

Deep Learning-Based Ophthalmic Disease Classification With Ensemble Modeling And Grad-CAM Visualization

Mukesh Kumar Ojha¹, Tarun Kumar², Amrita Rai³, Sindhu hak gupta⁴, Ajay Roy⁵

^{1,3,4} Department of Electronics and Communication Engineering, Greater Noida Institute of Technology, Greater Noida, India

² Department of Information Technology, International Institute of Information Technology, Pune, India

³ Galgotias college of Engineering and Technology, Greater Noida, India.

⁴ Amity School of Engineering and Technology, Amity University, Noida-201313, U.P., India

⁵ Department of Electrical Engineering, college of Engineering, University of Bisha, Saudi Arabia

Abstract: Cataract, diabetic retinopathy (DR), and glaucoma are ophthalmic diseases that cause vision impairment and blindness worldwide. For safeguarding timely treatment and preventing irreversible vision loss, early detection and diagnosis are vital. This study presents a comprehensive and interpretable deep learning framework that blends multiple pre-trained convolutional neural network architectures, ensemble learning, and Grad-CAM for accurate and transparent segmentation of retinal fundus images for multi-class classification of ophthalmic diseases. Four CNN models, i.e. ResNet-50, InceptionV3, GoogLeNet, and AlexNet were adjusted through transfer learning approaches and trained on a curated dataset comprising four categories: Cataract, Diabetic Retinopathy, Glaucoma, and Normal. Data augmentation approaches, such as rotation, translation, and scaling, were employed to enhance model robustness and generalization. Each model was evaluated in context to performance evaluation parameters such as accuracy, confusion matrices, and ROC curves. Further, an ensemble model was constructed to combine the strengths of individual networks, achieving improved diagnostic accuracy. ResNet-50 attained the highest classification accuracy of 93.80 % among the individual model, while the ensemble approach demonstrated superior overall performance with accuracy of 94.14 %. To address the limitations caused by opaque decision-making process of CNN, visualization methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) are employed to interpret class activation regions and validate model focus on clinically relevant areas to boost interpretability. The proposed framework establishes a promising foundation for clinical decision support systems and AI-assisted classification of ophthalmic diseases that not only achieves high diagnostic reliability but also bridges performance and explainability.

Keywords: Ophthalmic diseases, Fundus image classification, Deep learning, Convolutional neural networks, Ensemble model, Grad-CAM.

1. Introduction

Automated detection of ocular diseases and early diagnosis through deep learning plays a crucial role in early diagnosis and prevention of vision impairment caused due to disorders such as Cataract, Diabetic Retinopathy (DR), and Glaucoma. Conventional diagnostic methods rely on manual assessment of fundus or slit-lamp images, subjective, and dependent on clinical expertise. To address these limitations, deep learning-based computer-aided diagnostic frameworks have gained significant attention and demonstrated superior performance in medical image classification by automatically extracting complex hierarchical features. However, training CNNs from scratch requires high computational power and large annotated datasets, which are often unavailable in clinical practice. Transfer learning overcomes these challenges by adapting pretrained models such as ResNet, InceptionV3, and AlexNet to medical imaging tasks, that enable effective training with limited data. This study emphasise on implementing and comparing pretrained CNN models for automated classification of ophthalmic diseases, aiming to improve accessibility, diagnostic accuracy, and efficiency in ophthalmic screening (Kandel, I. and Castelli, M. (2020). Deep learning approaches

have been effective in automated classification of fundus images with high accuracy. CNN architectures are capable of learning complex spatial hierarchies of features directly from raw image data, thereby eliminating the requirement for handcrafted feature extraction as in case of classical machine learning methods. The lack of transparency in CNN-based predictions, however, can hinder clinical adoption. Incorporating visualization techniques enhances interpretability by highlighting key anatomical regions influencing predictions (Cai, L., Gao, J. and Zhao, D., 2020). Transfer learning has emerged as a practical solution to overcome these challenges by fine-tuning pretrained CNN models originally trained on large-scale datasets like Image Net for domain-specific medical applications. To overcome opaque decision-making process of CNNs, explainable AI techniques such as Grad-CAM have been developed to visualize class-discriminative regions, providing insights into model reasoning and highlighting disease-specific visual cues. Ensemble learning addresses this issue by amalgamating the benefits of multiple models, leading to improved robustness, stability and accuracy in classification predictions (Dubey, S.R. and Jalal, S., 2023).

In this context, the present study proposes retinal fundus images based ensemble deep learning framework with Grad-CAM interpretability for automated classification of ocular diseases from retinal fundus images. The following major considerations of current research work include:

1. This study explores the performance of multiple pretrained CNN architectures individually and develops an ensemble model.
2. Enhancement of the classification accuracy, robustness and generalization through ensemble integration.
3. Per-class metrics, ROC curves, confusion matrices, and Grad-CAM visualizations are analyzed to assess both predictive power and interpretability.
4. Improvement in model transparency by visualizing class-discriminative regions that correspond to pathological features.

The overall organization of this paper consists of six segments. The introduction and literature review of related research works are presented in Sections 1 and 2 respectively. Section 3 demonstrates the methodology and framework. Implementation of CNN classifier is discussed in Section 4. Section 5 presents the classification results. The summary and scope for future research is presented in Section 6.

2. Related Work

Various studies have demonstrated the application of CNN architectures in classification of retinal diseases. (Gulshan et al., 2016) introduced a deep learning-based method for diabetic retinopathy screening, demonstrating high sensitivity and specificity. Their work established deep CNNs as powerful tools for analyzing retinal fundus images. Since then, deep learning has emerged as the standard approach for ophthalmic image analysis. (Li et al., 2018) demonstrated glaucoma detection using deep transfer learning with high accuracy. Recent research has emphasized the improvement of transparency and clinician trust using Grad-CAM that is used to visualize pathological regions responsible for particular predictions. (Singh et al., 2022) explored DenseNet for multi-disease retinal classification, and EfficientNet for large-scale ocular datasets. Diverse CNNs models have been widely used in medical imaging for disease detection. Ensemble learning further improves predictive performance by combining complementary model strengths. Explainable AI methods like Grad-CAM provide insight into neural attention mechanisms, enabling clinical validation (Dietterich, T.G., 2000). In their research, (Pratt et al., 2016) demonstrated that CNN-based automated grading of diabetic retinopathy

significantly outperformed traditional handcrafted feature approaches in context to accuracy. (Agarwal et al.,2020)implemented transfer learning frameworkbased on ResNet-50 architecture achieving remarkable classification accuracy in glaucoma detection. (Li et al.,2021)deployed the InceptionV3 architecture for multi-class categorization of cataract grading.For providing comprehensive solutions for automated ophthalmic diagnosis across multiple retinal conditions, integrated frameworks such as DeepDR and EyeNet(Li, T., Gao, Y. and Wang, K., 2023) have introduced end-to-end deep learning architectures. (Tan et al.,2022)appliedpretrained deep learning architectures to classify the severity levels of diabetic retinopathy that leadsto substantial improvement in accuracy in contrast to traditional methods. (Muhammad et al.,2020) focused on optic disc analysis and segmentation through implementing deep CNN-based approach and achieving high sensitivity, specificity, and overall accuracy in distinguishing glaucomatous eyes from healthy ones. Research on automatic and reliable categorization of grade cataract severity is performed by (Li et al.,2022) by leveraging deep transfer learning approaches.

Research Gap and Novelty

Despite remarkable advancements in deep learning based ophthalmic image classification, following research gapsstill exists.

1. Thelittle transparency into their decision-making process owing to limited interpretability of deep learning models remains a significant barrier to their clinical implementation.
- 2.Theintegration ofvisualization methods of into comprehensive diagnostic frameworks for ophthalmic diseases is still underwhelmed.
3. Lack of standardized comparative analyses as related studies focuses on overall accuracy while neglecting per-class performance metrics such as sensitivity, specificity, and AUC, which are criticalclinical reliability.

To address these challenges, a unified and interpretable ensemble deep learning framework is introduced in current work.An ensemble model is then constructed for enhancing classification accuracy, generalization, and robustness by amalgamating the distinctive feature-learning capabilities of these architectures. Grad-CAM-based visualization is also incorporated to generate class-discriminative heatmaps, revealing the specific retinal regions. By combining quantitative evaluation metrics with visual interpretability, this research significantly demonstrates a reproducible, explainable, and high-performing framework for automated classification of ocular conditions.

3. Methodology

A deep learning framework based on four pre-trained CNN models ResNet50, InceptionV3, GoogleNet, and AlexNet are applied to automatically categorize retinal images consisting of Cataract, Diabetic Retinopathy, Glaucoma, and Normalclass. The presented approach integrates transfer learning, ensemble integration, and Grad-CAM-based interpretability that ensures accurate and explainable disease detection.

3.1 Experimental Workflow

1. Train and evaluate individual pre-trained CNN model implemented on the ophthalmic disease dataset.
2. Observe confusion matrices and ROC curves for each model for performance analysis.
3. Apply Grad-CAM visualizations to visualize decision regions to interpret model focus regions.
4. Develop and test an ensemble model integrating multiple CNNs

5. Compare ensemble performance with individual CNNs both quantitatively and visually.

3.2 Description of Pretrained CNN Architectures

The proposed classification framework utilizes pre-trained CNN architectures that were fine-tuned for multi-class classification of ocular diseases. Each network comprised of successive convolutional and pooling layers responsible for hierarchical feature extraction, followed by fully connected layers performing classification. To adapt the models for the present work, their original output layers were replaced with a customized classification module comprising of a dense layer, a softmax layer, and an output layer corresponding to four ophthalmic disease categories.

ResNet-50 is a 50-layer deep residual architecture designed to overcome gradient degradation in very deep networks. By utilizing skip and residual connections, it preserves essential feature information across layers and enables stable gradient propagation during training. It supports efficient convergence and reliable learning of identity mappings, making the classification model well optimized for capturing fine-level retinal and anterior-segment visual patterns. GoogLeNet is a 22-layer deep CNN that employs Inception modules to perform parallel convolution and pooling operations that enables effective multi-scale feature learning with fewer parameters. Its compact design and strong representational capability make GoogLeNet well suited for ophthalmic image analysis in computationally constrained environments. InceptionV3 builds upon the original Inception design by incorporating factorized convolutions, batch normalization and dimensionality reduction to enhancing learning efficiency while diminishing computational cost. Its parallel convolution paths with different kernel sizes enable effective capture of multi-scale spatial features and fine structural variations in images. Due to its balanced trade-off between accuracy and efficiency, InceptionV3 is optimal for retinal image analysis and ophthalmic disease classification. AlexNet is convolutional network that popularized deep learning for large-scale image classification having five convolutional layers followed by three fully connected layers, using ReLU activations, max pooling, and dropout to improve the efficiency and mitigate overfitting. Although AlexNet possess relatively shallow depth, it remains effective for transfer learning, particularly with small to medium-sized medical image datasets (Krizhevsky, A., et.al, 2017), (He, K., Zhang, X., et.al. 2016).

3.3 Ensemble Model Design

An ensemble model was designed to leverage diverse feature representations for enhancing diagnostic consistency and robustness by combining the outputs of all pre-trained networks. Soft voting ensemble averages the class softmax probabilities of all base models. Majority voting was used to determine the final class prediction from the predictions from individual CNNs. The performance of ensemble is analyzed through confusion matrices, ROC curves, and Grad-CAM visualizations to compare against individual models.

3.4 Grad-CAM Visualization for Interpretability

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the best-performing network that visualizes image regions that contribute most to the predicted class by overlaying heatmaps on the original fundus images. These visualizations align with clinical characteristics, indicating that the model captures meaningful disease relevant spatial information.

4. Classifier CNN Implementation and Architecture

Each network of the proposed transfer learning based deep CNN framework was fine-tuned on the target dataset to optimize feature representation and classification accuracy. Four pretrained architectures i.e. InceptionV3, ResNet50, GoogLeNet, and AlexNet were employed to evaluate and compare performance in recognition of four ocular disease conditions.

4.1 Dataset Description

The dataset utilized in this study consists of 1000 retinal fundus images for each class of Normal, Diabetic Retinopathy, Cataract and Glaucoma that are collected from various sources like IDRiD, Ocular recognition, etc. Each class consists of sufficient number of images systematically arranged into class-specific folders and Imported into MATLAB to enable effective training, validation, and testing of classifier models. The complete dataset was categorized into training (70%), validation (15%), and testing (15%) segments maintaining a balanced distribution of all disease categories. Prior to training, all images were rescaled to fit the input dimensions of the respective CNN architectures and normalized to reduce intensity variations.

4.2 Network Initialization and Selection

All pretrained CNN architectures were initialized using weights learned from the fundus image database, enabling the transfer of generic visual features to the medical imaging domain. InceptionV3 utilizes multi-scale Inception modules to capture hierarchical spatial features efficiently. AlexNet with shallow 8-layer architecture, served as a baseline model for comparison. ResNet50 mitigates vanishing gradient issues through skip connections that allow identity mapping. GoogLeNet adopts a similar inception-based design, providing tradeoff between predictive accuracy and computational cost.

4.3 Data Preprocessing and Augmentation

The preprocessing pipeline improved the dataset's diversity, ensuring reliable and stable model performance across all four disease categories involves resizing images to conform to the input specifications of each CNN model. The data augmentation approaches consisting rotation, flipping, and intensity normalization were applied to improve generalization and mitigate overfitting. These augmentation strategies introduced controlled variations in orientation, position, and scale, allowing the network to learn invariant and generalized features. Following augmentation, Images were rescaled to meet the input requisites of the respective CNN architectures.

4.4 Training Configuration

Training was carried out using the Adam optimization algorithm, configured with an initial learning rate of 1×10^{-4} and a mini-batch size of 16, over a maximum of 8 epochs, with validation executed after every 30 iterations. This training configuration ensured efficient learning, balanced performance, and stable optimization across all evaluated CNN models. The validation subset was used to monitor convergence and prevent overfitting by dynamically adjusting learning rates as per requirement. Data augmentation and early stopping regularization strategies were applied to enhance generalization. Depending on computational resources, the number of epochs can be extended for deeper convergence and improved performance. The fine-tuning process allowed the pretrained feature extraction layers to retain generic visual representations while adapting the final layers to problem-specific characteristics.

4.5 Evaluation Metrics

The model evaluation was performed using independent test dataset followed by training and performance was quantitatively assessed through metrics such as Overall accuracy and Confusion matrices. Among the tested models, the architecture achieving the highest classification accuracy and stable performance across different ocular disease classes was identified as the optimal CNN configuration for eye disease recognition.

5. Experimental Results and Discussion

The present work explores and assesses the performance of four pre-trained CNN architectures for multi-class ocular disease classification using retinal fundus images. The dataset comprises four diagnostic classes.

5.1 Model Performance Evaluation

Transfer learning techniques were applied to adapt the high-level features learned from the ImageNet dataset to domain-specific retinal characteristics. Due to differences in final layer naming conventions, conditional logic was implemented to modify the classification head dynamically. A comprehensive evaluation was conducted on a held-out test set is reported in Table 1. The present training convergence was observed within eight epochs, further enhanced like 20–30 epochs can yield further improvements with larger number of datasets and higher computational resources.

Table1 Performance metrics summary

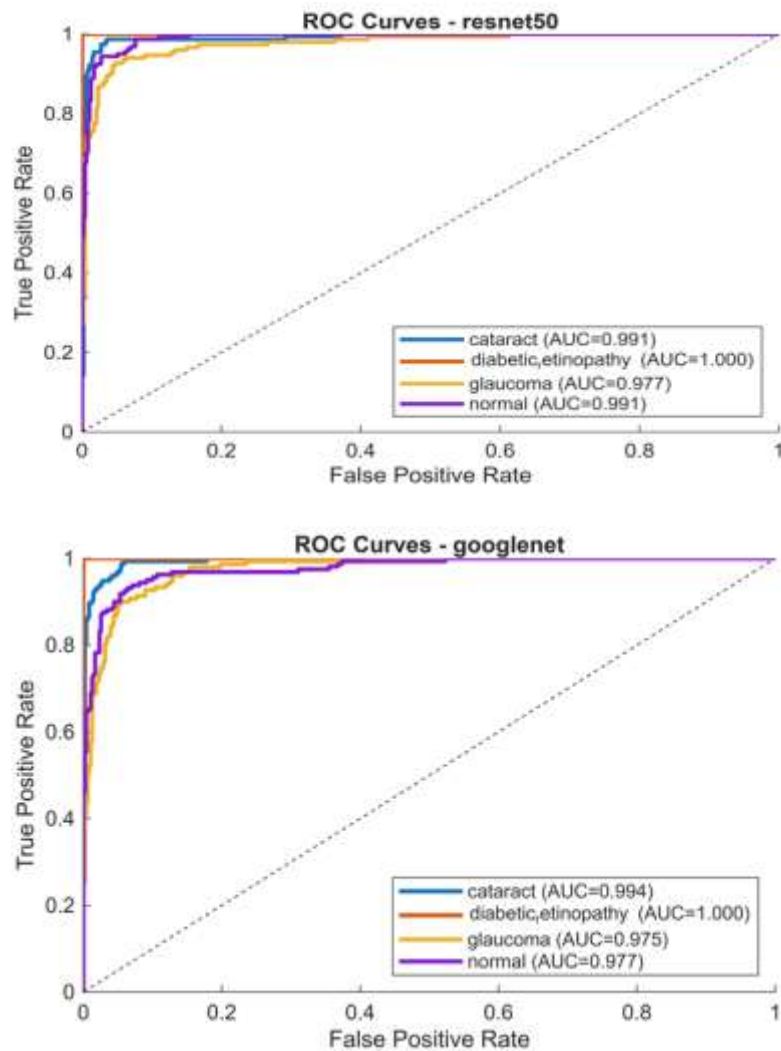
Class	Precision	Recall	F1-score	AUC
Cataract	0.948717	0.954838	0.951768	0.991040
Diabetic Retinopathy	1.0000	1.0000	1.0000	1.0000
Glaucoma	0.914285	0.847682	0.879725	0.977069
Normal	0.888888	0.944099	0.915662	0.991436

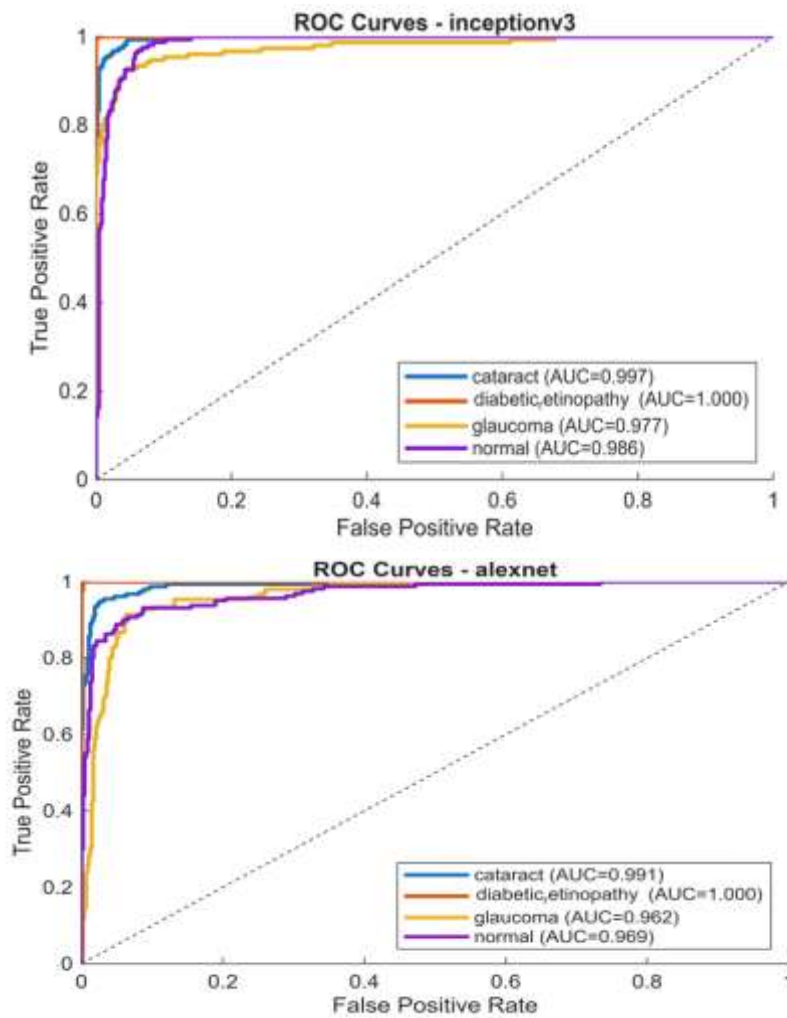
The ResNet-50 model achieved the highest classification accuracy among all implemented models, attributed to its residual learning framework, which alleviates vanishing gradient issues and supports deeper feature extraction. Although AlexNet is efficient in context to computations, achieves lower precision and recall. GoogLeNet exhibited efficient computation and enhanced sensitivity to fine-grained retinal details, though minor confusion was observed between Cataract and Normal classes due to similar textural patterns. InceptionV3 demonstrates effective multi-scale feature representation and comparable accuracy.

5.2 ROC Curve and AUC Analysis

For visual assessment of specificity and sensitivity, ROC curves were plotted and presented in Figure 1 (a)-(d). The ROC curves for all models demonstrated high separability across disease categories. The ResNet50 model recorded the maximum mean AUC, indicating superior sensitivity and specificity. The outcome substantiates the model's efficiency in mitigating misclassifications.

Figure 1. ROC Curve for every model





5.3 Quantitative Performance

Table 2 highlights the accuracies achieved by each pretrained CNN classifier models implemented for multi-class classification of ocular disease.

Table 2 Test Accuracy of Pretrained CNN Models

Classifier	Accuracy (%)
ResNet-50	93.82 %
InceptionV3	93.66 %
Googlenet	91.92 %
AlexNet	91.13%
Ensemble of ResNet-50, InceptionV3, googlenet, and AlexNet	94.14 %

Among the models, ResNet-50 demonstrates its capability to learn diverse feature hierarchies with computational efficiency by possessing maximum efficiency.

InceptionV3 and GoogLeNet performed comparably, demonstrating strong discriminative power, while AlexNet, though simpler, still achieved satisfactory classification performance.

5.4 Confusion Matrix Analysis

The confusion matrix for the best achieving model i.e. ResNet50 for which the most balanced classification performance is achieved with higher true positive rates across all four categories. Figure 2 indicates that misclassifications primarily occurred between Cataract and normal categories, suggesting overlapping visual patterns in some images. Normal and Glaucoma classes exhibited distinct feature boundaries with high precision and recall. Misclassifications primarily occurred between Cataract and Normal classes, likely due to mild lens opacities that are visually similar to healthy eyes. Glaucoma and Diabetic Retinopathy were classified with high precision, showing the effectiveness of deep CNNs in capturing disease-specific retinal features.

Figure 2. Confusion matrix for the best model

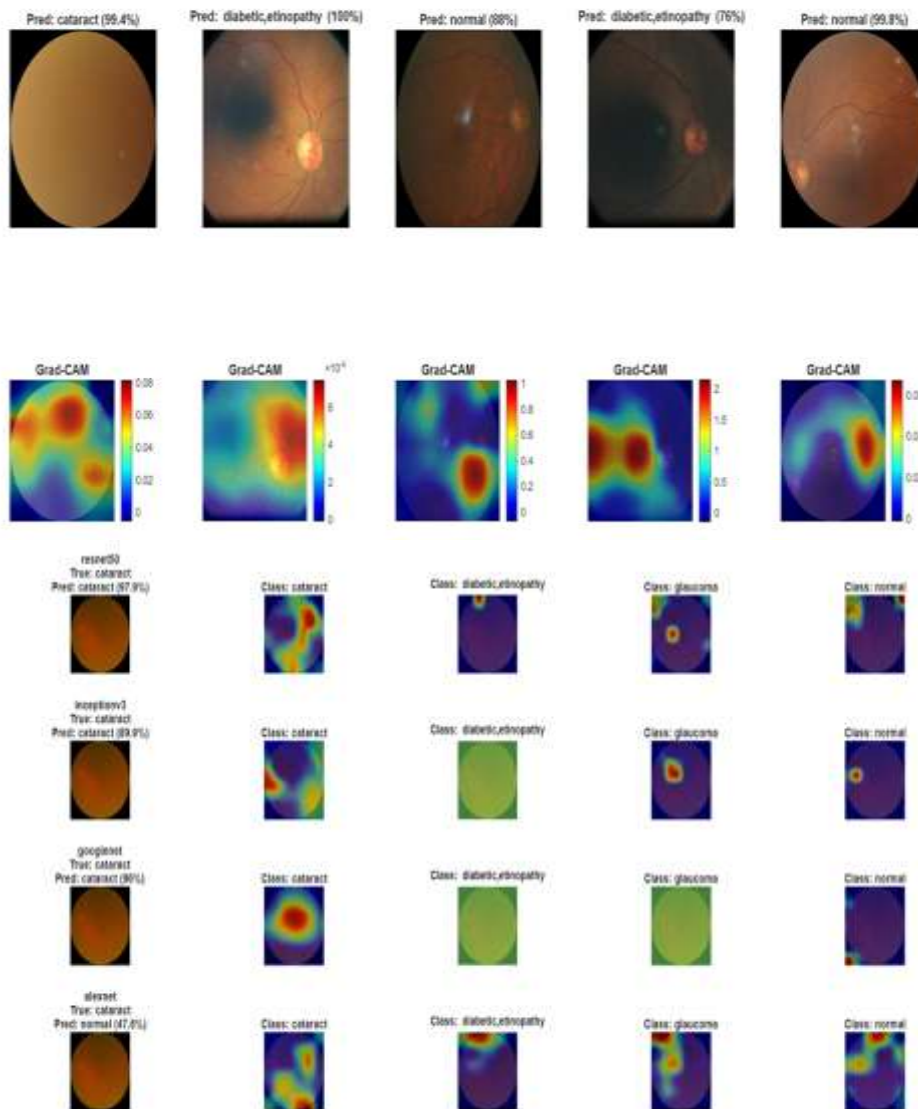
Confusion Matrix - resnet50

True Class	cataract	148		4	3
	diabetic,etinopathy		164		
	glaucoma	7		128	16
	normal	1		8	152
	Predicted Class	cataract	diabetic,etinopathy	glaucoma	normal

5.5 Grad-CAM Visualization and Model Interpretability

To boost the transparency and interpretability, Grad-CAM was applied by generating visual explanations of the discriminative regions that majorly influenced CNN predictions. The generated heatmap efficiently highlighted clinically relevant areas such as the optic disc, macula, hemorrhagic lesions, and lens opacity zones. For correctly classified samples, these activations aligned with disease-specific features such as optic disc deformation in glaucoma, and opacity regions in cataract cases. The ambiguous or overlapping patterns, offering insights into potential model confusion are revealed by misclassified images. Figure 3 presents the visualization highlights as lens opacity, hemorrhagic lesions, optic disc and cup areas, no pathological activation for Cataract, Diabetic Retinopathy, Glaucoma, and Normal respectively. This confirms that the models focus on clinically relevant retinal regions.

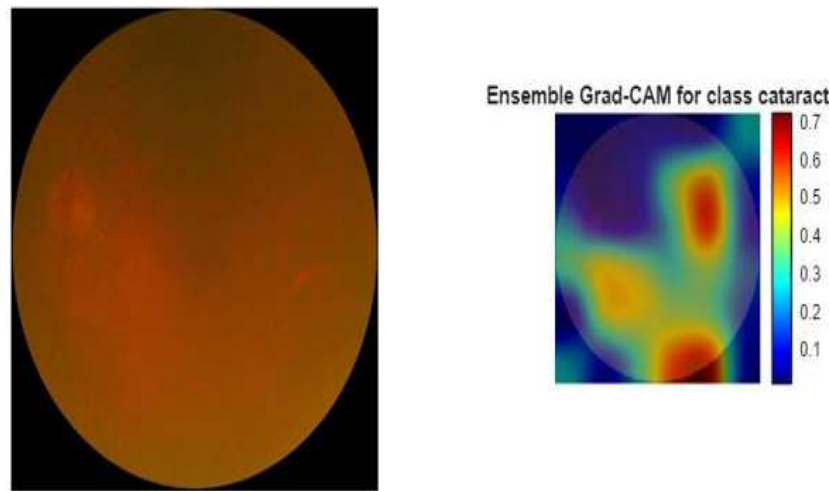
Figure 3. Grad-CAM visualization to interpret predictions



5.6 Ensemble Model Evaluation

By aggregating softmax probabilities through weighted averaging, the ensemble effectively reduced inter-class misclassifications, particularly in borderline cases such as glaucoma versus normal. Therefore, proposed ensemble model outperformed all individual CNN architectures, achieving the highest overall classification accuracy and demonstrating superior generalization and robustness. The ROC-AUC analysis further justified its enhanced discriminative capability across all classes of ocular diseases. Grad-CAM analysis verified that the ensemble predictions focus more precisely on object-relevant regions than individual CNNs. Grad-CAM analysis demonstrated that the ensemble reliably localized attention to pertinent retinal regions lens opacity in cataract, hemorrhagic lesions in diabetic retinopathy, and optic disc cup areas in glaucoma unlike some single models that occasionally emphasized background regions. The model's diagnostic coherence and reliability are confirmed by these interpretability outcomes.

Figure 4. Ensemble Grad-CAM visualization

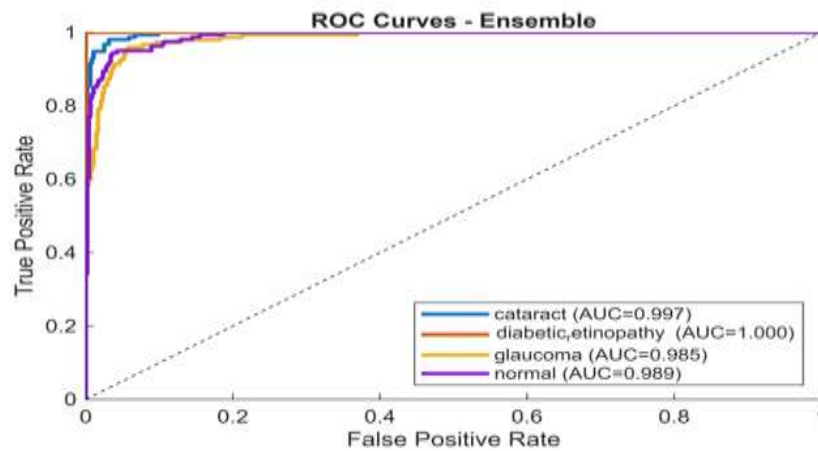


Figures 3 and 4 illustrates that Grad-CAM was utilized to visualize and interpret their predictions on both the best-performing individual model and the ensemble model respectively.

Figure 5. Training of ensemble model



While individual CNNs exhibited strong performance, they sometimes relied on irrelevant or background regions. The ensemble mitigated these tendencies, providing smoother and more confident probability distributions. The ensemble model achieved a consistent improvement in accuracy and AUC across all experiments. The networks were trained as highlighted in Figure 5. The ensemble outperformed individual models in overall accuracy, while maintaining high per-class AUC scores, demonstrating the benefits of multi-model integration. ROC curves of ensemble model as shown in Figure 6 demonstrate superior ensemble performance and higher balanced sensitivity across classes especially at lower false positive rates.

Figure 6. ROC curve of Ensemble

5.7 Comparative Discussion

Automated detection of ocular diseases is critical to prevent vision impairment. Four pretrained CNN architectures were fine-tuned through transfer learning for automated ophthalmic disease classification. Evaluation on a custom dataset revealed that ResNet-50 demonstrated superior accuracy of 93.82% followed by InceptionV3 (93.66%), GoogLeNet (91.92%), and AlexNet (91.13%).

Model comprehensive evaluation was performed using a separate test dataset, and classification performance metrics were computed to ensure reliable performance and generalization across disease categories. The best-performing model was further analyzed using Grad-CAM visualization. ResNet-50 demonstrated strong generalization across unseen test samples. InceptionV3 and GoogLeNet exhibited competitive results, effectively capturing multi-scale retinal features, while AlexNet, though shallower, achieved satisfactory performance with faster convergence. Table 3 summarizes recent research works focusing on deep learning-based eye disease classification.

Table 3. Comparative Summary of Related Research on Eye Disease Classification

Ref.	Application	Dataset Used	Model / Technique	Accuracy (%)	Remarks
Gulshan, V., Peng, L., Coram, M., et al. (2016)	Diabetic Retinopathy Detection	KaggleEyePACS	Custom Deep CNN	90.3	Achieved ophthalmologist-level diagnostic accuracy sensitivity and specificity
Singh, P., Bhatia, M. and Kumar, A.	Multi-Disease Retinal Classification	APTOS 2019	DenseNet-121	94.2	Multi-class CNN for DR and Glaucoma

(2022)					
Agarwal, A., Singh, R. and Gupta, S. (2020)	Glaucoma Classification	RIM-ONE Dataset	ResNet-50	90.8	Pretrained residual network enhanced accuracy. Improved performance using pre-trained residual networks
Li, X., Liu, R. and Chen, J. (2021)	Cataract Grading	Private Clinical Dataset	InceptionV3	92.5	Reliable cataract stage classification, Achieved reliable cataract stage classification
Zeng, Y., Xu, W., & Zhang, H. (2023)	Cataract Detection	Ocular Image Dataset	VGG-16 (Transfer Learning)	91.6	Transfer learning improved feature generalization. Demonstrated efficient feature reuse through transfer learning
Proposed Work	Ocular disease classification (Cataract, DR, Glaucoma, Normal)	Custom Mixed Dataset	ResNet-50, InceptionV3, GoogLeNet, AlexNet	94.14 % (Ensemble of ResNet-50, InceptionV3, GoogLeNet, AlexNet)	Best accuracy with ResNet-50 among pretrained CNNs; robust generalization, Ensemble model with

					Grad-CAM

6. Conclusion

This presented research work highlights explainable and comprehensive intelligent diagnostic framework for the automated classification of ophthalmic diseases i.e. Cataract, Diabetic Retinopathy, and Glaucoma using retinal fundus images. Four state-of-the-art CNN architectures i.e. ResNet50, InceptionV3, GoogLeNet, and AlexNet were systematically evaluated. ResNet50 demonstrated the highest classification accuracy (93.82%), among the individual model highlighting its superior feature representation and residual learning efficiency. For the upgradation of model's robustness and generalization, an ensemble model was implemented that achieved remarkable competence with an overall accuracy of 94.14%. The proposed framework emphasizes interpretability and clinical trust and enhances user confidence in AI-assisted decision-making through the amalgamation of Grad-CAM. The proposed ensemble-based deep learning system represents a scalable, reliable, cost effective and interpretable solution for automated ophthalmic disease screening systems.

Future work will incorporate extension of the dataset with multimodal imaging modalities such as slit-lamp photographs, integrating attention-guided and transformer-based architectures. Moreover, real-time implementation on embedded or cloud-based platforms will be explored to facilitate early diagnosis, remote screening, and improved accessibility in clinical and community healthcare settings.

Acknowledgment: The authors are thankful to the deanship of Graduate studies and Scientific Research at the university of Bisha for supporting this work through the fast track Research supporting Program. The authors also thankful to the management of Greater Noida Institute of technology for supporting this work.

References

1. Kandel, I. and Castelli, M. (2020) Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review. Applied Sciences, 10(6), 2021. <https://doi.org/10.3390/app10062021>
2. Cai, L., Gao, J. and Zhao, D. (2020) A review of the application of deep learning in medical image classification and segmentation. Ann Transl Med, 8(11), 713. doi: 10.21037/atm.2020.02.44
3. Dubey, S.R. and Jalal, S. (2023) A Comprehensive Survey on Deep Learning for Ocular Disease Classification. Artificial Intelligence in Medicine, 141, 102693.
4. Cho, H., Hwang, Y.H., Chung, J.K., Lee, K.B., Park, J.S., Kim, H.G. and Jeong, J.H. (2021) Deep Learning Ensemble Method for Classifying Glaucoma Stages Using Fundus Photographs and Convolutional Neural Networks. Curr Eye Res, 46(10), 1516–1524. doi: 10.1080/02713683.2021.1900268
5. Gulshan, V., Peng, L., Coram, M., et al. (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410.

- <https://doi.org/10.1001/jama.2016.17216>
6. Li, Z., He, Y., Keel, S., Meng, W., Chang, R.T. and He, M. (2018) Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*, 125(8), 1199–1206.
<https://doi.org/10.1016/j.ophtha.2018.01.023>
 7. Singh, P., Bhatia, M. and Kumar, A. (2022) Multi-class retinal disease classification using DenseNet and data augmentation. *Expert Systems with Applications*, 200, 117050.
 8. Dietterich, T.G. (2000) Ensemble methods in machine learning. *Multiple Classifier Systems*, 1–15.
 9. Krogh, A. and Vedelsby, J. (1995) Neural network ensembles, cross validation, and active learning. *NIPS*, 231–238.
 10. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.
 11. Pratt, H., Coenen, F., Broadbent, D.M., Harding, S.P. and Zheng, Y. (2016) Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90, 200–205.
 12. Agarwal, A., Singh, R. and Gupta, S. (2020) Glaucoma detection using transfer learning based deep CNN. *Biomedical Signal Processing and Control*, 62, 102150.
 13. Li, X., Liu, R. and Chen, J. (2021) Automatic cataract detection using InceptionV3 transfer learning. *Computer Methods and Programs in Biomedicine*, 204, 106050.
 14. Zhang, K., Liu, Y., Chen, J. and Chen, L. (2018) DeepDR: Detection of Diabetic Retinopathy Using Deep Convolutional Neural Networks. *IEEE Transactions on Medical Imaging*, 37(11), 2369–2380.
 15. Li, T., Gao, Y. and Wang, K. (2023) Multi-Disease Classification of Retinal Images Based on Deep Learning. *Biomedical Signal Processing and Control*, 79, 104098.
 16. Tan, M., Al-Bander, A. and Pratt, S. (2022) Transfer Learning for Diabetic Retinopathy Classification Using Pretrained Deep Convolutional Neural Networks. *Computers in Biology and Medicine*, 141, 105080.
 17. Muhammad, H., Fraz, S. and Barman, A. (2020) Automated Glaucoma Detection Using Deep Convolutional Neural Networks. *Expert Systems with Applications*, 158, 113486.
 18. Li, X., Hu, Y. and Zhao, W. (2022) Automatic Cataract Classification Using Deep Convolutional Neural Networks. *BMC Ophthalmology*, 22(1), 1–12.
 19. He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
 20. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826.
 21. Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6), 84–90.
 22. Ajitha, S. and Rani, S.S. (2019) Automated cataract detection using transfer learning with deep convolutional neural networks. *International Journal of Engineering and Advanced Technology*, 8(6S3), 540–544.
 23. Zeng, Y., Xu, W. and Zhang, H. (2023) Lightweight EfficientNet-based model for multi-disease eye image classification. *Biomedical Signal Processing and Control*, 85, 104950.