An Enhanced Hybrid Ensemble Deep Learning Framework for Automated Retinal Lesion Segmentation in Diabetic Retinopathy

Sirisha Veluri[1], Vijaya Chandra Jadala[2], Shivani Goel[3], Anshu Kumar Dwivedi[4]

[1]School of Computer Science & Artificial Intelligence, SR University, Ananthsagar, Hasanparthy, Telangana 506371, India, Email: 2303c50056@sru.edu.in

[2]School of Computer Science & Artificial Intelligence, SR University, Ananthsagar, Hasanparthy, Telangana 506371, India, Email: vijayachandra.j@sru.edu.in

[3]School of Computer Science & Artificial Intelligence, SR University, Ananthsagar, Hasanparthy, Telangana 506371, India, Email: shivani.goel@sru.edu.in

[4]School of Computer Science Engineering and Technology, Bennett University, Greater Noida, Uttar Pradesh, 201310, India, Email: anshucse.dwivedi@gmail.com

HOW TO CITE:

Sirisha Veluri, Vijaya Chandra Jadala, Shivani Goel, Anshu Kumar Dwivedi (2026). An Enhanced Hybrid Ensemble Deep Learning Framework for Automated Retinal Lesion Segmentation in Diabetic Retinopathy. International Journal of Special Education, 41(20s), 285-292

ABSTRACT:

Diabetic retinopathy (DR) is one of the leading causes of visual disability and early diagnosis and monitoring of the disease is critical for accurate segmentation of retinal lesions. Automated segmentation, however, is difficult because of the variations of lesions characteristics and image quality of the retina. In this study, an Enhanced Hybrid Ensemble Deep Learning Model (HEDLM) is proposed in order to automatically segment the retinal lesions using the hybrid model of U-Net, U-Net+, and U-Net++. A dataset of DR2 retinal fundus images with 757 images was tested using the framework. The image preprocessing steps involved resizing, normalization, and augmentation of the data. The three segmentation models were trained separately with a composite Dice and Binary Cross-Entropy loss function and their results fused by averaging over probabilities, weighted fusion and voting on majority method. The proposed ensemble outperformed the other models such as U-Net, U-Net+ and U-Net++ and recorded Dice coefficient 0.90, Intersection over Union 0.84, and a Segmentation accuracy of 96%. The results show the effectiveness of complementary segmentation architectures for enhancing lesion localization and robustness of segmentation. The framework proposed is a computer-aided method for analysis of retinal lesions that could help facilitate screening for diabetic retinopathy (DR) and aid in clinical decision making.

Keywords: Diabetic retinopathy, Retinal lesion segmentation, Deep learning, U-Net, Ensemble learning.

1. Introduction

DR is an ongoing eye problem of diabetes that can cause vision loss or blindness in individuals with diabetes if it is not caught and treated early. As diabetes is becoming a common problem, efficient retinal screening and computer-assisted diagnostic approaches are needed. In recent years, computer-aided diagnosis of DR based on the analysis of retinal fundus images has shown great promise thanks to rapid developments in artificial intelligence (AI) and deep learning (DL) technologies, with applications for DR detection, screening, grading, and monitoring [5], [7], [15].

Traditional DR diagnosis mainly depends on ophthalmologists' visual inspection of retinal fundus images. While clinical evaluation is important, manual evaluation can be labour intensive and difficult in large-scale screening programs. Thus, deep learning methods, especially convolutional neural networks (CNNs), have been increasingly used to automate the detection and classification of DR [9], [10],[11]. Most of the current methods, however, consider only disease classification at the image level and not at the level of the individual pixels. Accurate segmentation of the lesion is crucial, since it allows for the localization of pathological regions, and allows for more detailed information for disease assessment and monitoring [1],[2],[13]. In the medical image segmentation field, encoder-decoder architectures have performed well. In particular, U-Net was designed to fuse contextual information with localization in a single image due to its encoder-decoder structure and skip connections [14]. Following developments and variants have further enhanced feature propagation, multi-scale feature integration and boundary delineation [12],[20]. In recent years, several studies have employed optimized models based on U-Net to analyze retinal images, used multi-lesion detection networks, and other DL methods [3],[16],[17]. However, one individual segmentation architecture may not perform as well on all types of lesions, shapes, contrast and image quality. Ensemble and hybrid learning methods offer a chance to integrate different feature representations and overcome drawbacks in single models. The feasibility of feature fusion and ensemble techniques for the detection of diabetic retinopathy and retinal image analysis has been proven in the past [9],[16],[18]. In contrast, though, there is comparatively little work that focuses on a combination of several U-net-based architectures in the context of automated retinal lesion segmentation in a hybrid ensemble framework. Based on this, this study suggests an Enhanced Hybrid Ensemble Deep Learning Model (HEDLM) by combining U-Net, U-Net+ and U-Net++ for automatic segmentation of retinal lesions. The framework is an ensemble fusion of the three segmentation models trained independently, such as probability averaging, weighted fusion and majority voting. The proposed framework is tested with the DR2 retinal fundus image dataset [22] and it is compared with Dice coefficient, Intersection over Union (IoU), Precision, Recall, F1-score and Segmentation accuracy. The aim of this work is to investigate if complementary segmentation architectures can outperform single models with regards to retinal lesion localization, segmentation accuracy and robustness.

2. Methodology

2.1 Research Design

The proposed automated retinal lesion segmentation method in diabetic retinopathy used in this research was an Enhanced Hybrid Ensemble Deep Learning Model (HEDLM) with a quantitative experimental research design to determine the effectiveness of the method. The ensemble architecture proposed was compared with three single segmentation architectures (U-Net, U-Net+, and U-Net++). To ensure fairness and unbiased comparisons, the same pre-processing, partitioning of the dataset, training and evaluation were used across all models.

The aim of the experimental design was to investigate whether the performance of the retinal lesion localization and segmentation task could be enhanced from the individual models by using complementary segmentation architectures that are integrated via ensemble learning. Quantitative segmentation metrics, such as Dice coefficient, Intersection over Union (IoU) score, Precision, Recall, F1 score, and Pixel accuracy scores were used to evaluate the models.

2.2 Dataset and Sample

The dataset of retinal fundus images (RFIs) was acquired from publicly available DR2 dataset consisting of 757 colour retinal fundus images from a ZEISS Visucam 500 retinal imaging system [22]. The images have been clinically evaluated and classified into various stages of DR and a wide range of retinal images have been collected to test automated retinal image analysis.

The entire number of images present in the dataset were used for the experimental study. The retinal fundus images were captured and then resized and pre-processed to ensure consistency among all the experimental models prior to model training. The data set was randomly split into 70% training, 15% validation, and 15% test sets. The training subset was utilized to learn the parameters of the segmentation networks, and the validation subset was utilized to track model execution and help in the selection of the model. The testing set was set aside for the final testing of segmentation performance.

The same data split was used for all models, U-Net, U-Net+, U-Net++ and the HEDLM framework, consistent for the models, to prevent variations in the experimental data from affecting the results of the performance difference.

2.3 Tools and Experimental Setup

Resembling the original paper, the proposed HEDLM framework was realized in the deep learning environment TensorFlow–Keras. Three encoder-decoder based segmentation architectures (U-Net, U-Net+, U-Net++) were used in the experimental setup. The baseline segmentation architecture was U-Net, which has encoder-decoder architecture with skip connections, allowing the fusion of contextual and spatial information in segmenting biomedical images [14]. The advanced architecture was U-Net++ due to its re-designed nested and dense skip connections, which enable to pass through more features across the multi-scale feature maps in the encoder and decoder [20]. These models were trained with U-Net and a modified version of U-Net (U-Net+). The size of all retinal fundus images was resized to 256 × 256 pixels for uniformity of input to segmentation models and less computational complexity. To ensure numerical stability in the training of the model, the intensity values of the image pixels were normalized to [0,1]. The training images were augmented to ensure data generalization and avoid overfitting. All the augmentation techniques used were random rotation, horizontal flipping, vertical flipping, scaling, and zooming. To maintain spatial associations between input images and ground-truth lesion annotations, the segmentation masks were also geometrically transformed.

2.4 Experimental Procedure

The experimental methodology involved the following main stages: image preprocessing, training individual models, constructing the ensemble prediction, and model validation. The retinal fundus images and the ground-truth segmentation masks were pre-processed during the first stage. The images were resized to 256 × 256 pixels, normalized, and augmented to get this data more diverse. All three segmentation models were pre-processed with the same method. The second stage involved the independent training of U-Net, U-Net+ and U-Net++ with the same experimental conditions. The Adam optimizer [23] was used for optimizing all models starting with a learning rate of 1×10^{-4}. A composite loss function consisting of Dice Loss and Binary Cross-Entropy (BCE) Loss was used to define the optimization goal. A Dice-based loss component also helps to optimize segmentation by penalizing the difference between the predicted and ground-truth regions, especially in the case of an imbalance for the foreground and background regions [21].

A maximum of 100 epochs were used in training each model with a batch size of 16. To prevent overfitting, the Early Stopping technique was used to stop training if the results of the validation did not improve for a certain number of epochs. Model Checkpointing was implemented to save the model weights of the model achieving the maximum dice score for the validation set. Also, the ReduceLROnPlateau learning-rate scheduler was used to reduce the learning rate when the validation performance does not improve and to ensure better convergence in training. In the third stage, for each test image, each trained Segmentation Model produced a Probability Map, a map indicating the probability of each pixel being a retinal lesion. Then the predictions made by U-Net, U-Net+ and U-Net++ were fused with the proposed hybrid ensemble learning framework. Three ensemble fusion strategies—probability averaging, weighted fusion, and majority voting—were used to combine the individual model predictions. That led to an ensemble output which was thresholder to produce the final binary retinal lesion segmentation mask. The last step involved comparing the segmentation masks predicted by each of the models and the proposed HEDLM framework with their respective ground-truth masks. This allowed the architectures to be compared directly, and the proposed hybrid ensemble approach compared under the same experimental conditions.

2.5 Data Analysis

The different segmentation models and the proposed HEDLM framework were quantitatively assessed using Dice coefficient, Intersection over Union (IoU), precision, recall, F1-score and pixel accuracy. These metrics were chosen to evaluate overall segmentation overlap, lesion detection, lesion classification and overall pixel level accuracy.

To assess the overlap of the predicted lesion areas in relation to the ground truth lesion masks, the Dice coefficient and IoU were used. The proportion of correctly identified lesion pixels to all pixels identified as lesion (precision) and all pixels correctly detected as lesion (recall) were calculated. The F1 score gave a balanced assessment, taking into account precision and recall. Pixel accuracy was the overall percentage of correctly classified pixels. A comparative study is made between U-Net, U-Net+, U-Net++ and the proposed HEDLM framework. The analysis was also comparative because all models were trained and tested with the same partition of the dataset, preprocessing strategy, training configuration and evaluation metrics, thus allowing for a like-for-like evaluation of the contribution of the proposed hybrid ensemble framework to the performance of retinal lesion segmentation.

3. Results

The comparative analysis of the experimental results of the individual segmentation models and the proposed Enhanced Hybrid Ensemble Deep Learning Model (HEDLM) is shown in this section. Dice coefficient, Intersection over Union (IoU), precision, recall, F1-score and segmentation accuracy were used to evaluate the models.

3.1.1 Performance of Individual Segmentation Models

Following Table 1 shows the results of the 3 separate U-Net based segmentation architectures. The results of individual models (U-Net++, U-Net+ and U-Net) show that the best model is U-Net++, then U-Net+ and U-Net.

Table 1. Performance of Individual Segmentation Models

Model	Dice Coefficient	IoU	Segmentation Accuracy
U-Net	0.82	0.75	91%
U-Net+	0.85	0.78	93%
U-Net++	0.88	0.81	95%

The Dice coefficient for U-Net was 0.82 compared to 0.85 for U-Net+ and 0.88 for U-Net++. Similarly, IoU increased from 0.75 to 0.78 and 0.81, respectively.

3.2 Performance of the Proposed Hybrid Ensemble Model

The proposed HEDLM fused the predictions of U-Net, U-Net+ and U-Net++ with ensemble fusion techniques. Among all the tested models, the proposed one gained the top performance.

Table 2. Comparative Performance of Individual and Proposed Ensemble Models

Model	Dice Coefficient	IoU	Segmentation Accuracy
U-Net	0.82	0.75	91%
U-Net+	0.85	0.78	93%
U-Net++	0.88	0.81	95%
Proposed HEDLM	**0.90**	**0.84**	**96%**

The proposed HEDLM achieved Dice coefficient of 0.90, IoU of 0.84 and segmentation accuracy of 96%. The proposed model achieved improvements of 0.08 Dice coefficient, 0.09 IoU and 5 percentage points in segmentation accuracy compared with U-Net. HEDLM achieved improvements of 0.02 in Dice coefficient, 0.03 in IoU, and 1 percentage points in segmentation accuracy when compared to the best performing single model U-Net++.

3.3 Comparative Analysis

The comparative performance of each architecture of segmentation and proposed HEDLM is shown in figure 1.

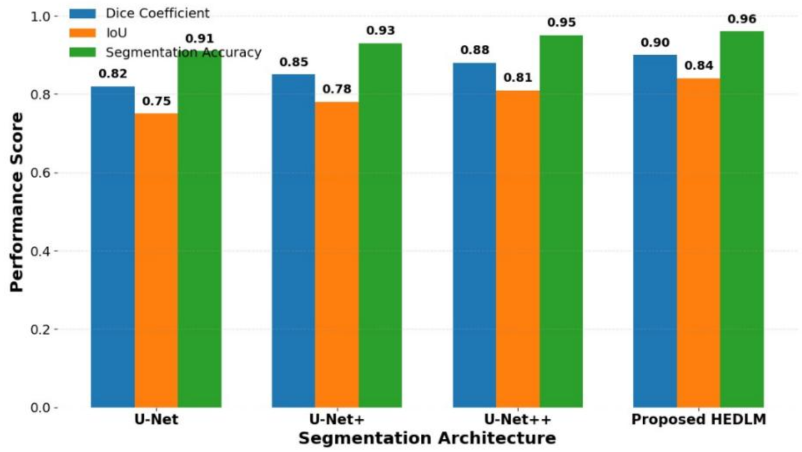

Figure 1. Performance comparison between U-Net, U-Net+, U-Net++ and the proposed HEDLM with Dice coefficient, IoU and segmentation accuracy.

The proposed HEDLM has achieved the highest scores for the three reported metrics, while the performance is progressive as U-Net is compared with U-Net++.

4. Discussion

The results of this study showed that the proposed Enhanced Hybrid Ensemble Deep Learning Model (HEDLM) outperformed the individual models of U-Net, U-Net+ and U-Net++. The proposed framework obtained a Dice coefficient of 0.90, an IoU of 0.84, and a segmentation accuracy of 96%. The results serve as an answer to the primary objective of the study, that is, to investigate if complementary segmentation architectures can help to better localize lesions and improve overall segmentation performance.

The individual models achieved the highest performance in the following order: U-Net++, U-Net+ and U-Net. The results suggest that the predictions from the three architectures contain complementary information, with the better performance of the proposed ensemble showing this. The skip connections of U-Net were effective to retain the spatial information, but super-improved U-Net based architectures such as U-Net++ have been proposed to better propagate features and multi-scale feature fusion [20]. These different representations of features are then fused together in the ensemble approach, which leads to better segmentation performance.

The results are in line with previous studies [2],[9],[16],[17],[18] which have reported the advantages of deep learning, hybrid learning, and ensemble methods for diabetic retinopathy analysis. Previous studies have shown that deep learning can be used to detect and segment retinal lesions and can be used to classify patients into a specific disease, but many of these studies are limited to specific architectures or specific disease-level classifications. The current result indicates that the use of multiple segmentation approaches can achieve better results for retinal lesion analysis at the pixel level. The proposed framework can be practically applied in computer-aided screening of diabetic retinopathy. The advancement of the segmentation of lesions can facilitate the process of detecting and locating pathological areas in fundus images of the retina, thereby providing more information that can help ophthalmologists in the evaluation and monitoring of the eye disease. This kind of AI-driven methods could also be applied to a large-scale screening setting where the analysis of the retinal image is required to be efficient [4],[5],[6],[7],[8]. Although the findings are these, the study is subject to some limitations. The proposed framework was tested with one retinal image dataset with 757 images. Thus, it remains to be studied how well it can be reproduced in images acquired from other populations, imaging devices, and clinical settings. Furthermore, ensemble models could require more computational resources than each segmentation model separately. The framework should be examined, using larger multi-centre data sets, and with external validation [12], [13],[19] in the future. Potential enhancements can also include the use of attention-based architectures and explainable AI techniques to further enhance model interpretability and clinical applicability.

5. Conclusion

The current work put forward an Enhanced Hybrid Ensemble Deep Learning Model (HEDLM) for automatic segmentation of retinal lesions in Diabetic Retinopathy. The framework united U-Net, U-Net+ and U-Net++ to take advantage of the complementary power of multiple encoder-decoder architectures. The results of the experimental evaluation on the DR2 retinal fundus image dataset showed that the proposed ensemble outperforms the individual models with Dice coefficient of 0.90, IoU of 0.84, and segmentation accuracy of 96%. The results show that using multiple segmentation architectures can further enhance the accuracy of lesion localization and segmentation performance than using a single architecture. The proposed framework thus has the potential as a computer-aided approach to assist in retinal image analysis and screening for DR.

The framework should be validated in future research with more diverse and larger multi-center datasets and images taken by multiple retinal imaging devices. Additional research could focus on attention-based segmentation models, explainable AI methods, computational optimizations, and clinical validation to enhance the interpretability, efficiency, and usability of the proposed system.

References

Lam, C., Wong, Y. L., Tang, Z., Hu, X., Nguyen, T. X., Yang, D., Zhang, S., Ding, J., Szeto, S. K. H., Ran, A. R., & Cheung, C. Y. (2024). Performance of artificial intelligence in detecting diabetic macular edema from fundus photography and optical coherence tomography images: A systematic review and meta-analysis. *Diabetes Care, 47*(2), 304–319. https://doi.org/10.2337/dc285-292

Abdelmaksoud, E., El-Sappagh, S., Barakat, S., Abuhmed, T., & Elmogy, M. (2021). Automatic diabetic retinopathy grading system based on detecting multiple retinal lesions. *IEEE Access, 9*, 15939–15960. https://doi.org/10.1109/ACCESS.2021.3052870

Asif, M., Ur Rehman, F., Rashid, Z., Hussain, A., Mirza, A., & Qureshi, W. S. (2025). An insight on the timely diagnosis of diabetic retinopathy using traditional and AI-driven approaches. *IEEE Access, 13*, 116869–116886. https://doi.org/10.1109/ACCESS.2025.3583647

Natarajan, S., Jain, A., Krishnan, R., Rogye, A., & Sivaprasad, S. (2019). Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone. *JAMA Ophthalmology, 137*(10), 1182–1188. https://doi.org/10.1001/jamaophthalmol.2019.2923

Gulshan, V., Rajan, R. P., Widner, K., Wu, D., Wubbels, P., Rhodes, T., Whitehouse, K., Coram, M., Corrado, G., Ramasamy, K., Raman, R., Peng, L., & Webster, D. R. (2019). Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmology, 137*(9), 987–993. https://doi.org/10.1001/jamaophthalmol.2019.2004

Rogers, T. W., Gonzalez-Bueno, J., Garcia Franco, R., Lopez Star, E., Méndez Marín, D., Vassallo, J., Lansingh, V. C., Trikha, S., & Jaccard, N. (2021). Evaluation of an AI system for the detection of diabetic retinopathy from images captured with a handheld portable fundus camera: The MAILOR AI study. *Eye, 35*(2), 632–638. https://doi.org/10.1038/s285-292-285-292

Beals, D., Simon, L., Rogers, F., & Pogroszewski, S. (2025). Revolutionizing diabetic retinopathy screening: Integrating AI-based retinal imaging in primary care. *Journal of CME, 14*(1). https://doi.org/10.1080/28338073.2024.2437294

Bhuiyan, A., Govindaiah, A., Deobhakta, A., Hossain, M., Rosen, R., & Smith, T. (2021). Automated diabetic retinopathy screening for primary care settings using deep learning. *Intelligence-Based Medicine, 5*, Article 100045. https://doi.org/10.1016/j.ibmed.2021.100045

Butt, M. M., Iskandar, D. N. F. A., Abdelhamid, S. E., Latif, G., & Alghazo, R. (2022). Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features. *Diagnostics, 12*(7), Article 1607. https://doi.org/10.3390/diagnostics12071607

Wan, S., Liang, Y., & Zhang, Y. (2018). Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Computers & Electrical Engineering, 72*, 274–282. https://doi.org/10.1016/j.compeleceng.2018.07.042

Nazir, T., Nawaz, M., Rashid, J., Mahum, R., Masood, M., Mehmood, A., Ali, F., Kim, J., Kwon, H.-Y., & Hussain, A. (2021). Detection of diabetic eye disease from retinal images using a deep learning-based CenterNet model. *Sensors, 21*(16), Article 5283. https://doi.org/10.3390/s21165283

Hossen, M. F., Sarkar, A., Majumder, T., Nabi, M. F., Ahamed, M. F., Shafi, F. B., ... & Chowdhury, M. E. (2026). A comprehensive review of U-Net architectures for medical image segmentation: Emerging trends and federated learning perspectives. *Neural Computing and Applications, 38*(12), Article 493. https://doi.org/10.1007/s285-292-285-292

CB, V., & Periasamy, P. (2026). Systematic review on diabetic retinopathy segmentation and detection using deep learning techniques: Research challenges and future directions. *Cogent Engineering, 13*(1), Article 2653310. https://doi.org/10.1080/23311916.2026.2653310

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/285-292-285-292285-292_28

Cao, F., Guo, X., Li, M., Li, S., & Peng, X. (2025). Development and validation of a deep learning model for early detection and screening of diabetic retinopathy. *BMC Medical Informatics and Decision Making, 25*(1), Article 315. https://doi.org/10.1186/s285-292-285-292

Nneji, G. U., Cai, J., Deng, J., Monday, H. N., Hossin, M. A., & Nahar, S. (2022). Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans. *Diagnostics, 12*(2), Article 540. https://doi.org/10.3390/diagnostics12020540

Gupta, A., & Rawat, A. S. (2026). Comparative analysis of deep learning models for lesion-wise segmentation of diabetic retinopathy. *IEEE MultiMedia*. Advance online publication. https://doi.org/10.1109/MMUL.2026.3673514

Singh, R. K., & Gorantla, R. (2020). DMENet: Diabetic macular edema diagnosis using hierarchical ensemble of CNNs. *PLOS ONE, 15*(2), Article e0220677. https://doi.org/10.1371/journal.pone.0220677

Chankhachon, S., Kansomkeat, S., Bhurayanontachai, P., & Intajag, S. (2026). Refined IDRiD: An enhanced dataset for diabetic retinopathy segmentation with expert-validated annotations and comprehensive anatomical context. *Data, 11*(2), Article 30. https://doi.org/10.3390/data11020030

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2020). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging, 39*(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and*

Multimodal Learning for Clinical Decision Support (pp. 240–248). Springer. https://doi.org/10.1007/285-292-285-292285-292_28

Castillo Benítez, V. E., Castro Matto, I., Mello Román, J. C., Vázquez Noguera, J. L., García-Torres, M., Ayala, J., Pinto-Roa, D. P., Gardel-Sotomayor, P. E., Facon, J., & Grillo, S. A. (2021). Dataset from fundus images for the study of diabetic retinopathy. *Data in Brief, 36*, Article 107068. https://doi.org/10.1016/j.dib.2021.107068

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

Acknowledgements

The authors would like to thank their institutions for support and also thank DR2 retinal fundus image set provided for this study. This research did not receive any external funding.

Author Declaration

The authors certify that this manuscript has not been published before and is not under consideration by another publication. Appropriate source citations and acknowledgements have been made for all sources used in this study. The authors also state that no competing interests or conflicts of interest existed that could have affected the research, analysis, interpretation of the results or preparation of this manuscript.