WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 5, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v5n1.17
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Lightweight Adaptive Grading Model for Diabetic Retinopathy Based on Ultra-Widefield Fundus Images

Xinpeng Miao [1], Shuaichao Feng [1], Jie Zhao [1], Guangping Zhuo [1,*], Fei Ma [2]

[1] School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China

[2] Department of Computer Science and Technology, Taiyuan University, Taiyuan 030032, China

ABSTRACT

Wide-field fundus images play a crucial role in the diagnosis of diabetic retinopathy. However, traditional convolutional neural networks face limitations when processing such images, struggling to balance local details and global structures. To address this, a novel lightweight feature extraction network, called the Adaptive Nonlinear Lightweight Retina Network (ANLR-Net), is proposed. The network consists of two key modules: the Adaptive Feature Mapping Module (AFMM) and the Lightweight Module (LEM). AFMM replaces traditional linear weight matrices with learnable nonlinear functions, enabling precise capture of both global structures and local details while reducing network parameters and improving learning efficiency. LEM performs dimensionality reduction and restoration using 1×1 convolutions, which reduces computational cost while maintaining the network's lightweight nature. ANLR-Net effectively captures multi-scale features in wide-field images, significantly reducing computational complexity and parameter count. Experimental results show that ANLR-Net achieves a classification accuracy of 95.05%, specificity of 98.64%, and an AUC of 0.9712 on the wide-field dataset, outperforming traditional models.

KEYWORDS

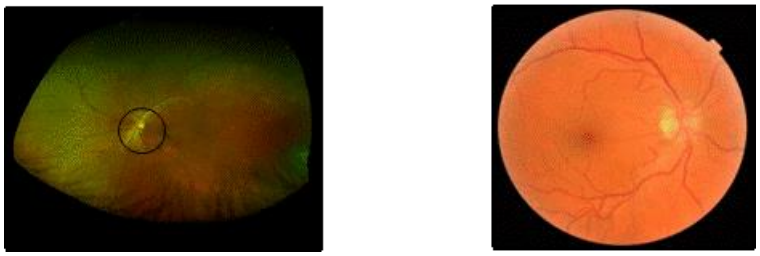
Kolmogorov-Arnold; Learnable nonlinear functions; Legendre polynomial; CLAHE; Lightweight

1. INTRODUCTION

Diabetes is a globally prevalent chronic disease, and its incidence has been rising steadily in recent years [1]. According to the World Health Organization (WHO), there are currently over 400 million people worldwide living with diabetes, and this number is expected to exceed 783 million by 2025. In China, the number of diabetes patients has reached 140 million [2]. Diabetic Retinopathy (DR) is one of the most common and severe complications among diabetic patients. Studies show that about one-third of diabetic patients develop DR, and nearly 10% of them will progress to sight-threatening retinal conditions [3]. Diabetic Retinopathy has become one of the leading causes of blindness in the working-age population globally, severely affecting patients' ability to live and work normally [4]. Clinical research indicates that early detection and treatment of diabetic retinopathy can significantly reduce the risk of vision loss [5]. However, due to the lack of obvious symptoms in the early stages of the disease, many patients only realize they have vision problems when the condition has already progressed significantly, which negatively impacts the effectiveness of treatment [6].

Ultra-widefield fundus images offer a broader retinal coverage compared to conventional fundus images, encompassing the peripheral regions of the retina. These peripheral areas are often where the early signs of diabetic retinopathy (DR) appear. Traditional fundus images may only capture the central portion of the retina, potentially missing lesions in the peripheral areas, which can lead to

reduced accuracy in early diagnosis [7-9]. The use of ultra-widefield fundus images improves the detection rate of diabetic retinopathy, aiding in early detection and timely intervention [10]. A comparison between conventional fundus images and ultra-widefield fundus images is shown in Figure 1.



(a) Ultra wide angle fundus image (b) Ordinary fundus image

Figure 1. Comparison between ultra-widefield retinal imaging and conventional fundus imaging

With the rapid development of deep learning technologies, researchers have begun exploring their applications in ophthalmic image analysis to improve the diagnostic accuracy of diabetic retinopathy. Many researchers have achieved promising results. Hemant et al. [11] conducted experiments using a DenseNet network model with transfer learning on a private dataset. By applying data augmentation to balance the dataset, the results showed that transfer learning improved the model's classification accuracy and specificity. Kandel et al. [12] pre-trained InceptionNet-V3 on ImageNet and fine-tuned the model parameters to complete a five-class classification task, achieving an accuracy of 90.9% on the Kaggle dataset. Kathiresan et al. [13] proposed a collaborative deep learning model for the automatic grading and classification of diabetic retinopathy images, covering several steps, including preprocessing, segmentation, and classification. On the Messidor dataset, this model demonstrated excellent performance with an accuracy of 99.28%, sensitivity of 98.54%, and specificity of 99.38%. Huang et al. [14] introduced a lesion patch-based contrastive learning framework for automating the grading of diabetic retinopathy (DR). This method showed outstanding performance in DR grading tasks on the EyePACS dataset. Hou et al. [15] proposed the Cross-Field Transformer (CrossFiT) method, which captures the relationships between two-field fundus images and uses the new DRTiD dataset, significantly improving the accuracy of diabetic retinopathy (DR) grading. Alberto Solano et al. [16] introduced the ConvMixer method, which combines image patch processing from ViT and convolutional foundations from U-Net, achieving excellent results in medical semantic segmentation tasks. Zhang et al. [17] proposed a deep graph convolutional network (DGCN) for the automated grading of diabetic retinopathy (DR) without the need for expert annotations. This method achieved accuracies of 89.9% and 91.8.

2. MODEL

This paper designs and introduces a novel convolutional neural network architecture based on the DenseNet model. In the benchmark model, the fixed convolutional layer activation functions may face challenges when processing complex data, especially with large training datasets, and are prone to issues such as vanishing or exploding gradients. To address these problems, this paper introduces an improved structure based on the Kolmogorov-Arnold idea, employing nonlinear combinations to enhance the flexibility and adaptability of feature extraction. In the original method, the spline functions used during training were difficult to optimize. In this study, Legendre polynomials are used instead of splines, thereby improving training stability and feature extraction capability. Additionally, to effectively handle the complex feature distributions in ultra-widefield fundus images, an adaptive feature mapping module is introduced. This module performs adaptive transformation of input features through learnable nonlinear activation functions, enabling it to flexibly capture complex features in the lesion areas. Combined with Legendre polynomials, this further enhances the global consistency and nonlinear expressiveness of the features, significantly improving classification

accuracy and robustness. To meet the diverse feature extraction needs in ultra-widefield images, dilated convolutions are also employed, expanding the receptive field of the network and allowing the model to capture a broader range of contextual information. Considering the large number of parameters in the benchmark model, depthwise separable convolutions are introduced to reduce the parameter count and improve computational efficiency. Through these improvements, the model demonstrates stronger feature extraction capabilities, higher computational efficiency, and better robustness in handling the ultra-widefield fundus image classification task.

2.1. Introduction to DenseNet Model

In 2017, Huang et al. [18] proposed an innovative convolutional neural network model called DenseNet. To address the issues of poor information flow and gradient vanishing in deep networks, the authors designed a densely connected network structure. Through this design, DenseNet ensures that each layer is directly connected to all preceding layers, facilitating more efficient information transmission and feature reuse. This model effectively reduces the number of parameters, enhances the model's expressive power, and improves the stability of training.

To further enhance performance, DenseNet adopts the concept of "dense connections," allowing each layer in the network to access feature information from all previous layers [19]. This structure not only improves feature diversity but also significantly enhances gradient propagation. Research results show that DenseNet demonstrates exceptional performance in multiple visual tasks, particularly in areas such as image classification and object detection.

The DenseNet network consists of an input layer, convolutional layers, pooling layers, dense blocks, fully connected layers, and an output layer. Its structure utilizes convolutional computations with dense connections, where each layer receives the outputs of all preceding layers as its input. This approach avoids redundant computations and improves the efficiency of feature propagation. By using convolutional kernels of different sizes, DenseNet is able to extract rich feature information and combine these features through dense connections, thereby enhancing the model's performance. In addition, to mitigate the gradient vanishing problem that may arise due to increased network depth, DenseNet employs skip connections and batch normalization, effectively promoting the efficient backpropagation of gradients. Although DenseNet performs excellently in image recognition, it still faces challenges such as high computational overhead, a large number of parameters, and increased model complexity. These issues require more computational resources and time during training. Moreover, because of the deep architecture, the receptive fields of the convolutional layers are relatively small, making it difficult for the network to capture subtle features of the samples, which affects the training effectiveness.

2.2. Adaptive Feature Mapping Module

The Adaptive Feature Mapping Module is designed based on the Kolmogorov-Arnold theorem, utilizing learnable nonlinear activation functions to dynamically adjust the input data and perform feature extraction. In the classification and detection of diabetic retinopathy in ultra-widefield images, the complexity of the images and the diversity of lesion feature distributions impose higher demands on the model's adaptability. The proposed Adaptive Feature Mapping Module combines Legendre polynomials to further enhance the network's feature representation capability, enabling it to more effectively handle complex images with varying features.

2.2.1. Kolmogorov Arnold theorem

The Kolmogorov-Arnold theorem states that any multivariable continuous function can be represented as a combination of multiple univariate functions, and its formula is as follows:

$$f(x) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p} \left(x_p \right) \right) \quad (1)$$

This theorem provides theoretical support for the representation and processing of high-dimensional data. Based on this, the model proposed in this paper performs nonlinear transformations on the input features, followed by a combination of activation functions to achieve efficient feature extraction. In the Adaptive Feature Mapping Module, we use nonlinear activation functions to dynamically adjust the input data, enabling the model to better handle complex image features.

2.2.2. Learnable nonlinear functions in adaptive feature mapping

The core of the Adaptive Feature Mapping Module lies in performing adaptive transformation of the input features using learnable nonlinear functions, enabling it to flexibly capture the complex features of the lesion areas. This not only enhances the flexibility of feature mapping but also effectively improves the accuracy and robustness of detection. The convolution formula is as follows:

$$\varphi = \omega_b \cdot b(x) + \tilde{\varphi}(x) \tag{2}$$

Here, the activation function $b(x)$ uses the SiLU function:

$$b(x) = SiLU(x) = \frac{x}{1+e^{-x}} \tag{3}$$

By using the SiLU function, the model is able to flexibly process the input data, thereby avoiding the common issue of gradient vanishing and enhancing the nonlinear expressive power during the feature mapping process.

To further enhance the feature representation capability of the Adaptive Feature Mapping Module, this paper introduces Legendre polynomials as the basis for nonlinear mapping. The Legendre polynomial $P_n(x)$ is an nth-degree orthogonal polynomial in

x, and its orthogonality ensures a non-redundant representation of features. In the Adaptive Feature Mapping Module, Legendre polynomials are used to globally approximate the input features, allowing the model to efficiently capture both global and local lesion features when processing ultra-widefield images. The orthogonality of Legendre polynomials is represented by the following formula:

$$\int_{-1}^{1} P_n(x)P_m(x)dx = 0 \qquad n \neq m \tag{4}$$

The convolution form in the network is as follows:

$$\tilde{\varphi}(x) = \sum_{i=0}^{N} \omega_i \cdot P_i(x) \tag{5}$$

This approach, combining Legendre polynomials, not only improves the global consistency of feature extraction but also reduces the number of model parameters, making the computation more efficient.

2.2.3. The advantages of adaptive feature mapping

Global Approximation Ability: The global orthogonality of Legendre polynomials enables them to efficiently approximate the lesion features in retinal ultra-widefield images, particularly in the peripheral regions of the retina. Compared to traditional spline functions, Legendre polynomials are better at capturing global patterns, thereby improving the accuracy of classification and detection.

Numerical Stability: The orthogonality of Legendre polynomials reduces redundancy and interference during the feature extraction process, significantly enhancing the numerical stability of the computations. This is particularly important for handling high-dimensional and complex image data, preventing issues like gradient vanishing or explosion, and ensuring the model's robustness.

Reduced Parameters, Improved Efficiency: Compared to the piecewise representation method using spline functions, Legendre polynomials maintain efficient feature extraction capabilities while reducing the number of parameters. This allows the KAN network to process ultra-widefield images more efficiently, reducing the consumption of computational resources.

2.2.4. Structure diagram of adaptive feature mapping module

The structure diagram of the Adaptive Feature Mapping Module (AFMM) is shown in Fig 2. It consists of two parallel paths: one path processes the input features through the base convolution layer (Base Conv) for preliminary processing, while the other path first applies a 1x1 convolution for dimensionality reduction, followed by a learnable nonlinear activation function $\tilde{\varphi}(x)$ for adaptive nonlinear transformation of the features, and finally performs a 1x1 convolution for dimensionality expansion. The feature maps from both paths are fused at the end through an element-wise addition operation. This module combines low-level feature extraction from the base convolution with the adaptive adjustment capability of the nonlinear activation function, effectively enhancing the flexibility and accuracy of feature representation, thereby improving the model's ability to capture complex lesion areas.

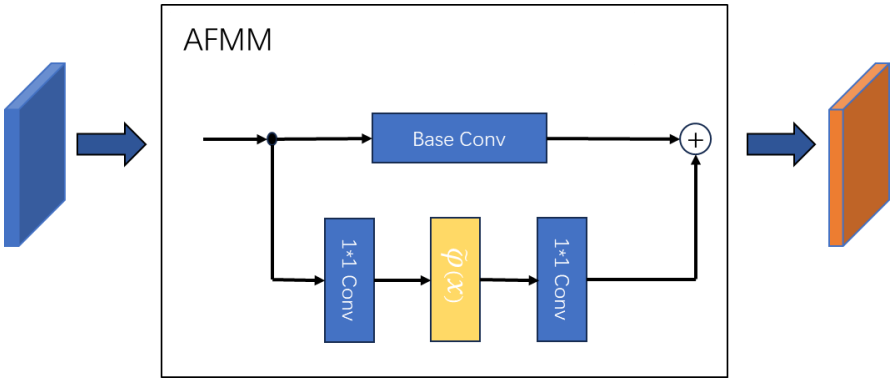


Figure 2. Structure diagram of adaptive feature mapping module

2.3. Lightweight Module

The bottleneck structure is a module design used to reduce the network's computational cost and parameters. The core idea is to compress information by reducing the dimensionality of the intermediate layers, while maintaining the consistency of the input and output dimensions. The basic form of the bottleneck structure consists of the following three stages:

(1) Dimensionality Reduction Layer: First, the feature dimensionality of the input is reduced using 1x1 convolutions or other dimensionality reduction operations. This step reduces the size of the feature maps in the intermediate calculations, thereby lowering both the number of parameters and the computational cost.

(2) Feature Extraction Layer: Next, a learnable nonlinear function is applied. The introduction of dimensionality reduction significantly reduces the computational complexity without excessively losing information.

(3) Dimensionality Expansion Layer: Finally, another 1x1 convolution restores the feature dimensionality to its original size. This step ensures that the network's input and output dimensions remain consistent.

Figure 3 illustrates the basic framework of the bottleneck structure. By compressing high-dimensional data into a lower-dimensional space, performing feature extraction, and then restoring it back to high-dimensional space, it effectively reduces computational cost.

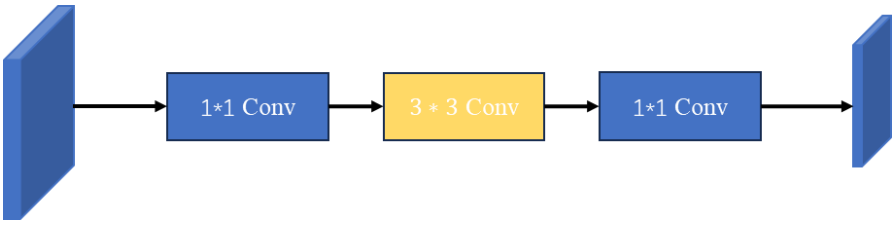


Figure 3. Bottleneck structure diagram

The advantages of the lightweight bottleneck structure include:

(1) Reducing Computational Complexity: The bottleneck structure reduces the computational load of intermediate layers by compressing the dimensionality of the feature maps. This operation is especially important for processing ultra-widefield images, as these images have large data sizes and high computational costs. With the bottleneck design, the network can process large-scale data more efficiently.

(2) Reducing Parameters and Improving Efficiency: The bottleneck structure reduces the number of parameters in the convolutional layers, making the network more lightweight and easier to deploy on resource-limited devices for real-time inference, such as mobile devices or embedded systems.

(3) Maintaining Performance: Despite the reduction in model parameters and computational load, the bottleneck structure ensures feature integrity during the compression and recovery processes through careful design. As a result, the model does not suffer significant losses in accuracy and is able to maintain high-quality feature extraction performance effectively.

2.4. ANLR Net Network Model Structure

Based on the above improvements, the network structure designed in this paper is shown in Fig 4. Compared to the original DenseNet, we have modified the convolutional layers in the Dense Block by removing the traditional linear weight matrices and using learnable nonlinear polynomial Legendre polynomials for adaptive feature mapping of the input images. The activation function in the Transition layer has also been replaced from RELU to SELU. SELU can adaptively keep the output within a certain range, which helps with the model's convergence. In addition, we have incorporated dilated convolutions and depthwise separable convolutions, which further enhance feature extraction capabilities and computational efficiency. With these improvements, the network performs more efficiently and stably when handling complex tasks.

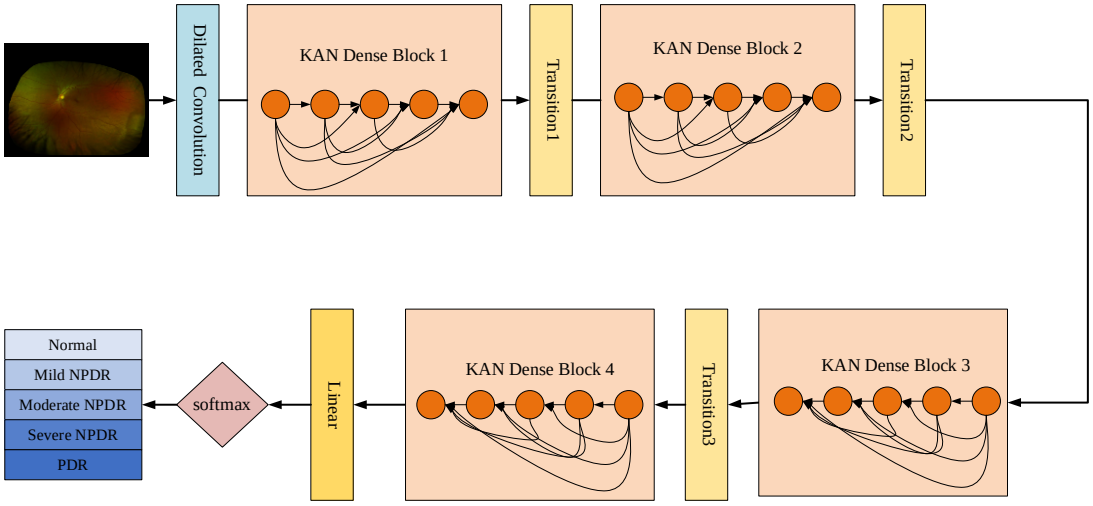


Figure 4. Diagram of the improved model structure

3. EXPERIMENTAL DATA

3.1. Dataset

The ultra-widefield image dataset used in this study comes from the ultra-widefield image database of a collaborating ophthalmology hospital, which contains a large number of high-resolution retinal images primarily used for research and diagnosis of diabetic retinopathy (DR) and other ophthalmic diseases. After initial screening, a total of 9,966 images were retained for model training. All images were captured by professional equipment and annotated by experienced ophthalmologists, ensuring the accuracy and reliability of the labels.

In addition, for comparative research, this paper also uses a dataset of regular retinal images from the diabetic retinopathy detection competition hosted on the Kaggle platform, which contains 35,126 regular retinal images. According to international standards, diabetic retinopathy is typically classified into five stages: no diabetic retinopathy, mild, moderate, severe non-proliferative retinopathy, and severe proliferative retinopathy. Example images for each category are shown in Figure 5, and the dataset distribution is provided in Tables 1 and 2.

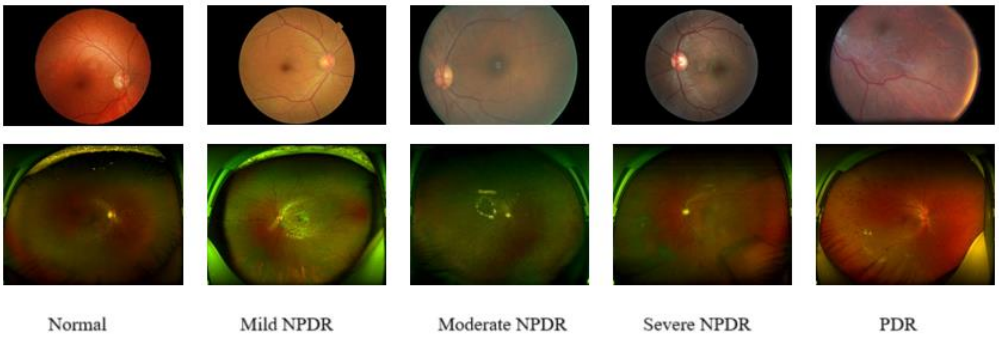


Figure 5. Five-category images of DR on two dataset

Table 1. Data distribution of ultra wide angle fundus image dataset (Unit: Images)

Data processing methods	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Ultra wide angle image dataset	7478	553	722	248	965
Data cleaning	7219	550	722	248	961
Data augmentation	2493	1430	1300	1044	961

Table 2. Data distribution of ordinary fundus image dataset (Unit: Images)

Data processing methods	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
EyePACS	25810	2443	5292	873	708
Data cleaning	3226	2443	2646	2620	2832
Data augmentation	3003	2317	2442	2444	2513

3.2. Data Preprocessing

Due to the wide-angle nature of ultra-wide field fundus images, combined with their high resolution, the number of parameters in the model significantly increases, resulting in reduced computational efficiency and a greater demand for computational resources. Additionally, the presence of many dark edges and noise around the images introduces a lot of irrelevant information, which not only occupies

the effective information area but may also interfere with the training and prediction results of the model.

To address these issues and optimize overall performance, we decided to use UNet for image segmentation to extract useful regions while appropriately reducing the resolution. This approach can not only minimize the interference from irrelevant information but also significantly lower computational complexity, thereby improving the processing efficiency of the model. The preprocessing workflow is illustrated in Figure 6.

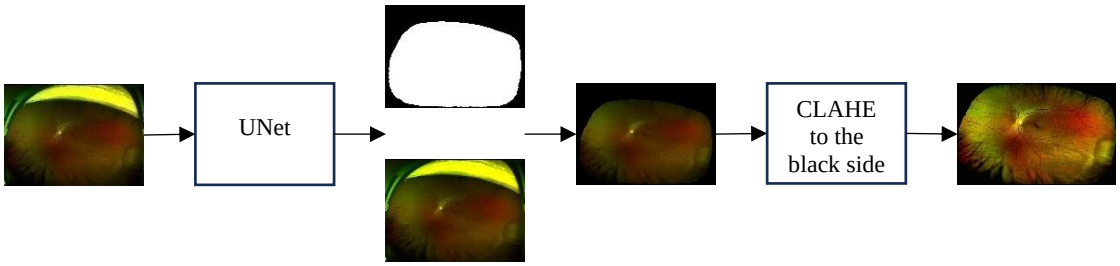


Figure 6. Data preprocessing flowchart

3.2.1. UNet

UNet is a convolutional neural network architecture widely used for medical image segmentation, proposed by Olaf Ronneberger [20] et al. in 2015. Its architecture is designed to handle small datasets while generating high-resolution segmentation structures. The UNet architecture can be divided into two main parts: the Encoder and the Decoder. The Encoder consists of multiple convolutional layers and pooling layers that progressively reduce the spatial dimensions of the feature maps. Each convolutional block typically includes two 3x3 convolutional layers followed by a 2x2 max pooling layer. Through this downsampling process, the network is able to extract high-level features from the images. The Decoder, on the other hand, gradually restores the spatial dimensions of the feature maps through transposed convolutional layers and convolutional layers. During each upsampling step, skip connections between the Decoder and Encoder incorporate high-resolution feature maps from the Encoder into the Decoder, thereby preserving more detailed information. Ultimately, the network outputs the segmentation results through a 1x1 convolutional layer, achieving precise localization of the target. The overall design philosophy of UNet is to extract rich features through the Encoder while progressively restoring image resolution via the Decoder, effectively merging low-level and high-level features to achieve accurate segmentation results that also contain contextual information.

In this study, UNet plays a crucial role in segmenting ultra-wide field fundus images to identify and extract relevant anatomical structures. Leveraging its ability to produce high-resolution segmentation maps, UNet enables the precise delineation of areas of interest, such as blood vessels and lesions, while minimizing the interference from irrelevant information like noise and dark edges. This segmentation facilitates subsequent analysis and enhances the overall performance of the model, providing a solid foundation for further image processing and interpretation tasks.

3.2.2. Contrast Limited Adaptive Histogram Equalization

To address issues such as low contrast in ultra-wide field fundus images, this paper employs Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the contrast of ultra-wide field fundus images, thereby providing higher-quality data for the classification and detection of diabetic retinopathy.

The processing steps of CLAHE include the following key stages:

(1) Image Partitioning: The input image is divided into multiple small, equally sized sub-regions. This step allows for independent histogram calculations for each local area, facilitating the enhancement of local details in the image.

(2) Local Histogram Calculation and Clipping Limit Setting: A histogram is computed for each sub-region, and a clipping limit is set to prevent excessive enhancement of certain grayscale levels in the histogram. The formula for calculating the clipping limit is:

$$\beta = \frac{MN}{L}\left[1 + \frac{\alpha}{100}(S_{max} - 1)\right] \tag{6}$$

Where β is the clipping limit, MN is the number of pixels in each sub-region, L is the number of gray levels, α is the clipping factor, and S_{max} is the maximum allowable slope. This approach controls the extent of contrast enhancement for each local area.

(3) Histogram Equalization and Reconstruction: For the grayscale levels that exceed the clipping limit, the frequencies of these levels are redistributed to other grayscale levels to balance the histogram distribution. Subsequently, an equalized histogram is generated for each local region, and the equalized results of the small regions are smoothly combined through interpolation to form a global equalized image.

(4) Interpolation Smoothing: Bilinear interpolation is applied at the boundaries of local regions to ensure that pixel values transition smoothly between different areas, avoiding discontinuities or abrupt changes in the image. The final resulting image achieves a balance between global and local details. The images before and after processing are shown in Figure 7.

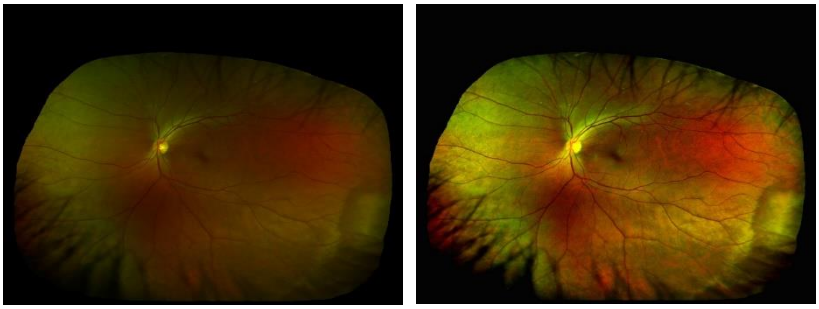


Figure 7. Comparison chart before and after CLAHE

In the classification and detection of diabetic retinopathy, images processed with CLAHE can provide the model with richer feature information. This allows the model to better learn key pathological features, such as microaneurysms, hemorrhages, and exudates, thereby improving its classification accuracy and generalization ability.

4. EXPERIMENTS AND DISCUSSION

4.1. Experimental Environment and Parameter Settings

In this experiment, the server configuration used was an Intel(R) Xeon(R) Silver 4316 CPU paired with an NVIDIA GTX 3090 GPU, and the memory was 192GB (8×24GB). The programming language chosen was Python, and the deep learning framework used was PyTorch. The initial learning rate (α) was set to 0.001, with a dynamic gradient decay factor (gamma) of 0.1, and a batch size of 2. After organizing the dataset, it was randomly divided into three parts: 70% for training, 20% for validation, and 10% for testing. The training process was conducted over a total of 200 epochs.

4.2. Experimental Evaluation Metrics

In the classification and detection of diabetic retinopathy, common model evaluation metrics include accuracy, precision, recall, F1-score, AUC (Area Under Curve), specificity, and confusion matrix. The metrics chosen for this paper are accuracy, specificity, AUC, and Kappa coefficient.

Accuracy $R_{Accuracy}$ represents the proportion of correctly classified samples among all samples and serves as an indicator of the model's overall classification performance. When dealing with imbalanced datasets, it is often used in conjunction with the confusion matrix to obtain a more comprehensive evaluation of the model.

$$R_{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

Specificity $R_{Specificity}$ is a crucial metric in medical diagnosis. It indicates the proportion of correctly predicted negative samples (i.e., samples without the disease) relative to the actual number of negative samples, and is typically used to assess a diagnostic test's ability to identify healthy individuals. Specifically, specificity reflects the proportion of patients without the disease that the test correctly identifies as disease-free.

$$R_{Specificity} = \frac{TN}{TN+FP} \tag{8}$$

Where:

TP is the number of true positive samples (actual positives correctly predicted as positive);

FN is the number of false negative samples (actual positives incorrectly predicted as negative);

FP is the number of false positive samples (actual negatives incorrectly predicted as positive);

TN is the number of true negative samples (actual negatives correctly predicted as negative).

AUC (Area Under Curve) refers to the area under the ROC curve. The ROC curve (Receiver Operating Characteristic Curve) illustrates the trade-off between the true positive rate and the false positive rate of a classification model at various thresholds. The AUC value ranges from 0 to 1, with values closer to 1 indicating better model performance and stronger ability to distinguish between positive and negative classes.

The Kappa coefficient can be used to evaluate whether the model's classification results are better than random guessing, especially in cases of imbalanced datasets. The Kappa coefficient ranges from -1 to 1, where 0 indicates no difference between the classification results and random guessing, 1 indicates perfect agreement, and negative values indicate consistency lower than random guessing.

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{9}$$

Where p_0 represents the observed agreement, or the proportion of results predicted by the model that align with expert annotations; p_e represents the expected random agreement, or the proportion of expected agreement across categories based on random guessing.

4.3. Comparative Experiments on Different Dataset

In this study, we conducted a comprehensive experimental analysis of diabetic retinopathy detection, evaluating the performance of models on both the EyePACS dataset and the ultra-widefield dataset. These two datasets have distinct characteristics: the Kaggle dataset consists of standard fundus images, while the ultra-widefield dataset contains high-resolution fundus images with a larger field of view, representing different image acquisition conditions and application scenarios. By comparing the performance of models on these two datasets, we can more thoroughly assess their applicability and advantages in various task settings. The following sections will present and analyze the experimental results for the Kaggle dataset and the ultra-widefield dataset separately.

4.3.1. Experimental results of EyePACS dataset

On the EyePACS dataset, the performance of multiple deep learning models was compared. ANLR-Net achieved the best performance, with an accuracy of 92.30%, a specificity of 95.75%, a Kappa value of 0.9231, and an AUC of 0.9612. Other models, such as Inception V3 and ResNet50, achieved accuracies of 86.17% and 85.16%, respectively, demonstrating good performance. Overall, the ANLR-Net model exhibited superior performance in ophthalmic disease detection. Detailed data are shown in Table 3.

Table 3. Comparison of classification results of the EyePACS dataset

Model	accuracy/%	specificity /%	Kappa	AUC
InceptionV3	86.17	93.88	0.8428	0.9420
ResNext50	85.16	94.01	0.8577	0.9410
DenseNet161	83.96	93.64	0.8556	0.9376
ANLR-Net	92.30	95.75	0.9231	0.9612

4.3.2. Experimental results of ultra wide angle dataset

In this section, we present the experimental results on the ultra-widefield dataset in detail, focusing on the performance of four models: InceptionV3, ResNext50, DenseNet161, and ANLR-Net. The performance of each model across different metrics is summarized in Table 4, and a thorough analysis is provided.

As shown in the table, the ANLR-Net model demonstrates outstanding performance across all metrics. Specifically, the model achieves an average accuracy of 95.05% and an average specificity of 98.64%, highlighting its exceptional capability in classification tasks. Additionally, the ANLR-Net model attains a Kappa value of 0.9484 and an AUC value of 0.9712, further confirming its superiority in model consistency and predictive performance. Detailed results are shown in Table 4.

Table 4. Comparison of classification results of the EyePACS dataset

Model	accuracy/%	specificity /%	Kappa	AUC
InceptionV3	89.46	95.38	0.8698	0.9610
ResNext50	88.26	95.51	0.8729	0.9523
DenseNet161	86.89	95.19	0.8752	0.9577
ANLR-Net	95.05	98.64	0.9484	0.9712

To provide a more intuitive illustration of the performance of each model during the training process, we have included three charts: first, Figure 8 shows a trend graph where the accuracy gradually increases as the training epochs progress; second, Figure 9 also presents a trend graph where the loss value gradually decreases with the increase in training epochs; and finally, Figure 10 displays the ROC curves for each class. It is worth noting that the ROC curves for all classes exhibit high true positive rates, with the AUC values for all classes exceeding 0.95, indicating strong classification performance across all categories. These charts effectively demonstrate the performance improvement of the models during training. ANLR-Net outperforms the comparison models across all performance metrics, proving its superiority and feasibility in the classification of diabetic retinopathy images.

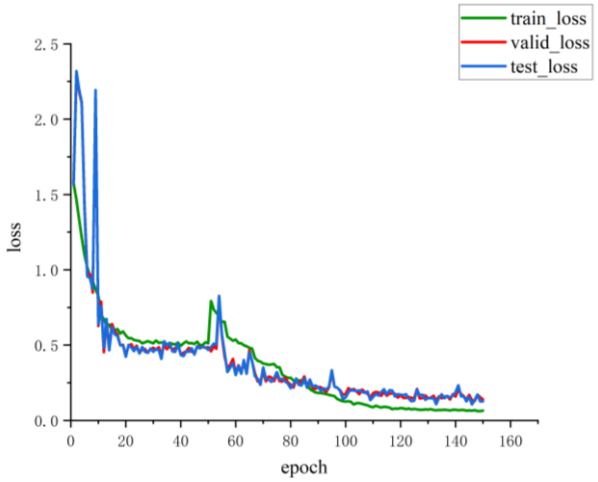


Figure 8. Accuracy Curve

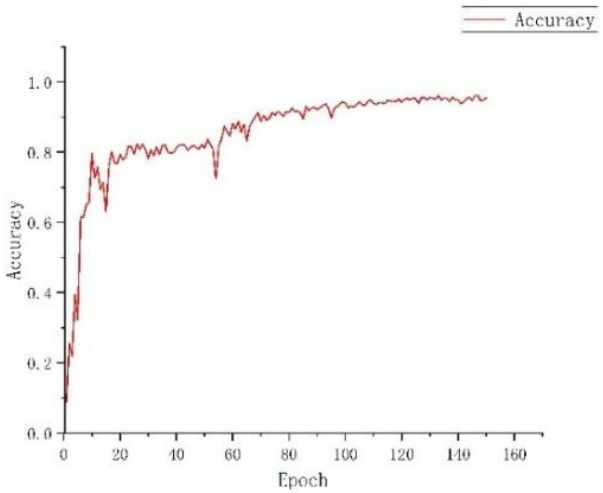


Figure 9. Loss value curve

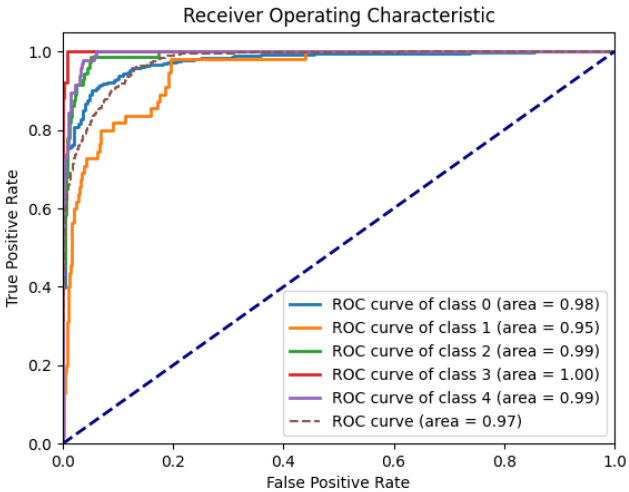


Figure 10. ROC curve of ANLR-Net model

5. CONCLUSION

This paper proposes a lightweight and novel feature extraction network. First, we use the UNet model to segment the data and extract effective regions. Then, data augmentation is performed to enrich the dataset, and the CLAHE method is applied to enhance image contrast, improving image quality and feature distinguishability. In the feature extraction phase, we modify the traditional convolutional layers in the DenseNet model by replacing the traditional linear weight matrices with learnable Legendre polynomials for adaptive feature mapping. This modification not only effectively reduces the number of network parameters but also enhances the efficiency of feature learning, enabling the model to more accurately capture the global structures and complex local details in ultra-widefield retinal images. Furthermore, this design effectively reduces the risk of overfitting.We also introduce a bottleneck structure to further reduce the parameter count and improve computational efficiency, optimizing resource usage while maintaining model performance. Additionally, the incorporation of dilated convolutions expands the receptive field, allowing the model to extract more comprehensive feature information. Ultimately, the model achieves automated diagnosis and classification of ultra-widefield retinal images. Comprehensive experimental validation on both the ultra-widefield dataset and the Kaggle EyePACS dataset demonstrates the robustness and generalization capabilities of the model across different datasets.The ANLR-Net model achieved excellent results on both datasets: an average accuracy of 95.05% and an average specificity of 98.64% on the ultra-widefield dataset, and an average accuracy of 92.30% and an average specificity of 95.75% on the Kaggle EyePACS dataset. In both datasets, the improved model outperformed traditional models in terms of average accuracy and specificity, verifying its effectiveness and accuracy in diabetic retinopathy detection and classification.Future research will focus on further optimizing the model architecture by introducing more efficient computational methods, such as model pruning and quantization techniques, to reduce computational complexity while maintaining high accuracy. Additionally, we will expand the diversity of the dataset by including more types of retinal images to enhance the model's adaptability and robustness in different scenarios. We anticipate that these improvements will advance the early diagnosis and automated classification of diabetic retinopathy, providing more efficient and accurate tools for clinical practice.

ACKNOWLEDGEMENTS

Shanxi Province Natural Science Foundation: Research on Core Scalable and Assurance Intelligent Service Model for Cross-Network Virtual Reality (201801D121147)

REFERENCES

[1] CHENG H-T, XU X, LIM P S, et al. Worldwide epidemiology of diabetes-related end-stage renal disease, 2000–2015 [J]. 2021, 44(1): 89-97.

[2] GUO L, XIAO X J A M. Guideline for the Management of Diabetes Mellitus in the Elderly in China (2024 Edition) [J]. 2024, 7(1): 5-51.

[3] TOMIĆ M, VRABEC R, LJUBIĆ S, et al. Patients with Type 2 Diabetes, Higher Blood Pressure, and Infrequent Fundus Examinations Have a Higher Risk of Sight-Threatening Retinopathy [J]. 2024, 13(9): 2496.

[4] ZHAO M, CHANDRA A, LIU L, et al. Investigation of the reasons for delayed presentation in proliferative diabetic retinopathy patients [J]. 2024, 19(2): e0291280.

[5] LI H, JIA W, VUJOSEVIC S, et al. Current Research and Future Strategies for the Management of Vision-Threatening Diabetic Retinopathy [J]. 2024: 100109.

[6] GREEN J, SIDDALL H, MURDOCH I J S S, et al. Learning to live with glaucoma: a qualitative study of diagnosis and the impact of sight loss [J]. 2002, 55(2): 257-67.

[7] ZHANG Z, DENG C, PAULUS Y M J B. Advances in Structural and Functional Retinal Imaging and Biomarkers for Early Detection of Diabetic Retinopathy [J]. 2024, 12(7): 1405.

[8] SORRENTINO F S, GARDINI L, FONTANA L, et al. Novel Approaches for Early Detection of Retinal Diseases Using Artificial Intelligence [J]. 2024, 14(7): 690.

[9] IKRAM A, IMRAN A, LI J, et al. A systematic review on fundus image-based diabetic retinopathy detection and grading: Current status and future directions [J]. 2024.

[10] NIGAM A, SUN J, SUBHASH V, et al. Deep Learning Approach to Identify Diabetic Retinopathy Severity and Progression Using Ultra-Wide Field Retinal Images [M]. AI for Health Equity and Fairness: Leveraging AI to Address Social Determinants of Health. Springer. 2024: 103-16.

[11] HEMANTH D J, DEPERLIOGLU O, KOSE U J N C, et al. RETRACTED ARTICLE: An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network [J]. 2020, 32(3): 707-21.

[12] KANDEL I, CASTELLI M J A S. Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review [J]. 2020, 10(6): 2021.

[13] SHANKAR K, SAIT A R W, GUPTA D, et al. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model [J]. 2020, 133: 210-6.

[14] HUANG Y, LIN L, CHENG P, et al. Lesion-based contrastive learning for diabetic retinopathy grading from fundus images; proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24, F, 2021 [C]. Springer.

[15] HOU J, XU J, XIAO F, et al. Cross-field transformer for diabetic retinopathy grading on two-field fundus images; proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), F, 2022 [C]. IEEE.

[16] SOLANO A, DIETRICH K N, MARTíNEZ-SOBER M, et al. Deep learning architectures for diagnosis of diabetic retinopathy [J]. 2023, 13(7): 4445.

[17] ZHANG G, SUN B, CHEN Z, et al. Diabetic retinopathy grading by deep graph correlation network on retinal images without manual annotations [J]. 2022, 9: 872214.

[18] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2017 [C].

[19] SAMAD S A, GITANJALI J J I A. Augmenting DenseNet: Leveraging Multi-Scale Skip Connections for Effective Early-Layer Information Incorporation [J]. 2024.

[20] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation; proceedings of the Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, F, 2015 [C]. Springer.