Optimizing Deep Learning for Diabetic Retinopathy Diagnosis

Krit Sriporn[1], Cheng-Fa Tsai[2]*, Li-Jia Rong[3], Paohsi Wang[4], Tso-Yen Tsai[5], Chih-Wen Chen[6]
Department of Digital Technology, Suratthani Rajabhat University, Surat Thani, Thailand[1]
Department of Management Information Systems, National Pingtung University of Science and Technology, Pingtung, Taiwan[2, 3]
Department of Food and Beverage Management, Cheng Shiu University, Kaohsiung, Taiwan[4]
Chunri Township Public Health Center-Public Health Bureau, Pingtung County Government, Pingtung, Taiwan[5]
Department of Emergency Medicine, Pingtung Christian Hospital, Pingtung, Taiwan[6]

***Abstract*—The detection of diabetic retinopathy traditionally requires the expertise of medical professionals, making manual detection both time- and labor-intensive. To address these challenges, numerous studies in recent years have proposed automatic detection methods for diabetic retinopathy. This research focuses on applying deep learning and image processing techniques to overcome the issue of performance degradation in classification models caused by imbalanced diabetic retinopathy datasets. It presents an efficient deep learning model aimed at assisting clinicians and medical teams in diagnosing diabetic retinopathy more effectively. In this study, image processing techniques, including image enhancement, brightness correction, and contrast adjustment, are employed as preprocessing steps for fundus images of diabetic retinopathy. A fusion technique combining color space conversion, contrast limited adaptive histogram equalization, multi-scale retinex with color restoration, and Gamma correction is applied to highlight retinal pathological features. Deep learning models such as ResNet50-V2, DenseNet121, Inception-V3, Xception, MobileNet-V2, and InceptionResNet-V2 were trained on the preprocessed datasets. For the APTOS-2019 dataset, DenseNet121 achieved the highest accuracy at 99% for detecting diabetic retinopathy. On the Messidor-2 dataset, InceptionResNet-V2 demonstrated the best performance, with an accuracy of 96%. The overall aim of this research is to develop a computer-aided diagnosis system for classifying diabetic retinopathy.**

Keywords*—Diabetic retinopathy; deep learning; image processing technologies; imbalanced image dataset; computer aided diagnosis

I. Introduction

Diabetes is a chronic disease and a major public health issue that significantly affects quality of life due to its rising incidence. According to the International Diabetes Federation's Diabetes Atlas, 537 million people worldwide have diabetes, and this number is expected to increase to 629 million by 2045. Most cases occur in low-to-middle-income countries, with more than half undiagnosed. The World Health Organization predicts that diabetes will become the seventh leading cause of death globally.

A common and often unnoticed complication of diabetes is diabetic retinopathy, which progresses slowly but can severely impact patients' quality of life, affecting their families, the economy, and society. The condition necessitates comprehensive care and regular screening. Diabetic retinopathy involves changes to the retina caused by diabetes, as elevated blood sugar levels lead to nerve and blood vessel damage. This results in complications such as retinal swelling, detachment, bleeding, and vision loss [1].

Many countries face shortages in medical resources and ophthalmologists, particularly in rural or remote areas, where patients may not receive timely diagnosis and treatment due to the uneven distribution of healthcare resources. Traditional manual diagnosis by specialists also presents challenges, as it can be time-consuming and early-stage symptoms are often subtle, increasing the risk of misdiagnosis even by a single expert. As a result, monitoring and treating diabetic retinopathy demands considerable time and human resources, while training professional ophthalmologists entails significant costs.

Given these challenges, recent years have seen numerous studies [2–4] proposing automated detection methods to assist medical teams and ophthalmologists in diagnosing diabetic retinopathy more efficiently. Automated diagnostic systems also help address the issue of low screening efficiency in areas with limited medical resources.

Research on automating diabetic retinopathy detection has focused on areas such as medical image enhancement, machine learning, and deep learning-based image recognition. Scholars have reviewed these studies, discussing the advantages and disadvantages of various methods and the factors that may influence recognition results.

The quality of fundus images for diabetic retinopathy is often inconsistent, making image preprocessing a crucial step in many studies. Different research efforts employ appropriate preprocessing techniques based on their specific objectives [5].

This study aims to develop a computer-aided diagnosis system for detecting diabetic retinopathy by addressing the imbalance in data characteristics within the diabetic retinopathy dataset and using categorical focal loss as the loss function instead of categorical cross-entropy loss. The research utilizes the Asia Pacific Tele-Ophthalmology Society-2019 Blindness Detection (APTOS-2019) and Messidor-2 datasets, containing 3,662 and 1,744 images respectively, to train and compare the performance of various convolutional neural network models in diabetic retinopathy classification against other studies. Six convolutional neural network (CNN) models ResNet50-V2, DenseNet121, Inception-V3, Xception, MobileNet-V2, and

InceptionResNet-V2 were used to determine the most suitable model for training with these datasets.

The remaining content research is organized as follows: materials and methodology are presented in Section II. implementation details are shown in Section III. Section IV presents the experimental results, while Section V presents the discussion, and finally, Section VI concluded the paper.

II. Materials and Methodology

This study divides the experimental procedures into four sections: image collection, image preprocessing, deep learning model training, and system performance evaluation. The following subsections detail the processes and methods involved in each step.

The first step is to confirm the dataset accessible under the experimental conditions. Due to concerns regarding medical privacy, diabetic retinopathy datasets often require annotation and grading by ophthalmic experts. To address these issues, a practical approach is to use existing publicly available datasets for experimentation. Public datasets offer several advantages: many previous studies have utilized them to evaluate their systems. By using the same publicly available datasets, researchers can assess the novelty of their approach and compare it with prior work. This also enhances the comparability and credibility of the current study.

The APTOS-2019 and Messidor-2 datasets are foundational resources in artificial intelligence research for the detection and classification of diabetic retinopathy. These datasets contain extensive collections of eye images, each annotated with a specific severity level of diabetic retinopathy, making them instrumental for training and evaluating computer vision models aimed at accurate disease diagnosis.

This research applies a variety of image processing techniques and deep learning models to train a classifier on the diabetic retinopathy dataset, as described in [5]. The classifier is responsible for categorizing images into five distinct stages of diabetic retinopathy: No Diabetic Retinopathy (NDR), Mild Diabetic Retinopathy (MiDR), Moderate Diabetic Retinopathy (MoDR), Severe Diabetic Retinopathy (SDR), and Proliferative Diabetic Retinopathy (PDR). Visual representations of these five stages are provided in Fig. 1.

In addition, the diabetic retinopathy datasets face issues with class imbalance. This imbalance can negatively affect system performance, as certain classes in the dataset may be overrepresented or underrepresented. As a result, the model tends to predict the more frequent classes more easily, while the less frequent classes are harder to predict and are more likely to be misclassified as one of the dominant classes.

In the second step, various image processing techniques were employed, including RGB color space conversion, commission internationale de l'eclairage (CIELAB), contrast-limited adaptive histogram equalization (CLAHE), multi-scale retinex with color restoration (MSRCR), and Gamma correction. These techniques enhance image features to improve the neural network's ability to learn features effectively. Following this, the dataset was split into training, validation, and testing sets, with 80% allocated to training and 20% for testing, as shown in Table I.

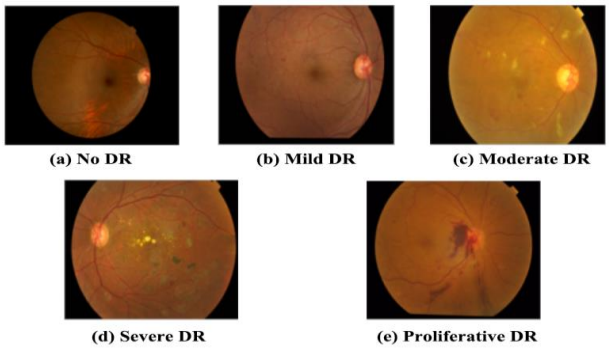

Fig. 1. Classification of the five levels of diabetic retinopathy.

TABLE I. Distribution of Image Quantities Across Disease Categories

Dataset Name	categories					Total
	NDR	MiDR	MoDR	SDR	PDR	
APTOS-2019	1,805	370	999	193	295	3,662
Messidor-2	1,020	264	347	77	36	1,744

After splitting the dataset, the training set undergoes data augmentation techniques, including horizontal flipping, vertical flipping, and pixel value scaling. Before being fed into the convolutional neural network (CNN) models for training, each image is resized to the optimal dimensions required by the model.

Fig. 2 presents a schematic representation of the hybrid image enhancement method used in this investigation, based on findings from previous studies [6–11]. Our proposed technique integrates the CLAHE algorithm with enhancements to the CIELAB color space. Empirical results show that this combined approach produces better results compared to using the hue, saturation, and value (HSV) color space or other color space transformations. In this framework, CLAHE is applied exclusively to the luminance (L) channel of the CIELAB color space. The process of splitting the V channel refers to isolating the Value component from the HSV color space, which allows for further processing or enhancement tasks such as sharpening or brightening.

In medical imaging, particularly in retinal imaging, researchers often convert images to the CIELAB color space to separate luminance (L channel) from color information (a and b channels). Subsequently, CLAHE is applied to enhance the luminance component in CIELAB, preserving the original color information. This approach effectively enhances image details without compromising color fidelity. The synergistic combination of CIELAB and CLAHE provides a robust solution for improving image clarity and quality, particularly when processing high-detail images.

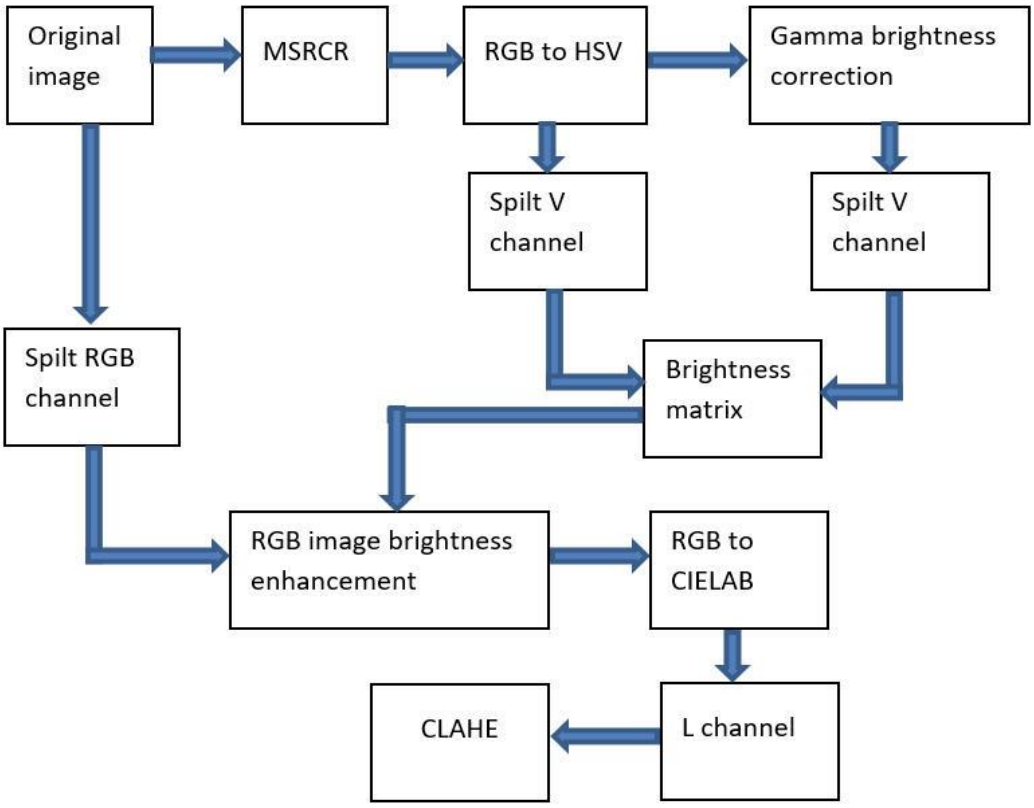

Fig. 2. The schematic of hybrid image enhancement method.

Fig. 3 presents a schematic of alternative image enhancement methods, which are standard preprocessing techniques used in diabetic retinopathy severity grading. The multi-scale retinex with color restoration (MSRCR) transformation is a pivotal aspect of the hybrid image enhancement strategy adopted in this research. MSRCR has been extensively applied in medical image segmentation and classification, including retinal vessel segmentation and arterial-venous classification, as evidenced by studies [12–15].

Data augmentation is a technique that enhances the amount of data by synthesizing new images from an original dataset. This method not only increases the dataset's size but also offers additional advantages, such as more efficient use of computer memory during model training and a regularization effect that helps mitigate overfitting. Common techniques for data augmentation include horizontal and vertical flipping, resizing, cropping, shifting, and scaling pixel values. While these transformations create variations of the original image that may look different to the model, they still allow humans to recognize them as the same image. This principle of generating multiple images from a single original image is the foundation of data augmentation [16].

In the third step, this study employs deep learning models, including ResNet50-V2 and InceptionResNet-V2, which were trained on diabetic retinopathy image datasets processed in the second step. Various CNN models were combined with image enhancement methods to identify the optimal combination and compare the advantages of each. The architecture of the models used in this study is illustrated in Fig. 4. All pretrained deep learning models mentioned serve as base models in this research, utilizing pretrained weights from the ImageNet image database. A custom dense neural network is connected beneath the base model, with each final layer of the CNNs configured to use a SoftMax activation function for five-class classification.

To ensure that model training achieves the expected classification performance, continuous testing and parameter tuning are conducted. Ultimately, other public datasets are used to validate the system's performance and the model's reliability. The evaluation metrics used in this study are described below, sourced from studies related to diabetic retinopathy severity classification [5, 17, 18]. Table II presents the definitions of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

TABLE II. MEANINGS OF TP, TN, FP, FN

Indicator Names:	Descriptions:
True Positive (TP)	True Positive (TP): Predicted positive and predicted correctly.
True Negative (TN)	True Negative (TN): Predicted negative and predicted correctly.
False Positive (FP)	False Positive (FP): Predicted positive but predicted incorrectly.
False Negative (FN)	False Negative (FN): Predicted negative but predicted incorrectly.

Accuracy is used to calculate the proportion of correctly predicted samples by the model among the total number of samples in a classification task. It is calculated as shown in Eq. (1).

$$\text{Accuracy} = （TP+TN）/\text{Total} \quad (1)$$

Precision is the proportion of correctly identified samples of a particular class among all samples predicted to belong to that class by the model. It is calculated as shown in Eq. (2).

$$\text{Precision} = TP/（TP+FP） \quad (2)$$

Sensitivity, also known as the True Positive Rate or Recall, refers to the proportion of true positive samples that are correctly predicted by the model among all actual positive samples. It is calculated as shown in Eq. (3).

$$\text{Sensitivity} = TP/（TP+FN） \quad (3)$$

Specificity, also known as the True Negative Rate, refers to the proportion of true negative samples that are correctly predicted by the model among all actual negative samples. It is calculated as shown in Eq. (4).

$$\text{Specificity} = TN/（FP+TN） \quad (4)$$

Additionally, the diabetic retinopathy datasets exhibit issues with class imbalance. This problem can adversely affect system performance, as certain classes in the dataset may be overrepresented while others are underrepresented. As a result, the more frequent classes are easier to predict, whereas the less frequent classes are harder to predict and more likely to be misclassified by the model as one of the more frequent classes.

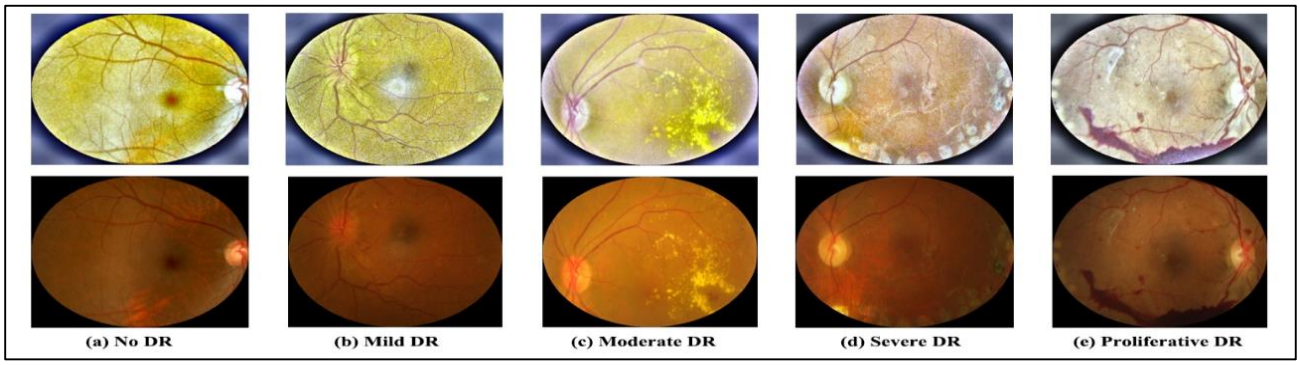

Fig. 3. The results of image enhancement methods.

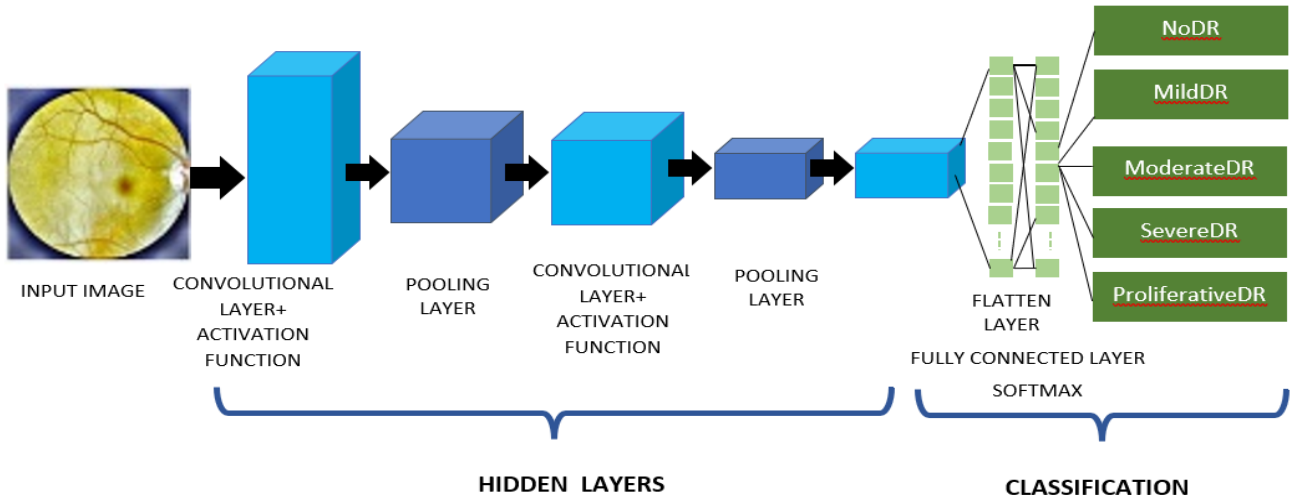

Fig. 4. Deep learning model.

III. Implementation Details

A. Experimental Environment

The configuration of the experimental environment in this paper is shown in Table III. The computer hardware used includes an Intel® Core™ i9-9900K CPU 3.60GHz as the central processing unit, an NVIDIA GeForce RTX 2080 Ti 11GB as the graphics processing unit, and 64GB of memory. In terms of software configuration, the operating system is Windows 10 Pro 64-bit. The platform used for programming is Jupyter Notebook 6.4.6, and the programming language used is Python 3.7.11. The deep learning frameworks utilized are TensorFlow-GPU 2.8.0 and Keras 2.8.0.

B. Model Parameter Settings

Each model was trained for a maximum of 100 epochs, with a callback function set to monitor the validation loss for early stopping, saving the model with the best training performance. In addition to early stopping, this study employed ModelCheckpoint to track and save the model weights with the highest validation specificity.

The batch size for all training processes was set to 8. Nadam [16], an adaptive learning rate optimizer, was employed to enhance training stability by adjusting the learning rate based on its variance. The class weight parameters were adjusted according to the proportion of classes from class 0 to class 4 to

balance the training weights for each class. Classes with more examples were assigned lower weights, while classes with fewer examples were assigned higher weights. Instead of using the traditional categorical cross-entropy loss, this study employed categorical focal loss as the loss function.

TABLE III. HARDWARE AND SOFTWARE ENVIRONMENT OF THE EXPERIMENT

experimental environment configuration	specifications
operating system	Windows 10 Professional 64-bit
processor	Intel (R) Core (TM) i9-9900K CPU 3.60GHz
graphic processor	NVIDIA GeForce RTX 2080 Ti 11G
memory	64GB
software development platform	Jupyter Notebook 6.4.6
programming language	Python 3.7.11
deep learning framework	Tensorflow-gpu 2.8.0、Keras 2.8.0

ResNet50-V2 is an improved version of ResNet50 [19], which outperforms both ResNet101 and ResNet50 in terms of system performance on the large-scale ImageNet database. The primary enhancement in ResNet50-V2 involves modifying the propagation formula within the basic building blocks of the Residual Network (ResNet). The architectural differences between the basic units of ResNet50-V1 and ResNet50-V2 are shown in Fig. 5 (with ResNet50-V1 on the left and ResNet50-V2 on the right). In ResNet50-V2, the batch normalization layer and ReLU activation function are moved before the weight layers, resulting in a pre-activation structure. The advantage of this structure is that it accelerates model convergence without changing the model's depth, and placing the batch normalization layer before the ReLU activation function and weight layers enhances the model's regularization effect.

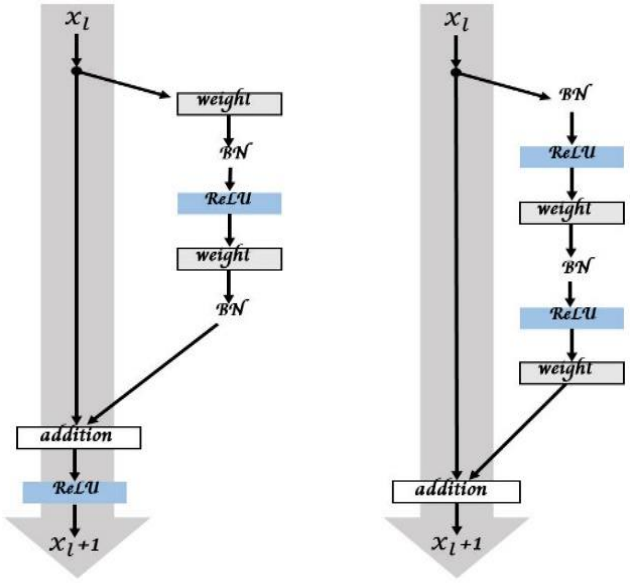

Fig. 5. Architectural differences between basic residual units of ResNet50-V1 and ResNet50-V2.

DenseNet (Densely Connected Convolutional Network) offers several advantages over ResNet, such as reducing network complexity, lowering the number of model parameters, mitigating the gradient vanishing problem, and enhancing feature propagation [20]. Due to its densely connected architecture, DenseNet efficiently improves feature utilization, reduces gradient vanishing, and aids model convergence. Reusing features means there is no need to learn new feature maps, leading to a reduction in the number of model parameters. Fig. 6 illustrates the architecture of a dense block in DenseNet.

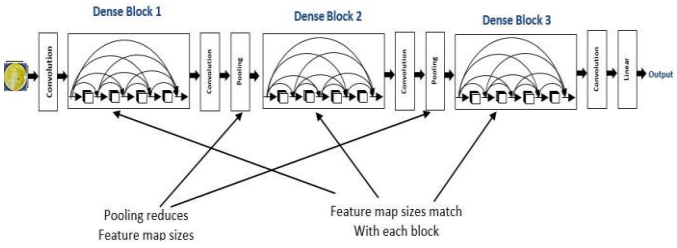

Fig. 6. Schematic diagram of dense block.

Inception-V3 was released in December 2015. The principle of Inception involves connecting convolutional layers of different sizes in parallel. The output from these convolutional layers is then used as input for the next inception block utilizing convolutional layers of varying sizes allows for the creation of better feature maps by extracting diverse feature information. Inception avoids bottleneck issues, which, although reducing the model's parameters, can result in the loss of important features and require slow dimensionality reduction to prevent excessive data loss [21]. Fig. 7 shows a schematic diagram of the inception block, including convolutional layers of different sizes.

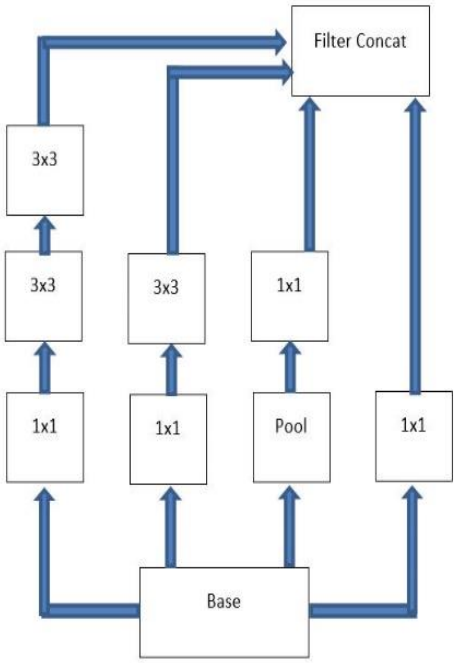

Fig. 7. Schematic diagram of inception block.

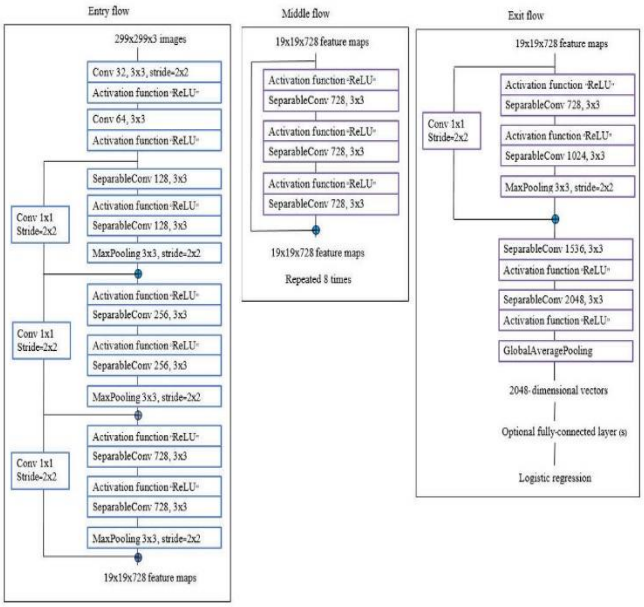

Fig. 8. Xception network architecture diagram.

Xception, short for extreme inception, is a network based on Inception-V3 [22]. It employs a depthwise separable Convolution architecture to replace the inception module structure in the original Inception-V3. depthwise separable convolution maintains the number of parameters of the original Inception-V3 network while reducing model complexity, increasing network width, and improving model accuracy, as illustrated in Fig. 8.

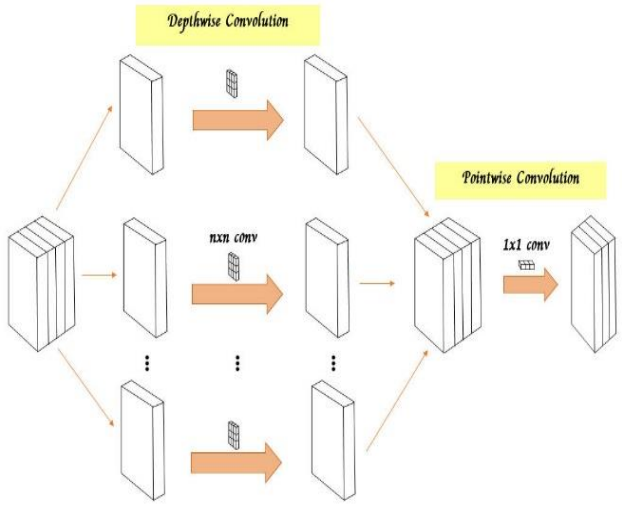

Fig. 9. Convolution operation flowchart of depthwise convolution.

MobileNet is a lightweight CNN model proposed by Google, designed to balance model size and computational speed for use in devices or mobile platforms [23]. MobileNet utilizes a depthwise separable convolution structure, which significantly reduces computational resources while maintaining good accuracy. Although both MobileNet and Xception use depthwise separable convolution, their goals differ: MobileNet focuses on model compression and computational speed, while Xception aims to enhance system performance with a parameter count similar to Inception-V3. The depthwise separable convolution consists of two parts: depthwise convolution and pointwise convolution. Depthwise convolution applies a k×k convolutional layer separately to each input channel, followed by pointwise convolution, which multiplies the output from the previous step using a 1×1 convolutional layer. The combined results provide a feature map, as illustrated in Fig. 9.

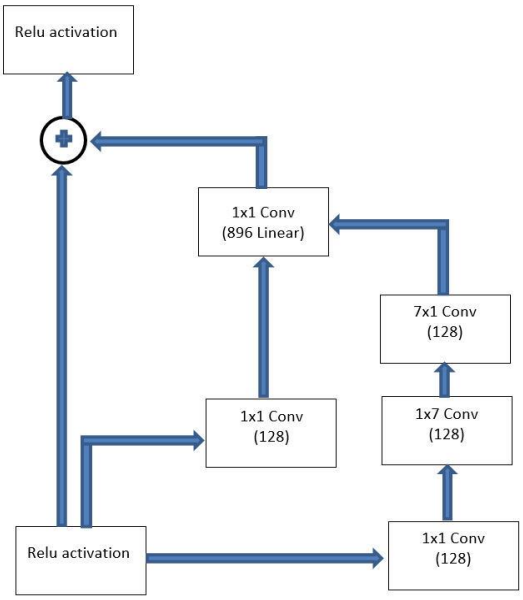

Fig. 10. Architecture diagram of inception block combined with residual connection.

InceptionResNet is a model developed by combining Inception-V3 and ResNet architectures [24]. The developers integrated residual connections into the inception architecture. Experimental results demonstrate that Inception networks with residual connections perform better in terms of system efficiency compared to inception networks without them. Additionally, residual connections significantly accelerate the training speed of the inception network. Fig. 10 illustrates the architectural diagram of an inception block integrated with a residual connection.

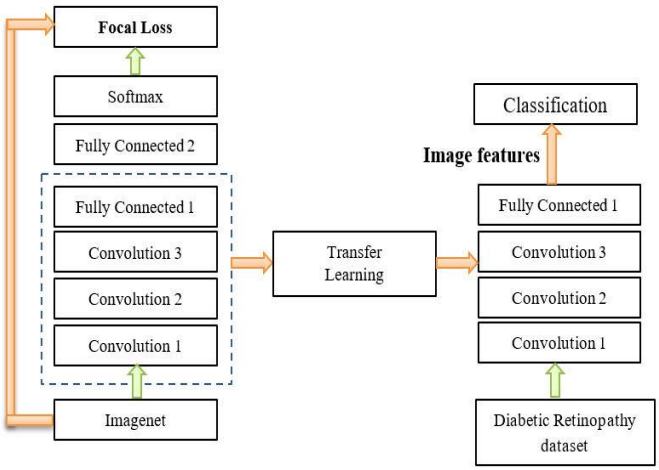

Fig. 11. Transfer learning technique.

Transfer learning involves training a neural network model on a related task and then adapting it to a new, similar problem. This technique leverages the model's pre-trained weights, often learned from large datasets like ImageNet. Transfer learning offers flexibility, enabling the use of pre-trained models as feature extractors or components of entirely new models. It has also been successfully applied in cancer subtype discovery, as illustrated in Fig. 11 [25].

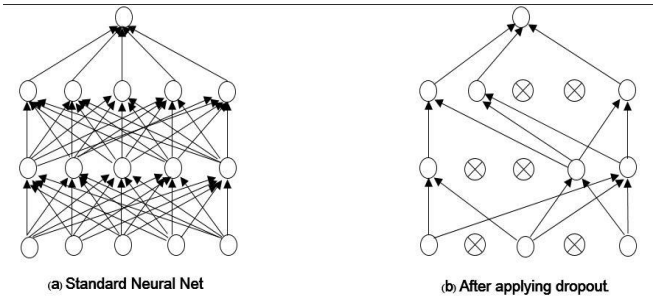

Fig. 12. Dropout technique.

To address overfitting, dropout is employed as a simple regularization technique that randomly deactivates neurons during training. In the context of CNNs, this means that some neurons are temporarily ignored, preventing them from sending signals to other neurons. A dropout rate of 0.5 in fully connected layers indicates that 50% of the neurons are deactivated. Dropout effectively mitigates overfitting by disrupting co-adaptations among neurons, thereby enhancing the model's ability to generalize and reducing its tendency to overfit the training data [25].

C. Categorical Focal Loss

The term α in the focal loss formula serves as a balancing factor that addresses class imbalance. It is calculated as shown in Eq. (5).

$$FL(p_t) = -\alpha(1-p_t)^{\gamma}.log(p_t) \quad (5)$$

Focal loss focuses on difficult examples by adjusting the loss based on the probability of correct classification (p_t). Hard examples, which have low p_t values, exert minimal influence on the modulating factor $(1-p_t)\ \gamma$, while easy examples with high pt values have a diminishing effect [26].

D. Optimizer

Nadam [23] is an optimization algorithm that utilizes past information to efficiently update model weights. It combines the strengths of both Adam and Nesterov momentum, addressing the decaying learning rate issue of Adagrad. This combination leads to faster convergence and reduced parameter instability, enabling more effective model training. Nadam is calculated as shown in Eq. (6).

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{V_t} + \varepsilon}\left(\beta_1 m_t + \frac{(1-\beta_1)g_t}{1-\beta_1^t}\right) \quad (6)$$

Nadam utilizes a learning rate of 0.002 (η) with an objective function (θ_t) defined by $\varepsilon = 1e^{-08}$ and $\beta_1 = 0.9$. These settings, inspired by [27], leverage v_t, g_t and m_t to enhance the optimizer's efficiency at each time step t.

IV. Experimental Results

Tables IV and V present the system evaluation results of the image enhancement fusion techniques and six deep learning model combinations on the APTOS-2019 and Messidor-2 datasets, respectively.

TABLE IV. System Evaluation Results of Image Enhancement Fusion Techniques and Six Deep Learning Model Combinations used in this Study on the APTOS-2019 Dataset

Models	Accuracy%	Precision%	Sensitivity %	Specificity %
ResNet50-V2	92.4	86.1	99.1	99.5
DenseNet121	93.0	86.7	99.8	99.7
Inception-V3	89.7	85.6	99.7	99.0
Xception	91.9	87.0	99.5	99.6
MobileNet-V2	91.2	83.0	99.7	99.5
Inception ResNet-V2	91.8	82.5	99.1	99.5

DenseNet121 achieved the highest performance on the APTOS-2019 dataset with an accuracy of 93.0%, precision of 86.7%, sensitivity of 99.8%, and specificity of 99.7%. Xception and ResNet50-V2 exhibited slightly lower accuracy compared to DenseNet121. Xception achieved an accuracy of 91.9%, precision of 87.0%, sensitivity of 99.5%, and specificity of 99.6%, while ResNet50-V2 had an accuracy of 92.4%, precision of 86.1%, sensitivity of 99.1%, and specificity of 99.5%. The remaining three models, InceptionV3, MobileNet-V2, and InceptionResNet-V2, showed the following performance: InceptionResNet-V2 achieved an accuracy of 91.8%, precision of 82.5%, sensitivity of 99.1%, and specificity of 99.5%; MobileNet-V2 achieved an accuracy of 91.2%, precision of 83.0%, sensitivity of 99.7%, and specificity of 99.5%; and Inception-V3 achieved an accuracy of 89.7%, precision of 85.6%, sensitivity of 99.7%, and specificity of 99.0%. Overall, the sensitivity and specificity of all six deep learning classification models were greater than 99.0%. In conclusion, the image preprocessing methods used in this study consistently demonstrated good performance across all models with minimal variation, as illustrated in Table IV.

TABLE V. System Evaluation Results of Image Enhancement Fusion Techniques and Six Deep Learning Model Combinations used in this Study on the Messidor-2 Dataset

Models	Accuracy %	Precision %	Sensitivity %	Specificity %
ResNet50-V2	84.0	80.0	99.1	90.6
DenseNet121	85.9	72.9	99.7	94.7
Inception-V3	81.4	57.2	98.9	87.3
Xception	86.9	79.5	99.1	96.2
MobileNet-V2	84.0	64.1	96.9	91.6
Inception ResNet-V2	87.9	75.7	99.4	97.0

InceptionResNet-V2 demonstrated the best performance on the Messidor-2 dataset, achieving an accuracy of 87.9%, precision of 75.7%, sensitivity of 99.4%, and specificity of 97.0%. The second and third best performing models were Xception and DenseNet121, respectively, with slightly lower accuracies. Xception had an accuracy of 86.9%, precision of 79.5%, sensitivity of 99.1%, and specificity of 96.2%. DenseNet121 achieved an accuracy of 85.9%, precision of 72.9%, sensitivity of 99.7%, and specificity of 94.7%. The remaining three models, ranked fourth to sixth, were ResNet50-V2, MobileNet-V2, and Inception-V3, respectively. ResNet50-V2 had an accuracy of 84.0%, precision of 80.0%, sensitivity of 99.1%, and specificity of 90.6%. MobileNet-V2 achieved an accuracy of 84.0%, precision of 64.1%, sensitivity of 96.9%, and specificity of 91.6%. Inception-V3 had the lowest performance with an accuracy of 81.4%, precision of 57.2%, sensitivity of 98.9%, and specificity of 87.3%. Overall, the sensitivity and specificity of all six deep learning classification models were greater than 97.0%. In conclusion, the image preprocessing methods used in this study consistently demonstrated good performance across all models with minimal variation, as illustrated in Table V.

TABLE VI. COMPARISON WITH OTHER STUDIES USING THE APTOS-2019 DATASET

Researches	Accuracy%	Precision %	Sensitivity %	Specificity %
[28]	86.5	85.7	86.1	85.9
[29]	92.0	85.0	99.0	99.0
[30] VGG19	88.1	90.3	79.3	94.0
[30] DenseNet121	89.5	83.8	92.1	87.7
[31]	84.2	-	98.5	98.8
[32]	83.4	69.7	67.7	67.0
This research methodology	**93.2**	**86.8**	**99.5**	**99.6**

Research in [28] utilized the APTOS-2019 Dataset to develop a diabetic retinopathy classification model. The model achieved an accuracy of 86.5%, precision of 85.7%, sensitivity of 86.1%, and specificity of 85.9%. Research [29] reported an accuracy of 92.0%, precision of 85.0%, sensitivity of 99.0%, and specificity of 99.0%. Research in [30] evaluated both VGG19 and DenseNet121 models, with VGG19 achieving an accuracy of 88.1%, precision of 90.3%, sensitivity of 79.3%, and specificity of 94.0, while DenseNet121 attained an accuracy of 89.5%, precision of 83.8%, sensitivity of 92.1%, and specificity of 87.7%, while research [31] showed an accuracy of 84.2%, sensitivity of 98.5%, and specificity of 98.8%. Research [32] reported an accuracy of 83.4%, precision of 69.7%, sensitivity of 67.7%, and specificity of 67.0. The proposed model demonstrated superior performance, achieving an accuracy of 93.2%, precision of 86.8%, sensitivity of 99.5%, and specificity of 99.6%. Overall, the model developed in this experiment demonstrated superior performance in classifying diabetic retinopathy compared to previous studies, as shown in Table VI.

Research in [30] utilized the Messidor-2 dataset to construct a diabetic retinopathy classification model using InceptionResNet-V2 and Inception-V3. The models' performance metrics were as follows: InceptionResNet-V2 achieved an accuracy of 69.2%, precision of 60.2%, sensitivity of 54.9%, and specificity of 75.9%. Inception-V3 obtained an accuracy of 72.8%, precision of 54.3%, sensitivity of 50.2%, and specificity of 80.6%. In contrast, research [33] reported a sensitivity of 81.0% and specificity of 86.1%, while research [29] showed an accuracy of 80.0%, precision of 85.0%, sensitivity of 89.0%, and specificity of 90.0%. Research [34] reported an accuracy of 89.6%, sensitivity of 91.4%, and specificity of 93.2. Comparing these results to the current experiment, which achieved an accuracy of 87.3%, precision of 75.2%, sensitivity of 99.6%, and specificity of 97.8%, it is evident that the current model exhibits high performance compared to other research and closely resembles the results of research [34]. The model in this experiment demonstrated a strong ability to detect positive cases due to its higher sensitivity compared to other studies, as shown in Table VII.

TABLE VII. COMPARISON WITH OTHER STUDIES USING THE MESSIDOR-2 DATASET

Researches	Accuracy%	Precision %	Sensitivity %	Specificity %
[29]	80.0	85.0	89.0	90.0
[30] Inception ResNet-V2	69.2	60.2	54.9	75.9
[30] Inception-V3	72.8	54.3	50.2	80.6
[33]	-	-	81.0	86.1
[34]	89,6	-	91.4	93.2
This research methodology	**87.3**	**75.2**	**99.6**	**97.8**

V. DISCUSSION

In this research, the diabetic retinopathy dataset consists of retinal images with varying resolutions and aspect ratios. The dataset has been adjusted to accommodate different CNN architectures. Additionally, background information or images unrelated to retinal diseases may impact classification accuracy, consistent with findings in prior research [5] [35]. Automated detection of diabetic retinopathy often involves preprocessing original images, including cropping, resizing, and adjusting resolution, to mitigate these issues and improve model training efficiency.

The imbalanced nature of the diabetic retinopathy dataset used to train the classification model resulted in suboptimal performance. To address this issue, categorical focal loss was employed to balance class weights instead of the standard categorical cross-entropy loss, aligning with findings from studies [36–38]. This approach demonstrated improved evaluation metrics. The dataset often suffers from class imbalance, where certain classes are overrepresented or underrepresented. This imbalance can bias the model toward predicting the majority class, making it challenging to accurately classify minority classes.

To enhance feature extraction, this research employed various image preprocessing techniques, including MSRCR, Gamma correction, CIELAB color space conversion, and CLAHE. MSRCR, based on retinex theory, simulates human visual perception by adjusting image brightness and color across multiple scales. Gamma correction controls brightness by

manipulating the gamma value, while CIELAB conversion transforms RGB images into a perceptually uniform color space. CLAHE enhances local contrast by adjusting histograms within image blocks. These techniques provide deeper insights and additional perspectives, contributing to a more comprehensive understanding of the underlying problem. Current research highlights the potential of deep learning techniques for feature extraction in medical image and signal processing.

Feature extraction relates to dimensionality reduction and is a crucial preprocessing step in deep learning, potentially leading to new discoveries. In this study, six models were used to extract features, along with an optimizer, which continuously adjusts model parameters (weights and biases) to improve accuracy. The main goal is to minimize the loss function, enhancing the model's predictive ability. The Nadam optimizer was employed, improving the measurement of the difference between predicted and actual results. The development of these techniques enhances classifier performance and aids in better decision-making for diagnosis. Evaluating large mixed input data requires significant memory and computational power. In this research, transfer learning and dropout techniques were applied to optimize model performance. Feature extraction ultimately helps save processing power and disk space while improving classification, thus enhancing the efficiency of diagnostic support systems. Therefore, developing effective algorithms for feature extraction is essential in building diagnostic support systems. A common method to eliminate irrelevant or redundant data is dimensionality reduction, with feature extraction being a popular approach.

Transfer learning was utilized by fine-tuning the weights of models trained on the APTOS-2019 dataset to adapt them to the Messidor-2 dataset. Although the results were slightly inferior, the best-performing model combination on the Messidor-2 dataset was InceptionResNet-V2, with a sensitivity of 99.4% and specificity of 97.0%. Sensitivity and specificity scores remained stable between 97% and 99%, demonstrating effective differentiation between diabetic retinopathy and normal retina. Compared to other research methods, these results were excellent, showcasing the effectiveness of transfer learning in this study.

VI. Conclusion

In recent years, the application of artificial intelligence-based diabetic retinopathy screening technologies has surged in the medical field. Automated screening methods effectively address the limitations of traditional manual diagnosis, enabling ophthalmologists to make quicker and more accurate assessments. Additionally, these technologies have played a crucial role in overcoming the challenges of inefficient screening in rural areas with limited medical resources, significantly enhancing the chances of early detection and treatment of diabetic retinopathy.

In this experiment, various image processing techniques were combined, including MSRCR, Gamma correction, CIELAB, CLAHE, and image enhancement. These techniques were integrated with six deep learning models and optimization techniques such as Nadam, transfer learning, and dropout. The models were trained and evaluated on the APTOS-2019 and Messidor-2 datasets, consisting of 3,662 and 1,744 images, respectively. These datasets exhibit class imbalance, with five severity levels: NDR, MiDR, MoDR, SDR, and PDR. The experimental results demonstrated that DenseNet121 achieved the highest performance on the APTOS-2019 dataset, with a sensitivity of 99.8% and specificity of 99.7%, outperforming previous studies.

Tables VI and VII present a comparison of the evaluation results between the best model from this research and other studies utilizing the same datasets (APTOS-2019 and Messidor-2). The data in these tables are directly cited from the best results of each study. For the APTOS-2019 dataset, the proposed method demonstrated comparable performance in terms of sensitivity and specificity, achieving values close to 100%. The model excelled across all metrics. In the case of the Messidor-2 dataset, although accuracy was not as optimal, sensitivity and specificity were more stable compared to other studies.

This study employed a fusion technique for image enhancement to improve the feature extraction efficiency of six deep learning models (ResNet50-V2, DenseNet121, Inception-V3, Xception, MobileNetV2, InceptionResNet-V2). The models were trained to classify healthy retinas and four different severities of diabetic retinopathy. Class-weighting techniques, including parameter setting and categorical focal loss, were utilized to enhance the model's accuracy in distinguishing various categories of retinal images, even when certain categories were underrepresented in the dataset.

To address the class imbalance problem and improve feature extraction performance, these models were trained to classify normal retinas and diabetic retinopathy across four different severity levels. categorical focal loss was adopted as the loss function to enhance classification accuracy for different types of retinal images, particularly those with smaller sample sizes. This loss function is specifically designed to address class imbalance issues in classification tasks, especially when the minority class has fewer samples than the majority class. By assigning more weight to samples from the minority class, categorical focal loss enables the model to better classify these underrepresented classes, which is a common challenge in real-world classification problems. This approach significantly contributes to the model's performance on datasets characterized by diabetic retinopathy. However, these methods can impact the model's ability to effectively classify classes with fewer samples. The greater the imbalance in sample proportions between classes, the more pronounced this effect becomes, particularly in improving classification performance for classes with minimal representation. The evaluation scores from this study's best results are compared with those from other research studies.

Acknowledgment

The authors would like to express their sincere gratitude to the anonymous reviewers for their useful comments and suggestions for improving the quality of this paper, and we thank National Pingtung University of Science and Technology Council of the Republic of China, Taiwan for financially supporting this research under Contract No. 111-2410-H-020-007.

REFERENCES

[1] J. Chantra, and S. Pratoom, “Factors Related to Diabetic Retinopathy in Type – 2 diabetes in Damnoensaduak Hospital, Ratchaburi Province, Thailand” J. Res. Improv. Qual. Life, vol. 3, pp. 25–36, 2023.

[2] V. Bellemo, G. Lim, T. H. Rim, G. S. W. Tan, C. Y. Cheung, S. Sadda, M. G. He, A. Tufail, M. L. Lee, W. Hsu, and D. S. W. Ting, "Artificial intelligence screening for diabetic retinopathy: the real-world emerging application," Curr. Diab. Rep., vol. 19, pp. 1-12, 2019.

[3] V. Bellemo, Z. W. Lim, G. Lim, D. Q. Nguyen, Y. Xie, M. Y. T. Yip, H. Hamzah, J. Ho, X. Q. Lee, W. Hsu, M. L. Lee, L. Musonda, M. Chandran, G. Chipalo-Mutati, M. Muma, G. S. W. Tan, S. Sivaprasad, G. Menon, T. Y. Wong, and D. S. W. Ting, "Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study," Lancet Digit. Health, vol. 1, no. 1, pp. e35-e44, 2019.

[4] A. Grzybowski, P. Brona, G. Lim, P. Ruamviboonsuk, G. S. Tan, M. Abramoff, and D. S. Ting, "Artificial intelligence for diabetic retinopathy screening: a review," Eye, vol. 34, no. 3, pp. 451-460, 2020.

[5] N. Tsiknakis, D. Theodoropoulos, G. Manikis, E. Ktistakis, O. Boutsora, A. Berto, F. Scarpa, A. Scarpa, D. I. Fotiadis, and K. Marias, "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," Comput. Biol. Med., vol. 135, pp. 104599, 2021.

[6] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian, "Color Retinal Image Enhancement Based on Luminosity and Contrast Adjustment," IEEE Trans. Biomed. Eng., vol. 65, no. 3, pp. 521-527, 2018.

[7] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization With Dual Gamma correction," IEEE Access, vol. 6, pp. 11782-11792, 2018.

[8] S. Sonali, S. Sahu, A. K. Singh, S. P. Ghrera, and M. Elhoseny, "An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE," Opt Laser Technol., vol. 110, pp. 87-98, 2019.

[9] M. R.Islam, L. F. Abdulrazak, M. Nahiduzzaman, M. O. F. Goni, M. S. Anower, M. Ahsan, J. Haider, and M. Kowalski, "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," Comput. Biol. Med., vol. 146, pp. 105602, 2022.

[10] [10] G. Cao, L. Huang, H. Tian, X. Huang, Y. Wang, and R. Zhi, "Contrast enhancement of brightness-distorted images by improved adaptive Gamma correction," Comput. Electr. Eng., vol. 66, pp. 569-582, 2018.

[11] M. Veluchamy and B. Subramani, "Image contrast and color enhancement using adaptive Gamma correction and histogram equalization," Optik, vol. 183, pp. 329-337, 2019.

[12] M. R. K. Mookiah, S. Hogg, T. J. MacGillivray, V. Prathiba, R. Pradeepa, V. Mohan, R. M. Anjana, A. S. Doney, C. N. A. Palmer, and E. Trucco, "A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification," Med. Image Anal., vol. 68, pp. 101905, 2021.

[13] Y. E. Almalki, N. A. Jandan, T. A. Soomro, A. Ali, P. Kumar, M. Irfan, M. U. Keerio, S. Rahman, A. Alqahtani, S. M. Alqhtani, M. A. Hakami, W. A. Aldhabaan, and A. S. Khairallah, "Enhancement of Medical Images through an Iterative McCann Retinex Algorithm: A Case of Detecting Brain Tumor and Retinal Vessel Segmentation," Appl. Sci., vol. 12, no. 16, pp. 8243, 2022.

[14] Y. Jiang, Z. Ma, C. Wu, Z. Zhang, and W. Yan, "RSAP-Net: joint optic disc and cup segmentation with a residual spatial attention path module and MSRCR-PT pre-processing algorithm," BMC Bioinform., vol. 23, no. 1, pp. 1-21, 2022.

[15] Q. Mirsharif, F. Tajeripour, and H. Pourreza, "Automated characterization of blood vessels as arteries and veins in retinal images," Comput. Med. Imaging Graph., vol. 37, no. 8, pp. 607-617, 2013.

[16] K. Sriporn, C. F. Tsai, C. E. Tsai, and P. Wang, “Analyzing lung disease using highly effective deep learning techniques,” Healthcare, vol. 8, pp.1-21, 2020.

[17] D. Nagpal, S. N. Panda, M. Malarvel, P. A. Pattanaik, and M. Zubair Khan, "A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 9, pp. 7138-7152, 2022.

[18] N. Salamat, M. M. S. Missen, and A. Rashid, "Diabetic retinopathy techniques in retinal images: A review," Artif. Intell. Med., vol. 97, pp. 168-188, 2019.

[19] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” arXiv, pp.1-15, 2016.

[20] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” arXiv, pp.1-9, 2018.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," arXiv, pp.1-10, 2015.

[22] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” arXiv, pp.1-8, 2017.

[23] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv, pp.1-9, 2017.

[24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," arXiv, pp.1-12, 2016.

[25] K. Sriporn, C. F. Tsai, C. E. Tsai, and P. Wang, “Analyzing malaria disease using effective deep learning approach,” Diagnostics, vol. 10, pp. 1-22, 2020.

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," arXiv, pp.1-10, 2018.

[27] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv, pp.1-14, 2017.

[28] M. Tian, H. Wang, Y. Sun, S. Wu, Q. Tang, and M. Zhang, "Fine-grained attention & knowledge-based collaborative network for diabetic retinopathy grading," Heliyon, vol. 9, no. 7, pp. e17217, 2023.

[29] M. Fatima, M. Imran, A. Ullah, M. Arif, and R. Noor, "A unified technique for entropy enhancement based diabetic retinopathy detection using hybrid neural network," Comput. Biol. Med., vol. 145, pp. 105424, 2022.

[30] S. El-Ateif and A. Idri, "Single-modality and joint fusion deep learning for diabetic retinopathy diagnosis," Sci. Afr., vol. 17, pp. e01280, 2022.

[31] A. Sugeno, Y. Ishikawa, T. Ohshima, and R. Muramatsu, "Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning," Comput. Biol. Med., vol. 137, pp. 104795, 2021.

[32] G. Yue, Y. Li, T. Zhou, X. Zhou, Y. Liu, and T. Wang, "Attention-Driven Cascaded Network for Diabetic Retinopathy Grading from Fundus Images," Biomed. Signal Process. Control., vol. 80, pp. 104370, 2023.

[33] G. Saxena, D. K. Verma, A. Paraye, A. Rajan, and A. Rawat, "Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets," Intell.-Based Med., vol. 3-4, pp. 100022, 2020.

[34] P. Modi and Y. Kumar, "Smart detection and diagnosis of diabetic retinopathy using bat based feature selection algorithm and deep forest technique," Comput. Ind. Eng., vol. 182, pp. 109364, 2023.

[35] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," Inform. Med. Unlocked, vol. 20, pp. 100377, 2020.

[36] M. Saini and S. Susan, "Deep transfer with minority data augmentation for imbalanced breast cancer dataset," Appl. Soft Comput., vol. 97, pp. 106759, 2020.

[37] M. Saini and S. Susan, "Diabetic retinopathy screening using deep learning for multi-class imbalanced datasets," Comput. Biol. Med., vol. 149, pp. 105989, 2022.

[38] H. Li, X. Dong, W. Shen, F. Ge, and H. Li, "Resampling-based cost loss attention network for explainable imbalanced diabetic retinopathy grading," Comput. Biol. Med., vol. 149, pp. 105970, 2022.