Journal of Health Informatics in Developing Countries
http://www.jhidc.org/
Vol. 19 No. 2, 2025

Submitted: May 28th, 2025 Accepted: Nov 13th, 2025

The Use of Explainable Deep Learning Models for Diabetic Retinopathy Management: From Screening to Severity Grading

Silas Majyambere[1, 2, *], Tony Lindgren[1], Celestin Twizere[2], Egide Gisagara[3]

[1] *Department of Computer and Systems Sciences, Stockholm University, Sweden.*

[2] *Center for Biomedical Engineering and E-health (CEBE), University of Rwanda, Rwanda.*

[3] *University Teaching Hospital of Kigali (CHUK), Rwanda.*

Abstract

Background: Diabetic Retinopathy (DR), a leading cause of blindness, is rising globally due to increasing diabetes prevalence, especially in low-resource settings lacking diagnostic tools and specialists. Early DR is often asymptomatic; timely and scalable screening is essential. This study introduces a cost-effective two-stage deep learning framework for DR detection: Stage 1 performs binary classification on Optical Coherence Tomography (OCT) images, while Stage 2 grades DR severity use multi-class classification. Shapley Additive exPlanations (SHAP) enhance model transparency and clinical trust. Trained on public datasets and validated by a CHUK ophthalmologist, the approach provides a cost-effective solution for DR management in the underserved diabetic population.

Methods: The study followed five phases: (1) image dataset preparation using a four-step preprocessing pipeline: blurring, denoising, augmentation, and vector transformation for CNN input; (2) training and evaluating five pre-trained CNNs (MobileNetV3, DenseNet121, InceptionV3, Xception, VGG19) with standard metrics; (3) applying SHAP to interpret predictions of a multi-class model; (4) validating performance on 270 CHUK specialist-annotated images and 270 from the Brazil dataset; and (5) deploying the top models in a web application.

Results: The proposed model achieved strong performance in both DR screening and severity grading. DenseNet121 performed best, reaching 97% accuracy for referable DR detection and class accuracy of 97% (healthy), 92% (mild), 89% (moderate), 94% (severe), and 92% (proliferative). Validation datasets yielded Cohen's Kappa scores of 0.771 and 0.680, demonstrating substantial agreement with expert grading. SHAP explanations enhanced interpretability across DR stages.

Conclusion: The applied preprocessing techniques improved image clarity and boosted model accuracy in grading DR severity. The proposed CNN framework enables end-to-end DR management. Combining expert validation, quality imaging, and interpretable deep learning offers a promising solution for early detection of DR and vision preservation in diabetic populations.

Keywords: Explainable Deep Learning, Diabetic Retinopathy, DR Screening, DR Grading, Image Denoising, SHAP.

*Silas Majyambere Department of Computer and Systems Sciences, Stockholm University, Sweden; Department of Computer Science, University of Rwanda, Rwanda; Email: majyambere@dsv.su.se

1. Introduction

Diabetes is a chronic condition that alters the normal functioning of the body. Despite being recognized for over a century, it remains a global health concern and a leading cause of morbidity and comorbidity [1]. When left uncontrolled, diabetes can lead to severe complications that are costly to manage and pose life-threatening risks to diabetic patients. One such complication is diabetic retinopathy, a progressive eye disease that is a leading cause of preventable vision loss and blindness worldwide [2]. The eye, often referred to as the light of our body, collaborates with the brain to maintain vision.

Diabetes primarily damages small blood vessels in various body organs, progressively leading to organ dysfunction. For this, it is known as a silent killer. It is also a significant contributor to diseases such as hypertension, heart disease, and stroke. chronic hyperglycemia accelerates damage to the small blood vessels in the retina, resulting in diabetic retinopathy [3]. Diabetic retinopathy is a progressive condition that is asymptomatic in its early stages. In the advanced stages of DR, patients may experience reduced vision, blurred vision, floaters, and frequent eye pain. Clinically, diabetic retinopathy is classified into two categories: non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) [4]. NPDR can be managed with timely intervention and at lower treatment costs, whereas PDR is more severe, costly to treat, and carries a high risk of vision loss and blindness. Preventive strategies include the regular screening and management of blood glucose levels among the diabetic population. To prevent the progression to PDR, diabetic patients, especially those with a history of diabetes exceeding five years, are strongly advised to undergo regular eye examinations each year.

1.1 DR Diagnosis

Diabetic retinopathy (DR) screening is crucial for early detection and vision preservation in diabetic patients. Diagnosis methods include both invasive and non-invasive approaches. Diabetic retinopathy (DR) is a leading cause of visual impairment and blindness among individuals with diabetes. Preventive measures include early detection and effective blood glucose management. However, since the early stages of DR are asymptomatic, individuals with diabetes for more than five years are advised to undergo annual screenings. In Low and Middle-Income Countries (LIMCs), limited healthcare resources pose significant challenges to large-scale DR screening.

(a) Invasive diabetic retinopathy screening

The invasive approach to diabetic retinopathy (DR) diagnosis, known as the mydriatic method [5], involves a dilated eye examination. Prior to retinal imaging, eye drops are administered to dilate the pupil. While effective, this method may cause side effects for the patient following the screening. Given the increasing prevalence of diabetic retinopathy and the shortage of ophthalmologists in low- and middle-income countries (LIMCs), the mydriatic method is not suitable for the timely detection and treatment of DR before it impacts vision.

(b) Non-invasive diabetic retinopathy screening

Optical coherence tomography (OCT) and fundus cameras are widely used non-invasive ophthalmic technologies for DR detection [6], providing high-resolution retinal images for detailed analysis by ophthalmologists. However, the high cost of these devices and their limited portability in remote areas hinder the widespread adoption of DR screening in LIMCs.

1.2. Grading of diabetic retinopathy severity

The Early Treatment Diabetic Retinopathy Study (ETDRS) is widely regarded as the gold standard for evaluating the progression of diabetic retinopathy (DR) [7]. It categorizes DR into two main types: Non-Proliferative DR (NPDR) and Proliferative DR (PDR). NPDR is further divided into three stages, including Mild, Moderate, and Severe, as illustrated in Figure 2. Mild NPDR is marked by microaneurysms alone, while Moderate NPDR includes additional hemorrhages and exudates. Severe NPDR is defined by the 4-2-1 rule: hemorrhages in four quadrants, venous beading in two, and intraretinal microvascular abnormalities (IRMAs) in one. PDR is characterized by neovascularization or vitreous hemorrhage. Accurate staging using the ETDRS framework is crucial for informed clinical decision-making and effective treatment planning. This study employs the International Clinical Diabetic Retinopathy Disease Severity Scale (ICDRS) [8], which classifies DR into five stages: No DR, Mild NPDR, Moderate NPDR, Severe NPDR, and PDR.

1.3. Deep Learning in DR Screening

Deep learning (DL) has demonstrated strong performance in medical image processing and analysis [9, 10, 11]. However, despite the reported performance in image processing, deep learning is associated with long-standing criticism of operating in a black box mode, which hinders its acceptance in the clinical domain, where predicted outcomes can significantly impact patient life [8]. Explainable AI (XAI) techniques, such as SHAP, enhance interpretability by highlighting retinal lesions and regions of interest that support DR severity predictions, thereby improving model transparency and fostering clinical trust.

1.4. Handheld and Smartphone-based DR screening

Handheld fundus cameras and modern smartphones equipped with fundus lenses are alternatives to OCT and Fundus cameras on the DR screen. In the studies [12, 13, 15], hand-held devices for DR detection demonstrated the effectiveness in detecting DR severity stages when powered by deep learning models. This approach is potentially useful for DR screening in lower- and middle-income countries

(LMICs), where DR screening is hindered by the high cost of OCT and fundus cameras, as well as challenges related to the portability of these devices in remote areas and the scarcity of ophthalmologists specializing in DR treatment.

1.5. Diabetes Retinopathy Screening in Rwanda

Rwanda is a country in the Low and Middle-Income Countries (LMICs) located in East Africa. In Rwanda, the Early diagnosis of diabetic retinopathy (DR) remains a major challenge due to the reliance on high-precision imaging technologies, such as Optical Coherence Tomography (OCT) and fundus cameras, which are approved by the World Health Organization (WHO) for non-invasive DR diagnosis [16]. Additionally, the shortage of trained ophthalmologists further limits DR screening and early treatment. While OCT and fundus cameras provide high-quality retina images for DR analysis, their high cost and portability issues hinder widespread DR screening in Rwanda. Additionally, the country has a low rate of diabetes screening [17], which increases the number of patients seeking DR screening.

1.6. Study objectives

This study proposes a low-cost diabetic retinopathy (DR) screening approach using deep learning (DL) with explainable AI (XAI) methods to enhance model transparency and clinical applicability. The model enables non-ophthalmologists, such as nurses in primary care settings, to detect referable DR cases, while specialists can assess DR severity using domain expertise, the outcome of prediction, and the model's generated explanations. Retinal images are classified into five DR stages based on stage-specific characteristics (Figure 1), supported by lesion detection using SHAP. A convolutional neural network (CNN) is trained on a public dataset using transfer learning and validated with retinal images from the Brazil Retinopathy Dataset (BRSET) and using a random sample of unlabeled retina images. These images are annotated by an ophthalmologist specializing in DR treatment from the CHUK hospital.

The key contributions are: (1) applying five pre-trained CNNs with transfer learning for DR screening; (2) using the same models to predict DR severity stages; (3) employing SHAP for model interpretability; and (4) deploying the best-performing models in a Python Django web application. The remainder of this paper is structured as follows: Section 2 details the methodology and recent related work, Section 3 presents experimental results, Section 4 discusses the findings, and Section 5 concludes our paper with future research directions.

2. Methodology

2.1.Study Design

The study was conducted in five phases, as illustrated in Figure 1. The workflow begins with image preprocessing and dataset augmentation, followed by training and evaluating multiple pre-trained models.

The best-performing model is then subjected to interpretability analysis using SHAP. External validation is performed using independent image datasets from two sources. The validated model is then deployed in a web-based application for practical use:

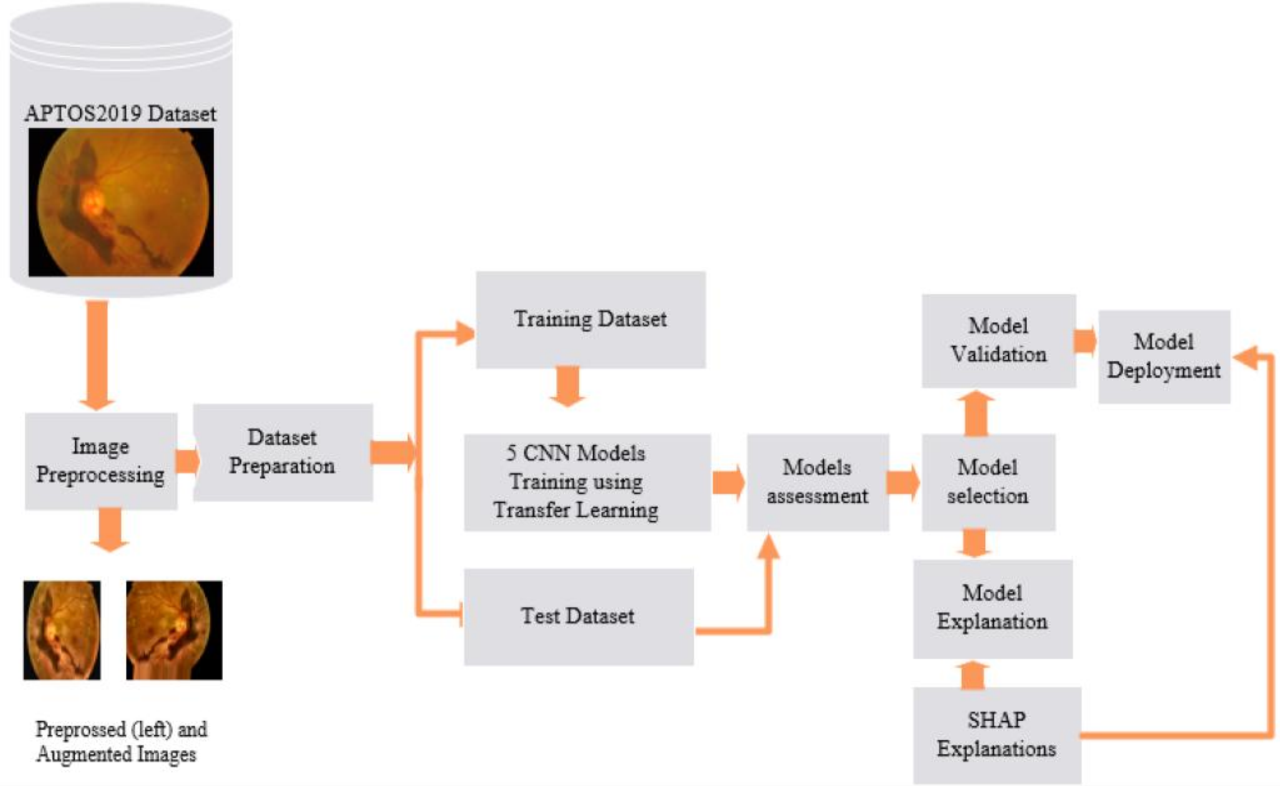


Figure (1) Study Design

2.2.Data Acquisition

This study utilized the publicly available Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 Blindness Detection dataset, which comprises 3,662 fundus images collected by Aravind Eye Hospital in rural India. Images were labeled by trained ophthalmologists according to the International Clinical Diabetic Retinopathy Disease Severity Scale (ICDRSS), with five DR categories: No DR, Mild, Moderate, Severe, and Proliferative DR. An additional 270 unlabeled images from the same dataset, originally released during the 2019 Kaggle challenge, were randomly selected and annotated by a DR specialist from CHUK for comparison with model predictions. Agreement was measured using Cohen's Kappa.

For external validation, a sample of 270 handheld fundus images from the Brazil dataset was used, categorized into the same five DR stages, and evaluated against predictions from the DenseNet121 model. Figure 2 illustrates the severity stages of DR, and Figure 3 visualizes the distribution of these stages in the original dataset.

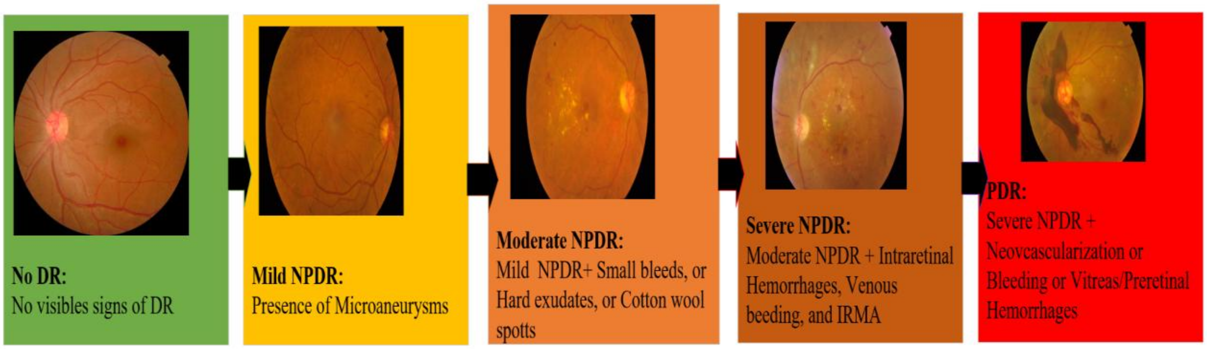

No DR:
No visibles signs of DR
Mild NPDR:
Presence of Microaneurysms
Moderate NPDR:
Mild NPDR+ Small bleeds, or Hard exudates, or Cotton wool spotts
Severe NPDR:
Moderate NPDR + Intraretinal Hemorrhages, Venous beeding, and IRMA
PDR:
Severe NPDR + Neovcascularization or Bleeding or Vitreas/Preretinal Hemorrhages

Figure (2) DR progression into severity stages

This study utilized the APTOS 2019 image dataset. Figure 3 illustrates the class distribution across five DR severity stages: NoDR, Mild, Moderate, Severe, and Proliferative DR (PDR). Descriptions of these stages are provided in Figure 2.

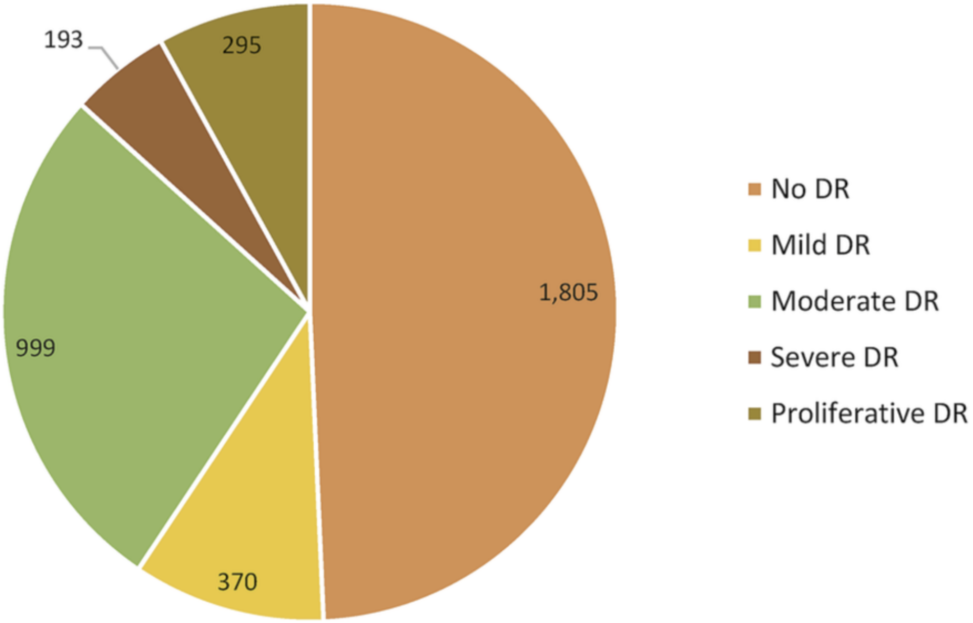


Figure (3) Description of APTOS 2019 Retinopathy Dataset

The original dataset was imbalanced, with a majority of No DR cases. This was addressed during the image augmentation phase. For binary classification, the four DR severity stages were combined into a single DR class, and No DR remained separate. After augmentation, the binary classification dataset consisted of 7,100 images (3,750 with No DR and 3,350 with DR). The multiclass dataset included 15,622 images across five DR stages. Figure 4 shows the class distribution in the multiclass dataset.

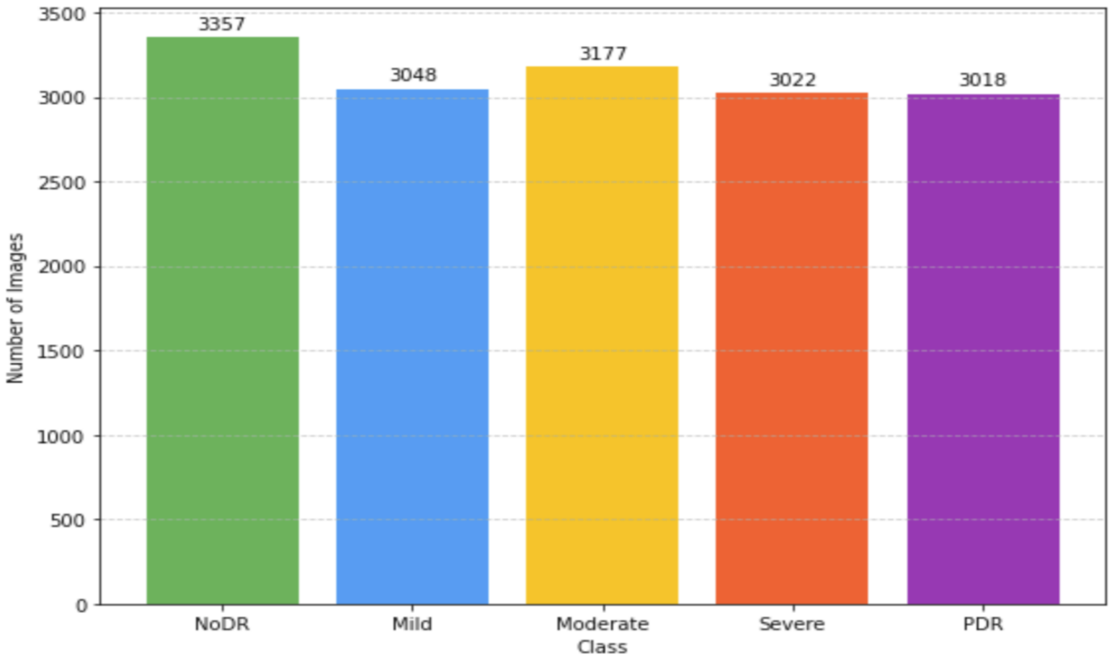


Figure (4) Class distribution in a multiclass dataset for DR severity stages prediction after image augmentation.

2.3 Methods

This study consisted of five stages: image preprocessing, DR classification using binary and multi-class pre-trained CNN models, explanation of the multi-class classifier using SHAP methods, model validation with CHUK image data, and deployment of both models in a Python-based Django web application.

2.3.1. Image Preprocessing

Image preprocessing is a critical step in classification tasks. Images were resized to (224, 224, 3), ensuring a square shape with three color channels. Enhancement was performed using Contrast Limited Adaptive Histogram Equalization (CLAHE) as used in [18], followed by Gaussian blurring and two-stage denoising with Non-Local Means (NLM) as detailed in [19] and Adaptive Median Filters (AMF) described in [20]. The final phase of preprocessing involved image augmentation through rotation, scaling, translation, shearing, and contrast adjustments, increasing the dataset from 3,662 to 15,622 images. Figure 5 shows the sample images after the preprocessing stages. The images after preprocessing meet the quality requirements of CNN models. Finally, images were normalized into a pixel array for model input. The resulting dataset was split into training (80%) and testing (20%) sets.

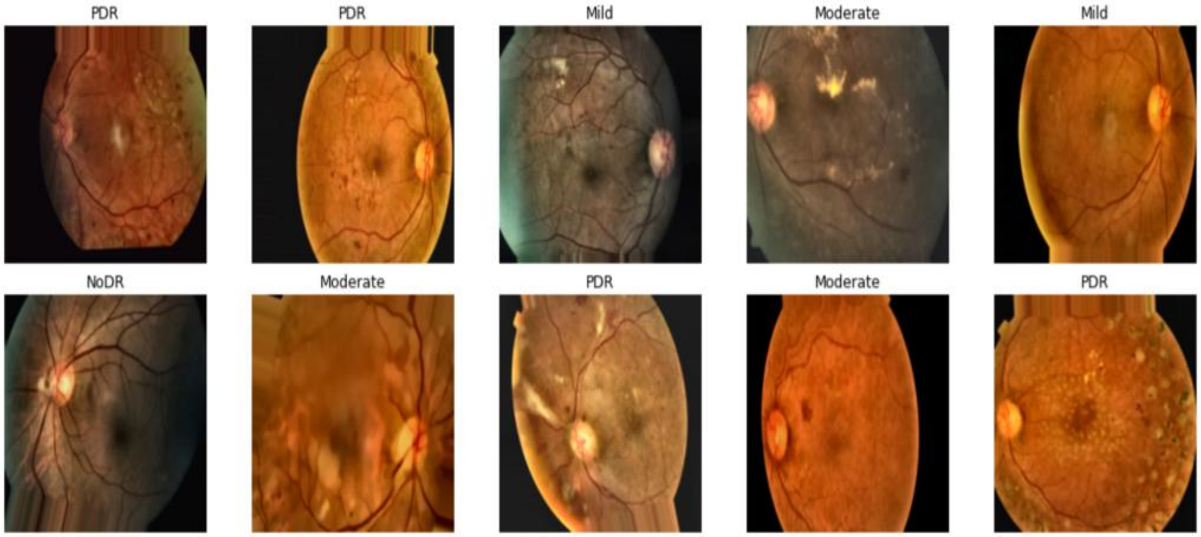


Figure (5). Sample of retina images in the training dataset

The combination of Gaussian filtering, CLAHE, blurring, NLM, and Adaptive Median Filter (AMF) produced denoised retinal images with preserved edges, essential for identifying regions of interest.

2.3.2. Retinopathy classification

This study employed five popular pre-trained CNN models, including DenseNet121, MobileNetV2, VGG19, InceptionV3, and Xception, for binary and multiclass DR classification. Binary classification identifies referable DR cases during screening, assuming images are uploaded from handheld devices or OCT machines at primary healthcare centers by non-specialists. Detected cases are then referred to DR specialists. In the second stage, multi-class classification determines the DR severity stage, helping specialists confirm diagnoses and schedule patient appointments for further treatment. The choice of these models is based on their widespread use in image classification, particularly their popularity in DR detection [21, 22].

2.3.3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [23] are designed to process grid-like data, such as images, using small filter kernels that perform convolutions. Figure 6 illustrates a typical CNN architecture, where an input image of shape (224, 224, 3) passes through convolutional layers, ReLU activations, and pooling layers. A dropout layer helps prevent overfitting, followed by a fully connected layer made of Flatten and Dense layers. The output layer uses Softmax to predict the most probable class, Proliferative Diabetic Retinopathy (PDR) in this CNN architecture.

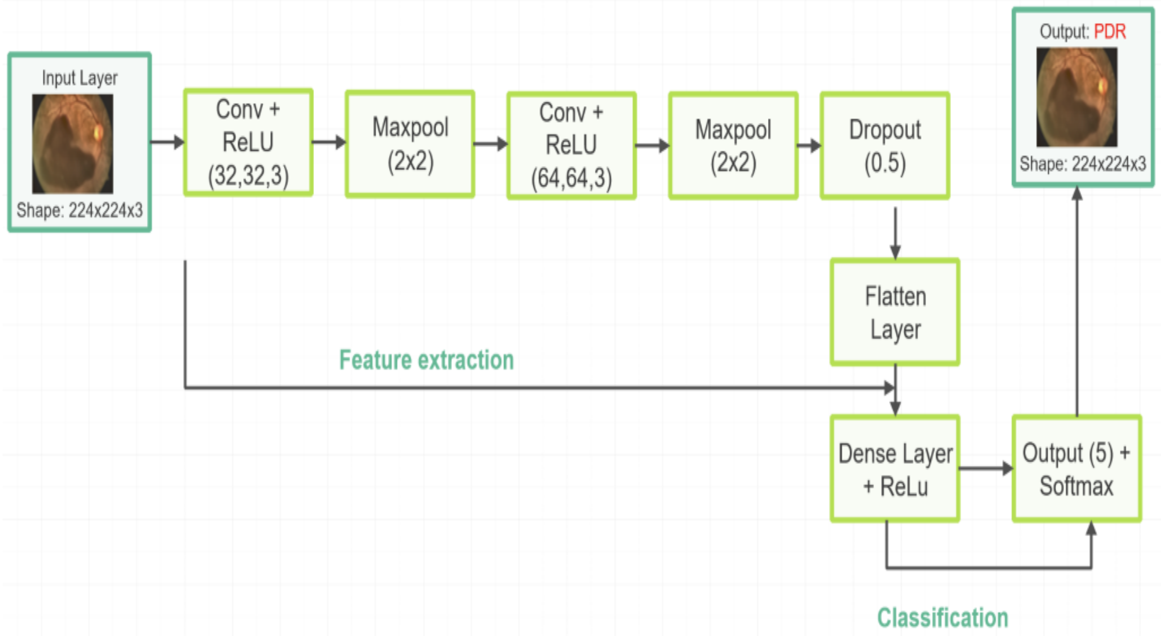


Figure (6) Architecture of a CNN model with input, feature extraction, and classification layers.

2.3.4. Transfer Learning

Transfer learning in deep learning involves reusing models trained on large datasets for related tasks [24]. In CNNs, it enables the adaptation of learned features by modifying the final layers. This approach is particularly effective when training data and computational resources are limited. In our study, we employed transfer learning by leveraging feature extraction layers from CNN models pre-trained on ImageNet [25]. We then replaced the top layers with custom classification and output layers tailored to retinal image analysis. This allowed efficient and accurate model adaptation to our specific diagnostic task.

2.4. Pre-trained CNN Models

Pre-trained models, such as those trained on ImageNet, can be adapted for various computer vision tasks, including retinal image classification. They have shown strong performance in extracting informative features from natural images [26]. To customize them for tasks like diabetic retinopathy screening, the top fully connected and output layers are replaced with new layers and retrained on the target dataset.

2.4.1. DenseNet121

DenseNet, short for Densely Connected Convolutional Networks, was introduced in [27] and is known for its high performance due to two key features: dense blocks, where each layer connects to all subsequent layers in a feed-forward manner, and bottleneck layers, which reduce parameters without compromising feature learning. DenseNet121, a specific variant of DenseNet architecture with 121 layers, includes an initial convolutional layer (1), dense layers (116), transition layers (3), and a classification layer (1), as illustrated in Figure 7.

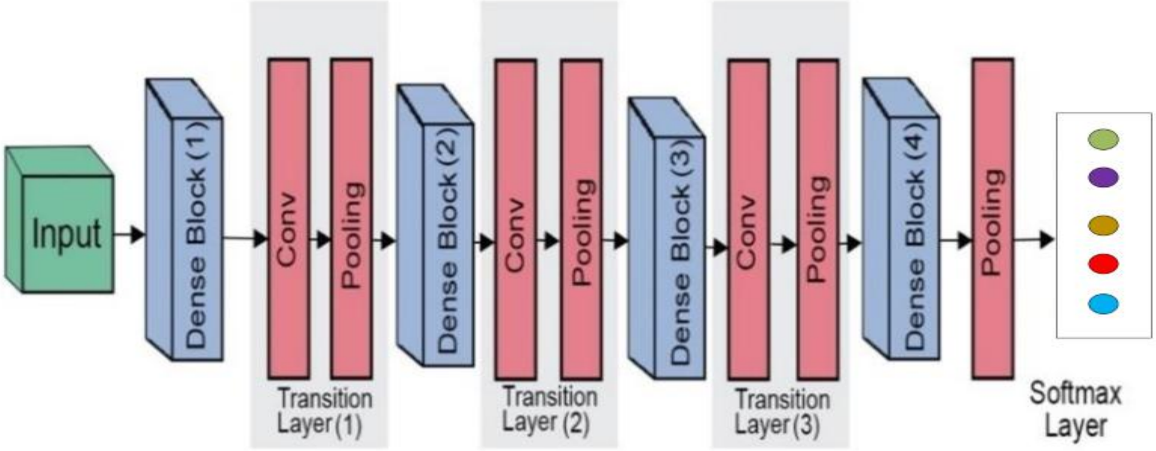


Figure (7) Architecture of DenseNet121 showing dense block layers, transition layers, and output layer with 5 class categories.

2.4.2. MobileNetV2

MobileNetV2 is a lightweight, pre-trained CNN model optimized for computer vision applications in smartphones and embedded devices. Building on MobileNetV1, it improves performance while maintaining low computational cost. Its efficiency is rooted from a bottleneck architecture that reduces parameters and computation compared to traditional CNNs [28]. As shown in Figure 8, MobileNetV2 consists of three main components: a stem block with an initial convolutional layer, a body of linear residual bottleneck blocks (each containing expansion, depth-wise, and projection convolutions), and a final classification layer built on extracted feature maps to be used in fully connected and output layers.

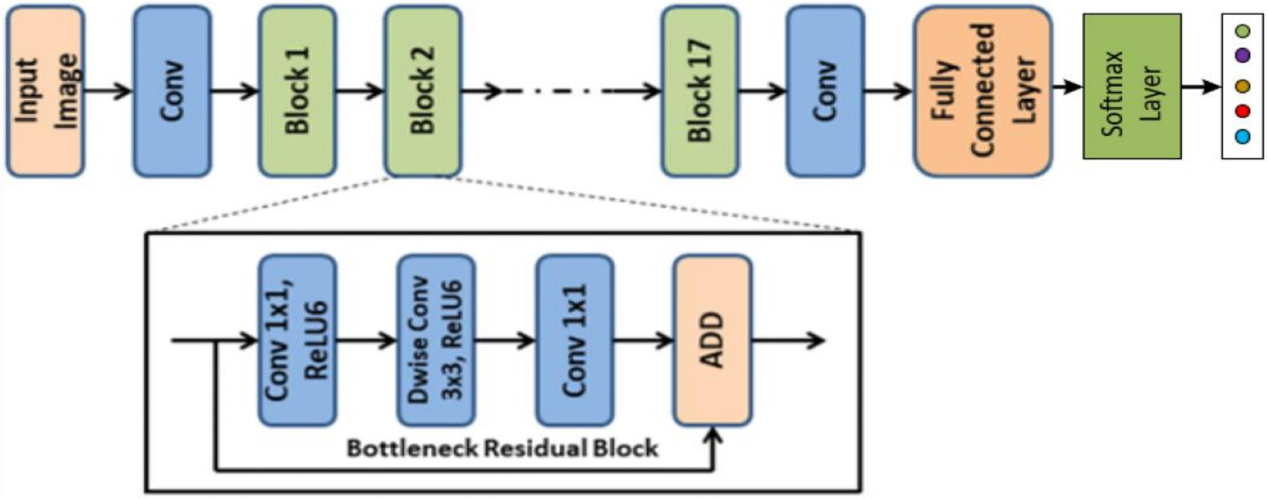


Figure (8) Architecture of MobileNet V2 showing Bottleneck Residual Layers

2.4.3. VGG19

VGG19 [29], an enhanced version of VGG16, was developed by the Visual Geometry Group (VGG). It features 16 convolutional layers, each followed by a 2×2 max pooling layer with a stride of 2, and 3 dense layers with 4096 neurons each, followed by an output layer. We employed a Softmax activation with 5 neurons to represent the 5 DR severity stages in our study. As illustrated in Figure 9, VGG19's deep architecture is expected to effectively capture low-level image features such as edges, textures, and regions of interest.

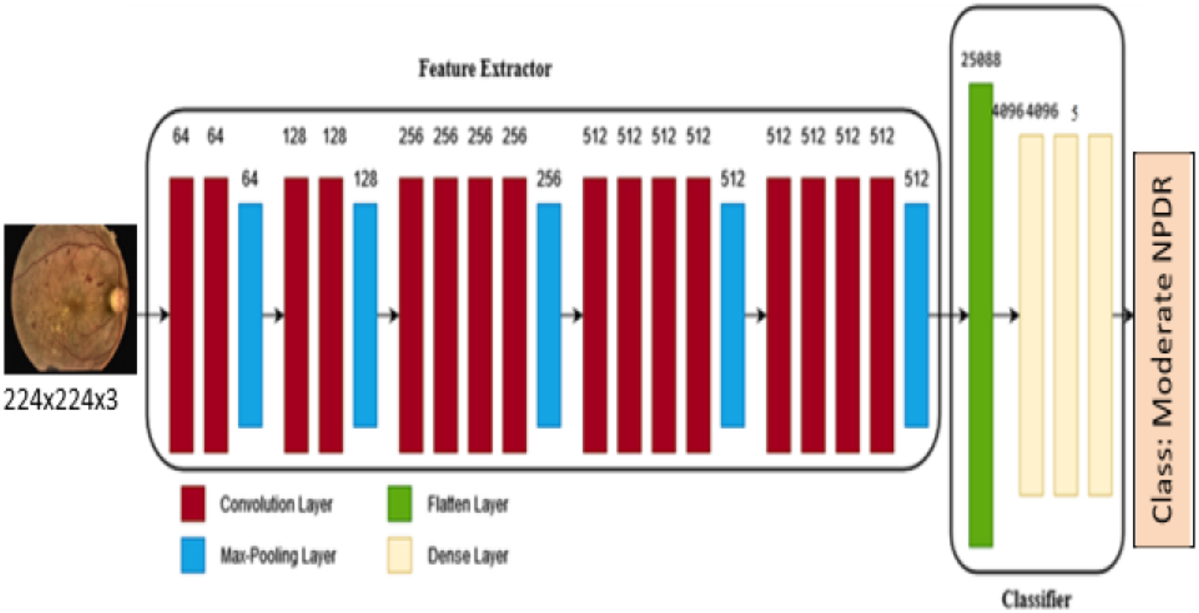


Figure (9) Architecture of VGG19 for predicting DR severity class category

2.4.4. InceptionV3

InceptionV3 [30], developed by Google as part of the GoogLeNet family, is a deep CNN architecture that employs parallel convolutional layers with varying kernel sizes to efficiently capture diverse image features. Its inception modules enhance performance and reduce computational cost through techniques such as factorization, batch normalization, and label smoothing. Figure 10 shows how the InceptionV3 architecture combines normal convolutions in the initial layers, followed by the inception modules. In our study, the retrained layers comprise the final pooling layer, the Fully Connected (FC) layer, and the output layer (Softmax).

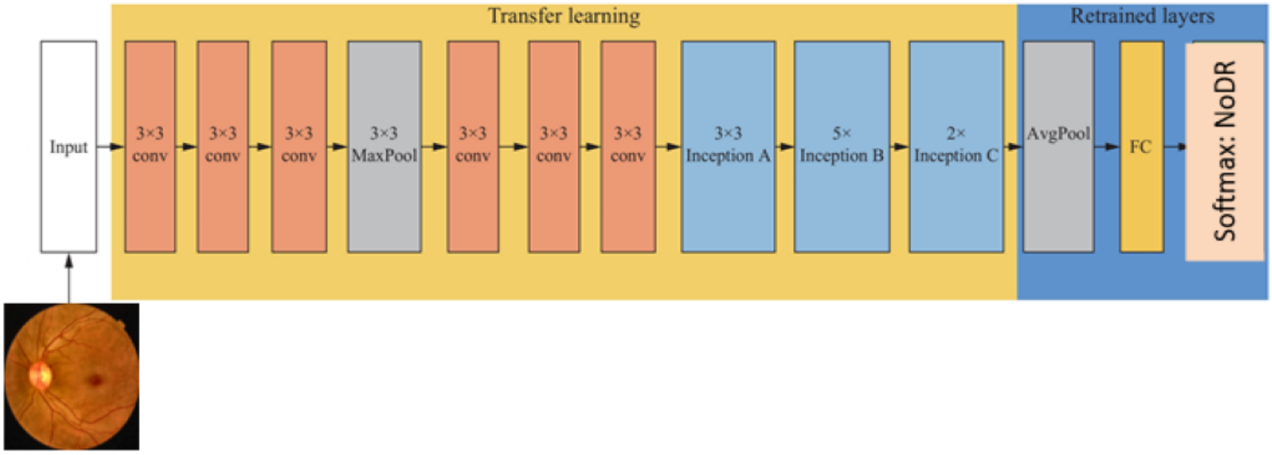


Figure 10. Architecture of InceptionV3 CNN model for DR severity classification

2.4.5. Xception

Xception [31], short for Extreme Inception, is a convolutional neural network (CNN) architecture that extends the principles of the Inception architecture by employing depthwise separable convolutions. This approach decouples spatial and cross-channel correlations by first applying depthwise convolutions to learn spatial relationships independently within each channel, followed by pointwise (1×1) convolutions to capture inter-channel correlations. Inspired by the residual connections in ResNet, Xception integrates skip connections between its modules to enhance gradient flow and learning stability. The architecture comprises 36 convolutional layers, organized into three main modules, as illustrated in Figure 11. The Entry Flow, Middle Flow, and Exit Flow. Unlike the parallel structure of Inception modules, Xception follows a linear stack of depth-wise separable convolution modules, each reinforced with residual connections. This design yields improved performance while significantly reducing computational complexity.

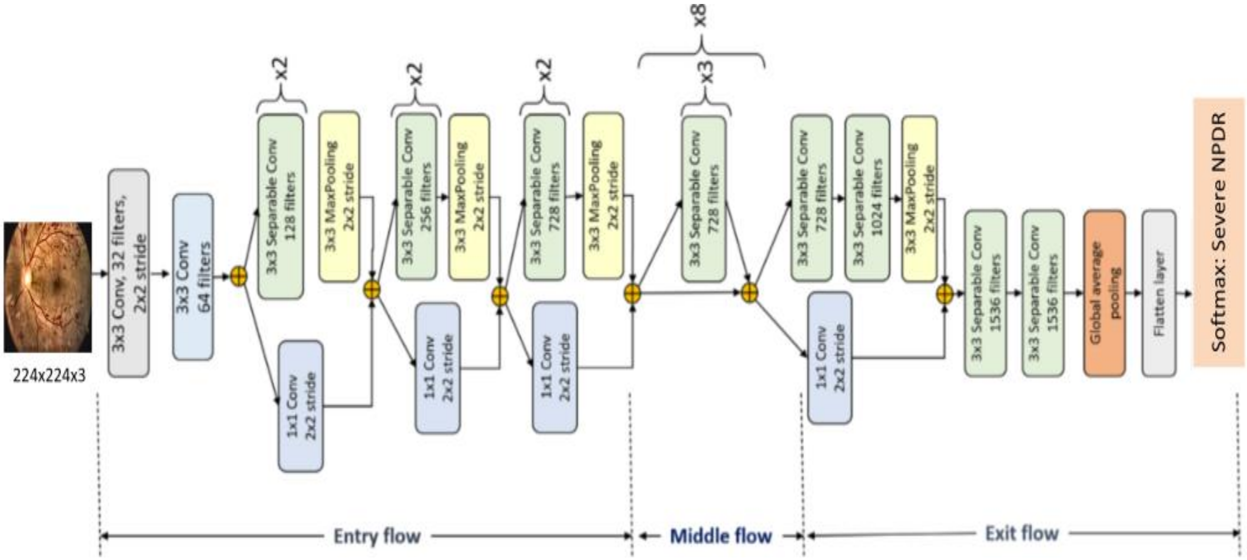

224x224x3
3x3 Conv, 32 filters, 2x2 stride
3x3 Conv 64 filters
x2
3x3 Separable Conv 128 filters
3x3 MaxPooling 2x2 stride
1x1 Conv 2x2 stride
x2
3x3 Separable Conv 256 filters
3x3 MaxPooling 2x2 stride
1x1 Conv 2x2 stride
x2
3x3 Separable Conv 728 filters
3x3 MaxPooling 2x2 stride
1x1 Conv 2x2 stride
x8
x3
3x3 Separable Conv 728 filters
3x3 Separable Conv 728 filters
3x3 Separable Conv 1024 filters
3x3 MaxPooling 2x2 stride
1x1 Conv 2x2 stride
3x3 Separable Conv 1536 filters
3x3 Separable Conv 1536 filters
Global average pooling
Flatten layer
Softmax: Severe NPDR
Entry flow
Middle flow
Exit flow

Figure (11) Architecture of Xception CNN Model

During the experimentation, the selected pre-trained CNN models were fine-tuned by adding a global average pooling layer, a dense layer of 512 neurons, a regularization factor to prevent overfitting, and an output layer as indicated in Table 1.

2.5. CNN Model Evaluation Metrics

Model performance was evaluated using six metrics: accuracy, precision, recall, F1-score, AUC, and Cohen's Kappa. The first five metrics evaluated the models on a test dataset comprising unseen images of the same quality as the training data. The equations (1-5) illustrate how the evaluation metrics are computed and how AUC and ROC-AUC curves are calculated using the probability distributions of True Positive Rate (TPR) and False Positive Rate (FPR), as shown in Figure 7. Cohen's Kappa is a statistical measure used to quantify agreement between two raters or graders and is commonly applied in machine learning to evaluate model predictions against ground truth [32]. In this study, Cohen's Kappa coefficient was used to assess the agreement between DR severity predictions from the best multiclass classifier and evaluations by an ophthalmologist specialized in DR treatment.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

Where TP is the true positive, TN the true negative, FP false positive, and FN is the false negative. In TF and TN, the model is in perfect agreement with the ground truth. FP and FN indicate the disagreement between the model and the ground truth.

$$k = \frac{(po - pe)}{1 - po} \tag{5}$$

where k is the Kappa statistic po is the observed agreement ratio considered as classification accuracy since it measures the number of predicted labels that agree with the true labels, and pe is the expected agreement when both graders assign labels randomly is calculated as follows:

$$pe = \frac{1}{N^2} \sum_{k=1} n_{k1} n_{k2} \tag{6}$$

N is the total number of classification attempts, K is the total class categories (5 categories in our study), n_{k1} the number of times the label k appears in the prediction and n_{k2} is the number of times k is the true label (label assigned by a specialist in this study).

2.6. Explainability Methods

The emergence of Explainable AI (XAI) has transformed healthcare by improving transparency and building trust in AI-driven disease diagnosis and treatment. SHAP, a widely used model-agnostic XAI technique based on cooperative game theory, belongs to the class of additive feature attribution methods [33]. It assigns each feature a contribution toward the prediction outcome using Shapley values, computed as follows:

$$\varphi_{i(f)} = \sum\nolimits_{S \subseteq \cup \mathrm{N} \setminus \{i\}} \frac{|s|!(N-|s|-1)!}{N!} [f_x(s \cup \{\mathrm{i}\}) - f_x(s)] \tag{7}$$

Here, $\boldsymbol{S}$ is a subset of features excluding feature $\boldsymbol{i}$, and $\boldsymbol{N}$ is the total number of features. $\boldsymbol{\phi_i(f)}$ represents the contribution of feature $\boldsymbol{i}$, calculated as the difference between the total prediction and the prediction without feature $\boldsymbol{i}$.

2.7. Model Validation

Model validation is essential for assessing the generalizability of machine learning systems. In this study, our convolutional neural network (CNN) models were validated using two independent datasets. The first consisted of 270 randomly selected retinal images from a publicly available Brazilian dataset obtained under a PhysioNet license. The second included 270 unlabeled images from the APTOS 2019 Kaggle challenge, which were subsequently annotated by a diabetic retinopathy (DR) specialist from CHUK with clinical experience in DR diagnosis. Both validation datasets underwent consistent preprocessing to ensure comparable image quality, with no augmentation applied. Model predictions on both sets were evaluated using standard performance metrics and compared against clinical diagnoses, with agreement assessed using Cohen's Kappa coefficient.

2.8. Model Deployment

The best-performing model for both binary and multiclass classification of diabetic retinopathy has been integrated into a web application. This platform allows non-ophthalmologists to upload retinal images and receive predictions indicating whether a case is referable or non-referable. Ophthalmologists can then review the uploaded images, examine the predicted severity stage, and interpret the accompanying visual explanations, supporting accurate clinical decision-making.

2.9. Related works

The growing availability of public retinal image datasets, improvements in computational resources, and the success of deep learning in image classification have driven significant advances in the detection of early diabetic retinopathy (DR). In the study [34], a deep learning model combining VGG and ResNet architectures was trained on a dataset of 76,370 retinal images from the Singapore Integrated Diabetic Retinopathy Program. The model was validated using 4,504 retinal images from diabetic patients in Zambia. It was designed to detect referable diabetic retinopathy, specifically cases starting at the Moderate NPDR stage according to the ICDR Severity Scale.

Model interpretability was achieved using the Integrated Gradients method. The model demonstrated strong performance, achieving an AUC of 0.983, a sensitivity of 98.8%, and a specificity of 82.0%. In 2021, a study was conducted to investigate the feasibility and acceptance of artificial intelligence-based diabetic retinopathy (DR) screening across four diabetes clinics in Rwanda [35] as part of the Rwanda Artificial Intelligence for Diabetic Retinopathy Screening (RAIDERS) project. A deep learning model was developed and trained on a dataset of 92,073 fundus images. The model was deployed for DR screening among patients who consented to participate in the study. A total of 827 individuals with type 1 and type 2 diabetes were enrolled. For the detection of referable DR cases, the model achieved a sensitivity of 92%, a specificity of 85%, and an AUC of 0.82.

The study also reported high levels of participant satisfaction with the AI-based screening process. An explainable deep learning model was developed in [36] for DR grading. This study used Gradient-based Class Activation Map (GradCAM) and SHAP explanation methods to visualize the predictions made by DL models built using the pre-trained CNN architectures in the form of ensemble models. The best model utilized the DenseNet121 CNN architecture and achieved an AUC of 97.80 on the APTOS 2019 dataset. The authors of the study [37] employed two methods of image quality enhancement, namely Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) and Contrast Limited Adaptive Histogram Equalization (CLAHE), to improve the visual quality of the APTOS image dataset before it was used in their proposed CNN architecture. The experiment was done in four scenarios, and the highest experimental results achieved an accuracy of 97.83% using 549 images from APTOS.

3 Results

3.1.Experimental Setup

The models developed for binary and multiclass classification of diabetic retinopathy utilize the same configurations as those presented in Table 1, with the exception of the activation function in the final classification layer. The sigmoid function was used for binary classification, and the Softmax function was used for 5-class DR severity classification.

Table (1) Configuration of pre-trained CNN models

S. No	Hyperparameter	Value
1.	Additional Layers	Global Average Pooling Layer, Dense Layer with 512 neurons and ReLU, Dropout Layer with 0.2
2.	kernel_regularizer	L2 with 0.001
3.	Optimizer	Adam with learning_rate=0.0001, epsilon=1e-07
4.	loss	binary_crossentropy and categorical_crossentropy
5.	EarlyStoping	monitor='val_loss', patience=10, restore_best_weights=True
6.	Epoch	50
7.	batch_size	64

3.2.Binary classification models used for DR Screening

We evaluated the performance of the models on identifying DR referable cases. Table 1 shows that the best model is a deep learning model that uses DenseNet121 architecture, and its confusion matrix is visualized in Figure 6.

Table (2) Models performance evaluation on 5 metrics

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
MobileNetV2	95.00	96.00	95.00	96.00	98.00
InceptionV3	95.00	96.00	94.00	95.00	98.00
Xception	95.00	94.00	96.00	95.00	99.00
VGG19	95.00	95.00	95.00	95.00	98.00
DenseNet121	**97.00**	**97.00**	**97.00**	**97.00**	**99.00**

The best-performing model accurately identified diabetic patients at risk of developing diabetic retinopathy with 97% accuracy. This high level of performance makes the model a valuable tool for use by non-specialists in broader DR screening efforts within the diabetic population.

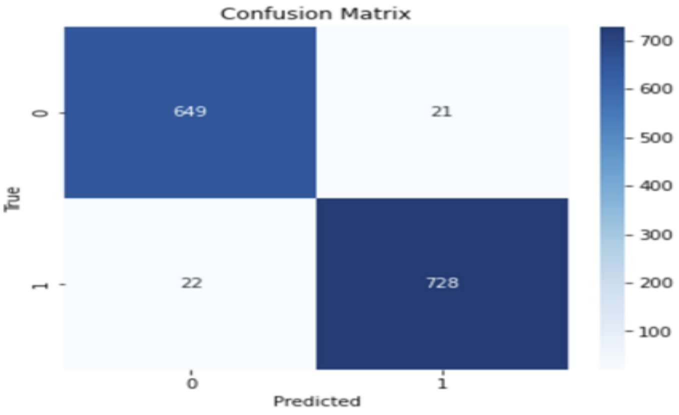


Figure (12) Confusion Matrix of DenseNet121 Model

3.3 Multiclass classification used for DR Severity Grading

As shown in Table 3, all selected models achieved an AUC of 99% in correctly identifying patients with normal retinas, which is as expected given the absence of DR indicators in such images. The models also performed well in detecting advanced DR stages (Severe and PDR), where clear pathological signs aid recognition. However, performance declined for the Mild and Moderate stages due to subtle differences that are harder to distinguish. This challenge is consistent with clinical observations, as noted by the DR severity grader from CHUK involved in the study.

Table (3). Area Under Curve (AUC) score for 5 models on predicting DR severity stages

	No DR (%)	Mild (%)	Moderate (%)	Severe (%)	PDR (%)
MobileNetV2	**99.00**	**96.00**	**92.00**	**98.00**	**96.00**
InceptionV3	99.00	92.00	83.00	92.00	89.00
Xception	99.00	93.00	85.00	95.00	92.00
VGG19	99.00	88.00	80.00	88.00	86.00
DenseNet121	**99.00**	**96.00**	**92.00**	**98.00**	**97**

DenseNet121 outperformed the other models, followed by MobileNetV2. It was selected for further evaluation, with its performance assessed across all metrics (Table 4) and illustrated by the five-class confusion matrix in Figure 14. The learning curves indicate good generalization on the validation set during training limiting the degree of overfitting as demonstrated in Figure 13.

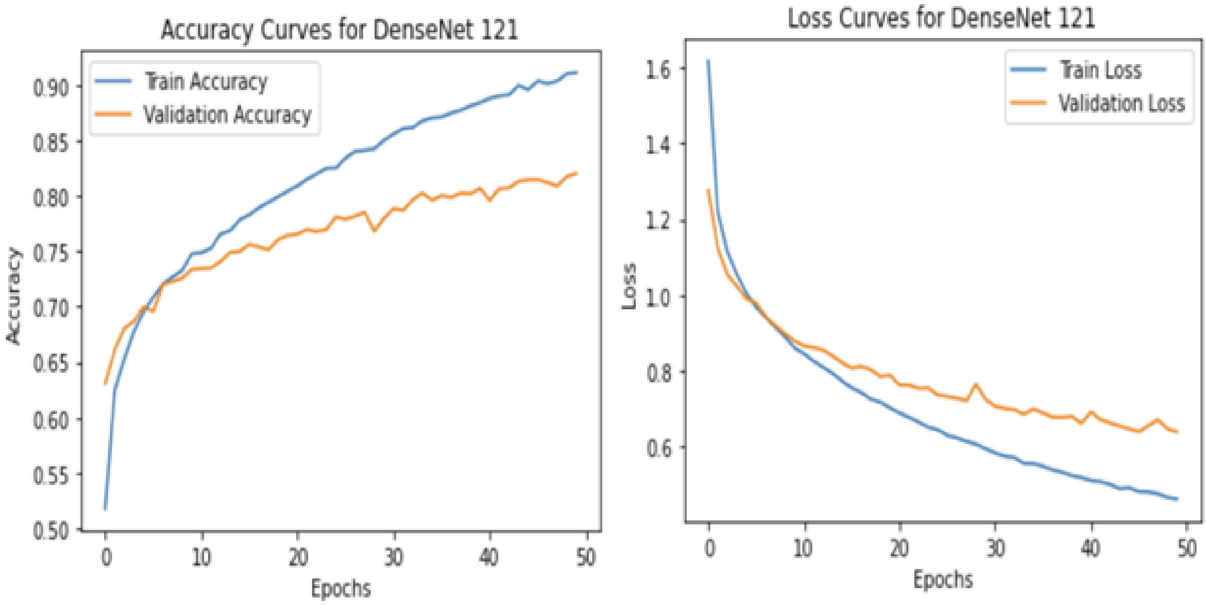


Figure (13). Learning Curve of DenseNet121 on Accuracy (Left) and Loss (Right)

The model demonstrates good performance in recognizing all DR severity stages, with slightly lower Recall and F1-score metrics for the Moderate NPDR stage, as highlighted in Table 4.

Table (4) DenseNet121 Model performance on five metrics for DR severity grading

	Accuracy	Precicision	Recall	F1-score	AUC
NoDR	97.00	92.00	96.00	94.00	99.00
Mild	92.00	76.00	84.00	80.00	96.00
Moderate	89.00	80.00	60.00	69.00	92.00
Severe	94.00	81.00	87.00	84.00	98.00
PDR	92.00	79.00	81.00	80.00	97.00

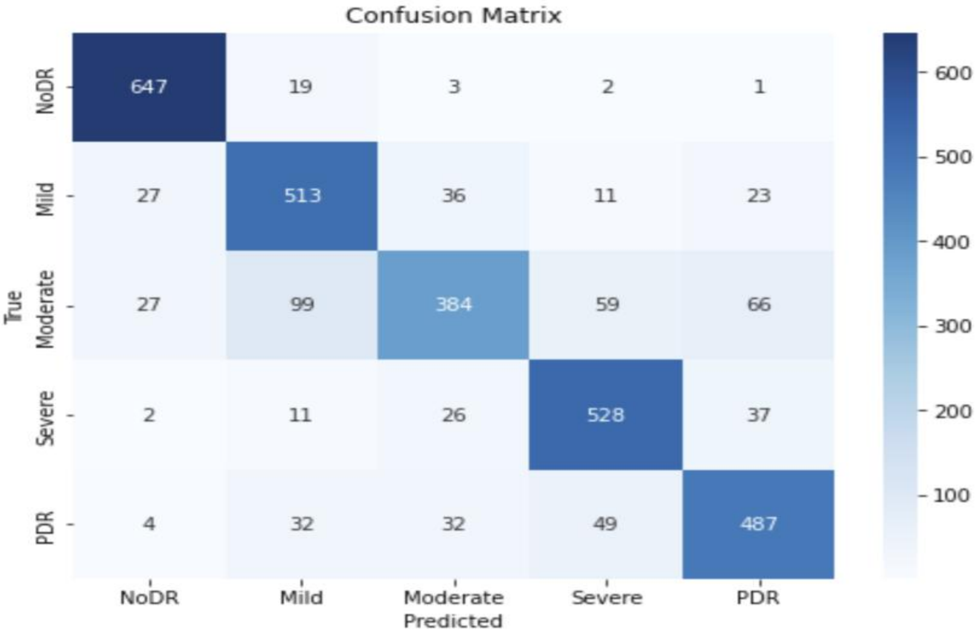


Figure (14) Confusion Matrix of DenseNet121 Model on DR Severity Grading

3.4 Model Validation

We conducted a two-phase validation of our model. In the first phase, we utilized a sample of 270 retinal images from the Brazil dataset (BRSET), which comprises 16,266 retinal images collected from 8,524 diabetic patients, along with associated sociodemographic data. These images were captured using handheld retinal imaging devices, and the severity of diabetic retinopathy (DR) was graded according to both the International Clinical Diabetic Retinopathy (ICDR) scale and the Scottish Diabetic Retinopathy Grading (SDRG) system. In the second phase, we randomly selected 270 retinal images from the APTOS 2019 Kaggle Challenge dataset, originally consisting of 1,928 unlabeled images. These selected images were subsequently annotated by an ophthalmologist specializing in diabetic retinopathy at CHUK Hospital. In both validation phases, we assessed the agreement between the model's predictions and expert annotations using Cohen’s Kappa coefficient.

3.4.1 Model validation using the Brazil dataset

Table 5 presents the model's performance on a validation subset of the Brazil retinopathy dataset, which includes 270 retinal images captured with handheld devices. The sample covers all DR severity stages and was preprocessed in a similar way to images used in the model training.

Table (5). DenseNet121 Model performance on validation dataset from Brazil

	Accuracy	Precision	Recall	F1-score	AUC
NoDR	96.00	86.00	93	89	98.00
Mild	90	77.00	77.00	77.00	90.00
Moderate	91	85.00	72.00	78.00	90.00
Severe	92.00	62	94.00	75.00	93.00
PDR	95.00	1.00	80.00	89.00	95.00

The model achieved a slightly similar performance with significant improvement in detecting Mild and Moderate NPDR stages. The confusion matrix of the model performance on the Brazil dataset is presented in Figure 15.

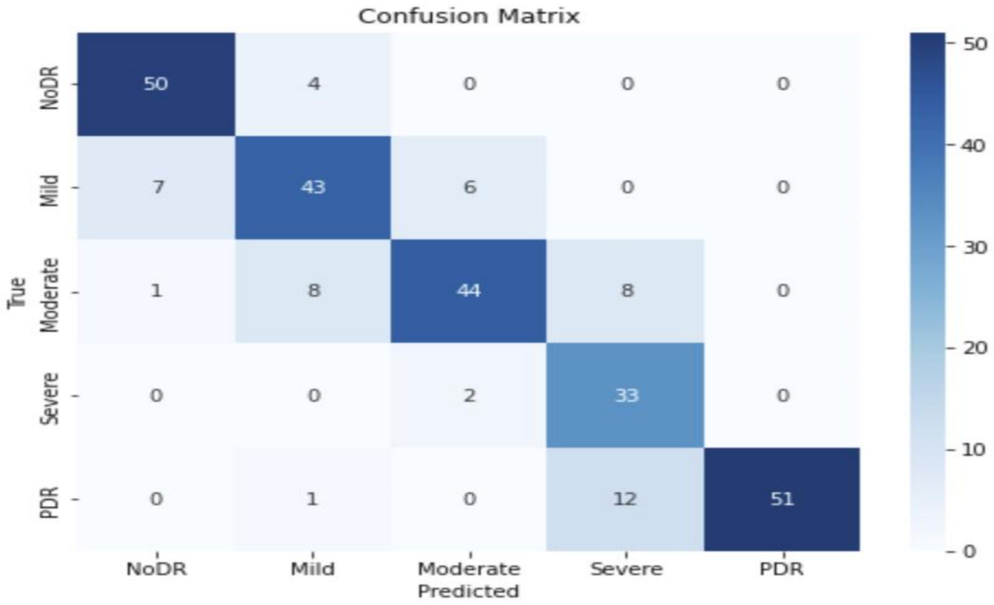


Figure (15) Confusion Matrix of DenseNet121 model on Brazil Dataset for DR grading.

Table 6 presents Cohen's Kappa results, indicating the level of agreement between the model's predictions and expert human grading of diabetic retinopathy severity.

Table (6) Cohen's Kappa for Brazil Dataset grading DR severity

DR Severity	**Cohen's Kappa**
NoDR	0.865
Mild	0.707
Moderate	0.721
Severe	0.704
PDR	0.857

An average per-class Cohen’s Kappa score of 0.771 indicates substantial agreement between our model and human graders, suggesting that the model's performance in DR severity classification is significantly better than random chance. A Kappa value of 1.00 represents perfect agreement between the model and human evaluation. High levels of agreement were observed at the extreme stages of the severity spectrum, particularly for the NoDR and PDR stages.

3.4.2 Model validation using APTOS test dataset

We evaluated the proposed model’s performance in predicting DR severity using a dataset of 270 retinal images annotated by a diabetic retinopathy (DR) specialist from CHUK. The expert annotations served as the ground truth for this second validation phase. Table 7 summarizes the model's performance across different DR severity levels. Notably, the model exhibited lower precision and recall for the Moderate and PDR stages.

Table (7) DenseNet121 Model performance on validation dataset annotated by a specialist from CHUK

	Accuracy	Precision	Recall	F1-score	AUC
NoDR	95.00	91.00	80.00	85.00	89.00
Mild	97	67.00	92.00	77.00	95.00
Moderate	81.00	52.00	89.00	65.00	84.00
Severe	87.00	90.00	73.00	81.00	84.00
PDR	89.00	82.00	52.00	64.00	75.00

Figure 16 presents the confusion matrix of the proposed model evaluated on the second validation dataset comprising 270 retinal images.

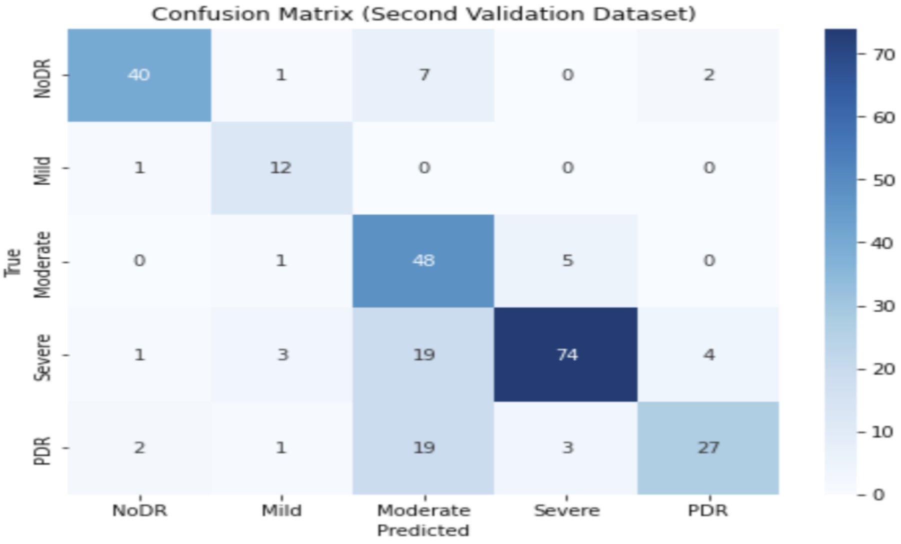


Figure (16) Confusion matrix of DenseNet121 model on the second validation dataset.

Table 8 presents the comparison between the proposed model and a CHUK specialist on DR severity grading for randomly selected retinal images from the Kaggle 2019 Challenge, using Cohen’s

Kappa as the evaluation metric.

Table (8) Cohen’s Kappa for grading DR severity on the dataset graded by a DR specialist from CHUK

DR Severity	Cohen’s Kappa
NoDR	0.820
Mild	0.761
Moderate	0.536
Severe	0.712
PDR	0.571

The average per-class Cohen’s Kappa score of 0.680 suggests that the model demonstrates moderate performance, with some improvement over random classification. Upon closer analysis of the lower Kappa scores for the Moderate DR and PDR categories, we found that the model often misclassified PDR cases as Moderate when neovascularization was present. This is likely due to the limited representation of such cases in the training dataset. Figure 17 illustrates an instance of disagreement between the DR specialist from CHUK and the model. Additionally, in certain cases, the model's prediction appeared to be more accurate than the human annotation, potentially due to human error, as highlighted in Figure 18.

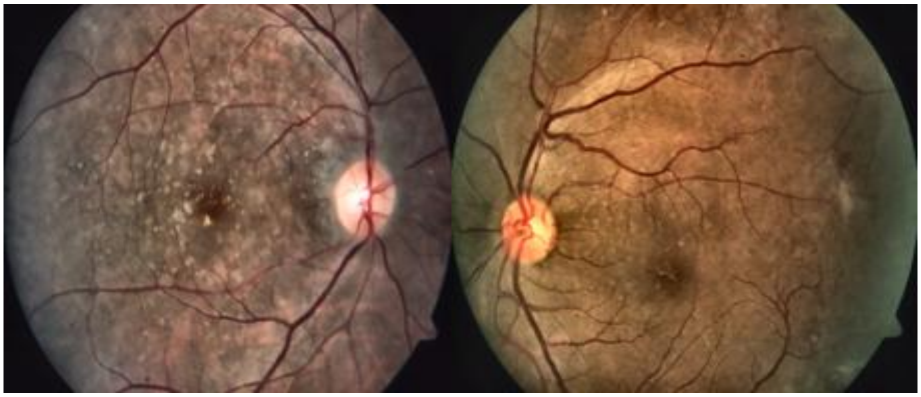


Figure (17) Misclassified images as Moderate while they are PDR due to neovascularization at the disk.

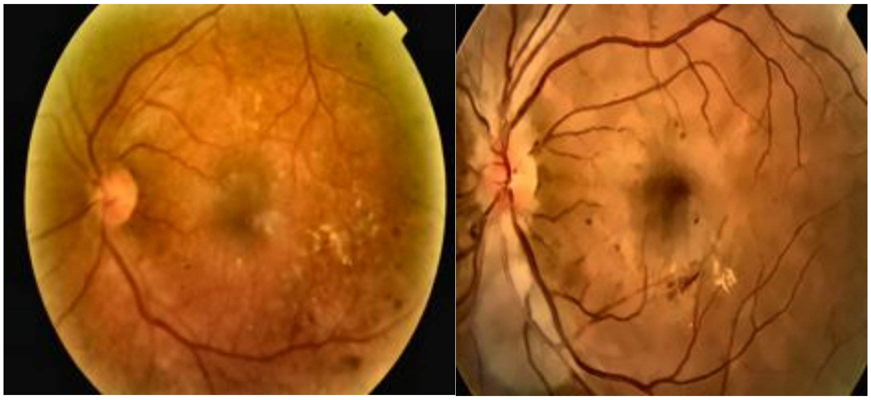
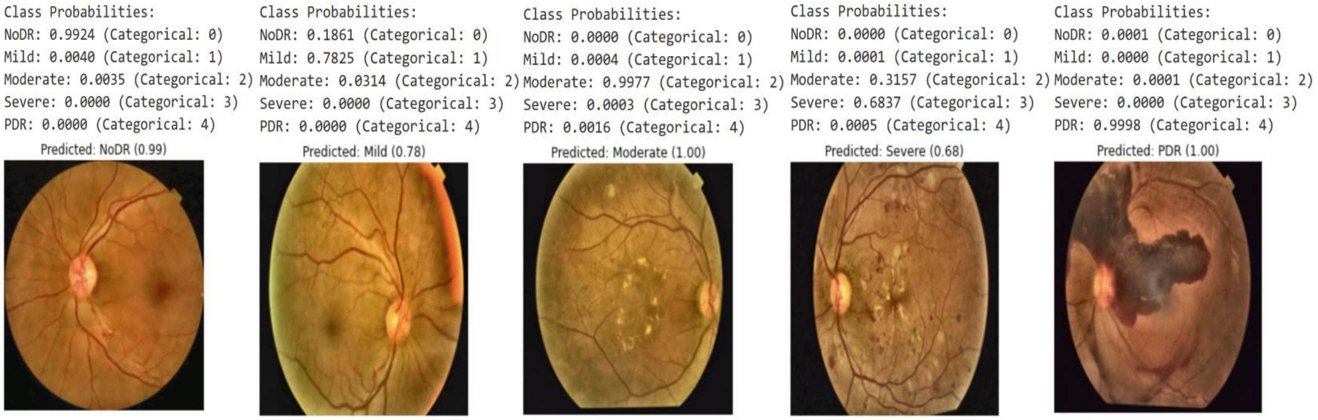


Figure (18) Images classified Moderate while human annotator said they are NoDR

The images presented in Figure 18 exhibit microaneurysms and exudates, which are key clinical indicators of the Moderate stage of diabetic retinopathy. This indicates that an ophthalmologist equipped with AI-powered assistive tools can minimize human error and efficiently diagnose DR severity stages.

The model developed using the DenseNet121 architecture effectively predicts the severity stages of diabetic retinopathy, as shown in Figure 19. The predicted five-stage classifications closely match the ground truth labels of the retinal images and are consistent with SHAP-based interpretations, as illustrated in Figure 20, which provides visual explanations for the same images.



```
Class Probabilities:
NoDR: 0.9924 (Categorical: 0)
Mild: 0.0040 (Categorical: 1)
Moderate: 0.0035 (Categorical: 2)
Severe: 0.0000 (Categorical: 3)
PDR: 0.0000 (Categorical: 4)
```

```
Class Probabilities:
NoDR: 0.1861 (Categorical: 0)
Mild: 0.7825 (Categorical: 1)
Moderate: 0.0314 (Categorical: 2)
Severe: 0.0000 (Categorical: 3)
PDR: 0.0000 (Categorical: 4)
```

```
Class Probabilities:
NoDR: 0.0000 (Categorical: 0)
Mild: 0.0004 (Categorical: 1)
Moderate: 0.9977 (Categorical: 2)
Severe: 0.0003 (Categorical: 3)
PDR: 0.0016 (Categorical: 4)
```

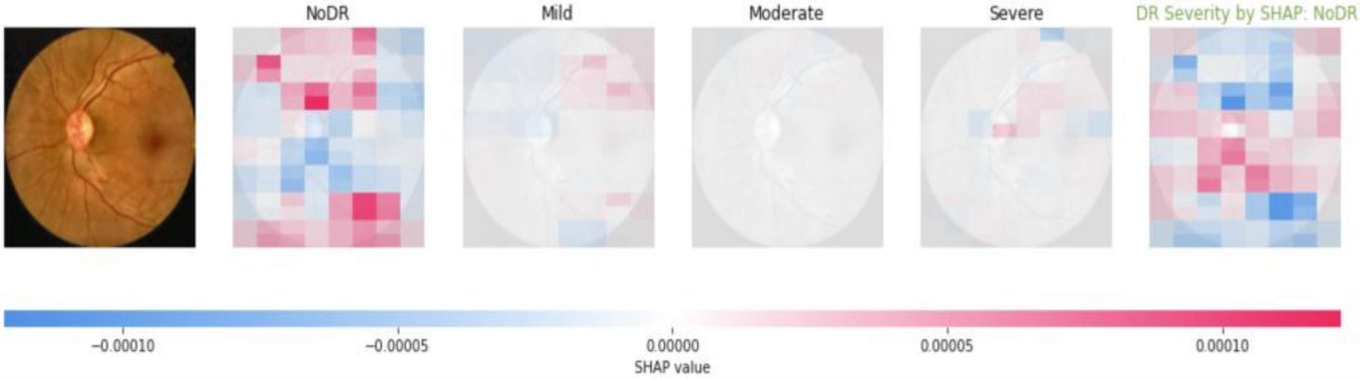
```
Class Probabilities:
NoDR: 0.0000 (Categorical: 0)
Mild: 0.0001 (Categorical: 1)
Moderate: 0.3157 (Categorical: 2)
Severe: 0.6837 (Categorical: 3)
PDR: 0.0005 (Categorical: 4)
```

```
Class Probabilities:
NoDR: 0.0001 (Categorical: 0)
Mild: 0.0000 (Categorical: 1)
Moderate: 0.0001 (Categorical: 2)
Severe: 0.0000 (Categorical: 3)
PDR: 0.9998 (Categorical: 4)
```

Figure (19) Diabetes retinopathy severity prediction (5 stages) and their probabilities

3.5 Model Explanations

To gain insight into how the proposed model performs diabetic retinopathy (DR) severity grading, we employed SHAP to compute Shapley values using an image masking technique with parameters mask_value="blur(128,128)" and shape=(224, 224, 3). Figure 20 illustrates the SHAP visualizations corresponding to each DR severity stage, ranging from NoDR to PDR. In these visualizations, red regions highlight areas that positively contribute to the model's prediction, while blue regions indicate features that oppose the predicted severity stage.



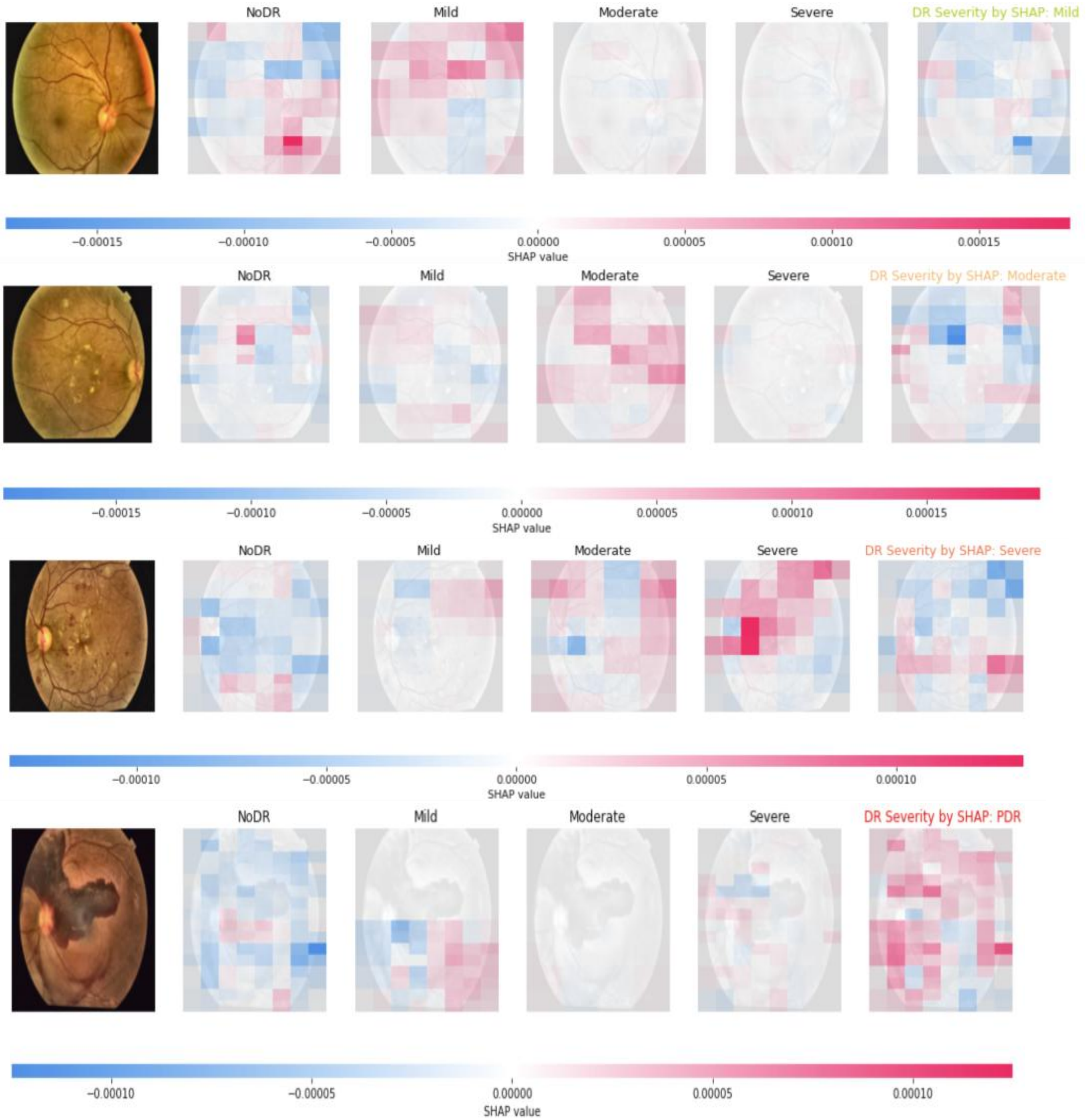


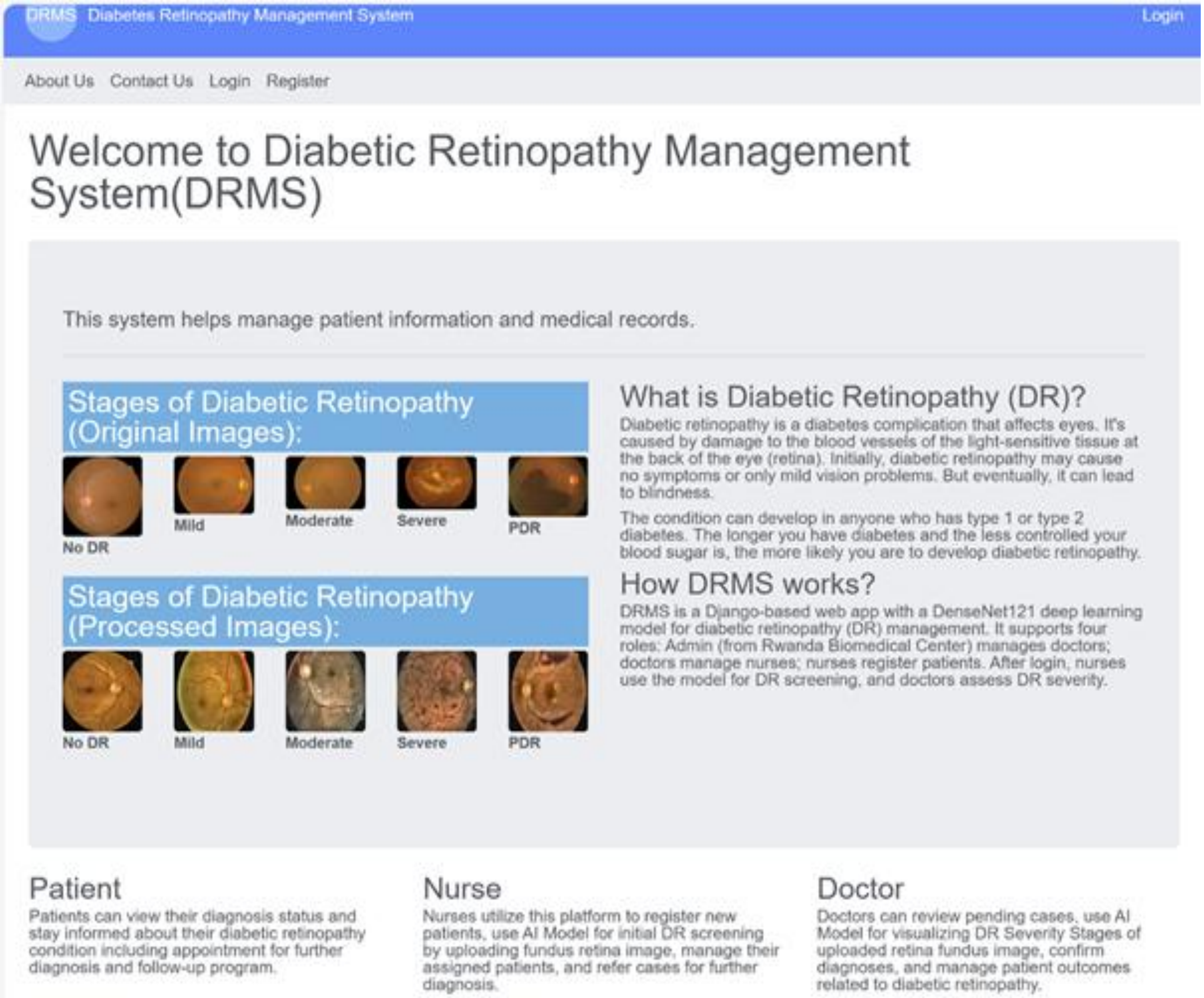
Figure (20) SHAP visual explanations of the proposed model on 5 DR severity stages.

This visualization using SHAP enhances transparency and aids DR specialists in understanding the model's decision-making process, thereby increasing clinical trust. SHAP was chosen over other explainability methods due to its strong theoretical foundation and ability to provide consistent, human-interpretable visual explanations. It also serves as an additional performance evaluation tool, offering insights into the model's internal reasoning.

3.6. Model Deployment

The top-performing models were integrated into a web application built using the Python Django framework and MySQL database server. This platform enables non-specialists to register diabetic patient data, including retinal images captured via OCT, fundus cameras, or handheld devices such as smartphones equipped with fundus lenses. A binary classification model is employed to detect cases of diabetic retinopathy that are referable. For confirmed cases, a multi-class classification model enables DR specialists to assess and visualize the disease's severity stages, facilitating informed clinical decisions. These outcomes are accessible to both nurses and patients within the same web application, which can be accessed through authentication using the username and password provided during the registration process. We are looking forward to implementing this application at CHUK.

Figure 21 shows the home page of the DR Management System (DRMS) web application, which integrates models for DR screening and severity grading.


DRMS
Diabetes Retinopathy Management System
Login
About Us
Contact Us
Login
Register
Stages of Diabetic Retinopathy (Original Images):
No DR
Mild
Moderate
Severe
PDR
Stages of Diabetic Retinopathy (Processed Images):
No DR
Mild
Moderate
Severe
PDR

Welcome to Diabetic Retinopathy Management System(DRMS)

This system helps manage patient information and medical records.

What is Diabetic Retinopathy (DR)?

Diabetic retinopathy is a diabetes complication that affects eyes. It's caused by damage to the blood vessels of the light-sensitive tissue at the back of the eye (retina). Initially, diabetic retinopathy may cause no symptoms or only mild vision problems. But eventually, it can lead to blindness.

The condition can develop in anyone who has type 1 or type 2 diabetes. The longer you have diabetes and the less controlled your blood sugar is, the more likely you are to develop diabetic retinopathy.

How DRMS works?

DRMS is a Django-based web app with a DenseNet121 deep learning model for diabetic retinopathy (DR) management. It supports four roles: Admin (from Rwanda Biomedical Center) manages doctors; doctors manage nurses; nurses register patients. After login, nurses use the model for DR screening, and doctors assess DR severity.

Patient

Patients can view their diagnosis status and stay informed about their diabetic retinopathy condition including appointment for further diagnosis and follow-up program.

Nurse

Nurses utilize this platform to register new patients, use AI Model for initial DR screening by uploading fundus retina image, manage their assigned patients, and refer cases for further diagnosis.

Doctor

Doctors can review pending cases, use AI Model for visualizing DR Severity Stages of uploaded retina fundus image, confirm diagnoses, and manage patient outcomes related to diabetic retinopathy.

Figure (21) Web application DR management using explainable deep Learning models

4. Discussion

4.1 Key Findings

For DR screening, a binary classification was applied using five models trained on the augmented APTOS dataset (7,100 images). DR cases (Mild to PDR) were combined into a single 'DR' class, while 'NoDR' remained unchanged. DenseNet121 achieved the highest performance with 97% accuracy and 99% AUC. All models exceeded 95% AUC, highlighting the effectiveness of the image preprocessing pipeline, which also improved image clarity, as noted by the DR expert during model validation. The same models were used for multiclass classification, trained on the augmented APTOS dataset (15,662 images) labeled across five DR severity stages: NoDR, Mild, Moderate, Severe NPDR, and PDR. As shown in Table 4, DenseNet121 achieved the best performance in DR severity grading, particularly in accurately identifying NoDR, Severe, and PDR stages, indicating effective feature extraction from clearly visible DR signs.

Performance was slightly lower for the Mild and Moderate stages, which are inherently more challenging to detect due to the asymptomatic nature of early DR, as illustrated in Figure 18. The two validation approaches demonstrated that the model's predictions were not random. The highest agreement between the model and human annotators was observed in the Brazil dataset, with an average Cohen’s Kappa of 0.771. However, a slightly lower Cohen’s Kappa of 0.680 was found in the validation dataset annotated by a DR specialist from CHUK. The model struggled to correctly identify the severity stage in retina images with neovascularization, a key indicator of PDR, often classifying these images as Moderate DR. This issue stemmed from the limited number of retina images with neovascularization in the training set.

SHAP visualizations, as shown in Figure 20, help clarify how the model makes predictions. To reduce the computational cost of Shapley values, we applied a (128,128) mask to the original (224,224,3) image. The results show that SHAP predictions align with the model’s prediction. Using red and blue, SHAP highlights image regions influencing the prediction. The web application, integrating the proposed models, was developed using Django and MySQL. It is intended for testing at CHUK pending internal ethical approval.

4.2. Comparison with Previous Research

Study	Image Enhancement	Prediction Method	Accuracy of Best Model	Best Model	Explainability
M. T. Herrero, et al. [38]	N/A	CNN + Transfer Learning (DenseNet121, InceptionV3,	94.6%	ResNet-50	SHAP

		MobileNet, Xception, VGG19, ResNet-50)			
M. Rosline and P. Kavitha [39]	N/A	Stacked Model (CNN-VGG16)	76.78%	CNN-VGG16	LIME and SHAP
A. S. Abdullah, et al. [40]	Images were already preprocessed	ResNet34, DenseNet121, GoogLeNet MobileViTv2, DeiT3	83.00%	MobileViTv2	SHAP
A..Hannan, et al. [41]	N/A	MobileNetV3-small, Efficientnet-b0, and DenseNet-169	83.20%	DenseNet-169	Grad-CAM
G. Deshpande, et al. [42]	N/A	InceptionV3	81.61%	InceptionV3	N/A
S. A. Karthik, et al. [43]	Gaussian Filter + CLAHE	proposed CNN framework	96.20%	proposed CNN framework	Grad-CAM
Z. Guanghua, et al. [44]	Gaussian Filter	VGG-16, ResNet-50	90.60%	ResNet-50	N/A
Robiul, et al. [45]	CLAHE	VGG16	91.32%	VGG16	N/A
Our Study	CLAHE + AMF	CNN + Transfer Learning (DenseNet121, InceptionV3, MobileNetV2, Xception, VGG19)	97.0%	DenseNet121	SHAP

4.3 Clinical usefulness of Proposed DR Management Model

The proposed model was trained and evaluated using Optical Coherence Tomography (OCT) images and subsequently validated with both OCT data and retinal images captured by handheld devices, including smartphones fitted with a fundus lens. This broad validation demonstrates the model's applicability for diabetic retinopathy detection in low-resource settings. The model exhibited strong predictive performance, offering a fast and reliable tool for identifying diabetic retinopathy in high-risk patients. It accurately grades the severity of diabetic retinopathy and serves as an assistive system for ophthalmologists by highlighting lesions in preprocessed images and providing interpretable explanations through SHAP methods. By integrating expert validation, high-quality imaging, and explainable deep learning, this study presents a promising approach for early detection of diabetic retinopathy and preservation of vision in diabetic populations.

4.4. Limitations

The proposed deep learning model achieved high accuracy in predicting the risk of diabetic retinopathy and effectively classified most disease stages. However, its ability to detect different variations of proliferative diabetics was limited, particularly in cases where retinal images exhibited neovascularization without accompanying hemorrhage. This limitation is likely due to the scarcity of representative training images for this specific presentation. Additionally, planned clinical testing of the model at a hospital in Rwanda was unable to be completed due to unexpected delays in obtaining ethical approval.

5. Conclusion

The proposed explainable deep learning models for DR management are not intended to replace diabetic retinopathy (DR) specialists, but to assist in DR screening and grading by providing reliable support for clinical decision-making. The models, trained on a large dataset of retinal images from both OCT (APTOS), demonstrated strong performance and generalizability when validated on an external image dataset collected using handheld devices (Brazil dataset). The enhanced image preprocessing techniques improved clarity and grading accuracy, contributing to the model's confidence and effectiveness. The proposed CNN-based framework enables end-to-end DR management, from screening to severity grading, and is integrated into a Django-MySQL web application designed for real-world deployment, pending ethical approval at CHUK. SHAP-based explainability enhances transparency by visually highlighting image regions that influence predictions. This study successfully met its objectives: accurately identifying referable DR cases, achieving high performance in severity grading, demonstrating the value of SHAP in enhancing model interpretability, and integrating the models in a web application. In practice, three outcomes were observed during DR grading: cases where the model was correct, where the human expert was correct, and where both agreed most often. These findings suggest that combining expert judgment with an explainable deep learning model can significantly improve DR screening and management.

6. Future work

We are planning to extend this study to the implementation stage. The implementation will be done in two phases. Phase 1 (July 2025) involves validating the model using existing retinal images of diabetic patients at CHUK, after obtaining internal ethical approval and training ophthalmologists on the use of the developed web application for DR screening and severity grading. Phase 2 (August 2025) focuses on monitoring the application in clinical use, evaluating its performance on images captured by OCT, fundus, and smartphone devices.

7. Declarations

7.1.Abbreviations

APTOS: Asia Pacific Tele-Ophthalmology Society

AMF: Adaptive Median Filter

BRSET: Brazilian Multilabel Ophthalmological Dataset of Retina Fundus Photos

CEBE: Center for Biomedical Engineering and E-health

CLAHE: Contrast Limited Adaptive Histogram Equalization

CHUK: Centre Hospitalier Universitaire de Kigali (in French)

CNN: Convolutional Neural Networks

DL: Deep Learning

DR: Diabetes Retinopathy

DRMS: Diabetes Retinopathy Management System

ICDRS: International Clinical Diabetic Retinopathy (ICDR) Scale

LIMCs: Low and Middle-Income Countries

OCT: Optical Coherence Tomography

PDR: Proliferative Diabetic Retinopathy

NLM: Non-Local Means

NPDR: Non-Proliferative Diabetic Retinopathy

SHAP: SHapley Additive exPlanations

ROC-AUC : Receiver Operating Characteristic Area Under the Curve

VGG: Visual Geometry Group

WHO: World Health Organization

XAI: Explainable Artificial Intelligence

7.2.Conflict of Interest Statement

The authors have no conflict of interests to declare.

7.3.Funding Disclosure

This research did not receive any specific grant from public, commercial, or not-for-profit funding agencies. However, it is part of the PhD training of Silas Majyambere, which is supported by the UR-Sweden Program.

7.4.Ethical Consideration

This study used publicly available retinopathy datasets (APTOS and BRSET), which contain no identifiable patient information; therefore, ethical approval and informed consent were not required.

7.5.Acknowledgement

The authors gratefully acknowledge the support of the University of Rwanda and Stockholm University. This research was funded by the UR-Sweden Research and Capacity Building Program through a scholarship awarded to Silas Majyambere. All authors contributed equally to writing the manuscript. Silas led research design, dataset collection, experimentation, and manuscript drafting. Tony and Celestin supervised the study, refined the manuscript, and analyzed the results. Gisagara, with over a decade of clinical experience in diabetic retinopathy management, validated the model by annotating the validation dataset and evaluating the clinical relevance of the findings. All authors reviewed and approved the final manuscript.

7.6.Data Availability

The image datasets used in this study are publicly available and intended for research purposes related to diabetic retinopathy.

8. References

[1]. International diabetes federation. IDF diabetes atlas, 10th edn (2021). Brussels, Belgium. Available at: https://www.diabetesatlas.org, 2021.

[2]. M. Kropp, et al. Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications-risks and mitigation. EPMA J.; 14(1):21-42. 2023. doi:10.1007/s13167-023-00314-8

[3]. U. V. Shukla and K. Tripath. Diabetic Retinopathy, PMID: 32809640, 2023

[4]. M. Porta and F. Bandello. Diabetic retinopathy: A clinical update, Diabetologia, pp. 1617–1634, 2002.

[5]. E. W. J. Dervan, P. D O'Brien, H. Hobbs, R. Acheson, and D.I Flitcroft. Targeted mydriasis strategies for diabetic retinopathy screening clinics, Nature Eye Vol. 24, pp. 207–1212, 2010.

[6]. T. Tan and T. Wong. Diabetic retinopathy: Looking forward to 2030, Frontiers in Endocrynology, 2022. DOI 10.3389/fendo.2022.1077669.

[7]. ETDRS Report. Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. Number 7. Ophthalmology. 1991;98(5 Suppl):741-756. doi:10.1016/s0161-6420(13)38009-9

[8]. Y. Zhengwei , T. Tien-En, S. Yan Shao, Y. W. Tien, and L. Xiaorong. Classification of diabetic retinopathy: Past, present and future, Frontiers in Endocrinology, 2022. https://doi.org/10.3389/fendo.2022.1079217.

[9]. H. T. María, R. O. Roberto, H. Roberto, C. G. T Gonzalo, I. L. María, and G. María. An explainable deep-learning model reveals clinical clues in diabetic retinopathy through SHAP, Biomedical Signal Processing and Control 102, 2025. https://doi.org/10.1016/j.bspc.2024.107328.

[10]. G. Andrzej, S. Panisa, and N. Onnisa. Artificial Intelligence for Diabetic Retinopathy Screening Using Color Retinal Photographs: From Development to Deployment, Ophthalmology and Therapy, pp. 1419–1437, 2023. https://doi.org/10.1007/s40123-023-00691-3.

[11]. H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma, and W. Qian. An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification, EMBC, 2019. doi: 10.1109/EMBC.2019.8857160.

[12]. R. H. A. Hamada, et al. A pilot study on diabetes detection using handheld fundus camera and mobile app development, Discovery in Applied Sciences, 2025. https://doi.org/10.1007/s42452-025-06460-0.

[13]. H. Naz, R. Nijhawan, and J. A. Neelu. Clinical utility of handheld fundus and smartphone-based camera for monitoring diabetic retinal diseases: a review study, International Ophthalmology vol. 44, 2024. https://doi.org/10.1007/s10792-024-02975-4.

[14]. T. Martina, V. Romano, D. Hendelja, K. Vilma, B. Tomislav, and R. Dario. Diagnostic Accuracy of Hand-Held Fundus Camera and Artificial Intelligence in Diabetic Retinopathy Screening, Biomedicines, 2023. https://doi.org/10.3390/biomedicines12010034

[15]. N. Mia, K. Feti, R.Nina, and S. H. Widihastha. Efficacy of Smartphone-based Fundus Photo in Vision Threatening Diabetic Retinopathy Screening: Developing Country Perspective, The Open Ophthalmology Journal, Vol. 18, 2024. http://dx.doi.org/10.2174/0118743641281527240116095349.

[16]. World Health Organization. Regional Office for South-East Asia, Strengthening diagnosis and treatment of Diabetic Retinopathy in SEA Region, 2020. https://iris.who.int/handle/10665/334224

[17]. M. B. Charlotte, M. Sanctus, P. C. Rutayisire, M. N. Loise, R. McQuillan, and H. W Sarah. Socio-demographic and clinical characteristics of diabetes mellitus in rural Rwanda: time to contextualize the interventions? A cross-sectional study, BMC Endocrine Disorders, 2020. https://doi.org/10.1186/s12902-020-00660-y

[18]. M. Hayati, et al. Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning, Procedia Computer Science 216, pp. 57–66, 2023.

[19]. U. K. Inam, et al. A Computer-Aided Diagnostic System to Identify Diabetic Retinopathy, Utilizing a Modified Compact Convolutional Transformer and Low-Resolution Images to Reduce Computation Time, Biomedicines Vol. 11, 2023. https://doi.org/10.3390/biomedicines11061566

[20]. S. B. Anuja and F. R. Dhanaseelan. Denoising of Diabetic Retinopathy Images Using Adaptive Median Filter, IJATEM pp. 122-131, 2023.

[21]. X. Luo, W. Wang, Y. Xu, Z. Lai, and J. Xiaopeng. A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence, CAAI Transactions on Intelligence Technology Vol. 9, pp. 153–166, 2023. https://doi.org/10.1049/cit2.12155

[22]. M. O. Odigie, G. O. George, E. C. Igodan, and K. C. Ukaoha. Detection of Diabetic Retinopathy Using VGG19 and ResNet 50 Models, EASJECS Vol. 7, 2024. https://doi.org/10.36349/easjecs.2024.v07i08.002

[23]. D. R. Sarvamangala and R. V. Kulkarni. Convolutional neural networks in medical image understanding: a survey, Evolutionary Intelligence Vol. 15, pp. 1-22, 2022. https://doi.org/10.1007/s12065-020-00540-3

[24]. K. Rajdeep, K. Rakesh, and M. Gupta. Review on Transfer Learning for Convolutional Neural Network, ICACCCN, 2021.

[25]. L. Tuggener, J. Schmidhuber, and T. Stadelmann. Is it enough to optimize CNN architectures on ImageNet?, Frontiers in Computer Science, 2022. https://doi.org/10.3389/fcomp.2022.1041703

[26]. J. E. Gutierrez, et al., Analysis of Pre-trained Convolutional Neural Network Models in Diabetic Retinopathy Detection Through Retinal Fundus Images, CISIM pp. 202-213, 2022.

[27]. G. Huang, L. Zhuang, P. Geoff, V. M. Laurens, and, K. Q. Weinberger. Convolutional Networks with Dense Connectivity, TPAMI, 2019. DOI 10.1109/TPAMI.2019.2918284

[28]. H. Chunjuan, S. Mohammad , and E. R. Adham. MobileNet-V2 /IFHO model for Accurate Detection of early-stage diabetic retinopathy, Heliyon Vol. 10, 2024. https://doi.org/10.1016/j.heliyon.2024.e37293

[29]. B. Rakesh, D. Ragavi, M. Kavya Reddy, and G. L. Sumalata. Detection and Classification of Non-Proliferation Diabetic Retinopathy using VGG-19 CNN Algorithm, ICAAIC, 2022. DOI: 10.1109/ICAAIC56838.2023.10141450

[30]. D. Gautam, G. Yash, and J. Anamika. Machine Learning-Based Diabetic Retinopathy Detection: A Comprehensive Study Using InceptionV3 Model, ICETSIS, 2024. DOI: 10.1109/ICETSIS61505.2024.10459541

[31]. F. Chollet. Xception: Deep learning with depthwise separable convolutions, CVPR, pp. 1251–1258, 2017. DOI 10.1109/CVPR.2017.195

[32]. V. Ashok, N. Hosmane, G. Mahagaonkar, A. Gudigar, and P. Anvith. Diabetic Retinopathy Detection using Retinal Images and Deep Learning Model, IJITEE Vol. 10 Issue-9, 2021.

[33]. K. M. Prasant, A. J. F. Sharmila, K. B. Rabindra, S. R. Diptendu, and J. S. Manob. Leveraging Shapley Additive Explanations for Feature Selection in Ensemble Models for Diabetes Prediction, Bioengineering Vol. 11, 2024.

[34]. V. Bellemo, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study, Lancet Digital Health Vol. 1, 2019.

[35]. N. Whitestone, et al. Feasibility and acceptance of artificial intelligence-based diabetic retinopathy screening in Rwanda, British Journal of Ophthalmology, 2022. https://doi.org/10.1136/bjo-2022-322683

[36]. M. Shorfuzzaman, M. S. Hossain, And A. E. Saddik. An Explainable Deep Learning Ensemble Model for Robust Diagnosis of Diabetic Retinopathy Grading, ACM Transactions on Multimedia Computing, Communications, and Applications, Vol. 17, 2021. https://doi.org/10.1145/3469841

[37]. G. Alwakid, W. Gouda, M. Humayun, and N. Z. Jhanjhi. Enhancing diabetic retinopathy classification using deep learning, Sage Digital Health Vol. 9, pp. 1-18, 2023. DOI: 10.1177/20552076231203676

[38]. M. T. Herrero, et al., An explainable deep-learning model reveals clinical clues in diabetic retinopathy through SHAP, Biomedical Signal Processing and Control, Vol. 102, 2025, 107328.

[39]. M. Rosline and P. Kavitha, Explainable AI for Diabetic Retinopathy Detection Based on a Hybrid-stacked Model, Journal of Intelligent & Fuzzy Systems, Vol. 49, Issue 3, 2025, pp. 703-719.

[40]. A. S. Abdullah, et al., Enhancing Diabetic Retinopathy Detection Through Transformer Based Knowledge Distillation and Explainable AI, IJCNN, 2024.

[41]. A..Hannan, Z. Mahmood, R. Qureshi, and H. Ali, Enhancing diabetic retinopathy classification accuracy through dual-attention mechanism in deep learning, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2025, Vol. 13, No. 1, 2539079

[42]. G. Deshpande, Y. Govardhan, and A. Jain, Machine Learning-Based Diabetic Retinopathy Detection: A Comprehensive Study Using InceptionV3 Model, ICETSIS, 2024.

[43]. S. A. Karthik, et al., Early Detection and Severity Classification of Diabetic Retinopathy Using Convolutional Neural Networks, SN Computer Science, 2025, https://doi.org/10.1007/s42979-025-04361-y.

[44]. Z. Guanghua, et al., Multi-Model Domain Adaptation for Diabetic Retinopathy Classification, Frontiers in Physiology, 2022, Vol. 13, 918929.

[45]. I. Robiul, H. Mehedi, and A. Sayeed, Transfer Learning based Diabetic Retinopathy Detection with a Novel Preprocessed Layer, IEEE TENSYMP, 2020.