



OPEN A deep learning based model for diabetic retinopathy grading

Samia Akhtar¹, Shabib Aftab¹✉, Oualid Ali², Munir Ahmad^{3,4}, Muhammad Adnan Khan⁵✉, Sagheer Abbas⁶ & Taher M. Ghazal⁷

Diabetic retinopathy stands as a leading cause of blindness among people. Manual examination of DR images is labor-intensive and prone to error. Existing methods to detect this disease often rely on handcrafted features which limit the adaptability and classification accuracy. Thus, the aim of this research is to develop an automated and efficient system for early detection and accurate grading of diabetic retinopathy severity with less time consumption. In our research, we have developed a deep neural network named RSG-Net (Retinopathy Severity Grading) to classify DR into 4 stages (multi-class classification) and 2 stages (binary classification). The dataset utilized in this study is Messidor-1. In preprocessing, we have used Histogram Equalization to improve image contrast and denoising techniques to remove noise and artifacts which enhanced the clarity of the fundus images. We applied data augmentation techniques to our preprocessed images in order to tackle class imbalance issues. Augmentation techniques involve flipping, rotation, zooming and adjustment of color, contrast and brightness. The proposed RSG-Net model contains convolutional layers to perform automatic feature extraction from the input images and batch normalization layers to improve training speed and performance. The model also contains max pooling, drop out and fully connected layers. Our proposed RSG-Net model achieved a testing accuracy of 99.36%, specificity of 99.79% and a sensitivity of 99.41% in classifying diabetic retinopathy into 4 grades and it achieved 99.37% accuracy, 100% sensitivity and 98.62% specificity in classifying DR into 2 grades. The performance of RSG-Net is also compared with other state-of-the-art methodologies where it outperformed these methods.

Keywords Diabetic retinopathy, Convolutional neural network, Deep learning, Augmentation, Optimization algorithm

Diabetes mellitus is a chronic medical condition where the body either doesn't produce enough insulin or cannot use it effectively which leads to high blood sugar levels¹. This disease has become a global health concern affecting millions of people worldwide. According to the International Diabetes Federation, as of 2021, approximately 537 million adults are living with diabetes and this number is expected to increase significantly in the coming years². Diabetes puts a heavy strain on healthcare systems and can lead to serious complications if not properly managed. One such complication is Diabetic Retinopathy (DR). Among those with diabetes, approximately one-third develop DR.

Diabetic retinopathy is a micro vascular complication of diabetes that affects the eyes³. It happens when high blood sugar levels damage the blood vessels in the retina. Retina is the part of the eye that detects light and sends signals to the brain. Over time, excessive glucose weakens the vessel walls causing microaneurysms, fluid leakage, and swelling of the retinal tissues⁴. As the disease advances, blocked vessels restrict blood flow which deprives the retina of oxygen and nutrients. This triggers the growth of new abnormal blood vessels in later stages. These fragile vessels may rupture and lead to hemorrhages and retinal scarring or detachment which progressively worsen vision. Without timely treatment, this damage caused by DR leads to severe vision loss and eventually blindness⁵.

As the disease progresses, symptoms may include blurred or fluctuating vision, dark or empty areas in the visual field, poor night vision, and seeing spots or floaters⁶. In advanced stages, the condition can lead to

¹Department of Computer Science, Virtual University of Pakistan, Lahore 54000, Pakistan. ²Computer Sciences Department, College of Arts & Science, Applied Science University, P.O.Box 5055, Manama, Kingdom of Bahrain.

³School of Computer Science, National College of Business Administration and Economics, Lahore 54000, Pakistan.

⁴University College, Korea University, Seoul 02841, Republic of Korea. ⁵Department of Software, Faculty of Artificial Intelligence and Software, Gachon University, Seongnam-si 13557, South Korea. ⁶Department of Computer Science, Prince Mohammad Bin Fahd University, Al-Khobar, Dhahran 34754, Saudi Arabia. ⁷Department of Networks and Cybersecurity, Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan.

✉email: shabib.aftab@vu.edu.pk; adnan@gachon.ac.kr

complete vision loss if not properly managed. Some prominent symptoms of DR include soft/hard exudates (EX), micro aneurysms (MA), hemorrhages (HM) and neovascularization⁷.

Early screening of diabetic retinopathy is crucial in preventing severe complications⁸. It is essential for detecting diabetic retinopathy before it progresses to more advanced stages where treatment becomes complex and less effective. During the early stages of the disease (when symptoms are absent or mild), timely intervention can prevent severe vision loss by addressing the problems before irreversible damage occurs. Physicians can slow or halt disease progression through interventions like better glycemic control or laser therapy.

This reduces the risk of patients developing proliferative diabetic retinopathy or macular edema. Thus, prioritizing early screening and leveraging advanced technologies can significantly reduce the burden of diabetic retinopathy, enhance patient outcomes, and prevent the vision loss that affects millions of individuals worldwide. Regular eye exams are recommended for individuals with diabetes, as early-stage DR often lacks noticeable symptoms. Physicians typically use retinal scans, such as fundus photography and optical coherence tomography (OCT), to detect changes in the retina^{9–11}. These imaging techniques allow for a detailed view of the retinal structure and any abnormalities that may indicate the presence of diabetic retinopathy.

The symptoms of diabetic retinopathy can vary depending on the stage and severity of the disease. Diabetic retinopathy (DR) has two primary stages: Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR)¹². NPDR can be classified into mild, moderate, and severe stages. In mild NPDR, microaneurysms develop in the retinal blood vessels; however, these changes often go unnoticed as other significant symptoms are typically absent. As the disease advances to moderate NPDR, blocked blood vessels can cause blurred or fluctuating vision and clear signs of microaneurysms are present. Severe NPDR leads to more significant blood vessel blockage which leads to the presence of exudates and retinal hemorrhages. In PDR, abnormal blood vessels grow on the retina, leading to severe complications like vitreous hemorrhage, retinal detachment, and potential blindness if left untreated.

The interpretation of retinal scans is a critical aspect of detecting diabetic retinopathy¹³. Traditionally, ophthalmologists and trained specialists examined these images manually to identify signs of the disease¹⁴. However, the increasing prevalence of diabetic retinopathy has highlighted the need for more efficient and accurate diagnostic tools. This necessity has driven the development of computer-aided diagnosis (CAD) systems, which utilize artificial intelligence (AI), deep learning and machine learning algorithms to analyze retinal images^{15,16}.

Deep learning (DL) has significantly transformed the landscape of DR screening and diagnosis. By training neural networks on large datasets of retinal images, these systems can learn to identify patterns and abnormalities associated with various stages of diabetic retinopathy¹⁷. Furthermore, deep learning and AI have proven beneficial in detecting various diseases globally including glaucoma, age-related macular degeneration and diabetic macular edema. DL has enhanced diagnostic capabilities across multiple specialties^{18–20}. Studies have shown that deep learning-based CAD systems can achieve high sensitivity and specificity in detecting diabetic retinopathy reducing the risk of human error and enhancing the efficiency of screening programs²¹.

The integration of AI and deep learning into diabetic retinopathy screening has several advantages over traditional methods. Firstly, these technologies can process vast amounts of data quickly. This enables large-scale screening programs that are essential for early detection and intervention^{22–24}. Secondly, deep learning models continuously improve as they are exposed to more data, surpassing human capabilities in identifying subtle retinal changes. Lastly, the use of CAD systems can alleviate the workload of ophthalmologists, allowing them to focus on complex cases and patient care. In remote and underserved areas, where there may not be enough specialized doctors, these technologies can be used through telemedicine platforms. With the help of AI, even healthcare workers who are not specialists can perform initial screenings²⁵.

Research contributions

This research proposes a framework for 4-stage and 2-stage classification of diabetic retinopathy using deep learning. We have developed our own Convolutional Neural Network (CNN) model named (Retinopathy Severity Grading) RSG-Net. The motivation behind developing RSG-Net stems from the limitations of methods.

that often rely on handcrafted features which reduce adaptability and classification accuracy. Some existing approaches struggle to handle class imbalance issues. RSG-Net addresses these gaps by offering a more accurate and automated solution for diabetic retinopathy severity grading. It automates diabetic retinopathy detection by extracting features and identifying patterns in retinal images without the need for manual examination. It rapidly analyzes and localizes abnormalities thus, enables faster, consistent and more accurate diagnosis. By addressing class imbalance through data augmentation and enhancing image quality with the Histogram Equalization (HE) technique, our model achieved improved generalization ensuring better performance across varying conditions and class distributions. The dataset for training and testing the model used in this research is Messidor-1. It contains a total of 1200 fundus images labeled by the experts into 4 grades from 0 to 3. Various preprocessing techniques are applied to prepare the dataset for more accurate model training. Data augmentation is applied to up-sample the minority classes. The main contributions of this research are:

1. The noise in the images was removed by denoising technique called Gaussian Blur. Histogram Equalization (HE) was applied for improving overall image contrast along with making the fine details of the features clearer. These techniques played a major role in helping the CNN model train well on quality images.
2. Augmentation techniques were used to address the skewness in the dataset. Number of samples were increased in the minority classes to prevent the model from giving biased results.
3. A deep learning CNN, RSG-Net is developed for grading diabetic retinopathy. This model contains convolutional and pooling layers to extract features from retinal images. Then it classifies the images using fully

- connected layers. We have added batch normalization layers to improve the training speed, stability and performance of the model.
- The performance of our model is compared with other state-of-the-art techniques. Despite the simplicity, RSG-Net achieved superior performance and accuracy in both, 4-stage and 2-stage classification compared to other complex techniques.

Research organization

This research is organized as follows: “Related work” section provides an overview of previous literature on the detection and grading of diabetic retinopathy. “Methodology” section details the proposed classification framework. “Results and discussion” section analyzes the results of the suggested methodology and compares them with other state-of-the-art methodologies using performance evaluation metrics. Finally, “Conclusion” section concludes the study and offers insights into future work.

Related work

Many researchers have dedicated their efforts to experimenting with diverse techniques and methods to detect diabetic retinopathy. Their dedication has significantly advanced the prediction of this disease.

In²⁶, the researchers proposed a system to classify diabetic retinopathy into five categories using automated diagnostic techniques. They integrated two convolutional neural network models, CNN512 and a YOLOv3-based CNN to handle DR stage classification and lesion localization. The study utilized the APTOS 2019 and DDR datasets. It implemented preprocessing techniques such as CLAHE, noise removal, cropping, and augmentation before training the models. CNN512 was specifically trained for classification tasks, while the YOLOv3 model was used for detecting lesions. This combined approach improved the classification accuracy to 89%, achieving a sensitivity of 89% and a specificity of 97.3%. In²⁷, the authors introduced a deep transfer learning framework for the automatic classification and detection of diabetic retinopathy stages. They utilized the APTOS 2019 Blindness Detection dataset, divided it into training and testing sets and then applied augmentation techniques to the training data. The preprocessing involved five steps including Gaussian Filtering, cropping, rescaling, and the application of CLAHE (Contrast Limited Adaptive Histogram Equalization). The preprocessed images were then fed into three models: ResNet152, DenseNet201 and VGGNet19. Among these models, DenseNet201 achieved the highest test accuracy of 82.7% while ResNet152 attained the highest AUC value of 94.1%.

The authors of²⁸ proposed a binary classification framework for detecting diabetic retinopathy using transfer learning. They combined PySpark with deep learning leveraging the capabilities of Big Data tools to process the IDRiD dataset (Indian Diabetic Retinopathy Image Dataset). For preprocessing, the dataset underwent cleaning where dark images were removed followed by cropping to eliminate unwanted spaces and resizing. The classification was performed using a logistic regression (LR) classifier. Authors employed DL Pipelines on Apache Spark for this purpose, splitting the dataset into 80% for training and 20% for testing. Results indicated that among the three transfer learning models utilized, InceptionV3 demonstrated the best performance achieving an accuracy of 95%, an AUC of 94.98%, and an F1-Score of 95%. Authors of²⁹ introduced a binary classification framework for diabetic retinopathy detection using Convolutional Neural Networks (CNNs). They adopted the VGG (Visual Geometry Group) architecture which consists of four blocks, each containing a

convolutional layer followed by maximum pooling layers. ReLU was used as the activation function in these layers while the output layer employed the sigmoid function for binary classification. The loss function employed was binary cross-entropy with Adam serving as the optimizer. The dataset was sourced from diabetic-retinopathy-classified via Kaggle with image augmentation applied to enhance diversity and generalization. The proposed model in this study achieved an accuracy of 95.5%.

In³⁰, a blockchain-based healthcare framework is proposed for the detection of diabetic retinopathy using deep learning. The research utilizes the publicly available IDRiD dataset from IEEE DataPort. Preprocessing of the dataset involves median filtering followed by lesion segmentation. Hyperparameter tuning is conducted using the Taylor African Vulture Optimization (AVO) algorithm. The most relevant features are then selected and input into the SqueezeNet classifier to predict the occurrence of diabetic retinopathy. The resulting output is securely stored within the blockchain architecture which is accessible to the Electronic Health Record (EHR) manager. Comparative analysis against previous research demonstrated that the proposed model outperformed achieving an accuracy of 94.2%, sensitivity of 94.8%, and specificity of 93.4%. In their paper³¹, researchers introduced an ensemble learning framework for categorizing the severity of diabetic retinopathy. Their methodology incorporates a gray level intensity algorithm alongside a decision tree-based ensemble learning approach. The APTOS 2019 dataset is utilized in their research. Before classification, the authors implemented several preprocessing techniques including feature extraction and selection. Images were resized and to tackle class imbalance issues, augmentation techniques were applied. Additionally, sharpening and contrast enhancement methods were also employed. For classification, the ensemble algorithm XGBoost was utilized. This ensemble approach demonstrated promising outcomes achieving an accuracy of 94.20%, along with an F-measure of 93.51% and a recall of 92.69%.

In their work³², researchers introduced a computer-aided diagnostic (CAD) system tailored for detecting non-proliferative diabetic retinopathy through convolutional neural networks (CNNs). Their CNN is optimized for optical coherence tomography (OCT) imaging. They trained the model on an OCT dataset implementing crucial preprocessing steps to extract retina patches for CNN training. Transfer learning principles and effective feature combination techniques were also employed to improve performance. Utilizing the AlexNet CNN with an input size of 227×227 , the methodology achieved the highest accuracy with minimal computational complexity. By combining output features from two independently trained CNNs, the system attained 94% accuracy, 100% recall, and 88% specificity. In³³, the authors introduced three distinct hybrid models for classifying diabetic

retinopathy (DR) into five classes: Hybrid-a, Hybrid-f, and Hybrid-c. These hybrids models integrate five base CNN models: NASNetLarge, EfficientNetB5, EfficientNetB4, InceptionResNetV2, and Xception. Two loss functions are used to train the base models and then their outputs train the hybrid models. Experiments are conducted on three datasets: APTOS, EyePACS, and DeepDR. Preprocessing involves both initial image enhancement techniques and those applied during training. Among these hybrid models, Hybrid-c achieved the highest accuracy of 86.34%.

In³⁴, the authors introduce an innovative approach for diabetic retinopathy (DR) classification based on deep learning. They propose a hybrid model combining ResNet and GoogleNet for feature extraction, enhanced by adaptive particle swarm optimization (APSO). These features are then classified using machine learning techniques such as SVM, random forest, decision tree, and linear regression. The study uses the EyePACS dataset and employs preprocessing steps like image resizing, green channel extraction, and top-hat/bottom-hat transformations. Their hybrid model achieved an impressive 94% accuracy, surpassing existing binary and multiclass DR detection techniques. Future work includes testing on diverse datasets to ensure broader applicability. In³⁵, researchers tackle the challenges associated with determining the severity of Diabetic Retinopathy (DR) by presenting a novel convolutional neural network (CNN) architecture known as the attention-guided CNN (AG-CNN) which was characterized by its dual-branch design. They utilize the APTOS 2019 DR Dataset, emphasizing the necessity of initial preprocessing to enhance image quality. The AG-CNN features a global branch for overall attention and a local branch that targets important localized features. Their experimental results indicate that the baseline model, DenseNet-121, achieves an accuracy of 97.46% and an AUC of 0.995, while the AG-CNN significantly improves these metrics, reaching an accuracy of 98.48% and an AUC of 0.998.

In³⁶, researchers introduced a deep learning framework designed for the detection and classification of diabetic retinopathy stages using a single fundus retinal image. Utilizing transfer learning, the study fine-tunes two advanced pre-trained models, ResNet50 and EfficientNetB0, on a comprehensive multi-center dataset which includes the APTOS 2019 dataset. Various image augmentation techniques were applied during preprocessing to enhance image quality and increase dataset diversity. The proposed model achieved a notable 98.50% accuracy in binary classification, with sensitivity at 99.46% and specificity at 97.51%. In stage grading, it recorded an accuracy of 89.60% and a quadratic weighted kappa of 93.00%. This dual-branch transfer learning approach proved to be a robust tool for the detection and classification of diabetic retinopathy greatly enhancing clinical decision-making and patient care. In³⁷, researchers developed DeepDR Plus, a deep learning model aimed at predicting diabetic retinopathy (DR) progression over five years using fundus images. The model was pre-trained on 717,308 images from 179,327 diabetic individuals and validated with a diverse set of 118,868 images. It achieved impressive concordance indexes between 0.754 and 0.846, indicating strong predictive capability. According to researchers, integrating this system into clinical practice could extend average screening intervals from 12 months to approximately 32 months with only a 0.18% rate of delayed detection of vision-threatening DR. The model focused on high-risk patients utilizing features such as retinal vascular patterns and foveal characteristics. Authors of³⁸ have presented a deep learning methodology for evaluating the quality of ultra-widefield optical coherence tomography angiography (UW-OCTA) images which is essential for diabetic retinopathy analysis. Due to the scarcity of UW-OCTA datasets and high associated costs, the researchers first pre-trained a vision transformer (ViT) model on smaller OCTA images followed by fine-tuning it on larger UW-OCTA datasets. The proposed approach outperformed traditional methods, achieving an area under the curve (AUC) of 0.9026 and a Kappa value of 0.7310. This innovative method enhanced the precision and efficiency of image quality assessments and effectively addressed the challenges in clinics that may lack specialized personnel. A significant future application of this system included its integration into UW-OCTA devices. It allows instant notifications for image re-acquisition when substandard images are detected, ultimately improving the diagnosis and management of retinal disease.

Unlike many previous methodologies that focused on either binary or multi-class classification, our RSG-Net effectively handles both tasks offering a more versatile approach. While several studies relied on pre-trained models or complex/hybrid CNN architectures, we developed our CNN from scratch achieving superior performance with a simpler architecture. This design not only enhances computational efficiency but also ensures high accuracy. In contrast to other models, our preprocessing techniques including Histogram Equalization and denoising improved image clarity for better feature extraction. Data augmentation tackled class imbalance while batch normalization boosted training stability and speed contributing to a more robust performance.

Methodology

This study introduces a framework for grading diabetic retinopathy into four and two classes using a convolutional neural network (CNN) named RSG-Net. Multi-class classification aids in grading DR severity for precise treatment while binary classification identifies DR presence for large-scale screening. In our study, we performed both tasks—multi-class and binary classification—within the same framework but as distinct tasks. These two tasks were carried out independently using the same CNN architecture (RSG-Net) but each was trained and evaluated separately. This allowed us to leverage the same model's strengths while addressing both classification needs with high accuracy. The framework involves a benchmark dataset to train and test the performance of our model. The proposed framework is illustrated in Fig. 1. It systematically outlines each step involved in the process.

Dataset description

Messidor-1 is a benchmark dataset publicly available at ADCIS³⁹. It comprises around 1200 color fundus images that are carefully annotated by the expert ophthalmologists. The dataset consists of images collected from three ophthalmologic departments. About 800 images were captured with pupil dilation and 400 were captured

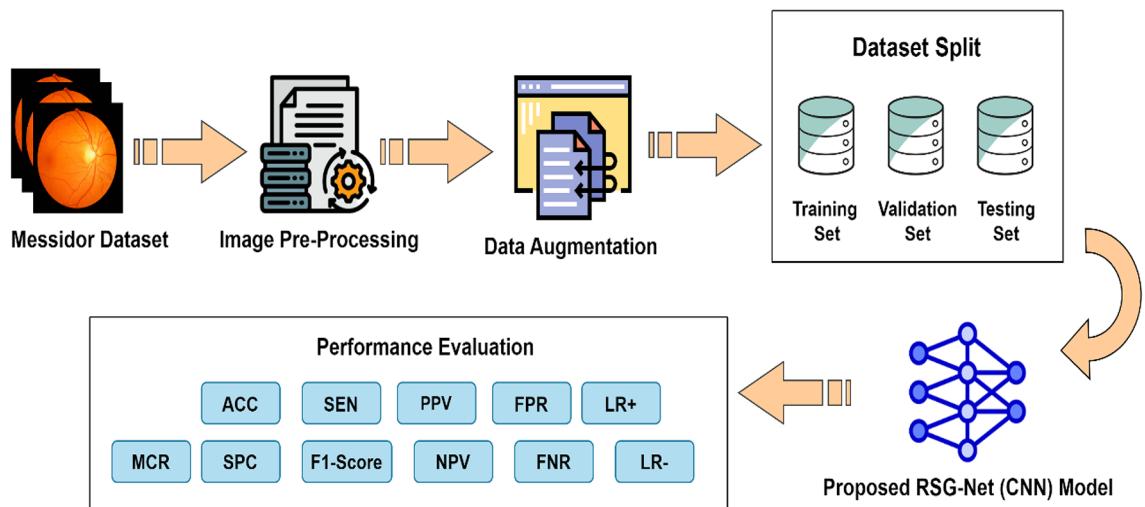


Fig. 1. Proposed framework of this research.

Retinopathy stage	Grade/level	Description	No. of images per grade (4-stage)	No. of images per grade (2-stage)
Normal	0	($\mu\text{A} = 0$) AND ($H = 0$)	546	546
Mild	1	($0 < \mu\text{A} \leq 5$) AND ($H = 0$)	153	654
Moderate + Severe	2	(($5 < \mu\text{A} < 15$) OR ($0 < H \leq 5$)) AND ($NV = 0$)	247	
Proliferative	3	($\mu\text{A} \geq 15$) OR ($H \geq 5$) OR ($NV = 1$)	254	

Table 1. Image distribution among each grade of Messidor-1 dataset.

Symbols	Meaning
μA	Number of micro aneurysms
H	Number of hemorrhages
$NV = 0$	No neovascularization
$NV = 1$	Neovascularization

Table 2. Detail of symbols.

without dilation. These images are categorized into four grades based on retinopathy severity. Risk of macular edema grading is also done based on hard exudates present in the retina. Images which were captured using a color video 3CCD camera mounted on a Topcon TRC NW6 with a 45-degree field of view, were recorded at either 1440*960, 2240*1488 or 2304*1536 pixels with 8 bits per color plane. As the images were captured at different resolutions, these variations do not provide uniformity across all images. The resolution of the images depends on factors such as the imaging device settings and the specific conditions under which the images were captured. Additionally, while some images were taken with pupil dilation and others were captured without it, this led to differences in image clarity and quality. These variations in both resolution and imaging conditions such as differences in lighting, calibration, and the use of dilation, contributed to the overall heterogeneity in the dataset. Each image is assigned a severity class corresponding to diabetic retinopathy. The Messidor-1 dataset includes 546 images from DR level 0, 153 from DR level 1, 247 from DR level 2, and 254 from DR level 3. The dataset has unbalanced distribution of images per grade/level. For multi-class classification, we considered all 4 stages of the dataset. For binary classification, the dataset was divided into 2 classes: 0 and 1. DR level 0 (having no diabetic retinopathy) containing normal images while grade 1, 2 and 3 were merged together into DR level 1 (having diabetic retinopathy) to fulfil the purpose of 2-stage classification.

The detail of each grade of this dataset is given in Table 1 where μA means number of micro aneurysms, H means number of hemorrhages, $NV = 0$ means no neovascularization and $NV = 1$ means neovascularization. Detail of these symbols is given in Table 2.

Image preprocessing

Data preprocessing is a crucial step in ensuring the success of deep learning processes by improving the quality of data to facilitate accurate and reliable model predictions⁴¹. The aim is to solve issues such as missing values, noise, inconsistencies, and extraneous data. Preprocessing is essential for optimizing image data before neural

network input. It ensures that neural networks receive well-conditioned data, leading to improved model performance and robustness⁴². In this research, the images are preprocessed in 4 steps: (1) Image Cropping, (2) Image Denoising, (3) Histogram Equalization (HE) and (4) Image Resizing. Preprocessing steps are the same for both 4 and 2-stage classification.

Image cropping

Cropping is the first step of preprocessing. The images in the dataset have an undesirable black background. Fundus images typically contain a large amount of black background surrounding the actual eye structure such as the retina, optic nerve, and blood vessels. Cropping out irrelevant background ensures that analysis is performed only on the desired structure that is the eye and this leads to more accurate results. For this purpose, we have used Bounding Box cropping method. Images were converted to grayscale. Then Thresholding was applied using cv2.threshold() with a low threshold value of 10 and a maximum value of 255. This operation segments the black background from the rest of the image and produces a binary image.

The eye region appears as white pixels against a black background once the image is binarized which easily isolates our region of interest (ROI) from unwanted background. The formula for Thresholding is below,

$$T(x, y) \begin{cases} 2 & \text{if } I(x, y) > \text{threshold55} \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $I(x, y)$ represents the intensity of the pixel at location (x, y) and $T(x, y)$ represents the thresholded output. After applying thresholding, the contours were detected using cv2.findContours() on the thresholded image. Among the detected contours, the contour with the maximum area was chosen as the eye region. Formula for calculating Area of Contour is,

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_i + 1 - x_i + 1 y_i) \quad (2)$$

Once the largest contour (the eye area) is identified, we calculated a bounding rectangle around it using cv2.boundingRect(). The bounding rectangle provides the coordinates (x, y) of the top-left corner as well as the width w and height h of the rectangle.

$$(x, y, w, h) = (\text{top-left x coordinate}, \text{top-left y coordinate}, \text{width}, \text{height}) \quad (3)$$

This operation extracts the region of interest (ROI), which in our case is the eye region. After getting the ROI, images were converted from grayscale back to RGB color space.

Image denoising

The second step of preprocessing is Denoising. Noise can obscure details and distort features in images, reducing their clarity and interpretability. By removing noise, we can improve the overall visual quality of the image. To remove noise from our dataset, Gaussian blur is applied using cv2.GaussianBlur(). Gaussian blur is a widely used method for image smoothing which replaces each pixel's value with a weighted average of its neighboring pixels.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (4)$$

where $G(x, y)$ represents the Gaussian kernel, x and y are the distances from the kernel center, and σ is the standard deviation. The size of the Gaussian kernel affects the amount of smoothing applied to the image. We experimented with various kernel sizes (e.g., 1×1 , 3×3 , 5×5 , and 7×7). While larger kernels provided more smoothing, they resulted in significant loss of crucial features in the retinal images. The 3×3 kernel effectively balanced noise reduction and detail preservation. In our research, a kernel size of 3×3 is chosen, indicating that the weighted average is calculated using a 3×3 neighborhood around each pixel. We chose Gaussian Blur for denoising because it effectively reduces noise without significantly altering the edges and important features in the retinal images. This is particularly crucial in medical images, where preserving details such as blood vessels and lesions is essential. While other denoising methods like median or bilateral filtering were considered, we found that Gaussian Blur provided the best balance between noise reduction and feature preservation. The superior performance of our model with Gaussian Blur in terms of accuracy and image quality justified its use. Also our model's results demonstrate the efficacy of Gaussian Blur in achieving high test accuracies.

Histogram equalization (HE)

For third step, Histogram Equalization (HE) is applied on the images. Histogram equalization is used to enhance the contrast of an image by redistributing pixel intensities. It works by computing the cumulative distribution function (CDF) of pixel intensities in the image and then stretching this function to cover the entire intensity range^{43–45}. Some images in the dataset are dull and have low contrast due to variations in illumination which makes it difficult to see the small details of retinal structures like micro aneurysms and tiny blood vessels. By applying histogram equalization, these images are enhanced to reveal fine details and improve overall quality of the image. For applying this technique, we have first converted our images from the RGB color space to the YUV color space using cv2.cvtColor(). When HE is applied, it turns the images to grayscale but we wanted to keep the images in RGB color space and enhance the contrast of the image without altering its actual.

color characteristics. That's why we used YUV color spacing. The YUV color space separates the image into three components: Y (luminance), U (chrominance), and V (chrominance). We applied Histogram Equalization

to the Y channel (luminance) only in the YUV color space using OpenCV's equalizeHist() function. This method adjusts the contrast by redistributing pixel intensity values. It effectively enhances the brightness and details of the image without affecting the color information which is stored in the U and V channels. After the application of HE, the enhanced luminance channel (Y) is merged with the original chrominance channels (U and V). This combines the enhanced contrast of the luminance with the original color information from the chrominance channels. Let $H(k)$ denote the histogram of the input image intensities, where k ranges from 0 to 255. The cumulative distribution function (CDF) $C(k)$ was computed as,

$$C(k) = \sum_{i=0}^k H(i) \quad (5)$$

The equalized intensity $I'(x, y)$ of a pixel with intensity $I(x, y)$ is computed as,

$$I'(x, y) = \frac{(C(I(x, y)) - C_{min})}{M \times N - C_{min}} \times 255 \quad (6)$$

Where $M \times N$ is the total number of pixels in the image and C_{min} is the minimum nonzero value in the cumulative histogram.

While we considered other contrast enhancement techniques as well, HE proved to be the most effective in improving image quality for our research. By applying HE to the luminance channel in the YUV color space, we enhanced the visibility of fine details like blood vessels and micro aneurysms without altering the color balance. This approach yielded the best results in terms of both image clarity and model performance. As a result, we selected Histogram Equalization (HE) for this research.

Image resizing

Fourth step of preprocessing is resizing the images of the dataset. Images are resized to ensure they all have uniform dimensions of $200 \times 200 \times 3$ pixels. Here, the first two numbers represent the height and width of the pixels while the third number signifies the presence of RGB (red, green, and blue) color channels in the image. By resizing the images to a consistent resolution, we ensure uniformity in the input data format. The steps of preprocessing are shown in Fig. 2.

Data augmentation

After preprocessing of the images, they are augmented using numerous techniques. As shown in Table 1, there is an uneven distribution of images per grade of Messidor-1 dataset³⁹. Grade 0 has the most images and Grade 1 has the least. The imbalanced distribution of classes in the Messidor dataset presents challenges for both 4-stage and 2-stage classification tasks. In multi-class setup, our model may struggle to detect minority classes particularly higher severity grades (like 2, 3 and 4) leading to missed critical cases. In binary classification, the bias toward the majority class (grade 0) can result in false negatives which may affect performance metrics like sensitivity and specificity. This skewness in the dataset can bias the model's performance towards predicting the majority class. This happens when certain grades or classes are represented by a disproportionately large number of images compared to others, the model may become biased towards the over-represented classes during training. As a result, the CNN may struggle to accurately classify instances from the under-represented classes, leading to decreased performance and reliability.

So to address the skewness in the dataset, we have applied data augmentation to increase the number of images in the minority classes. Data augmentation is a technique used to artificially increase the size and diversity of a dataset by applying various transformations to the existing data samples^{51,53}. It increases the number of samples in the minority class by generating new synthetic samples through transformations^{57–59}. In our framework, we have utilized both, geometric and photometric augmentation techniques. Photometric augmentation involves applying transformations to the pixel values of images to change their appearance while preserving their semantic content. Geometric augmentation involves applying transformations that modify the spatial arrangement or geometry of objects within images. The techniques used in this research are zooming, flipping, rotation, brightness adjustment, color adjustment and contrast adjustment. Geometric techniques (rotation flipping and zooming) not only increased sample diversity in our dataset but also helped the model learn to recognize features from different orientations making it less sensitive to image alignment. Photometric techniques (adjusting brightness, contrast, and color) helped the model become more robust to variations in lighting and imaging conditions.

These augmentation techniques enhanced our model performance by increasing data diversity, improving generalization and enabling the model to learn robust features that are invariant to variations in imaging.

conditions and orientations. Table 3 shows the augmentation techniques used in this research. Table 4 shows the new number of images per grade after applying augmentation. The new number of images for 4-stage classification became 8304 while for 2-stage classification, it became 4800. Figure 2 shows a sample of each augmentation technique. These techniques enhanced the model's performance on underrepresented classes. Analysis of the confusion matrix in “Results and discussion” section reveals that while sensitivity for minority classes improved, there remains a slight gap in accuracy compared to majority classes which can be improved in future by refining augmentation strategies and further analyzing class-specific performance to optimize model reliability across all classes.

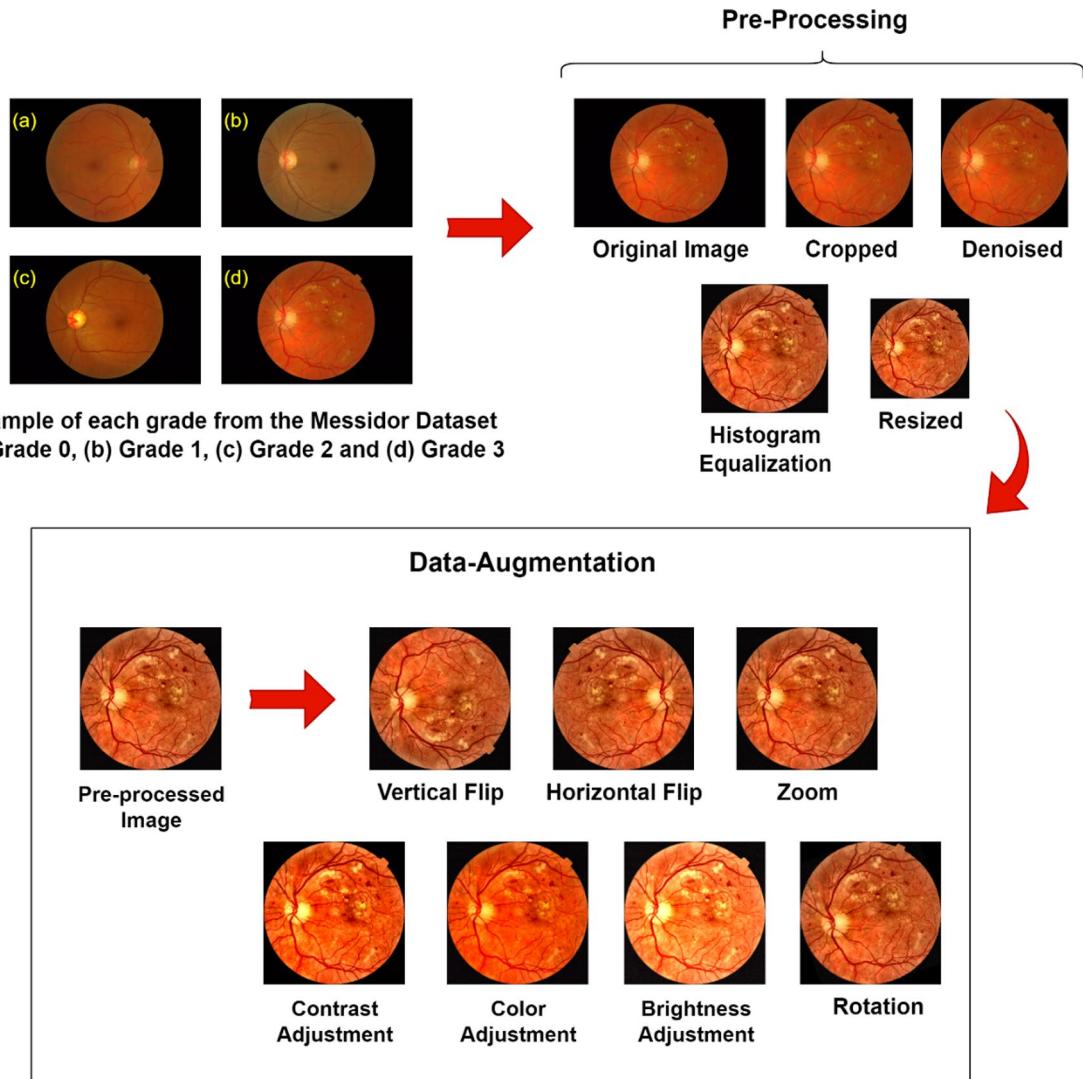


Fig. 2. A sample from each grade of Messidor-1 dataset is shown in the figure. It also demonstrates the preprocessing steps applied to these images and different augmentation techniques used on the preprocessed images.

Techniques	Value
Zoom range	0.2
Rotation range	10
Flip	Horizontal + Vertical
Brightness Adjustment	(0.5, 1.5)
Color Adjustment	(0.5, 1.5)
Contrast Adjustment	(0.5, 1.5)

Table 3. Augmentation techniques with their values.

Dataset split

Before model training, the augmented dataset is partitioned into training, validation, and testing sets. The dataset division in this research follows a ratio of 70:10:20, with 70% assigned to the training set, 10% to the validation set, and 20% to the testing set. The training set is utilized for model training, while the testing set evaluates the trained model's performance. Throughout the training phase, the validation set serves to assess the model's performance.

Retinopathy grade	Original image count	New no. of images after augmentation (4-stage)	New no. of images after augmentation (2-stage)
0	546	2184	2184
1	153	2000	2616
2	247	2080	
3	254	2040	
Total	1200	8304	4800

Table 4. New image count after augmentation.

Proposed RSG-Net (CNN) model

A Convolutional Neural Network (CNN) is a deep learning architecture that is widely used in computer vision⁶⁰. It is specifically designed to process structured grid data such as images. The key idea behind CNNs is to use convolution operations to automatically learn spatial hierarchies of features from the input data. AlexNet is an example of a CNN model that was developed by Alex Krizhevsky and his team in 2012⁶¹. It was a groundbreaking CNN that demonstrated exceptional performance in the Image Net Large Scale Visual Recognition Challenge, marking a significant advance in deep learning and computer vision⁶². A CNN generally comprises multiple layers: convolutional layers, pooling layers, and fully connected layers. The convolutional layers are the main components which use learnable filters (or kernels) that scan the input image to create feature maps. Pooling layers, like max pooling or average pooling, reduce the size of these feature maps while keeping the most important information. Finally, fully connected layers take these features to perform high-level reasoning and classification. The network learns the best filters through back propagation and gradient descent allowing it to accurately recognize patterns and objects in images.

The CNN proposed in this research is RSG-Net (Retinopathy Severity Grading Net). The layers used to build up this model include convolutional layers for feature extraction, max pooling layers for dimensionality reduction, fully connected layers for high-level reasoning, and regularization techniques like dropout and batch normalization to improve generalization and prevent over fitting. We added a total of 4 convolutional layers, 2 max pooling, 1 flatten layer, 1 fully connected, 1 batch normalization and 1 dropout layer. We have divided our CNN into 3 blocks.

The first block contains two convolutional layers, each with 32 filters and a kernel size of 3×3 . These layers are responsible for automatically extracting various features from the input retinal images with Rectified Linear Unit (ReLU) activation. They perform feature extraction from the input images by detecting edges, textures, and patterns. The ReLU activation function introduces non-linearity into the model. It allows the model to capture complex relationships in the data. The convolutional layers in this block are designed to learn low-level features such as edges and basic shapes that are fundamental for recognizing more complex patterns in deeper layers. By applying filters to local regions of the image, these layers create feature maps that highlight important structures in the input. After these two layers, we added a max pooling layer with a 2×2 pool size to reduce the spatial dimensions, decrease computational complexity and mitigate overfitting by down sampling the feature maps. The pooling layers retain the most important features for further processing. By summarizing the feature maps, max pooling aids in achieving translation invariance, allowing the model to recognize patterns regardless of their position in the input.

The second block has the same layer configuration as the first one only with slight changes. It contains two convolutional layers with 64 and 128 filters, respectively. In previous convolutional layers of Block 1, we used smaller filter value to detect low-level features like edges, textures, and basic shapes from the input images. As we moved deeper into the CNN network, we used larger filters (64, 128) to capture more abstract, high-level features like patterns, structures, or objects. This progression allowed the CNN network to gradually learn more complex and meaningful representations of the input data. After these layers we added the max pooling layer with a pool size of 2×2 .

Then comes the third block. It contains a flatten layer which flattens the feature maps into a one-dimensional array. The flatten layer converts the 2D feature maps from the convolutional layers into a 1D array, making the data compatible for input into fully connected layers. This transformation is necessary because fully connected layers require a flat input to perform high-level reasoning and classification. It acts as a bridge between the convolutional and dense layers which allows final decision-making in the network. After that a fully connected layer (dense layer) with 128 neurons and ReLU activation was added. Here, the dense layer serves to integrate all the features extracted by the convolutional layers and enables high-level reasoning by combining the learned features into a comprehensive understanding of the input data. Batch normalization layer was added right after the dense layer to normalize activations. Batch normalization helps stabilize the learning process by reducing internal covariate shift, allowing the network to train more efficiently and leading to faster convergence. Next, we added a dropout layer with a rate of 0.1 to randomly deactivate 10% of the neurons during each training iteration. This regularization technique prevents over fitting by ensuring the network doesn't become overly dependent on specific neurons, thereby improving its ability to generalize to new unseen data. Finally, a fully connected output layer was added to produce class probabilities using either softmax or sigmoid activation, depending on the classification task. The output layer combines the features from the previous layers and converts them into probabilities for each class. For 4-stage classification, the softmax activation function was used which outputs a probability distribution across all four classes. For 2-stage classification the sigmoid activation function was applied, which outputs a probability between 0 and 1 for each input representing the likelihood of the input

belonging to one of two classes that is the presence or absence of retinopathy. The detail on the layers used in RSG-Net is given in Fig. 3. The overall architecture of the model is given in Fig. 4.

Performance evaluation metrics

The performance of our model is evaluated using various metrics. These evaluation metrics are used to gauge the efficiency and accuracy of a RSG-Net predictions. When evaluating both the training and testing sets in this study, each data instance can fall into one of four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A True Positive denotes the model correctly predicting a positive

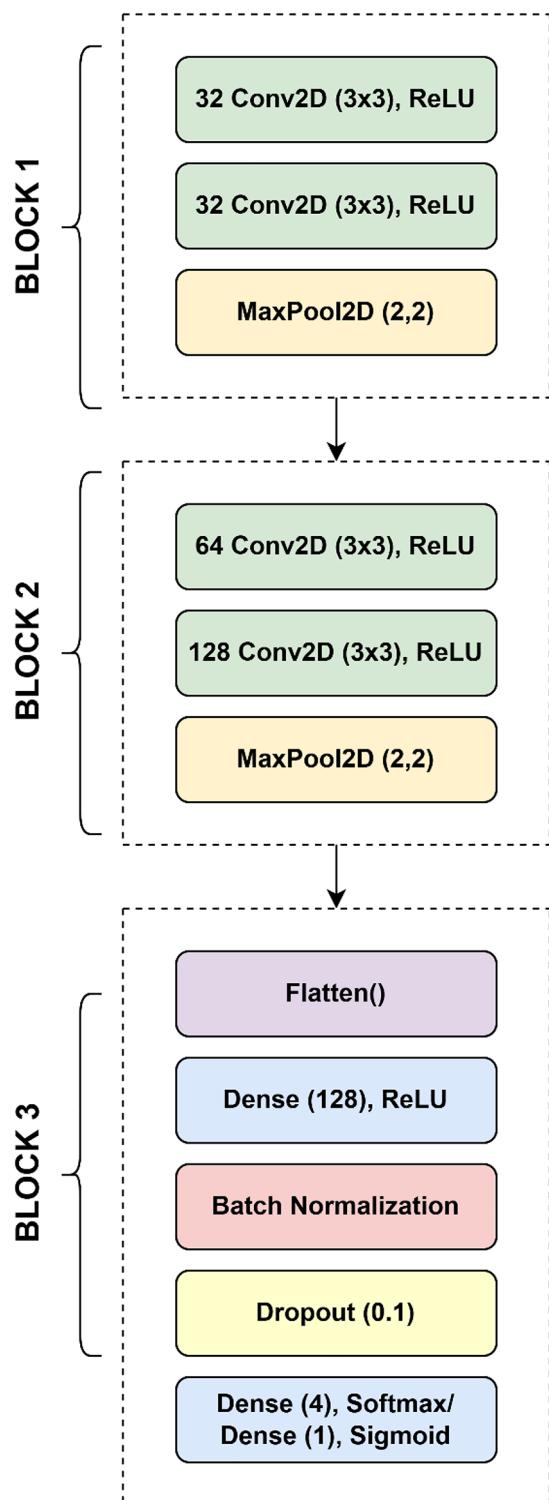


Fig. 3. Detail on the layers used for RSG-Net.

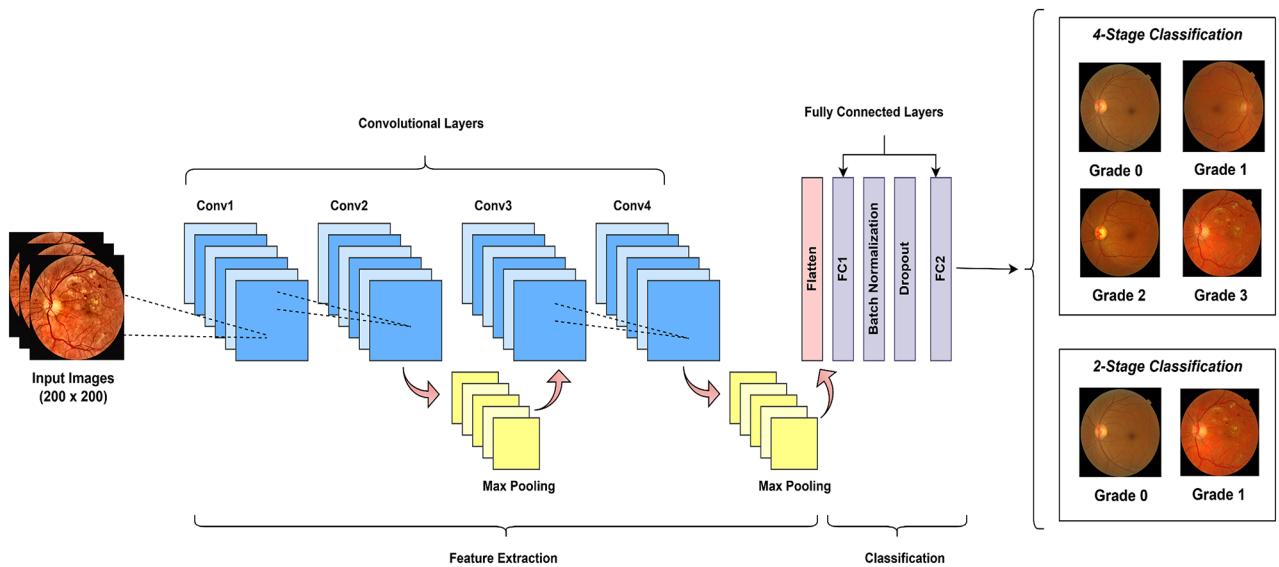


Fig. 4. RSG-Net architecture.

instance, while a True Negative signifies accurate prediction of a negative instance. Conversely, a False Positive indicates an incorrect prediction of a positive instance, and a False Negative represents an incorrect prediction of a negative instance. The focus of our research is on accuracy as a key performance metric. This is done to ensure reliable and effective diagnosis which is crucial for timely interventions in clinical settings. In the context of diabetic retinopathy screening, high accuracy signifies that the model is proficient in correctly classifying the majority of cases which is essential for better diagnosis and patient management. We also calculated additional metrics to gain deeper insights into the model's behavior. Our approach achieved a higher accuracy compared to the previous existing techniques thus it can help reduce the number of missed diagnoses, improve patient outcomes and lower healthcare costs. The performance of the proposed model is assessed on both train and test sets using following measures:

$$\text{Accuracy (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Misclassification Rate (MCR)} = \frac{FP + FN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Specificity (SPC)} = \frac{TN}{TN + FP} \quad (9)$$

$$\text{Sensitivity (SEN)} = \frac{TP}{TP + FN} \quad (10)$$

$$F1 - Score (F1) = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

$$\text{Positive Prediction Value (PPV)} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{Negative Prediction Value (NPV)} = \frac{TN}{TN + FN} \quad (13)$$

$$\text{False Positive Ratio (FPR)} = 1 - \text{Specificity} \quad (14)$$

$$\text{False Negative Ratio (FNR)} = 1 - \text{Sensitivity} \quad (15)$$

$$\text{Positive Likelihood Ratio (LR+)} = \frac{\text{Sensitivity}}{(1 - \text{Specificity})} \quad (16)$$

$$\text{Negative Likelihood Ratio (LR-)} = \frac{(1 - \text{Sensitivity})}{\text{Specificity}} \quad (17)$$

Results and discussion

After preprocessing and augmentation, the dataset was divided into training, validation and testing sets at a ratio of 70:10:20, respectively. After the images were up-sampled using augmentation, the total number of images became 8304 in 4-stage classification. So, the distribution of images across each set is as follows: the training set

Retinopathy grade	Image distribution per grade		
	Training set	Validation set	Testing set
0	1585	169	430
1	1473	151	376
2	1482	166	432
3	1438	179	423
Total	5978	665	1661

Table 5. Image distribution per grade for 4-stage classification after dataset split.

Retinopathy grade	Image distribution per grade		
	Training set	Validation set	Testing set
0	1572	175	437
1	1884	209	523
Total	3456	384	960

Table 6. Image distribution per grade for 2-stage classification after dataset split.

contains 5978 images (70%), while the validation and testing sets consist of 665 images (10%) and 1661 images (20%), respectively. The number of images per grade of these sets is given in Table 5. For 2-stage classification, out of 4800 images, the training set contains 3456 images (70%), while the validation and testing sets consist of 384 images (10%) and 960 images (20%), respectively. This distribution is shown in Table 6.

Experimental setting

To execute our model, we have used the Python programming language via Kaggle Notebooks. These notebooks offer extensive resources such as storage capacity of up to 73GB, 29GB of RAM, and a GPU with 15GB memory. It is capable of supporting heavy deep learning models. Various libraries including TensorFlow, OpenCV, Pandas, and Matplotlib were employed in this research to build up our framework. The computational efficiency of our model was evaluated by measuring the inference time during testing. The testing time for evaluating the model on the test set for both 4-stage and 2-stage was 1 s, with an average inference time of approximately 23 milliseconds per batch across 52 batches. Moreover, the training time for 4-stage and 2-stage classification was 12 min and 8 min respectively. This indicates that RSG-Net operates efficiently, making it suitable for practical deployment.

Prior to training our CNN model, the training parameters were adjusted to control aspects of the learning process and model architecture. Same parameters were used for 4 and 2-stage classification. Only Loss Function was different. Categorical cross entropy was used for 4-stage and binary cross entropy was used for 2-stage classification. A batch size of 64 was employed, and training was conducted over 30 epochs.

To evaluate the generalizability of our model, we adopted several strategies to ensure robust performance on unseen data.

1. Validation data monitoring: During training, validation loss and accuracy metrics were monitored to ensure the model's learning was generalizable and not overfitting. The model's loss and accuracy on the validation set were calculated at the end of each epoch, providing key indicators of overfitting or underfitting. If validation loss started to increase while training loss continued to decrease, this indicated overfitting, prompting us to implement regularization techniques like *EarlyStopping* and *ReduceLROnPlateau*. Additionally, the validation accuracy was tracked to ensure consistent improvement, guiding adjustments to the model architecture and hyperparameters as needed.
2. Regularization techniques: To avoid further overfitting we used regularization techniques including *EarlyStopping* and *ReduceLROnPlateau* methods. These mechanisms are configured to terminate the training process when the model achieves optimal performance to prevent over fitting. *EarlyStopping* monitored validation loss and halted the training process if the validation loss remained stable or improved for 10 consecutive epochs. Conversely, the *ReduceLROnPlateau* method reduced the learning rate by a factor of 0.3 if the validation accuracy remained unchanged for 2 successive epoch.
3. Experimentation with optimizers: For optimization, we used Stochastic Gradient Descent (SGD) optimizer with a learning rate set to 0.001. We carried out extensive experiments to evaluate different optimizers (Adam and RMSProp) from which we selected Stochastic Gradient Descent (SGD) for its effectiveness, and stable training performance on our specific architecture and dataset. We tracked convergence speed by monitoring the loss and accuracy metrics during training. Among the other optimizers, SGD with a learning rate of 0.001 demonstrated better long-term performance and generalization on the test set by achieving a higher accuracy and lower overfitting. It consistently provided better results overall making it the optimal choice for our RSG-Net model.

- Dropout and batch normalization layers: Through experimentation, we determined how adding/removing these layers affected overfitting and stability. Their inclusion improved generalizability by reducing overfitting and ensuring stable training. The hyper parameters utilized in the study are outlined in Table 7.

Experimental evaluation

The confusion matrix for both training and testing of 4-stage classification using RSG-Net model is provided in Fig. 5. During the 4-stage training phase, RSG-Net achieved a training accuracy of 99.96%, with a sensitivity of 99.96% and specificity of 99.98%. This indicates that the model was effectively trained and achieved high accuracy. It can be seen in Fig. 5 that out of 1585 images of grade 0 in the training set, the model correctly predicted all the images. Similarly it correctly predicted 1473 images of grade 1 and 1438 images of grade 3. While our model was able to classify all images of grade 0, 1, and 3 correctly in training, it misclassified 2 images in grade 2 and predicted 1480 images out of 1482 images correctly. When evaluated on the test set, RSG-Net demonstrated a test accuracy of 99.36%, with sensitivity and specificity of 99.41% and 99.79%, respectively. On the test set, the model misclassified 5 images across different grades, with the majority of errors occurring in grades 0 and 2. From Fig. 5, it can be seen that model successfully predicted all the images of grade 1 whereas it predicted 425 out of 430 images of grade 0, 430 out of 432 images for grade 2 and 420 out of 423 images for grade 3.

The confusion matrix for both training and testing of 2-stage classification using RSG-Net model is provided in Fig. 6. During the 2-stage training phase, RSG-Net achieved a training accuracy of 100%, with a sensitivity of 100% and specificity of 100%. The model excelled at training phase. It can be seen in Fig. 6 that model accurately predicted all the images of grade 0 and 1. When evaluated on the test set, RSG-Net demonstrated a test accuracy of 99.37%, with sensitivity and specificity of 100% and 98.62%, respectively. On the test set, the model successfully predicted all the images of grade 1 whereas it predicted 431 out of 437 images of grade 0. The model misclassified 6 images, all belonging to grade 0. The high sensitivity of 100% indicates the model is excellent at correctly identifying positive instances (grade 1), but the slightly lower specificity of 98.62% suggests it occasionally misclassified grade 0 instances as grade 1.

To address the issue of overfitting, several strategies were implemented in our framework including data augmentation, dropout, and batch normalization, all aimed at improving the model's generalization and performance on unseen data. Additionally, we carefully tuned the hyperparameters such as learning rate and batch size. Early stopping was applied to monitor the validation loss and halt training when performance stopped improving. Despite these efforts, the misclassification seen in 4-stage and 2-stage classification indicates that our model slightly struggled to differentiate between those particular grades potentially due to class overlap where adjacent stages of diabetic retinopathy have similar features or imbalanced data distribution. Additionally, overfitting resulting from perfect training scores suggests that the model also contributed to misclassification. To address this, future efforts will focus on refining classification thresholds and further balancing the dataset.

Table 8 presents the performance evaluation metrics for the proposed model computed for both the training and testing datasets. These metrics demonstrate the model's effectiveness in grading diabetic retinopathy on the fused dataset. The table reveals that our model achieved an accuracy (ACC) of 99.36%, a misclassification rate (MCR) of 0.0060, a specificity (SPC) of 99.79%, a sensitivity (SEN) of 0.9527, an F1-Score of 95.30%, a positive predictive value (PPV) of 95.30%, a negative predictive value (NPV) of 99.41%, a false positive rate (FPR) of 0.002, and a false negative rate (FNR) of 0.0058 on the test set during 4-stage classification. The high classification accuracy and low misclassification rate indicate excellent performance of RSG-Net on the test set. With strong performance across all metrics, it can be inferred that our proposed model excels at the classification of diabetic retinopathy. Similarly RSG-Net showed outstanding performance in 2-stage classification as well. Model achieved an accuracy (ACC) of 99.37%, a misclassification rate (MCR) of 0.0062, a specificity (SPC) of 98.62%, a sensitivity (SEN) of 100%, an F1-Score of 99.42%, a positive predictive value (PPV) of 98.86%, a negative predictive value (NPV) of 100%, a false positive rate (FPR) of 0.0137, and a false negative rate (FNR) of 0.0 on the test set during 2-stage classification which can be seen in Table 8.

Figure 7 provides a graphical representation of the training and validation accuracy+loss for RSG-Net for both (4-stage and 2 stage) classifications. The graphs illustrate that the model was successfully trained, as evidenced by the minimal validation loss observed. Our model achieved a validation accuracy of 98.80% and a validation loss of 0.0638 during 4-stage classification whereas it achieved a validation accuracy of 99.48% and a validation loss of 0.0460 during 2-stage classification. The validation set was used during the training of this CNN model to assess its performance. The high validation accuracy indicates that the model was well-trained.

Hyper parameters	Value
Optimizer	Stochastic gradient descent (SGD)
Learning rate	0.001
Batch size	64
No. of epochs	30
Loss function	Categorical cross entropy, binary cross entropy
EarlyStopping	Patience = 10
ReduceLROnPlateau	Factor = 0.3 Patience = 2

Table 7. Training parameters used in the research.

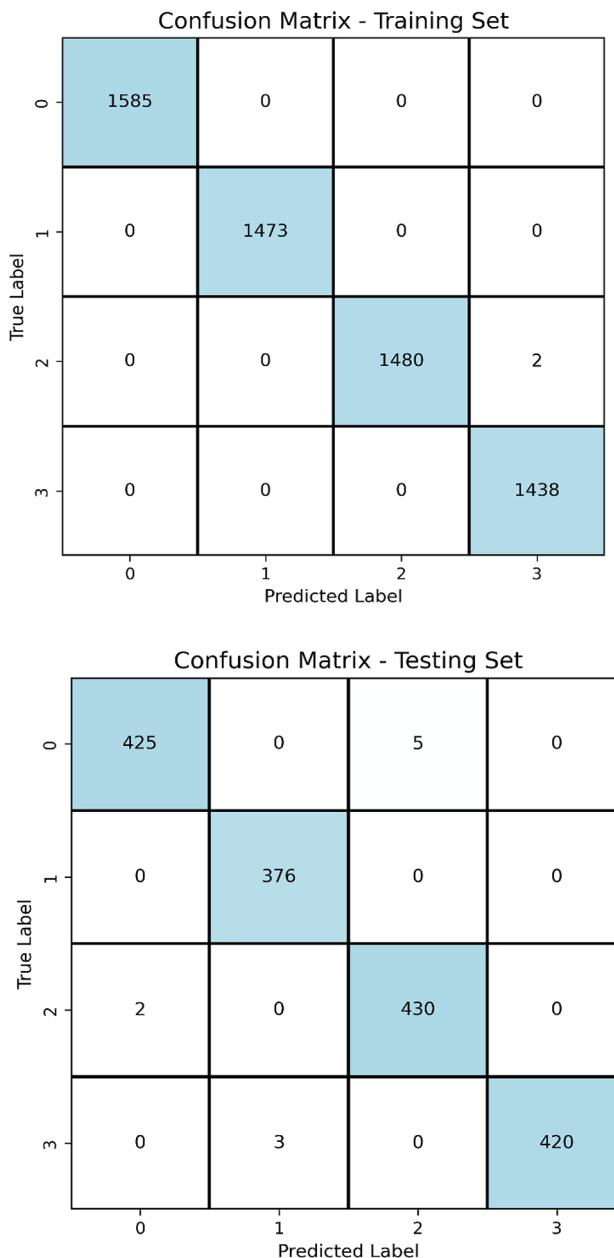


Fig. 5. Confusion matrix of training and testing set using RSG-Net for 4-stage classification.

Additionally, ROC curves were generated for both the training and testing sets, and the Area Under the Curve (AUC) was calculated from these curves. The AUC provides a summary of the ROC curve, indicating the model's ability to distinguish between different classes. An AUC value close to 1 signifies excellent performance in differentiating between positive and negative classes. In the ROC curve, the x-axis represents the false positive rate (1 - specificity), and the y-axis represents the true positive rate (sensitivity). Figure 8 displays the ROC curves for both the training and testing sets. The ROC curves are close to the y-axis, indicating a true positive rate near 1 and a false positive rate near 0. In 4-stage classification, the AUC obtained for the training set is 99.99%, while the testing set achieved an AUC of 99.98%. In 2-stage classification, the AUC obtained for the training set is 100%, while the testing set achieved an AUC of 99.98%. While the ROC curves for both the training and testing sets exhibit near-perfect AUC values, subtle deviations from the ideal curve highlight areas of potential underperformance particularly in minority classes. To address this, classification thresholds can be optimized, and further augmentation may improve model robustness.

The results obtained by the RSG-Net are compared with other state-of-the-art techniques used in previous researches in terms of test accuracy. The authors in^{40,48,52,54,56}, and⁶² proposed their own Convolutional Neural Network (CNN) architectures for classification of DR. Authors of⁴⁷ worked with stacked CNN structures

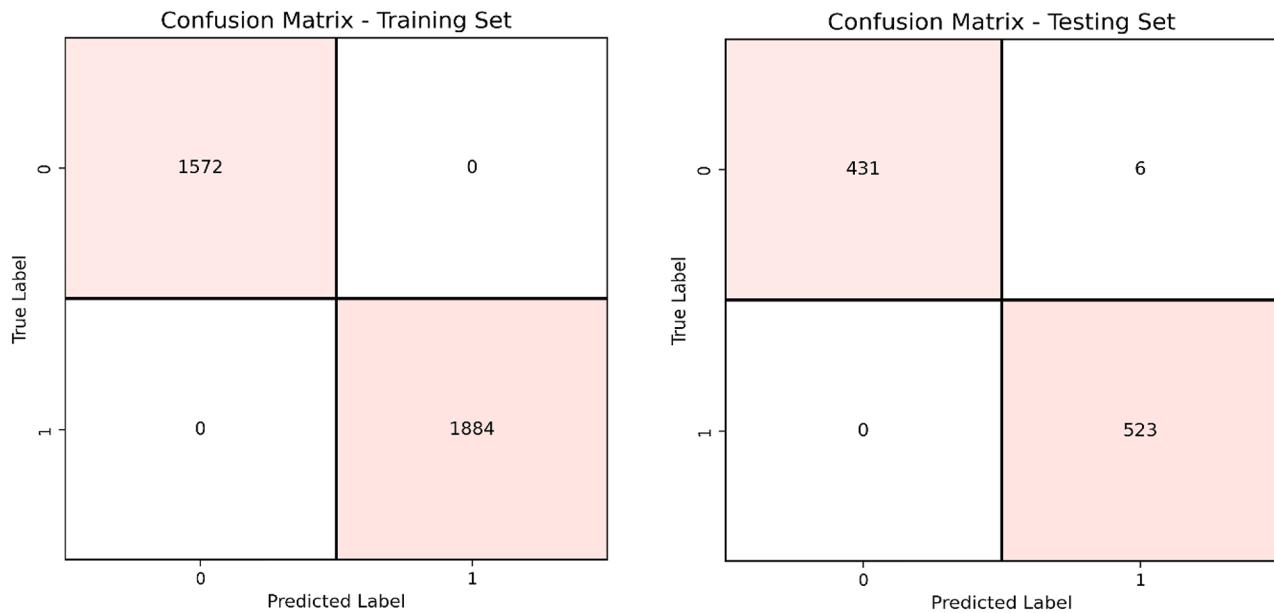


Fig. 6. Confusion matrix of training and testing set using RSG-Net for 2-stage classification.

Performance metrics	4-Stage classification (multi-class)		2-Stage classification (binary)	
	Training	Testing	Training	Testing
Accuracy (ACC)	0.9996	0.9936	1.0	0.9937
Misclassification rate (MCR)	0.0003	0.0060	0.0	0.0062
Specificity (SPC)	0.9998	0.9979	1.0	0.9862
Sensitivity (SEN)	0.9996	0.9941	1.0	1.0
F1-score (F1)	0.9994	0.9944	1.0	0.9942
Positive prediction value (PPV)	0.9996	0.9939	1.0	0.9886
Negative prediction value (NPV)	0.9998	0.9979	1.0	1.0
False positive ratio (FPR)	0.0001	0.0020	0.0	0.0137
False negative ratio (FNR)	0.0003	0.0058	0.0	0.0
Positive likelihood ratio (LR+)	∞	∞	∞	72.833
Negative likelihood ratio (LR-)	0.0003	0.0058	0.0	0.0

Table 8. Performance evaluation of RSG-Net model using various measures.

and in⁴⁶, authors created an intermediate fusion architecture using CNN. Hybrid architectures combining various deep learning models were explored in⁵⁵. Additionally, researchers in^{44,49} and^[35] utilized pre-trained models to achieve optimal accuracy. Among all these researchers, some have worked on binary classification of diabetic retinopathy while others worked on multi-class classification. Our proposed model demonstrated the highest accuracy by utilizing the proposed preprocessing and data augmentation techniques on both types of classifications. It outperformed the previous methods and predicted DR for 2-stage and 4-stage with high accuracy. Table 9 provides a detailed comparison between our RSG-Net model and previous research techniques.

The superior performance of RSG-Net is largely due to the rigorous approach we took in addressing class imbalance and carefully developing our CNN architecture. Unlike the compared existing methodologies which did not efficiently handle the inherent imbalance in diabetic retinopathy datasets, we ensured that the minority classes were adequately represented preventing the model from becoming biased toward the majority classes. Some of the existing techniques lack the use of regularization or proper preprocessing techniques due to which their models were unable to achieve higher accuracy. Our framework enabled RSG-Net to generalize better and achieve significantly higher accuracy compared to other state-of-the-art methods that lacked such structured approaches to imbalance and regularization. Our model can be efficiently deployed in hospitals and clinics where it will help reduce the workload for ophthalmologists while providing more accurate and early detection of diabetic retinopathy. Its deployment may face challenges such as ensuring proper training for healthcare staff or variability in image quality. Aside from this, our model can improve patient outcomes in real-world settings. In addition, RSG-Net showcased impressive computational efficiency by reducing training and testing

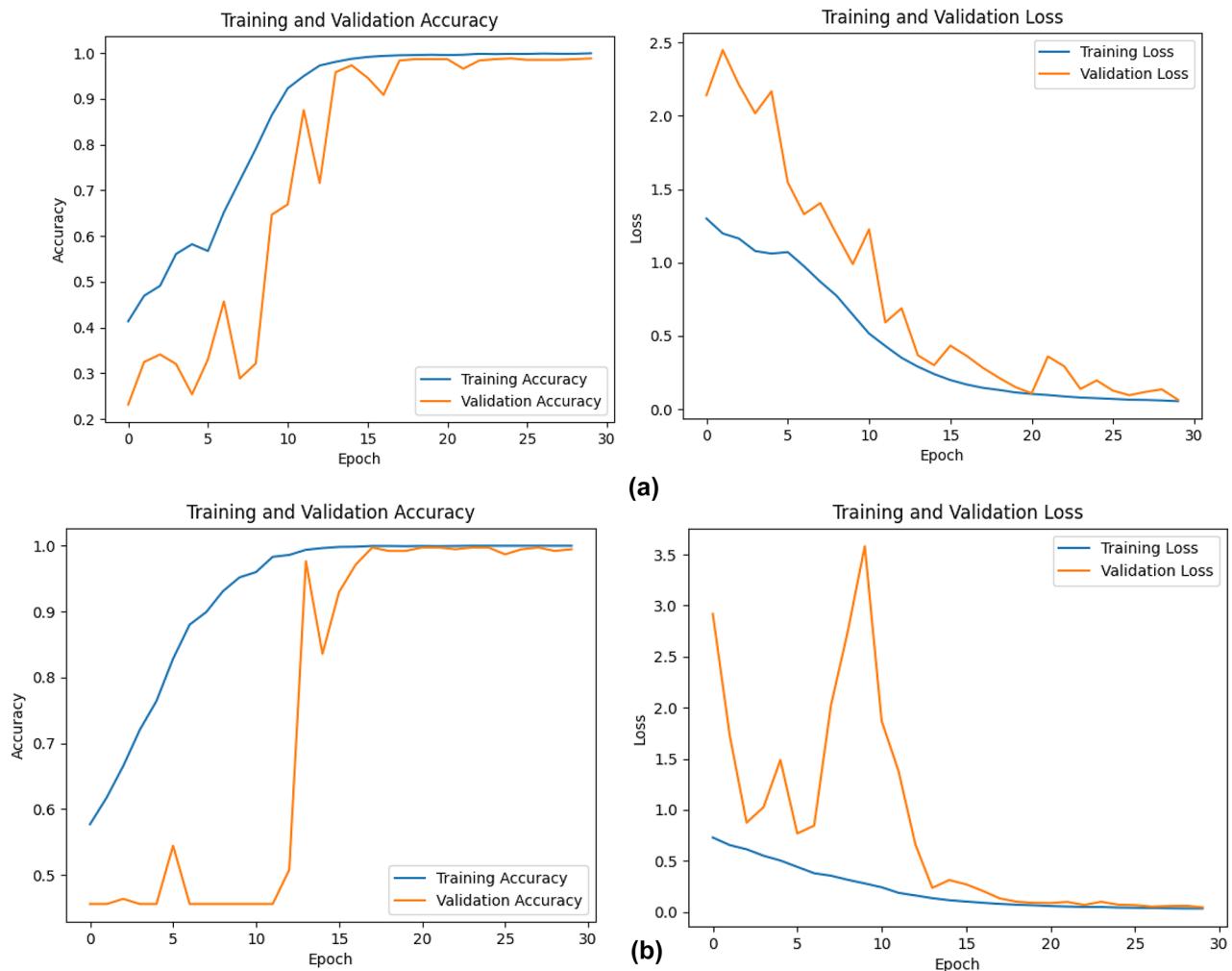


Fig. 7. (a) Graph for Training and Validation Accuracy + Loss of RSG-Net (4-stage classification). (b) Graph for Training and Validation Accuracy + Loss of RSG-Net (2-stage classification).

times compared to existing complex methods. This combination of high accuracy and lower computational cost enhanced the overall effectiveness of our approach.

Conclusion and future work

In this research we proposed RSG-Net, a convolutional neural network, for automating the detection and classification of diabetic retinopathy into 4 and 2 grades. The key contribution of this work is the development of an optimized CNN architecture that not only achieves high classification accuracy but also effectively addresses challenges like class imbalance through data augmentation and improves image quality using Histogram Equalization and Gaussian Blur. Using the benchmark Messidor-1 dataset, RSG-Net achieved a test accuracy of 99.36% for 4-stage classification and 99.37% for binary classification, outperforming previously existing models in terms of accuracy. These advancements and results suggest that RSG-Net has strong potential for clinical applications and it provides a reliable, automated solution for timely diabetic retinopathy detection. Compared to existing models, RSG-Net demonstrated superior performance particularly in handling the complexity of multi-class classification where subtle differences between DR stages can be challenging to detect. Our model's robust architecture combined with effective preprocessing techniques enabled more accurate feature extraction and classification.

A potential limitation of the study is the risk of overfitting due to the high training accuracy observed. Although the model performs well on the test set, future work would try out more regularization techniques such as increasing dropout rates, to further enhance its generalization on unseen data. Additionally, the model has only been tested on the Messidor-1 dataset and its generalizability across diverse datasets with different imaging conditions remains to be explored. In future work, we intend to apply our framework to other diverse DR datasets to evaluate and validate the approach across broader data. Furthermore, we plan to incorporate pre-trained models and ensemble techniques. We will also refine our augmentation strategies and focus on improving class-specific performance to optimize model reliability. Moreover, we will incorporate statistical hypothesis testing and confidence intervals to further enhance the analytical rigor and validation of our results.

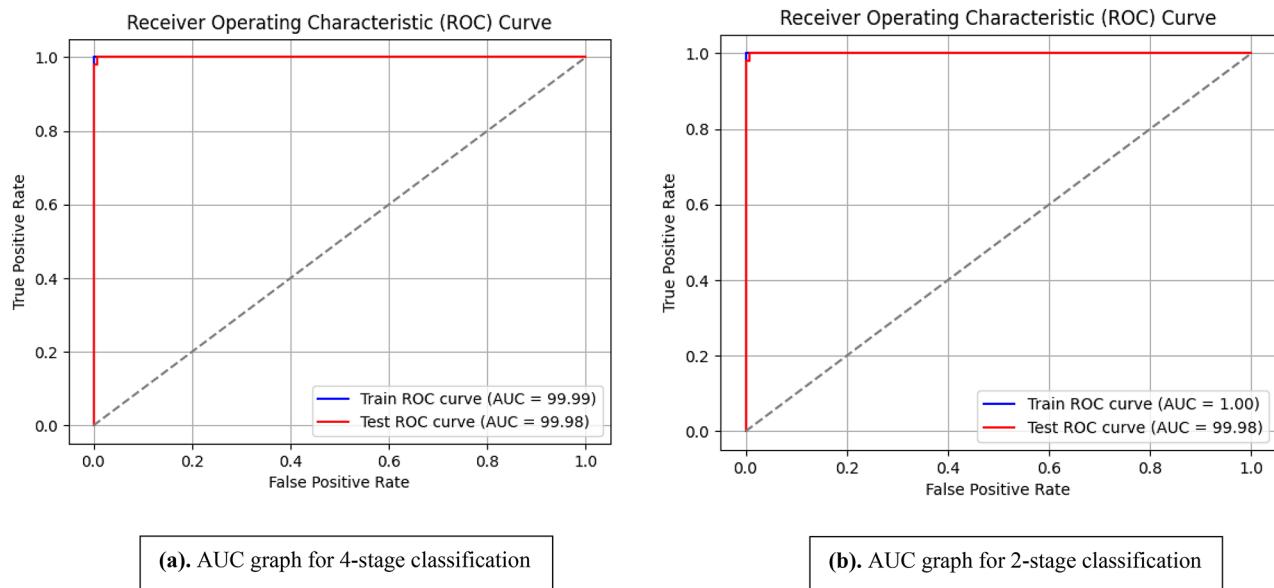


Fig. 8. Graphs of AUC obtained on training and testing using RSG-Net for both, 4-stage and 2-stage classification.

Reference	Year	Dataset	Classification	Methodology	Accuracy
Raja Kumar et al. ⁴⁰	2021	Kaggle Dataset	Multi	CNN Model	94.44%
Ebrahimi et al. ⁴⁶	2023	Private	Multi	CNN (intermediate fusion) Architecture	92.65%
Kaushik et al. ⁴⁷	2021	EyePACS	Binary + Multi	Stacked CNNS	Binary = 97.92% Multi = 87.45%
Hemanth et al. ⁴⁸	2019	Messidor-1	Multi	CNN Model	97%
Abbood et al. ⁴⁹	2022	Messidor-1	Multi	ResNet-50	93.6%
Mondal et al. ⁵⁰	2022	APTOs2019 + DIARETDB1	Binary + Multi	EDLDR	Binary = 96.98% Multi = 86.08%
Mushtaq et al. ⁴⁴	2021	DR detection 2015 + APTOS2019	Multi	DenseNet-169	90%
Moustari et al. ³⁵	2024	APTOs2019	Multi	DenseNet-121	98.4%
Nahiduzzaman et al. ⁶²	2023	Kaggle DR 2015 Competition + APTOS2019	Multi	ELM (Parallel CNN) Model	97.27%
Shaban et al. ⁵²	2020	APTOs2019	Multi	CNN Model	88–89%
Nasir et al. ⁵⁴	2022	IDRID	Binary	CNN Model	96%
Menaouer et al. ⁵⁵	2022	APTOs2019	Multi	Hybrid models (CNN, VGG16 and VGG19)	90.60%.
Jian et al. ⁶³	2023	APTOs2019	Multi	Triple-DRNets	92.08%
Khan et al. ⁵⁶	2019	Messidor-1	Multi	CNN Model	98.15%
Current Study	2024	Messidor-1	Binary + Multi	RSG-Net (CNN) Model	Binary = 99.37% Multi = 99.36%

Table 9. Comparison of proposed model with previous studies in terms of test accuracy.

Data availability

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Received: 23 August 2024; Accepted: 16 January 2025

Published online: 30 January 2025

References

- Chilukoti, S. V., Maida, A. S. & Hei, X. Diabetic retinopathy detection using transfer learning from pre-trained convolutional neural network models. *IEEE J. Biomed. Heal Inf.* **20**, 1–10 (2022).
- Al-Smadi, M., Hammad, M., Baker, Q. B. & Sa'ad, A. A transfer learning with deep neural network approach for diabetic retinopathy classification. *Int. J. Electr. Comput. Eng.* **11** (4), 3492 (2021).
- Akhtar, S., Aftab, S., Ahmad, M. & Ihnaini, B. A transfer learning based framework for diabetic retinopathy detection using data fusion. In *2024 2nd International Conference on Cyber Resilience (ICCR)*. 1–5. (IEEE, 2024).

4. Akhtar, S. & Aftab, S. A framework for diabetic retinopathy detection using transfer learning and data fusion. *Int. J. Inform. Technol. Comput. Sci. (IJITCS)*. **16** (6), 61–73 (2024).
5. Sebti, R., Zroud, S., Kahloul, L. & Benharzallah, S. A deep learning approach for the diabetic retinopathy detection. In *The Proceedings of the International Conference on Smart City Applications*. 459–469. (Springer, 2021).
6. Gangwar, A. K. & Ravi, V. Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications* (FICTA 2020). Vol. 1. 679–689. (Springer, 2021).
7. Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z. & Jabbar, A. Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sci.* **12** (5), 535 (2022).
8. Ghosh, S. & Chatterjee, A. *Transfer-Ensemble Learning based Deep Convolutional Neural Networks for Diabetic Retinopathy Classification*. arXiv preprint: arXiv:2308.00525 (2023).
9. Bilal, A., Liu, X., Shafiq, M., Ahmed, Z. & Long, H. NIMEQ-SACNet: A novel self-attention precision medicine model for vision-threatening diabetic retinopathy using image data. *Comput. Biol. Med.* **171**, 108099 (2024).
10. Akhtar, S., Aftab, S., Ahmad, M. & Ihnaini, B. A Classification framework for diabetic retinopathy detection using transfer learning. In *2024 2nd International Conference on Cyber Resilience (ICCR)*. 1–5. (IEEE, 2024).
11. Ahmad, M., Alfayad, M., Aftab, S., Khan, M. A., Fatima, A., Shoaib, B. & Elmitwal, N. S. Data and machine learning fusion architecture for cardiovascular disease prediction. *Comput. Mater. Contin.* **69**(2) (2021).
12. AbdelMaksoud, E., Barakat, S. & Elmogy, M. A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. *Med. Biol. Eng. Comput.* **60** (7), 2015–2038 (2022).
13. Bora, A., Balasubramanian, S., Babenko, B., Virmani, S., Venugopalan, S., Mitani, A. & Bavishi, P. Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit. Health* **3**(1), e10–e19 (2021).
14. Li, F., Wang, Y., Xu, T., Dong, L., Yan, L., Jiang, M. & Zou, H. Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. *Eye* **36**(7), 1433–1441 (2022).
15. Akhtar, S., Aftab, S., Ahmad, M. & Akhtar, A. Diabetic retinopathy severity grading using transfer learning techniques. *Int. J. Eng. Manuf. (IJEM)*. **14** (6), 41–53 (2024).
16. Singh, L. K., Garg, H. & Khanna, M. An artificial intelligence-based smart system for early glaucoma recognition using OCT images. In *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*. 1424–1454. (IGI Global, 2023).
17. Singh, L. K., Khanna, M. & Thawkar, S. A novel hybrid robust architecture for automatic screening of glaucoma using fundus photos, built on feature selection and machine learning-nature driven computing. *Expert Syst.* **39**(10), e13069 (2022).
18. Abbas, S. et al. Data and ensemble machine learning fusion based intelligent software defect prediction system. *Computers Mater. Contin.* **75**(3) (2023).
19. MunishKhanna, Singh, L. K. & Garg, H. A novel approach for human diseases prediction using nature inspired computing & machine learning approach. *Multimed. Tools Appl.* **83** (6), 17773–17809 (2024).
20. Singh, L. K., Garg, H. & Pooja Automated glaucoma type identification using machine learning or deep learning techniques. *Adva. Mach. Intell. Interact. Med. Image Anal.* 241–263 (2020).
21. Singh, L. K., Khanna, M., Thawkar, S. & Singh, R. Nature-inspired computing and machine learning based classification approach for glaucoma in retinal fundus images. *Multimed. Tools Appl.* **82** (27), 42851–42899 (2023).
22. Kassani, S. H. et al. Diabetic retinopathy classification using a modified xception architecture. In *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. 1–6. (IEEE, 2019).
23. Bilal, A., Zhu, L., Deng, A., Lu, H. & Wu, N. AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning. *Symmetry* **14** (7), 1427 (2022).
24. Ebin, P. M. & Ranjana, P. A variant binary classification model for no-DR mild-DR detection using CLAHE images with transfer learning. In *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*. 1–5. (IEEE, 2022).
25. Bhardwaj, C., Jain, S. & Sood, M. Transfer learning based robust automatic detection system for diabetic retinopathy grading. *Neural Comput. Appl.* **33** (20), 13999–14019 (2021).
26. Alyoubi, W. L., Abulkhair, M. F. & Shalash, W. M. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors* **21** (11), 3704 (2021).
27. Çınarer, G., Kılıç, K. & Parlar, T. A deep transfer learning framework for the staging of diabetic retinopathy. *J. Sci. Rep.-A*. **051**, 106–119 (2022).
28. Kotiyal, B. & Pathak, H. Diabetic retinopathy binary image classification using PySpark. *Int. J. Math. Eng. Manag. Sci.* **7** (5), 624 (2022).
29. Skouta, A., Elmoufidi, A., Jai-Andalousi, S. & Ochetto, O. Automated binary classification of diabetic retinopathy by convolutional neural networks. In *Advances on Smart and Soft Computing: Proceedings of ICACIn 2020*. 177–187. (Springer Singapore, 2021).
30. Uppamma, P. & Bhattacharya, S. Diabetic retinopathy detection: A blockchain and African vulture optimization algorithm-based deep learning framework. *Electronics* **12** (3), 742 (2023).
31. Sikder, N. et al. Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images. *Symmetry* **13** (4), 670 (2021).
32. Ghazal, M., Ali, S. S., Mahmoud, A. H., Shalaby, A. M. & El-Baz, A. Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks. *IEEE Access*. **8**, 34387–34397 (2020).
33. Liu, H. et al. Hybrid model structure for diabetic retinopathy classification. *J. Healthc. Eng.* **2020** (2020).
34. Jabbar, A. et al. *A Lesion-Based Diabetic Retinopathy Detection Through Hybrid Deep Learning Model*. (IEEE Access, 2024).
35. Moustari, A. M., Brik, Y., Attallah, B. & Bouaouina, R. Two-stage deep learning classification for diabetic retinopathy using gradient weighted class activation mapping. *Automatika* **65** (3), 1284–1299 (2024).
36. Shakibania, H., Raoufi, S., Pourafkham, B., Khotanlou, H. & Mansoorizadeh, M. Dual branch deep learning network for detection and stage grading of diabetic retinopathy. *Biomed. Signal Process. Control.* **93**, 106168 (2024).
37. Dai, L., Sheng, B., Chen, T., Wu, Q., Liu, R., Cai, C. & Jia, W. A deep learning system for predicting time to progression of diabetic retinopathy. *Nat. Med.* **30**(2), 584–594 (2024).
38. Jin, Y. et al. Deep learning-driven automated quality assessment of ultra-widefield optical coherence tomography angiography images for diabetic retinopathy. *Vis. Comput.* 1–11 (2024).
39. <https://www.adcis.net/en/third-party/messidor/>.
40. Raja Kumar, R., Pandian, R., Jacob, P., Pravin, T. & Indumathi, P. A. Detection of diabetic retinopathy using deep convolutional neural networks. In *Computational Vision and Bio-Inspired Computing: ICCVBIC 2020*. 415–430. (Springer Singapore, 2021).
41. Chen, P. N. et al. General deep learning model for detecting diabetic retinopathy. *BMC Bioinform.* **22**, 1–15 (2021).
42. Bilal, A., Sun, G., Mazhar, S. & Imran, A. Improved grey wolf optimization-based feature selection and classification using CNN for diabetic retinopathy detection. In *Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2021*. 1–14. (Springer Singapore, 2022).
43. Martínez-Murcia, F. J., Ortiz-García, A., Ramírez, J., Górriz-Sáez, J. M. & Cruz, R. *Deep Residual Transfer Learning for Automatic Diabetic Retinopathy Grading*. (2021).
44. Mushtaq, G. & Siddiqui, F. Detection of diabetic retinopathy using deep learning methodology. In *IOP Conference Series: Materials Science and Engineering*. Vol. 1070(1). 012049. (IOP Publishing, 2021).

45. Butt, M. M., Iskandar, D. A., Abdelhamid, S. E., Latif, G. & Alghazo, R. Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features. *Diagnostics* **12** (7), 1607 (2022).
46. Ebrahimi, B. et al. Optimizing the OCTA layer fusion option for deep learning classification of diabetic retinopathy. *Biomed. Opt. Exp.* **14** (9), 4713–4724 (2023).
47. Kaushik, H. et al. Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models. *IEEE Access* **9**, 108276–108292 (2021).
48. Hemanth, D. J., Deperlioglu, O. & Kose, U. An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. *Neural Comput. Appl.* **32** (3), 707–721 (2020).
49. Abbood, S. H. et al. Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model. *IEEE Access* **10**, 73079–73086 (2022).
50. Mondal, S. S., Mandal, N., Singh, K. K., Singh, A. & Izonin, I. Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy. *Diagnostics* **13** (1), 124 (2022).
51. Mutawa, A. M., Al-Sabti, K., Raizada, S. & Sruthi, S. A deep learning model for detecting diabetic retinopathy stages with discrete wavelet transform. *Appl. Sci.* **14** (11), 4428 (2024).
52. Shaban, M., Ogur, Z., Mahmoud, A., Switala, A., Shalaby, A., Abu Khalifeh, H. & El-Baz, A. S. A convolutional neural network for the screening and staging of diabetic retinopathy. *Plos one* **15**(6), e0233514 (2020).
53. Bilal, A., Sun, G., Li, Y., Mazhar, S. & Khan, A. Q. Diabetic retinopathy detection and classification using mixed models for a disease grading database. *IEEE Access* **9**, 23544–23553 (2021).
54. Nasir, N. et al. M., Deep DR: Detection of diabetic retinopathy using a convolutional neural network. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)*. 1–5. (IEEE, 2022).
55. Menaouer, B., Dermange, Z., Houda Kebir, E. & Matta, N. Diabetic retinopathy classification using hybrid deep learning approach. *SN Comput. Sci.* **3** (5), 357 (2022).
56. Khan, S. H., Abbas, Z. & Rizvi, S. D. Classification of diabetic retinopathy images based on customised CNN architecture. In *2019 Amity International Conference on Artificial Intelligence (AICAI)*. 244–248. (IEEE, 2019).
57. Masood, S., Luthra, T., Sundriyal, H. & Ahmed, M. Identification of diabetic retinopathy in eye images using transfer learning. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*. 1183–1187. (2017).
58. Salvi, R. S., Labhsetwar, S. R., Kolte, P. A., Venkatesh, V. S. & Bareto, A. M. Predictive analysis of diabetic retinopathy with transfer learning. In *2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*. 1–6. (IEEE, 2021).
59. Elsharkawy, M., Sharafeldeen, A., Soliman, A., Khalifa, F., Ghazal, M., El-Dayamony, E. & El-Baz, A. A novel computer-aided diagnostic system for early detection of diabetic retinopathy using 3D-OCT higher-order spatial appearance model. *Diagnostics* **12**(2), 461 (2022).
60. Gao, Z. et al. Automatic interpretation and clinical evaluation for fundus fluorescein angiography images of diabetic retinopathy patients by deep learning. *Br. J. Ophthalmol.* **107** (12), 1852–1858 (2023).
61. Bilal, A., Sun, G. & Mazhar, S. Diabetic retinopathy detection using weighted filters and classification using CNN. In *2021 International Conference on Intelligent Technologies (CONIT)*. 1–6. (IEEE, 2021).
62. Nahiduzzaman, M. et al. Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier. *Expert Syst. Appl.* **217**, 119557 (2023).
63. Jian, M., Chen, H., Tao, C., Li, X. & Wang, G. Triple-DRNet: A triple-cascade convolution neural network for diabetic retinopathy grading using fundus images. *Comput. Biol. Med.* **155**, 106631 (2023).

Author contributions

Samia Akhtar, Shabib Aftab, and Munir Ahmad have collected data from different resources and contributed to writing—original draft preparation. Sagheer Abbas, Samia Akhtar, and Muhammad Adnan Khan performed formal analysis and Simulation, Munir Ahmad, Muhammad Adnan Khan and Samia Akhtar; writing—review and editing, Sagheer Abbas and Muhammad Adnan Khan; performed supervision, Shabib Aftab, Munir Ahmad, Oualid Ali and Samia Akhtar.; drafted pictures and tables, Taher M. Ghazal, Oualid Ali and Muhammad Adnan Khan.; performed revisions and improve the quality of the draft. All authors have read and agreed to the published version of the manuscript.

Funding

This manuscript is funded by Prince Mohammad Bin Fahd University, Al-Khobar, Dhahran, 34754, Saudi Arabia.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.A. or M.A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025