[1]Anand M
[2]Dr. Meenakshi Sundaram A

Hybrid Deep Transfer Learning and Feature Fusion Architecture for Diabetic Retinopathy Classification and Severity Grading

Journal of Electrical Systems

Abstract: - Diabetic Retinopathy (DR) is the leading cause of blindness among individuals with diabetes. Automating the diagnosis of DR has the potential to greatly benefit patients by enabling early detection and intervention, thus reducing the risk of blindness. The primary objective of this research is to develop a robust approach for the classification of DR and to analyze its severity grading. By achieving this, we aim to provide an effective tool for accurate diagnosis and assessment of DR, contributing to improved patient care and outcomes. The current literature review analysis reported the importance of deep learning in computer vision based applications. Moreover, plenty of pre-trained models are also present which can be used for classification tasks. Therefore, we present a hybrid DL classification approach by combining Inception V3, VGG-19 and ResNet 50. The proposed architecture uses transfer learning, and feature fusion model to produce the weighted feature vector which is used for classification analysis. The proposed approach is experimented on publicly available datasets APTOS-2019 and Messidor. The performance is measured in terms of accuracy, precision, recall and F1-score.

Keywords:- Diabetic Retinopathy, Deep Learning, Computer Vision, Classification, Severity Grading

1. Introduction

Human health is a complex and multifaceted issue that affects every aspect of our lives. From physical well-being to mental health and emotional wellness, our health is crucial to our ability to function and thrive in the world around us. Unfortunately, there are a multitude of factors that can negatively impact our health which leads to several diseases including chronic and non-chronic diseases. Non-chronic diseases are usually acute, meaning they have a sudden onset and a short duration. These diseases can be caused by viruses, bacteria, or other pathogens e.g. Influenza, flu etc. In contrast, Chronic diseases are long-term medical conditions that can be managed but not cured e.g. Heart disease, Arthritis, Diabetes etc.

Recently, diabetes mellitus is identified as on one of the rapidly increasing chronic disease and it has reported serious health issues globally. Over the last twenty years, there has been a concerning rise in the number of individuals impacted by diabetes. The IDF Diabetes Atlas [1] reports that approximately 500 million people worldwide, spanning all age groups, have received a diagnosis of diabetes. However, projections suggest that this number is expected to rise significantly to more than 780 million by 2045. This disease has several diverse impacts on human organs including heart, liver, kidney eyes etc. Moreover, for individuals under 50 years of age, diabetes is the leading cause of blindness. The individuals who are affected by the diabetes, have higher risk of developing the diabetic retinopathy [2].

DR) is an advanced disorder characterized by changes in the small blood vessels of the retina, leading to retinal ischemia, increased retinal permeability, the formation of abnormal new blood vessels, and macular edema [3]. If left untreated, DR can result in severe visual impairment [4]. In developed countries, DR is the primary cause of blindness among the working-age population [5], imposing a significant economic burden on society, particularly on healthcare structures [6]. Generally, DR is characterized a complex consequence of diabetes mellitus where the blood vessels connected to eye are blocked dye to glucose. This blockage leads to swelling and leakage of blood and other fluids from eye which can lead to severe eye injury. Given that appropriate management of DR can prevent more than 90% of cases of visual loss, it is crucial to accurately categorize, classify, and stage the severity of DR to facilitate appropriate therapeutic procedures. According to Mamtora et al. [7] DR accounts for approximately 2.6% of all cases of blindness globally. High haemoglobin A1c, and high blood pressure are the most known reasons associated with the increase of DR.

Therefore, regular screening is highly recommended for diabetic patients to ensure that the early detection of DR in its initial stage. This screening includes examination of retinal imaging to analyze the shape and appearance of various lesions that includes Microaneurysms (MA), Haemorrhages (HM), soft and hard exudates (EX) [8, 9]. Microaneurysms are small, localized bulges that occur in the tiny blood vessels called capillaries. They can appear as small red dots or dots with tiny white centers on the retina. However, they may not always cause noticeable symptoms in the early stages. The dots are characterized by distinct boundaries, exhibiting sharp margins that do not exceed a size of 125 micrometers. Haemorrhages, also known as bleeding or blood vessel ruptures, occur when blood escapes from the circulatory system and accumulates in surrounding tissues.

[1]*Corresponding author: Research Scholar, School of Computer Science & Engineering, REVA University, Bengaluru
Email: r21pcs01@reva.edu.in
[2]Author 2 Affiliation: Associate Professor, School of Computer Science & Engineering, REVA University,Bengaluru.
Email: meenakshi.sa@reva.edu.in

HMs are identified through the observation of sizable spots on the retina, exhibiting irregular margins that exceed 125 micrometers in size. Hard exudates are a manifestation of retinal changes characterized by the presence of yellow or white deposits in the retinal tissue. These deposits typically have a hardened or waxy consistency and can be observed during a retinal examination. Soft exudates, also known as cotton wool spots or retinal nerve fibre layer infarcts, are white or greyish patches that appear on the retina. They are characterized by their soft and fluffy appearance. Soft exudates are typically seen as localized areas of swelling or accumulation of fluid in the retinal tissue.

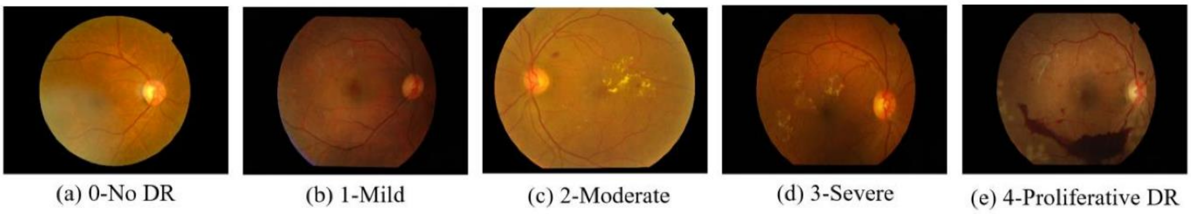

Fig .1. Sample images of different severity grading

Based on these lesions, the diabetic retinopathy detection is classified into five different stages which are no DR, mild DR, moderate DR, severe DR, and proliferative DR. Figure 1 shows sample images of these cases. Therefore, early detection plays important role to prevent the diverse impacts of DR. The manual detection of DR require high skilled practitioners to carry out the assessments. Moreover, high skilled profession ophthalmologists face severe challenges due to inconsistences due to inter and intra grades. Therefore, researchers have suggested to adopt the automated DR image analysis methods to overcome these issues. Currently, several methods have been presented to analyze, segment and classify the biomedical images such as threshold based, region based, texture feature, color feature analysis etc. moreover, traditional machine learning classification schemes such as SVM, RF, DT etc are also used widely. These algorithms face several challenges therefore, deep learning based schemes have gained huge attention in various computer vision based applications such as in [11] AlexNet architecture is used for DR analysis. Farrag et al [16] used DenseNet169, and Liu et al. [18] used combination of several architectures including, EfficientNetB5, and Xception etc. These methods have reported significant performance in DR classification. However, limited works have been carried out for DR severity grading and accuracy also remain a challenging issue in this context. Motivated by the performance of Deep learning methods, we focus on developing a new deep learning architecture for DR image classification as DR and No-DR along with its severity grading. The main contributions of this approach are listed below:

- To develop a combined hybrid deep learning approach for deep feature extraction by using well-known pre-trained DL architectures
- To incorporate deep transfer learning to fully utilize the features captured by pre-trained model
- To present a feature fusion model to produce the weighted feature vector.

Rest of the manuscript is organized in following sections: section II describes the recent works in this domain of DR classification, section III elaborates proposed hybrid deep learning classification approach, section IV presents comparative analysis of proposed approach ad section V presents the concluding remarks of the proposed approach.

2. literature survey

This section presents the brief overview of existing methods of DR classification. As discussed before, the main aim of this approach is to develop the deep learning approach therefore we discuss most recent deep learning based methods here.

Gayathri et al. [10] presented CNN based lightweight classification method where CNN is employed for extracting the significant features from fundus images. Further, these features are trained by using various machine learning classifiers such as SVM, RF, AdaBoost, NB and J48 where J48 classifier outperformed. However, several pre-trained models are also presented by various research communities based on the standards of deep learning such as AlexNet, GoogleNet and ResNet50. These models have been used in various applications and reported significant classification performance. Caicho et al. [11] employed these models to classify the five different stages of DR. The experimental analysis shows the significant performance of AlexNet by achieving the classification accuracy as 93.56%. Yaqoob et al. [12] adopted ResNet model to extract the deep features and employed random forest classifier for DR detection. Martinez-Murcia et al. [13] used the concept of transfer learning. This approach is mainly based on the deep residual convolutional neural network which is used to extract the discriminatory features without any complex transformation to enhance the image quality. Further, transfer learning model is also applied to reuse the deep learning layers.

Currently, attention based mechanisms also have gained huge attention which computes the attention weights of input and output sequence. The obtained weights are used further to compute the weighted sum which is known as attention context vector. This vector captures the important information from input to output. Al-Antary et al. [14] used attention mechanism and reported the importance of high-level attributes. This model extracts mid and

high level attributes by using encoder network. Additionally, a multi-scale feature pyramid is incorporated to capture the retinal structure across different regions. Moreover, a multi-scale attention mechanism is applied to the high-level representation to further boost its discriminative capability. This architecture uses cross entropy loss function. Alahmadi et al. [15] also reported the issue of lack of semantic dependency and suggested to incorporate attention mechanism to identify the important region of given image resulting in boosting the performance of the model. According to this approach, the input data is passed through the encoder module where high-level and semantic features are encoded. Additionally, the representation space is decomposed as texture, semantic and contextual features. The texture attention module employs high pass filtering to highlight the information. After attention mechanism, the feature fusion model is applied to combined the features. Later, decoder module performs classification task. Farag et al. [16] used single Color Fundus photograph (CFP) to develop the deep learning based DR grading system. This system uses DenseNet169 encoder module to generate the visual embedding of input. Further, Convolutional Block Attention Module (CBAM) is also presented on the encoder side to improve its discriminative power. Li et al. [17] introduced lesion-attention pyramid network which is used in integrating different resolutions to produce multi-scale features. The low-resolution network computes the lesion activation map and leads the high-resolution network to emphasize on lesion regions.
Liu et al. [18] focused on improving the DR classification performance. Therefore, authors introduced a hybrid deep learning architecture which uses EfficientNetB4, EfficientNetB5, NASNetLarge, Xception, and InceptionResNetV2 network architectures as base models which are trained with the help of cross-entropy and enhance cross entropy loss function. Sikder et al. [19] presented a novel ensemble machine learning method by using gray level intensity and texture feature extraction methods. This model achieved accuracy of 94.23%. Riaz et al. [20] presented a combination of densely connected and deep learning approach to classify retinal images. The dense connections between convolution layers facilitate deep supervision between layers which helps to improve the accuracy of the system. Gadekallu [21] reported that traditional methods neglected the aspects of data pre-processing and dimensionality reduction which can lead to biased results. Therefore, to overcome this issue, authors introduced principal component analysis for feature selection and firefly approach for dimension reduction. Finally, the obtained dataset is trained by using deep neural networks. Das et al. [22] reported that the existing methods use handcrafted features but achieving the desired accuracy is a challenging task due to uncertainty of features. Therefore, authors adopted CNN based deep learning approach and included genetic algorithm for optimizing the CNN parameters. Finally, SVM classifier is used to perform the classification.

3. Proposed Model

This section presents the proposed deep learning based model for diabetic retinopathy classification along with its severity grading. Both tasks belong to the classification problem where DR classification is treated as binary classification and severity grading is considered as multiclass classification problem. In literature review, we have discussed several deep learning based framework for biomedical image processing, segmentation and classification. Several works have been carried out in this domain of image classification where pretrained deep learning architectures have been considered a major breakthrough to mitigate the issue of poor classification. Therefore, in this work, we adopt the pretrained models for deep feature extraction. However, these architectures are trained by using different models therefore we incorporate transfer learning mechanism to improve the classification accuracy.
In other words, instead of training a neural network from scratch, a model that has already been trained on a large dataset is used as a starting point. This saves time and resources, as the initial layers of the network have already learned how to identify basic features such as edges and corners. The pretrained deep learning model is then fine-tuned on a smaller dataset of images specifically related to diabetic retinopathy. This involves adjusting the weights of the existing layers in the network to better fit the new dataset. By doing so, the model is able to learn and identify unique features associated with diabetic retinopathy that it may not have encountered during its initial training. Below given figure 2 depicts the overall architecture of proposed deep learning model for DR classification.

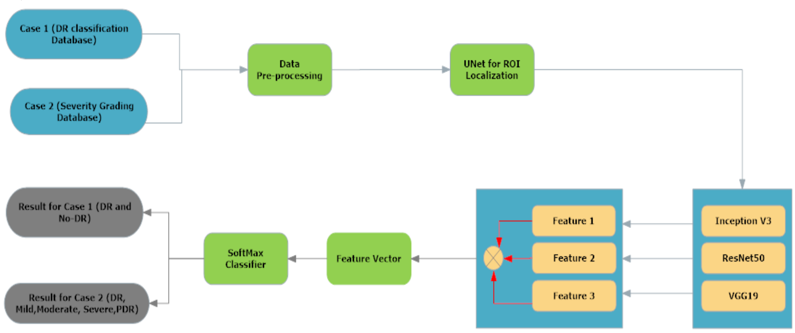

Fig. 2. Proposed hybrid deep transfer learning and feature fusion architecture for DR classification and severity grading.

According to this approach, the proposed model uses data pre-processing, ROI extraction, combination of Inception, VGG-19 model and ResNet for deep feature extraction. Moreover, a transfer learning model is also incorporated to capture the entire feature information. In next stage, a feature fusion model is added to formulate the final feature vector. Finally, softmax layer is added to obtain the classification results.

Each block of proposed architecture is described below:

Input dataset: in this work we have considered retinal fundus image as input which are divided into two cases such as case 1 and case 2. Case 1 denotes the input image set corresponding to binary classification whereas case 2 denotes the input image set corresponding to different severity grading.

Data pre-processing: this stage plays important role in computer vision and CAD based applications. generally, the input images vary in size therefore standardization is required to process all images. The size of exudate patches vary from 25×25 to 286×487 resolution which affects the performance of retinal image analysis in automated systems. Therefore, we consider smallest patch size and divide all images in 25×25 patches.

Region of Interest localization: Exudates are characterized as vivid abnormalities that appear as radiant patches and spots in cases of diabetic retinopathy. These lesions stand out prominently in the yellow color plane of the color fundus image, exhibiting complete contrast. Prior to performing feature extraction through region of interest (ROI) localization, exudate segmentation is employed. Several methods have been used for segmentation such as edge based, ROU based, fuzzy logic, neural network etc. However, deep learning based UNet segmentation model is widely adopted for segmentation and ROI extraction. To improve the segmentation accuracy, we adopted channel and attention mechanism based Deep UNet architecture [cite p2].

Pre-trained CNN Models: classification problem is widely studied and several promising solutions have been presented. Therefore, we use well-known pre-trained deep learning models in this work. This model uses Inception, ResNet and VGGNet architectures. These architectures are described below:

- **Inception-V3:**

The Inception v3 network employs a deep architecture with multiple layers of convolutional and pooling operations. It is designed to capture complex patterns and hierarchical representations in images by utilizing a combination of 1x1, 3x3, and 5x5 convolutions. One notable feature of the Inception v3 architecture is the use of inception modules, which are composed of multiple parallel convolutional layers with different filter sizes. To reduce the computational complexity and improve efficiency, Inception v3 also incorporates dimensionality reduction techniques, such as 1x1 convolutions, which help reduce the number of input channels while preserving valuable information. Additionally, the network employs auxiliary classifiers at intermediate layers to encourage the learning of more discriminative features. Below given figure 3 depicts the complete architecture of Inception v3 model.

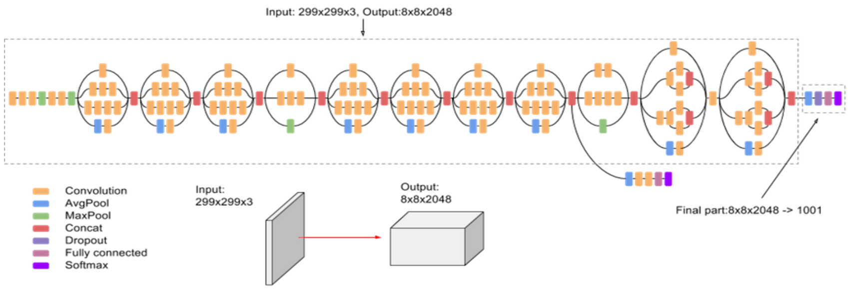

Fig. 3. Inception v3

The inception v3 uses factorized convolution operations which helps to minimize the total number of parameters used in network and it also impacts on network accuracy. in this network architecture, the bigger convolution operations are replaced by smaller convolutions thus its operational speed is increased.

ResNet 50: ResNet-50 consists of 50 layers, including convolutional layers, pooling layers, fully connected layers, and skip connections, which are the key innovation of the ResNet architecture. The skip connections, also known as residual connections, allow for the flow of information from earlier layers directly to later layers. This helps alleviate the degradation problem, where deeper networks tend to perform worse than shallower networks. By introducing skip connections, ResNet-50 enables the network to learn residual mappings and effectively retain information from previous layers, leading to improved gradient flow and easier training of deep networks. ResNet-50 utilizes a building block called the "residual block." Each residual block consists of multiple convolutional layers with shortcut connections. These shortcut connections add the original input of the block to the output of the block, allowing the network to learn the residual information Below given 4 figure depicts the architecture of resnet50.

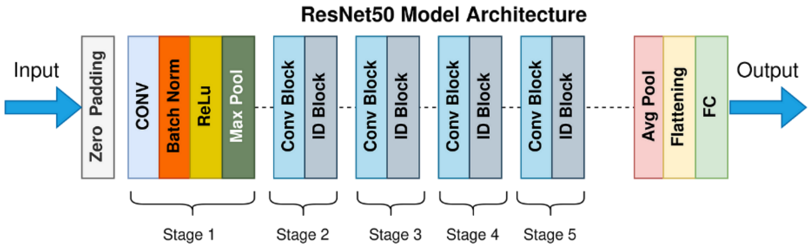

Fig 4. ResNet 50 architecture

Generally, the increase in depth leads to saturate the accuracy of deep learning and further increase can degrade the overall accuracy. This issue is called as "degradation problem". In order to overcome this issue, the ResNet module uses residual mapping mechanism. Rather than relying on the expectation that a few consecutive layers will perfectly learn the desired underlying mapping, the Residual Network takes a different approach. It explicitly allows these layers to learn a residual mapping, effectively capturing the difference between the desired output and the current representation. Figure 5 shows the structure of residual mapping block.

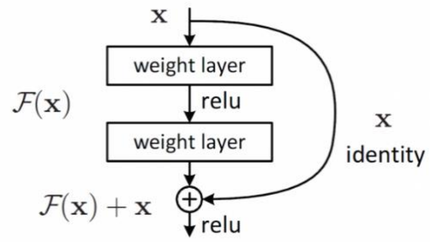

Fig. 5. Residual mapping

VGG-19: The VGG19 network is characterized by its simplicity and uniformity. It consists of a series of convolutional layers, each using a 3x3 filter size, followed by max pooling layers for downsampling. It consists of total 19 layers, including convolutional, pooling, and fully connected layers. The architecture of VGG19 follows a repetitive pattern, which makes it easy to understand and implement. It is known for its deep stack of convolutional layers, which allows it to capture complex visual features in images. The use of small filter sizes helps to preserve fine-grained details. Below given figure 6 shows the architecture of VGG 19.

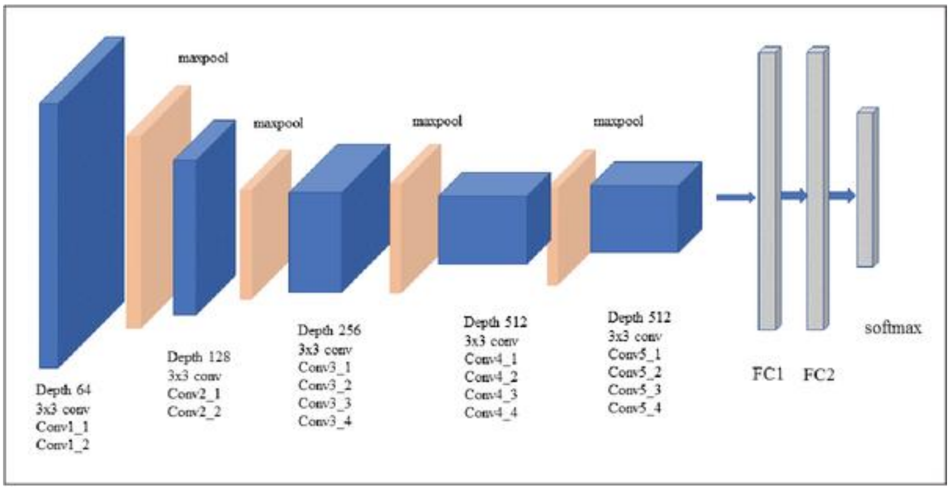

Fig 6. VGG 19 architecture

The process of VGG 19 is as follows: It considers a fixed input size of (MxN) RGB image thus the input shape becomes MXNX3. In next stage it performs pre-processing where mean RGB is subtracted from each pixel. In order to cover the entire notion of image, it uses 3x3 kernel with stride size of 1 pixel and uses spatial padding to preserve the spatial resolution of the image. Similarly, max-pooling operation is performed over 2x2 pixel window with stride 2. This operation is followed by Rectified linear unit (ReLU) to incorporate the non-linearity in training process. Finally, we incorporate the transfer learning module which is used to learn the contextual information to solve the problem and employ this process to solve another related problem. The transfer learning approach involves training a network specifically for a particular task using a related dataset. Once this initial training is completed, the network is then fine-tuned or further trained using the target dataset for the objective task [23].

In next stage, we focus on feature fusion model by using global average pooling and convolution operations. Below given figure 7 shows the architecture of feature fusion module.

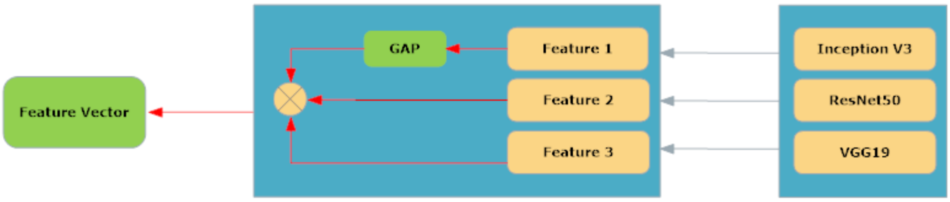

Fig 7. Feature fusion architecture

According to this module, the feature weights are computed as follows:

$$F_W = \frac{F_a}{Conv\left(\frac{1}{H \times W}\sum_H \sum_W F_a\right)} \tag{1}$$

Where F_a denotes the output of pre-trained networks which is subjected to global average pooling operation with a size of 1x1 to mitigate the issue of overfitting. The $Conv\left(\frac{1}{H\times W}\sum_H \sum_W F_a\right)$ operation helps to obtain the redistributed weights. This operation is performed on the all three pre-trained models and final feature weights for model are obtained as F_W^0, F_W^1,and F_W^2. Further, these feature weights are concatenated by using softmax and final feature weights are obtained as $W^0, W^1$and W^2. The weights can be expressed as:

$$W^i = \frac{F_w^i}{\sum_j F_w^i} \tag{2}$$

The final fusion vector can be obtained as:

$$O = W^0.F^0 + W^1.F^1 + W^2.F^2 \tag{3}$$

Where F^i denotes the convolution layer and W^i denotes the weights. The obtained feature vectors are processed through the softmax to obtain the classification result.

During the training process, we use dropout of 0.3 to mitigate the issue of overfitting and it also helps in reducing computational complexity. The DR dataset suffers from a highly imbalanced distribution, where the number of fundus images without disease far exceeds those with DR. Consequently, traditional loss functions like cross-entropy struggle to effectively differentiate between categories due to the equal weight assigned to each sample. As a result, they yield unsatisfactory results when applied to DR images. To address this issue and enhance training speed while minimizing accuracy loss, we opt for the adoption of Focal Loss as our loss function. Focal Loss is expressed as follows:

$$\mathcal{L}(\hat{y}, y) = -\sum_{i=1}^{n=2\ or\ 5} y_i \alpha (1-\hat{y}_i)^{\gamma}.\log \hat{y}_i \tag{4}$$

Where α is a weighting factor, γ is the penalty factor, $\hat{y}$ and y denotes the predicted result and actual labels of input.

4. Results And Discussion

4.1. Performance Measurement parameters

The outcome of proposed DL based DR classification is measured with the help of confusion matrix. Below given table 1 shows the basic representation of confusion matrix in this case of DR classification.

Table. 1. Confusion matrix

	DR	No DR
DR	TP	FN
No DR	FP	TN

True positive (TP): it shows that the classifier correctly predicts the positive class from the given test set.

True Negative: it shows that the classifier model correctly predicts the negative class from the given test set.

The true negative and true positive values show the accuracy of classifier. However, these categories should match the actual values of TP and TN.

False positive: denotes the classifier model incorrectly predicts the positive class.

False negative: denotes that the classifier mistakenly predicted the negative class.

This confusion matrix helps to evaluate the performance by computing precision, recall, F1-Score. The symbol for accuracy stands for the rate of accurate classification. It is calculated by dividing the total number of predictions by the percentage of true predictions. It may be computed as follows:

$$Acc = \frac{\text{TruePositive} + \text{TrueNegative}}{TP + TN + FP + FN} \tag{5}$$

Similarly, the Recall performance also can be computed with the help of true negative and false positive values. This can be computed as follows:

$$Recall = \frac{TruePositive}{TruePositive + TN} \tag{6}$$

Then, we measure the performance in terms of Precision which is computed based on true positive, and false positive. It is expressed as:

$$P = \frac{TruePositive}{TruePositive + FP} \quad (7)$$

Finally, we compute F-measure which is computed based on sensitivity and precision i.e. it is the mean of precision and sensitivity performance. It is computed as:

$$F = \frac{2 * P * Sensitivity}{P + Sensitivity} \quad (8)$$

Based on these parameters, we compute the performance of proposed approach and compared with other supervised classifiers.

4.2. Severity Grading Classification

4.2.1. Performance analysis for APTOS-2019 dataset

This section presents the outcome of proposed deep learning based model. The proposed deep learning architecture is trained using APTOS-2019 dataset. This dataset contains five different classes such classes which are characterized as severity grading such as no-DR (Class 1), Mild DR (Class 2), Moderate DR (Class 3), Severe DR (Class 4) and proliferative DR (Class 5) which contains total 1625,333, 899, 174 and 265 images, respectively. For testing, 180, 37, 100, 19, and 30 images are allocated to no-DR, mild, moderate, severe, and proliferative DR cases, respectively. The classification performance is evaluated with the help of confusion matrix. Below given table depicts the obtained confusion matrix for existing CNN, hybrid CNN-SVD and proposed hybrid ensemble deep learning model. Below given table 2 demonstrates the confusion matrix obtained by applying CNN classifier.

Table 2. Severity grading confusion matrix by using CNN classifier

	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	179	0	0	0	1
Class 2	0	35	2	0	0
Class 3	1	0	98	1	0
Class 4	0	0	2	17	0
Class 5	0	0	0	0	30

According to this confusion matrix the total number of true positive cases are 358. Based on this confusion matrix, we measure performance of proposed approach in terms of Accuracy, Precision, Recall, and F1-score. Below given table 3 shows the obtained performance of CNN classifier for severity grading classification.

Table 3. Statistical performance for CNN classifier

Class	Accuracy	Precision	Recall	F1-score
Class 1	99.18%	0.99	0.98	0.99
Class 2	99.18%	0.95	0.97	0.96
Class 3	98.36%	0.98	0.96	0.97
Class 4	99.18%	0.89	0.94	0.92
Class 5	99.73%	1.00	0.97	0.98

Similarly, below given table 4 shows the confusion matrix of CNN-SVD classifier for severity grading classification.

Table 4. Statistical performance for CNN-SVD classifier

	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	179	1	0	0	0
Class 2	0	35	2	0	0
Class 3	2	0	96	1	1
Class 4	0	0	2	17	0
Class 5	1	0	0	0	30

The confusion matrix reported that the total number of true positive samples are 359 which improves accuracy when compared with traditional CNN classifier. Further, we use this confusion matrix to measure other statistical parameters as mentioned in below given table 5.

Table 5. Statistical performance for CNN-SVD classifier

Class	Accuracy	Precision	Recall	F1-score
No DR	98.92%	0.99	0.98	0.99
Mild	99.19%	0.95	0.97	0.96
Moderate	97.83%	0.96	0.96	0.96
Severe	99.19%	0.89	0.94	0.92
PDR	99.46%	0.97	0.97	0.97

Finally, we evaluate the performance of proposed model. The obtained confusion matrix is presented in below given table 6.

Table 6. Confusion matrix by using proposed model

	Class 1	Class 2	Class 3	Class 4	Class 5
No DR	179	0	0	0	0
Mild	0	35	1	0	0
Moderate	1	0	99	0	0
Severe	0	0	2	18	0
PDR	0	0	0	0	30

This confusion matrix reported the total true positive entities as 361 which achieves the average accuracy of 99.18%. Based on this confusion matrix, we measure other performance evaluation parameters as mentioned in below given table 7.

Table 7. Statistical performance of proposed model

Class	Accuracy	Precision	Recall	F1-score
Class 1	99.37%	1.00	0.99	1.00
Class 2	99.73%	0.97	1.00	0.99
Class 3	99.18%	0.99	0.98	0.99
Class 4	99.73%	0.95	1.00	0.97
Class 5	100%	1.00	1.00	1.00

4.2.2. Performance Analysis for MESSIDOR-2 Dataset

This section describes the outcome of existing and proposed approach for MESSIDOR-2 dataset. In this experiment, the number of no-DR, mild, moderate, severe and proliferative DR are considered as 203, 54, 69, 15, and 7, respectively. Below given table 8 shows the confusion matrix obtained by employing CNN based scheme.

Table 8. Confusion matrix by using CNN for MESSIDOR-2 dataset

	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	179	0	0	0	0
Class 2	0	35	1	0	0
Class 3	1	0	99	0	0
Class 4	0	0	2	18	0
Class 5	0	0	0	0	30

Further, we use this confusion matrix to evaluate other performance metrics such as Accuracy, Precision, Recall, and F1-score. The measured performance is presented in below given table 9.

Table 9. Statistical performance of CNN for Messidor-2 dataset

Class	Accuracy	Precision	Recall	F1-score
Class 1	95.40%	0.97	0.95	0.96
Class 2	97.41%	0.93	0.91	0.92
Class 3	96.55%	0.88	0.94	0.91
Class 4	100%	1.00	1.00	1.00
Class 5	99.71%	0.86	1.00	0.92

Further, we considered the confusion matrix obtained by CNN-SVD model as presented in below given table 10. According to this confusion matrix, the total number of true positive entities is obtained as 332.

Table 10. Confusion matrix by using CNN-SVD for MESSIDOR-2 dataset

	Class 1	Class 2	Class 3	Class 4	Class 5
No DR	200	3	0	0	0
Mild	2	50	3	0	0
Moderate	8	0	61	0	0
Severe	0	0	2	15	0
PDR	1	0	0	0	6

With the help of this confusion matrix, we measured statistical performance in terms of accuracy, precision, recall and F1-score. The obtained performance is reported in table 11.

Table 11. Statistical performance of CNN-SVD for Messidor-2 dataset

Class	Accuracy	Precision	Recall	F1-score
Class 1	95.40%	0.97	0.95	0.96
Class 2	97.41%	0.93	0.91	0.92
Class 3	96.55%	0.88	0.94	0.91
Class 4	100%	1.00	1.00	1.00
Class 5	99.71%	0.86	1.00	0.92

Finally, we measured the performance of proposed approach based on the confusion matrix which is presented in below given table 12.

Table 12. Confusion matrix by using proposed method for MESSIDOR-2 dataset

	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	209	0	0	0	0
Class 2	2	53	2	0	0
Class 3	8	0	64	0	0
Class 4	0	0	2	15	0
Class 5	1	0	0	0	6

According to this experiment, the average accuracy of proposed model is obtained as 97.47%. Moreover, we also measure the other evaluation parameters as mentioned in below given table 13.

Table 13. Statistical performance of proposed method for Messidor-2 dataset

Class	Accuracy	Precision	Recall	F1-score
Class 1	98.03%	1.00	0.97	0.98
Class 2	99.44%	0.96	1.00	0.98
Class 3	99.75%	0.91	0.97	0.94
Class 4	100%	1.00	1.00	1.00
Class 5	99.72%	0.86	1.00	0.92

This experiment shows the proposed approach achieves average accuracy as 97.47%

4.3. Comparative Analysis

This subsection describes the classification performance of CNN, CNN-SVD and proposed deep learning classifier for diabetic retinopathy as diabetic or non-diabetic i.e. this performance doesn't predict the severity grading. First of all, we consider the APTOS-2019 dataset where 1625 images belong to no- DR and 1671 belong to DR category. The performance of these binary classifier models can be measured with the help of confusion matrix. Below given table shows the comparative analysis.

Table 14. Comparison performance of proposed method with existing classifiers

Authors	Classes	Dataset	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)
Binary Class (APTOS Dataset)						
[24]	2	APTOS-2019	97	97	97.41	-
[31]	2	APTOS-2019	98	98	97.82	98
Hybrid CNN-SVD	2	APTOS-2019	99	99	99.32	99.50
Proposed Model	2	APTOS-2019	99.15	99.21	99.40	99.60
Severity Grading (Messidor Dataset)						

[25]	5	Messidor -2	-	92	-	88
[26]	5	Messidor -2	-	94	-	90
[27]	5	Messidor -2	-	93	-	94
Hybrid CNN-SVD	5	Messidor -2	98	94	96.26	98.24
Proposed Model	5	Messidor -2	98.50	95.10	97.25	98.50
Severity Grading (APTOS Dataset)						
[24]	5	APTOS-2019	80	81	81.70	-
[28]	5	APTOS-2019	91.37	-	86.34	-
[31]	5	APTOS-2019	82	83	82.54	79
[29]	5	APTOS-2019	76	77	77.90	-
[30]	5	APTOS-2019	87	88.24	83.09	91.80
[32]	5	APTOS-2019	89	-	89	97.90
[33]	5	APTOS-2019	94.34	92.69	94.20	-
Hybrid CNN-SVD	5	APTOS-2019	95	94	96.61	98.58
Proposed Model	5	APTOS-2019	96.20	95.80	97.50	99.10

The comparative analysis shows the importance of proposed model for binary classification and severity grading for different datasets. The proposed approach shows significant improvement in the overall classification performance.

5. Conclusion

In this work, we have focused on development of a deep learning based architecture for DR classification. Several methods have been introduced in literature but traditional feature extraction and classification methods fail to achieve the robust performance. Currently, deep learning based approaches are utilized in these tasks. Therefore, we consider a combination of pre-trained deep learning models and presented a transfer learning framework along with feature fusion to improve the classification accuracy.

Compliance and Ethical Standards

Funding: No funding received for this research work.

***Conflict of interest*:** The authors declare there is no conflict of Interest

6. Acknowledgement

This endeavour would not have been possible without the generous support from REVA University, Bengaluru. Words cannot express my gratitude to my Research Supervisor Dr. MeenakshiSundaram A, Associate Professor, School of Computer Science & Engineering, REVA University, Bengaluru and chair of my committee for her invaluable patience and feedback. I also could not have undertaken this journey without my defense committee, who generously provided knowledge and expertise.

REFERENCES

[1] IDF Diabetes Atlas 9th Edition. Available online: https://diabetesatlas.org/atlas/ninth-edition

[2] Ting, D. S. W., Cheung, G. C. M., & Wong, T. Y. (2016). Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. *Clinical & experimental ophthalmology*, *44*(4), 260-277.

[3] Shukla, A. K., Pandey, R. K., &Pachori, R. B. (2020). A fractional filter based efficient algorithm for retinal blood vessel segmentation. Biomedical Signal Processing and Control, 59, 101883.

[4] Mahmoud, M. H., Alamery, S., Fouad, H., Altinawi, A., & Youssef, A. E. (2021). An automatic detection system of diabetic retinopathy using a hybrid inductive machine learning algorithm. *Personal and Ubiquitous Computing*, 1-15.

[5] Kropp, M., Golubnitschaja, O., Mazurakova, A., Koklesova, L., Sargheini, N., Vo, T. T. K. S., ...&Thumann, G. (2023). Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—risks and mitigation. *EPMA Journal*, *14*(1), 21-42.

[6] Udaondo, P., Parravano, M., Vujosevic, S., Zur, D., & Chakravarthy, U. (2022). Update on current and future management for diabetic maculopathy. Ophthalmology and therapy, 11(2), 489-502.

[7] S. Mamtora, Y. Wong, D. Bell, and T. Sandinha, ‘‘Bilateral birdshot retinochoroiditis and retinal astrocytoma,’’ Case Rep. Ophthalmolog. Med., vol. 2017, pp. 1–4, Feb. 2017

[8] Sebastian, A., Elharrouss, O., Al-Maadeed, S., &Almaadeed, N. (2023). A Survey on Diabetic Retinopathy Lesion Detection and Segmentation. *Applied Sciences*, *13*(8), 5111.

[9] Saleh, R. W., & Abdullah, H. N. (2022, April). Smart Healthcare System for Early Diabetic Retinopathy Detection. In *2022 2nd International Conference on Computing and Machine Intelligence (ICMI)* (pp. 1-5). IEEE.

[10] Gayathri, S., Gopi, V. P., & Palanisamy, P. (2020). A lightweight CNN for Diabetic Retinopathy classification from fundus images. Biomedical Signal Processing and Control, 62, 102115.

[11] Caicho, J., Chuya-Sumba, C., Jara, N., Salum, G. M., Tirado-Espín, A., Villalba-Meneses, G., ... & Almeida-Galárraga, D. A. (2022, March). Diabetic retinopathy: detection and classification using AlexNet, GoogleNet and

ResNet50 convolutional neural networks. In Smart Technologies, Systems and Applications: Second International Conference, SmartTech-IC 2021, Quito, Ecuador, December 1–3, 2021, Revised Selected Papers (pp. 259-271). Cham: Springer International Publishing.

[12] Yaqoob, M. K., Ali, S. F., Bilal, M., Hanif, M. S., & Al-Saggaf, U. M. (2021). ResNet based deep features and random forest classifier for diabetic retinopathy detection. Sensors, 21(11), 3883.

[13] Martinez-Murcia, F. J., Ortiz, A., Ramírez, J., Górriz, J. M., & Cruz, R. (2021). Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy. Neurocomputing, 452, 424-434.

[14] Al-Antary, M. T., &Arafa, Y. (2021). Multi-scale attention network for diabetic retinopathy classification. *IEEE Access*, *9*, 54190-54200.

[15] Alahmadi, M. D. (2022). Texture attention network for diabetic retinopathy classification. *IEEE Access*, *10*, 55522-55532.

[16] Farag, M. M., Fouad, M., & Abdel-Hamid, A. T. (2022). Automatic severity classification of diabetic retinopathy based on denseNet and convolutional block attention module. IEEE Access, 10, 38299-38308.

[17] Li, X., Jiang, Y., Zhang, J., Li, M., Luo, H., & Yin, S. (2022). Lesion-attention pyramid network for diabetic retinopathy grading. Artificial Intelligence in Medicine, 126, 102259.

[18] Liu, H., Yue, K., Cheng, S., Pan, C., Sun, J., & Li, W. (2020). Hybrid model structure for diabetic retinopathy classification. Journal of Healthcare Engineering, 2020.

[19] Sikder, N., Masud, M., Bairagi, A. K., Arif, A. S. M., Nahid, A. A., &Alhumyani, H. A. (2021). Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images. Symmetry, 13(4), 670.

[20] Riaz, H., Park, J., Choi, H., Kim, H., & Kim, J. (2020). Deep and densely connected networks for classification of diabetic retinopathy. Diagnostics, 10(1), 24.

[21] Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P. K. R., Ra, I. H., &Alazab, M. (2020). Early detection of diabetic retinopathy using PCA-firefly based deep learning model. Electronics, 9(2), 274.

[22] Das, S., & Saha, S. K. (2022). Diabetic retinopathy detection and classification using CNN tuned by genetic algorithm. Multimedia Tools and Applications, 81(6), 8007-8020.

[23] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ...& Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, *35*(5), 1285-1298.

[24] V. Naralasetti, S. N. Shareef, and S. Hakak, ‘‘Blended multi-modal deep convnet features for diabetic retinopathy severity prediction,’’ Electronics, vol. 9, no. 6, p. 914, 2020.

[25] C. I. Sánchez, M. Niemeijer, A. V. Dumitrescu, M. S. Suttorp-Schulten, M. D. Abràmoff, and B. van Ginneken, ‘‘Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data,’’ Investigative Ophthalmol. Vis. Sci., vol. 52, no. 7, pp. 4866–4871, Jun. 2011.

[26] L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet, and J. M. P. Langlois, ‘‘Red lesion detection using dynamic shape features for diabetic retinopathy screening,’’ IEEE Trans. Med. Imag., vol. 35, no. 4, pp. 1116–1126, Apr. 2015.

[27] R. Gargeya and T. Leng, ‘‘Automated identification of diabetic retinopathy using deep learning,’’ Ophthalmology, vol. 124, no. 7, pp. 962–969, 2017.

[28] [53] H. Liu, K. Yue, S. Cheng, C. Pan, J. Sun, and W. Li, ‘‘Hybrid model structure for diabetic retinopathy classification,’’ J. Healthcare Eng., vol. 2020, pp. 1–9, Oct. 2020.

[29] V. Dondeti, J. D. Bodapati, S. N. Shareef, and N. Veeranjaneyulu, ‘‘Deep convolution features in non-linear embedding space for fundus image classification,’’ Rev. Intell. Artif., vol. 34, no. 3, pp. 307–313, 2020.

[30] 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. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT), Dec. 2019, pp. 1–6.

[31] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, ‘‘Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification,’’ J. Ambient Intell. Hum. Comput., vol. 12, pp. 1–15, Jan. 2021.

[32] W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, ‘‘Diabetic retinopathy fundus image classification and lesions localization system using deep learning,’’ Sensors, vol. 21, no. 11, p. 3704, May 2021.

[33] 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, p. 670, Apr. 2021.

6.