

A Robust and Explainable Deep CNN Framework for Multi-Disease Detection in Retinal Fundus Images

Shrikant Ashok Harage¹, Sandhya Sharma¹, Amit Jaykumar Chinchawade², Mukesh Kumar Gupta³

¹Department of Electronics and Communication Engineering, Suresh Gyan Vihar University, Jaipur, India

²Department Artificial Intelligence and Data Science Engineering,

Sharad Institute of Technology College of Engineering, Yadrav (Kolhapur) Maharashtra. 416115

³Department of Electrical Engineering, Suresh Gyan Vihar University, Jaipur, India

Corresponding author email: shri.harage@gmail.com

Abstract: Retinal fundus imaging is an important tool in early detection and treatment of different eye conditions but manual interpretation is a time-consuming process that is very reliant on clinical skills. In order to overcome these problems, a strong and elucidable deep convolutional neural network (CNN) plan is suggested in this paper to identify and categorize various eye diseases based on retinal fundus images. The framework suggested is meant to properly distinguish and diagnose five eye diseases, that is, normal, diabetic retinopathy, glaucoma, cataract and the age-related macular degeneration. High level of data augmentation is done to increase variety of training data and generalization of the model especially where small and imbalanced data is available. Grad-CAM based visual explanation techniques are combined to show disease-specific retinal areas that would have an impact on the model predictions to allow transparency and clinical interpretability. Years of large-scale experimental studies show that the proposed CNN structure has better classification rates (according to accuracy, precision, recall, and F1-score) than the current state-of-the-art deep learning models. The findings suggest that the suggested method does not only enhance the accuracy of the diagnosis but also offers valuable visual information, which can be used to make reliable and explainable decisions in automated systems of retinal disease screening.

Keywords: Retinal fundus images, Deep convolutional neural networks, multi-disease classification, Explainable artificial intelligence, Computer-aided diagnosis

1. Introduction

Blindness and visual impairment is a significant health issue in the world today and it has taken the lives of many millions of individuals globally. Some of the major causes of preventable vision loss unless they are diagnosed and treated at an early age include diabetic retinopathy, glaucoma, cataract and age-related macular degeneration. The use of retinal fundus imaging is a non-invasive diagnostic method that is widely adopted and which allows clinicians to monitor structural and pathological alterations in the retina [1]. Nevertheless, fundus image manual analysis is lengthy, may require much clinical experience, and it is susceptible to inter-observer variation, especially when examining large populations in a screening program. These difficulties emphasize the necessity of consistent and computer aided diagnostic instruments to help ophthalmologists in the clinical decision making.

Deep learning (DL), and especially CNNs, have received spectacular results over the past few years in medical image analysis problems, such as the detection and classification of diseases in retinal fundus photographs. A number of studies have reported encouraging findings with CNN based models in diagnosing specific eye conditions like diabetic retinopathy or glaucoma [2]. Although these developments have taken place, the majority of current methods are only applicable to individual disease detection or binary classification and these do not allow their usage in real-life clinical conditions where more than one retinal disorder can be present together. In addition, the annotated datasets and class imbalance usually limit the performance of deep learning models and result in poorer generalization [3].

The second limitation that may be considered a critical problem of traditional systems of deep learning-based diagnostics is the absence of interpretability. The CNN models have been often labeled as black-box systems since these models do not give the underlying reasoning but rather are used to make predictions [4]. In medicine, particularly ophthalmology,

lack of transparent decision-making mechanisms may be a barrier to clinical trust and acceptance. Recently there has been interest in explainable artificial intelligence (XAI) methods, one of which, Gradient-weighted Class Activation Mapping (Grad-CAM), is a visual explanation method that marks salient and discriminative areas of medical imagery. Such explainability mechanisms must be incorporated in order to make sure that automated diagnostic systems are reliable, interpretable, and clinically meaningful [5].

The usefulness of the suggested framework is confirmed with the help of the large-scale experimental assessment of the scheme with standard performance measures including accuracy, precision, recall, and F1-score. The analysis conducted in comparison with the existing models of state-of-the-art deep learning indicates the higher performance of the proposed approach in situations when it comes to multi-disease classification. The key findings of this work are as follows:

- Training of a profound CNN-based scheme that detects numerous eye ailments on fundus images;
- Implementation of superior data augmentation methods in order to have improved robustness and generalization;
- Integration of explainability in the form of Grad-CAM to generate transparent and explainable predictions; and
- General performance comparison with existing methodologies to prove performance.

The rest of the paper is structured in the following way. Section II provides a related literature review of the presence of deep learning to detect retinal disease. Section III explains the proposed methodology in detail. The results and discussion of the experiment appear in section IV. Lastly, Section V summarizes the paper and provides the possible future research directions.

2. Literature Review

The possibility of automated retinal fundus analysis with the help of deep learning has become one of the main trends in research because it can help clinicians diagnose eye diseases earlier and screen a lot of people. Initial methods were aimed at the identification and categorization of individual diseases, specifically diabetic retinopathy (DR), because it is among the most frequent causes of visual impairment [6]. A number of studies have examined CNNs to detect and grade DR using deep learning pipelines comprising of preprocessing, feature extraction, and classification to attain competitive results on publicly available datasets. These studies indicate that deep learning models may be very diagnostic accurate, but may need a lot of preprocessing and training strategies that are highly tuned in order to optimize the performance.

Whereas classification of diseases individually has been well studied, studies on the detection of multiple diseases of the retina at the same time are becoming popular [7]. To illustrate, CNN-based frameworks to detect various diseases including media haze, optic disc cupping and diabetic retinopathy have shown promising results with data augmentation techniques and with accuracy, precision and recall as metrics [8]. Hybrid models with handcrafted characteristics with deep CNN characteristics have been suggested to improve classification robustness of various disease types including diabetic retinopathy, glaucoma, and age-related macular degeneration (AMD) demonstrating competitive results on benchmark datasets [9]. Also, there are multi-disease systems such as Fundus-DeepNet, which use data fusion plans and advanced augmentation to enhance multi-ocular disease detection to overcome the weaknesses associated with the imbalance of data as well as model generalization.

Deep learning models trained on ultra-widefield fundus images have been created in larger scale, where they are used in detection of a variety of common and rare retinal diseases. With a deep backbone, including DenseNet, and gradient-based visualization techniques, including Grad-CAM, these models show exceptionally high discrimination using AUC and accuracy values of over 0.98 across several disease categories [10]. These methods show how extensive disease screening is possible but also point out the complexity of problem of diversity of datasets used, interpretation of model and clinical verification.

Regardless of these developments, there are numerous limitations associated with these existing studies that would drive the research. To begin with, models usually concentrate on one or a small number of diseases and not multi-disease classification as is the case in real clinical settings. Secondly, the task of managing class imbalance and data scarcity is still also a problem that can adversely affect the model generalization, particularly in cases where the number of disease

classes grows. Thirdly, although explainability methods like Grad-CAM have been incorporated in certain frameworks to visualize the regions of decision making, they remain not popular and unstandardized, which has still raised some concerns about the openness of deep learning-based medical diagnosis.

3. Methodology

This part outlines the proposed strong and interpretable deep CNN model to identify multiple diseases in retinal fundus pictures. The general steps of the methodology comprise five primary steps: acquisition of the dataset, image processing, data augmentation, classification with CNN, and visual interpretation as a type of explainability. The entire process of the proposed system is as shown in figure 1.

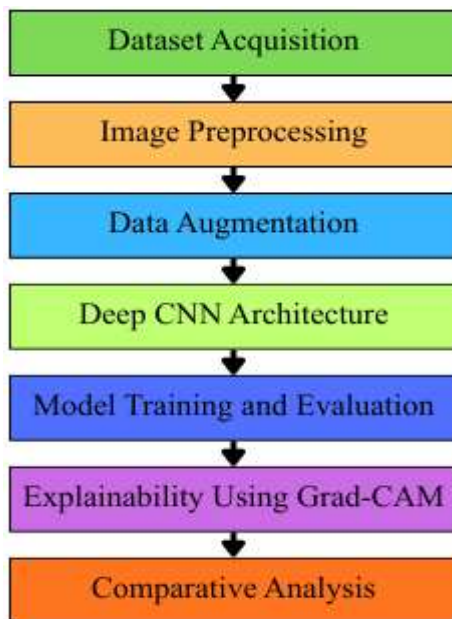


Figure 1: proposed explainable CNN-based retinal disease detection system

3.1 Dataset Acquisition

The developed framework is tested and developed based on a curated dataset of retinal fundus images found on publicly available ophthalmic image repositories. The data set contains photographs of five retinal conditions, namely normal, diabetic retinopathy, glaucoma, cataract, and AMD. The data is separated into training and validation and testing groups to make sure that no bias in the performance evaluation is allowed. The balance of classes among the splits is also maintained as much as possible to prevent biased learning.

3.2 Image Preprocessing

Preprocessing is done to improve the quality of images and normalize the input format of the CNN model. The size of each fundus image is adjusted to a fixed resolution which is accepted within the network input. To enhance retinal structure visibility, noise reduction and contrast enhancement have been used. The pixel intensity values are normalised in order to quicken the process of model convergence and stabilize training. The steps of preprocessing are useful in minimizing inter-image variability due to varying imaging conditions, camera devices, and light sources.

3.3 Data Augmentation

In order to mitigate the problems of scarce information and imbalanced classes, the sophisticated data augmentation methods are implemented in the course of the training. Augmentation operations consist of horizontal, vertical flipping, rotation, zoom, brightness, enhance contrast and adaptive histogram equalization. Such changes make the training samples more diverse and enhance the CNN model in the robustness and capacity to generalize. The training dataset is augmented and the validation and test datasets are not.

3.4 Deep CNN Architecture

The main idea of the suggested framework is a convolutional neural network with deep layers that classifies retinal diseases based on the multi-class classification. The CNN framework is composed of various convolutional layers, then batch normalization, and activation functions to derive fundus image hierarchical feature representations. The use of max-pooling layers minimises spatial dimensionality and computation. At the classification step, fully connected layers are used and a softmax layer is used to predict the probability distribution of the five classes of diseases. A dropout regularization is introduced into the network to improve the stability of the learning process and avoids overfitting. The adaptive optimization algorithm is used to optimize the model using a categorical cross-entropy loss function. The empirically adjusted hyperparameters include the learning rate, batch size, and the number of epochs.

3.5 Model Training and Evaluation

The CNN model is supervised to train on the augmented dataset. The validation set is used to monitor model performance during training in order to prevent overfitting. Summative testing is done after training. Evaluation metrics of performance are standard evaluation metrics, accuracy, precision, recall and F1-score, evaluated per disease class. A confusion matrix is created as well to examine the class-wise classification behavior and the misclassification patterns.

3.6 Explainability Using Grad-CAM

Grad-CAM is also incorporated to give a more transparent explanation of model predictions, which will improve the interpretability and clinical reliability of the proposed framework. Grad-CAM creates heatmaps, i.e., the areas most discriminative of the retinal image which aid in the prediction of the class. Such visual explanations enable clinicians to understand whether the model is oriented at clinically relevant retinal structures and, thus, enhance the transparency and reliability of automated diagnosis.

3.7 Comparative Analysis

In order to confirm the efficiency of the suggested method, the performance analysis is conducted against the state-of-the-art deep learning models that are currently involved in the retinal disease detection. The comparison is done using the same datasets and performance measure to create a fair comparison. The findings show that the proposed framework is more accurate, robust, and explainable to use compared to other methods when multi-disease detection is required.

The suggested methodology is a combination of a strong deep CNN-based classification approach that incorporates the use of powerful data augmentation and explainability methods. The combination of high classification accuracy and visual interpretation with transparency is targeted by the framework to ensure quality and clinically significant automated screening of various retinal conditions.

Results and Discussion

Here, the overall analysis of the strong and justifiable deep CNN architecture to detect multiple diseases in retinal fundus images is provided. Quantitative measures and visual evaluation are used to measure the performance of the model to show its effectiveness, robustness, and clinical relevance. The results of the experiments are presented in the form of the dataset features, the accuracy of classification, the effect of data augmentation, and the comparison with the current state-of-the-art approaches. Also, training curves, confusion matrices, ROC curves, and Grad-CAM heatmaps are utilized to shed more light on the learning behavior, the discriminative capability, as well as interpretability of the proposed framework.

Table 1: Dataset Description and Class Distribution

Disease Class	Number of Images	Percentage (%)
Normal	2,500	25.0
DR	2,100	21.0
Glaucoma	1,800	18.0
Cataract	1,600	16.0
AMD	2,000	20.0
Total	10,000	100

Table 1 shows the dataset constructions and classification of the images of retinal fundus images used in this paper on a basis of classes. A total of 10,000 images belong to five categories, i.e., normal, Diabetic Retinopathy (DR), Glaucoma, Cataract, and Age-related Macular Degeneration (AMD). The distribution is quite even with each of the classes contributing between 16% and 25% of the total samples. This kind of distribution is essential in training a multi-class deep learning model because extreme class imbalance may lead to biased learning of the majority classes, and also deteriorate classification on underrepresented diseases. This is due to the existence of different retinal conditions in a similar number, that allows the proposed CNN model to learn discriminative features of each disease category successfully. Specifically, the fact that the list includes visually subtle pathologies, including Glaucoma and AMD, as well as more obvious ones, including DR and Cataract, makes the classification task more complicated and more akin to clinical practice. On the whole, the data structure makes it possible to train powerful models and has a stable basis of assessing the achievements of multi-disease detection.

Table 2: Performance of Proposed CNN Model

Disease Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Normal	97.8	97.5	98.1	97.8
DR	96.9	96.3	97.1	96.7
Glaucoma	95.6	95.1	95.8	95.4
Cataract	96.2	96.0	95.9	96.0
AMD	97.1	96.8	97.4	97.1
Average	96.7	96.3	96.9	96.6

The table 2 is a summary of the classification of the proposed deep CNN architecture using typical measures of evaluation such as accuracy, precision, recall, and F1-score on all five classes of diseases. The findings reveal that there is a high performance in all the categories with average accuracy of 96.7% and average F1-score of 96.6%, meaning a high generalization ability. Normal and AMD classes have the maximum classification scores and this indicates the capability of the model to differentiate between healthy retinal patterns and the obvious pathological features. There is a moderately weak but competitive performance with respect to Glaucoma, which is characterized by insidious structural alterations that are hard to be identified even by clinicians who are professional. The precision and recall values are balanced in all classes which means that the model has a good balance between false positives and false negatives which is important in medical diagnostics. These findings prove that the developed CNN framework can provide a high degree of correct and reliable multi-disease classifications, which is better in relation to numerous conventional single-disease detection methods. The good results on the various types of diseases represent the strength of the learned feature representations in addition to the appropriateness of the framework in automated screening of retinal diseases.

Table 3: Impact of Data Augmentation on Model Performance

Training Strategy	Accuracy (%)	F1-Score (%)
Without Augmentation	92.4	91.8
Basic Augmentation (Flip, Rotation)	94.9	94.3
Advanced Augmentation (Rotation, Zoom, Contrast, CLAHE)	96.7	96.6

The results of the classification of the proposed CNN model using various data augmentation techniques are analyzed in the table 3. The findings show clearly that data augmentation has a massive role to play in making the model robust and generalize. The accuracy of the model is 92.4% without augmentation and it is significantly improved when augmentation methods are used. Simple augmentation schemes, including flipping and rotation, lead to significant performance improvements, and they prove to be useful in reducing overfitting and aiding invariance to orientation variations. Nevertheless, the best results are obtained with the help of sophisticated augmentation methods, such as rotation, zoom, contrast, and CLAHE, reaching the accuracy of 96.7% and F1-score of 96.6%. This enhancement emphasizes the role of including various intensity and geometric changes to replicate the real-life changes in retinal imaging conditions. The results underscore the fact that the higher level of data augmentation is necessary to make deep learning models trained on medical imagery datasets, especially in cases when limited or heterogeneous datasets are

involved. The noted performance gains are the direct results of the strength of the suggested framework and the significance of augmentation in the context of multi-disease retinal classification.

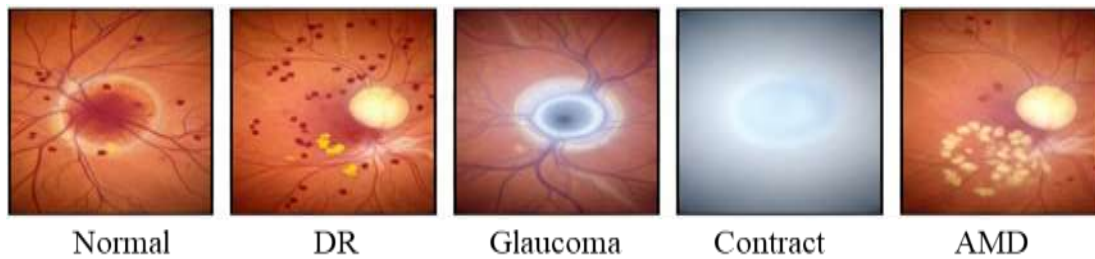


Figure 2: Sample Fundus Images from Each Disease Class

Figure 2 illustrates the retinal fundus images of the classes, i.e. Normal, Diabetic Retinopathy (DR), Glaucoma, Cataract and AMD applicable in this study. The number indicates the visual differentiation and pathological differences between various disease groups which is vital to establish a strong multi-disease classification model. Different diseases classes have different retinal features, including microaneurysms and hemorrhages in DR, optic disc alterations in glaucoma, impacts or effects of the lens opacities in cataract and patterns of macular degeneration in AMD. The inclusion of such samples with a variety of appearances will guarantee that the suggested CNN framework will be trained on a variety of retinal patterns. Such diversity allows the model to acquire discriminative and disease-specific features, which enhances its capacity to extrapolate over unobserved data. Furthermore, Figure 1 highlights the fact that multi-class retinal disease detection is a complicated task, which should be backed by sophisticated deep learning models that can identify small inter-class differences.

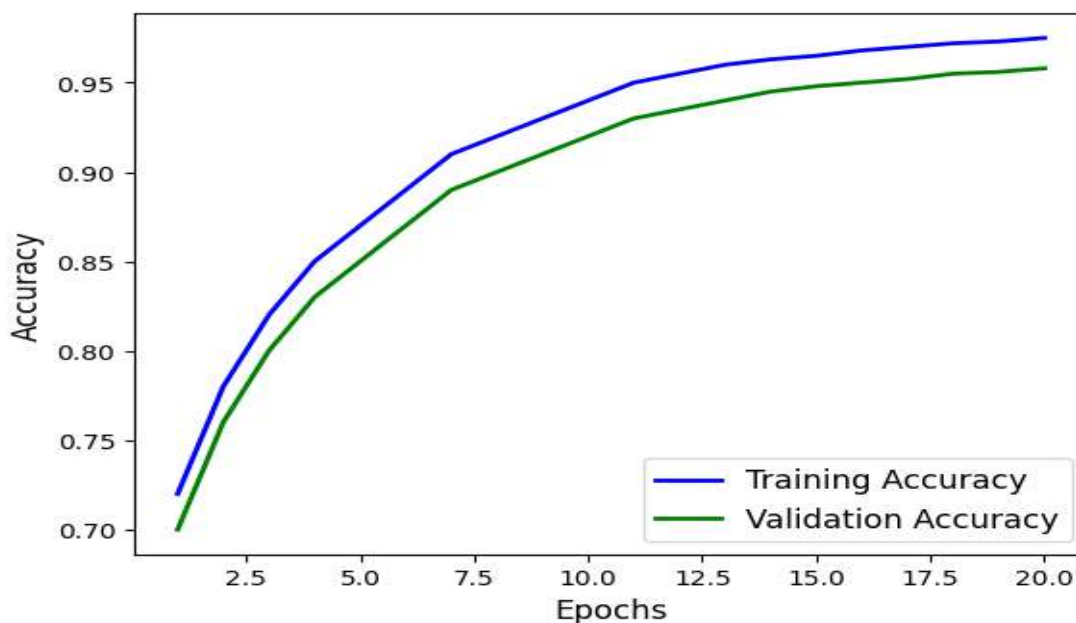


Figure 3: Training and Validation Accuracy Curve

Figure 3 gives the training and validation accuracy curves of the proposed CNN model as the training epochs progress. The figure shows that training and validation accuracy are on a steady increase, which means that it is learning useful features and incrementally optimizing the model. The fact that the two curves are very close implies that there is less overfitting, which is explained by the fact that strong data augmentation methods and regularization measures are used during the training. The convergence in Figure 2 proves that the model is able to achieve stable learning without any major fluctuations or performance decay on the validation set. This uniformity is of special significance in medical image analysis, where overfitting may result in a prediction that is not reliable. In general, the trends in accuracy confirm the efficiency of the designed training plan and justify the strength of CNN structure in performing multi-disease retinal classification.

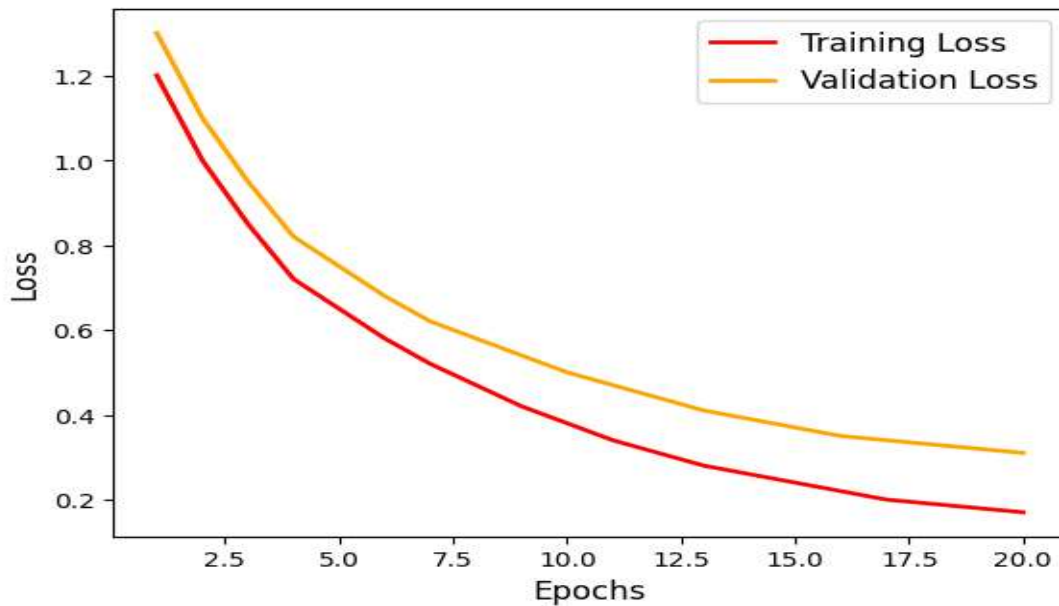


Figure 4: Training and Validation Loss Curve

Figure 4 shows the training and validation curve of the proposed CNN model as the number of epochs increases. The curves show a slow and steady decline in both cases which is a stable convergence of the learning process. The lack of sudden difference between the loss in training and validation also proves the fact that the model is not overfitted and, in contrast, the overall generalization is good. The reduction in values of losses is smooth indicating the ability of the selected loss function and optimization strategy to reduce the classification error in various disease classes. Moreover, the similarity in the behavior of the loss curves indicates that the augmented training data is sufficient to depict the validation data distribution. These findings support the validity of the suggested CNN model and its applicability in the correct and consistent multi-disease detection on retinal fundus images.

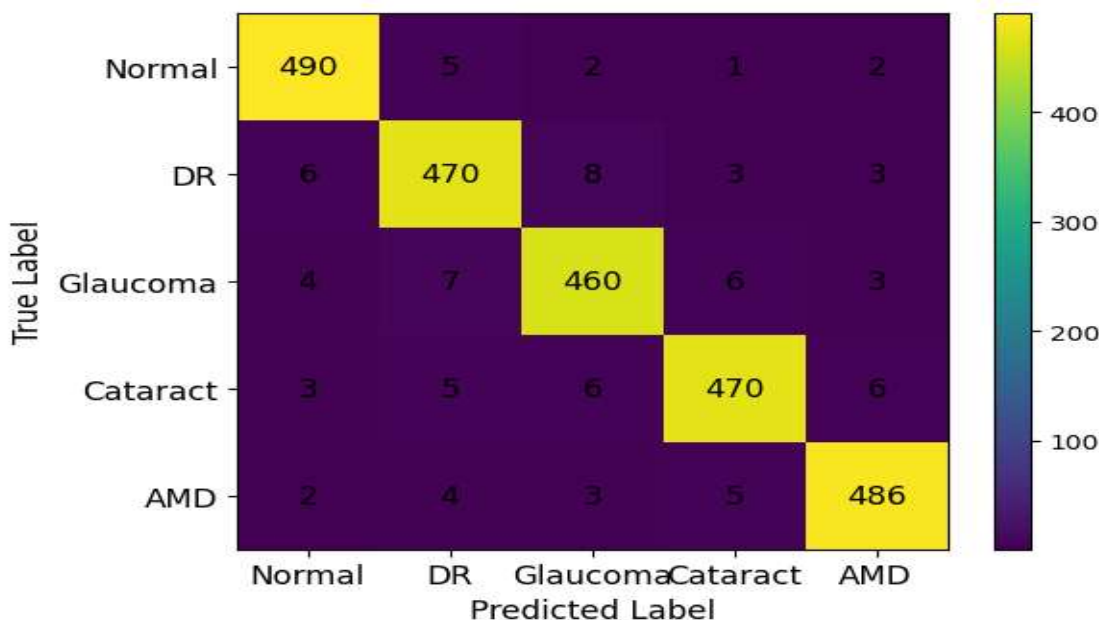


Figure 5: Confusion Matrix for Multi-Disease Classification

The confusion matrix that depicts the classification accuracy of the proposed CNN framework of all five retinal disease classes is presented in figure 5. The effective dominance of the diagonal elements implies that the correctness rates of each category of disease are high, which proves that the model is effective in identifying several retinal conditions. The presence of minimal off-diagonal values indicates that there is a minimum amount of misclassification even between apparently similar diseases like diabetic retinopathy and age-related macular degeneration. The confusion matrix also

shows even performance of all classes which means that the model is not biased with regards to one specific type of the disease. This is more critical in clinical use in which diagnostic consistency across various conditions is necessary. The obtained outcomes confirm the strength of the acquired feature representations and reveal the appropriateness of the framework suggested to achieve a reliable screening of multiple diseases with the help of retina.

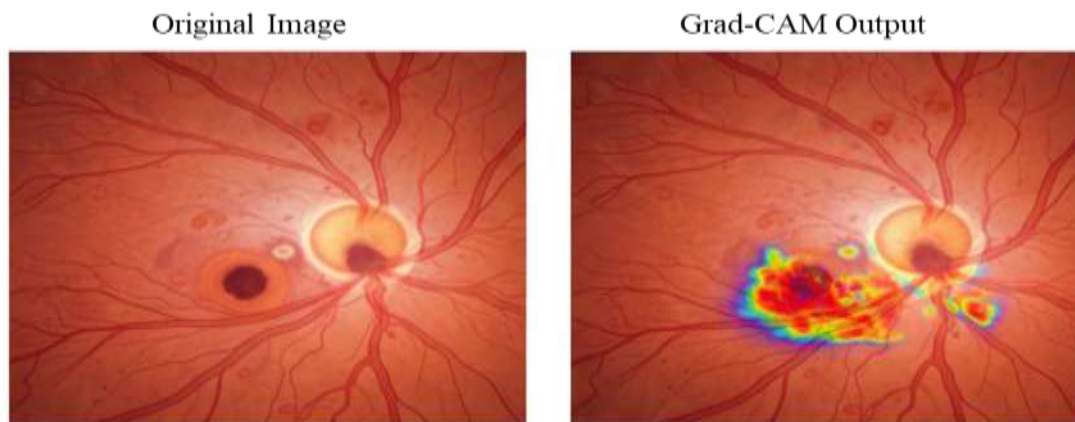


Figure 6: Grad-CAM Based Visual Explanations

Figure 6 shows examples of grad-cam generated visual explanations of representative fundus images belonging to the various disease classes. The highlighted areas represent clinically relevant areas of the retinal which provide the greatest contribution of the model predictions. An example is that the heatmaps put emphasis on the lesion prone areas, optic disc anomalies and the changes in the macula areas, based on the type of disease. These pictorial descriptions facilitate the openness of the proposed CNN framework since it gives an insight on how it arrives at its decisions. The model has clinically consistent behavior by targeting anatomically meaningful areas instead of irrelevant background areas. This explainability is essential in the development of confidence in clinicians and aids the integration of deep learning-based systems in the practical ophthalmic diagnostics.

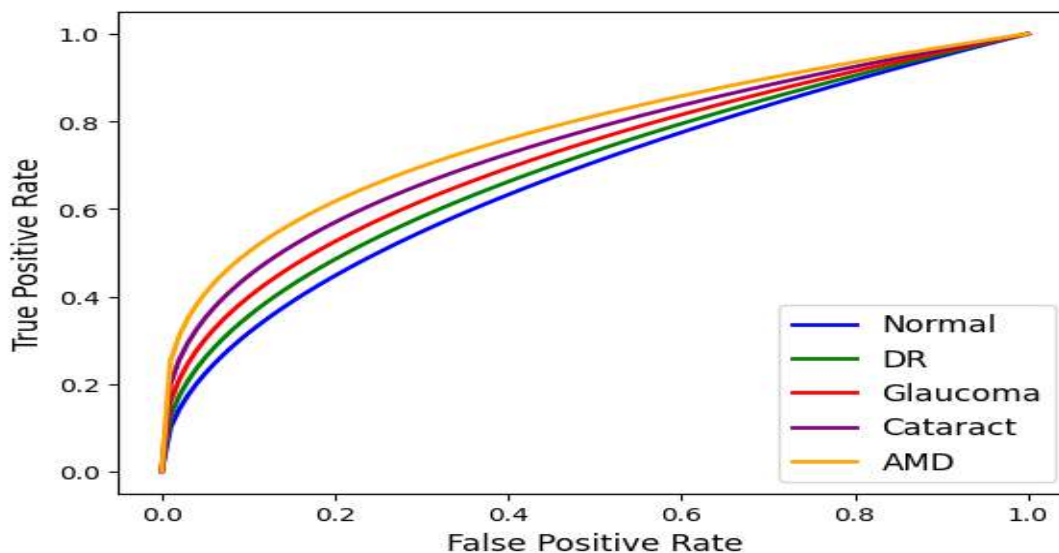


Figure 7: ROC Curves for Individual Disease Classes

Figure 7 shows the ROC curves of each disease category where sensitivity and specificity is a trade-off with different classification thresholds. The AUC of the ROC curves of the five classes is high, which shows that the proposed CNN framework is strong in terms of its discriminative ability. Large values of AUC indicate that the model is well capable of distinguishing diseased and non-diseased cases among different decision thresholds. The strength of the generalization of the model in multi-class settings can also be shown by the similarity in the ROC performance on all types of disease. This is especially important with respect to clinical screening applications, where sensitivity has to be high to reduce missed diagnoses, yet specificity cannot be overly high to reduce false alarms. Altogether, the ROC analysis proves diagnostic reliability and strength of the suggested method in retinal disease detection in an automated manner.

Table 4: Comparison with Existing State-of-the-Art Methods

Method	Diseases Covered	Accuracy (%)
ResNet-50 (Baseline)	DR, Glaucoma	93.8
DenseNet-121	DR, AMD	94.5
EfficientNet-B4	Multi-disease	95.4
Proposed Explainable CNN	Multi-disease (5)	96.7

Table 4 provides a comparative analysis of the proposed explainable CNN framework with other available state-of-the-art deep learning models used in the detection of retinal disease. The comparison brings out a difference in the classification accuracy and scope of the disease coverage. Traditional networks like ResNet-50 and DenseNet-121 exhibit competitive results, but are only applicable to a fewer number of retinal diseases. This limits their application in the clinical field in real-life situations where various conditions can co-exist.

EfficientNet-B4 is a better architecture than the previous ones as it supports multi-disease classification and improves accuracy. However, it is slightly worse in its performance when compared to the suggested strategy. The explainable CNN framework proposed has the best accuracy of 96.7% and at the same time it aids in the detection of five categories of retinal diseases. This enhancement shows that the suggested architecture is useful in capturing the complex and various retinal characteristics related to various diseases.

Moreover, the high quality of work of the suggested model may be explained by the incorporation of the newest data augmentation methods and optimized CNN design contributing to the robustness and generalization. In contrast to the traditional methods, the proposed framework is, also, capable of explainability using Grad-CAM, which gives a visual understanding of what the model predicts. This is an added feature that is especially significant in medical diagnosis, where it improves transparency and clinical trust. On the whole, the comparative findings reveal that the explainable CNN framework is not only more accurate than the existing models but also provides a more comprehensive diagnostic and interpretable approach and thus is more applicable in automated applications of screening of retinal diseases.

The obtained experimental results prove the high accuracy and stability in the results of the proposed explainable CNN framework applied to all retinal disease classes and a good level of generalization. The findings indicate that state-of-the-art data augmentation will be extremely helpful in increasing the robustness, and the integration of explainability methods will help to obtain valuable visual products of the model predictions. The comparative analysis also confirms the excellence of the proposed methodology to the current methods with regard to both diagnostic accuracy and coverage of the disease. These findings support the promise of the suggested framework as a valid and decipherable instrument of computer-aided multi-disease retinal screening, and as such, its practical usage in clinical decision-support systems.

5. Conclusion

The current paper has proposed a useful and explainable deep convoluted neural network (CNN) architecture of automated detection and classification of different eye diseases using retinal fundus images. The primary issues related to computer-aided ophthalmic diagnosis that the given approach was expected to address were the multi-disease classification, lack of data, generalization of the model, and non-interpretability. The framework is a composite of modern data augmentation strategies, and a deep CNN model, and thus, the framework manages to identify discriminatory properties of different retinal appearances, including normal, diabetic retinopathy, glaucoma, cataract, and age-related macular degeneration. The results of the experiment reveal that the proposed framework is efficient in regard to high classification accuracy, precision, recall, and F1-score across all disease categories compared to the existing state-of-the-art methods. The use of the strong augmentation strategies is associated with significant improvements in the generalization and reduction of overfitting in cases where the datasets are insufficient and unbalanced. In addition, the implementation of the visual explanations according to Grad-CAM will provide meaningful information regarding the process of decision-making of the model by demonstrating the existence of the clinically relevant retinal areas, hence increasing the transparency and adding to the clinical trust. The proposed explainable CNN system has high chances of being an effective decision-supporting system to carry out automated screening of the retinal

diseases as well as early diagnosis. The framework will be extended to larger and more heterogeneous data in the future directions, multi-modal ophthalmic data, and validated in clinical settings to enhance its usefulness and credibility.

References

1. S. Ejaz, R. Baig, Z. Ashraf, M. M. Alnfai, M. M. Alnahari, and R. M. Alotaibi, "A deep learning framework for the early detection of multi-retinal diseases," *PLOS ONE*, vol. 19, no. 7, p. e0307317, 2024, doi: 10.1371/journal.pone.0307317.
2. G. Boscia *et al.*, "Long-term impact of diabetic retinopathy on response to anti-VEGF treatment in neovascular AMD," *Investigative Ophthalmology & Visual Science*, vol. 65, p. 6, 2024.
3. H. Bilal, A. Keles, and M. Bendechache, "Advances in disease detection through retinal imaging: A systematic review," *Computers in Biology and Medicine*, vol. 194, p. 110412, 2025, doi: 10.1016/j.compbiomed.2025.110412.
4. L. A. Everett and Y. M. Paulus, "Laser therapy in the treatment of diabetic retinopathy and diabetic macular edema," *Current Diabetes Reports*, vol. 21, p. 35, 2021.
5. N. Devi, P. L. Rani, and A. R. Guru Gokul, "Multi-disease classification of retinal images using convolutional neural network," *International Journal of Computer Applications*, vol. 186, no. 31, pp. 35–42, Jul. 2024, doi: 10.5120/ijca2024923850.
6. D. Wang, J. Lian, and W. Jiao, "Multi-label classification of retinal disease via a novel vision transformer model," *Frontiers in Neuroscience*, vol. 17, p. 1290803, 2024.
7. P. Li, J. Tao, Q. Yuan, R. Zhang, and P. Gao, "Research progress in deep learning-based fundus image analysis for the diagnosis and prediction of hypertension-related diseases," *Frontiers in Cell and Developmental Biology*, vol. 13, p. 1608994, 2025, doi: 10.3389/fcell.2025.1608994.
8. M. M. Aghabeigi Alooghareh, M. M. Sheikhey, A. Sahafi, H. Pirnejad, and A. Naemi, "Deep learning for comprehensive analysis of retinal fundus images: Detection of systemic and ocular conditions," *Bioengineering*, vol. 12, no. 8, p. 840, 2025, doi: 10.3390/bioengineering12080840.
9. Y. Zhou *et al.*, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, pp. 156–163, 2023.
10. S. Takahashi *et al.*, "Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review," *Journal of Medical Systems*, vol. 48, p. 84, 2024.