



EfficientNet-B0-Based Deep Learning Framework for Four-Class Cataract Severity Classification Using Fundus Images

Mahavir Shantinath Kasar^{1,2*} , Pradeep B. Mane¹ , Dnyandeo K. Shedge¹ 

¹ Department of Electronics & Telecommunication Engineering, AISSMS Institute of Information Technology, Savitribai Phule Pune University, Pune 411001, India

² Department of Electronics & Telecommunication Engineering, Bharati Vidyapeeth's College of Engineering for Women, Pune 411043, India

Corresponding Author Email: Mahavir.Kasar@bharatividyaapeeth.edu

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310913>

ABSTRACT

Received: 11 March 2026

Revised: 15 August 2026

Accepted: 2 September 2026

Available online: 30 September 2026

Keywords:

cataract severity classification, fundus image analysis, deep learning, EfficientNet, medical image classification, transfer learning

Cataract remains one of the leading causes of avoidable blindness worldwide, and accurate severity assessment is essential for timely screening and clinical management. However, access to ophthalmic specialists remains limited in many resource-constrained regions, creating a need for reliable computer-aided analysis systems. This study develops an EfficientNet-B0-based deep learning (DL) framework for four-class cataract severity classification using fundus images. The proposed framework classifies images into Normal, Mild, Moderate, and Severe categories and systematically investigates the influence of different training strategies under a patient-independent evaluation protocol. Experiments were conducted using an Ocular Disease Intelligent Recognition 5K (ODIR-5K)-derived dataset containing 3,868 fundus images from 2,029 patients. Three EfficientNet-B0 configurations, including Frozen Baseline, Severe-Focused, and Final Frozen models, were evaluated using overall performance, class-wise metrics, discrimination measures, statistical analysis, and computational characteristics. The Final Frozen EfficientNet-B0 achieved an accuracy of 88.38%, weighted F1-score of 89.29%, macro F1-score of 56.33%, and macro receiver operating characteristic-area under the curve (ROC-AUC) of 91.60% on an independent test set of 585 images. The results demonstrate that EfficientNet-B0 provides a feasible approach for multiclass cataract severity classification, while also highlighting the influence of class imbalance on minority-category recognition. The findings emphasize the importance of patient-independent evaluation and comprehensive performance assessment rather than relying solely on overall accuracy. Further validation with larger and more balanced datasets is required before clinical implementation.

1. INTRODUCTION

One of the most common causes of impairment of vision and blindness all over the world is cataract [1]. This develops gradually as the lens becomes cloudy and the quality of vision and activities in the normal course of daily living decrease [2]. A trained ophthalmologist is typically necessary for clinical assessment, and specialist ophthalmic service may be less readily available in rural and resource-poor areas [3]. The difficulties have spurred computer-aided cataract screening and cataract severity assessment [4]. Figure 1 illustrates the global burden of vision impairment and cataract, based on estimates reported by the World Health Organization [5].

Fundus photography is widely used in ophthalmic screening and provides an accessible image-based modality for automated analysis. Although cataract is primarily associated with lens opacity, it can influence the appearance and quality of retinal photographs. Consequently, previous studies have investigated fundus images for automated cataract detection and severity grading. Global estimates for adults aged 50 years and older identify cataract as a major contributor to blindness,

as shown in Figure 2 [6]. In India, untreated cataract was reported as the leading cause of blindness among adults aged 50 years and older, accounting for 66.2% of blindness cases in the National Blindness and Visual Impairment Survey 2015–2019; the distribution of major causes is shown in Figure 3 [7].

The Ocular Disease Intelligent Recognition 5K (ODIR-5K) dataset comprises 10,000 color fundus photographs from 5,000 patients, including images from both eyes and diagnostic information for multiple ocular conditions [8]. Class imbalance is an important challenge in multi-class medical image classification because overall accuracy may not adequately reflect minority-class performance. Therefore, the present study uses class-wise and macro-averaged measures alongside overall accuracy and interprets severe-class performance cautiously because of its limited representation in the independent test set.

Researchers have investigated fundus photography for automated cataract assessment using conventional machine learning (ML) and deep learning (DL) approaches [9–13]. Convolutional Neural Network (CNN)-based and transfer learning (TL) methods have further demonstrated the

feasibility of cataract detection and severity grading from ophthalmic images [14-17]. However, differences in datasets, severity definitions, class distributions, model architectures, and evaluation protocols limit direct comparison between reported results. Reliable evaluation of minority severe-cataract cases also remains challenging.

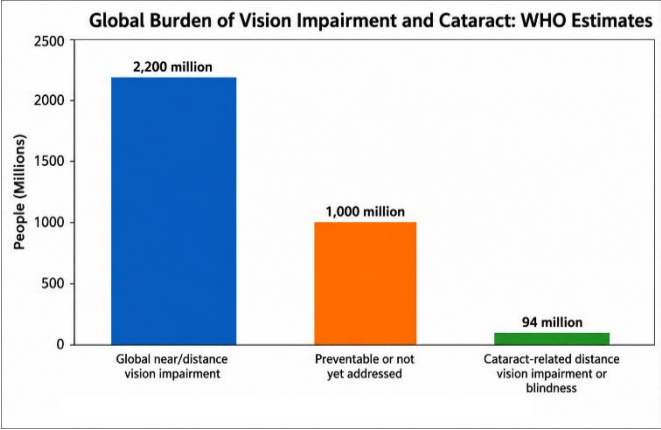


Figure 1. Global burden of vision impairment and cataract: WHO estimates (2026)
Source: World Health Organization (WHO), Blindness and vision impairment, Fact Sheet, 2026.

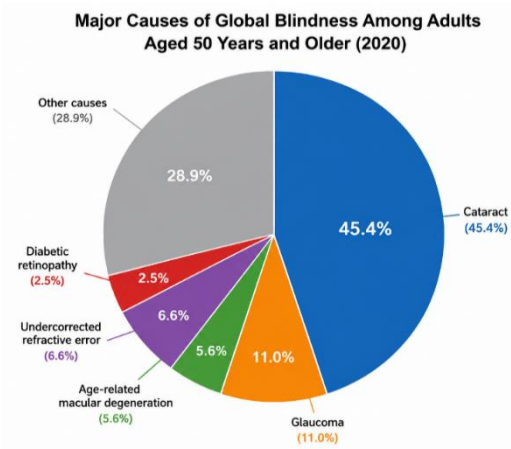


Figure 2. Global causes of blindness among adults aged 50 years and older in 2020
Source: GBD 2019 Blindness and Vision Impairment Collaborators [6].

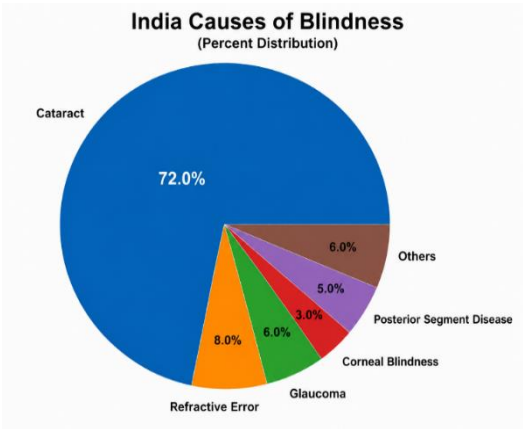


Figure 3. Leading causes of blindness among adults aged 50 years and older in India
Source: The National Blindness and Visual Impairment Survey 2015–2019 [7].

EfficientNet-B0 provides a balance between classification performance and computational requirements through compound scaling of network depth, width, and input resolution [18]. Accordingly, this study investigates EfficientNet-B0 for four-class cataract severity classification from fundus images. Three configurations—Frozen Baseline EfficientNet-B0, Severe-Focused EfficientNet-B0, and Final Frozen EfficientNet-B0—are evaluated using the same independent test observations. Their performance is examined using overall, macro, weighted, class-wise, agreement-based, statistical, and computational analyses. This research highlights an EfficientNet-B0-based approach to classify fundus images into four classes: Normal, Mild, Moderate, and Severe cataract.

The study focuses on independent-test and minority-class analysis, with particular attention to the underrepresented Severe category. Statistical and computational evaluation includes paired statistical testing and measurements of parameter count, FLOPs, model size, inference time, and throughput. Finally, the Final Frozen EfficientNet-B0 configuration is selected based on its overall predictive performance, class-wise behavior, and computational characteristics while explicitly recognizing the uncertainty associated with the limited number of Severe cases.

In Section 2, the existing literature in the area of cataract analysis is reviewed, and the research gap to be addressed by this work is set. Section 3 explains the methodological approach, which includes data preparation, model configurations, network design, training methods, and performance evaluation. The experimental results are presented in Section 4, with an analysis of the results of the various EfficientNet-B0 architectures. The paper ends with a summary of the main findings, a discussion of the contribution and limitation(s) of the study, and suggested directions for future studies in Section 5.

2. LITERATURE REVIEW

2.1 Ocular Disease Intelligent Recognition 5K and dataset considerations

Advancements in automated approaches for retinal image analysis have been facilitated by the availability of large and diverse datasets of ophthalmic images. The ODIR-5K dataset includes fundus images from several individuals, where each image is from both the left and right eyes, and their diagnostic information for various eye diseases. Because of its diversity and wide coverage of ophthalmic anomalies, ODIR-5K has proven to be a valuable tool for computer-aided identification and classification of eye diseases research.

Variability in ophthalmic photographs is relevant to DL research because image quality, illumination, contrast, camera characteristics, and acquisition conditions can affect model performance. Consequently, preprocessing and evaluation procedures should be carefully defined when developing models using ODIR-5K-derived data.

2.2 Conventional and deep learning approaches for cataract assessment

Early computer-aided cataract assessment approaches primarily relied on manually designed image features combined with conventional ML classifiers. These methods generally extracted characteristics such as texture, intensity,

wavelet coefficients, edge information, or other statistical descriptors from ocular images and subsequently used classifiers to distinguish cataract categories. Guo et al. [9] investigated image features extracted from fundus photographs for automated cataract classification and grading, demonstrating that information available in retinal images could be used for computer-aided cataract assessment. Yang et al. [10] subsequently investigated an ensemble learning framework for cataract detection and grading, combining multiple feature representations and classifiers to improve classification performance.

Hierarchical classification strategies were also investigated to address the difficulty of distinguishing different cataract severity levels. Cao et al. [11] proposed an approach based on an improved Haar-wavelet representation and hierarchical classification, where multiclass cataract grading was decomposed into a sequence of classification decisions. Pathak et al. [12] investigated texture-based cataract detection and grading with consideration of portable and tele-ophthalmology applications. However, the imaging modality was not consistent across all studies, with some investigations using retinal fundus photographs and others using lens or retro-illumination images. Therefore, their results cannot always be directly compared. Gao et al. [13] proposed a computer-aided cataract detection method using enhanced texture features extracted from retro-illumination lens images and reported an accuracy of 84.8% on a clinical database containing 4,545 pairs of images. However, the use of retro-illumination lens images limits direct comparison with fundus-image-based approaches.

Shin et al. [14] demonstrated the applicability of deep CNNs and TL to medical image analysis, establishing the usefulness of pretrained deep architectures when medical datasets are relatively limited.

Deep-learning methods have subsequently been investigated for cataract detection and grading. Pratap and Kokil [15] investigated deep transfer learning for computer-aided cataract diagnosis. Imran et al. [16] developed a hybrid convolutional and recurrent neural network for four-class cataract classification, while Zhang et al. [17] investigated automatic cataract grading using deep learning with multi-feature fusion and stacking, combining deep CNN features with a Random Forest classifier for cataract detection and grading. Tan and Le [18] introduced EfficientNet-B0 based on compound scaling of network depth, width, and input resolution. Pratap and Kokil [19] investigated efficient network selection for computer-aided cataract diagnosis under noisy fundus-image conditions, while Khan et al. [20] investigated cataract detection using a VGG-19-based convolutional neural network.

More recent research has focused on improved feature representation and multiclass cataract assessment. Xu et al. [21] proposed a global-local representation approach for cataract grading, while deep learning-based cataract type classification and severity grading using fundus photographs have also been investigated [22]. Yadav and Yadav [23, 24] investigated severity classification using retinal fundus images and considered four categories: Normal, Mild, Moderate, and Severe. Junayed et al. [25] further investigated automated cataract detection from fundus images using a DL framework. Wu et al. [26] investigated an artificial-intelligence-based approach for cataract diagnosis with emphasis on anti-interference capability, whereas Xiong et al. [27] examined image blurriness associated with vitreous opacity as a factor

relevant to cataract diagnosis from retinal images.

Overall, the literature demonstrates a progression from handcrafted features and conventional classifiers to CNNs, TL, hybrid architectures, and multiclass severity classification. However, differences in datasets, imaging modalities, preprocessing procedures, severity definitions, class distributions, and evaluation protocols limit direct comparison among reported studies.

2.3 Fundus-based cataract severity classification

Cataract assessment from fundus photographs is relevant to screening-oriented applications because fundus cameras are widely used for retinal imaging. While the appearance of an opaque lens is the hallmark of cataract, changes in the cataract can affect the appearance of the retina and the quality of the photograph, thus promoting the application of fundus images for automated detection and severity assessment.

Several studies have moved beyond binary classification toward multiclass cataract severity grading. Other studies have similarly investigated multiclass grading using DL, hierarchical, or ensemble approaches.

The present study uses an ODIR-5K-derived fundus-image dataset, providing the image source for the four-class experimental framework described in Section 3.

The distinction between detection and severity grading is important. Unlike binary detection, severity classification must distinguish between adjacent categories such as mild, moderate, and severe. Therefore, four-class evaluation requires class-wise analysis in addition to overall accuracy.

Several survey studies have reviewed the broader development of automated cataract detection and classification. Morales-Lopez et al. [28] surveyed cataract detection and classification systems based on computational intelligence. Shaheen et al. [29] reviewed the application of artificial intelligence in cataract management, including automated diagnosis and grading using ophthalmic images. Zhang et al. [30] provided a survey of ML approaches for cataract classification and grading across ophthalmic imaging modalities.

2.4 Class imbalance and minority-class evaluation

Class imbalance is an important challenge in multiclass medical-image classification. When one category substantially outnumbers others, a model may become biased toward the majority class, and overall accuracy may not adequately represent minority-class performance.

This issue is particularly important in the present study, where the independent test set contains 511 Normal, 31 Mild, 38 Moderate, and only 5 Severe cases. Thus, the severe category represents a very small proportion of the test observations.

Therefore, evaluation is not restricted to accuracy and includes class-wise, macro-averaged, weighted, discrimination, and agreement-based measures. The Final Frozen EfficientNet-B0 was selected as the final configuration based on the overall balance of predictive performance, class-wise behavior, statistical analysis, and computational characteristics, while explicitly considering the uncertainty associated with the limited number of Severe cases in the independent test set.

Additional conventional and multimodal approaches include texture-based classification by Fuadah et al. [31],

Hough-transform-based cataract localization and grading in slit-lamp images by Liu et al. [32], and deep-learning-based cataract detection using ocular B-ultrasound images by Wang et al. [33]. The development of CNNs introduced a data-driven alternative to handcrafted feature extraction. Another important advancement in DL is residual learning, in which shortcut connections enable effective information flow across deeper network layers and support the training of very deep convolutional architectures [34]. Grad-CAM has been widely used to provide visual explanations of deep neural network predictions by highlighting image regions that contribute to the model's output [35]. In related work, Kasar and Shedge [36] presented a survey of cataract detection and classification systems, followed by a ResNet-based approach for cataract detection and classification [37]. The present study extends this work through EfficientNet-B0-based four-class cataract severity classification using a patient-independent evaluation protocol.

The reviewed literature demonstrates the feasibility of automated cataract detection and severity classification; however, dataset variability, severe-class imbalance, differences in evaluation protocols, and computational efficiency remain important challenges.

2.5 EfficientNet-B0 for computationally efficient classification

Medical imaging DL models can be memory-intensive and consume high computational power; thus, the efficiency of these models is important for real-world application in resource-constrained healthcare settings. Compound scaling is a systematic method of scaling up the network capacity while

maintaining performance and computational cost. EfficientNet-B0 is the backbone architecture of this family, and it is a relatively compact backbone for TL applications. In this study, EfficientNet-B0 is employed to classify the severity of cataract in fundus photographs into four classes. In addition to classification performance, its computational characteristics such as the number of parameters, FLOPs, model size, inference time, and throughput are also evaluated.

2.6 Comparative analysis of existing studies

The second step was to critically examine previous studies. TL, class-balancing strategies, and advanced DL architectures have been explored for automated cataract detection and severity classification, yielding promising results. However, differences in datasets, class definitions, class distributions, preprocessing methods, and evaluation protocols make direct comparison of reported results difficult. Class-wise evaluation and independent testing are particularly important for four-class cataract classification because of severe class imbalance.

A summary of representative studies related to fundus-based detection and severity classification of cataracts is summarized in Table 1. Differences in the datasets, in the definition of severity, in the distribution of classes, and in how the results are evaluated, however, prevent direct comparison of reported results.

The motivation for the present study lies in these gaps: it systematically assesses various EfficientNet-B0 configurations under an identical patient-independent experimental protocol, focusing on overall performance as well as minority-class behavior and statistical comparison, along with computational characteristics.

Table 1. Comparison of representative fundus image-based cataract studies

Ref.	Focus Area	Dataset	Model Used	Accuracy (%)	Key Contribution
[9]	Cataract detection and grading	Fundus images	Handcrafted features + ML	77.10	Automated cataract grading using image features
[10]	Detection and four-level grading	Fundus images	Ensemble learning	84.50	Combined feature representations and classifiers
[11]	Four-class severity grading	Retinal/fundus images	Haar-wavelet + hierarchical classifier	85.98	Hierarchical four-class severity classification
[12]	Cataract detection and grading	Fundus images	Texture features + ML	83.80	Portable and tele-ophthalmology-oriented assessment
[15]	Four-stage classification	Fundus images	Transfer learning + SVM	92.91	Normal, Mild, Moderate, and Severe classification
[16]	Four-class cataract classification	Fundus images	Hybrid CNN-RNN + transfer learning	97.39	Hybrid convolutional and recurrent feature representation for four-class classification
[21]	Cataract grading	Fundus images	Global-local CNN	83.47	Combined global and local feature representations
[22]	Cataract type and severity grading	Fundus images	Dual-Stream Cataract Evaluation Network (DCEN)	97.03	Joint cataract type classification and severity grading using fundus photographs
[23]	Four-class severity classification	Retinal fundus images	DL	93.10	Four-stage cataract severity assessment
Present study	Four-class cataract severity classification	ODIR-5K-derived fundus images	Proposed Frozen EfficientNet-B0 Framework	88.38	Independent 585-image test set, severe-class evaluation, statistical and computational analysis

Note: CNN = Convolutional Neural Network, ML = machine learning, DL = deep learning, TL = transfer learning, ODIR-5K = Ocular Disease Intelligent Recognition 5K, SVM = Support Vector Machine.

3. METHODOLOGY

The classification task includes Normal, Mild, Moderate, and Severe categories. Three configurations—Frozen Baseline

EfficientNet-B0, Severe-Focused EfficientNet-B0, and Final Frozen EfficientNet-B0—are evaluated under a common experimental protocol using the same independent test set. Training and validation data are used for model development

and parameter selection, while the independent test set is reserved for final evaluation to prevent test information from influencing model development. Performance is assessed through predictive, class-wise, statistical, and computational analyses.

3.1 Dataset and class distribution

The present study uses a curated ODIR-5K-derived fundus-image dataset for four-class cataract severity classification. The final dataset contains 3,868 labelled fundus images from 2,029 unique patients. The four-class labels were established during dataset curation based on the available diagnostic information and predefined categorization criteria used for the experimental dataset. The labels were used as experimental categories and were not independently established through a separate clinical grading procedure or ophthalmologist validation. Duplicate, invalid, and unusable samples were removed during dataset preparation.

Let the complete dataset be represented as:

$$D = \{(x_i, y_i)\}_{i=1}^N \tag{1}$$

where, x_i denotes the i -th fundus image, y_i denotes its corresponding severity class, and $N = 3868$. The resulting

dataset comprised 3,466 Normal, 195 Mild, 179 Moderate, and 28 Severe images, indicating substantial class imbalance. A patient-independent partitioning strategy was adopted, ensuring that images from the same patient were not assigned to different subsets. The final partition included 2,698 training images from 1,437 patients, 585 validation images from 296 patients, and 585 independent test images from 296 patients, representing 2,029 unique patients overall, with no patient overlap among the subsets. The training and validation sets were used for model development, while the independent test set served as a common evaluation set for all three EfficientNet-B0 configurations. The proposed EfficientNet-B0-based framework for four-class cataract severity classification is shown in Figure 4. Table 2 summarizes the distribution of images across the training, validation, and independent test subsets.

No resampling or augmentation was applied to the validation or independent test subsets. The independent test set contained 511 Normal, 31 Mild, 38 Moderate, and 5 Severe images. Owing to the limited representation of the Severe category, class-wise and macro-averaged metrics were evaluated alongside overall accuracy, and the Severe-class results were interpreted cautiously. Representative fundus images from the four severity categories are presented in Figure 5.

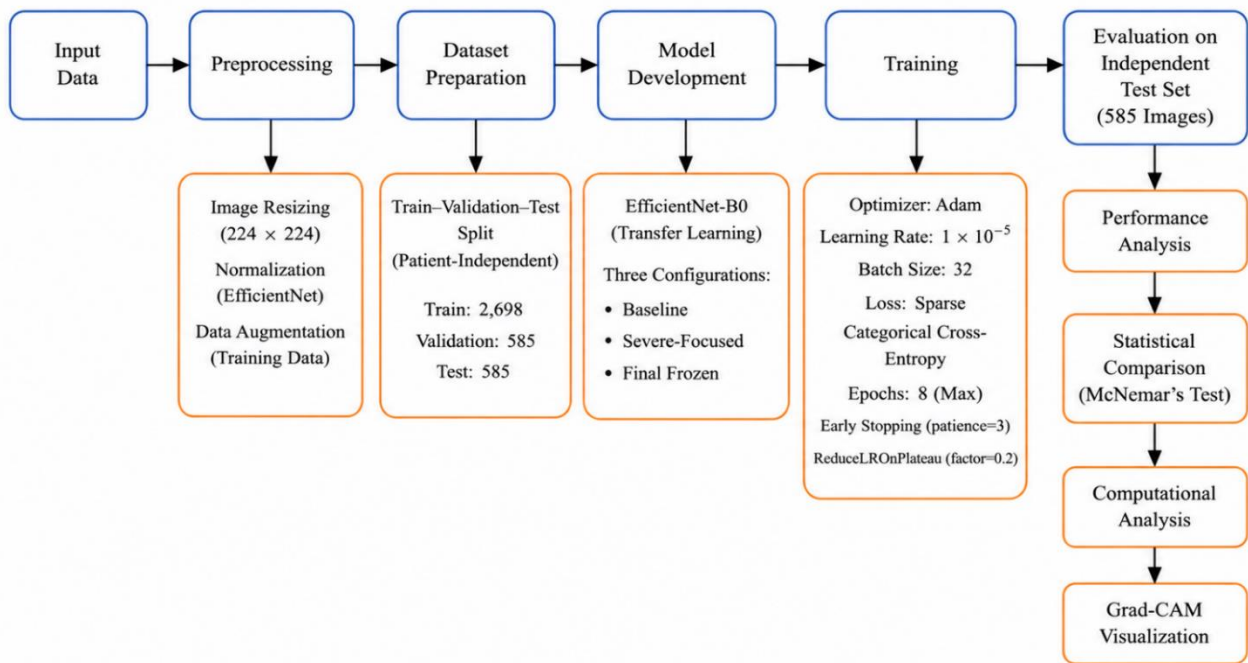


Figure 4. Proposed experimental framework for cataract severity classification

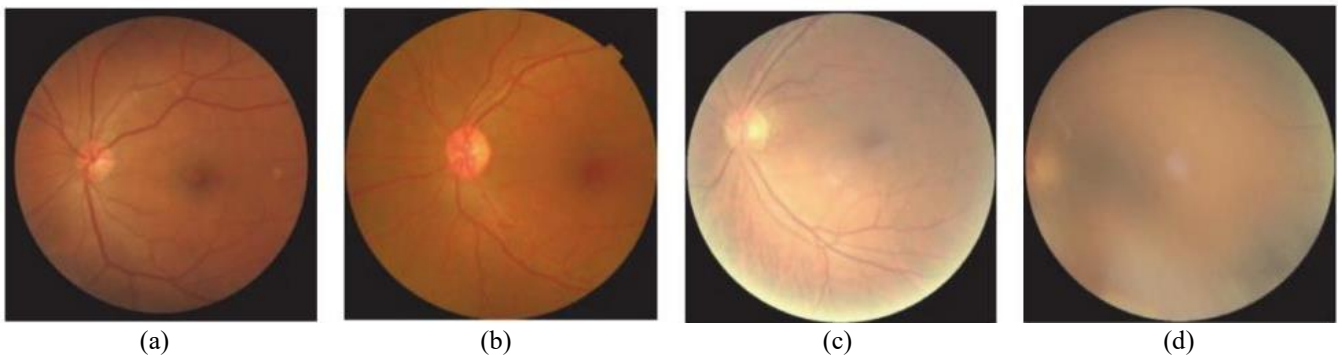


Figure 5. Fundus images: (a) Normal, (b) Mild, (c) Moderate, (d) Severe

Table 2. Dataset overview and class distribution

Category	Training	Validation	Independent Test	Total	Description
Normal	2,444	511	511	3,466	Reference category representing fundus images assigned to the Normal experimental class. This category constitutes the majority of the dataset.
Mild	133	31	31	195	Lower-severity experimental category used to evaluate the model's ability to distinguish Mild cases from the remaining categories.
Moderate	103	38	38	179	Intermediate experimental category used to evaluate discrimination of Moderate cases from adjacent severity categories.
Severe	18	5	5	28	Highest-severity experimental category and the least represented class. Only five Severe images are included in the independent test set.
Total	2,698	585	585	3,868	Complete image set used for model training, validation, and patient-independent evaluation.

3.2 Image quality assessment and selection

To further characterize the quality of the RGB fundus images, Natural Image Quality Evaluator (NIQE) and Perception-based Image Quality Evaluator (PIQE) were used as no-reference image quality measures. In general, lower values of these metrics correspond to comparatively better perceptual image quality. For the quality assessment performed in this study, images satisfying both NIQE and PIQE were categorized within the selected acceptable-quality range, while samples not satisfying these criteria were treated as comparatively lower-quality images. The resulting quality-score distributions and selection criteria are illustrated in Figure 6, with Figure 6(a) presenting the threshold-based assessment and Figure 6(b) showing the corresponding NIQE–PIQE scatter distribution.

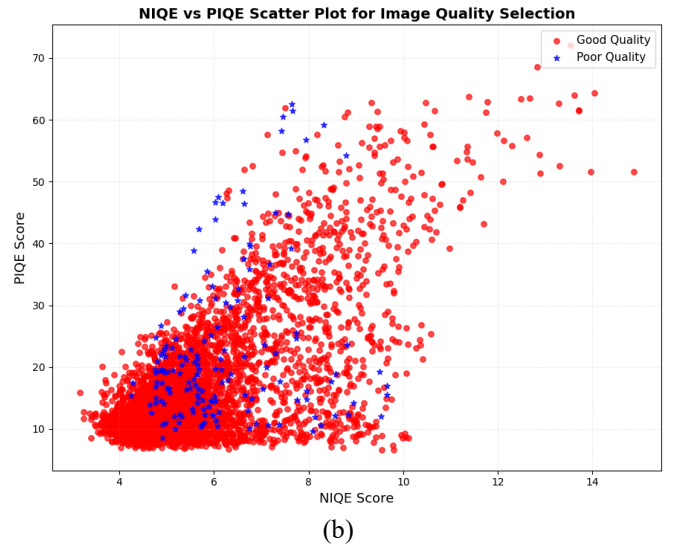


Figure 6. Resulting quality-score distributions and selection criteria: (a) threshold-based fundus image quality selection, (b) scatter plot showing the quality distribution of fundus images

As illustrated in Figure 7, region *P* represents the selected acceptable quality region in the NIQE–PIQE space. Images with coordinate points falling within region *P* satisfy the defined quality criteria, whereas those outside the region are considered comparatively poor-quality samples. This NIQE–PIQE analysis was used solely as an additional image quality assessment and visualization step and did not alter the patient-independent data partitioning, model training procedure, or the reported evaluation results of the Final Frozen EfficientNet-B0 configuration.

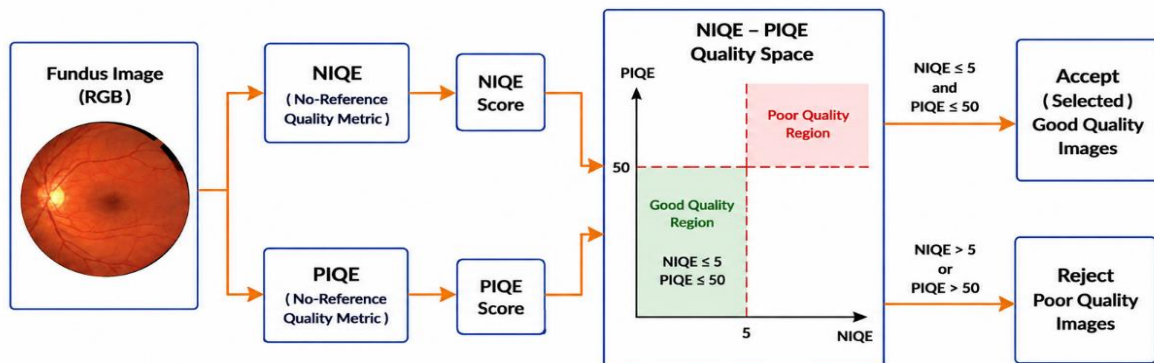


Figure 7. Image quality assessment analysis of the dataset

3.3 Image preprocessing

The fundus images were prepared using a consistent preprocessing pipeline before being provided to the EfficientNet-B0 model. All images were resized to 224×224 pixels to match the input dimensions of the backbone network and were converted into numerical tensors.

$$x_i^r = R(I_i, 224, 224, 3) \quad (2)$$

where, I_i is the original fundus image, and $R(\cdot)$ denotes the resizing operation.

Each input image is therefore represented as an RGB image with three color channels. The RGB representation and its individual R-, G-, and B-channel components are illustrated in Figure 8.

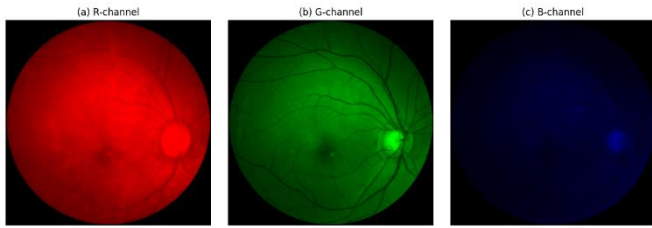


Figure 8. RGB components of a fundus image

Each input image is therefore represented as:

$$x_i^r \in R^{224 \times 224 \times 3}$$

The images were then processed using the preprocessing function associated with EfficientNet-B0. The four severity

categories were encoded as integer labels representing Normal, Mild, Moderate, and Severe.

During model development, data augmentation and controlled minority-class sampling were applied only to the training data to reduce the effect of class imbalance. Training data were shuffled and batched, while validation and test data were used in their original patient-independent form. Prefetching was applied to improve data-delivery efficiency during training and evaluation.

$$\tilde{x}_i = T(x_i; \theta_i) \quad (3)$$

where, $T(\cdot)$ represents the augmentation transformation, and θ_i denotes the transformation parameters. T includes the augmentation operations actually used in the final experiment, while validation and test images are not augmented.

Following image resizing and training-only augmentation, the preprocessed images were passed through the preprocessing function associated with EfficientNet-B0. Let x_p^i denote the resized and augmented image. The EfficientNet-B0 preprocessing operation can be represented as

$$x_i^e = P_{EffNet}(x_p^i) \quad (4)$$

where, $P_{EffNet}(\cdot)$ denotes the preprocessing function implemented for the EfficientNet-B0 input pipeline and x_i^e represents the resulting model-ready image.

This preprocessing pipeline provided a consistent input representation for all three EfficientNet-B0 configurations.

The validation and independent test sets remained unchanged throughout model development and final evaluation, ensuring that augmentation and minority-class sampling did not influence the reported test performance.

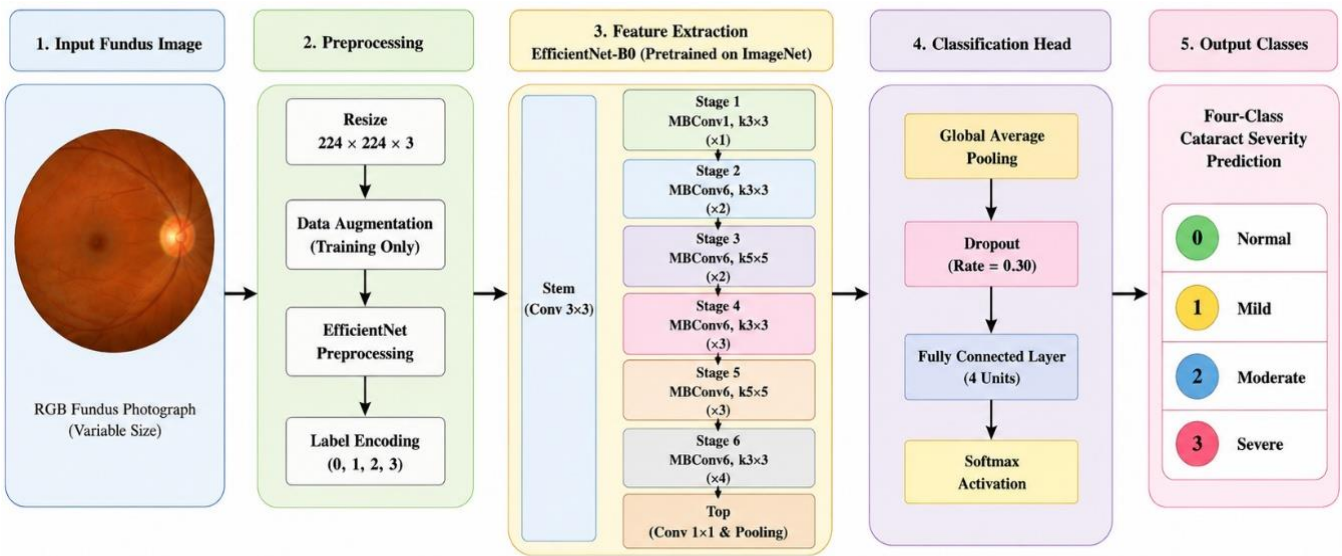


Figure 9. Overall architecture of the proposed Final Frozen EfficientNet-B0 framework

3.4 Proposed Final Frozen EfficientNet-B0 architecture

EfficientNet-B0 is used as the backbone network for the proposed four-class cataract severity classification framework. The architecture was selected because it provides an effective balance between feature representation capability and computational requirements through compound scaling of network depth, width, and input resolution. In the present

study, the ImageNet-pretrained EfficientNet-B0 backbone is adapted to classify fundus images into four severity categories: Normal, Mild, Moderate, and Severe. The overall architecture of the proposed Final Frozen EfficientNet-B0 framework is illustrated in Figure 9.

The input to the network is a preprocessed fundus image of size. The EfficientNet-B0 backbone progressively extracts hierarchical visual features from the input image. The feature

extraction process can be represented as:

$$F = f_{\theta}(X) \quad (5)$$

where, X represents the preprocessed fundus image, where f_{θ} denotes the EfficientNet-B0 feature extractor, and θ represents its learnable parameters, and F denotes the extracted feature maps.

The feature maps generated by the backbone are subsequently aggregated using Global Average Pooling (GAP), which summarizes the spatial information in each channel into a single feature value and produces a compact feature representation. For the k -th feature channel, this operation is expressed as:

$$g_k = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W F_k(i, j), \quad (6)$$

where, H and W denote the height and width of the feature map, $F_k(i, j)$ represents the activation at location (i, j) within the k -th channel and g_k is the resulting pooled feature value. By averaging the activations over the spatial dimensions, GAP produces a lower-dimensional representation while preserving the information encoded within each feature channel.

The resulting pooled features are subsequently passed through a dropout layer with a rate of 0.30, which is used during training to improve regularization and reduce the likelihood of overfitting.

The dropout operation can be represented as:

$$\tilde{g} = m \odot g \quad (7)$$

where, g represents the pooled feature vector, m denotes the dropout mask, $\odot$ represents element-wise multiplication, and $\tilde{g}$ is the regularized feature vector.

The resulting feature vector is subsequently passed to the final four-class classification layer.

A fundus image is resized to $224 \times 224 \times 3$, followed by data augmentation during training and EfficientNet-specific preprocessing. The preprocessed image is then passed through the EfficientNet-B0 backbone, consisting of an initial convolutional stem and multiple MBConv blocks for hierarchical feature extraction. The extracted features are processed using Global Average Pooling and dropout before being passed to a fully connected layer with four output units. A softmax activation function produces the predicted probabilities for the Normal, Mild, Moderate, and Severe categories.

3.5 Four-class classification

The final classification layer maps the regularized feature representation to the four cataract severity categories: Normal, Mild, Moderate, and Severe. The output logits for an input image are expressed as

$$z = W\tilde{g} + b \quad (8)$$

where, $\tilde{g}$ represents the regularized feature vector after dropout, where W and b correspond to the learnable weight matrix and bias vector of the final fully connected layer, respectively, and z denotes the resulting vector of class logits. These logits are subsequently transformed into class

probabilities using the softmax function:

The probability associated with the c -th class is computed by applying the softmax transformation to the output logits:

$$P(y_c | X) = \frac{e^{z_c}}{\sum_{j=0}^3 e^{z_j}}, c \in \{0, 1, 2, 3\} \quad (9)$$

The model assigns the input image to the class with the highest predicted probability among the Normal, Mild, Moderate, and Severe categories:

$$\hat{y} = \arg\max_c P(y = c | X) \quad (10)$$

3.6 Loss function and training strategy

3.6.1 Loss function

Since the classification task contains four mutually exclusive classes and integer class labels are used, sparse categorical cross-entropy is employed as the training loss.

The loss for N training samples is defined as

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \log(\hat{p}_{i, y_i}) \quad (11)$$

where, y_i is the ground-truth class label of the i -th image and $\hat{p}_{i, y_i}$ is the predicted probability assigned to its class. Optimization of this loss function promotes higher predicted probabilities for the correct cataract severity category.

3.6.2 Model optimization

The trainable model parameters are optimized by minimizing the classification loss:

$$\theta^* = \arg\min_{\theta} L_{CE}(\theta) \quad (12)$$

where, θ^* represents the optimized model parameters.

The models were optimized using sparse categorical cross-entropy as the objective function and the Adam optimizer. The optimization process was monitored using validation performance, with early stopping and adaptive learning-rate adjustment employed to support stable convergence and reduce unnecessary training.

3.6.3 Computational environment

Experiments were implemented using Python 3.12.13, TensorFlow 2.20.0, Keras, and scikit-learn, and executed in Google Colab using an NVIDIA Tesla T4 GPU.

3.6.4 Training and freezing strategy

The three EfficientNet-B0 configurations were developed using the same basic optimization framework but differed in their training and parameter-freezing strategies. Frozen Baseline EfficientNet-B0 served as the reference configuration, while Severe-Focused EfficientNet-B0 incorporated a training strategy intended to improve recognition of the underrepresented severe class. Final Frozen EfficientNet-B0 used the selected layer-freezing strategy, in which designated layers remained non-trainable while the remaining parameters were updated during optimization.

The three configurations were subsequently evaluated using the same patient-independent test set to examine the effect of the different training strategies on four-class cataract severity

classification.

3.6.5 Training configuration and hyperparameters

All experiments used an EfficientNet-B0 backbone initialized with ImageNet-pretrained weights. Each input fundus image was resized to $224 \times 224 \times 3$ and processed using the preprocessing procedure associated with EfficientNet-B0. Training was performed with a learning rate of 1×10^{-5} , a batch size of 32, and a maximum of eight epochs. Early stopping was configured with a patience of three epochs, while the learning rate was reduced by a factor of 0.2 after two consecutive epochs without improvement in the monitored validation performance. The three experimental configurations followed this common training protocol and differed primarily in their training-data sampling procedures and backbone layer-freezing strategies.

3.7 Evaluation metrics

Because the dataset is highly imbalanced, particularly for the severe category, model performance is evaluated using multiple complementary metrics rather than overall accuracy alone.

3.7.1 Accuracy

Overall accuracy represents the proportion of test observations that are correctly classified by the model:

$$\text{Accuracy} = \frac{N_{\text{correct}}}{N_{\text{total}}} \quad (13)$$

For the four-class problem, N_{correct} represents the number of test images for which the predicted class matches the ground-truth class.

Although accuracy provides an overall measure of classification performance, it can be dominated by the majority normal class in an imbalanced dataset. Therefore, additional class-balanced measures are considered.

3.7.2 Precision

Precision quantifies the reliability of predictions assigned to a given class. Specifically, for class c , it is defined as the fraction of correctly classified samples among all samples predicted as belonging to that class.

For class, precision is defined as:

$$\text{Precision}_c = \frac{TP_c}{(TP_c + FP_c)} \quad (14)$$

where, TP_c indicates the number of correctly identified samples for class c and FP_c denotes the number of samples incorrectly predicted as belonging to that class.

Precision indicates the reliability of the model when assigning a sample to a particular cataract severity category.

3.7.3 Recall

Recall measures the model's ability to correctly identify samples belonging to a given class. For class c , it represents the proportion of actual class c samples that are correctly recognized by the model. It is defined as:

$$\text{Recall}_c = \frac{TP_c}{(TP_c + FN_c)} \quad (15)$$

where, FN_c denotes the false-negative count for class c . Recall is particularly important for the present study because it indicates how effectively the model identifies the less-represented cataract severity categories.

3.7.4 F1-score

The F1-score was calculated separately for each of the four classes to assess class-wise classification performance. Because the Severe category is substantially underrepresented, overall accuracy alone may not adequately reflect minority-class performance. Therefore, class-wise results were interpreted alongside macro- and weighted-averaged measures.

3.7.5 Macro-F1 and weighted-F1

Because of the substantial class imbalance, both macro-averaged and weighted F1-scores were considered. Macro-F1 assigns equal importance to each of the four classes and is therefore useful for examining minority-class performance. In contrast, weighted-F1 accounts for the number of observations in each class and reflects the influence of the overall class distribution.

Both measures were reported to provide complementary perspectives on model performance in the imbalanced four-class classification setting.

3.7.6 Specificity and receiver operating characteristic-area under the curve

Specificity was calculated on a class-wise basis to measure the ability of the model to correctly identify samples that do not belong to a particular class. Receiver operating characteristic-area under the curve (ROC-AUC) was calculated using a one-vs-rest strategy for the four-class classification problem and summarized using the averaging methods reported in the experimental results.

3.7.7 Cohen's Kappa

Cohen's Kappa quantifies the agreement between the predicted and ground-truth class labels while accounting for agreement that may occur by chance.

3.7.8 Matthews Correlation Coefficient

The multiclass Matthews Correlation Coefficient (MCC) was calculated as an additional balanced measure of classification performance based on the complete pattern of correct and incorrect predictions across all four classes.

3.7.9 Confusion matrix

Prediction patterns were analyzed by using the confusion matrix. The diagonal elements depict correct classification of observations, while the off-diagonal elements depict misclassifications between severity categories. This analysis was particularly apt for studying the confusion between Mild, Moderate, and Severe cases.

3.7.10 Statistical and computational evaluation

The same set of 585 independent observations was used to assess all three configurations. Therefore, pairwise comparisons were made with exact McNemar's test obtained from the same observations of test items. The total, trainable, and non-trainable parameters, trainable parameter ratio, model size, inference time, and throughput were used to evaluate computational characteristics.

3.7.11 Interpretability using Gradient-weighted Class Activation Mapping

Class-specific heat maps were then created using Gradient-weighted Class Activation Mapping (Grad-CAM), and the regions of the image that were relevant to the predicted cataract severity were visualized [35]. For qualitative interpretation, the heatmaps were overlaid on the respective fundus photographs. Because the fundus changes that may occur in cataract patients do not always represent a localized retinal lesion, Grad-CAM was not used as a means of clinical localization, but rather as a qualitative interpretability tool.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents and interprets the experimental results for four-class cataract severity classification using fundus images. The four categories are Normal, Mild, Moderate, and Severe. The patient-independent experimental protocol consisted of 2,698 training images, 585 validation images, and 585 independent test images. The same independent test set was used to evaluate all three EfficientNet-B0 configurations. Adam was used for model optimization with a learning rate of 1×10^{-5} , a batch size of 32, and sparse categorical cross-entropy was chosen as the loss function. To control the training process and facilitate model convergence, early stopping and adaptive learning-rate reduction were added.

4.1 Overall performance comparison

The three configurations listed in Table 3 were implemented using TensorFlow/Keras with an ImageNet-pretrained EfficientNet-B0 backbone and evaluated using the same independent test set.

Given the highly imbalanced distribution of the independent test set, class-wise performance was examined across all four severity categories. The Normal category achieved the strongest performance across the three configurations, consistent with its majority representation in the test set.

For the Final Frozen EfficientNet-B0, the Normal class achieved a precision of 97.31%, a recall of 92.17%, and an F1-score of 94.67%. Table 4 presents the overall performance comparison of the evaluated configurations.

Table 3. Experimental configurations

Configuration	Description
Frozen Baseline EfficientNet-B0	Baseline configuration using a frozen EfficientNet-B0 backbone for four-class cataract severity classification.
Severe-Focused EfficientNet-B0	Configuration incorporating a severe-class-focused training strategy to improve recognition of the underrepresented Severe category.
Final Frozen EfficientNet-B0	Final configuration using controlled class sampling, mild Severe-class augmentation, and the selected freezing and fine-tuning strategy.

Table 4. Overall performance comparison of the evaluated configurations

Metric	Frozen Baseline EfficientNet-B0	Severe-Focused EfficientNet-B0	Final Frozen EfficientNet-B0
Accuracy (%)	88.21	86.32	88.38
Macro Precision (%)	44.34	53.65	53.86
Macro Recall (%)	49.07	58.95	60.92
Macro F1-score (%)	45.84	53.41	56.33
Weighted F1-score (%)	88.55	87.75	89.29
Macro Specificity (%)	92.20	92.87	93.12
Macro ROC-AUC (%)	90.75	91.49	91.60
Weighted ROC-AUC (%)	89.80	89.99	89.96
Cohen's Kappa	0.5245	0.5146	0.5655
MCC	0.5277	0.5311	0.5745

Note: ROC-AUC = Receiver operating characteristic-area under the curve, MCC = Matthews Correlation Coefficient.

Table 5. Class-wise performance comparison of the evaluated configurations

Class	Metric	Frozen Baseline EfficientNet-B0	Severe-Focused EfficientNet-B0	Final Frozen EfficientNet-B0
Normal	Precision (%)	96.79	97.47	97.31
	Recall (%)	94.32	90.41	92.17
	F1-score (%)	95.54	93.81	94.67
	Specificity (%)	78.38	83.78	82.43
Mild	Precision (%)	41.18	33.78	43.14
	Recall (%)	67.74	80.65	70.97
	F1-score (%)	51.22	47.62	53.66
	Specificity (%)	94.58	91.16	94.77
Moderate	Precision (%)	39.39	50.00	50.00
	Recall (%)	34.21	44.74	60.53
	F1-score (%)	36.62	47.22	54.76
	Specificity (%)	96.34	96.89	95.80
Severe	Precision (%)	0.00	33.33	25.00
	Recall (%)	0.00	20.00	20.00
	F1-score (%)	0.00	25.00	22.22
	Specificity (%)	99.48	99.66	99.48

4.2 Class-wise performance analysis

For the Mild category, Final Frozen EfficientNet-B0 achieved 43.14% precision, 70.97% recall, and 53.66% F1-score. Its F1-score was higher than those obtained with both Frozen Baseline EfficientNet-B0 and Severe-Focused EfficientNet-B0 on the same independent test set. Overall performance was evaluated using accuracy, macro-averaged precision, recall, F1-score, weighted F1-score, specificity, ROC-AUC, Cohen's Kappa, and MCC.

The precision, recall, F1, and specificity of those three configurations for each class are shown in Table 5. Hence, the results should not be taken as proof of solid clinical-level classification of severe cataract. The results, however, show better classification for some minority classes and acceptable performance for intermediate severity but poor performance for the severe class, which needs to be tested with a larger, balanced training set.

The Severe category requires particularly cautious interpretation because only five severe images were available in the independent test set. Final Frozen EfficientNet-B0 correctly identified one of the five severe test images, resulting in 20.00% recall, 25.00% precision, and 22.22% F1-score. Although severe-class recall was higher than the 0% obtained with Frozen Baseline EfficientNet-B0, the limited number of severe observations prevents reliable conclusions regarding severe-class recognition.

4.3 Receiver operating characteristic-area under the curve analysis

The one-vs-rest receiver operating characteristic (ROC) curves for the four classification categories are shown in Figure 10 with the EfficientNet-B0 as the final network configuration.

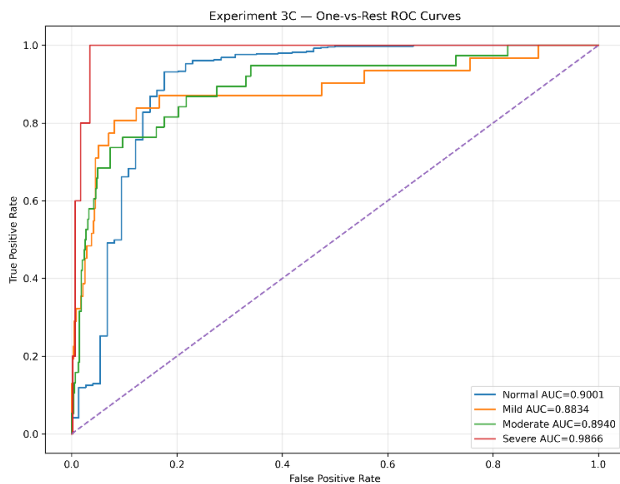


Figure 10. One-vs-rest receiver operating characteristic (ROC) curves for the Final Frozen EfficientNet-B0 configuration

The ROC curves show the compromise between the true-positive rate and false-positive rate at different classification levels. The corresponding AUC values are 0.9001 for Normal, 0.8834 for Mild, 0.8940 for Moderate, and 0.9866 for Severe. The AUC value was the greatest for the Severe category, which shows good separation from the other categories when evaluated in a ranking manner. However, this result should be interpreted cautiously because only five Severe observations

were available in the independent test set.

The Mild category had the lowest AUC, indicating that it was the most difficult to distinguish from the other categories based on the ROC analysis., suggesting that this category was the most difficult to distinguish from other categories. Overall, the ROC analysis shows that the system discriminates between the four classes well, but class-wise ROC analysis should be considered in conjunction with precision, recall, and F1-score for each class, especially for the Severe class.

4.4 Confusion matrix analysis

The confusion matrices give additional information on the prediction behavior of the tested configurations by class. In the Final Frozen EfficientNet-B0, 471/511 Normal images were correctly classified; 22/31 Mild images were correctly classified; 23/38 Moderate images were correctly classified; 1/5 Severe images were correctly classified. For the Severe category, four of the five severe observations were classified as Moderate. This pattern should be interpreted cautiously because of the extremely small number of severe observations available for independent evaluation.

Compared with Frozen Baseline EfficientNet-B0, the number of correctly classified Moderate cases increased from 13 to 23, consistent with the improvement in Moderate-class recall observed in the class-wise analysis. The number of correctly classified Severe cases increased from 0 to 1.

Overall, the confusion matrix shown in Figure 11 indicates improved recognition of the Moderate category for Final Frozen EfficientNet-B0, while Severe-class recognition remains limited.

True Class	Normal	Mild	Moderate	Severe
Normal	471	24	16	0
Mild	6	22	3	0
Moderate	7	5	23	3
Severe	0	0	4	1
Predicted Class				

Figure 11. Confusion matrix of the Final Frozen EfficientNet-B0 configuration on the independent test set

These results should therefore be interpreted alongside the class-wise precision, recall, and F1-score rather than overall accuracy alone.

4.5 Training and validation performance

The three EfficientNet-B0 configurations were developed using the prescribed training and validation protocol, while the independent test set was excluded from model development. The Final Frozen EfficientNet-B0 was trained for a maximum of eight epochs, with early stopping based on validation performance, and the best-performing model checkpoint was retained.

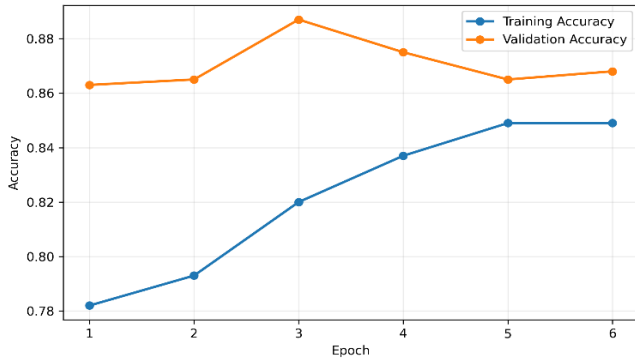


Figure 12. Training and validation accuracy of the Final Frozen EfficientNet-B0

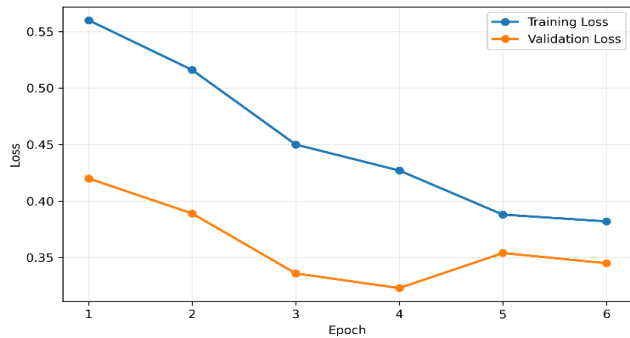


Figure 13. Training and validation loss of the Final Frozen EfficientNet-B0

Generally, the curves show satisfactory learning without significant differences between training and validation performance and reasonable convergence. Figure 12 presents the training and validation accuracy curves, whereas Figure 13 presents the corresponding training and validation loss curves.

The training accuracy increased progressively from approximately 78.2% to 84.9%, while validation accuracy reached its maximum of approximately 88.7% at epoch 3.

Training loss decreased consistently throughout the training period, whereas validation loss reached its minimum of approximately 0.32 at epoch 4 before showing a slight increase.

These curves are used to examine training convergence and the relationship between training and validation performance.

4.6 Gradient-weighted Class Activation Mapping interpretability analysis

Grad-CAM was applied to the Final Frozen EfficientNet-

B0 configuration to qualitatively examine the image regions contributing to the model's predictions.

The activation pattern was distributed throughout wider areas of the fundus image, not focused on a specific anatomical structure. Because the presented example was misclassified, the highlighted regions should not be interpreted as definitive severity-specific pathological features.

Instead, the visualization provides qualitative insight into the image regions that influenced the model's prediction. Further expert assessment and quantitative localization analysis would be required to establish the clinical relevance of these activation patterns. Figure 14 presents a representative severe cataract case that was incorrectly classified as Mild, together with its corresponding Grad-CAM visualization.

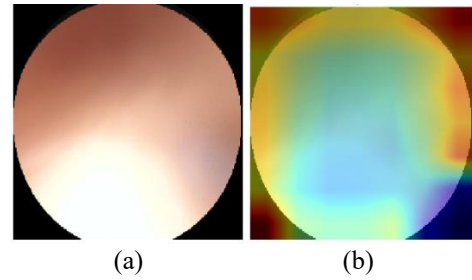


Figure 14. Severe cataract misclassified as mild: (a) Original image and (b) Gradient-weighted Class Activation Mapping (Grad-CAM) map

4.7 Statistical analysis

The three EfficientNet-B0 configurations were evaluated on the same 585 independent test observations. Since the predictions correspond to the same test samples, paired statistical comparison was performed using McNemar's test.

Table 6 presents the results of the pairwise comparisons. The comparison between Frozen Baseline EfficientNet-B0 and Severe-Focused EfficientNet-B0 yielded a p -value of 0.070756, indicating no statistically significant difference at the 0.05 significance level.

Similarly, Frozen Baseline EfficientNet-B0 and Final Frozen EfficientNet-B0 yielded a p -value of 1.000000, indicating no statistically significant difference in their paired prediction outcomes.

In contrast, the comparison between Severe-Focused EfficientNet-B0 and Final Frozen EfficientNet-B0 yielded a p -value of 0.022656, indicating a statistically significant difference at the 0.05 significance level. Thus, a statistically significant difference was observed only between the Severe-Focused and Final Frozen configurations.

Table 6. McNemar's test results

Model Comparison	A Correct, B Incorrect	A Incorrect, B Correct	Discordant Pairs	Exact (p)-Value
Frozen Baseline EfficientNet-B0 vs Severe-Focused EfficientNet-B0	21	10	31	0.070756
Frozen Baseline EfficientNet-B0 vs Final Frozen EfficientNet-B0	12	13	25	1.000000
Severe-Focused EfficientNet-B0 vs Final Frozen EfficientNet-B0	6	18	24	0.022656

4.8 Computational analysis

The computational characteristics of the three EfficientNet-

B0 configurations were evaluated in terms of parameter counts, computational complexity, model size, CPU inference latency, and throughput. All configurations use the same

underlying EfficientNet-B0 architecture with an input size of $224 \times 224 \times 3$ and contain 4,054,695 total parameters. Each model requires approximately 800,795,439 FLOPs (0.801 GFLOPs) and 400,397,719.5 MACs (0.400 GMACs) per inference.

Because the underlying architecture is identical across the three configurations, the total parameter count and computational complexity remain unchanged. The differences among the configurations are primarily related to the number of trainable and non-trainable parameters resulting from their respective freezing strategies and to the empirically measured inference characteristics under the experimental environment.

The Final Frozen EfficientNet-B0 has a model size of 39.63 MB and achieved a CPU inference latency of 61.98 ms per image, corresponding to a throughput of 16.13 images/s. The Severe-Focused EfficientNet-B0 showed the lowest measured

CPU latency of 57.02 ms per image and the highest throughput of 17.54 images/s.

These latency and throughput differences should be interpreted as empirical measurements obtained under the specific implementation and CPU-based measurement environment rather than as inherent differences in the computational architecture. Overall, the results indicate that the evaluated EfficientNet-B0 configurations maintain relatively low computational complexity while providing four-class cataract severity classification. Table 7 summarizes selected related approaches by dataset or image source, number of classes, algorithm, TL strategy, evaluation protocol, and reported accuracy. Because the datasets and experimental protocols differ across studies, the reported results should not be interpreted as direct or statistically controlled comparisons.

Table 7. Comparison of the present study with previous studies on cataract classification

Study	Dataset / Image Source	Classes	Algorithm / Method	Transfer Learning (TL)	Evaluation Protocol	Accuracy (%)
Cao et al. [11]	1,355 retinal images	4	Improved Haar Wavelet + hierarchical classification	No	Train/test evaluation	85.98
Pratap and Kokil [15]	800 fundus images from multiple open-access datasets	4	Pretrained CNN feature extraction + SVM	Yes	Train/test evaluation	92.91
Yadav and Yadav [23]	Fundus images from multiple open-source datasets	4	2D-DFT spectrogram + CNN + softmax	No	Training and testing evaluation	93.10
Yadav and Yadav [24]	Fundus images from multiple open-access datasets	4	Pretrained AlexNet + VGGNet + ResNet + ensemble of SVMs	Yes	Training and testing evaluation	96.25
Present study	3,868 ODIR-5K-derived fundus images	4	Final Frozen EfficientNet-B0	Yes	Patient-independent 70:15:15 train/validation/test split with an independent test set of 585 images	88.38

Note: TL = Transfer learning; Cls. = Number of classes, CNN = Convolutional Neural Network, ODIR-5K = Ocular Disease Intelligent Recognition 5K, SVM = Support Vector Machine.

4.9 Discussion

The Final Frozen EfficientNet-B0 achieved an overall accuracy of 88.38% on the independent patient-level test set containing 585 fundus images. To provide a comprehensive evaluation under the highly imbalanced four-class distribution, additional performance measures were considered. The model achieved 53.86% macro precision, 60.92% macro recall, 56.33% macro F1-score, 89.29% weighted F1-score, 93.12% macro specificity, and 91.60% macro ROC-AUC, along with a Cohen's Kappa of 0.5655 and an MCC of 0.5745.

Among the four severity categories, the Final Frozen EfficientNet-B0 demonstrated improved recognition of the Moderate category compared with the Frozen Baseline configuration. However, performance for the Severe category remained limited. Because only five Severe images were available in the independent test set, the corresponding class-wise performance measures are associated with substantial uncertainty and should therefore be interpreted cautiously.

Pairwise statistical analysis using the exact McNemar's test showed a statistically significant difference between the Severe-Focused EfficientNet-B0 and the Final Frozen EfficientNet-B0, whereas no statistically significant difference

was observed between the Frozen Baseline EfficientNet-B0 and the Final Frozen EfficientNet-B0. Therefore, the selection of the Final Frozen EfficientNet-B0 is based on its overall balance of predictive performance, class-wise behavior, and computational characteristics rather than on a statistically significant improvement over the Frozen Baseline configuration.

Previous studies have reported strong results for cataract detection and grading using conventional ML, DL, TL, ensemble, and ranking-based approaches. However, direct numerical comparisons with the present study should be interpreted cautiously because of differences in imaging modalities, dataset composition, class definitions, sample distributions, preprocessing procedures, data-partitioning strategies, and evaluation protocols.

Accordingly, the present findings are primarily interpreted through the controlled comparison of the three EfficientNet-B0 configurations evaluated using the same patient-independent test set. The main contribution of this study is not the introduction of a new neural network architecture, but rather the systematic evaluation of different EfficientNet-B0 training configurations for four-class cataract severity classification.

This evaluation includes class-wise performance analysis,

statistical comparison of two classes, computational evaluation, and explicit treatment of the uncertainty of the underrepresented Severe class.

5. CONCLUSION

The present study evaluated three EfficientNet-B0 training configurations for four-class cataract severity classification from fundus images comprising the Normal, Mild, Moderate, and Severe categories. Among the evaluated configurations, the Final Frozen EfficientNet-B0 provided the most favorable overall balance of predictive performance, class-wise behavior, and computational characteristics. It achieved an accuracy of 88.38%, a macro F1-score of 56.33%, and a macro ROC-AUC of 91.60% on the independent patient-level test set.

The results indicate improved recognition of the underrepresented Moderate category compared with the Frozen Baseline configuration, whereas recognition of the severely underrepresented Severe category remained limited. Grad-CAM analysis provided qualitative insight into the image regions associated with model predictions. Pairwise comparison using McNemar's test showed a statistically significant difference between the Severe-Focused and Final Frozen EfficientNet-B0 configurations, whereas no statistically significant difference was observed between the Frozen Baseline and Final Frozen EfficientNet-B0 configurations. The Final Frozen EfficientNet-B0 showed the most favorable overall balance among the evaluated configurations, while recognition of the severely underrepresented Severe category remained limited.

Overall, the findings demonstrate the potential of EfficientNet-B0 for four-class cataract severity classification under the present experimental protocol. However, the limited number of Severe images in the independent test set and the absence of external validation restrict the generalizability of the findings. Future work should evaluate the approach using larger, more balanced, clinically annotated, and externally validated multicenter datasets, with particular emphasis on improving the classification of the underrepresented Severe category.

AUTHOR CONTRIBUTIONS

The first author carried out the primary research activities, including conceptualization, data preparation, experimental design, fundus image preprocessing, model development, training, performance evaluation, result analysis, visualization, and preparation of the original manuscript draft. The second author provided technical guidance, methodological supervision, critical review of the experimental design and results, and contributed to manuscript review and revision. The third author provided overall research guidance, academic supervision, critical feedback, and contributed to the review and refinement of the manuscript. All authors reviewed and approved the final version of the manuscript.

REFERENCES

[1] Flaxman, S.R., Bourne, R.R.A., Resnikoff, S., et al.

- (2017). Global causes of blindness and distance vision impairment 1990-2020: A systematic review and meta-analysis. *The Lancet Global Health*, 5(12): e1221-e1234. [https://doi.org/10.1016/s2214-109x\(17\)30393-5](https://doi.org/10.1016/s2214-109x(17)30393-5)
- [2] Pascolini, D., Mariotti, S.P. (2012). Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 96(5): 614-618. <https://doi.org/10.1136/bjophthalmol-2011-300539>
- [3] Allen, D., Vasavada, A. (2006). Cataract and surgery for cataract. *BMJ*, 333(7559): 128-132. <https://doi.org/10.1136/bmj.333.7559.128>
- [4] Gali, H.E., Sella, R., Afshari, N.A. (2019). Cataract grading systems. *Current Opinion in Ophthalmology*, 30(1): 13-18. <https://doi.org/10.1097/icu.0000000000000542>
- [5] World Health Organization. (2026). Blindness and vision impairment. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>.
- [6] GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease Study. (2021). Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study. *The Lancet Global Health*, 9(2): e144-e160. [https://doi.org/10.1016/S2214-109X\(20\)30489-7](https://doi.org/10.1016/S2214-109X(20)30489-7)
- [7] Vashist, P., Senjam, S.S., Gupta, V., et al. (2022). Blindness and visual impairment and their causes in India: Results of a nationally representative survey. *PLOS ONE*, 17(7): e0271736. <https://doi.org/10.1371/journal.pone.0271736>
- [8] ODIR-2019. (2019). Ocular disease intelligent recognition (ODIR-5K) dataset. Peking University International Competition on Ocular Disease Intelligent Recognition.
- [9] Guo, L., Yang, J.J., Peng, L., Li, J., Liang, Q. (2015). A computer-aided healthcare system for cataract classification and grading based on fundus image analysis. *Computers in Industry*, 69: 72-80. <https://doi.org/10.1016/j.compind.2014.09.005>
- [10] Yang, J.J., Li, J., Shen, R., et al. (2016). Exploiting ensemble learning for automatic cataract detection and grading. *Computer Methods and Programs in Biomedicine*, 124: 45-57. <https://doi.org/10.1016/j.cmpb.2015.10.007>
- [11] Cao, L., Li, H., Zhang, Y., Zhang, L., Xu, L. (2020). Hierarchical method for cataract grading based on retinal images using improved Haar wavelet. *Information Fusion*, 53: 196-208. <https://doi.org/10.1016/j.inffus.2019.06.022>
- [12] Pathak, S., Raj, R., Singh, K., Verma, P.K., Kumar, B. (2022). Development of portable and robust cataract detection and grading system by analyzing multiple texture features for tele-ophthalmology. *Multimedia Tools and Applications*, 81(16): 23355-23371. <https://doi.org/10.1007/s11042-022-12544-5>
- [13] Gao, X., Li, H., Lim, J.H., Wong, T.Y. (2011). Computer-aided cataract detection using enhanced texture features on retro-illumination lens images. In 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, pp. 1565-1568. <https://doi.org/10.1109/icip.2011.6115746>
- [14] Shin, H.C., Roth, H.R., Gao, M., et al. (2016). Deep

- convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5): 1285-1298. <https://doi.org/10.1109/tmi.2016.2528162>
- [15] Pratap, T., Kokil, P. (2019). Computer-aided diagnosis of cataract using deep transfer learning. *Biomedical Signal Processing and Control*, 53: 101533. <https://doi.org/10.1016/j.bspc.2019.04.010>
- [16] Imran, A., Li, J.Q., Pei, Y., Akhtar, F., Mahmood, T., Zhang, L. (2021). Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network. *The Visual Computer*, 37(8): 2407-2417. <https://doi.org/10.1007/s00371-020-01994-3>
- [17] Zhang, H.Y., Niu, K., Xiong, Y.M., Yang, W.H., He, Z.Q., Song, H.X. (2019). Automatic cataract grading methods based on deep learning. *Computer Methods and Programs in Biomedicine*, 182: 104978. <https://doi.org/10.1016/j.cmpb.2019.07.006>
- [18] Tan, M.X., Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105-6114. <https://proceedings.mlr.press/v97/tan19a.html>
- [19] Pratap, T., Kokil, P. (2021). Efficient network selection for computer-aided cataract diagnosis under noisy environment. *Computer Methods and Programs in Biomedicine*, 200: 105927. <https://doi.org/10.1016/j.cmpb.2021.105927>
- [20] Khan, M.S.M., Ahmed, M., Rasel, R.Z., Khan, M.M. (2021). Cataract detection using convolutional neural network with VGG-19 model. In *2021 IEEE World AI IoT Congress (AIoT)*, Seattle, WA, USA, pp. 0209-0212. <https://doi.org/10.1109/aiiot52608.2021.9454244>
- [21] Xu, X., Zhang, L., Li, J., Guan, Y., Zhang, L. (2020). A hybrid global-local representation CNN model for automatic cataract grading. *IEEE Journal of Biomedical and Health Informatics*, 24(2): 556-567. <https://doi.org/10.1109/jbhi.2019.2914690>
- [22] Gao, W.H., Shao, L., Li, F., et al. (2024). Fundus photograph-based cataract evaluation network using deep learning. *Frontiers in Physics*, 11: 1235856. <https://doi.org/10.3389/fphy.2023.1235856>
- [23] Yadav, S., Yadav, J.K.P.S. (2023). Automatic cataract severity detection and grading using deep learning. *Journal of Sensors*, 2023(1): 2973836. <https://doi.org/10.1155/2023/2973836>
- [24] Yadav, J.K.P.S., Yadav, S. (2022). Computer-aided diagnosis of cataract severity using retinal fundus images and deep learning. *Computational Intelligence*, 38(4): 1450-1473. <https://doi.org/10.1111/coin.12518>
- [25] Junayed, M.S., Islam, M.B., Sadeghzadeh, A., Rahman, S. (2021). CataractNet: An automated cataract detection system using deep learning for fundus images. *IEEE Access*, 9: 128799-128808. <https://doi.org/10.1109/access.2021.3112938>
- [26] Wu, X., Xu, D., Ma, T., et al. (2022). Artificial intelligence model for antiinterference cataract automatic diagnosis: A diagnostic accuracy study. *Frontiers in Cell and Developmental Biology*, 10: 906042. <https://doi.org/10.3389/fcell.2022.906042>
- [27] Xiong, L., Li, H., Xu, L. (2017). An approach to evaluate blurriness in retinal images with vitreous opacity for cataract diagnosis. *Journal of Healthcare Engineering*, 2017: 5645498. <https://doi.org/10.1155/2017/5645498>
- [28] Morales-Lopez, H., Cruz-Vega, I., Rangel-Magdaleno, J. (2021). Cataract detection and classification systems using computational intelligence: A survey. *Archives of Computational Methods in Engineering*, 28(3): 1761-1774. <https://doi.org/10.1007/s11831-020-09440-2>
- [29] Shaheen, I., Tariq, A. (2018). Survey analysis of automatic detection and grading of cataract using different imaging modalities. In *EAI/Springer Innovations in Communication and Computing*, Springer International Publishing, pp. 35-45. https://doi.org/10.1007/978-3-319-96139-2_4
- [30] Zhang, X.Q., Hu, Y., Xiao, Z.J., Fang, J.S., Higashita, R., Liu, J. (2022). Machine learning for cataract classification/grading on ophthalmic imaging modalities: A survey. *Machine Intelligence Research*, 19(3): 184-208. <https://doi.org/10.1007/s11633-022-1329-0>
- [31] Fuadah, Y.N., Setiawan, A.W., Mengko, T.L.R. (2015). Performing high accuracy of the system for cataract detection using statistical texture analysis and K-Nearest Neighbor. In *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Indonesia, pp. 85-88. <https://doi.org/10.1109/isitia.2015.7219958>
- [32] Liu, X.Y., Jiang, J.W., Zhang, K., et al. (2017). Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. *PLOS ONE*, 12(3): e0168606. <https://doi.org/10.1371/journal.pone.0168606>
- [33] Wang, Y., Tang, C.W., Wang, J., Sang, Y.S., Lv, J.C. (2021). Cataract detection based on ocular B-ultrasound images by collaborative monitoring deep learning. *Knowledge-Based Systems*, 231: 107442. <https://doi.org/10.1016/j.knosys.2021.107442>
- [34] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [35] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 618-626. <https://doi.org/10.1109/iccv.2017.74>
- [36] Kasar, M.S., Shedje, D.K. (2023). Survey of cataract detection and classification systems. In *7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, Pune, India, pp. 1-6. <https://doi.org/10.1109/ICCUBEA58933.2023.10392102>
- [37] Kasar, M.S., Mane, P., Shedje, D.K. (2026). ResNet for cataract detection and classification: A deep learning approach. In *International Conference on Sustainable Futuristic Technologies (ICSFT)*, Pune, India, pp. 1-5. <https://doi.org/10.1109/ICSFT67370.2026.11566448>