

# Intelligent diagnostic model for pterygium by combining attention mechanism and MobileNetV2

Mao-Nian Wu<sup>1,2</sup>, Kai He<sup>1,3</sup>, Yi-Bei Yu<sup>1</sup>, Bo Zheng<sup>1,2</sup>, Shao-Jun Zhu<sup>1,2</sup>, Xiang-Qian Hong<sup>4</sup>, Wen-Qun Xi<sup>4</sup>, Zhe Zhang<sup>4</sup>

<sup>1</sup>School of Information Engineering, Huzhou University, Huzhou 313000, Zhejiang Province, China

<sup>2</sup>Zhejiang Province Key Laboratory of Smart Management and Application of Modern Agricultural Resources, Huzhou University, Huzhou 313000, Zhejiang Province, China

<sup>3</sup>School of Mathematical Information, Shaoxing University, Shaoxing 312000, Zhejiang Province, China

<sup>4</sup>Shenzhen Eye Institute, Shenzhen Eye Hospital, Jinan University, Shenzhen 518040, Guangdong Province, China

**Correspondence to:** Wen-Qun Xi and Zhe Zhang. Shenzhen Eye Institute, Shenzhen Eye Hospital, Jinan University, Shenzhen 518040, Guangdong Province, China. wtongxxiwq125@126.com; hypotato@126.com

Received: 2023-09-24 Accepted: 2024-04-01

## Abstract

• **AIM:** To evaluate the application of an intelligent diagnostic model for pterygium.

• **METHODS:** For intelligent diagnosis of pterygium, the attention mechanisms—SENet, ECANet, CBAM, and Self-Attention—were fused with the lightweight MobileNetV2 model structure to construct a tri-classification model. The study used 1220 images of three types of anterior ocular segments of the pterygium provided by the Eye Hospital of Nanjing Medical University. Conventional classification models—VGG16, ResNet50, MobileNetV2, and EfficientNetB7—were trained on the same dataset for comparison. To evaluate model performance in terms of accuracy, Kappa value, test time, sensitivity, specificity, the area under curve (AUC), and visual heat map, 470 test images of the anterior segment of the pterygium were used.

• **RESULTS:** The accuracy of the MobileNetV2+Self-Attention model with 281 MB in model size was 92.77%, and the Kappa value of the model was 88.92%. The testing time using the model was 9ms/image in the server and 138ms/image in the local computer. The sensitivity, specificity, and AUC for the diagnosis of pterygium using normal anterior segment images were 99.47%, 100%, and 100%, respectively; using anterior segment images in the observation period were 88.30%, 95.32%, and 96.70%,

respectively; and using the anterior segment images in the surgery period were 88.18%, 94.44%, and 97.30%, respectively.

• **CONCLUSION:** The developed model is lightweight and can be used not only for detection but also for assessing the severity of pterygium.

• **KEYWORDS:** deep learning; attention mechanism; pterygium; intelligent diagnosis

**DOI:10.18240/ijo.2024.07.02**

**Citation:** Wu MN, He K, Yu YB, Zheng B, Zhu SJ, Hong XQ, Xi WQ, Zhang Z. Intelligent diagnostic model for pterygium by combining attention mechanism and MobileNetV2. *Int J Ophthalmol* 2024;17(7):1184-1192

## INTRODUCTION

Pterygium, a common ocular surface disease, is an upper type of fibrovascular tissue that grows from the bulbar conjunctiva towards the cornea<sup>[1]</sup>. As the condition of the pterygium worsens, abnormal tissue encroaches the corneal region, leading to visual impairment. Therefore, pterygium requires early diagnosis and interventional measures to reduce the growth of fibrovascular tissue<sup>[2]</sup>. Some studies<sup>[3-5]</sup> have shown that the prevalence of pterygium is 12% worldwide, with high prevalence in populations experiencing frequent sun exposure, and living in rural areas (e.g., fishermen and farmers). Nearly 109 million people in China have pterygium<sup>[6]</sup> while there are only a few ophthalmologists. These physicians are mostly distributed in large urban hospitals and there is a dearth of ophthalmic medical resources in primary hospitals, such as county-level hospitals and community hospitals. In this study, we propose an intelligent assisted diagnosis model that can assist primary care physicians and young inexperienced physicians with the initial diagnosis of pterygium and, thus, help pterygium to be detected early, patients to be referred in time, and medical resources to be allocated rationally.

With great progress in computer vision and deep learning technology, the intelligent diagnosis and treatment mode of “artificial intelligence+medical imaging” has been widely studied and promoted. Computer-aided diagnosis can provide

doctors with an effective and accurate clinical diagnosis while also reducing the time, cost, and workload of disease screening. Traditional image feature extraction methods rely on large amounts of prior knowledge and have several limitations. In contrast, deep learning techniques can select and extract image features autonomously and have excellent performance. Convolutional neural networks (CNNs) have become important models in deep learning techniques because of their powerful feature representation capability and are widely used in tasks requiring image classification, target detection, and image segmentation. A series of classical networks based on CNN structures were proposed earlier, such as AlexNet<sup>[7]</sup>, VGGNet<sup>[8]</sup>, and ResNet<sup>[9]</sup>, which laid the foundation for subsequent network improvements. Later, MobileNet<sup>[10-12]</sup> used depthwise separable convolutions to build lightweight deep neural networks that are more suitable for mobile and embedded vision applications. Notably, all of the aforementioned network models have reasonable image detection capabilities.

In ophthalmology, deep learning techniques have been applied to intelligently detect diseases including diabetic retinopathy (DR)<sup>[13-15]</sup>, age-related macular degeneration (AMD)<sup>[16-17]</sup>, retinopathy of prematurity (ROP)<sup>[18-19]</sup>, glaucoma<sup>[20-21]</sup>, and some studies have even achieved classification of multiple fundus diseases<sup>[22-23]</sup>. Studies applying deep learning techniques to pterygium have focused on intelligent classification and segmentation<sup>[24-25]</sup>. Lopez and Aguilera<sup>[26]</sup> were the first to propose the use of deep learning techniques for pterygium classification. Studies<sup>[27-32]</sup> conducted many pterygium-related studies between 2018 and 2021 based on the pterygium datasets such as UBIRIS, MILES, Brazilian pterygium (BP), and Australian pterygium (AP). In 2018, to distinguish between pterygium and non-ptyerygium images, an automatic detection system was proposed which consisted of four modules: image preprocessing, corneal segmentation, feature extraction, and classification. The classification module uses support vector machines (SVMs), artificial neural networks, and traditional machine learning methods<sup>[27]</sup>. In 2019, using deep learning techniques, a Pterygium-Net algorithm based on a fully convolutional neural network was proposed to automatically classify and localize pterygium-infected tissue<sup>[28]</sup>. Pterygium-Net consists of a three-layer convolutional neural network and a three-layer fully connected network. Classification training was performed using the ImageNet dataset<sup>[33]</sup> to pretrain the weights of the convolutional neural network layers, and localization was based on supervised learning from expert-labeled image data. In 2020, six convolutional neural networks based on AlexNet, VggNet16, VggNet19, GoogleNet, ResNet101, and DenseNet201 were evaluated for pterygium detection. Also, the VggNet16-wbn two-classification

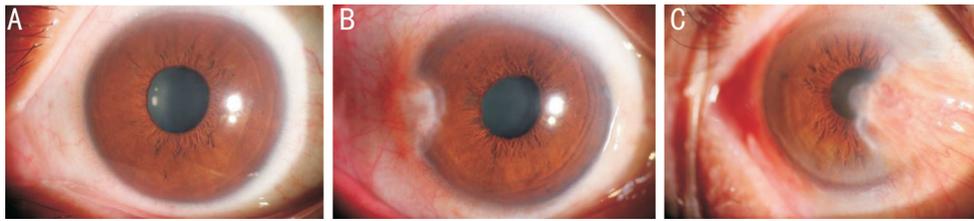
model was improved using VggNet16, with an accuracy, sensitivity and, specificity of 99.22%, 98.45%, and, 100%, respectively<sup>[30]</sup>. The Group-PPM-Net segmentation model was proposed in 2021 by integrating the spatial pyramid module and group convolution into a deep learning segmentation network, which achieved segmentation of pterygium-infected tissues with an average accuracy of 93.3% and an average intersection over union of 86.4%<sup>[32]</sup>. The latest study in 2022 Hung *et al*<sup>[34]</sup> proposed a deep learning grading system for pterygium, which first trained a two-classification model to detect the pterygium image, segmented the lesion region of the pterygium image and classified it into three grades according to the severity, trained a two-classification model between each two grades, and finally obtained four two-classification models to achieve intelligent grading. Presently, research on the intelligent classification of pterygium mainly consists of two-classification models to detect pterygium, and of these models most use deep learning classical models, which are complex, use a large number of parameters and are difficult to be deployed in low configuration devices.

In this study, an intelligent lightweight pterygium-assisted diagnosis model was constructed based on MobileNetV2 by combining it with an attention mechanism. The model can be used to classify and detect normal images, pterygium observation phase images and pterygium surgery phase images with high classification accuracy and easy application. The model can be deployed for use in the department of ophthalmology in primary hospital. A comparative study with the classical model was also conducted as described below.

## MATERIALS AND METHODS

**Data Acquisition** In this study, 1220 images of the anterior segment of the pterygium with a resolution of 5184×3456 were provided by the Eye Hospital of Nanjing Medical University and were classified into three categories according to medical classification criteria: 439 images of the normal anterior segment, 421 images of the observational anterior segment, and 360 images of the surgery-period anterior segment. The classification criteria was as follows<sup>[35]</sup>: 1) normal preoptic node images: no significant conjunctival congestion or proliferative bulge, transparent cornea (Figure 1A); 2) observation of preoptic node images: the head tissue of pterygium invades the horizontal length of the corneal margin <3 mm (Figure 1B); 3) surgery preoptic node images: the head tissue of pterygium invades the horizontal length of the corneal margin ≥3 mm (Figure 1C). All images were captured in the same environment with the same level of equipment, and all relevant patient information was removed to avoid violating patient privacy.

All the pterygium anterior segment images were divided into training, validation, and test sets, using random assignment. The 225 images from each of the three categories of pterygium



**Figure 1** Images of the anterior segment of three types of pterygium A: Normal preoptic node images; B: Observation of preoptic node images; C: Urgery preoptic node images.

(normal, observational, and surgical anterior segment) were used as the training set; 25 images from each as the validation set; the remaining 189 images of the normal anterior segment, 171 images of the observational anterior segment, and 110 images of the surgical anterior segment were used as the test set. To ensure generalization of the model results, the original training set images were first horizontally flipped. Then, the horizontally flipped images and original training set were rotated by  $-3^\circ$  and  $3^\circ$ , respectively. The final pterygium training set used for model training comprised 4050 images, the validation set comprised 75 images, the test set comprised 470 images (Table 1).

**Methods** The MobileNet network is a lightweight network, proposed by the Google team that focuses on mobile and embedded devices. MobileNetV1<sup>[10]</sup> includes depthwise separable convolution, which greatly reduces model parameters and operations with only a small reduction in accuracy compared to traditional convolutional neural networks. MobileNetV2<sup>[11]</sup> introduced the inverse residual and linear bottleneck structure based on MobileNetV1 which resulted in higher accuracy, smaller model, and faster computation time. As shown in Table 2, the MobileNetV2 network structure mainly comprises a convolutional layer, bottleneck layer, and average pooling layer, where  $t$  denotes the multiplicity of the  $1 \times 1$  convolutional channel up-dimensioning in the inverse residual structure,  $c$  is the number of output feature matrix channels,  $n$  is the number of bottleneck layer repetitions, and  $s$  denotes the step size.

The inverted residual structure of the bottleneck layer is shown in Figure 2. First, the input data is up-dimensioned by a  $1 \times 1$  convolution, then the image features are extracted by a depthwise separable convolution, and finally down-dimensioned by a  $1 \times 1$  convolution. Linear activation is used in the down-dimensioning process to avoid information loss. When the step size is 1 and the input and output channels are the same, the input and output are summed together.

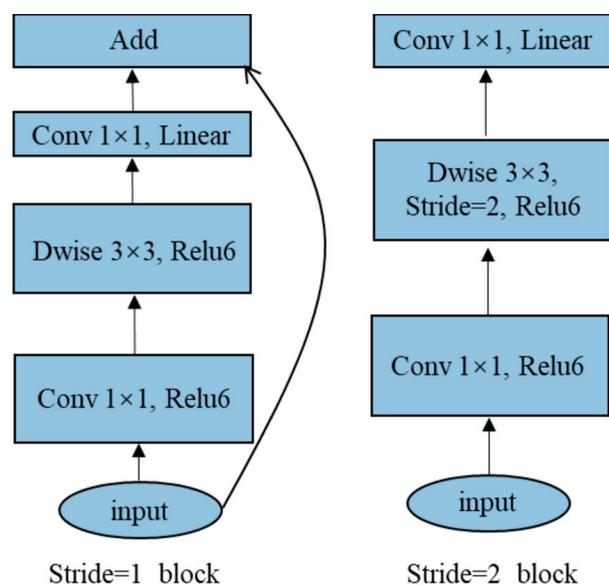
The attention mechanism in neural networks is a resource-allocation scheme that solves the information-overload problem with limited computational power and allocates computational resources to important tasks. In neural network learning, the attention mechanism can use limited attention

**Table 1** Experimental data division

Pterygium types	Original data			Enhanced data		
	Training set	Validation set	Test set	Training set	Validation set	Test set
Normal	225	25	189	1350	25	189
Observe	225	25	171	1350	25	171
Surgery	225	25	110	1350	25	110
Sum	675	75	470	4050	75	470
Total		1220			4595	

**Table 2** MobileNetV2 network structure

Input	Operator	$t$	$c$	$n$	$s$
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d1x1	-	1280	1	1
$7^2 \times 1280$	avgpool7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d1x1	-	k	-	-



**Figure 2** Bottleneck.

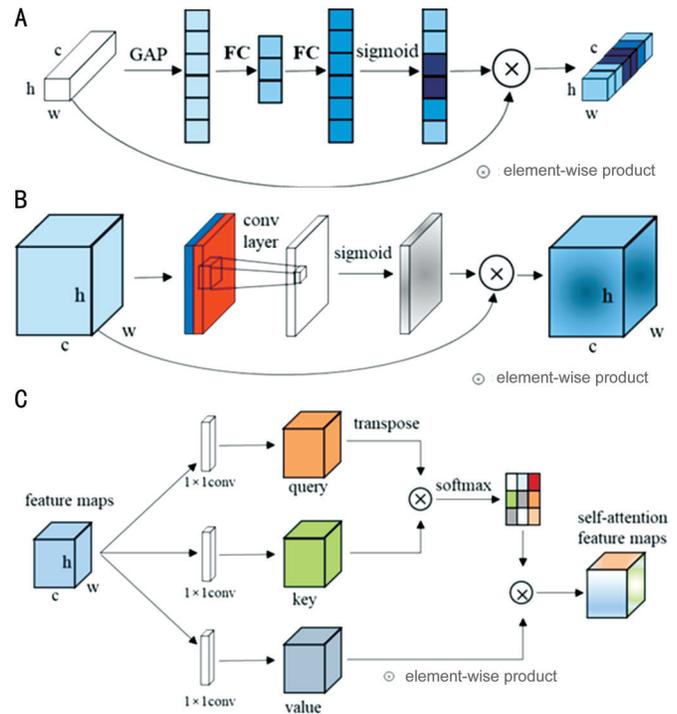
resources to quickly filter out high-value information from a large amount of data. In computer vision tasks when the

network extracts image features, the overall performance of the network can be improved by adding attention mechanisms between the convolutional layers. There are four main types of attention mechanisms: channel attention, spatial attention, mixed spatial and channel attention, and self-attention.

The channel attention mechanism aims to establish correlations between different feature map channels and assign different weight coefficients to each channel so as to reinforce important features and suppress non-important ones. SENet<sup>[36]</sup> and ECANet<sup>[37]</sup> are typical channel-attention mechanisms. As shown in Figure 3A, SENet first performs global average pooling (GAP) on the feature map to obtain the global compressed feature volume, and then passes through two fully connected layers to obtain the weights of each channel in the feature map. The sigmoid activation function ensures that the weight values are distributed between 0 and 1, and the values of different channels are multiplied by different weight values so that attention to key channels can be enhanced. ECANet uses a one-dimensional convolution operation to optimize the fully connected operation in SENet which substantially reduces the number of parameters while maintaining the same network performance.

The spatial attention mechanism aims to improve the feature performance of crucial regions, generate weights for each spatial location, and improve the target regions of interest while weakening irrelevant background regions. The convolutional block attention module (CBAM)<sup>[38]</sup> combines channel attention and spatial attention mechanisms. As shown in Figure 3B, the spatial attention mechanism first performs GAP and global maximum pooling operations based on channels to generate two feature maps representing different information, and then merges them to obtain the weights for each spatial location in the feature map by convolution. The sigmoid activation function ensures that the weight values are distributed between 0 and 1. Finally, the input feature maps are multiplied by the weight values, thus enabling the target region to be enhanced.

The self-attention mechanism<sup>[39]</sup> allows the machine to notice the correlation between different parts of the input, capture the internal correlation of data or features, and reduce dependence on external information. The CNN only considers pixels within a  $K \times K$  receptive field, whereas the receptive field of the self-attention mechanism is the whole image. As shown in Figure 3C, first, the input feature maps are convolved  $1 \times 1$  to obtain the query, key, and value matrices, respectively. Then, the query matrix is transposed and element-wise produced with the key matrix, and the softmax activation function is normalized to process the data and element-wise produced with the value matrix to obtain the self-attention feature map as the output. Key, query, and value triplets provide an efficient way to capture global contextual information for modeling.



**Figure 3 Attentional mechanisms** A: SENet channel attention mechanism; B: Spatial attention mechanism in CBAM; C: Self-Attention mechanism.

The MobileNetV2 network structure contains a feature layer and classification layer. The ingenious design of depthwise separable convolution in the feature layer greatly reduces the model parameters, operations and loses in feature information. Although the proposed inverse residual and linear bottleneck structures can reduce information loss, the accuracy of the algorithm is low. In this study, the attention mechanism module was added after the feature layer of the MobileNetV2 model, and four different attention mechanisms, SENet, CBAM, ECANet, and Self-Attention, were used to construct four pterygium tri-classification models; the simple model structures are shown in Figure 4. The initialization parameters for model training were: pre-trained weights on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset, propagation learning using the Adam optimizer with an initial learning rate of 0.0001, and a total of 50 epochs. In this study, VGG16, ResNet50, MobileNetV2, and EfficientNetB7 classical deep learning classification models were used as comparison networks to train models on the same pterygium dataset. These four models were initialized with parameters that were pre-trained on the ILSVRC dataset and only the final output of the classification layer was updated to three classes while the remaining model structure remained unchanged. Finally, all three pterygium classification models were validated and compared using the test set.

The programming language used in this study was Python and the deep learning framework used was PyTorch. Model training and testing were performed on a server, and a local

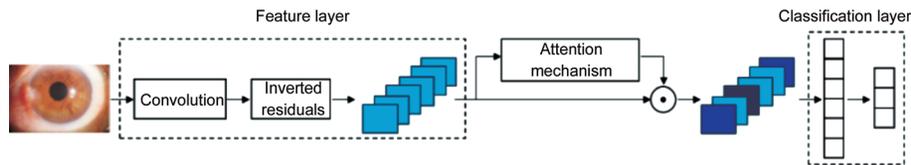


Figure 4 MobileNetV2 network architecture with combined attention mechanism.

Table 3 Experimental results

Evaluation Parameters	Model	VGG16	ResNet50	MobileNetV2	MobileNetV2+SENet	MobileNetV2+CBAM	MobileNetV2+ECANet	MobileNetV2+Self-Attention	EfficientNetB7
Sensitivity (%)	Normal	98.41	100.00	100.00	99.47	100.00	98.94	99.47	100.00
	Observe	84.80	83.63	84.80	90.06	83.63	83.04	88.30	91.81
	Surgery	91.82	92.73	89.09	85.45	90.91	90.91	88.18	89.09
Specificity (%)	Normal	99.64	99.29	97.86	99.29	98.22	98.93	100.00	99.64
	Observe	96.32	97.32	96.32	94.31	96.66	95.99	95.32	95.99
	Surgery	92.78	92.78	94.17	95.83	93.61	92.78	94.44	96.39
False positive rate (%)	Normal	1.59	0	0	0.53	0	1.06	0.53	0
	Observe	15.20	16.37	15.20	9.94	16.37	16.96	11.70	8.19
	Surgery	8.18	7.27	10.91	14.55	9.09	9.09	11.82	10.91
False negative rate (%)	Normal	0.36	0.71	2.14	0.71	1.78	1.07	0	0.36
	Observe	3.68	2.68	3.68	5.69	3.34	4.01	4.68	4.01
	Surgery	7.22	7.22	5.83	4.17	6.39	7.22	5.56	3.61
AUC (%)	Normal	100.00	100.00	100.00	99.80	99.70	100.00	100.00	100.00
	Observe	96.60	96.30	96.80	97.30	96.40	96.60	96.70	96.80
	Surgery	97.50	97.60	97.70	97.60	97.50	97.00	97.30	97.10
Accuracy (%)		91.91	92.34	91.91	92.77	91.91	91.28	92.77	94.47
Kappa value (%)		87.68	88.32	87.62	88.88	87.64	86.69	88.92	91.51
Model size (MB)		841	521	239	241	242	240	281	1919
Model parameters (million)		134	23	2	2	2	2	8	63
Time-S (ms)		10	12	11	9	10	9	9	60
Time-C (ms)		606	328	91	91	91	93	138	4570

Time-S: The time of testing an image on the server; Time-C: The time of testing an image on the local computer.

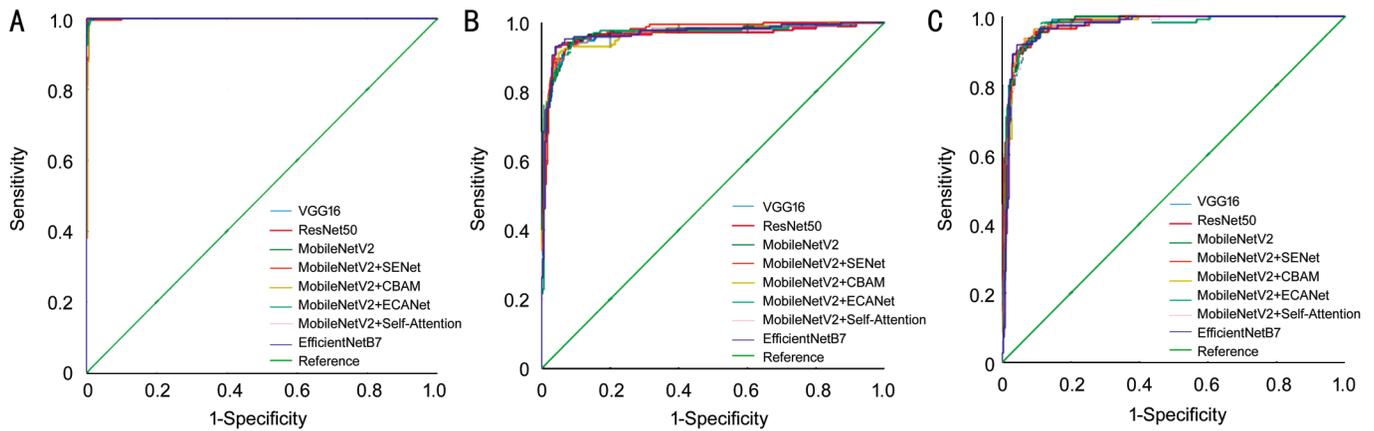
computer was also used for model testing in consideration of future practical deployment in primary hospitals. The server device is Intel(R) Xeon(R) Gold 5118 CPU, Tesla V100 graphics card, and the operating system is Linux. The local computer device is Intel(R) Core(TM) i5-4200M CPU, NVIDIA GeForce GT 720M x graphics card, and the operating system is Windows 10.

**Statistical Analysis** The results were analyzed using the SPSS 25.0 statistical software. The evaluation metrics used to measure the performance of the three classification models for pterygium were sensitivity, specificity, accuracy, Kappa value, test time, receiver operator characteristic curve (ROC), and area under curve (AUC). High sensitivity, low underdiagnosis rate, high specificity, and low misdiagnosis rate. Kappa value is used to assess the agreement between the expert and model; with 61% to 80% being significant agreement and greater than 80% being high agreement. An AUC value between 70%–85% indicates general effectiveness while a value greater than 85% indicates good diagnostic value.

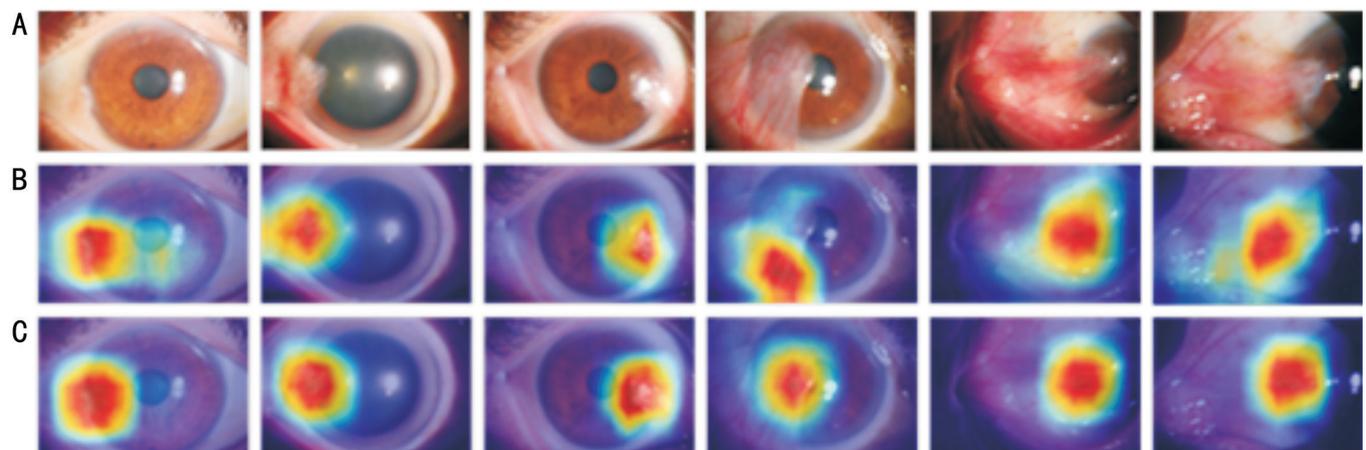
**RESULTS**

**Evaluation Metrics Results** The test set of this study consisted of 470 images, of these 189 were normal anterior

ocular segments, 171 were pterygium in the observation phase, and 110 were pterygium in the surgery phase. The results of the eight pterygium tri-classification models are shown in Table 3 and Figure 5. The MobileNetV2+SENet, MobileNetV2+CBAM, MobileNetV2+ECANet, and MobileNetV2+Self-Attention models constructed by the combined attention mechanism compared with the MobileNetV2 model yielded the following results; the Kappa values improved by 1.26%, 0.02%, -0.93%, and 1.3%; the mean values of sensitivity improved by 0.36%, 0.21%, -0.34%, and 0.68%; the mean values of specificity improved by 0.36%, 0.04%, -0.22%, and 0.47%; and the mean values of AUC improved by 0.07%, -0.3%, -0.3%, and -0.17%, respectively. MobileNetV2+Self-Attention had the best classification result, with 92.77% accuracy, 88.92% Kappa value, 91.98% sensitivity mean, 96.59% specificity mean, 8.02% false positive mean, 3.41% false negative mean, and 98.00% AUC mean, and the model has favorable diagnostic value. Among the four different attention mechanisms fused on the MobileNetV2 network structure, only the model classification results are reduced after adding the ECANet attention module, which is analyzed as the difference in the



**Figure 5 Receiver operating characteristic (ROC) curves of the eight models** A: ROC curve of the normal; B: ROC curve of the observe; C: ROC curve of the surgery.



**Figure 6 Heat map visualization** A: Original image; B: MobileNetV2 model heat map; C: MobileNetV2+Self-Attention model heat map.

K value of the one-dimensional convolutional kernel size affects the channel weight results, leading to poor classification results. Adding CBAM, SENet and Self-Attention attention module model classification results are improved, of which MobileNetV2+Self-Attention classification results are the best, analyzed as SENet and CBAM calculate two-dimensional attention weights, Self-Attention calculates three-dimensional attention weights. Self-Attention can pay more attention to the correlation between different parts of the input data, and can extract more comprehensive and effective image feature information, and the model training effect is good.

When compared with the VGG16, ResNet50, and EfficientNetB7 pterygium classification models, the MobileNetV2+Self-Attention model showed higher classification results than those of the VGG16 and ResNet50 models and lower than that of the EfficientNetB7 model. The results were the smallest in model size, model parametric quantity, and test time, at 281 MB, 8 million, 9ms and 138ms. The local computer configuration in the experiment matches well with the configuration of medical equipment in primary hospitals, and its test time can be used as a reference. The MobileNetV2+Self-Attention model classifies and recognizes pterygium with good results, and the model is small and the test time is short, which is

suitable for primary hospitals to apply it on local computers or mobile devices, and it has a good application value.

**MobileNetV2+Self-Attention Model Diagnosis Results** The model was constructed by combining the attention mechanism Self-Attention and MobileNetV2, which diagnosed 188 images of the normal anterior ocular segment, 165 images of pterygium in the observation phase, and 117 images of pterygium in the surgery phase (Table 4).

**Visualization Results** The gradient-weighted class activation map (Grad-CAM)<sup>[40]</sup> is a commonly used model visualization method that uses the form of a heat map to show the focused regions that the model focuses on when predicting an image as a target class. This study visualized the heat map for MobileNetV2+Self-Attention and MobileNetV2 pterygium classification models (Figure 6). The highlighted areas on the heat map match the actual lesion sites. Notedly, the Self-Attention mechanism results show better coverage of the lesion areas, more focused pockets of model attention, and lower noise.

## DISCUSSION

The occurrence and recurrence of pterygium is closely related to the environment and is highly prevalent in people who work outdoors. Bikbov *et al*<sup>[41]</sup> analyzed the prevalence of

**Table 4 MobileNetV2+Self-Attention model diagnosis results**

Ophthalmologist diagnosis	MobileNetV2+Self-Attention diagnosis			Total
	Normal	Observe	Surgery	
Normal	188	1	0	189
Observe	0	151	20	171
Surgery	0	13	97	110
Total	188	165	117	470

pterygium in more than 5000 local pterygium patients in 2019 and concluded that the prevalence of pterygium was higher in people living in rural areas. For rural areas, where specialized medical resources in ophthalmology are lacking, intelligent models can provide an effective diagnostic tool for primary hospitals and patients. In this study, we use deep learning image classification technology to construct a lightweight intelligent-assisted diagnostic model with full consideration of the equipment the intended population has access to. The model can, not only, intelligently assist in mass screening of pterygium patients in primary hospitals but also provide self-examination and treatment recommendations for outdoor populations.

In this study, four different attention mechanism modules were added to the MobileNetV2 network structure, of these MobileNetV2+Self-Attention achieved the best classification results. There are several implementations of the attention mechanism, but the core is expressed in weights, and important features are assigned larger weights. SENet, ECANet, and CBAM calculate attention weights in two dimensions, whereas Self-Attention calculates attention weights in three dimensions. The Self-Attention mechanism can pay more attention to the correlation between different parts of the input, as a result, it can extract more comprehensive and effective image feature information, resulting in better model training.

In this study, the classical image classification models VGG16, ResNet50, and EfficientNetB7 were selected for comparison. According to the results in Table 3, we see that the EfficientNetB7 model has the best classification rate. The MobileNetV2+Self-Attention model is compared with the EfficientNetB7 model; the accuracy difference is 1.7%; the Kappa value difference is 2.59%. However, in terms of model size, MobileNetV2+Self-Attention has 1/7, the number of parameters and uses approximately 1/8, the testing time on the server is approximately 1/7, and the testing time on the local computer is approximately 1/33. The EfficientNetB7 model is large, with a longer testing time and high-performance requirements for hardware devices, whereas the MobileNetV2+Self-Attention model is small, has a shorter testing time, and is more suitable for deployment in primary hospitals in mobile and embedded devices. A shortcoming of this study is that the classification effect of the lightweight

MobileNetV2+Self-Attention model still needs to be improved. Optimizing the model structure, may improve the classification effect.

Few studies have been conducted to achieve intelligent grading of pterygium using deep learning techniques. The available studies mainly focus on constructing two-classification models to detect pterygium images. The latest study in 2022 by Hung *et al*<sup>[34]</sup> proposed a deep learning grading system to detect pterygium images with an accuracy of 91.67%, sensitivity of 91.67%, and specificity of 91.67%. The three two-classification models trained among the three grades showed a mean accuracy of 88.62%, mean sensitivity of 83.7%, and mean specificity of 97.22%. In this study, pterygium disease images were divided into observation and surgery phases. If the observation and surgery phase images were grouped into the pterygium category, the MobileNetV2+Self-Attention model detected pterygium disease images with 99.79% accuracy, 99.47% sensitivity, 100% specificity, and 100% AUC value. The MobileNetV2+Self-Attention model recognized the observed and surgical phases of pterygium with 92.77% accuracy, 91.98% mean sensitivity, and 96.59% mean specificity. Compared to the results of Hung *et al*<sup>[34]</sup>, the MobileNetV2+Self-Attention model proposed in this study was able to detect pterygium images better. In terms of grading recognition, the accuracy and sensitivity were better, but the specificity was slightly worse.

Deep learning models for pterygium can provide ophthalmologists with intelligent diagnostic decisions; however, erroneous decisions may prevent physicians from fully trusting these models. A heat map is a presentation form used to visualize the model decision. A heat map can highlight the image focus areas with which the model decision is concerned. The Grad-CAM visualization method does not need to modify the original network structure and the algorithm to generate a simple and accurate heat map. As a result, heat maps are widely used in different network models. The MobileNetV2+Self-Attention model proposed in this study provides intelligent diagnostic decisions using visualized heat maps to further assist the diagnosis. The heat maps highlight areas that match the actual lesion area, which is consistent with the ophthalmologist’s diagnostic approach and facilitates the physician’s understanding and trust in the model.

The limitation of this study is that the classification effect of the lightweight MobileNetV2+Self-Attention model still needs to be improved. In the future, our team will collaborate with multiple hospitals to conduct multi center research, and optimize the model structure to improve classification performance.

In conclusion, this study combined the Self-Attention mechanism and MobileNetV2 to develop a model for intelligent diagnosis

of pterygium disease. The MobileNetV2+Self-Attention model can be used not only for the detection of pterygium, but also for grading pterygium severity. The model has excellent application value because of its high classification accuracy, small size, short testing time on low-configuration computers, and straightforward visualization heat map. This study is hoped to bring a useful tool for intelligent diagnosis of pterygium, which will bring convenience to primary care physicians and patients and obtain good social benefits.

#### ACKNOWLEDGEMENTS

**Authors' contributions:** Wu MN and He K wrote the manuscript. Zhu SJ guided the experiments. Zheng B, Xi WQ, and Zhang Z planned experiments and the manuscript. Wu MN, Xi WQ, and Zhang Z reviewed the manuscript. He K trained the model. Yu YB trained the model and edited the manuscript. Hong YQ planned experiments and reviewed the manuscript. All authors issued final approval for the version to be submitted.

**Foundations:** Supported by the National Natural Science Foundation of China (No.61906066); Scientific Research Fund of Zhejiang Provincial Education Department (No. Y202147191); Huzhou University Graduate Research Innovation Project (No.2020KYCX21); Sanming Project of Medicine in Shenzhen (SZSM202311012); Shenzhen Science and Technology Program (No.JCYJ20220530153604010).

**Conflicts of Interest:** Wu MN, None; He K, None; Yu YB, None; Zheng B, None; Zhu SJ, None; Hong XQ, None; Xi WQ, None; Zhang Z, None.

#### REFERENCES

- 1 Chu WK, Choi HL, Bhat AK, Jhanji V. Pterygium: new insights. *Eye (Lond)* 2020;34(6):1047-1050.
- 2 Shao Y, Jie Y, Liu ZG, Expert Workgroup of Guidelines for the application of artificial intelligence in the diagnosis of anterior segment diseases (2023); Ophthalmic Imaging and Intelligent Medicine Branch of Chinese Medicine Education Association; Ophthalmology Committee of International Association of Translational Medicine; Chinese Ophthalmic Imaging Study Groups. Guidelines for the application of artificial intelligence in the diagnosis of anterior segment diseases (2023). *Int J Ophthalmol* 2023;16(9):1373-1385.
- 3 Rezvan F, Khabazkhoob M, Hooshmand E, Yekta A, Saatchi M, Hashemi H. Prevalence and risk factors of pterygium: a systematic review and meta-analysis. *Surv Ophthalmol* 2018;63(5):719-735.
- 4 Ke HQ, Zhang WJ, Liu H, et al. Prevalence and risk factors of pterygium in Zhuang and Miao nationality adults aged 40 and over in Wenshan prefecture, Yunnan Province. *Guoji Yanke Zazhi(Int Eye Sci)* 2022;22(2):347-351.
- 5 Akbari M. Update on overview of pterygium and its surgical management. *J Popul Ther Clin Pharmacol* 2022;29(4):e30-e45.
- 6 Yang M, Guan Y, Kang LH, et al. Meta-analysis of the prevalence of pterygium in people aged 40 years and above in China. *Chinese Journal of Experimental Ophthalmology* 2019;4(03):190-196.
- 7 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(6):84-90.
- 8 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014:arXiv:1409.1556. <http://arxiv.org/abs/1409.1556>
- 9 He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. IEEE, 2016:770-778.
- 10 Howard AG, Zhu ML, Chen B, Kalenichenko D, Wang WJ, Weyand T, Andreetto M, Adam H. MobileNets: efficient convolutional neural networks for mobile vision applications. 2017:arXiv:1704.04861. <http://arxiv.org/abs/1704.04861>
- 11 Howard A, Zhmoginov A, Chen LC, et al. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. 2018:arXiv:1801.04381. <https://arxiv.org/pdf/1801.04381>
- 12 Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3.2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019:1314-1324.
- 13 Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318(22):2211-2223.
- 14 Yang WH, Zheng B, Wu MN, Zhu SJ, Fei FQ, Weng M, Zhang X, Lu PR. An evaluation system of fundus photograph-based intelligent diagnostic technology for diabetic retinopathy and applicability for research. *Diabetes Ther* 2019;10(5):1811-1822.
- 15 Verbraak FD, Abramoff MD, Bausch GCF, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. *Diabetes Care* 2019;42(4):651-656.
- 16 Sotoudeh-Paima S, Jodeiri A, Hajizadeh F, Soltanian-Zadeh H. Multi-scale convolutional neural network for automated AMD classification using retinal OCT images. *Comput Biol Med* 2022;144:105368.
- 17 Peng YF, Dharssi S, Chen QY, Keenan TD, Agrón E, Wong WT, Chew EY, Lu ZY. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology* 2019;126(4):565-575.
- 18 Greenwald MF, Danford ID, Shahrawat M, et al. Evaluation of artificial intelligence-based telemedicine screening for retinopathy of prematurity. *J AAPOS* 2020;24(3):160-162.
- 19 Wang J, Ji J, Zhang MZ, et al. Automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening. *JAMA Netw Open* 2021;4(5):e218758.
- 20 Chang J, Lee J, Ha A, et al. Explaining the rationale of deep learning glaucoma decisions with adversarial examples. *Ophthalmology* 2021;128(1):78-88.
- 21 Li F, Yan L, Wang YG, Shi JX, Chen H, Zhang XD, Jiang MS, Wu ZZ, Zhou KQ. Deep learning-based automated detection of glaucomatous optic neuropathy on color fundus photographs. *Graefes Arch Clin Exp Ophthalmol* 2020;258(4):851-867.

- 22 Cen LP, Ji J, Lin JW, *et al.* Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat Commun* 2021;12(1):4828.
- 23 Zheng B, Jiang Q, Lu B, He K, Wu MN, Hao XL, Zhou HX, Zhu SJ, Yang WH. Five-category intelligent auxiliary diagnosis model of common fundus diseases based on fundus images. *Transl Vis Sci Technol* 2021;10(7):20.
- 24 He K, Wu MN, Zheng B, *et al.* Research on the automatic classification system of pterygium based on deep learning. *Guoji Yanke Zazhi(Int Eye Sci)* 2022;22(5):711-715.
- 25 Zhu SJ, Fang XW, Zheng B, *et al.* Research on segmentation of pterygium lesions based on convolutional neural networks. *Guoji Yanke Zazhi(Int Eye Sci)* 2022;22(6):1016-1019.
- 26 López Y, Aguilera LR. *Automatic classification of pterygium-non pterygium images using deep learning.* In: Tavares J, Natal Jorge R. (eds) VipIMAGE 2019. Lecture Notes in Computational Vision and Biomechanics, vol 34. Springer, Cham.
- 27 Wan Zaki WMD, Mat Daud M, Abdani SR, Hussain A, Mutalib HA. Automated pterygium detection method of anterior segment photographed images. *Comput Methods Programs Biomed* 2018;154: 71-78.
- 28 Zulkifley MA, Abdani SR, Zulkifley NH. Pterygium-Net: a deep learning approach to pterygium detection and localization. *Multimed Tools Appl* 2019;78(24):34563-34584.
- 29 Abdani SR, Zulkifley MA, Hussain A. Compact convolutional neural networks for pterygium classification using transfer learning. *2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. Kuala Lumpur, Malaysia. IEEE, 2019:140-143.
- 30 Zamani NSM, Zaki WMDW, Huddin AB, Hussain A, Mutalib HA, Ali A. Automated pterygium detection using deep neural network. *IEEE Access* 2022;10:191659-191672.
- 31 Abdani SR, Zulkifley MA, Moubark AM. *Pterygium tissues segmentation using densely connected deeplab.* 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Malaysia, 2020:229-232.
- 32 Abdani SR, Zulkifley MA, Zulkifley NH. Group and shuffle convolutional neural networks with pyramid pooling module for automated pterygium segmentation. *Diagnostics (Basel)* 2021;11(6):1104.
- 33 Deng J, Dong W, Socher R, Li LJ, Kai L, Li FF. ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA. IEEE, 2009:248-255.
- 34 Hung KH, Lin C, Roan J, *et al.* Application of a deep learning system in pterygium grading and further prediction of recurrence with slit lamp photographs. *Diagnostics* 2022;12(4):888.
- 35 Gumus K, Erkilic K, Topaktas D, Colin J. Effect of pterygia on refractive indices, corneal topography, and ocular aberrations. *Cornea* 2011;30(1):24-29.
- 36 Hu J, Shen L, Sun G. *Squeeze-and-excitation networks.* 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018:7132-7141.
- 37 Wang QL, Wu BG, Zhu PF, Li PH, Zuo WM, Hu QH. *ECA-net: efficient channel attention for deep convolutional neural networks.* 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020:11531-11539.
- 38 Woo S, Park J, Lee JY, *et al.* CBAM: convolutional block attention module. 2018:arXiv:1807.06521. <https://arxiv.org/pdf/1807.06521>
- 39 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. 2017:arXiv:1706.03762. <https://arxiv.org/pdf/1706.03762>
- 40 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-CAM: visual explanations from deep networks via gradient-based localization.* 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017:618-626.
- 41 Bikbov MM, Zainullin RM, Kazakbaeva GM, *et al.* Pterygium prevalence and its associations in a Russian population: the Ural eye and medical study. *Am J Ophthalmol* 2019;205:27-34.