Computer Assisted Grade Diagnosis Using Feature Fusion and Network Algorithm Flow

Yongqi Nie, Chong Ma, Peng Qiu, Hui Cao*
School of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China

*Corresponding author Email: yjsc@sduetcm.edu.cn

Abstract. Diabetic retinopathy is one of the important causes of visual impairment and blindness. In order to strengthen the early prevention and treatment of diabetic retinal diseases and prevent the occurrence of serious diseases, feature fusion diagnosis model that combines traditional features and depth features was proposed. In the diagnostic model, four kinds of traditional algorithms were used to extract imaging features, and VGG16 and VGG19 networks were used to extract deep features. Then k-nearest neighbor, random forest, support vector machine and extreme learning machine were used to grade the degree of diabetic retinopathy based on the fused feature vectors. The diagnostic model adopted the depth feature extracted by VGG19 network and the traditional feature fusion method to achieve the best grade effect on the support vector machine, with the accuracy of 84.44%, which was better than comparative feature fusion and classifier combination methods, and also performed well in the precision rate and recall rate. The diagnostic model based on the fusion of traditional features and depth features provides a new detection method for the grading prediction of diabetic retinopathy, and at the same time improves the accuracy of clinical diagnosis.

1. Introduction
Diabetic retinopathy (DR) is a chronic disease that spreads all over the world, and it is also the main cause of visual impairment and blindness in the world today [1]. According to statistics from the World Health Organization (WHO), in the next two decades, global health problems related to diabetes will affect 700 million people, since one third of diabetic patients have potential DR, this would cause about 250 million people worldwide to suffer from DR [2]. If DR is not treated in a timely manner, it may lead to a sharp decline in the patient's vision at a later stage, and eventually lead to blindness. According to research, the longer you have diabetes, the higher your risk of diabetic retinopathy, but effective early prevention and treatment can reduce the incidence of patients by 90% [3]. With the increase in the amount of DR screening, the overloaded reading work has caused great pressure on doctors, and the manual reading speed is slow, unable to provide timely feedback on the patient's condition, and the number of misdiagnosis is increasing [4]. If in the early screening of DR, fundus images can be used for automatic grade, it can save doctors' diagnosis time and reduce the number of missed diagnoses.

At this stage, machine learning and deep learning methods have been widely used in DR detection, and have achieved good classification results [5]. In recent years, "feature expression + classifier" has been widely used in target detection and classification tasks [6]. Retina classification based on color
fundus images is also applicable. The key point is efficient and accurate feature extraction and construction of a reasonable classifier. Traditional feature extraction methods require experts to design features for a specific lesion based on prior knowledge. Although they are extremely representative, they are time-consuming to implement, such as scale-invariant feature transform (SIFT) and speeded up robust features (SURF). In addition, in the pattern recognition, k-nearest neighbor (KNN), random forest (RF), support vector machine (SVM) and extreme learning machine (ELM) are widely used in fields such as portrait recognition and text classification. Kamil et al. [7] used the bag of word (BOW) technology to extract SURF features, and selected SVM with RBF kernel as the classifier to complete the four-class task of retinopathy and achieve an accuracy of 0.94 at the same time. Ramon et al. [8] also used the BOW technique to extract the global features of the entire image, and extracted the SIFT and SURF features to represent the region of interest, and after fusing the two, the SVM classifier was used to achieve the two classification of normal and diseased fundus retinal images. Experiments have proved that the method of feature fusion has achieved better classification accuracy than single feature, and can achieve better classification results. With the development of deep learning, the feature extraction method based on deep learning algorithm avoids the cumbersome problem of manually designing features and provides a good solution for DR detection. Abroad, the deep learning system developed by Abrámoff et al. [9] to diagnose DR lesions was the first to be approved by the US Food and Drug Administration. Google [10] was based on the Inception V3 model to realize the judgment research of retinopathy referral. Domestically, Ding et al. [11] independently designed the CompactNet to classify the degree of diabetes after preprocessing the diabetic retinal image, and compared it with the LeNet and AlexNet classic networks. The classification accuracy rates under the three networks reached 69%, 42% and 62%.

Based on the research of traditional algorithms and deep learning networks, this paper used five kinds of fundus retinal images with different DR lesions as samples, fuses traditional features and deep features, and used the fusion features to grade the retinal fundus image lesions.

2. Material and Method

2.1. Diabetic retina dataset

This article used the Kaggle-diabetic-retinopathy-224x224-gaussian-filtered data set, which contained high-resolution color retinal images of the fundus of multiple patients. All images were processed by Gaussian filtering and the size of all pictures was fixed to 224*224 resolution. The data set was divided into five grades: NO_DR, mild, moderate, severe and Proliferate_DR according to the degree of DR lesions. The lesion features are mainly manifested in hard exudates, bleeding points, and microvascular aneurysms. Label 0 indicates healthy and normal state, label 4 indicates the most severe state of the disease, and the fundus images of the retina dataset with different disease severity are shown in Fig. 1.

![Fig. 1 Retina images with different degrees of disease](image)

2.2. Feature extraction

The traditional feature extraction algorithm needs to be designed according to the characteristics of the detected object in the image. It mainly extracts the color, boundary, shape, texture and other features of the target area, and it is relatively weak to reflect the essential features of the image. At the same time, due to the complexity and diversity of medical image information, traditional features cannot accurately and comprehensively extract the information in the image, and the evaluation ability is limited. With the
emergence of the ImageNet Challenge, deep learning has entered the public eye. It simulates the process of human brain data analysis, extracts nonlinear relationships in massive data, automatically generates features that reflect image information, and effectively solves the problems of intra-class variation and inter-class similarity. But deep learning also has its limitations. For example, the feature representation caused by the deep learning framework structure is not comprehensive and it is difficult to take into account global information. As a result, traditional features and deep features can be merged to increase the diversity of feature expression and obtain more comprehensive features of retinal DR lesions. In this paper, 117-dimensional traditional features were extracted from each gray-scale DR image, including first-order statistical features, LBP features, Gabor features and gray-level co-occurrence matrix features; 4096-dimensional depth features were extracted from VGG16 and VGG19 networks respectively.

(1) Traditional features

Before the feature extraction of the DR lesion image, the color image was pre-processed. Firstly, the RGB image was converted into a grayscale image, as shown in the Fig.2.

![Fig. 2 Image pre-processing](image)

The feature extraction of DR lesions based on traditional algorithms is a key step in grading the severity of lesions. By extracting massive features from images to quantify chronic diseases such as DR lesions, it can effectively solve the problem that the degree of lesions is difficult to quantitatively evaluate. There was a total of 117-dimensional traditional features such as first-order statistical features, local binary patterns (LBP), Gabor features and gray-level co-occurrence matrix (GLCM). Specific characteristics are described in Table. 1.

| Type                        | Feature dimension | Feature description                                                                                     |
|-----------------------------|-------------------|--------------------------------------------------------------------------------------------------------|
| First-order statistical     | 6                 | Mean, standard deviation, variance, skewness, kurtosis and information entropy                           |
| features                    |                   | Equivalent LBP: When the cyclic binary number corresponding to a certain LBP has at most two transitions from 0 to 1 or from 1 to 0, the binary is defined as an equivalent mode class, and other modes are defined as mixed mode classes. The feature vector extracted from 8 sampling points in the 3*3 neighborhood is reduced from 256 dimensions to 59 dimensions |
| Texture feature             | 59                |                                                                                                        |
| Gabor                       | 48                | Gabor: According to the characteristics of the Gabor filter, design a set in four different scales (ν=0, π/6, π/3, π/2, 2π/3, 5π/6) Gabor filter bank, convolve the input gray image with the time domain and frequency domain of each filter, and finally obtain the mean and variance of the gray value of the image |
| GLCM                        | 4                 | GLCM: Contrast, energy, entropy, homogeneity                                                           |
(2) Depth characteristics
The data set used in this experiment was processed by Gaussian filtering, and the picture size was uniformly set to 224*224 pixels. The input size of neural networks such as the VGG series [12] is 224*224, and the image does not need to be reduced in size before input, so it will not lose part of the image information. Because the VGG series network uses multiple smaller convolution kernels, this makes the receptive field of the network layer smaller, which will affect the judgment of the image's global features. However, the image information extracted by traditional algorithms is mostly global features and is less affected by image size. Therefore, traditional features can complement the retinopathy features extracted by the VGG series network to make the feature expression more perfect. The depth features of this experiment were based on the VGG19 and VGG16 network extraction. Fig. 3 shows the feature extraction process of the fundus retinal image in the VGG19 network.

![Feature map changes of the retina image in the VGG19 network](image)

2.3. Feature Fusion
There have been many researches using traditional methods and deep learning methods for fundus retinal image classification. However, there are not many studies on the combination of the two methods, and the methods of feature fusion are also different. The fusion methods are mainly additive fusion and cascade fusion and maximum fusion. The core of the combination of traditional methods and deep learning methods is how to effectively and reasonably perform feature fusion. It is particularly critical to select a suitable fusion location and fusion method. In this paper, the traditional feature of a single image extracted by traditional algorithms was a one-dimensional feature vector, while in the VGG series network, only the fully connected layer was a one-dimensional feature vector. Therefore, the fully connected layer was selected as the location for feature fusion. The traditional feature extracted by the traditional algorithm was a one-dimensional feature vector of 117 dimensions, and the feature extracted by the convolutional neural network was a one-dimensional feature vector of 4096 dimensions. Because the two feature extraction methods are different but the functions are the same, the features extracted by the two different methods can be retained more completely by using the cascade fusion method. The specific feature dimensions and fusion process are shown in Fig.4.
2.4. Feature selection

As the step of feature pre-processing, firstly, the feature data set was normalized to [0,1] to avoid the features in the small value interval from being dominated by the features in the large value interval [13]. In order to improve the calculation speed, UMAP was used to reduce the dimensionality of the feature data. UMAP algorithm is one of the most commonly used nonlinear dimensionality reduction methods. It measures the low-dimensional optimization process through cross-entropy, so that it has the fuzzy topological representation as close as possible to the construction. UMAP operation steps are as follows:

Step 1. For a specific high-dimensional spatial distribution, use the nearest neighbor algorithm to obtain the k nearest neighbor set of each high-dimensional data;

Step 2. Identify for each \( x_i \);

Step 3. Select cross entropy as the cost function;

Step 4. Optimize the fuzzy topology of the low-dimensional representation to obtain the final reduced-dimensional matrix.

2.5. Construction of diagnostic model

First, the principal component analysis method was used to determine the target embedding dimension of the fusion feature, and the feature dimensionality reduction was performed through UMAP, and then the normalized feature data set was randomly divided into a training set (70%) and a test set (30%), and finally based on the training set of fusion features, four different diagnostic models were established by using the classifier for verification on the test set. The classifiers are as follows: KNN, RF, SVM and ELM. KNN finds the K samples most similar to the data of a single sample, and determines the classification of a single sample according to the category of most of the data in the k sample. RF uses the idea of ensemble learning to integrate the basic unit of decision tree into a forest for classification [14]. SVM uses kernel functions to map feature data from a lower-dimensional space to a high-dimensional feature space for classification according to the mapping space measure [15]. ELM is a single hidden layer neural network, which improves the algorithm speed on the basis of maintaining the learning accuracy [16]. The algorithm flow of this paper is shown in Fig.5.
2.6. Model evaluation criteria

This experiment was done on Windows 10, python3.6, Opencv2, Inter(R) Core (TM) i9-9820X CPU and 128GB RAM, NVIDIA GeForce RTX 2080 Ti GPU environment. In order to evaluate the effectiveness of the feature fusion method and diagnosis model, Precision (P), Recall (R) and Accuracy(A) were selected for evaluation, as follows:

\[ P = \frac{TP}{TP + FP} \]  
\[ R = \frac{TP}{TP + FN} \]  
\[ A = \frac{TP + TN}{TP + TN + FP + FN} \]

3. Test Results and Discussions

Principal component analysis was used to extract the feature dimensions of VGG16+traditional feature vector and VGG19+traditional feature vector whose cumulative contribution rate were greater than 95%. The target embedding dimensions after dimensionality reduction were 102 and 104 dimensions respectively, and then UMAP was used to reduce the dimensionality of the data. In order to maintain a good non-linear relationship in the data after dimensionality reduction, the neighbors and min_dist parameters selected for different classifiers are shown in Table 2.

| Parameter | VGG16 | VGG19 |
|-----------|-------|-------|
| neighbors | Min_dist | neighbors | Min_dist |
| KNN       | 10    | 0.5   | 30     | 0.2 |
| RF        | 10    | 0.3   | 30     | 0.3 |
| ELM       | 20    | 0.1   | 20     | 0.1 |
| SVM       | 25    | 0.01  | 10     | 0.5 |
In order to improve classification accuracy, various model parameters were optimized. KNN parameter setting: K value was an even value between 0 and 36. RF parameter setting: the number of decision tree setting was between 50 and 2000 decision trees, with 50 trees as the interval. SVM parameter setting: Linear kernel function would be used to perform 50% cross validation through network search method. ELM parameter setting: the number of nodes in the hidden layer was between 100 and 2000, with 50 nodes as the interval. Fig.6 (a) shows the impact of K value size on KNN classification performance. Fig.6 (b) shows the impact of decision trees on the performance of RF classification performance. Fig.6 (c) shows the impact of the number of hidden layer nodes on ELM classification performance. The optimal parameters when the model reaches the highest accuracy are shown in Table. 3.

**Fig. 6 Model parameter optimization**

**Table. 3** Optimal parameters

| feature                | Evaluation index | classifier      |  |  |  |
|------------------------|------------------|-----------------|---|---|---|
| traditional            | Optimal parameters | KNN | 28 | 700 | 300 |
| VGG16+ traditional     | Optimal parameters | RF  | 19 | 1700 | 700 |
| VGG19+ traditional     | Optimal parameters | ELM | 18 | 1500 | 400 |
| UMAP_ VGG16+ traditional | Optimal parameters |       | 31 | 900  | 100 |
| UMAP_ VGG19+ traditional | Optimal parameters |       | 23 | 850  | 100 |
After parameter optimization of the model, the prediction P, R and A of KNN, RF, SVM and ELM models based on traditional features, VGG16+ traditional features, VGG19+ traditional features, UMAP_VGG16+ traditional features and UMAP_VGG19+ traditional features are shown in the Table 4.

Table 4 grade results of diagnosis model

| Model | Evaluation index | Traditional | VGG16+ traditional | VGG19+ traditional | UMAP_VGG16+ traditional | UMAP_VGG19+ traditional |
|-------|------------------|-------------|--------------------|--------------------|------------------------|------------------------|
| KNN   | P                | 0.7129±2.21 | 0.7528±2.25        | 0.7751±2.27        | 0.7757±2.68            | 0.7785±1.49            |
|       | R                | 0.7543±2.45 | 0.7944±2.55        | 0.7862±2.50        | 0.8025±2.75            | 0.7989±1.52            |
|       | A                | 0.7470±1.25 | 0.7907±1.69        | 0.7889±1.32        | 0.7980±1.73            | 0.7925±0.84            |
| RF    | P                | 0.7008±2.43 | 0.7874±2.55        | 0.8026±2.32        | 0.7685±2.87            | 0.7660±1.84            |
|       | R                | 0.7416±2.57 | 0.7916±2.74        | 0.8016±2.78        | 0.7853±2.98            | 0.7843±2.51            |
|       | A                | 0.7585±1.69 | 0.8053±1.72        | 0.8098±1.53        | 0.7939±1.85            | 0.7798±1.63            |
| SVM   | P                | 0.7495±3.10 | 0.8445±3.20        | 0.8258±3.48        | 0.7042±3.48            | 0.7395±2.39            |
|       | R                | 0.7461±3.01 | 0.8462±2.94        | 0.8289±3.54        | 0.7461±3.57            | 0.7907±3.15            |
|       | A                | 0.7548±1.98 | 0.8053±1.82        | 0.8444±2.31        | 0.8016±1.99            | 0.7970±1.89            |
| ELM   | P                | 0.6902±2.52 | 0.7407±2.54        | 0.7590±2.38        | 0.7443±2.59            | 0.7460±1.98            |
|       | R                | 0.7307±2.63 | 0.7561±2.85        | 0.7898±2.36        | 0.7944±2.81            | 0.7753±2.05            |
|       | A                | 0.7493±1.88 | 0.7776±1.76        | 0.7731±1.67        | 0.7900±1.85            | 0.8027±1.25            |

As can be seen from Table 4, in the diagnostic model, the feature fusion method is superior to the traditional feature fusion method in all the indexes of the four diagnostic models. Among them, the grade effect based on VGG16 network feature fusion is better than that based on VGG19 network feature fusion when combined with KNN and ELM, while the grade accuracy is slightly lower when combined with RM and SVM classifier. This indicates that different feature fusion methods will directly affect the grading effect of the diagnostic model. Secondly, it also can be seen from the table, based on feature fusion VGG16 network and VGG19 network data after UMAP dimension reduction, grade in the KNN, RF and ELM classifier is better than before data dimension reduction, and grade in SVM classifier effect was poorer. This shows that the combination of different feature dimensions + classifier will have an impact the result of the grade. Among them, the grade effect of KNN and RF diagnosis model is the best after dimension reduction of fusion feature based on VGG16 network. After dimensionality reduction based on VGG19 network feature fusion, the grade effect in ELM diagnostic model is the best. This shows that feature dimensionality reduction can not only speed up operation, but also remove redundant data and improve grade accuracy.

Compared with the grade effects of the same classifier under different features in Table 4, the UMAP_VGG16+ traditional features have the best effect in KNN classification. VGG19+ traditional features have the best effect in both RF and SVM classification, followed by VGG16+ traditional features have better grade effect in these two classifiers. UMAP_VGG19+ traditional features have the best grade effect in ELM classification. The highest-grade accuracy of KNN, RF, SVM and ELM fusion models was 79.80%, 80.98%, 80.44% and 80.27%, respectively. Meanwhile, by comparing the grade effect of the same feature in the table under different classifiers, the traditional feature in RF has the best grade effect. VGG16+traditional features, VGG19+traditional features, UMAP_VGG16+ traditional features and UMAP_VGG19+ traditional features have the best grade effect in SVM. Based on traditional features, VGG16+ traditional features, VGG19+ traditional features, UMAP_VGG16+ traditional features and UMAP_VGG19+ traditional feature fusion model, the highest-grade accuracy was 75.85%, 80.53%, 84.44%, 80.16% and 80.27%, respectively.
In order to further evaluate the performance of the model, the best results in the literature with the same five-level classification of the degree of disease were compared. Among them, LeNet, AlexNet and CompactNet were the deep learning networks adopted in literature [11], the experimental results are shown in Table 5. Comparing the accuracy of the various diagnostic models in the table, it can be seen that the "multi-feature fusion + classifier" method have a better classification effect and is practical.

| diagnostic models          | accuracy |
|---------------------------|----------|
| LeNet                     | 42%      |
| AlexNet                   | 62%      |
| CompactNet                | 69%      |
| VGG19+ traditional +SVM   | 84%      |

4. Conclusion
Based on the existing image classification methods, this paper summarized and improved the grade method of diabetic retinopathy, combined the traditional feature method of artificial design and the deep learning method, fused the traditional retinal features and the deep retinal features, and established a variety of diagnosis models to obtain better grade results. At the same time, the results find that: the diagnosis model of feature fusion have achieved better grade results. Under different dimensions and different convolutional neural networks, the manually designed traditional features and the depth features extracted by the deep learning framework have different grade effects under different diagnostic models, but the four feature fusion models based on the SVM classifier have achieved the best grade effect. Finally, compared with other diagnostic models, the grade effect of feature fusion model is improved to a certain extent. If there are better feature fusion methods and diagnostic models, and the characteristics of fusion are fully utilized, the grade effect of diagnostic models will be further improved. Therefore, in the next research work, we will design and extract more different traditional features, constantly improve the deep learning network, and explore more methods suitable for the classification of diabetic retinopathy degree and disease classification references.

Acknowledgments
The corresponding author of this article is Professor Hui Cao. This work was funded by the National Natural Science Foundation (No. 81973981, No. 81473708).

References
[1] Liang Ping, Xiong Biao, et al. Classification methods for diabetic retinopathy fromretinal images [J]. Journal of Shenzhen University (Science & Engineering), 2017, 34(3):290-299 (in chinese).
[2] Ni Jiayuan. Research and design expert system for diabetes diagnosis diagnosis based on decision tree [D]. Hefei: Anhui University, 2016 (in chinese).
[3] Doshi D, Shenoy A, et al. Diabetic retinopathy detection using deep convolutional neural networks [C]// International Conference on Computing, Analytics and Security Trends. IEEE, 2017:261-266.
[4] Yang Ling, Shen Xi. Correlation between Diabetic Retinopathy and Dry Eye. International Journal of Ophthalmology, 2018, 18(4): 744-747 (in chinese).
[5] Sun Yuchen, Liu Yuhong, et al. Diagnosis of diabetic retinopathy based on deep learning [J/OL]. Progress in Laser and Optoelectronics, 2020: 1-14 (in chinese).
[6] Guo Zhiqiang, Hu Yongwu, et al. Outdoor weather image classification based on feature fusion [J]. Computer Applications, 2020, 40(04): 1023-1029 (in chinese).
[7] Kamil R, Al-Saedi K, et al. (2018) An accurate system to measure the diabetic retinopathy using svm classifier. Ciência e Técnica Vitivinícola 33:135–139.
[8] Lambin P, Rios-Velazquez E, et al. Radiomics: Extracting more information from medical images using advanced feature analysis [J]. European Journal of Cancer, 2007, 43(4): 441-446.
[9] Abrámoff MD, Lavin PT, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018; 1(1): 39.
[10] Gulshan V, Peng L, et al. Development And Validation of A Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. Jama,2016, 316(22): 2402-2410
[11] Ding Pengli, Li Qingyong, et al. Diabetic retinal image classification method based on deep neural network [J]. Journal of Computer Applications,2017,37(03):699-704(in chinese).
[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL] [2019-04-10]. https://arxiv.org/pdf/1409.1556.pdf.
[13] Ferreira J, José R, et al. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases [J]. Computer Methods and Programs in Biomedicine, 2018, 159:23-30.
[14] Breiman L, Breiman L, et al. Random Forests Machine Learning [J]. Journal of Clinical Microbiology, 2001, 2:199-228.
[15] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20(3):273-297.
[16] Huang GB, Zhu QY, et al. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70(1): 489-501.