CHAPTER 6

ENSEMBLE CLASSIFICATION AND REGRESSION MODEL USING SPECIAL GENERATIVE ADVERSARIAL NETWORK GENERATED IMAGES FOR CLASSIFYING THE STAGES OF DIABETIC RETINOPATHY

6.1 Introduction

The common reason for eyesight loss is due to DR disease which is identified in advanced stages and has varying levels of severity. The DR disease affects the eyeball where the light is turned as an electronic signal to generate images. The DR destroys the blood vessels that surround the retina which is filled with oxygen and nutrients. This leads to shortage of blood supply to the retina which causes vision damage. Microaneurysms (MAs), hemorrhages and exudates (EXs) are instances of arteries that can grow moderately, Nakayama, L. F. (2022). DR disease advances to proliferative retinopathy that leads to severe retinal damage than non-proliferative retinopathy. The right quantity of blood flow is not produced when the retina is depressed which leads to vision degradation and vision loss risk. The DR classification by skilled ophthalmologist is time consuming, expensive and it's less accurate in most cases when the disease does not possess visible signs. Hence there is a need for refined automated techniques to identify the DR disease criticalities which eases the task of an ophthalmologist for further care and prediction.

To identify DR many computer-aided methods were created over the years to find the lesions at pixel level. Computer-Aided Diagnosis (CADx) differentiates DR at image level (Vikramathithan A.C. *et al.* (2022); Pires R *et al.* (2019)). Machine and deep learning algorithms are introduced to implement other strategies of CADx models. Machine and deep learning techniques are commonly used concepts for developing an automatic DR disease classification system. CNN DL model contains millions of learners and training involves numerous RF images. The DR classification and pretrained CNN architecture issues are discussed in these literatures (Tariq H *et al.* (2021); Chen PN *et al.* (2021); Lam C *et al.* (2018)). The advanced pre-trained CNN models such as ImageNet, VGG, ResNet, Dual Path Network (DPN) are frequently learnt to alter low-level into high level features. El Houby EM (2021); Chen Y *et al.* (2017); Salman H *et al.* (2020). The required results are not obtained due to two reasons: i) a vital portion of the DR fundus image is not accessible to understand

the CNN framework, ii) to optimize the pre-learned designs a brute force policy is applied to modify the classifier blocks. The optimized model uses minimal quantity of RF images to enhance the overall performance. The learning algorithms leverage the RF images to integrate pre-trained CNN frameworks.

A transfer learning model using pre-learned CNN framework such as VGG19, ResNet152 and DPN 107 was introduced by Saeed F *et al.* (2021). The lesions in retinal images are needed for DR grading. The retinal image is given as input to analyze the framework and the model differentiates between the healthy and unhealthy stages of DR. The framework learns the patterns of low- and high-level features. In ResNetGB model, ROI collected from the retinal images are used to reinitialize the upper most layer of the pre-trained CNN framework. E-optha was explored for lesion which consists of pixel-level tumor labeling. The framework was optimized to learn about the low-level blocks and the other forms of tumor and normal areas. The fully connected layer translates the high-level features based on the PCA which is used to obtain various features of retinal images in an unsupervised learning method.

In the first model of this study, an MSA strategy with ResNetGB model has been used to improve the performance of DR stage classification in comparison with the other CNN-based pretrained models. Valarmathi S. and Vijayabhanu R. (2022) designed a model to train high level features of RF images for DR stage classification based on severity levels. The encoder network operates to provide RF images with high-level interpretational space by facilitating the fusion of middle and high-level data to improve interpretability. The model was built to define retinal image patterns with various positions where the MSA was applied on high-level interpretation to enhance feature analysis. Cross-entropy loss is estimated for training and testing loss. Since the MSA-ResNetGB model was trained on less varied images, the model suffers from lack of generalization ability which makes it difficult to effectively classify the DR stages. The diversity in retinal fundus images is limited by factors such as imaging conditions and techniques. For instance, if the images are predominantly captured under similar lighting conditions or using a specific type of retinal imaging technology, the dataset's diversity is restricted, potentially hindering the model's performance when faced with variations in real-world imaging scenarios.

Hence, a novel model named Special Generative Adversarial Network with Ensemble Classification Regression (SGAN-ECR) was proposed to overcome the limitations and

enhance the classification accuracy of MSA-ResNetGB model. The more realistic and high-quality retinal images are synthesized using suitable labels. The proposed SGAN enables the generation of images with higher resolution and finer details. This contributes to a more realistic appearance, especially when it comes to reproducing complex textures, intricate patterns, and sharp edges. The key objective of the SGAN-ECR model is to generate high-contrast retinal fundus images and to provide better generalization ability for classifying DR stages based on severity levels.

The SGAN consists of a pixel-2-pixel GAN with a UNet++ generator and a patch-GAN discriminator. A pixel-2-pixel GAN generator is used to train the shape of dissimilarities of the retinal fundus images. A patch-GAN discriminator is used to enhance the feature extraction with accurate local contrast patterns. In the SGAN model, high-contrast synthetic RF images are generated, which are utilized for training the MSA-ResNet regression and classification structures. The inter-dependencies among the various stages of DR severity are learned in the regression structure. The discriminative features of the RF images are learned in the classification structure. Finally, the features are concatenated and given as input to the MLP classifier for classification.

The following sections in this chapter are organized as follows. Recent works related to the GAN-based retinal fundus image augmentation model are discussed. A detailed description related to the SGAN-ECR model based on retinal image augmentation, the structure of SGAN, and DR stage classification using an ensemble classification regression model is defined in the subsequent sections. Experimental results are analyzed based on the performance metrics, and the outcome analysis and assessment of the proposed technique are presented conclusively.

6.2 Related work on GAN-based Retinal Fundus Image Augmentation Models

A Multiple-Channels-Multiple-Landmarks (MCML) scheme was proposed by Yu *et al.* (2019), which was based on the GANs to produce a color retinal fundus image with vessel, optic disc, and optic cup channels. The objective is to map various landmarks created from the channels and actual retinal fundus images. A Conditional GAN (CGAN) model was proposed by HaoQi & Ogawara (2020) to augment the images and images with vessel partition. A learning model is created to train the images, which are combined with actual images to train the complete network and to partition the retinal images. A Cycle-consistent

GAN (CycleGAN) model was created by Yoo *et al.* (2020) to produce images without the corresponding paired images. The artifacts are eliminated automatically from the retinal images. A SUD-GAN model by Yang *et al.* (2020) distinguished the blood vessels from the retinal images. The generator adopts the encoder-decoder structures, and the connection units were added with convolution layers to prevent gradient dispersion. A dense connection structure was included in the middle part of the CNN to enhance the differentiation ability. A Wasserstein GAN (WGAN) was proposed by Magister L.C. *et al.* (2021) to create synthetic retinal image data by synthesizing the images to semantically paint the retinal images based on context and perceptual loss of pixels.

Chen *et al.* (2021) proposed a GAN (RF-GAN) model, which consists of RF-GAN1 and RF-GAN2 to synthesize the images. The translation of retinal images from the semantic partition database to the domain of the APTOS database is performed by RF-GAN1. The semantic partition model is trained with the translated images, and the trained images are applied to capture the structure and lesion marks with the help of the RF-GAN2 model to grade the DR images. Andreini P. *et al.* (2021) proposed a design for a 2-phase GAN to synthesize high-quality images to the semantic label map. GAN is trained to produce semantic label maps to define vasculature. The translation method is applied to obtain realistic images from the vasculature. The actual and synthesized images are combined to train an entire network to enhance the partition efficiency.

The above literature was based on the GAN-retinal image synthesis and augmentation, which specifies that there are a few limitations in image augmentation for DR classification. The limitations include i) GAN requires an annotated database, which is expensive for creating high-level quality images. ii) the image quality was not considered, and low-quality images were created; iii) the model fails to learn enormously thin vessels while generating synthetic retinal images. The above-specified limitations are eliminated in the proposed SGAN model, which is an image augmentation scheme to enhance image quality. Using this SGAN, high-contrast images are generated and used to train and classify the DR images based on severity levels.

6.3 The Proposed SGAN-ECR Model

In this section, the proposed SGAN-ECR model is presented in detail, followed by SGAN-based image augmentation, the structure of SGAN and DR stage classification using

ensemble classification regression model. Figure 6.1 portrays the workflow of the SGAN-ECR Model to categorize different stages of DR disease based on severity levels. Initially, various retinal images are considered, and the SGAN model is applied to produce high-quality images. The original dataset images, along with SGAN-generated synthetic images, are used for training ensemble classification and regression models using the MSA-ResNet structure, which consists of an MLP classifier. The trained model is applied to assess the samples for DR stage classification.

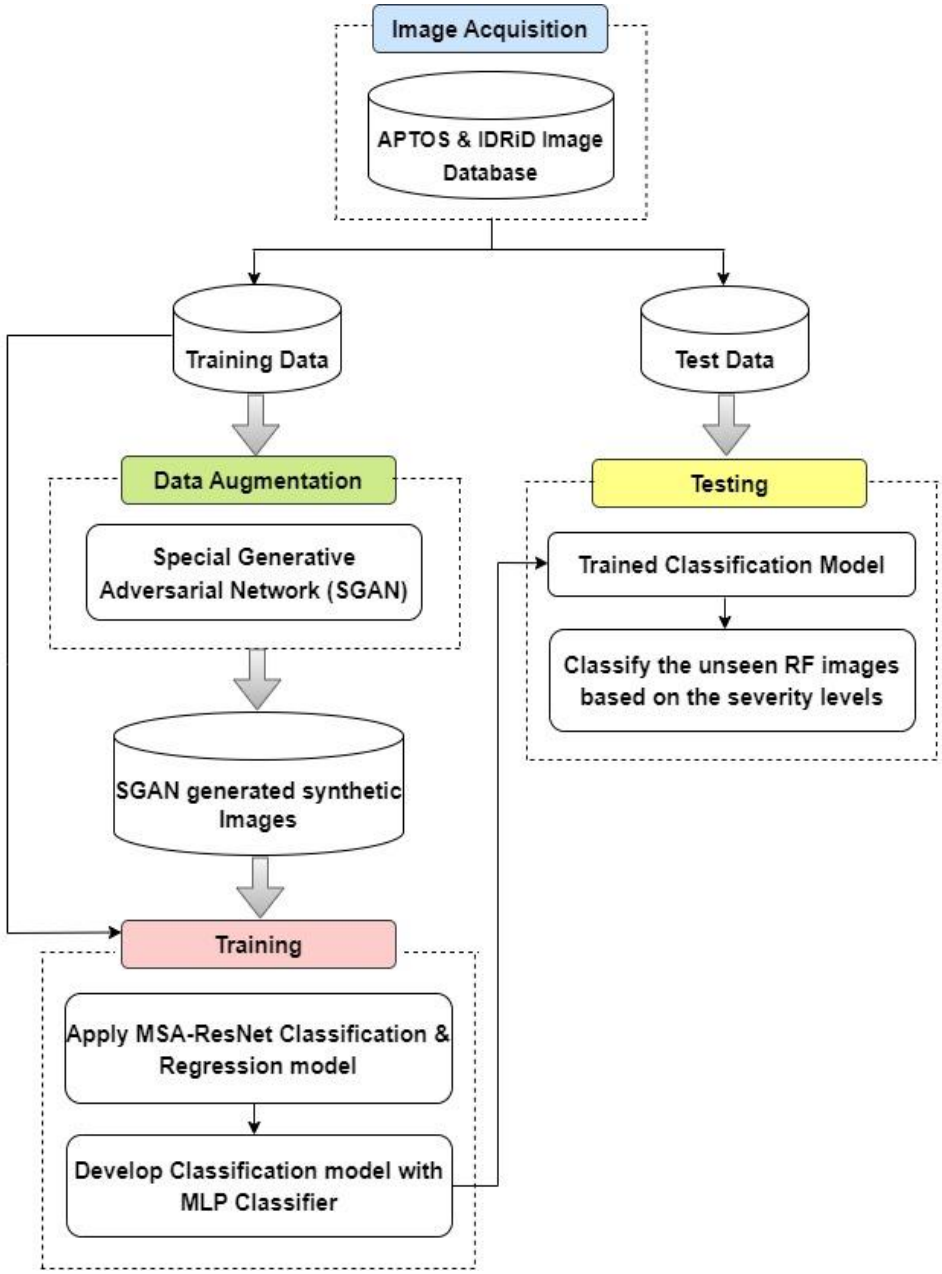

Figure 6.1: Workflow of SGAN-ECR Model for Diabetic Retinopathy Stage Classification

6.3.1 Retinal Fundus Image Augmentation using SGAN

The SGAN model aims to synthesize the low-contrast retinal images by executing the deep frameworks along with the high-contrast retinal images. The deep framework is trained with a system consisting of image distortion fields and textual features from the retinal image database using the GAN model. The SGAN model eliminates the learning complexities of the images with the high-contrast framework present in the GAN model. Figure 6.2 portrays the GAN structure for retina image augmentation. The GAN is a generative model that learns the mapping from the random noise vector z to the final image y denoted as: $z \rightarrow y$. The generator (G) in Figure 6.2 converts the noise vector (z) from the distribution $p(z)$, which is the latent space, to a synthesized RF image(i_s). The discriminator (D) distinguishes the synthetic and actual retinal images according to the distributions of i_s and i_r. The produced retinal images make a distribution of $p_s(i)$, which approximates $p_r(i)$ from the actual retinal image after the current learning process.

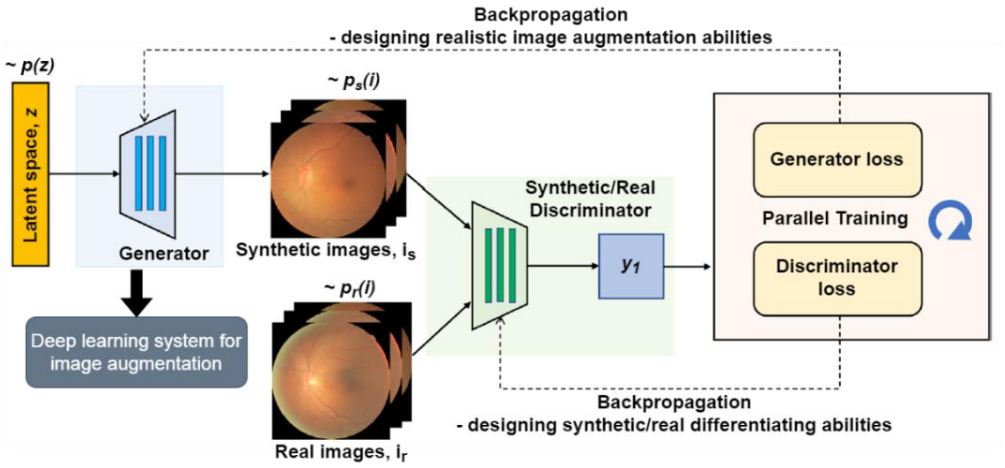

Figure 6.2: GAN Structure for Retinal Image Augmentation

The SGAN structure is aimed at customizing the CNN-GAN model in super-resolution based on deep generative neural networks, as shown in Figure 6.3. The SGAN consists of a patch-GAN model and a pixel-2-pixel GAN model with skip links. The network includes Instance Regularization (IR), convolution and Rectified Linear Unit (ReLU) layers. The generator network (G) aids in accelerating the patch-GAN model that develops an image and decodes the pixels of the synthetic image. The G network acts as an independent process for all the subsequent images. The pixel-2-pixel GAN with the UNet++ model is trained to decode a feature map pixel to generate the target images.

A. Pixel-2-Pixel GAN with UNet++ as Generator

The UNet++ architecture, an extension of the original UNet model, represents a sophisticated neural network designed for image segmentation tasks. But due to its encoder & decoder properties, it is also highly used in feature extraction and serves as the generator in the GAN model. This component takes an input image and generates an augmented or translated image. Comprising five convolutional and deconvolutional layers, each employing a 3x3 kernel and varying neuron sizes (128, 64, 32, 16, and 8), the model integrates skip connections to enhance feature representation. The architecture begins with an initial block, consisting of two consecutive convolutional layers with 128 neurons each, followed by max-pooling for downsampling. Subsequent blocks, labeled Block 2 to Block 5, follow a similar pattern with decreasing neuron sizes. Notably, a bridge connects the encoder and decoder pathways, composed of two convolutional layers with 16 neurons each. The decoder pathway starts with Block 5, where the output of Block 5 is concatenated with the bridge output and processed through two convolutional layers with eight neurons each, followed by upsampling. This process is iteratively repeated for Blocks 4 to 1, with each block concatenating the output of its corresponding encoder block with the output from the previous decoder block. The architecture concludes with a convolutional layer featuring a single neuron and a sigmoid activation function, facilitating the generation of the final output. This comprehensive architecture, characterized by its intricate skip connections and multi-level feature integration, is well-suited for binary segmentation tasks, offering superior performance in capturing complex image structures.

The Pixel-2-Pixel Generative Adversarial Network (GAN) architecture, described as the discriminator, evaluates both the source image (original) and the generated output from the UNet++. The discriminator's role is to distinguish between real and generated images. The Pixel-2-Pixel (GAN) is implemented as the generator in the described architecture. This generator aims to augment images, enhancing their quality through an encoder-decoder structure. The encoder employs Conv2D operations with increasing filter sizes and strides, creating downsampling layers labeled from d1 to d7. This process compresses the input images into a latent space representation. An additional convolutional layer further condenses the encoded information, introducing rectified linear units (ReLU) for activation. The decoder component, comprised of Conv2DTranspose operations labeled from u1 to u7, mirrors the structure of the encoder through upsampling layers. The output layer utilizes

Conv2DTranspose with a (4, 4) kernel size and a (2, 2) stride to reconstruct the image with three channels. A hyperbolic tangent activation function (tanh) ensures pixel values are constrained within the range [-1, 1]. The discriminator in this architecture distinguishes between real and generated images and takes both source and target images as input. It consists of several convolutional layers incorporating leaky rectified linear units (LeakyReLU) and batch normalization with a value of 1. The output layer employs a sigmoid-activated Conv2D operation, producing a binary classification indicating the authenticity of the input images.

B. Patch GAN Model as Discriminator

The adversarial learning in this framework is bolstered by incorporating context-related Markovian patches, which refer to image patches displaying Markovian properties. These properties indicate that a patch's characteristics are dependent solely on its immediate surroundings rather than the entire image. Essentially, the information within a Markovian patch is considered sufficient for analysis when contextualized with neighboring patches. Instead of unifying the context and images, the mapping of multiple images with similar contexts is trained. Notably, the patch dictionary during the iterative adjacent search is replaced by the discriminator network (D), trained to differentiate between real feature patches and inappropriately generated images using VGG19 with ReLU3. Subsequent evaluations employ VGG19 encoding with the abstract layer ReLU5, contributing to the deconvolution process in the patch-GAN, where the deconvolutional image generator is obtained. The SGAN structure, depicted in Figure 6.3, integrates both the generator and discriminator networks. The discriminator (D) network is driven by adversarial learning, discerning between neutral patches from synthetic images and sampled patches from reference images. Real patch categorization scores ($s=\pm1$) are provided for all neutral patches, and the texture loss is minimized for sample patches from the generated image, optimizing the D network with IR and leaky ReLU. The VGG19 network is not fine-tuned, while the D and G networks are optimized to enhance the efficiency of the generator (G), with D acting as an opponent.

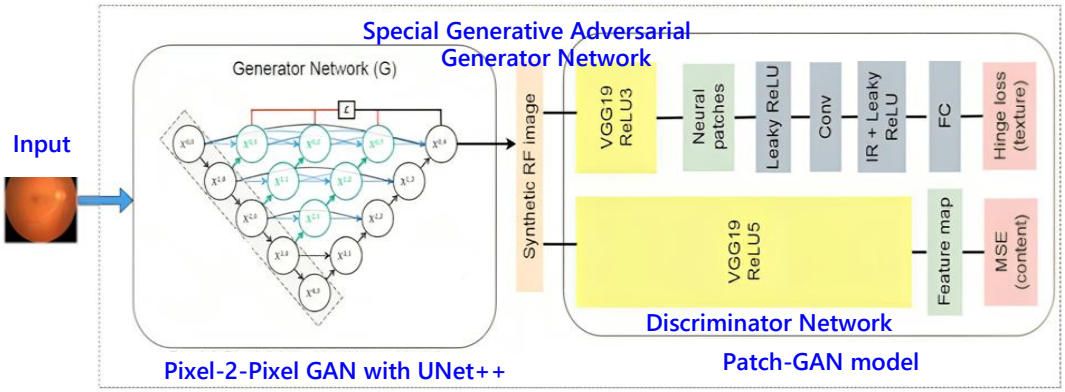

Figure 6.3: SGAN Structure with generator and discriminator network

In terms of architecture, the encoder in this network is composed of several blocks, with the initial block featuring a convolutional layer with 64 filters and a 7x7 kernel, excluding batch normalization. Subsequent encoder blocks utilize 3x3 convolutional layers with an increasing number of filters (128, 256, 512) and include batch normalization for regularization. The final two encoder blocks maintain 512 filters each. The bottleneck layer compresses the representation through a 4x4 convolutional layer with 512 filters and a ReLU activation function. Each block corresponds to an upsampling step, and the first decoder block employs a transposed convolutional layer with 512 filters, a 3x3 kernel, batch normalization, and ReLU activation moving to the decoder section. Subsequent decoder blocks mirror the encoder's architecture, progressively reducing the number of filters. The final layer of the decoder employs a transposed convolutional layer with a 4x4 kernel, generating three output channels for RGB color.

A hyperbolic tangent activation function is applied to ensure pixel values are within the valid range. Importantly, skip connections are strategically incorporated throughout the architecture, concatenating feature maps from the encoder to the corresponding decoder layers, facilitating the retention of spatial details. Simultaneously, a discriminator model is introduced for adversarial training. This discriminator processes pairs of source and target images, aiming to distinguish between real and generated images. Comprising convolutional layers with leaky ReLU activation and instance normalization, the discriminator progressively down samples input to produce a binary classification indicating the authenticity of image pairs. Trained to minimize binary cross-entropy loss, the entire adversarial network, including the generator and discriminator, is optimized using the Adam optimizer with a learning rate of 0.0002 and a beta parameter of 0.5. This simultaneous

training of both components fosters a competitive learning process, ultimately enhancing the quality of the generated images.

Let $i_t \in \mathbb{R}^{w_t * h_t}$ denotes the reference texture image and $i_s \in \mathbb{R}^{w_t * h_t}$ denotes the generated image. In the model i_s is initialized with random noise for an unguided generation and $i_c \in \mathbb{R}^{w_t * h_t}$ denotes the content image for guided generation. The deconvolution iteratively changes i_s as shown in Equation (6.1).

$$i_s = \arg_{i_s} \min E_t(\Phi(i_s), \Phi(i_t)) + \alpha_1 E_c(\Phi(i_s), \Phi(i_c) + \alpha_2 \gamma(i_s) \quad (6.1)$$

Where E_t denotes texture loss, $\Phi(i_s)$ denotes feature map in ReLU layer of VGG19. $\gamma(i_s)$ denotes the additional normalization term to define smoothness in pixels. E_t and $\gamma(i_s)$ denotes the random textures in E network. The content loss E_c is reduced, and SGAN generates an image and i_c denotes the guidance image that is associated. The content loss is estimated using mean squared error between (i_s) *and* (i_c).The weights are assigned as α_1 =1 and α_2 = 0.0001. The equation (6.1) is reduced with an optimizer with 0.01 training rate and 0.2 momentum. The neutral patches obtain the final gradient through the backpropagation of the D network.

The patches from $\Phi(i_s)$ and E_t are set to 1, as shown in Equation (6.2)

$$E_t(\Phi(i_s), \Phi(i_t)) = \frac{1}{N} \sum_{x=1}^{N} \max(0, 1 - 1 * s_x) \quad (6.2)$$

Where s_x denotes the categorization score of xth neutral path and N denotes sample patches in $\Phi(i_s)$. The D network is trained randomly using initialized variables and altered after deconvolutions to enhance the generation results.

6.3.2. Training SGAN

The combined GAN model connects the Pixel-2-Pixel generator and discriminator. During training, the generator produces an output given a source image, and the discriminator evaluates both the source image and the generated output. The model is trained using binary cross-entropy for the discriminator and mean absolute error for the generator. An Adam optimizer is utilized with specified loss weights of [1, 100] to balance the adversarial and

perceptual objectives. The training loop involves iteratively updating the discriminator and generator. Discriminator training considers both real and fake samples, optimizing weights based on binary cross-entropy loss. Generator training updates weights based on both discriminator classification and the mean absolute error between the generated and real images, contributing to the overall goal of creating high-quality augmented images. In total, 4500 synthetic images are generated by SGAN from both APTOS 2019 and IDRiD datasets to mitigate the class imbalance. Figure 6.4 shows the SGAN augmented images in APTOS and IDRiD datasets, respectively. The synthetic images that are created will have the same image size as the dataset images to fool the Discriminator. The quality of SGAN-generated synthetic images is evaluated using the metric called Structural Similarity Index Measure (SSIM). It can be calculated using equation 6.3.

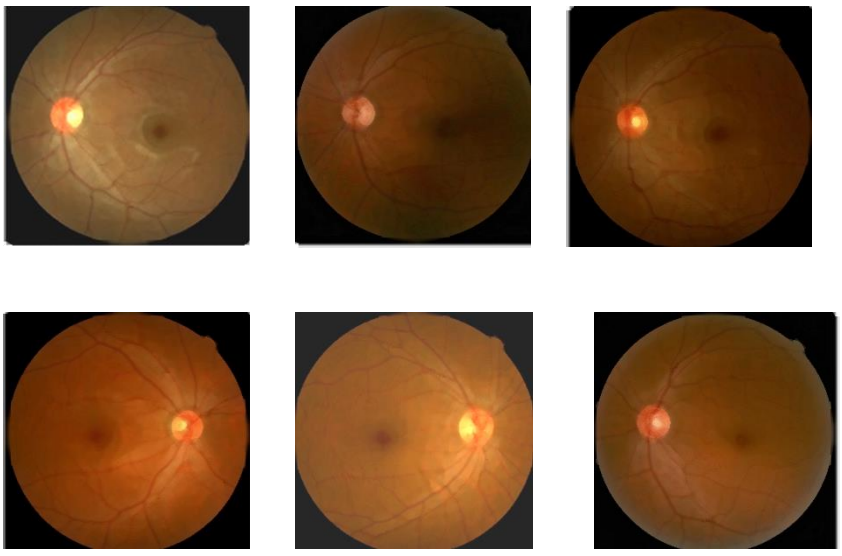

Figure 6.4: SGAN Augmented images in APTOS and IDRiD datasets

$$SSIM(I, I') = \frac{(2\mu_I \mu_{I'} + c_1)(2\sigma_{II'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1)(\sigma_I^2 + \sigma_{I'}^2 + c_2)} \quad (6.3)$$

where $\mu_I = \sum_{i=1}^{N} w_i x_i$ and $\sigma_I = \left(\sum_{i=1}^{N} w_i (x_i - \mu_I) \right)^{1/2}$ denotes the gray values in average and I variance. $\sigma_{II'} = \sum_{i=1}^{N} w_i (x_i - \mu_I)(y_i - \mu_{I'})$ denotes the covariance between the images I and I'. $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ denotes the two constants. The average SSIM values of the generated SGAN images were found to be 0.7231. Table 6.1 shows the balancing image samples for APTOS 2019 and IDRiD datasets. The total image count in APTOS 2019 dataset is 7177. In this, 2930 are the training images and 4247 are SGAN augmented image, whereas the total image count in IDRiD dataset is 670. In this, 410 are the training images and 260 are SGAN augmented images.

Table 6.1: Balancing image samples in dataset classes

Class / Severity level	APTOS 2019 Dataset			IDRiD Dataset		
	Training	SGAN Augmented	Total	Training	SGAN Augmented	Total
0 – Normal	1435	0	1435	134	0	134
1 - Mild NPDR	322	1113	1435	18	116	134
2 – Moderate NPDR	791	644	1435	131	3	134
3 - Severe NPDR	146	1289	1435	53	81	134
4 - Proliferative DR	234	1201	1435	74	60	134
Total no. of images	2930	4247	7177	410	260	670

6.3.3 DR Stage Classification using SGAN-ECR Model

The SGAN generated synthetic retinal images along with the benchmark images are used for training MSA-ResNet classification and Regression model. To obtain the subsequent feature maps, the MSA employs a high-level feature map with multiple scales that are concatenated with the attention maps. The MSA-ResNet regression model is equipped to

categorize the different stages of DR by simultaneously learning the progressive features between DR severity levels. The dependency among the DR severity levels is based on the severity stages such as higher and lower.

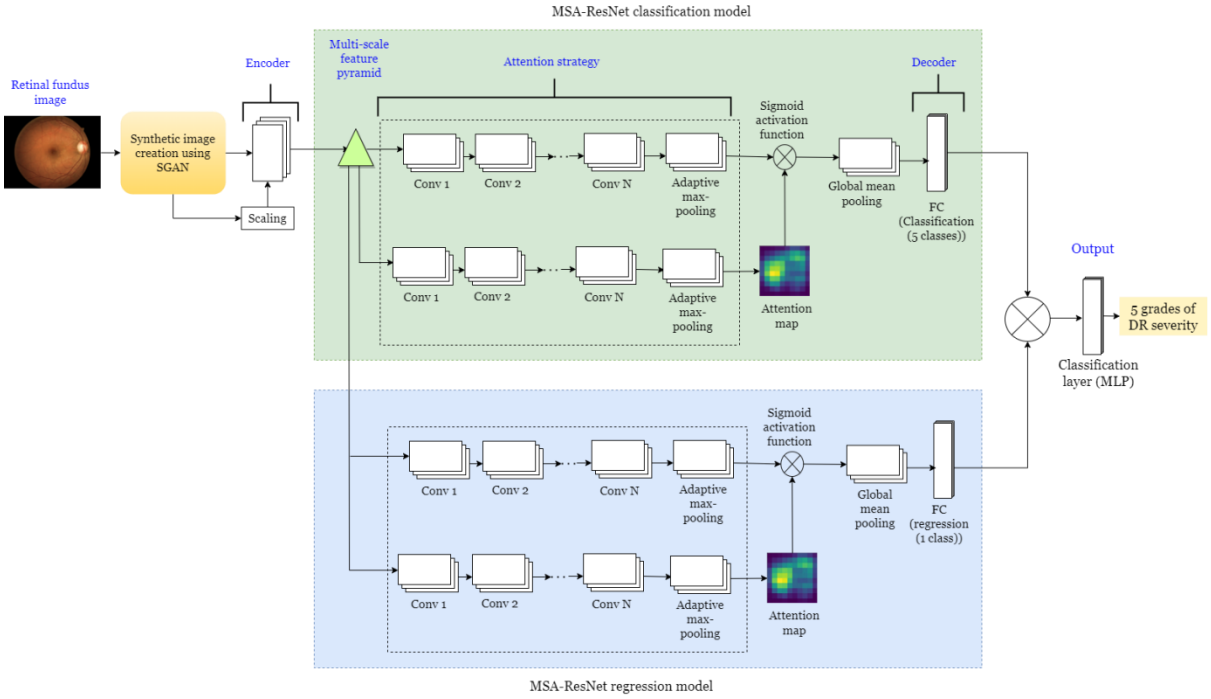

Figure 6.5: SGAN-ECR Model Structure

Figure 6.5 illustrates the structure of the SGAN-ECR model. The selective features between different DR stages are learnt by the MSA-ResNet classification model and the inter-dependency features among the different levels are learnt by the MSA-ResNet regression model. The efficiency of classification is enhanced by cross-entropy loss factor for classification shown in Equation (6.4) and mean square error loss factor for regression process shown in Equation (6.5). The total loss for the model will be weighted sum of the classification and regression losses.

$$Cross\ entropy = -\sum_i^M y_i \log(\hat{y}_i) \tag{6.4}$$

$$Mean\ square\ error = \frac{1}{M}\sum_i^M (y_i - \hat{y}_i)^2 \tag{6.5}$$

Where $\hat{y}_i$ denotes the expected value, y_l denotes the actual value, M denotes the class count. The cross entropy and mean square error value ranges between 0 and 1. The retinal images

use linear and SoftMax activation functions in the classification model and its trained independently. Equation (6.6) shows the SoftMax activation function.

$$Softmax(i) = \frac{e^{\delta(i)}}{\sum_{j=1}^{K} e^{\delta(j)}}, j = 1, \dots, i, \dots, K \quad (6.6)$$

where, K denotes the class count and δ denotes the fully connected unit's result.

The output of the SGAN-ECR model will consist of two parts: the classification part, which corresponds to the number of classes and is followed by a softmax activation to produce probabilities, and the regression part, which produces a single continuous value without any activation function. There are five probability scores associated with five DR grades in the classification model. The severity stage of DR produces a single outcome in the regression model. The model is trained with result labels like 0 indicates no DR, 0.2 indicates mild DR, 0.4 indicates moderate DR, 0.6 indicates severe DR and 0.8 indicates proliferative DR. The learnt features of the model are combined and specified as input to the concluding classification layer. The final categorization of the five DR grades is obtained using the MLP classifier instead of GB classifier. The pseudo code of SGAN-ECR model is presented.

Pseudo code: The Proposed SGAN-ECR model for DR Stage Classification

Input: *Retinal images from APTOS & IDRiD dataset along with SGAN generated synthetic images* $I_1, \dots, I_n$

Output: *Different DR Classes/Stages*

Begin

Get the dataset images for DR stage classification;

Categorize the image database into training and test set;

Execute SGAN model on training set to create synthetic RF images;

Add images to original training set to balance the dataset classes;

Initialize ResNetGB encoder parameters such as learning rate, iteration count, $t \leftarrow 0$*, maximum accuracy test set* $max_{acc} \leftarrow 0$ *and maximum iteration* T_{max}*;*

$\boldsymbol{while}(t < T_{max})\ \ t \leftarrow t + 1;$

$\boldsymbol{for}(training\ set)$

Train MSA-ResNet classification model and MSA-ResNet regression model;

Calculate loss factor using Equation (5.3) & (5.4); adjust the training variables;

MLP classifier applied to get the result of the trained classification model

end for

Classify test images (DR stages) using trained model;

Determine the classification accuracy (acc) *on the test set;*

$\boldsymbol{if}(a > max_{acc})$

$max_{acc} \leftarrow acc;$

Set the optimal training variables of ResNetGB;

End if

End while

End

6.4 Experimental Result and Analysis

This section discusses about the results of the SGAN-ECR model. The proposed and the existing classification methods comparison and the confusion matrices are elaborated. The test results for the SGAN-ECR model are discussed. The structure is trained with 100 epochs, batch size of 3 and training frequency is 10^{-4}.

6.4.1 Result Analysis of SGAN-ECR model on the APTOS dataset

The performance metrics of the proposed SGAN-ECR model are trained using APTOS and IDRiD to classify the DR stages. Multi-class classification is carried out and the dataset is classified into five classes starting from 0 to 4. The APTOS dataset, out of 3662 training samples, 20% of the samples are considered for test data. In the classification part, categorical labels corresponding to each image are typically stored as one-hot encoded vectors, whereas in the regression part, continuous values corresponding to each image are stored as floating-point numbers. Table 6.2 shows the existing models for DR Stage Classification on APTOS dataset in the literature. Table 6.3 shows the performance comparison of the proposed models. Accuracy metric is one of the frequently used metrics and the other metrics are also presented. CNN is the commonly used DL method for medical image analysis.

Table 6.2: Existing models for DR Stage Classification on the APTOS 2019 dataset

Reference	Classification method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Qummar S. *et al.* (2019)	Extra Tree model	91.07	90.40	89.54	89.97
S. Sheikh and U. Qidwai (2021)	DenseNet121 model	90.50	93.00	90.00	88.47
N. Sikder *et al.* (2021)	Tuned XGBoost model	94.20	94.34	92.68	93.51

Table 6.3: Comparison of the Proposed Models for DR Stage Classification on the APTOS 2019 dataset

Reference	Classification Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SrinivasanV & Rajagopal V (2022)	MSA-ResNetGB Model	94.40	94.53	94.40	94.43
Valarmathi Srinivasan & Vijayabhanu Rajagopal (Proposed - 2023)	**SGAN-ECR model**	**97.81**	**95.95**	**97.98**	**96.90**

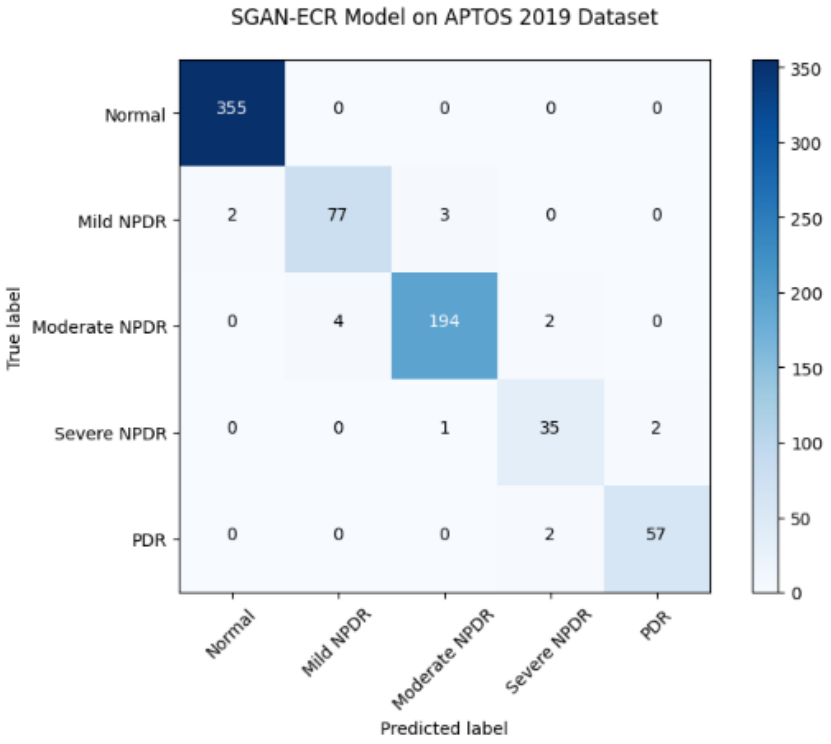

Figure 6.6: Confusion matrix for SGAN-ECR model on APTOS 2019 dataset

The accuracy of the proposed MSA-ResNetGB model obtained 94.40 % and it is 3.49% less compared to SGAN-ECR model. Figure 6.6 shows the confusion matrix of the

SGAN-ECR model on APTOS dataset. The sample distribution by classes and the accurate ratio of classification and misclassification are demonstrated in the matrix. For example, the precisely identified sample count is 355 for class 0 (no DR). The precisely identified sample count is 77 for class 1 (mild DR). For class 2 (moderate DR), the precisely identified sample count is 194. For class 3 (Severe DR), the precisely identified sample count is 35 and finally for class 4 (PDR), the precisely identified sample count is 57.

6.4.2 Result Analysis of SGAN-ECR model on the IDRiD dataset

Multi-class classification is performed on IDRiD dataset, and it is classified into five classes starting from 0 to 4. The IDRiD dataset, out of 516 training samples, 20% of the samples are considered for test data. Table 6.4 shows the proposed models performance on IDRiD dataset. Accuracy metric is one of the frequently used metrics and the other metrics that are not stated in the literature are represented as '-'. The performance metrics are analyzed for SGAN-ECR model on IDRiD dataset. The accuracy of the proposed model is 2.07% better than MSA-ResNetGB model.

Figure 6.7 shows the confusion matrix of SGAN-ECR model on IDRiD dataset. The sample distribution by classes and the accurate ratio of classification and misclassification are demonstrated in the matrix. For example, the precisely identified sample count is 34 for class 0 (no DR). For class 1 (mild DR), the precisely identified sample count is 3. For class 2 (moderate DR), the precisely identified sample count is 30. For class 3 (Severe DR), the precisely identified sample count is 13 and finally for class 4 (PDR), the precisely identified sample count is 19.

Table 6.4: Comparison of the Proposed Models for DR Stage Classification on the IDRiD dataset

Reference	Classification method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Valarmathi S and Vijayabhanu R (2022)	MSA-ResNetGB model	94.18	91.48	91.57	91.45
Valarmathi Srinivasan & Vijayabhanu Rajagopal (Proposed - 2023)	**SGAN-ECR model**	**96.12**	**96.31**	**93.73**	**94.77**

6.4.3 Results obtained from the SGAN-ECR model

This section analyses the results obtained by the SGAN-ECR. The outcome of SGAN-ECR model, applied on the APTOS and IDRiD outperform the literature models. The local and global feature representation aids in the learning process of the DR structure. Table 6.5 compares the SGAN-ECR model on the APTOS and IDRiD. The SGAN-ECR model on the APTOS achieved 97.81% accuracy, 95.95% precision, 97.98% recall, and 96.90% F1-Score. The SGAN-ECR model on IDRiD dataset has obtained 96.12% accuracy, 96.31% precision, 93.73% recall, and 94.77% F1-Score. The performance of the APTOS dataset is exceeding the performance of the IDRiD dataset. When both the datasets are compared there is a performance loss that occurs in the IDRiD dataset due to the lack of image samples. Data augmentation techniques enable to produce effective outcomes.

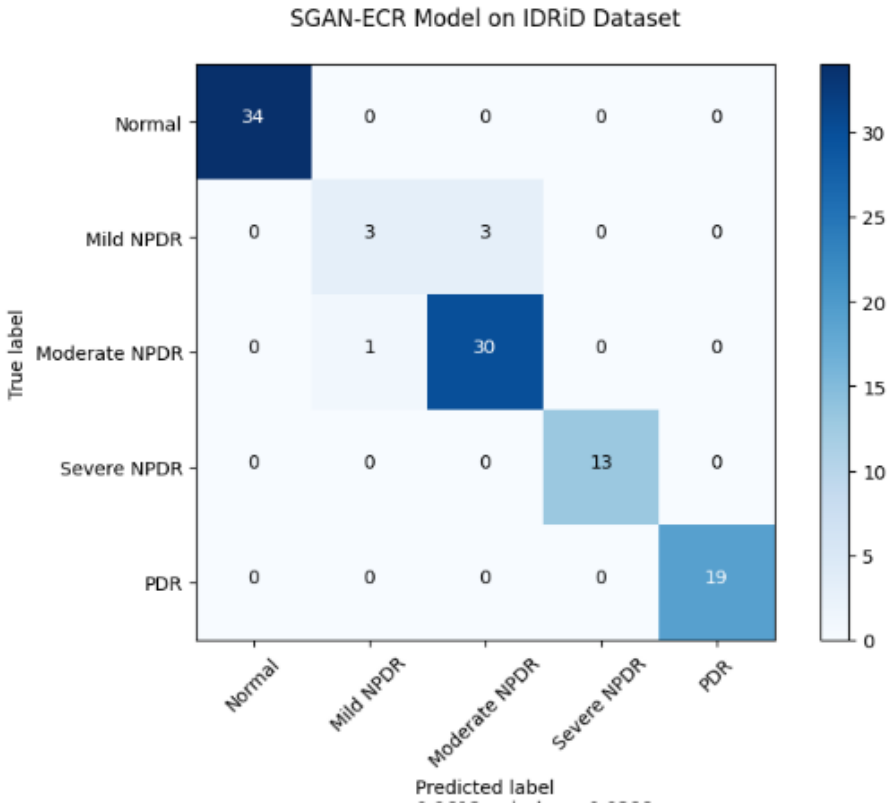

Figure 6.7: Confusion matrix for SGAN-ECR model on the IDRiD dataset

Table 6.5: Performance Comparison of SGAN-ECR model on the APTOS and IDRiD datasets

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
APTOS	97.81	95.95	97.98	96.90
IDRiD	96.12	96.31	93.73	94.77

Figure 6.8 represents the accuracy analysis of the SGAN-ECR model on APTOS and IDRiD datasets. The datasets are labeled as five classes ranging from 0 to 4. For the APTOS dataset, the accuracy achieved is 96%, 93%, 92%, 99% and 93% for the classes 0 to 4. For the IDRiD dataset, the accuracy achieved is100%, 94%, 94%, 92%, and 95% for the classes 0 to 4. The plot shows that the performance of class 3 is better than class 2 for the APTOS dataset. In the case of IDRiD dataset, the performance of class 0 is better than class 3.

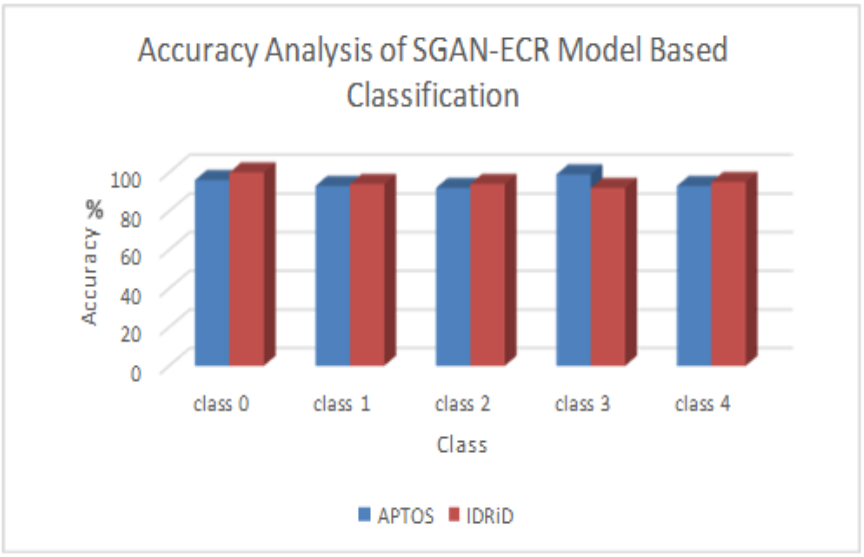

Figure 6.8: Accuracy analysis of SGAN-ECR model on the APTOS and IDRiD datasets

Figure 6.9 represents the precision analysis of SGAN-ECR model on APTOS and IDRiD datasets. The datasets are labeled as five classes ranging from 0 to 4. For the APTOS dataset, the precision achieved is 95%, 92%, 92%, 98% and 95% for the classes 0 to 4. For the IDRiD dataset, the precision achieved is 90%, 92%, 90%, 95%, and 96% for the classes 0 to 4. The plot shows that the performance of class 3 is better than the class 4 for the APTOS dataset. In the case of IDRiD dataset, the performance of class 3 and class 4 are better.

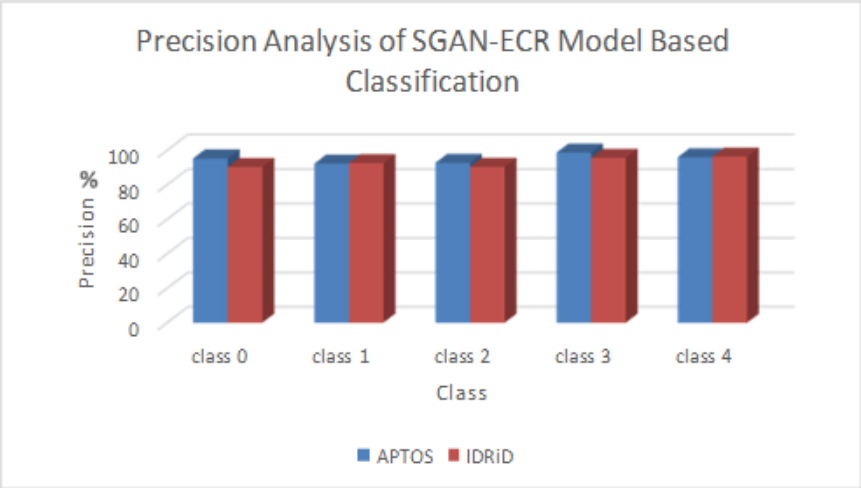

Figure 6.9: Precision analysis of SGAN-ECR model on the APTOS and IDRiD datasets

Figure 6.10 represents the recall analysis of the SGAN-ECR model on APTOS and IDRiD datasets. The datasets are labeled as five classes ranging from 0 to 4. For the APTOS dataset, the recall achieved is 96%, 92%, 94%, 99% and 97% for the classes 0 to 4. For the IDRiD dataset, the recall achieved is 88%, 90%, 91%, 92%, and 93% for the classes 0 to 4. The plot shows that the performance of class 0 is better for the APTOS dataset. In the case of IDRiD dataset, the performance of class 3 and class 4 are better.

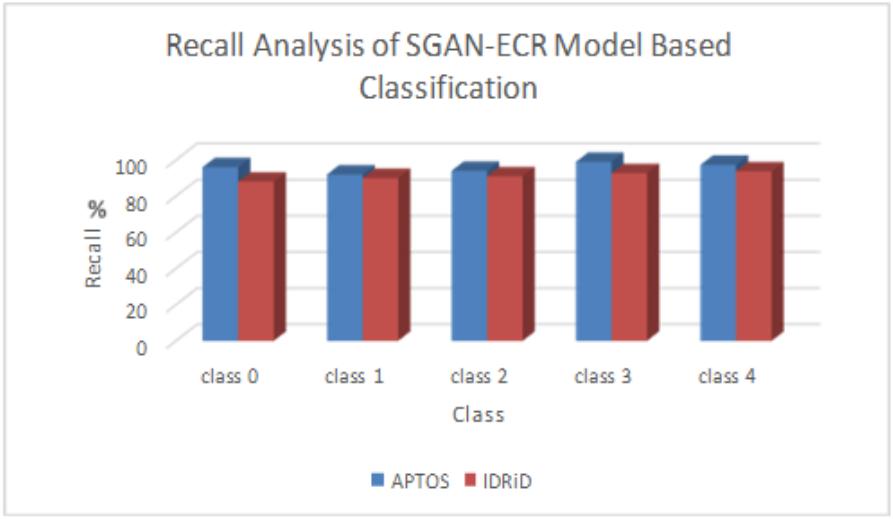

Figure 6.10: Recall analysis of SGAN-ECR model on the APTOS and IDRiD datasets

Figure 6.11 represents the F1-score analysis of the SGAN-ECR model on APTOS and IDRiD datasets. The datasets are labeled as five classes (0-4). For APTOS dataset, the F1-Score achieved is 92%, 93%, 95%, 98% and 96% for the classes 0 to 4. For the IDRiD dataset, the F1-Score achieved is 89%, 90%, 88%, 92%, and 94% for the classes 0 to 4. The

plot illustrates that the class 3 performance is superior for the APTOS dataset. In the case of IDRiD dataset, the performance of class 0 and class 4 are better.

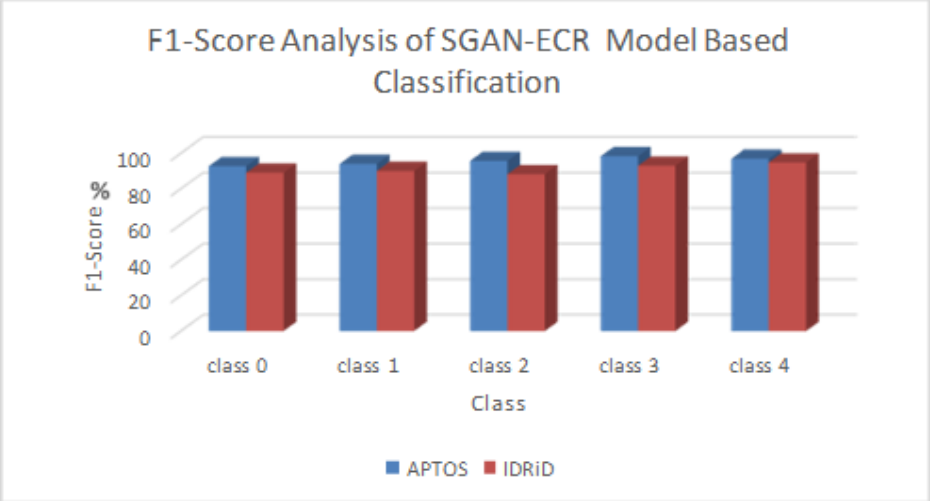

Figure 6.11: F1-Score analysis of SGAN-ECR model on the APTOS and IDRiD datasets

Figure 6.12 compares the proposed MSA-ResNetGB and SGAN-ECR models performance metrics on APTOS dataset. The structure SGAN and ECR models are combined to train the classification and regression model where the retinal images are specified as input to the MSA-ResNet model. The output is processed in the classification layer where the retinal images are categorized into DR stages based on the severity levels. The results of the SGAN-ECR model showed optimal performance compared to the MSA-ResNetGB model on APTOS dataset.

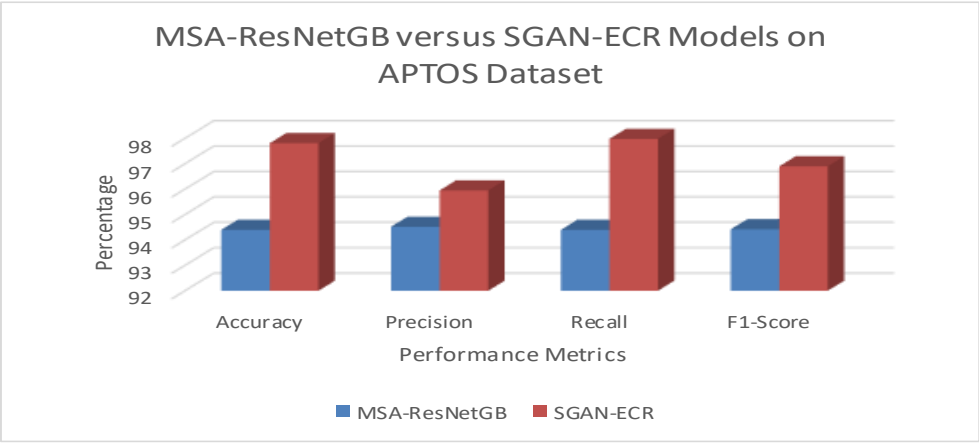

Figure 6.12: Comparison of MSA-ResNetGB and SGAN-ECR model on the APTOS dataset

Figure 6.13 compares the proposed MSA-ResNetGB and SGAN-ECR models performance metrics on IDRiD dataset. The structure SGAN and ECR models are combined

to train the classification and regression model where the retinal images are specified as input to the MSA-ResNet model. The output is processed in the classification layer where the retinal images are categorized based on the DR stages. The results of the SGAN-ECR model showed optimal performance compared to the MSA-ResNetGB model on IDRiD dataset.

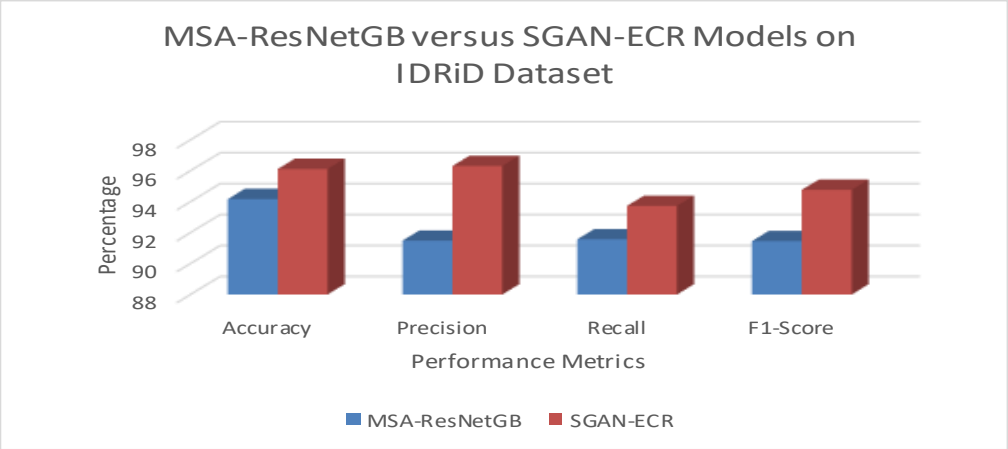

Figure 6.13: Comparison of MSA-ResNetGB and SGAN-ECR model on the IDRiD dataset

The proposed SGAN-ECR model's training ability is represented in Figure 6.14 which shows the accuracy plot for the SGAN-ECR model on both APTOS and IDRiD datasets. The accuracy of the model increases when the number of iterations increases. Figure 6.15 shows the loss plot for the proposed model that is good to fit both APTOS and IDRiD datasets. The training loss drops at a point of constancy that is optimal for the model. The average execution time (in seconds) of the proposed SGAN-ECR model model is 1.58 for APTOS dataset and 1.62 for IDRiD dataset.

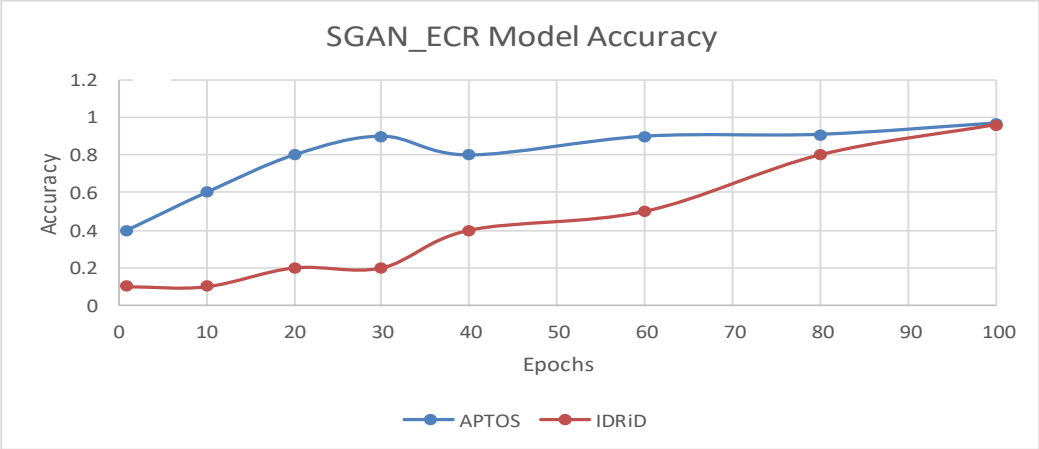

Figure 6.14: SGAN-ECR model Accuracy Plot

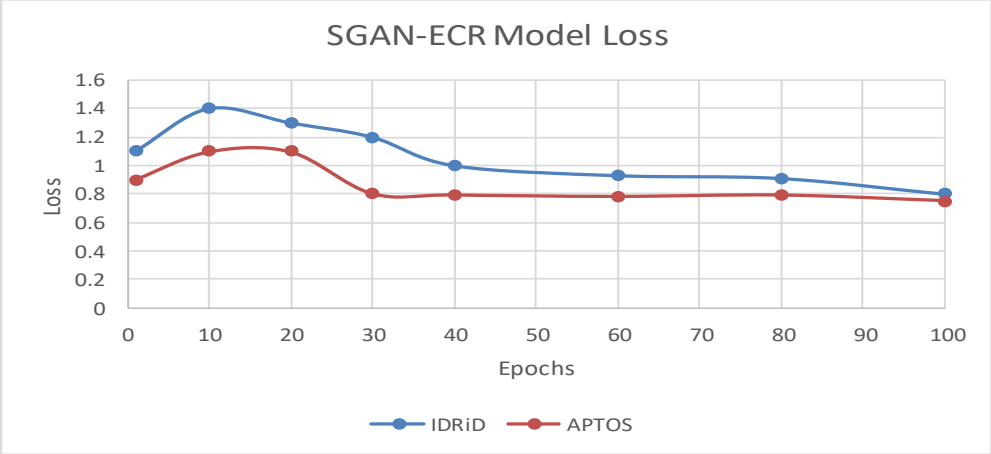

Figure 6.15: SGAN-ECR model Loss Plot

Table 6.6 represents the performance metrics of the SGAN-ECR model with and without preprocessing on the APTOS datasets. The performance metrics of the SGAN-ECR model without preprocessing produce better results compared to the results produced with preprocessing.

Table 6.6: Performance of SGAN-ECR model with and without preprocessing on APTOS dataset

Metric	SGAN-ECR Model with Preprocessing	SGAN-ECR Model without Preprocessing
Accuracy (%)	91.33	**97.81**
Precision (%)	90.12	**95.95**
Recall (%)	91.20	**97.98**
F1-Score (%)	90.66	**96.90**

Table 6.7 represents the performance metrics of the SGAN-ECR model with and without preprocessing on the IDRiD datasets. The performance metrics of the SGAN-ECR model without preprocessing produce better results compared to the results produced with preprocessing. The performance metrics of the SGAN-ECR model without preprocessing produce better results for the APTOS and IDRiD datasets.

Table 6.7: Performance of SGAN-ECR model with and without preprocessing on IDRiD dataset

Metric	SGAN-ECR Model with preprocessing	SGAN-ECR Model without preprocessing
Accuracy (%)	86.06	**96.12**
Precision (%)	85.03	**96.31**
Recall (%)	86.64	**93.73**
F1-Score (%)	85.69	**94.77**

6.5 Summary

In this chapter, the SGAN-ECR model was proposed to improve the DR stage classification effectively based on severity levels. First, to create the synthetic retinal images, SGAN was implemented to train the retinal images using the pixel-to-pixel GAN and patch-GAN. These images and the real retinal images are trained by the MSA-ResNet regression and MSA-ResNet classification models which have inter-dependency features and discriminative features based on the DR stages. Both these features are merged and given to MLP classification unit to categorize the different stages of DR. Finally, the cross-entropy error and mean square error define the DR severity levels. Experimental results are validated and analyzed based on the performance metrics, the outcome analysis and assessment of the proposed technique are compared with the literature work. Thus, the accuracy of 97.81% and 96.12% is achieved in the SGAN-ECR model on APTOS and IDRiD datasets respectively.