applied sciences

MDPI

Article

Hybrid Deep Learning Model for Cataract Diagnosis Assistance

Zonghong Feng [1,*], Kai Xu [1], Liangchang Li [1] and Yong Wang [2,3,*]

[1] School of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou 730070, China; 12231769@stu.lzjtu.edu.cn (K.X.); 12231771@stu.lzjtu.edu.cn (L.L.)
[2] School of Sciences, Southwest Petroleum University, Chengdu 610500, China
[3] Key Laboratory of Numerical Simulation of Sichuan Provincial Universities, College of Mathematics and Information Sciences, Neijiang Normal Univeristy, Neijiang 641000, China
* Correspondence: fzh@mail.lzjtu.cn (Z.F.); wangyong@swpu.edu.cn (Y.W.)

Abstract: With the population aging globally, cataracts have become one of the main causes of vision impairment. Early diagnosis and treatment of cataracts are crucial for preventing blindness. However, the use of deep learning models for assisting in the diagnosis of cataracts is limited due to reasons such as scarce data labeling, small sample size, uneven distribution, and poor generalization ability in the field. Therefore, this paper proposes a hybrid deep learning network for assisting in the diagnosis of cataract fundus images, attempting to solve the above problems and limitations. The network is based on the idea of transfer learning for feature extraction of fundus images, and introduces the Squeeze-and-Excitation (SE) module and prototype network for feature enhancement and classification, improving the model's generalization ability for new disease samples. Finally, this paper verifies the role of each part of the model through ablation experiments, especially the significant contribution of the SE_block module and the prototype network classifier in enhancing the model's performance. The experimental results show that the proposed model achieves excellent performance in the task of cataract fundus image recognition, with an accuracy of 0.9875, AUC value of 0.9984, and F1 score of 0.9855. The establishment of this hybrid model not only provides an effective tool for the auxiliary diagnosis of cataracts but also provides a new perspective and method for the application of deep learning in the field of ophthalmic disease recognition.

Keywords: cataract; ancillary care; transfer learning; Squeeze and Excitation; prototype network

Citation: Feng, Z.; Xu, K.; Li, L.; Wang, Y. Hybrid Deep Learning Model for Cataract Diagnosis Assistance. *Appl. Sci.* **2024**, *14*, 11314. https://doi.org/10.3390/app142311314

Academic Editor: Christos Bouras

Received: 29 October 2024
Revised: 23 November 2024
Accepted: 25 November 2024
Published: 4 December 2024

1. Introduction

1.1. Research Background

Medical imaging disease classification has always been a key aspect of clinical care and related teaching activities, and ophthalmic disease recognition is an important branch of the field. According to statistics, about 2.2 billion people worldwide suffer from visual impairment, nearly half of which can be prevented by timely detection and treatment [1]. Ocular disease recognition utilizes imaging techniques such as fundus photography and optical coherence tomography (OCT) for early diagnosis and classification of ocular diseases. OCT technology captures a detailed cross-sectional view of the retina by non-invasive means. The human retina is a layer of light-sensitive tissue at the back of the eye. Incident light is converted into neural signals by receptors in the retina and processed by the visual cortex of the brain to produce images. Visual function may be impaired by a variety of pathological changes in the retina [2]. Analysis of images using optical coherence tomography (OCT) can be effective in diagnosing a variety of eye diseases including glaucoma (GLA), diabetic fundopathy (DR), and age-related macular degeneration (AMD). In recent years, with the development of deep learning techniques, research on ophthalmic disease recognition has made significant progress, and convolutional neural networks (CNNs) play a particularly important role in ophthalmic disease recognition. They can automatically learn and recognize disease features from fundus images, greatly improving the

efficiency and accuracy of diagnosis. CNNs show great potential in the screening and diagnosis of common ophthalmic diseases such as diabetic retinopathy, assisting doctors in understanding the diagnostic process by providing visual feedback of lesions.

Although deep learning shows great potential in assisted diagnosis of ophthalmic diseases, the high cost of annotation, small data volume, and uneven sample distribution of medical image data limit the training and generalization ability of the model. In addition, the differences in the characteristics of fundus images captured by different devices, as well as in people of different ages and occupations, put higher requirements on the generalization and accuracy of models. How to be able to effectively deal with these challenges is a future research direction.

1.2. Research Issues

Reduced light transmission or color change in the lens is called a cataract, which is a degenerative eye disease [3]. This lesion can weaken vision and in severe cases lead to complete blindness. It is associated with a number of factors, including aging, genetic factors, external injuries, and metabolic disorders. The prevalence of cataracts continues to rise as the global population ages. According to the World Health Organization [4], the number of people blinded by cataracts may increase to 40 million by 2025. Cataracts are the leading cause of vision loss globally and are responsible for half of all blindness in low- and middle-income countries [5,6]. Cataract fundus vs. normal fundus images are shown in Figure 1.

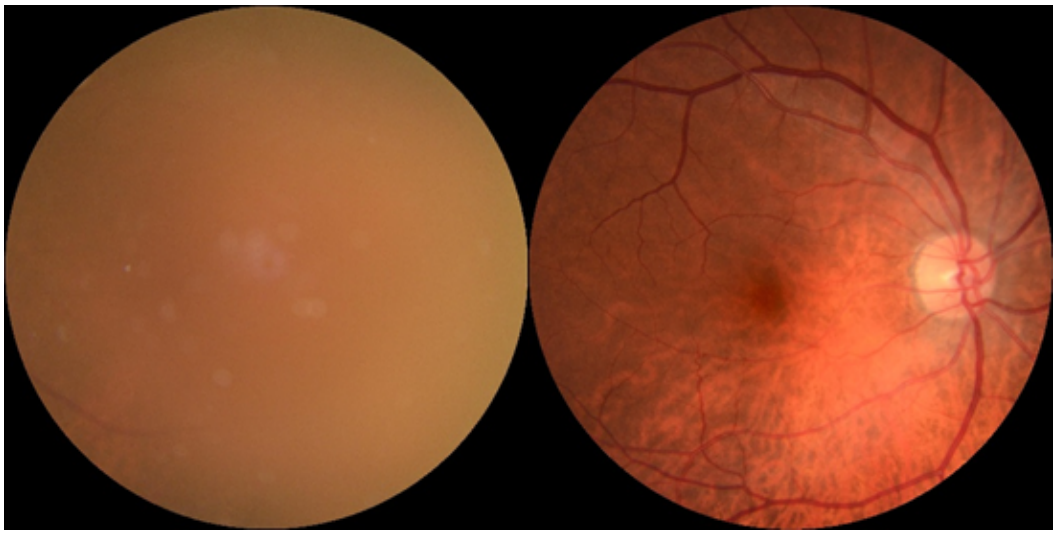

Figure 1. Cataract fundus image (**left**) vs. normal fundus (**right**).

Currently, there are many issues as well as challenges regarding deep learning models for assisted diagnosis of cataract disease:

(1) Scarcity and imbalance of datasets: Obtaining large datasets of cataract fundus images may be difficult because cataract diagnosis usually requires specialized medical equipment and doctors with specialized skills. In addition, the collection and use of such data requires extra care and handling due to privacy and ethical considerations [7]. Meanwhile, the number of samples of healthy fundus images in a given cataract image dataset is much larger than the number of samples of diseased images. This imbalance may result in deep learning models having better recognition ability for a larger number of categories and insufficient recognition ability for a smaller number of categories.

(2) Model generalization ability: Existing diagnostic methods rely heavily on the subjective judgment of professional ophthalmologists, which is a limiting factor in resource-poor areas, and the characteristics of cataract fundus images may vary across regions

and populations. How to design deep learning models to ensure that they have good generalization ability and can adapt to image features in different devices and populations is an extremely critical issue.

(3) Interpretability of models: Deep learning models are often considered as "black box" models, and their decision-making process lacks transparency. Through effective extraction and enhancement of features, and combined with actual clinical practice, the interpretability of deep learning models can be improved to truly achieve the purpose of assisted diagnosis, so that doctors and patients can understand the diagnostic basis of the model.

The main contributions of this paper can be summarized as follows:

1. In this paper, we propose a new hybrid deep learning network for cataract fundus image-assisted diagnosis, which aims to enhance the importance of cataract fundus image channel features to improve the generalization ability and classification accuracy of the model.
2. By integrating multiple open-source cataract fundus image classification datasets and applying data augmentation technology, a more diverse and representative fundus image dataset was constructed to solve the problems of unbalanced image labeling and quantity distribution in a single dataset.
3. Through ablation experiments, the effectiveness of the backbone network and each component used in the model was verified, especially the key role of the SE_block and prototype networks in improving the performance of the model, and excellent performance was achieved in the cataract fundus image recognition task.
4. In terms of model interpretation, the model enhances the sensitivity of the network to key information features, highlights the feature channels that have the greatest impact on the classification decision, and finally the distance-based classification method also provides an intuitive explanation for the model's decision making.

2. Related Work

2.1. Application of Deep Learning in Medical Image Recognition

In recent years, ophthalmic diseases have become the leading cause of vision loss, and the development of deep learning technology has brought new opportunities for early diagnosis and treatment. Deep learning models such as convolutional neural networks (CNNs) perform well in image recognition tasks, so they are widely used in ophthalmic medical image analysis.

Google's medical imaging research group has developed a deep learning-based algorithm that can automatically detect lesions such as microhemangiomas, hemorrhages, and exudations in retinal images. The algorithm first uses transfer learning to pre-train the model foundation on a massive general image dataset, and then fine-tunes it to adapt to a smaller medical image dataset. In terms of detecting DR, its performance is comparable to that of professional ophthalmologists, with a TP value of 0.95 and a TN value of 0.94 [8]. While these results are encouraging, the generalization ability of this algorithm in different populations and pathological changes must be carefully considered, especially when the data volume is small and diversity is limited. The deep learning model developed by Mimoun et al. achieves high accuracy in the detection and classification of AMD at different stages, with an AUC value of 0.89 through fundus image analysis, combined with local details and global contextual information, and multi-scale feature and depth feature fusion technology [9]. Still, this technology falls short when it comes to the recognition of subtle changes caused by AMD at different stages; especially in the early and late stages of disease progression, the accuracy of the model may be limited by data noise or image quality, which can affect the decision support for disease prognosis and treatment options. The deep learning model developed by Poplin et al. utilizes an attention mechanism that enables the model to focus on areas of the image that are closely related to glaucoma development, with a detection accuracy of 0.90, a TP value of 0.87, and a TN value of 0.92 [10]. However, the complex pathological characteristics of glaucoma and individual differences may affect

the accuracy of the model, especially in the early stages of the lesion, where the changes in the optic nerve head are subtle and may not be captured by the algorithm. Therefore, while these deep learning models demonstrate impressive diagnostic capabilities, their generalization capabilities and stability need to be treated with caution in practical applications. Future studies can validate models with large-scale, multicenter, multi-ethnic medical image datasets to ensure their reliability and interpretability in clinical applications.

Deep learning techniques are promising for application to ophthalmic disease recognition and have demonstrated their effectiveness in several disease areas. Wang et al. [11] integrated the SE module into CNNs such as ResNet and DenseNet to enhance the focus on tumor regional features in the study of breast cancer diagnosis, thereby improving the model's classification performance for benign and malignant tumors. However, the channel attention mechanism of the SE module focuses on the weight allocation of global features, and may ignore some subtle features in small or marginal regions of tumor images. Zhang et al. [12] proposed SE-UNet in their study, applying the SE module to the encoding and decoding module of U-Net to improve the segmentation effect of skin lesions, but the effect may be reduced on low-resolution or noisy images. Li et al. [13] used the SE module to improve the ResNet structure for distinguishing X-ray images from normal images in COVID-19 patients, as well as images from other lung diseases, and achieved better diagnostic results than traditional CNNs. However, the model focuses on the feature distribution of the training dataset, which leads to the poor generalization ability of the model under new data or changes in imaging conditions.

2.2. Cataract Fundus Image Recognition

In medical diagnosis, the diagnosis of cataract usually relies on slit lamp microscopy to observe the degree of opacity of the lens, combined with comprehensive information such as the patient's visual condition and medical history. However, due to the lack of resources of professional ophthalmologists and the large number of patients, many patients are not able to get treatment in a timely manner. In 2010, a study [14] first proposed a method of using AI to recognize slit lamp microscope images to diagnose nuclear cataracts, and established a 38-point shape model to identify the nuclear region of the lens and extract key features from it for grading. Grading is based on a comparison of the patient image to four standard images. The results show that only under the premise of accurate structure recognition can the features be extracted to achieve automatic grading. Unlike previous traditional methods of averaging the entire lens, the system is able to automatically identify nuclear regions in slit lamp images, and test results in more than 5000 images show that about 0.95 of the images can be automatically diagnosed. However, this technology still faces certain limitations, especially when image quality is limited or the nuclear region is occluded. In addition, the system is designed with user intervention to deal with special cases caused by inaccurate focal length, small pupils or ptosis, etc., although the system adaptability is enhanced, but the manual intervention also reflects the strong dependence of the algorithm on specific samples, which limits its ability to fully automate.

With the advancement of artificial intelligence, cataract diagnosis algorithms have been further improved and optimized. For example, in 2019, researchers proposed a novel "multi-feature fusion" strategy based on deep learning [7,15], which can classify cataracts into six different grades based on fundus images. Firstly, the deep neural network is used to extract fundus image features, and texture features are extracted from the original image and vascular image, and the classification method of multi-model training and ensemble learning is used to reduce the classification error and improve the accuracy of hierarchical diagnosis. The final results showed that the accuracy of the method for the six-level classification of cataracts reached an average of 0.9266, and the accuracy of the four-level classification was as high as 0.9475, which was at least 0.0175 higher than that of the traditional method. However, this method is effective when the image dataset is rich and the training is sufficient, but it has high requirements for data distribution and quality, and the effect may be reduced for poor images or noisy samples.

Although these new methods provide significant improvements in diagnostic efficiency, it is still necessary to consider their generalization ability and stability under different populations and fundus structural differences. The multi-feature fusion strategy performs well in the diagnosis of moderate-to-severe cataracts, but the detection of mild cataracts is uncertain, and there are adaptability problems in the environment, with limited shooting conditions. In addition, over-reliance on AI systems may lead doctors to overlook some subtle pathological features [16]. In the future, such algorithms should be further validated in multi-ethnic and multi-environment databases, and the reliability of clinical application should be improved based on doctors' experience to ensure that they are both efficient and robust in practical applications.

3. Method

3.1. Dataset

ODIR-5K dataset—The ODIR-5K dataset is a multi-label classification dataset of fundus images released by a Chinese team, which was released in the "Wisdom Eye" competition held by Peking University in 2019, containing a total of 5000 paired fundus image data of the left and right eyes, of which 3500 labeled datasets were released as training data. Compared with other fundus datasets, in addition to the differences in tasks (multi-label classification), the most important difference of ODIR-5K is that it provides paired left and right eye data and related descriptions, which can be used as 7000 individual data or as 3500 pairs of special data for some consistency exploration. In the ODIR-5K, a total of 8 categories of labels are provided, which are normal (N), diabetic retinopathy (D), glaucoma (G), cataract (C), age-related macular degeneration (A), hypertensive retinopathy (H), myopia (M), and other diseases/abnormalities (O). These diseases include many common diseases in ophthalmology, and there are also long-tail problems in the dataset that can be explored.

Retina dataset—The Retina dataset contains 601 retinal fundus images in 4 categories: 300 normal images, 100 cataract images, 101 glaucoma images, and 100 retina disease images. The data are available from GitHub—yiweichen04/retina_dataset: The Retina dataset contains images of (1) normal, (2) cataract, (3) glaucoma, (4) retina disease.

Cataract dataset—The Cataract dataset contains 2112 retinal fundus images in 2 categories, including 1074 normal images and 1038 cataract images. The Cataract dataset is available at https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification (accessed on 24 November 2024).

In this paper, normal fundus images and cataract fundus images were extracted from the three datasets to complete the experiment.

3.2. Data Preprocessing

Identification of ocular diseases based on fundus image datasets is an important branch in medically assisted diagnosis, especially in the diagnosis and research of ophthalmic diseases such as cataracts. However, due to the quality and scale of fundus image datasets, there are many challenges in auxiliary diagnosis.

When it comes to data quality, images can have artifacts, or be out of focus, underexposed, or overexposed due to shooting conditions, equipment quality, or patient cooperation. At the same time, due to the need for professional knowledge for image annotation, the same image may appear to be labeled with different annotations. In addition, some common eye diseases (e.g., myopia) may have far more image samples than other diseases, resulting in an imbalance of categories in the dataset, which may affect the model's ability to identify a small number of categories.

Data size issues: High-quality fundus image datasets may be limited by the difficulty and cost of data collection, resulting in a limited number of samples being available; datasets need to be representative of real-world diversity, including samples of different ages, genders, ethnicities, and stages of disease. If the dataset lacks diversity, this may affect the generalization ability of the model; as medical knowledge advances and diagnostic

criteria are updated, existing datasets may need to be updated and maintained regularly to maintain their relevance and validity.

In order to solve the above problems, this paper collects three datasets and extracts cataract-related fundus images to form the final dataset. Table 1 shows the distribution of the datasets.

Table 1. Dataset distribution.

Dataset	Diseased Fundus	Normal Fundus	Data Source
ODIR-5K	500	588	Competition
Cataract dataset	1038	1074	Kaggle
Retina	100	300	GitHub
Total	1638	1962	3600

In order to solve the above problems, this paper combines three datasets, and then extracts cataract-related fundus images and adopts data augmentation operations. Data augmentation is an important technique to improve the generalization ability and performance of models. The dataset is enriched by a series of transformations of the original image to generate a new image sample. In the field of medical imaging, the application of data augmentation is particularly important due to the high cost of acquiring a large number of annotated images.

There are a variety of strategies that can be employed when performing data augmentation of cataract fundus images. Common data augmentation operations include rotating, scaling, clipping, flipping, and adding noise, which can increase the diversity of data while better simulating real-world clinical scenarios and maintaining the same image content. For example, rotation can simulate different angles of fundus images caused by the natural tilt of the patient's head, and zoom can simulate imaging at different focal lengths.

When applying data augmentation, this paper adopts a series of image transformation operations to enrich the diversity of data samples and improve the generalization ability of the model. These data enhancement operations include rotation, scaling, cropping, flipping, and brightness and contrast adjustments, etc., and in order to avoid introducing excessive noise or distortion, the amplitude scale of each enhancement operation is set to 0.2. Specifically, the angle range of the rotation operation is set to ±20°, the zoom is limited to 0.2, and the brightness and contrast adjustments do not vary by more than 0.2. These settings enable the enhanced image to retain the key features of the original data, avoiding the distortion of image features caused by excessive transformation, so that the model can effectively learn the key features of the disease. In addition, by controlling the amplitude of the enhancement, it is ensured that the data transformation is carried out within a reasonable range, so as to avoid the model learning irrelevant features or noise [17]. Excessive data augmentation may make it difficult for the model to identify the true features of the lesion area, and even cause the model to overfit the noise. Therefore, setting the amplitude scale to 0.2 can effectively balance image diversity and data authenticity, and enhance the adaptability of the model to different images.

After the data enhancement is completed, the enhanced image is re-labeled to ensure the accuracy of the label information; that is, the label of each enhanced image is consistent with the original image. In this way, the model can learn consistent disease features on different forms of images, which helps to improve the model's ability to generalize unseen data. In addition, training and validation experiments are carried out on the enhanced dataset, and the results show that the moderate data augmentation strategy significantly improves the accuracy and robustness of the model, so that the model has better performance under different lighting, angles, and imaging equipment.A partial data-enhanced fundus image is shown in Figure 2.

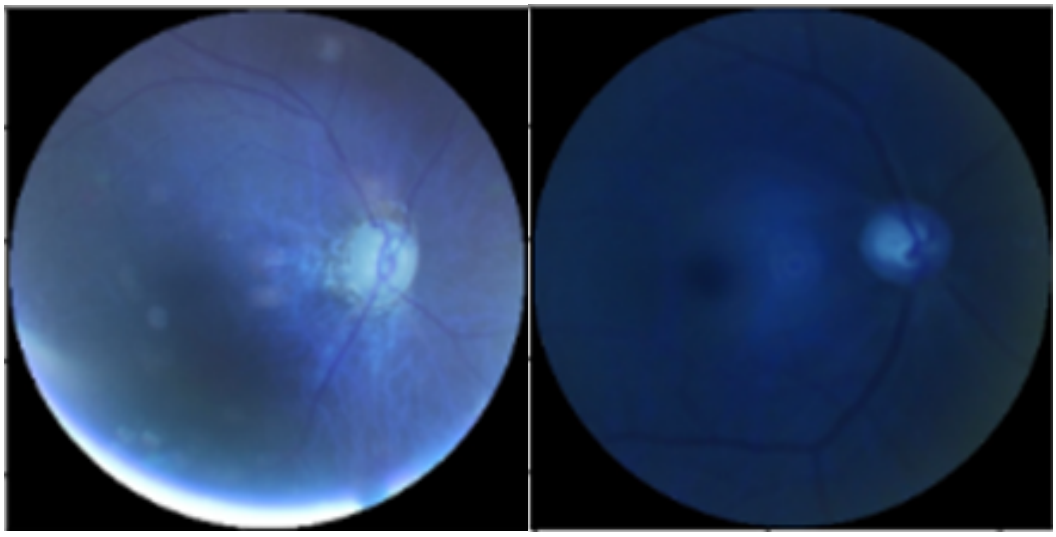

Figure 2. Adjustment of fundus image pixels and contrast.

3.3. Model Architecture

ResNet50 is a deep convolutional neural network, and its the core of its design is aimed at solving the problems of gradient vanishing and gradient explosion in deep networks, as well as the problem of network performance degrading with increasing depth. ResNet50 allows the network to learn the residual mapping between inputs and outputs by introducing "residual learning", rather than directly learning the unprocessed feature mapping. The network consists of multiple residual blocks, each containing a convolutional layer, batch normalization, and activation functions. The design of the residual block allows signals in the network to bypass some layers and propagate directly, which helps the gradient backpropagate more efficiently during training. The model also uses the residual block of the bottleneck structure, which is compressed and expanded by a convolutional layer of 1×1 and a convolutional layer of 3×3 is in the middle, which effectively reduces the amount of computation and the number of parameters [18].

The complete structure of the ResNet50 model is shown in Figure 3.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 3. ResNet50 model layer structure.

For cataract fundus images, ResNet50 can extract rich features from the images, including low-level edge and texture information and high-level semantic information, which provides great performance for cataract recognition in fundus images. At the same time, ResNet50 is flexible and scalable, and when used as the backbone network of other complex models, it can adjust and optimize the network structure according to specific tasks to better adapt to new tasks.

Therefore, in the experiment in this paper, the ResNet50 model was used as the backbone network to extract different types of fundus image features, pre-trained on the ImageNet dataset, and then part of the dataset was extracted for parameter fine-tuning to better adapt to the cataract ocular disease recognition task.

3.3.1. Feature Enhancement

The Squeeze-and-Excitation block (SE_block) mainly consists of two parts, which are Squeeze and Excitation [19].

The input for $\mathbf{F}_{tr}$ in the structure is $\mathbf{X}$, and $\mathbf{X} \in \mathbb{R}^{H' \times W' \times C'}$, and the output is U, and $U \in \mathbb{R}^{H \times W \times C}$, where F_{tr} is considered as a simple convolution operation, denoted by the formula $V = [v_1, v_2, \ldots, v_c]$, in which v_c represents the c convolution kernel; and the output is represented by the formula $U = [u_1, u_2, \ldots, u_c]$; then, there is

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s \tag{1}$$

where $*$ denotes the convolution, $v_c = [v_c^1, v_c^2, \ldots, v_c^{C'}]$, $X = [x^1, x^2, \ldots, x^{C'}]$, and $u_c \in \mathbb{R}^{H \times W}$, v_c^s is a 2D spatial convolution, where the bias term is simplified.

Typically, the correlation between channels is implicitly expressed by the summing of all channels, and this expression is intertwined with the capture of local spatial features. The sensitivity of the network to key information features cna be enhanced by explicitly defining dependencies between channels, which can then be effectively utilized at a deeper level of the network. Through this design, the network can not only learn local features, but also understand the global relationship between channels, so as to improve the ability of feature expression, and ultimately improve the performance of the entire network. The roles of the two phases (Squeeze and Excitation) are detailed below.

The role of the Squeeze part is to obtain the global information embedding for each channel of the feature representation U. In the SE_block, this is achieved by global mean pooling (GAP), where the average value of the features of each channel $c \in 1, 2, \ldots, C$ is

$$z_c = \mathbf{F}_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j) \tag{2}$$

The eigenvalues obtained by the above equation are global, and the resulting eigenvectors are global.

The role of the Excitation section is to learn the feature weights of each channel in C through z_c, and there are three main design principles: flexibility, simplicity, and non-exclusivity. Flexibility ensures that the weights obtained from learning have practical application value and can be adapted to different feature importance; simplicity ensures the training efficiency of the network after adding SE_blocks. Finally, we recognize that the relationships between channels are not completely independent, so our goal is to make the learned weights reinforce those features that are critical to the task while suppressing those that are less important. Based on these principles, the SE_block adopts a gating mechanism consisting of two fully connected layers, which dynamically adjusts the contribution of each channel. The gating unit s (i.e., the $1 \times 1 \times C$ eigenvector in Figure 4) is calculated as follows:

$$s = \mathbf{F}_{ex}(z, W) = \sigma(g(z, \mathbf{W})) = \sigma(g(\mathbf{W}_2 \delta(\mathbf{W}_1 z))) \tag{3}$$

where δ is the ReLU function and σ is the sigmoid function, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are the weight matrices of the two fully connected layers, respectively, and r is the number of hidden-layer nodes in the middle layer.

After obtaining the gating unit s, the final output $\tilde{X}$ is represented as the vector product of s and U, the operation $F_{scale}(\cdot,\cdot)$ in Figure 4:

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(u_c, s_c) = u_c \cdot s_c \tag{4}$$

where $\tilde{\mathbf{x}}_c$ is a feature representation of a feature channel of $\tilde{\mathbf{X}}$, and s_c is a constant in the gating unit s.

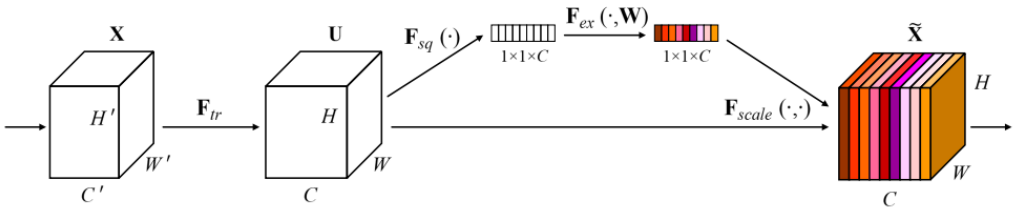

Figure 4. SE_block structure.

Spatial attention focuses on the importance of different areas in an image, i.e., how much a pixel value at a particular location contributes to a task. With spatial attention, the model can learn to focus attention on specific areas of the image to capture and emphasize important spatial features in the image, such as textures, shapes, edges, and more. Spatial attention enables the model to perform more fine-grained processing of local areas of the image, which helps to improve the model's ability to understand the shape and structure of objects. Channel attention focuses on the importance of different channels in the image; that is, the degree to which the feature map of the different channels contributes to the task. With channel attention, the model can learn to focus attention on a channel-specific feature map to capture and emphasize important channel features in the image, such as color, texture, frequency, and so on. Channel attention enables the model to better understand the relationship between different channels in the image, which helps to improve the model's semantic understanding of objects [20].

For fundus images, channel attention is significantly more important, which is mainly manifested in the following:

(1) Fundus images may contain a wealth of information, such as blood vessels, spots, retina, etc. The channel attention mechanism can help the model automatically learn and pay attention to the most relevant and important feature channels, thereby improving the performance of the model.
(2) There may be a large amount of redundant information in the fundus image, and the channel attention mechanism can help the model to suppress those channels that are irrelevant or unimportant for the classification task, so as to improve the generalization ability and robustness of the model.
(3) The channel attention mechanism can improve the interpretability of the model, making the results predicted by the model easier to understand and accept. This is especially important for applications in the medical field, where physicians need to understand how models make predictions and make clinical decisions based on the predictions.

3.3.2. Classifier

The underlying classification module in this experiment is improved and reorganized on the basis of the classical few-shot learning network, the prototypical network. The prototype network was proposed by Jake Snell et al., and its basic principle is to make vectors of the same type closer together and different types of vectors farther apart [21]. The mean of the features of each category is called the prototype of that category (c_k), and by calculating the distance between the test samples and each prototype, the prototype with the smallest distance is taken as the category of the test sample. The class prototype calculation formula is shown in (5):

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i) \tag{5}$$

where $f_\phi(x_i)$ is the mapping function $f_\phi : \mathbb{R}^D \rightarrow \mathbb{R}^M$, which is used to map the D dimensional sample data to the M-dimensional space. $S = (x_1, y_1, \ldots, (x_N, y_N))$ is a set of small-scale N-type label-supported datasets in FSL learning, x is a vector representation of D-dimensional sample data, y is its corresponding category, and S_k represents a dataset of category k.The class prototype diagram is shown in Figure 5. Then, given the distance function $d : \mathbb{R}^M \times \mathbb{R}^M \rightarrow [0, \infty]$, the prototype network generates a class distribution of query point x based on softmax at the distance to the prototype in the embedded space:

$$p_\phi(y = k|x) = \frac{exp(-d(f_\phi(x), c_k))}{\sum_k exp(-d(f_\phi(x), c_{k'}))} \tag{6}$$

Learning is performed by minimizing the negative logarithmic probability of the real class k by SGD, and the loss function is

$$J(\phi) = -\log p_\phi(y = k|x) \tag{7}$$

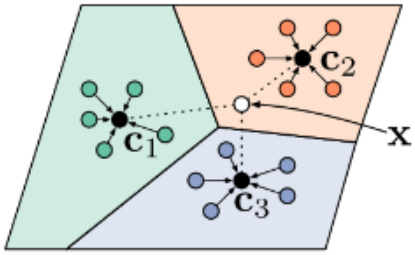

Figure 5. Schematic representation of a class archetype in Prototypical Networks.

In this experimental study, Prototypical Networks is used as the underlying classifier, and in terms of feature representation, it enhances the feature representation by learning the center or prototype of the category, which is particularly important when processing medical images with complex patterns [22]. Secondly, Prototypical Networks has strong generalization ability and can be well generalized to unseen data, which is crucial for accuracy and reliability in practical clinical applications. In addition, Prototypical Networks is insensitive to noise and outliers and shows good robustness, which is very beneficial for medical image processing that may contain noise. The decision-making process of the model is based on the category center, which not only improves the prediction accuracy of the model, but also enhances its explanatory power, helping doctors understand the diagnostic basis of the model. In terms of computational efficiency, Prototypical Networks typically have lower computational complexity than other complex classifiers, which makes them more practical in resource-constrained medical settings. At the same time, the adaptability of Prototypical Networks allows it to be adapted to different datasets and task requirements to improve classification performance [23].

4. Results

4.1. Details

The purpose of this experiment was to construct a deep learning-based model for identifying and classifying cataract fundus images. The neural network model used in the experiment uses ResNet50 as the backbone network, and integrates the Squeeze-and-Excitation (SE) module and the prototype network for feature extraction and classification. The model architecture diagram is shown in Figure 6.

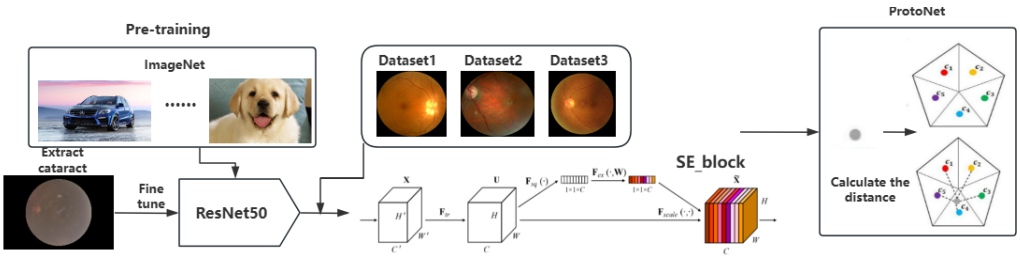

Figure 6. Schematic diagram of the model architecture.

During the experiment, the backbone network was selected by using five pre-trained models, including ResNet50, ResNet152, and the InceptionV3 architecture. The constructed dataset is divided into training set and validation set in proportions of 0.8 and 0.2, respectively. The results show that the accuracy of the ResNet50 model can reach 0.9622, and the AUC value and F1 score are 0.9878 and 0.9797, respectively, which are ahead of the backbone network model. After selecting the backbone network model, the ResNet50 pre-trained weights without the top fully connected layer is first loaded to take advantage of the feature extraction capabilities obtained from the pre-training on the ImageNet dataset. All layers are set to be trainable, 0.05 of the dataset is extracted for parameter fine-tuning, and then the weight of the pre-trained parameters is updated to better adapt to this task. Next, an SE_block is defined to enhance the feature representation. The SE module learns the relationship between channels through global average pooling and two fully connected layers, and then recalibrates features using the sigmoid activation function to reinforce important features and suppress unimportant features. The output of the backbone network is further extracted through a convolutional layer, and then the SE_block is applied again for feature enhancement. Finally, the data are prepared for classification through batch normalization and global average pooling layers.

In the low-level classification, i.e., prototype network part, a category prototype is generated from a fully connected layer with 128-dimensional output, and then processed using L2 normalization to measure the distance between the test sample and the prototype using the Euclidean distance in subsequent classifiers. The classifier is a single-class binary output layer that uses a sigmoid activation function and is suitable for binary classification problems. The entire model was compiled using the Adam optimizer and the binary cross-entropy loss function.

In the model training, the dropout method was used, the value was set to 0.5, and 0.5 of the neurons were randomly discarded to avoid model overfitting. In order to improve the generalization ability of the model and avoid overfitting, the model checkpoint and EarlyStopping callbacks were set up experimentally. Model checkpoints are used to preserve the best model when the accuracy of the validation set is improved, while early stop terminates training when the accuracy of the validation set has not been improved for several epochs in a row. Finally, the model was trained on 100 epochs at a batch size of 32 samples and monitored using a validation set during the training process. At the end of the training, the model was evaluated using the test set and output loss and accuracy. All experimental work was carried out on an NVIDIA T4 GPU with 16 GB of memory.

4.2. Analysis

In this paper, the accuracy, AUC value, and F1 score index are used to evaluate the performance of the proposed model. Accuracy refers to "the number of samples correctly predicted ÷ total number of samples"; the AUC curve is a performance estimator for classification problems; an AUC value close to 1 means that the model performs better in classifying disease labels. Similarly, the F1 score is defined as the harmonized average of the recall and accuracy values, as shown in Equations (8) and (9). The specific construction formulae are as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{9}$$

where $precision = \frac{TP}{FP+TP}$, and $recall = \frac{TP}{FN+TP}$.

After model training, testing, and verification on the validation set, the final accuracy, AUC, and F1 values can reach 0.9875, 0.9984, and 0.9855, respectively, which are 2.53, 0.58, and 1.06 percentage points higher than the initial backbone network model, respectively, indicating that the designed model architecture has a certain effect on this study, and also shows the effectiveness of the SE_block and prototype network classifiers for fundus image recognition. By looking at the training curve, it was found that the model designed in this experiment was better than other models in terms of convergence speed and final accuracy.The model training curve and loss curve are shown in Figure 7.

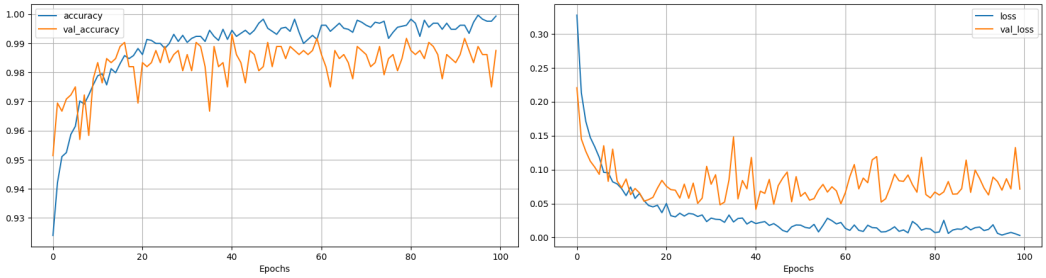

Figure 7. Model training curves.

From the calculated confusion matrix, it can be seen that the model performs well in both the cataract and normal categories, and can effectively reduce the risk of misdiagnosis and missed diagnosis in practical applications. The model was tested and the results of the confusion matrix are shown in Figure 8.

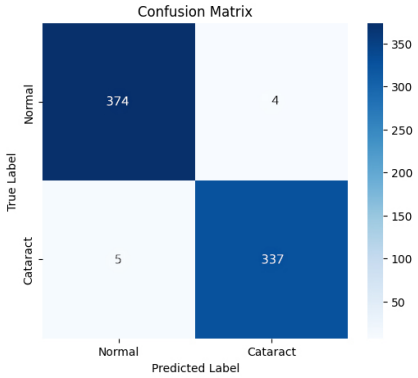

Figure 8. Confusion matrix for the cataract detection model.

5. Ablation Experiments

In this paper, ablation experiments were performed to evaluate the contribution of different components in the model to the overall performance. By progressively removing or modifying parts of the model, we can observe the impact of these changes on the performance of the model and understand the role and importance of each component. The ablation experiments in this paper mainly focus on the selection of backbone networks and the impact of the addition of the SE_block and prototype network classifiers on the performance of the model.

First, we performed ablation experiments on different backbone networks (see Table 2), including CNN, VGG16, ResNet50, ResNet152, and InceptionV3. These networks were trained and tested on the same dataset, and the results are shown in Table 2. As can be seen from the table, ResNet50 is better than other backbone networks in terms of accuracy, AUC (area under the curve), and F1 score, which indicates that ResNet50 has a strong ability

in feature extraction and can learn features that are helpful for classification from fundus images more effectively.

Table 2. Comparison of the performance of different models.

Model	Accurary	AUC	F1
CNN	80.4	76.32	84.22
VGG16	95.67	96.32	95.88
ResNet50	**96.22**	**98.78**	**97.97**
ResNet152	95.67	97.23	96.51
InceptionV3	86.94	84.12	87.65

The bold font is the best accuracy result under the same conditions.

In the ablation experiment, we not only looked at the impact of the backbone network model selection model, but also looked at the performance when two modules were added to other backbone networks. This comparative analysis helps us to understand more comprehensively the role and effect of the SE_block and prototype network classifiers in different model frameworks.

Next, we performed ablation experiments on the ResNet50 model, the ResNet152 model, and the InceptionV3 model to evaluate the impact of the SE_block and prototype network classifiers. The experimental results are shown in Table 3. The original ResNet50 model already performed well in terms of accuracy, AUC, and F1 score, but the performance of the model was further improved with the addition of the SE_block and prototype network classifiers. Specifically, the accuracy of the ResNet50 model increased from 0.9622 to 0.9875, the AUC increased from 0.9878 to 0.9984, and the F1 score increased from 0.9797 to 0.9855. At the same time, according to the obtained results, it can be seen that the performance of the model also significantly improved after adding the SE_block and prototype network classifiers to the two backbone networks: ResNet152 and InceptionV3. For the ResNet152 model, the original model had an accuracy of 0.9567, an AUC of 0.9723, and an F1 score of 0.9651. After adding the SE_block and prototype network classifiers, the accuracy rate increased by 1.16 percentage points and the F1 score increased by 0.74 percentage points. For the InceptionV3 model, the accuracy of the original model was 0.8694, the AUC was 0.8412, and the F1 score was 0.8765. After adding the SE_block and prototype network classifiers, the accuracy increased by 2.61 percentage points, and the AUC increased by 3.1 percentage points. This indicates that the introduction of the SE_block and prototype network classifiers not only enhances the expressive ability of features, but also improves the classification accuracy of cataract fundus images in the model.

From these ablation experiments, we can draw the following conclusions:

(1) By introducing the channel attention mechanism, the SE_block can enhance the expression ability of features in different models and improve the sensitivity of the model to key information.
(2) As a small-shot learning method, the prototype network classifier can improve the generalization ability of the model to new samples, especially in the field of medical images, which helps to improve the accuracy of classification and the explanatory nature of the model.
(3) Different backbone networks can achieve performance improvement after fusing the SE_block and prototype network classifiers, but the magnitude of the improvement may vary depending on the model. This may be related to the differences in the structural characteristics and feature extraction capabilities of each backbone network.

In future research, we can further explore the application of the SE_block and prototype network classifiers in other types of deep learning models, as well as their performance on different datasets and tasks. In addition, it is possible to investigate how to optimize the structure and parameters of the two modules to achieve more efficient feature learning and classification performance.

Table 3. Comparison of the performance of different deep learning models.

Model	Accurary (%)	AUC (%)	F1 Score (%)
ResNet152	95.67	97.23	96.51
ResNet152++	**96.83**	97.03	**97.25**
InceptionV3	86.94	84.12	87.65
InceptionV3++	**89.25**	**87.22**	87.12
ResNet50	96.22	98.78	97.97
ResNet50++	**98.75**	**99.84**	**98.55**

The bold font is the best accuracy result under the same conditions.

5.1. Supplementary Experiment 1

In order to demonstrate the effectiveness of the model on an external dataset, we performed two complementary experiments.

In supplementary experiment 1, we took different numbers of cataract patients of different age groups (35–50, 51–60, 61–70, 71–80, and over 81 years old) in the OIA-ODIR dataset, and also predicted the extracted images according to gender classification. The specific projections are shown in Table 4.

Table 4. Supplementary experiment 1 sample table.

Age Group	Sample Size	Gender Ratio (Male–Female)
35–50	10	6:4
51–60	25	13:12
61–70	43	14:29
71–80	41	20:21
80+	16	6:10

From the results of supplementary experiment 1, we analyzed the fundus images with detailed stratification by age and sex ratio, and showed a high accuracy rate for patients aged 61–80 years, which may be related to the obvious cataract characteristics in this age group; and another reason may be that cataract symptoms are more common in older patients, and the number of fundus images is relatively large, and thus they are easier to collect. In addition, the analysis of sex ratio shows that gender difference has little effect on the predictions of our model, with the prediction accuracy for male patients being slightly higher than that for female patients. This slight bias in gender recognition is worthy of further research and optimization. The results are shown in Figure 9.

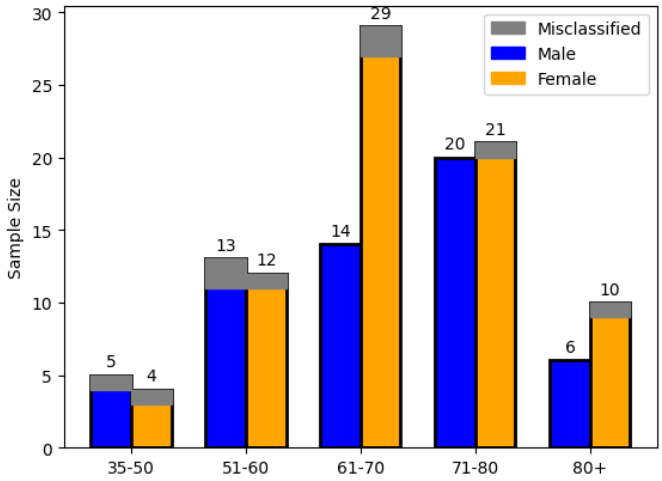

Figure 9. Supplementary experiment 1 results.

5.2. Supplementary Experiment 2

In Supplementary Experiment 2, we used an external dataset collected by kaggle [24], and 100 images of cataract disease and 100 images of normal fundi were taken to predict the disease. A confusion matrix was generated to show the classification effect, and the confusion matrix is shown in Figure 10. According to the classification of the confusion matrix, there were a total of 187 fundus images with a classification accuracy of 0.935, a precision calculation of 0.948, and a recall rate and F1 score of 0.92 and 0.934, respectively.

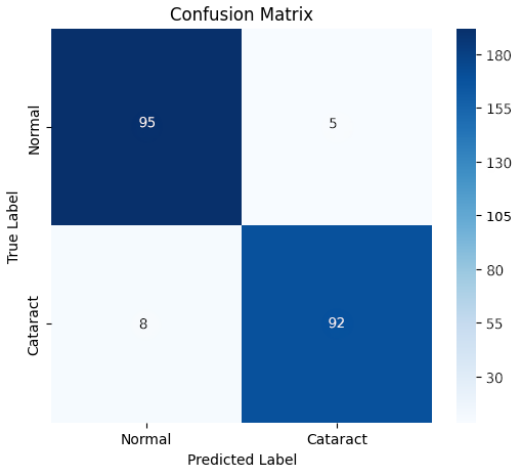

Figure 10. Supplementary experiment 2—confusion matrix.

6. Discussion

The hybrid deep learning model proposed in this paper has achieved remarkable results in assisting cataract diagnosis. The experimental results show that the model achieves high accuracy in the recognition task of cataract retinal images by using ResNet50 as the backbone network and integrating an SE_block and prototype network classifier. The SE module enhances the ability of feature expression through the channel attention mechanism, enabling the model to capture key image information more sensitively. At the same time, the prototype network classifier performs well in its ability to generalize to new samples, which is suitable for different sample distributions and shows strong robustness. These characteristics play an important role in the processing of medical images, especially cataract images, and greatly improve the practicability of the model.

The contribution of this study is the use of transfer learning and the combination of SE_blocks with prototype networks, which improves the adaptability of the model to new data while maintaining high accuracy, which has important implications for clinical settings. In the case of limited resources, this model can provide effective auxiliary support for physicians, thereby improving the coverage and efficiency of screening. In addition, the ablation experiments further verified the effectiveness of each module, and confirmed that both the SE_block and the prototype network significantly improved the recognition performance of the model.

However, there are some limitations to this study. First, the diversity of the dataset is limited, and more images of patients of different ages, genders, and ethnicities need to be included in the future to ensure the generalizability of the model in a wider population. In addition, the explanatory nature of the model is also key in medical applications, especially in the process of auxiliary diagnosis; medical staff need to understand the decision-making basis of the model. Therefore, future research may consider exploring more explanatory network structures or visualization methods to help doctors better understand the diagnostic results of the model. In terms of practical application, the feasibility and stability of the model in the clinical setting still need to be further verified. In the future, the model can be deployed in practical diagnostic work by working with experts, and the model structure and parameters can be optimized based on feedback to improve its performance in complex

medical scenarios. At the same time, it is also worth exploring further the applicability of SE_blocks and prototype networks in other medical imaging diagnostic tasks.

7. Conclusions

In this study, a hybrid model based on deep learning is proposed to assist cataract diagnosis. By introducing the SE module and the prototype network classifier, the feature extraction and classification capabilities are further enhanced on the basis of ResNet50. Experimental results show that the proposed model has achieved excellent performance in the cataract retinal image recognition task, with an accuracy of 0.9875, an AUC value of 0.9984, and an F1 score of 0.9855, which verifies the effectiveness of the model in auxiliary diagnosis. At the same time, the ablation experiments show that the SE module and the prototype network significantly improve the performance of the model, and these characteristics enhance the generalization ability of the model to new samples.

With the aging of the global population, the incidence of cataracts is increasing year by year. The model developed herein can effectively support doctors to make rapid and accurate diagnosis under the condition of limited resources, and improve the efficiency and coverage of cataract screening. However, future research still needs to further improve the diversity and interpretability of the model to adapt to the differences between different populations, the left and right eyes, and the need for model transparency in the medical field. By being deployed in a clinical setting, the model constructed in this study is expected to provide effective support in actual diagnosis.

Author Contributions: Conceptualization, Z.F. and K.X.; Methodology, K.X.; Software, K.X.; Validation, K.X., L.L. and Z.F.; Formal analysis, Z.F.; Investigation, K.X.; Data curation, L.L.; Writing—original draft, K.X.; Writing—review & editing, K.X., L.L. and Y.W.; Visualization, K.X.; Supervision, Z.F. and Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 12162020), the Young Scholars Science Foundation of Lanzhou Jiaotong University (No. 2020022), the Sichuan Provincial Natural Science Foundation (No. 2023NSFSC0428), the Central Government Funds of Guiding Local Scientific and Technological Development (No. 2023ZYD0004), the Sichuan National Applied Mathematics Center open fund (No. 2024-KFJJ-01-01), the Sichuan Province Science and Technology Innovation Seedling Engineering Cultivation Project (No. MZGC20240034), and the open fund for Key Laboratory of Numerical Simulation of Sichuan Provincial Universities (No. KLNS-2024SZFZ001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets were collected from publicly available sources and can all be found within the article, and further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors sincerely thank the referees for their valuable comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Khan, E.; Rehman, M.Z.U.; Ahmed, F.; Alfouzan, F.A.; Alzahrani, N.M.; Ahmad, J. Chest X-ray classification for the detection of COVID-19 using deep learning techniques. *Sensors* **2022**, *22*, 1211. [CrossRef] [PubMed]
2. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef] [PubMed]
3. Asbell, P.A.; Dualan, I.; Mindel, J.; Brocks, D.; Ahmad, M.; Epstein, S. Age-related cataract. *Lancet* **2005**, *365*, 599–609. [CrossRef] [PubMed]
4. Wang, W.; Yan, W.; Fotis, K.; Prasad, N.M.; Lansingh, V.C.; Taylor, H.R.; Finger, R.P.; Facciolo, D.; He, M. Cataract surgical rate and socioeconomics: A global study. *Investig. Ophthalmol. Vis. Sci.* **2016**, *57*, 5872–5881. [CrossRef] [PubMed]
5. Waltz, K.L.; Featherstone, K.; Tsai, L.; Trentacost, D. Clinical outcomes of TECNIS toric intraocular lens implantation after cataract removal in patients with corneal astigmatism. *Ophthalmology* **2015**, *122*, 39–47. [CrossRef] [PubMed]
6. West, S. Epidemiology of cataract: Accomplishments over 25 years and future directions. *Ophthalmic Epidemiol.* **2007**, *14*, 173–178. [CrossRef] [PubMed]

7. Zhang, H.; Niu, K.; Xiong, Y.; Yang, W.; He, Z.; Song, H. Automatic cataract grading methods based on deep learning. *Comput. Methods Programs Biomed.* **2019**, *182*, 104978. [CrossRef]
8. Ting, D.S.W.; Cheung, C.Y.L.; Lim, G.; Tan, G.S.W.; Quang, N.D.; Gan, A.; Hamzah, H.; Garcia-Franco, R.; San Yeo, I.Y.; Lee, S.Y.; 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*, 2211–2223. [CrossRef] [PubMed]
9. Xu, K.; Huang, S.; Yang, Z.; Zhang, Y.; Fang, Y.; Zheng, G.; Lin, B.; Zhou, M.; Sun, J. Automatic detection and differential diagnosis of age-related macular degeneration from color fundus photographs using deep learning with hierarchical vision transformer. *Comput. Biol. Med.* **2023**, *167*, 107616. [CrossRef] [PubMed]
10. Poplin, R.; Varadarajan, A.V.; Blumer, K.; Liu, Y.; McConnell, M.V.; Corrado, G.S.; Peng, L.; Webster, D.R. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2018**, *2*, 158–164. [CrossRef] [PubMed]
11. Wang, X.; Chen, Y. Breast Cancer Diagnosis Using Squeeze-and-Excitation Networks with ResNet and DenseNet Backbones. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1917–1926.
12. Zhang, Z.; Li, X.; Yang, J. Squeeze-and-Excitation UNet for Skin Lesion Segmentation. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 1–10.
13. Li, L.; Qin, L.; Xu, Z.; Zhang, H. COVID-19 Chest CT Image Segmentation Using Deep Learning. *J.-Comput.-Assist. Tomogr.* **2020**, *44*, 566–572.
14. Li, H.; Lim, J.H.; Liu, J.; Mitchell, P.; Tan, A.G.; Wang, J.J.; Wong, T.Y. A computer-aided diagnosis system of nuclear cataract. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1690–1698. [PubMed]
15. Zhang, X.Q.; Hu, Y.; Xiao, Z.J.; Fang, J.S.; Higashita, R.; Liu, J. Machine learning for cataract classification/grading on ophthalmic imaging modalities: A survey. *Mach. Intell. Res.* **2022**, *19*, 184–208. [CrossRef]
16. Ismail, W.N.; Alsalamah, H.A. A novel CatractNetDetect deep learning model for effective cataract classification through data fusion of fundus images. *Discov. Artif. Intell.* **2024**, *4*, 54. [CrossRef]
17. Imran, A.; Li, J.; Pei, Y.; Mokbal, F.M.; Yang, J.J.; Wang, Q. Enhanced intelligence using collective data augmentation for CNN based cataract detection. In *Frontier Computing: Theory, Technologies and Applications (FC 2019)*; Springer: Singapore, 2020; Volume 8, pp. 148–160.
18. Koonce, B.; Koonce, B.E. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: New York, NY, USA, 2021; pp. 109–123.
19. Theckedath, D.; Sedamkar, R.R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **2020**, *1*, 79. [CrossRef]
20. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
21. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
22. Mohammadi, S.F.; Sabbaghi, M.; Hadi, Z.; Hashemi, H.; Alizadeh, S.; Majdi, M.; Taee, F. Using artificial intelligence to predict the risk for posterior capsule opacification after phacoemulsification. *J. Cataract. Refract. Surg.* **2012**, *38*, 403–408. [CrossRef] [PubMed]
23. Laenen, S.; Bertinetto, L. On episodes, prototypical networks, and few-shot learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24581–24592.
24. Kaggle. (n.d.). *Cataract Dataset*. Available online: https://www.kaggle.com/datasets/kershrita/cataract (accessed on 24 November 2024).