

A Feature-Fusion Approach Combining Domain-Specific Descriptors and RNNs for Automated Diabetic Retinopathy Classification

Shilpa Chindamwar¹, Dr. Rajendra Rewatkar²

¹Research Scholar, Department of Computer Science, Engineering, (AIDS), FEAT, DMIHER, Wardha, India

²HOD, Biomedical Engineering, FEAT, DMIHER, Wardha, India

HOW TO CITE: ABSTRACT:

Shilpa Chindamwar, Rajendra Rewatkar (2026). A Feature-Fusion Approach Combining Domain-Specific Descriptors and RNNs for Automated Diabetic Retinopathy Classification.

International Journal of Special Education, 41(17s), 309-324

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication under the terms and conditions

of the Creative Commons Attribution

(CC BY)

license (<http://creativecommons.org/licenses/by/4.0/>).

The increasing prevalence of diabetes has significantly raised the incidence of diabetic retinopathy (DR) is essential for preventing irreversible vision impairment and blindness. This study presents a robust binary classification framework for automated DR detection using the Augmented Messidor retinal fundus image dataset. The proposed methodology integrates a heterogeneous set of handcrafted features, including matched filter based vascular descriptors, Gray Level Co-occurrence Matrix (GLCM) texture features, discrete wavelet transform (DWT) based multiresolution features, Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP) descriptors, patch-level contrast features, and blind image quality assessment metrics. These complementary features are fused into a unified feature representation that captures both structural and textural characteristics of retinal images. The fused feature vector is subsequently classified using a customized Recurrent Neural Network (RNN), which effectively models sequential relationships within the high dimensional feature space to enhance classification performance. Experimental evaluation demonstrates that the proposed framework achieves an overall classification accuracy of 96.11%, with a sensitivity of 95.95%, specificity of 94.20%, precision of 94.78%, and an F1 score of 95.35%, outperforming or matching existing state of the art approaches. Furthermore, binary classification across individual DR severity grades yields accuracies of 94.44% (Grade 0), 97.08% (Grade 1), 95.83% (Grade 2), and 97.78% (Grade 3), demonstrating the robustness of the proposed framework under varying disease conditions. The findings confirm that the integration of carefully engineered handcrafted features with a customized deep learning architecture significantly enhances diabetic retinopathy detection and provides a reliable foundation for future multi class DR grading systems and real world clinical decision support applications.

Keywords: Diabetic retinopathy, fundus image analysis, handcrafted features, recurrent neural network, binary classification, medical image processing, computer-aided diagnosis.

1. Introduction

Diabetic Retinopathy (DR) is among the leading causes of visual impairment and is one of the most commonly recognized complications in a diabetic patient with diabetes [1]. DR is a progressive microvascular disease of the retina that results from sustained hyperglycemia,

leading to damage to the retinal vasculature. A cardinal hurdle to overcome clinically is that, in the early stages of DR, the presentation of clinical symptoms is usually asymptomatic; i.e., a patient will not experience any visible discomfort or visual impairment until a more severe phase of DR [2]. By the time this happens, irreversible damage could be done to

their retina, greatly increasing either vision impairment or even total blindness. Therefore, accurate early diagnosis at this stage is essential to prevent disease progression and to protect vision. For the monitoring and grading of DR in clinical practice, ophthalmologists mainly rely on colour fundus images (FIs), which allow visual inspection of the retinal structures and have proved effective in identifying the classic lesions (e.g. microaneurysms, haemorrhages, and exudates).

The increasing incidence of diabetes worldwide has resulted in a significant rise in the number of patients needing DR screening. Although accurate, manual inspection of fundus images for diagnostic purposes requires a lot of time and expertise, and is not only labourious but also costly in clinical resources. In some resource-restricted healthcare systems, population-level screening is hard to maintain. In addition, DR is mostly not symptomatic in its early stage and thus, the disease often goes unnoticed until it develops into more severe forms with severe damage that cannot be easily treated, and treatment alternatives are limited. This highlights the compelling demand for automatic systems to detect this disorder on an early, accurate and in bulk basis on an urgent basis that is of great necessity. Automated DR classification systems could effectively aid ophthalmologists by mitigating diagnostic burden, screening in a timely manner and supporting early treatment interventions thanks to reducing costs and ease of diagnosis. Thus, the incorporation of intelligent, automatic diagnostic tools into therapy processes offers a major advance in targeted treatment and the early management of DR.

Recently, there has been much attention on the diabetic retinopathy (DR) grading in clinical work by researchers, especially using deep learning and AI for the evaluation task. The grading exercise is usually treated as an image classification task, and the current method is broadly classified into Rule- and End-to-end approaches. Rule-based methods use handcrafted features to describe the textural and structural characteristics of fundus images (FIs), which are subsequently captured and then classified by machine learning or deep learning

models. U-Net has previously been used for vessel and optic disc separation, followed by CNN and SVM-based classifiers [3], and Naïve Bayes was developed for the classification of three DR grades from textural factors [4]. They have also devised dual model systems to detect abnormality and DR grading [5], automate the analysis of colour FIs for retinopathy detection [6] and ML-based identification of microaneurysms and exudates with high accuracy [7]. Taken together, these rule-based strategies focus on the analysis of high-level retinal structures (i.e., optic disc, vasculature, lesion patterns) with an objective of automating the development of early DR grading, thus enabling the ability to intervene early to reduce the probability of vision loss.

With the recent developments in deep learning, advanced end-to-end architectures for hierarchical DR recognition are now accessible. Zhang et al. [8] propose an Efficient Net model enhanced by an attention mechanism and transfer learning for DR grade classification to deliver accurate feature extracts targeting lesions for DR classification. Ramos et al. [9] use a semi-supervised method to classify with improved accuracy by means of using labelled and unlabelled images for self-supervised pre-training, curated fragments of both images for fine-tuning and distillation to improve the accuracy. Likewise, DenseNet coupled with attention was used for highlighting subtle lesion features that are essential for DR progression [10], whereas a second model that synthesized discrete wavelet characteristics with CNN outperformed current methods [11]. Khanna et al. [12] further validated the effectiveness of deep learning in DR grades with the use of three complementary nets to categorise DR into five grades. Together, these examples illustrate how AI holds great potential for DR diagnosis, and they emphasize the need for innovative end-to-end and hybrid CNN architectures to improve the accuracy, efficiency and patient outcome.

Current research in diabetic retinopathy (DR) aims to achieve better accuracy as well as better throughput. Despite the development of a variety of feature-based deep learning architectures, this challenge remains: the accurate mapping of fundus images (FIs) to

detect the characteristics of the subtle lesions, which are essential for DR characterization. Additionally, the dependency of ophthalmologists on manual lesion labelling not only increases the time spent during diagnosis, but also adds to the clinical burden. In this context, the need for an intelligent DR classification system is critical. Such a system should be able to process FIs itself for extracting clinically relevant features and classifying disease severity with little or no human involvement [2]. A dedicated intelligent DR grading module of this type is required to facilitate rapid large-scale screening, relieve the load of ophthalmologists and enable early diagnosis of DR, which would result in positive and meaningful patient outcomes and facilitate public health action for diabetic eye care.

2. Literature Review

Several frameworks for the diabetic retinopathy (DR) grade classification methods have been introduced in recent years, which indicate the growing importance of artificial intelligence for the analysis of ophthalmic images. Gupta et al. [13] proposed a hybrid method combining deep learning and Fennec Fox optimization for DR grading. Their framework consisted of median filtering in order to suppress noise, feature extraction using Inception-ResNet-v2, and classification based upon a fine-tuned gated recurrent unit (GRU) with validation performed on the Kaggle benchmark dataset. Zhang et al. [14] proposed a dual-path multi-module network using augmentation and colour-correction with multi-scale fusion techniques for increasing lesion visibility. Multi-path multiplexing was achieved with a set of variable convolutional kernels, and the classification was validated using multi-feature fusion in the public datasets and a locally acquired fundus dataset. Biswas and Banik [15] suggested an ensemble deep learning model consisting of a pretrained CNN for feature extraction and a recurrent neural network (RNN) for spatial modelling with 96% precision for DR grading. Similarly, Srilakshmi et al. [16] used augmented data and dynamic weighted optimisation. The approach involved lesion localization by Otsu thresholding and morphological operations, feature extraction using GLCM, optimization by dynamic Flemingo optimization and final classification with cascaded voting ensemble network. Feng et al. [17] advanced DR classification by modifying

ResNet with an external attention mechanism to improve global feature representation and introducing multi-scale convolution within the residual branch. The Adam optimizer was replaced with the Sophia optimizer for enhanced discriminative capacity. Applied to an augmented Kaggle dataset, the modified ResNet achieved a 4.36% improvement in accuracy over the baseline ResNet. Ahmed [18] adopted transfer learning with the Vision Transformer (ViT-Base-Patch16-224) for classifying DR and age-related macular degeneration on the APTOS and IChallenge-AMD datasets, following preprocessing and tensor standardization. Building on this, Kavitha and Kasthuri [19] proposed a two-way cascaded CNN designed for the detection of non-proliferative DR, combining enhanced preprocessing techniques with handcrafted feature fusion to achieve a four-grade classification. Prakash and Vinoth Kumar [20] introduced an ensemble learning methodology and combined the outputs from five pretrained networks (ResNet50, Inception-ResNet-v2, DenseNet121, Xception, VGG19) achieving a classification accuracy of approximately 92%. Zhang et al. [21] proposed a two-stage system of lesion detection supplemented by DR classification. They applied an improved neural process for lesion-sliced detection and an enhanced deep forest classifier for interpretable grading in their new model, which outperformed the performance of conventional forest and residual network models on the IDRiD, E-optha, and Kaggle datasets.

In this paper, we deploy a custom-trained, multilayer machine learning model (RNN) based on hand-crafted features from a variety of sources, such that it can handle a wide variety of images with different datasets and conditions. By merging vascular, textural, and structural descriptors – including matched filter-based vessel features, GLCM, wavelet, HOG, and LBP – with patch-based contrast and blind image quality features, the proposed framework captures both interpretable and abstract cues that are pivotal for effective DR classification. This approach provides clinically relevant, interpretable features for lesion detection, leading to more trustworthy networks in comparison to purely black-box CNNs. Furthermore, feature fusion alone on the single RNN substantially reduces computation requirements compared to deep ensemble methods and still satisfies the competitive performance requirements. Table 1 shows a

summary of some recent work in DR

classification discussed in the literature review.

Table 1 –Summary of recent Literature work of DR classification

Reference	Methodology	Dataset(s)	Key Results / Highlights
Gupta et al. [13]	Deep learning + Fennec Fox optimization; noise removal with median filter; features via Inception-ResNet-v2; classification with fine-tuned GRU	Kaggle benchmark dataset	Effective optimization; validated GRU-based classification
Zhang et al. [14]	Dual-path multi-module network; augmentation with color correction + multi-scale fusion; multi-path convolution + feature fusion	Two public datasets + local hospital dataset	Improved lesion enhancement; optimized local features; robust classification
Biswas and Banik [15]	Ensemble CNN + RNN; CNN for feature extraction, RNN for spatial dependencies	Large-scale retinal dataset (public)	Achieved 96% precision in DR grade classification
Srilakshmi et al. [16]	Augmentation + dynamic weighted optimisation; lesion localization via Otsu thresholding + morphology; GLCM + dynamic Flemingo optimization; cascaded voting ensemble	Fundus image datasets (not specified)	Improved robustness and generalization; ensemble improved accuracy
Feng et al. [17]	Modified ResNet with external attention and multi-scale convolution; Sophia optimizer for better feature discrimination	Kaggle dataset (augmented: 20,184 images from 3,662)	4.36% higher accuracy than baseline ResNet
Ahmed [18]	Vision Transformer (ViT-Base-Patch16-224) with transfer learning; preprocessing + tensor standardization	APTOS + IChallenge-AMD datasets	Classified DR and AMD; strong performance with transfer learning
Kavitha and Kasthuri [19]	Preprocessing (enhancement, denoising, defect removal); two-way cascaded CNN; handcrafted + deep feature fusion	Retinal images (non-proliferative DR detection)	Four-grade DR classification; improved interpretability
Prakash and Vinoth Kumar [20]	Ensemble of 5 pretrained CNNs (ResNet50, Inception-ResNet-v2, DenseNet121, Xception, VGG19)	Public fundus datasets	Five-grade classification (normal to severe); accuracy ~92%
Zhang et al. [21]	Two-stage framework: lesion detection (improved neural process) + DR classification (enhanced deep forest)	IDRiD, E-optha, Kaggle	Outperformed residual networks and forest models; improved interpretability and generalization

Although DR classification systems provide a promising experience for our DR classification framework, many limitations remain. An excessive number of models are heavily reliant on optimization strategies, heavy ensembles, or complex multi-stage pipelines, which has raised their computational burden and reduces the scalability for real-time deployment. Several studies have indicated large levels of accuracy, but their fidelity is typically tuned to particular

datasets (e.g., Kaggle), giving rise to considerations at the transferability levels to other types of clinical workloads. Others concentrate on particular DR phases (e.g., non-proliferative grades), and hence, there is a lack of generalizability up to the whole DR spectrum. Furthermore, methods dependent on heavy manual augmentation or handcrafted preprocessing steps may not function well on heterogeneous fundus images.

Taken together, these limitations emphasize the necessity for more balanced paradigms obtaining high diagnostic accuracy, efficiency and generalizability with low computational and operational burden. In addition to accuracy, the proposed method has high sensitivity and specificity as well as F1-score values (reducing false negatives), thus satisfying a significant requirement during the clinical DR screening process. The framework is designed mainly to be used for binary classification, but has demonstrated strong per-grade classification performance, thus ensuring scalability to multi-class grading without degrading performance.

3. Methods and Materials

The current study has been made with a total of 1,200 fundus images from the Messidor dataset [22] specifically curated for medical diagnosis.

Each image is presented with two separate expert diagnoses, describing the retinopathy grade and the risk of macular edema. Images were collected from three different ophthalmologic departments, which generated different resolution segments (1440×960, 2240×1488, 2304×1536) in four annotated zip folders. Of the images available, 400 were taken without pupil dilation, whereas the other 800 were obtained with a single drop of 0.5% Tropicamide after pupil dilation. Although the fundus images were initially classified into four grades (0 to 3) according to the findings of microaneurysms (MA), haemorrhages (H), and neovascularization (N), current work applies a re-categorization in order to facilitate comparison of current state-of-the-art methods. The distribution of samples across the retinopathy grades in the complete dataset is summarized in Table 2.

Table 2 – Description of the Messidor dataset

Retinopathy Grade	Criteria	Number of samples
0 (Normal)	MA = 0, H = 0	546
1	$0 < MA \leq 5, H = 0$	153
2	$5 < MA < 15$ or $0 < H < 5, N = 0$	247
3 (Severe)	$MA \geq 15$ or $H \geq 5$ or $N = 1$	254

Figure 1 provides a brief overview of the overall layout of the developed DR grading scheme. The system is built with three phases. The Preprocessing phase prepares the fundus images by enhancing quality and reducing noise, ensuring that the relevant retinal structures are emphasized. Clinically relevant features, such as vascular, textural, and structural descriptors, are

systematically sampled in the Feature Extraction step to provide important lesion data. To conclude, the Classification phase employs a customized recurrent neural network to categorize the images into appropriate DR grades, effectively integrating the extracted features for accurate and robust assessment.

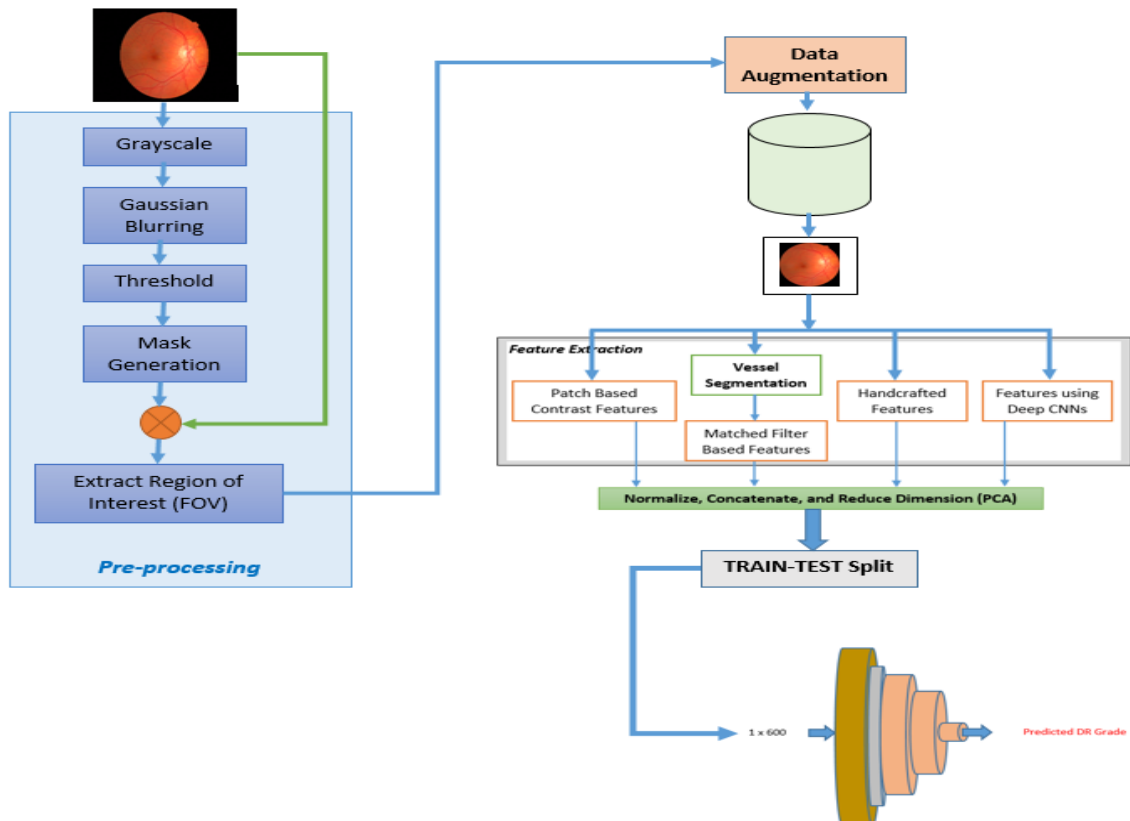


Figure 1 – The proposed DR Grade Classification System

3.1 Data Augmentation

Having just 1,200 fundus images for the Messidor dataset limits the generalizability of deep learning models and can increase the chances of overfitting in that the network would memorize training examples rather than learning representations that are robust and useful. This limitation is counterbalanced by the method of data augmentation, which synthetically augments the data through transformations like rotations, horizontal flips, contrast adjustments, and controlled noise. This gives the model more robustness towards variability and increased generalization ability when the dataset is

expanded. This is especially important for fundus images, where identifying subtle features related to DR or other ocular conditions through diverse imaging parameters is critical. In the current research, each original Messidor image was rotated 100° clockwise and counterclockwise, flipped horizontally, and stretched for contrast, which made five augmented samples per image. As a result, the effective dataset size grew to 7,200 images ($1,200 \times 6$). Figure 2 shows the five augmented samples collected from a single original image, with the first panel containing a Messidor image without modification.

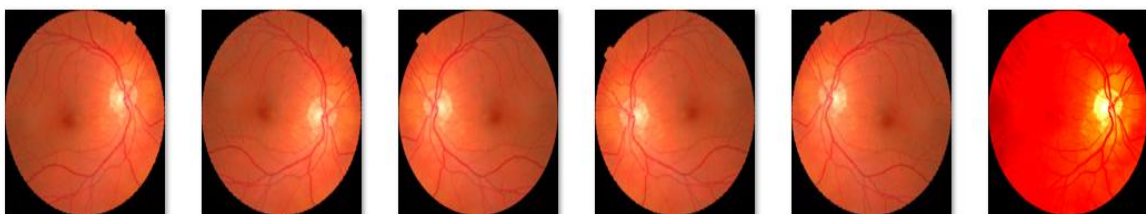


Figure 2 – From left to right: the original fundus image (FOV), the image rotated 100° clockwise, the horizontally flipped image, the horizontally flipped image rotated 100° counter-clockwise, the horizontally flipped image rotated 100° clockwise, and the contrast-stretched image.

3.2 Preprocessing

Initially the image of fundus is grayscale and is used for Gaussian blur with a 3×3 kernel (which provides the minimum discriminative noise and retains important structure). From the analysis of all 1,200 images in the dataset, we threshold the blur image at a number of 80 to produce a binary mask that separates foreground (intensity 1) from background (intensity 0), thereby lessening the computational load. Using this mask the original image is then multiplied so that all regions extraneous to the Field of View (FOV) are in the background. Afterward, the extreme non-zero pixels of all four borders are identified to eliminate residual background, and then the image is cropped accordingly. To achieve uniformity for feature extraction, all images are resized to 299×299 pixels. The output of the preprocessing stage in the sequence is shown in Figure 3.

The preprocessing step is essential, aiming to generate images of higher quality, remove background areas that are no longer relevant, standardized dimensions, and retain important retinal architectures which deliver a clean and stable input for feature extraction and classification with the aim of enhancing the accuracy and reliability of the DR grading system. The automatic DR severity classification of retinal fundus images forms an integral part of computer-aided ophthalmic diagnosis. To a patient, the representation of the characteristic features is important in obtaining its pathological variations, similar to microaneurysms, exudates, and other lesions. Bringing together human-perceptible, handcrafted, and blind features, a combined heterogeneous, complementary set of clinically relevant cues can be collected that enhances DR-related changes, enhancing the model's ability to pick out small changes related to DR and improves the accuracy of grading.

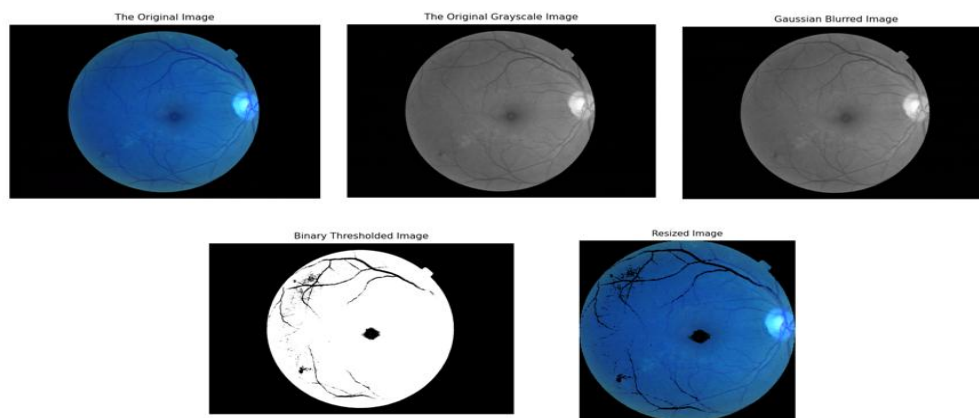


Figure 3 – Output at each stage of preprocessing

3.3 Local Contrast-based Features

A variety of state-of-the-art contrast weighting algorithms have been suggested for improved feature representation in retinal images [23]. Among these, Simone et al. [24] proposed a new scheme based on a modified Difference of Gaussian (mDoG) to quantify image contrast levels effectively. Their approach is driven by the recognition that human visual perception is affected not only by retinal illumination but also by stimulus contrast. To capture this behaviour, a receptive field model was developed by the authors which accounts for the spatial sensitivity of the human visual system. Specifically, for a two-dimensional circular and symmetric

Gaussian distribution with unit magnitude, the centre of the receptive field determines the spatial sensitivity (S_s). The central component of this receptive field is mathematically expressed by the following expression (1).

$$\text{Middle component } (mm, nn) = \exp \left[-\left(\frac{mm}{s_c} \right)^2 - \left(\frac{nn}{s_c} \right)^2 \right] \quad (1)$$

The factor in the denominator (s_c) represents the radius. If ' S ' denotes the value of the highest sensitivity, then s_c denotes the length over which S_s reduces to $\frac{S_s}{e}$. This formulation allows the mDoG-based technique to emphasize regions of

interest in fundus images where subtle contrast variations, such as those caused by microaneurysms or haemorrhages, are critical for accurate DR grading.

Similarly, the neighbouring component is given by expression (2).

$$\text{Neighbouring component } (mm, nn) = 0.85 \left(\frac{s_c}{s_s} \right)^2 \exp \left[- \left(\frac{mm}{r_s} \right)^2 - \left(\frac{nn}{s_s} \right)^2 \right] \quad (2)$$

The scaling factor $[0.85 * (\frac{s_c}{s_s})^2]$ governs the volume of the Neighbouring component to 85% of the middle receptive field.

Thus, the universal strategy to measure the image perceptual contrast is evaluated by expression (3).

$$C(mm, nn) = \frac{R_c(mm, nn) - R_s(mm, nn)}{R_c(mm, nn) + R_s(mm, nn)} \quad (3)$$

Therefore, the middle component takes the form,

$$R_c(mm, nn) = \sum_{ii=mm-3s_c}^{ii=mm+3s_c} * \sum_{jj=nn-3s_c}^{jj=nn+3s_c} \text{Middle}(ii - mm, jj - nn) I(ii, jj) \quad (4)$$

The neighbouring component is,

$$R_s(mm, nn) = \sum_{ii=mm-3s_c}^{ii=mm+3s_c} * \sum_{jj=nn-3s_c}^{jj=nn+3s_c} \text{Neighbor}(ii - mm, jj - nn) I(ii, jj) \quad (5)$$

The FOV image was used to compute patch-level contrast weights via the mDoG approach [25–26] with a patch size of $w = 13 \times 13$. To handle boundary effects, the image was padded with $(\frac{w}{2})$ pixels on all sides. Patch-based contrast weights effectively capture salient characteristics of the fundus image, and averaging over a 3×3 neighbourhood further reduces feature dimensionality while preserving local structure. It improves the contrast of local structures in fundus images with the goal of emphasizing subtle retinal lesions like microaneurysms, haemorrhages, and exudates. The resulting contrast-weighted image serves as an input for subsequent feature extraction, improving the

sensitivity of the DR grading system to clinically significant variations. The integration of the physiological basis of human contrast perception in the mDoG technique helps the model to focus on visually salient features, minimizing the effect of background variations and enhancing the robustness of automated DR severity classification. The FOV image is used to compute the patch-level contrast weights using the mDoG approach introduced in [25–26]. We used a patch size of $w=13 \times w=13$ for contrast weights.

To compensate for the effect at the boundary, the FOV image was padded with int ($w/2$) pixels from all sides, where ‘ $w=13$ ’ is the window or patch size over which the contrast is computed. The proposed patch-based contrast weights show significant representative characteristics of the FI. Since the neighbourhood pixels show similar characteristics, we computed the mean of contrast values using a 3×3 neighbourhood, eventually reducing the feature dimension.

3.4 Conventional Features

Alongside patch-based contrast, global handcrafted features are extracted from the FOV for delicate retinal structure enhancement. The FOV was resized to 260×260 and the green channel employed for feature extraction. Global contrast is calculated and corrected using Tadmor and Tolhurst method. This enhancement focuses attention on diagnostics-important details to ensure the subtle lesions are easier to identify, an important attribute for correct DR grading.

The contrast-corrected fundus image (C_c) can be calculated in a single formulation (expression 6):

$$C_c = \frac{259(255c_m=255)}{255(259-255c_m)} (FI - 128) + 128 \quad (6)$$

Here, FI denotes the original fundus image, c_m is the measured contrast, and C_c represents the contrast-enhanced output.

To suppress weak and aliasing textures and simultaneously enhance edge structures, a denoising filter proposed by Wetzler and Kimmel [27] was used to smooth the local regions. We trained this filter on local pixels in a spectral 5D space $\{x, y, R, G, B\}$ to provide edge-preserving denoising at the patch level. Two parameters can be employed to fine-tune the

method: the number of iterations (E); and time-step stability (δ). $E=5$ and $\delta=0.25$ were used in this study to create the filter, which improved vessels and microaneurysms without any loss of retinal details. Next, the edge-enhanced image was processed utilizing the Frangi vesselness filter [28], which utilizes the eigenvalues of the Hessian matrix to increase multiscale tubular structures, which enhances their efficiency as a method of detection of blood vessels in the fundus images. The processed output with the vessel enhancement was binarized at a threshold value of 0.01, and subsequently deleted

unwanted pixels were deleted at the border of the image.

These preprocessing steps, including contrast enhancement, denoising, vesselness filtering and boundary removal, not only improve the detection of small pathological features, but they further provide a clean and robust input for downstream feature extraction, which substantially contributes to making the accuracy and prediction of DR severity in DR classification. Figure 4 illustrates the sequential outcomes of these operations.

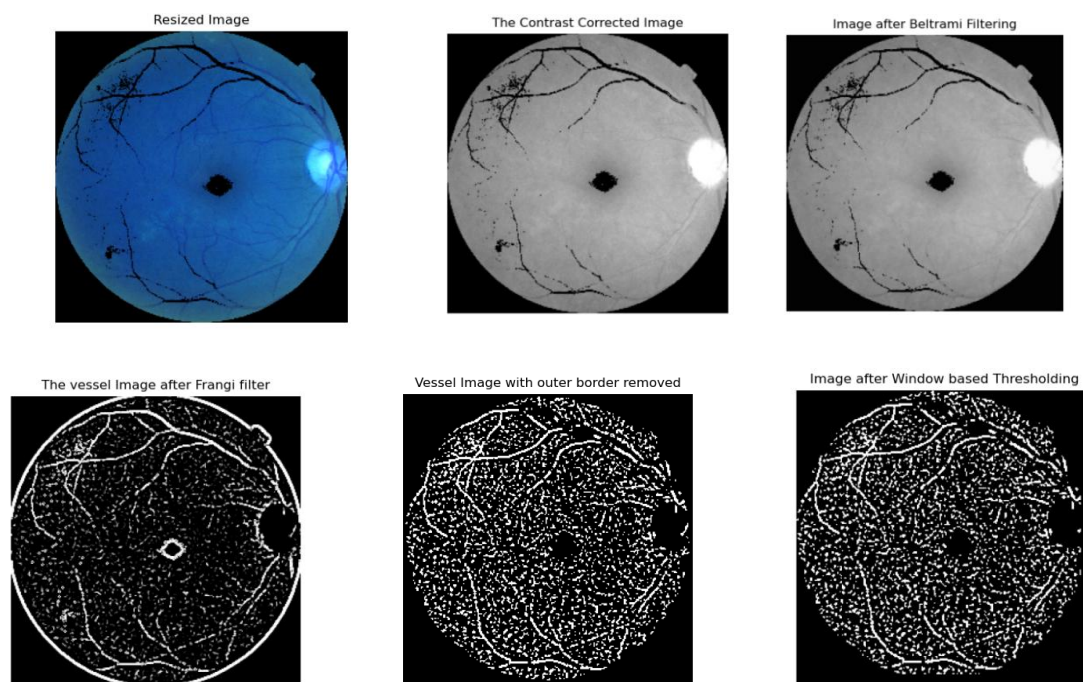


Figure 4 – The resized FOV, contrast-corrected image, Denoised image, filtered binarized image using Frangi filter, vessel image without border, and image after thresholding.

Compared to typical filtering, matched filtering allows for to identification of structures with a well-defined geometric structure in noisy images. 32 features were derived from the vessel image in 12 orientations (0° – 180°) and combinations of the length of filters [9, 11, 13, 15] and standard deviation [0.5, 1.0, 1.5, 2] with the Sobel operator and its gradient magnitude, capturing vessels with different sizes and directions. Another 32 features were obtained from the magnitude of first-order derivatives throughout the two dimensions of the image to improve the detection of changes in intensity across vessel orientations.

Textural features - varying, harmonious, micro-pattern, slow-changing, etc. were documented employing 6 GLCM features. In addition, the

local micro-textures were encoded with the aid of Uniform Local Binary Patterns (LBP) with a spatial radius of 3 and row- and column-wise features were collected, and 520 descriptive descriptors were achieved to efficiently identify fine retinal structures. Finally, the strength and structure of the textural variations (versus vertical and diagonal) were transformed in a 1-D discrete wavelet transform with six mother wavelets ['bior3.1', 'bior3.5', 'bior3.7', 'db3', 'sym3', 'haar'].

The features based on wavelet are found from the vertical W_v and diagonal W_d components of the wavelet transform, and they are computed using the unified formulation in (7):

$F =$

$$\frac{1}{i,j} \sum_{a=1}^i \sum_{b=1}^j \begin{cases} W_x, & \text{for magnitude} \\ |W_x|^2 & \text{for energy} \end{cases} \quad (7)$$

Where i and j denote the dimensions of each wavelet component. W_x represents either the vertical W_v or diagonal W_d wavelet components. The former captures the average strength of local variations, reflecting overall feature intensity, while the latter measures the local energy, emphasizing high-contrast regions and fine structural details, which are essential for detecting microaneurysms and subtle retinal abnormalities.

Consequently, 24 features capturing the average strength and energy of the image are incorporated into the fundus image representation vector. Additionally, edge orientation information from localized regions,

enhanced through preprocessing, is encoded using 324 Histogram of Oriented Gradients (HoG) features, providing robust descriptors of vessel and lesion structures. The gradient magnitude $\nabla I(x, y)$ at each pixel (x, y) can be computed using the finite difference approximation along both axes, as is given in expression (8):

$$G_x(x, y), G_y(x, y) = (I(x + 1, y) - I(x - 1, y), I(x, y + 1) - I(x, y - 1)) \quad (8)$$

The handcrafted features depicting the fundus image are described in Table 3. LBP generates the bulk of the total, where GLCM adds fewer yet useful descriptors. Matched filter features are extracted from the vessel image, whereas all other features are obtained from the grayscale FOV image.

Table 3 – Handcrafted Features employed for Retinal Image Representation

Feature Descriptor	Parameters	Representation dimension
Matched Filter (vessel image)	Sobel with Orientations	32
Matched Filter (vessel image)	Magnitude of I-order derivative	32
GLCM (grayscale FOV)	Six global features	6
LBP (grayscale FOV)	Row and Column Sum	520
Wavelet (grayscale FOV)	Magnitude and Energy using 6 different mother wavelets	24
HoG (grayscale FOV)	9 Orientations and 64 pixels per cell	324

3.5 Features using deep CNNs

The representation of the fundus image is subsequently enhanced with deep learning enriched with high-level semantic and abstract features derived from both the two pretrained CNNs - ResNet50 and VGG19. Both structures are highly competent feature extractors from medical image analysis algorithms, respectively. ResNet50 is effective at extracting discriminative features, with better generalization capabilities, whereas VGG19 performs better in image extraction of hierarchical texture and shape information. For feature extraction, we remove the top classification layers in both networks and resize the input fundus images to 224×224 pixels and scale the pixel values as normal for an ideal feature quality. ResNet50 gives 2048-dimensional feature vectors, and VGG19 returns 512-dimensional representations.

Thus, this unique feature vector fuses discriminative global features, such as ResNet50 and hierarchical textural/shape descriptions, VGG19, to enrich fundus image representation for robust DR classification that utilizes pretrained semantic knowledge.

The combined 2560-dimensional high-level feature vector F_{CNN} can be represented by expression (9) as follows:

$$F_{CNN} = [Res50_{2048}(Norm(FI_{224 \times 224})) || Vgg19_{512}(Norm(FI_{224 \times 224}))] \quad (9)$$

Here, the input image is adjusted as per the network requirement. That is, the FI is resized to 224×224 ($FI_{224 \times 224}$). Further, the pixel intensities in the image are normalized ($Norm(.)$) to obtain better quality of features. Two pretrained networks (Resnet50 and VGG19) are

incorporated for representing the FI, allowing their capabilities to extract high and low-level features from the image. A total of 2048 and 512 F_{CNN} features are used, respectively, from both networks, separately contributing $2048+512 = 2560$ textural attributes using the concatenation (“||”).

The “blind” deep features incorporated with handcrafted descriptors, i.e., matched filter, GLCM, LBP, wavelet, and HoG features, improve the richness of the image representation. Handcrafted elements provide interpretable, clinically relevant cues to local textures, vessel structures, and microlesions, while deep features encode high-level, abstract patterns that traditional descriptors did not capture explicitly. This fusion boosts the robustness, generalization, and discriminative power of the feature set, therefore improving DR severity classification ability among a wide range of fundus images.

3.6 The classifier

The complete feature set, excluding the CNN-based features and features based on contrast, was max-normalized. The work considers binary classification by combining samples from the first two DR grades versus a combination of the remaining two DR grades. The approach carried out in [29] was followed for a compatible comparison of the evaluation. Combining two DR grades yielded 699 samples in the first category, while 501 were in the second category. The class imbalance and scarce data sample issues were compensated for using data

augmentation. The augmentation increased the class 0 and 1 samples to 4194 and 3006, respectively. Owing to the complexity of the Messidor dataset as reported in several studies, only 10% samples were considered for assessment, while the remaining samples were used for training the RNN network.

This classification model adopts a Recurrent Neural Network (RNN), which captures sequential dependencies and relationships between heterogeneous features and is suited to incorporate handcrafted, contrast-based, and deep features. The network was trained for 1,500 epochs with a batch size of 128, and model weights were optimized using Stochastic Gradient Descent (SGD), chosen for its low memory footprint and strong generalization capabilities. The final aggregated feature vector was developed employing patch-based contrast (1,155), handcrafted features (938), and deep features of ResNet50 (2,048) and VGG19 (512), yielding a total dimension of 4,653. In order to minimize computation while preserving salient information, Principal Component Analysis (PCA) was utilized, which resulted in a compact 600-dimensional representation.

Performance was assessed via standard metrics such as accuracy, sensitivity, specificity, and F1-score. The RNN-based integration of multiple feature types allows for a resilient DR severity classification system that utilizes interpretable handcrafted cues and high-level semantic patterns captured by pretrained networks, leading to enhanced generalization and clinical utility (Figure 5).

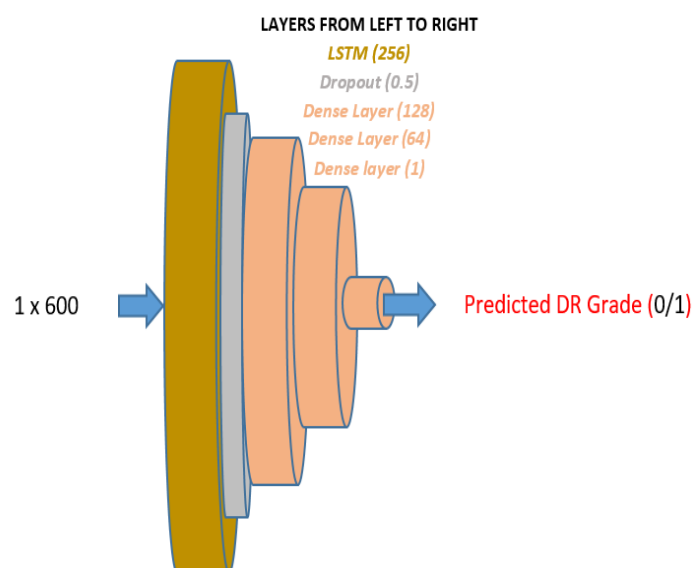


Figure 5 – DR severity Grade classification network

4. Experimental Results and Discussion

The performance of the proposed DR severity grade classification framework is detailed in Table 4 in comparison with existing state-of-the-art approaches. The experimental arrangement follows previous studies, with grades 0 and 1 merged into the normal class, and grades 2 and 3 grouped into the DR class, for a fair evaluation. Performance details for competing methods were retrieved directly from the respective reports, with ‘-’ representing unavailable results. Our method achieved binary classification accuracy of 96.11%, quite close to the 96.34% reported in [29]. Moreover, Zhang et al. [8] maintain the highest accuracy in binary DR classification, but the attention mechanism in their EfficientNet-based model leads to an additional computational cost, potentially limiting real-time applicability in resource-

constrained clinical settings. Likewise, the hybrid attention design (channel + multi-head self-attention) in [29] results in architectural complexity and computational cost, which may impede deployment on less powerful hardware.

The proposed framework, while never claiming superiority over the best-performing methods, can compete on performance levels. A robust and interpretable solution combining a tailored RNN with widely ranging handcrafted, patch-based contrast, and deep features minimizes the compromise of accuracy and computational power. It demonstrates the potential of such a method of DR screening within a practical setting where it is desirable to achieve a well-balanced performance-resource trade-off in applications with constraints imposed by deployment.

Table 4 – Comparison with other competing DR classification models

Model and Reference	Year	% Sensitivity (Sn)	% Specificity (Sp)	% F1-score (F1)	% Accuracy (A)
Attention mechanism plus Efficient Network [8]	2020	95.60	98.70	95.20	97.20
Statistical method plus Bayesian classifier [4]		91.94	88.24	92.44	93.44
Integration of segmentation into classification [5]	2022	90.00	-	91.00	92.23
Asymmetric deep learning features[3]	2023	87.04	95.67	-	94.80
Transfer learning and attention mechanism[10]		88.75	89.63	97.00	83.69
Cascaded hybrid attention module [29]	2024	95.94	94.24	95.36	96.34
Hybrid Deep Learning [30]	2025	-	98.31	97.71	94.04
Proposed approach		95.95	94.20	95.35	96.11

The proposed DR classification framework was evaluated over 20 independent runs, with random shuffling of training and testing samples in each iteration, and the average performance metrics were reported. For the highest performance configuration (described in Table 4), the corresponding training accuracy and loss curves are illustrated in Figures 6 and 7. Across almost all runs, the training accuracy was near 100% while the training loss remained consistently low, approaching zero,

demonstrating the effective convergence of the model. The separation of DR grades 1 and 3 is more distinctly shown in Table 5 than that of grades 0 and 2, which contributes to the overall binary classification accuracy of 96.11% when the classification system classes 0–1 as normal, while grades 2–3 is grouped as DR (Table 4). The confusion matrix shown in Figure 8 also elucidates the classification response, indicating that DR grade 3 can be clearly distinguished from the aggregate set of grades 0–2, albeit with

occasional classification errors on adjacent grades suffering similar pathological characteristics. These conclusions highlight the difficulty of separating early-stage DR from

healthy images and validate the performance trends we observe in the binary classification system.

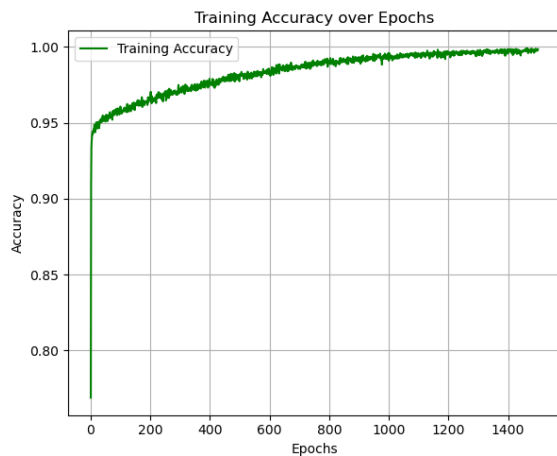


Figure 6 – Training accuracy

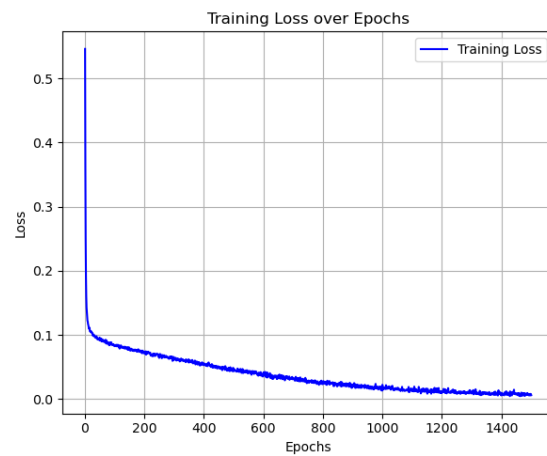


Figure 7 – Training loss

Table 5 – Performance of Binary classification using individual DR grade samples

Binary categories	Sn	Sp	F1	A
Grade 0	93.06	95.83	94.37	94.44
Grade 1	97.22	96.94	97.09	97.08
Grade 2	96.11	95.56	95.84	95.83
Grade 3	97.50	98.06	97.77	97.78

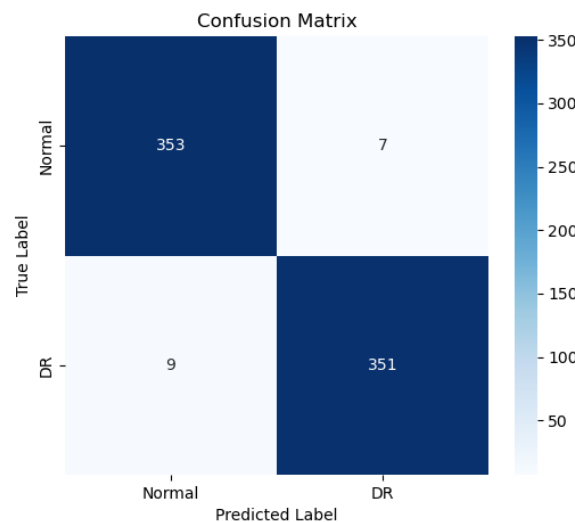


Figure 8–Confusion matrix for configuration DR grades 0-2 versus DR grade 3.

In recent years, DR grading has gained use through AI-based techniques and has shown itself to be an important tool for ophthalmic

diagnostics. Previous methods used handcrafted features pulled from diverse descriptors, whilst modern deep learning-based approaches account

for a large share of work. These approaches tend to largely rely on high-level semantic descriptors or necessitate detailed lesion annotations at the pixel or patch level, which results in suboptimal diagnostic performance or adds strain to ophthalmologists. As a result, there are still tremendous obstacles to discovering an efficient and reliable method of DR grading. In this work, we tackled this issue by aiming to derive strong, representative features with less dependence on sophisticated classifier architectures. An efficient, lightweight, custom Recurrent Neural Network (RNN) was built to capture the semantic associations and classify fundus images for proper DR grades.

The proposed framework harmonizes the global and local image statistics in computational speed and is capable of obtaining good results and a low-cost network. The performance of the model on the binary classification task was competitive with other state-of-the-art methods (taken together in Table 4). Notwithstanding the promising outcomes, there were still aspects to be further improved. First, while data augmentation enhanced the Messidor dataset from 1,200 to 7,200 images, class imbalance remains and may harm generalization.

Second, this study is limited to a reliable DR grading model and does not propose a fully deployable clinical system. Third, despite its benefit derived from an array of micro- and macro-descriptors for extracting clinically relevant features, the framework could also be adapted to classify other retinal pathologies, making it more clinically valuable.

References

- Holl'o, G., Aung, T., Cantor, L. B., & Aihara, M. (2020). Cystoid macular edema related to cataract surgery and topical prostaglandin analogs: mechanism, diagnosis, and management. *Survey of Ophthalmology*, 65(5), 496–512.
- Tawfik, A., Pistilli, M., Maguire, M., Chen, Y., Yu, Y., Greiner, J., Asbell, P., & Ying, G. S. (2023). Association of dry eye symptoms and signs in patients with dry eye disease. *Ophthalmic Epidemiology*, 1–9.
- Jena, P. K., Khuntia, B., Palai, C., Nayak, M., Mishra, T. K., & Mohanty, S. N. (2023). A novel approach for diabetic retinopathy screening using asymmetric deep learning features. *Big Data and Cognitive Computing*, 7(1), 25.
- Kang, Y., Fang, Y., & Lai, X. (2020). Automatic detection of diabetic retinopathy with statistical method and Bayesian classifier. *Journal of Medical Imaging and Health Informatics*, 10(5), 1225–1233.
- Andersen, J., Hubel, M., Rasmussen, M., Grauslund, J., & Savarimuthu, T. (2022). Automatic detection of abnormalities and grading of diabetic retinopathy in 6-field retinal images: Integration of segmentation into classification. *Translational Vision Science and Technology*, 11, 19.

5. Conclusion

Here, we established a robust binary DR grading framework using a combination of diverse handcrafted features (vessel, texture, wavelet, HOG, LBP descriptors) combined with patch-based contrast and blind quality features, along with a customized RNN. By effectively capturing global and local retinal characteristics, the RNN exploited temporal-like dependencies to perform discriminative learning. Model classification results obtained competitive performance (accuracy-wise 96.11%, sensitivity-wise 95.95%, specificity-wise 94.20%, F1-score = 95.35%) based on the DR grading structure, showing better separation of DR grades with distinct pathological features. This framework is scalable and low-cost, combining interpretable handcrafted descriptors with deep learning to allow for automated DR screening, decreasing the workload of ophthalmologists and providing opportunities for earlier intervention. The proposed approach will be applied in future efforts to multi-class DR grading, multimodal retinal imaging, as well as to support the validation of its performance across a wide range of populations and clinical settings to help facilitate translational adoption in practical real-world teleophthalmology systems. In addition, multiple imaging devices and multiple populations will need to be broadly validated to characterize robustness and generalizability.

Conflict of interest – The authors have no conflict of interest to declare.

Funding – The work is not funded by any organization or agency.

Data Availability - Data sharing does not apply to this article as no datasets were generated during the current study. The data was obtained from Messidor, which is publicly available.

- Mukherjee, N., & Sengupta, S. (2023). A hybrid CNN model for deep feature extraction for diabetic retinopathy detection and gradation. *The International Journal on Artificial Intelligence Tools*, 32(8), Article 2350036.
- Malhi, A., Grewal, R., & Pannu, H. S. (2023). Detection and diabetic retinopathy grading using digital retinal images. *International Journal of Intelligent Robotics and Applications*, 7, 426–458.
- Zhang, Z., Liu, M., & Zhu, D. (2020). Automatic recognition and classification of diabetic retinopathy images by combining an attention mechanism and an efficient network. *Journal of Image and Graphics*, 25(8), 1708–1718.
- Ramos, J. M. A., Perd'omo, O., & Gon'zalez, F. A. (2022). Deep semi-supervised and self-supervised learning for diabetic retinopathy detection. *arXiv preprint arXiv: 2208.02408*.
- Dinpajhoush, M., & Seyyedsalehi, S. A. (2023). Automated detecting and severity grading of diabetic retinopathy using transfer learning and attention mechanism. *Neural Computing & Applications*, 1–13.
- Ashwini, K., & Ratnakar, D. (2023). Grading diabetic retinopathy using multiresolution based CNN. *Biomedical Signal Processing and Control*, 86, Article 105210.
- Khanna, M., Singh, L. K., Thawkar, S., & Goyal, M. (2023). Deep learning based computer-aided automatic prediction and grading system for diabetic retinopathy. *Multimedia Tools and Applications*, 82, 39255–39302.
- Indresh Kumar Gupta, Shruti Patil, Supriya Mahadevkar, Ketan Kotecha, Awanish Kumar Mishra, and Joel J. P. C. Rodrigues, "Retinal fundus imaging-based diabetic retinopathy classification using transfer learning and fennec fox optimization," *MethodX*, 14, 2025, 103232.
- Zhang L, Gang J, Liu J, Zhou H, Xiao Y, Wang J, Guo Y. Classification of diabetic retinopathy algorithm based on a novel dual-path multi-module model. *Med Biol Eng Comput*. 2025 Feb;63(2):365-381. doi: 10.1007/s11517-024-03194-w. Epub 2024 Sep 25. PMID: 39320579.
- Ankur Biswas and Rita Banik, "Advancing diabetic retinopathy classification using ensemble deep learning approaches," *Biomedical Signal Processing and Control*, Volume 106, August 2025, 107804.
- Uppalapati Srilakshmi, K. Vinay Kumar, Sirisha Korimilli, Shiramshetty Goutham, Jose Mary Golamari, and Putta Brundavani, "Data Augmentation-Based Diabetic Retinopathy Classification and Grading with the Dynamic Weighted Optimization Approach," *Technologies and Their Effects on Real-Time Social Development*, Volume 167, No. A2(S), 2025.
- Feng, M.; Cai, Y.; Yan, S. Enhanced ResNet50 for Diabetic Retinopathy Classification: External Attention and Modified Residual Branch. *Mathematics* 2025, 13, 1557.
- Ahmed, Faisal, "OcuViT: Automated Detection of Diabetic Retinopathy and AMD Using a Hybrid Vision Transformer Approach," (February 20, 2025). Available at SSRN: <https://ssrn.com/abstract=5166835> or <http://dx.doi.org/10.2139/ssrn.5166835>
- N. S. Kavitha and N. Kasthuri, "An efficient automatic diabetic retinopathy grading using a two-way cascaded convolution neural network," *Discover Computing*, 28:7, 2025.
- Prakash J., and Vinoth Kumar, "An ensemble approach for classification of diabetic retinopathy in fundus image," *Multimedia Tools and Applications*, 84, 10567–10586 (2025).
- Chen Zhang, Shaoqi Dong, Ziyun Song, Liming Liu, Jiaxu Ning, Bin Zhang, and Changsheng Zhang, "A two-stage deep learning based method for diabetic retinopathy classification," *Connection Science*, Volume 37, No. 1, 2025, 2507182.
- <https://www.adcis.net/en/third-party/messidor/>
- Tadmor Y. and Tolhurst D., "Calculating the contrasts that retinal ganglion cells and LGN neurons encounter in natural scenes," *Vision Research*, volume 40(22), 2000, pp. 3145–3157.
- Gabriele Simone, Marius Pedersen, Jon Yngve Hardeberg, "Measuring perceptual contrast in digital images," *Journal of Visual Communication and Image Recognition*, Volume 23, 2012, pp. 491-506.
- Rodieck, R. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5, 583–601.
- Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, 187, 517552.
- Wetzler and Kimmel R. (2012). Efficient Beltrami flow in patch space. In *Scale Space and Variational Methods in Computer Vision*, volume 6667 of *Lecture Notes in Computer Science*, pp. 134-143. Springer Berlin/Heidelberg.
- Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. (1998). *Multiscale vessel enhancement filtering*. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI'98* (pp. 130–137). Springer, Berlin, Heidelberg.

Yanxia Liu, "Automatic grading diabetic retinopathy in color fundus image: Cascaded hybrid attention module," Journal of Radiation Research and Applied Sciences, 17, 2024, 101160.

Mishmala Sushith, A. Sathiya, V. Kalaipoonguzhali & V. Sathya, "A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images," Scientific Reports, (2025) 15:15166.