Received: April 24, 2022. Revised: June 10, 2022.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Multi-Scale Attention-Based Mechanism in Gradient Boosting Convolutional Neural Network for Diabetic Retinopathy Grade Classification

Valarmathi Srinivasan[1]* **Vijayabhanu Rajagopal[1]**

[1]*Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India*
* Corresponding author's Email: valarmathis313@gmail.com

Abstract: Diabetic Retinopathy (DR) is a common complication of Diabetes Mellitus (DM) that produces retinal abnormalities and can lead to blindness if not diagnosed and treated on time. To address this concern, an adaptive Convolutional Neural Network (CNN) model with Gradient Boosting (GB) called ResNetGB has been used from the literature where a Principal Component Analysis (PCA) based Fully Connected (FC) layer is used to capture the discriminative characteristics from the Retinal Fundus (RF) samples. It is essential to extract more effective features to categorize the DR grades. Hence, in this article, the Multi-Scale Attention (MSA) strategy is incorporated into the ResNetGB model for effective DR grade classification. First, the encoder network is used to embed the RF image in a high-level interpretational space in which the mixture of mid and high-level characteristics is considered to enhance the representation. Then, a Multi-Scale Feature Pyramid (MSFP) is added to define the retinal pattern in various localities and the MSA strategy is applied to the high-level interpretation. Moreover, the entire MSA-ResNetGB framework is trained by the cross-entropy loss to categorize the patients with respective DR grades. Finally, the experimental analysis exhibit that the MSA-ResNetGB model achieves the 94.40% and 94.17% accuracy on two benchmark datasets: Kaggle-APTOS and IDRiD, respectively compared to the cutting-edge models.

Keywords: Diabetic retinopathy, CNN, Gradient boosting, Multi-scale feature pyramid, Attention strategy.

1. Introduction

Diabetes Mellitus is one of the most prevalent causes of DR, which produces vision impairment. It involves varying degrees of severity and is treated when identified ahead of time [1]. DR affects the retina, which is in charge of turning light into an electronic signal that can be processed to produce an image. A network of blood vessels surrounds the retina, supplying it with oxygen and nutrients. Diabetes damages blood vessels, resulting in a shortage of blood supply to the retina. This deteriorates the health of the retina and, as a result, affects visual acuity. The most severe kind of DR is Non-proliferative retinopathy. DM does not influence vision at this level, although it does alter blood vessels. The arteries can expand moderately (microaneurysms – MAs, exudates - EXs, or retinal hemorrhages - HEs) [2].

In the final phase, DR progresses to proliferative retinopathy, which causes greater damage to the retina than non-proliferative retinopathy. When the majority of the retina is deprived of sufficient blood flow, it raises the possibility of visual impairment and poses a danger of vision loss. Manual detection of DR by ophthalmologists or skilled graders is costly. Diagnosis of a huge proportion of diabetic patients for probable DR incidence places a significant strain on ophthalmologists or evaluators, reducing their reliability and avoiding DR diagnoses. The screening is highly subjective, with various graders emerging with many interpretations [3].

To tackle this concern, modified and pre-learned CNN structures are used for the DR classification [4]. The most advanced pre-learned CNN structures, such as VGG and Residual Network (ResNet) [5] and Dual Path Network (DPN) [6] are often learned on ImageNet [7] and convert the low-level

characteristics to high-level characteristics. As DR grading is dependent on the existence of lesions in RF images such as MAs, EXs, and HMs, Saeed et al. [8] developed a 2-phase transfer learning employing a pre-learned CNN structure such as VGG19, ResNet152 and DPN107. The approach encompasses a RF scan as input, analyzes it using the modified framework, and then classifies it as healthy or DR. The CNN structure trains the domain-specific structure of low and high-level characteristics. First, the foremost layer of the pre-trained CNN structure called ResNetGB is reinitialized by the Regions-of-Interest (ROI) of tumors retrieved from the labeled RF samples. For lesion ROI extraction, E-optha is considered since it involves pixel-level tumor labeling. Then, the structure is optimized, wherein the low-level units train the neighborhood patterns of the tumor and healthy areas. Because the FC units convert high-level characteristics, these are substituted by the novel FC unit depending on the PCA and apply it in an unsupervised scheme to capture discriminative characteristics from the RF samples. Also, the GB-based classifier unit is added to estimate the DR scores of RF scans. Although it extracts high-level features to identify and classify the retina lesions into DR classes, highly effective features are essential to categorize the DR severity stage. The following are the proposed work's significant contributions:

1. The ResNetGB with MSA approach is presented to increase the accuracy of DR grade categorization. To begin, the encoder network is utilized to place the RF image in a high-level interpretational space, with a blend of mid and high-level data added to reinforce the interpretation. The MSFP is also built to specify the RF pattern in various locations.
2. The MSA strategy is employed in the high-level interpretation to improve the discriminative power of the feature interpretation.
3. Using the cross-entropy loss, the MSA-ResNetGB model is trained to detect DR patients based on their DR grades. As a result, distinguishing between normal and DR patients is simplified.
4. The proposed MSA-ResNetGB model is tested on two public benchmark datasets: Kaggle-APTOS 2019 dataset and Indian Diabetic Retinopathy Image Dataset (IDRiD).
5. The MSA-ResNetGB model is validated using four performance metrics: Accuracy, Precision, Recall, and F1-Score.

The rest of this paper is prepared as follows: The recent work linked with the DR detection and classification model is discussed in Section 2. Section 3 explains the ResNetGB-MSA model, while Section 4 demonstrates its efficacy. Section 5summarizes this paper and suggests its possible improvement.

2. Related work

A novel modified Xception-based feature mining and Multi-Layer Perceptron (PLP) classification model [9] have been developed to categorize DR criticality and diagnose them efficiently. An ensemble model [10] has been designed, which combines CNN and classical hand-crafted features into a single structure to classify RF images. A framework [11] was developed by considering a 3-channel RF image as input and resulting in the criticality of DR. Also, transfer learning was applied to the pre-trained MobileNetV2 and a weighted loss function was utilized.

A new Cross-disease Attention Network (CANet) [12] was designed to jointly grade DR and diabetic macular edema by finding the internal correlation among the diseases with only image-level supervision. A modified EfficientNet structure [13] was suggested to classify the early and advanced grades of the DR disease.

An improved cross-entropy loss function and three hybrid CNN models [14] have been developed to classify DR. A new CNN structure [15] was introduced to capture characteristics from RF scans. A new technique [16] was presented for DR prognosis depending on the gray-level pixels and fine details from the RF scans by the decision tree-based ensemble training. A composite Deep Neural Network (DNN) structure [17] was integrated with a gated-attention strategy for automated prognosis of DR. Deep transfer learning [18] was investigated based on the AlexNet, GoogleNet, InceptionV4, Inception ResNetV2 and ResNext50 to automatically diagnose DR.

CNN-based computer-aided diagnosis system [19] was developed to categorize RF images into different grades of DR. Two deep learning-based models [20] were developed: a CNN512 was utilized to categorize the RF image into different grades of DR, as well as, an adopted YOLOv3 was utilized to identify and localize the DR lesions.

A multi-task deep learning system [21] was developed using a modified Squeeze Excitation densely connected DNN and Xecption network as a multitasking scheme. Also, the MLP was used as a classification to categorize the DR grades. A new deep learning hybrid model [22] was developed, in which transfer learning was applied on pre-trained Inception-resNetV2 and a custom module of CNN layers was added on top of Inception-resNetV2 to recognize DR disorders.

A new early blind recognition technique [23] was designed using the color information obtained from RF images based on the ensemble learning scheme such as Extra tree model. Different deep learning-based classifiers [24] have been analyzed such as VGG16, ResNet50, InceptionV3 and DenseNet121 to partition the retina areas and categorize DR severity grades. The MSA Network (MSA-Net) [25] based on the encoder and decoder structure was designed for DR categorization. A Graph Neural Network (GNN) model [26] was designed to categorize DR severity.

2.1 Problem definition

From the above-studied related works, the issues in classifying the DR severity grades are:

- The number of training and testing images was limited, which impacts the accuracy of classifying the DR grades.
- Also, the accuracy was degraded because of imbalanced classes in the RF image databases.
- The imbalanced classes affect the significance of the feature maps during training and classification.
- Some CNN structures trained by only image-level supervision, which makes the model very difficult to recognize the exact abnormal signs like soft EXs, hard EXs, MAs and HEs.
- The classification efficiency mainly relies on the RF image resolution and the optimal hyperparameters of CNN structures such as learning rate, number of epochs, number of layers and batch size.
- Though high-level features were extracted, additional features are essential to enhance the accuracy of classifying the different DR grades.

2.2 Research contribution

This research focuses on enhancing the accuracy of classifying the different grades of DR severity.

Table 1. List of notations

Notations	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

To achieve this task, a MSA-based ResNetGB model is proposed. First, the RF image database is obtained and fed to the encoder network to create the interpretation tensor. Then, the MSFP is applied to represent high-level features at various scales. Also, an attention strategy is performed to get the attention maps and multiply them with the high-level feature map representations to obtain the final feature interpretation. Further, the global interpretation is created and classified into various classes of DR grades.

3. Materials and methods

In this research work, the ResNetGB model [8] is used as a baseline model under different system settings. This section explains the MSA-ResNetGB model for DR severity categorization briefly. Table 1 presents the notations used in this study.

Pseudo code for the proposed MSA-ResNetGB model:

Input: ROI lesions from the E-Optha dataset and RF images $I_1, \ldots, I_n$ (from Kaggle-APTOS and IDRiD)

Output: DR grade classification (0-No DR, 1-Mild DR, 2-Moderate DR, 3-Severe DR and 4-Proliferative DR)

***Step 1*:** Collect the annotated RF images and split the images into training and test set.

Step 2: In the training set, embed the collected images into the high-level representational space using the ResNetGB encoder; generate the interpretation tensor F_{ec} as:

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

Step 3: Add the attention strategy to the multi-scale feature interpretation;

Step 4: Improve the discriminative ability of high-level feature interpretation; Use the point-wise convolution among the pyramid representations.

Step 5: Generate the attention tensor $\mathcal{A}$ as:

$$\mathcal{A}^{h\times w\times 64} = \left(F_{all}^{h\times w\times 4c} * \mathbb{K}_{pw}\right) \quad (2)$$

Step 6: Obtain the global interpretation by the GMP and create the final feature interpretation vector F as:

$$F^{1\times 1024} = \frac{GMP\left(\sigma\left(F^{h\times w\times c} \odot \mathcal{A}^{h\times w\times 1}\right)\right)}{GMP\left(\mathcal{A}^{h\times w\times 1}\right)} \quad (3)$$

Step 7: Map the feature vectors using the PCA layer to the required outcomes;

Step 8: Train the classification layer using GB classifier and calculate the loss factor $\mathcal{L}\ (\theta,\ \varphi)$ and adjust the training variables.

Step 9: Classify the test images based on DR severity grades using the trained MSA-ResNetGB model.

3.1 ResNetGB as encoder

The initial unit of the model is the encoder module. As shown in Fig. 1, the ResNetGB structure is utilized as the encoder in this model to insert the RF images in the high-level interpretational space. The model efficiency is enhanced with the help of the number of units in the deep learner. On the other hand, there is a challenge in this network called vanishing gradients. So, this challenge is resolved by the short links in the ResNetGB model, i.e. direct paths among the outcome of all layers including the input of the nearby unit. The ResNetGB trains the residuals.

As the ResNetGB is comparatively simple to adopt, the efficiency is improved by incorporating additional units. The initial unit is $7 * 7$ convolutional layer followed by the four different units which contain 3, 8, 12 and 3 residual blocks, correspondingly. The last unit comprises a mean pooling layer. It is observed that the ResNetGB architecture is utilized without an FC layer, i.e. PCA layer as the encoder of the MSA model.

The encoder structure G having an encoder variable θ_1 in this ResNetGB-MSA model considers the RF scan I and creates the interpretation tensor (F_{ec}) given in Eq. (1).

3.2 Multi-scale feature extraction and interpretation

The characteristics are obtained from the series of residual blocks in the encoder module. Such obtained characteristics near the input scans include greater resolution and so contain more data regarding the neighborhood characteristics when the characteristics near the final unit have multiple semantic data. Then, multi-level features such as mid and high-level characteristics are concatenated to use both types of data in the consecutive phases. As such characteristics contain varying spatial resolutions, a scaling method is applied to generate them all of the same size. Next, a set of various features is passed through an atrous convolution to retrieve features with varied scales. Convolutional filters with varying view dimensions are used in the atrous convolution. The network can encode more

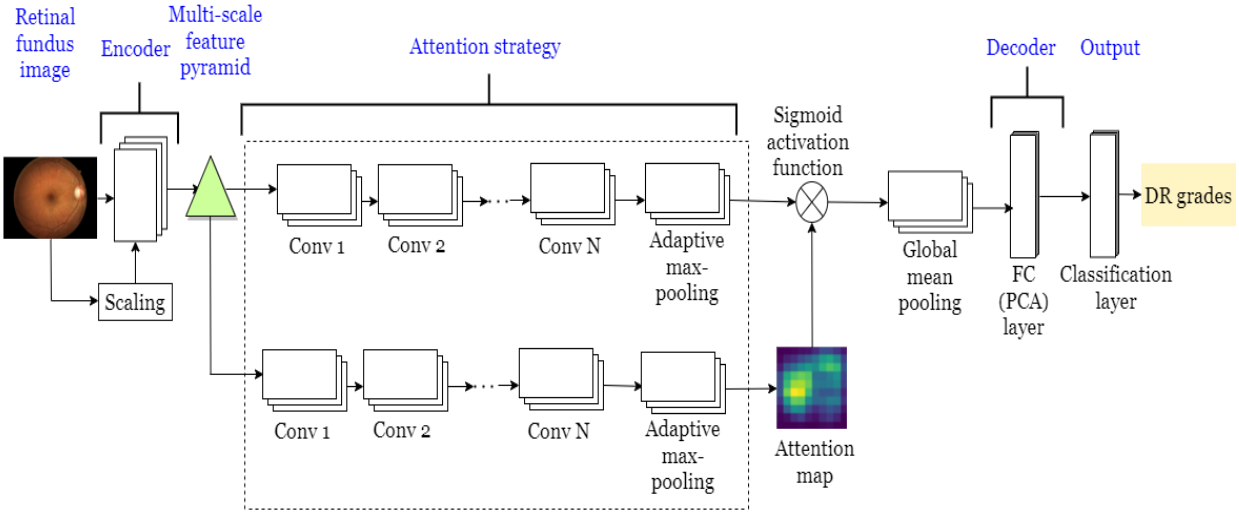

Figure. 1 Multi-scale attention-based mechanism for diabetic retinopathy grade categorization

local details by using a limited view for feature extraction. But, more global details are considered for a larger image context. The resulting multi-scale feature interpretation encapsulates the input scan in a small subspace, allowing it to train DR symptoms of varying sizes, localization and severity.

3.3 MSA strategy

Based on the DR criticality grade, the pattern of the retina is altered. The alteration can induce specific impairment to the RF. To detect such impairments, the high-level feature interpretation is considered which differentiates various classes. However, its efficiency is not satisfactory because of the lack of diabetic patterns. Therefore, the attention strategy is added to the multi-scale feature interpretation to improve the discriminative ability of high-level feature interpretation. The main goal of this attention strategy is to learn where to search for visual loss and adjust the feature space. Particularly, the attention strategy in this MSA-ResNetGB structure focuses on the abnormal regions in the RF images while ignoring the healthy areas.

In this MSA-ResNetGB model, the attention strategy is a sequence of convolution units used for the multi-scale feature interpretation. Initially, the point-wise convolution is used among the pyramid representations to generate a small interpretation as given in Eq. (2). In Eq. (2), $\mathcal{A}$ indicates the attention tensor, F is the interpretation tensor created using the multi-scale unit, $\mathbb{K}_{pw}$ is the point-wise convolution kernel, h, w are the height and width of the feature maps and c is the number of channels. Subsequently, the created small interpretation is applied to the sequence of convolution to produce the attention map $\mathcal{A}^{h\times w\times 1}$ which is multiplied by the high-level interpretation $F^{h\times w\times c}$ to limit the interpretation. To normalize the outcome range from 0 to 1, sigmoid activation (σ) is used. Moreover, the global interpretation is retrieved by the Global Mean Pooling (GMP) of the final interpretation, which is normalized by the GMP data of $\mathcal{A}^{h\times w\times 1}$. So, the created absolute feature interpretation vector F as given in Eq. (3).

In Eq. (3), $\odot$ defines point-wise multiplication. Significantly, the learning variables for the multi-scale and attention units are defined as θ_2. The MSA strategy enhances the training and helps to increase the precision of RF image classification depending on DR criticality grade because it uses the outcomes of the preceding units with different significance. It differentiates this presented framework from the ResNetGB or other CNN structures that do not examine discrepancy.

3.4 Decoder and classification units

The decoder module has the FC layers called PCA layers to map the feature vectors to the required outcomes. The FC units of the pre-learned structure are trained from the ImageNet database to convert the global high-level characteristics, which are not appropriate to the usual and tumor characteristics in the RF scans. Also, the FC units include a massive amount of trainable variables. To solve this problem, all FC units are discarded from the modified structure and included the PCA unit has 153 neurons to minimize the training difficulty. Such 153 neurons are chosen by the greedy method depending on ROIs and the modified ResNet model. Thus, the PCA unit retrieved the characteristics associated with the common and tumor patterns in the RF scans. Those obtained characteristics are then classified by the categorization unit that applies the GB classifier to predict whether the given RF scans belong to healthy people or the DR patients with their severity grades.

The key intention of these units is to categorize the RF images into healthy and DR with severity grades. So, the loss factor $\left(\mathcal{L}(\theta,\varphi)\right)$ is represented as the categorization error for the MSA-ResNetGB structure with encoder and attention variables $(\theta = \theta_1 + \theta_2)$ along with the categorization branch variable φ. The cross-entropy error is implemented in both the estimated and the actual labels. Moreover, the non-learnable weight is added in $\left(\mathcal{L}(\theta,\varphi)\right)$ to reduce the significance of all class errors on the absolute error rates. The error aims to limit the impact of imbalanced scans during the learning phase.

Because the vital goal of automated DR identification is to support the ophthalmologist and decrease the screening complexity, this framework is developed to assist the physician in detecting DR. Also, labeling the normal and abnormal without accurate DR grades is much simpler than proper grading for the physicians. So, this weak labeling is handled conveniently. The secondary process is learned with variables $\left(\mathcal{L}(\theta,\varphi)\right)$ by the cross-entropy error factor. As well, the image augmentation schemes including resizing and flipping are involved in the training phase to prevent overfitting. For this reason, the ROIs around the lesions with various dimensions (16×16, 32×32 and 64×64) are initially extracted, which comprise various context data. After that, those are resized to the equal dimension and all ROIs are rotated in multiple directions $\left[40^{\circ}, 120^{\circ}, 180^{\circ}, 275^{\circ}\right]$ as well as flipped horizontally.

4. Experimental analysis

The MSA-ResNetGB model is evaluated using four distinct measures, including Accuracy, Precision, Recall, and F1-Score, on two different benchmark datasets, Kaggle-APTOS and IDRiD. The system is learned for 120 epochs employing adam optimization with a batch of 3 and a training rate of 10^{-4}.

4.1 Dataset description

Two benchmark datasets Kaggle-APTOS and IDRiD were used in this experimental analysis section to evaluate the proposed model. This subsection further describes the dataset in detail, including the number of collected samples, annotated samples, DR grades and ground truths. The sample RF images from both datasets are shown.

4.1.1. APTOS 2019 blindness detection dataset:

The RF images from the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 Blindness Detection dataset were used in this study [27]. This Kaggle image collection comprises 3662 samples obtained from a diverse range of rural Indian people. Aravind Eye Hospital in India collected and organized the data. However, a panel of medical experts analyzed and classified the collected samples using the International Clinical Diabetic Retinopathy Disease Severity Scale (ICDRSS). The Kaggle-APTOS dataset samples are classified into five groups on the scale of 0-4, No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. The first classification comprises healthy RF samples that do not have DR. Each of the subsequent classifications contains more defective retinas than the previous class. The last classification, proliferative DR, includes samples with vitreous or pre-retinal HEs. Fig. 2 shows the RF samples from each class in Kaggle-APTOS.

4.1.2. Indian diabetic retinopathy image dataset (IDRiD):

IDRiD sub-challenge 2 from the IEEE ISBI - 2018 has been employed in this study [28]. It has 516 images with a range of clinical states of DR and DME, including 413 and 103 training and test images, respectively under Disease grading. IDRiD is the first dataset to represent an Indian population. Fig. 3 shows the RF samples from each class in the IDRiD dataset. Each sample in the IDRiD collection is annotated with Diabetic Retinopathy and Diabetic Macular Edema severity grades at a pixel level. Based on the severity scale, the DR grade is labeled into five classes on the scale of 0 – 4, categories similar to the Kaggle-APTOS dataset.

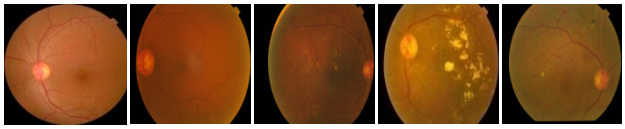

Figure. 2 APTOS 2019 blindness detection dataset samples from each class (0-4)

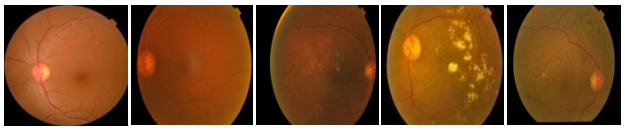

Figure. 3 Indian diabetic retinopathy image dataset (IDRiD) samples from each class (0-4)

4.2 Performance metrics

The performance of the MSA-ResNetGB model is examined with the cutting-edge models using the following measures: Accuracy in Eq. (4), Precision in Eq. (5), Recall in Eq. (6) and F1-Score in Eq. (7). The following are the mathematical formulae utilized to compute the metrics:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Whereas True Positive (TP) is the accurate categorization of the number of samples as positive, True Negative (TN) is the correct classification of the number of samples as negative. Further, False Positive (FP) denotes the ratio of negative class samples categorized as a positive class, while False Negative (FN) denotes the ratio of positive class samples classified as negative class.

4.3 Evaluation

In this section, the proposed MSA-ResNetGB model is validated using the performance measures: Accuracy, Precision, Recall and F1-Score on two public-benchmark datasets. The MSA-ResNetGB model was trained on two different datasets to classify the DR images based on the severity grades.

Table 2. Performance analysis of the proposed model and the cutting-edge models for DR classification on the Kaggle-APTOS 2019 blindness detection dataset

Reference	Year	Classification method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[19]	2019	CNN model	77.00	-	-	-
[9]	2019	Modified Xception model	83.09	-	88.24	-
[17]	2021	Composite gated attention DNN	82.54	82.00	83.00	82.00
[20]	2021	CNN512 model	84.10	-	-	-
[21]	2021	Xception Multitask model	86.00	77.00	70.00	73.00
[22]	2021	Hybrid Inception ResNet-v2	82.18	-	-	-
[23]	2019	ExtraTree model	91.07	90.40	89.54	89.97
[11]	2020	MobileNetV2 model	78.47	68.66	60.01	64.04
[24]	2021	DenseNet121 model	90.50	93.00	90.00	88.47
[16]	2021	Tuned XGBoost model	94.20	94.34	92.68	93.51
[25]	2021	MSA-Net model	84.60	-	91.00	-
Proposed	**2022**	**MSA-ResNetGB model**	**94.40**	**94.52**	**94.40**	**94.42**

In both datasets, multi-class classification is performed as the dataset comprises five classes (0 – 4). Further, APTOS and IDRiD experimental results are discussed in subsequent subsections, which include the performance comparison table and confusion matrices. The class-wise performance evaluation on APTOS and IDRiD are discussed.

4.3.1. APTOS results:

In this subsection, the proposed model has trained on Kaggle-APTOS 2019 Blindness Detection dataset and is evaluated to obtain DR classification results. The Kaggle-APTOS dataset consists of 3662 training samples in which 10% of labeled samples are taken as test data. The multi-class model is validated by comparing it with the latest literature work. Table 2 shows the performance analysis of the proposed model and the cutting-edge models for DR classification on the Kaggle-APTOS 2019 Blindness Detection dataset. The most commonly used metric in the literature is accuracy and some of the other metrics which are not reported in the literature are denoted as '-'.

Table 2 aggregates and compares the recent research work on the Kaggle-APTOS dataset. It is found from the literature that CNN is the most popularly used deep learning method for medical image analysis. In Table 2, eleven distinct classification models from recent literature were utilized to compare the performance of the MSA-ResNetGB model on the Kaggle-APTOS dataset using four metrics: Accuracy, Precision, Recall, and F1-Score. According to Table 2, the proposed method surpassed the cutting-edge models on the Kaggle-APTOS. CNN models such as CNN [19] and CNN512 [20] obtained 77% and 84% accuracy, respectively, which is 17.4% and 10.3% less than the proposed model. The CNN variant models mentioned in the literature [9, 11, 17] and [21-25] performed better but not greater than the proposed model.

Moreover, the proposed model surpassed the Hybrid Inception ResNet-v2 [22] by 12.22%. This proves the proposed model's robustness. The confusion matrix for the proposed MSA-ResNetGB model on the APTOS dataset is shown in Fig. 4. It provides the distribution of samples by class, the ratio of accurately classified and the misclassified samples. In class 0, 70 samples were correctly identified as having no DR, 70 samples were correctly classified in class 1 as having mild DR, 71 samples were correctly classed in class 2 as having moderate DR and 69 samples were correctly classified in class 3 as having severe DR, as well as, 70 samples were correctly classified in class 4 as having PDR.

4.3.2. IDRiD results:

In this subsection, the proposed model is trained on IDRiD and is evaluated to obtain DR classification results. The IDRiD consists of 516 samples in which 20% of labeled samples are taken as test data. The multi-class model is validated by comparing it with the latest literature work.

Table 3 shows the performance analysis of the proposed model and the cutting-edge models for DR classification on the IDRiD. The most commonly used metric in the literature is accuracy and some of the other metrics which are not reported in the literature are denoted as '-'.

Table 3 aggregates and compares the recent research work on the IDRiD. In Table 3, four

Table 3. Performance analysis of the proposed model and the cutting-edge classification models for DR classification on the IDRiD dataset

Reference	Year	Classification method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[10]	2019	AlexNet model	90.07	-	-	-
[12]	2020	CANet model	65.10	-	-	-
[13]	2021	EfficientNet B0 model	86.00	-	-	-
[26]	2019	GNN-based model	79.30	-	-	-
Proposed	**2022**	**MSA-ResNetGB model**	**94.17**	**91.48**	**91.57**	**91.45**

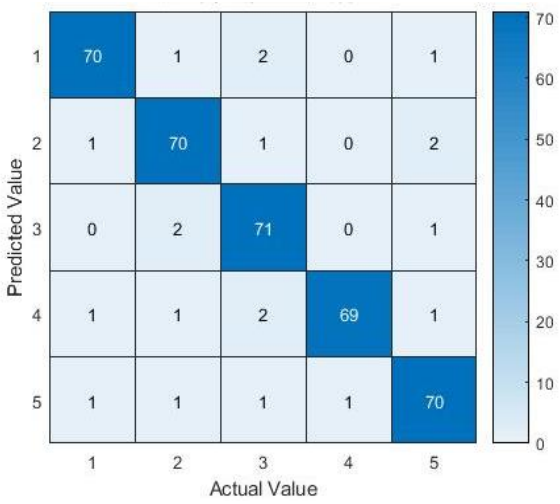

Figure. 4 Confusion matrix for MSA-ResNetGBmodel on the Kaggle-APTOS 2019blindness detection dataset

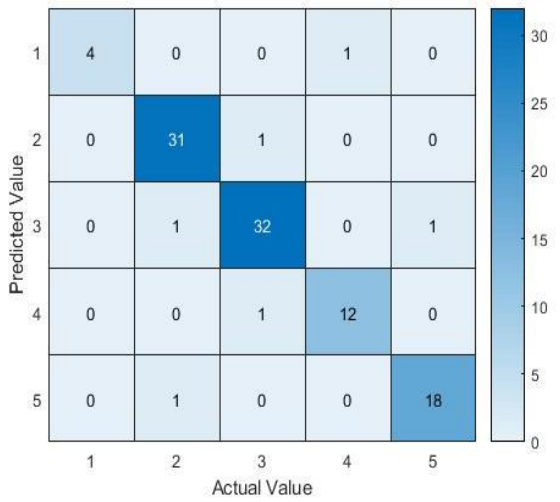

Figure. 5 Confusion matrix for MSA-ResNetGB model on the IDRiD

distinct classification models from recent literature were used to analyze the performance of the MSA-ResNetGB model on the IDRiD using four metrics: Accuracy, Precision, Recall, and F1-Score. According to Table 3, CNN and the pre-trained models [10, 12, 13, 26] surpassed the proposed model on the IDRiD. The proposed model obtained 4.1% greater than AlexNet model [10], 29.07% greater than CANet model [12], 8.17% greater than EffiecientNet B0 model [13] and 14.87% greater than GNN-based model [26]. The confusion matrix for the MSA-ResNetGB model on the IDRiD is shown in Fig. 5.

The matrix shown in Fig. 5 provides the distribution of image samples by class, as well as the ratio of accurately classified and misclassified samples. In class 0, 4 samples were correctly identified as having no DR, 31 samples were correctly classified in class 1 as having mild DR, 32 samples were correctly classed in class 2 as having moderate DR and 12 samples were correctly classified in class 3 as having severe DR, as well as, 18 samples were correctly classified in class 4 as having PDR.

4.4 Discussion

The empirical values of the proposed MSA ResNetGB model on the Kaggle-APTOS and IDRiD datasets are discussed and validated in this section. The model performance on the two datasets was analyzed and evaluated individually with the cutting-edge models in the previous sections. According to the previous section, it is discovered that both benchmark datasets outperformed the literature models. The model learns the DR structure with varied feature selection and localization due to the combination of local and global feature representations, resulting in improved performance. The multi-scale attention strategy, which is implemented in the high-level interpretation space, focuses on the key region in identifying the DR severity levels. Table 4 compares the performance of the MSA-ResNetGB model on the Kaggle-APTOS and IDRiD datasets.

The MSA-ResNetGB model on the Kaggle-APTOS dataset achieved 94.40% accuracy, 94.53% precision, 94.40% recall, and 94.43% F1-Score, which is higher than the performance on the IDRiD dataset, which achieved 94.18% accuracy, 91.48% precision, 91.57% recall, and 91.45% F1-Score.

When compared to the Kaggle-APTOS dataset, the performance loss in IDRiD is mostly due to a lack of image samples. For computer vision and

Table 4. Performance of MSA-ResNetGB model on the Kaggle-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

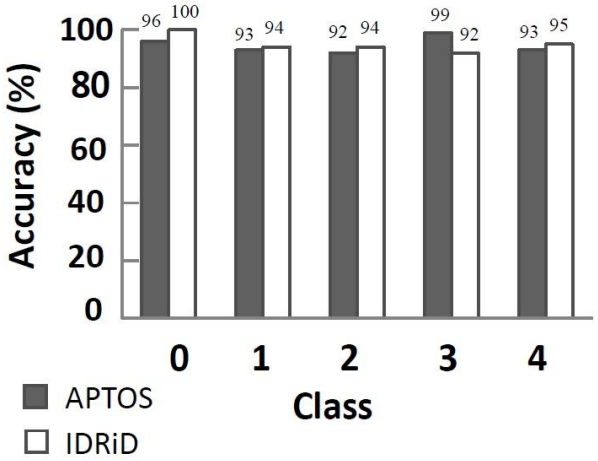

Figure. 6 Class-wise performance evaluation of MSA-ResNetGB model on Kaggle-APTOS and IDRiD dataset

images, data augmentation strategies produce effective outcomes. Fig. 6 represents the class-wise performance evaluation of the MSA-ResNetGB model on the Kaggle-APTOS and IDRiD datasets.

As previously stated, these datasets have five distinct classes labeled from 0 to 4. According to Fig. 6, on the Kaggle-APTOS dataset, 96%, 93%, 92%, 99% and 93% of accuracy were attained in classes 0, 1, 2, 3, and 4, respectively. On the IDRiD dataset, accuracy rates of 100%, 94%, 94%, 92%, and 95% were reached for classes 0, 1, 2, 3, and 4, respectively. The Kaggle-APTOS dataset shows that class 3 performs better while class 2 performs less. In the IDRiD dataset, however, class 0 generates better results while class 3 produces lower results.

5. Conclusion

In this paper, the MSA-ResNetGB model was proposed to detect and classify RF samples based on DR severity grades. Initially, the encoder network was used to integrate the RF picture into the high-level feature interpretation space. The MSFP model was then used to describe the RF pattern at various scales. In addition, the MSA strategy was used on the high-level interpretation to increase the feature interpretation's differentiation efficiency. Further, the whole MSA-ResNetGB structure was trained on the cross-entropy error to correctly characterize the DR criticality grades. Finally, when validated with cutting-edge models to classify DR severity classes, the MSA-ResNetGB test results on Kaggle-APTOS and IDRiD revealed 94.40% and 94.17% accuracy, respectively.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, methodology, software, validation, Valarmathi Srinivasan; formal analysis, investigation, Vijayabhanu Rajagopal; resources, data curation, writing—original draft preparation, Valarmathi Srinivasan; writing—review and editing, Vijayabhanu Rajagopal; visualization; supervision, Vijayabhanu Rajagopal;

References

[1] D. J. Eszes, D. J. Szabo, G. Russell, C. Lengyel, T. Varkonyi, E. Paulik, and B. E. Petrovski, "Diabetic retinopathy screening in patients with diabetes using a handheld fundus camera: the experience from the south-eastern region in Hungary", *Journal of Diabetes Research*, pp. 1-9, 2021.

[2] S. D. Solomon and M. F. Goldberg, "ETDRS grading of diabetic retinopathy: still the gold standard?", *Ophthalmic Research*, Vol. 62, No. 4, pp. 190-195, 2019.

[3] K. S. Sreejini and V. K. Govindan, "Retrieval of pathological retina images using bag of visual words and pLSA model", *Engineering Science and Technology, an International Journal*, Vol. 22, No. 3, pp. 777-785, 2019.

[4] Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, and J. Zhong, "Diagnosis of diabetic retinopathy using deep neural networks", *IEEE Access*, Vol. 7, pp. 3360-3370, 2018.

[5] M. Aatila, M. Lachgar, H. Hrimech, and A. Kartit, "Diabetic retinopathy classification using ResNet50 and VGG-16 pretrained networks", *International Journal of Computer Engineering and Data Science*, Vol. 1, No. 1, pp. 1-7, 2021.

[6] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks", In: *Proc. of Advanced Neural Information Processing Systems*, pp. 1-11, 2017.

[7] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?", In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pp. 2661-2671, 2019.

[8] F. Saeed, M. Hussain, and H. A. Aboalsamh, "Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network", *IEEE Access*, Vol. 9, pp. 41344-41359, 2021.

[9] S. H. Kassani, P. H. Kassani, R. Khazaeinezhad, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Diabetic retinopathy classification using a modified xception architecture", In: *Proc. of the IEEE International Symposium on Signal Processing and Information Technology*, pp. 1-6, 2019.

[10] B. Harangi, J. Toth, A. Baran, and A. Hajdu, "Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features", In: *Proc. of the 41st Annual International Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 2699-2702, 2019.

[11] L. Wang and A. Schaefer, "Diagnosing diabetic retinopathy from images of the eye fundus", *Cs230. Stanford. Edu*, 2020.

[12] X. Li, X. Hu, L. Yu, L. Zhu, C. W. Fu, and P. A. Heng, "CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading", *IEEE Transactions on Medical Imaging*, Vol. 39, pp. 1483-1493, 2020.

[13] E. Abdelmaksoud, S. Barakat, and M. Elmogy, "Diabetic retinopathy grading system based on transfer learning", *International Journal of Advanced Computer Research*, Vol. 11, No. 52, pp. 1-12, 2021.

[14] H. Liu, K. Yue, S. Cheng, C. Pan, J. Sun, and W. Li, "Hybrid model structure for diabetic retinopathy classification", *Journal of Healthcare Engineering*, Vol. 2020, pp. 1-9, 2020.

[15] S. Gayathri, V. P. Gopi, and P. Palanisamy, "A lightweight CNN for diabetic retinopathy classification from fundus images", *Biomedical Signal Processing and Control*, Vol. 62, pp. 1-11, 2020.

[16] N. Sikder, M. Masud, A. K. Bairagi, A. S. M. Arif, A. A. Nahid, and H. A. Alhumyani, "Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images", *Symmetry*, Vol. 13, No. 4, pp. 1-26, 2021.

[17] J. D. Bodapati, N. S Shaik, and V. Naralasetti, "Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification", *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-15, 2021.

[18] H. Tariq, M. Rashid, A. Javed, E. Zafar, S. S. Alotaibi, and M. Y. I. Zia, "Performance analysis of deep-neural-network-based automatic diagnosis of diabetic retinopathy", *Sensors*, Vol. 22, No. 1, pp. 1-15, 2022.

[19] O. Dekhil, A. Naglah, M. Shaban, M. Ghazal, F. Taher, and A. Elbaz, "Deep learning based method for computer aided diagnosis of diabetic retinopathy", In: *Proc. of the IST 2019-IEEE International Conf. on Imaging Systems and Techniques*, pp. 1-4, 2019.

[20] W. L. Alyoubi, M. F. Abulkhair, W. M. Shalash, "Diabetic retinopathy fundus image classification and lesions localization system using deep learning", *Sensors*, Vol. 21, No. 11, pp. 1-22, 2021.

[21] S. Majumder and N. Kehtarnavaz, "Multitasking deep learning model for detection of five stages of diabetic retinopathy", *IEEE Access*, Vol. 9, pp. 123220-123230, 2021.

[22] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning", *Evolution in Computational Intelligence*, pp. 679-689, 2021.

[23] N. Sikder, M. S. Chowdhury, A. S. M. Arif, and A. A. Nahid, "Early blindness detection based on retinal images using ensemble learning", In: *Proc. of the 22nd International Conf. on Computer and Information Technology*, pp. 1-6, 2019.

[24] S. Sheikh and U. Qidwai, "Smartphone-based diabetic retinopathy severity classification using convolution neural networks", In: *Proc. of the Advances in Intelligent Systems and Computing*, Vol. 1252, pp. 469-481, 2021.

[25] M. T. A. Antary and Y. Arafa, "Multi-Scale attention network for diabetic retinopathy classification", *IEEE Access*, Vol. 9, pp. 54190-54200, 2021.

[26] A. Sakaguchi, R. Wu, and S. I. Kamata, "Fundus image classification for diabetic retinopathy using disease severity grading", In: *Proc. of the 9th International Conf. on Biomedical Engineering and Technology*, pp. 190-196, 2019.

[27] APTOS 2019 Blindness Detection. [Online] Av ailable: https://www.kaggle.com/c/aptos2019-bl indness-detection/

[28] P. Prasanna, P. Samiksha, K. Ravi, K. Manesh, D. Girish, S. Vivek, and M. Fabrice, "Indian diabetic retinopathy image dataset (IDRiD)", *IEEE Dataport*, 2018.