

An Enhanced Multi-View Deep-Broad Learning Framework for Reliable Retinal Disease Classification

Thorat Priya Somnath¹, Prof. N. M. Kolhe², Prof. D. S. Kshirsagar³, Dr. H. B. Jadhav³

Student, Department of Computer Engineering¹

Guide, Department of Computer Engineering²

Co-Guide, Department of Computer Engineering³

HOD, Department of Computer Engineering⁴

Adsul's Technical Campus, Chas, Ahilyanagar, India

Abstract: Optical coherence tomography (OCT) provides cross-sectional views of retinal structures and is widely used to assess conditions that affect the macula. Automated analysis may help clinicians review large volumes of images, but a model's performance can be affected by image quality, differences between datasets, class imbalance, and uncertainty in its predictions. This paper proposes a reliability-oriented framework for classifying OCT images into four categories: choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, and normal retina. The proposed pipeline combines basic image-quality screening, three complementary image representations, feature extraction, attention-guided fusion, and a broad learning classification stage. A confidence-based referral step is also included so that unsuitable or uncertain cases can be flagged for expert review rather than automatically assigned a definitive label. Recent research from 2024 to 2026 is reviewed to place the proposal in context. In particular, a 2024 study has already investigated multi-view deep-broad learning for retinal disease assessment, while later OCT studies have examined robustness, hybrid CNN–Transformer architectures, lesion-aware fusion, and calibration. Therefore, the potential contribution of this work is framed around evaluating reliability and referral behaviour within a clearly controlled OCT classification experiment, not around claiming that multi-view deep-broad learning is a new idea. No experimental results are claimed in this manuscript; implementation and validation are required before the framework's effectiveness can be established.

Keywords: Optical coherence tomography, retinal disease classification, multi-view representation, broad learning system, feature fusion, model calibration, uncertainty, referral.

I. INTRODUCTION

Retinal diseases can affect central vision and may require timely clinical assessment. Optical coherence tomography (OCT) supports this process by producing cross-sectional images of retinal tissue, allowing clinicians to examine changes in retinal layers and structures. Reviewing many scans can be demanding, and image interpretation requires specialized knowledge. These factors have encouraged research into computer-assisted OCT analysis.

Deep-learning methods can learn image features directly from training data. Earlier work showed that convolutional neural networks could distinguish common retinal conditions in OCT images [1], and clinically oriented systems have explored diagnosis and referral using retinal imaging [2]. Residual networks and densely connected networks are widely used architectures for learning visual representations [3], [4]. However, strong results on one dataset do not automatically establish reliable performance on images collected from different devices, clinical settings, or patient groups.



Recent OCT research increasingly considers robustness and model reliability alongside classification accuracy. Zhang et al. introduced FlexiVarViT in 2025 to address variation in OCT scan structure and resolution [6]. In 2026, published studies investigated hybrid CNN–Transformer classification [7], lesion-attentive fusion across OCT volumes [8], and approaches that explicitly consider calibration and statistical validation [9]. These developments suggest that evaluation should look beyond a single accuracy score.

This paper proposes a reliability-oriented multi-view framework. Each OCT image is represented in its original form, through a contrast-enhanced view, and through a texture/edge view. Features from these views are combined using learned attention weights and passed to a broad learning classifier. A quality gate and confidence-based referral mechanism are included as components to be tested. The purpose is not to claim that these components are individually novel, but to define an experiment that measures whether they provide measurable value for the selected four-class task.

A. Objectives and Contributions

The main objectives of this study are:

- To define a reproducible pipeline for four-class OCT image classification: CNV, DME, drusen, and normal retina.
- To compare a single-view model, multi-view feature concatenation, and attention-guided multi-view fusion under the same data split.
- To investigate whether a broad learning output stage offers a useful accuracy–cost trade-off against conventional classification heads.
- To evaluate image-quality screening and confidence-based referral using explicit measures, rather than treating confidence as proof of correctness.
- To report limitations transparently and avoid claims about performance until the proposed method has been implemented and tested.

II. EXISTING SYSTEM

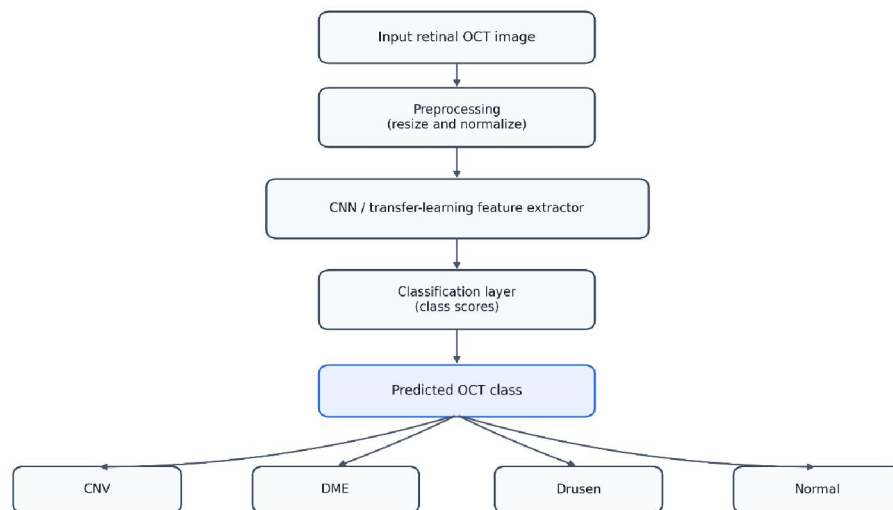


Fig. 1: General architecture of a conventional retinal OCT image classification system

A common automated OCT classification pipeline begins with labelled images, applies preprocessing, extracts features using a convolutional neural network or a transfer-learning model, and predicts a class through a final classification



layer. Residual networks and DenseNet are examples of architectures that can serve as baselines [3], [4]. Other approaches focus on segmentation or on clinical referral rather than image-level classification alone [2], [5].

A. Challenges in Existing Pipelines

Several risks should be considered when designing automated OCT classification systems:

- Image quality: Blur, noise, artefacts, and low contrast may obscure relevant retinal structures.
- Limited representation: A single image representation may not emphasize every useful structural or textural cue.
- Class imbalance: A model can favour categories that appear more often in the training data.
- Generalization: Performance may change when test images differ from training images in acquisition protocol, device, or population.
- Uncertain predictions: A conventional classifier may return its highest-scoring class even when all class scores are unreliable.
- Resource requirements: Larger networks may increase training time, memory use, and inference cost.

These points describe risks that should be checked, not defects that necessarily affect every existing model. Their practical importance depends on the dataset, preprocessing, training procedure, and evaluation design.

III. REVIEW OF RESEARCH PUBLISHED IN 2024–2026

The recent literature shows that retinal OCT classification is evolving quickly. Researchers are exploring multi-view learning, hybrid neural-network architectures, robustness to image variation, and prediction reliability. The studies below are relevant to the proposed framework, but they do not all use the same dataset, class definitions, or evaluation protocol. Their reported metrics should therefore not be compared as if they came from one controlled experiment.

A. Recent Research Studies

1. Zhang et al. (2024)

Topic: A novel artificial-intelligence-based approach for automatic assessment of retinal disease images using a multi-view deep-broad learning network.

This study is particularly relevant because it already combines multi-view representations with a deep-broad learning network for retinal disease assessment.

Relevance to this paper: It is the closest conceptual predecessor to the proposed framework. Therefore, this paper should not claim that combining multi-view learning and deep-broad learning is entirely new. The proposed study must demonstrate its differences through its experimental design, reliability evaluation, and measured results.

2. Zhang et al. (2025)

Topic: A robust deep-learning classifier for screening multiple retinal diseases on optical coherence tomography.

This research investigates the use of deep learning for retinal disease screening while considering variation in OCT images, including differences in scan structure and resolution.

Relevance to this paper: It highlights the importance of robustness and generalization. The proposed framework should evaluate whether its performance remains consistent across suitable test conditions instead of relying only on accuracy measured on one dataset.

3. Liu et al. (2026)

Topic: MergeNeXt: A hybrid CNN–Transformer model for retinal OCT image classification.

This work explores a hybrid architecture that combines convolutional neural networks with Transformer-based processing and attention mechanisms.

Relevance to this paper: It demonstrates that feature extraction and feature fusion remain active research areas. However, its reported performance cannot be assumed for the proposed EMV-DBL framework because the models and experimental conditions differ.



4. Ko et al. (2026)

Topic: Retinal disease classification from spectral-domain optical coherence tomography using lesion-attentive fusion.

This study investigates feature fusion that gives greater attention to informative retinal lesions across OCT data.

Relevance to this paper: It emphasizes the importance of identifying useful visual information. The proposed framework instead evaluates three image representations and attention-guided fusion. A fair comparison would require consistent datasets and evaluation conditions.

5. Vijayan et al. (2026)

Topic: A spatial adaptive multi-scale ConvNeXt framework for robust, calibrated, and statistically validated multi-class retinal OCT classification.

This work focuses on robustness and calibration in addition to classification performance.

Relevance to this paper: It supports evaluating whether predicted probabilities are trustworthy, whether performance is consistent, and how the model behaves when predictions are uncertain.

6. Al Zuabi et al. (2026)

Topic: Automated retinal disease classification from optical coherence tomography images using deep learning.

This study compares pretrained convolutional neural networks for multiple OCT diagnostic categories and reports measures beyond accuracy, including class-wise performance and calibration-related evaluation.

Relevance to this paper: It provides a useful reference for experimental design. Its seven-class results, however, should not be directly compared with the proposed four-class classification task.

B. Synthesis of the Literature

Three main observations emerge from the reviewed studies.

First, multi-view deep-broad learning already has a published retinal-imaging precedent in 2024. Consequently, the proposed study must establish a specific and measurable contribution rather than relying on the combination of these techniques as its main novelty.

Second, recent research explores different ways to preserve anatomical detail and combine local and global visual features. These include variable-resolution processing, hybrid CNN–Transformer architectures, and lesion-aware fusion.

Third, reliability is increasingly assessed through class-wise performance, calibration, patient-level data separation, and independent validation—not accuracy alone.

The studies use different datasets, class taxonomies, image units, and test protocols. Their published scores provide context, but they cannot establish how the proposed method will perform. A fair comparison requires implementation on the same dataset with identical patient-level splits and consistent evaluation rules.

IV. RESEARCH GAP AND MOTIVATION

The literature does not support claiming that multi-view learning, attention-based fusion, or deep-broad learning is individually new for retinal image analysis. In particular, the 2024 study by Zhang et al. already proposes a multi-view deep-broad learning network for retinal disease image assessment.

A more defensible research question is whether a clearly specified multi-view pipeline, combined with explicit quality screening and confidence-based referral, can improve the balance between classification performance, computational cost, and reliability for a defined four-class OCT task.

The proposed research gap is therefore focused on four objectives:

- Reproducible representations: Define the image transformations and feature-fusion procedure clearly enough for other researchers to reproduce the experiment.
- Controlled comparison: Compare the proposed framework against single-view and conventional deep-learning baselines using identical data partitions.



- Reliability evaluation: Measure calibration and referral behaviour in addition to conventional classification metrics.
- Transparent reporting: Report performance variability, computational requirements, and limitations.

This research gap must be reassessed after a full-text comparison of the 2024 predecessor and the selected 2025–2026 studies. If the implemented system does not demonstrate a measurable and distinct improvement, the contribution should be narrowed accordingly.

V. PROPOSED SYSTEM

The proposed framework has six stages:

- Input validation.
- Image preprocessing.
- Multi-view generation.
- Feature extraction.
- Attention-guided feature fusion.
- Classification with confidence assessment.

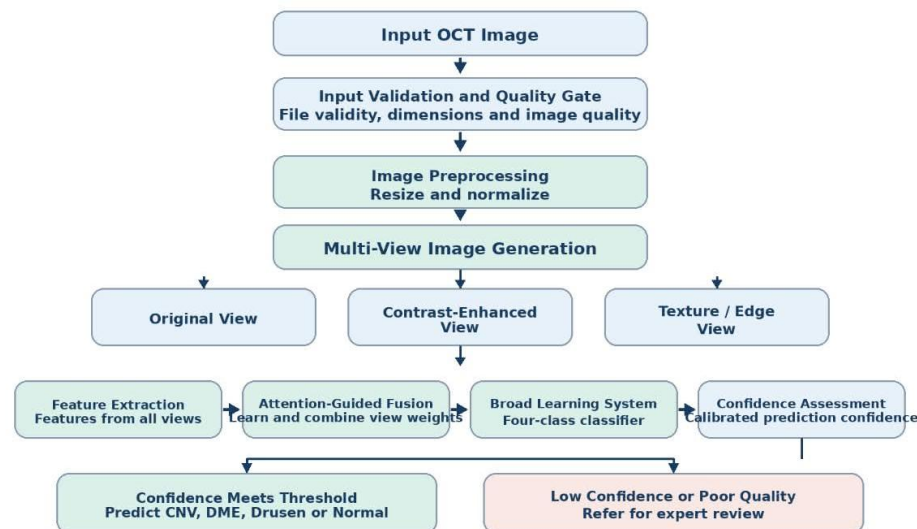


Fig.2: Proposed reliability-oriented multi-view framework for retinal OCT classification.

A. Input Validation and Preprocessing

Each input image is checked for file readability, valid dimensions, and predefined quality problems that can be detected consistently. The quality rules must be specified before testing.

Images that fail the checks are flagged for review rather than being passed automatically to the classifier. Accepted images are resized only as required by the selected feature extractor and normalized using one documented procedure.

Additional operations such as denoising or contrast enhancement should be selected carefully because they may remove or exaggerate clinically relevant structures. Any preprocessing parameters must be established without using test labels.

Data augmentation, if used, should be limited to training data and should preserve the anatomical meaning of the image.



B. Multi-View Representation

Three representations are proposed for each source OCT image.

1. Original view

The normalized image preserves the source intensity and structural information. It provides the reference representation for the classification process.

2. Contrast-enhanced view

This image is processed using one predefined contrast-enhancement method to emphasize local intensity differences. It may help reveal structural patterns that are less prominent in the original representation.

3. Texture/edge view

This representation is produced by a fixed texture or edge operator to describe local patterns and structural boundaries. The exact enhancement and edge/texture operators must be named and fixed in the final implementation. These views are derived from the same image and should not be described as independent clinical observations. Their contribution must be verified through ablation experiments.

C. Feature Extraction

Each view is passed through a feature extractor to obtain a numerical representation.

A shared backbone with separate view inputs or separate branches may be used. However, the final paper must state the selected architecture, pretrained weights, input size, layer used for feature extraction, and whether the backbone is frozen or fine-tuned.

Let h_1 , h_2 , and h_3 represent the feature vectors from the original, contrast-enhanced, and texture/edge views, respectively.

The feature vectors are projected to compatible dimensions before fusion. The feature extractors are trained using the training partition, while architecture and hyperparameter selection are performed using the validation partition.

D. Attention-Guided Feature Fusion

The fusion module assigns a learned score s_v to each view v . A softmax operation converts these scores into non-negative weights that sum to one:

$$\alpha_v = \frac{\exp(s_v)}{\sum_{j=1}^V \exp(s_j)}$$

Where:

s_v is the learned score for view v .

α_v is the attention weight assigned to that view.

V is the total number of views.

The fused representation is calculated as:

$$z = \sum_{v=1}^V \alpha_v h_v$$

where h_v represents the feature vector for view v , and z is the resulting fused feature vector.

The implementation must explain how s_v is generated, how the fusion module is trained, and whether attention scores are computed for each image or in another manner.



The attention weights are intended to learn which representations are useful for classification. However, attention weights should not automatically be interpreted as clinically valid explanations.

E. Broad Learning Classification

The fused representation is passed to a broad learning system (BLS), which uses feature and enhancement nodes rather than relying only on adding more sequential neural-network layers.

For a regularized least-squares output stage, let Z denote the training design matrix produced by the broad-learning representation and Y the one-hot encoded class labels. The output weights may be estimated as:

$$W = (Z^T Z + \lambda I)^{-1} Z^T Y$$

where:

W is the output-weight matrix.

Z is the broad-learning design matrix.

Y contains the training labels.

λ is the regularization parameter.

I is the identity matrix.

In the implementation, a numerically stable linear solver should be used rather than explicitly computing a matrix inverse.

The final paper must document the number and type of nodes, activation functions, regularization selection, and how class scores are converted into probabilities.

F. Quality Gate and Confidence-Based Referral

The system combines the quality-check outcome with the classifier's prediction confidence.

Images that fail the predefined quality gate are referred for review. For acceptable images, a confidence threshold is selected using the validation set only. Predictions below this threshold are flagged for expert review rather than treated as reliable automatic classifications.

The threshold and the meaning of confidence must be documented. Probability estimates should be calibrated if they are used to guide referral decisions.

Referral performance should be reported using:

The proportion of images referred.

The proportion retained for automatic classification, known as coverage.

Classification performance among the retained cases.

A confidence score alone does not prove that a prediction is correct, and referral is not a substitute for clinical validation.

VI. EXPERIMENTAL DESIGN AND EVALUATION

A. Dataset and Class Definitions

The experiment requires a publicly available, appropriately licensed OCT dataset that contains the four intended categories:

Choroidal neovascularisation (CNV).

Diabetic macular edema (DME).

Drusen.

Normal retina.

Before training, the authors must verify the dataset version, source, image counts per class, label definitions, patient identifiers if available, and permitted use.

These details should be reported only after checking the actual dataset. No dataset size is assumed in this manuscript.



If multiple images originate from the same patient, all images from one patient should remain in a single partition. Patient-level separation reduces the risk that related images appear in both training and test sets. If patient identifiers are unavailable, that limitation must be stated and the data-splitting procedure described clearly.

B. Data Splits and Training Protocol

The dataset should be divided into training, validation, and held-out test sets before model development.

Training set: Used to fit the model.

Validation set: Used to select architectures, hyperparameters, calibration methods, and the referral threshold.

Test set: Used once for final evaluation.

All baselines must use the same partitions.

The final paper should document:

Image size.

Optimizer and learning rate.

Batch size and number of epochs.

Early-stopping rule.

Random seeds.

Hardware and software versions.

Class-balancing strategy, if used.

C. Baseline Models

The following models are proposed for comparison:

A CNN trained using the original view only.

A residual-network transfer-learning baseline.

A DenseNet transfer-learning baseline.

Three-view feature concatenation without attention.

Attention-guided multi-view fusion with a conventional classification head.

The complete proposed system, including the broad learning classifier and referral process.

Where feasible, a faithful implementation of the 2024 multi-view deep-broad learning method, with differences from the published method documented.

The comparison with the 2024 method is particularly important because it is the closest conceptual predecessor. If a faithful reproduction is not feasible, the paper should explain why and avoid implying that the proposed method outperforms it.

D. Evaluation Metrics

The experiment should report the following measures.

1. Accuracy and balanced accuracy

Accuracy measures the proportion of correctly classified images. Balanced accuracy gives equal importance to the recall of each class.

2. Precision, recall, and F1 score

These measures should be reported for each class, together with macro-averaged values.

3. Multiclass area under the receiver operating characteristic curve (AUROC)

The averaging strategy and calculation method should be specified.

4. Confusion matrix

The confusion matrix shows correct and incorrect predictions for each category and helps identify which diseases the model finds difficult to distinguish.



5. Calibration metrics

Expected calibration error (ECE) or another suitable calibration measure can help assess whether predicted probabilities agree with observed correctness. The calibration method should be reported.

6. Computational cost

Inference latency, parameter count or model size, and relevant hardware details should be included.

7. Referral metrics

For the confidence-based referral mechanism, report the referral rate, coverage, and classification performance among cases retained for automatic prediction.

When feasible, confidence intervals or variation across repeated runs should be reported. A single point estimate can hide instability, particularly when the test set is small or class counts are unequal.

E. Ablation Study

Ablation experiments are needed to identify which components contribute to the final system.

The following configurations are proposed:

Original view only.

All three views combined by simple concatenation.

All three views combined by attention-guided fusion.

Attention-guided fusion with a conventional classification head.

Attention-guided fusion with the broad learning classifier.

The full framework with quality screening and confidence-based referral.

The results should establish whether each additional component improves performance or reliability enough to justify its computational cost. If a component does not provide a measurable benefit, that result should be reported rather than omitted.

F. Results and Reporting Status

This manuscript does not contain experimental results because the proposed framework has not yet been implemented and evaluated in the supplied draft.

After running the experiments, the authors should add:

A results table comparing all baseline models.

Confusion matrices.

Class-wise precision, recall, and F1 scores.

Calibration results.

Referral analysis.

Ablation-study findings.

Computational cost measurements.

All reported values must be generated from the documented experiment. Results from other papers must remain clearly attributed to those papers.

VII. DISCUSSION

The central question is not whether combining several representations sounds beneficial, but whether the combination improves classification or reliability under a controlled evaluation.

A multi-view model may capture complementary cues, yet it can also introduce redundant features, extra computation, or sensitivity to the chosen image transformations. Similarly, broad learning may reduce the complexity of the classification stage in some settings, but its practical advantage must be measured against strong and fairly trained baselines.



The recent literature also cautions against relying on one test dataset or a single headline score. Studies from 2025 and 2026 address input variability, hybrid feature learning, volume-level lesion weighting, and calibration.

The proposed study should therefore make its dataset scope explicit and distinguish internal test performance from external generalization. It should not claim clinical readiness without independent validation and review by eye-care professionals.

VIII. CONCLUSION AND FUTURE WORK

This paper has outlined a reliability-oriented framework for four-class retinal OCT image classification. It combines image-quality screening, three complementary image views, feature extraction, attention-guided fusion, a broad learning classifier, and confidence-based referral.

The review of studies published from 2024 to 2026 shows that multi-view deep-broad learning already has a retinal-imaging precedent and that recent OCT research increasingly examines robustness, fusion, calibration, and validation.

The next step is to implement the framework and evaluate it against appropriate baselines using identical, preferably patient-level, data partitions. Ablation studies should establish the value of each component, while calibration and referral metrics should test whether the system can identify uncertain cases.

Until those experiments are completed, no claim of improved accuracy, reduced cost, or clinical benefit can be made. Any future clinical use would require independent validation, assessment across devices and populations, and oversight by qualified professionals.

ACKNOWLEDGMENT

The authors acknowledge the researchers whose published work on OCT imaging, retinal disease classification, deep learning, and broad learning systems informed the proposed study. Any dataset used in future experiments should be acknowledged according to its licence and citation requirements.

REFERENCES

- [1] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.
- [2] J. De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Medicine*, vol. 24, pp. 1342–1350, 2018, doi: 10.1038/s41591-018-0107-6.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [4] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.
- [5] C. S. Lee, A. J. Tying, S. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," *Biomedical Optics Express*, vol. 8, no. 7, pp. 3440–3448, 2017, doi: 10.1364/BOE.8.003440.
- [6] P. Zhang, G. Quellec, S. Matta, et al., "A robust deep learning classifier for screening multiple retinal diseases on optical coherence tomography," *Scientific Reports*, vol. 15, art. 35334, 2025, doi: 10.1038/s41598-025-19286-y.
- [7] J. Liu, Y. Yuan, S. Xue, et al., "MergeNeXt: a hybrid CNN–transformer model for retinal OCT image classification," *Scientific Reports*, 2026, doi: 10.1038/s41598-026-70044-0.
- [8] J. Ko, J. Han, J. Yoon, et al., "Retinal disease classification from spectral domain optical coherence tomography using lesion attentive fusion," *Scientific Reports*, 2026, doi: 10.1038/s41598-026-64770-8.
- [9] M. Vijayan, G. Veerapu, and N. J. S., "A spatial adaptive multi-scale ConvNeXt framework for robust, calibrated, and statistically validated multi-class retinal OCT classification," *Frontiers in Artificial Intelligence*, vol. 9, 2026, doi: 10.3389/frai.2026.1898978.



- [10] A. Zhang, X. Qian, C. Xu, and J. Zhang, "A novel artificial-intelligence-based approach for automatic assessment of retinal disease images using multi-view deep-broad learning network," IEEE Access, vol. 12, pp. 13248–13259, 2024, doi: 10.1109/ACCESS.2024.3356824.
- [11] B. A. Al Zuabi et al., "Automated retinal disease classification from optical coherence tomography images using deep learning," International Journal of Retina and Vitreous, 2026, doi: 10.1186/s40942-026-00929-9.
- [12] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 1, pp. 10–24, 2018, doi: 10.1109/TNNLS.2017.2716952.

