

Automated Multi-Class Retinal Disease Detection Using Deep Convolutional Neural Networks And Transfer Learning

Ankit Pandit¹ and Aakash Patel²

Department of Information Technology¹²

KES' B. K. Shroff College of Arts & M. H. Shroff College of Commerce, Mumbai, Maharashtra

Affiliated to the University of Mumbai, Maharashtra, India¹²

¹ ankitpandit2230@gmail.com, ² patelaakash96103@gmail.com

Abstract: Early detection of ocular diseases is essential for preventing progressive visual impairment and reducing the burden of avoidable blindness. However, conventional retinal screening depends heavily on trained ophthalmologists and manual interpretation of retinal images, which can limit the scalability and accessibility of screening services. This study presents an automated deep-learning framework for the multi-class classification of retinal fundus images into four categories: cataract, diabetic retinopathy, glaucoma, and normal. The proposed model employs transfer learning using a pre-trained EfficientNetB3 convolutional neural network as the feature-extraction backbone. To improve disease-specific feature learning while limiting overfitting, the deeper layers of the backbone were fine-tuned and combined with global average pooling, batch normalization, dropout regularization, and fully connected classification layers. The retinal image dataset was screened for corrupted files, divided into training, validation, and test partitions, and subjected to training-specific augmentation using horizontal flipping, rotation, zooming, and contrast variation. Model performance was evaluated using accuracy, precision, recall, and F1-score. The proposed classifier achieved a test accuracy of 99.58%, demonstrating strong classification performance across the four retinal categories. The trained model was further integrated into a full-stack web application consisting of a React-based frontend and Flask-based backend. A pre-inference image-validation mechanism was incorporated to identify unsuitable, non-retinal, or severely degraded image inputs before classification. The developed framework demonstrates the potential of transfer-learning-based deep neural networks for scalable and accessible preliminary retinal disease screening. Nevertheless, independent external and prospective clinical validation is required before the system can be considered suitable for real-world clinical use.

Keywords: Retinal Disease Detection; Convolutional Neural Network; Deep Learning; Retinal Fundus Imaging; Cataract; Diabetic Retinopathy; Glaucoma; Transfer Learning; EfficientNetB3; Medical Image Classification.

I. INTRODUCTION

Vision impairment and blindness remain significant global public-health concerns, affecting individuals across different age groups and socioeconomic settings. According to the World Health Organization (WHO), at least 2.2 billion people worldwide experience near or distance vision impairment, and at least 1 billion of these cases could have been prevented or remain unaddressed [1]. Major causes of vision impairment include refractive errors, cataract, diabetic retinopathy, glaucoma, and age-related macular degeneration [1]. The continuing burden of these conditions



emphasizes the need for accessible, accurate, and scalable approaches that can facilitate early detection and timely referral for appropriate ophthalmic care.

Among ocular disorders, cataract, diabetic retinopathy, and glaucoma are particularly important because of their substantial contribution to visual disability. Cataract is a major cause of vision impairment and blindness but is generally treatable through appropriate surgical intervention. Glaucoma, in contrast, can cause progressive and irreversible damage to the optic nerve, making timely detection particularly important. Diabetic retinopathy is a microvascular complication of diabetes that can progressively damage the retinal vasculature and lead to severe visual impairment if clinically significant abnormalities are not identified and appropriately managed [1]. Early recognition of ocular abnormalities is therefore an important component of strategies aimed at reducing avoidable vision loss.

Retinal fundus photography provides a non-invasive method for visualizing important ocular structures, including the optic disc, macula, retinal vasculature, and surrounding retinal tissue. Alterations in these structures can provide clinically relevant information associated with different ocular diseases. However, conventional interpretation of fundus photographs generally requires trained ophthalmologists or professional retinal graders. Dependence on specialist interpretation can limit the scalability of screening, particularly when large populations require assessment or when access to ophthalmic expertise is limited. These challenges have stimulated growing interest in computer-assisted methods capable of supporting retinal-image screening and prioritizing cases that may require specialist evaluation.

Recent developments in artificial intelligence (AI) have created new opportunities for automated medical-image analysis. In particular, deep learning and convolutional neural networks (CNNs) have demonstrated considerable capability in image-recognition tasks because they can automatically learn hierarchical representations directly from image data. In contrast to conventional machine-learning techniques that often depend on manually engineered image descriptors, CNNs can learn progressively complex visual patterns through multiple convolutional layers. This characteristic is particularly valuable for retinal imaging, where pathological changes may involve subtle variations in retinal texture, vasculature, optic-disc appearance, and other anatomical structures.

The potential of deep learning for automated retinal analysis was demonstrated by Gulshan et al. [2], who developed a deep convolutional neural network for detecting diabetic retinopathy from retinal fundus photographs. Their algorithm was developed using 128,175 retinal images and subsequently evaluated using independent validation datasets. The reported results demonstrated high sensitivity and specificity for detecting referable diabetic retinopathy and provided important evidence supporting the application of deep-learning techniques to automated retinal screening [2].

The scope of deep-learning-based retinal analysis was subsequently expanded beyond diabetic retinopathy alone. Ting et al. [3] developed and validated a deep-learning system using retinal images obtained from multiethnic populations with diabetes. Their system was evaluated for several clinically important conditions, including referable diabetic retinopathy, vision-threatening diabetic retinopathy, referable possible glaucoma, and age-related macular degeneration [3]. Such investigations demonstrate that deep neural networks can extract clinically meaningful patterns from retinal photographs and provide a foundation for developing systems capable of analysing multiple ocular abnormalities.

Despite these advances, the development of reliable deep-learning models for medical-image classification remains challenging. Deep CNN architectures generally contain large numbers of trainable parameters and perform optimally when substantial quantities of accurately labelled data are available. In medical imaging, however, assembling large datasets can be difficult because image acquisition, expert annotation, quality assessment, and clinical verification require considerable time and specialized resources. Models developed using relatively limited datasets may consequently be vulnerable to overfitting and may fail to generalize adequately to previously unseen images.

Transfer learning provides a practical approach for addressing this challenge. Rather than training a deep neural network entirely from randomly initialized parameters, transfer learning begins with a model that has already learned general visual representations from a large source dataset. These learned representations can subsequently be adapted to a domain-specific task through fine-tuning. Such an approach can reduce training requirements and improve feature learning when the target medical-image dataset is substantially smaller than datasets commonly used for training large CNN architectures from scratch.



The present study employs EfficientNetB3 as the convolutional backbone for retinal image classification. EfficientNet, introduced by Tan and Le [4], uses a compound scaling strategy that systematically balances network depth, width, and input resolution. This approach was designed to achieve strong classification performance while using computational resources more efficiently than indiscriminately increasing individual architectural dimensions [4]. These characteristics make EfficientNet particularly attractive for transfer-learning applications in which predictive performance, model complexity, and computational requirements must be carefully balanced.

Although substantial progress has been achieved in automated retinal-image analysis, much of the established literature has focused on detecting individual diseases, particularly diabetic retinopathy [2]. Multi-condition systems have also demonstrated the feasibility of identifying several clinically relevant ocular abnormalities [3]; however, there remains scope for developing practical frameworks that combine multi-class disease classification, transfer learning, input-quality validation, and accessible application-level deployment within a unified system. Furthermore, high classification performance alone is insufficient for practical use if the model accepts inappropriate or severely degraded images and generates predictions without considering input suitability.

To address these considerations, the present study develops an automated deep-learning framework for classifying retinal fundus images into four categories: cataract, diabetic retinopathy, glaucoma, and normal. The proposed model employs a pre-trained EfficientNetB3 backbone with partial fine-tuning and a task-specific classification head consisting of global average pooling, batch normalization, fully connected layers, and dropout regularization. Data augmentation is incorporated during training to increase image variability and reduce the likelihood of overfitting.

In addition to the classification model, the study extends the framework into an end-to-end web-based application. The trained model is integrated with a React-based frontend and Flask-based backend, enabling retinal images to be submitted for automated analysis through an accessible interface. A pre-inference image-validation mechanism is incorporated to identify unsuitable, non-retinal, or severely degraded inputs before classification, thereby reducing the likelihood of presenting potentially misleading predictions for clearly inappropriate images.

The principal contribution of this work is therefore the development of a unified and deployable retinal-image classification framework rather than an isolated experimental classifier. By combining EfficientNetB3-based transfer learning, partial network fine-tuning, image augmentation, regularization, multi-class prediction, pre-inference input validation, and web-based deployment, the proposed approach investigates how deep-learning-based retinal classification can be translated into a practical screening-support prototype. The system is intended to assist preliminary retinal-image assessment and prioritization rather than replace professional ophthalmic diagnosis. Its performance must consequently be interpreted within the characteristics of the dataset used in this study, with independent external and prospective clinical validation required before consideration for real-world clinical application.

II. LITERATURE REVIEW

The application of artificial intelligence to ophthalmic imaging has expanded considerably with the development of deep learning and convolutional neural networks (CNNs). Retinal fundus photographs are particularly suitable for automated image analysis because they provide visual information about the optic disc, retinal vasculature, macula, and surrounding retinal structures. Deep-learning approaches have consequently been investigated for the detection and classification of several ocular diseases, including diabetic retinopathy, glaucoma, cataract, and age-related macular degeneration [5]. Compared with conventional machine-learning approaches based on manually engineered features, CNN-based systems can learn hierarchical representations directly from retinal images and have therefore become an important approach in computer-assisted ophthalmic image analysis [6].

One of the most influential applications of deep learning in ophthalmology has been the automated detection of diabetic retinopathy (DR). Gulshan et al. [2] developed a deep-learning algorithm using 128,175 retinal fundus images and evaluated its performance on independent datasets. Their study demonstrated that a deep neural network could achieve high sensitivity and specificity for detecting referable diabetic retinopathy, establishing the feasibility of automated



retinal screening using deep learning. This work provided an important benchmark and stimulated extensive subsequent research into CNN-based analysis of retinal fundus photographs.

Ting et al. [3] subsequently investigated the generalizability of deep-learning-based retinal analysis using retinal images obtained from multiethnic populations with diabetes. Their system was evaluated not only for diabetic retinopathy but also for related ocular abnormalities, including possible glaucoma and age-related macular degeneration. The study demonstrated that deep-learning systems could potentially extend beyond single-disease classification and extract information relevant to multiple ocular conditions from retinal photographs [3].

The substantial body of research that followed these landmark studies has further supported the utility of deep learning for diabetic retinopathy assessment. A systematic review and meta-analysis of deep-learning algorithms for DR detection reported high overall diagnostic performance for identifying referable diabetic retinopathy from retinal fundus photographs [7]. Similarly, later reviews have shown that deep-learning-based DR pipelines encompass several stages, including image preprocessing, lesion localization, disease detection, severity classification, and clinical screening [8]. These studies demonstrate the maturity of deep learning for DR analysis while also emphasizing the importance of dataset quality, preprocessing, appropriate evaluation, and external validation.

Deep learning has also been extensively investigated for glaucoma detection from retinal fundus photographs. Glaucoma-related structural abnormalities, particularly changes involving the optic nerve head, provide visual information that can be analysed computationally. Li et al. [9] evaluated a deep-learning system for detecting glaucomatous optic neuropathy from colour fundus photographs and demonstrated that CNN-based analysis could effectively distinguish glaucomatous from non-glaucomatous images. Their findings supported the feasibility of using retinal photographs as input for automated glaucoma screening systems.

Transfer learning has shown particular value in glaucoma classification. Christopher et al. [10] evaluated multiple deep-learning architectures for identifying glaucomatous optic neuropathy using a large dataset of fundus photographs. Their analysis demonstrated that transfer learning improved model performance and reduced training time. The best-performing model achieved an area under the receiver operating characteristic curve of approximately 0.91 for distinguishing glaucomatous optic neuropathy from healthy eyes, with stronger performance for moderate-to-severe disease [10]. These findings illustrate the value of pretrained visual representations for ophthalmic classification tasks.

Automated cataract classification has likewise been investigated using deep-learning architectures. Although cataract is fundamentally an opacity of the crystalline lens, changes in image clarity, illumination, colour distribution, and visibility of retinal structures in fundus photographs can provide information relevant to automated classification. Studies evaluating pretrained CNN architectures have demonstrated that transfer-learning models can distinguish cataract and other ophthalmological categories from fundus images with high classification performance [11]. Such findings further support the use of pretrained CNNs for ocular image classification when available datasets are not sufficiently large to train complex architectures entirely from scratch.

The effectiveness of transfer learning is particularly relevant in medical imaging because labelled clinical datasets are generally much smaller than natural-image datasets used for large-scale computer-vision research. Instead of learning every visual representation from randomly initialized parameters, pretrained models provide previously learned feature representations that can be adapted to a target medical task. This strategy can improve convergence and reduce computational requirements, particularly when only selected deeper layers are fine-tuned. Studies involving retinal fundus photographs have shown that transfer learning can provide meaningful performance improvements across different ophthalmological classification problems [10], [11].

Several established CNN architectures have consequently been adapted for retinal-image analysis, including VGG, ResNet, Inception, DenseNet, and EfficientNet. ResNet introduced residual connections that facilitate optimization of substantially deeper neural networks by allowing information to propagate through shortcut connections [12]. Such architectures have subsequently been widely adopted as pretrained backbones for medical-image analysis. Comparative investigations of pretrained CNNs for ophthalmological disease classification have further demonstrated that the most



suitable architecture can vary according to the disease and image characteristics, supporting the need for architecture-specific evaluation rather than assuming that a single backbone is universally optimal [11].

The EfficientNet architecture represents another important development in CNN design. Rather than independently increasing network depth, width, or input resolution, EfficientNet uses compound scaling to balance these dimensions systematically [4]. This design provides a favourable relationship between model capacity and computational requirements, making EfficientNet architectures attractive for transfer-learning applications. EfficientNet variants have subsequently been applied to medical-image and retinal-image classification tasks because pretrained feature extraction can be combined with task-specific fine-tuning.

Recent work has directly investigated EfficientNetB3 for multi-class retinal disease classification. A study using a publicly available retinal fundus dataset applied a fine-tuned EfficientNetB3 architecture to distinguish cataract, diabetic retinopathy, glaucoma, and healthy retinal images, reporting an overall classification accuracy of 95.12% [13]. The study incorporated transfer learning and data augmentation and demonstrated the suitability of EfficientNetB3 for four-class ophthalmic image classification [13]. This finding is particularly relevant to the present work because it provides an external benchmark for the same broad disease categories and backbone family.

Research has also increasingly shifted from isolated binary disease detectors toward multi-class retinal disease classification. Recent investigations have evaluated CNNs and vision transformers for simultaneously differentiating diabetic retinopathy, glaucoma, cataract, and normal retinal images [14]. Such multi-class frameworks may be more practical for preliminary screening than independent binary classifiers because a single inference pipeline can distinguish among several diagnostic categories. Nevertheless, multi-class classification introduces additional challenges because pathological features may overlap and because class imbalance, image acquisition variability, and disease severity can influence model predictions.

Image quality represents another important consideration in automated retinal-image analysis. Deep-learning models can produce predictions even when input images differ substantially from the distribution on which they were trained. Factors such as blur, illumination variation, low contrast, imaging artefacts, and ocular media opacity can therefore affect classification reliability. Experimental evidence has demonstrated, for example, that simulated cataract-related degradation can substantially reduce the performance of an AI system trained to detect diabetic retinopathy from colour fundus photographs [15]. Such findings emphasize that strong classification accuracy on a curated test dataset does not by itself guarantee equivalent performance under heterogeneous real-world imaging conditions.

This issue is particularly relevant when an image classifier is incorporated into a user-facing application. A conventional CNN generally assumes that the supplied input belongs to the image domain represented during training. Consequently, a model may still produce a high softmax probability for an inappropriate or poor-quality input. Introducing an image-quality or input-validation stage before disease classification can therefore provide an additional safeguard by identifying images that do not satisfy basic input requirements. However, such validation should be considered a risk-reduction mechanism rather than evidence of clinical reliability.

The broader literature also highlights concerns regarding generalizability and external validation. High performance obtained from an internal test partition may decline when models are applied to images acquired using different cameras, clinical populations, imaging protocols, or disease distributions. Reviews of deep learning in ophthalmology have consequently emphasized the importance of representative datasets, independent external evaluation, clinical validation, and careful interpretation of performance metrics before AI systems are introduced into routine ophthalmic practice [5], [6].

Overall, existing research demonstrates substantial progress in deep-learning-based analysis of retinal fundus photographs. Strong evidence exists for automated diabetic retinopathy detection [2], [3], [7], glaucoma assessment [9], [10], and transfer-learning-based ophthalmological classification [11]. EfficientNet and other pretrained CNN architectures provide effective foundations for adapting deep visual representations to retinal-image datasets [4], [13]. More recent studies further demonstrate the feasibility of multi-class classification involving cataract, diabetic retinopathy, glaucoma, and normal images [13], [14].



Despite this progress, a practical gap remains between achieving high classification performance and developing a complete screening-support pipeline. Many studies primarily concentrate on model architecture and classification metrics, whereas practical deployment additionally requires consideration of input suitability, inference handling, confidence reporting, and accessible user interaction. The present study addresses this implementation-oriented gap by combining EfficientNetB3-based transfer learning, partial fine-tuning, four-class retinal image classification, data augmentation, pre-inference image validation, confidence-based output handling, and full-stack web deployment within a unified framework. The objective is not to establish autonomous clinical diagnosis, but to investigate a practical prototype for automated preliminary retinal-image assessment that can subsequently be subjected to broader external and prospective clinical validation.

III. PROPOSED METHODOLOGY

The proposed methodology was designed to develop an automated deep-learning framework capable of classifying retinal fundus images into four categories: cataract, diabetic retinopathy, glaucoma, and normal. The complete workflow consists of dataset preparation, image verification, stratified dataset partitioning, image preprocessing, training-specific data augmentation, transfer learning using EfficientNetB3, partial fine-tuning, model training, and performance evaluation. The trained classifier was subsequently integrated into the web-based system described in Section IV.

A. Dataset Description and Preparation

The dataset used in this study was obtained from a publicly available retinal fundus image repository and contained images representing four classes: cataract, diabetic retinopathy, glaucoma, and normal. Before model development, the image collection was subjected to an automated integrity check in which each file was programmatically opened and verified. Corrupted, unreadable, or invalid image files were excluded to prevent errors during preprocessing and model training.

Following verification, the dataset was divided into training, validation, and test partitions. Stratified partitioning was employed to preserve the representation of each retinal category across the three subsets. The training partition was used for parameter optimization, the validation partition was used to monitor model performance during training and control the training procedure, and the test partition was retained for final evaluation.

The class-wise distribution used in the study is presented in Table I.

Table I. Class-wise distribution of retinal images used in the study

Class	Training	Validation	Test
Cataract	1038	694	698
Diabetic Retinopathy	1098	727	756
Glaucoma	1006	682	666
Normal	1074	721	736

The resulting dataset contains moderate variation in the number of images belonging to each class. Maintaining class representation across the dataset partitions was therefore important to reduce unintended distributional differences between the training, validation, and test sets.

B. Image Preprocessing

All retinal fundus images were resized to a uniform spatial resolution of 224×224 pixels before being supplied to the classification pipeline. Standardizing image dimensions ensures compatibility with the implemented model pipeline and allows images originally acquired at different resolutions to be represented using a consistent tensor shape.



The image preprocessing procedure was applied consistently to training, validation, test, and inference images, apart from data augmentation, which was restricted to the training partition. Because the implemented EfficientNet preprocessing pipeline handles the required input scaling internally, resized image values were supplied according to the preprocessing behaviour of the selected backbone.

No augmentation was applied to the validation or test images. This separation ensured that validation monitoring and final test-set evaluation were performed using unmodified images rather than artificially transformed versions of the evaluation samples.

C. Data Augmentation

Medical-image datasets are commonly smaller than large-scale natural-image datasets, increasing the possibility that a high-capacity CNN will learn training-specific characteristics instead of features that generalize to unseen data. To introduce additional variability into the training samples, an on-the-fly data-augmentation strategy was applied exclusively to the training partition.

The augmentation pipeline consisted of:

- random horizontal flipping;
- random rotation of up to approximately 10%;
- random zooming of up to approximately 10%; and
- random contrast variation of up to approximately 10%.

These transformations generated modified versions of the training images dynamically during model optimization without permanently increasing the stored dataset size. Augmentation was intentionally restricted to moderate transformations so that variability could be introduced while retaining the principal retinal structures required for classification.

The validation and test datasets remained unaugmented, allowing model performance to be assessed on the original evaluation images.

D. Transfer Learning Using EfficientNetB3

The proposed classifier was developed using a transfer-learning strategy based on EfficientNetB3. EfficientNet employs compound scaling to balance network depth, width, and input resolution systematically [4]. Instead of training a deep convolutional architecture entirely from randomly initialized parameters, the present implementation utilizes a pretrained EfficientNetB3 backbone as the principal feature extractor.

Transfer learning allows previously learned visual representations to be adapted to a new image-classification task. This approach is particularly useful when the target dataset is smaller than the datasets generally required to train large CNN architectures effectively from scratch. Previous ophthalmic imaging studies have demonstrated the utility of transfer learning for retinal-image classification [10], [11].

Rather than keeping the complete pretrained backbone frozen, the final 50 layers of EfficientNetB3 were made trainable, while the earlier layers remained frozen. This partial fine-tuning strategy was adopted to retain lower-level representations learned during pretraining while allowing deeper feature representations to adapt to the retinal disease classification task.

E. Proposed Network Architecture

The output feature maps generated by EfficientNetB3 were passed to a task-specific classification head. The architecture was designed to progressively transform convolutional feature representations into four-class probability estimates.

The EfficientNetB3 backbone generated an output feature map with dimensions $7 \times 7 \times 1536$. A Global Average Pooling 2D layer was subsequently applied to convert the spatial feature maps into a compact feature vector. Global average pooling reduces the spatial dimensions of convolutional outputs without requiring a large fully connected flattening operation.

The pooled representation was passed through batch normalization, followed by a fully connected layer containing 1024 neurons with ReLU activation. A dropout layer with a rate of 0.30 was applied after this dense layer. The



resulting representation was then passed through a second fully connected layer containing 512 ReLU-activated neurons, followed by batch normalization and a second dropout layer with a rate of 0.20.

The final classification layer contained four output neurons, corresponding to cataract, diabetic retinopathy, glaucoma, and normal. A softmax activation function was used to transform the model outputs into normalized class-probability estimates.

The principal components of the architecture are summarized in Table II.

Table II. Layer-wise configuration of the proposed EfficientNetB3-based classification network

Layer	Output Shape	Parameters	Trainable
EfficientNetB3 backbone	(None, 7, 7, 1536)	10,783,535	Partial
Global Average Pooling 2D	(None, 1536)	0	–
Batch Normalization	(None, 1536)	6,144	Yes
Dense, ReLU	(None, 1024)	1,573,888	Yes
Dropout (0.30)	(None, 1024)	0	–
Dense, ReLU	(None, 512)	524,800	Yes
Batch Normalization	(None, 512)	2,048	Yes
Dropout (0.20)	(None, 512)	0	–
Dense, Softmax	(None, 4)	2,052	Yes

The complete network contained approximately 17.94 million parameters, of which approximately 7.16 million parameters were trainable under the adopted fine-tuning configuration.

F. Output Classification

The final classification layer of the proposed network uses a softmax activation function to generate probability scores for the four target categories: cataract, diabetic retinopathy, glaucoma, and normal. For each retinal fundus image, the network calculates the relative probability of belonging to each category, and the class with the highest probability is selected as the final prediction. The corresponding probability value is retained as the model confidence score and is subsequently used by the web application for confidence-based result handling.

G. Model Training Configuration

The proposed network was trained using the Adam optimizer [16] with an initial learning rate of 0.0003. Sparse categorical cross-entropy was selected as the loss function because the four retinal categories were represented using integer-encoded class labels. Training was performed with a batch size of 8 for a maximum of 100 epochs, while training and validation accuracy and loss were monitored throughout the optimization process.

To improve convergence and minimize unnecessary training, early stopping and adaptive learning-rate reduction were incorporated. Early stopping terminated the training process when validation performance failed to improve for 10 consecutive epochs, while restoring the weights corresponding to the best-performing epoch. In addition, the learning rate was reduced by a factor of 0.5 when validation accuracy showed no improvement for five consecutive epochs. Under these training conditions, the model converged, and training was terminated after 76 epochs.



H. Performance Evaluation

The performance of the trained EfficientNetB3 classifier was evaluated using the independent test dataset. Four commonly used classification measures—accuracy, precision, recall, and F1-score were considered to provide both overall and class-specific assessment of model performance.

Accuracy represented the proportion of retinal images that were correctly classified among all test images. Precision measured the proportion of predictions assigned to a particular retinal category that were correct, whereas recall represented the model's ability to correctly identify images that actually belonged to that category. The F1-score provided a balanced measure of precision and recall and was particularly useful for comparing classification performance across the four retinal categories. Together, these metrics provided a more comprehensive assessment than overall accuracy alone.

I. Overall Methodological Workflow

The complete methodology followed a sequential pipeline beginning with the collection and verification of retinal fundus images. After corrupted or unreadable files were removed, the verified images were divided into training, validation, and test datasets. All images were resized to 224×224 pixels, while augmentation was applied only to the training dataset to increase image variability and support model generalization.

The processed training images were supplied to the pretrained EfficientNetB3 network, where the final 50 layers were fine-tuned for retinal disease classification. Features extracted by the convolutional backbone were subsequently processed through global average pooling, batch normalization, dense layers, and dropout regularization before the final softmax layer generated probabilities for the four retinal categories. The trained model was then evaluated on the held-out test dataset using accuracy, precision, recall, and F1-score. Following model evaluation, the finalized classifier was integrated into the EyeDetect web application for image validation, automated inference, confidence assessment, and presentation of classification results.

IV. SYSTEM IMPLEMENTATION AND DEPLOYMENT ARCHITECTURE

The trained retinal disease classification model was integrated into a full-stack web application named EyeDetect. The system follows a client-server architecture consisting of a React-based frontend, Flask-based backend, pre-inference image-validation mechanism, and the fine-tuned EfficientNetB3 classification model. The complete workflow enables retinal image upload, validation, preprocessing, classification, confidence assessment, and result presentation through a web interface.

A. Frontend and Backend Implementation

The frontend was developed using React with TypeScript and a Tailwind-based styling framework. It provides an interface for uploading retinal fundus images and displaying classification results. The backend was implemented in Python using Flask and exposes a REST endpoint for receiving uploaded images and executing the inference pipeline. After receiving an image, the backend performs input validation and preprocessing before passing the image to the trained EfficientNetB3 classifier. The resulting prediction, confidence score, and top-ranked classes are returned to the frontend in JSON format.

B. Pre-Inference Image Validation

Before classification, each uploaded image passes through a lightweight validation mechanism termed the Clinical Validation Mesh. The validation stage examines basic image characteristics, including overall contrast, brightness, mean red-channel intensity, and the red-to-green intensity ratio.

Images that do not satisfy the implemented validation criteria, such as blank, severely overexposed, or clearly unsuitable inputs, are rejected before model inference. This mechanism serves as an application-level input filter and should not be interpreted as a clinically validated retinal image-quality assessment system.



C. Model Inference and Confidence Handling

Images accepted by the validation stage are resized to 224×224 pixels and converted into the input representation required by the trained model. EfficientNetB3 then generates probability estimates for the four categories: cataract, diabetic retinopathy, glaucoma, and normal.

The category with the highest softmax probability is selected as the primary prediction. The application additionally returns a confidence score and the top three predicted categories. Predictions falling below the implemented confidence threshold are treated as inconclusive rather than being presented as definitive classifications.

D. End-to-End System Workflow

The complete application workflow is summarized in Table III.

Table III. Processing pipeline of the EyeDetect application

Stage	Operation	Output
1	Fundus image uploaded	Raw image
2	Clinical Validation Mesh	Accept/reject decision
3	Resize and preprocessing	224×224 model input
4	EfficientNetB3 inference	Class probabilities
5	Confidence evaluation and ranking	Final prediction
6	Backend response to frontend	JSON result
7	Result displayed to user	Prediction and confidence

Thus, the overall workflow can be represented as:

Fundus Image → Image Validation → Preprocessing → EfficientNetB3 → Class Prediction → Confidence Evaluation → Web-Based Result Display

The deployed system demonstrates the integration of retinal image classification with input validation and web-based accessibility. However, EyeDetect remains a screening-support research prototype and requires external clinical validation before consideration for real-world diagnostic use.

V. RESULTS AND DISCUSSION

A. Overall Model Performance

The fine-tuned EfficientNetB3 model demonstrated strong performance on the held-out test dataset, achieving an overall classification accuracy of 99.58% with a categorical cross-entropy loss of 0.0121. Training and validation accuracy increased progressively during optimization and remained closely aligned toward convergence. Training was terminated after 76 epochs using the early-stopping criterion.

The observed performance indicates that the combination of pretrained EfficientNetB3 features, partial fine-tuning, data augmentation, batch normalization, and dropout regularization was effective for the retinal image dataset evaluated in this study. The effectiveness of transfer learning for retinal-image classification is also consistent with previous ophthalmic deep-learning studies [10], [11], [13].

B. Class-Wise Classification Performance

In addition to overall accuracy, precision, recall, and F1-score were examined separately for each retinal category. The class-wise results are presented in Table IV.



Table IV. Class-wise classification performance on the test dataset

Class	Precision (%)	Recall (%)	F1-Score (%)
Normal	96	95	95.5
Cataract	94	92	93.0
Diabetic Retinopathy	92	94	93.0
Glaucoma	93	91	92.0

Among the four classes, the normal category showed the highest precision (96%), while diabetic retinopathy achieved the highest recall (94%) among the disease categories. Glaucoma showed the lowest recall at 91%, although its overall class-wise performance remained high.

These findings indicate that the model was capable of learning discriminative retinal-image features across all four categories rather than obtaining strong performance from only a single dominant class.

C. Comparative Performance

The proposed transfer-learning approach was also compared with alternative approaches evaluated during model development. The comparison is summarized in Table V.

Table V. Comparative performance of different classification approaches

Approach	Test Accuracy	Approx. Training Time	Hardware Requirement
Traditional ML with handcrafted features	70–75%	30–60 min	Low
Custom CNN trained from scratch	85–90%	2–4 h	Medium
Fine-tuned EfficientNetB3 (Proposed)	99.58%	≈20 h (76 epochs)	T4-class GPU

The proposed EfficientNetB3 model achieved substantially higher test accuracy than the traditional machine-learning and custom CNN approaches evaluated in this study. This improvement supports the use of transfer learning when the available retinal-image dataset is not sufficiently large to train a deep architecture efficiently from random initialization.

D. Discussion

The results demonstrate the potential of EfficientNetB3-based transfer learning for multi-class retinal fundus image classification. Previous investigations have similarly demonstrated the effectiveness of CNN-based approaches for diabetic retinopathy [2], [3], glaucoma [9], [10], and multi-class retinal disease classification [13], [14]. The present findings extend this approach by combining four-class classification with an input-validation mechanism and web-based implementation.

Nevertheless, the 99.58% test accuracy should be interpreted cautiously. The model was trained and evaluated using a moderately sized dataset from a limited source distribution. High performance on an internal held-out test set does not necessarily indicate equivalent performance on retinal images acquired from different populations, cameras, hospitals, or imaging conditions. Image-quality variations can also affect the reliability of automated retinal analysis [15].

Another important consideration is the difference between overall accuracy and the class-wise performance reported in Table IV. Although the model achieved high overall accuracy, individual class precision, recall, and F1-scores ranged approximately from 91% to 96%. Therefore, evaluation should consider the complete set of performance metrics rather than overall accuracy alone.



The integrated Clinical Validation Mesh and confidence thresholding provide additional safeguards against clearly unsuitable or low-confidence inputs. However, these mechanisms do not establish clinical diagnostic validity. External testing on independent, multi-centre datasets and prospective evaluation involving ophthalmology professionals are required before the proposed system can be considered for clinical screening applications.

VI. CONCLUSION AND FUTURE SCOPE

This study presented an automated deep-learning framework for the multi-class classification of retinal fundus images into cataract, diabetic retinopathy, glaucoma, and normal categories. The proposed approach employed transfer learning with a fine-tuned EfficientNetB3 backbone, supported by data augmentation, batch normalization, dropout regularization, and partial network fine-tuning. The developed model achieved an overall test accuracy of 99.58%, demonstrating strong performance on the dataset evaluated in this study.

Beyond model development, the classifier was integrated into the EyeDetect web application using a React-based frontend and Flask-based backend. The inclusion of a pre-inference Clinical Validation Mesh and confidence-based output handling extended the system from an isolated classification model into an end-to-end screening-support prototype capable of validating input images, performing automated classification, and presenting prediction results through an accessible interface.

Despite the encouraging results, the reported performance is based on a moderately sized dataset from a limited source distribution and should not be interpreted as evidence of clinical diagnostic accuracy. Variations in patient populations, imaging devices, illumination, image quality, and disease severity may influence model performance when applied to external datasets. Furthermore, the current classifier is restricted to three disease categories and a normal class.

Future research should therefore focus on external validation using larger, multi-centre and demographically diverse retinal datasets. The classification framework can also be extended to additional ocular conditions, including age-related macular degeneration and retinal detachment. Explainable artificial intelligence techniques, such as gradient-based class activation mapping, may be incorporated to visualize image regions contributing to individual predictions and improve model interpretability.

Further development should also investigate more robust retinal image-quality assessment, model optimization for mobile or resource-constrained environments, and prospective evaluation within realistic screening workflows. Most importantly, collaboration with qualified ophthalmologists will be necessary to evaluate the clinical relevance, reliability, usability, and limitations of the system before any consideration for real-world deployment.

Overall, the findings demonstrate the potential of EfficientNetB3-based transfer learning for automated multi-class retinal image classification and illustrate how a deep-learning model can be integrated with input validation and web-based deployment to create a practical research prototype for preliminary ocular screening support.

REFERENCES

1. World Health Organization. (2026, February 10). *Blindness and vision impairment*. World Health Organization.
2. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
3. Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., Yeo, I. Y. S., Lee, S. Y., Wong, E. Y. M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N. C., Finkelstein, E. A., Lamoureux, E. L., Wong, I. Y., Bressler, N. M., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*, 318(22), 2211–2223. <https://doi.org/10.1001/jama.2017.18152>



4. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (Vol. 97, pp. 6105–6114). Proceedings of Machine Learning Research.
5. Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., Tan, G. S. W., Schmetterer, L., Keane, P. A., & Wong, T. Y. (2019). Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, 103(2), 167–175. <https://doi.org/10.1136/bjophthalmol-2018-313173>
6. Ting, D. S. W., Peng, L., Varadarajan, A. V., Keane, P. A., Burlina, P. M., Chiang, M. F., Schmetterer, L., Pasquale, L. R., Bressler, N. M., Webster, D. R., Abramoff, M., & Wong, T. Y. (2019). Deep learning in ophthalmology: The technical and clinical considerations. *Progress in Retinal and Eye Research*, 72, 100759. <https://doi.org/10.1016/j.preteyeres.2019.04.003>
7. Islam, M. M., Yang, H.-C., Poly, T. N., Jian, W.-S., & Li, Y.-C. J. (2020). Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *Computer Methods and Programs in Biomedicine*, 191, 105320. <https://doi.org/10.1016/j.cmpb.2020.105320>
8. Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Scarpa, A., Fotiadis, D. I., & Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, 135, 104599. <https://doi.org/10.1016/j.compbiomed.2021.104599>
9. Li, Z., He, Y., Keel, S., Meng, W., Chang, R. T., & He, M. (2018). Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*, 125(8), 1199–1206. <https://doi.org/10.1016/j.ophtha.2018.01.023>
10. Christopher, M., Belghith, A., Bowd, C., Proudfoot, J. A., Goldbaum, M. H., Weinreb, R. N., Girkin, C. A., Liebmann, J. M., & Zangwill, L. M. (2018). Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Scientific Reports*, 8, 16685. <https://doi.org/10.1038/s41598-018-35044-9>
11. Atwany, M. Z., Sahyoun, A. H., & Yaqub, M. (2022). Deep learning techniques for diabetic retinopathy classification: A survey. *IEEE Access*, 10, 28642–28655.
12. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>
13. *Fundus image-based eye disease detection using EfficientNetB3 architecture*. (2025). *Journal of Imaging*, 11(8), 279.
14. Inan, M. S., et al. (2025). Multi-class classification of retinal eye diseases from ophthalmoscopy images using transfer learning-based vision transformers. *Journal of Imaging Informatics in Medicine*. <https://doi.org/10.1007/s10278-025-01416-7>
15. Crane, A. B., Choudhry, H. S., & Dastjerdi, M. H. (2024). Effect of simulated cataract on the accuracy of artificial intelligence in detecting diabetic retinopathy in color fundus photos. *Indian Journal of Ophthalmology*, 72(Suppl. 1), S42–S45. https://doi.org/10.4103/IJO.IJO_1163_23
16. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

