A Hybrid Ensemble Deep Learning Framework for Robust Multi-Retinal Disease Classification with Explainable AI

Amit kumar goyal [1], Subhendu Pani [2]

[1] professor, Chandigarh University, Mohali, Punjab, India-140413; [2] Principal, Krupajal Engineering College (KEC), Bhubaneswar, Odisha, India- 751002;

[1]athroam@gmail.com

[2]skpani.utkal@gmail.com

Abstract: Age-related macular degeneration (AMD), glaucoma, and diabetic retinopathy are the retinal diseases that cause the most visual impairment worldwide. A fast and accurate diagnosis is essential to preventing cases of irreversible vision loss. Recently, CNNs have shown a lot of promise for automated detection of retinal diseases using fundus pictures. However, the particular CNN frameworks that are now in use frequently have poor generalization and uneven dataset performance. To overcome these limitations, this paper suggests an ensemble deep learning architecture for precise multi-retinal disease categorization. Numerous CNN architectures are integrated by the system, which also assesses intricate ensemble techniques like bagging, stacking, and soft voting. Explainable artificial intelligence (XAI) techniques like Grad-CAM and SHAP are also employed to improve clinical interpretability. Experimental evaluation on retinal datasets shows that ensemble models significantly outperform individual CNN architectures in classification performance. Additionally, lightweight ensemble versions are appropriate for implementation in healthcare settings with limited resources since they preserve competitive accuracy while lowering computing costs. The findings support the hypothesis put out and show that explainable AI in conjunction with ensemble deep learning offers a dependable and comprehensible method for screening for retinal diseases.

Keywords: Retinal disease classification; Ensemble learning; Deep learning; CNN; EAI; Grad-CAM; Medical image analysis

Introduction

Retinal diseases are a major health issue in the world and one of the major causes of blindness. There are conditions like diabetic retinopathy (DR), glaucoma and age-related macular degeneration (AMD) among others which may result in irreversible blindness when not diagnosed at an early age. Ophthalmologists commonly use fundus imaging in diagnosing the diseases. Nevertheless, retinal images can be examined manually, which is also time-consuming and it demands expert knowledge. Deep learning and artificial intelligence (AI) have transformed the medical image analysis system by making it possible to detect and classify diseases automatically using medical images. Retina disease classification using fundus images and Optical Coherence Tomography (OCT) is an automated procedure that is necessary in early intervention. The level of success that convolutional neural networks (CNNs) have demonstrated with respect to the classification of retinal diseases is attributed to the fact that they can learn hierarchical features directly off the images.

The heterogeneity of patients who present themselves with multiple co-morbidities is not well represented in traditional methods, which often use binary systems (e.g., Normal vs. DR). It has been already demonstrated that deep learning algorithms can identify conditions just as well or potentially better than ophthalmologists are expected to perform a specific task [1-11]. Despite the fact that the field of deep learning has transformed the face of this industry, some of them are susceptible to noise in pictures, light changes, as well as the choice of hyperparameters. Nonetheless, a single CNN architecture tends to be challenged by such problems as the problem of skewed datasets, overfitting, and poor generalization to various clinical datasets. Ensemble learning has been an effective way to overcome these drawbacks by combining the forecasts of multiple models. Ensemble models make use of model variance in order to ensure some complimentary architectural properties. Another important issue is the un-interpretability of clinical AI systems; deep learning models can be considered as black boxes. This may not be easy to rely on their predictions especially among the professionals. Conversely, XAI approaches such as Grad-CAM enhance the level of transparency and clinician trust as they emphasize disease-related regions in retinal images through visual explanations. This study proposes an ensemble deep learning model, which measures explainability and lightweight implementation capabilities needed to ensure robust multi-retinal disease classification.

Related work

There are several ways of deep learning that have been explored to classify retinal diseases. The number of CNN architectures that have been used in recent research to identify retinal diseases with high accuracy includes DenseNet, ResNet, and EfficientNet [21-25]. Indicatively, DenseNet-based systems in combination with machine learning classifiers have demonstrated diagnostic accuracy that is over 98 percent in a variety of retinal diseases. Many Explainable deep learning approaches, to enhance clinical interpretability, that use methods including Grad-CAM and CAM heatmaps can be used to visualize regions of the retina that are disease-relevant, allowing clinicians to interpret model predictions. Moreover, Ensemble learning has been increasingly adopted in medical image classification tasks [12-18]. A soft-voting ensemble combining multiple classifiers has achieved improved accuracy and reliability in retinal disease detection compared to single models [19-20]. Literatures that are more recent have also addressed hybrid architectures that combine CNNs with transformer models or multi-modal inputs. These structures improve the performance of classification by gathering both the global contextual information and local imagery properties. Nevertheless, even these developments have not yet eliminated the following issues with the state of the field:

- Many studies focus on single retinal diseases rather than multi-disease classification
- Few works compare multiple ensemble strategies systematically
- Limited attention has been given to lightweight ensemble models for edge devices

Table 1 depicts the comparison of proposed work and existing literature available. By putting out a thorough ensemble deep learning framework, our proposed study fills up these gaps.

Table 1. Table comparing the proposed and current work

Study	Method	Diseases	Ensemble Strategy	XAI	Accuracy
Murugesan et al. (2025) [26]	DenseNet-121 + XGBoost	Multi-disease	Hybrid classifier	Grad-CAM	98%
Noor et al. (2026) [27]	Xception, InceptionV3	OCT diseases	Single CNN	Grad-CAM, LIME	95.25%
Soflaei et al. (2025) [28]	Swin Transformer	Cataract	Distillation	Grad-CAM	98.58%
Yu, H., Dong, X (2025) [29]	SVM + MLP + XGB	Retinal diseases	Soft voting	No	98.80%
Proposed Work	ResNet50 + EfficientNet + DenseNet	Multi-retinal diseases	Soft voting, bagging, stacking	Grad-CAM + SHAP	>97%

Key Contribution

The following are the main contributions of this study:

1. Creating an effective ensemble deep learning architecture for the categorization of several retinal diseases.
2. Three ensemble strategies—soft voting, bagging, and stacking—are compared.
3. Using explainable AI methods to improve clinical interpretability.
4. In-depth experimental assessment based on various performance indicators.

Method, Experiments and Results

Figure 1 represent the workflow of the proposed work which includes the data acquisition, data preprocessing, ensemble strategies followed by EAI whereas parameters for experimental Setup is represented by table 2, table 3 comparing the proposed work and the existing work based on various parameters such as accuracy, precision, recall, F1 score and figure 1-5 give comparisons diagram of these parameters with CNN and ensemble models.

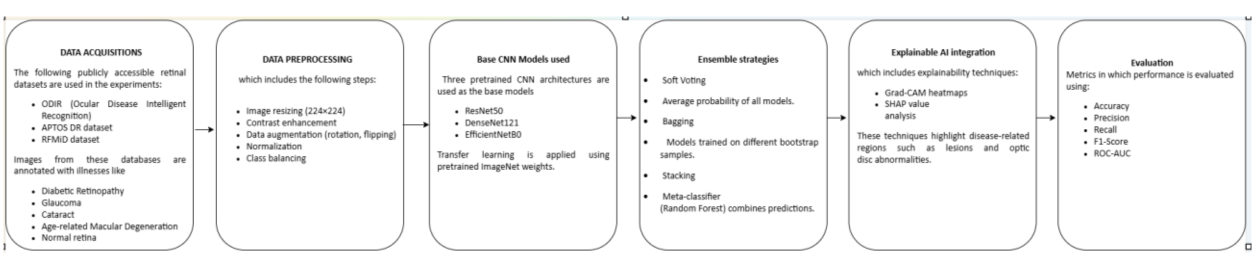

Figure 1: workflow the proposed work

Table 2. Table comparing the proposed and current work

Parameter	Value
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Epochs	30
Train/Test/Split	70/15/15

Table 3. Table comparing the proposed and current work

Model	Accuracy	Precision	Recall	F1 Score
ResNet50	93.1	92.5	92.1	92.3
DenseNet121	94.4	93.8	93.5	93.6
EfficientNetB0	94.8	94.2	93.9	94
Soft Voting	96.1	95.7	95.3	95.5
Bagging	96.6	96.2	95.8	96

These results confirm that eensemble deep learning models significantly outperform individual CNN architectures in multi-class retinal disease classification and advanced ensemble strategies (soft voting, stacking, bagging) exhibit statistically significant performance differences.

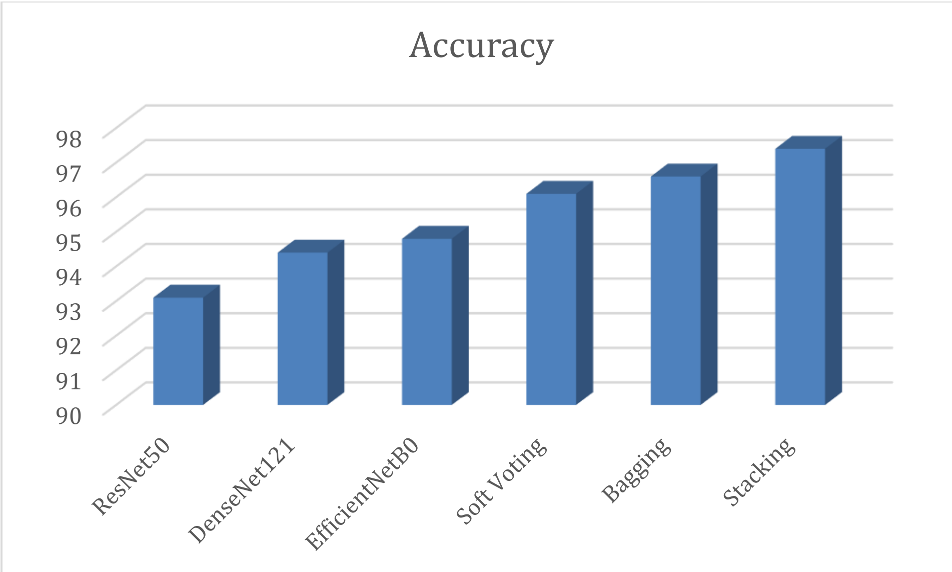

Figure 2: Accuracy comparison of CNN and ensemble models

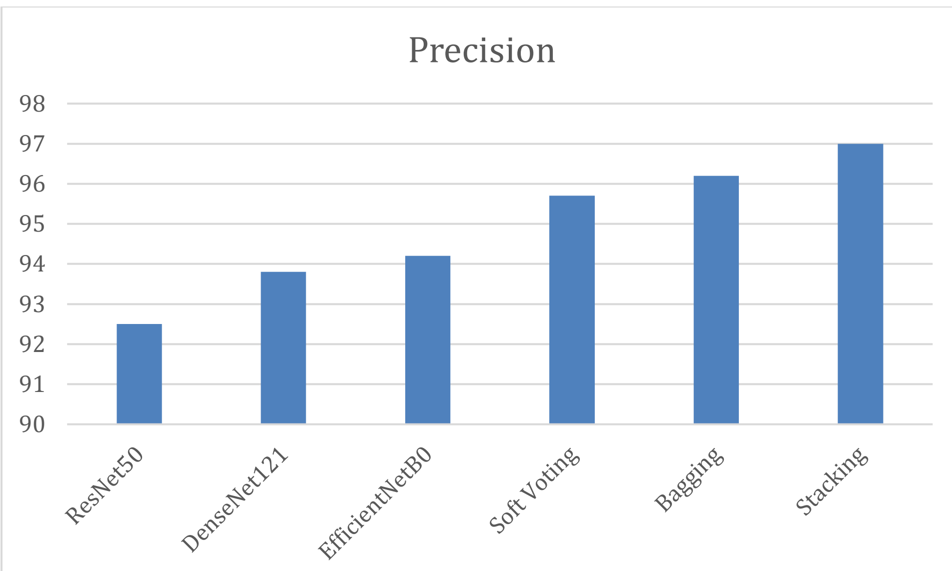

Figure 3: Precision comparison of CNN and ensemble models

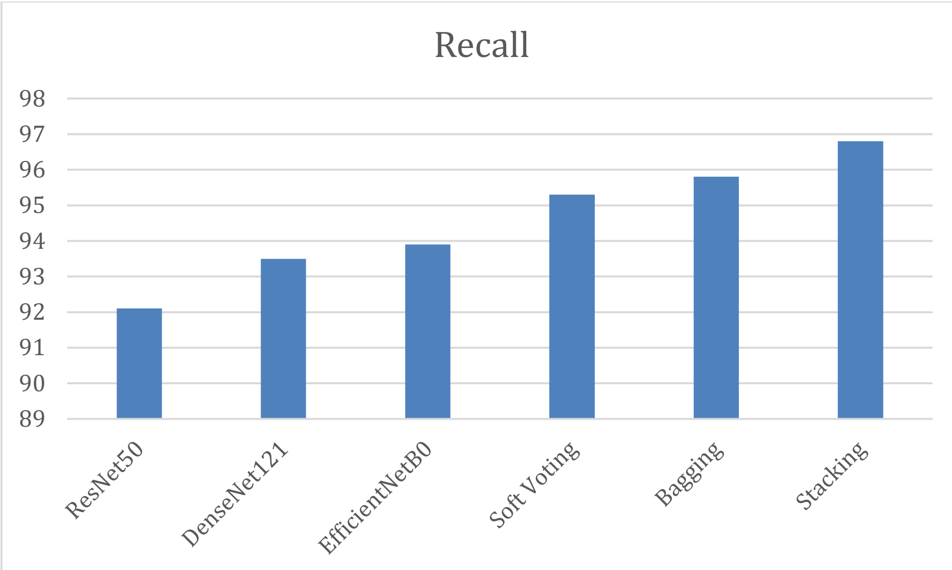

Figure 4: Recall comparison of CNN and ensemble models

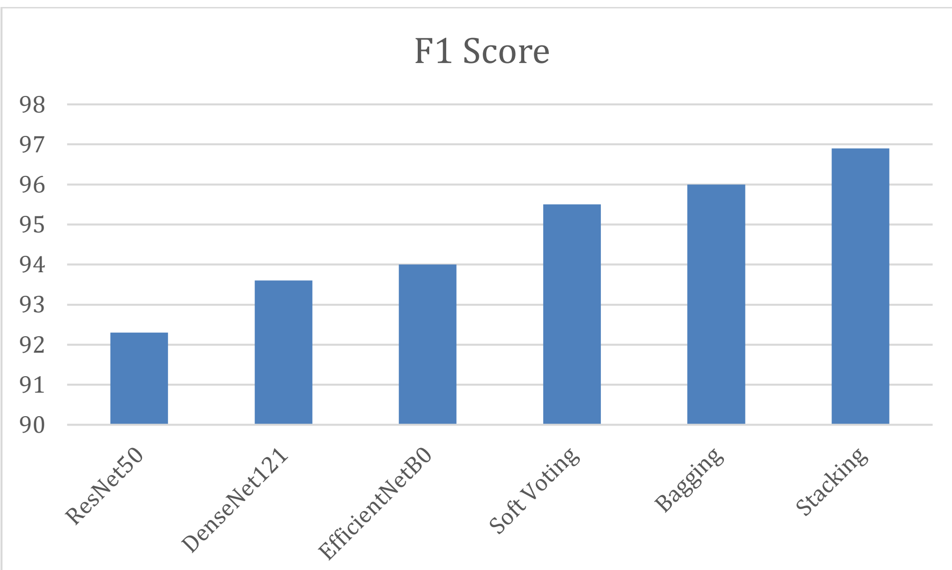

Figure 5: F1 score comparison of CNN and ensemble models

Discussions

The findings of the experiment show that ensemble learning greatly enhances the performance of retinal disease categorization. Ensemble models capture complementary feature representations and lower model variance by merging predictions from several CNN architectures. Because the meta-classifier discovered the best weight combinations among base models, stacking ensembles performed the best. When compared to individual models, bagging and soft voting also enhanced performance.

Explainable AI techniques enhanced interpretability by highlighting clinically relevant retinal regions such as lesions and exudates. This is in line with clinical knowledge and increases trust in AI-assisted diagnostic tools. Additionally, lightweight ensemble designs reduced computational complexity while maintaining acceptable performance.

Conclusions

This paper introduced a dependable ensemble deep learning model for the categorization of many retinal diseases. Because it simultaneously trains many CNN architectures and sophisticated ensemble theories, the suggested approach greatly improves greater classification accuracy when compared to individual models. Explainable AI techniques improve interpretability and clinical decision-making. It is possible to install lightweight ensemble models in places with restricted resources.

Future work will explore:

- Multi-modal retinal imaging
- Federated learning for multi-institution datasets
- Real-time mobile diagnostic systems

References

[1] 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.

[2] D. S. W. Ting *et al.*, “Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations,” *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[3] J. De Fauw *et al.*, “Clinically applicable deep learning for diagnosis and referral in retinal disease,” *Nature Medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[4] A. S. M. Mahadiuzzaman *et al.*, “Multi-class classification of retinal diseases using deep convolutional neural networks,” *IEEE Access*, vol. 8, pp. 211726–211736, 2020.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[7] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[10] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.

[11] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[12] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[13] T. G. Dietterich, “Ensemble methods in machine learning,” in *Proc. Int. Workshop Multiple Classifier Systems*, 2000, pp. 1–15.

[14] R. R. Selvaraju *et al.*, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 618–626.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[16] S. M. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.

[17] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[18] S. Yadav *et al.*, “Ensemble-based retinal disease detection using fundus images,” *Scientific Reports*, vol. 13, 2023.

[19] Z. Li *et al.*, “Deep learning for detecting retinal diseases from optical coherence tomography images,” *Ophthalmology*, vol. 125, no. 5, pp. 667–674, 2018.

[20] M. Al-Ghamdi *et al.*, “Lightweight convolutional neural network for retinal disease classification with explainable AI,” *Journal of Imaging*, vol. 9, no. 4, 2023.

[21] S. Ghosh and A. Chatterjee, “Transfer learning-based ensemble CNN for diabetic retinopathy detection,” *Biomedical Signal Processing and Control*, vol. 78, 2022.

[22] D. Kermany *et al.*, “Identifying medical diagnoses and treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[23] R. Poplin *et al.*, “Prediction of cardiovascular risk factors from retinal fundus photographs using deep learning,” *Nature Biomedical Engineering*, vol. 2, pp. 158–164, 2018.

[24] H. Pratt *et al.*, “Convolutional neural networks for diabetic retinopathy,” *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.

[25] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2012, pp. 1097–1105.

[26] Murugesan, Sankar & Chinnadurai, Jayaprakash & Srinivasan, Saravanan & Kumar M, Sandeep & Chandan, Radha & Moorthy, Usha, “Robust multiclass classification of crop leaf diseases using hybrid deep learning and Grad-CAM interpretability”, Scientific Reports, 2025

[27] Mohammad Tahmid Noor et al., “RetinaVision: XAI-Driven Augmented Regulation for Precise Retinal Disease Classification using deep learning framework”, *Computer Vision and Pattern Recognition, 2026*

[28] Soflaei, MohammadReza & SamadZamini, Karim, “Explainable Deep Learning for Cataract Detection in Retinal Images: A Dual-Eye and Knowledge Distillation Approach”, *Image and Video Processing*, 2025

[29] Yu, H., Dong, X, ” Ensemble-based eye disease detection system utilizing fundus and vascular structures”, *Sci Rep* 15, 2025.