International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

A 2D-Structured Dilation Based Hierarchical CNN for the Detection of Diabetic Retinopathy Grade Levels

[1]Mrs. Sharmila EMN and [2]Dr.Prof. R. Suchitra
[1]Research Scholar and [2]Professor & Research Guide, Chirashree Institute of Research and Development & RPA First Grade College, (Approved by University of Mysore)

[1]emnsharmila@gmail.com and [2]suchithra.suriya@gmail.com

Abstract: *This paper introduces a diabetic retinopathy severity grading approach thatuses a 2D-structured dilation-based hierarchical convolutional neural network (CNN). In this approach, the pre-processed fundus image is utilized to segment the regions such as the optic disc, blood vessels, and lesion regions. The optic disc, blood vessel, and lesion regions combine to form the region of interest. The proposed 2D-structured dilation-based Hierarchical CNN (2D-SDHCNN) has a parallel section of Lstages that use different dilated masks in a hierarchical structure. The dilation is also applied to the convolutional filters of each subnetwork and the region that corresponds to the dilated mask of the global feature is also utilized in the hierarchical network. The hierarchical network can able to extract deep features near the vessels and lesion candidates. Datasets such as Kaggle APTOS and Messidor-2 are utilized for evaluating the suggested 2D-SDHCNN approach. The suggested approach performance highly depends on the dilation factor used in the 2D-SDHCNN. The 2D-SDHCNN approach yields a precision, Mathews correlation coefficient (MCC), and accuracy of 97.61%, 97.03%, and 97.73% respectively when evaluated using the Kaggle APTOS dataset. Also, the suggested scheme when evaluated utilizing Messidor-2 provides precision, MCC, and accuracy of 93.30%,93.42%, and 95.39% respectively.*

Keywords: *Diabetic retinopathy, Severity grading, Convolutional neural network, dilation, Blood vessels*

I. INTRODUCTION

Due to the inability to secrete sufficient blood insulin by the pancreas the glucose level in blood rises leading to a condition named diabetes mellitus [1]. World Health Organization (WHO) reports that the number of diabetes incidents cases will be around 700 million by the year 2045 [2]. This increase in diabetes incidence cases will be much higher since the number of diabetes incidents in the year2014 is 422 million. The diabetes diseases can cause diabetic retinopathy (DR) where the retinal capillaries get blocked and start bleeding. As a result of bleeding new blood vessels start to grow that leads to vision impairment. For individuals under the age of 50, diabetes is the common cause of blindness [3]. Early diagnosis and identification of DR severity are essential to avoid the complications caused by DR. The ophthalmologists examine the retinal images based on the appearance and structure of lesion regions. This lesion region may be hard exudates, soft exudates, hemorrhages, and microaneurysms [4]. The yellowish-white deposit caused due to protein and lipid leakage in the retina forms the hard exudates. The retinal nerve fiber layer infracts and forms fluffy white cotton wool spots known as soft exudates. The lesion haemorrhages have irregular margin size which is the bleeding vessels. The lesion microaneurysms is caused due to the dilation of blood vessels that create red coloredsmall round spots.

Based on this formation of lesions and new blood vessels, the DR severity grades can be non-proliferative grades (mild, moderate, and severe),and proliferative grades. The non-proliferative (NPDR) stage is characterized by the formation of exudates,hemorrhages, and microaneurysms, while the proliferative (PDR) stage is characterized by the formation of new abnormal blood vessels.The abnormality in fundus image and optical coherence tomography (OCT) images [5] are commonly detected using deep learning approaches [6] that perform the descriptor extraction process and classification process. The descriptors that are obtained from various layers of the Xception structure are aggregated to construct multi-level descriptors [7]. These descriptors are classified using multi-layer perceptron to detect the severity level. This Xception-based DR classification attains an accuracy of 83.09% using the Kaggle APTOS dataset which is higher than the traditional Xception classifier.

Different lesion candidate descriptors based on statistics, intensity, and shape are used [8] to identify the lesion candidates. The lesion candidates are classified using a hybrid classifier formed with the Gaussian mixture model and m-Mediods models. The severity levels are classified as stages 1, 2, 3, and healthy with the use of modified architecture [9] that uses ReLU and soft-max activations.

Antary et al. [10] extracted both high and mid-level descriptors by embedding the descriptors in a high-level representation which also improves the power of differentiating the lesion severity levels. This approach applies the attention process to the descriptors that are obtained at different scales. The capsule layer is utilized by Kalyani et al. [11] which classifies the descriptors extracted by the primary capsule network. The usage of two capsule networks increases the complexity of this approach. Machine learning models such as AdaBoost and support vector machine (SVM) were used to classify the CNN features extracted from the fundus picture [12]. Maximal principal curvature [13] was utilized to detect the blood vessel branches that also use the Hessian matrix. The squeeze and excitation process was incorporated with the CNN to differentiate the abnormal and normal fundus images. Features related to contextual, semantic, and textures are extracted using the encoder part of the deep learning structure [14]. The classification was performed by the decoder part after passing through the attention and fusion process. The fusion process combines contextual, semantic, and texture descriptors. The authors Mussarat et al. [15] used geometric and statistical descriptors after enhancing using a Gabor filter. This approach can segment exudates region leaving the non-exudates to another class.

Models such as ResNeXt, and DeneNet101 are combined [16] to detect the DR images. This approach uses a split-transform combine approach along with a tacking layer to handle the descriptor. The use of DenseNet provides higher performance than the ReNeXt due to the use of concatenation operation in the dense block. The authors Wafaa et al. [17] used YOLOV3 and CNN512 to classify the severity level. The actual result was obtained by fusing the results of the two models. A coarse-to-fine classification approach was proposed [18] by modifying the CNN structure. In this approach, the background features are suppressed while the lesion-related features are enhanced by the attention module. The fine network classifies the classes of DR, while the coarse network classifies the fundus image as DRand No-DR classes. The local and global descriptors in fundus images are collected using a multi-path CNN [19], where the features are classified using the random forest and SVM models. A synergic deep learning approach [20] was used to collect and classify the descriptors from the lesion regions which are segmented using a histogram-based approach. The CNN model was utilized to derive the active deep learning approach [21] where the patches that have more information are detected from which the classification was done using the active deep learning (active-DL) approach. The complexity of the active DL approach is higher than the CNN model.

Circular Hough transform with fuzzy rule is used as a pre-processing [22]. From the pre-processed picture retinal localization is performed from which the descriptors are collected. This approach can classify the abnormality such as maculopathy and diabetic retinopathy. A multi-resolution-based attention [23] was proposed by Sandeep et al. that has depth-wise filters. The multi-resolution images are generated at different dilation rates. The authors report that the use of multi-resolution-based attention can collect subtle descriptors. Finally, an SVM classifier is utilized to classify the multi-resolution descriptors. The vascular structures are extracted [24] using an architecture that is derived from the U-Net. Instead of using 4 down and up-sampling process, 2 down and up-sampling are used in the encoder and decoder sections. A regularized random walker-based approach was used to improve the structural connectivity of the blood vessels that are broken. Though the U-Net-based scheme provides reasonable performance, the complexity is higher.

The methods that are discussed in the literature do not use a dilation-based approach that hasan increasing area near the region of interest for extracting the descriptors. Therefore, the suggested scheme introduces a 2D structured-based dilation whose dilation factor differsbetween the subnetwork and within the subnetwork that collects deep descriptors near the region of interest. The contributory work of the suggested scheme is as follows:

(i) The work extracts two different features namely the hierarchical descriptor and the global descriptor where the hierarchical descriptor is utilized for actual classification. The proposed approach uses the blood vessels, Optic disc, and lesion as the region of interest.

(ii) The approach uses the maximal principal curvature for detecting the blood vessels, and the Circular Hough transform for detecting the Optic disc. The lesion regions are detected using thresholding and morphological operations.

(iii) The hierarchical descriptor extraction process uses a 2D structured dilation, where dilation is performed based on two categories namely the inter-subnetwork dilation and intra-subnetwork dilation.

(iv) The inter-subnetwork and intra-subnetwork dilation helps to extract deep descriptors from the lesion region and the region that surrounds it. An embedding layer is used that combines the features from different channels including the global features and preceding layer features.

(v) Finally, for evaluating the suggested 2D-SDHCNN approach the datasets namely Kaggle APTOS and Messidor-2 are utilized.

The paper has subsequent sections. Section 2 enumerates the working of the suggested2D-SDHCNN-based DR severity grading approach, while Section 3 provides a brief analysis of the suggested scheme in terms of the classification evaluation measures used in evaluating the deep learning approaches. Lastly, the paper was concluded with the key findings obtained during the analysis of the suggested 2D-SDHCNN-based grading approach.

2. PROPOSED METHOD

The 2D-SDHCNN-based DR severity classification approach includes the major process namely preprocessing, region of interest estimation, and 2D-SDHCNN structure for extracting the descriptor and classifying the severity levels as illustrated in Fig .1.

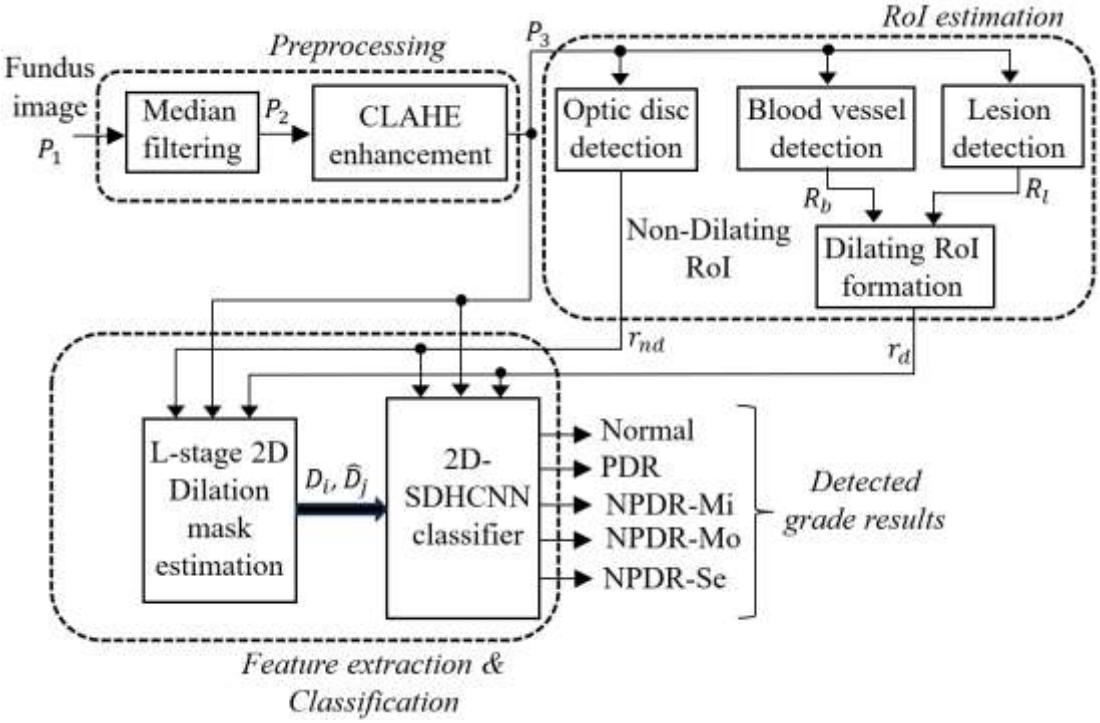

Fig. 1: Diagrammatic representation of suggested 2D-SDHCNN-based DR grading approach

(a) Preprocessing

The fundus image$P_1(x, y)$is initially preprocessed by median filtering and contrast limited adaptive

histogram equalization (CLAHE) image enhancement algorithm [25]. The median filtering removes the noisy pixels, while the CLAHE improves the appearance of the lesion and blood vessels suitable for segmentation. Let the median filtered outputestimated for the picture input P_1 be represented as P_2. The CLAHE-based algorithm is applied to the H,S,V channels of the image after converting the RGB image $P_2$to HSV form. The CLAHE-applied HSV channels are again converted back to RGB to obtain the pre-processed picture P_3.

(b) Region of interest (RoI) estimation

The proposed approach uses regions such as optic disc, blood vessels, and lesion regions as the regions of interest since these three regions play a crucial role in categorizing the DR grade levels.

(i) Blood vessel detection

The approach uses a maximal principal curvature-based blood vessel detection approach [26] to detect the blood vessels. Let $P_{3,G}(x,y)$ represent the green channel of the image $P_3(x,y)$. The green channel is utilized to detect the blood vessels since the retinal vessels are more prominent in this channel than the blue and red channels. The Gaussian filter isthen applied to $P_{3,G}(x,y)$ to reduce the fine noisy details that perform smoothing operations. Thus, the filtered image is expressed as,

$$\hat{P}_{3,G}(x,y) = \frac{1}{2\pi e^2} exp\left(\frac{-x^2+y^2}{2e^2}\right) * P_{3,G}(x,y) \quad (1)$$

Here, $*$ resembles the convolution operator and e is the factor to control the smoothing which resembles the Gaussian kernels standard deviation. The blood vessels have higher intensity variation than other regions which is represented by the Hessian matrix as,

$$H(x,y) = \begin{bmatrix} \frac{\partial^2 \hat{P}_{3,G}(x,y)}{\partial x^2} & \frac{\partial^2 \hat{P}_{3,G}(x,y)}{\partial x \partial y} \\ \frac{\partial^2 \hat{P}_{3,G}(x,y)}{\partial x \partial y} & \frac{\partial^2 \hat{P}_{3,G}(x,y)}{\partial y^2} \end{bmatrix} \quad (2)$$

The highest eigenvalue computed on the Hessian matrix gives the maximal principal curvature expressed as,

$$\hat{P}_H(x,y) = e_{max}\big(H(x,y)\big) \quad (3)$$

Where $e_{max}(.)$ resembles the highest eigenvalue computed on the matrix $H(x,y)$. The contrast of the maximal principal curvature is then enhanced using the relation,

$$\hat{P}_E(x,y) = \frac{\hat{P}_H(x,y) - min\big(\hat{P}_H(x,y)\big)}{max\big(\hat{P}_H(x,y)\big) - min\big(\hat{P}_H(x,y)\big)} \times (a_{max} - a_{min}) + a_{min} \quad (4)$$

Here a_{min} and a_{max} resembles the lower and upper thresholds.The intensity-based thresholding (ISO data thresholding) is then used to detect the blood vessels as,

$$R_b(x,y) = \begin{cases} 1 & \hat{P}_E(x,y) > \delta_1 \\ 0 & \hat{P}_E(x,y) \le \delta_1 \end{cases} \quad (5)$$

Here δ_1 resembles the optimal threshold to segment the blood vessel. Let R_b resembles the segmented blood vessels.

(ii) Optic disc and lesion detection

Circular Hough transform-based optic disc detection [27] approach is utilized to detect the optic disc. The edges are initially detected after converting $P_3(x,y)$ to gray-scale, and from each edge point, the potential circle centers are estimated. The points that havethe highest rating are considered as the optic disc region which is then segmented. The segmented optic disc region is represented as r_{nd}.The lesion region is estimated by the process such as thresholding, and morphological operations. The thresholding segments all possible lesion regions using the relation,

$$r_l = \begin{cases} 1 & P_3(x,y) < \delta_2 \\ 0 & P_3(x,y) \geq \delta_2 \end{cases} \quad (6)$$

Where δ_2 resembles the threshold. The regions whose area is less than 50 pixels are eliminated which contains the noisy region. From the resulting segmented regions dilation is performed and the regions whose area is less than 80 pixels are again eliminated to obtain the lesion-segmented imageR_l. The regions R_b and R_l constitute the dilating region of interest while the optic disc segmented region r_{nd} constitute the non-dilating region.

(c) 2D-SDHCNN classifier

The suggested 2D-SDHCNN approach extracts two different types of descriptors namely the hierarchical descriptor and the global descriptors. The global descriptors are not directly used in the classification process. However, these features are used to performfeature embedding in the hierarchical feature extraction process. Instead of using the fundus image, the approach also uses a dilated version of the region of interest. The region of interest is constructed using the segmented blood vessel region R_b, lesion region R_l and the optic disc region r_{nd}. The optic disc region is not involved in dilation, while the segmented blood vessel region r_b and the lesion region r_l undergo the dilation process. There are three mask images used in each stage represented as D_i, where $i = 1,2\ and\ 3$. Let α_1 be the dilation factor used in each section of layers (intra-sub-network), while α_2 be the dilation factor used in each stage (inter-sub-network). The initial RoI can be estimated as

$$D = r_{nd} \cup R_b \cup R_l \quad (7)$$

Here $\cup$ resembles the set union operator. The above expression can also be expressed as $D = r_{nd} \cup r_d$, since the dilating region of interest can be estimated $r_d = R_b \cup R_l$. Let the structuring element to perform dilation can be represented as Ω. The equation for dilation between the sections (intra-subnetwork)$i-1$ and i in the dilating region can be expressed as

$$r_d^{(i)} = r_d^{(i-1)} \oplus \alpha_1\Omega = \{q|(\alpha_1\Omega)_q \cap r_d^{(i-1)} \neq 0\} \quad (8)$$

Here α_1 resembles the dilation factor used in different filter sections. Also, $\alpha_1\Omega$ resembles the structuring element Ω scaled by α_1 and $(\alpha_1\Omega)_q$ resembles the translation of the scaled structuring element $\alpha_1\Omega$ centered at pixel q. Thus, the dilated RoI in the section i can be estimated as,

$$D_i = S\left(r_d^{(i)}, s\right) \cup S(r_{nd}, s/i) \quad (9)$$

Here $S(r_{nd}, s/i)$ resembles the scaling operation performed on the non-dilated RoI r_{nd} with scaling factor s/i. Similarly,$S\left(r_d^{(i)}, s\right)$ resembles the scaling operation performed on the dilated RoI $r_d^{(i)}$ with scaling factor s.The scaling factor used is $s = 0.5$ and the structuring element used is 'disk' type. Since there arethree embedded layers, i ranges between 1 to 3 as provided in Fig. 2.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

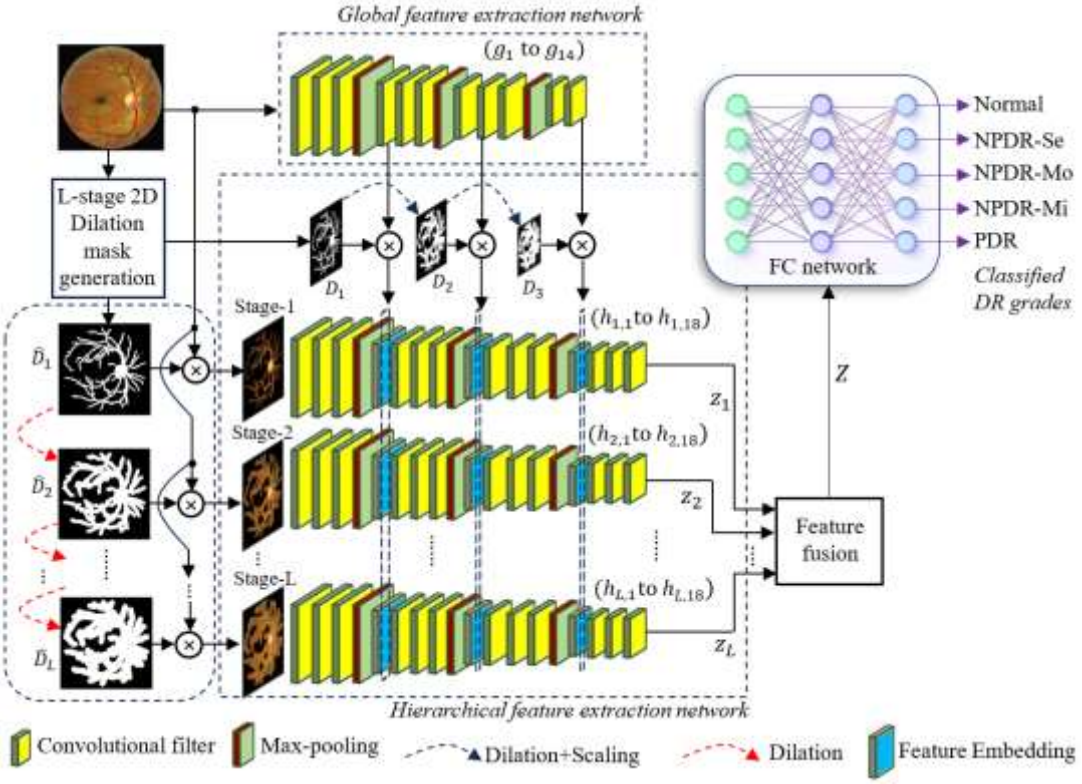

Fig. 2: Architecture of the 2D-SDHCNN in classifying the DR grade levels

The equation for dilation between stages$j-1$ and j in the dilating region (inter-subnetwork dilation) can be expressed as

$$\hat{r}_d^{(j)} = \hat{r}_d^{(j-1)} \oplus \alpha_2 \Omega = \{p | (\alpha_2 \Omega)_p \cap \hat{r}_d^{(j-1)} \neq 0\} \quad (10)$$

Here Ω and α_2 resembles the structuring element and the dilation factor respectively. Also,$\alpha_2 \Omega$ resembles the structuring element Ω scaled by α_2 and$(\alpha_2 \Omega)_p$ resembles the translation of the scaled structuring element $\alpha_2 \Omega$ centered at pixel p. Thus, the dilated RoI in the stage $j$can be estimated as

$$\widehat{D}_j = \hat{r}_d^{(j-1)} \cup r_{nd} \quad (11)$$

Here $j = 1,2 \ldots . L$. For stage 1, $\hat{r}_d^{(j-1)} = \hat{r}_d^{(0)} = r_d$.The input to each stage can be expressed as $P_3 \& \widehat{D}_j$. Here,& resembles the logical and operator. The 2D-SDHCNN architecture has four sections of convolutional filters with 3 max-pooling functions in the global feature extraction network. Each stage of the hierarchical network also has 4 sections of convolutional filters, 3 max-poolinglayers, and 3-feature embedding layers. The feature embedding layer in jth stage will combine the descriptors from the global feature map f having c channels and the features from first jthstage having c_j channels. The embedding layer will use $c/2$ channels from f that have higher energy component and $c_i/2^j$ channel each from first jth channels as depicted in Fig. 3.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

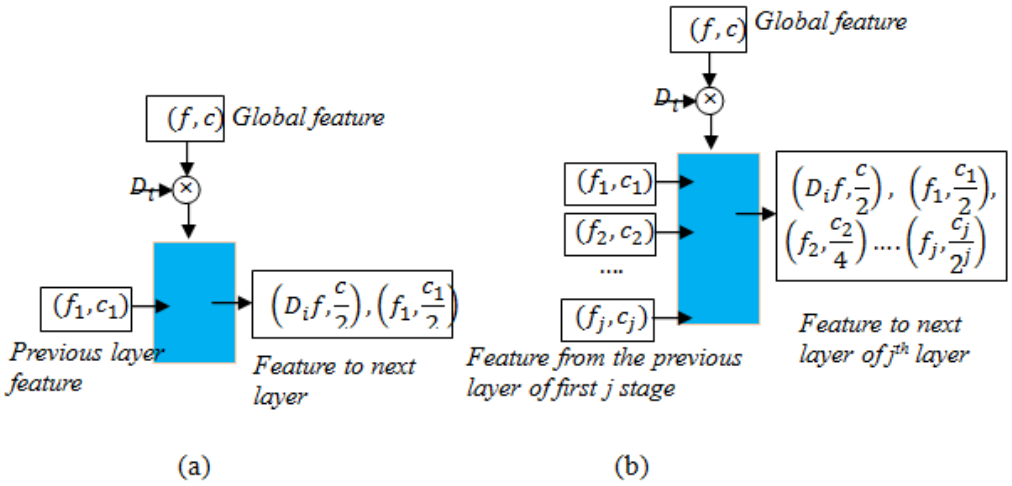

Fig. 3: Function of embedding layer (a) embedding layer in stage-1 (b) embedding layer in stage-j

From$c$ number of channelsof the global feature maps f, $c/2$ number of channels are selected based on higher energy. For performing this operation, the feature maps f from c channels are initially multiplied using the mask D_i. The average energy is estimated on the resulting feature map $D_i f$ using the relation

$$E(f) = \frac{1}{\tau_{D_i}} \sum_{(x,y)\in\{D_i=1\}} |D_i f(x,y)|^2 \qquad (12)$$

τ_{D_i}resembles the number of pixels in the mask D_i having logic '1'. Thus the$c/2$channels having higher energy can be estimated as

$$\underset{(f,\frac{c}{2})}{\operatorname{argmax}}\{E(f)\} = \underset{(f,\frac{c}{2})}{\operatorname{argmax}}\left\{\frac{1}{\tau_{D_i}} \sum_{(x,y)\in\{D_i=1\}} |D_i f(x,y)|^2\right\} \qquad (13)$$

The same process is repeated to select $c_j/2^j$ channels from the feature maps f_j having c_j channel. Thus the$c_j/2^j$channels having higher energy can be estimated as

$$\underset{(f_j,\frac{c_j}{2^j})}{\operatorname{argmax}}\{E(f_j)\} = \underset{(f_j,\frac{c_j}{2^j})}{\operatorname{argmax}}\left\{\frac{1}{\tau_{D_i}} \sum_{(x,y)\in\{D_i=1\}} |f_j(x,y)|^2\right\} \qquad (14)$$

The output of the embedding layer has the feature maps from different channels as illustrated in Fig. 3. Let $\{z_1, z_2 \ldots \ldots z_L\}$ be the features obtained by the L stage in the hierarchical network. These features are combined using the fusion process and the combined feature can be estimated as $Z = [z_1, z_2 \ldots \ldots z_L]$. The feature Z is flattened and used as the input to the fully connected network. The description of layers used in the 2D-SDHCNN approach is provided in Table I.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

Table I: Description of layers used in the 2D-SDHCNN approach

Layer	Description	Output size
g_1- g_3	Conv-5 × 5	256 × 256, 16
g_4	max-pool	128 × 128
g_5-g_7	Conv-5 × 5	128 × 128, 32
g_8	max-pool	64 × 64
g_9-g_{11}	Conv-5 × 5	64 × 64, 64
g_{12}	max-pool	32 × 32
g_{13}-g_{14}	Conv-5 × 5	32 × 32, 32
$\bar{D}_j$	mask	256 × 256
D_1	mask	128 × 128
D_2	mask	64 × 64
D_3	mask	32 × 32

Layer	Description	Output size
$h_{j,1}$-$h_{j,3}$	Conv-3× 3	256 × 256, 32
$h_{j,4}$	max-pool	128 × 128
$h_{j,5}$	Embedding layer	128 × 128, $(16 + 32/2^j)$
$h_{j,6}$-$h_{j,8}$	Conv-3× 3	128 × 128, 64
$h_{j,9}$	max-pool	64 × 64
$h_{j,10}$	Embedding layer	64 × 64, $(32 + 64/2^j)$
$h_{j,11}$-$h_{j,13}$	Conv-3× 3	64 × 64, 128
$h_{j,14}$	max-pool	32 × 32, 128
$h_{j,15}$	Embedding layer	32 × 32, $(64 + 128/2^j)$
$h_{j,16}$-$h_{j,18}$	Conv-3× 3	32 × 32, 64

The max-pooling in the suggested 2D-SDHCNN uses a kernel of size 2×2. For updating of weights and bias in the convolutional filter and the fully connected (FC) layer, the cross entropy-based loss function is used. The output layer of the FC network has 5 outputs that can able to estimate the predicted probability for the DR grades namely normal, PDR, NPDR-Se, NPDR-Mo, and NPDR-Mi.

3. EXPERIMENTAL RESULTS

Fundus images from the Messidor-2 [28] and Kaggle APTOS [29] datasets are utilized to evaluate the performance of the 2D-SDHCNN-based DR severity classification approach. Five image categories namely Normal images, NPDR-severe (NPDR-Se), NPDR-moderate (NPDR-Mo), NPDR mild (NPDR-Mi), and PDR from the two datasets are utilized to evaluate the grading performance of 2D-SDHCNN. The evaluation scales viz. accuracy (Acc), precision (Pre), Mathew's correlation coefficient (MCC), specificity (Spe), recall (Rec), and F1-score (F1)are employed to estimate the grading performance of the 2D-SDHCNN model which can be estimated using,

$$Acc\ (\%) = \frac{\mu_{tn}+\mu_{tp}}{\mu_{fp}+\mu_{tp}+\mu_{tn}++\mu_{fn}} \times 100 \quad (15)$$

$$Pre\ (\%) = \frac{\mu_{tp}}{\mu_{fp}+\mu_{tp}} \times 100 \quad (16)$$

$$MCC\ (\%) = \frac{\mu_{tp}\times\mu_{tn}-\mu_{fn}\times\mu_{fp}}{\sqrt{(\mu_{tn}+\mu_{fp})\times(\mu_{tp}+\mu_{fp})\times(\mu_{tn}+\mu_{fn})\times(\mu_{tp}+\mu_{fn})}} \times 100 \quad (17)$$

$$Spe\ (\%) = \frac{\mu_{tn}}{\mu_{fp}+\mu_{tn}} \times 100 \quad (18)$$

$$Rec\ (\%) = \frac{\mu_{tp}}{\mu_{tp}+\mu_{fn}} \times 100 \quad (19)$$

$$F1(\%) = \frac{\mu_{tp}}{(\mu_{fp}+\mu_{fn})\times\frac{1}{2}+\mu_{tp}} \times 100 \quad (20)$$

Here μ_{tp}, μ_{fn}, μ_{fp} and μ_{tp} resembles the true positive, false negative, false positive, and true positives attained during the 2D-SDHCNN modelclassification.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

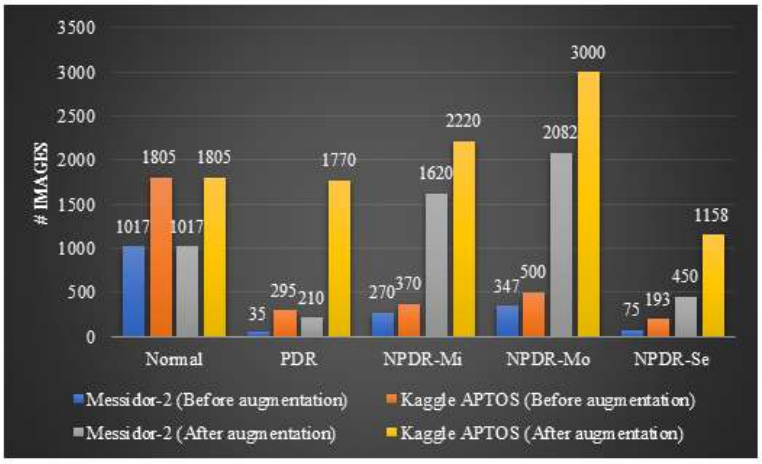

Fig. 4: Distribution of images used for analysis after and before augmentation

Table II: Data distribution in different severity classes for training and testing the 2D-SDHCNN approach

Phase	Dataset	Normal	PDR	NPDR-Mi	NPDR-Mo	NPDR-Se
Training	Messidor-2	712	147	1134	1457	315
	Kaggle APTOS	1264	1239	1554	2100	811
Testing	Messidor-2	305	63	486	625	135
	Kaggle APTOS	541	531	666	900	347

The number of images taken for analysis after and before augmentation is provided in Fig.4.To minimize the overfitting problem, fundus images are augmented by alterations such as darkening by 50, brightening by 50, and rotations by 90°, 180°, and 270°.Image augmentation was not performed for the class Normal, since this class hasa sufficient number of images to train the model.

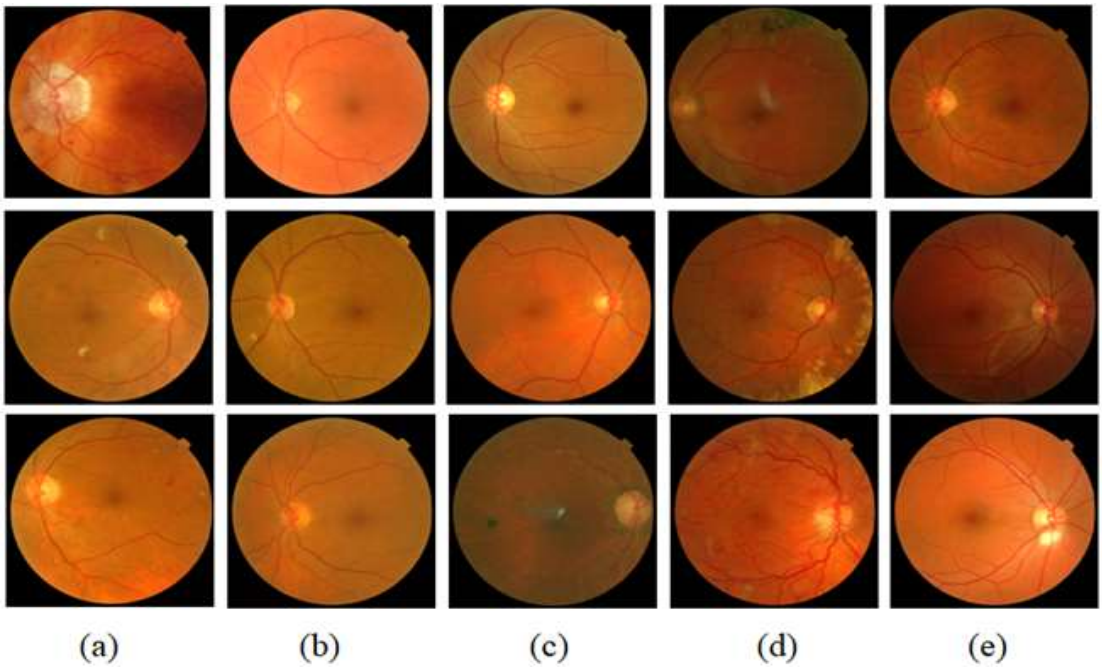

Fig. 5: Sample images utilized for analysis of 2D-SDHCNN approach (a) NPDR-Se (b) NPDR-Mo (c) NPDR-Mi (d) PDR (e) Normal

In the augmented fundus pictures, 30% of the pictures are randomly chosen and used for testing the classification performance. The other 70% of pictures are utilized in learning the 2D-SDHCNN model.

The distribution of images in each grade is depicted in Table II. A few images from these five DR severity categories are illustrated in Fig. 5.

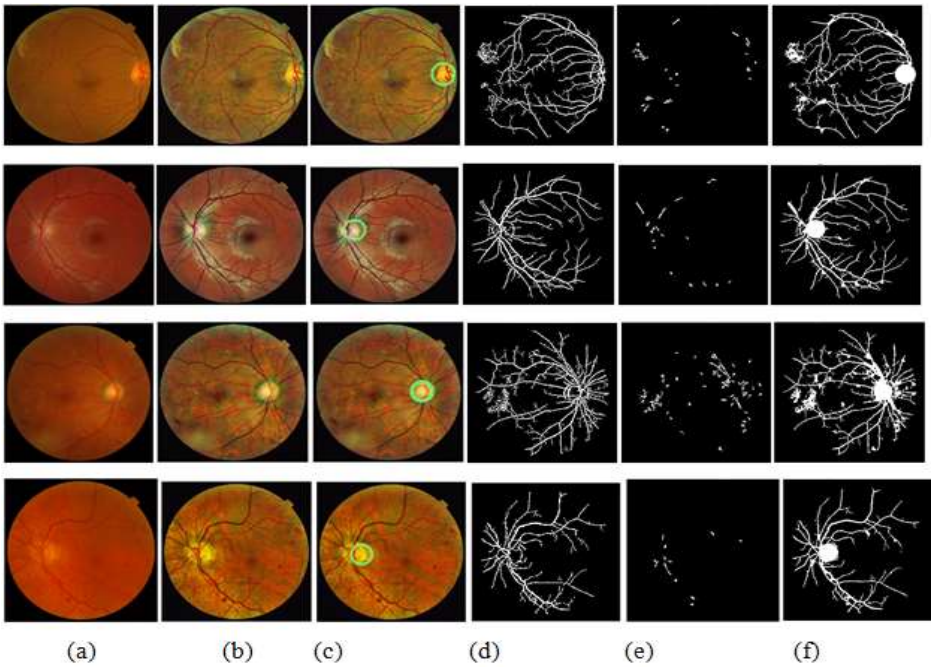

Fig. 6: Sample experimental outputs obtained during the 2D-SDHCNN-based severity detection process (a) Input fundus (b) Pre-processed fundus (c) Detected optic disc (d) Detected blood vessels (e) Detected lesion region (f) Region of interest

Fig. 6 provides the sample outputs attained during the severity detection process. It shows the pre-processed image, optic disc detection result, blood vessel detection result, and lesion detection result. For the detection of blood vessels, the maximal principal curvature algorithm uses the Gaussian filter standard deviation and a kernel size of unity and 3×3 respectively. The contrast enhancement process in blood vessel detection uses the factors a_{min} and a_{max} as 0 and 255 respectively. The circular Hough detection algorithm uses a canny edge detector to detect the edges. The minimum and maximum radius of the optic disc to detect is set between 20 to 60. The RoI image provided in Fig. 6(f) includes the three regions namely the optic disc region, the lesion region, and the blood vessel regions.

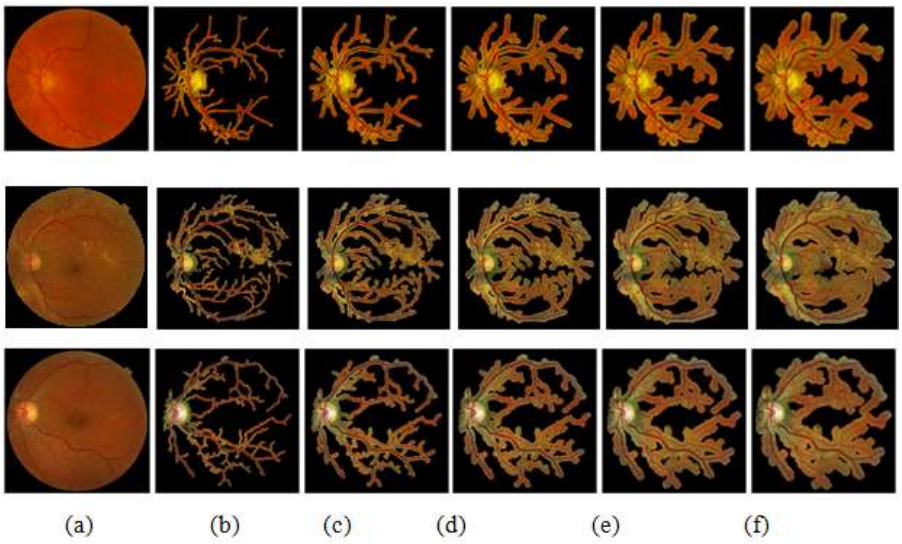

Fig. 7: Representation of dilated mask regions used in hierarchical descriptor extraction (a) Input fundus picture (b)-(f) Stage-1 to Stage-5 dilated regions

Fig. 7 shows the dilated images used by the hierarchical network for extracting the hierarchical descriptors. For this evaluation, the number of levels used is$L$ =5. The dilation is performed only on the blood vessel and the lesion region, where the optic disc region is left non-dilated. The hierarchical descriptor extracts more deep features from stage 5, while less deep features are extracted from stage1.The 2D-SDHCNN model was trained using the Adam optimizer with an epoch of 80, $L = 5$, a learning rate of 0.001, and a batch size of 32.

Table III: Evaluation results assessed for each DR grade using the Kaggle APTOS dataset

Severity	Acc (%)	Pre (%)	MCC (%)	Spe(%)	Rec(%)	F1(%)
Normal	98.23	97.69	97.44	99.69	98.23	97.96
PDR	96.31	98.53	96.80	99.89	96.31	97.40
NPDR-Mi	97.86	98.00	97.26	99.61	97.86	97.93
NPDR-Mo	98.04	97.82	96.92	99.20	98.04	97.93
NPDR-Se	98.24	96.03	96.72	99.69	98.24	97.13

The class-specific performance attained by the 2D-SDHCNN approach when analyzed using the Kaggle APTOS dataset is provided in Table III. In the case of the Kaggle APTOS dataset, the accuracy for the DR grades Normal, PDR, NPDR-Mi, NPDR-Mo, and NPDR-Se was estimated as 98.23%, 96.31%, 97.86%, 98.04%, and 98.24% respectively. The suggested 2D-SDHCNN approach provides a maximum accuracy for the DR grade NPDR-Se, while lower accuracy for the DR grade PDR.

Table IV: Evaluation results assessed for each DR grade using the Messidor-2 dataset

Severity	Acc (%)	Pre (%)	MCC (%)	Spe(%)	Rec(%)	F1(%)
Normal	98.36	97.09	97.19	99.31	98.36	97.72
PDR	88.89	80.00	83.66	99.10	88.89	84.21
NPDR-Mi	96.91	97.52	96.02	98.94	96.91	97.21
NPDR-Mo	96.48	98.37	95.82	98.99	96.48	97.42
NPDR-Se	96.30	93.53	94.43	99.39	96.30	94.89

Similarly, the evaluation results assessed using the Messidor-2 dataset for each DR gradeare provided in Table IV. The 2D-SDHCNN approach yields an accuracy of 98.36%, 88.39%, 96.91%, 96.48%, and 96.30% for the class Normal, PDR, NPDR-Mi, NPDR-Mo, and NPDR-Se respectively. The 2D-SDHCNN-based scheme provides higher accuracy for the DR grade Normal classes and lower accuracy for the PDR grade. Also, the suggested 2D-SDHCNN scheme results in an MCC of 97.19%, 83.66%, 96.02%, 95.82%, and 94.43% respectively for the DR grades Normal, PDR, NPDR-Mi, NPDR-Mo, and NPDR-Se respectively. The accuracy estimated in the Messidor-2 dataset is 7.42%, 0.94%, 1.56%, and 1.94% lower than the Kaggle APTOS dataset for the DR grades PDR, NPDR-Mi, NPDR-Mo, and NPDR-Se respectively. However, the accuracy attained in the Messidor-2 dataset is 0.13% higher than the APTOS dataset for the Normal class.

Table V: Comparison of performance between 2D-SDHCNN and other recent DR severity grading approaches when assessed using the Kaggle APTOS dataset

Methods	Acc (%)	Pre (%)	MCC (%)	Spe(%)	Rec(%)	F1(%)
MA-SL [30]	94.09	93.94	93.15	95.85	94.06	93.88
TA-Net [14]	93.48	93.49	92.66	95.67	93.63	93.67
DL-attention [31]	95.42	95.17	94.81	97.38	95.54	95.23
IR-CNN [32]	94.51	94.31	93.99	96.69	94.60	94.82
Vision transformer [33]	96.06	95.89	95.69	98.00	96.22	95.96
DeepRetiNet [34]	96.95	96.90	95.98	98.79	96.90	96.68
Proposed	97.73	97.61	97.03	99.61	97.73	97.67

Recent DR grading approaches namely vision transformer [33], DL-attention [31], DeepRetiNet [34], IR-CNN [32], TA-Net [14], and MA-SL [30] are utilized for comparison. In the case of the Kaggle

APTOS dataset, the 2D-SDHCNN approach results in an average accuracy, precision, MCC, specificity, recall, and F1-score of 97.73%, 97.61%, 97.03%, 99.61%, 97.73%, and 97.67% respectively as depicted in Table V. The accuracy, precision, MCC, specificity, recall, and F1-score found by the suggested 2D-SDHCNN approach is 0.78%, 0.71%, 1.04%, 0.82%, 0.83%, and 0.98% respectively higher than the DeepRetiNet model.

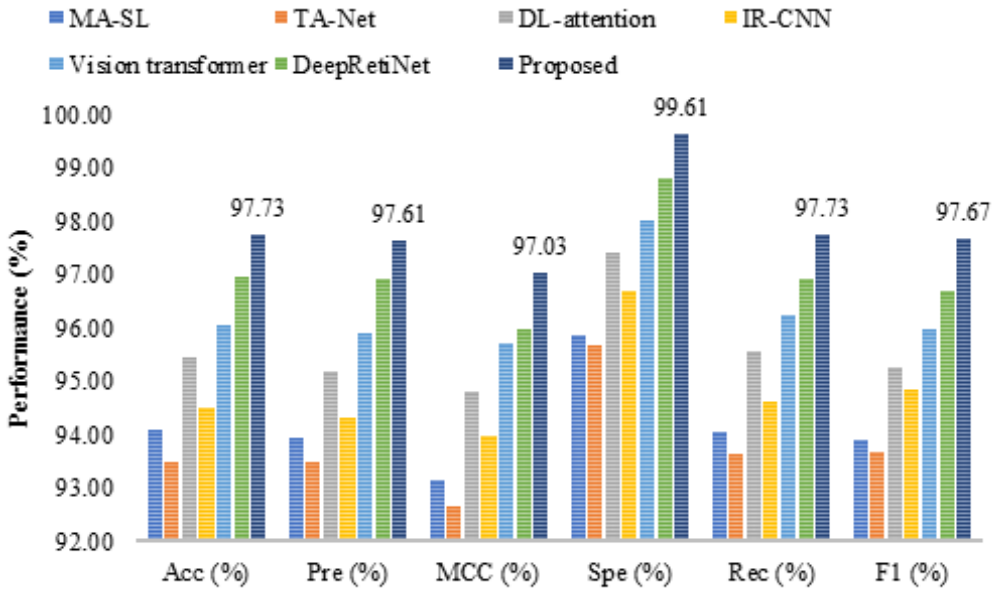

Fig. 8: Graphical comparison using Kaggle APTOS data between the 2D-SDHCNN and other recent DR grading approaches

The accuracy estimated by the 2D-SDHCNN approach using the APTOS dataset is0.78%, 1.67%, 3.22%, and 2.31% higher than the DeepRetiNet, Vision transformer, IR-CNN, and DL attention approaches as exemplified in Fig. 8.

Table VI: Comparison of performance between 2D-SDHCNN and other recent DR severity grading approaches when assessed using the Messidor-2 dataset

Methods	Acc (%)	Pre (%)	MCC (%)	Spe(%)	Rec(%)	F1(%)
MA-SL[30]	90.49	88.72	88.84	95.02	90.74	89.79
TA-Net [14]	90.52	88.38	88.46	94.96	90.31	89.45
DL-attention [31]	91.99	90.04	89.88	96.35	92.45	91.30
IR-CNN [32]	91.31	89.41	89.39	95.88	91.65	90.10
Vision transformer [33]	92.72	90.87	90.98	97.19	92.75	91.86
DeepRetiNet [34]	93.88	91.43	91.83	98.24	93.96	92.39
Proposed	95.39	93.30	93.42	99.15	95.39	94.29

The comparison was also made between the suggested 2D-SDHCNN approach and a few recent DR grading approaches using the Messidor-2 dataset and the results are presented in Table VI. In the case of the Messidor-2 dataset, the suggested approach yields an average accuracy, precision, MCC, specificity, recall, and F1-score of 95.39%, 93.30%, 93.42%, 99.15%, 95.39%, and 94.29% respectively which is 1.51%, 1.87%, 1.60%, 0.91%, 1.43%, and 1.90% more than the DeepRetiNet approach.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

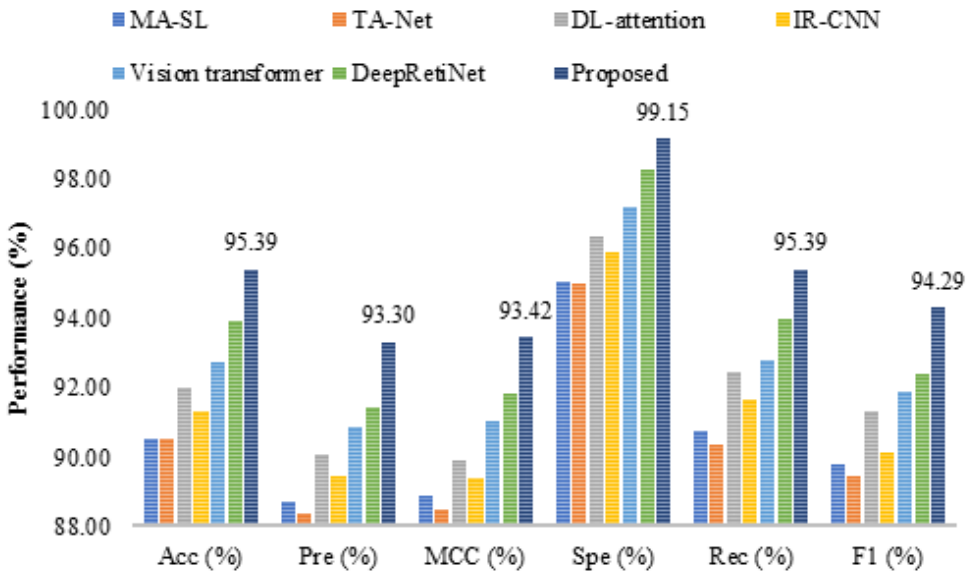

Fig. 9: Graphical comparison using Messidor-2 data between the 2D-SDHCNN and other recent DR grading approaches

The graphical comparison provided in Fig. 9 shows the increase in performance by the 2D-SDHCNN approach over the recent DR grading approaches in classifying the DR grades when evaluated using the Messidor-2 data. The accuracy estimated by the 2D-SDHCNN approach is 1.51%, 2.67%, and 4.08% more than the DeepRetiNet, Vision transformer, and DL-attention approach respectively when evaluated using the Messidor-2 dataset.

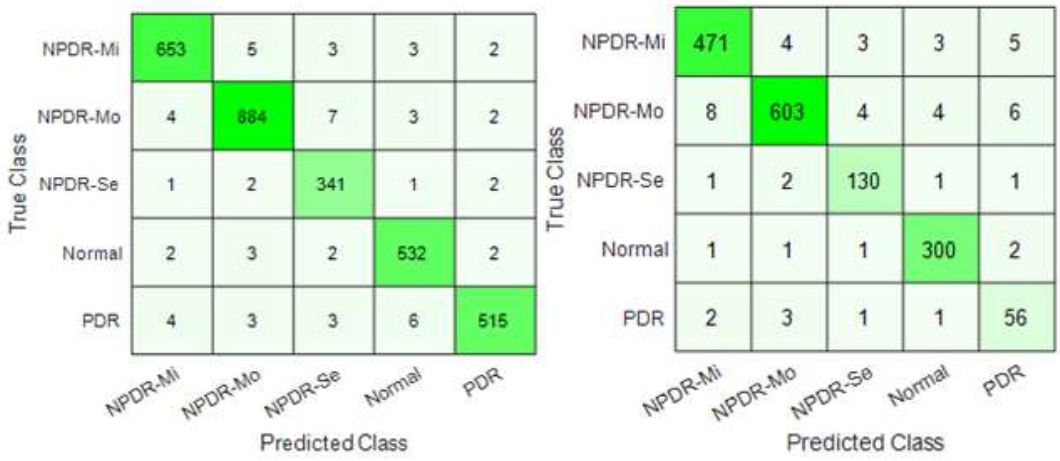

Fig. 10: Confusion matrix estimated by 2D-SDHCNN during the testing phase (a) Kaggle APTOS (b) Messidor-2

The number of fundus images classified under each category of DR grades is plotted by the confusion matrix shown in Fig. 10. The confusion plot illustrates that the 2D-SDHCNN provides a higher true positive result that improves the performance in DR grading.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7s, 2025
https://www.theaspd.com/ijes.php

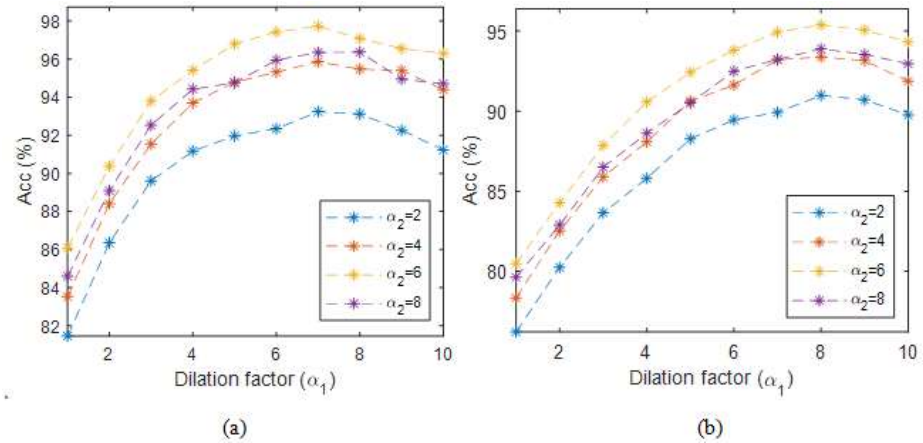

Fig. 11: Impact of dilation factors α_1, α_2 in DR grading accuracy (a) Kaggle APTOS dataset (b) Messidor-2 dataset

The performance of the 2D-SDHCNN approach with respect to the dilation factors α_1, α_2 is illustrated in Fig. 11.In the case of the Kaggle APTOS dataset, as the dilation factor α_1 is varied from 1, the accuracy increases and attains a maximum for the dilation factor $\alpha_1 = 7$. For further increase in dilation factor α_1, the accuracy reduces. Thus, while using the Kaggle APTOS dataset, the maximum performance is attained for $\alpha_1 = 7$ and $\alpha_2 = 6$. In the case of the Messidor-2 dataset, the dilation factor α_1 is varied from 1, the accuracy increases and attains a maximum for the dilation factor $\alpha_1 = 8$. For further increase in dilation factor α_1, the accuracy reduces. Thus, while using the Messidor-2 dataset, the maximum performance is attained for $\alpha_1 = 8$ and $\alpha_2 = 6$. Generally, the proposed 2D-SDHCNN approach can be used with a dilation factor $\alpha_1 = 7\ or\ 8$ and $\alpha_2 = 6$.

4. CONCLUSION

This work introduced a DR grading approach 2D-SDHCNN that uses dilated images in extracting the hierarchical descriptors. The approach initially detects three regions namely the blood vessels, optic disc, and lesion regions as RoI. For detecting the blood vessels maximal principal curvature-based approach is used, while for detecting the optic disc, a circular Hough transform is used. Thresholding and morphological operations are used to detect the possible lesion regions. From these three regions, blood vessels and lesion regions are used as dilating regions while the optic disc is used as the non-dilating region. Two networks namely the global feature extraction network and the hierarchical network are used to extract the descriptors. The Hierarchical network uses different dilated images for feature extraction, where each section of the subnetwork uses a different number of channels. Also, dilation is performed for each section of the subnetwork to obtain more deep features. Datasets namely Kaggle APTOS and Messidor-2 are utilized for evaluating the suggested approach. The suggested 2D-SDHCNN yields an accuracy of 97.73% and 95.39% when evaluated using the Kaggle APTOS and Messidor-2 datasets. In the case of the Kaggle APTOS dataset, the suggested scheme attains anaverage accuracy, precision, MCC, specificity, recall, and F1 score of 97.73%, 97.61%, 97.03%, 99.61%, 97.73%, and 97.67% respectively. The evaluation results show that the suggested 2D-SDHCNN approach can better detect the severity grades in DR clinical studies.

REFERENCES

[1]. Alam, U., Asghar, O., Azmi, S., & Malik, R. A. (2014). General aspects of diabetes mellitus. *Handbook of clinical neurology*, *126*, 211-222.

[2]. Tomic, D., Shaw, J. E., & Magliano, D. J. (2022). The burden and risks of emerging complications of diabetes mellitus. *Nature Reviews Endocrinology*, *18*(9), 525-539.

[3]. Teo, Z. L., Tham, Y. C., Yu, M., Chee, M. L., Rim, T. H., Cheung, N., ... & Cheng, C. Y. (2021). Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*, *128*(11), 1580-1591.

[4]. Curtis, T. M., Gardiner, T. A., & Stitt, A. W. (2009). Microvascular lesions of diabetic retinopathy: clues towards understanding pathogenesis?. *Eye*, *23*(7), 1496-1508.

[5]. Suchetha, M., Ganesh, N. S., Raman, R., & Dhas, D. E. (2021). Region of interest-based predictive algorithm for subretinal hemorrhage detection using faster R-CNN. *Soft Computing*, *25*(24), 15255-15268.

[6]. Krishnan, S. H., Vishwa, C., Suchetha, M., Raman, A., Raman, R., Sehastrajit, S., & Dhas, D. E. (2023). Comparative performance of deep learning architectures in classification of diabetic retinopathy. *International Journal of Ad Hoc and Ubiquitous Computing*, *44*(1), 23-35.

[7]. Kassani, S. H., Kassani, P. H., Khazaeinezhad, R., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2019, December). Diabetic retinopathy classification using a modified xception architecture. In 2019 IEEE international symposium on signal processing and information technology (ISSPIT) (pp. 1-6). IEEE.

[8]. Akram, M. U., Khalid, S., Tariq, A., Khan, S. A., & Azam, F. (2014). Detection and classification of retinal lesions for grading of diabetic retinopathy. Computers in biology and medicine, 45, 161-171.

[9]. Shanthi, T., &Sabeenian, R. S. (2019). Modified Alexnet architecture for classification of diabetic retinopathy images. Computers & Electrical Engineering, 76, 56-64.

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

[11]. Kalyani, G., Janakiramaiah, B., Karuna, A., & Prasad, L. N. (2023). Diabetic retinopathy detection and classification using capsule networks. Complex & Intelligent Systems, 9(3), 2651-2664.

[12]. 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.

[13]. Das, S., Kharbanda, K., Suchetha, M., Raman, R., & Dhas, E. (2021). Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. Biomedical Signal Processing and Control, 68, 102600.

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

[15]. Amin, J., Sharif, M., Yasmin, M., Ali, H., & Fernandes, S. L. (2017). A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions. Journal of Computational Science, 19, 153-164.

[16]. Mondal, S. S., Mandal, N., Singh, K. K., Singh, A., &Izonin, I. (2022). Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy. Diagnostics, 13(1), 124.

[17]. Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. Sensors, 21(11), 3704.

[18]. Wu, Z., Shi, G., Chen, Y., Shi, F., Chen, X., Coatrieux, G., ... & Li, S. (2020). Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network. Artificial Intelligence in Medicine, 108, 101936.

[19]. Gayathri, S., Gopi, V. P., & Palanisamy, P. (2021). Diabetic retinopathy classification based on multipath CNN and machine learning classifiers. Physical and engineering sciences in medicine, 44(3), 639-653.

[20]. Shankar, K., Sait, A. R. W., Gupta, D., Lakshmanaprabu, S. K., Khanna, A., & Pandey, M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recognition Letters, 133, 210-216.

[21]. Qureshi, I., Ma, J., & Abbas, Q. (2021). Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. Multimedia Tools and Applications, 80(8), 11691-11721.

[22]. Rahim, S. S., Palade, V., Shuttleworth, J., & Jayne, C. (2016). Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing. Brain informatics, 3, 249-267.

[23]. Madarapu, S., Ari, S., & Mahapatra, K. (2024). A multi-resolution convolutional attention network for efficient diabetic retinopathy classification. Computers and Electrical Engineering, 117, 109243.

[24]. Sivapriya, G., Devi, R. M., Keerthika, P., & Praveen, V. (2024). Automated diagnostic classification of diabetic retinopathy with microvascular structure of fundus images using deep learning method. Biomedical Signal Processing and Control, 88, 105616.

[25]. Musa, P., Al Rafi, F., &Lamsani, M. (2018, October). A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods to help the application of face recognition. In *2018 third international conference on informatics and computing (ICIC)* (pp. 1-6). IEEE.

[26]. Zana, F., & Klein, J. C. (2001). Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE transactions on image processing*, *10*(7), 1010-1019.

[27]. Abdullah, M., Fraz, M. M., & Barman, S. A. (2016). Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm. *PeerJ*, *4*, e2003.

[28]. Abràmoff, M. D., Folk, J. C., Han, D. P., Walker, J. D., Williams, D. F., Russell, S. R., ... &Niemeijer, M. (2013). Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA ophthalmology, 131(3), 351-357.

[29]. Karthik, Maggie, and Sohier Dane. APTOS 2019 Blindness Detection. https://kaggle.com/competitions/aptos2019-blindness-detection, 2019. Kaggle.

[30]. Zhang, C., Chen, P., & Lei, T. (2023). Multi-point attention-based semi-supervised learning for diabetic retinopathy classification. Biomedical Signal Processing and Control, 80, 104412.

[31]. Romero-Oraá, R., Herrero-Tudela, M., López, M. I., Hornero, R., & García, M. (2024). Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading. Computer Methods and Programs in Biomedicine, 249, 108160.

[32]. Ali, G., Dastgir, A., Iqbal, M. W., Anwar, M., & Faheem, M. (2023). A hybrid convolutional neural network model for automatic diabetic retinopathy classification from fundus images. IEEE Journal of Translational Engineering in Health and Medicine, 11, 341- 350.

[33]. Oulhadj, M., Riffi, J., Khodriss, C., Mahraz, A. M., Yahyaouy, A., Abdellaoui, M., ... & Tairi, H. (2024). Diabetic retinopathy prediction based on vision transformer and modified capsule network. Computers in Biology and Medicine, 175, 108523.

[34]. Chellaswamy, P., & Kamalam, C. J. R. N. R. (2025). Attention-enhanced DeepRetiNet for robust hard exudates detection in diabetic retinopathy. Biomedical Signal Processing and Control, 100, 106903.