
Fundus Image-Based Assisted Diagnosis Using Deep Learning Technique

Rohit Thanki¹, Dan Chen²

¹Portland State University, Portland, USA

²Portland State University, Portland, USA

*Corresponding author: Dan Chen; danc2784@pdx.edu

Abstract: Existing fundus image classification approaches mostly focus on grading a single type of ocular disease, while many deep neural networks still suffer from large parameter sizes and high computational cost. To address these issues, this work develops a multi-label fundus image assisted diagnosis system built on the lightweight SqueezeNet architecture. The model is trained and evaluated on the public ODIR dataset, and the results demonstrate that the proposed system exhibits strong feature extraction capability. It significantly accelerates inference and reduces memory consumption without causing substantial performance degradation. Owing to its compact structure and high efficiency, the system is well suited for deployment on resource-constrained hardware platforms and provides useful insights for integrating artificial intelligence into computer-aided medical diagnostic devices.

Keywords: deep learning, fundus imaging, lightweight model, hardware deployment

1. Introduction

In ophthalmic clinics, ophthalmologists can evaluate and diagnose the ocular health of patients through fundus images. Early detection, early diagnosis, and early treatment are essential for managing ophthalmic diseases, and most importantly, regular fundus examinations are required. According to authoritative WHO statistics, by 2025, the number of individuals with visual impairments in China had reached as high as 150 million, and an average ophthalmic specialist would need to handle visual health issues for more than 5,000 patients. Faced with the growing demand for evaluating and diagnosing large volumes of fundus images, challenges such as slow manual interpretation, significant variability, high labor intensity, and insufficient workforce have become increasingly evident.[1] Therefore, the research and development of intelligent and automated fundus disease-assisted diagnostic systems have great clinical value for early screening, postoperative evaluation, and treatment monitoring of ophthalmic diseases and related conditions.

With the development of big data, AI-assisted diagnosis based on medical datasets has become increasingly common. Intelligent diagnostic systems using fundus images have shown excellent sensitivity and specificity in screening and diagnosing diseases such as diabetic retinopathy (DR), glaucoma, cataracts, refractive errors, age-related macular degeneration, and other retinal disorders, significantly improving ophthalmologists' diagnostic accuracy [2-6] while greatly reducing screening costs. Deep learning originates from the study of artificial neural networks, which learn patterns and hierarchical representations from large-scale datasets. Such patterns have been successfully leveraged in image classification, object detection, and speech processing, and have far surpassed the performance of traditional techniques. In 2018, the world's first

FDA-approved autonomous AI diagnostic system for DR, known as IDx-DR, was launched in the United States [7]. It can be used to screen patients aged 22 or older for DR severity. In 2024, China's National Medical Products Administration approved two DR fundus image-assisted diagnostic software products, which adopt neural network-based autonomous deep learning structures. These systems are trained and validated using large-scale fundus datasets with labeled categories.

Through research on domestic fundus image-assisted diagnostic systems, it has been found that current systems mainly address binary or ternary classification problems and focus primarily on differentiating disease severity levels. Their ability to support ophthalmologists in initial screening remains limited. At the same time, the rapidly increasing size of medical image datasets and extended model inference times introduce additional challenges. Therefore, this study establishes and designs a lightweight deep learning model to achieve a multi-classification intelligent fundus diagnostic system on mobile devices. By reducing the number of model parameters, the system improves diagnostic efficiency and accuracy, offering valuable support for the screening and diagnosis of fundus diseases.

2. Data Acquisition and Preprocessing for the Deep Learning Based Fundus Image Assisted Diagnosis System

In a typical workflow of a deep learning based fundus image assisted diagnosis system, large amounts of raw data, data preprocessing, model design, training, and evaluation are all indispensable. The quality of the dataset largely determines the performance of the deep learning model and the effectiveness of the applied algorithm. Therefore, sufficient

data collection, cleaning, and annotation are essential to improve system performance and enhance the practical value of deep learning models in intelligent diagnosis of ophthalmic diseases.

The dataset used in this study originates from the Ocular Disease Intelligent Recognition (ODIR) competition held at Peking University in 2024. The dataset contains more than 5000 pairs of color fundus images collected from collaborating hospitals through regular ophthalmic examinations. Each sample is labeled by ophthalmologists with professional diagnostic information, including patient ID, age, gender, diagnostic keywords, and diagnosis results. For each patient, both left and right eyes are labeled with one or more disease categories. The labels include: N for Normal, D for Diabetic Retinopathy, G for Glaucoma, C for Cataract, A for Age-related Macular Degeneration, H for Hypertensive Retinopathy, M for Pathologic Myopia, and O for Others. The corresponding disease label sample distribution is shown in Figure 1.

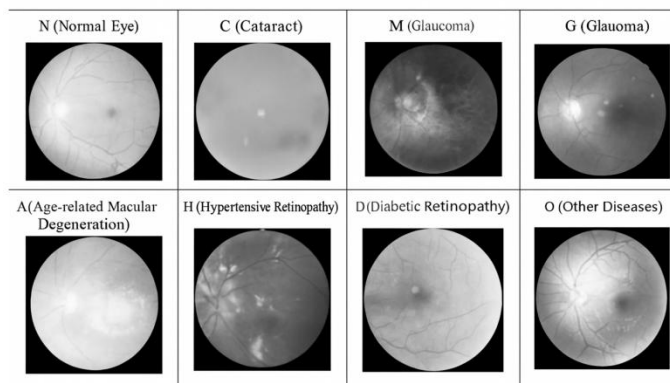


Figure 1. Sample Distribution of Disease Labels in the Dataset

To ensure better applicability of the dataset in neural network training, this study performs a series of preprocessing and augmentation procedures. The main workflow includes: original data cleaning, label re-generation, training-validation set split, and image resizing and augmentation.

2.1 Label Re-generation

The original dataset provides labels for both left and right eyes, meaning each pair of images corresponds to two separate labels. However, some diagnosis descriptions involve both eyes together, which may lead to inconsistent mappings between the textual description and the individual eye labels. Therefore, we re-generate paired labels based on the diagnostic keywords to ensure the reliability of training samples. While this process increases the labeling workload, it strengthens the logical consistency between classification categories and corresponding labels, further enhancing the generalization capability of the deep learning model.

The re-generated label process in this study is summarized as follows:

(1) Based on professional diagnostic information in the ODIR dataset, we construct a mapping table between 8 disease categories and their corresponding labels. A total of 98 disease subtypes were identified under label O (Others). Since these

subtypes have weak correlation and indistinct classification characteristics, they are not used directly in this study. Additionally, four noise categories were created, including poor image quality, image blurring, camera overexposure, and invalid fundus images, resulting in 26 unusable images that were removed.

(2) Disease keywords for each left and right eye were extracted separately and mapped to the corresponding disease category according to the disease-to-label mapping table. If the textual descriptions contained multiple disease types, multi-label mappings were generated to ensure consistency between disease types and labels.

(3) The mappings from steps (1) and (2) were merged to produce the final ID-based mapping table. Table 1 shows the sample statistics for different disease categories. The total dataset contains 10536 images. Among all fundus disease types, Normal and Diabetic Retinopathy account for the highest proportions, with 40.45% and 24.13% respectively. Label re-generation is an essential part of preprocessing, as it acts as the key step in forming reliable and logically consistent category labels.

2.2 Dataset Splitting for Training

In general, deep learning models are trained on labeled datasets, and the trained model is then evaluated on a validation set to assess its performance before being applied to predict labels for unseen data. To ensure that the model trained on the dataset possesses sufficient generalization capability, a randomized dataset splitting strategy is adopted. Considering that the input to the network framework in this study consists of single fundus images, we randomly allocate 10% of the samples from each disease category as the validation set, while the remaining 90% are used as the training set.

Table 1: Sample Counts and Proportions of Each Disease Category

Disease Category	Sample Count	Proportion/%
N	4262	40.45
D	2542	24.13
G	449	4.26
C	453	4.3
A	400	3.8
H	277	2.63
M	345	3.27
O	1808	17.16

Note: If a patient is associated with multiple diseases, the sample of that patient will be counted under each corresponding disease category.

2.3 Image Augmentation

Image augmentation is a technique used to modify training images so that the dataset can better adapt to practical scenarios and improve the overall training process[8].By performing augmentation operations, we can enhance

informative features, suppress undesired noise, reduce overfitting, and improve the representation learning capability of the network. In this study, several augmentation methods such as geometric transformation and cutout are employed. The main steps are as follows.

(1) The validation images are cropped from the center to remove excessive black backgrounds while preserving the essential fundus regions. The cropped images are then uniformly resized to 224×224 as required by the network input format. For the training images, random cropping is applied at a ratio of 0.6-1.0 of the original region, as illustrated in Figure 2. Furthermore, random geometric transformations such as horizontal flipping are performed to enrich the diversity of training samples.

(2) The images from both the validation set and the training set are converted into tensor format and normalized to accelerate network training and improve experimental accuracy.

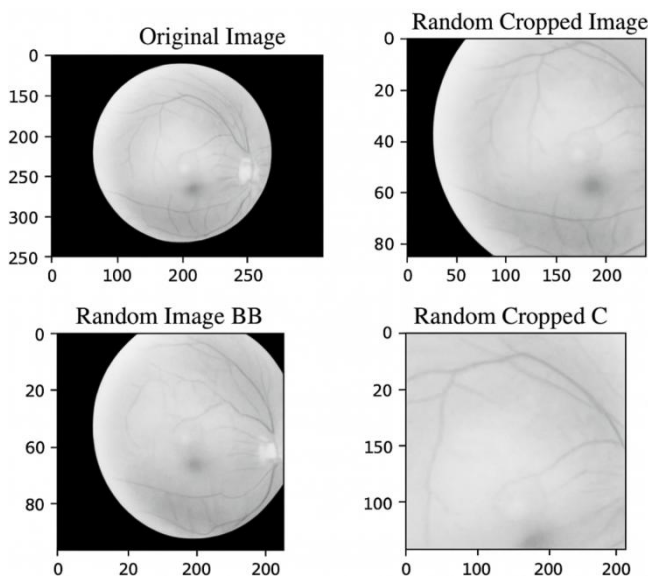


Figure 2. Random Cropping on the Scaled Images

3. Design and Implementation of the Deep Learning Based Fundus Image Assisted Diagnosis System

3.1 Model Selection

Most current fundus retinal image classification methods focus on identifying a single disease, while auxiliary diagnostic systems for joint multi-disease classification are still rarely applied in clinical practice. In existing studies on multi-disease classification, deep learning models typically adopt ResNet as the backbone network for feature extraction[9].

As shown in Figure 3, the ResNet-50 network is composed of multiple residual units, pooling layers, and fully connected layers. It possesses strong deep feature modeling capability and can effectively alleviate the vanishing gradient problem in deep neural networks.

The study in [10] provides important inspiration for our model design. Its use of graph attention mechanisms and temporal modeling strategies for complex medical data offers an effective architectural reference for our work. Building on these ideas, we construct a ResNet-50-based feature extraction network for multi-disease classification and further integrate attention mechanisms to enhance feature selection capability.

However, ResNet-based deep networks still suffer from large parameter scales and high computational costs, which limits their applicability on embedded hardware or resource-constrained platforms. Therefore, designing computationally efficient and more lightweight model architectures remains a key challenge for current multi-disease fundus image diagnostic systems. The ResNet-50 architecture adopted in this study is illustrated as follows.

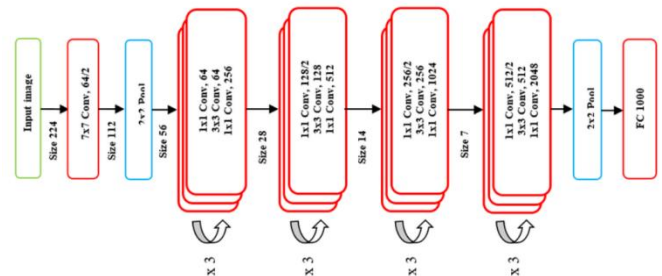


Figure 3. Network Architecture of ResNet50

Based on the above analysis, this study adopts a lightweight convolutional neural network model, SqueezeNet. SqueezeNet was proposed by Iandola [11] as a lightweight network architecture capable of achieving performance comparable to AlexNet on the ImageNet dataset while requiring 50 times fewer parameters. In our experiments, we compare SqueezeNet with ResNet50. ResNet50 occupies 89.9 MB of storage space, whereas SqueezeNet requires only 2.78 MB, making it approximately 32 times smaller. Due to its considerably reduced number of parameters, SqueezeNet can achieve comparable recognition accuracy while greatly decreasing model storage requirements and accelerating computation. This makes SqueezeNet more suitable for deployment on embedded platforms such as FPGAs (Field Programmable Gate Arrays) and ASICs (Application Specific Integrated Circuits).

The SqueezeNet architecture consists of convolutional layers, pooling layers, Fire modules, and a softmax classifier. A Fire module (the core component of SqueezeNet) performs feature extraction using a combination of 1×1 and 3×3 convolutions. A squeeze layer is first used to reduce the number of input channels, thereby compressing the feature representation and lowering the computational cost. Each Fire module contains two parts: a squeeze layer and an expand layer, as illustrated in Figure 4. The squeeze layer is composed of a set of continuous 1×1 convolution kernels that reduce the number of channels, while the expand layer is responsible for restoring feature dimensionality using continuous 1×1 and 3×3 convolution kernels.

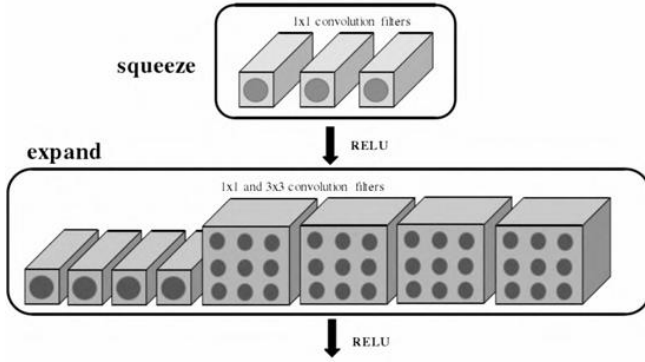


Figure 4. Schematic Diagram of the Fire Module [12]

3.2 System Architecture

The system architecture of the fundus image assisted diagnosis model designed in this study is illustrated in Figure 5. To classify seven types of conditions (six disease categories and one normal category), seven independent lightweight SqueezeNet deep learning models are introduced. Each SqueezeNet model is trained separately for its corresponding binary classification task. The outputs of the seven SqueezeNet models are then aggregated in the “Integrated Analysis Module,” where necessary data processing and analytical operations are performed. Based on the analysis results, the system finally outputs multi-label disease information.

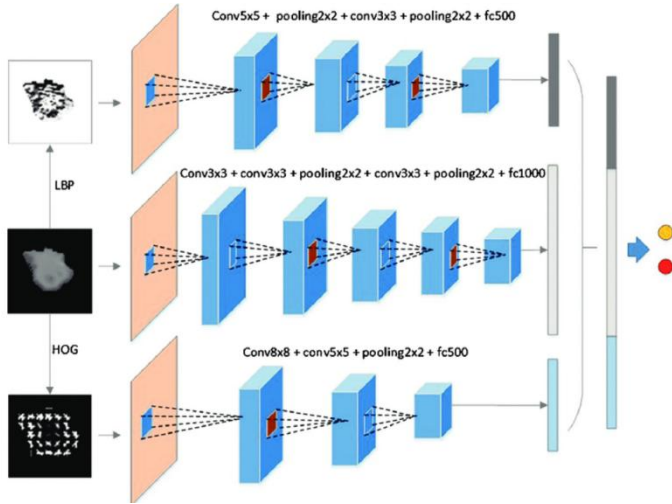


Figure 5. System Architecture Diagram

In deep learning training, a sufficiently large training dataset is typically required to continuously optimize the model parameters. When the dataset is relatively small but the number of model parameters is large, overfitting may occur. To mitigate the risk of overfitting caused by limited training data, this study applies transfer learning during the training process. Specifically, the pretrained SqueezeNet weights from the torchvision library (configured as weights='SqueezeNet1_1_Weights.DEFAULT') are imported into the project. The final classification module of each SqueezeNet branch is then replaced with a new module tailored to the binary classification task of this study, while keeping the

feature extraction layers unchanged. This approach allows the model to retain the pretrained feature extraction capability without requiring extensive retraining of the entire network.

The pretrained SqueezeNet model is trained on the ImageNet dataset, which contains 1,000 object categories and approximately 1.2 million images. This large-scale pretraining enables the model to learn highly robust visual features, which helps prevent overfitting and enhances the generalization capability of the model when applied to fundus disease classification.

3.3 System Performance Evaluation

1) Training and Testing Environment

The experimental training and testing environment used in this study is configured as follows. The system runs on Windows 10 (Chinese version) with an Intel i5-8265U CPU (4 cores and 8 threads). The Python version used is 3.9.12, and the PyTorch version is 1.12.0. The model training parameters are set as follows: the loss function is CrossEntropyLoss, the optimizer is SGD (stochastic gradient descent) with a learning rate of 0.001, momentum of 0.9, batch size of 64, and the total number of epochs is 50.

2) Evaluation Metrics

Common evaluation metrics for medical image classification include accuracy A (Accuracy), precision P (Precision), recall R (Recall), and the F-measure (F-Measure). The F-measure used here is the F1-score. A higher value indicates better classification performance. The specific formulas are given below:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$P = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$R = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

Taking DR binary classification as an example, TP (True Positive) refers to the number of DR samples correctly classified as DR samples, TN (True Negative) refers to the number of non-DR samples correctly classified as non-DR, FP (False Positive) refers to the number of non-DR samples incorrectly classified as DR, and FN (False Negative) refers to the number of DR samples incorrectly classified as other categories.

3) Results and Analysis

From the perspective of multi-label disease category prediction, the performance varies across single diseases, as shown in Figure 2. The best prediction results are obtained for cataract (C) and pathological myopia (M), which may be due to the more distinct lesion characteristics of these two diseases compared with others. The prediction accuracy for age-related macular degeneration reaches 88.09%, which is close to the

88.4% accuracy of the deep convolutional neural network (DCNN)-based intelligent fundus image analysis method reported in the literature [13].

From the perspective of precision (A), the prediction results for normal eyes and diabetic retinopathy (D) differ significantly. This is mainly because the number of samples in some disease categories is relatively small. If the model lacks the ability to identify such categories and predicts all samples as negative, the evaluation metrics for these categories will appear artificially high, introducing bias in assessing the model's performance for these diseases. Based on precision and F1-score metrics, the model finds it most challenging to learn and identify the lesion features of glaucoma (G) and hypertensive retinopathy (H), making these categories more prone to misclassification.

Table 2: Evaluation Metrics for Multi-Label Disease Prediction

Disease Category	Accuracy (A)/%	Precision (P)/%	Recall (R)/%	F1-Score/%
N	81.23	78.41	80.92	79.65
D	82.35	79.69	78.63	79.16
G	86.15	71.87	79.47	75.48
C	89.47	83.94	87.26	85.57
A	88.09	78.52	76.54	77.52
H	84.72	61.7	71.38	66.19
M	88.93	81.16	82.69	81.92

To verify the computational efficiency of the lightweight SqueezeNet model, this study compares the processing speed of SqueezeNet and ResNet50 in single-disease prediction. The model parameter size, storage footprint, and system runtime were evaluated, and the results are shown in Table 3. The experimental results indicate that the accuracy difference between SqueezeNet and ResNet50 is 5.4%. ResNet50 occupies 89.9 MB of storage, whereas SqueezeNet requires only 2.78 MB, making the SqueezeNet model 32 times smaller. When using a quad-core CPU to process 9000 images, the runtime of ResNet50 is 68.22 minutes, while SqueezeNet completes the task in only 8.7 minutes, achieving a speedup of 6.6 times. Despite the 5.4% difference in accuracy, SqueezeNet requires only 3.09% of the storage capacity and 13.0% of the runtime compared with ResNet50.

Table 3: Comparison of Runtime Metrics Between SqueezeNet and ResNet50

Model	Accuracy/%	Storage Space/MB	Runtime/min
SqueezeNet	81.2	2.78	8.97
ResNet50	86.6	89.9	68.22

Therefore, under the premise of maintaining reasonable classification accuracy, the lightweight SqueezeNet model substantially improves computational efficiency and enhances system usability when compared to the much larger ResNet50 model.

4. Conclusion

In summary, this study first re-generated the labels of the ODIR dataset to obtain a clean and reliable single-eye dataset. Image augmentation techniques were then employed to enrich the training samples. Based on these data, a lightweight SqueezeNet deep learning model was developed to construct a multi-label fundus image assisted diagnosis system. Experimental results demonstrate that the proposed model achieves outstanding multi-label classification performance across six disease categories and normal fundus images, with the highest single-label classification accuracy reaching 89.47%. This verifies the strong feature extraction capability of SqueezeNet.

Furthermore, SqueezeNet requires only 3.09% of the storage space and 13.14% of the runtime compared with the ResNet50 model, while achieving only a 5.4% reduction in accuracy. This significant improvement in model efficiency highlights the suitability of SqueezeNet for deployment in hardware-constrained environments.

Based on the lightweight deep learning framework constructed in this study, a multi-label fundus image assisted diagnosis system was designed and implemented. The system features a small number of model parameters and high computational efficiency, making it suitable for medical hardware deployment. To further enhance system performance, future work may explore model fusion strategies, attention mechanisms, and the integration of FPGA-based acceleration to optimize and refine the system.

References

- [1] Fu H, Cheng J, Xu Y, et al. Disc-aware ensemble network for glaucoma screening from fundus images. *IEEE Transactions on Medical Imaging*, 2018, 37(11): 2493-2501.
- [2] Roger S T W, Jaccard N, Carbonaro F, et al. Evaluation of an AI system for the automated detection of glaucoma from stereoscopic optic disc photographs: the European Optic Disc Assessment Study. *Eye (London)*, 2019, 11: 1791-1797.
- [3] Ting D S W, Pasquale L R, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, 2019, 2: 167-175.
- [4] N. Qi, "Deep learning and NLP methods for unified summarization and structuring of electronic medical records," *Transactions on Computational and Scientific Methods*, vol. 4, no. 3, 2024.
- [5] R. Hao, W. C. Chang, J. Hu and M. Gao, "Federated learning-driven health risk prediction on electronic health records under privacy constraints," [Journal/Conference Name], [vol./conf. info if available], 2025. Long E, Lin H, Liu Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nature Biomedical Engineering*, 2017, 2: 5-17.
- [6] Topol E J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 2019, 25(1): 44-56.
- [7] Salamon J, Bello J. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 2017, 3: 279-283.

- [8] Chea N, Nany M. Classification of fundus images based on deep learning for detecting eye diseases. *Computers, Materials & Continua*, 2021, 1: 411-426.
- [9] P. N. Karthikayan, Y. S. V. Varshan, H. G. Kattamuri and U. Jayaraman, "Explainable AI: Comparative Analysis of Normal and Dilated ResNet Models for Fundus Disease Classification," arXiv preprint arXiv:2407.05440, 2024.
- [10] X. Zhang and Q. Wang, "EEG anomaly detection using temporal graph attention for clinical applications," *Journal of Computer Technology and Software*, vol. 4, no. 7, 2025.
- [11] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 2014, 27: 3320-3328.
- [12] Burlina P M, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmology*, 2017, 135: 1170-1176.