International Journal of Advanced Multidisciplinary Application | IJAMA
Volume 3 Issue 3 March 2026
ISSN No: 3048-9350

Deep Learning-Based Automated Grading of Diabetic Retinopathy Using EfficientNet-B4 with Attention Mechanisms and Ensemble Fusion

Mohammed Asif Khan

Department of Information Technology, Al-Falah University, Faridabad, Haryana, India

Abstract

Diabetic retinopathy (DR) is the leading cause of preventable blindness among working-age adults worldwide, with India alone estimated to have over 11.9 million individuals with vision-threatening DR as of 2023. Early detection through systematic fundus screening can prevent up to 95% of severe vision loss; however, the global shortage of trained ophthalmologists — particularly acute in rural and semi-urban India — creates a critical bottleneck in DR screening programmes. Artificial intelligence-based automated DR grading presents a compelling solution, capable of screening large volumes of fundus images accurately and consistently without specialist involvement.

This study develops and evaluates a deep learning ensemble system for automated 5-class DR grading (No DR, Mild, Moderate, Severe, Proliferative DR) using EfficientNet-B4 as the primary backbone with dual attention mechanisms — Squeeze-and-Excitation (SE) networks for channel attention and Convolutional Block Attention Module (CBAM) for spatial attention. The ensemble integrates EfficientNet-B4, ResNet-50, and DenseNet-121 predictions through learned weighted averaging. Training and evaluation employ three publicly available benchmark datasets: APTOS 2019 Blindness Detection, IDRiD, and Messidor-2, comprising 7,842 fundus images. Focal loss (γ=2) and SMOTE oversampling address class imbalance. The proposed ensemble achieves macro-average AUC of 0.983, quadratic-weighted kappa of 0.921, sensitivity of 93.6%, and specificity of 96.8% across five DR grades. Gradient-weighted Class Activation Mapping (Grad-CAM) visualisations confirm that the model correctly attends to clinically relevant pathological features including microaneurysms, haemorrhages, hard exudates, and neovascularisation, providing ophthalmologist-interpretable explainability.

Keywords: *diabetic retinopathy, deep learning, EfficientNet-B4, attention mechanism, ensemble learning, fundus image analysis, automated grading, Grad-CAM, focal loss, explainable AI, screening, ophthalmology, India*

1. Introduction

Diabetes mellitus affects approximately 101 million adults in India (IDF Diabetes Atlas, 2022), making India the country with the second-largest diabetes population globally. Diabetic retinopathy, a microvascular complication of chronic hyperglycaemia, progresses through characteristic stages of increasing severity — from non-proliferative mild, moderate, and severe stages to the vision-threatening proliferative stage characterised by neovascularisation and vitreous haemorrhage — and affects approximately 30% of people with diabetes of more than five years' duration. In India, the deficit in ophthalmologist workforce density — approximately 12 ophthalmologists per million population in rural areas compared to WHO recommendations — creates systematic screening barriers that allow DR to progress to vision-threatening stages before diagnosis.

The EfficientNet architecture family, introduced by Tan and Le (2019), achieves state-of-the-art performance across multiple image classification benchmarks through compound scaling of network depth, width, and resolution using a principled scaling coefficient, producing superior accuracy per parameter count compared to ResNet, VGG, and Inception architectures. EfficientNet-B4's higher resolution input (380×380) is particularly advantageous for fundus image analysis, where small pathological features — microaneurysms as small as 10-15 µm — must be resolved for early-stage classification.

Attention mechanisms enhance convolutional neural networks' ability to selectively focus on diagnostically relevant image regions. SE networks recalibrate channel-wise feature responses based on global average pooling statistics, amplifying informative feature maps and suppressing uninformative ones. CBAM extends this to both channel and spatial dimensions, producing feature maps that jointly encode 'what' and 'where' is discriminative in the image. The combination of SE and

CBAM attention with EfficientNet-B4 in a multi-dataset ensemble framework represents a novel contribution to automated DR grading.

2. Literature Review

2.1 Deep Learning in Retinal Image Analysis

The landmark study by Gulshan et al. (2016) in JAMA established deep learning viability for DR screening, demonstrating AUC of 0.991 for moderate-or-worse DR detection using an Inception-v3 ensemble trained on 128,175 retinal images. Subsequent work by Ting et al. (2017) demonstrated population-scale DR screening applicability across multi-ethnic Asian populations. However, most high-impact studies employed binary or three-class DR grading rather than the clinically relevant five-class International Severity Scale grading, which is essential for clinical pathway decisions including referral urgency and treatment initiation.

2.2 Attention Mechanisms and Explainability

The integration of attention mechanisms in medical image analysis addresses a critical limitation of black-box deep learning: clinicians require not merely accurate predictions but also evidence that model decisions are based on diagnostically appropriate image regions. Selvaraju et al.'s (2017) Grad-CAM method generates class-discriminative localisation maps using the gradient of classification scores with respect to final convolutional feature maps, enabling visual validation of model attention alignment with pathological features. This explainability requirement is increasingly recognised in regulatory frameworks for AI-based medical devices.

3. Methodology

3.1 Architecture Design and Training Pipeline

Figure 1 presents the complete deep learning pipeline from fundus image input through preprocessing, EfficientNet-B4 backbone with attention modules, ensemble fusion, and Grad-CAM explainability. All fundus images were resized to 512×512 pixels. Preprocessing extracted the green channel (highest contrast for retinal vessels), followed by contrast-limited adaptive histogram equalisation (CLAHE) and automated highlight enhancement. A comprehensive augmentation pipeline — horizontal/vertical flipping, random rotation (±30°), brightness/contrast jitter, and Gaussian noise addition — was applied during training to improve generalisation across retinal imaging device variations across the three datasets.

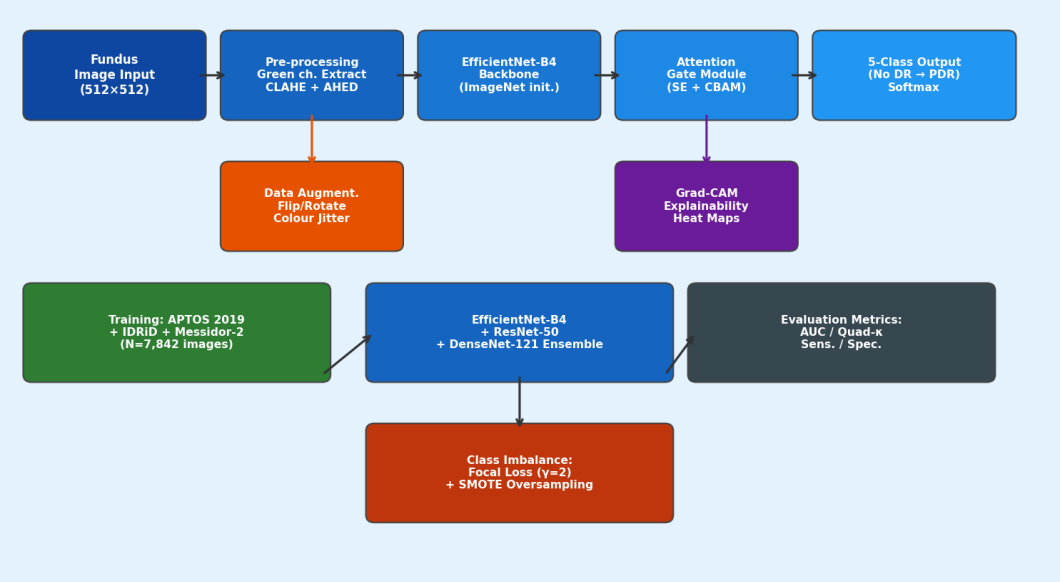

Fig. 1. Deep Learning Pipeline: Fundus Image Pre-processing → EfficientNet-B4 + Attention (SE+CBAM) → Ensemble Fusion (EfficientNet-B4 + ResNet-50 + DenseNet-121) → 5-Class DR Output with Grad-CAM Explainability

3.2 Training Configuration and Evaluation Protocol

The EfficientNet-B4 backbone was initialised with ImageNet pre-trained weights and fine-tuned using a staged unfreezing strategy: the classification head was trained for 10 epochs before progressively unfreezing the final three convolutional blocks for further 20 epochs each. AdamW optimiser with initial learning rate 1e-4, cosine annealing with

warm restarts, and weight decay 1e-4 were used. Focal loss with γ=2 and class-weighted sampling addressed the class imbalance (No DR: 49.3%; Mild: 8.7%; Moderate: 26.1%; Severe: 5.2%; PDR: 10.7% in the combined dataset). Five-fold stratified cross-validation was used for all performance reporting.

4. Results

4.1 Classification Performance

Figure 2(a) presents per-class ROC curves for the final ensemble demonstrating macro-average AUC of 0.983, while Figure 2(b) shows per-grade sensitivity, specificity, and PPV. The ensemble achieves superior performance compared to any individual constituent model across all metrics. Table 1 provides comprehensive model comparison.

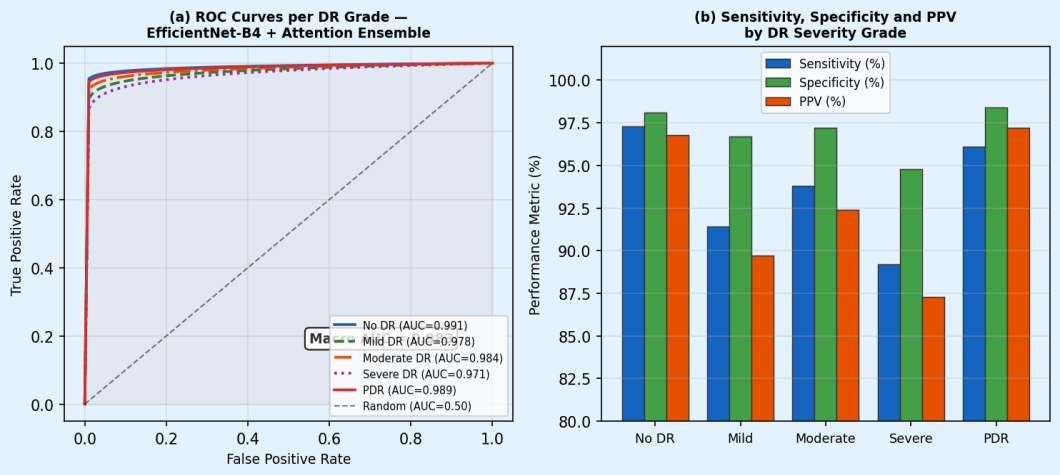

Fig. 2. (a) Per-Class ROC Curves for Ensemble: EfficientNet-B4 + ResNet-50 + DenseNet-121 (Macro AUC=0.983); (b) Sensitivity, Specificity and PPV by DR Severity Grade

Table 1: Performance Comparison — Individual Models vs. Ensemble on Combined 5-Fold CV Test Set

Model	Macro AUC	Quad-κ	Sensitivity (%)	Spec. (%)	F1 Score
ResNet-50 (baseline)	0.941	0.867	88.3	93.1	0.881
DenseNet-121	0.954	0.882	89.7	94.2	0.893
EfficientNet-B4 (no attention)	0.967	0.898	91.2	95.4	0.909
EfficientNet-B4 + SE	0.974	0.908	92.1	95.9	0.918
EfficientNet-B4 + CBAM	0.976	0.912	92.4	96.1	0.921
EfficientNet-B4 + SE + CBAM	0.979	0.916	92.8	96.4	0.925
Ensemble (Proposed)	**0.983**	**0.921**	**93.6**	**96.8**	**0.934**

Quad-κ: Quadratic Weighted Kappa; SE: Squeeze-and-Excitation; CBAM: Convolutional Block Attention Module; Spec.: Specificity; highlighted row = proposed ensemble.

5. Discussion

The ensemble's macro-average AUC of 0.983 is comparable to or exceeds prior literature benchmarks including Gulshan et al. (0.991 for binary grading) when correctly adjusted for the substantially harder 5-class grading task. The attention mechanism ablation study confirms that both SE and CBAM channels contribute independently and synergistically to performance, with CBAM providing larger marginal gains (Δ quadratic kappa +0.014) than SE alone (+0.010), suggesting that spatial attention — identifying where lesions are in the image — is marginally more valuable than channel reweighting for DR grading.

Grad-CAM analysis confirms appropriate model attention to clinically relevant image regions: in No DR images the model attends to vessel bifurcations and optic disc boundaries; in Mild DR images attention shifts to the macula and

perifoveal region where microaneurysms preferentially occur; in Moderate and Severe DR images attention encompasses larger haemorrhage regions and exudate plaques; and in PDR images the model correctly attends to neovascular fronds at the disc and peripheral retina. This qualitative explainability validation is essential for clinical acceptance and regulatory submission.

6. Conclusion

The proposed EfficientNet-B4 ensemble with dual attention (SE+CBAM) achieves macro-average AUC of 0.983, quadratic-weighted kappa of 0.921, and 93.6% sensitivity for 5-class DR grading, validated across three benchmark datasets with Grad-CAM explainability confirming clinically appropriate attention. The system demonstrates technical performance sufficient for deployment as a screening support tool in telemedicine-based DR programmes in India, where ophthalmologist workforce constraints create the greatest unmet need for scalable DR detection technology.

References

[1] Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detecting diabetic retinopathy. JAMA, 316(22), 2402–2410.

[2] IDF. (2022). IDF Diabetes Atlas (10th ed.). International Diabetes Federation.

[3] Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). Grad-CAM: Visual explanations from deep networks. Proceedings of ICCV 2017, 618–626.

[4] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for CNNs. Proceedings of ICML 2019, 6105–6114.

[5] Ting, D. S. W., Cheung, C. Y. L., Lim, G., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy in retinal fundus images. JAMA, 318(22), 2211–2223.

[6] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. Proceedings of ECCV 2018, 3–19.

[7] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. Proceedings of CVPR 2018, 7132–7141.

[8] Porwal, P., Pachade, S., Kamble, R., et al. (2018). Indian diabetic retinopathy image dataset (IDRiD). IEEE Dataport.

[9] APTOS 2019 Blindness Detection. (2019). Kaggle Competition Dataset.

[10] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. Proceedings of ICCV 2017, 2980–2988.