

Robust Diabetic Retinopathy Classification via EfficientNet on APTOS Dataset

Saba Fatima¹, Vijayalaxmi S D²

¹Student, ²Assistant Professor

Department of Computer Science and Engineering
VTU's CPGS, Kalaburagi, Karnataka, India

Abstract

Untreated diabetic retinopathy (DR) may cause irreversible blindness if not caught in its early stages. The efficiency and accuracy of screenings may be greatly improved by automating the identification and categorisation of DR using retinal fundus pictures. This research uses EfficientNet, an architecture for convolutional neural networks that aims to optimise both accuracy and computational efficiency, to establish a DL-based DR detecting system. No DR, Mild, Moderate, Severe, and Proliferative DR are the five severity categories of retinal pictures that the model is taught to categorise. Healthcare providers and academics may now input retinal scans and get real-time forecasts using a user-friendly online application built on Streamlit. System outputs both the raw model probabilities and the rounded predicted class, facilitating transparent decision-making. Experimental evaluation demonstrates efficacy of given method in accurately classifying DR severity, portraying its probability for real-world clinical deployment and automated screening programs.

Index Terms

Diabetic Retinopathy, EfficientNet, Retinal Fundus Imaging, Medical Image Classification, APTOS 2019

I. INTRODUCTION

As a severe complication of diabetes, DR is among top causes of blindness worldwide. Early diagnosis via retinal imaging is crucial in preventing blindness. As Diabetes, a deadly disease, is characterised by body's failure to control blood sugar levels. Quantity of glucose inside the blood cells gets raised when these situations are considered. Diabetic complications include renal failure, cardiovascular disease, nerve damage, and blindness. In 2022, it was found that almost 9% of the world's population had diabetes. Forecasts indicate that the prevalence of diabetes will rise to around 12% by 2030. A comparison of the retinas of healthy and diabetic eyes is shown in Fig.1. [1].

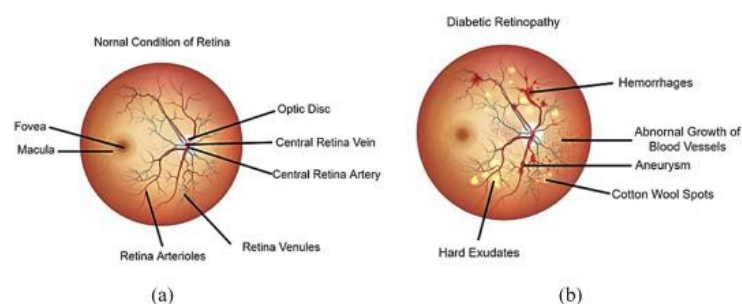


Fig. 1. Diagrams of (a) healthy and normal condition of retina and (b) diabetic retinopathy [2].

The presence of various lesions upon retinal scan may be utilised for diagnosing DR. The lesions in question include MA, HM, and exudates (EX). [3].

- The first symptom of DR is microaneurysms (MA), which are caused by the fragility of the walls of the blood vessels and show up on the retina as little red spherical dots. The size is under 125 μm and the edges are crisp, according to reference 3.
- Spots upon retina that are over 125 μm in size and with an uneven border are known as haemorrhages (HM). Two varieties of HM exist: flame, which is more surface, and blot, which is deeper [3].
- Hard exudates, which appear as bright yellow patches on the retina, are a consequence of plasma leakage. These pointed structures are part of retina's outer layers. [3].

- Retinal white spots, sometimes called cotton wool or soft exudates, are caused by expansion of nerve fibres. Oval or circular forms characterise it. [3].

Figure 2 shows phases of DR, and Table 1 shows the five severity categories recommended by International Clinical Diabetic Retinopathy Disease Severity Scale [4].

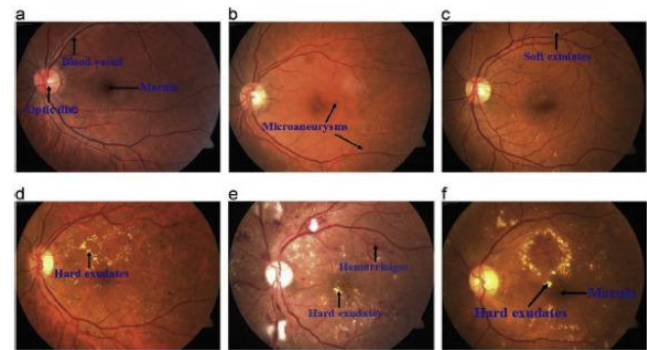


Fig. 2. DR stages: (a) Normal Retinal (b) Mild DR, (c) Moderate DR, (d) Severe DR, (e) Proliferative DR, (f) Macular Edema [5].

II. RELATED WORK

[1]Senapati et al. (2024)

Methods as per CNNs and transfer learning were shown to be the most effective in this investigation of artificial intelligence (AI) techniques for diabetic retinopathy identification. For improving diagnostic accuracy, research highlights need of preprocessing, augmentation, and ensemble approaches.

[2] Sebastian et al. (2023)

The authors surveyed deep-learning-based DR classification methods, comparing various CNNs, transfer learning, and hybrid architectures. They also evaluated commonly used datasets (EyePACS, APTOS, Messidor) and reported challenges in handling class imbalance and clinical validation.

[3] Alyoubi et al. (2020)

This review focused on DL methods for DR detection, summarizing CNN-based methods, preprocessing pipelines, and challenges in image quality and class imbalance. The paper concludes

DR Severity Level	Findings upon Dilated Ophthalmoscopy
0 = No DR	No abnormalities
1 = mild DR	Micro-aneurysms only
2 = Moderate DR	More than micro-aneurysms but less than NPDR
3 = Severe DR	Any of the following and no signs of proliferative retinopathy: 1. More than 20 intraretinal haemorrhages in each of four quadrants 2. Definite venous beading in two or more quadrants 3. Prominent IRMA in one or more quadrants
4 = Proliferative DR	One or both of the following: 1. Neovascularization 2. Vitreous/preretinal haemorrhage

Table 1. International clinical DR disease severity scale (ICDRDSS).

Note: (1) Very severe NPDR is defined as the presence of two or more symptoms indicative of severe NPDR in a patient. Two types of PDR exist: those with a high risk and those with a low risk.

that transfer learning and hybrid deep models yield promising performance.

[4] Wilkinson et al. (2003)

The authors proposed ICDR and Diabetic Macular Edema Disease Severity Scales, which remain the global standard for DR grading. These scales classify DR into severity levels that guide both clinical diagnosis and AI-based classification tasks.

[5] Mookiah et al. (2013)

This early review examined computer-aided diagnosis systems for DR before deep learning became dominant. It covered traditional machine learning methods, feature extraction from retinal images, and clinical integration challenges. The work provides a foundation for comparing classical and modern AI approaches.

[6] Huang et al. (2016)

Presented DenseNet, a CNN design that enhances feature reuse and gradient flow via dense layer connections. The use of DenseNet as a foundation for medical picture categorisation, particularly DR detection, has grown in recent years.

[7] Chaturvedi et al. (2020)

This study proposed an automated DR grading system using DenseNet-121 on retinal fundus images. Their experiments demonstrated strong classification accuracy, proving DenseNet's suitability for hierarchical DR severity grading. The work validates DenseNet as a competitive backbone for DR detection.

[8] Mika30 (2021)

The applied EfficientNet models to APTOS 2019 dataset for DR detection, highlighting the advantages of compound scaling, transfer learning, and extensive data augmentation. Research proved that EfficientNet designs are good for real-time deployment because they reduce computing complexity while achieving excellent accuracy.

[9] Abbas et al. (2023)

The proposed HDR-EfficientNet, an optimized EfficientNet variant for retinal image classification. By leveraging depthwise separable convolutions and adaptive scaling, model attained superior efficacy in predicting DR severity compared to traditional CNNs [9].

III. PROPOSED SYSTEM

An efficient and accurate multi-stage deep-learning pipeline is suggested for automated identifying and categorisation of diabetic retinopathy severity. EfficientNetB5 model, a cutting-edge CNN selected for its exceptional scalability and parameter efficiency, is the backbone of this system. A powerful feature extractor, this architecture was pretrained using massive ImageNet dataset to identify subtle and complex patterns in retinal fundus pictures that point to the development of illness.

The methodology begins with a meticulous preprocessing phase to standardize and normalize the input data. During the model training phase, we deviate from standard optimization techniques by employing the RAdam (Rectified Adam) optimizer. RAdam's unique combination of Adam's adaptive learning rates with a rectified learning rate ensures faster and more stable convergence, mitigating risk of getting stuck in suboptimal local minima. This is particularly

critical for medical image analysis tasks where achieving optimal performance on a limited dataset is paramount.

A key innovation and crucial component of our system lies in the post-processing stage, which employs a custom Optimized rounder utility. This component, rather than relying on arbitrary or standard thresholding (e.g., 0.5), automatically determines the optimal decision boundaries for each classification class. It achieves this by iteratively searching for rounding thresholds that maximize the validation set's QWK score. QWK metric is the official evaluation criterion for this task, and by directly optimizing for it, our system ensures the highest possible performance on the final metric. By applying these optimized thresholds to the model's output logits, we significantly enhance the final classification accuracy and the model's alignment with the official evaluation metric. The integration of these techniques—a powerful CNN backbone, an advanced optimizer, together with a post-processing phase that is particular to metrics—forms a thorough and very successful framework for automatic and accurate categorisation of retinal pictures showing diabetic retinopathy.

IV. METHODOLOGY

Utilising APTOS 2019 dataset, suggested technique aims to build a deep learning system that can automatically identifying and grade severity of DR. Methods include cleaning and preparing datasets, choosing an architecture for model, developing training plan, and finally, evaluating results.

A. Dataset

APTOS has made available a dataset called APTOS 2019 Blindness Detection. It includes 3,662 retinal fundus pictures that have been classified using ICDR grading system:

- 0: No DR
- 1: Mild DR
- 2: Moderate DR
- 3: Severe DR
- 4: Proliferative DR

The data set is severely skewed towards two extremes: "No DR" and "Moderate DR," where the vast majority of samples fall. To address this imbalance and achieve better model generalisation, suitable preprocessing and augmentation procedures must be used. Figure 3 shows that there are 1,805 photographs with no DR, 370 images with mild DR, 999 pictures having moderate DR, 193 pictures having severe DR, and 295 pictures having proliferative DR. Additionally, there are five severity categories for the images.

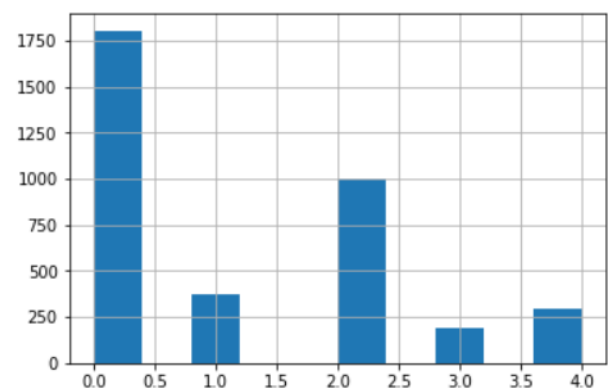


Fig. 3. Class distribution of the APTOS 2019 dataset across diabetic retinopathy severity levels.

B. Data Preprocessing

To guarantee consistent input and improving model's efficacy, our technique begins with a number of crucial data pretreatment processes. For purpose of improving computational performance and standardising input for neural network, all pictures are first shrunk to a consistent dimension, usually 512x512 pixels. After that, training dataset is artificially expanded using a thorough data augmentation method. During this procedure, small photometric modifications to brightness and contrast are made, in addition to random geometric alterations like flipping and rotations. To avoid overfitting and greatly enhance model's capacity to generalise to new data, this approach is vital.

C. Model Architecture

Proposed model is based on transfer learning approach using EfficientNetB5 as the backbone architecture. EfficientNetB5, pre-trained on ImageNet, was employed to leverage its strong feature extraction capabilities, despite the domain difference from medical imaging. On top of the base network, a set of custom layers were added to adapt it for regression-based severity prediction of diabetic retinopathy. The architecture can be summarized as follows:

1. **Base Model (EfficientNetB5):** The convolutional feature extractor outputs feature maps of shape $(15 \times 15 \times 2048)$.
2. **Layer for Global Average Pooling:** Condenses spatial data into a 2048-by-2048 feature vector.
3. To increase generalisation and decrease overfitting, a dropout layer is used with a rate of 0.5.
4. **Fully Connected Dense Layer (5 units, ELU activation):** Introduces nonlinearity

and allows learning of higher-level representations specific to the task.

5. **Final Dense Layer (1unit, Linear activation):** Outputs a continuous regression score corresponding to disease severity, which is later rounded to the nearest integer class (0–4).

The RAdam optimizer has been utilised to optimise the model with learning rate of 0.00005, and it was built utilising MSE loss. Retinal image dataset allowed for fine-tuning since EfficientNetB5's all layers were configured as trainable. To avoid overfitting, training process included early halting and reducing learning rate on plateau. Model's checkpoints were kept according to best validation QWK score. To further enhance classification performance, predictions were optimized using threshold tuning with the OptimizedRounder method, yielding a maximum validation QWK of 0.93712 and validation accuracy of 86.53%.

D. Training

Training the model were 3,662 retinal fundus pictures from the APTOS 2019 dataset. We utilised data augmentation methods such random angle rotations, horizontal and vertical flips, and pixel intensity normalisation by dividing each pixel value by 128 to make up for the little dataset and make the generalisations work better.

A regression-based formulation was used for the training, with the target diabetic retinopathy severity levels (0–4) being regarded as continuous data. To maximise QWK score whenever predictions are rounded to integer categories, the model was optimised utilising MSE loss function. This is because minimising MSE is aligned with maximising QWK. For reliable convergence, the RAdam optimizer was utilised, which has a learning rate of 0.00005.

We used early stopping and monitored validation loss for four epochs to avoid overfitting. Another

thing that was done was to utilise ReduceLROnPlateau to cut the learning rate in half when validation performance stopped improving. In order to preserve model checkpoints, the highest validation QWK score was used. To improve classification performance after training, decision thresholds were fine-tuned using the OptimizedRounder approach.

Outperforming traditional optimisation methods focused on accuracy, this training technique allowed the model to get a validation QWK score of 0.93712 and an accuracy of 86.53%.

E. Evaluation Metrics

Utilising many known criteria, we objectively assessed the efficacy of the suggested paradigm. What we call MSE is

$$MSE = (1/N) \sum (y_i - \hat{y}_i)^2$$

a value of 0.1095, which means that the projected severity scores ($\hat{P}_i$) and the associated ground truth values (y_i) have a low average squared deviation. This ordinal grading exercise used QWK to compare participants' levels of agreement.

$$\kappa = 1 - (\sum W_{ij} O_{ij}) / (\sum W_{ij} E_{ij}),$$

where $W_{ij} = (i-j)^2 / (k-1)^2$

This, with penalties given to bigger categorisation disparities, obtained a score of 0.93712, indicating a very high degree of consistency between predictions and real labels. With an accuracy of 86.53%, it was determined that most samples were accurately identified. With a macro-average of 0.87 for the Precision metric ($TP / (TP + FP)$), class-wise performance was good, and with a macro-average of 0.86 for the Recall metric ($TP /$

$(TP + FN)$), real instances were discovered with a high degree of success. Lastly, the classification job demonstrated a strong compromise between sensitivity and specificity as F1-score, which is defined as $F1 \text{ Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$, came out at 0.86% (macro average).

V. RESULTS

To show how the suggested diabetic retinopathy detection model works in practice, a web app was built utilising the Streamlit framework. The interface, presented on the main page (Fig. 4), enables users to upload retinal fundus images for real-time analysis. Upon uploading, the model computes raw prediction probabilities for each diabetic retinopathy severity class, which are subsequently rounded to two decimal places to determine the final predicted stage.



Fig. 4. Main page of the DR detection application.

Two illustrative instances were utilised for assessing system's performance. First, the model identified no diabetic retinopathy in the uploaded retinal picture (Fig. 5), classifying it as No DR. Second, at most advanced stage of the illness, the model indicated that uploaded picture was of Severe Proliferative Diabetic Retinopathy (Fig. 6). By displaying the final category prediction alongside the raw probabilistic outputs, the

application helps users understand the model's confidence and decision-making process. For the purpose of both clinical evaluation and research-oriented examination of the severity of diabetic retinopathy, this Streamlit-based application offers a clear and interactive platform for assessing retinal pictures.

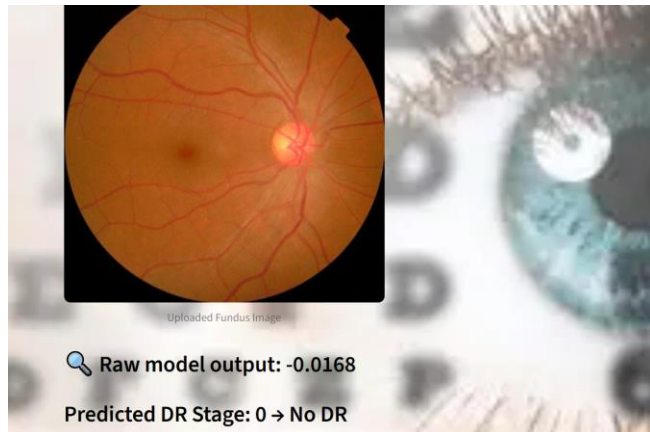


Fig. 5. Prediction output showing No DR for an uploaded retinal image.

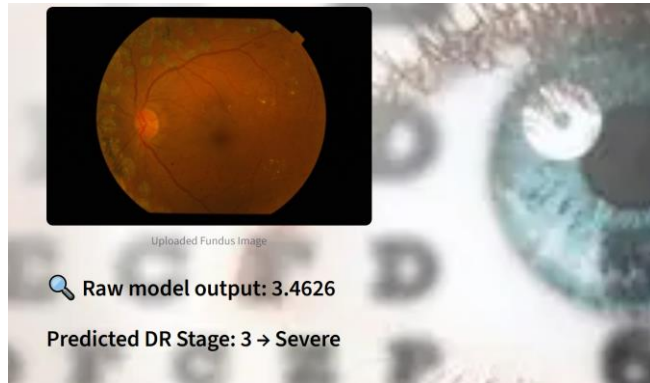


Fig. 6. Prediction output showing Severe Proliferative Diabetic Retinopathy for an uploaded retinal image.

VI. DISCUSSION

To accurately categorise retinal fundus pictures into five severity categories ranging from No DR to Severe Proliferative DR, the proposed diabetic retinopathy detection system uses the EfficientNet architecture. The model successfully balances computational economy with predicted accuracy

via the use of transfer learning, allowing it to attain excellent performance even with minimal training data.

By using the Streamlit-based interface, medical practitioners may easily input retinal scans and get instantaneous forecasts via real-time image processing. It is more transparent and easier for clinicians to understand when both the raw probability results and the rounded categorical forecasts are shown. The system can successfully differentiate between various phases of DR, according to evaluation on example photos (Figs. 2-3), which include both No DR and Severe Proliferative DR instances. Problems still exist, especially with low-quality photos or when there is a disparity in the classes, indicating that data augmentation and model ensembling could provide even more robust results.

VII. CONCLUSION

Automated diabetic retinopathy identification using EfficientNet models and a Streamlit-based web application is shown in this work. With the system's precise predictions, early diagnosis and action may be supported in real-time. The method makes it easier to utilise for research and therapeutic purposes by merging high-performance deep learning with an intuitive interface. The model may be improved for use in DR screening programmes by adding more retinal biomarkers, enhancing preprocessing methods, and increasing the size of available datasets.

REFERENCES

- [1] Senapati A, Tripathy HK, Sharma V, Gandomi AH. Artificial intelligence for diabetic retinopathy detection: A systematic review. *Informatics Med Unlocked*. 2024;45:101445.
- [2] Sebastian A, Elharrouss O, Al-Maadeed S, Almaadeed N. A survey on deep-learning-based diabetic retinopathy classification. *Diagnostics*. 2023;13(3):345.

- [3] Alyoubi WL, Shalash WM, Abulkhair MF. Diabetic retinopathy detection through deep learning techniques: A review. *Informatics Med Unlocked*. 2020;20:100377. doi:10.1016/j.imu.2020.100377.
- [4] Wilkinson CP, Ferris FL III, Klein RE, Lee PP, Agardh CD, Davis M, Dills D, Kampik A, Pararajasegaram R, Verdager JT. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110(9):1677-1682.
- [5] Mookiah MRK, Acharya UR, Kuang C, Min LC, Ng EYK, Laude A. Computer-aided diagnosis of diabetic retinopathy: a review. *Comput Biol Med*. 2013;43(12):2136-2155.
- [6] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *arXiv preprint*. 2016;arXiv:1608.06993.
- [7] Chaturvedi, S., et al. (2020). Automated Diabetic Retinopathy Grading using DenseNet-121. *arXiv preprint arXiv:2004.06334*.
- [8] Mika30, "EfficientNets for Diabetic Retinopathy Detection," Kaggle Notebook, 2021. [Online]. Available: <https://www.kaggle.com/code/mika30/efficientnets-for-diabetic-retinopathy-detection/notebook>
- [9] Q. Abbas, A. Rehman, and M. I. Khan, "HDR-EfficientNet: Classification of hypertensive and diabetic retinopathy using optimized EfficientNet architecture," *Diagnostics*, vol. 13, no. 20, p. 3236, 2023.
- [10] Gulshan, V., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410
- [11] Kaggle. (2019). APTOS 2019 Blindness Detection Dataset. Asia Pacific Tele-Ophthalmology Society (APTOS).
- [12] Ting, D. S. W., et al. (2019). Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, 103(2), 167–175.
- [13] Selvaraju, R. R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE ICCV*, 618–626.
- [14] Li, X., et al. (2021). Multi-hospital validation of deep learning-based diabetic retinopathy grading. *Frontiers in Medicine*, 8, 737.