electronics

Article

Enhancing Diabetic Retinopathy Detection Using Pixel Color Amplification and EfficientNetV2: A Novel Approach for Early Disease Identification

Yi-Hsuan Kao and Chun-Ling Lin *

Department of Electrical Engineering, Ming Chi University of Technology, New Taipei City 24301, Taiwan; m10128012@o365.mcut.edu.tw

* Correspondence: ginnylin@mail.mcut.edu.tw; Tel.: +886-2-2908-9899 (ext. 4819)

Abstract: Diabetic retinopathy (DR) is a severe complication of diabetes, causing damage to retinal blood vessels due to high blood sugar levels. Early detection is crucial but often requires significant time and expertise from ophthalmologists. While artificial intelligence (AI) and image recognition hold promise for DR detection, inconsistent image quality poses a challenge. Our study presents a novel technique that integrates pixel color amplification and EfficientNetV2 to enhance fundus image attributes, aiming to address issues related to image quality and achieving superior performance in DR detection. Leveraging EfficientNetV2, an advanced convolutional neural network (CNN) architecture, we achieve 84% multiclass accuracy and 99% binary accuracy, surpassing various other CNN models, including VGG16-fc1, VGG16-fc2, NASNet, Xception, Inception ResNetV2, EfficientNet, InceptionV3, MobileNet, and ResNet50. Our research tackles the critical challenge of early detection of DR, essential for preventing vision loss. This advancement holds the potential to enhance the efficiency and accuracy of DR classification, potentially alleviating the burden on medical professionals and ultimately improving the quality of life for individuals at risk of vision loss.

Keywords: DR; artificial intelligence; deep learning; EfficientNetV2; pixel color amplification; image preprocessing; convolutional neural network; medical imaging; fundus images; disease detection

Citation: Kao, Y.-H.; Lin, C.-L. Enhancing Diabetic Retinopathy Detection Using Pixel Color Amplification and EfficientNetV2: A Novel Approach for Early Disease Identification. *Electronics* **2024**, *13*, 2070. https://doi.org/10.3390/electronics13112070

Academic Editor: Hyunjin Park

Received: 29 April 2024
Revised: 21 May 2024
Accepted: 22 May 2024
Published: 27 May 2024

1. Introduction

1.1. Background and Problem Statement

Diabetic retinopathy (DR) is an ocular condition caused by diabetes, primarily affecting the retina. It stands as a leading cause of vision impairment and blindness in developed countries. Shockingly, statistics reveal that, as of 2020, approximately 103.12 million individuals worldwide were impacted by DR, with projections reaching 160.5 million by 2045 [1]. These figures underscore the substantial threat DR poses to global vision health. Early detection and intervention are crucial for effectively managing DR and preventing vision loss. However, in routine examinations of diabetic patients, clinicians may struggle to identify such cases, potentially due to the inconspicuous early symptoms of DR or the challenges of thorough examination within busy clinical environments. Due to the subtle and inconspicuous nature of early symptoms of DR, along with the requirement for specialized expertise and significant time investment from ophthalmologists for fundus image evaluation, early detection and diagnosis pose a challenge.

1.2. Advances in AI for DR Detection

To address this challenge, numerous studies are exploring the application of Artificial Intelligence (AI) for automated DR detection [2,3]. AI technologies, leveraging image recognition and deep learning techniques, can aid in identifying early signs of DR, thereby alleviating the workload of ophthalmologists. Among these technologies, convolutional

neural networks (CNNs) are extensively employed in DR detection studies due to their remarkable performance in image processing and classification tasks [4–6].

1.3. Related Workd

J. D. Bodapati et al. aimed to enhance DR recognition models through the optimization of retinal image representations [4]. They integrated features extracted from pretrained ConvNet models using a multimodal fusion module, and then trained a deep neural network (DNN) for both DR identification and severity prediction. Their method notably improved the quality of representations through 1D and cross pooling fusion, surpassing the performance of uni-modal ConvNet features. On the Kaggle APTOS 2019 dataset, their model achieved impressive results, achieving 97.41% accuracy and a kappa statistic of 94.82 for DR identification, as well as 81.7% accuracy and a kappa statistic of 71.1% for severity prediction. Interestingly, their DNN model, incorporating dropout at the input layer, exhibited quicker convergence when trained with blended features. Their study focused on obtaining comprehensive retinal image representations and significantly enhanced the performance of DR prediction. The success of their model was confirmed through experiments, highlighting average pooling as the most effective method in terms of performance and convergence speed.

S. H. Kassani et al. introduced an innovative feature extraction method using a modified Xception architecture for diagnosing DR [5]. Their approach involved deep layer aggregation, which combined multilevel features from various convolutional layers within the Xception architecture. The extracted features were then fed into a multilayer perceptron (MLP) for training in DR severity classification. The authors evaluated the method's performance by comparing it with four other deep feature extractors: InceptionV3 [7], MobileNet [8], ResNet50 [9], and the original Xception architecture [10]. The proposed deep CNN layer aggregation effectively merged features, enhancing the learning process compared to the conventional Xception architecture. Their strategy also incorporated transfer learning and hyper-parameter tuning to further improve classification accuracy. When validated on the Kaggle APTOS 2019 dataset, the modified Xception deep feature extractor demonstrated an enhanced DR classification accuracy of 83.09% (compared to 79.59% with the original Xception), sensitivity of 88.24% (versus 82.35%), and specificity of 87.00% (versus 86.32%). This highlighted the effectiveness of the modified Xception architecture in enhancing the diagnosis of DR.

1.4. Our Previous Work

The paper published in 2023 represents our team's research on DR classification [6]. It emphasized the impact of varying image quality due to different conditions on the efficacy of training models to classify DR stages, spanning from 0-No DR to 4-Proliferative DR. A proposed preprocessing method enhanced image features, effectively expanding the training dataset. Additionally, enhancements to the EfficientNet model were implemented to enhance performance, resulting in an accuracy increase from 77.27% to 79.20% for the test dataset. Notably, the refined EfficientNet achieved a superior average AUC (0.926) across five classes compared to MobileNet (0.54) and the original EfficientNet (0.922). Ultimately, the creation of an API-based system empowers users to autonomously upload fundus images and acquire DR assessment results

1.5. Current Study Contributions

In our paper presented at IEEE ICASI 2023 [11], we acknowledged the challenge posed by variations in imaging specifications on the effectiveness of DR detection. To tackle this issue, we introduced an approach utilizing the EfficientNetV2 model with pixel enhancement preprocessing to aid in DR detection. However, in our current journal research, we conducted more comprehensive comparisons of various modules and methodologies. We applied pixel color enhancement techniques to enhance image features, thereby mitigating discrepancies arising from diverse imaging devices and environments. This process im-

proves image quality and accentuates subtle pathological features, facilitating DR detection. Additionally, we optimized and fine-tuned the EfficientNetV2 model to ensure precise differentiation among the five distinct stages of DR. The proposed approach integrates pixel color enhancement techniques and model adjustments, offering a dependable and accurate method for DR detection. This has the potential to alleviate the workload of medical professionals while enhancing detection efficiency and accuracy. Our study introduces several key innovations that advance the field of DR detection:

1. Comprehensive Preprocessing Workflow: Our workflow includes precise cropping of images using the Gradient Hough Transform, pixel color amplification, and data augmentation techniques to improve the robustness and generalizability of the model;
2. Pixel Color Amplification: We introduce a novel application of the Dark Channel Prior (DCP) method [12,13] to enhance the visibility of retinal features in fundus images, addressing issues of underexposure and overexposure commonly encountered in retinal photography;
3. EfficientNetV2 Optimization: We optimized the EfficientNetV2 model for DR detection using transfer learning and fine-tuning, achieving higher accuracy and faster convergence compared to traditional models. This optimization includes early stopping, adaptive learning rates, and label smoothing to prevent overfitting, dynamically adjust learning rates, and enhance model generalization and stability;
4. Integrated Predictive System: We develop a web-based system that allows for real-time DR detection, providing a practical tool for ophthalmologists to use in clinical settings.

1.6. Structure of the Paper

The remainder of this paper is structured as follows:

1. Section 2 discusses the methodology, including data sources, preprocessing techniques, and model architecture;
2. Section 3 presents the experimental results and analysis;
3. Section 4 details the implementation of the predictive system;
4. Section 5 provides a discussion on the findings and future work;
5. Section 6 concludes the paper with a summary of the contributions and potential applications.

2. Methods

This section outlines the methodologies and workflow employed in the entire study, with a focus on the innovative aspects of pixel color amplification and EfficientNetV2 model optimization. It covers aspects such as data sources, image preprocessing, model design, methodology validation, and the creation of a predictive system to assist medical professionals. Figure 1 shows the flowchart of this study, highlighting the key innovations. The dataset required preprocessing, involving crop methods using the Gradient Hough Transform and pixel color amplification using the Dark Channel Prior method. These preprocessing steps significantly enhance image quality, addressing issues of underexposure and overexposure, and making subtle retinal features more discernible. Data augmentation was applied to further improve the robustness and generalizability of the training dataset. Subsequently, this paper adopted the EfficientNetV2 model as the training architecture. The EfficientNetV2 model was optimized through techniques such as transfer learning and fine-tuning. This included advanced methods like early stopping, adaptive learning rates, and label smoothing to enhance performance and prevent overfitting. Hyperparameter tuning was also incorporated to further refine the model's generalization capabilities. Following these steps, the performance of the trained EfficientNetV2 model was evaluated using test data. A range of evaluation metrics and model interpretability methods, including Grad-CAM, were employed to assess the model's effectiveness and ensure that the model accurately captured disease-relevant features. The programming in this study was conducted within the PyCharm 2023.1.2 software environment.

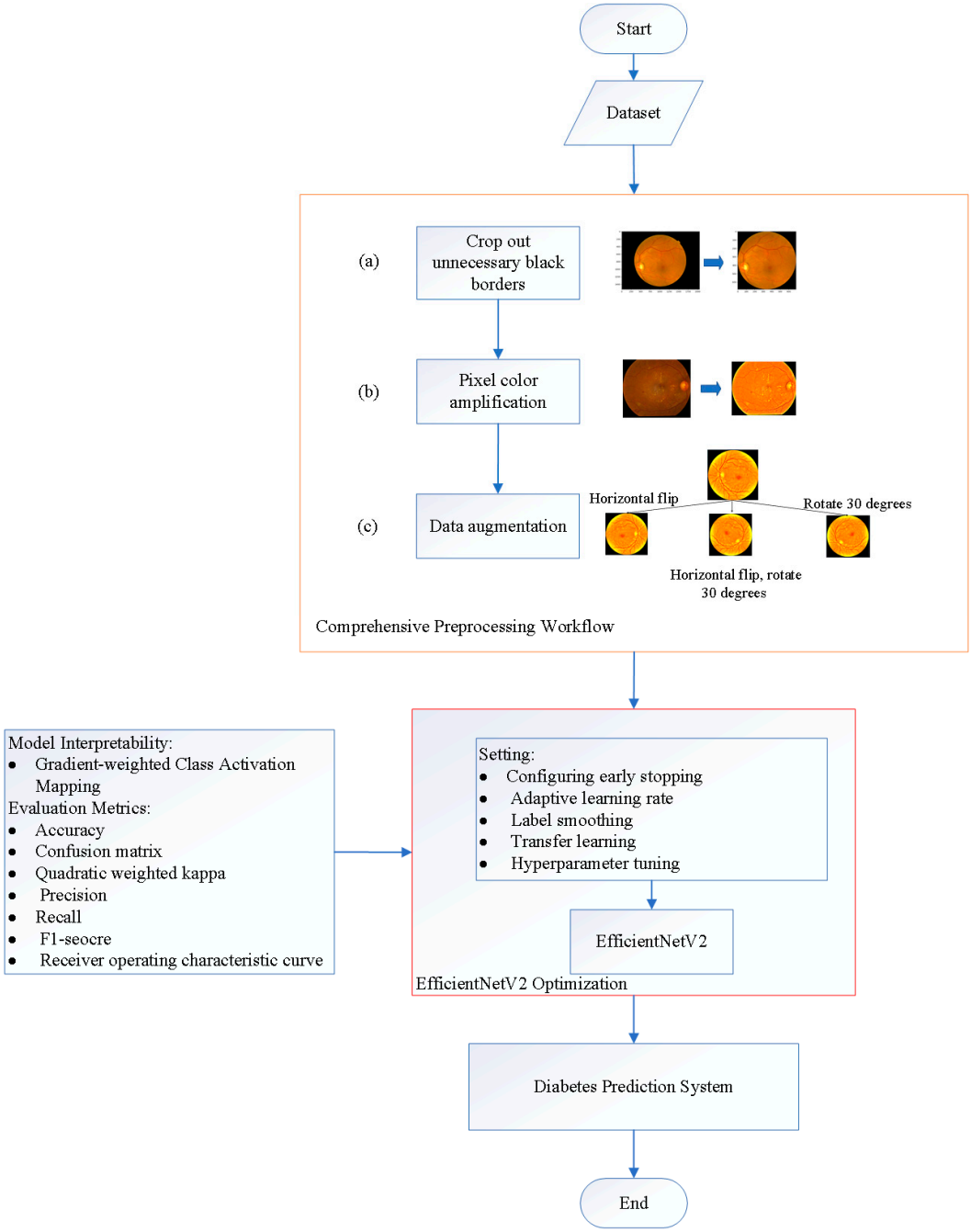


Figure 1. Shows the flowchart of this study, which encompasses a comprehensive preprocessing workflow, including (**a**) a crop method using Gradient Hough Transform, (**b**) pixel color amplification using Dark Channel Prior, and (**c**) data augmentation. It also includes model design, model interpretability, evaluation metrics for methodology validation, and the development of a predictive system. The key innovative steps, such as pixel color amplification and EfficientNetV2 model optimization, are highlighted to demonstrate their significant contributions to improving the accuracy and efficiency of DR detection.

Ultimately, the well-performing EfficientNetV2 model, integrated with these innovative preprocessing and optimization techniques, can be incorporated into a diabetes prediction system. This system provides medical professionals with a reliable tool for early detection and classification of DR, potentially alleviating their workload and improving patient outcomes. Figure 1 illustrates the complete workflow, emphasizing the innova-

tive steps that enhance the overall detection accuracy and efficiency. The forthcoming subsections will elaborate on the intricate methods utilized in each stage.

2.1. Dataset

In this study, the retinal image data primarily originated from the APTOS (Asia Pacific Tele-Ophthalmology Society) 2019 competition hosted on the Kaggle online platform (https://www.kaggle.com/c/aptos2019-blindness-detection, accessed on 2 February 2023). The competition provided a dataset of 3662 retinal images collected from diverse patients in the Indian region. A team of professional ophthalmologists meticulously evaluated and manually labeled these images to ensure accurate disease identification. The images were classified into five distinct categories: Healthy (0), Mild DR (1), Moderate DR (2), Severe DR (3), and Proliferative DR (4), based on varying degrees of severity.

This study partitioned the retinal image dataset based on a predefined ratio. Precisely, the dataset was split into three subsets: 80% for training data, 10% for validation data, and 10% for testing data, as depicted in Figure 2. This partitioning played a pivotal role in evaluating the model's performance throughout the training and testing phases. The training subset facilitated model training, enabling it to learn intricate features and patterns present within retinal images. The validation subset was instrumental in refining the model's hyperparameters, including learning rate and regularization parameters, to enhance its performance and generalization capacity. Lastly, the testing subset was utilized to gauge the model's ultimate performance, providing insights into its real-world applicability.

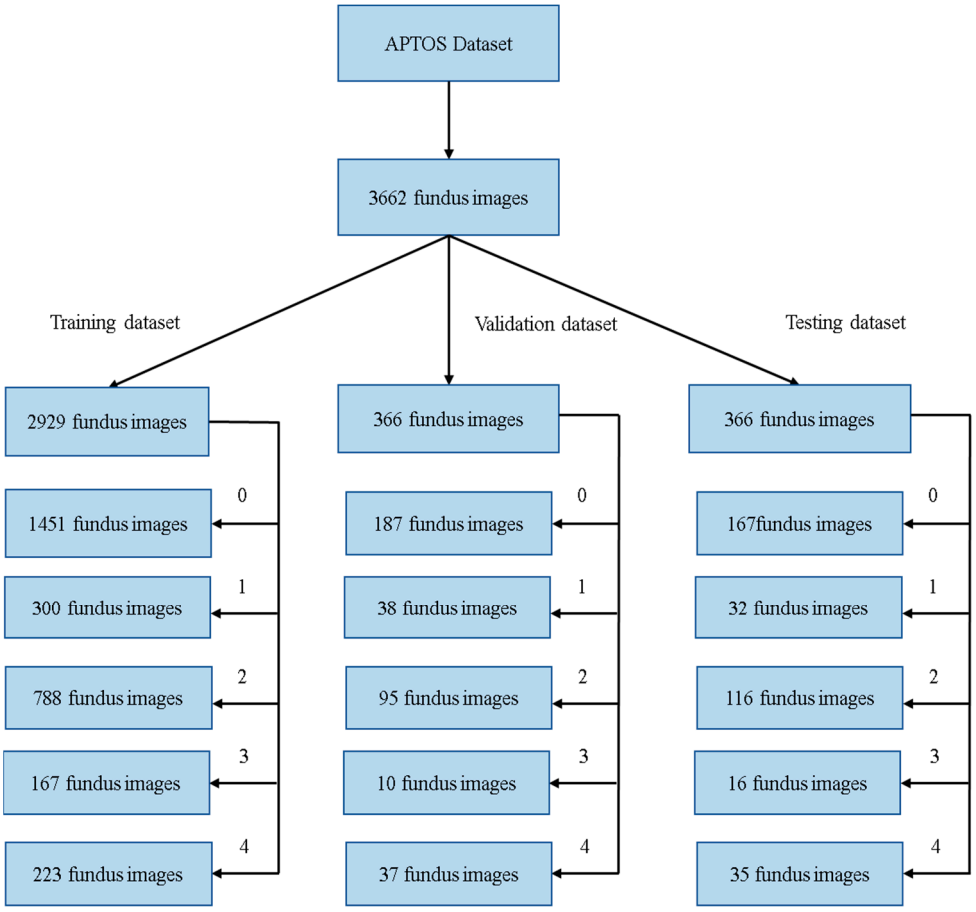


Figure 2. Dataset categorization resulted in five distinct classes: Healthy (0), Mild DR (1), Moderate DR (2), Severe DR (3), and Proliferative DR (4).

2.2. Comprehensive Preprocessing Workflow

In the realm of deep learning image recognition, image preprocessing holds paramount importance. Adequate preprocessing, such as the elimination of superfluous black borders around images through cropping, can attenuate irrelevant information, bolster recognition performance, and curtail computational resource demands. Proficient preprocessing can elevate the model's efficacy and precision. Figure 1 provides a visual depiction of the suggested preprocessing workflow. Our preprocessing steps create a robust training dataset that enhances model performance. This combination of techniques ensures that the model can generalize well across different imaging conditions and patient populations. The preprocessing workflow includes the following key steps:

1. Cropping: Removing superfluous black borders around images to focus on the relevant retinal area;
2. Pixel Color Amplification: Applying the Dark Channel Prior (DCP) method to enhance the visibility of retinal features;
3. Data Augmentation: Utilizing techniques such as rotation, scaling, and flipping to increase the diversity of the training dataset.

These preprocessing steps are crucial for creating a robust training dataset that enhances model performance by ensuring consistency and clarity in the input images.

2.2.1. Gradient Hough Transform for Cropping

In this study, the retinal image dataset used was sourced from different entities, and due to the utilization of diverse imaging devices, the images might exhibit varying specifications, as illustrated in Figure 3. These surplus black borders were irrelevant to disease discrimination, and their elimination can lead to reduced memory consumption, faster data retrieval, and increased model computation speed, ultimately enhancing training efficiency.

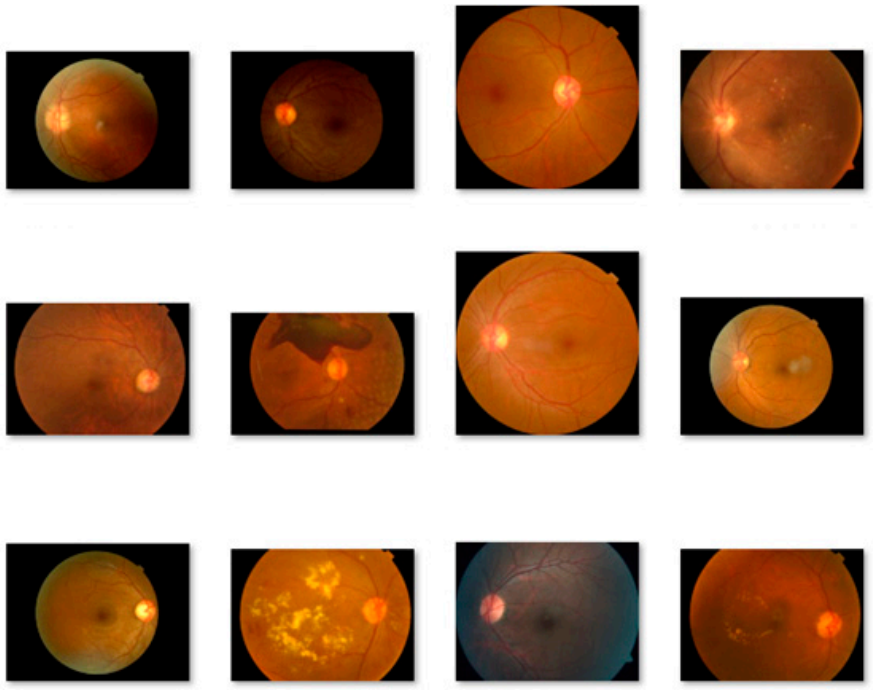


Figure 3. The original data contain excess black borders.

To address this issue, this study employed the Gradient Hough Transform [14] to mitigate the excessive black borders. The Gradient Hough Transform is a technique utilized for detecting straight lines within an image. Its principal mechanism involves edge detection to identify non-zero points in the image, subsequently calculating the gradient direction and intensity of each edge point. By considering the gradient direction and the

maximum and minimum distances of non-zero points, a line segment can be identified on the image. The intersection points of multiple line segments are then employed to ascertain the center's position (Point C), as illustrated in Figure 4.

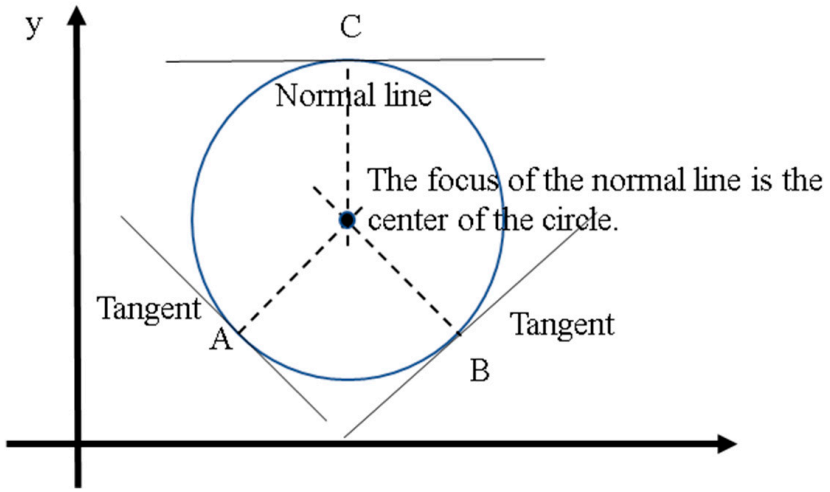


Figure 4. Detecting circles by identifying their center and radius. A: Edge point detected by the Gradient Hough Transform. B: Gradient direction of the edge point. C: Line segment identified by the Gradient Hough Transform using the edge points and their gradient directions. The intersection of these line segments determines the center position (Point C).

In retinal images, this study applied the Gradient Hough Transform to identify circular structures such as the retina. Once the circular structure of the retina was detected, this study could utilize its center and radius information for cropping, thereby removing unnecessary portions as shown in Figure 1a. This step reduces memory consumption and computational load while improving the accuracy of feature extraction.

2.2.2. Pixel Color Amplification

In the process of capturing retinal images, retinal photography devices are employed. The illumination apparatus of these devices emits specific light sources, such as visible light or infrared light, which are directed towards the back of the eye to illuminate the retinal structures. The reflected light is then captured by the retinal photography device, passing through a lens and entering the imaging system, where it is transformed into digital images. These images contain vital structures such as the macula, retinal vessels, and lens, which physicians use to assess ocular structures and overall eye health.

However, the acquisition of retinal images involves external illumination equipment, making them susceptible to various external factors like environmental conditions and human intervention. These factors can result in issues such as underexposure, overexposure, and unclear contours, as illustrated in Figure 5. To mitigate the impact of these external factors, this study employed a technique known as Dark Channel Prior (DCP) [12,13].

The use of the DCP method in our preprocessing pipeline significantly enhances image quality by addressing the illumination variations in retinal images. This technique is particularly innovative in its application to medical imaging, where maintaining the integrity of subtle features is crucial. By applying DCP, we improve the visibility of retinal features, ensuring clearer and more consistent images that are essential for accurate diagnosis and assessment. This enhancement process not only mitigates the negative effects of variable lighting conditions but also accentuates subtle pathological features, thereby facilitating more reliable and accurate DR detection.

Dark Channel Prior can be utilized for mitigating the haze effect in images. It can be expressed as shown in Equation (1):

$$\mathrm{I(x)} = \mathrm{J}(x)t(x) + \mathrm{A}(1 - t(x)) \tag{1}$$

where $\mathrm{I(x)}$ is the original image and $\mathrm{J}(x)$ represents the undistorted radiance image, reflecting the authentic colors and intensity of the scene untouched by haze or distortion. x represents each individual pixel location. The term $t(x)$ denotes the transmission rate, employed to measure the extent of light ray attenuation emitted from the observed surface at each pixel location x. A signifies the natural light present in the air, illustrating the uniformly dispersed color of light due to haze-induced scattering. The goal of Dark Channel Prior is to recover $\mathrm{J}(x)$, A, and $t(x)$ from $\mathrm{I(x)}$. By applying Equation (2), the undistorted image $\mathrm{J}(x)$ can be obtained, following the specific calculation method outlined as follows:

$$\mathrm{J}(x) = \frac{\mathrm{I(x)} - \mathrm{A}}{\max(\mathrm{t(x)}, \mathrm{t0})} + \mathrm{A} \tag{2}$$

where $\mathrm{t0} = 10^{-8}$ is used as a lower bound for the transmission map.

Figure 5. The images in the dataset were underexposed and had unclear contours.

By utilizing Equation (3), the transmission rate $t(x)$ can be calculated to generate the transmission map. This methodology enables the computation of the final transmission rate $t(x)$, as illustrated in Equation (3):

$$\begin{gathered} t(x) = \begin{cases} \text{solve\_t}(1 - \mathrm{I}, \mathrm{A} = 1) \\ \text{solve\_t}(\mathrm{I}, \mathrm{A} = 1) \\ 1 - \text{solve\_t}\ (\mathrm{I}, \mathrm{A} = 1) \\ 1 - \text{solve\_t}\ (1 - \mathrm{I}, \mathrm{A} = 1) \end{cases} \\ \text{solve\_t}\ (\mathrm{I}, \mathrm{A} = 1) = 1 - \min_{c} \min_{y \in \Omega_{(\mathbf{x})}} \frac{I^c(\mathrm{y})}{A^c} \\ \text{solve\_t}\ (1 - \mathrm{I}, \mathrm{A} = 1) = 1 - \min_{c} \min_{y \in \Omega_{(\mathbf{x})}} \frac{1 - I^c(\mathrm{y})}{A^c} \end{gathered} \tag{3}$$

where $c \in [R, G, B]$ is the color channel index. $\min_{c}$ and $\max_{c}$ denote the minimum and maximum values of the RGB color space at each pixel location, respectively. $\Omega_{(\mathbf{x})}$ is a local patch centered at pixel x. $I^c(\mathrm{y})$ represents the pixels within a square window $\Omega_{(\mathbf{x})}$ centered at x.

Using Equation (3), this study extended to four distinct variations of the transmission rate, targeting enhancements for both dark and bright regions through strong and weak amplifications. This resulted in eight distinct enhancement effects, as depicted in Figure 6,

originating from Figure 1b. The specific procedures for enhancing the transmission rate were as follows: for strong enhancement in dark regions, the value of the transmission rate $t(x)$ was increased, as shown in Figure 6c; for weak enhancement in dark regions, the value of the transmission rate $t(x)$ was slightly elevated, as depicted in Figure 6a. For strong enhancement in bright regions, the value of the transmission rate $t(x)$ was decreased, as demonstrated in Figure 6d; and for weak enhancement in bright regions, the value of the transmission rate $t(x)$ was slightly decreased, as illustrated in Figure 6b.

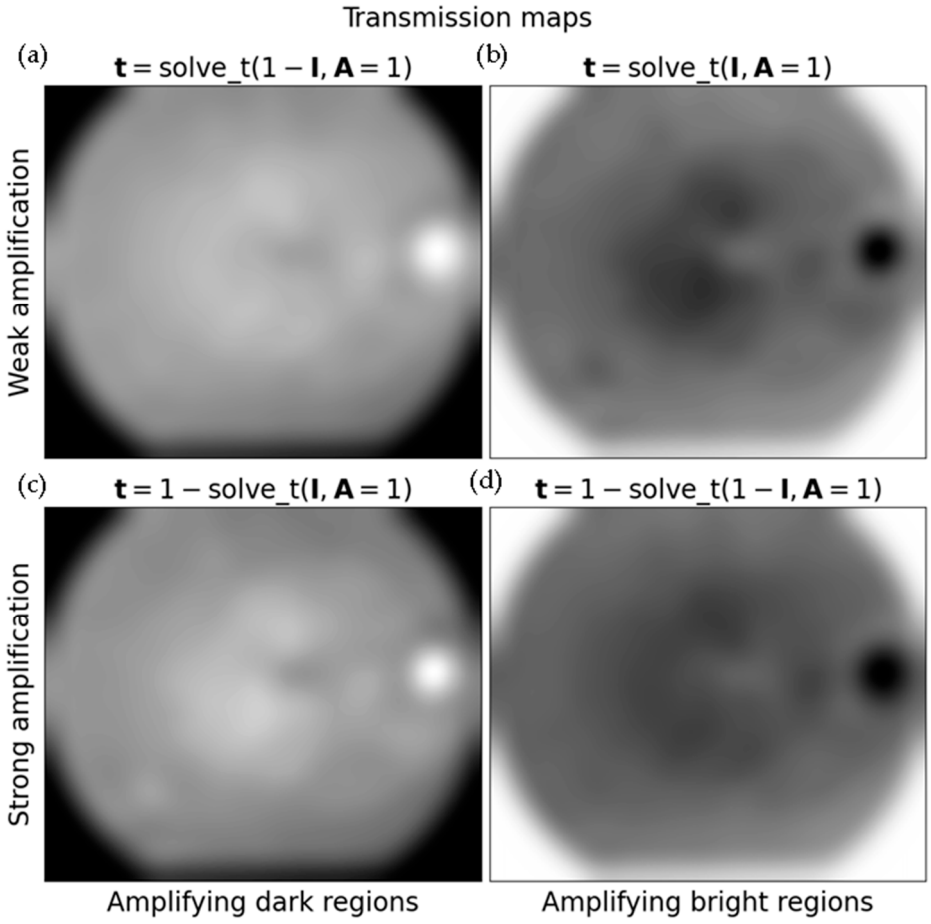


Figure 6. Four representations of transmission rate maps: (**a**) Weak enhancement in dark regions, (**b**) weak enhancement in bright regions, (**c**) strong enhancement in dark regions, and (**d**) strong enhancement in bright regions.

Combining the four transmission-rate enhancement methods (as shown in Figure 6 and described by Equation (3)) with Equation (2), where setting A = 1 resulted in brighter images and setting A = 0 resulted in darker images, generates a total of eight unique enhancement effects as seen in Figure 1b. These effects are illustrated in Figure 7. Through this approach, this study modulated the transmission rate to different extents, guided by the features of dark and bright regions within the images. This led to an enhancement in image quality. The specific selection of the enhancement method is contingent upon the specific application and requirements.

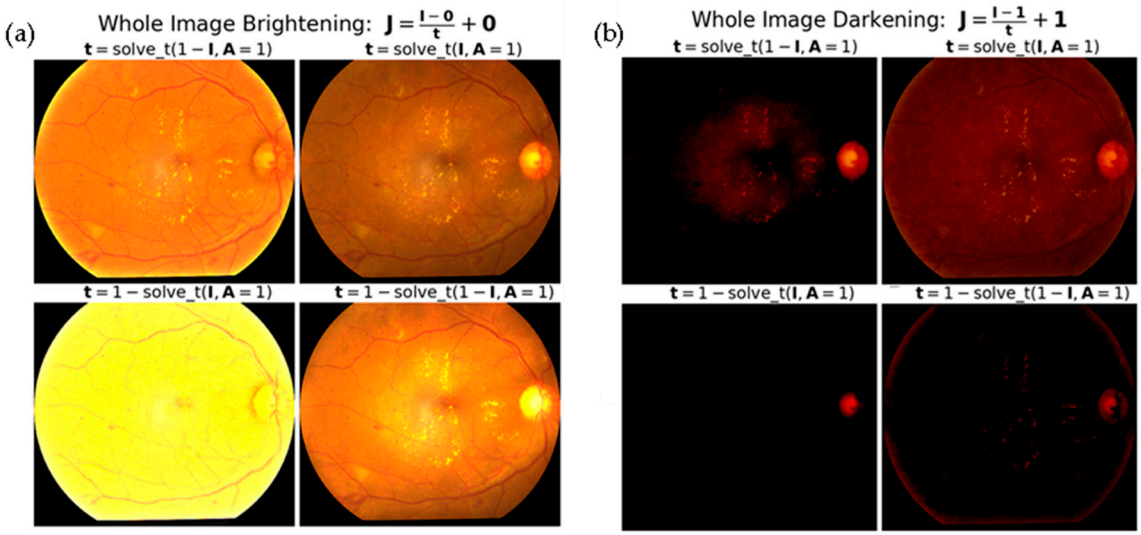


Figure 7. (**a**) Four enhanced images can be generated using four transmission-rate enhancement methods based on the bright images. (**b**) Four enhanced images can be produced using four transmission-rate enhancement methods based on the dark images.

2.2.3. Data Augmentation

Given the limited availability of fundus images for DR and the associated concerns regarding data privacy, the effective utilization of accessible data becomes of paramount importance. To address this, the present study employs the ImageDataGenerator technique provided by Keras (https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator, accessed on 2 February 2023) to implement data augmentation. ImageDataGenerator encompasses a range of image enhancement methods, including rotation, translation, scaling, cropping, and flipping, which collectively amplify the scope of the training dataset. Data augmentation entails modifying and enriching existing images to generate supplementary training samples. This strategy expands both the size and diversity of the training dataset, consequently bolstering the performance and generalizability of deep learning models.

In this study, three distinct techniques were employed (Figure 1c): image horizontal flipping, a 30-degree image rotation, and a combination of horizontal flipping followed by a 30-degree rotation. Image horizontal flipping involves horizontally mirroring the image, thereby creating a new image that is symmetrically opposite to the original. A 30-degree image rotation entails rotating the image counterclockwise by 30 degrees, resulting in a slightly altered orientation.

2.3. Proposed Diabetic Prediction System

EfficientNet [15], introduced by Google in 2019, is a convolutional neural network architecture that leverages a technique known as Compound Scaling. This technique uniformly scales the depth, width, and image resolution of the neural network. In conjunction with this, it incorporates Neural Architecture Search (NAS) to amplify network performance. Compound Scaling involves simultaneously adjusting the depth, width, and resolution of the network during model construction. Depth refers to the number of layers within the network, width pertains to the number of channels, and resolution signifies the dimensions of input images. Through this methodology, EfficientNet achieves impressive accuracy while utilizing a comparatively smaller number of parameters, a significant advantage it possesses.

EfficientNetV2 [16], introduced by Google in 2021 as an enhancement and refinement of the EfficientNet architecture, builds upon the foundational principles of EfficientNet. In contrast to EfficientNet, EfficientNetV2 implements improvements by replacing depthwise convolutional layers with regular convolutions in certain MBConv structures. This

innovative configuration is termed Fused-MBConv. This alteration stems from the recognition that while MBConv reduces parameter counts, it does not optimize GPU utilization, consequently resulting in reduced GPU efficiency. Architectural features of EfficientNetV2:

1. Fused-MBConv Layers: EfficientNetV2 replaces some depthwise separable convolutions with regular convolutions in the MBConv blocks. This adjustment, known as Fused-MBConv, improves GPU utilization by enhancing parallelism during computation, leading to faster training times and better performance;
2. Progressive Learning: EfficientNetV2 introduces a progressive learning approach, which gradually increases image size and regularization during training. This strategy helps the model adapt to larger images and more complex patterns over time, enhancing both training efficiency and final model accuracy;
3. Advanced Model Scaling: Like its predecessor, EfficientNetV2 uses compound scaling to adjust network width, depth, and resolution. However, EfficientNetV2 further refines this scaling technique, allowing for more precise control over these dimensions, which contributes to better performance with fewer computational resources;
4. Optimized Training Pipeline: EfficientNetV2 benefits from an optimized training pipeline that includes advanced techniques such as mixed precision training, which utilizes both 16-bit and 32-bit floating-point operations. This not only speeds up training but also reduces memory usage, making it feasible to train larger models on available hardware;
5. Neural Architecture Search (NAS): EfficientNetV2 continues to leverage NAS to find the optimal network architecture. This automated process explores various architectural configurations to identify the most efficient and effective design, tailored specifically for the given tasks.

Tan et al. also demonstrate the discrepancy in training time between EfficientNetV2 and EfficientNet [16]. Notably, EfficientNetV2 showcases substantial advancements in training time, highlighting its capability for expeditious model training. This acceleration contributes to swifter model development and experimentation. These observations collectively underscore the advantages of EfficientNetV2 in terms of both performance and efficiency.

This study chose the EfficientNetV2 model as the training framework due to its status as a pretrained model trained on extensive datasets, demonstrating high accuracy and exceptional feature extraction capabilities. During training, EfficientNetV2 combines Neural Architecture Search (NAS) and model scaling techniques. In comparison to other pretrained models, EfficientNetV2 achieves higher accuracy while maintaining a faster training pace within the same parameter count. This confers upon EfficientNetV2 a competitive advantage in the realm of retinal image classification.

In the context of this study, the selection of EfficientNetV2 as the training architecture effectively enhanced the model's performance and precision. Furthermore, an important benefit of EfficientNetV2 lies in its capability to promptly adapt and update with new retinal image data. As new retinal image data are collected, EfficientNetV2 can rapidly incorporate and learn from new features, leading to a continuous improvement in model performance as the dataset grows. In the upcoming results section, this study will conduct a comparative analysis of EfficientNetV2 against other models to further evaluate its effectiveness in retinal image classification tasks. This study utilized the pretrained EfficientNetV2S as the foundational architecture. The model training was carried out through transfer learning. Throughout the training process, various techniques, including early stopping, adaptive learning rate adjustment, label smoothing, and hyperparameter tuning (fine-tuning), were incorporated to enhance the model's overall performance.

2.3.1. EfficientNetV2 Optimization

We leveraged the EfficientNetV2 model, optimizing it for DR detection through transfer learning and fine-tuning, which resulted in higher accuracy and faster convergence compared to traditional models. This optimization includes early stopping, adaptive learn-

ing rates, label smoothing, and hyperparameter tuning tailored to the specific requirements of DR classification.

Early Stopping

To boost training efficiency and overall performance, we implemented the early stopping technique. Early stopping involves monitoring the performance on a validation set throughout the training process. If performance improvement stagnates, the training process is terminated, effectively preventing overfitting and conserving training time. The mechanism entails defining a threshold. If, after a certain number of iterations, the model's performance fails to improve or meet a predetermined threshold, training is halted prematurely. Subsequently, the saved optimal model parameters are employed for subsequent predictions or applications.

Adaptive Learning Rate

The learning rate directly influences the extent of weight updates during each iteration. To expedite the quest for the optimal solution, it is common practice to set a relatively high learning rate. However, excessively high learning rates can result in overshooting the optimal solution as the training progresses. To tackle this challenge, we adopted an adaptive learning rate approach, which automatically fine-tunes the learning rate based on the progress of training [17]. As illustrated in Equation (4), where lr denotes the learning rate, factor represents the reduction factor, and lr' stands for the freshly computed learning rate; this adaptive learning rate dynamically adjusts in response to the number of iterations. This technique serves to counteract overfitting and attains the optimal outcomes.

$$\mathrm{lr}' = \mathrm{lr} \times \mathrm{factor} \tag{4}$$

The adaptive learning rate dynamically adjusts in response to the number of iterations, counteracting overfitting and attaining optimal outcomes. In this study, the learning rate was reduced by a factor of 0.1 if no improvement was observed after every 5 epochs.

Label Smoothing

Clear data categorization is of utmost importance for both model training and predictions, particularly in the context of DR grading, which requires the evaluation of specialized ophthalmologists. Different medical professionals may provide differing judgments for the same retinal image, leading to label inconsistencies. As highlighted by Google Developers during the TensorFlow Dev Summit 2017 presentation [18,19], label smoothing techniques become particularly vital when training deep learning models. Label smoothing serves to diminish the model's excessive confidence in training data labels, thereby enhancing the model's generalization capacity and stability. The calculation for label smoothing is provided by Equation (5), where ϵ represents the smoothing factor, y stands for the original model's prediction values, and n is the reciprocal of the total number of categories. Assuming a smoothing factor of 0.2 and an initial prediction of [0 1], utilizing the label smoothing formula would yield an output y' of [0.1 0.9].

$$y' = (1 - \epsilon) \times y + \epsilon \times n \tag{5}$$

Transfer Learning and Fine-Tuning

By applying transfer learning and fine-tuning specifically for DR detection, our optimized EfficientNetV2 model demonstrates superior performance in terms of both accuracy and training efficiency. This study utilized the pretrained EfficientNetV2S as our foundational architecture. The approach of transfer learning was employed [20], training the model in two distinct phases.

In the first phase, all model parameters were held constant, while solely the last 5 layers were unfrozen and trained, encompassing newly introduced fully connected layers. This step was taken as the top layers of the pretrained model had typically captured

abstract and advanced features that were likely highly effective for the training task at hand. Consequently, the freezing of these top layers prevented excessive adjustments to their weights. Conversely, the lower layers typically learned more generalized feature representations, making them better suited for adaptation to new target tasks. Thus, in this phase, we focused on fine-tuning these last few layers while introducing L1 and L2 regularization as well as dropout techniques in the fully connected layers to counteract overfitting. Furthermore, ELU activation functions and categorical cross-entropy loss functions were set, all of which are commonly employed configurations for classification tasks.

In the second phase, we unfroze all the parameters of the trained model, allowing the training of any layer in the model. This unfreezing of all layers enhanced the model's adaptability, enabling it to better fit the training data. Simultaneously, the model continued to leverage the general features learned during pretraining, thereby assisting in adapting to new data. This phase served to mitigate the risk of overfitting, as a greater number of parameters could be adjusted, thus enhancing the model's expressiveness. Furthermore, the aforementioned model-training approaches (early stopping, adaptive learning rate, and label smoothing) were introduced in the second phase.

By undergoing these two-stage training procedures, we were able to fully leverage the feature extraction capabilities of the pretrained model while adjusting for new target tasks, ultimately enhancing the model's performance and generalization capabilities. This fine-tuning approach is widely utilized in transfer learning, allowing for effective utilization of the advantages of pretrained models while tailoring them to specific target tasks.

2.4. Model Interpretability

The research and development of model interpretability have experienced rapid advancements in recent years. Many initially opaque machine learning algorithms, such as Random Forest, Gradient Boosting, and even Deep Learning Models, have gradually evolved to produce human-understandable outcomes. In this study, the Grad-CAM (Gradient-weighted Class Activation Mapping) method was employed to assess whether the model learned accurate disease indicators.

CAM [21] replaces the flattening fully connected layer typically used with a Global Average Pooling (GAP) layer, connecting the weights of the classification layer with the feature maps of the convolutional layer. By weighting each feature map according to its corresponding weight, CAM is obtained.

Grad-CAM [22], an enhanced CAM method, extends beyond the Global Average Pooling layer of the final convolutional layer. Grad-CAM calculates the partial derivatives of neuron activations with respect to the feature maps of the last convolutional layer. This is achieved using backpropagation to compute gradients, as shown in Equation (13). Here, w_k^c represents the weight of class c for the k-th feature map in the last convolutional layer, ∂Y^c denotes the output gradient, and ∂A_{ij}^k signifies the gradient of the feature map in the final convolutional layer.

$$w_k^c = \sum_i \sum_j \frac{\partial Y^c}{\partial A_{ij}^k} \tag{6}$$

Through this approach, it becomes possible to visualize and comprehend the neural network's attention to feature maps for specific classes based on gradient weights. Such visualizations aid in understanding the decision-making process of the neural network, providing insights into which regions play a crucial role in identifying specific classes.

2.5. Evaluation Metrics

When evaluating model performance, it is essential to utilize evaluation metrics to quantify its effectiveness. Evaluation metrics serve as standards for assessing a model's performance across different tasks, helping this study determine its accuracy, stability, and reliability. This section introduces common evaluation metrics and provides the corresponding calculation methods [23]. These metrics include accuracy, confusion matrix,

quadratic weighted kappa (QWK), precision, recall, F1-score, cross-entropy loss, and ROC curve (Receiver Operating Characteristic curve). The formulas for these evaluation metrics are provided as follows:

The accuracy calculation yields values between 0 and 1, with higher values indicating stronger predictive capabilities.

$$accuracy = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{Total\ Samples} \tag{7}$$

The confusion matrix compares the model's predicted results with actual labels. Each column of the matrix represents the predicted results for a class, while each row represents an actual class. By examining the confusion matrix, this study can directly visualize the model's prediction results.

The QWK is an evaluation metric designed for multi-class problems, especially suitable for labels with ordinal or hierarchical relationships. It performs well in handling imbalanced datasets. The calculation principle involves using the confusion matrix and multiplying it by a weight matrix. The weight matrix is defined based on the square root of the differences between predicted results. This approach penalizes larger discrepancies between predictions and actual outcomes, allowing it to represent model performance on imbalanced data. By calculating the weighted error, an evaluation metric ranging from -1 to 1 is obtained. A value closer to 1 indicates higher consistency between model predictions and true values. The weighted kappa matrix is computed using Equations (8) and (9), where i and j, respectively, represent the indices of the classes. N represents the total number of classes, $O_{i,j}$ represents the predicted confusion matrix, and $E_{i,j}$ represents the actual true results:

$$w_{i,j} = \frac{(i-j)^2}{(N-1)^2} \tag{8}$$

$$Kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}} \tag{9}$$

The precision calculation measures the accuracy of positive predictions among predicted positive samples, emphasizing how many predicted positives are true positives. Recall, also known as sensitivity, measures the model's ability to correctly predict positives among actual positive samples.

Precision and recall often trade off against each other, leading to the development of another metric, the F1-score, which calculates the harmonic mean of precision and recall:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FP} \tag{11}$$

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

Cross-entropy loss measures the dissimilarity between the predicted probability distribution and the true label distribution. In the context of a multi-class classification problem with N classes, the cross-entropy loss is defined as follows:

$$loss = -\sum_{i=1}^{N} y_i \times \log(p_i)$$

where N is the number of classes. y_i is the true label for class i (either 0 or 1). p_i is the predicted probability of class i from the model's output.

The ROC curve is a graphical representation that illustrates the variation between True Positive Rate (TPR) and False Positive Rate (FPR) across different decision thresholds. TPR refers to the rate of correctly identifying positives, while FPR indicates the rate of

incorrectly identifying negatives as positives. The Area Under the ROC Curve (AUC ROC) represents the area under the ROC curve and ranges from 0.1 to 1. A higher AUC value indicates better model performance.

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

$$FPR = \frac{TP}{TP + FN} \tag{14}$$

3. Experimental Results and Analysis

In the initial validation phase of this study, due to the limited memory capacity of the local NVIDIA GeForce RTX 3060 GPU (12 GB), a smaller batch size (8) and fewer epochs were used for preliminary preprocessing. This was conducted to ensure that the model could operate within the available memory capacity and undergo preliminary model training and evaluation.

However, as the research progressed and the complexity of the model increased, larger memory capacity became necessary to support a larger batch size (32) and more epochs. To meet these requirements, the experiments were transferred to the Colab environment provided by Google, utilizing the NVIDIA Tesla T4 GPU with a larger memory capacity. This allowed for efficient utilization of computational resources and enabled larger-scale model training and evaluation.

3.1. Comparison Results of Image Preprocessing

In this section, we compare the performance of our DR detection model with different image preprocessing techniques, focusing on the impact of pixel color amplification using the Dark Channel Prior (DCP) method. The goal is to analyze how each preprocessing step influences the overall model performance, particularly in terms of accuracy and robustness. In this study, we employed the pretrained EfficientNetV2S as the foundational architecture. The approach of transfer learning was utilized [20] involving two distinct phases of model training, as described in Section Transfer Learning and Fine-Tuning. Table 1 presents the configurations of hyperparameters and structures for the first and second phases of EfficientNetV2S.

Table 1. The configurations of hyperparameters and structures for the first and second phases of EfficientNetV2S.

Description	Output Shape of First Phases	Output Shape of Second Phases
Image_height	256	256,256
Image_width	256	256
Image_channels	3	3
lr	0.001	0.00002 (initiation value) adjusted by Equation (4)
Layers	512	512
Epoch	5	50 with configuring early stopping
Batch_size	16	16
Activation	[elu,softmax]	[elu,softmax]
Dropout	0.5	0.5
Label	[0, 1, 2, 3, 4]	Label smoothing by Equation (5)

To assess the influence of employing pixel color augmentation techniques on DR detection, this section conducts a performance comparison between two models: one with the utilization of pixel color augmentation and the other without it, during the second phase. To ensure experiment fairness, both models share identical design and hyperparameter selection throughout the training process, differing solely in the implementation of pixel color augmentation techniques in the second phase. As observed in Figure 8a, the model

trained without pixel color augmentation exhibits significant oscillations during training and may encounter issues of overfitting. In contrast, Figure 8b presents the training results of the model using pixel color augmentation, showcasing smoother training progress and relatively faster convergence speed. Moreover, when early stopping techniques are employed, the model with pixel color augmentation demonstrates not only shortened training time but also higher accuracy. Table 2 provides the relevant results for EfficientNetV2S with and without pixel color augmentation technique.

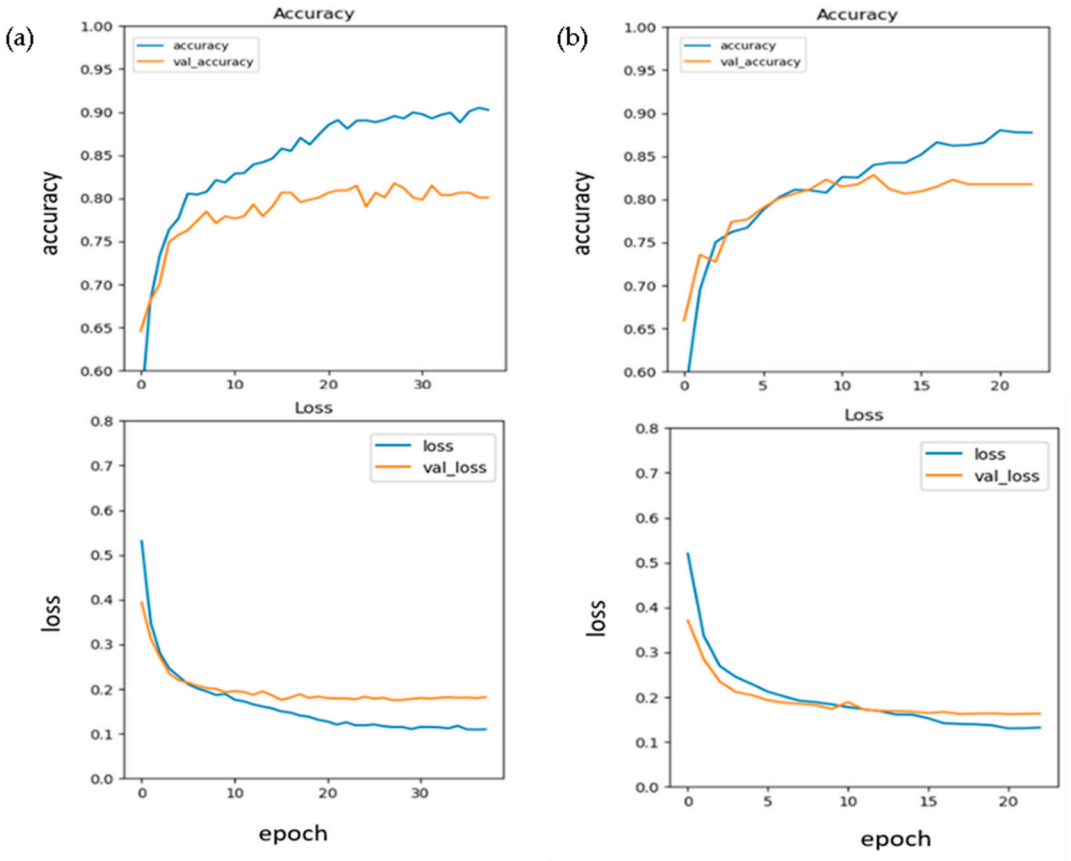


Figure 8. (**a**) Performance without pixel color augmentation and (**b**) performance with pixel color augmentation.

Table 2. Comparison results of EfficientNetV2S with and without pixel color augmentation technique.

	Accuracy/Loss	**Performance without Pixel Color Augmentation**	**Performance with Pixel Color Augmentation**
Training dataset	Accuracy	0.9027	0.8774
	Loss	0.1100	0.1323
Validation dataset	Accuracy	0.8011	0.8274
	Loss	0.1819	0.1633
Testing dataset	Accuracy	0.8279	0.8443
	Loss	0.1688	0.1489

3.1.1. Detailed Comparison

The preprocessing workflow consists of three main techniques: cropping, pixel color augmentation, and data augmentation. To assess the impact of each technique, we conducted experiments with different combinations of preprocessing steps. Table 2 presents the accuracy and loss for models trained with and without pixel color augmentation.

Looking solely at the accuracy and loss may not provide a comprehensive representation of the models' performance. Table 3 presents additional metrics that are computed based on the testing dataset, offering a more detailed comparison and assessment. These metrics offer a holistic view to evaluate the performance differences between the two models.

Table 3. Precision, QWK score, recall, precision, and F1-score of testing dataset.

Testing Dataset	Performance without Pixel Color Augmentation	Performance with Pixel Color Augmentation
QWK	0.8804	0.8913
Precision	0.7211	0.7641
Recall	0.6572	0.6711
F1-score	0.6738	0.6997

By considering the additional metrics provided in Table 3, a more thorough evaluation of the models' performance disparities can be achieved. These metrics may include precision, QWK score, recall, precision, and F1-score, among others. Through these metrics, the substantial benefits of using pixel color augmentation for diabetic retinal lesion classification become evident.

3.1.2. Analysis of Results

The results indicate that the inclusion of pixel color augmentation leads to improved model performance on the validation and testing datasets. Specifically:

1. Improved Accuracy and Reduced Loss: The model trained with pixel color augmentation achieved higher accuracy and lower loss across both validation and testing datasets compared to the model without this preprocessing step. This suggests that the pixel color augmentation technique enhances the model's ability to generalize to new data, thereby reducing overfitting;
2. Smoother Training Convergence: As shown in Figure 8b, the training process with pixel color augmentation exhibited fewer oscillations and a more stable convergence. This stability can be attributed to the enhanced image quality, which provides the model with clearer and more consistent features to learn from;
3. Enhanced Performance Metrics: The additional metrics in Table 3, such as precision, QWK score, recall, and F1-score, further confirm the effectiveness of pixel color augmentation. These metrics show that the model with pixel color augmentation not only predicts more accurately but also captures a higher proportion of relevant positive cases, leading to a more balanced performance.

3.1.3. Reasons for Improvement

The primary reasons for the observed improvements include the following:

1. Enhanced Image Quality: Pixel color augmentation improves the visibility of retinal features by mitigating issues of underexposure and overexposure. This allows the model to learn more relevant features, which are crucial for accurate DR detection;
2. Robustness to Variability: By addressing illumination variations in retinal images, the pixel color augmentation technique ensures that the model is exposed to a more consistent dataset. This consistency helps the model to generalize better across different imaging conditions.

The application of pixel color augmentation significantly enhances the performance of our DR detection model. The improvements in accuracy, loss, and additional performance metrics, coupled with smoother training convergence, highlight the importance of this preprocessing step in developing robust and reliable medical imaging models.

3.2. Model Comparison

This subsection provides a comprehensive comparison between our optimized EfficientNetV2 model and other commonly used models such as EfficientNetB5, VGG16, NASNet, and ResNet50. We detail how specific optimization techniques (e.g., transfer learning, fine-tuning, early stopping, adaptive learning rates, and label smoothing) contribute to the superior performance of our model. This comparison highlights the advantages of our approach in terms of accuracy, training efficiency, and robustness across different datasets.

3.2.1. Detailed Comparison

Considering the need for regular model updates as the dataset grows, this study opted for the EfficientNetV2S model within the EfficientNetV2 architecture, which offers a balance of minimal complexity and high accuracy. A comparison was conducted between EfficientNetV2S and EfficientNetB5, which achieves similar accuracy [11]. Both models were trained on the same preprocessed data under identical conditions as in Table 1. Figure 9a illustrates the parameter count for EfficientNetB5, while Figure 9b displays the parameter count for EfficientNetV2S.

```
global_average_pooling2d_5 (Gl  (None, 2048)        0           ['top_activation[0][0]']
obalAveragePooling2D)

dropout_10 (Dropout)           (None, 2048)         0           ['global_average_pooling2d_5[0][0
                                                                ]']

dense_5 (Dense)                (None, 512)          1049088     ['dropout_10[0][0]']

dropout_11 (Dropout)           (None, 512)          0           ['dense_5[0][0]']

pred (Dense)                   (None, 5)            2565        ['dropout_11[0][0]']

==================================================================================================
Total params: 29,565,180
Trainable params: 29,392,437
Non-trainable params: 172,743
__________________________________________________________________________________________________
```

(**a**)

```
top_activation (Activation)    (None, 8, 8, 1280)   0           ['top_bn[0][0]']

global_average_pooling2d_3 (Gl  (None, 1280)        0           ['top_activation[0][0]']
obalAveragePooling2D)

dropout_6 (Dropout)            (None, 1280)         0           ['global_average_pooling2d_3[0][0
                                                                ]']

dense_3 (Dense)                (None, 512)          655872      ['dropout_6[0][0]']

dropout_7 (Dropout)            (None, 512)          0           ['dense_3[0][0]']

pred (Dense)                   (None, 5)            2565        ['dropout_7[0][0]']

==================================================================================================
Total params: 20,989,797
Trainable params: 20,835,925
Non-trainable params: 153,872
__________________________________________________________________________________________________
```

(**b**)

Figure 9. (**a**) The parameter counts of EfficientNetB5 and (**b**) the parameter counts of EfficientNetV2S.

3.2.2. Analysis of Results

These findings suggest that EfficientNetV2S outperforms EfficientNetB5 in terms of accuracy and is more efficient in terms of training time under the same settings. The primary goal was to create a system that offers rapid and precise retinal lesion diagnosis for diabetic patients. EfficientNetV2S demonstrated superior performance in terms of accuracy, convergence speed, and training efficiency compared to other options like EfficientNetB5.

1. Improved Accuracy and Reduced Loss: According to Table 4, EfficientNetV2S achieved a test accuracy and loss of 0.8443 and 0.1489, respectively, while EfficientNetB5 attained a test accuracy and loss of 0.8361 and 0.1676, respectively. This suggests that the EfficientNetV2S model, with its advanced optimization techniques, enhances the model's ability to generalize to new data, thereby reducing overfitting;
2. Training Efficiency: EfficientNetV2S exhibited approximately 40% faster overall training time compared to EfficientNetB5, as indicated by the training process shown in Figure 9a,b. This efficiency is crucial for regular model updates as the dataset grows.

Table 4. Comparison results between EfficientNetB5 and EfficientNetV2S models.

	Accuracy/Loss	**EfficientNetB5**	**EfficientNetV2S**
Training dataset	Accuracy	0.8539	0.8774
	Loss	0.1451	0.1323
Validation dataset	Accuracy	0.8229	0.8274
	Loss	0.1713	0.1633
Testing dataset	Accuracy	0.8361	0.8443
	Loss	0.1676	0.1489
Time per epoch		77 s	43 s

3.2.3. Reasons for Improvement

The primary reasons for the observed improvements include the following:

1. Transfer Learning and Fine-Tuning: These techniques allow the EfficientNetV2S model to leverage pretrained weights, providing a strong starting point and enabling faster convergence and higher accuracy;
2. Advanced Optimization Techniques: Early stopping, adaptive learning rates, and label smoothing help in preventing overfitting, improving generalization, and stabilizing training. These techniques ensure that the model learns efficiently and effectively;
3. Model Architecture: EfficientNetV2S's compound scaling approach optimally balances network depth, width, and resolution, resulting in a more powerful and efficient model compared to others.

Furthermore, considering the continuous data collection and periodic retraining of the model to enhance its performance over time, the choice of EfficientNetV2S offered an optimal configuration that facilitated efficient model updates. This approach enabled the system to adapt and improve as new data became available, supporting the objective of providing accurate and up-to-date diagnosis results to ophthalmologists. Therefore, the selection of EfficientNetV2S as the model architecture in this study was a strategic decision made in alignment with the goal of creating a practical and effective automated diagnosis system for diabetic retinal lesions.

3.2.4. Comparison with Prior Research

The earlier results have already covered the performance comparison between EfficientNetV2S and EfficientNetB5, as well as clarified the rationale for choosing EfficientNetV2S for this study. This study proceeds to a comparison with prior research that utilized the same ATPOS dataset for training. The paper published in 2023 represents our team's research on DR classification [6], in which we employed preprocessing quality filters to

eliminate low-quality images and trained an EfficientNet model. S. H. Kassani et al. [5] employed minimal pooling preprocessing on retinal images and employed a variety of pretrained models for training. J. D. Bodapati et al. [4] utilized multiple pretrained models for feature extraction and incorporated a pooling technique for fusion.

Table 5 displays a binary classification of DR (presence vs. absence of disease), while Table 6 offers a comparison of predicting the five DR categories using diverse models and preprocessing approaches. By examining Tables 5 and 6, it becomes evident that when utilizing the same training data, the utilization of pixel color augmentation in conjunction with EfficientNetV2S, as employed in this study, results in higher predictive accuracy. These comparisons highlight the substantial benefits of using pixel color augmentation and EfficientNetV2S for diabetic retinal lesion classification, demonstrating higher predictive accuracy compared to previous methods.

Table 5. Comparison accuracy for predicting DR as presence vs. absence of disease.

	Year	Model	Accuracy
The proposed method in this study	2024	EfficientNetV2	99.4%
J. D. Bodapati et al. [4]	2020	VGG16-fc1	97.27%
		VGG16-fc2	97.32%
		NASNet	97.14%
		Xception	97.41%
		Inception ResNetV2	97.34%

Table 6. Comparison accuracy for predicting DR into five categories.

	Year	Model	Accuracy
The proposed method in this study	2024	EfficientNetV2	84.42%
Our team's previous research [6]	2022	EfficientNet	79.20%
S. H. Kassani et al. [5]	2019	Xception	79.59%
		InceptionV3	78.72%
		MobileNet	79.01%
		ResNet50	74.64%
		Modified Xception	83.09%
J. D. Bodapati et al. [4]	2020	Blended features + DNN	80.96%

3.2.5. Confusion Matrix and ROC Curves

Additionally, Figure 10a presents the confusion matrix of EfficientNetV2S for each retinal image in the testing dataset, displaying the predicted results alongside their corresponding true classes. This illustration allows us to comprehensively observe the model's predictive performance in this study. However, due to the data imbalance where the number of samples in each class varies, a simple observation of the confusion matrix based on images alone might not accurately assess the model's overall performance. To gain a better understanding of the model's performance, this study employs a percentage-based representation of the confusion matrix as depicted in Figure 10b. Through the percentage-based confusion matrix shown in Figure 10b, predictions for each class can be observed. With the exception of classes three (Severe) and four (Proliferative) that pose challenges due to limited training data, the remaining three classes exhibit an accuracy rate of nearly 80%.

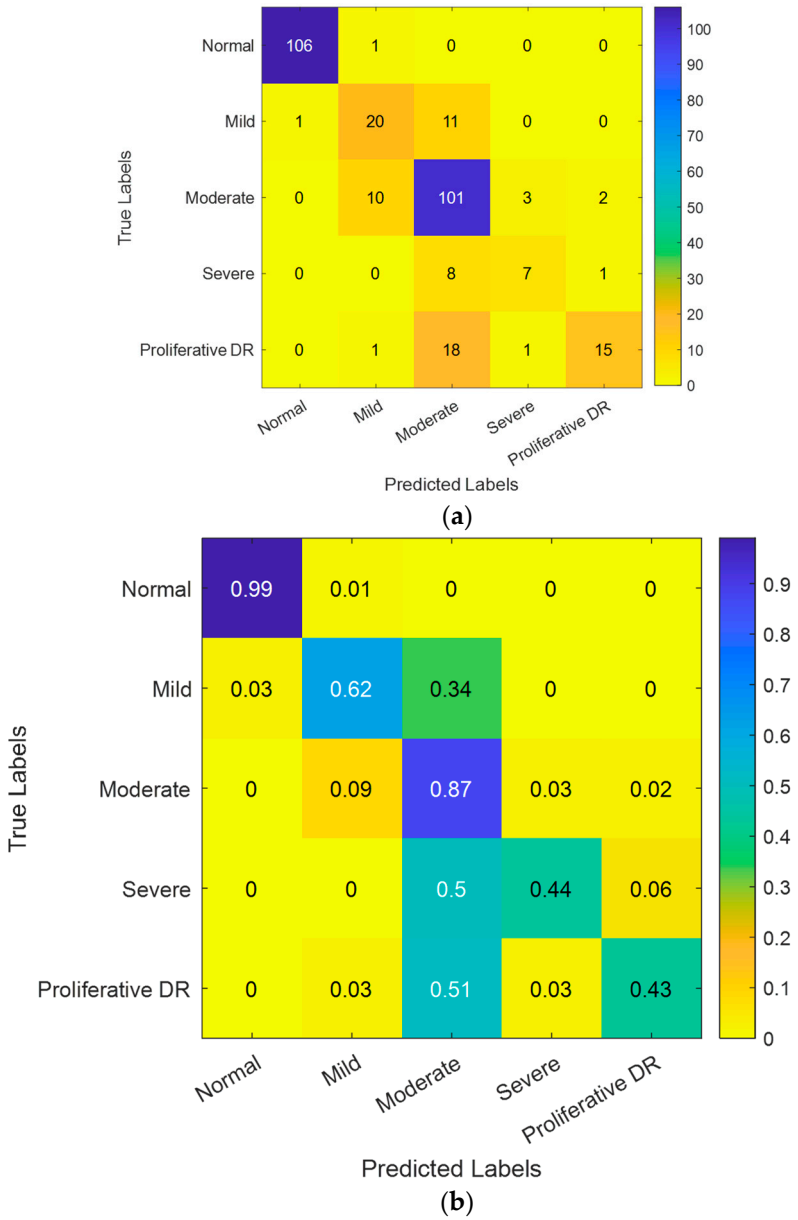


Figure 10. (**a**) Confusion matrix of the EfficientNetV2S model predictions in the testing dataset and (**b**) percentage-based representation of the confusion matrix of the EfficientNetV2S model predictions in the testing dataset.

When simplifying DR of EfficientNetV2S into two categories in the testing dataset: "disease" and "no disease", as shown in Figure 11, the model achieves an impressive accuracy of 99.4%. Only one healthy retinal image was mistakenly classified as having a disease, showcasing the model's strong performance in early detection of disease symptoms.

Figure 12a illustrates the ROC curves of EfficientNetV2S for the five categories of DR in the testing dataset, with their corresponding AUC scores: 1 for "Normal", 0.951 for "Mild", 0.948 for "Moderate", 0.955 for "Severe", and 0.927 for "Proliferative_DR". These AUC scores reflect the high classification capabilities, as each category exhibits AUC values close to 1. Figure 12b displays the ROC curve of EfficientNetV2S for the binary classification of DR (disease present vs. absent) in the testing dataset. In this binary classification scenario, the AUC reaches 0.992, indicating the model's strong discriminatory ability in this two-class classification.

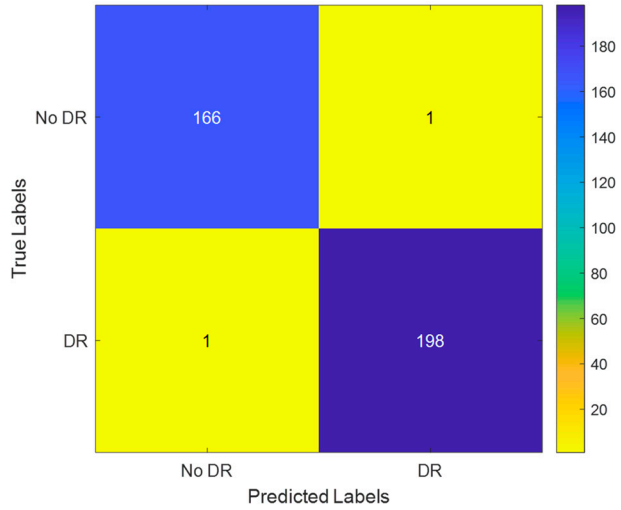
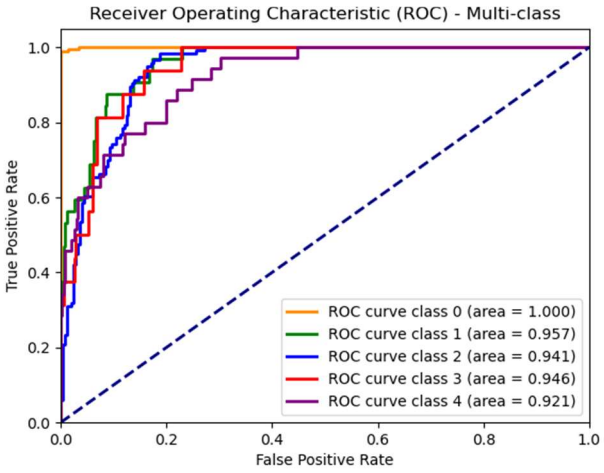
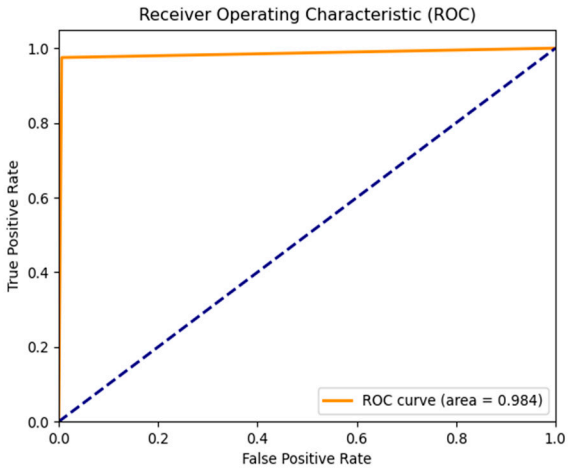


Figure 11. Confusion matrix for DR classification.



(**a**)

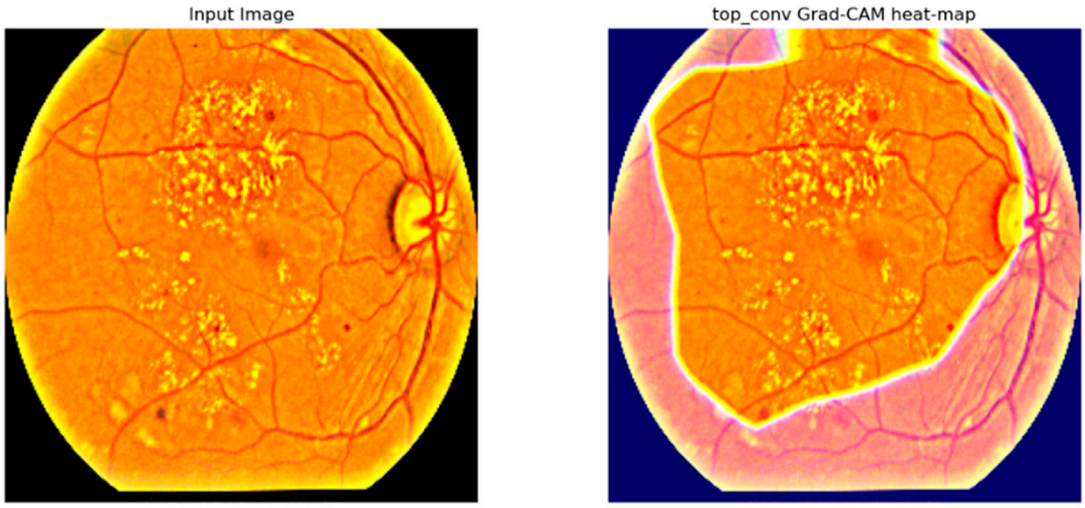


(**b**)

Figure 12. (**a**) ROC curves for the five categories of DR and (**b**) ROC curve for the binary classification of DR.

3.3. Visualization of Model Results

By employing the Grad-CAM technique [22], heatmaps are overlaid onto the original input images, allowing for the observation of whether the model captures disease-related features during classification. As shown in Figure 13, the model accurately captures exudate features in retinal images. This further substantiates the model's learning process moving in the correct direction. The heatmap provided by the Grad-CAM technique offers an intuitive way to comprehend the regions and critical features that the model focuses on during the classification process. This serves as a valuable tool for doctors and researchers, aiding them in better understanding the basis of the model's decisions and providing more accurate diagnoses and research outcomes.



(**a**) (**b**)

Figure 13. Observing the features captured by the model: (**a**) Model training input image. (**b**) Image after applying the heatmap visualization.

4. System Implementation for Diabetic Prediction System

To provide practical assistance to ophthalmologists, this study has developed a comprehensive and automatically updatable DR classification system. The aim of this system is to offer physicians a tool for assistance and preliminary diagnosis of the disease. The system comprises a web interface where users can upload fundus images. These images are then transmitted to the server-side through a Web API for disease assessment. The results are returned and stored in a database. After collecting sufficient new data, the model will be retrained, and if the accuracy of the new model is higher, the system will be updated. Figure 14 illustrates the workflow of the prediction system, while the subroutine in Figure 14 depicts the automatic updating process of the system. This study primarily utilized the Django web framework (Django 4.2.3) to establish a Web API [24] and invoke the pretrained EfficientNetV2 model. It integrated SQL server management studio (SSMS) 2019 [25] to create a database for storing retinal images and prediction results. The programming was conducted within the PyCharm 2023.1.2 software environment, involving Python scripting to implement the Django web framework for the Web API and to facilitate the invocation of the pretrained EfficientNetV2 model. The configuration and setup of the SSMS database were accomplished using structured query language (SQL).

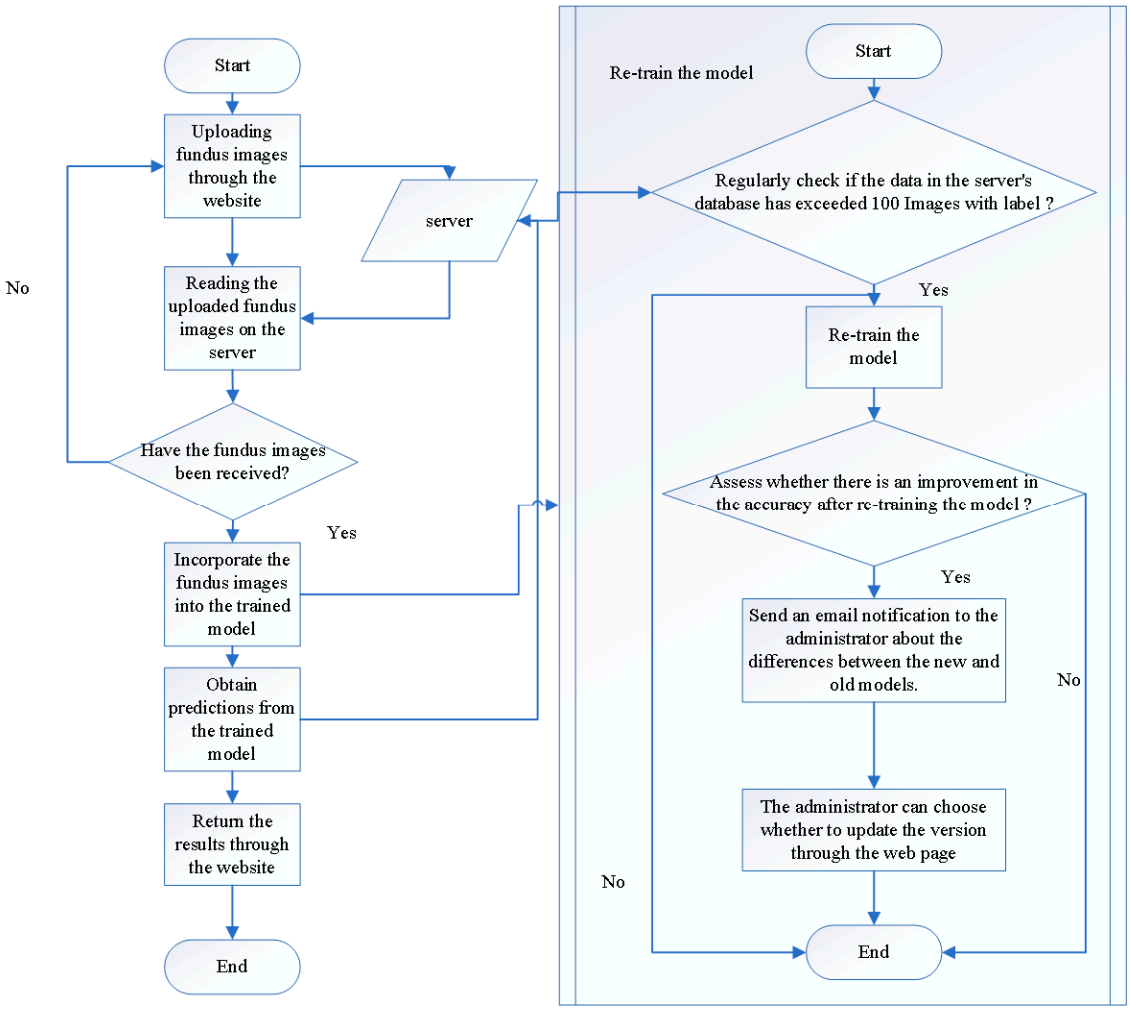


Figure 14. Workflow of the prediction system.

As depicted in Figure 14, users can upload retinal images to the server's database through the webpage. Subsequently, the website retrieves the uploaded retinal images from the server, applies the proposed preprocessing techniques for retinal images, and feeds them into the trained EfficientNetV2 model for classification. After obtaining the predicted outcomes of DR, these results are stored in the server's database and are also promptly displayed on the webpage.

Therefore, the system developed in this paper aims to provide a fast and accurate diagnosis of retinal lesions, enabling ophthalmologists to effectively diagnose retinal lesions in diabetic patients. To maintain the accuracy and performance of the system, new retinal image data will be collected and incorporated into the system on a regular basis. These data will be used to retrain the model, aiming to enhance the accuracy and performance of the system. When the accuracy of the new model surpasses that of the old model, the system will be automatically updated, allowing both doctors and users to benefit from the latest model. As outlined in the subroutine illustrated in Figure 14, the system undertakes regular assessments of the volume of data residing within the server's database, ensuring that it meets a minimum threshold of 100 entries. Prior to this step, each retinal image needs to be evaluated by a professional ophthalmologist to determine the accuracy of its classification. The physician can refer to the predictions made by the trained EfficientNetV2 model for this assessment. If the prediction is correct, no modifications are needed; however, if an error is detected, corrections would be applied to the classification. Once the database contains at least 100 entries, the system would proceed to adjust and retrain the previously trained EfficientNetV2 model. The accuracy of both the previous and new versions of the trained

EfficientNetV2 model will be evaluated, and the results will be sent to the administrator. The decision to update to the new version of the trained EfficientNetV2 model would be made by the administrator based on the assessment results.

Through this system, ophthalmologists can conveniently utilize automated tools to assist in their diagnostic work. This not only saves time but also provides accurate and reliable results, contributing to improved diagnostic efficiency and disease prevention.

Diabetic Prediction System Website

Through the web-based retinal image upload functionality, the process begins by converting the uploaded images into array format in the backend. These arrays are then transmitted to the server for the preprocessing steps proposed in this study and for the classification of DR levels. On the server side, a pre-established model is utilized to classify the retinal images and generate corresponding diagnostic results. These results are sent back to the user and displayed on the web page, as illustrated in Figure 15a. Concurrently, the diagnostic outcomes along with the retinal images are stored in the database, as depicted in Figure 15b. This approach ensures the association of diagnostic results with their corresponding retinal images, facilitating future queries and tracking analyses. The webpage in Figure 15a was built within the local network of Ming Chi University of Technology (MCUT). The establishment of this webpage was solely for the purpose of validating the functionality of the DR prediction system. Consequently, the website of this page cannot be accessed externally from the MCUT.

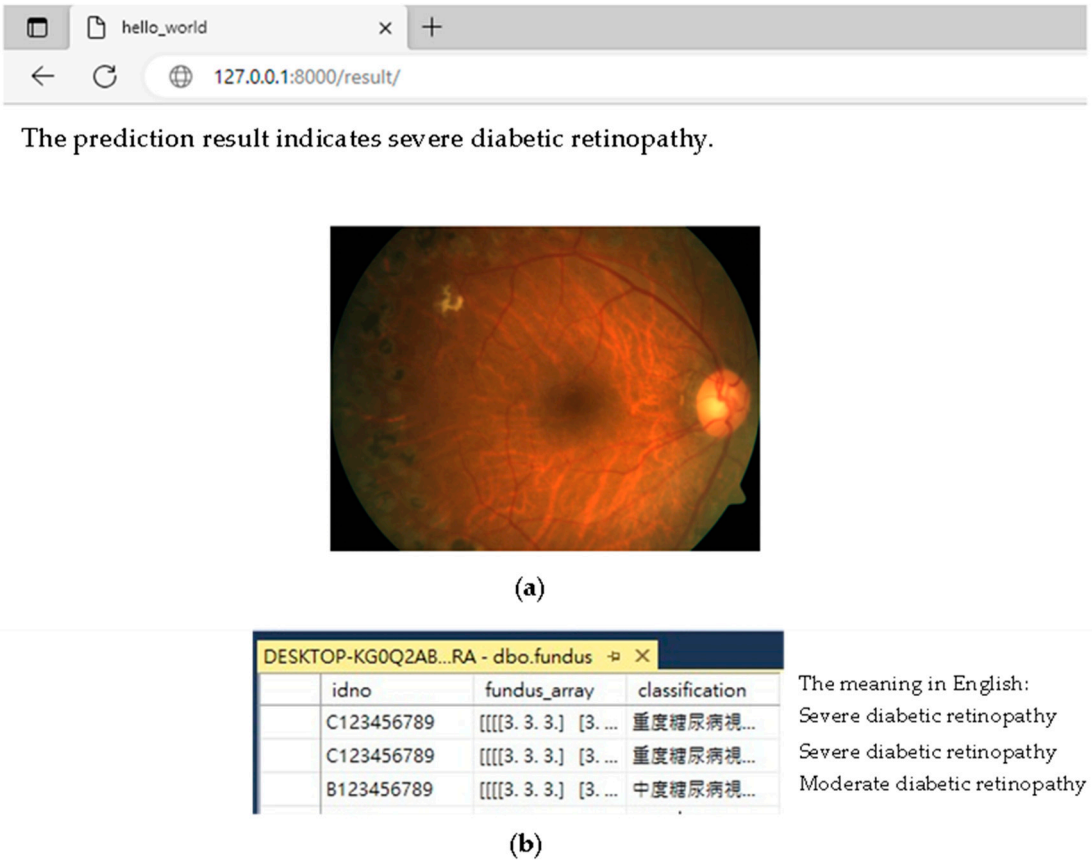


Figure 15. (**a**) Server-side result return. (**b**) Illustration of writing into the database.

Prior to retraining the model, this study enlisted the expertise of professional ophthalmologists to evaluate the accuracy of the model's predictive outcomes in their routine patient diagnostic procedures. In cases where discrepancies were identified between the model's predictions and actual classifications, the relevant data in the database were up-

dated to ensure the accuracy of the training dataset and mitigate potential adverse effects on subsequent training. Given the imbalanced distribution of disease severity levels within the existing training data, the initial focus was on collecting data for mild DR, severe DR, and proliferative DR to achieve dataset balance. Scheduled tasks within the Windows system were employed to periodically scan the database on the 1st day of each month. If the newly added data in the database reached a count of 100 entries, the system would initiate the process of retraining the model, as illustrated in Figure 16a. Following each round of model retraining, if the newly trained model outperformed the old model, the classification model on the server was updated. Simultaneously, the system sent an email notification to the administrator, as depicted in Figure 16b,c, indicating the completion of the update. Through this mechanism, this study ensured the gradual improvement of future model performance to better support the diagnostic tasks of ophthalmologists. Figure 16 illustrates the system setup within a Traditional Chinese system environment. Consequently, the diagram includes English annotations tailored for the respective Chinese contexts.

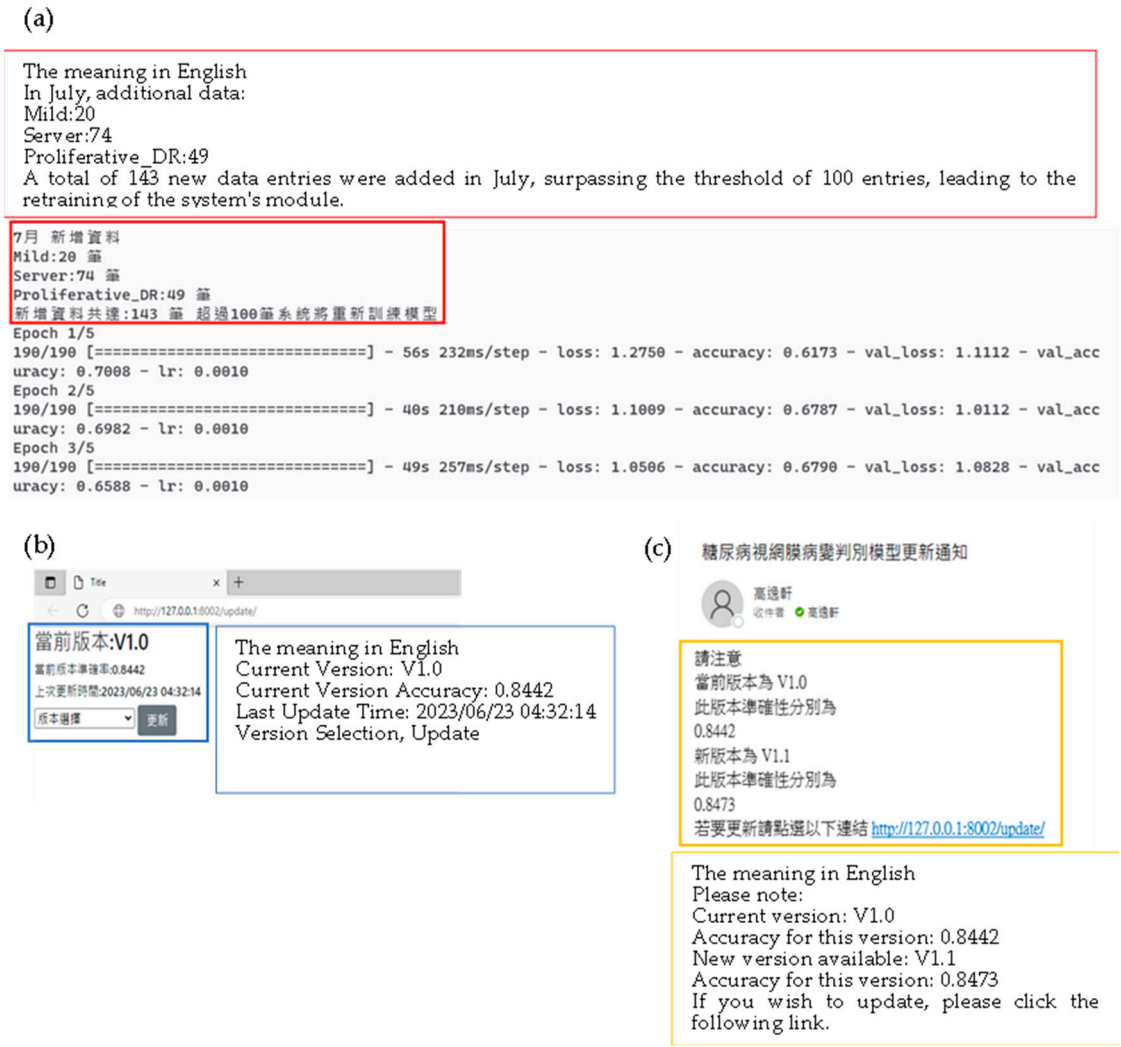


Figure 16. (**a**) Diagram depicting automated model training and system update process. (**b**) Interface for user to select model version in the system update diagram. (**c**) Notification email for new version update.

In order to validate the functionality of the system and confirm whether adding data to the minority classes enhances model accuracy, this study incorporated the IDRiD (Indian DR Image Dataset) dataset provided by IEEE. The IDRiD dataset consists of 20 instances of

mild DR, 74 instances of severe DR, and 49 instances of proliferative DR. After integrating the IDRiD dataset [26] and meeting the conditions for model updates and retraining, the model's accuracy improved to 0.8473 compared to the original 0.8442. This enhancement demonstrates that even with a small amount of new data added to the minority classes, the data imbalance can be alleviated, leading to a notable improvement in the model's performance. Figure 17 illustrates the confusion matrix of the new model after retraining.

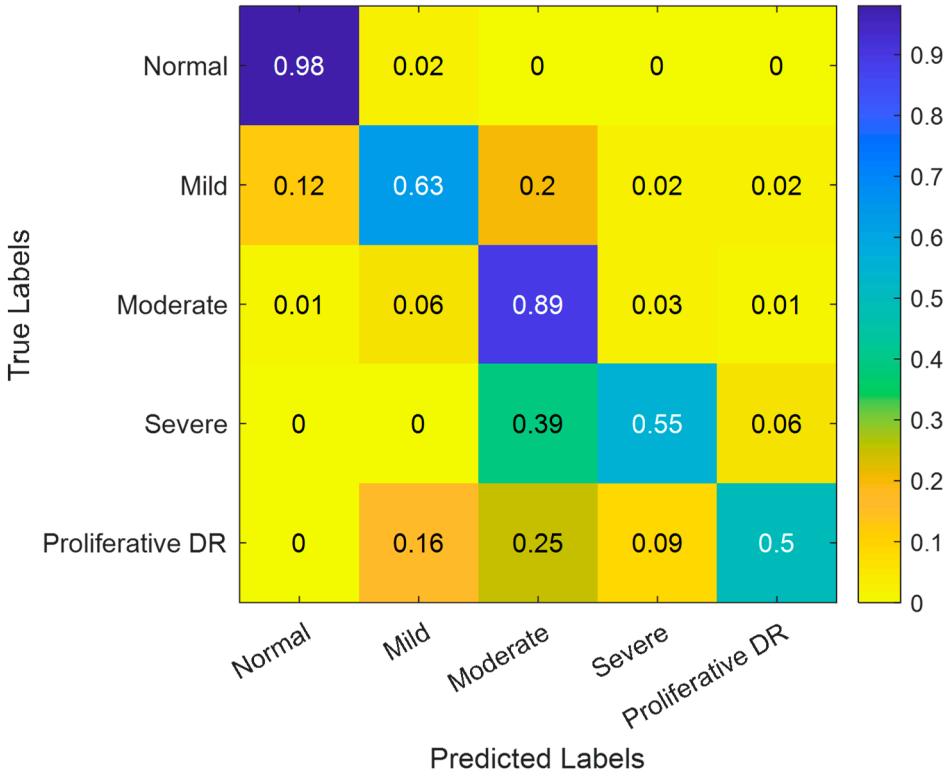


Figure 17. The confusion matrix of the new model after retraining when the data count reaches 100 or more entries.

5. Discussion

5.1. Discussion on Findings

The results of our study demonstrate that the integration of pixel color amplification and EfficientNetV2 significantly improves the performance of DR (DR) detection models. Our approach addressed several key challenges.

1. Enhanced Image Quality: The use of Dark Channel Prior (DCP) for pixel color amplification effectively mitigated issues of underexposure and overexposure, leading to clearer and more consistent retinal images. This enhancement is crucial for accurate diagnosis, as it highlights subtle pathological features that are often missed in standard images;
2. Improved Model Performance: The optimized EfficientNetV2 model outperformed other models such as EfficientNetB5, VGG16, NASNet, and ResNet50 in terms of accuracy and training efficiency. The combination of transfer learning, fine-tuning, and advanced techniques like early stopping, adaptive learning rates, and label smoothing contributed to this improvement;
3. Robustness and Generalizability: Our comprehensive preprocessing workflow, including precise image cropping and extensive data augmentation, ensured that the model generalized well across different imaging conditions and patient populations. This robustness is essential for real-world clinical applications where image quality and patient demographics can vary widely.

5.2. Comparison with Previous Studies

Our approach demonstrated superior performance compared to previous studies. For instance, the modified Xception architecture by S. H. Kassani et al. achieved an accuracy of 83.09% for DR classification, while our EfficientNetV2 model achieved an accuracy of 84.42% [5]. Additionally, our model's binary classification accuracy of 99.4% is significantly higher than the results obtained by J. D. Bodapati et al. using various deep CNN models [4]. These comparisons highlight the effectiveness of our pixel color amplification technique and the EfficientNetV2 architecture in improving DR detection accuracy.

5.3. Limitations and Challenges

Despite the promising results, several limitations remain:

1. Imbalanced Dataset: The imbalanced distribution of DR severity levels in the training dataset may impact the model's ability to accurately classify less-common stages of the disease. Future work should focus on collecting more balanced datasets to address this issue;
2. Generalization to Different Populations: While our model generalized well within the dataset used, its performance on images from different populations or imaging devices remains to be validated;
3. Model Interpretability: Although Grad-CAM provided insights into the model's decision-making process, further work is needed to enhance the interpretability and explainability of the model to ensure its acceptance in clinical practice;
4. Scalability and Real-Time Processing: The proposed method may face challenges in scalability and real-time processing, which are crucial for practical deployment in clinical settings [27];
5. Incomplete Circular Structures in Images: Some images in the dataset do not have a full circular structure, which could potentially impact training efficiency and accuracy. Although our preprocessing techniques, such as cropping and data augmentation, help mitigate this variability, a more detailed analysis is needed to fully understand and address this issue.

5.4. Future Work

Future research directions include the following:

1. Hyperparameter Optimization: Exploring automated techniques like grid search or evolutionary algorithms for hyperparameter tuning to further enhance model performance;
2. Data Augmentation and Collection: Collaborating with medical institutions to gather additional data, particularly for underrepresented DR stages, and implementing advanced data augmentation techniques;
3. Cloud Deployment: Developing cloud-based systems for broader data collection and real-time DR detection, which could facilitate continuous model improvement and provide valuable resources for researchers and practitioners [28];
4. Integration with Clinical Workflows: Testing and integrating the predictive system within clinical workflows to assess its practical utility and gather feedback from ophthalmologists for further refinements [29];
5. Incomplete Circular Structures in Images: Some images in the dataset do not have a full circular structure, which could potentially impact training efficiency and accuracy. Although our preprocessing techniques, such as cropping and data augmentation, help mitigate this variability, a more detailed analysis is needed to fully understand and address this issue.

6. Conclusions

6.1. Summary of Contributions

Our study presents a novel approach to DR detection by integrating pixel color amplification with the EfficientNetV2 model. The key contributions include the following:

1. Innovative Preprocessing Techniques: The use of Dark Channel Prior (DCP) for pixel color amplification significantly improved retinal image quality, addressing common issues of underexposure and overexposure;
2. Optimized Deep Learning Model: The EfficientNetV2 model, enhanced through transfer learning, fine-tuning, and advanced optimization techniques, demonstrated superior accuracy and efficiency in DR detection;
3. Robust Predictive System: The development of a comprehensive, web-based predictive system enables real-time DR detection, providing a practical tool for ophthalmologists.

6.2. Potential Applications

The proposed system holds significant potential for clinical application:

1. Early Disease Detection: The high accuracy and efficiency of the model can aid in early detection of DR, preventing severe outcomes like blindness and improving patient outcomes;
2. Clinical Decision Support: The web-based system provides ophthalmologists with a reliable tool for preliminary diagnosis, reducing their workload and enhancing diagnostic efficiency;
3. Continuous Improvement: The automatic update feature ensures that the model evolves with new data, maintaining high accuracy and adapting to changing clinical needs.

In conclusion, our approach represents a significant advancement in the field of DR detection, offering a highly accurate, efficient, and practical solution for early disease identification. Future work will focus on addressing the identified limitations and further refining the system for broader clinical adoption.

Author Contributions: Conceptualization, C.-L.L. and Y.-H.K.; methodology, C.-L.L. and Y.-H.K.; software, Y.-H.K. validation, C.-L.L. and Y.-H.K.; formal analysis, Y.-H.K.; data curation, Y.-H.K.; writing—original draft preparation, C.-L.L. and Huang; writing—review and editing, C.-L.L.; supervision, C.-L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This study utilized the APTOS 2019 Blindness Detection dataset from the Kaggle online platform, made accessible for advancing ophthalmology and blindness detection research. No direct interaction with human subjects or animals occurred during this study, as the dataset was anonymized and de-identified in compliance with ethical standards. The dataset's purpose is to advance DR diagnosis research, and ethical considerations regarding human subjects and data privacy were addressed by the dataset creators and Kaggle. This study strictly adhered to ethical usage terms set by Kaggle and APTOS organizers, acknowledging their contribution to ophthalmology research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Teo, Z.L.; Tham, Y.C.; Yu, M.; Chee, M.L.; Rim, T.H.; Cheung, N.; Bikbov, M.M.; Wang, Y.X.; Tang, Y.; Lu, Y.; et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology* **2012**, *128*, 1580–1591. [CrossRef] [PubMed]
2. Grzybowski, A.; Brona, P.; Lim, G.; Ruamviboonsuk, P.; Tan, G.S.W.; Abramoff, M.; Ting, D.S.W. Artificial intelligence for diabetic retinopathy screening: A review. *Eye* **2020**, *34*, 451–460. [CrossRef] [PubMed]
3. Sheng, B.; Chen, X.; Li, T.; Ma, T.; Yang, Y.; Bi, L.; Zhang, X. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. *Front. Public Health* **2022**, *10*, 971943. [CrossRef] [PubMed]
4. Bodapati, J.D.; Naralasetti, V.; Shareef, S.N.; Hakak, S.; Bilal, M.; Maddikunta, P.K.R.; Jo, O. Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. *Electronics* **2020**, *9*, 914. [CrossRef]

5. Kassani, S.H.; Kassani, P.H.; Khazaeinezhad, R.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Diabetic retinopathy classification using a modified xception architecture. In Proceedings of the 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, UAE, 10–12 December 2019; pp. 1–6.
6. Lin, C.L.; Jiang, Z.X. Development of preprocessing methods and revised EfficientNet for diabetic retinopathy detection. *Int. J. Imaging Syst. Technol.* **2023**, *33*, 1450–1466. [CrossRef]
7. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
8. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
9. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
10. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
11. Kao, Y.-H.; Lin, C.-L. Detection of Diabetic Retinopathy via Pixel Color Amplification Using EfficientNetV2. In Proceedings of the 2023 9th International Conference on Applied System Innovation (ICASI), Chiba, Japan, 21–25 April 2023; pp. 148–150.
12. Gaudio, A.; Smailagic, A.; Campilho, A. Enhancement of retinal fundus images via pixel color amplification. In *Image Analysis and Recognition: 17th International Conference, ICIAR 2020, Póvoa de Varzim, Portugal, 24–26 June 2020, Proceedings, Part II*; Springer: Cham, Switzerland, 2020; pp. 299–312.
13. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [PubMed]
14. Jiang, L. A fast and accurate circle detection algorithm based on random sampling. *Future Gener. Comput. Syst.* **2021**, *123*, 245–256. [CrossRef]
15. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML) 2019, Long Beach, 9–15 June 2019; pp. 6105–6114.
16. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International Conference on Machine Learning (ICML), Virtual, 18–24 July 2021; pp. 10096–10106.
17. Lin, C.-L.; Wu, K.-C. Development of revised ResNet-50 for diabetic retinopathy detection. *BMC Bioinform.* **2023**, *24*, 157. [CrossRef] [PubMed]
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
19. Chen, M.X.; Firat, O.; Bapna, A.; Johnson, M.; Macherey, W.; Foster, G.; Jones, L.; Parmar, N.; Schuster, M.; Chen, Z.; et al. The best of both worlds: Combining recent advances in neural machine translation. *arXiv* **2018**, arXiv:1804.09849.
20. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [CrossRef]
21. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
22. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
23. Vujović, Ž. Classification model evaluation metrics. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 599–606. [CrossRef]
24. Vincent, W.S. Django for APIs: Build Web APIs with Python and Django; WelcomeToCode: 2022. Available online: https://www.amazon.com/Django-APIs-Build-web-Python/dp/1093633948 (accessed on 21 May 2024).
25. Dewson, R. Sql server management studio. In *Beginning SQL Server 2008 for Developers: From Novice to Professional*; Apress: New York, NY, USA, 2008; pp. 25–50.
26. Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Meriaudeau, F. Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research. *Data* **2018**, *3*, 25. [CrossRef]
27. Faber, K.; Corizzo, R.; Sniezynski, B.; Japkowicz, N. Lifelong Continual Learning for Anomaly Detection: New Challenges, Perspectives, and Insights. *IEEE Access* **2024**, *12*, 41364–41380. [CrossRef]
28. Ding, A.; Qin, Y.; Wang, B.; Cheng, X.; Jia, L. An elastic expandable fault diagnosis method of three-phase motors using continual learning for class-added sample accumulations. *IEEE Trans. Ind. Electron.* **2023**, *71*, 7896–7905. [CrossRef]
29. Shao, L.; Zhu, F.; Li, X. Transfer learning for visual categorization: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *26*, 1019–1034. [CrossRef] [PubMed]