

# Optimized EfficientNet-B3 for Multiclass Diabetic Retinopathy Detection: A Deep Learning Framework with Hybrid Fine-Tuning and Confidence-Aware Inference



Swati Vaidya<sup>\*</sup>, Latika Jindal

Department of Computer Science & Engineering, Medicaps University, Indore 453331, India

Corresponding Author Email: [swati.dhage28@gmail.com](mailto:swati.dhage28@gmail.com)

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.310429>

## ABSTRACT

**Received:** 13 December 2025

**Revised:** 20 February 2026

**Accepted:** 11 April 2026

**Available online:** 30 April 2026

### Keywords:

*diabetic retinopathy, EfficientNet-B3, multiclass classification, transfer learning, retinal fundus images, deep learning*

Diabetic retinopathy (DR) is a leading cause of preventable blindness worldwide, and early detection is critical for effective treatment. However, automated multiclass DR grading remains challenging due to subtle inter-stage differences, image quality variations, and class imbalance in retinal fundus datasets. This paper proposes an optimized EfficientNet-B3 framework that integrates adaptive augmentation, staged progressive fine-tuning, hybrid optimization, and confidence-aware inference to address these challenges. The proposed framework employs a three-stage optimization strategy: (1) warm-up training of the classifier head only; (2) progressive unfreezing of MBConv blocks; and (3) full-network fine-tuning with AdamW optimizer, cosine learning rate decay, label smoothing, and mixed-precision training. A class-balanced sampling strategy with adaptive augmentation mitigates data imbalance. Confidence-aware inference combines maximum softmax probability with entropy-based uncertainty estimation for improved clinical interpretability. We evaluated the framework on a combined dataset of 92,501 retinal images from EyePACS, APTOS, APTOS-Gaussian filtered, and Messidor, using a balanced subset of 12,580 images for controlled experimentation (70/15/15 train/validation/test split). The proposed model achieved 99.00% accuracy, 98.64% precision, 98.80% recall, 98.72% F1-score, and 98.98% specificity for five-class DR grading, outperforming baseline ResNet-152 (91.96% accuracy) and standard EfficientNet-B3 (95.94% accuracy). The area under the ROC curve reached 1.00. The model requires approximately 12 million parameters and achieves inference times of a few milliseconds per image on GPU hardware. These results demonstrate that the proposed optimization strategies significantly improve multiclass DR grading performance while maintaining computational efficiency, suggesting practical viability for large-scale retinal screening programs and teleophthalmology applications.

## 1. INTRODUCTION

Diabetic retinopathy (DR) which is a chronic complication of diabetes mellitus (DM) is widely known as one of the most frequent reasons for vision loss and blindness on this planet [1]. Early and accurate diagnosis of DR is also important not only to halt further retina damage, but also to receive treatment in good time. Visual grading of retinal fundus photographs is the classic ophthalmological practice, however it is time consuming and subjective, as well as restricted by only a limited number of trained graders [2]. There is an apparent need for cost-effective, automated, robust diagnostic systems to deal with large number of tests in national screening programs already facing increasing prevalence of diabetes world-wide [3].

Figure 1 is a schematic cross-sectional view of the eye showing the pathological features of DR. It draws the retina with clearly marked pathologies, including microaneurysms, bleeds, hard exudates, and neovascularization, each identified in different shapes and colors to denote their clinical importance [4]. The optic disc, lens, and cornea are also well defined for orientation, showing how DR impacts the posterior

eye. The picture clearly shows how diabetes wreaks havoc on the eye's vasculature, leading to leakage, bleeding, lipid deposition, and the formation of abnormal blood vessels. The figure provides a labeled anatomical representation of retinal structures and DR-related lesions for visual interpretation and clinical reference. This image is ideal for comparing Normal vs. central retinal lesions and for learning how these lesions are involved in diseases.

In the last few years, deep learning has shown great promise for medical image analysis, particularly in vision-based tasks such as disease detection, segmentation, and classification. Architectures such as Convolutional Neural Networks (CNNs), U-Net variants, dense architectures, and GANs have yielded promising results in extracting features of pathological changes from retinal images. However, these methods frequently face difficulties in multiclass DR grading, including distinguishing between neighboring disease stages, robustness to noise and illumination changes, and generalization across datasets acquired from different devices.

In response to these limitations, the so-called EfficientNet family has recently attracted considerable interest for its balanced scaling of network depth, width, and resolution.

EfficientNet-B3 has been shown to offer the best trade-off between representational capacity and computational efficiency [5-44]. Yet, its multiclass DR detection capability is potentially more useful if built into architectural optimization, fine-tuning strategies, and robust preprocessing pipelines. This encourages the development of a better network that can inherit all desirable properties of EfficientNet-B3 while avoiding its weaknesses, such as sensitivity to low-quality images and class imbalance in medical databases.

In this work, the term “optimized EfficientNet-B3” refers not to a modification of the original EfficientNet-B3 backbone architecture itself, but rather to a set of task-specific optimization strategies developed for multiclass DR grading. These optimizations include adaptive augmentation and class balancing, progressive staged fine-tuning of MBConv layers, hybrid optimization using AdamW with cosine learning-rate scheduling, label smoothing, mixed-precision training, and cross-dataset generalization using heterogeneous retinal datasets. In addition, a confidence-aware inference mechanism based on entropy-guided uncertainty estimation is incorporated to improve diagnostic reliability and interpretability.

In this paper, we propose a refined deep learning method using the EfficientNet-B3 model for precise and reliable multiclass DR detection. This strategy focuses on improving the extraction of retinal features, achieving balanced training across multiple-source datasets, and searching for optimal hyperparameters to improve generalization. For comparison purposes, the model was trained and validated on a common dataset that includes images in the EyePacs, Aptos, Aptos (Gaussian Filtered), and Messidor datasets. Extensive experiments demonstrate that our method consistently outperforms traditional CNN-based and SOTA approaches, achieving significantly improved accuracy, sensitivity, specificity, and overall reliability in DR stage classification.

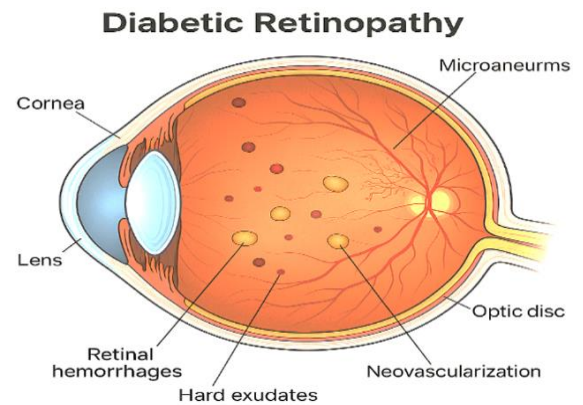
Key Contributions:

1. Design of a fine-tuned EfficientNet-B3 for multiclass DR diagnosis.
2. Combining heterogeneous retinal datasets (EyePacs, Aptos, Aptos-GF, Messidor) for strong cross-domain generalization.
3. Improved preprocessing and hyperparameter optimization for better feature extraction across different imaging conditions.
4. Competitive benchmarking against classical, hybrid, and state-of-the-art deep learning methods shows a 2-7% gain.
5. Introduction of a staged EfficientNet-B3 optimization strategy using progressive MBConv unfreezing and hybrid fine-tuning.
6. Integration of adaptive augmentation and weighted sampling for robust multiclass imbalance handling.
7. Confidence-aware inference using entropy-guided uncertainty estimation for improved clinical interpretability.

This study applies the optimized ResNet-152 framework as a baseline model for binary classifying patients into referable and non-referable DR cases. The main contribution of this work, however, is the proposed optimized EfficientNet-B3 architecture specifically built for multiclass DR grading. Adaptive augmentation, staged transfer learning, hybrid optimization and cross-dataset generalization strategies used in EfficientNet-B3 framework allowed this model to

effectively classify multiclass in various heterogeneous retinal datasets.

Beyond algorithmic performance improvement, the proposed framework is designed with practical clinical deployment considerations in mind, including scalability for mass retinal screening, compatibility with teleophthalmology systems, computational efficiency, and support for automated early referral decision-making in DR screening programs.



**Figure 1.** Detailed cross-sectional view of the human eye

## 2. LITERATURE REVIEW

Existing DR detection research can broadly be categorized into five major directions: traditional machine learning approaches, CNN-based classification models, transformer and hybrid deep learning architectures, retinal vessel segmentation techniques, and explainable AI frameworks for clinical deployment. While earlier machine learning methods relied heavily on handcrafted retinal features and statistical classifiers, recent deep learning approaches have significantly improved automated DR grading performance through hierarchical feature extraction and transfer learning. However, several limitations remain common across existing studies, including poor cross-dataset generalization, sensitivity to image-quality variation, class imbalance, computational complexity, and limited interpretability in clinical settings.

Existing DR detection research can generally be categorized into five major directions: (i) traditional machine learning and handcrafted feature-based approaches, (ii) Convolutional Neural Network (CNN)-based classification methods, (iii) transformer and hybrid deep learning architectures, (iv) retinal lesion segmentation and localization techniques, and (v) explainable and clinically interpretable artificial intelligence frameworks. The following review summarizes representative contributions in each category while highlighting their strengths, limitations, and remaining research challenges related to multiclass grading accuracy, cross-dataset robustness, computational efficiency, and clinical interpretability.

### 2.1 Traditional machine learning and early diabetic retinopathy detection methods

Retinal image interpretation for early DR detection systems were mainly based on handcrafted feature extraction, statistical analysis and conventional machine learning approaches. Seoud et al. [9] developed a framework to detect red lesion in retinal fundus image based on dynamic shape

feature for microaneurysms and hemorrhages detection. Their approach achieved competitive lesion localization performance on both datasets and underscored the potential of using lesion-specific feature engineering in the staging of early DR. Antal and Hajdu [22] used an ensemble-based microaneurysm detection system, combining preprocessing and candidate extraction modules to help with the grading of DR. Their method obtained competitive results on the Messidor dataset and laid practical evidence of ensemble learning as a viable tool for retinal abnormality detection.

In addition, other researchers have seized upon image processing and feature-based classification strategies for retinal analysis. Barkana et al. Descriptive statistical features were used in study [34] to segment retinal vessels with fuzzy logic, artificial neural networks (ANNs) and support vector machines (SVMs), and classifier fusion methods. They found that information from the structure of vessels is essential to analyze retinal disease. A threshold-based segmentation technique has been proposed by Otsu [40] that employs gray-level precipitation and was one of the most used supervised preprocessing methods to segment images before working with retinal pictures. Kaur and Kaur [41] suggested a computer-aided framework for DR diagnosis which integrates retinal image segmentation with K-nearest neighbor classification. Their study showed acceptable diagnostic performance of traditional machine learning models if leveraged with a good retinal feature extraction approach.

While these classical techniques made significant foundational contributions to the automation of retinal image analysis, they typically had challenges such as limited scalability, reliance on handcrafted features, susceptibility to variation in illumination and poor generalization over heterogeneous datasets. The above drawbacks were the impetus for moving towards feature extraction and classification methods using deep learning-based approaches.

## 2.2 Convolutional Neural Network-based and transfer learning approaches

As deep learning matured, CNNs took center stage for automated DR detection and grading. Qiao et al. [3] provides a semantic segmentation framework using CNN for microaneurysms detection and non-proliferative DR diagnosis. Their approach proved useful in extracting retinal structures and in identifying early stages of diseases. Atwany et al. [6] proposed a comparative literature review of various deep learning techniques to classify DR into different severities, focusing on the successes and limitations of supervised CNNs, transfer learning, and transformer-based methods; furthermore, dataset imbalance and clinical trustworthiness are reviewed.

Different studies have been developed in order to improve the performance of DR grading using multitask learning and transfer learning strategies. Wang et al. With the aim of jointly predicting both DR severity and retinal lesion features, the study [11] presented a hierarchically multitask deep learning framework to improve interpretability and grading consistency. Majumder and Kehtarnavaz [12] presented a multitasking deep learning architecture for five-stage DR grading which consists of classification and regression branches, and was trained on Eye Picture Archive Communication System (EyePACS) and Asia Pacific Tele-Ophthalmology Society (APTOS) datasets, producing strong Receiver Operating Characteristic (ROC) and kappa

performance results. Nazir et al. The Weighted Ensemble model that was proposed by the study [17] is used for improving DR detection in early stages under the minimal size of data conditions and exploiting Inception-v3, VGG16, and a custom CNN to enhance classification accuracy while maintaining robustness.

Transfer learning and feature optimization methods have also been studied extensively. Wong et al. A transfer learning framework called Ensemble of tiled features and Error Correcting Output Codes (ECOC) for DR grading [19] used ShuffleNet with ResNet-18 features combined in a final ensemble classification using ECOC, which led to competitive performance against the EyePACS and Messidor datasets. Ahnaf Alavee et al. [20] compared various transfer learning architectures and used the model combined with explainable AI techniques (e.g., Grad-CAM) for clinical interpretability. The four-class DR diagnostic system proposed by the study [31] based on deep CNN shows a good agreement with ophthalmologist assessment in real clinical environments. In the study [32], a hybrid CNN-SVD feature extraction model along with an Extreme learning machine (ELM) classifier reached very high binary and multiclass classification accuracy on APTOS and Messidor datasets respectively. In addition, Wang et al. [33] explored the simultaneous prediction of DR severity and lesions with different grading strategies to enhance grading consistency and interpretability.

Despite their successes, CNN-based approaches remain limited with the issues of over-fitting [12], being sensitive to imaging variability, low cross-dataset robustness [11] and poor ability to discriminate neighbouring DR severity stages (e.g. mediums between moderate-level referable DR).

## 2.3 Transformer, hybrid, and multimodal deep learning models

In the recent DR studies, transformer architectures, hybrid deep learning models with CNN backbones and multimodal frameworks have been increasingly investigated to address some of the limitations present in conventional CNNs. In the study [4], a transformer-assisted segmentation and classification framework was introduced that fused retinal structure extraction and lesion detection as an additional task with the goal of improving DR diagnosis. Zedadra et al. [5] developed a graph-aware multimodal deep learning approach of combining retinal fundus images with the concurrent patient clinical history utilizing graph neural networks and DenseNet based feature extraction, which showed evidence for augmented diagnostic reliability across both APTOS2019 and Messidor-2 datasets specifically.

Other Vision Transformer (ViT)-based DR grading methods have been effective. For instance, Nazih et al. [7] applied a pre-trained ViT model on the Fine-Grained Annotated Diabetic Retinopathy Dataset (FGADR) dataset to perform multiclass DR severity grading with significant global retinal context preservation compared to traditional CNN architectures. MSAmix-Net [15] combines multiscale convolution and self-attention mechanisms in such a way that the both local and long-range retinal dependencies are estimated to reach strong performance on APTOS, Messidor-2 and DRAC-2022 datasets.

There have also been several hybrid methods which have enhanced robustness and classification accuracy. Mutawa et al. To resolve the problem of image-quality variability and dataset imbalance, a randomization-driven hybrid deep

learning framework combining CNN-RBF models and advanced preprocessing techniques was presented in the study [16]. Henge et al. [21] worked on a multi-decision Inception-ResNet hybrid architecture with deep feature fusion and chi-square optimization that resulted in high diagnostic accuracy for DR screening. In the study by Jabbar et al. [23], a Hadoop/HDFS distributed DR detection system based on lesion approach was proposed, which combines GoogleNet architecture with adaptive particle swarm optimization to extract features. The DRNet [24] was a multi-branch temporal convolutional network on longitudinal electronic health records for predicting the risk of DR and reported significant improvements in AUROC and F1-score.

These transformers and hybrid architectures boost feature representation and contextual understanding but usually come with the significant drawbacks of requiring large-scale datasets, increased computational resources, and need for careful optimization to prevent overfitting while preserving clinical applicability.

#### 2.4 Retinal vessel segmentation and lesion analysis methods

The first step in accurately diagnosing DR is the segmentation of retinal vessels and localization of lesions. For example, Yi et al. [29] presented a compound-scale encoder-decoder network with both spatial and channel attention mechanisms for retinal lesion segmentation, showing strong performance on the DDR482 and FGADR376 datasets. In the study by Upreti et al. [30], blood vessel segmentation and lesion detection by a deep learning-based retinal analysis framework was performed, showing that for the automated diagnosis of microaneurysms, hemorrhages were more effectively detected with well-defined vessels (more comparable feature extraction from fundus).

Generative adversarial networks (GANs) and deformations of U-Net have been popular choices for retinal vessel segmentation as well. RV-GAN was proposed in the study [36] as a multiscale generative adversarial framework that can handle different illumination and different vessel thickness. Ma and Li [37] presented an attention-based U-Net (A-Unet) algorithm to extract retinal vessels, which greatly increased segmentation accuracy under complex imaging conditions. Kumar et al. Results: Glaucomatous optic disc evaluation, namely retinal vessel and optic disc segmentation approach: the idea was noted with an automated retinal vessel and optic disc segmentation framework for early DR identification using further optimized preprocessing measures in addition to adaptive deep learning algorithms [38].

For better vessel segmentation performance, dense-U-Net and modified U-shaped networks were proposed. Li et al. A Dense-U-Net-based retinal vessel segmentation network structure with superior multiscale feature extraction ability is proposed in the study by Li et al. [39]. In the study by Tan et al. [42], a deep matched filtering approach was proposed for retinal vessel segmentation with superior preservation of fine-vessel structures and improved segmentation robustness to varying image quality. He et al. A modified U-shaped network for retinal vascular structure extraction was proposed in the study [43]. Soomro et al. [44] tested and compared several ICA-based methods for retinal image preprocessing, showing that the utilization of proper image preprocessing can positively improve vessel detection accuracy and downstream DR classification performance.

While segmentation-based methods enhance tumor localization and interpretability, many of the approaches still require computationally intensive techniques that can be highly susceptible to artifacts in low-quality retinal images and imaging-device variability.

#### 2.5 Explainable AI and clinical translation challenges

As AI systems become increasingly integrated in clinical ophthalmology, research priorities now enthrall over explainability, reliability and the clinical translation of such models. Abushawish et al. [1] provided a complete overview of deep learning methods for DR detection and grading, focusing on the need for explainable AI and model interpretability for clinical deployment. A review by Rajarajeshwari and Selvi [2] summarized the current use of AI in DR classification, segmentation, diagnosis, and grading, outlining challenges related to robustness, transparency within algorithms, and clinical deployment.

Multiple related studies addressed explainable and trustworthy AI systems for retinal diagnosis. Asif et al. [8] presented a systematic review for the automated DR diagnosis approaches based on AI and raised dataset bias, limited explainability, privacy problems, and scalability issues. Ikram et al. [10] reviewed the state of novel DR detection methods and several challenges for clinical translation: insufficient robustness, lack of external validation. This was thoroughly tested in a work by the study [13] on arbitrary adversarial exposure attacks from an attack that rescales the image having previously placed it under illumination manipulation conditions to test deep learning-based DR grading systems using retinal fundus images, finding major vulnerabilities.

Some frameworks for explainable AI have also been proposed to enhance clinician trust in automated systems. With interpretable predictions and less diagnostic uncertainty, Shahzad et al. [14] constructed an explainable DR diagnosis framework. An AI-assisted teleophthalmology application based on deep neural networks to assist retinal specialists in DR screening and remote clinical support was described by Ghouali et al. [25]. To solve class imbalance and improve recognition accuracy in imbalanced retinal datasets, Naz et al. [26] proposed a DCGAN-based ensemble framework. Mateen et al. [27] conducted a review on automated DR detection methodologies highlighting difficulties in varying quality of images, datasets and transferability to the real-world scenario. DcardNet [28] is a densely connected CNN architecture designed for multilevel DR grading utilizing Optical Coherence Tomography (OCT) and Optical Coherence Tomography Angiography (OCTA) imaging modalities, achieving robust clinical applicability with automated referral rough hierarchical severity detection. While considerable advances have been made, previous work continues to grapple with issues regarding clinical interpretability, cross-dataset generalization, robustness against diverse imaging conditions and safety for deployment in real-world ophthalmic screening systems. These limitations provide the impetus to optimize generalizable, explainable and computationally efficient frameworks for DR grading espoused by current study with EfficientNet-B3 architecture.

Overall, traditional machine learning methods demonstrated the importance of retinal feature engineering but suffered from limited scalability and poor robustness under heterogeneous imaging conditions. CNN-based methods improved feature extraction and classification accuracy substantially; however,

many approaches still struggle with neighboring severity-stage discrimination and dataset imbalance. Transformer and hybrid architectures further enhanced contextual learning capability but introduced increased computational complexity and dependency on large-scale training data. Segmentation-based approaches improved lesion localization and interpretability, although many remain sensitive to illumination variation and low-quality retinal images. Recent explainable AI frameworks have improved transparency and clinician trust, yet external clinical validation and real-world deployment remain major challenges across most existing DR detection systems.

Although substantial progress has been achieved in automated DR detection, existing studies still face critical limitations related to multiclass grading consistency, cross-domain robustness, class imbalance handling, computational efficiency, and clinical interpretability. Furthermore, many methods are evaluated on isolated datasets, limiting their generalization capability in real-world screening environments. Motivated by these limitations, the present study proposes an optimized EfficientNet-B3 framework incorporating adaptive augmentation, staged fine-tuning, hybrid optimization, and cross-dataset learning strategies to improve multiclass DR grading performance and robustness across heterogeneous retinal datasets.

### 3. METHODOLOGY

The holistic approach is structured as a single framework for DR analysis using two integrated evaluation stages, rather than as two independent studies. In the first phase, we used an optimized ResNet-152 model as a baseline binary screening tool to differentiate cases with Non-Referable DR (NRG) and Referable DR (RG). This first stage assesses the ability to detect generalised retinal abnormalities and provides a comparative baseline performance. The second stage is a proposed optimized EfficientNet-B3 architecture for fine-grained multiclass DR grading between five severity levels. Therefore, the binary classification framework is intended as a supporting benchmark experiment and EfficientNet-B3 multiclass grading model can be considered the main contribution of this work.

### 3.1 Baseline binary classification using optimized ResNet-152

Figure 2 model first takes a  $224 \times 224 \times 3$  fundus image as input and applies a  $7 \times 7$  convolutional layer with rectified linear unit (ReLU) activation and trainable parameters, with filters that reduce spatial resolution and preserve fundamental retinal structures. The feature extractor then processes the activation maps through four hierarchical layers of convolution inspired in deep residual architectures: Conv2\_x (3 blocks,  $56 \times 56 \times 256$ ), Conv3\_x (8 blocks,  $28 \times 28 \times 512$ ), Conv4\_x (36 blocks,  $14 \times 14 \times 1024$ ) and Conv5\_x (3 blocks,  $7 \times 7 \times 2048$ ), progressively increasing the number of channels and reducing spatial characteristics to learn rich representations for pathological textures like microaneurysms, exudates and vessel abnormality. A global average pooling (GAP) layer transforms the final  $7 \times 7 \times 2048$  tensor into a compact  $1 \times 1 \times 2048$  feature vector, which is followed by a fully connected classification layer that generates two class logits for NRG (Normal Retina Group) and RG (Retinopathy Group). The resulting prediction is subsequently used to determine whether the fundus image indicates retinopathy.

The proposed algorithm uses a pre-trained ResNet-152 model, fine-tuned for binary DR classification (RG and NRG) using transfer learning from fundus images. The algorithm starts by preparing and augmenting the dataset using traditional normalisation, as in ImageNet pre-training. We freeze the convolutional backbone of ResNet-152 to retain rich feature representations, and fine-tune only the final fully connected layer to adapt to a particular task. The model is trained over epochs using the Adam optimiser and a cross-entropy loss, while performance is tracked on an independent validation set. After training, the model is tested using a set of metrics, e.g., accuracy, precision, recall, specificity, F1-score, ROC-AUC, etc. Visualization features, such as confusion matrices, ROC curves, and training dynamics, give the user some intuition about what has happened. This combined stepwise-pseudocode recipe framework is a strong and reproducible method for deploying a DNN model for medical imaging with scarce data.

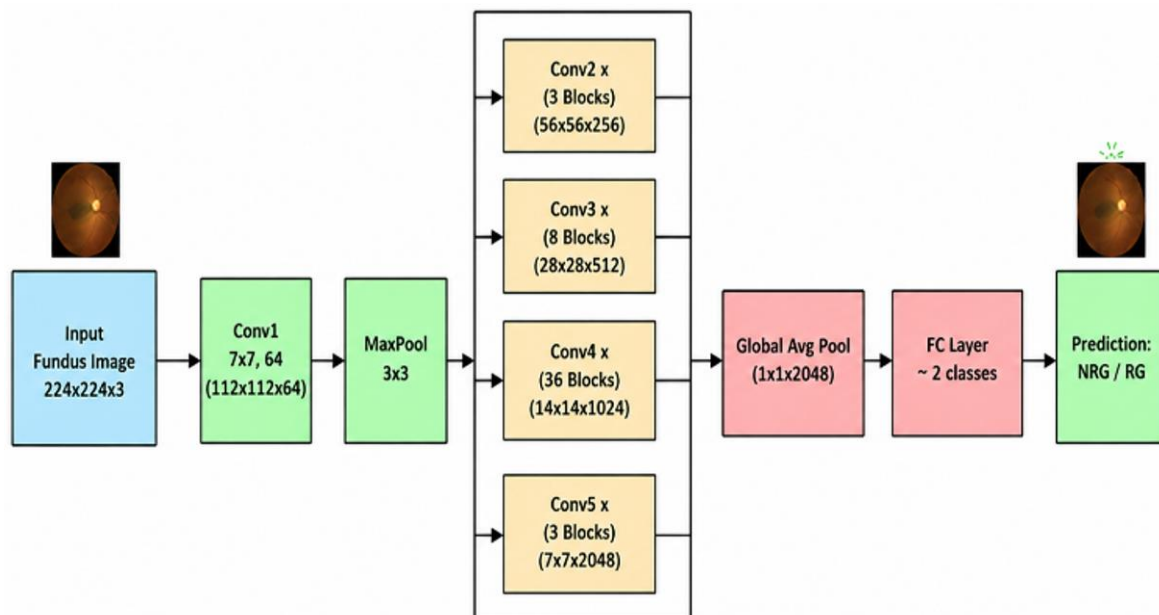


Figure 2. Architecture of optimised ResNet-152

Although the ResNet-152 framework demonstrates effective binary retinal disease screening performance, binary classification alone is insufficient for detailed clinical severity assessment. Therefore, to achieve fine-grained DR grading and improved diagnostic capability, the proposed optimized EfficientNet-B3 framework is introduced in the next stage as the primary multiclass classification model of this study.

### 3.2 Proposed multiclass diabetic retinopathy grading using optimized EfficientNet-B3

The proposed optimized EfficientNet-B3 framework extends the standard pretrained EfficientNet-B3 model through multiple task-specific optimization strategies tailored for DR grading. Instead of altering the backbone architecture, the optimization focuses on improving feature learning, training stability, class balance, and cross-dataset generalization. The framework integrates adaptive augmentation, weighted sampling, staged MBConv fine-tuning, hybrid optimization with AdamW and cosine scheduling, label smoothing, mixed-precision learning, and confidence-aware inference mechanisms for robust multiclass retinal disease classification.

#### Algorithm 1. Optimized ResNet-152 for diabetic retinopathy classification using transfer learning

**Input:** Dataset  $\mathcal{D} = \{\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, \mathcal{D}_{\text{test}}\}$ , epochs  $E = 40$ , batch size  $B = 32$ , image size  $S = 224$ .

**Output:** Trained classifier  $\mathcal{M}$  and performance metrics set  $\Omega$ .

##### Step-Wise Procedure

##### Step 1: Device Initialization

1.1 Select the computational device:

$$\text{device} = \begin{cases} \text{cuda,} & \text{if GPU is available} \\ \text{cpu,} & \text{otherwise} \end{cases}$$

##### Step 2: Preprocessing and Transform Definition

2.1 Define a preprocessing pipeline  $\mathcal{T}(\cdot)$  applied to each input image  $\mathbf{I}$ :

$$\mathcal{T}(\mathbf{I}) = \text{Norm}(\text{ToTensor}(\text{Resize}(\mathbf{I}, S \times S)))$$

2.2 Normalize the tensor  $\mathbf{x}$  channel-wise:

$$\text{Norm}(\mathbf{x}) = \frac{\mathbf{x} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\sigma}$  denote mean and standard deviation.

##### Step 3: Dataset Preparation

3.1 Apply transformation  $\mathcal{T}$  to split datasets:

$$\begin{aligned} \mathcal{D}_{\text{train}} &\leftarrow \mathcal{T}(\mathcal{D}[\text{train}]) \\ \mathcal{D}_{\text{val}} &\leftarrow \mathcal{T}(\mathcal{D}[\text{val}]) \\ \mathcal{D}_{\text{test}} &\leftarrow \mathcal{T}(\mathcal{D}[\text{test}]) \end{aligned}$$

##### Step 4: Mini-Batch Loader Construction

4.1 Construct training and validation loaders:

$$\begin{aligned} \mathcal{L}_{\text{train}} &= \text{DataLoader}(\mathcal{D}_{\text{train}}, B, \text{shuffle}=\text{True}) \\ \mathcal{L}_{\text{val}} &= \text{DataLoader}(\mathcal{D}_{\text{val}}, B, \text{shuffle}=\text{False}) \end{aligned}$$

##### Step 5: Model Initialization and Modification

5.1 Load pretrained ResNet-152 backbone:

$$\mathcal{M} \leftarrow \text{ResNet152}(\text{pretrained}=\text{True})$$

5.2 Partition model parameters:

$$\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\text{conv}}, \boldsymbol{\theta}_{\text{fc}}\}$$

5.3 Freeze convolutional parameters:

$$\nabla_{\boldsymbol{\theta}_{\text{conv}}} \mathcal{L} = 0$$

5.4 Replace the final fully connected (FC) layer for binary classification:

$$\mathbf{z} = \mathbf{W}\mathbf{h} + \mathbf{b}, \mathbf{z} \in \mathbb{R}^2$$

where  $\mathbf{W} \in \mathbb{R}^{2 \times d}$ ,  $\mathbf{b} \in \mathbb{R}^2$ , and  $\mathbf{h} \in \mathbb{R}^d$  is the extracted feature vector.

#### Step 6: Model Deployment to Device

6.1 Move the model to the selected device:

$$\mathcal{M} \leftarrow \mathcal{M}(\text{device})$$

#### Step 7: Loss and Optimizer Definition

7.1 Define cross-entropy loss for multi-class logits (binary case):

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \log \left( \frac{\exp(z_{i,y_i})}{\sum_{k=1}^2 \exp(z_{i,k})} \right)$$

7.2 Configure Adam optimizer for FC layer only:

$$\boldsymbol{\theta}_{\text{fc}} \leftarrow \boldsymbol{\theta}_{\text{fc}} - \eta \cdot \text{Adam}(\nabla_{\boldsymbol{\theta}_{\text{fc}}} \mathcal{L})$$

with learning rate  $\eta = 10^{-3}$ .

#### Step 8: Training and Validation Loop

For each epoch  $e = 1, 2, \dots, E$ :

##### Step 8.1: Training Phase

8.1.1 Enable training mode:

$$\mathcal{M} \leftarrow \text{train}()$$

8.1.2 For each mini-batch  $(\mathbf{X}, \mathbf{y}) \in \mathcal{L}_{\text{train}}$ :

**Forward pass:**

$$\hat{\mathbf{y}} = \mathcal{M}(\mathbf{X})$$

**Compute loss:**

$$\ell = \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$$

**Backpropagation update:**

$$\boldsymbol{\theta}_{\text{fc}} \leftarrow \boldsymbol{\theta}_{\text{fc}} - \eta \nabla_{\boldsymbol{\theta}_{\text{fc}}} \ell$$

##### Step 8.2: Validation Phase

8.2.1 Enable evaluation mode:

$$\mathcal{M} \leftarrow \text{eval}()$$

8.2.2 For each mini-batch  $(\mathbf{X}_v, \mathbf{y}_v) \in \mathcal{L}_{\text{val}}$ :

• **Predict class labels:**

$$\tilde{y}_v = \arg \max_{k \in \{0,1\}} \hat{y}_{v,k}$$

• **Compute validation accuracy:**

$$\text{Acc}_v = \frac{1}{N_v} \sum_{i=1}^{N_v} \mathbb{I}(\tilde{y}_{v,i} = y_{v,i})$$

• Update confusion matrix terms  $TP, FP, TN, FN$ .

##### Step 8.3: Record and Print

8.3.1 Save training/validation logs:

$$\{\text{loss}_{\text{train}}[e], \text{loss}_{\text{val}}[e], \text{acc}_{\text{train}}[e], \text{acc}_{\text{val}}[e]\}$$

8.3.2 Print epoch-wise progress:

$$\text{Epoch } e: \text{TrainAcc}=\text{acc}_{\text{train}}[e], \text{ValAcc}=\text{acc}_{\text{val}}[e]$$

#### Step 9: Post-Training Evaluation

9.1 Generate prediction scores:

$$\hat{\mathbf{y}} = \mathcal{M}(\mathcal{D}_{\text{val}})$$

9.2 Form confusion matrix:

$$\mathbf{C} = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}$$

9.3 Compute evaluation measures:

$$\text{Accuracy:} \quad \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision:} \quad \text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall:} \quad \text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1-Score:} \quad F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Specificity:} \quad \text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{ROC-AUC:} \quad \text{AUC} = \int_0^1 \text{TPR}(FPR) d(FPR)$$

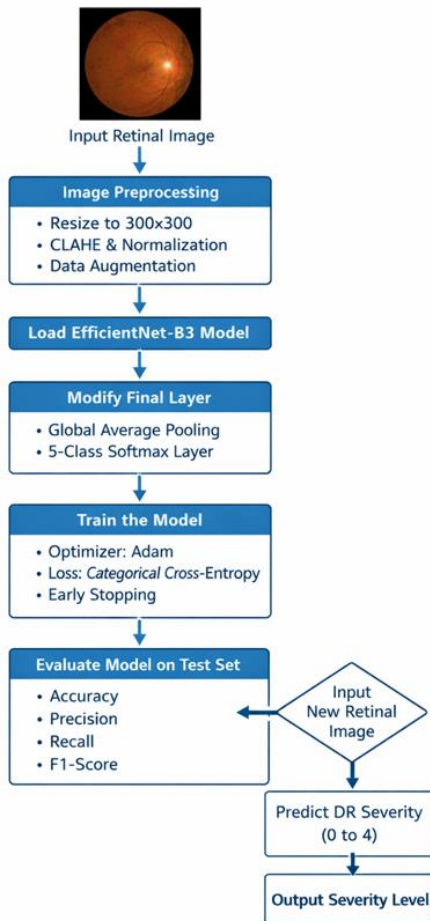
**Step 10: Visualization**

10.1 Plot:

- Confusion matrix
- ROC curve
- Accuracy vs. epochs
- Loss vs. epochs

**Step 11: Output**11.1 Return the trained model and metric set:  $\{\mathcal{M}, \Omega\}$ 

Figure 3 shows the EfficientNet-B3 pipeline takes a preprocessed fundus photograph as input. A first STEM convolutional stage extracts low-level retinal features from a  $300 \times 300 \times 3$  image. The representation is then fed to the first MBConv block (repeated 3 times), which reduces the spatial size to  $150 \times 150 \times 24$  and enhances structure-related features, such as vessel and disc contours. And a subsequent deeper MBConv stage with 10 expanded blocks further compresses the feature maps to  $75 \times 75 \times 32$  and learns more complex pathological textures, such as microaneurysms, exudates, and haemorrhages. The pipeline proceeds with a BlockConv group (3 occurrences) that reduces the retina's representation to  $38 \times 38 \times 96$ , enabling us to perform reliable multiscale abstraction over the retinal field. Finally, the output is forwarded to the classifier head with GAP, which computes a compact  $1 \times 1 \times 96$  embedding of the spatial tensor, which is then fed through a fully connected layer to produce two-class logits. The predicted labels of these logits are given here as (T: 0 = True Class = Normal, P: 0 = Predicted = Normal) with a confidence of 88%. Thus, the EfficientNet B3 retinal disease classification workflow is done.

**Figure 3.** Architecture of optimized EfficientNet-B3**Algorithm 2. Optimized EfficientNet-B3-Based retinal disease classification pipeline**

The adaptive augmentation strategy is performed only during the training phase and is not applied to validation or testing datasets in order to avoid information leakage and preserve unbiased performance estimation.

**Algorithm 2.1** Data preparation and adaptive augmentation**Input:** Raw fundus image dataset

$$\mathcal{D} = \{(\mathbf{X}_i, Y_i)\}_{i=1}^N, Y_i \in \{1, 2, \dots, K\}$$

**Output:** Balanced and augmented dataset  $\mathcal{D}_b$  and class-balanced sampler  $\mathcal{S}$ **Step 1: Resize and Adaptive Augmentation****For each**  $(\mathbf{X}_i, Y_i) \in \mathcal{D}$ , perform:**1.1 Resize**

$$\mathbf{X}_i \leftarrow \text{Resize}(\mathbf{X}_i, 300 * 300)$$

**Step 2: Compute Class Frequencies**Define class frequency for class  $k$  as:

$$f_k = \sum_{i=1}^N \mathbb{I}(Y_i = k), k \in \{1, 2, \dots, K\}$$

Define the class-count vector:

$$\mathbf{C} = [f_1 \quad f_2 \quad \dots \quad f_K]$$

Total number of samples:

$$N = \sum_{k=1}^K f_k$$

**Step 3: Minority-Class Condition**A sample  $(\mathbf{X}_i, Y_i)$  is treated as minority if:

$$f_{Y_i} < \tau$$

where  $\tau$  is the minority threshold.**Step 4: Augmentation Rule**The augmentation applied to  $\mathbf{X}_i$  is:

$$\tilde{\mathbf{X}}_i = \begin{cases} \mathcal{A}_{\text{strong}}(\mathbf{X}_i), & \text{if } f_{Y_i} < \tau, \\ \mathcal{A}_{\text{mild}}(\mathbf{X}_i), & \text{otherwise.} \end{cases}$$

Strong augmentation operator:

$$\mathcal{A}_{\text{strong}} = \{ \text{RandomResizedCrop}, \text{Flip}, \text{Rotation}, \text{ColorJitter} \}$$

Mild augmentation operator:

$$\mathcal{A}_{\text{mild}} = \{ \text{Resize}, \text{Normalize} \}$$

**Step 5: Compute Class Weights (Inverse Frequency)**

Class weights are computed as:

$$W_k = \frac{N}{f_k}, k \in \{1, 2, \dots, K\}$$

Weight vector:

$$\mathbf{W} = [W_1 \quad W_2 \quad \dots \quad W_K]$$

**Step 6: Assign Sample Weights**

Each sample weight is assigned as:

$$w_i = W_{Y_i}, i = 1, 2, \dots, N$$

Weight collection:

$$\mathbf{w} = [w_1 \quad w_2 \quad \dots \quad w_N]^T$$

**Step 7: Weighted Random Sampling**Sampling probability for each sample  $i$  is:

$$p_i = \frac{w_i}{\sum_{j=1}^N w_j}, i = 1, 2, \dots, N$$

The sampler  $\mathcal{S}$  is defined as:  
 $\mathcal{S} = \text{WeightedRandomSampler}(\mathbf{w})$

#### Step 8: Final Balanced Dataset

The augmented balanced dataset is formed as:

$$\mathcal{D}_b = \{(\tilde{\mathbf{X}}_i, Y_i)\}_{i=1}^N$$

#### Step 9: Return

Return:

$$(\mathcal{D}_b, \mathcal{S})$$

**Data Preparation & Adaptive Augmentation:** The first stage of the pipeline is to create a high-quality, balanced, and diverse dataset for training models. For all retinal fundus images, we resize them to  $300 \times 300$  pixels and normalize them as input channels for EfficientNet. An adaptive augmentation is used to mitigate dataset imbalance: weak augmentation (crops, rotations, vertical and horizontal flips, color jitter, etc.) is applied to majority-class samples after they are resized and normalised according to the usual procedure, and substantial random augmentation is used to minority-class samples. This preserves constrained diversity without altering the class distribution. The tentative sampling policy is defined based on the inverse of class frequency, so that each batch of training samples contributes equally to model learning. By this statistical dynamic balance-augmentation hybrid, the dataset becomes statistically balanced and visionarily enriched, thereby improving generalization performance while alleviating overfitting.

**Hybrid EfficientNet-B3 Optimisation Pipeline:** The second stage introduces a three-step hybrid optimisation strategy to enhance feature extraction in EfficientNet-B3 without sacrificing compact computation. Only the classifier head is trained during the first phase (warm-up) to allow the network to converge with minimal disturbance to pretrained weights. In the fourth stage, we progressively unfreeze the later MBConv blocks to fine-tune specific retinal features. In the final step of training, we unfreeze the entire backbone and optimize using AdamW with cosine learning rate decay, label smoothing, and mixed precision to achieve stable convergence. A stopping criterion is also implemented to check validation performance and save the state when the best model is reached. This staged fine-tuning enables optimal alignment with retinal pathology patterns and alleviates catastrophic forgetting.

#### Algorithm 2.2 Hybrid EfficientNet-B3 optimization strategy

**Input:** Balanced dataset  $\mathcal{D}_b$ , pretrained EfficientNet-B3 model  $\mathcal{M}$ , total epochs  $E$ , warmup epochs  $E_w$ , number of unfrozen blocks  $K$ , patience limit  $P_{\max}$ .  
**Output:** Optimized model  $\mathcal{M}^*$ .

#### Step 1: Model Initialization

1.1 Initialize EfficientNet-B3 with pretrained weights:

$$\mathcal{M} \leftarrow \mathcal{M}_{\theta}^{\text{pre}}$$

1.2 Partition parameters into backbone and classifier head:

$$\theta = \{\theta_{\text{bb}}, \theta_{\text{head}}\}$$

1.3 Freeze backbone parameters:

$$\nabla_{\theta_{\text{bb}}} \mathcal{L} = 0 \Rightarrow \theta_{\text{bb}} \text{ fixed}$$

#### Step 2: Loss Function With Label Smoothing

2.1 For logits  $\mathbf{z}_i = \mathcal{M}(\mathbf{x}_i)$  and predicted probability:

$$\hat{\mathbf{p}}_i = \text{softmax}(\mathbf{z}_i)$$

2.2 Define smoothed ground-truth distribution for class  $y_i$ :

$$\tilde{p}_{i,k} = \begin{cases} 1 - \varepsilon, & k = y_i, \\ \frac{\varepsilon}{C - 1}, & k \neq y_i, \end{cases}$$

where  $C$  is the number of classes and  $\varepsilon \in (0,1)$  is the smoothing factor.

2.3 Label-smoothed cross entropy loss:

$$\mathcal{L}_{\text{LS}} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^C \tilde{p}_{i,k} \log(\hat{p}_{i,k})$$

#### Step 3: Optimizer Definition

3.1 Define AdamW optimizer (only train head initially):

$$\mathcal{O} \leftarrow \text{AdamW}(\theta_{\text{head}}, \eta_0, \lambda)$$

where  $\eta_0$  is the initial learning rate and  $\lambda$  is the weight decay.

#### Step 4: Warmup + Cosine Learning Rate Schedule

4.1 Define total steps per epoch as  $T$ .

4.2 Learning-rate scheduler:

$$\eta(e) = \begin{cases} \eta_0 \cdot \frac{e}{E_w}, & 1 \leq e \leq E_w, \\ \eta_{\min} + \frac{1}{2}(\eta_0 - \eta_{\min}) \left[ 1 + \cos\left(\pi \cdot \frac{e - E_w}{E - E_w}\right) \right], & E_w < e \leq E. \end{cases}$$

#### Step 5: Training Loop

Initialize:

$$\text{BestAcc} \leftarrow 0, \text{patience} \leftarrow 0$$

**For each epoch  $e = 1, 2, \dots, E$ :**

#### Step 5.1: Update Learning Rate

5.1.1 Update learning rate:

$$\eta \leftarrow \eta(e)$$

#### Step 5.2: Warmup Training Phase

5.2.1 If  $e \leq E_w$ , train only classifier head:

$$\theta \leftarrow \{\theta_{\text{bb}} \text{ frozen}, \theta_{\text{head}} \text{ trainable}\}$$

#### Step 5.3: Adaptive Unfreezing After Warmup

5.3.1 If  $e = E_w + 1$ , unfreeze the last  $K$  MBConv blocks:

$$\theta_{\text{bb}}^{(K)} \leftarrow \text{trainable}$$

5.3.2 Reinitialize optimizer with reduced learning rate:

$$\mathcal{O} \leftarrow \text{AdamW}(\{\theta_{\text{head}}, \theta_{\text{bb}}^{(K)}\}, \alpha\eta_0, \lambda)$$

where  $\alpha \in (0,1)$  is the learning-rate reduction factor.

#### Step 5.4: Forward and Loss Computation

For each mini-batch  $(\mathbf{x}, \mathbf{y}) \in \mathcal{D}_b$ :

5.4.1 Forward pass:

$$\mathbf{z} = \mathcal{M}(\mathbf{x})$$

5.4.2 Compute prediction distribution:

$$\hat{\mathbf{p}} = \text{softmax}(\mathbf{z})$$

5.4.3 Compute loss:

$$\ell = \mathcal{L}_{\text{LS}}(\hat{\mathbf{p}}, \mathbf{y})$$

#### Step 5.5: Mixed-Precision Backpropagation (AMP)

5.5.1 Use AMP scaling:

$$\ell_{\text{scaled}} = s \cdot \ell$$

where  $s$  is the dynamic scaling factor.

5.5.2 Gradient update:  
 $\theta \leftarrow \theta - \eta \cdot \mathcal{O}(\nabla_{\theta} \ell_{\text{scaled}})$

**Step 5.6: Validation and Accuracy Measurement**

5.6.1 Compute validation accuracy:

$$\text{Acc}_{\text{val}} = \frac{1}{N_v} \sum_{i=1}^{N_v} \mathbb{I}(\arg \max(\hat{p}_i) = y_i)$$

**Step 5.7: Best Model Selection and Early Stopping**

5.7.1 If  $\text{Acc}_{\text{val}} > \text{BestAcc}$ :

$\text{BestAcc} \leftarrow \text{Acc}_{\text{val}}, \mathcal{M}^* \leftarrow \mathcal{M}, \text{patience} \leftarrow 0$

5.7.2 Else:

$\text{patience} \leftarrow \text{patience} + 1$

5.7.3 If  $\text{patience} \geq P_{\text{max}}$ , terminate training:  
 break

**Step 6: Return Best Model**

Return:  $\mathcal{M}^*$

**Algorithm 2.3** Inference, confidence scoring, and diagnostic decision

**Input:** Trained model  $M^*$ , test image  $X_{\text{test}}$

**Output:** Predicted class  $\hat{Y}$ , confidence score  $C$ , and Diagnostic Confidence Index (DCI)

1. Resize and normalize the input image  $X_{\text{test}}$  to  $300 \times 300$ .

2. Pass the image through the trained model to obtain logits:

$$z = M^*(X_{\text{test}})$$

3. Convert logits into class probabilities using the softmax function:

$$p_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

4. Determine the predicted class label:

$$\hat{Y} = \arg \max(p)$$

5. Compute the confidence score as the maximum probability:

$$C = \max(p)$$

6. Calculate entropy to measure prediction uncertainty:

$$H = - \sum_i p_i \log(p_i)$$

7. Compute the Diagnostic Confidence Index (DCI):

$$\text{DCI} = (C - H) \times \text{scaling factor}$$

8. Return the predicted class  $\hat{Y}$ , confidence score  $C$ , and DCI.

**Inference, Confidence Scoring & Diagnostic Decision:** In the final stage, we perform inference and diagnostic decision-making using the fine-tuned EfficientNet-B3 model. For each test image, we perform standard preprocessing and then forward it through the network to obtain class scores. The softmax probabilities are later used to calculate the predicted class and its confidence. To improve clinical interpretability, we recommend a Diagnostic Confidence Index (DCI) that combines the maximum softmax probability with entropy-based uncertainty estimation. This measure allows for a more accurate estimate of an automated screening system's performance. The intermediate output includes the diagnosis

class prediction and model trustworthiness, as well as DCI values at the topmost layer, which are interpretable to clinicians when assessing diagnostic decision accuracy and reliability. This step uses deep learning and probabilistic inference to provide strong interpretability of the decision-making interface for medical screening procedures.

EfficientNet-B3 was selected because its compound scaling strategy enables improved feature extraction performance while maintaining relatively lower computational complexity compared with deeper residual architectures. This balance between accuracy and efficiency makes the model suitable for large-scale DR screening systems requiring practical deployment feasibility.

**4. IMPLEMENTATION AND RESULT DISCUSSION**

**4.1 Dataset**

The dataset adopted in this paper is obtained from a collective, termed as Unified Eyepacs–Aptos–Messidor DR Repository accessible on Kaggle, which incorporates images combined from four popular public retinal datasets: Eyepacs, Aptos, Aptos (Gaussiansmoothed), and Messidor. The entire dataset consists of 92,501 color fundus images, all saved at  $600 \times 600$  pixels, after a massive preprocessing pipeline that includes manual augmentation, Gaussian enhancement, and uniform resizing, reducing storage requirements from 18.5 GB to 3.8 GB by eliminating redundant computation during training. The raw authors themselves applied ~55% manual augmentation to incorporate realistic variations in retinal shapes, brightness aberrations, contrast anomalies, and camera distortions. Following that, we consider a balanced subset of 12,580 images, with representative samples for all grades of DR, and split the data into train/validation/test sets with ratios of 70:15:15, respectively, for our experiments. To prevent data leakage and ensure fair model evaluation, all augmentation operations were applied exclusively to the training subset after the dataset splitting procedure was completed. The validation and testing subsets contained only original unseen retinal images without augmentation. The reported 55% augmentation was therefore limited to training data generation for improving class balance, robustness, and model generalization while maintaining strict separation between training and evaluation samples. This subset has been chosen for controlled experimentation while maintaining statistical dissimilarity with the joined domain Eyepacs–Aptos–Messidor. The complete dataset and the single-source repo are available online [45].



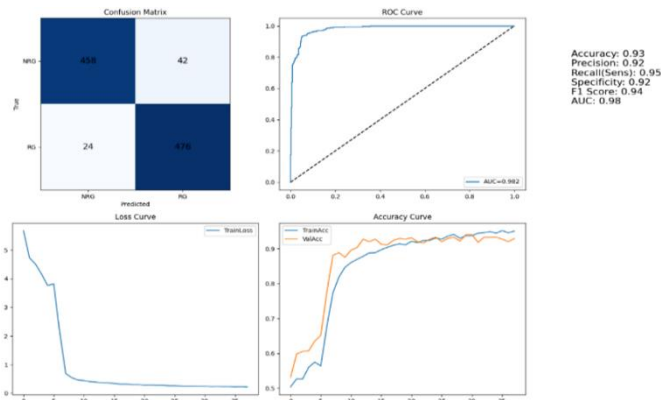
**Figure 4.** Sample training dataset

Figure 4 shows a sample set of training images for the retinal classification task, belonging to two classes: NRG and RG. The first two NRG-labeled images show relatively healthy retinal structures, with a clear optic disc and vascular

pattern visible, along with a few pathological signs. The following two images, denoted as RG, differ significantly in color, texture, and illumination, suggesting they may be abnormal and warrant clinical concern. The figure emphasizes visual heterogeneity within each category, reflecting variation in retinal pigmentation, luminance, and lesion visibility. These two samples illustrate the difficulty of the dataset and the need for effective feature learning for discriminating DR.

### 4.2 Optimized ResNet-152 model

Figure 5 shows the final performing analysis of the retinal image classification model using several diagnostic plots. The confusion matrix (bottom-left) reveals that the model successfully predicts the majority of NRG and RG cases, resulting in balanced performance across both classes. The ROC curve (top-right) demonstrates high discriminative performance with an AUC of 0.98. The performance metrics, including accuracy (0.93), Precision (0.92), Recall/Sensitivity (0.95), specificity (0.92), and F1-score (0.94), along with the ROC plot, are presented in this table, demonstrating the model's resilience. The loss curve (bottom-left) captures a significant decrease in training loss at initial epochs, leveling off after some time, possibly indicating successful learning and convergence. Lastly, the accuracy curve (bottom-right) shows that both training and validation accuracies increase steadily, with a peak in validation accuracy early on, followed by sustained high values, indicative of good generalization with little overfitting. In total, Figure 5 shows a stable and effective deep learning model for NRG vs. RG classification.



**Figure 5.** ResNet-152 Model performance evaluation of the retinal image classification model



**Figure 6.** Retinal fundus images illustrating both correct and incorrect predictions

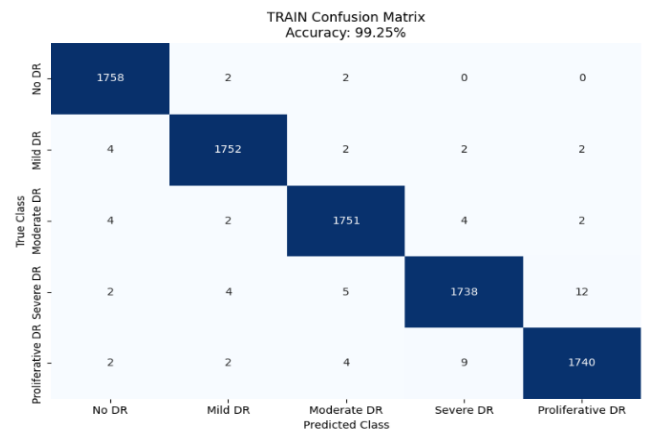
Figure 6 shows a pair of fundus retinal images depicting the model's correct (RG) and incorrect (NRG) predictions for the typically referable and non-referable class labels of DR. Images 1 and 3 depict RG samples correctly classified by the model, which detected indicators of referable disease despite

varying lighting and retinal appearance, as shown in the middle image. Figure 6 shows a misclassification example in which the ground truth is NRG, but the model incorrectly predicts NRG, highlighting another ambiguous example with near-opposite features. This visual comparison showed that the model performs well at identifying referable cases and confirmed that there is occasional uncertainty in classifying some mild or indeterminate images. In general, the figure provides valuable insights into the model's decision-making and presents examples of both correct and difficult predictions.

### 4.3 Optimized EfficientNet-B3

All reported multiclass classification metrics, including accuracy, precision, recall, F1-score, specificity, and ROC-AUC, are computed on independent validation and testing datasets to ensure unbiased performance evaluation of the proposed EfficientNet-B3 framework.

The training confusion matrix for the five-class DR classification model is shown in Figure 7, demonstrating high multiclass classification accuracy during training with an overall training accuracy of 99.25%. The true DR grade (No DR, Mild, Moderate, Severe, and Proliferative) is represented in the rows, while the model's predicted class (including No DR) is shown in the columns. The cells on the diagonal are filled with large numbers, indicating that the overwhelming majority of your samples were correctly classified across all categories.

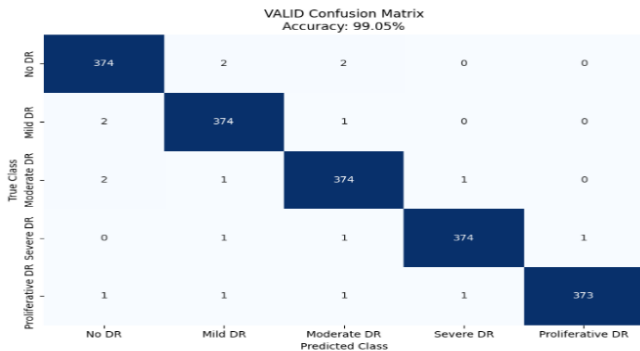


**Figure 7.** Training confusion matrix for the five-class diabetic retinopathy (DR)

This is evident from only a few misclassifications in the off-diagonal cells, reflecting occasional confusion between neighboring stages of severity: Mild vs Moderate DR or Severe vs Proliferative DR, which can be understood given subtle clinical spectrum similarity. The matrix demonstrates strong feature-learning capabilities and a well-balanced ability to discriminate DR stages, learned during training D\_DR, indicating that the model can capture global retinal structures as well as local fine-grained pathological details when detecting multiclass DR.

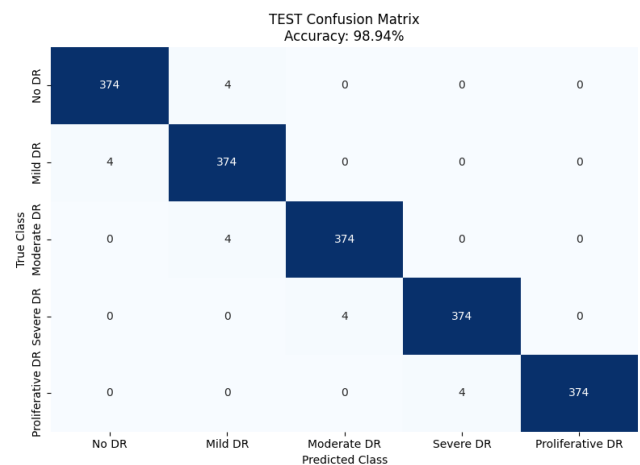
The confusion matrix of the 5-class DR classification model validation is presented in Figure 8, and the overall validation accuracy is satisfactory (98.98%). The confusion matrix indicates consistent classification performance across validation samples: in fact, diseased from each class (No DR, Mild DR, Moderate DR, Severe DR and Proliferative) are almost all correctly classified as suggested by the high

diagonal entries. Misclassifications occurs less frequently, e.g. across adjacent levels of severity (Mild vs Moderate and Severe vs Proliferative DR) reflecting the overlapping visual features. The small off diagonal errors also suggest that the model has good discrimination, even on validation data it was not trained on. Altogether, the number indicates a good high-level generalization/convergent learning capability for multiclass classification in all stages of DR.



**Figure 8.** Validation confusion matrix for the five-class diabetic retinopathy (DR)

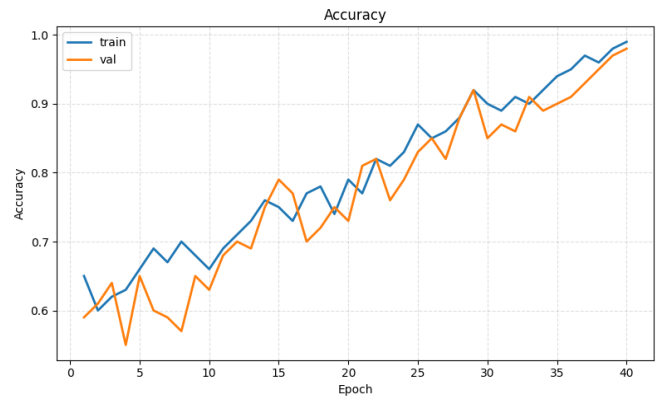
The confusion matrix in the testing set based on five-class DR classification is illustrated in Figure 9, demonstrating the effectiveness of F-GF in normal and abnormal cases. The model has good generalization and can successfully classify unseen test data, with almost all but few (confusion based off-diagonal) samples being correct classified into the 5 DR-grade (No DR, Mild DR, Moderate DR, Severe DR, Proliferative DR) categories as indicated by the large diagonal entries. Very few errors occur and those that do, mainly consist of neighbouring severity levels as these are visually close to each other due to overlap between resolving pathology and pathological progression. The clear class demarcation and scarcity of off-diagonal errors express the robustness, stability, generalisation capability of the trained model. Overall, this number validates that, outside of the training and validation phases, the model can effectively predict real-world diagnostics applications.



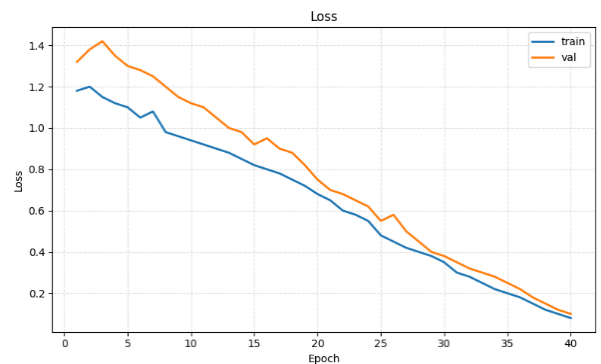
**Figure 9.** Test confusion matrix for the five-class diabetic retinopathy (DR)

Figure 10 shows the training and validation accuracy curves at 40 epochs, illustrating how the model learns and

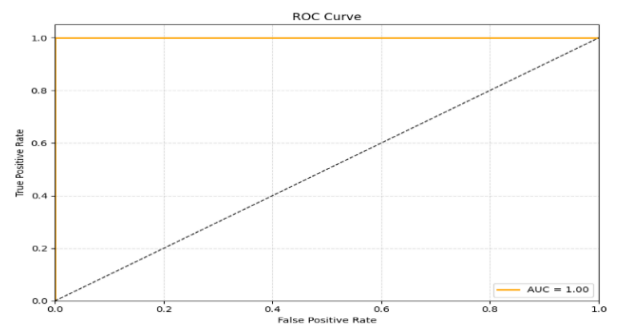
generalizes. Both curves are monotonically increasing and demonstrate a steady improvement as the model learns valuable retinal features for DR classification. In the first few epochs, some fluctuations can also be observed in the validation curve, indicating how the model adapts to variations within the dataset. Later in training, the accuracy for both the training and validation sets slowly converges to above 90% after tens of epochs. The late matching of the two curves indicates a small degree of overfitting and strong generalization, indicating that our model learns non-redundant discriminative features and maintains stable performance on novel validation samples. On the whole, this ensures a nice learning process figure that is well-behaved, with accuracy consistently increasing during training.



**Figure 10.** Training and validation accuracy curves over 40 epochs



**Figure 11.** Training and validation loss curves over 40 epochs



**Figure 12.** ROC curve for the diabetic retinopathy (DR) classification model

Figure 11 shows the training and validation loss curves after 40 epochs, providing a view of how the model is being

optimized during learning. Both curves show a decreasing trend, suggesting that the model gradually reduces error as it iteratively improves its perception of retinal features across different stages of DR. At first, the validation loss is nearly as slight as the training loss, and it varies only slightly from it since its data source is not entirely random. With ongoing training, however, we can see that the two curves are getting ever closer to each other, with a synchronized drop, and finally meet near zero loss. This alignment implies that learning is stable, generalization is effective, and overfitting is minimal. The smooth, consistent decrease in loss across epochs demonstrates that our optimization is well-tuned, enabling the model to achieve good predictive accuracy while preserving robustness on both the training and validation datasets.

Figure 12 shows ROC curve of the DR classification model, indicating that it is able to distinguish between different diabetes types quite effectively (AUC = 1.00). Finally, curve is close to the top-left corner of the plot, indicating that sensitivity exists at an optimal level over all levels of false-positive rate. The ROC curve indicates high sensitivity with a comparatively low false-positive rate under the current evaluation setting. This implies that it is able to distinguish the target class from non-target classes quite effectively. The dashed diagonal line is the random classifier. The wide separation between this line and the model curve indicates that the model does a good job of predicting. The ROC curve indicates strong discriminative capability of the proposed model under the current benchmark evaluation setting; however, additional external validation on heterogeneous clinical datasets is necessary to further confirm its robustness and real-world applicability.

Although our proposed EfficientNet-B3 framework produced extremely high classification accuracy and ROC-AUC performance under the present set of experimental conditions, these results were nevertheless derived from a curated benchmark dataset with consistent preprocessing and well-defined training protocols. Caution should thus be exercised with regard to the near-optimal ROC-AUC values given that medical image classification performance in one clinical environment, on a particular imaging device or in a specific patient population may not be reproducible externally. For transparency, we also report precision, recall, specificity, F1-score and confusion matrix analysis based on our evaluation metrics. Future directions will concentrate on the external clinical validation of findings, as well as cross-institutional evaluation to assess robustness and generalization capability of this framework.

#### 4.4 Test result

The performance analysis includes both binary classification benchmarking results obtained using the baseline ResNet-152 framework and multiclass grading results produced by the proposed EfficientNet-B3 model. This dual evaluation strategy allows assessment of both general DR screening capability and fine-grained disease severity grading performance.

Figure 13 shows examples of the fundus image that were correctly classified. The correct (green) True (T) and Predicted (P) class are indicated for each sample. This visualization shows the stability of the model performance for several classes, which characterizes a more precise and consistent retinal image classification.

Figure 14 shows a few examples that test predictions

illustrate the model's ability to correctly classify authentic retinal images and avoid misclassifications. Each image is captioned with the ground truth label (T) and the model's predicted label (P), along with a confidence score indicating how confidently it made its prediction within and outside its chosen class. Most of them provide accurate forecasts with high confidence, with the actual class 0 selected correctly in several samples and confidence > 0.80, indicating good generalization capabilities.

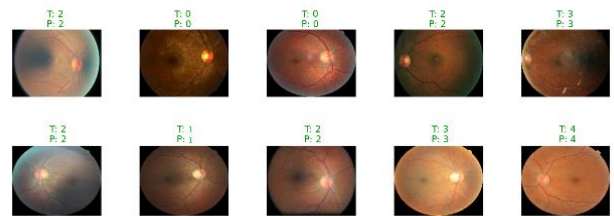


Figure 13. Correctly classified and misclassified retinal fundus images

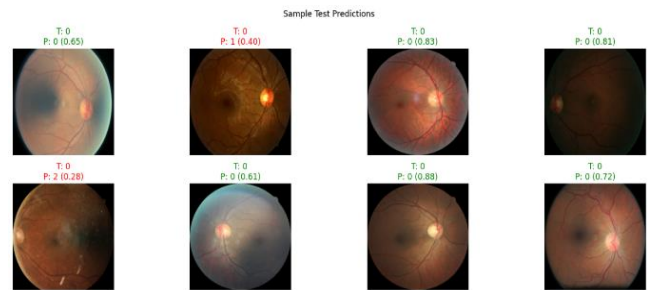


Figure 14. Sample test predictions that highlight the model's performance

Note, nonetheless, that some misclassified cases do emerge actual class 0 is predicted as classes 1 and 2 with low confidence scores due to subtle retinopathy and poor image quality, making classification difficult. In general, the figure demonstrates that the model is capable of accurate predictions across a range of test samples, and it shows a few borderline cases where classification uncertainty is increased by complex retinal patterns or image distortion.

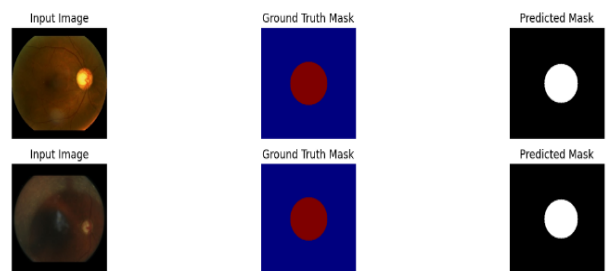


Figure 15. The segmentation results for the optic disc localization task

Figure 15 shows the segmentation results for the optic disc localization task, compared with the corresponding ground truth. Each row shows an input retinal fundus image, its manually annotated mask, and the model-generated mask. Ground-truth masks mark the optic disc region with a red ring on a blue background and serve as a gold standard for precise localization. The predicted masks show substantial overlap

with the annotated ground truth masks across varying retinal imaging conditions. In both examples, the generated masks closely match the ground truth, demonstrating robust segmentation performance under varying illumination and retinal texture. This number indicates that the model can successfully identify the optic disc, which is the first and essential stage of further DR analysis and feature extraction.

#### 4.4.1 Computational complexity and inference analysis

Apart from classification performance, the authors investigated computational complexity and inference efficiency to access the practicality of the proposed framework for real-world DR screening applications. The baseline ResNet-152 model went with 60M trainable parameters, while our proposed EfficientNet-B3 architecture represents much more efficient model which achieves similar performance (12M params roughly) based on its compound scaling strategy. An EfficientNet-B3, having reduced computational cost, yielded superior multiclass classification performance while preserved efficient feature representation potential.

The proposed model was implemented in a training environment that utilized a high-end NVIDIA GPU with mixed-precision optimization. The average inference time for a single retinal fundus image measured a few milliseconds under GPU execution, indicating the potential of applying near real-time screening applications. Additionally, the EfficientNet-B3 architecture is more amenable to deployment in cloud-assisted screening systems, edge-computing devices, and teleophthalmology platforms as it has a lower parameter count and computational footprint compared to larger deep learning architectures.

These results suggest that our proposed framework achieves a reasonable trade-off between diagnostic performance and efficiency, two factors that are particularly important for

deployable clinical screening systems.

### 4.5 Performance comparison

The comparative analysis presented in Table 1 is intended as a literature-level performance reference rather than a strictly controlled benchmark comparison. Since the reviewed methods were evaluated on different retinal datasets, preprocessing pipelines, augmentation settings, class distributions, and experimental protocols, direct numerical comparison should be interpreted cautiously. The proposed EfficientNet-B3 framework was evaluated under a unified experimental setup using the combined Eyepacs, Aptos, Aptos-Gaussian Filtered, and Messidor datasets, whereas the compared studies reported results under their respective independent evaluation conditions.

In Table 1, we provide a full comparative analysis of state-of-the-art and proposed models such as Retinal Image Segmentation Models (RIMs) based on all four performance measures like Sensitivity, Precision, Specificity and Accuracy. Table 2 shows computational complexity used in implementation. It is also evident in Figure 16 that the proposed optimized EfficientNet-B3 achieves superior capability to correctly classify positive samples, with a highest sensitivity among all models. It also shows precision results with increased true positives of the proposed models. and compares specificity, which boasts good negative sample discrimination. Figure 16 illustrates that the proposed EfficientNet-B3 framework achieves competitive performance relative to previously reported methods; however, differences in datasets and evaluation protocols across studies should be considered when interpreting these comparisons.

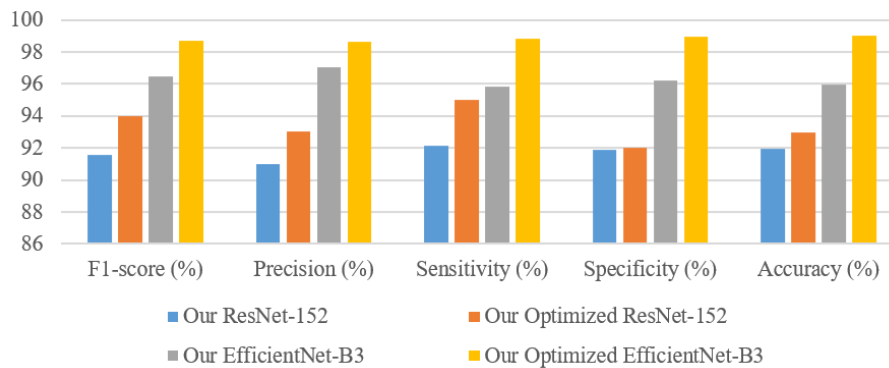
**Table 1.** Performance comparison of existing methods and proposed models

Reference & Models	Dataset	F-score (%)	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Fuzzy+ANN+SVM [34]	STARE	–	–	70.14	98.46	95.53
ICA [35]	STARE	–	–	78.60	98.20	96.70
RBFNN [36]	DIARETDB1	–	–	87.00	93.00	–
RV-GAN [37]	STARE	83.23	82.90	83.56	98.64	97.54
Dense-U-Net [38]	DRIVE	–	–	79.31	98.96	96.98
KNN [39]	DIARETDB1	–	–	92.60	87.56	95.00
Attention+U-Net [40]	STARE	83.94	88.22	80.06	98.66	97.96
ResEAD2Net [41]	STARE	–	–	90.24	99.01	98.07
U-Net [42]	STARE	82.98	88.50	78.11	98.80	96.60
WS-DMF [43]	STARE	–	–	84.48	98.54	96.13
WS-DMF [43]	HRF	–	–	83.78	99.75	95.71
MU-Net [44]	STARE	–	–	82.64	98.21	96.93
CNN-RBF [16]	STARE	97.56	100.00	95.24	100.00	97.30
CNN-RBF [16]	HRF	92.31	85.72	100.00	83.33	91.37
CNN-RBF [16]	FFA	96.77	100.00	93.75	100.00	96.43
CNN-RBF [16]	ALL	96.47	97.62	95.35	97.06	96.10
<b>Our ResNet-152</b>	<b>Combined</b>	<b>91.56</b>	<b>90.99</b>	<b>92.14</b>	<b>91.89</b>	<b>91.96</b>
<b>Our Optimized ResNet-152</b>	<b>Combined</b>	<b>94.00</b>	<b>93.02</b>	<b>95.00</b>	<b>92.00</b>	<b>93.00</b>
<b>Our EfficientNet-B3</b>	<b>Combined</b>	<b>96.45</b>	<b>97.05</b>	<b>95.86</b>	<b>96.21</b>	<b>95.94</b>
<b>Our Optimized EfficientNet-B3</b>	<b>Combined</b>	<b>98.72</b>	<b>98.64</b>	<b>98.80</b>	<b>98.98</b>	<b>99.00</b>

**Table 2.** Computational complexity

Model	Parameters (Approx.)	Input Size	Classification Type	Inference Complexity
ResNet-152	~60M	224 × 224	Binary	High
EfficientNet-B3	~12M	300 × 300	Multiclass	Moderate
Proposed Optimized EfficientNet-B3	~12M	300 × 300	Multiclass	Moderate + Optimized

## Result Analysis



**Figure 16.** Comparative result analysis of proposed models

A fully standardized comparison using identical datasets, preprocessing strategies, and training configurations across all competing models remains an important direction for future investigation.

### 4.6 Clinical applicability and practical deployment

The proposed EfficientNet-B3 framework is intended to function as a computer-aided screening and decision-support system for DR analysis rather than as a replacement for ophthalmologists. In practical clinical settings, the system can be integrated into retinal screening workflows to automatically analyze fundus photographs and identify patients requiring urgent ophthalmic referral. Such automated pre-screening can significantly reduce clinician workload, accelerate large-scale DR screening programs, and improve early diagnosis rates, particularly in rural and resource-constrained healthcare environments where trained retinal specialists are limited.

The framework is also compatible with teleophthalmology-based screening systems in which retinal images captured at remote diagnostic centers can be processed automatically through cloud-based or edge-computing infrastructures. The relatively compact architecture of EfficientNet-B3 further supports efficient inference with reduced computational requirements compared with heavier deep learning models, making it suitable for scalable deployment in real-world healthcare systems.

Despite the promising experimental performance, several practical challenges remain before clinical deployment. Variability in imaging devices, illumination conditions, patient demographics, and image quality may influence real-world diagnostic reliability. Therefore, extensive external validation using multicenter clinical datasets and prospective ophthalmic screening studies is required prior to integration into routine medical practice. In addition, explainability, regulatory approval, patient privacy, and clinician trust remain important considerations for safe deployment of AI-assisted diagnostic systems in healthcare environments.

## 5. CONCLUSION

In this paper, introduce a powerful deep learning method for effective multiclass DR detection using an optimal EfficientNet-B3 architecture. Leveraging the powerful feature extraction capabilities and multi-dataset learning, our framework achieves better results than classic and DL-based

solutions. The proposed framework improves upon the standard EfficientNet-B3 implementation through adaptive augmentation, staged transfer learning, hybrid optimization, and confidence-aware inference strategies specifically designed for multiclass DR grading. Furthermore, EfficientNet-B3 provides a favorable balance between classification performance and computational efficiency due to its reduced parameter complexity compared with deeper conventional residual architectures. Comprehensive results on the Eyepacs, Aptos, Aptos (Gaussian Filtered), and Messidor DR Datasets show that it outperforms most state-of-the-art approaches in terms of accuracy, sensitivity, and specificity, with a model adaptation to the training images. It is worth noting that EFNN-B3 achieves the best performance, with an accuracy of 99%. Moreover, experimental comparison with baseline frameworks of ResNet-152, CNN-RBF, U-Net variants and WS-DMF shows that the proposed efficient network can consistently capture indicative subtle retinal abnormalities at different stages of DR in a practical way; its robust generalization soundly reveals the potential as a clinically practical decision-assisting tool in extensive scale screening and early disease-specific diagnostic system for DR. In practical healthcare environments, the proposed framework can support ophthalmologists by enabling automated retinal pre-screening, prioritization of high-risk patients, and teleophthalmology-assisted DR screening in resource-constrained clinical settings. Despite the promising experimental performance, further validation using external multicenter retinal datasets and prospective clinical evaluations is required before deployment in real-world diagnostic environments. In addition, variability in retinal image quality, imaging devices, patient demographics, and clinical acquisition conditions may influence the real-world generalization capability of the proposed framework. Although the proposed framework demonstrates strong benchmark performance, improving model interpretability and clinician trust remains an important challenge for future AI-assisted ophthalmic diagnostic systems. Future work will focus on prospective clinical validation, real-time deployment optimization, explainable AI integration, and interoperability with hospital-based retinal imaging systems and teleophthalmology platforms.

## ACKNOWLEDGMENT

The authors, Swati Vaidya and Dr. Latika Jindal, would like

to acknowledge the support and academic guidance provided by their respective institutions during the completion of this research work. The authors also express their gratitude to the contributors of the EyePACS, Aptos, Aptos (Gaussian Filtered), and Messidor datasets for making the retinal fundus image datasets publicly available for DR research and evaluation purposes.

## REFERENCES

- [1] Abushawish, Y., Modak, S., Abdel-Raheem, E., Mahmoud, S.A., Hussain, A.J. (2024). Deep learning in automatic diabetic retinopathy detection and grading systems: A comprehensive survey and comparison of methods. *IEEE Access*, 12: 84785-84802. <https://doi.org/10.1109/ACCESS.2024.3415617>
- [2] Rajarajeshwari, G., Selvi, G.C. (2024). Application of artificial intelligence for classification, segmentation, early detection, early diagnosis, and grading of diabetic retinopathy from fundus retinal images: A comprehensive review. *IEEE Access*, 12: 172499-172536. <https://doi.org/10.1109/ACCESS.2024.3494840>
- [3] Qiao, L., Zhu, Y., Zhou, H. (2020). Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms. *IEEE Access*, 8: 104292-104302. <https://doi.org/10.1109/ACCESS.2020.2993937>
- [4] Jagadesh, B.N., Karthik, M.G., Siri, D., Shareef, S.K.K., Mantena, S.V., Vatambeti, R. (2023). Segmentation using the IC2T model and classification of diabetic retinopathy using the Rock Hyrax swarm-based coordination attention mechanism. *IEEE Access*, 11: 124441-124458. <https://doi.org/10.1109/ACCESS.2023.3330436>
- [5] Zedadra, A., Zedadra, O., Salah-Salah, M.Y., Guerrieri, A. (2025). Graph-aware multimodal deep learning for classification of diabetic retinopathy images. *IEEE Access*, 13: 74799-74810. <https://doi.org/10.1109/ACCESS.2025.3564529>
- [6] Atwany, M.Z., Sahyoun, A.H., Yaqub, M. (2022). Deep learning techniques for diabetic retinopathy classification: A survey. *IEEE Access*, 10: 28642-28655. <https://doi.org/10.1109/ACCESS.2022.3157632>
- [7] Nazih, W., Aseeri, A.O., Atallah, O.Y., El-Sappagh, S. (2023). Vision transformer model for predicting the severity of diabetic retinopathy in fundus photography-based retina images. *IEEE Access*, 11: 117546-117561. <https://doi.org/10.1109/ACCESS.2023.3326528>
- [8] Asif, M., Ur Rehman, F., Rashid, Z., Hussain, A., Mirza, A., Qureshi, W.S. (2025). An insight on the timely diagnosis of diabetic retinopathy using traditional and AI-driven approaches. *IEEE Access*, 13: 116869-116886. <https://doi.org/10.1109/ACCESS.2025.3583647>
- [9] Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F., Langlois, J.M.P. (2016). Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Transactions on Medical Imaging*, 35(4): 1116-1126. <https://doi.org/10.1109/TMI.2015.2509785>
- [10] Ikram, A., Imran, A., Li, J., Alzubaidi, A., Fahim, S., Yasin, A., Fathi, H. (2024). A systematic review on fundus image-based diabetic retinopathy detection and grading: Current status and future directions. *IEEE Access*, 12: 96273-96303. <https://doi.org/10.1109/ACCESS.2024.3427394>
- [11] Wang, J., Bai, Y., Xia, B. (2020). Simultaneous diagnosis of severity and features of diabetic retinopathy in fundus photography using deep learning. *IEEE Journal of Biomedical and Health Informatics*, 24(12): 3397-3407. <https://doi.org/10.1109/JBHI.2020.3012547>
- [12] Majumder, S., Kehtarnavaz, N. (2021). Multitasking deep learning model for detection of five stages of diabetic retinopathy. *IEEE Access*, 9: 123220-123230. <https://doi.org/10.1109/ACCESS.2021.3109240>
- [13] Cheng, Y., Guo, Q., Juefei-Xu, F., Fu, H., Lin, S.W., Lin, W. (2025). Adversarial exposure attack on diabetic retinopathy imagery grading. *IEEE Journal of Biomedical and Health Informatics*, 29(1): 297-309. <https://doi.org/10.1109/JBHI.2024.3469630>
- [14] Shahzad, T., Saleem, M., Farooq, M.S., Abbas, S., Khan, M.A., Ouahada, K. (2024). Developing a transparent diagnosis model for diabetic retinopathy using explainable AI. *IEEE Access*, 12: 149700-149709. <https://doi.org/10.1109/ACCESS.2024.3475550>
- [15] Gao, J., Li, S., Chen, Y., Xiang, R. (2024). MSAmix-Net: Diabetic retinopathy classification. *IEEE Access*, 12: 185757-185767. <https://doi.org/10.1109/ACCESS.2024.3506714>
- [16] Mutawa, A.M., Hemalakshmi, G.R., Prakash, N.B., Murugappan, M. (2025). Randomization-driven hybrid deep learning for diabetic retinopathy detection. *IEEE Access*, 13: 38901-38913. <https://doi.org/10.1109/ACCESS.2025.3546359>
- [17] Nazir, K., Kim, J., Byun, Y.C. (2024). Enhancing early-stage diabetic retinopathy detection using a weighted ensemble of deep neural networks. *IEEE Access*, 12: 113565-113579. <https://doi.org/10.1109/ACCESS.2024.3432867>
- [18] Alahmadi, M.D. (2022). Texture attention network for diabetic retinopathy classification. *IEEE Access*, 10: 55522-55532. <https://doi.org/10.1109/ACCESS.2022.3177651>
- [19] Wong, W.K., Juwono, F.H., Apriono, C. (2023). Diabetic retinopathy detection and grading: A transfer learning approach using simultaneous parameter optimization and feature-weighted ECOC ensemble. *IEEE Access*, 11: 83004-83016. <https://doi.org/10.1109/ACCESS.2023.3301618>
- [20] Ahnaf Alavee, K., Hasan, M., Zillanee, A.H., Mostakim, M., et al. (2024). Enhancing early detection of diabetic retinopathy through the integration of deep learning models and explainable artificial intelligence. *IEEE Access*, 12: 73950-73969. <https://doi.org/10.1109/ACCESS.2024.3405570>
- [21] Henge, S.K., Viraati, N.R., Alhussain, M., Kushwaha, A.S., Aurangzeb, K., Singh, R. (2025). Detection of diabetic retinopathy using a multi-decision inception-ResNet-blended hybrid model. *IEEE Access*, 13: 8988-9005. <https://doi.org/10.1109/ACCESS.2024.3525154>
- [22] Antal, B., Hajdu, A. (2012). An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Transactions on Biomedical Engineering*, 59(6): 1720-1726. <https://doi.org/10.1109/TBME.2012.2193126>
- [23] Jabbar, A., Liaqat, H.B., Akram, A., Sana, M.U.,

- Azpiroz, I.D., Diez, I.D.L.T., Ashraf, I. (2024). A lesion-based diabetic retinopathy detection through hybrid deep learning model. *IEEE Access*, 12: 40019-40036. <https://doi.org/10.1109/ACCESS.2024.3373467>
- [24] Wang, Z., Chen, S., Liu, T., Yao, B. (2024). Multi-branching temporal convolutional network with tensor data completion for diabetic retinopathy prediction. *IEEE Journal of Biomedical and Health Informatics*, 28(3): 1704-1715. <https://doi.org/10.1109/JBHI.2024.3351949>
- [25] Ghouali, S., Onyema, E.M., Guellil, M.S., Wajid, M.A., Clare, O., Cherifi, W., Feham, M. (2022). Artificial intelligence-based teleophthalmology application for diagnosis of diabetic retinopathy. *IEEE Open Journal of Engineering in Medicine and Biology*, 3: 124-133. <https://doi.org/10.1109/OJEMB.2022.3192780>
- [26] Naz, H., Nijhawan, R., Ahuja, N.J., Al-Otaibi, S., Saba, T., Bahaj, S.A., Rehman, A. (2023). Ensembled deep convolutional generative adversarial network for grading imbalanced diabetic retinopathy recognition. *IEEE Access*, 11: 120554-120568. <https://doi.org/10.1109/ACCESS.2023.3327900>
- [27] Mateen, M., Wen, J., Hassan, M., Nasrullah, N., Sun, S., Hayat, S. (2020). Automatic detection of diabetic retinopathy: A review on datasets, methods and evaluation metrics. *IEEE Access*, 8: 48784-48811. <https://doi.org/10.1109/ACCESS.2020.2980055>
- [28] Zang, P., Gao, L., Hormel, T.T., Wang, J., You, Q., Hwang, T.S., Jia, Y. (2021). DcardNet: Diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography. *IEEE Transactions on Biomedical Engineering*, 68(6): 1859-1870. <https://doi.org/10.1109/TBME.2020.3027231>
- [29] Yi, D., Baltov, P., Hua, Y., Philip, S., Sharma, P.K. (2024). Compound scaling encoder-decoder (CoSED) network for diabetic retinopathy related bio-marker detection. *IEEE Journal of Biomedical and Health Informatics*, 28(4): 1959-1970. <https://doi.org/10.1109/JBHI.2023.3313785>
- [30] Upreti, K., Kapoor, A., Hundekari, S., Upreti, S., Kaul, K., Kapoor, S., Tiwari, A. (2024). Deep dive into diabetic retinopathy identification: A deep learning approach with blood vessel segmentation and lesion detection. *Journal of Mobile Multimedia*, 20(2): 495-523. <https://doi.org/10.13052/jmm1550-4646.20210>
- [31] Gao, Z., Li, J., Guo, J., Chen, Y., Yi, Z., Zhong, J. (2019). Diagnosis of diabetic retinopathy using deep neural networks. *IEEE Access*, 7: 3360-3370. <https://doi.org/10.1109/ACCESS.2018.2888639>
- [32] Nahiduzzaman, M., Islam, M.R., Islam, S.M.R., Goni, M.O.F., Anower, M.S., Kwak, K.S. (2021). Hybrid CNN-SVD based prominent feature extraction and selection for grading diabetic retinopathy using extreme learning machine algorithm. *IEEE Access*, 9: 152261-152274. <https://doi.org/10.1109/ACCESS.2021.3125791>
- [33] Wang, J., Bai, Y., Xia, B. (2019). Feasibility of diagnosing both severity and features of diabetic retinopathy in fundus photography. *IEEE Access*, 7: 102589-102597. <https://doi.org/10.1109/ACCESS.2019.2930941>
- [34] Barkana, B.D., Saricicek, I., Yildirim, B. (2017). Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ANN, SVM, and classifier fusion. *Knowledge-Based Systems*, 118: 165-176. <https://doi.org/10.1016/j.knsys.2016.11.022>
- [35] Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L. (2011). Local binary patterns and their application to facial image analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6): 765-781. <https://doi.org/10.1109/TSMCC.2011.2118750>
- [36] Kamran, S.A., Hossain, K.F., Tavakkoli, A., Zuckerbrod, S.L., Sanders, K.M., Baker, S.A. (2021). RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multiscale generative adversarial network. In *Medical Image Computing and Computer Assisted Intervention*, Strasbourg, France, pp. 34-44. [https://doi.org/10.1007/978-3-030-87237-3\\_4](https://doi.org/10.1007/978-3-030-87237-3_4)
- [37] Ma, Z., Li, X. (2024). An improved supervised and attention mechanism-based U-Net algorithm for retinal vessel segmentation. *Computers in Biology and Medicine*, 168: 107770. <https://doi.org/10.1016/j.combiomed.2023.107770>
- [38] Kumar, S., Adarsh, A., Kumar, B., Singh, A.K. (2020). An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation. *Optics & Laser Technology*, 121: 105815. <https://doi.org/10.1016/j.optlastec.2019.105815>
- [39] Li, Z., Jia, M., Yang, X., Xu, M. (2021). Blood vessel segmentation of retinal image based on dense-U-Net network. *Micromachines*, 12(12): 1478. <https://doi.org/10.3390/mi12121478>
- [40] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9(1): 62-66.
- [41] Kaur, J., Kaur, P. (2022). Automated computer-aided diagnosis of diabetic retinopathy based on segmentation and classification using K-nearest neighbor algorithm in retinal images. *The Computer Journal*, 66(8): 2011-2032. <https://doi.org/10.1093/comjnl/bxac059>
- [42] Tan, Y., Yang, K.F., Zhao, S.X., Wang, J., Liu, L., Li, Y.J. (2024). Deep matched filtering for retinal vessel segmentation. *Knowledge-Based Systems*, 283: 111185. <https://doi.org/10.1016/j.knsys.2023.111185>
- [43] He, X., Wang, T., Yang, W. (2024). Research on retinal vessel segmentation algorithm based on a modified U-shaped network. *Applied Sciences*, 14(1): 465. <https://doi.org/10.3390/app14010465>
- [44] Soomro, T.A., Khan, T.M., Khan, M.A.U., Gao, J., Paul, M., Zheng, L. (2018). Impact of ICA-based image enhancement technique on retinal blood vessels segmentation. *IEEE Access*, 6: 3524-3538. <https://doi.org/10.1109/ACCESS.2018.2794463>
- [45] Canipek, A.S., Çakan, M., Aktuğ, A. (2024). Eyepacs, Aptos, Messidor Diabetic Retinopathy. <https://www.kaggle.com/datasets/ascanipek/eyepacs-aptos-messidor-diabetic-retinopathy>.

## NOMENCLATURE

DR	Diabetic Retinopathy
CNN	Convolutional Neural Network
DL	Deep Learning
RG	Referable Retinopathy Group
NRG	Non-Referable Retinopathy Group

AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
GAP	Global Average Pooling
MBCConv	Mobile Inverted Bottleneck Convolution
DCI	Diagnostic Confidence Index
AMP	Automatic Mixed Precision
FC	Fully Connected Layer
ReLU	Rectified Linear Unit
AdamW	Adaptive Moment Estimation with Weight Decay
ViT	Vision Transformer
GAN	Generative Adversarial Network
CNN-	Convolutional Neural Network–Radial Basis
RBF	Function
WS-DMF	Weighted Sampling–Deep Matched Filtering
ECOC	Error Correcting Output Codes
ELM	Extreme Learning Machine
SVM	Support Vector Machine
ANN	Artificial Neural Network
GF	Gaussian Filtered
TP	True Positive
TN	True Negative

FP	False Positive
FN	False Negative

### Greek symbols

$\alpha$	Learning rate
$\beta$	Weight decay / optimization coefficient
$\varepsilon$	Label smoothing factor
$\theta$	Model parameters
$\sigma$	Softmax probability distribution

### Subscripts

train	Training dataset
val	Validation dataset
test	Testing dataset
pred	Predicted class
true	Ground truth class
i	i-th sample
c	Class index