Ophthalmic Disease Detection via Deep Learning With a Novel Mixture Loss Function

Xiong Luo, *Senior Member, IEEE*, Jianyuan Li, Maojian Chen, Xi Yang, and Xiangjun Li

***Abstract*—With the popularization of computer-aided diagnosis (CAD) technologies, more and more deep learning methods are developed to facilitate the detection of ophthalmic diseases. In this article, the deep learning-based detections for some common eye diseases, including cataract, glaucoma, and age-related macular degeneration (AMD), are analyzed. Generally speaking, morphological change in retina reveals the presence of eye disease. Then, while using some existing deep learning methods to achieve this analysis task, the satisfactory performance may not be given, since fundus images usually suffer from the impact of data imbalance and outliers. It is, therefore, expected that with the exploration of effective and robust deep learning algorithms, the detection performance could be further improved. Here, we propose a deep learning model combined with a novel mixture loss function to automatically detect eye diseases, through the analysis of retinal fundus color images. Specifically, given the good generalization and robustness of focal loss and correntropy-induced loss functions in addressing complex dataset with class imbalance and outliers, we present a mixture of those two losses in deep neural network model to improve the recognition performance of classifier for biomedical data. The proposed model is evaluated on a real-life ophthalmic dataset. Meanwhile, the performance of deep learning model with our proposed loss function is compared with the baseline models, while adopting accuracy, sensitivity, specificity, Kappa, and area under the receiver operating characteristic curve (AUC) as the evaluation metrics. The experimental results verify the effectiveness and robustness of the proposed algorithm.**

***Index Terms*—Deep learning, Ophthalmic disease detection, Convolutional neural network (CNN), Loss function.**

Manuscript received February 3, 2021; revised May 10, 2021; accepted May 19, 2021. Date of publication May 25, 2021; date of current version September 3, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants U1836106 and 81961138010, in part by the Beijing Natural Science Foundation under Grants 19L2029 and M21032,Beijing Intelligent Logistics System Collaborative Innovation Center under Grant BILSCIC-2019KF-08, in part by the Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB, under Grant BK20BF010, and in part by the Fundamental Research Funds for the University of Science, and Technology Beijing under Grant FRF-BD-19-012 A. *(Corresponding author: Xiong Luo.)*

Xiong Luo, Jianyuan Li, and Maojian Chen are with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China, and with the Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China, and also with the Shunde Graduate School, University of Science and Technology Beijing, Foshan 528399, China (e-mail: xluo@ustb.edu.cn; B20200341@xs.ustb.edu.cn; B20190321@xs.ustb.edu.cn).

Xi Yang is with Beijing Intelligent Logistics System Collaborative Innovation Center, Beijing 101149, China (e-mail: yangxi@bwu.edu.cn).

Xiangjun Li is with China Mobile Information Security Center, Beijing 100053, China (e-mail: lixiangjun@chinamobile.com).

Digital Object Identifier 10.1109/JBHI.2021.3083605

I. Introduction

OPHTHALMIC diseases can cause visual impairment and even blindness. Due to the increasing number of patients with eye diseases, the effective computer-aided diagnosis (CAD) methods for automatic diseases detection have caught the interest of both the academic and industrial communities [1], [2]. Specifically, the deep learning methods have been widely utilized in CAD to analyze biomedical data, while achieving ophthalmic diseases detection.

Generally speaking, among all kinds of ophthalmic data, retinal fundus color images are usually employed to assist clinician in diagnosing disease. The retina is the tissue at the back of the eyeball, including optic disc, macula, and vessel. Here, Fig. 1 demonstrates the retinal fundus images, which are fundus photograph showing retinal morphologies and pathologies. The morphological variation of retinal fundus images may reveal various eye diseases, including cataract, glaucoma, and age-related macular degeneration (AMD).

In general, cataract is diagnosed by many reasons such as aging, genetics, immune and metabolic abnormalities, trauma, radiation, and some others, which can cause lens metabolism disorders, leading to lens protein denaturation and opacity. Meanwhile, the light is obstructed by opaque lens and cannot be projected on the retina, resulting in blurred vision. The fundus image of a cataract eye is presented in Fig. 1(b). Glaucoma is a group of ocular diseases, the manifestations of which including optic disc atrophy and depression, visual field defects and visual loss. Pathologic hypertension and optic nerve blood supply deficiency are the main risk factors for the onset of glaucoma. The optic cup is a bright zone at the center of optic disc. In a retinal fundus image of a glaucomatous eye, such as Fig. 1(c), The ratio of the area of the optic cup to the optic disc is abnormal. AMD is an aging change in the structure of the macular area, which can cause vision impairment for the elderly. The main symptoms of macular spots are decreased central vision, central dark spots, and distortion of vision. The fundus has yellow-gray exudative lesions and oval hemorrhages in the macula, along with unclear borders and slight uplifts. Fig. 1(d) presents an AMD eye. Then, the classification tasks related to cataract, glaucoma, and AMD are considered in this article.

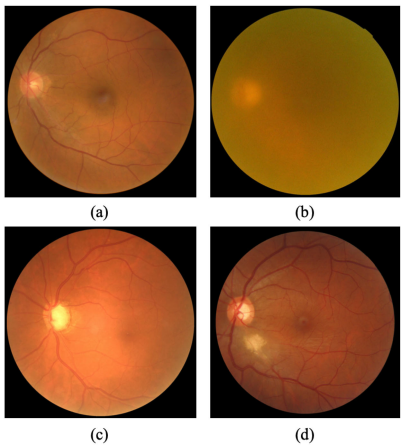

Fig. 1. Retinal fundus color image of different condition: (a) Normal; (b) Cataract; (c) Glaucoma; (d) AMD.

Deep learning is a subarea of machine learning [3], and it has been increasingly applied to process various medical tasks including retinal image analysis. Grassmann *et al.* defined 13 classes of AMD based on severity, and proposed a classification strategy based on six common convolutional neural networks (CNNs) to assess the stage of AMD [4]. Peng *et al.* presented a deep learning classification architecture, and achieved high accuracy on detection and prediction of AMD by simulating the human grading process [5]. The model firstly analyzes each AMD risk factor separately and then calculates the final severity score of AMD. Yan *et al.* analyzed both genetic information and image data with deep learning model to predict AMD progression [6]. A deep transfer learning strategy was proposed by Pratap *et al.* to detect various stage of cataract [7]. The classification is based on pre-trained CNN and support vector machine (SVM) classifier. Zhang *et al.* combined SVM and fully connected neural network (FCNN) to extract important features automatically and grade six-level cataract [8]. A deconvolution network method was implemented by Xu *et al.* to analyze global feature extraction process [9], while discussing the importance of detail information and proposing a cataract grading model based on both global feature and local feature. Li *et al.* recruited 21 ophthalmologists to diagnosis glaucoma based on color fundus images, and then investigated the efficacy of an inception-based CNN model for the detection of glaucomatous optic neuropathy [10]. An *et al.* combined SVM and five separate CNN models to differentiate glaucomatous and healthy subjects using optical coherence tomography (OCT) and color fundus photographs [11]. Bajwa *et al.* proposed a two-stage deep neural network to detect glaucoma [12]. The functions of two stages are optic disc extraction and subject classification, respectively.

Although significant progresses have been made in the deep learning-based processing and diagnostic analysis within this field, some existing deep learning methods still cannot fully evade the ill effect from data imbalance and outliers in fundus images, and then the satisfactory performance may not be given in some cases. To avoid such limitations and develop a more practical approach on the basis of retinal fundus images for eye diseases detection, we propose a deep learning-based classification network through the combination with focal loss and correntropy-induced loss functions. Focal loss is suitable to keep up with the ever-increasing demand for the imbalance issue, and it can increase the weight of hard samples. Additionally, compared with common classification loss functions, e.g., cross entropy loss or mean square error (MSE), correntropy loss has better generalization and robustness to noise and outlier.

The main contributions of this article are elaborated as follows:

1) In consideration of the effectiveness and robustness of focal loss and correntropy loss functions in addressing complex dataset with class imbalance and outliers, they are hereby combined to design a mixture loss function as the substitute of classic loss function for processing biomedical data.
2) After analyzing retinal fundus images, we develop a novel deep learning model to detect ophthalmic diseases through the combination of afore mentioned mixture loss function. The proposed model is specifically evaluated on a real-life ophthalmic dataset, and then some metrics, including accuracy, sensitivity, specificity, Kappa, and area under the receiver operating characteristic curve (AUC), are used to fully test the performance of this model. Furthermore, the classification performance of our model in addressing ophthalmic diseases data is accordingly discussed.

The remaining content is summarized as follows. In the second section, we give a simple introduction for relevant technologies. On the basis of afore mentioned technologies, our deep learning model is proposed in Section III. The results of experiments and the related discussion are provided in Section IV. Finally, we conclude this article in the last section.

II. Background

A related technology, i.e., CNN, is simply presented in this section. Generally, one of the most popular deep learning-based models is CNN [13]. Since CNN is adept in learning hierarchical features from images, it has been widely utilized in image analysis, such as image classification [14] and image segmentation [15].

Specifically, CNN has become increasingly prevalent in the research field of medical computer vision. Liu *et al.* developed a dermatologic diagnostic system that takes both image feature processed by CNN and clinical data into account [16]. A fused CNN was proposed to detect COVID-19 using CT images [17]. More recently, CNN-based transfer learning method is also widely used in biomedical image classification [18], [19].

In this article, we achieve the biomedical image classification on the basis of EfficientNet [20]. Compared with other prevalent CNNs, such as VGG [21] and GoogLeNet [22], EfficientNet can reduce the computational cost while guaranteeing the accuracy, which has been confirmed on the most popular benchmark dataset, ImageNet. EfficientNet simultaneously intensifies the depth, width, and resolution of the network in accordance with

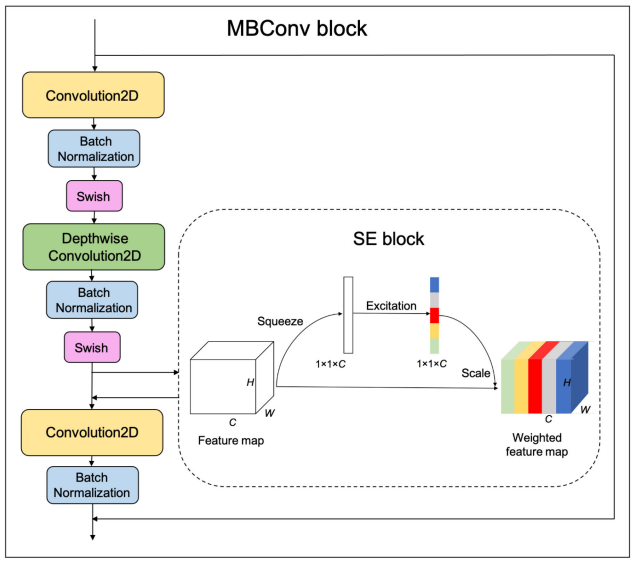

Fig. 2. Architecture of MBconv block.

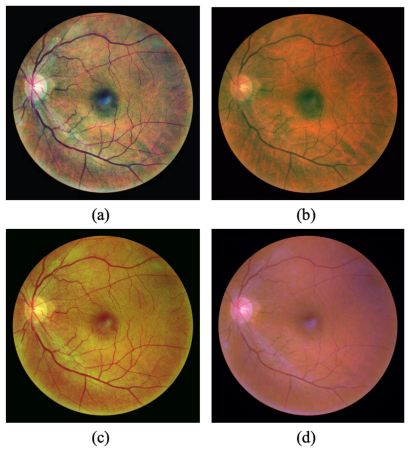

Fig. 3. Enhanced retinal fundus image: (a) Applying CLAHE to R, G, B channels; (b) Applying CLAHE to R channel; (c) Applying CLAHE to G channel; (d) Applying CLAHE to B channel.

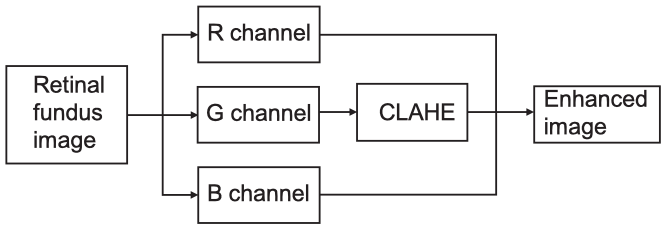

Fig. 4. Image enhancement process using CLAHE.

a certain scaling rule, to improve the performance of network and reduce the computation complexity.

The MBconv block is the representative structure of EfficientNet [23]. It consists of depthwise separable convolution and squeeze-and-excitation network (SENet). The detail of this block is shown in Fig. 2. Compared with the conventional convolution process, the depthwise convolution operation has less parameters and operation cost. SENet has been widely applied in computer vision field [24], and the channel-wise attention mechanism in it is verified as an effective optimization technique for the deep learning model. Generally, the attention mechanism can filter out effective factors from an enormous amount of information and suppress the effect of other useless factors. SENet block acquires the weights of each feature channel by means of learning, and it increases the weight of useful features according to this importance and inhibits the insignificance features. Firstly, the features are squeezed to turn three-dimensional feature map into one-dimensional feature channel descriptor. The second step is excitation, which generates a weight value w for each feature channel, where w denotes the importance of each feature channel. Then, the original features are recalibrated by scale operation. After going through the block, the processed feature maps are concatenated and then sent to the next block.

III. The Proposed Method

In our proposed deep learning method, the key modules, including the image processing, learning model, and optimization function, are analyzed in this section.

A. Image Enhancement

Considering that some biomedical images have low resolution and it makes the features of lesions and blood vessel indistinct, the contrast enhancement technique is used to enhance the quality of original images. Adaptive histogram equalisation (AHE) [25] is a common image enhancement techniques via local contrast amplification. However, image distortion and noise amplification problems caused by the AHE is intolerant. Here, an advanced AHE technique, contrast limited adaptive histogram equalization (CLAHE) [26], is utilized to suppress noise and improve the quality of the processed images.

In order to get adequate quality of retinal fundus color images, CLAHE is applied on R channel, G channel, B channel, and all three channels. According to Fig. 3, the processed image of using CLAHE to G channel has the clearest feature and the highest contrast between vessel and other tissues. Therefore, in this study, the strategy of applying CLANE to G channel is utilized for all experimental images.

The flow diagram of image enhancement process is presented in Fig. 4. Firstly, the input RGB image is split into R, G, and B channels, respectively. Then, CLAHE is only conducted on G channel. After merging the enhanced G channel and original R and B channels, an enhanced image is reconstructed.

Fig. 5 shows the comparison between a low contrast fundus image and the enhanced one. As illustrated, compared with the original image, the vascular outline in the enhanced one is more distinct for visual observation.

B. Deep Learning Model

The developed deep learning model is shown in Fig. 6. Firstly, we use CLAHE to enhance the input image. Then, the data generator function in the Keras library is used for data augmentation.

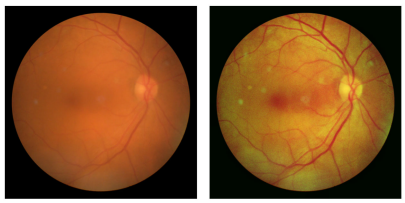

Fig. 5. Comparison between original image and enhanced image.

After preprocessing the input image, EfficientNet is adopted to extract the image feature. Specifically, in consideration of the computing power and model performance, we take EfficientNet-B3 as the baseline model. The processed feature goes through the global average pooling layer, dropout layer and dense layer, and finally gets the classification and visualization results.

C. Loss Function

The loss function is utilized to evaluate the difference between the predicted value and the ground truth, therefore, the network in deep learning model aims to minimize the expected loss. In this article, to effectively address complex biomedical dataset with class imbalance and outliers, we propose a novel mixture loss function, which is the synthesis of focal loss [27] and correntropy loss [28].

1) Focal Loss: The focal loss is modified from binary cross entropy. Usually, binary cross entropy is defined as:

$$L_{\mathrm{ce}}(\widehat{y}, y) = -y \log \widehat{y} - (1-y)\log(1-\widehat{y}) = \begin{cases} -\log \widehat{y}, & y = 1, \\ -\log(1-\widehat{y}), & \text{otherwise.} \end{cases} \tag{1}$$

where y refers to the true label, $\widehat{y}$ is the predicted result for the class with $y = 1$. For further deduction, p is defined as:

$$p = \begin{cases} \widehat{y}, & y = 1, \\ 1-\widehat{y}, & \text{otherwise.} \end{cases} \tag{2}$$

Hence, the cross entropy can be rewritten as:

$$L_{\mathrm{ce}}(p) = -\log(p). \tag{3}$$

Focal loss function adds a modulating factor $(1-p)^{\gamma}$ on the basis of cross entropy to reduce the weight of easy examples and make the training focus on the hard ones. Then, it is defined as:

$$L_{\mathrm{fl}}(p) = -(1-p)^{\gamma}\log(p), \tag{4}$$

where γ is a focusing parameter. As the focusing parameter increases, the contribution of easy samples decreases. Additionally, a weighting factor α is incorporated into focal loss to address the class imbalance problem. Thus, the focal loss can be further redefined as:

$$L_{\mathrm{fl}}(p) = -\alpha(1-p)^{\gamma}\log(p). \tag{5}$$

In addition, by introducing the definition of p, the focal loss can also be written as:

$$L_{\mathrm{fl}}(\widehat{y}, y) = \begin{cases} -\alpha(1-\widehat{y})^{\gamma}\log \widehat{y}, & y = 1, \\ -(1-\alpha)\widehat{y}^{\gamma}\log(1-\widehat{y}), & \text{otherwise.} \end{cases} \tag{6}$$

2) Correntropy Loss: The advantage of correntropy for classification is its adaptability to the sample distances via the different L norm adjustments, which makes correntropy loss robust to noise and outliers [29]–[31]. For very small error, correntropy have the property of L_2 norm. As the error value increases, correntropy loss conducts like L_1 norm, and eventually reaches L_0 norm when the error value is large enough. Correntropy is defined as:

$$V(X,Y) = \mathbb{E}[k_{\sigma}(X-Y)] = \int k_{\sigma}(x-y)dF_{XY}(x,y), \tag{7}$$

where X and Y are two random variables, $k_{\sigma}(\cdot)$ is kernel function, $F_{XY}(x,y)$ is the joint distribution function of variables X and Y, and σ refers to the bandwidth. On the basis of correntropy, correntropy-induced loss function (C-loss) can be defined as:

$$L_{\mathrm{C}}(\widehat{y}, y) = 1 - k_{\sigma}(y-\widehat{y}). \tag{8}$$

Here, we take Gaussian kernel to calculate correntropy. Hence, the C-loss is written as:

$$\begin{aligned} L_{\mathrm{C}}(\widehat{y}, y) &= 1 - G_{\sigma}(y-\widehat{y}) \\ &= 1 - \exp\left(-\frac{(y-\widehat{y})^2}{\sigma^2}\right). \end{aligned} \tag{9}$$

Furthermore, the C-loss function can improve the model performance without increasing the computational cost [32]. Thus, when using Gaussian kernel, the computational complexity of C-loss is the same as cross entropy loss.

3) The Proposed FC-Loss: In this article, through the combination of focal loss and C-loss, we propose a mixture loss function, called FC-loss, which is defined as:

$$L_{\mathrm{FC}}(\widehat{y}, y) = \begin{cases} L_{\mathrm{fl}}(\widehat{y}, y), & 0 < t \leqslant n, \\ L_{\mathrm{C}}(\widehat{y}, y), & n < t < N, \end{cases} \tag{10}$$

where t denotes the current epoch and n is a pre-defined threshold. We first use focal loss to train n epochs, and then change to C-loss to complete the training. Here, N refers to the total number of epochs.

The C-loss has many advantages, such as performing better generalization and robustness to the noise, however, it is non convex, which may lead to local minima. Thus, we use focal loss to pretrain the model for several epochs. The focal loss is adopted to address the class imbalance problem and increase the weight of hard samples. The pseudo-code of proposed model is shown in Algorithm 1.

IV. Experimental Results

The experiments are conducted on a real-life ophthalmic dataset. Here, our experiments are performed with Python and Keras library [33] running on a GeForce GTX 1080 Ti GPU.

A. Dataset

Shanggong Medical Technology Co., Ltd. collected a real-life ophthalmic dataset of 5000 patients in China with retinal fundus images and clinical information, named OIA-ODIR [34]. In this dataset, the patients are classified into different categories

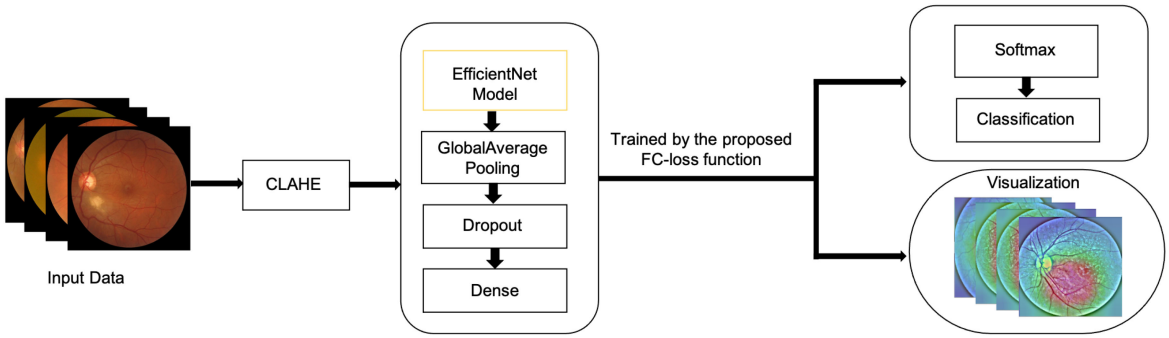

Fig. 6. The proposed deep learning model.

Algorithm 1: FCL-EfficientNet-B3.

Input: Training dataset; test dataset; parameters α, n, γ, σ; epochs N; batch size; learning rate; $t = 0$.
Output: Loss, Accuracy (Acc)

```
Load training dataset.
Initialize the model.
for each epoch within [1, N] do
    Process forward propagation and calculate the loss.
    If the current epoch t ⩽ n, calculate the focal loss;
    else, calculate the correntropy loss.
    Calculate gradients and process backpropagation.
end for
End training.
Process test dataset, return loss and Acc.
```

TABLE I
INFORMATION OF THE DATASET

Disease	AMD	Glaucoma	Cataract	Normal
Number of images	344	259	370	1,000

based on their condition. In addition, annotations are recorded by trained ophthalmologist. It is noted that in this article we focus on detecting glaucoma, cataract, and AMD, thus, only four classes of fundus images in the original dataset are used. They are normal, glaucoma, cataract, and AMD. All glaucoma, cataract, and AMD images contained in the dataset are used. We randomly select 1000 normal samples for training. The number of images in each class is shown in Table I.

B. Metrics

To evaluate the effectiveness of our proposed algorithm, five common metrics, including accuracy (Acc), sensitivity (Sen), specificity (Spe), Kappa, and AUC, are calculated to analyze the experimental results. The definitions are as follows.

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{11}$$

$$\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{12}$$

$$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}. \tag{13}$$

TABLE II
PARAMETERS OF THE MODELS ON THREE DISEASE DATASET

Disease	Model	Parameters				
		N	n	α	γ	σ
AMD	EfficientNet-B3 FL-EfficientNet-B3	600	–	0.25	2	0.80
	BCL-EfficientNet-B3 FCL-EfficientNet-B3	600	200	0.25	2	0.80
Glaucoma	EfficientNet-B3 FL-EfficientNet-B3	500	–	0.30	3	0.80
	BCL-EfficientNet-B3 FCL-EfficientNet-B3	500	150	0.30	3	0.80
Cataract	EfficientNet-B3 FL-EfficientNet-B3	50	–	0.25	2	0.70
	BCL-EfficientNet-B3 FCL-EfficientNet-B3	50	15	0.25	2	0.70

Here, four basic statistics, i.e., true positive (TP), true negative (TN), false positive (FP), and false negative (FN), are used to calculate the metrics. Specifically, in this article, TP denotes the number of patients who are correctly classified as sick, TN denotes the number of healthy persons who are correctly classified as healthy, FP denotes the number of healthy persons who are incorrectly classified as sick, and FN denotes the number of patients who are incorrectly classified as healthy. Accuracy is defined as the value of the amount of data that is correctly predicted compared to the amount of data involved in the prediction. The recognition ability of classifier for positive subjects and negative subjects is measured by sensitivity and specificity, respectively.

Kappa is an indicator used for consistency evaluation [35]. When the number of samples of each class is imbalanced, Kappa is often used to assist in evaluating the performance of classifier. AUC refers to the area under the receiver operating characteristic (ROC) curve [36], and the classifier with larger AUC performs better.

C. Results and Discussion

Here, Adam is used as the optimizer for the CNN models. In addition, the batch size is set to 16, the initial learning parameter is set to 0.0001 and gradually decreases. The input size is 300×300, and dropout rate is set to 0.3. The parameters of each model are shown in Table II.

The baseline methods include EfficientNet-B3 implemented with binary crossentropy loss, as well as FL-EfficientNet-B3

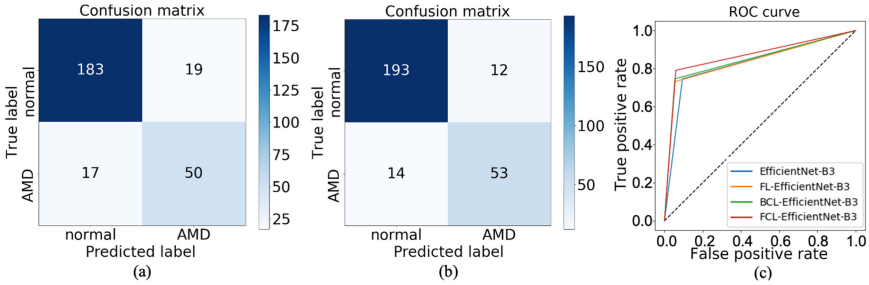

Fig. 7. Classification result of AMD: (a) Confusion matrix of EfficientNet-B3; (b) Confusion matrix of our proposed method; (c) ROC curve.

TABLE III
AMD CLASSIFICATION RESULTS

Model	Acc(%)	Sen(%)	Spe(%)	Kappa	AUC
EfficientNet-B3	87.72	91.91	75.12	0.6751	0.8351
FL-EfficientNet-B3	88.11	92.40	75.12	0.6801	0.8376
BCL-EfficientNet-B3	88.96	93.89	74.12	0.6977	0.8400
FCL-EfficientNet-B3	**90.08**	**94.71**	**76.16**	**0.7276**	**0.8542**

TABLE IV
CATARACT CLASSIFICATION RESULTS

Model	Acc(%)	Sen(%)	Spe(%)	Kappa	AUC
EfficientNet-B3	99.02	99.49	97.86	0.9760	0.9867
FL-EfficientNet-B3	98.90	98.98	98.72	0.9732	0.9884
BCL-EfficientNet-B3	**99.45**	**99.74**	98.72	0.9865	0.9922
FCL-EfficientNet-B3	99.38	99.49	**99.14**	**0.9910**	**0.9935**

TABLE V
GLAUCOMA CLASSIFICATION RESULTS

Model	Acc(%)	Sen(%)	Spe(%)	Kappa	AUC
EfficientNet-B3	82.39	85.90	68.63	0.4991	0.7726
FL-EfficientNet-B3	83.92	86.81	72.55	0.5525	0.7884
BCL-EfficientNet-B3	84.52	**88.05**	70.68	0.5502	0.8017
FCL-EfficientNet-B3	**84.78**	87.23	**75.16**	**0.5698**	**0.8119**

and BCL-EfficientNet-B3. Here, FL-EfficientNet-B3 refers to EfficientNet-B3 model combined with focal loss function. BCL-EfficientNet-B3 refers to the model which first uses binary crossentropy loss function, then switches to correntropy loss function. FCL-EfficientNet-B3 is our proposed model, which takes the combination of focal loss and C-loss as its mixture loss function. The accuracy, specificity, sensitivity, Kappa, and AUC are obtained by 5-fold cross validation. We randomly take 80% of total samples as the training samples, the remaining 20% samples as the test set. Results of detecting AMD, cataract, and glaucoma with each model are presented in Tables III, IV, and V, respectively.

From those results, among all those deep learning models implemented to detect three eye diseases including glaucoma, AMD, and cataract, FCL-EfficientNet-B3 model outperforms other baseline methods. Compared with baseline methods, owing to the contribution of the proposed mixture loss function, when the weight of hard sample increases, FCL-EfficientNet-B3 performs better. BCL-EfficientNet-B3 performs even better than FL-EfficientNet-B3, which shows that the strategy of switching to C-loss function midway through the training can actually improve the generalization of classifier. Moreover, using focal loss instead of binary cross entropy to pretraining for several epochs further optimizes the model. For glaucoma and cataract classification tasks, our method does not achieve the best accuracy and sensitivity, but the specificity, Kappa, and AUC have been significantly improved as a result of the incorporation of our proposed loss function, which shows that our model has better ability to recognize the positive samples.

The confusion matrices and ROC curves of those experimental results are shown in Figs. 7, 8, and 9. Compared with the baseline method EfficientNet-B3, our proposed FCL-EfficentNet-B3 achieves better generalization.

Fig. 10 shows the examples of correctly classified samples and misclassified samples using our method. For AMD subjects, we can see dense bright spots or a relatively larger and distinct bright lesion area, while in the misclassified image, the contrast of lesion and other tissues are relatively low or there is no obvious macular area. Cataracts affect the transparency of the eye's lenses, thus, in the cataract fundus images, the retina is usually a blur, and blood vessels cannot be seen clearly. We believe that the cataract is easy to be recognized due to its overall low contrast. However, if the disease is not severe, it is hard to be classified. Other possible factor for misclassification is the noise, and the image enhancement technique enhances the noise and creates the bright area on the fundus image. In glaucoma fundus image, the ratio of optic cup and optic disc is abnormal, which is hard to be classified. In the well classified images, the optic cup is bigger and brighter, while in the misclassified images, the contrast of optic cup and optic disc is not visible.

In sum, the experimental results validate the performance of our proposed method, and also indicate the glaucoma and AMD are two hard diseases for machine learning recognition. To address this problem, the research on further feature extraction technologies to assist the detection of glaucoma and AMD is our future work.

D. The Evaluation With Class Activation Map (CAM)

Specifically, the class activation map (CAM) [37] is implemented to further comprehend the learning ability of our proposed method. CAM generates class activation heat maps for specified input images and indicates the important area of ophthalmic images.

The predictive CNN combines feature extractor and classifier. In feature extraction process, convolutional layers are utilized to extract local features of images, and those features are given different weights in accordance with the importance. The CAM

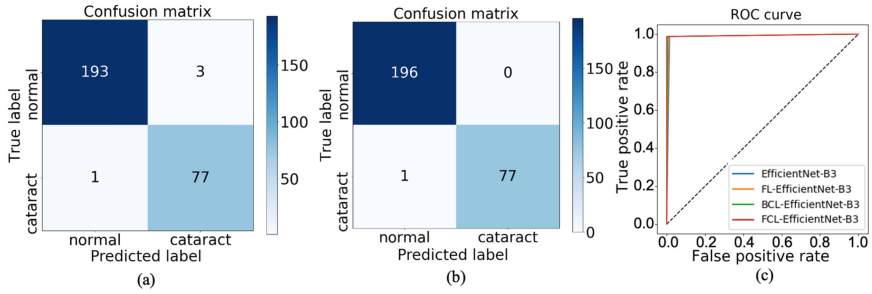

Fig. 8. Classification result of cataract: (a) Confusion matrix of EfficientNet-B3; (b) Confusion matrix of our proposed method; (c) ROC curve.

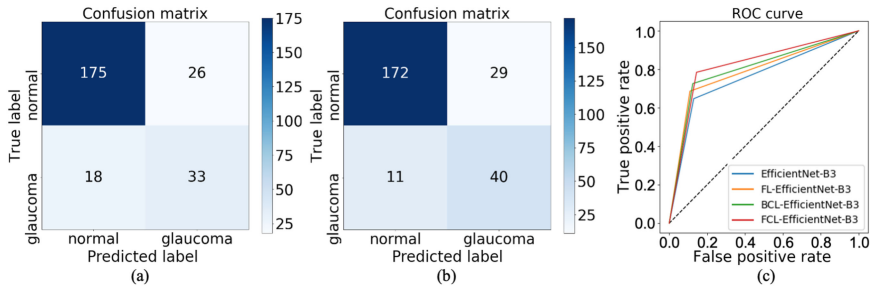

Fig. 9. Classification result of glaucoma: (a) Confusion matrix of EfficientNet-B3; (b) Confusion matrix of our proposed method; (c) ROC curve.

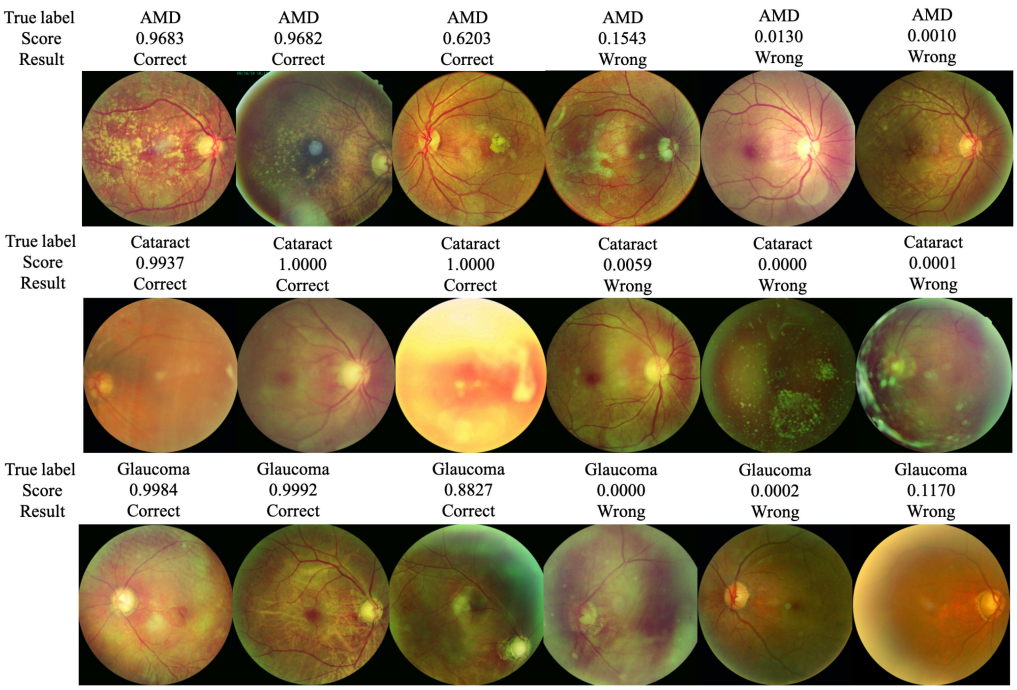

Fig. 10. Examples of correctly classified and misclassified samples in three types of ophthalmic diseases.

uses the weights of feature maps to obtain heat maps, which allows us to have a good interpretation of the CNN model. The resulting maps are shown in Fig. 11.

In this figure, the most discriminative parts of AMD, cataract and glaucoma images obtained by learning process are highlighted. As shown in Fig. 11(a), the highlight region covers macular disease area, which indicates that the deep learning model can localize the lesion precisely. However, the highlight region in Fig. 11(b) is not limited to a certain area of retina. According to the analysis, it is believed the reason for the erratic localisation is that the severe cataract eyes are usually blurs because of the same features in each area of the images. In Fig. 11(c), the heat maps focus on optic disc and optic cup area, which are in line of the diagnosis criteria.

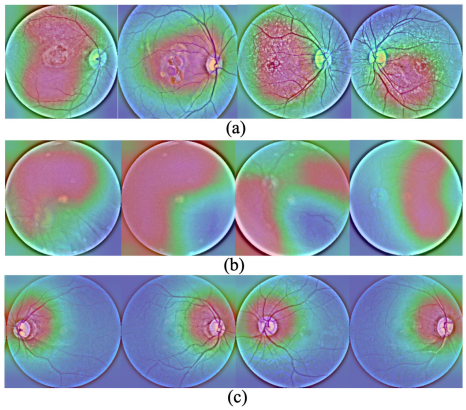

Fig. 11. The class activation maps by using the proposed deep learning method on retinal fundus images: (a) AMD; (b) Cataract; (c) Glaucoma.

V. Conclusion

This article proposes a deep learning-based CAD model to deal with a challenging issue in classification of ophthalmic images, through the analysis of retinal fundus images. Starting from the general introduction of related method on CNN, we propose a deep learning-based ophthalmic disease detection method, which is a CNN combined with our proposed FC-loss function. Meanwhile, some key technologies in relation to our method, including image enhancement technique CLAHE, CNN, and loss function, are analyzed. Furthermore, in the experiments, comparing to other traditional deep learning-based classifiers, our proposed method achieves higher classification accuracy, sensitivity, specificity, Kappa, and AUC, due to the incorporation of a mixture loss function. The experimental results verify the effectiveness and robustness of our proposed deep learning method on ophthalmic diseases classification task.

References

[1] N. Asiri, M. Hussain, F. A. Adel, and N. Alzaidi, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey," *Artif. Intell. In Med.*, vol. 99, 2019, Art. no. 101701.

[2] I. Qureshi, J. Ma, and Q. Abbas, "Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning," *Multimedia Tools Appl.*, vol. 80, no. 8, pp. 11691–11721, 2021.

[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep Learn.," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] J. F. Grassmann *et al.*, "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography," *Ophthalmology*, vol. 125, no. 9, pp. 1410–1420, 2018.

[5] Y. Peng *et al.*, "DeepSeeNet: A deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs," *Ophthalmology*, vol. 126, no. 4, pp. 565–575, 2019.

[6] Q. Yan *et al.*, "Deep-learning-based prediction of late age-related macular degeneration progression," *Nat. Mach. Intell.*, vol. 2, no. 2, pp. 141–150, 2020.

[7] T. Pratap and P. Kokil, "Computer-aided diagnosis of cataract using deep transfer learning," *Biomed. Signal Process. Control*, vol. 53, 2019, Art. no. 101533.

[8] H. Zhang *et al.*, "Automatic cataract grading methods based on deep learning," *Comput. Methods Programs Biomed.*, vol. 182, 2019, Art. no. 104978.

[9] X. Xu, L. Zhang, J. Li, Y. Guan, and L. Zhang, "A hybrid global-local representation CNN model for automatic cataract grading," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 556–567, Feb. 2020

[10] Z. Li *et al.*, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, 2018.

[11] G. An *et al.*, "Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images," *J. Healthcare Eng.*, vol. 2019, pp. 1–9, 2019.

[12] M. N. Bajwa *et al.*, "Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning," *BMC Med. Informat. Decis. Mak.*, vol. 19, no. 1, pp. 1–16, 2019.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE Proc. IRE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[14] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically designing CNN architectures using the genetic algorithm for image classification," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020.

[15] F. Cao and Q. Bao, "A survey on image semantic segmentation methods with convolutional neural network," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng.*, 2020, pp. 458–462.

[16] Y. Liu *et al.*, "A deep learning system for differential diagnosis of skin diseases," *Nature Med.*, vol. 26, no. 6, pp. 900–908, 2020.

[17] S.-H. Wang *et al.*, "COVID-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Inf. Fusion*, vol. 67, pp. 208–229, 2021.

[18] M. F. Aslan *et al.*, "CNN-based transfer learning-BiLSTM network: A novel approach for COVID-19 infection detection," *Appl. Soft Comput.*, vol. 98, 2021, Art. no. 106912.

[19] S.-H. Wang *et al.*, "COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis," *Inf. Fusion*, vol. 68, pp. 131–148, 2021.

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 10691–10700.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations*, 2014.

[22] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[23] A. G. Howard *et al.*, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 10691–10700.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[25] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Comput. Vis., Graph., Image Process.*, vol. 38, no. 3, pp. 355–368, 1987.

[26] S. Pizer *et al.*, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visual. Biomed. Comput.*, 1990, pp. 337–345.

[27] T. Y. Lin *et al.*, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[28] W. Liu, P. Pokharel, and J. C. Principe, "Correntropy: A localized similarity measure," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2006, pp. 4919–4924.

[29] A. Gunduz and J. C. Principe, "Correntropy as a novel measure for nonlinearity tests," *Signal Process.*, vol. 89, no. 1, pp. 14–23, 2009.

[30] X. Luo *et al.*, "Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4963–4971, Nov. 2018.

[31] X. Luo *et al.*, "Towards improving detection performance for malware with correntropy-based deep learning method," *Digit. Commun. Netw.*, to be published, 2021. doi: 10.1016/j.dcan.2021.02.003.

[32] A. Singh, R. Pokharel, and J. Principe, "The C-loss function for pattern classification," *Pattern Recognit.*, vol. 47, no. 1, pp. 441–453, 2014.

[33] Keras Team, "Keras API reference,," 2021. [Online]. Available: https://keras.io/

[34] Nkicsl, "nkicsl/OIA-ODIR," 2021. [Online]. Available: https://github.com/nkicsl/OIA-ODIR

[35] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[36] K. A. Spackman, "Signal detection theory: Valuable tools for evaluating inductive learning," in *Proc. 6th Int. Workshop Mach. Learn.*, 1989, pp. 160–163.

[37] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.