CHAPTER 5

MULTI-SCALE ATTENTION-BASED RESIDUAL NETWORK WITH GRADIENT BOOSTING FOR CLASSIFYING THE STAGES OF DIABETIC RETINOPATHY

5.1 Introduction

Multi-class Classification of DR stages in medical diagnosis is vital due to the severity of the disease, such as vision impairment. Blood vessels surrounding the retina provide oxygen and essential nutrients. The blood vessels are damaged due to diabetes, which leads to a lack of blood supply to the retina. Identifying the damaged blood vessels automatically using a screening system is necessary. An automated system classifies retinal images as normal and abnormal with the help of technical advancement. In this study, an adaptive residual neural network model named ResNetGB, which incorporates Gradient Boosting (GB), has been developed to synergize with the Multi-Scale Attention (MSA) approach named MSA-ResNetGB model for improved performance in classifying the stages of DR disease. A pre-trained Residual Neural Network (ResNet-32) has been used as a base network. A high-level interpretational space is utilized to implant the RF image, facilitating the fusion of mid and high-level data to improve interpretability. The extracted features are given as input to the classifier model to classify the retinal image as normal and abnormal images within the DR stages. This is accomplished by creating a MSFP to represent the RF image across locations. Also, to further boost the discriminative capacity of feature interpretation, the MSA method is applied within the high-level interpretational space. A detailed description related to the ResNetGB, multi-scale feature extraction analysis, MSA strategy, and classification units are defined in the subsequent sections of this chapter. Experimental results are analyzed based on the performance metrics, and the outcome analysis and assessment of the proposed technique are presented conclusively.

5.2 Multi-Scale Attention-Based Residual Network with Gradient Boosting Model

The proposed MSA-ResNetGB model is presented in detail. The encoder module of the ResNetGB model is described. The multi-scale feature extraction and analysis, MSA strategy, decoder and classification units are explained in the subsequent sections. In the architecture model, the preprocessing strategies are applied to the input image to regularize

and remove the image noise, as discussed in Chapter 4. The retinal images (pre-processed) are used to train the MSA-ResNetGB model to classify DR stages. The MSA-ResNetGB model consists of four components: the ResNetGB as an encoder component that encodes the input retina image into a high-level representation space, next is the MS feature extraction and analysis component, then a MSA tool, the decoder component that helps in classifying the different stages of DR. In the subsequent subsections, each component is discussed in further detail.

5.2.1. Residual Network along with the Gradient Boosting

In the proposed network architecture, the initial component, known as the encoder module, features the ResNetGB architecture, a combination of ResNet-34 and Gradient Boosting inspired by ResNet50. This encoder serves as the down-sampling part of the model, aiming to efficiently extract hierarchical features from input images with reduced memory consumption and faster feature extraction. The use of pre-trained ResNet50 facilitates the extraction of hierarchical features from images. To mitigate overfitting, Gradient Boosting is incorporated alongside the ResNet model, reweighting the input data and enhancing the model's generalization capability. The ResNetGB encoder employs the ResNet-34 as a feature extractor. The convolutional layers, batch normalization, ReLU activations, and residual blocks collectively work to extract hierarchical features, resulting in high-level feature maps capturing abstract patterns and representations learned during training. The obtained feature maps are then flattened or pooled into vectors, creating fixed-size representations for each input image. These vectors serve as feature sets, encapsulating relevant information. The Gradient Boosting model is subsequently trained on labeled data to establish relationships between ResNet-34 features and the target variable, effectively capturing complex interactions and non-linearity in the data.

The input data with the dimension of 256x256x3 is set with an applied convolutional layer of 7x7 with 64 filters. Each convolutional layer is followed by the Batch Normalization (BN) layer to reduce the overfitting of the model during the feature extraction process. The model is built with the ReLu activation function, which helps in the reduction of the spatial dimension of the images. This will help in increasing the processing speed of the model. Each convolutional layer ends with the max pooling layer with a kernel size of 3x3 with a stride value of 2. So, the feature extraction is performed with the size of 2x2 regions of the images.

The series of residual blocks is comprised of two 3x3 convolutional layers followed by the BN and ReLu activation functions. The skip connections are added to address the vanishing gradient problem, and it helps store the feature information. After each residual block, the global average pooling is used to set the standard size for the feature extraction. The feature vector obtained is taken as the input to the gradient boosting model. The model is trained with gradient boosting to label the feature data of the images. In the training set, with the help of the ResNetGB encoder, the collected images are embedded into the representational space as represented in Equation (5.1) where F_{ec} generation represents the feature extracted, and G represents the encoder, i.e., ResNetGB model and:

$$F_{ec} = G(\theta_1;1) \tag{5.1}$$

The ResNetGB network model is utilized without a fully connected layer, which is a PCA layer as the decoder. Figure 5.1 shows the Network Architecture of the Proposed MSA-ResNetGB model.

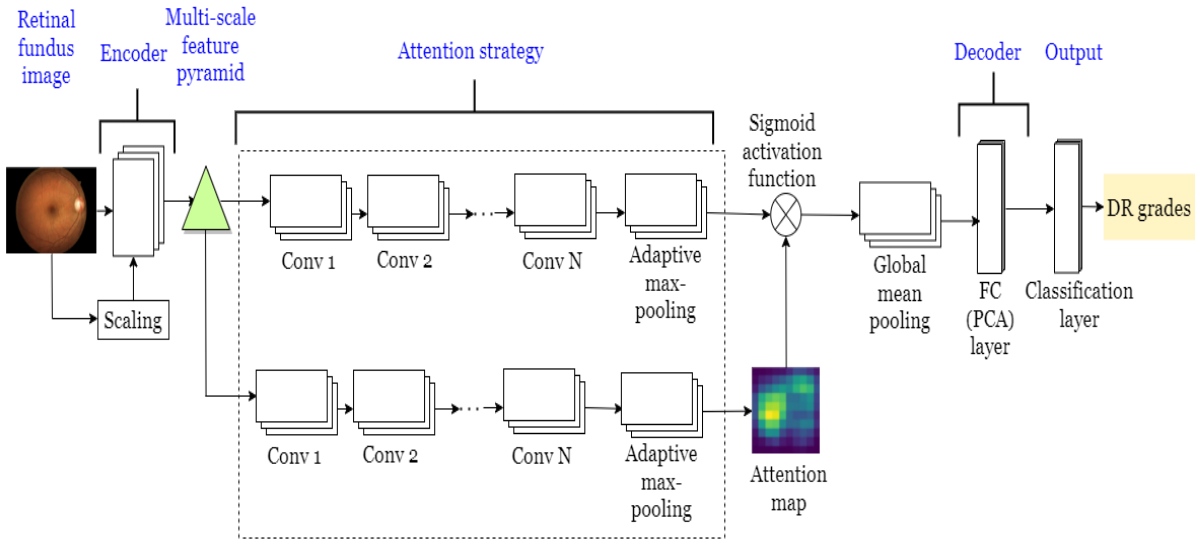

Figure 5.1: Network Architecture of the Proposed MSA-ResNetGB model

5.2.2. Multi-Scale Feature Extraction and Analysis

The encoder extracts feature through successive residual blocks. The image resolution is higher for images near the input, facilitating feature extraction. Semantic information related to local features is prominent in the final layer. The next stage involves multilevel features combined with mid and high-level features. A scaling mechanism imparts semantic

learning to features with various spatial resolutions. Multi-scale feature extraction, a technique in image processing, captures information at different spatial scales. It uses convolutional kernels with different scales, decomposing the image into multiple scales. Pyramid decomposition generates different scales, and features are independently extracted at each scale, capturing fine details and coarse structures. The extracted features from various scales are joined to hierarchical representations in DL. The input to the feature sets are an Atrous convolution, extracting information with various scales. Convolutional filters with different dimensions are applied, encoding local information with a small field of view and global information considering a larger image context (Azad R *et al.*, 2019).

The Multi-Scale feature interpretation encodes the input image in a compressed representation space proficient in learning signs of diabetics with localization, different scales, and severities. The initial layers comprise convolutional blocks using atrous convolution, batch normalization, and ReLU activation. These layers progressively reduce spatial dimensions, resulting in feature maps of size 128x128. Atrous convolutional blocks with varying dilation rates capture features at different scales, the final convolutional layer outputs feature maps of size 128x128 with a specified number of filters. Subsequently, the Atrous Spatial Pyramid Pooling (ASPP) module is applied, employing atrous convolutions with multiple dilation rates in parallel. Outputs from parallel convolutions are concatenated to form a 128x128 feature map with a combined set of filters. Additional convolutional layers refine these concatenated features. An upsampling layer increases spatial resolution, bringing the feature map back to 256x256. Skip connections are incorporated by concatenating features from an earlier encoder layer.

5.2.3. Multi-Scale Attention strategy

In the proposed MSA-ResNetGB model, a sequence of convolution units is used for MS feature interpretation. The retinal structure can be deformed based on the DR stages. The RF can be damaged during the deformation process. A DNN model is applied to classify the damage for the high-level representation. The high-level representation distinguishes the different classes proficiently. Due to the shortage of DR patterns, the efficiency of identifying the damage is limited. The proposed model enhances the discriminative power of this representation by including the attention strategy on the multi-scale representation. This strategy aims to learn how to identify the retina damage and scale the representation space.

The model focuses on the unwell parts in the retinal image and does not emphasize the normal regions. The pyramid representations are used to create a small interpretation for point-wise convolution, as shown in Equation (5.2).

$$\mathrm{A}^{hxwx64} = (F_{all}^{h*w*4c} * \mathrm{K}_{pw}) \tag{5.2}$$

Pseudo Code: The Proposed MSA-ResNetGB model for DR Stage Classification

Input: *Retinal Fundus images from APTOS & IDRiD dataset along with Synthetic images* $I_1, \ldots, I_n$

Output: *Different DR Classes/Stages*

Begin

Step 1: *Split the Dataset (training & test sets).*

Step 2: *In training set, with the help of the ResNetGB encoder, the collected images are embedded into the high-level realistic space.*

Equation (5.1) shows the interpretation of tensor F_{ec} *generation:*

$$F_{ec} = G(\theta_1;1) \tag{5.1}$$

Step 3: *The attention strategy is included in the MS feature interpretation.*

Step 4: *High-level feature analysis is improved; the point-wise convolution and pyramid representations are used.*

Step 5: *Create the attention tensor* A *. Equation (5.2) shows the attention tensor:*

$$\mathrm{A}^{hxwx64} = (F_{all}^{h*w*4c} * \mathrm{K}_{pw}) \tag{5.2}$$

Step 6: *Acquire the global interpretation by Global Mean Pooling (GMP) and create feature analysis vector F. Equation (5.3) shows the feature interpretation vector:*

$$F^{1x1024} = \frac{GMP(\sigma(F^{h*w*c} \odot A^{h*w*1}))}{GMP(A^{h*w*1})} \tag{5.3}$$

Step 7: *Feature vectors are mapped using the PCA layer to the necessary outcomes.*

Step 8: *Classification layer is trained using the GB classifier. Estimate the loss factor and adjust the variables.*

Step 9: *Classify test images (DR stages) using the trained MSA-ResNetGB model.*

End

'A' denotes attention tensor, and 'F' denotes analysis tensor produced using the MS unit. 'K_{pw}' denotes the point-wise. 'h and w' denote the feature map's height and width, and finally, 'c' denotes the channel count. The produced small interpretation is implemented to the convolution sequence to generate an attention map $A^{(h\times w\times 1)}$ and is multiplied by the high-level interpretation $F^{(h\times w\times c)}$ to limit the interpretation. The sigmoid activation (σ) is normalized in the range from 0 to 1. The Global Mean Pooling (GMP) is retrieved in the final interpretation by normalizing the GMP data $A^{(h\times w\times 1)}$. Then, the generated absolute feature interpretation vector 'F' is given in Equation (5.3).

$$F^{1x1024} = \frac{GMP(\sigma(F^{h*w*c} \odot A^{h*w*1}))}{GMP(A^{h*w*1})} \quad (5.3)$$

The '⊙' denotes point-wise multiplication. The MS learning variables and attention units are denoted as θ_2. The MSA strategy improves the training and aids in increasing the accuracy of RF image classification. The output of the preceding units differentiates the created framework from the ResNetGB or other CNN structures.

A tensor can be a generic structure that can be used for storing, representing, and changing data. Tensors are the fundamental data structure used by all machine and deep learning algorithms.

Table 5.1: List of Notations in the MSA-ResNetGB model

Notation	Description
G	Encoder structure
θ_1	Encoding variable
I	Retinal fundus image
F_{ec}	Interpretation tensor
$\mathcal{A}$	Attention tensor
$\mathbb{K}_{pw}$	Point-wise convolution kernel
h	Height of the feature map
w	Width of the feature map
c	Number of channels
σ	Sigmoid activation
$\odot$	Point-wise multiplication
F	Absolute feature interpretation vector
θ_2	Learning variables for the multi-scale and attention units
$\mathcal{L}(\theta, \varphi)$	Loss factor

5.2.4. Decoder and classification units

The final part of the network in the proposed model is the decoder unit, which consists of fully connected layers to map the feature vectors to the outcome in the principal component analysis layer. The retinal image classification is the objective of the proposed network, where the fully connected layers with the pre-learned structure are trained. There are a vast number of variables to be trained in the fully connected layers. The modified structures are included in the principal component analysis layer that has neurons, which aids in minimizing the training efforts. The modified ResNet model retrieves the characteristics related to the common patterns from the retinal images. These characteristics are fed into the GB classifier that classifies the retinal fundus images of DR patients based on severity levels. The aim of automatic DR classification is to assist the ophthalmologist in the identification of DR stages in patients based on the classification result. $L(\theta, \varphi)$ denotes the loss function for the classification loss model along with the encoder and attention parameters, $\theta = \theta_1 \cup \theta_2$ and the φ denotes the branch parameter for classification. The predicted and the actual class are denoted as cross-entropy loss.

5.3. Experimental Result and Analysis

This section discusses the performance metrics of the MSA-ResNetGB model, such as accuracy, precision, recall, and F1-score. The evaluation of the proposed model using datasets APTOS and IDRiD are described. The proposed and the existing classification methods comparison and the confusion matrices are elaborated.

5.3.1. Training and Testing

The images in the dataset are resized to 256x256 pixels with three color channels (RGB). During training, a batch size of 32 is employed, indicating that the model's parameters are updated after processing 32 images at once. The learning rate is set at 0.001, controlling the step size for parameter updates. The Adam optimizer is chosen for its efficiency in handling noisy or sparse gradients, which is particularly suitable for medical images. Categorical cross-entropy is the selected loss function, comparing predicted outputs with ground truth labels. The model undergoes 100 epochs, signifying that the entire training dataset is iterated through the model 100 times for training.

5.3.2. Performance Evaluation

The performance of the proposed MSA-ResNetGB model is examined using the following measures: Accuracy (A), Precision (p), Recall (r), and F1-score (FS). The mathematical formulation of the metrics is computed as shown in Equation (5.4) for accuracy, Equation (5.5) for precision, Equation (5.6) for recall and Equation (5.7) for F1-Score.

$$A = \frac{TP+TN}{TP+TN+FP+FN} \tag{5.4}$$

$$p = \frac{TP}{TP+FN} \tag{5.5}$$

$$r = \frac{TN}{TN+TP} \tag{5.6}$$

$$FS = \frac{2*p*r}{p+r} \tag{5.7}$$

Where True Positive (TP) is the categorization accuracy of positive sample count, True Negative (TN) is the classification of negative sample count, and False Positive (FP) represents the negative class sample ratio, which is categorized under positive class. In contrast, False Negative (FN) represents the positive class sample ratio, which is categorized under the negative class.

5.3.3 Result Analysis of MSA-ResNetGB model on the APTOS dataset

The performance metrics of the MSA-ResNetGB model are trained on 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. In the APTOS dataset, out of 3662 training samples, 20% of the samples are considered for test data. Table 5.2 shows the existing works for DR Stage Classification on the APTOS 2019 blindness detection dataset in the literature work, including the Modified Xception model, MobileNetV2 model, Composite gated attention, Xception Multitask model, Hybrid Inception ResNet-v2, ExtraTree model, and DenseNet121 model. Table 5.3 compares the performance of the proposed model with the existing CNN variants under the same setting. Accuracy metric is the frequently used metric, and the other metrics that are not stated in the literature are represented as '-.' CNN is the commonly used DL method for medical image analysis. When compared to the proposed

MSA-ResNetGB model, the CNN variant models underperform, resulting in an accuracy difference of 11.69% and 24.32%, respectively. This demonstrates the effectiveness of the proposed model on retinal fundus images.

Table 5.2: Existing works for DR Stage Classification on the APTOS 2019 dataset

Reference	Classification method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
O Dekhil *et al.* (2019)	CNN model	77.00	-	-	-
S. H. Kassani *et al.* (2019)	Modified Xception model	83.09	-	88.24	-
J. D. Bodapati *et al.* (2021)	Composite gated attention DNN	82.54	82.00	83.00	82.00
W. L. Alyoubi *et al.* (2021)	CNN512 model	84.10	-	-	-
S. Majumder and N. Kehtarnavaz (2021)	Xception Multitask Model	86.00	77.00	70.00	73.00
A. K. Gangwar and V. Ravi (2021)	Hybrid Inception ResNet-v2	82.18	-	-	-
N. Sikder *et al.* (2019)	ExtraTree model	91.07	90.40	89.54	89.97
L. Wang and A. Schaefer (2020)	MobileNetV2 model	78.47	68.66	60.01	64.04
S. Sheikh and U. Qidwai (2021)	DenseNet121 model	90.50	93.00	90.00	88.47
N. Sikder (2021)	Tuned XGBoost model	94.20	94.34	92.68	93.51
M. T. Al-Antary and Y. Arafa (2021)	MSA-Net model	84.60	-	91.00	-

Table 5.3: Performance comparison of the Proposed Model with the existing CNN variants for DR Stage Classification on the APTOS 2019 dataset

Reference	Classification Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Srinivasan V & Rajagopal V (Proposed, 2022)	**MSA-ResNetGB Model**	**94.40**	**94.53**	**94.40**	**94.43**
Srinivasan V & Rajagopal V (Existing)	ResNet-50 Model	82.71	82.65	81.82	82.23
Srinivasan V & Rajagopal V (Existing)	VGG-19 Model	70.08	71.45	69.27	70.59

Figure 5.2 shows the confusion matrix of the MSA-ResNetGB model on the APTOS dataset. The matrix illustrates the scattering of samples by classes and the ratio of accurate

classification and misclassified samples. For example, the precisely identified sample count is 351 for class 0. For class 1, the precisely identified sample count is 70. For class 2, the precisely identified sample count is 187. For class 3, the precisely identified sample count is 29 and finally, for class 4, the precisely identified sample count is 54.

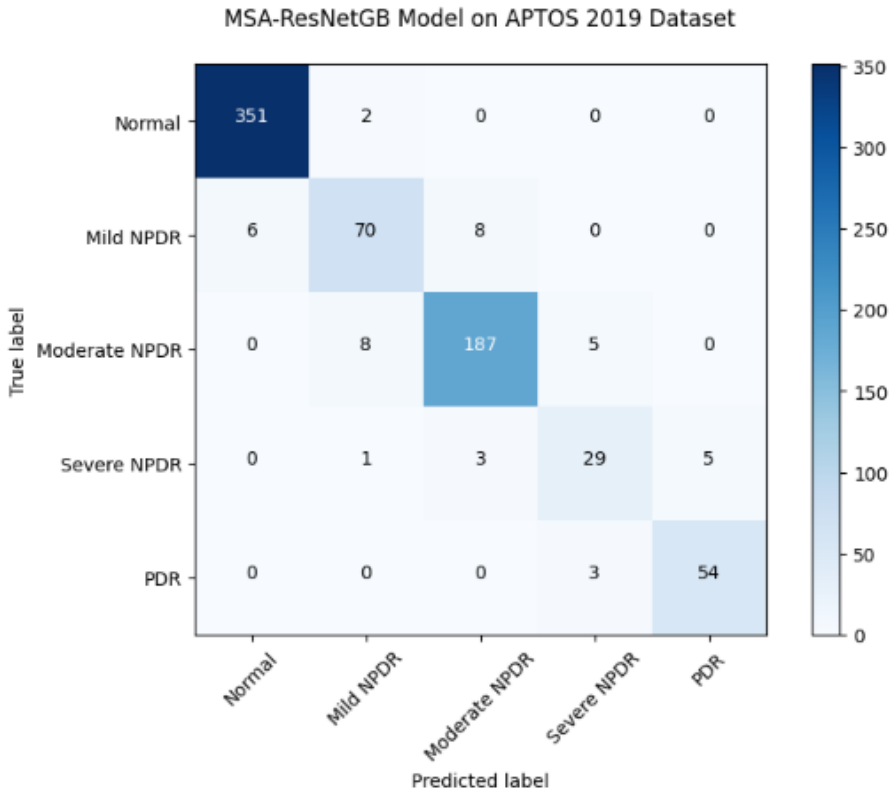

Figure 5.2: Confusion matrix for MSA-ResNetGB model on the APTOS 2019 dataset

5.3.4. Result Analysis of MSA-ResNetGB model on the IDRiD dataset

Multi-class classification is performed on the IDRiD, having five classes starting from 0 to 4. In the IDRiD dataset, out of 516 training samples, 20% of the samples are considered for test data. Table 5.4 shows the existing works for DR Stage Classification on the IDRiD dataset in the literature work, including the AlexNet model, CANet model, EffiecientNet B0 model and GNN-based models. Table 5.5 shows the performance of the proposed model. Accuracy metric is the frequently used metric, and the other metrics that are not stated in the literature are represented as '-.'

Table 5.4: Existing works for DR Stage Classification on the IDRiD dataset

Reference	Classification method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
B. Harangi *et al.* (2019)	AlexNet model	90.07	-	-	-
X. Li (2020)	CANet model	65.10	-	-	-
E. Abdelmaksoud (2021)	EfficientNet B0 model	86.00	-	-	-
A. Sakaguchi (2019)	GNN-based model	79.30	-	-	-

Table 5.5: Performance comparison of the Proposed Model with the existing CNN variants for DR Stage Classification on the IDRiD dataset

Reference	Classification Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Srinivasan V & Rajagopal V (Proposed, 2022)	**MSA-ResNetGB Model**	**94.18**	**91.48**	**91.57**	**91.45**
Srinivasan V & Rajagopal V (Existing)	ResNet-50 Model	78.24	78.56	77.82	78.30
Srinivasan V & Rajagopal V (Existing)	VGG-19 Model	67.98	66.50	66.04	67.21

The performance metrics are analyzed for the MSA-ResNetGB model on the IDRiD dataset. When compared to the proposed MSA-ResNetGB model, the CNN variant models underperform, resulting in an accuracy difference of 15.94% and 26.20%, respectively. This demonstrates the effectiveness of the proposed model on retinal fundus images. Figure 5.3 shows the confusion matrix of the MSA-ResNetGB model on the IDRiD dataset. The matrix illustrates the distribution of samples by classes and the ratio of accurate classification and misclassified samples. For example, the precisely identified sample count is 34 for class 0. For class 1, the precisely identified sample count is 2. For class 2, the precisely identified sample count is 29. For class 3, the precisely identified sample count is 13 and finally, for class 4, the precisely identified sample count is 19.

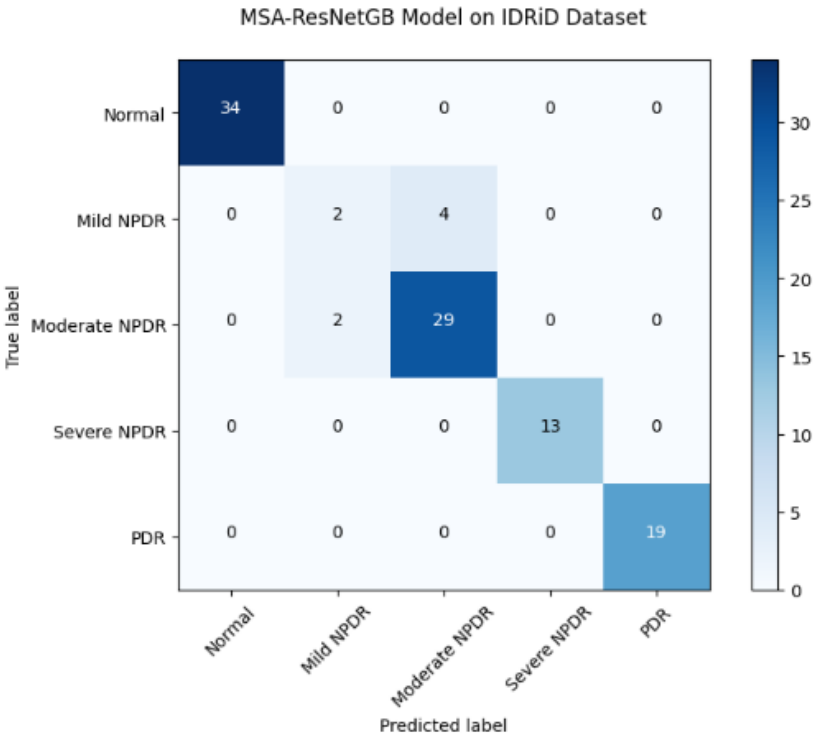

Figure 5.3: Confusion matrix for MSA-ResNetGB model on the IDRiD dataset

5.3.5 Results obtained from the MSA-ResNetGB model

The results obtained by the proposed MSA-ResNetGB model applied on the APTOS and IDRiD datasets outperform the other CNN variant models. The local and global feature representation aids in the learning process of the DR structure with different feature selections and locations, leading to enhanced performance. The MSA strategy is applied to the high-level interpretation to focus on the vital areas to identify the DR stages. Table 5.6 compares the performance of the MSA-ResNetGB model on the APTOS and IDRiD datasets. The MSA-ResNetGB model on the APTOS dataset achieved 94.40% accuracy, 94.53% precision, 94.40% recall, and 94.43% F1-Score. The MSA-ResNetGB model on the IDRiD dataset has obtained 94.18% accuracy, 91.48% precision, 91.57% recall, and 91.45% F1-Score. The performance of the APTOS dataset exceeds the performance of the IDRiD dataset. When both datasets' performance is compared, there is a performance loss in the IDRiD dataset since fewer image samples are used for training the MSA-ResNetGB model compared to the APTOS dataset. Data augmentation techniques enable the production of effective and unbiased outcomes.

Table 5.6: Performance Comparison of MSA-ResNetGB model on the APTOS and IDRiD dataset

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
APTOS	94.40	94.53	94.40	94.43
IDRiD	94.18	91.48	91.57	91.45

Figure 5.4 represents the accuracy analysis of the MSA-ResNetGB model on the 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, respectively. For the IDRiD dataset, the accuracy achieved is 100%, 94%, 94%, 92%, and 95% for the classes 0 to 4, respectively. The plot shows that the performance of class 3 is better than that of class 2 for the APTOS dataset. In the case of the IDRiD dataset, the performance of class 0 is better than that of class 3.

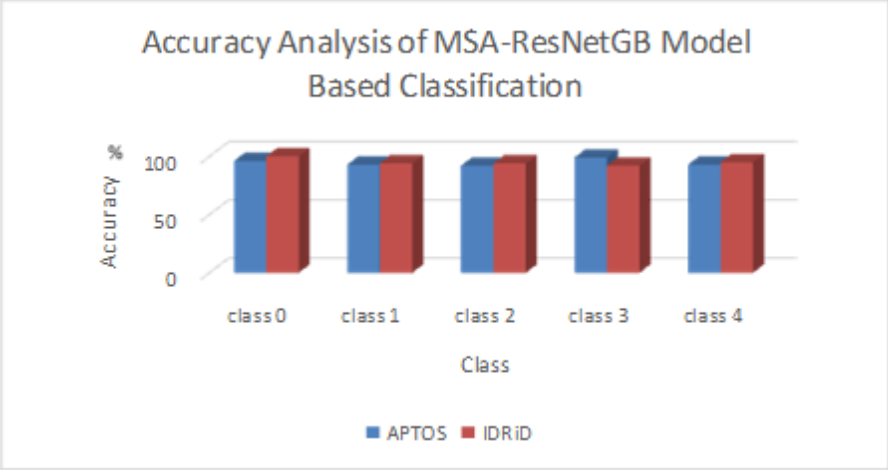

Figure 5.4: Accuracy Analysis of MSA-ResNetGB model on the APTOS and IDRiD dataset

Figure 5.5 represents the precision analysis of the MSA-ResNetGB model on the APTOS and IDRiD datasets. The datasets are labeled as five classes ranging from 0 to 4. For the APTOS dataset, the precision achieved is 96%, 93%, 92%, 98% and 94% for the classes 0 to 4, respectively. For the IDRiD dataset, the precision achieved is 88%, 89%, 90%, 91%, and 91% for the classes 0 to 4, respectively. The plot shows that the performance of class 3 is

better than that of class 2 for the APTOS dataset. In the case of the IDRiD dataset, the performance of class 3 and class 4 is better.

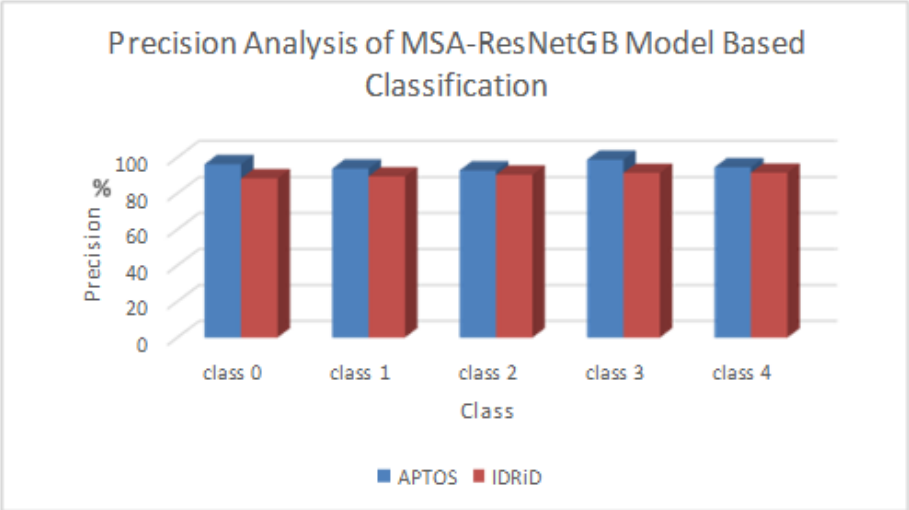

Figure 5.5: Precision Analysis of MSA-ResNetGB model on the APTOS and IDRiD dataset

Figure 5.6 represents the recall analysis of the MSA-ResNetGB model on the APTOS and IDRiD datasets. The datasets are labeled as five classes ranging from 0 to 4. For the APTOS dataset, the recall achieved is 95%, 93%, 93%, 100% and 94% for the classes 0 to 4, respectively. For the IDRiD dataset, the recall achieved is 87%, 88%, 90%, 91%, and 91% for the classes 0 to 4, respectively. The plot shows that the performance of class 3 is better for the APTOS dataset. In the case of the IDRiD dataset, the performance of class 3 and class 4 is better.

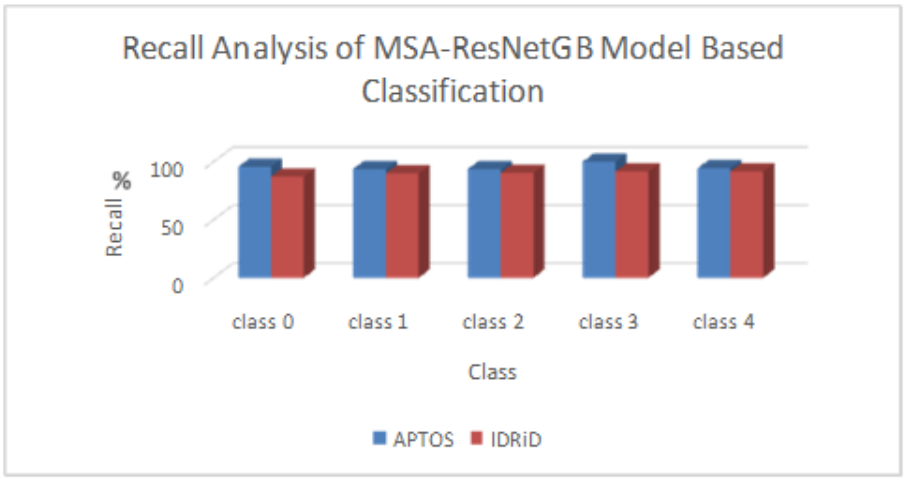

Figure 5.6: Recall Analysis of MSA-ResNetGB model on the APTOS and IDRiD dataset

Figure 5.7 represents the F1-Score analysis of the MSA-ResNetGB model on the APTOS and IDRiD datasets. The datasets are labeled as five classes ranging from 0 to 4, respectively. For the APTOS dataset the F1-Score achieved is 92%, 93%, 94%, 99% and 94% for the classes 0 to 4, respectively. For the IDRiD dataset, the F1-Score achieved is 94%, 98%, 90%, 92%, and 91% for the classes 0 to 4, respectively. The plot shows that the performance of class 3 is better for the APTOS dataset. In the case of the IDRiD dataset, the performance of class 1 is better.

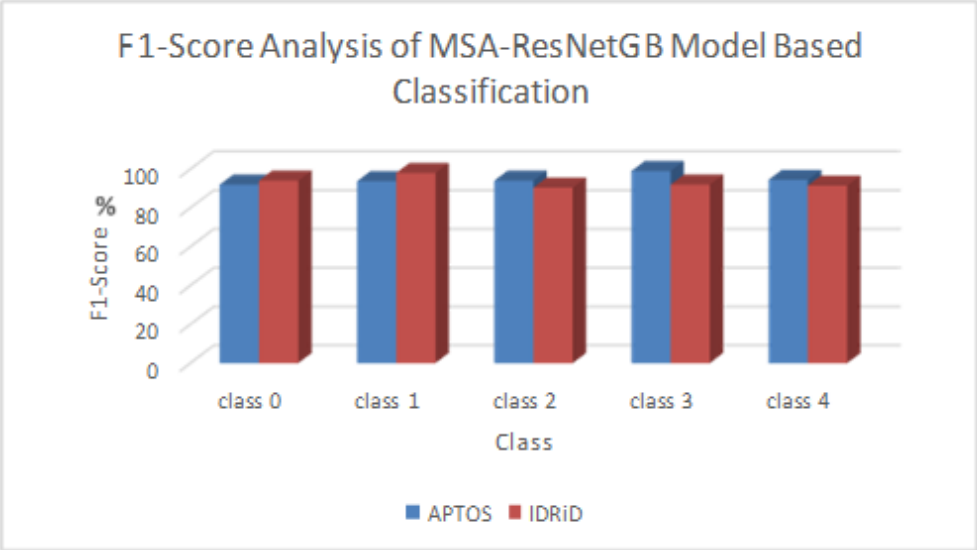

Figure 5.7: F1-Score Analysis of MSA-ResNetGB model on the APTOS and IDRiD dataset

The proposed MSA-ResNetGB model's training ability is represented in Figure 5.8, which shows the accuracy plot for the MSA-ResNetGB model on both APTOS and IDRiD datasets. The accuracy of the model increases when the number of iterations increases. Figure 5.9 shows the loss plot for the proposed model that is good to fit on both APTOS and IDRiD datasets. Cross-entropy loss has been used in this work as it is a multi-class classification. The training loss drops at a point of constancy that is optimal for the model. The average execution time (in seconds) of the proposed MSA-ResNetGB model is 1.72 for APTOS dataset and 1.84 for IDRiD dataset.

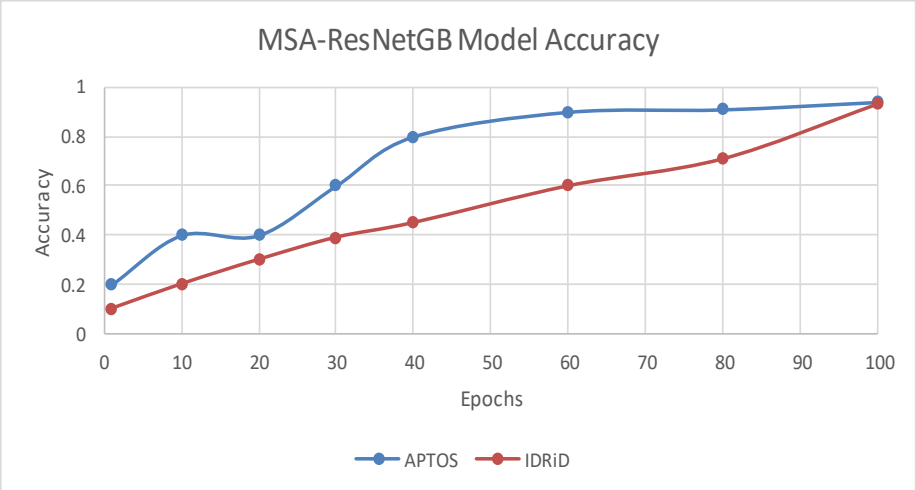

Figure 5.8: MSA-ResNetGB model Accuracy Plot

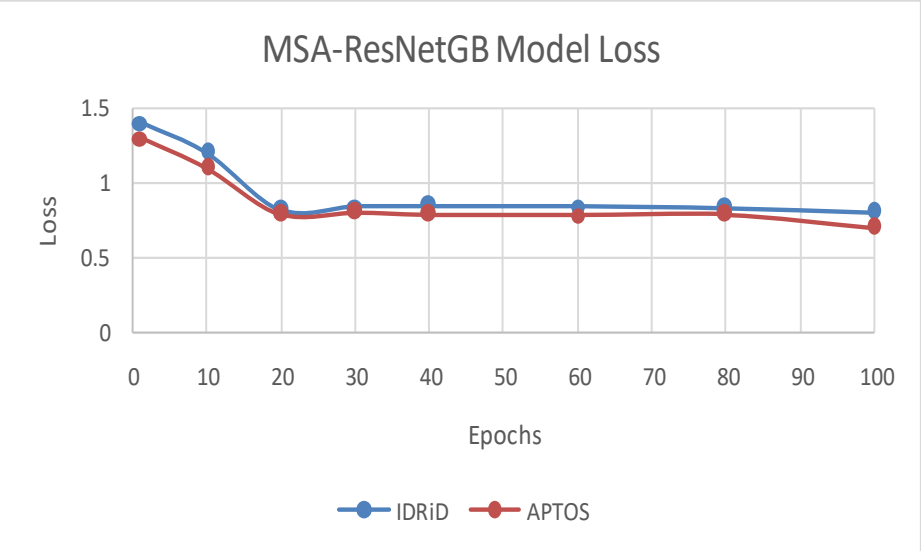

Figure 5.9: MSA-ResNetGB model Loss Plot

Table 5.7 represents the performance metrics of the MSA-ResNetGB model with and without preprocessing on the APTOS datasets. The performance metrics of the MSA-ResNetGB model with preprocessing produced better results compared to the results produced without preprocessing.

Table 5.7: Performance of MSA-ResNetGB model with and without Preprocessing on the APTOS dataset

Metric	MSA-ResNetGB Model with preprocessing	MSA-ResNetGB Model without preprocessing
Accuracy (%)	**94.40**	94.01
Precision (%)	**94.53**	93.65
Recall (%)	**94.40**	93.73
F1-Score (%)	**94.43**	93.21

Table 5.8 represents the performance metrics of the MSA-ResNetGB model with and without preprocessing on the IDRiD datasets. The performance metrics of the MSA-ResNetGB model with preprocessing produced better results compared to the results produced without preprocessing.

Table 5.8: Performance of MSA-ResNetGB model with and without Preprocessing on IDRiD dataset

Metric	MSA-ResNetGB Model with preprocessing	MSA-ResNetGB Model without preprocessing
Accuracy (%)	**94.18**	92.31
Precision (%)	**91.48**	91.02
Recall (%)	**91.57**	90.25
F1-Score (%)	**91.45**	90.69

5.4. Summary

In this chapter, the MSA-ResNetGB model was proposed to classify the stages of DR disease based on the severity levels. A ResNetGB encoder is used to incorporate retinal images into high-level feature interpretation locations. The model describes the retinal image patterns to scale various locations with the help of multi-scale feature extraction analysis. The feature interpretation efficiency is enhanced on the high-level interpretation in the MSA strategy. The decoder module performs the mapping of fully connected layers to the feature vectors to produce the outcome in the principal component analysis layer. Finally, the

structure was trained on the cross-entropy error to define the DR classification properly. Experimental results are validated and analyzed based on the performance metrics, and the outcome analysis and assessment of the proposed technique are compared with the literature work. Thus, the accuracy of 94.40% and 94.17% is achieved in the MSA-ResNetGB model on APTOS and IDRiD datasets, respectively.