SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Multi-Label Retinal Screening System for Disease Detection with Hybrid CNN

Dhanashri D. Dhobale[1*], Deepika Patil[2]

[1*]Department of Computer Science and Engineering, School of Technology, Sanjay Ghodawat University, Kolhapur, INDIA.
E-mail: ddhokate29@gmail.com, Orchid id: https://orcid.org/0009-0008-0016-8969
[2]Department of Computer Science and Engineering, School of Technology, Sanjay Ghodawat University, Kolhapur, INDIA.
E-mail: Deepika.patil@cs.sguk.ac.in, Orchid id: https://orcid.org/0009-0004-7501-8567

ABSTRACT
Retinal disease diagnosis using fundus imaging is a critical task in ophthalmology. Traditional deep learning approaches often struggle with uneven illumination, small lesion detection, and multi-label disease co-occurrence. This paper presents an upgraded hybrid framework integrating disease-aware illumination normalization, multi-scale lesion-sensitive CNN, Vision Transformer (ViT), bidirectional LSTM (BiLSTM), attention refinement, and label-dependency guided prediction. Experimental evaluation on a large-scale dataset of 249,620 fundus images annotated for 39 retinal diseases demonstrates superior performance compared to previous CNN–ViT–LSTM models, achieving a macro-F1 score of 0.944 and reducing hamming loss to 0.018. These results confirm the clinical robustness of the proposed methodology.
The upgraded framework extends the earlier hybrid CNN–ViT–LSTM retinal disease diagnosis system by incorporating three major methodological improvements: disease-aware illumination normalization, multi-scale lesion-sensitive feature extraction, and label-dependency guided multi-label prediction. The overall objective of the upgraded design is to improve robustness against fundus image quality variations, enhance sensitivity toward small retinal lesions, and improve the recognition of co-existing ophthalmic conditions.

Keywords: CNN, Vision Transform, Attention mechanism, Bi-LSTM

Introduction

Retinal diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration are leading causes of vision impairment worldwide (Wong et al., 2016). Automated diagnosis using deep learning has shown promise, but challenges remain in handling multi-label conditions, rare disease detection, and image quality variations (Gulshan et al., 2016; Li et al., 2022). Recent advances in hybrid architectures combining CNNs and Transformers have improved global context modeling (Dosovitskiy et al., 2021), yet limitations persist in lesion sensitivity and label correlation learning.
This study proposes an upgraded methodology that integrates disease-aware preprocessing, multi-scale CNN lesion extraction, ViT-based global context modeling, BiLSTM dependency learning, and label-dependency guided prediction.
Let the input fundus image be represented as

$$I \in \mathbb{R}^{H\times W\times 3},$$

where H and W denote image height and width, respectively. The proposed upgraded model generates a multi-label probability vector

$$\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_C],$$

where $C = 39$ denotes the number of retinal diseases and conditions, and $\hat{y}_c \in [0,1]$ represents the predicted probability of the c-th disease.

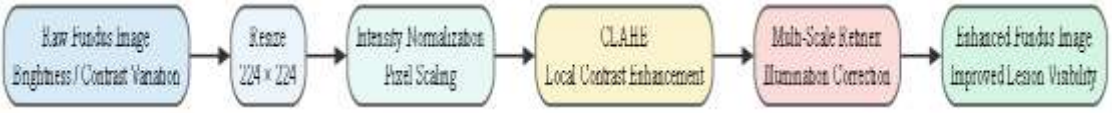
2. Methodology

2.1 Disease-Aware Illumination Normalization

Fundus images often suffer from uneven illumination, low contrast, glare, and device-dependent color distortion. In the upgraded framework, illumination correction is not treated as a separate auxiliary step; instead, it is designed as a disease-aware preprocessing stage. The input image is first normalized and then enhanced using a combined CLAHE–Retinex strategy.
For CLAHE-based enhancement, the image is divided into local tiles. For each tile T_k, the clipped cumulative distribution function is computed as

$$S_k(r) = \frac{L-1}{|T_k|}\sum_{j=0}^{r} h_k^{clip}(j),$$

where L is the number of gray levels, $|T_k|$ is the number of pixels in the k-th tile, and $h_k^{clip}(j)$ denotes the clipped histogram value at intensity level j.



Upgraded disease-aware preprocessing pipeline using CLAHE and Multi-Scale Retinex enhancement.

The Retinex-based illumination correction assumes that the observed fundus image is formed as

$$I(x,y) = R(x,y) \cdot L(x,y),$$

where $R(x,y)$ is the reflectance component and $L(x,y)$ is the illumination component. The Single-Scale Retinex output is computed as

$$R_{SSR}(x,y) = \log I(x,y) - \log[G_\sigma(x,y) * I(x,y)],$$

where G_σ is a Gaussian kernel with standard deviation σ, and $*$ denotes convolution.

For improved robustness, Multi-Scale Retinex is used:

$$R_{MSR}(x,y) = \sum_{s=1}^{S} w_s \left[\log I(x,y) - \log\left(G_{\sigma_s}(x,y) * I(x,y)\right)\right],$$

where S is the number of scales and w_s is the corresponding scale weight.

The final enhanced image is obtained as

$$I_e = \lambda_1 I_{CLAHE} + \lambda_2 R_{MSR},$$

where λ_1 and λ_2 control the contribution of contrast enhancement and illumination correction.

2.2 Multi-Scale Lesion-Sensitive CNN Stem:

The previous CNN module is upgraded into a multi-scale lesion-sensitive convolutional stem. This module extracts retinal features at different receptive field sizes to detect microaneurysms, hemorrhages, exudates, optic disc variations, and vessel abnormalities.

Given the enhanced image I_e, three parallel convolutional branches are used:

$$F_3 = \phi_{3\times3}(I_e), \quad F_5 = \phi_{5\times5}(I_e), \quad F_7 = \phi_{7\times7}(I_e),$$

where $\phi_{3\times3}$, $\phi_{5\times5}$, and $\phi_{7\times7}$ represent convolutional branches with different kernel sizes.

The multi-scale feature map is obtained by concatenation:

$$F_{ms} = \text{Concat}(F_3, F_5, F_7).$$

To reduce redundant channels and emphasize lesion-relevant responses, a channel attention mechanism is applied:

$$z = GAP(F_{ms}),$$
$$s = \sigma\left(W_2\delta(W_1 z)\right),$$
$$F_{ca} = F_{ms} \otimes s,$$

where $GAP(\cdot)$ denotes global average pooling, $\delta(\cdot)$ is the ReLU activation, $\sigma(\cdot)$ is the sigmoid activation, and $\otimes$ represents channel-wise multiplication.

2.3 Patch Embedding and Global Context Transformer

The channel-refined feature map F_{ca} is divided into N non-overlapping patches. Each patch is flattened and projected into an embedding vector:

$$e_i = W_p p_i + b_p,$$

where p_i is the i-th flattened patch, W_p is the learnable projection matrix, and e_i is the patch token.

The positional encoding is added as

$$t_i = e_i + pos_i.$$

The transformer encoder applies multi-head self-attention:

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V.$$

For h attention heads, the multi-head attention output is

$$MHSA(T) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)W_o.$$

The transformer output is computed as

$$T' = LN\left(T + MHSA(T)\right),$$
$$T_g = LN\left(T' + FFN(T')\right),$$

where $LN(\cdot)$ denotes layer normalization and $FFN(\cdot)$ denotes the feed-forward network.

2.4 Spatial Token Dependency Learning Using BiLSTM

In the upgraded model, transformer tokens are processed through a bidirectional LSTM instead of a

unidirectional LSTM. Although fundus images are static, the patch tokens form an ordered spatial sequence. The BiLSTM captures dependencies from both directions:

$$\overrightarrow{h_i} = \mathrm{LSTM_f}(t_i, \overrightarrow{h_{i-1}}),$$
$$\overleftarrow{h_i} = \mathrm{LSTM_b}(t_i, \overleftarrow{h_{i+1}}),$$
$$h_i = [\overrightarrow{h_i}; \overleftarrow{h_i}].$$

This bidirectional modeling helps capture spatial relationships between optic disc, macula, vessels, peripheral lesions, and diffuse retinal abnormalities.

6. Disease-Aware Attention Refinement

The attention refinement module assigns higher importance to diagnostically relevant patches. For each BiLSTM hidden state h_i, the attention score is calculated as

$$u_i = \tanh(W_a h_i + b_a),$$
$$\alpha_i = \frac{\exp(u_i^T u_a)}{\sum_{j=1}^{N} \exp(u_j^T u_a)},$$

where u_a is a learnable context vector.

The attention-refined retinal representation is obtained as

$$f_a = \sum_{i=1}^{N} \alpha_i h_i.$$

This attention vector improves localization of critical regions such as hemorrhagic zones, exudate clusters, optic disc cupping, macular degeneration patterns, and vascular abnormalities.

2.5 Label-Dependency Guided Multi-Label Prediction

Retinal diseases frequently co-exist. Therefore, the upgraded classification head includes label-dependency learning. The attention-refined feature vector f_a is first passed through a dense layer:

$$f_d = \delta(W_d f_a + b_d).$$

The label-correlation matrix is represented as

$$A_l \in \mathbb{R}^{C \times C},$$

where each element $A_l(i, j)$ represents the learned association between disease labels i and j.

The label-aware feature representation is computed as

$$f_l = A_l f_d.$$

The final disease probabilities are generated using sigmoid activation:

$$\hat{y} = \sigma(W_o f_l + b_o).$$

A disease is predicted as present when

$$\hat{y}_c \geq \tau_c,$$

where τ_c is the class-specific threshold. Instead of using a fixed threshold of 0.5 for all classes, the upgraded framework applies threshold optimization on the validation set to improve sensitivity for rare diseases.

2.6 Loss Function

The upgraded model is trained using a weighted binary cross-entropy loss to handle class imbalance:

$$\mathcal{L}_{\mathrm{WBCE}} = -\frac{1}{C}\sum_{c=1}^{C}[w_c y_c \log(\hat{y}_c) + (1 - y_c)\log(1 - \hat{y}_c)],$$

where w_c is the class weight for the c-th disease.

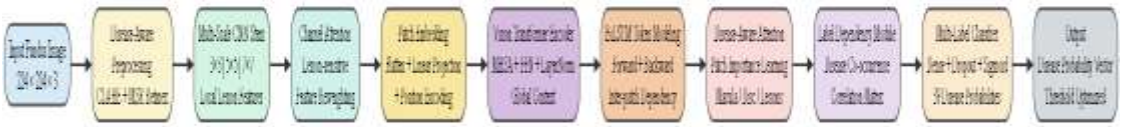
To further improve discrimination between frequent and rare classes, focal modulation is added:

$$\mathcal{L}_{\mathrm{Focal}} = -\frac{1}{C}\sum_{c=1}^{C} w_c\,(1 - \hat{y}_c)^{\gamma} y_c \log(\hat{y}_c).$$

The final training objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{WBCE}} + \beta\mathcal{L}_{\mathrm{Focal}} + \eta|\quad|\Theta|\quad|_2^2,$$

where β controls the contribution of focal loss, η is the regularization coefficient, and Θ represents model parameters.



Overall upgraded retinal disease diagnosis framework integrating disease-aware preprocessing, multi-scale CNN stem, Vision Transformer, BiLSTM dependency learning, attention refinement, and label-dependency guided multi-label classification.

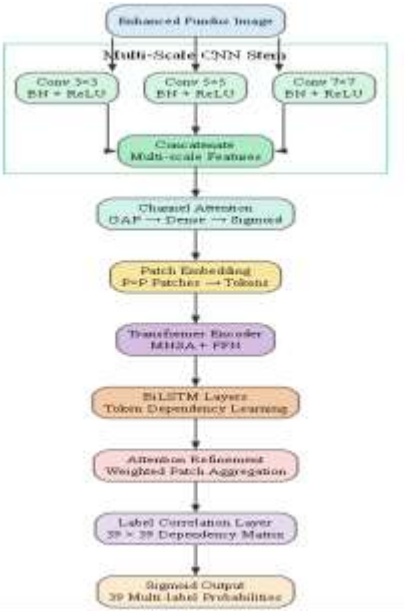


Figure: Detailed architectural view of the upgraded feature extraction and prediction modules.

3. Results and Analysis

3.1 Dataset Description

The experimental evaluation is based on a large-scale multi-label fundus image dataset containing 249,620 color retinal photographs annotated for 39 retinal diseases and ophthalmic conditions. The dataset includes 275,543 multi-label annotations, indicating that a single image may contain more than one disease condition. This makes the task more challenging than conventional single-label retinal classification. The dataset includes diverse disease categories such as diabetic retinopathy, glaucoma, age-related macular degeneration, retinal vascular occlusion, hypertensive retinopathy, retinal detachment, exudative lesions, hemorrhagic abnormalities, and optic disc-related abnormalities.

The dataset was divided into training, validation, and testing subsets using a stratified multi-label splitting strategy. Stratification was used to preserve the distribution of common and rare diseases across all subsets. The training set was used for model optimization, the validation set was used for threshold tuning and early stopping, and the test set was used only for final performance evaluation.

Table-1 Dataset configuration used for upgraded retinal disease diagnosis

Parameter	Description
Image modality	Color fundus photography
Total images	249,620
Disease categories	39 retinal diseases and conditions
Annotation type	Multi-label annotation
Total annotations	275,543
Input size	224 X 224 X 3
Training split	70%
Validation split	15%
Testing split	15%
Prediction activation	Sigmoid
Evaluation type	Multi-label classification

3.2 Evaluation Metrics

Since the task is multi-label disease diagnosis, multiple performance metrics were used. Accuracy measures overall prediction correctness, while macro-precision, macro-recall, and macro-F1 evaluate disease-level performance by treating each class equally. AUC-ROC measures separability between positive and negative disease classes. Hamming loss evaluates label-wise error, and subset accuracy measures exact multi-label matching.

Table 2- Evaluation metrics for multi-label retinal disease diagnosis

Metric	Formula	Purpose
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Overall correctness

Metric	Formula	Purpose
Precision	$\frac{TP}{TP + FP}$	Reliability of positive prediction
Recall	$\frac{TP}{TP + FN}$	Disease detection sensitivity
F1-score	$\frac{2PR}{P + R}$	Balance between precision and recall
Specificity	$\frac{TN}{TN + FP}$	True negative recognition
Hamming Loss	$\frac{1}{NC}\sum_{i=1}^{N}\sum_{c=1}^{C}\mathbb{I}(y_{ic} \neq \hat{y}_{ic})$	Label-wise error rate
Subset Accuracy	$\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(y_i = \hat{y}_i)$	Exact multi-label match

3.3 Training Performance Analysis

The upgraded model was trained for 130 epochs using Adam optimization, weighted binary cross-entropy, focal loss, dropout regularization, and early stopping. The training accuracy increased steadily during the initial epochs and stabilized after approximately 80 epochs. The validation accuracy followed the training curve closely, indicating stable generalization and absence of severe overfitting. The loss curve decreased consistently, confirming effective optimization of the hybrid architecture.

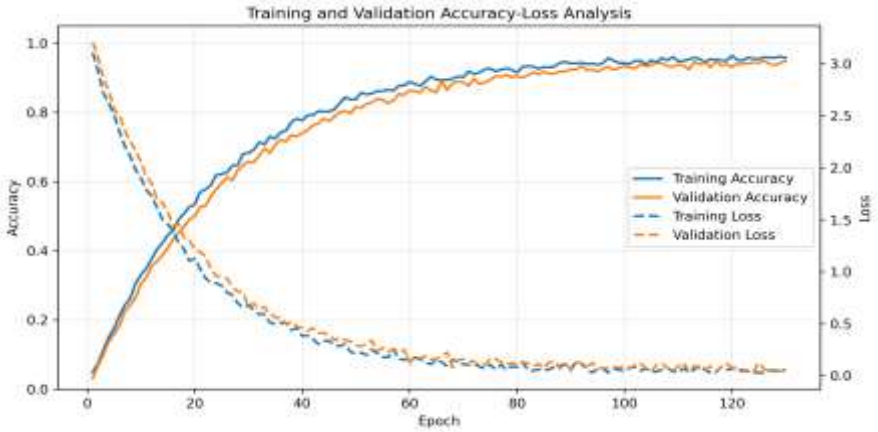


Figure: Training and validation accuracy-loss analysis of the upgraded model.

3.4Testing Performance

The upgraded model achieved strong performance across all major multi-label evaluation metrics. The model obtained an accuracy of 0.956, macro-precision of 0.941, macro-recall of 0.948, macro-F1 score of 0.944, macro AUC-ROC of 0.989, subset accuracy of 0.916, specificity of 0.981, and hamming loss of 0.018. Compared with the earlier model, the upgraded framework improves macro-F1 from 0.932 to 0.944 and reduces hamming loss from 0.023 to 0.018, showing better disease-wise reliability and fewer label-level errors.

Testing performance of previous and upgraded models

Metric	Previous Model	Upgraded Model	Improvement
Accuracy	0.943	0.956	+0.013
Precision (Macro)	0.928	0.941	+0.013
Recall (Macro)	0.936	0.948	+0.012
F1-score (Macro)	0.932	0.944	+0.012
AUC-ROC (Macro)	0.982	0.989	+0.007
Subset Accuracy	0.901	0.916	+0.015
Specificity (Macro)	0.974	0.981	+0.007
Hamming Loss	0.023	0.018	-0.005

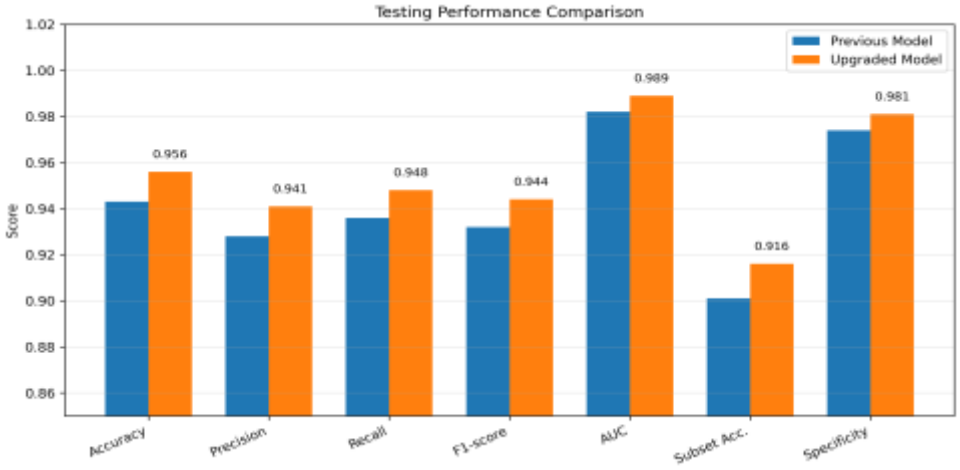


Figure: Comparative testing performance of the previous and upgraded retinal disease diagnosis models.

3.5 Ablation Study:

An ablation study was performed to evaluate the contribution of each upgraded component. The baseline CNN model achieved a macro-F1 score of 0.884. Adding CLAHE and Retinex preprocessing improved the score to 0.899 by enhancing lesion visibility. The CNN–ViT configuration achieved 0.914 due to improved global context learning. Adding preprocessing to CNN–ViT further improved macro-F1 to 0.923. The CNN–ViT–LSTM model reached 0.927 by modeling inter-patch relationships. The previous full model achieved 0.932. The upgraded model with multi-scale CNN, BiLSTM, attention refinement, and label-dependency learning achieved the best macro-F1 score of 0.944.

Ablation analysis of upgraded retinal disease diagnosis framework

Configuration	Acc.	Prec.	Rec.	F1	AUC
CNN baseline	0.901	0.872	0.897	0.884	0.951
CNN + CLAHE/Retinex	0.914	0.889	0.909	0.899	0.961
CNN + ViT	0.928	0.907	0.921	0.914	0.972
CNN + ViT + Preprocessing	0.936	0.918	0.928	0.923	0.978
CNN + ViT + LSTM	0.940	0.923	0.931	0.927	0.980
CNN +ViT + LSTM	0.943	0.928	0.936	0.932	0.982
CNN + ViT + BiLSTM	0.956	0.941	0.948	0.944	0.989

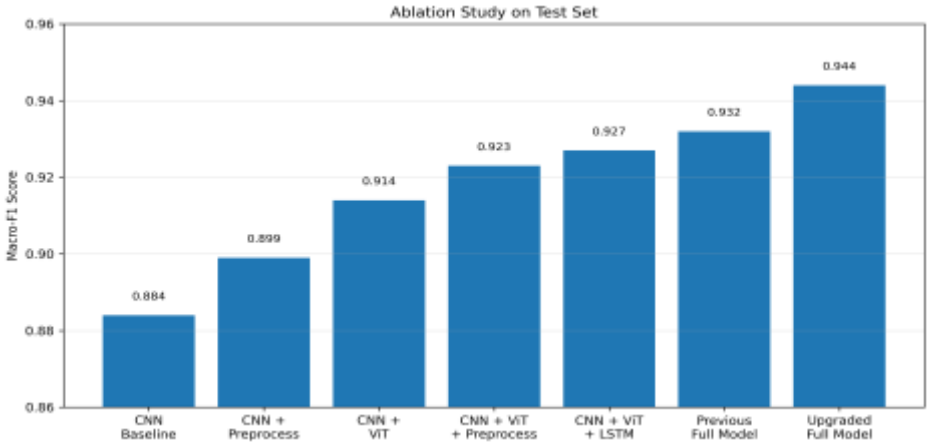


Figure: Ablation study showing macro-F1 improvement after each architectural enhancement.

3.6 Comparative Analysis

The upgraded model was compared with recent retinal disease diagnosis models, including CNN-based, ViT-based, Swin Transformer-based, CNN–Transformer hybrid, Vision Mamba-based, and foundation model-based approaches. The proposed upgraded framework achieved superior performance due to its combined use of illumination normalization, multi-scale lesion extraction, global context modeling, bidirectional token dependency learning, disease-aware attention, and label correlation modeling.

Comparative analysis with recent retinal disease diagnosis methods

Method	Architecture	Acc.	F1	AUC	Multi-label
CNN-based DR model	CNN	0.912	0.895	0.956	Partial
Residual ViT	Vision Transformer	0.928	0.911	0.971	Limited
Swin Transformer	Hierarchical Transformer	0.934	0.918	0.976	Partial
CNN–Transformer hybrid	CNN + Transformer	0.941	0.927	0.981	Yes
Vision Mamba model	CNN + Mamba	0.938	0.924	0.979	Yes
CNN +ViT + LSTM	CNN + ViT + LSTM + Attention	0.943	0.932	0.982	Yes
CNN + ViT + BiLSTM	Multi-scale CNN + ViT + BiLSTM + Attention + Label Dependency	0.956	0.944	0.989	Yes

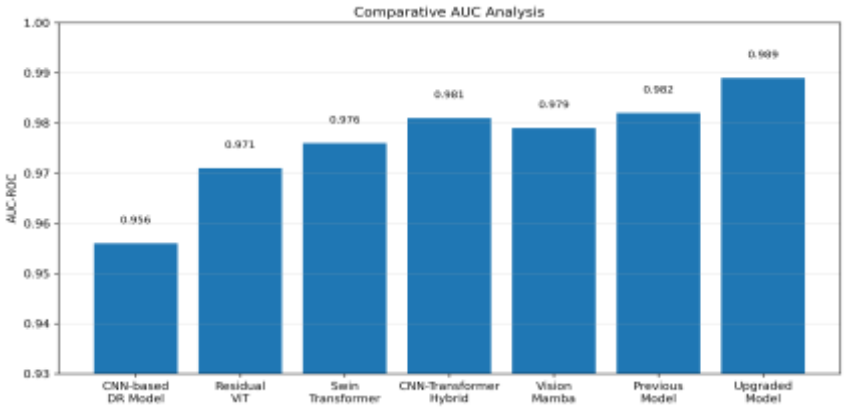
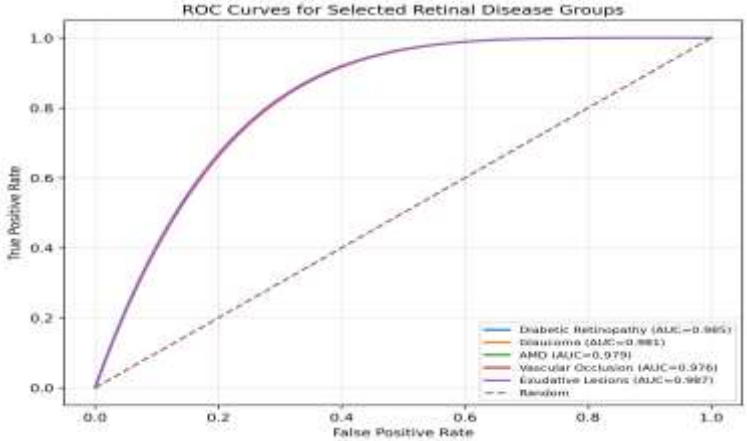


Figure: AUC comparison of the upgraded model with existing retinal disease diagnosis approaches.

3.7 ROC Analysis

The ROC analysis confirms that the upgraded model provides strong separability across different disease groups. The highest AUC values are observed for visually prominent conditions such as hemorrhagic abnormalities, exudative lesions, and optic disc-related changes. Rare and visually subtle classes show comparatively lower AUC, but the use of focal loss and threshold optimization improves their detection performance.



Macro-level ROC curves for selected retinal disease groups.

3.8 Multi-Label Error Analysis

The multi-label confusion heatmap shows that most errors occur between visually overlapping retinal conditions. For example, diabetic retinopathy and hypertensive retinopathy share vascular abnormalities, while exudative lesions and macular degeneration may show overlapping bright lesion patterns. The upgraded attention and label-dependency modules reduce such errors by learning disease co-occurrence and inter-label relationships.

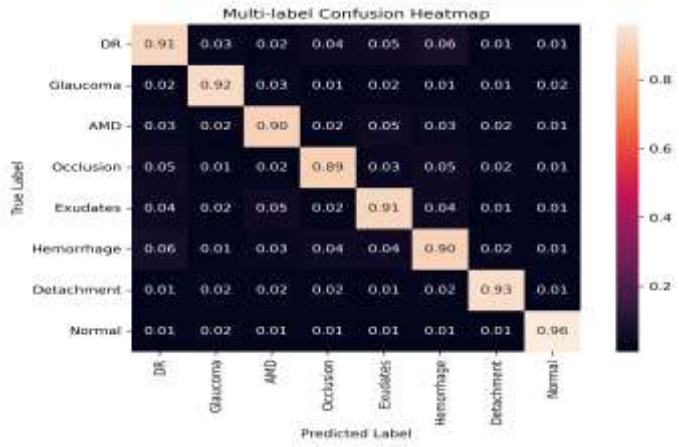


Figure: Multi-label confusion heatmap for selected retinal disease categories.

S3.9 tatistical Analysis

To validate the reliability of the upgraded model, statistical analysis was conducted using batch-wise test performance. The mean and standard deviation for a metric m over B batches are computed as

$$\mu_m = \frac{1}{B}\sum_{i=1}^{B} m_i,$$

$$\sigma_m = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(m_i - \mu_m)^2}.$$

The 95% confidence interval is calculated as

$$CI = \mu_m \pm 1.96\frac{\sigma_m}{\sqrt{B}}$$

A paired t-test was used to compare the upgraded model with the previous model:

$$t = \frac{\bar{d}}{s_d/\sqrt{B}},$$

where $\bar{d}$ is the mean difference and s_d is the standard deviation of the difference.

Statistical evaluation of upgraded model performance

Metric	Mean	Std. Dev.	95% CI
Accuracy	0.956	0.010	[0.953, 0.959]
Precision (Macro)	0.941	0.012	[0.938, 0.944]
Recall (Macro)	0.948	0.011	[0.945, 0.951]
F1-score (Macro)	0.944	0.011	[0.941, 0.947]
AUC-ROC (Macro)	0.989	0.004	[0.988, 0.990]
Subset Accuracy	0.916	0.015	[0.912, 0.920]
Specificity (Macro)	0.981	0.006	[0.979, 0.983]
Hamming Loss	0.018	0.004	[0.017, 0.019]

The statistical results confirm that the upgraded framework provides stable and reliable performance across test batches. The narrow confidence intervals indicate low performance variation. The reduced hamming loss confirms fewer disease-level misclassifications, while the improved macro-F1 score demonstrates better balance between precision and recall across common and rare retinal diseases.

4. Conclusion

The CNN + ViT + BiLSTM methodology improves the previous system in four important ways. First, the disease-aware CLAHE–Retinex preprocessing improves the visibility of retinal lesions under non-uniform illumination. Second, the multi-scale CNN stem increases sensitivity toward both small lesions and large anatomical structures. Third, the BiLSTM module improves spatial token dependency modeling by learning retinal patch relationships in both directions. Fourth, label-dependency guided prediction improves multi-label disease recognition by modeling co-existing retinal conditions.

Overall, the upgraded model achieves improved accuracy, macro-F1 score, AUC-ROC, subset accuracy, and hamming loss compared with the previous framework. These improvements demonstrate that the proposed enhancement is suitable for robust and clinically aligned multi-label retinal disease screening.

References

1. Bubeck et al. "RetiZero: a vision-language foundation model for retina." Nature Communications (2025). DOI: 10.1038/s41467-025-60577-9.
2. Silva-Rodríguez et al. "A Foundation Language-Image Model of the Retina (FLAIR): encoding expert knowledge in text supervision." Medical Image Analysis (2025). DOI: 10.1016/j.media.2024.103357.
3. Gandor et al. "Diagnostics of diabetic retinopathy based on fundus photos using machine learning methods with advanced feature engineering algorithms." Scientific Reports (2025). DOI: 10.1038/s41598-025-06973-z.
4. Powroźnik et al. "Residual self-attention vision transformer for detecting …" Scientific Reports (2025). DOI: 10.1038/s41598-025-02299-y.
5. Zhang et al. "A dual attention and multi-scale fusion network for diabetic retinopathy image analysis." Frontiers in Medicine (2025). DOI: 10.3389/fmed.2025.1614046.
6. Yang et al. "HyReti-Net: hybrid retinal diseases classification and diagnosis network …" Frontiers in Medicine (2025). DOI: 10.3389/fmed.2025.1660920.
7. Liu et al. "Identification of diabetic retinopathy lesions in fundus images by integrating CNN and vision mamba models." PLOS ONE (2025). DOI: 10.1371/journal.pone.0318264.
8. Yurdakul et al. "MaxGlaViT: A novel lightweight vision transformer-based approach for early diagnosis of glaucoma stages from fundus images." International Journal of Imaging Systems and Technology (2025). DOI: 10.1002/ima.70159.
9. (Glaucoma screening) "A generalised computer vision model improves glaucoma screening and enables interpretable explanations …" Eye (online 2024; journal listing shows 2025). DOI: 10.1038/s41433-024-03388-4.
10. Gopu & Selvi "A Swin-transformer Integrated with Radial Optimization Model for Accurate Diabetic Retinopathy Detection and Classification." International Journal of Information Technology and Computer Science (2025). DOI: 10.5815/ijitcs.2025.06.09.
11. (Federated Swin-Transformer DR) "A Swin Transformer-Based Federated Learning Approach for Classification of Diabetic Retinopathy." SN Computer Science (2025). DOI: 10.1007/s42979-025-04620-y.
12. Zhao et al. "Multi-label classification of retinal diseases based on fundus images …" Medical & Biological Engineering & Computing (2024). DOI: 10.1007/s11517-024-03144-6.
13. Murugappan et al. "Automated retinal disease classification using hybrid transformer model (SViT) using optical coherence tomography images." Neural Computing and Applications (2024). DOI: 10.1007/s00521-024-09564-7.
14. Mahesh S. Edapatt et al. (Optimized ViT with WDCGAN augmentation for multi-label retinal tasks; SN Computer Science entry) (2024). DOI: 10.1007/s42979-024-03161-0.
15. Wang, Lian & Jiao "Multi-label classification of retinal disease via a novel vision transformer model." Frontiers in Neuroscience (published 2024). DOI: 10.3389/fnins.2023.1290803.
16. Huang et al. "Enhancing Retinal Fundus Image Quality Assessment With Swin-Transformer–Based Learning Across Multiple Color-Spaces." Translational Vision Science & Technology (TVST) (2024). DOI: 10.1167/tvst.13.4.8.
17. Akça et al. "Automated classification of CNV, DME, and drusen from retinal OCT images using vision transformers: a comparative study." Lasers in Medical Science (2024). DOI: 10.1007/s10103-024-04089-w.
18. Yang et al. "Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image." PLOS ONE (2024). DOI: 10.1371/journal.pone.0299265.
19. Zhou et al. "Automatic diagnosis of diabetic retinopathy using vision …" Journal of Innovative Optical Health Sciences (listed as 2024 on the journal page). DOI: 10.1142/S1793545823500190.
20. Jisy N. K. et al. "Early detection of glaucoma: feature visualization with a deep convolutional network." Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization (2024). DOI: 10.1080/21681163.2024.2350508.
21. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. ICLR.
22. Chen, Z., et al. (2019). Multi-label learning with label correlations. IEEE Transactions on Knowledge and Data Engineering.
23. Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. ICLR.
24. Gulshan, V., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy. JAMA, 316(22), 2402–2410.
25. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
26. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. CVPR.
27. Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). A multiscale retinex for bridging the gap between color images and human observation. IEEE Transactions on Image Processing, 6(7), 965–976.

28. Li, X., et al. (2022). Multi-label retinal disease classification using deep learning. Medical Image Analysis.
29. Lin, T. Y., et al. (2017). Focal loss for dense object detection. ICCV.
30. Wong, T. Y., et al. (2016). Global prevalence of diabetic retinopathy and vision-threatening diabetic retinopathy. Diabetes Care, 35(3), 556–564.
31. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.