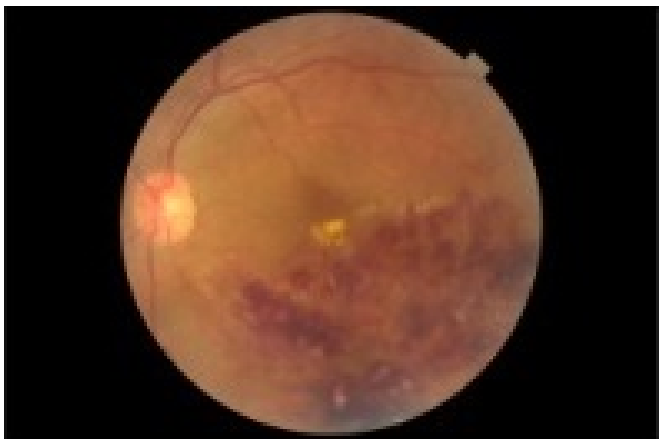


Fig. 2. Fundus image showing features of BRVO characterized by soft exudates, retinal edema, and dot hemorrhages.

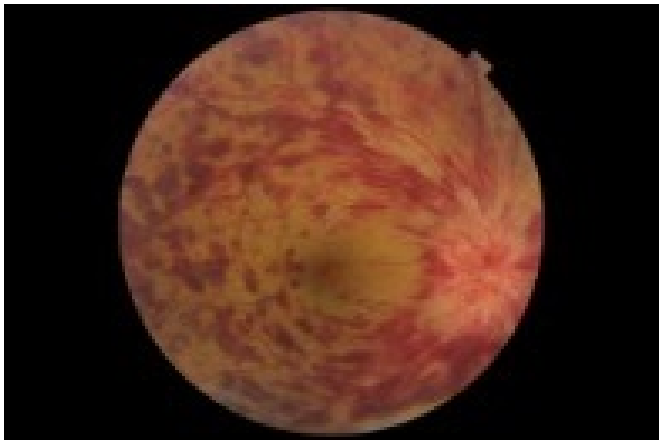


Fig. 3. Fundus image showing flame-shaped hemorrhages caused by CRVO.

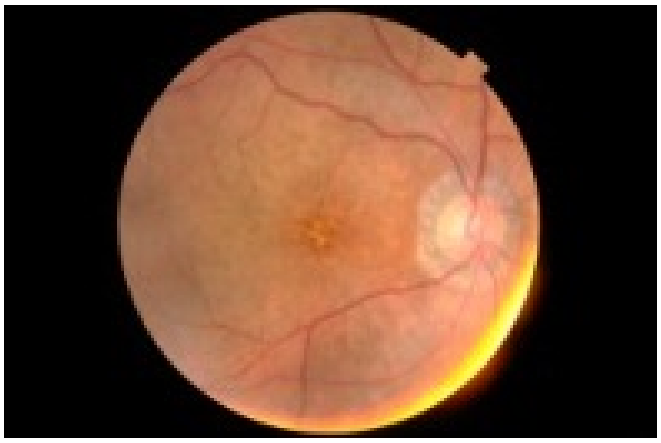


Fig. 4. Fundus image illustrating AMD with visible geographic atrophy and multiple drusen in the macular region.

A. Data Augmentation with Multi-Spectral Generative Adversarial Networks

GANs provide a general framework for generating high-quality, natural-looking images. In this paper, we propose an MSGAN for generating synthetic data to overcome the class imbalance problem. The general architecture of the proposed MSGAN is shown in Figure 5.

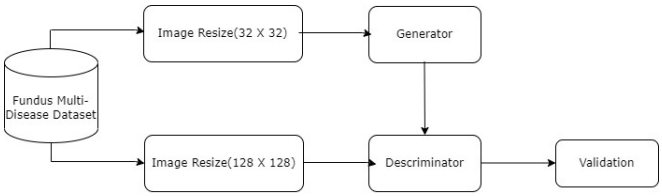


Fig. 5. General architecture of the MSGAN.

The main reason for choosing MSGAN is its ability to handle multi-class imbalances and generate high-resolution, realistic images. It incorporates residual blocks and perceptual loss for improved image quality and feature representation, addressing limitations in existing datasets like ARIA and RFMiD. The proposed MSGAN generates high-quality synthetic fundus images to address class imbalance in datasets. The generator transforms low-resolution images (32×32) into high-resolution outputs (64×64) using 16 residual blocks with Conv2D layers, batch normalization, and PReLU activation, followed by two upscaling layers. The discriminator classifies synthetic and real images using Conv2D layers with Leaky ReLU activation and dense layers for binary classification. Performance is evaluated using content loss (VGG16-based pixel similarity) and adversarial loss (guiding realism), combined as perceptual loss for better quality. This method enhances dataset diversity and supports robust fundus disease classification across underrepresented classes. This resizing facilitates efficient processing and improves the model's ability to generate high-quality synthetic images. The model is optimized using the Adam optimizer and employs perceptual loss to ensure that the generated images are realistic. The details of the generator and discriminator design are described below and illustrated in Figure 6.

```
Generator model design:
Input: A low-resolution image of 32×32
1. Number of residual blocks = 16
2. Layers= use Conv2D layer over the input
   image
     a. Feature maps = 64
     b. Kernel size = 9×9
     c. Activation function = pReLU
3. Residual block in building generator
   model
   For in range (residual blocks):
     a. Use two Conv2D layers over layers
     b. Feature maps = 64
     c. Normalization= Batch normalization
     d. Kernel size = 3×3
     e. Activation function = pReLU
     f. Return layers
4. Layers = use Conv2D layer over layers
     a. Feature maps = 64
     b. Kernel size = 3×3
     c. Normalization = Batch normalization
5. Upscaling block in building generator
   model
   Upscaling is done twice to match the
   size of the high-resolution image
   Layers= use Conv2D layer over layers
     a. Feature maps = 256
     b. Kernel size = 3×3
     c. Activation function = pReLU
     d. Optimizer = Adam
     e. Return upscaled model

Discriminator model design:
Input: A high-resolution image of size
128×128
1. Number of filters: fts = 64
2. Number of convolutions: con = 6
3. For I in range(con):
     a. Use Conv2D layer over the input
        image
     b. Kernel size = 3×3
     c. Activation function= Leaky ReLU
     d. Strides = 2
     e. Fts = fts*2
     f. Optimizer = Adam
     g. Return discriminator
4. Disc = Flatten the return discriminator
   layers
5. Disc = Dense(fts*16)*disc, Activation
   function = Leaky ReLU
6. Validity=Dense(1, activation
   function=sigmoid) disc
7. Output the validity
```

1) Loss Functions in Performance Evaluation

The use of loss functions in the generator and discriminator is critical in evaluating the performance of the proposed

MSGAN. In the proposed model, we have used the sum of the content loss and the adversarial loss. called perceptual loss.

- Content loss: The mean square error calculated for the pixels in each image is called the content loss. The content loss is obtained in the VGG16 model using the PReLU activation function. Here, $FM_{i,j}$ is a feature map defined over the j-convolution and i-max pooling layer in the VGG16 network. VGG loss is defined as the Euclidean distance between the synthetic image and the high-resolution image features. The content loss is defined in (1).

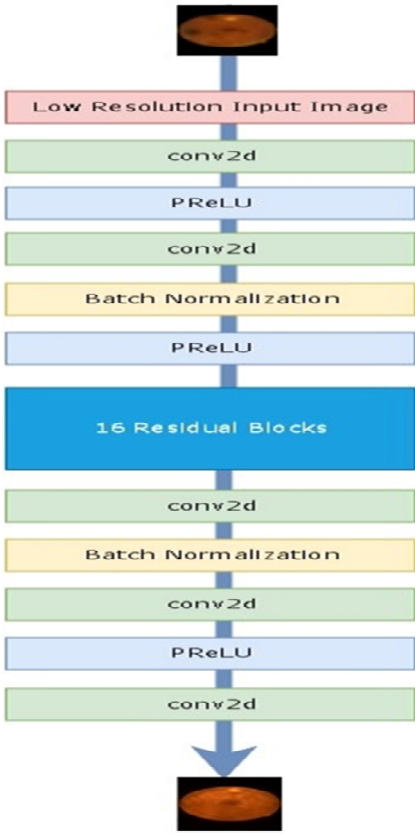
$$content_{loss_{i,j}} = y\frac{1}{SI_{i,j}HI_{i,j}} \cdot \sum_{x=1}^{SI_{i,j}} \sum_{y=1}^{HI_{i,j}} \left(FM_{i,j}\left(S_{I_{x,y}}\right) - FM_{i,j}\left(H_{I_{x,y}}\right)\right)^2 \quad (1)$$

- Adversarial loss: The adversarial loss is mainly used to fool the discriminator network. The adversarial loss, which is mainly defined as the generator loss over the probabilities of the misleading discriminator, discriminates synthetic images with authentic images on all the training examples. Here, $\mathrm{D}(G_{SI})$ denotes the probability of the generated synthetic image, and G_{SI} is the probability of a high-resolution raw image. The adversarial loss is defined in (2).

$$Adversarial\_loss_{SI} = \sum_{n=1}^{N} -\log(\mathrm{D}(G_{SI}) \quad (2)$$

The perceptual loss is defined as the sum of the content loss and adversarial loss.

(a)



(b)

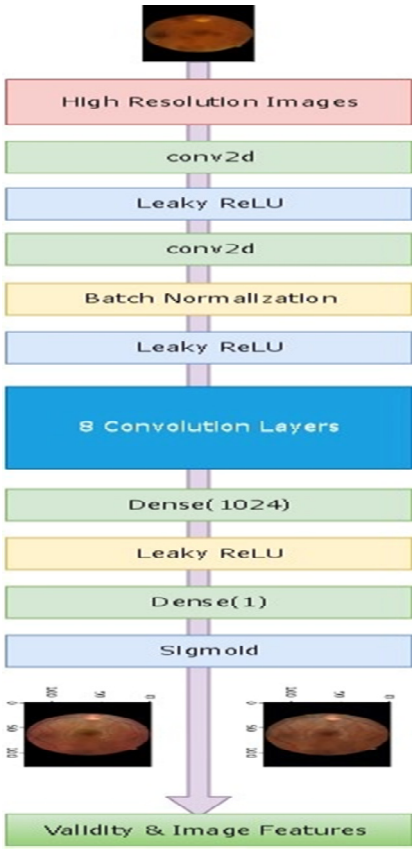


Fig. 6. Data augmentation with image augmentor: (a) generator and (b) discriminator.

IV. RESULTS

The proposed model utilizes MSGANs to balance fundus datasets across 4 imbalanced classes. Each class varies in quantity and resolution, and MSGANs are applied to generate synthetic images that resemble natural images. Key contributions include the creation of a balanced dataset for underrepresented ocular pathologies, overcoming the limitations of datasets such as ARIA and RFMiD. The model produces synthetic images that are validated by perceptual and adversarial loss metrics. Advanced augmentation techniques, including rotation and elastic distortion, further diversify the dataset. This approach supports comprehensive multi-disease detection models, facilitating early diagnosis of ocular diseases. Future work will improve the accuracy of GAN-based augmentation.

Figure 7 shows the comparison of original and synthetic fundus images for DR. The first two images (original-DR1 and original-DR2) represent the original fundus images, whereas the last two images (synthetic-DR1 and synthetic-DR2) are synthetic images generated using MSGANs. Figure 8 shows the comparison of original and synthetic fundus images for BRVO. The first two images (original-BRVO1 and original-BRVO2) represent the original fundus images showing features such as soft exudates, retinal edema, and dot hemorrhages. The last two images (synthetic-BRVO1 and synthetic-BRVO2) are synthetic images generated using MSGANs. The performance

analysis of the proposed MSGAN was evaluated using content loss, adversarial loss, and perceptual loss. The content loss decreased significantly during training, starting from 0.0185 and converging to 0.0043 after 50 epochs. This indicates that the synthetic images closely resemble the high-resolution real images in terms of pixel-level features. The adversarial loss, which measures the generator's ability to fool the discriminator, decreased from 0.12 to 0.02 over the training period. This demonstrates the generator's improvement in producing realistic images indistinguishable from the real ones. Combining content and adversarial loss, the perceptual loss was reduced from 0.138 to 0.0245, highlighting the generator's overall performance in producing high quality and realistic synthetic images.

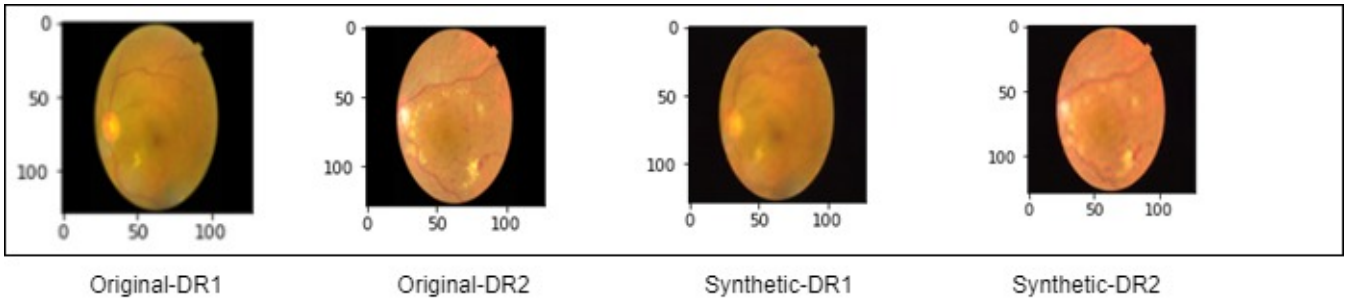


Fig. 7. Comparison of original and synthetic fundus images for DR.

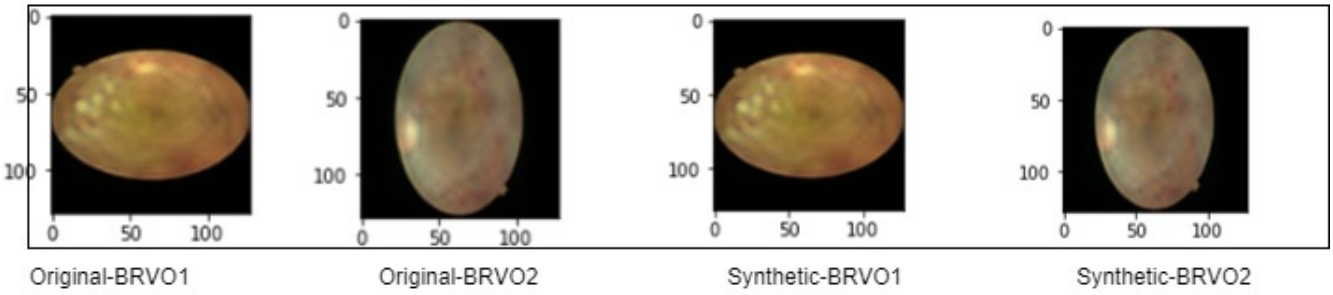


Fig. 8. Comparison of original and synthetic fundus images for BRVO.

V. CONCLUSION

The proposed study addresses the critical issue of class imbalance and limited representation in existing fundus image datasets for multi-disease detection. By developing the Fundus Multi-Disease Dataset (FMDD) with 4 classes of ocular pathologies, this research fills a significant gap in ophthalmic diagnostic datasets. Leveraging Multi-Spectral Generative Adversarial Networks (MSGANs), we have generated high quality synthetic images to improve the diversity and balance of the dataset. This ensures robust training and improved generalizability of deep learning models. The main drawback of the proposed method is its dependence on pre-existing datasets, which limits the diversity and requires more computational time to train the model. These issues will be addressed in the next study.

DATA AVAILABILITY

The dataset we utilized in our study can be found at https://drive.google.com/drive/folders/1hjyK-ew2FvmVi_rlzxIh0vnPuOJPGKgW.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, https://doi.org/10.1145/3065386.

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 2242–2251, https://doi.org/10.1109/ICCV.2017.244.

[3] S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, "Improving Automated Detection of Cataract Disease through Transfer Learning using ResNet50," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17541–17547, Oct. 2024, https://doi.org/10.48084/etasr.8530.

[4] C. Nithyeswari and G. Karthikeyan, "An Effective Heuristic Optimizer with Deep Learning-assisted Diabetic Retinopathy Diagnosis on Retinal Fundus Images," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14308–14312, Jun. 2024, https://doi.org/10.48084/etasr.7004.

[5] S. Pachade *et al.*, "Retinal Fundus Multi-disease Image Dataset (RFMiD)." IEEE Dataport, Nov. 25, 2020, https://doi.org/10.21227/s3g7-st65.

[6] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004, https://doi.org/10.1109/TMI.2004.825627.

[7] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, Mar. 2000, https://doi.org/10.1109/42.845178.

[8] M. M. Fraz *et al.*, "CHASE_DB1 retinal vessel reference dataset," Kingston University Research, Jun. 2012, [Online]. Available: https://researchdata.kingston.ac.uk/96/.

[9] K. S and M. Durgadevi, "Generative Adversarial Network (GAN): a general review on different variants of GAN and applications," in *2021 6th International Conference on Communication and Electronics Systems*, Coimbatre, India, 2021, pp. 1–8, https://doi.org/10.1109/ICCES51350.2021.9489160.