

Eye Disease Detection Using Multiclass Classification of Retinal Images

Kaustubh Shinde*, Vinay Mortole, Shreenath Patil, Suraj More, Manish Gawas

¹Department of Computer Engineering, Sinhgad Institute of Technology and Science, Narhe, Pune, India

Abstract

The increasing global prevalence of myopia and its progression to Pathologic Myopia (PM) pose a significant risk of irreversible vision loss, making early and accurate detection essential. This paper presents a deep learning-based framework for multi-class classification of retinal fundus images into Normal, Myopia, and Pathologic Myopia categories. The proposed approach employs EfficientNet-B3 with a custom classification head and a two-phase transfer learning strategy that combines AdamW optimization, weighted loss, label smoothing, MixUp augmentation, and cosine annealing learning rate scheduling. To improve interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is integrated to visualize the regions influencing model predictions.

The model was trained and evaluated on a large-scale retinal image dataset drawn from multiple sources, including the PALM dataset, using a custom 380×380 input resolution. Experimental results show that EfficientNet-B3 achieved an overall test accuracy of 96.08%, outperforming EfficientNet-B0 (89.3%) and ResNet50 (82%). The model also demonstrated strong class-wise performance, including perfect detection of Pathologic Myopia with 100% precision, recall, and F1-score, while significantly improving discrimination between Normal and Myopia cases. The framework is further deployed as a Django-based web application supporting real-time prediction and automated report generation. These results indicate that the proposed system is a clinically relevant decision-support tool for automated retinal screening.

Keywords—Retinal Image Classification, Myopia Detection, Pathologic Myopia, EfficientNet-B3, EfficientNet-B0, ResNet50, Transfer Learning, Grad-CAM, Medical Image Analysis, Clinical Decision Support System.

1 Introduction

Myopia has become an increasingly important public health issue worldwide, with its prevalence rising steadily across both pediatric and adult populations. Although conventional myopia can usually be corrected with optical means, a sub-set of cases progresses to Pathologic Myopia (PM), a vision-threatening condition associated with progressive retinal degeneration. Because this progression can lead to irreversible visual impairment, timely recognition is essential for preventing severe outcomes.

Retinal fundus imaging is widely used to examine structural abnormalities linked to myopia and its pathological progression. In practice, however, reliable interpretation of these

images depends on ophthalmic expertise, which is often limited in routine screening environments and underserved regions. This creates a clear need for automated systems that can support early detection while remaining scalable and dependable.

Deep learning, especially Convolutional Neural Networks (CNNs), has shown strong capability in medical image classification by learning discriminative spatial features directly from image data. Despite this progress, much of the existing literature on retinal disease analysis concentrates on binary classification settings, such as disease versus normal, rather than finer-grained multi-class prediction. As a result, the clinically meaningful distinction among Normal, Myopia, and Pathologic Myopia remains relatively underexplored.

Another important limitation of many deep learning systems is their limited interpretability. In medical applications, a model's prediction is not sufficient on its own; clinicians also need to understand which regions of the image influenced the decision. Without such transparency, confidence in automated diagnosis may remain low, particularly for severe conditions such as Pathologic Myopia.

To address these issues, this study proposes a multi-class retinal fundus classification framework based on EfficientNet-B3. The method uses a two-phase transfer learning strategy to adapt the model to the target task and incorporates Gradient-weighted Class Activation Mapping (Grad-CAM) to provide visual explanations of model predictions. A comparative evaluation against EfficientNet-B0 and ResNet50 is also performed to assess the benefit of the proposed architecture.

The main contributions of this work are as follows:

- A three-class retinal fundus image classification framework using EfficientNet-B3 for Normal, Myopia, and Pathologic Myopia recognition.
- A two-phase transfer learning strategy combined with label smoothing and MixUp augmentation to improve generalization and robustness.
- Grad-CAM-based explanation of predictions to improve transparency and support clinical interpretation.
- A comparative study with EfficientNet-B0 and ResNet50 to demonstrate the effectiveness of the proposed approach.
- Deployment of the trained system as a web-based decision-support tool for practical use.

2 Related Work

Deep learning has played a major role in advancing retinal image analysis, particularly through the use of Convolutional Neural Networks (CNNs). By learning hierarchical representations directly from fundus images, CNNs are able to detect both low-level texture patterns and higher-level anatomical structures associated with myopic changes. This capability has made them highly effective for retinal disease recognition and classification.

A number of studies have addressed myopia-related prediction and classification from retinal or multimodal image data. Qi et al. (2024) proposed a CNN-based framework combined with metadata fusion to forecast myopia onset over multiple years, reporting an AUC of 0.908. Xing et al. (2025) introduced a lightweight attention-based model with depth-wise separable convolutions, achieving competitive accuracy while reducing computational cost.

Much of the published work still focuses on binary classification, separating healthy from diseased cases. Li et al. (2023) and Kumar et al. (2024) explored CNN-based architectures such as ResNet and VGG for screening tasks of this type, showing promising performance in detecting abnormal cases. However, binary formulations do not capture the intermediate distinction between Myopia and Pathologic Myopia, which is important for clinically meaningful grading.

Research on Pathologic Myopia detection has also grown in recent years. Wang et al. (2025) used an ensemble of MobileNetV2 and DenseNet models on PALM and custom datasets, obtaining strong classification results. Chen et al. (2026) proposed a dual autoencoder-based method to identify pathological patterns, improving sensitivity at the expense of additional model complexity. Zhao et al. (2025) examined EfficientNet-based multi-class retinal classification and achieved high accuracy, although the interpretability of the decision process was not emphasized. Patel et al. (2025) investigated multimodal learning with OCT and fundus images, which improved diagnostic coverage but increased implementation complexity.

Despite these contributions, several gaps remain. First, many studies still treat retinal screening as a binary problem rather than a three-class classification task involving Normal, Myopia, and Pathologic Myopia. Second, a large number of approaches provide high accuracy but limited explanation of model decisions, which restricts clinical trust. Third, several methods are not designed for practical deployment, making them less suitable for real-world decision-support use. These limitations motivate the development of an end-to-end framework that combines accurate multi-class classification, interpretable predictions, and deployable implementation.

These observations motivate the proposed framework, which integrates an EfficientNet-B3-based architecture with advanced training strategies, class imbalance handling, Grad-CAM-based explainability, and a web-based deployment pipeline.

Table 1: Summary of representative prior studies

Study	Method	Main Limitation
Qi et al. (2024)	CNN with metadata fusion	Predictive rather than diagnostic
Xing et al. (2025)	Attention-based lightweight CNN	Limited dataset scale
Li et al. (2023)	CNN / ResNet	Binary classification only
Wang et al. (2025)	MobileNetV2 + DenseNet ensemble	Limited deployment focus
Chen et al. (2026)	Autoencoder-based framework	Higher complexity, lower interpretability
Zhao et al. (2025)	EfficientNet-based classifier	Explainability not addressed
Kumar et al. (2024)	VGG-based model	Small dataset, binary setup
Patel et al. (2025)	Multimodal learning	Greater deployment complexity

3 Methodology

This section describes the proposed framework for automated multi-class classification of retinal fundus images. The methodology covers the overall study design, dataset preparation, preprocessing, model architecture, training strategy, evaluation metrics, and deployment pipeline. The proposed system is developed to balance classification accuracy, robustness, interpretability, and practical usability in a clinical screening setting.

3.1 Study Design

This study presents an end-to-end deep learning framework for multi-class classification of retinal fundus images into three categories: Normal, Myopia, and Pathologic Myopia. The system integrates preprocessing, model training, evaluation, and deployment into a unified pipeline. A comparative analysis is also performed between baseline models and the proposed EfficientNet-based approach to assess performance improvements.

The classification task is formulated as a supervised multi-class problem, where each input retinal image is assigned to one of the three predefined classes. Particular attention is given to the detection of Pathologic Myopia because of its clinical importance, and the evaluation strategy is designed to measure both overall performance and class-wise behavior.

3.2 Dataset Description

The dataset used in this study comprises approximately 132,000 retinal fundus images gathered from multiple sources, including the PALM dataset. The images are organized into three clinically relevant classes: Normal, Myopia, and Pathologic Myopia. As is common in medical imaging datasets, the class distribution is imbalanced, with Pathologic Myopia representing a comparatively smaller subset of the samples.

To support reliable evaluation, the dataset is divided into training, validation, and testing subsets. This split enables model selection and final assessment on unseen data. Because

of the class imbalance, particular care is taken during training to reduce bias toward the majority classes and improve sensitivity to the minority class.

3.3 Preprocessing and Data Augmentation

All images are resized to 380×380 pixels to match the input requirements of the proposed EfficientNet-B3 model. The images are then normalized using the ImageNet mean and standard deviation so that the pretrained backbone can be fine-tuned effectively.

During training, data augmentation is applied to improve robustness and reduce overfitting. The augmentation pipeline includes horizontal and vertical flipping, random rotation, perspective transformation, color jittering, and random erasing. These transformations simulate realistic variations in fundus image acquisition, such as changes in orientation, illumination, and device-specific characteristics, thereby helping the model generalize better to unseen clinical data.

3.4 Model Architecture

The proposed framework is built upon EfficientNet-B3 as the feature extraction backbone. EfficientNet adopts compound scaling to balance network depth, width, and input resolution in a coordinated manner, which allows the model to achieve strong representational capacity while remaining computationally efficient.

For the present classification task, the pretrained EfficientNet-B3 backbone is adapted by replacing its original classification layer with a custom prediction head. The head is designed to improve regularization and class separation through a combination of dropout, normalization, and nonlinear activation.

- Dropout (rate = 0.4)
- Fully connected layer (backbone output \rightarrow 512)
- Batch Normalization
- SiLU activation function
- Dropout (rate = 0.3)
- Fully connected layer (512 \rightarrow 3)

This design preserves the expressive power of the pretrained backbone while allowing the classifier to adapt to the retinal image classification task.

3.5 Training Strategy

A two-stage transfer learning strategy is employed to improve adaptation while retaining useful pretrained representations.

In the first stage, the EfficientNet-B3 backbone is frozen and only the custom classification head is trained. This enables the model to learn task-specific decision boundaries without immediately altering the pretrained visual features.

In the second stage, selected deeper layers of the backbone are unfrozen and fine-tuned together with the classifier. A smaller learning rate is used for the backbone to avoid disrupting the pretrained weights, whereas a relatively larger learning rate is assigned to the classification head to support faster adaptation.

Table 2: Training configuration of the model

Parameter	Phase 1	Phase 2
Epochs	10	20
Optimizer	AdamW	AdamW
Learning Rate	1×10^{-3}	5×10^{-6} (backbone), 5×10^{-5} (classifier)
Weight Decay	1×10^{-3}	1×10^{-3}
Label Smoothing	0.1	0.15
MixUp	Disabled	Enabled ($\alpha = 0.4$)
Scheduler	Cosine Annealing ($T_{max} = 10$)	Cosine Annealing ($T_{max} = 20$)
Early Stopping Patience	5	7

To further improve generalization, label smoothing is applied to reduce overconfident predictions, while MixUp is introduced in the second stage to encourage smoother decision boundaries. Cosine annealing is used to gradually adjust the learning rate and promote stable convergence throughout training.

3.6 Class Imbalance Handling

The dataset exhibits an imbalanced class distribution, with Pathologic Myopia appearing less frequently than Normal and Myopia. If left unaddressed, this imbalance may bias the model toward the majority classes and reduce its ability to detect clinically important minority cases.

To reduce this effect, the training process uses a weighted cross-entropy loss function in which each class contributes according to its frequency in the dataset. By assigning a larger penalty to misclassification of underrepresented classes, the model is encouraged to learn decision boundaries that are less sensitive to class imbalance.

This strategy is reflected in the final results, where the model achieves perfect recall for Pathologic Myopia. Such performance suggests that the weighting scheme helps preserve sensitivity for high-risk cases while still maintaining strong overall classification accuracy.

3.7 Explainability using Grad-CAM

To improve interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated into the framework. Grad-CAM produces class-discriminative visual explanations by backpropagating gradients from the predicted output to the last convolutional feature maps.

The resulting heatmaps are superimposed on the original fundus images to reveal the image regions that most strongly influence the prediction. In retinal analysis, this allows inspection of whether the model is attending to meaningful anatomical structures such as the optic disc, macular region, and surrounding retinal tissue.

The qualitative Grad-CAM results indicate that the network generally focuses on clinically relevant regions rather than irrelevant background areas. This supports the reliability of the predictions and strengthens the suitability of the model for clinical decision-support applications by improving transparency and trust.

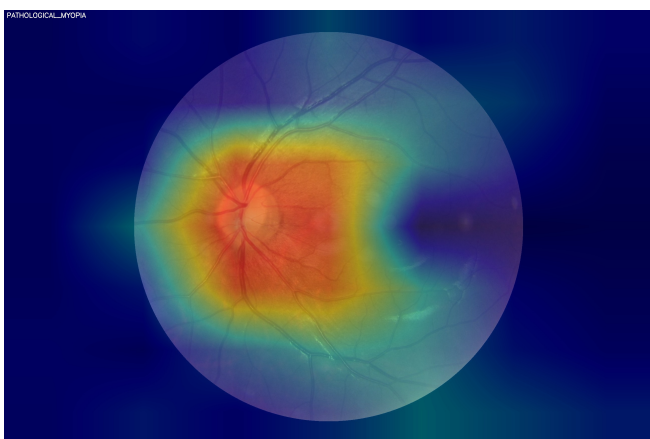


Figure 1: Grad-CAM visualization for a Pathologic Myopia prediction. The highlighted regions indicate that the model attends to clinically relevant retinal structures during inference.

3.8 Baseline Model: ResNet50

To assess the effectiveness of the proposed EfficientNet-B3-based approach, a ResNet50 model is included as a baseline. ResNet50 is a deep residual network that uses skip connections to facilitate optimization and improve gradient flow during training.

The baseline model is evaluated under the same dataset split, preprocessing pipeline, training protocol, and performance metrics as the proposed model to ensure a fair comparison. This controlled setup allows differences in performance to be attributed more reliably to the model architecture rather than to experimental variation.

The experimental results show that ResNet50 performs well, but its overall accuracy and macro F1-score remain below those of EfficientNet-B3. In particular, ResNet50 exhibits more confusion between the Normal and Myopia classes, suggesting that its feature representation is less effective at capturing subtle retinal differences. These results support the advantage of EfficientNet-B3's compound scaling design for fine-grained medical image classification.

3.9 System Deployment

To examine practical usability, the trained model is deployed within a web-based application built using the Django frame-

work. The system provides role-based access for patients and clinicians so that sensitive information is available only to authorized users.

When a retinal image is uploaded, the backend processes it and returns the predicted class, confidence information, and Grad-CAM-based visual explanations. The application also stores patient details, uploaded images, and prediction outcomes in a lightweight SQLite database, enabling organized record management within the deployed environment.

In addition, the system generates PDF reports containing the diagnosis and supporting details for clinical review. This makes the framework more practical for real-world screening workflows and improves its usefulness as a decision-support tool.

3.10 Evaluation Metrics

Model performance is measured using standard multi-class classification metrics, including accuracy, precision, recall, and F1-score. Both macro-averaged and weighted variants are reported to provide a balanced view of performance across classes, especially under class imbalance.

Table 3: Evaluation Metrics of EfficientNet-B3

Metric	Value (%)
Accuracy	95.12
Precision	95.68
Recall	97.59
F1-Score	96.46

A confusion matrix is used to examine class-wise prediction patterns in greater detail. The evaluation shows that the model achieves perfect recall for Pathologic Myopia, which is important for identifying clinically critical cases. At the same time, some Normal samples are still misclassified as Myopia, indicating partial overlap in visual features between these categories.

Table 4: Confusion matrix of EfficientNet-B3

Actual / Predicted	Myopia	Normal	Pathologic
Myopia	9496	0	0
Normal	667	8557	0
Pathologic	0	0	748

Together, these metrics provide a more complete assessment of the model's behavior and help clarify both its strengths and the remaining sources of error in a clinically relevant setting.

4 Results

The experimental evaluation is designed to assess the classification performance of the proposed model under a controlled and reproducible testing protocol. This section summarizes the dataset split, evaluation procedure, and implementation settings used to obtain the reported results. The goal is to

provide a clear basis for interpreting the performance metrics presented in the following subsections.

4.1 Experimental Setup and Evaluation Protocol

The proposed model is evaluated on a large-scale retinal fundus image dataset comprising approximately 131,048 images, including samples from the PALM dataset. The training set contains three classes: Normal (43,049 samples), Myopia (44,299 samples), and Pathologic Myopia (4,783 samples).

For testing, the model is assessed on a separate unseen set with the following class distribution: Myopia (9,496), Normal (9,224), and Pathologic Myopia (748). This partitioning provides a realistic measure of generalization by ensuring that the reported results are obtained on data not used during model fitting.

Training follows a two-phase transfer learning strategy with early stopping to reduce overfitting. The implementation also uses ImageNet-based normalization, channels last memory format, and cuDNN benchmarking to support efficient computation during experimentation.

4.2 Overall Performance

The proposed EfficientNet-B3 model achieves an overall test accuracy of 96.08%, indicating strong performance in multi-class retinal image classification.

Table 5: Overall performance metrics of EfficientNet-B3

Metric	Value
Accuracy	0.9608
Macro Precision	0.975
Macro Recall	0.972
Macro F1-score	0.973
Weighted Precision	0.964
Weighted Recall	0.961
Weighted F1-score	0.961

The high macro F1-score indicates that the model performs consistently across classes, while the weighted metrics show that strong performance is maintained even under class imbalance.

4.3 Class-wise Performance Analysis

The class-wise results provide a clearer view of how the model behaves across the three diagnostic categories.

Table 6: Class-wise performance metrics of EfficientNet-B3

Class	Precision	Recall	F1-score	Support
Myopia	0.926	1.000	0.961	9496
Normal	1.000	0.918	0.957	9224
Pathologic Myopia	1.000	1.000	1.000	748

The model achieves perfect recall for Pathologic Myopia,

which is especially important because missing such cases can delay diagnosis and treatment. This result suggests that the network remains highly sensitive to clinically critical retinal patterns.

Performance on the Normal class is lower than for the other categories, with recall of 0.918. This indicates that some Normal images are still assigned to the Myopia class, reflecting the visual similarity between normal retinal appearance and mild myopic changes. Even so, the class-wise scores remain strong overall, showing that the model generalizes well across all three categories.

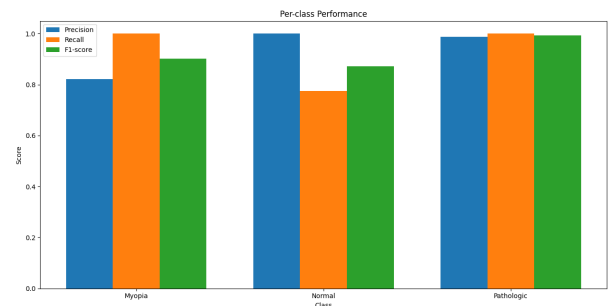


Figure 2: Per-class precision, recall, and F1-score for EfficientNet-B3. The model performs strongly across all classes, with perfect classification of Pathologic Myopia.

Figure 2 shows that the proposed model maintains high precision and recall for all categories, with the most noticeable improvement seen in the Normal class compared with the earlier baseline model.

4.4 Confusion Matrix Analysis

The confusion matrix offers a detailed view of the prediction errors made by the model.

Table 7: Confusion matrix of EfficientNet-B3

Actual / Predicted	Myopia	Normal	Pathologic
Myopia	9496	0	0
Normal	667	8557	0
Pathologic	0	0	748

The confusion matrix shows that Pathologic Myopia is classified without error, confirming strong sensitivity for the most critical category. Most of the remaining mistakes occur between the Normal and Myopia classes, where a subset of Normal images is assigned to Myopia. This pattern suggests that the model is conservative in its predictions and tends to favor detecting myopic features rather than overlooking them.

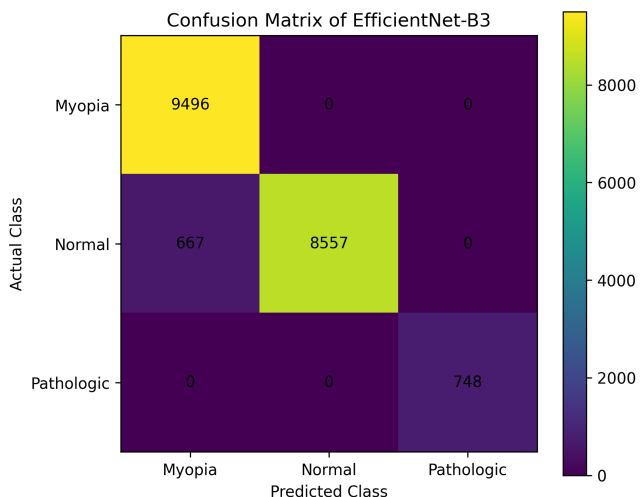


Figure 3: Confusion matrix heatmap of EfficientNet-B3. The model correctly identifies all Pathologic Myopia cases, while the remaining errors mainly occur between Normal and Myopia classes.

4.5 Qualitative Analysis using Grad-CAM

To better understand the basis of the model’s predictions, Grad-CAM visualizations are examined.

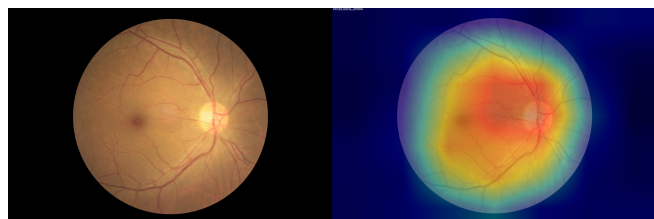


Figure 4: Grad-CAM visualization showing the image regions that influence the model prediction. The highlighted areas correspond to clinically meaningful retinal structures.

As illustrated in Fig. 4, the model tends to attend to anatomically relevant regions such as the optic disc and surrounding retinal structures. This behavior suggests that the network is using meaningful retinal cues rather than spurious background information, which supports the reliability of its predictions.

4.6 Comparative Analysis with ResNet50

To evaluate the benefit of the proposed architecture against a conventional residual network, EfficientNet-B3 is compared with ResNet50 under the same experimental setup. Both models are trained and tested using the same dataset split, preprocessing pipeline, and evaluation protocol, allowing a fair assessment of architectural effectiveness.

The performance of ResNet50 is summarized in Table 8. Although the model provides a reasonable baseline, its results indicate a lower level of class separation than the proposed EfficientNet-B3 model.

Table 8: Overall performance metrics of ResNet50

Metric	Value
Accuracy	0.86
Macro Precision	0.89
Macro Recall	0.91
Macro F1-score	0.89
Weighted Precision	0.89
Weighted Recall	0.86
Weighted F1-score	0.86

The class-wise behavior of ResNet50 is given in Table 9. The results show that while Pathologic Myopia is detected reliably, the Normal class remains more difficult to distinguish from Myopia. This is further reflected in the confusion matrix in Table 10, where a considerable number of Normal images are misclassified as Myopia.

Table 9: Class-wise performance metrics of ResNet50

Class	Precision	Recall	F1-score	Support
Myopia	0.79	1.00	0.88	9496
Normal	1.00	0.72	0.84	9224
Pathologic Myopia	0.89	1.00	0.94	475

Table 10: Confusion matrix of ResNet50

Actual / Predicted	Myopia	Normal	Pathologic Myopia
Myopia	9494	2	0
Normal	2534	6634	56
Pathologic Myopia	0	0	475

In comparison, EfficientNet-B3 achieves substantially better classification performance. Table 11 shows that the proposed model improves both accuracy and macro F1-score over ResNet50, while preserving perfect recall for Pathologic Myopia.

Table 11: Performance gain of EfficientNet-B3 over ResNet50

Metric	ResNet50	EfficientNet-B3	Improvement
Accuracy	0.8200	0.9608	+0.1408
Macro F1-score	0.8900	0.9730	+0.0830
PM Recall	1.0000	1.0000	0.0000

The improvement can be attributed to the stronger representational capacity of EfficientNet-B3 together with its compound scaling design, which appears better suited to capturing fine retinal patterns in this task.

4.7 Comparative Analysis with EfficientNet-B0

To further evaluate the effectiveness of the proposed model, EfficientNet-B3 is compared with EfficientNet-B0, which serves as the earlier EfficientNet-based baseline in this study. Both models are trained and evaluated under the same dataset split, preprocessing pipeline, and experimental protocol, ensuring that the observed differences are attributable to architectural and input-resolution changes rather than training variation.

The overall performance of EfficientNet-B0 is summarized in Table 12. Although the model achieves respectable performance, it still shows room for improvement in separating visually similar classes.

Table 12: Overall performance metrics of EfficientNet-B0

Metric	Value
Accuracy	0.893
Macro Precision	0.936
Macro Recall	0.925
Macro F1-score	0.923
Weighted Precision	0.912
Weighted Recall	0.893
Weighted F1-score	0.892

The class-wise behavior of EfficientNet-B0 is reported in Table 13. The results indicate strong sensitivity for Myopia and Pathologic Myopia, but reduced recall for Normal due to confusion with Myopia. This is also evident in the confusion matrix shown in Table 14, where a substantial number of Normal samples are predicted as Myopia.

Table 13: Class-wise performance metrics of EfficientNet-B0

Class	Precision	Recall	F1-score	Support
Myopia	0.821	1.000	0.902	9496
Normal	1.000	0.775	0.873	9224
Pathologic Myopia	0.988	1.000	0.994	748

Table 14: Confusion matrix of EfficientNet-B0

Actual / Predicted	Myopia	Normal	Pathologic
Myopia	9493	3	0
Normal	2070	7145	9
Pathologic	0	0	748

In comparison, EfficientNet-B3 shows a clear improvement across the main performance metrics. Table 15 highlights the gain achieved over EfficientNet-B0, while Table 16 provides the overall comparison among all three models evaluated in this study.

Table 15: Performance gain of EfficientNet-B3 over EfficientNet-B0

Metric	Improvement
Accuracy	+6.78 percentage points
Weighted F1-score	+0.069
Macro F1-score	+0.050
Myopia Precision	+0.105
Myopia F1-score	+0.059
Normal Recall	+0.143

Table 16: Performance comparison of ResNet50, EfficientNet-B0, and EfficientNet-B3

Model	Accuracy	Macro F1	PM Recall
ResNet50	0.8200	0.8900	1.0000
EfficientNet-B0	0.8930	0.9230	1.0000
EfficientNet-B3	0.9608	0.9730	1.0000

The class-wise results show that the most noticeable gain is observed in the Normal class, where EfficientNet-B0 has

lower recall due to confusion with Myopia. In contrast, EfficientNet-B3 reduces this misclassification significantly while maintaining perfect recall for Pathologic Myopia. This balance between improved general accuracy and strong sensitivity for the clinically critical class makes EfficientNet-B3 more suitable for the present three-class classification task.

The improvement appears to be linked to the compound scaling strategy of EfficientNet-B3, which provides a better balance of depth, width, and resolution for capturing subtle retinal variations.

4.8 Deployment Performance

The proposed model is deployed through a local Django-based web application intended for practical screening scenarios. After a retinal image is uploaded, the system returns the predicted class, confidence score, Grad-CAM visualization, and a downloadable PDF report.

The average inference time is approximately 2–3 seconds per image, although this can vary slightly depending on the input size and the hardware environment. By integrating explanation into the prediction pipeline, the application provides results that are both responsive and interpretable at the point of use.

4.9 Summary of Findings

The experimental results indicate that the proposed EfficientNet-B3 model performs strongly on the multi-class retinal classification task. In particular, it achieves perfect sensitivity for Pathologic Myopia while also improving discrimination between Normal and Myopia compared with the earlier baseline models.

Although some overlap remains between Normal and Myopia, the overall performance is consistent and high across the test set. Combined with Grad-CAM-based explanation and web-based deployment, the framework demonstrates practical potential as a clinical decision-support tool.

5 Discussion

This section examines the experimental results in greater depth and discusses their implications for multi-class retinal image classification. Particular attention is given to the effect of class imbalance, the distribution of prediction errors, and the comparative behavior of the evaluated architectures. The discussion further considers the clinical significance of the proposed EfficientNet-B3-based framework.

5.1 Overall Performance Interpretation

The proposed EfficientNet-B3 model achieves an overall accuracy of 96.08% with a macro F1-score of 0.973, indicating strong and well-balanced performance across all classes. The

close agreement between macro and weighted F1-scores suggests that the model remains robust despite the uneven class distribution in the dataset.

The macro recall of 0.972 further shows that performance is not concentrated only on the majority classes. Instead, the model demonstrates consistent recognition ability across Normal, Myopia, and Pathologic Myopia, which is important for a clinically relevant screening task.

5.2 Impact of Class Imbalance Handling

Class imbalance is addressed through a weighted cross-entropy loss function, where classes contribute differently to the training objective according to their relative frequency. This design gives greater emphasis to less frequent cases and helps the model remain attentive to clinically important minority samples.

The effect of this strategy is visible in the final results, particularly in the perfect recall achieved for Pathologic Myopia. From a clinical perspective, this is especially valuable because missing pathological cases may delay treatment and increase the risk of vision loss.

Overall, the re-weighting strategy helps the model prioritize sensitivity for critical cases while still preserving strong performance across the full class set. This indicates that loss re-weighting is an effective component of the proposed framework.

5.3 Error Distribution and Confusion Analysis

The confusion matrix shows that most remaining errors occur between the Normal and Myopia classes. A portion of Normal images is predicted as Myopia, which suggests that the model detects subtle retinal patterns associated with myopic change even when the ground truth label is Normal.

This behavior lowers the recall of the Normal class relative to the other categories, but it also reflects a conservative screening tendency. In medical classification, such a bias can be preferable to missing early signs of disease, especially when the goal is early detection rather than strict rejection of borderline cases.

The overlap between Normal and Myopia classes indicates that these categories are visually close and not always easy to separate, even for a deep model. The results suggest that the proposed architecture captures clinically meaningful differences well, but that some boundary cases remain inherently difficult.

5.4 Effectiveness of EfficientNet Architecture

The comparative study with ResNet50 highlights the strengths of the proposed EfficientNet-B3 architecture in this task. Although both models achieve perfect recall for Pathologic Myopia, EfficientNet-B3 delivers superior overall accuracy and a higher macro F1-score, indicating more balanced

classification performance.

ResNet50 shows greater confusion between the Normal and Myopia classes, which suggests weaker discrimination of subtle retinal differences. By comparison, EfficientNet's compound scaling approach provides a more effective balance of depth, width, and input resolution, enabling the model to capture finer retinal patterns with greater consistency.

This architectural advantage contributes to better generalization and more stable decision boundaries, particularly in cases where the visual distinction between classes is limited.

5.5 Interpretability and Clinical Relevance

Grad-CAM visualizations offer qualitative support for the model's predictions. The heatmaps consistently emphasize clinically meaningful regions, including the optic disc and nearby retinal structures.

This agreement between the highlighted regions and relevant anatomical areas suggests that the model is learning useful retinal representations rather than relying on irrelevant background cues. Such transparency is important in medical applications because it helps clinicians understand and trust the system's output.

5.6 Deployment and Practical Considerations

The deployed system demonstrates practical feasibility, with an average inference time of approximately 2–3 seconds per image. This is appropriate for real-time or near real-time screening scenarios.

The inclusion of Grad-CAM visual output and automated PDF report generation improves usability by producing results that are both interpretable and easy to share. In addition, the use of a local SQLite database with role-based access supports controlled handling of patient data, which is relevant for privacy-aware medical deployment.

5.7 Limitations and Future Scope

Despite the strong results, some limitations remain. The lower recall for the Normal class indicates that distinguishing Normal images from early myopic changes is still challenging.

The relatively limited number of Pathologic Myopia samples may also restrict the model's exposure to broader pathological variation. Future work can address this by expanding the dataset, exploring more advanced feature extraction methods, and evaluating ensemble-based approaches.

Further validation on external datasets and with additional performance measures would also strengthen confidence in the generalizability of the proposed system.

6 Conclusion

This work presents a deep learning framework for multi-class retinal fundus image classification into Normal, Myopia, and Pathologic Myopia. The proposed approach is based on EfficientNet-B3 and incorporates a two-phase transfer learning strategy, optimization techniques designed to improve generalization, and Grad-CAM for visual explanation of model predictions.

The experimental results show that the proposed model achieves strong performance on the test set, with an accuracy of 96.08% and a macro F1-score of 0.973. In particular, the model attains perfect recall for Pathologic Myopia, which is important for reliably identifying clinically critical cases. The comparative study with ResNet50 demonstrates that EfficientNet-B3 provides better class separation and stronger overall classification performance.

Grad-CAM analysis shows that the model focuses on clinically meaningful retinal regions, which improves the transparency of the system. In addition, deployment through a Django-based web application demonstrates that the framework is practical for real-time prediction, visualization, and report generation in a screening setting.

Although the model performs well, some confusion remains between Normal and Myopia classes. Future work may address this by expanding the dataset, exploring alternative training strategies, and testing the framework on external datasets to further improve robustness and generalization.

Acknowledgement

The authors would like to express their sincere gratitude to their project guide and faculty members for their guidance, encouragement, and valuable feedback throughout the development of this work.

We also acknowledge the use of publicly available retinal datasets, including the PALM dataset, which were essential for training and evaluating the proposed model. In addition, we thank our peers and collaborators for their support during experimentation and system development.

Finally, we are grateful for the institutional resources and computational support that made this research possible.

References

- [1] Qi et al., "Prediction of Myopia Onset Using Deep Learning with Fundus Images and Metadata," 2024.
- [2] Xing et al., "Lightweight Attention-Based Network for Myopia Severity Classification," 2025.
- [3] Li et al., "Deep Learning-Based Binary Classification of Myopia Using Retinal Images," 2023.
- [4] Wang et al., "Pathologic Myopia Detection Using Ensemble Deep Learning Models," 2025.
- [5] Chen et al., "Dual Autoencoder-Based Deep Learning Framework for Pathologic Myopia Identification," 2026.
- [6] Zhao et al., "EfficientNet-Based Multi-Class Retinal Disease Classification," 2025.
- [7] Kumar et al., "Myopia Detection in Indian Population Using Transfer Learning with VGG16," 2024.
- [8] Patel et al., "Multi-Modal Deep Learning for Retinal Disease Detection Using OCT and Fundus Images," 2025.
- [9] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of ICML*, 2019.
- [10] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of CVPR*, 2016.
- [11] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of ICCV*, 2017.
- [12] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *NeurIPS*, 2019.
- [13] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 2011.
- [14] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [15] A. Clark, "Pillow (Python Imaging Library Fork)," 2015.