A deep learning based fine-grained classification algorithm for grading of visual impairment in cataract patients*

JIANG Jiewei[1], ZHANG Yi[1], XIE He[2], YANG Jingshi[1], GONG Jiamin[1], and LI Zhongwen[3]**

1. School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

2. School of Ophthalmology and Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325000, China

3. Ningbo Eye Hospital, Wenzhou Medical University, Ningbo 315000, China

(Received 20 March 2023; Revised 12 July 2023)

Recent advancements in artificial intelligence (AI) have shown promising potential for the automated screening and grading of cataracts. However, the different types of visual impairment caused by cataracts exhibit similar phenotypes, posing significant challenges for accurately assessing the severity of visual impairment. To address this issue, we propose a dense convolution combined with attention mechanism and multi-level classifier (DAMC_Net) for visual impairment grading. First, the double-attention mechanism is utilized to enable the DAMC_Net to focus on lesions-related regions. Then, a hierarchical multi-level classifier is constructed to enhance the recognition ability in distinguishing the severities of visual impairment, while maintaining a better screening rate for normal samples. In addition, a cost-sensitive method is applied to address the problem of higher false-negative rate caused by the imbalanced dataset. Experimental results demonstrated that the DAMC_Net outperformed ResNet50 and dense convolutional network 121 (DenseNet121) models, with sensitivity improvements of 6.0% and 3.4% on the category of mild visual impairment caused by cataracts (MVICC), and 2.1% and 4.3% on the category of moderate to severe visual impairment caused by cataracts (MSVICC), respectively. The comparable performance on two external test datasets was achieved, further verifying the effectiveness and generalizability of the DAMC_Net.

Document code: A **Article ID:** 1673-1905(2024)01-0048-10

DOI https://doi.org/10.1007/s11801-024-3050-4

Cataract is a prevalent visual impairment disease with a high risk of blindness, primarily characterized by progressive loss of vision. According to statistics, cataract-induced blindness affects a substantial number of individuals, ranging from 16 million to 21 million worldwide[1], accounting for approximately 60% of all cases of blindness. Notably, the incidence of cataract-related blindness is higher in populations with lower socioeconomic status and in developing countries[2]. Therefore, early detection and prompt treatment of cataract have become pressing priorities that require immediate attention.

Cataract surgery is a routine ophthalmological procedure and remains the most effective treatment for cataract[3]. The standard surgical approach involves cataract phacoemulsification combined with intraocular lens implantation[4,5]. However, certain clinical criteria restrict the eligibility of patients for cataract surgery, and it is reserved for individuals who meet specific conditions[3]. First of all, visual acuity requirements must be met, typically necessitating a best-corrected visual acuity of less than 0.3. Additionally, there should be no active inflammation in the eyes, such as conjunctivitis or keratitis, and any systemic diseases like diabetes or hypertension must be carefully monitored and controlled during the surgical procedure. On the contrary, patients with a best-corrected visual acuity equal to or greater than 0.3 are recommended to undergo conservative drug treatment instead. Therefore, accurately assessing the degree of visual impairment during the diagnosis stage is of utmost importance in determining the appropriate treatment plan. According to the aforementioned criteria[6], the degree of visual impairment caused by cataracts can be categorized into two groups: mild visual impairment caused by cataracts (MVICC) with a best-corrected visual acuity equal to or greater than 0.3, and moderate to severe visual impairment caused by cataract (MSVICC) with best-corrected visual acuity less than 0.3.

The grading of visual impairment in cataract patients usually requires a skilled ophthalmologist to examine

* This work has been supported by the National Natural Science Foundation of China (Nos.62276210, 82201148 and 61775180), and the Natural Science Basic Research Program of Shaanxi Province (No.2022JM-380).

** E-mail: zhangyi03110214@163.com

patients' best corrected visual acuity by methods of a routine vision examination. However, there is current and expected future shortfall of ophthalmologists in both developing and developed countries[7]. In particular, in developing countries like China, there are large differences in medical resources among different regions. More cataract patients in remote areas may become blind due to a lack of timely treatment. Moreover, the grading of visual impairment by ophthalmologists has several drawbacks, including being time-consuming, labor-intensive, and susceptible to subjective factors.

Recent developments in deep learning algorithms have resulted in the widespread applications of artificial intelligence (AI) in various healthcare domains, including disease diagnosis, medical image segmentation, and lesion localization[8-11]. In particular, the combination of deep learning with medical big data has provided potential opportunities for the automatic diagnosis and grading of eye disease. In the field of ophthalmology, numerous studies have successfully developed high-accuracy AI system using fundus images for automatic disease screening, including diabetic retinopathy, glaucoma, retinal exudation, and retinal hemorrhage[12-14]. In addition, there have been investigations utilizing anterior segment images for disease diagnosis and severity grading, such as keratitis, pterygium, and eye tumor[15,16]. Several studies have specifically focused on the development of deep learning-based systems for the automatic diagnosis and grading of cataracts[17-19]. However, compared to previous studies, the grading of visual impairment based on fundus images presents unique characteristics. The MVICC and MSVICC categories share many similarities in phenotypes, which easily result in a lower sensitivity in the diagnostic system. Furthermore, due to the turbidity of the lens of cataract patients, the varying degrees of damage to blood vessels and optic disc tissues, exacerbating the difficulty of grading visual impairment. Especially for patients with visual acuity around 0.3, the distinguishability is even lower.

To address the above-mentioned issues, this study proposes a dense convolution combined with attention and multi-level classification network (DAMC_Net) for fine-grained assessment of visual impairment in cataract patients. First, the dense convolution combined with double-attention network is proposed to extract fundus features from input images. At the same time, to optimize model parameters and reduce calculation time, the depthwise separable convolution is employed instead of the traditional 3×3 convolution in the dense layer. Second, by constructing two-level classification tasks and assigning different weight coefficients, the extracted features of visual impairment are classified. The first-level classifier distinguishes between normal and cataract, while the second-level classifier differentiates among normal, MVICC, and MSVICC. Notably, a cost-sensitive is employed in the second-level classifier to address the problem of high false negative rates caused by training with imbalanced dataset. Lastly, this study explores and compares the impact of different weight coefficients of two tasks on the performance of the DAMC_Net for fine-grained assessment of normal, MVICC, and MSVICC.

In this study, a total of 7 686 fundus images derived from routine examinations between April 2019 and September 2022 at Zhejiang Eye Hospital at Wenzhou (ZEHWZ) were used to develop the DAMC_Net. An additional external test set comprising 1 398 fundus images were collected from Ningbo Eye Hospital (NEH) and Zhejiang Eye Hospital at Hangzhou (ZEHHZ). Each fundus image was examined, discussed, and labeled by three experienced ophthalmologists, and then classified into three categories: normal, MVICC, and MSVICC. The fundus images from the ZEHWZ dataset were randomly divided into three groups: 70% for training (5 354 images), 15% for validation (1 158 images), and 15% for the internal test dataset (1 174 images). The training and validation datasets were employed to develop the DAMC_Net and the internal test dataset was used to evaluate its performance. Further details regarding the datasets obtained from ZEHWZ, NEH, and ZEHHZ are summarized in Tab.1.

Tab.1 Distribution of cataract fundus images

Type	ZEHWZ dataset			NEH dataset	ZEHHZ dataset
	Train	Val	Test		
Normal	2 120	461	477	107	405
MVICC	1 933	425	417	103	560
MSVICC	1 301	272	280	91	132
Total	5 354	1 158	1 174	301	1 097

As shown in Fig.1, the overall framework of the DAMC_Net primarily consists of three stages: input image preprocessing, feature extraction, and multi-level classification. In stage I, data augmentation techniques[15,20] are employed to preprocess the input fundus images, including random cropping, random rotations around the image center, and data normalization. These preprocessing techniques are beneficial in enhancing the diversity of the dataset, improving the network convergence speed, and preventing overfitting and bias problems during training. In stage II, the proposed dense convolution combined with the double-attention network is utilized to extract meaningful fundus features from the input images. In the stage III, the extracted features are processed by multi-level classifiers to achieve a fine-grained assessment of visual impairment in cataract patients.

The fundus images contain both underlying texture features and high-level semantic information, which are crucial for accurately assessing visual impairment in cataract patients. To improve the accuracy of this grading task, we introduce the double-attention mechanism in the convolutional neural network. The dense convolutional network (DenseNet)[21] is chosen as backbone network due to its exceptional performance with reduced computations and increased effectiveness. In this study, the double-attention mechanism[22] is further applied in the DenseNet to enhance the performance and generalization ability of fine-grained assessment of visual impairment. Specifically, a channel attention module is embedded before the depthwise separable convolution following the 1×1 convolution operation, and a spatial attention module is concatenated after the depthwise separable convolution, as depicted in Fig.1.

The channel attention map is generated by leveraging the inter-channel relationship of features. Each channel in the feature map is considered as a feature detector, enabling the channel attention to focuses on ‘what’ is meaningful given an input image. To compute the channel attention efficiently, we squeeze the spatial dimension of the input feature map. Specifically, as shown in Fig.2(a), the spatial information of the input feature map is aggregated using both average-pooling and max-pooling operations, resulting in the generation of two distinct spatial context descriptors: $F^{\mathrm{c}}_{\mathrm{avg}}$ and $F^{\mathrm{c}}_{\max}$. Both descriptors are then fed into a shared network to produce our channel attention map M_{c}. The shared network is composed of multi-layer perceptron (MLP) with one hidden layer. After applying the shared network to each descriptor, the output feature vectors are merged using element-wise summation followed by a sigmoid activation function. The channel attention can be formalized as

$$M_{\mathrm{c}}(F)=\sigma(\mathrm{MLP}(\mathrm{AvgPool}(F))+\mathrm{MLP}(\mathrm{MaxPool}(F)))=$$
$$\sigma(W_1(W_0(F^{\mathrm{c}}_{\mathrm{avg}}))+W_1(W_0(F^{\mathrm{c}}_{\max}))), \qquad (1)$$

where σ, AvgPool, MaxPool, F denote the sigmoid function, average-pooling operation, max-pooling operation, and input feature map, respectively. W_0 and W_1 are the weights of MLP.

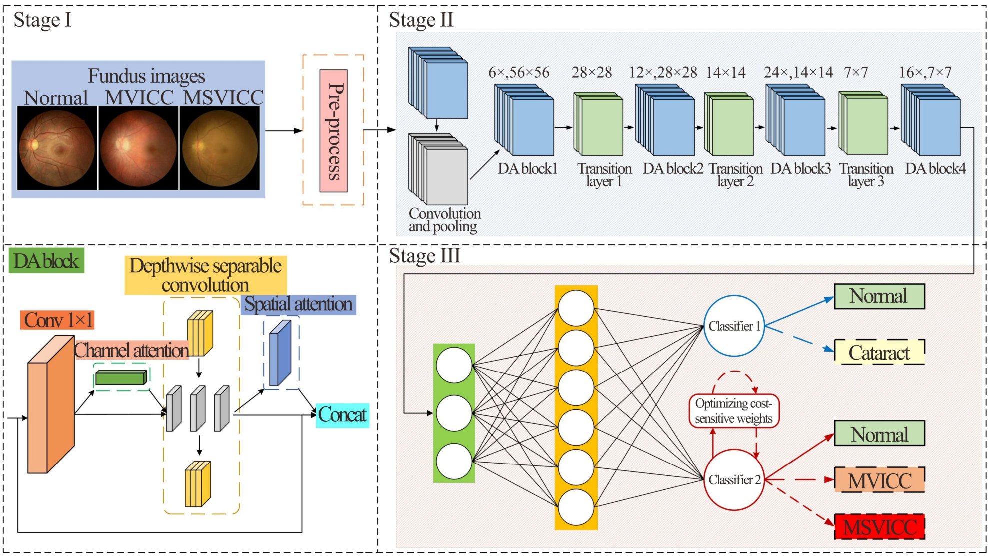


Fig.1 The overall framework of the DAMC_Net

The spatial attention map is generated by utilizing the inter-spatial relationship of features. Unlike the channel attention, the spatial attention focuses on ‘where’ is an informative part, which is complementary to the channel attention. As shown in Fig.2(b), max-pooling and average-pooling are employed to aggregate information of feature map, generating two 2D maps: $F^{\mathrm{s}}_{\mathrm{avg}}$ and $F^{\mathrm{s}}_{\max}$. The two 2D maps denote the average-pooled features and max-pooled features across the channel, respectively. Subsequently, these feature maps are concatenated and convolved using a standard convolution layer, yielding the spatial attention map. The spatial attention can be formalized as

$$M_{\mathrm{s}}(F)=\sigma(f^{7\times7}([\mathrm{AvgPool}(F);\ \mathrm{MaxPool}\ (F)]))=$$
$$\sigma(f^{7\times7}([F^{\mathrm{s}}_{\mathrm{avg}};\ F^{\mathrm{s}}_{\max}])), \qquad (2)$$

where $f^{7\times7}$ represents a convolution operation with the filter size of 7×7.

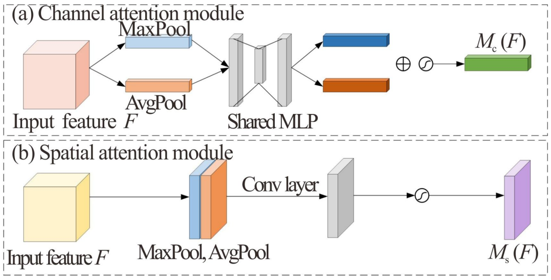


Fig.2 Diagram of the double-attention mechanism

The structure of the improved feature extraction network is complex. To reduce the number of model parameters and the calculation time, a depth separable convolution method[23] is introduced in the backbone network of DenseNet. As shown in Fig.1, we replaced 3×3 convolution in DenseLayer with a depthwise separable convolution module. This module divides a traditional convolution into two operations: a depthwise convolution and a pointwise convolution. First, the depthwise convolution performs a separate convolution operation on each channel of the input feature map, and then the resulting convolution output is concatenated to capture comprehensive information. Then, the pointwise convolution is employed to weight and combine the feature maps obtained from the previous step using a 1×1 convolution operation. By leveraging feature information from different channels at the same spatial position, the pointwise convolution facilitates effective feature fusion and representation learning. The use of depthwise separable convolution significantly reduces the number of model parameters and improves the computational efficiency by reducing the cross-channel interactions and parameter sharing across different spatial locations. Also, this design maintains model accuracy by preserving important spatial and channel-wise information.

As shown in Tab.1, the ZEHWZ dataset is an imbalanced dataset. Although, the number of normal samples and MVICC is equivalent, the number of MSVICCs is lower compared to normal and MVICCs samples. This imbalanced dataset can easily cause higher false-negative rates of classifiers. To effectively address this imbalanced dataset problem, the cost-sensitive approach[24] is adopted to adjust the weights of different classes in the loss function. Specifically, we discriminatively determine the cost of misclassification of different classes and assign a larger weight to the MSVICC class. During each iterative training stage, *n* samples are selected at random to form a training dataset $\{[x^{(1)}, y^{(1)}], [x^{(2)}, y^{(2)}], \dots, [x^{(n)}, y^{(n)}]\}$, where $x^{(i)} \in R^{l}$ and $y^{(i)} \in \{1,\dots,k\}$. Here, $x^{(i)}$ denotes the features of the *i*th sample and $y^{(i)}$ represents the category label. The cost-sensitive loss function is computed according to

$$F(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} I\left\{y^{(i)} = j\right\} * CS\left\{y^{(i)} = \text{ MSVICC }\right\} * \log\frac{\mathrm{e}^{\theta_j^T x^{(i)}}}{\sum_{s=1}^{k}\mathrm{e}^{\theta_S^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=1}^{n}\theta_{ij}^2, \tag{3}$$

where *n*, *m*, *k*, and *θ* denote the number of training samples, the number of input neurons, the number of classes, and trainable parameters, respectively. *I*{*y*(*i*)=*j*} represents the indicator function (*I*{*y*(*i*) is equal to *j*}=1 and *I* {*y*(*i*) is not equal to *j*}=0) while *CS*{*y*(*i*)=MSVICC} is the cost-sensitive weight function (*CS*{*y*(*i*) is the MSVICC class lable}=*C* and *CS*{*y*(*i*) is not the MSVICC class lable}=1). Using a grid-search procedure, we determined that the effective cost-sensitive weight parameter *C* falls within the interval [2—4]. $\frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=1}^{n}\theta_{ij}^2$ is a weight decay term applied to penalize larger trainable weights. To obtain the optimal trainable weights *θ** (as shown in Eq.(4)), we aim to minimize the objective function *F*(*θ*) using mini-batch gradient descent (Mini-batch-GD) shown as

$$\theta^{*} = \arg\min_{\theta} F(\theta), \tag{4}$$

$$\nabla_{\theta_j} F(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left[CS\left\{y^{(i)} = \text{MSVICC }\right\} \times x^{(i)} \times \left(I\left\{y^{(i)} = j\right\} - p\left(y^{(i)} = j | x^{(i)}; \theta\right)\right)\right] + \lambda\theta_j . \tag{5}$$

In the grading of visual impairment in cataract patients, the normal samples exhibit better separability. However, for patients with MVICC and MSVICC, the inter-class variance between these two visual impairment grades is small, resulting in high image similarity and posing challenges for accurate grading. To address this issue, a multi-level classifier is proposed to achieve a fine-grained assessment of cataract visual impairment by introducing a multi-task learning approach. Specifically, the extracted fundus features are fed into a two-level classifier. The first level is dedicated to the screening of cataracts, distinguishing between normal and cataract cases. The second level focuses on the fine-grained assessment of cataract visual impairment grade, further categorizing them as normal, MVICC and MSVICC. In our study, two-level loss functions for the two-level classifier are constructed during the training process, which can be expressed mathematically as

$$Loss = \alpha * level_{1\text{loss}} + (1-\alpha) * level_{2\text{loss}} . \tag{6}$$

By adjusting the weight coefficient *α*, we can dynamically manage the correlation between the two-level of loss functions. One loss function is utilized for the screening task, which calculates the loss value for the screening of normal and cataract visual impairment. Another loss function is employed for grading the severity of cataract visual impairment. These two loss functions are combined to train the model parameters. As a result, the trained model can simultaneously perform both screening and grading tasks in parallel. In addition, to address the problem of sample imbalance between MVICC and MSVICCs cases, we have introduced a cost sensitive algorithm into the second-level loss function. This algorithm assigns a greater weight factor to MSVICCs, thereby improving the recognition rate of the minority class of severe patients in the grading task.

In this study, the training procedure of DAMC_Net methods was conducted in parallel using four NVIDIA TITAN RTX graphics processing unit (GPU). The implementation was based on the Pytorch deep learning framework running on Ubuntu 18.04 LTS. The initial learning rate was set at 10^{-4} and progressively decreased by one tenth of the original value every 20 epochs. The total number of epochs was set to 80. A batch size of 64 was utilized on each GPU. During the training process, the performance of the model was evaluated using the validation set, and the model with the highest accuracy on the validation set was saved as the optimal model.

To evaluate the superiority of the DAMC_Net method compared to conventional methods, we calculated the confusion matrix and several evaluation indicators, including accuracy (*Acc*), sensitivity (*Sen*), and specificity (*Spe*), as described below

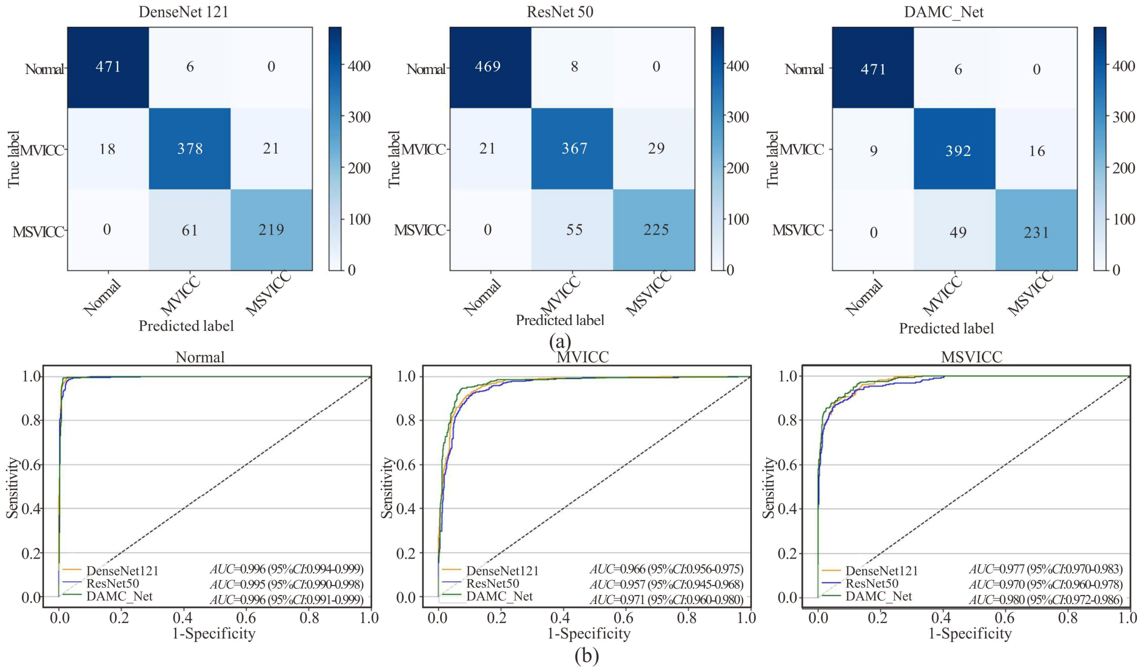
$$Acc=\frac{TP+TN}{TP+FP+FN+TN}, \tag{7}$$

$$Sen=\frac{TP}{TP+FN}, \tag{8}$$

$$Spe\ =\frac{TN}{TN+FP}, \tag{9}$$

where true positives (*TP*) represents the number of samples correctly predicted to be of a specific grade, false positives (*FP*) represents the number of samples incorrectly predicted to be of a specific grade when they actually belong to other grades, true negative (*TN*) represents the number of samples correctly predicted to be of other grades when they indeed belong to other grades, and false negative (*FN*) represents the number of samples incorrectly predicted to be of other grades when they should be classified as the specific grade. Accuracy, sensitivity, and specificity are the most commonly used evaluation metrics. Additionally, two important objective measures, the receiver operating characteristic (ROC) curve and the area under the ROC curve (*AUC*) were used for comparison and analysis.

To further assess the performance of DAMC_Net in fine-grained grading of visual impairment in cataract patients, two conventional convolutional neural networks (CNNs) were selected for comparison experiments in this study, including dense convolutional network 121 (DenseNet121) and ResNet50. The performance on the internal test dataset is shown in Fig.3, illustrating that the performance of DAMC_Net was superior to those of other classical CNNs. Tab.2 provides detailed information on the accuracies, sensitivities, and specificities of these three methods. The DAMC_Net method exhibited remarkable performance in distinguishing normal images from abnormal images (including MVICC and MSVICC images), with an *AUC* of 0.996 (95% confidence interval (*CI*), 0.991—0.999), a sensitivity of 98.7% (95% *CI*, 97.7—99.7), and a specificity of 98.7% (95% *CI*, 97.9—9.5). The DAMC_Net discriminated MVICC images from normal images and MSVICC images with an *AUC* of 0.971 (95% *CI*, 0.960—0.980), a sensitivity of 94.0% (95% *CI*, 91.7—96.3), and a specificity of 92.7% (95% *CI*, 90.9—94.6). Also, our method discriminated MSVICC images from normal images and MVICC images with an *AUC* of 0.980 (95% *CI*, 0.972—0.986), a sensitivity of 82.5% (95% *CI*, 78.0—87.0), and a specificity of 98.2% (95% *CI*, 97.3—99.1).



(a)

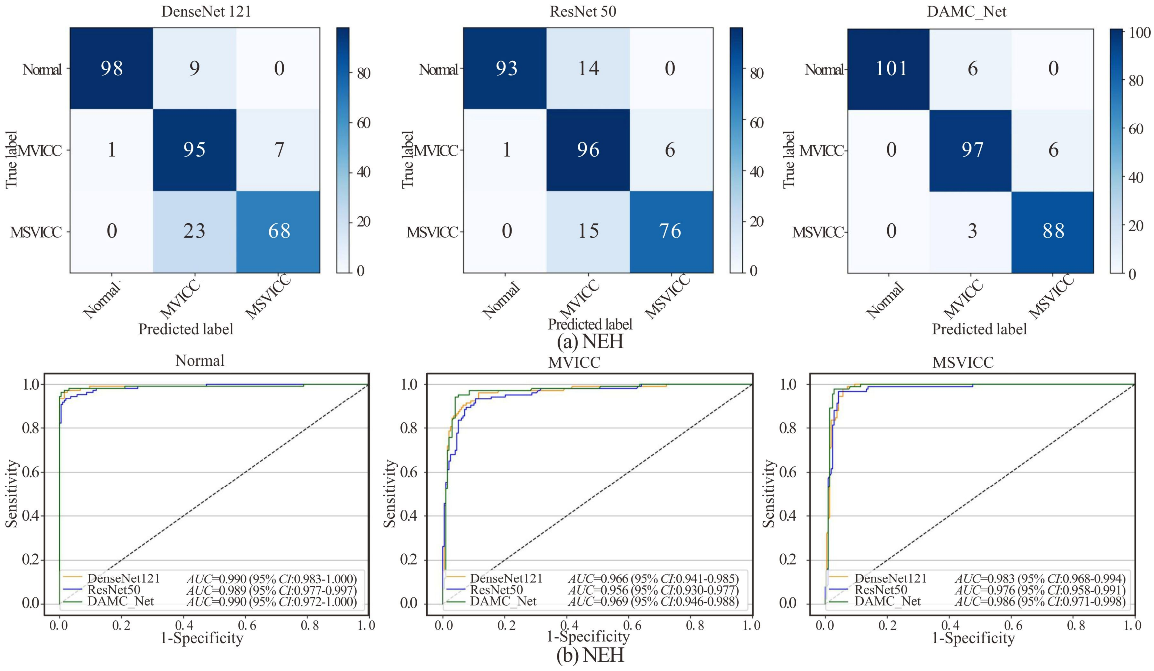
(b)

Fig.3 (a) Confusion matrices and (b) ROC curves of DAMC_Net and conventional CNNs in the internal test dataset for discriminating Normal, MVICC, and MSVICC

To validate the generalization ability of the DAMC_Net, we conducted further evaluations using two external test datasets. The performance on these external test datasets is presented in Fig.4, further confirming its superiority over other conventional CNNs. In the NEH external test dataset, the DAMC_Net achieved *AUCs* of 0.990 (95% *CI*, 0.972—1.000), 0.969 (95% *CI*, 0.946—0.988), and 0.986 (95% *CI*, 0.971—0.988) for distinguishing normal, MVICC, and MSVICC, respectively. Similarly, in the ZEHHZ dataset, the DAMC_Net demonstrated excellent performance with *AUCs* of 0.998 (95% *CI*, 0.995—0.999), 0.943 (95% *CI*, 0.928—0.957), and 0.934 (95% *CI*, 0.915—0.950) for distinguishing normal, MVICC, and MSVICC, respectively. Tab.2 presented the detailed performance of the DAMC_Net and other conventional CNNs on the external datasets. In the NEH external test dataset, the sensitivities of the DAMC_Net for distinguishing normal, MVICC, and MSVICC were 94.4% (95% *CI*, 90.0—98.8), 93.2% (95% *CI*, 88.3—98.1), and 91.2% (95% *CI*, 85.4—97.0), respectively. In the ZEHHZ external test dataset, the sensitivities of the DAMC_Net for distinguishing normal, MVICC, and MSVICC were 95.8% (95% *CI*, 93.8—97.8), 85.0% (95% *CI*, 82.0—88.0), and 72.7% (95% *CI*, 65.1—80.3), respectively. These results further validated the robustness and generalization ability of the DAMC_Net method across different external test datasets.

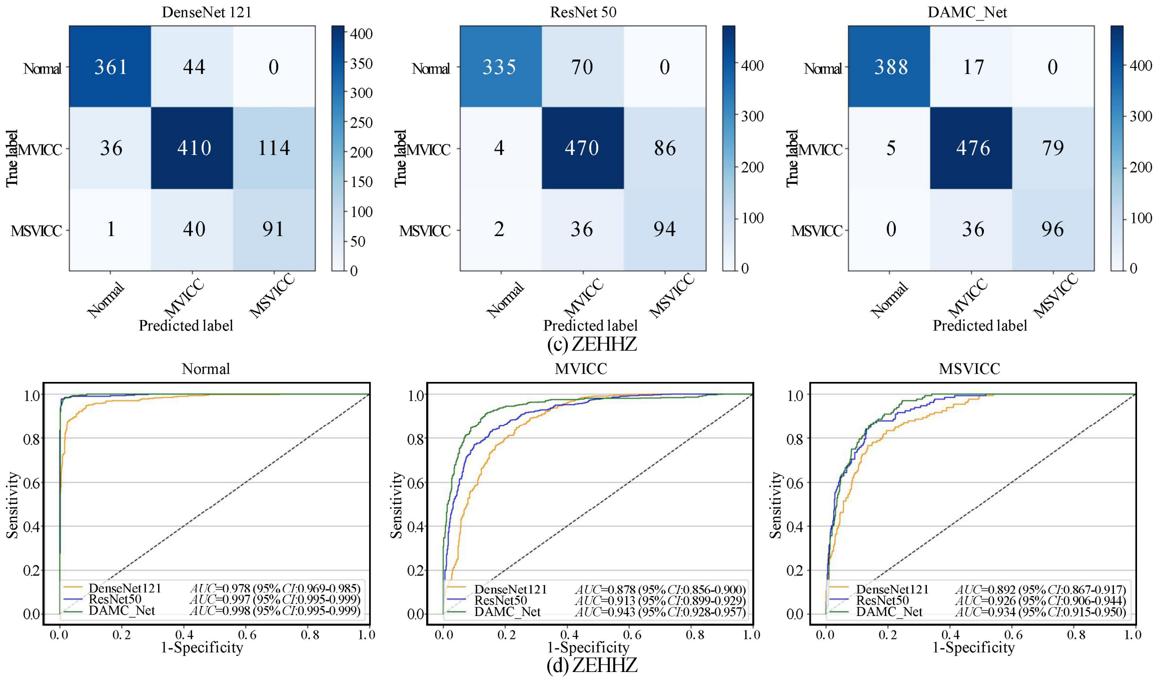
Tab.2 Performance comparison of DAMC_Net and conventional CNNs in the internal and external test datasets

One vs. rest classification	ZEHWZ internal test dataset			NEH external test dataset			ZEHHZ external test dataset		
	Acc (%) (95% *CI*)	*Sen* (%) (95% *CI*)	*Spe* (%) (95% *CI*)	*Acc* (%) (95% *CI*)	*Sen* (%) (95% *CI*)	*Spe* (%) (95% *CI*)	*Acc* (%) (95% *CI*)	*Sen* (%) (95% *CI*)	*Spe* (%) (95% *CI*)
Normal vs. MVICC + MSVICC									
DenseNet121	98.0 (97.1-98.8)	98.7 (97.7-99.7)	97.4 (96.2-98.6)	96.7 (94.7-98.7)	91.6 (86.3-96.8)	99.5 (98.5-100)	92.6 (91.1-94.2)	89.1 (86.1-92.2)	94.7 (93.0-96.3)
ResNet50	97.5 (96.6-98.4)	98.3 (97.2-99.5)	97.0 (95.7-98.3)	95.0 (92.6-97.5)	86.9 (80.5-93.3)	99.5 (98.5-100)	93.1 (91.6-94.6)	82.7 (79.0-86.4)	99.1 (98.4-99.8)
DAMC_Net	98.7 (98.1-99.4)	98.7 (97.7-99.7)	98.7 (97.9-99.5)	97.7 (96.0-99.4)	94.4 (90.0-98.8)	99.5 (98.5-100)	98.0 (97.2-98.8)	95.8 (93.8-97.8)	99.3 (98.6-99.9)
MVICC vs. Normal + MSVICC									
DenseNet121	91.0 (89.3-92.6)	90.6 (87.9-93.4)	91.1 (89.1-93.2)	86.7 (82.9-90.5)	92.2 (87.1-97.4)	83.8 (78.7-89.0)	78.7 (76.2-81.1)	73.2 (69.5-76.9)	84.4 (81.3-87.4)
ResNet50	90.4 (88.7-92.1)	88.0 (84.9-91.1)	91.7 (89.7-93.6)	88.0 (84.4-91.7)	93.2 (88.3-98.1)	85.4 (80.4-90.3)	82.1 (79.9-84.4)	83.9 (80.9-87.0)	80.3 (76.9-83.6)
DAMC_Net	93.2 (91.7-94.6)	94.0 (91.7-96.3)	92.7 (90.9-94.6)	93.0 (90.1-95.9)	93.2 (88.3-98.1)	92.9 (89.4-96.5)	87.5 (85.6-89.5)	85.0 (82.0-88.0)	90.1 (87.6-92.7)
MSVICC vs. Normal + MVICC									
DenseNet121	93.0 (91.6-94.5)	78.2 (73.4-83.0)	97.7 (96.7-98.6)	90.0 (86.6-93.4)	74.7 (65.8-83.7)	96.7 (94.2-99.1)	85.9 (83.8-87.9)	68.9 (61.0-76.8)	88.2 (86.2-90.2)
ResNet50	92.8 (91.4-94.3)	80.4 (75.7-85.0)	96.8 (95.6-97.9)	93.0 (90.1-95.9)	83.5 (75.9-91.1)	97.1 (94.9-99.4)	88.7 (86.8-90.6)	71.2 (63.5-78.9)	91.1 (89.3-92.9)
DAMC_Net	94.5 (93.2-95.8)	82.5 (78.0-87.0)	98.2 (97.3-99.1)	95.3 (93.0-97.7)	91.2 (85.4-97.0)	97.1 (94.9-99.4)	89.5 (87.7-91.3)	72.7 (65.1-80.3)	91.8 (90.1-93.5)



(a) NEH

(b) NEH



(c) ZEHHZ

(d) ZEHHZ

Fig.4 Confusion matrices and ROC curves of DAMC_Net and conventional CNNs in two external test datasets: (a, c) Confusion matrices in NEH and ZEHHZ external test datasets; (b, d) ROC curves for discriminating Normal, MVICC, and MSVICC in NEH and ZEHHZ external test datasets

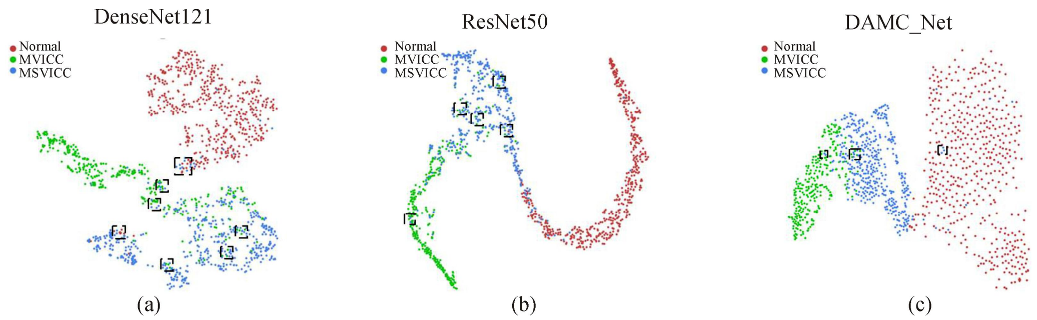
To determine the high efficiency of the proposed method, we conducted a comparison of the efficiency and resource utilization of the DAMC_Net and conventional CNNs, including the model size, number of trainable parameters, and running time of training and testing. As shown in Tab.3, the model size of the DAMC_Net was only 70.04 MB, which was substantially smaller than DenseNet121 by 10.37 MB and ResNet50 by 199.41 MB. Also, the number of trainable parameters of the DAMC_Net was lower than the other models. Moreover, the training time required for the DAMC_Net was a mere 1.47 h, resulting in significant time savings when training on a local server. This experimental finding demonstrated that the proposed method outperformed other conventional CNNs in terms of efficiency and resource utilization.

Tab.3 Efficiency comparison of different classification methods

Model	Size	Parameters	Training time	Testing time
DenseNet121	80.41 MB	8.0×10^6	1.56 h	0.195 s
ResNet50	269.46 MB	2.6×10^7	1.62 h	0.204 s
DAMC_Net	70.04 MB	7.2×10^6	1.47 h	0.187 s

The t-distributed stochastic neighbor embedding (t-SNE) was employed to intuitively analyze whether the characteristics of each category learned by the deep learning model were discriminative in a two-dimensional space. Specifically, we removed the classification layer from the network and utilized the output of the previous layer, just before the classification layer, as the final extracted feature. Subsequently, the t-SNE technology was applied to visualize the extracted features and assess their separability. Visualized maps of the high-level features extracted from DAMC_Net and other conventional CNNs were displayed in Fig.5. The red points represent normal samples, the blue points represent MVICC samples, and the green points represent MSVICC samples. Compared with DenseNet121 and ResNet50, the separability of the DAMC_Net features was improved markedly. Notably, the proportion of easily misdiagnosed samples (dotted square in Fig.5) is significantly reduced in the DAMC_Net model. This analysis showed that the DAMC_Net exhibited a superior capability in separating high-level features for fine-grained assessment of visual impairment grading in cataract.

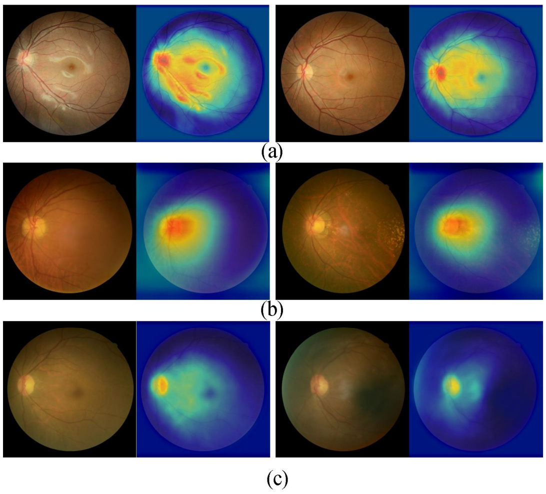
To visualize the fundus regions contributing most to the DAMC_Net, we generated heatmaps using the gradient-weighted class activation mapping (Grad-CAM) method. For abnormal fundus images (including MVICC and MSVICC), heatmaps effectively highlighted the capillary area around the optic disc. Compared to the MVICC, the heatmaps of the MSVICC exhibited fewer highlighted areas, which can be attributed to the increased turbidity of the vessels around the optic disc as the severity of visual impairment deepens. On the other hand, for normal images, heatmaps display highlighted visualization on the clear retinal vessels. Typical examples of the heatmaps for MVICC, MSVICC, and normal images were presented in Fig.6.



(a) (b) (c)

Fig.5 Visualization by t-SNE of the separability for the features learned by DAMC_Net and conventional CNNs: (a) DenseNet121; (b) ResNet50; (c) DAMC_Net

Since the proposed multi-level classifier of DAMC_Net contains two levels of classification tasks, the performance of the model can be influenced by different combinations of loss functions for these two levels. Therefore, we assigned weight coefficients to the loss functions of these two levels, and their weighted sum is equal to 1. In total, five different combinations were tested and compared in detail (Fig.7). From the comparative experiments, it was observed that the DAMC_Net achieved the best performance when the weight coefficients for the two levels were set to 0.2 and 0.8, respectively.



(a)

(b)

(c)

Fig.6 Typical examples of the heatmaps for three grades of visual impairment: (a) Normal; (b) MVICC; (C) MSVICC

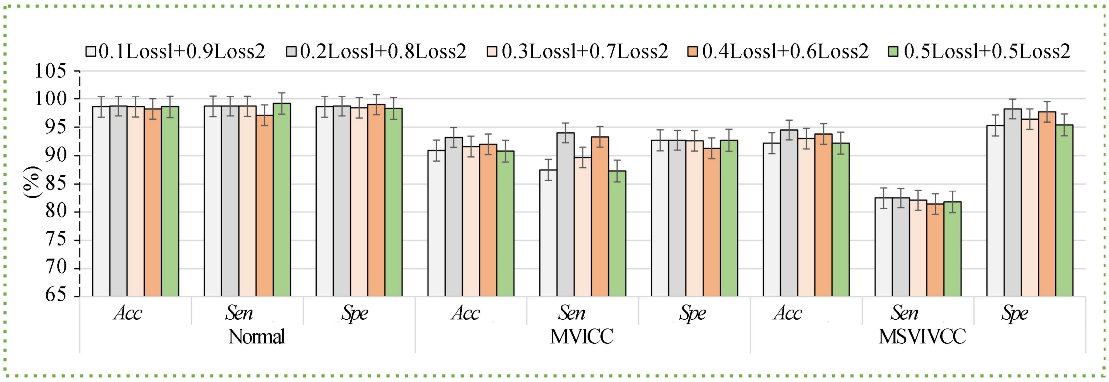


Fig.7 Performance comparison of different weight coefficients combinations in the multi-level classifier

The outstanding performance of DAMC_Net for visual impairment grading in cataract patients can be attributed to three key reasons. First, the double-attention mechanism was performed to enable the DAMC_Net to integrate fundus features from two dimensions of spatial attention and channel attention. Second, a multi-level classifier was proposed to complete the fine-grained assessment of visual impairment grading in cataract, which is beneficial to improve the recognition rate of the MVICC from MSVICC. Third, the cost-sensitive was adopted to facilitate the DAMC_Net to focus more on the minority category of MSVICC. When compared to ResNet50 and DenseNet121, DAMC_Net demonstrated a higher sensitivity in distinguishing MVICC, with an increase of 6.0% and 3.4%, respectively. Similarly, in differentiating MSVICC, DAMC_Net achieved an increase of 2.1% and 4.3% in sensitivity compared to ResNet50 and DenseNet121, respectively. This improved performance was also validated on two external datasets, indicating the superior generalization ability of the DAMC_Net model. Furthermore, the integration of depthwise separable convolution in DAMC_Net effectively reduced the number of parameters, training time, and testing time, thus improving the efficiency and resource utilization of the model.

Several limitations exist in this study. First, although the DAMC_Net provided a feasible method for fine-grained assessment for visual impairment grading, it currently focused on two levels of severity. This may not be sufficient for clinical applications that require more precise grading. Future research will explore the development of a more fine-grained grading system with multiple levels. Second, the training of our deep learning model heavily relied on high-quality dataset, which was often challenging to obtain in the real-world clinical settings. Meta-learning techniques may be advantageous in transferring knowledge from large-scale natural image datasets to the medical domain with limited data availability. Third, relying solely on fundus images as input data for our model may not fully capture the complexity and heterogeneity of visual impairment caused by cataracts. Visual impairment is a multifaceted condition influenced by various factors, including age, medical history, and underlying diseases, in addition to fundus images. With the accumulation of more data, the application of multimodal in the assessment of visual impairment will be investigated to provide more accurate and comprehensive diagnoses in clinic.

This study presented a feasible fine-grained classification algorithm DAMC_Net for assessing visual impairment in cataract patients. The algorithm combined dense convolution network, double-attention mechanism, depthwise separable convolution, cost-sensitive learning, and multiple levels classifiers to accurately grade the severities of visual impairment. Experimental results and comparison analyses verified that the proposed method outperformed other conventional CNNs. The robustness and generalizability of DAMC_Net were validated by its excellent performance on two external test datasets. This research could provide a valuable reference for the analysis of other fine-grained assessment of medical images and promote the application of artificial intelligence techniques in clinical settings.

Ethics declarations

Conflicts of interest

The authors declare no conflict of interest.

Source code and data sharing statement

The code and data used in this study can be accessed at GitHub (https://github.com/jiangjiewei/DAMC_Net).

References

[1] FLAXMAN S R, BOURNE R R, RESNIKOFF S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis[J]. The lancet global health, 2017, 5(12): e1221-e1234.

[2] LAM D, RAO S K, RATRA V, et al. Cataract[J]. Nature reviews disease primers, 2015, 1(1): 1-15.

[3] DAY A C, FINDL O. Femtosecond laser-assisted vs conventional cataract surgery[J]. The lancet, 2020, 395(10219): 170-171.

[4] HE Y, ZHANG R, ZHANG C, et al. Clinical outcome of phacoemulsification combined with intraocular lens implantation for primary angle closure/glaucoma (PAC/PACG) with cataract[J]. American journal of translational research, 2021, 13(12): 13498.

[5] SCHWEITZER C, BREZIN A, COCHENER B, et al. Femtosecond laser-assisted versus phacoemulsification cataract surgery (FEMCAT): a multicentre participant-masked randomised superiority and cost-effectiveness trial[J]. The lancet, 2020, 395(10219): 212-224.

[6] World Health Organization. World report on vision[R]. Geneva: WHO, 2019.

[7] RESNIKOFF S, FELCH W, GAUTHIER T M, et al. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners[J]. British journal of ophthalmology, 2012, 96(6): 783-787.

[8] SHEN D, WU G, SUK H I. Deep learning in medical image analysis[J]. Annual review of biomedical engineering, 2017, 19: 221.

[9] KHOJASTE-SARAKHSI M, HAGHIGHI S S, GHOMI S F, et al. Deep learning for Alzheimer's disease diagnosis: a survey[J]. Artificial intelligence in medicine, 2022: 102332.

[10] CHEN S, QIU C, YANG W, et al. Combining edge guidance and feature pyramid for medical image segmentation[J]. Biomedical signal processing and control, 2022, 78: 103960.

[11] LOTTER W, DIAB A R, HASLAM B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach[J]. Nature medicine, 2021, 27(2): 244-249.

[12] GRZYBOWSKI A, BRONA P, LIM G, et al. Artificial intelligence for diabetic retinopathy screening: a review[J]. Eye, 2020, 34(3): 451-460.

[13] YU K H, BEAM A L, KOHANE I S. Artificial intelligence in healthcare[J]. Nature biomedical engineering, 2018, 2(10): 719-731.

[14] SENGUPTA S, SINGH A, LEOPOLD H A, et al. Ophthalmic diagnosis using deep learning with fundus images-a critical review[J]. Artificial intelligence in medicine, 2020, 102: 101758.

[15] LI Z, JIANG J, CHEN K, et al. Preventing corneal blindness caused by keratitis using artificial intelligence[J]. Nature communications, 2021, 12(1): 1-12.

[16] LI Z, QIANG W, CHEN H, et al. Artificial intelligence to detect malignant eyelid tumors from photographic images[J]. NPJ digital medicine, 2022, 5(1): 1-9.

[17] ZHANG H, NIU K, XIONG Y, et al. Automatic cataract grading methods based on deep learning[J]. Computer methods and programs in biomedicine, 2019, 182: 104978.

[18] JUNAYED M S, ISLAM M B, SADEGHZADEH A, et al. CataractNet: an automated cataract detection system using deep learning for fundus images[J]. IEEE access, 2021, 9: 128799-128808.

[19] XU X, ZHANG L, LI J, et al. A hybrid global-local representation CNN model for automatic cataract grading[J]. IEEE journal of biomedical and health informatics, 2020, 24(2): 556-567.

[20] BLOICE M D, ROTH P M, HOLZINGER A. Biomedical image augmentation using augmentor[J]. Bioinformatics, 2019, 35(21): 4522-4524.

[21] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, USA. New York: IEEE, 2017: 4700-4708.

[22] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018, 11211: 3-19.

[23] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, USA. New York: IEEE, 2017: 1800-1807.

[24] JIANG J, WANG L, FU H, et al. Automatic classification of heterogeneous slit-illumination images using an ensemble of cost-sensitive convolutional neural networks[J]. Annals of translational medicine, 2021, 9(7): 550.