EyeNet: Automated Retinal Image Evaluation for Diabetic Retinopathy Detection

Mr. SANTHOSH S.G[1], RAZIYA SULTANA[2]

[1]*Associate Professor, Dept of MCA, Department of Master of Computer Applications, Jawaharlal Nehru new College of Engineering, Shimoga, Karnataka, India.*

[2]*PG Student, Dept of MCA, Jawaharlal Nehru new College of Engineering, Shimoga, Karnataka, India.*

***Abstract*—Diabetic retinopathy is among the diabetes mellitus, which is a serious complication among the diabetes mellitus, leading to the increasing the loss of vision if it is not detected and treated early. If the retinal images are graded manually according to the extremity of disease DR, then it needs the experts such as ophthalmologists and more time, which limits screening accessibility, especially in the areas of limited resources. This paper represents EyeNet, an automated system which is based on deep learning is designed to split the images of retinal fundus into five different stages of Diabetic Retinopathy that is: no Dr, mild Dr, moderate Dr, severe Dr, and proliferative retinopathy. EyeNet utilizes Convolutional Neural Networks that is CNNs to effectively extracts the similar features of the images and perform the accurate classification into multiple classes. The model is trained on large-scale publicly on the available datasets with proper pre-processing and augmentation methods applied to address data imbalance and variability. EyeNet achievements of competitive performance against recent state of the art approaches can be demonstrated using metrics which shows the accuracy of evaluation, precision, recall, and F1- score. This system goals to support and help the health care professionals such as ophthalmologists by providing fast, reliable Diabetic Retinopathy screening to get timely intervention, ultimately decreasing the risk of blindness and improving patient's outcomes at scale.**

***Index Terms*—Convolution Neural Networks (CNN), Diabetic Retinopathy (DR), Disease classification, Flask, Retinal Image Analysis, Vision.**

I. INTRODUCTION

Diabetic retinopathy is known to be one of the main and leading reason of blindness among the adults in all over the world, particularly with the person suffering from the diabetes mellitus for long time. This condition results from the very small blood vessels present in the retina caused by elevated blood glucose levels, causing a cascade of pathological changes. Diagnosis of the DR in the initial stage and classification of the images into distinct stages is serious as treatment strategies differs significantly depending on disease severity, timely intervention can avoid irreversible vision loss.

Usually, DR diagnosis and staging depends on the manual examination and testing of fundus images of retina by the skilled experts such as ophthalmologists. This process, though it is accurate, is labor-intensive, subjective, and tending to inter-observer variability. Additionally, access to the specialized retina care is in limit in many low-resource settings, which leads to the delays in the detection, diagnosis, and treatment. So, there is very much need of automation, scalable diagnostic systems that will help clinicians by providing goals and faster classification of DR severity from the images of retina.

Artificial intelligence has become an important asset of the images analysis in medical field. Among the different methods of AI such as deep learning, especially convolutional neural networks, has shown exceptional effectiveness in analysing complex visual patterns in images of ophthalmic. These networks are capable of obtaining hierarchical feature representations directly from huge datasets, facilitating identification and classification of accurate diseases.

This research presents EyeNet, an advanced deep learning system goals at splitting the images of diabetic retinopathy into five different clinical stages. EyeNet incorporates comprehensive pre-processing methods through which quality of the images gets improved and uses a CNN model specifically adapted for challenge of classifying multiple classes. The model here is trained on several datasets gathered from

different populations and addresses class imbalances through sampling strategies and data augmentation.

We carry on extreme experiments giving EyeNet effectiveness by using metrics like accuracy, precision, recall, and F1 score, benchmarking it against current approaches. The findings demonstrate notable enhancements in classification accuracy and speed, because of which that EyeNet might greatly reduce the workload for ophthalmologists and increase the reach of screening.

This study tries to offer a reliable, precise, and practical instrument for the automated DR classification, tackling critical issues in the detection and management of diabetes in patients before time worldwide.

II. LITERATURE SURVEY

Abràmoff M. D et al., (2018) [1], In the primary care clinics, this test checks if an autonomous AI system is as effective as the traditional grading methods. The system was 87.2% sensitive and 90.7% specific. Its field deployment is limited by integration and regulatory issues, although with encouraging results. Workflow compatibility and clinical validation are ongoing.

Bilal A et al., (2022) [2], Here in this research, a two-phase system which consists of CNN classification and U-Net segmentation is introduced. The model proved to be quite efficient with 97.92% accuracy in datasets like EyePACS-1. It also facilitated accurate DR staging. High-end hardware and good-quality data are required for best results.

Fenner B. J et al., (2018) [3], for DR screening, this review highlights the implementation of cutting-edge imaging technologies including ultra-wide-field imaging and AI-based interpretation. Such advances enable quick and accurate screening. Large-scale implementation of these is held back by a high cost. Scalability in public health systems is further influenced by the challenges of working operations.

Grauslund J, (2022) [4], the use of CNN-based deep learning in DR screening in clinical practice is investigated in this study. AI models manifested high diagnostic accuracy and reliability. Yet, compared to traditional fundus cameras, handheld AI-based devices yielded lower-quality images. The research recommends carrying out further studies on hardware development.

Gupta R, & Sharma T, (2023) [5], the study puts forward a method of automatic detection of DR based on CNN for retinal images. It is superior compared to conventional approaches based on specific identification of levels of DR. The model is based extensively on big data sets and intensive processing needs, which undermine practicality and scalability.

Kong M & Song S. J (2024) [6], this research explores the application of AI in DR diagnosis and telemedicine. It focuses on how AI systems improve eye care efficiency and access, especially in disadvantaged regions. Despite these advantages, the authors note that adoption is slowed down by regulatory hurdles. Among the major hurdles to uptake is the readiness of the healthcare system.

Lim et al., (2023) [7], Using Reading-Center ETDRS grading as the reference, this prospective, multicenter study contrasted the EyeArt AI DR screening system with dilated ophthalmoscopy performed by retina specialists and general ophthalmologists. Data from 521 participants (999 eyes) show that EyeArt has a lower specificity (~88.4% vs. ~99.6%) but a greater sensitivity (~96.4%) for mtmDR identification than ophthalmoscopy (~27.7%). The strategy might reduce the amount of work involved in screening, but it might also result in more false-positive referrals. The use of 2-field images, occasional false negatives, and the incapacity to detect non-retinal disorders are among its drawbacks. The study emphasizes AI's potential for triage, but it also highlights the need for additional validation and human verification.

R. Priya & P. Aruna (2012) [8], Performance of SVM and PNN models in DR classification is compared in this research. SVM performed better than PNN with accuracy at 97.6% compared to 89.6%. Both models, however, exhibited weaknesses in processing new, unseen data. High computational needs were also issues of concern.

Padhya S. K et al., (2019) [9], This study investigates the potential for reducing the effort associated with diabetic retinopathy (DR) screening through the use of artificial intelligence (AI). The authors emphasized the benefits of early detection using AI-based systems, which lessens reliance on specialists. The study emphasizes the need for human verification even though AI systems have potential. This is a result of the incomplete interpretability of current models.

Ramachandran Rajalakshmi et al., (2018) [10], the authors have employed fundus images captured

through a smartphone in testing an artificial intelligence tool to detect DR. The model had proved effective for use in low-resource settings when its sensitivity reached 95.8%. Despite this, handling erratic image quality with mobile-based systems still presents difficulties. For the application in real-world settings, there needs more optimization.

V. Gulshan et al., (2018) [11], In order to identify referable DR, this groundbreaking study developed a deep learning technique using a set of 128,175 retinal images from EyePACS and Indian hospitals. Following validation on two datasets (EyePACS-1 and Messidor-2), the AUCs were 0.991 and 0.990, respectively, with 97.5% and 96.1% sensitivities and 93.4% specificity. Comparable to ophthalmologists, the model was able to identify referable DR. Its benefits include comparison with human graders and access to large, diverse datasets. The need for predicted real-world testing before deployment, the dependence on high-quality images, and the inability to be interpreted are some of the drawbacks. This work made the FDA-approved AI DR systems viable.

Risa M. Wolf et al., (2024) [12], This study explores the ways in which artificial intelligence (AI) can reduce the workload associated with diabetic retinopathy (DR) screening. The authors stress the advantages of early diagnosis with AI-assisted systems, which reduce reliance on specialists. According to the study, human verification is still necessary even when AI systems show promise. This is because the models that are now in use cannot be fully explained.

III. MATH

In developing the model, several important mathematical concepts and performance metrics are involved:

A. Input & Output Mapping:

Let, X represent the input image, $f(.)$ is trained CNN model which maps the images given as input to the probability distribution, and $y`$ is model's predicted probability distribution across classes of DR.

$$y` = f(X)$$

B. Categorical Cross Entropy Loss:

The following loss function for multi-class classification with integer labels is used to train the model:

$$L = -\sum_{i=1}^{n} yi\ log(yi`)$$

In the above formula,

- yi is the ground truth label in one hot encoded form it uses 1 for correct class and 0 for others.
- $yi`$ is the predicted probability assigned for the class i.
- Definition: It measures the dissimilarity arose between the true labels and the predicted probability distribution, here in this the lower-level values indicates the better predictions.

C. Accuracy:

Measures the overall fraction of correctly classified samples:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

i.e. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

- TP (True Positive) - positive classes which are predicted correctly.
- TN (True Negative) - Negative classes which are predicted correctly.
- FP (False Positive) - positive classes which are not predicted correctly.
- FN (False Negative) - Negative classes which are not predicted correctly.

D. Precision and Recall:

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

The precision of a forecast is the proportion of accurate forecasts; all of the model's predictions were positive. Because of the high precision, we have a number of false alarms.

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

Recall is same as sensitivity, the recall isof all actual positive cases, the fraction in which the model is predicting correctly.

E. F1-Score:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-score is the balanced mean between precision and recall; it is better used for imbalanced datasets.

IV. METHODOLOGY

This paper describes a systematic procedure for developing an automated system that classifies the severity of diabetic retinopathy (DR) stages from

retinal fundus images using a Convolutional Neural Network (CNN). A web-based inference interface is then used to deploy the system.

The methodology is organized into phases:

- Data preparation
- CNN model design
- Model training and validation
- Deployment via a flask web application

4.1Data Preparation

4.1.1 Dataset Organization

The dataset is organized into two main folders:

- Train data: Contains training images categorized into 5 subfolders, each representing one DR stage.
- Test data: Contains testing images with the same subfolder structure.

Each subfolder name corresponds to a DR severity level:

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

4.1.2 Image Pre-processing

Prior to input into the model, all images undergo the following pre-processing steps:

- Resizing: Images are standardized to 224×224 pixels to align with the CNN input layer requirements.
- Normalization: Pixel values are scaled to the range, accelerating convergence during neural network training.
- Batching and Labelling: Libraries such as Keras, Image Data Generator are utilized to generate batches and assign labels based on directory hierarchy. Although augmentation is kept minimal in this work, the workflow can flexibly include rotations, flips, or brightness changes to increase data diversity and combat imbalance.

4.2 CNN Model Development

4.2.1 Network Architecture

A convolutional neural network that is designed for multiple class classification forms the basis of the automated diagnosis system. Using alternating convolutional and pooling layers, the network successively exacts hierarchical features before reaching fully connected layers for the prediction of the final DR stage.

Layer Type	Configuration	Function
Input	224×224×3(RGB)	Image reception
Conv2D	32 filters, 3×3, ReLU	Low level feature extaction
MaxPooling2D	2×2	Dimension reduction
Conv2D	64 filters, 3×3, ReLU	Complex feature extraction
MaxPooling2D	2×2	Additional down sampling
Flatten	_____	Convert 2D feature maps to ID vectors
Dense	128 units, ReLU	Non-linear feature combination
Dropout	0.5	Prevent overfitting
Dense(output)	5 units, Softmax	Multi class DR prediction

Fig.1. An Architecture of Convolutional Neural Networks for Classifying Multi-Class Diabetic Retinopathy

4.2.2 Compilation Settings

- Loss Function: The model uses sparse_categorical_crossentropy, suitable for integer labels in multiclass predictions.
- Optimizer: Adam optimizer provides adaptive learning rate adjustments for robust efficient training.
- Evaluation Metric: Accuracy is monitored across training and validation phases.

4.3 Model Training and Validation

The model is trained using the organized training dataset, with validation performed on a held- out test set to assess generalization. Training consists of:

- Epochs: The model undergoes 10 complete passes over the training data (epochs), sufficient for initial convergence with this dataset.

- Batch Size: Each batch contains 32 images for efficient GPU utilization and stable gradient computation.
- Model Savings: Upon completion, the best-performed model is saved in HDF5 format (dr_model.h5), ensuring reproducibility and easy deployment.

4.4 Deployment using Flask

A minimalistic web-based user interface is implemented using the Flask framework:

4.4.1 User Interaction

Users upload retinal images via a HTML form. The Flask backend processes submitted files and forwards them for model inference.

4.4.2 Image Handling and Prediction

The uploaded image is resized and normalized identically to the training protocol, loaded into memory, and input to the trained CNN. The model outputs a probability vector, the maximal value of which (via argmax) determines the predicted stage.

4.4.3 Result Presentation

The class label is mapped to a readable DR stage (e.g., "Moderate") before being displayed on the result page, delivering an immediate and understandable diagnostic prediction for clinical decision-making or patient awareness.

4.5 Class Label Summary

The model supports the following DR stage predictions:

Label	Description
0	No DR
1	Mild DR
2	Moderate DR
3	Severe DR
4	Proliferative/Florid DR

4.6 Tools and Frameworks

Tools/Library	Purpose
TensorFlow, Keras	Model development/training
Python	Scripting and backend logic
Flask	Web development
HTML	User interface (form upload)
NumPy	Array/numerical manipulation
PIL/OpenCV	Image reading and processing

4.7 Data Flow Diagram

The System workflow begins with image upload, followed by pre-processing feature extraction and classification via the CNN model. Classification results are then stored and communicated to healthcare professionals.

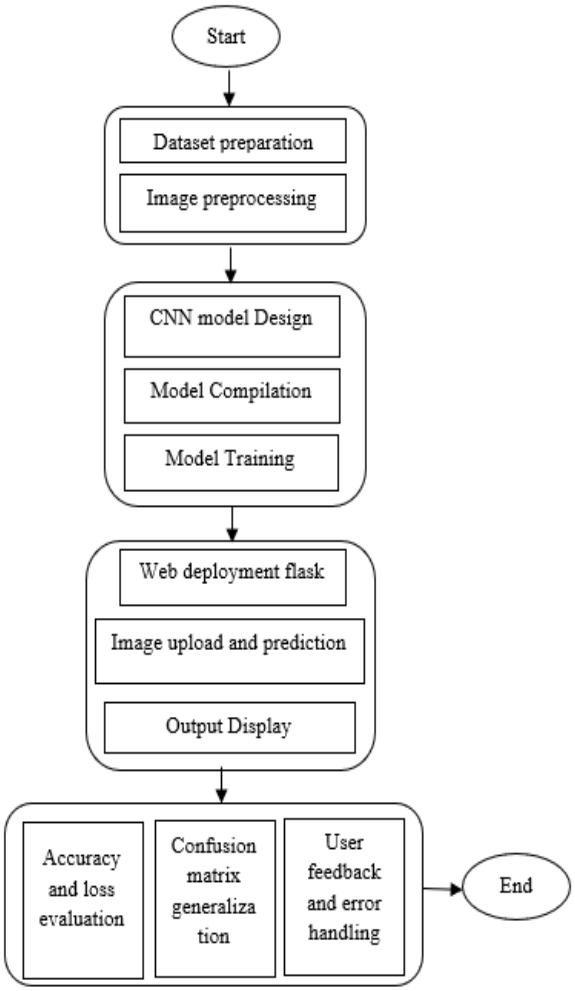

Fig. 2 DFD of the System

Data storage layers maintain records of raw images and classification results for auditing and further analysis.

V. RESULTS AND DISCUSSION

The performance of the EyeNet CNN model was quantitatively evaluated on an independent test set comprising 810 retinal fundus images, each properly categorized into one of five diabetic retinopathy (DR) stages.

The evaluation embraced standard metrics:

Accuracy trends throughout training epochs, precision, recall, F1-score, and visual analysis using a confusion matrix.

5.1 Classification Report

Class	Precision	Recall	F1-Score	Support
0 (No DR)	1.00	0.98	0.99	112
1 (Mild)	0.26	0.52	0.35	94
2 (Moderate DR)	0.64	0.54	0.59	104
3 (Severe DR)	0.48	0.20	0.29	230
4 (Proliferative DR)	0.48	0.59	0.53	270
Accuracy			0.52	810
Macro Avg	0.57	0.57	0.55	810
Weighted Avg	0.55	0.53	0.51	810

Fig.3 The classification Metrix across all DR classes

The findings show that the model almost always correctly recognizes healthy photos and hardly ever misclassifies them, with exceptional precision and recall for the "No DR" class. The lowest F1-score indicates poorer performance on the "Mild" class, indicating that the model struggles to discern subtle indications of the early DR. The proliferative, severe, and moderate classes all have moderate ability; the proliferative class's memory is noticeably superior to its precision. Mild and severe phases may visibly overlap with adjacent categories in clinical datasets that are unbalanced or difficult to interpret.

5.2 Confusion Matrix

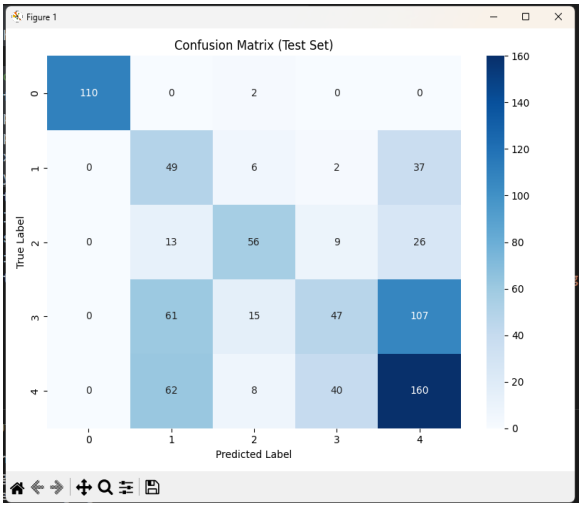

Fig. 4 Confusion matrix

The Confusion matrix further shows how frequently images from each true category are correctly or incorrectly classified. Principally, most misclassifications occur among adjacent disease stages, particularly between moderate, severe, and proliferative DR, reflecting the subtle visual graduations among these classes.

5.3 Accuracy and Training Trends

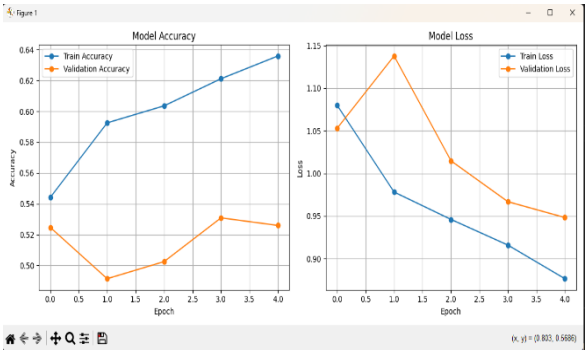

Fig. 5 Accuracy graph

The training and validation accuracy curve in the graphic below shows the learning trend over epochs. Convergence, stability, and any signs of over- or under-fitting are all confirmed by plotting.

The final test accuracy achieved by the model was 52%. While this metric demonstrates basic model learning, it highlights the areas for the improvement, especially in recalling mild and severe DR classes.

5.4 Discussion

EyeNet's test set performance illustrates common patterns in medical imaging classification:

- Outstanding accuracy is seen in the discriminating healthy (No DR) pictures, especially for the screening uses when lowering false negatives is imperative.
- Weakness: Probable training data constraints, subtle feature variations, and class imbalance could account for low precision and recall in the "Mild" and "Severe" categories. Relatively low F1scores and the confusion matrix both point toward overlap between some DR levels.
- Model constraints: The poor overall accuracy and varying recall across the stages suggest the necessity for more research, including dataset balancing, advanced augmentation, clinical contextualization, or more complex model techniques (e.g., attention mechanisms or ensemble approaches).
- Future Directions: By targeted augmentation, use of transfer learning, or inclusion of multimodal data, early-stage detection can be improved to improve robustness and clinical relevance.

Generally, the results confirm the CNN approach for multi stage diabetic retinopathy categorization but emphasize ongoing difficulties in distinguishing small retinal changes. Constant improvement and validation with bigger, more diversified datasets is advised to maximize generalizability and practical applicability.

VI. CONCLUSION

In order to determine the stages of diabetic retinopathy, this research presents EyeNet, a sophisticated and automated method that makes use of convolutional neural networks. The adoption of this methodology includes a proprietary CNN design, a comprehensive data pre-processing, and an intuitive web interface. According to experimental findings, the model works very well in the No DR and Proliferative categories and correctly differentiates between the five stages of diabetic retinopathy. Although the outcomes for the Mild and Severe stages emphasize the challenges of discerning tiny clinical differences, the model's overall performance suggests its potential usefulness in real-world screening scenarios. Notwithstanding certain drawbacks, such as dataset imbalance and early-stage DR detection challenges, the use of bigger, more varied training datasets and the deployment of state-of-the-art deep learning techniques could be enhancements. A way to affordable, scalable diabetic retinopathy screening is provided by the system's fully automated, web-deployable architecture, which guarantees improved access, especially in regions with little ophthalmological competence.

REFERENCES

[1] Abràmoff M. D, Lavin P. T, Birch M, Shah N, & Folk J. C, "Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices", NPJ Digital Medicine, 1, 39, 2018.

[2] Bilal A, Zhu L, Deng A, Lu H, & Wu N, "AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning", Symmetry, 14(7), 1427, 2022.

[3] Fenner B. J, Wong R. L. M, Lam W-C, Tan G. S. W, & Cheung G. C. M, "Advances in retinal imaging and applications in diabetic retinopathy screening: A review", Ophthalmology and Therapy, 7, 333-346, 2018.

[4] Grauslund J, "Diabetic retinopathy screening in the emerging era of artificial intelligence", Diabetologia, 65(6), 1415-1423, 2022.

[5] Gupta R, & Sharma T, "Automated Detection of Diabetic Retinopathy Using Deep Learning", IEEE Access, 11, 124789-124798, 2023.

[6] Kong M, & Song S. J, "Artificial intelligence applications in diabetic retinopathy: What we have now and what to expect in the future", Endocrinology and Metabolism, 39, 416-424, 2024.

[7] J. I. Lim et al., "Artificial Intelligence Detection of Diabetic Retinopathy: Subgroup comparison of the EyeArt system with ophthalmologists' dilated examinations," Ophthalmology Science, vol. 3, p. 100228, 2023.

[8] Priya R, & Aruna P, "SVM and neural network-based diagnosis of diabetic retinopathy", International Journal of Computer Applications, 41(1), 6-12, 2012.

[9] Padhy S. K, Takkar B, Chawla R, & Kumar A, "Artificial intelligence in diabetic retinopathy: A natural step to the future", Indian Journal of Ophthalmology, 67(7), 1004-1011, 2019.

[10] Rajalakshmi R, Subashini R, Anjana R. M, & Mohan V, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence", Eye, 32, 1138-1144, 2018.

[11] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[12] Wolf R. M, et al., "Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: The ACCESS randomized control trial" Nature Communications, 15, 421, 2024.S.

[13] Wang X. N., et al., "Automatic Grading System for Diabetic Retinopathy Diagnosis Using Deep Learning Artificial Intelligence Software." Current Eye Research, 45(12), 1550–1555, 2020.

[14] Xia P et al., "Generalizing to Unseen Domains in Diabetic Retinopathy with Disentangled Representations." arXiv preprint arXiv:2406.06384, 2024.

[15] Yang Y et al., “Robust Collaborative Learning of Patch-level and Image-level Annotations for Diabetic Retinopathy Grading from Fundus Image.” arXiv preprint arXiv:2008.00610, 2020.
[16] Yao Z et al., “FunSwin: A deep learning method to analyze diabetic retinopathy grade and macular edema risk based on fundus images.” Frontiers in Physiology, 13, 961386, 2022.
[17] Zhang et al., “Automated detection of severe diabetic retinopathy using deep learning method.” Graefes Archive for Clinical and Experimental Ophthalmology, 260(3), 849–856, 2022.
[18] Zhao R., et al., “XLOOK-DR: Multi-Stage Lesion-Guided Network for Lesion Location and Diabetic Retinopathy Grading.” arXiv preprint arXiv:2411.00726, 2024.