# Ocular diseases classification using a lightweight CNN and class weight balancing on OCT images

**Neha Gour[1] · Pritee Khanna[1]** 

## Abstract

Optical coherence tomography (OCT) is a non-invasive technique to capture cross-sectional volumes of the human retina. OCT images are used for the diagnosis of various ocular diseases. However, OCT datasets generally suffer from the problem of class imbalance. This work aims to leverage CNN capability for OCT images classification in the presence of class imbalance. A lightweight convolutional neural network (CNN) with class weight balancing (CWB) is proposed for OCT image classification. Training of CNN is done while penalizing the classes having a higher number of samples using the CWB method. The performance of the proposed method is evaluated on spectral-domain OCT (SD-OCT) images from two publicly available datasets, namely, ZhangLab and Duke. The performance of the proposed method is evaluated on the confusion matrix-based parameters like accuracy, sensitivity, specificity, and $F1-score$. The proposed method achieved 99.17% and 98.46% accuracy for ZhangLab and Duke datasets, respectively. It is observed that the proposed method performs better as compared to most of the state-of-the-art OCT classification methods The generalizability and interpretability of the proposed method are also evaluated to improve the understanding of the CNN model.

**Keywords** Ocular disease classification · Optical coherence tomography ·
Convolutional neural networks · Class weight balancing

## 1 Introduction

The development of better image acquisition techniques and reduced cost of storage and computational facilities lead to advancements in medical image analysis. Optical coherence

✉ Pritee Khanna
pkhanna@iiitdmj.ac.in

Neha Gour
g.neha@iiitdmj.ac.in

1 Computer Science and Engineering, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India

tomography (OCT) is being increasingly used to diagnose ocular diseases in the human retina. Textural and morphological variations in retinal layers captured by OCT imaging suggest the presence of abnormalities like diabetic macular edema (DME), age-related macular degeneration (AMD), glaucoma, diabetic retinopathy, and choroidal neovascularization (CNV) in the retina [46]. These diseases are the leading cause of irreversible vision loss.

With the availability of a large amount of data and image augmentation techniques, it is possible to implement convolutional neural network (CNN) based models to develop a better computer-aided diagnosis system for ocular diseases. However, medical image datasets usually suffer from class imbalance problem. Unequal distribution creates a bias towards the majority class and interferes with the classification performance of the CNN models. This work aims to develop an ocular disease identification algorithm for the classification of OCT images using CNN. The novelty of this work lies in developing a lightweight CNN, which needs fewer parameters to be trained and is also capable of dealing with the issue of class imbalance for ocular disease classification. The feature maps learned through CNN have important localization and classification information which can be harnessed using the global average pooling (GAP) layer as the last layer of the architecture. GAP layer helps in computing class activation maps (CAM) and highlights the important features of the image responsible for the particular disease class. Class weight balancing (CWB) is used to address the class imbalance problem in the proposed method. The proposed method is implemented on two publicly available datasets, ZhangLab and Duke. ZhangLab dataset consists of 83, 484 OCT images for classification into CNV, DME, and Drusen ocular diseases, and the Duke dataset comprises 3231 OCT images for classification into AMD and DME diseases.

The work is organized into five sections. The state-of-the-art methods for the classification of OCT images are summarized in Section 2. The proposed methodology for ocular disease classification using class weight balancing in CNN is described in Section 3. Results are discussed in Section 4 and compared with other state-of-the-art methods. Finally, the work is concluded in Section 5.

## 2 Related work

Ocular disease classification using OCT images is a recently explored research problem in medical image processing. OCT imaging is a non-invasive technique that uses light waves to capture cross-sectional images of the retina [11]. OCT imaging technology was first introduced by Huang et al. [20] to capture important retinal structures like choroid, retinal pigment epithelium, posterior vitreous, and retina. The OCT images are mainly captured through time-domain (TD) and spectral-domain (SD) imaging techniques. TD-OCT is the conventional technique for acquiring retinal images, whereas advancements in SD-OCT imaging have been made in recent years. The SD-OCT technique gives three times more depth estimation as compared to TD-OCT technology [8]. Due to this, SD-OCT is a popular imaging technique for diagnosing AMD, CNV, and DME diseases.

In literature, OCT classification methods are divided into two main categories: handcrafted feature extraction and deep learning-based methods. The techniques proposed for OCT image classification are mainly implemented on two publicly available datasets, namely Duke [44] and ZhangLab [28], for three and four-class classification of ocular diseases. The first method on the Duke dataset was proposed by Srinivasan et al. [44] using histogram of oriented gradient (HOG) features and Support Vector Machine (SVM) classifier. Hussain et al. [21] used retinal layer segmentation based features to extract ocular

disease related abnormalities indicated by changes in the retinal layer for the classification task. Canny edge detection algorithm is used for the segmentation of retinal layer structure. The features are extracted from segmented layers and classified using random forest classifier into AMD, DME, and Normal classes. Venhuizen et al. [48] extracted interest points by thresholding the top 3% values of first-order vertical Gaussian gradient of OCT images of the Duke dataset [44]. Principal component analysis (PCA) is used to reduce the dimensionality of the features. Unsupervised learning is used to classify the final feature vector into disease class labels. A local binary pattern (LBP) feature based method was proposed by Lemaitre et al. [31] in combination with the bag of word technique for classification using an SVM classifier. A CAD system for OCT images of the Duke dataset is developed using a multi-scale convolutional mixture of expert (MCME) ensemble model [40]. The success of hand-crafted feature extraction techniques lies in finding the suitable trade-off between classification accuracy and computational cost. Hand-crafted features represent selected characteristics of a particular set of data. Important features relevant to the task may not be included in the final feature vector by employing a particular hand-crafted feature extraction technique [37].

Recently, deep learning-based methods have been a growing trend in data analysis and computer vision [15]. In particular, CNN has proven to be a powerful tool for medical imaging tasks like segmentation and classification. Various methods proposed for the classification of OCT images into ocular diseases include transfer learning from a network trained for a general image dataset, combining hand-crafted features with learned features, and training the network from scratch [7, 15]. Qingee et al. [23] proposed a transfer learning-based approach using the Inception V3 model by extracting middle-level features and training them with new CNN architecture. The performance of the trained model was evaluated on the Duke dataset for ocular disease classification. Another transfer learning-based method using Google Net was proposed by Karri et al. [25], where the network was fine-tuned for OCT images of the Duke dataset. A wavelet-based convolutional neural network (WCNN) was proposed to generate representative OCT CNN-codes in the spatial-frequency domain. The presence of abnormalities was scored over these features [39]. Najeeb et al. [36] extracted the region of interest of OCT images using binary thresholding. A computationally inexpensive single layer CNN was applied for the classification of OCT images from the ZhangLab dataset. Kermany et al. [27] proposed a transfer learning-based method for the classification of OCT images into four ocular disease classes for the ZhangLab dataset. Alqudah [1] proposed a 19 layer CNN architecture to classify SD-OCT images into five disease categories. The images are obtained from ZhangLab and a dataset released by Duke University. Ibrahim et al. [22] implemented a hybrid architecture where the region of interest based and CNN layer based features are concatenated together to classify OCT images of the ZhangLab dataset. Thomas et al. [47] proposed a multi-scale CNN architecture to classify OCT images into AMD and normal classes. The method was implemented on four publicly available datasets. Sunija et al. [45] implemented a lightweight CNN architecture on the earlier version of the ZhangLab dataset. Wu et al. [49] implemented an attention-based CNN model using ResNet50 architecture to extract features related to ocular classes of ZhangLab and a private dataset.

The sample distribution in the image dataset is an important property to be considered for classification problems. If the number of samples in one class is significantly higher than in the other classes, then the dataset is termed imbalanced [3]. Class imbalance is a relatively less explored and important research problem in medical image analysis. Generally, medical datasets comprise a higher number of normal samples (majority class) as compared to diseased samples (minority class) [12]. The same is true for OCT datasets like are Duke [44],

ZhangLab [28], OCTID [13], and EUGENDA [48]. The imbalance present in training set results in over-classification of the majority class due to its higher prior bias. It leads to mis-classification of minority class and less accurate overall prediction accuracy [24]. Various techniques like sampling based methods [16], cost-sensitive learning [29], and active learning [9] are implemented to reduce the effect of class imbalance in medical image analysis. Sampling based methods include random undersampling of majority class, oversampling of the minority class, and synthetic minority over-sampling technique (SMOTE) [4, 35]. Instead of sampling the data to create a balanced dataset, cost-sensitive learning defines the cost of misclassifying images of different classes [34]. This approach penalizes the misclas-sification of the minority class by giving more priority to its samples [29]. The process of random oversampling of minority classes is equivalent to class weight balancing [6]. The advantage of using class weight balancing over image augmentation lies in the reduction of training time required. The addition of the class weight balancing technique does not affect the number of training parameters, the number of floating-point operations, the size of the input data generator, and the time per epoch. The usage of data augmentation increases the training dataset size, which increases training time. Active learning is an iterative approach which deals with a small training set. It is an iterative approach that selects data from unla-beled or synthesized samples. It reduces the efforts of experts in labeling the medical data [43].

In general, publicly available OCT datasets have normal class as majority class, class imbalance issue needs to be addressed for OCT image classification. The contribution of this work lies in the embedding of a cost-sensitive learning-based CWB with a lightweight CNN proposed in this work which is trained from scratch. The performance of the proposed model is observed on not only ZhangLab and Duke datasets but also cross-validated on these datasets. The final results are compared with state-of-the-art methods discussed in this section for OCT classification performance analysis.

## 3 Materials and methods

### 3.1 Datasets used for experimentation

Creating a well labeled, balanced, and large image dataset is one of the major challenges in the medical domain. Expert annotation processes are expensive, and the number of images with disease is scarce compared to normal images [15]. OCT datasets by Zhang et al. [28] and Duke et al. [44] are used to implement and evaluate the proposed method. Kermany et al. [27] introduced a well labeled and suitably large OCT dataset named as ZhangLab. ZhangLab dataset was created at Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Asso-ciates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center. The SD-OCT images in the ZhangLab dataset are captured using Spectralis OCT imaging protocol. The images were graded and verified for quality and labels of the diseases. This dataset com-prises of total 83, 484 images with 37, 216 images of CNV, 11, 348 images of DME, 8, 617 images of Drusen, and 26, 316 images of normal categories. Sample images from each class are shown in Fig. 1. The number of images of CNV and normal class are higher than DME and Drusen categories. This class imbalance tends to reduce the performance of the clas-sification models. The dataset contains 1000 separate images for testing purpose with 250 images from each category.
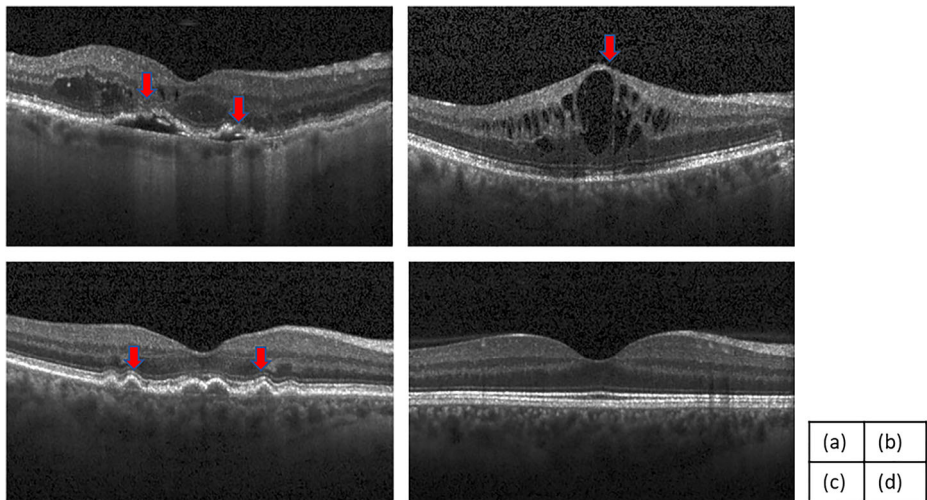
**Fig. 1** OCT image examples from ZhangLab dataset [28], where red arrows are showing abnormalities: (a) choroidal neovascularization (CNV) (b) diabetic macular edema (DME) (c) drusen, and (d) normal image

Similar to ZhangLab dataset, Duke dataset comprises of SD-OCT volumetric scans of 45 patients captured with Spectralis SD-OCT imaging at Duke, Harvard, and Michigan Universities. The scans are divided into three classes having 15 volumes each of normal, dry AMD, and DME patients. A sample image from each class is shown in Fig. 2. Each volume contains varying B-scans (31 − 97) making total number of OCT images in the dataset as 3231. Duke dataset consists of three classes with 723, 1101, and 1407 images of AMD, DME, and normal categories. Duke dataset is a relatively small dataset and might not be suitable for deep learning. The number of images in each class is relatively balanced in comparison to the ZhangLab dataset.

## 3.2 Convolutional Neural Networks (CNN)

CNN is a specialized neural network for processing data with a grid-like structure, e.g., time-series data and images. CNN was first introduced by Le Cun et al. [30] to classify hand-written digit image data. CNN employs mathematical spacial convolution operation in at
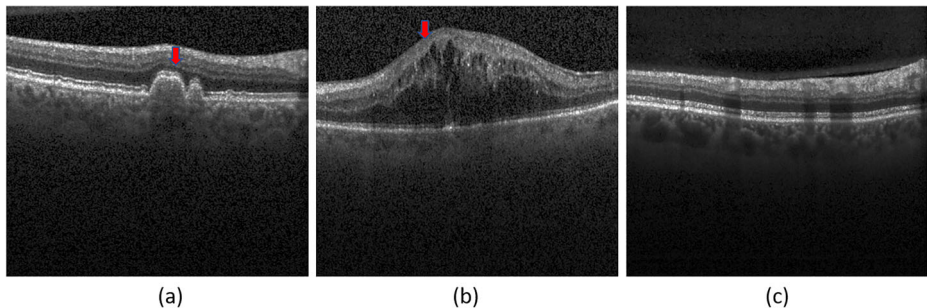


**Fig. 2** OCT image examples from Duke dataset [44], where red arrows are showing abnormalities: (a) age-related macular degeneration (AMD) (b) diabetic macular edema (DME), and (c) normal image

least one of the layers [14]. The basic architecture of CNN comprises convolutional (Conv), pooling, rectified linear unit (ReLU), fully connected (FC), batch normalization (BN), and global pooling (GP) layers. CNN architectures incorporate techniques like local receptive fields, weight sharing, and sub-sampling [2]. The convolutional filters are randomly initialized and learned through training CNN. These techniques give CNN an advantage over hand-crafted feature based methods to a great extent. Convolution operation combined with neural networks provides high performance in computer vision models for applications like segmentation, feature extraction, denoising, enhancement, and classification [15]. Although improved classification performance is observed with deep learning methods, but interpretability of the models reduces. The classification prediction relies on the millions of parameters, which makes it difficult to interpret as compared to hand-crafted features and basic machine learning classifiers. Interpretability is very crucial in the medical imaging applications to validate the correctness of trained models. Zhou et al. [50] proposed technique of obtaining activation maps through class specific output weights called class activation mapping (CAM). The important parts of medical images are projected back using the weights of the final layer of CNN on the last convolutional feature maps.

### 3.3 The proposed CWB based CNN architecture

Based on the traditional CNN architectures, a lightweight CNN model is proposed to classify OCT images into three and four classes of ocular diseases. The OCT images are resized to $256 \times 256 \times 3$ dimensions for fast implementation of CNN architecture. Each pixel of the OCT image is normalized to bring the intensities to a common scale. As depicted in Fig. 3, the proposed CNN architecture comprises four convolutional layers with 64, 64, 128, and 256 filters for each subsequent layer. Each convolutional layer has a $3 \times 3$ kernel with ReLU activation function. The size of the feature maps is selected based on architectures available in the literature, which depicted state-of-the-art performance. The architecture like LeNet, AlexNet, VGG16, and InceptionV3 follow architecture trends with increasing feature maps or filters in subsequent layers. In the retinal image analysis domain, implementing lightweight or shallow CNN capable of good classification accuracy is important. This information is utilized in the proposed method to prepare 4 layers of lightweight CNN architecture for OCT classification. The convolution layers are combined with batch normalization and max-pooling layers.
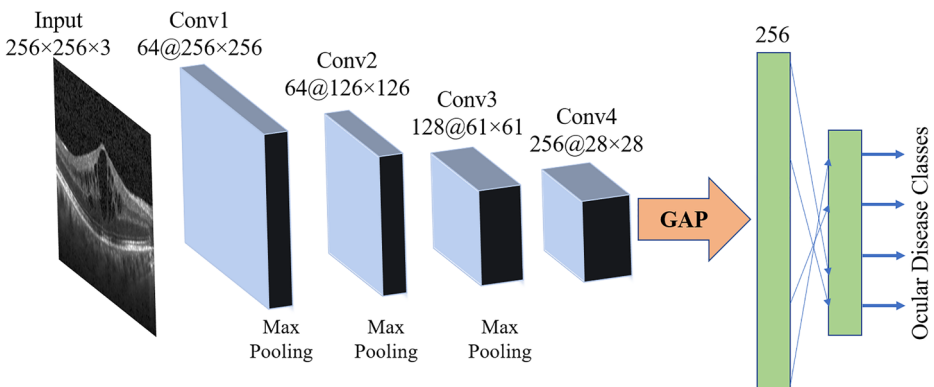


**Fig. 3** Proposed class weight balancing based CNN for ophthalmological disease classification using OCT images

The fully connected layer is usually used as the final layer of the CNN architecture for the classification task, which gives predicted probabilities against each class label. But the use of fully connected layers does not give any insight into the image's features or particular area responsible for the prediction. The global average pooling (GAP) layer is more native to the convolutional structure of the CNN and hence enforces good correspondence between feature maps and classes [33]. It is the main reason the GAP layer contributes to the construction of class-specific activation maps. During the training and validation steps, the GAP layer does not contribute any parameters to optimize and hence helps in reducing the chance of overfitting. Therefore, the global average pooling (GAP) layer is the last layer of the proposed CNN architecture in place of a fully connected layer [18]. Each feature map constructed by the last convolutional layer is averaged to get a single value as a feature to classify OCT images. Therefore, many filters, i.e., 256 with size $28 \times 28$, are used in the last layer of the proposed method so that the important features (information) loss does not happen during the GAP step.

Further, the proposed architecture uses CWB to address the issue of class imbalance. In this technique, higher weights are assigned to the classes with less number of OCT image samples and vice versa. The class weights for each class are calculated using the (1) given by King et al. [29]:

$$W_j = \frac{N}{k * N_j} \tag{1}$$

where $W_j$ is the weight of class $j$, $N$ is the total number of OCT images, $k$ is the total number of classes, and $N_j$ is the number of OCT images in class $j$. The class weights are applied during CNN weight learning and penalize classification of minority class into majority class. The final features obtained through the GAP layer are applied to the dense layer for classification using the softmax activation function.

## 3.4 Experimental setup

The proposed model is implemented using the Keras library on the TensorFlow backend. In the proposed architecture, the final feature vector obtained from the GAP layer is applied to the dense layer with four neurons for each class of the ZhangLab dataset. The training and testing sets are pre-partitioned in the ZhangLab dataset. The training set of the ZhangLab dataset is further divided into training and validation split with 80% and 20% OCT images in each set, respectively. The number of images in each set is mentioned in Table 1.

The same CNN architecture is used for the three-class classification problem of the Duke dataset, where the output layer comprises three neurons. The number of images in the Duke dataset is relatively balanced as compared to the ZhangLab dataset. The literature observed that experiments are performed with different partitions of the Duke dataset, as predefined training and testing sets are not given with the Duke dataset. In this work, the images of the

**Table 1** Ocular disease class distribution for ZhangLab dataset

| Dataset | Set | CNV | DME | Drusen | Normal | Total |
|---------|-----|-----|-----|--------|--------|-------|
| ZhangLab | Training | 29764 | 9079 | 6893 | 21052 | 66788 |
| | Validation | 7441 | 2269 | 1723 | 5263 | 16696 |
| | Testing | 250 | 250 | 250 | 250 | 1000 |

**Table 2**  Ocular disease class distribution for Duke dataset

| Dataset | Set | AMD | DME | Normal | Total |
|---|---|---|---|---|---|
| Duke (60-40) | Training | 412 | 624 | 786 | 1822 |
| | Validation | 274 | 416 | 524 | 1214 |
| | Testing | 37 | 61 | 97 | 195 |
| Duke (70-30) | Training | 480 | 728 | 917 | 2125 |
| | Validation | 206 | 312 | 393 | 911 |
| | Testing | 37 | 61 | 97 | 195 |
| Duke (80-20) | Training | 549 | 832 | 1048 | 2429 |
| | Validation | 137 | 208 | 262 | 607 |
| | Testing | 37 | 61 | 97 | 195 |
| Duke (90-10) | Training | 617 | 936 | 1179 | 2732 |
| | Validation | 69 | 104 | 131 | 304 |
| | Testing | 37 | 61 | 97 | 195 |

$15^{th}$ volume of each class of the Duke dataset are kept separate for testing, and the remaining volumes are used as the training set. The training set is further divided into four partition combinations (starting from 60% in the increment of 10%) for training and validation. Table 2 shows the number of training, validation, and testing images for each partition.

The model is trained by taking batch size as 32 for 100 epochs with early stopping on the best validation accuracy. The model is trained using the categorical cross-entropy loss function. Adam optimizer is used for back-propagation with a learning rate of 0.01. Xavier uniform initializer is used for initializing the weight of various layers involved in the proposed architecture, and the biases are initialized as zero. During the training phase, training and validation image samples are applied along with the class weights. The learning is penalized according to the weights assigned for each class. The best model with the highest validation accuracy is selected for testing without applying CWB. Finally, the GAP layer is used to construct CAMs.

## 4  Results and discussion

The proposed method is applied on 83, 484 OCT images for the four-class classification of the ZhangLab dataset [28] and 3231 OCT images for the three-class classification of the Duke dataset [44] with the distribution as given in Tables 1 and 2 . The performance is evaluated and compared on metrics based on the confusion matrix. The classification results obtained from the proposed method are compared with the existing state-of-the-art methods earlier discussed in Section 2. A multi-class confusion matrix is constructed on the predicted labels of the testing set for each of the datasets. Predicted labels are compared with the ground truth provided by the experts for each OCT image of test datasets. Performance metrics like accuracy (ACC), sensitivity (SENS), specificity (SPEC), and $F1 - score$ are calculated based on the multi-class confusion matrix [38]. The receiver operating characteristics (ROC) curve is plotted, and the area under the curve (AUC) of the plot is calculated [10]. The training is done for 100 epochs with an early stopping criterion (monitoring up to 10 epochs for improvement). Figure 4 shows the training and validation accuracy per epoch for both datasets. Due to early stopping criteria, the best models were obtained at $60^{th}$ and
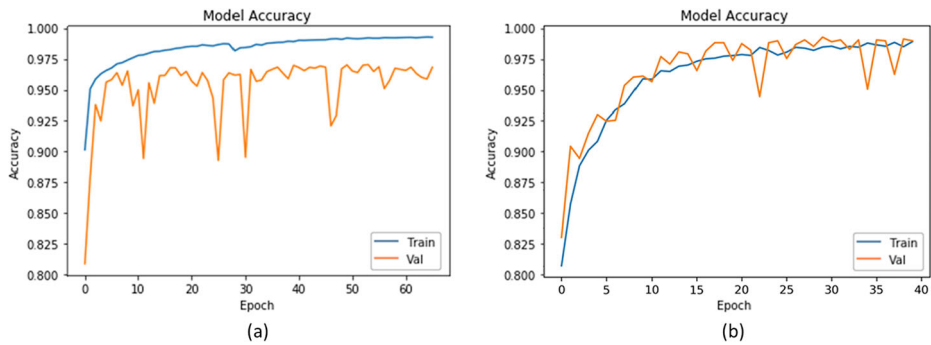
**Fig. 4** Plots depicting the performance of OCT classification in the training and validation sets: (a) ZhangLab Dataset (early stopping at epoch number 60) and (b) Duke Dataset (early stopping at epoch number 40)

$40^{th}$ epochs for ZhangLab and Duke datasets, respectively. The training accuracy increases with the increase in validation accuracy, which indicates that models are not overfitting the OCT image data.

**Performance on the ZhangLab dataset:** The CNN models are usually tested on metrics like $F1 - score$, Accuracy, Sensitivity, and Specificity, but optimized using loss function like categorical cross-entropy. Sanyal et al. [42] proposed optimization of non-decomposable loss functions like $F1 - score$ for better performing deep learning models developed for imbalanced datasets. Therefore, for the comparison purpose, this method is implemented with the proposed architecture where CNN is optimized on $F1 - score$ as loss function. Results for different combinations of CWB and $F1 - score$ loss function optimization experimented on the ZhangLab dataset are summarized in Table 3. The basic model (OCT_GAP + CCE) using categorical-cross entropy (CCE) loss function performs better in comparison to model (OCT_GAP + F1) with $F1 - score$ loss function optimization. The $F1 - score$ loss function combined with the CWB technique (OCT_GAP + CWB + F1) shows clear improvement in the results compared to $F1 - score$ loss function optimization (OCT_GAP + F1). This improvement is an indicator of the benefit of using class weight balancing for the imbalanced ZhangLab dataset. Although $F1 - score$ loss function optimization with class weight balancing (OCT_GAP + CWB + F1) model shows improvement, it underperforms compared to class weight balancing used with the basic model (OCT_GAP + CWB + CCE).

For the ZhangLab dataset, the proposed method performs better than the existing OCT image classification methods presented in Table 4. Zhang et al. [28] presented a transfer

**Table 3** Performance of the proposed algorithm with CWB and $F1 - score$ optimization on ZhangLab dataset

| Method | ACC | SENS | SPEC | F1-score | AUC |
|---|---|---|---|---|---|
| OCT_GAP + CCE | 98.76 | 98.76 | 99.58 | 0.99 | 0.97 |
| OCT_GAP + F1 | 97.31 | 97.41 | 99.10 | 0.97 | 0.97 |
| OCT_GAP + CWB + F1 | 99.07 | 99.07 | 99.68 | 0.99 | 0.95 |
| **OCT_GAP + CWB + CCE** | **99.17** | **99.27** | **99.72** | **0.99** | **0.99** |

**Table 4** Performance of the proposed method on ZhangLab dataset

| Author | Method | ACC | SENS | SPEC | *F1-score* |
|---|---|---|---|---|---|
| Huang et al. [19] | LGCNN | 88.40 | 84.60 | 82.90 | - |
| Zhang et al. [28] | Inception V3 TL | 96.60 | 97.80 | 97.40 | - |
| Najeeb et al. [36] | CNN Scratch | 95.66 | 95.65 | 98.55 | 0.95 |
| Kaymak et al. [26] | AlexNet TL | 97.10 | 99.60 | 98.40 | - |
| Li et al. [32] | VGG 16 TL | 98.60 | 98.60 | 99.53 | 0.99 |
| Chetoui et al. [5] | EfficientNet-B7 | 98.46 | 98.68 | 98.37 | - |
| Proposed Method | OCT_GAP + CWB + CCE | 99.17 | 99.27 | 99.72 | 0.99 |

learning-based method using the InceptionV3 model comprising of a large number of training parameters. Li et al. [32] also proposed a VGG16 transfer learning based method. A layer-guided CNN was implemented by Huang et al. [19] for the classification of OCT images. An EfficientNet-B7 model was proposed by Chetoui et al. [5] for the classification of OCT images. As proposed by Najeeb et al. [36], a CNN model trained from scratch also underperforms compared to the proposed method. The proposed lightweight CNN has 0.41 M training parameters in comparison to the methods given by Zhang et al. [28], Kaymak et al. [26], Li et al. [32], Chetoui et al. [5], and Najeeb et al. [36] with 21.8 M, 91.7 M, 15.2 M, 64.8 M, and 102.7 M training parameters, respectively. These techniques do not consider the class imbalance problem in the ZhangLab dataset. The proposed method performs better even without using class weight balancing (OCT_GAP + CCE in Table 3). However, the results are further improved by addressing the class imbalance issue using class weight balancing. Thus, given consideration to the class imbalance problem, the proposed method has an added advantage over other methods existing in the literature.

**Performance on the Duke dataset:** As the model (OCT_GAP + CWB + CCE) performs well for ZhangLab dataset, the same combination is applied to the Duke dataset. The results are summarized in Table 5. The partition with 70% training data with CWB produced the best results for the Duke dataset, and the same is used to compare the performance with the existing methods on this dataset. The results summarized in Table 6 clearly show that deep learning-based methods give better results in comparison to hand-crafted feature-based techniques given by Venhuizen et al. [48], Srinivasan et al. [44], Hussain et al. [21], and Hani et al. [17]. The proposed method is also compared with recently proposed deep learning-based methods. Qingee et al. [23] proposed a transfer learning (TL) based approach using the InceptionV3 model, whereas Rasti et al. [39] and Rong et al. [41] proposed specialized CNN architectures. The proposed method performed better than all the hand-crafted feature-based methods in terms of $ACC$, $SENS$, $SPEC$, and $F1 - score$ except for [44] in terms

**Table 5** Performance of the proposed method on Duke dataset for different partitions

| Partition (Training-Validation) | ACC | SENS | SPEC | *F1-score* | AUC |
|---|---|---|---|---|---|
| 60-40 | 83.07 | 70.27 | 91.79 | 0.83 | 0.87 |
| **70-30** | **98.46** | **98.40** | **99.32** | **0.98** | **0.99** |
| 80-20 | 95.89 | 96.08 | 98.12 | 0.96 | 0.97 |
| 90-10 | 97.96 | 98.06 | 99.04 | 0.98 | 0.99 |

**Table 6** Performance of the proposed method on Duke dataset

| Author | Method | ACC | SENS | SPEC | *F1-score* |
|--------|--------|-----|------|------|-----------|
| Venhuizen et al. [48] | Layer Features | 70.00 | 71.42 | 68.75 | 0.70 |
| Srinivasan et al. [44] | HOG | 95.55 | 100 | 86.67 | 0.93 |
| Qingee et al. [23] | Inception V3 TL | 98.96 | 98.30 | 99.15 | - |
| Rasti et al. [39] | Wavelet CNN | 97.98 | 99.11 | - | - |
| Hussain et al. [21] | Layer Features | 96.89 | 96.67 | 100 | 0.98 |
| Rong et al. [41] | Surrogate CNN | 92.05 | 90.59 | 93.71 | - |
| Hani et al. [17] | Texture Based Features | 87.45 | 92.36 | 89.11 | - |
| Proposed Method | OCT_GAP + CWB + CCE | 98.46 | 98.40 | 99.32 | 0.98 |

of $SENS$ and $SPEC$, respectively. It also performed better than Rong et al. [41]. The CWB based lightweight architecture proposed in this work performed better than [41] and yielded comparable performance to [23] and [39]. The advantage of the proposed method over these methods lies in the fewer training parameters compared to transfer learning based methods. The proposed method uses 0.41 M training parameters compared to the transfer learning based method by Qingee et al. [23] with 21.8 M training parameters.

The ROC curve in the classification problem depicts the trade-off between sensitivity and specificity. This trade-off is important for medical image analysis. The ROC curves for ZhangLab and Duke datasets are shown in Fig. 5. ROC curve for each class is plotted separately, and the area under the curve is calculated, which is depicted in the bottom right corner. Different values of True positive and false positive rates are plotted for different classification thresholds to get the ROC curve. ROC curve having AUC value 1 is considered as the best model. The AUC achieved through the proposed method is 0.99 for both ZhangLab and Duke datasets, respectively.

**Cross-validation Performance on Datasets:** To check the generalization capability of the proposed CNN model, an experiment is performed for cross-dataset validation by only keeping common classes, i.e., DME and normal class, from both the datasets for training and testing purposes. The model trained on the images of the ZhangLab dataset showed ACC,
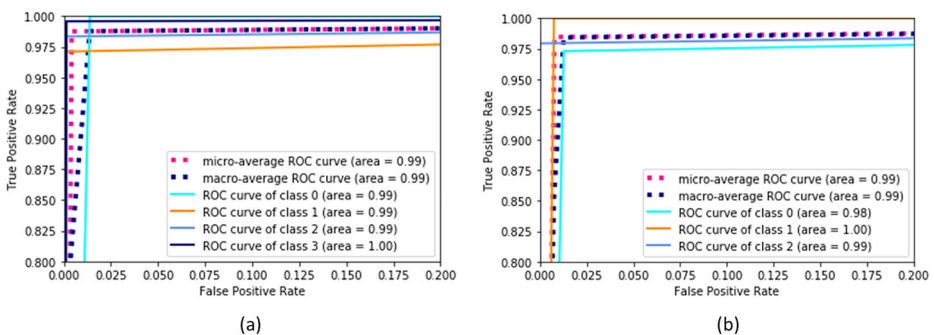


**Fig. 5** ROC plots: (a) ZhangLab dataset - CNV (class-0), DME (class-1), drusen (class-2), and normal (class-3) and (b) Duke dataset - AMD (class-0), DME (class-1), and normal (class-2)

SENS, SPEC, F1-score, and AUC as 90.78%, 83.10%, 96.80%, 0.90, and 0.91, respectively. The model trained on Duke dataset and tested on ZhangLab dataset yielded 77.68%, 82.64%, 73.14%, 0.78, and 0.78 values for ACC, SENS, SPEC, F1-score, and AUC, respectively. This analysis shows that the model trained on the ZhangLab dataset generalized better for OCT images of DME and normal classes than the model trained on the Duke dataset. This shows the effectiveness of the larger dataset and its better generalization capability.

**Model Interpretation using CAMs:** The output weights learned during the training phase are projected back to the last convolutional layer filters to get the class activation maps of some example images from ZhangLab and Duke dataset as shown in Figs. 6 and 7 respectively. In the case of OCT images, the important features responsible for ocular diseases are found near the retinal layers. These important features are highlighted in red and yellow colors. The dip in the retinal layer in a normal image is visible in Figs. 6(a) and 7(a). The rise in the retinal layers in the case of a CNV and Drusen images can be seen in Fig. 6(b) and (c). DME, Drusen, and AMD class activation maps also show abnormality in the intermediate layers of the retina.
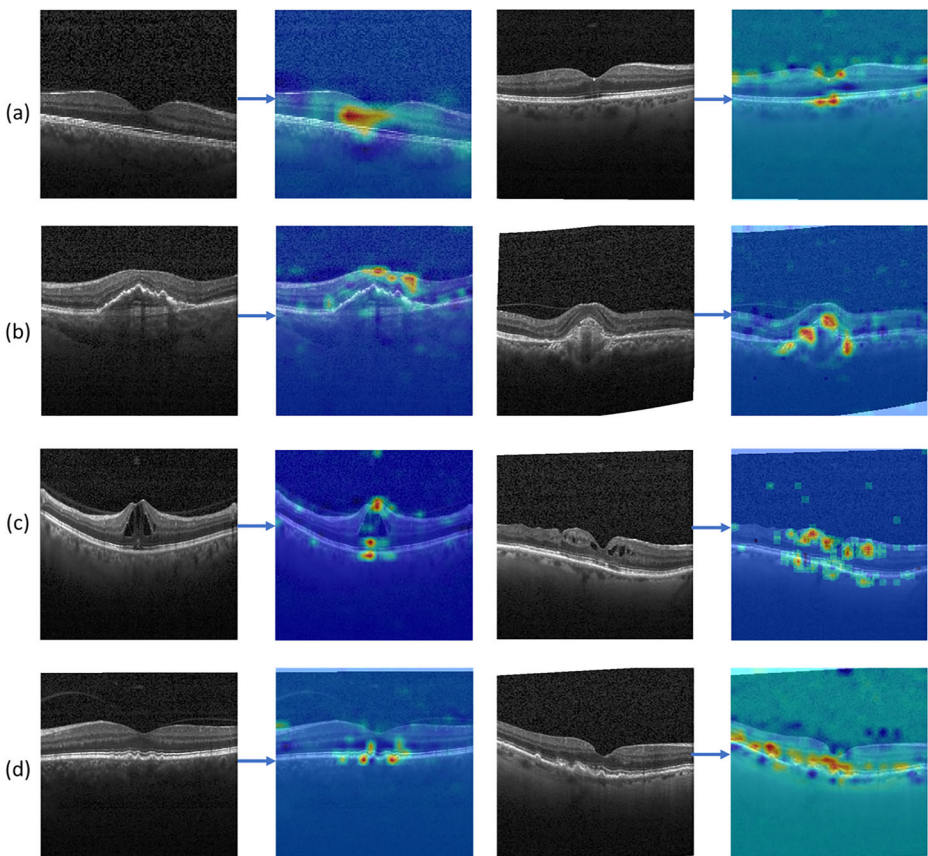


**Fig. 6** Class activation maps obtained for OCT images of ZhangLab dataset: (a) normal class images, (b) CNV class images, (c) DME class images, and (d) Drusen class images
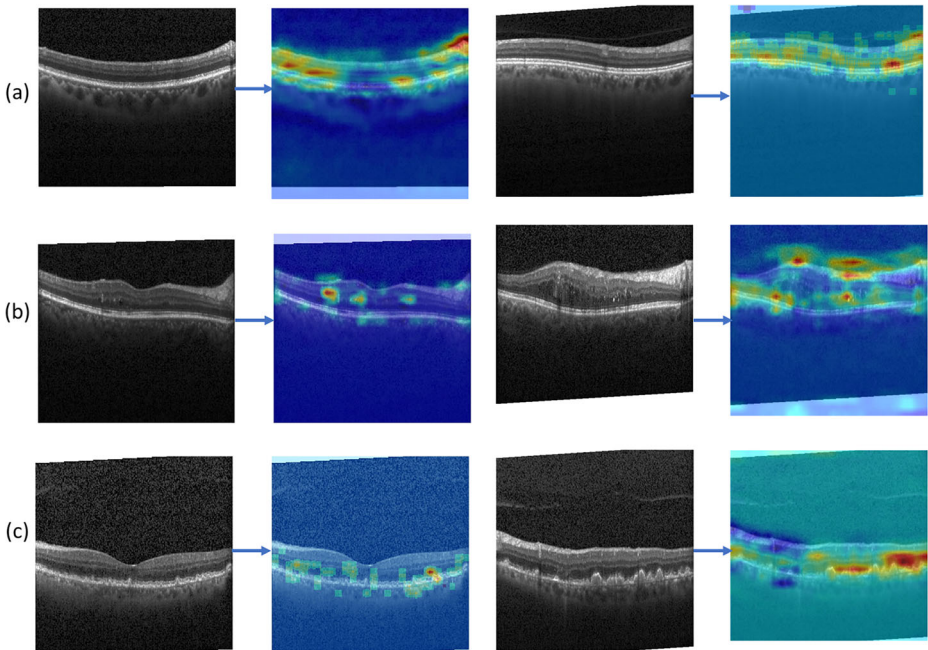
**Fig. 7** Class activation maps obtained for OCT images of Duke dataset: (a) normal class images, (b) DME class images, and (c) AMD class images

# 5 Conclusion

This work presents an automated OCT classification method using deep learning-based CNN with CWB for ocular diseases. Medical imaging datasets suffer from an imbalance in the number of samples for normal and diseased classes. Addressing the class imbalance issue is crucial in CAD systems for the diagnosis of ocular diseases. A global average pooling based model is proposed in this work to calculate feature maps from CNN. The model is trained by applying class weights that penalize the majority class. The trained model is finally tested on publicly available ZhangLab and Duke datasets for OCT classification. The final results are evaluated using confusion matrix based parameters. The proposed method yielded 99.17% and 98.46% classification accuracy and 0.99 and 0.98 $F1 - score$ for ZhangLab and Duke datasets. The results show improvement in the testing performed on the imbalance ZhangLab dataset with the model using the CWB technique compared to the model without CWB. The class activation maps calculated using the final convolutional layer show important features responsible for a particular disease. The overall advantage of the proposed method over other state-of-the-art methods discussed here lies in the light CNN architecture, CWB, and class activation map construction. Such systems can be used to assist Ophthalmological experts. Network compression methods for well-known CNN architectures, different class balancing techniques, and interpretation of CNN for OCT images through various visualization techniques can be explored in the future.

## Declarations

**Conflict of Interests** The authors declare that they have no conflicts of interest associated with this publication.

## References

1. Alqudah AM (2020) Aoct-net: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. Med Biol Eng Comput 58(1):41–53
2. Bishop CM (2006) Pattern recognition and machine learning springer
3. Buda M, Maki A, Mazurowski MA (2018) A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 106:249–259
4. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
5. Chetoui M, Akhloufi MA (2020) Deep retinal diseases detection and explainability using oct images. In: International conference on image analysis and recognition, Springer, pp 358–366
6. Daumé H (2017) A course in machine learning Hal daumé III
7. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D et al (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Med 24(9):1342–1350
8. Drexler W, Fujimoto JG (2008) State-of-the-art retinal optical coherence tomography. Prog Retin Eye Res 27(1):45–88
9. Ertekin S, Huang J, Giles CL (2007) Active learning for class imbalance problem. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp 823–824
10. Fawcett T (2006) An introduction to roc analysis. Pattern Recognit Lett 27(8):861–874
11. Fujimoto J, Drexler W (2008) Introduction to optical coherence tomography. In: Optical coherence tomography, Springer, pp 1–45
12. Gao L, Zhang L, Liu C, Wu S (2020) Handling imbalanced medical image data: A deep-learning-based one-class classification approach. Artif Intell Med 108:101935
13. Gholami P, Roy P, Parthasarathy MK, Lakshminarayanan V (2020) Octid: Optical coherence tomography image database. Comput Electr Eng 81:106532
14. Goodfellow I, Bengio Y, Courville A (2016) Deep learning MIT press
15. Greenspan H, Van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35(5):1153–1159
16. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G (2017) Learning from class-imbalanced data: Review of methods and applications. Expert Syst Appl 73:220–239
17. Hani M, Ben Slama A, Zghal I, Trabelsi H (2020) Appropriate identification of age-related macular degeneration using oct images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp 1–11
18. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
19. Huang L, He X, Fang L, Rabbani H, Chen X (2019) Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network. IEEE Signal Process Lett 26(7):1026–1030
20. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, Hee MR, Flotte T, Gregory K, Puliafito CA et al (1991) Optical coherence tomography. Science 254(5035):1178–1181
21. Hussain MA, Bhuiyan A, Luu CD, Smith RT, Guymer RH, Ishikawa H, Schuman JS, Ramamohanarao K (2018) Classification of healthy and diseased retina using sd-oct imaging and random forest algorithm. PloS one 13(6):e0198281
22. Ibrahim MR, Fathalla KM, Youssef SM (2020) Hycad-oct: a hybrid computer-aided diagnosis of retinopathy by optical coherence tomography integrating machine learning and feature mapslocalization. Appl Sci 10(14):4716
23. Ji Q, He W, Huang J, Sun Y (2018) Efficient deep learning-based automated pathology identification in retinal optical coherence tomography images. Algorithms 11(6):88
24. Johnson JM, Khoshgoftaar TM (2019) Survey on deep learning with class imbalance. J Big Data 6(1):27

25. Karri SPK, Chakraborty D, Chatterjee J (2017) Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. Biomed Opt Express 8(2):579–592

26. Kaymak S, Serener A (2018) Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. In: 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), pp 265–269

27. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122–1131

28. Kermany D, Zhang K, Goldbaum M (2018) Large dataset of labeled optical coherence tomography (oct) and chest x-ray images, Mendeley Data, v3. https://doi.org/10.17632/rscbjbr9sj 3.

29. King G, Zeng L (2001) Logistic regression in rare events data. Polit Anal 9(2):137–163

30. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1(4):541–551

31. Lemaître G, Rastgoo M, Massich J, Cheung CY, Wong TY, Lamoureux E, Milea D, Mériaudeau F, Sidibé D (2016) Classification of sd-oct volumes using local binary patterns: experimental validation for dme detection, Journal of ophthalmology

32. Li F, Chen H, Liu Z, Zhang X, Wu Z (2019) Fully automated detection of retinal disorders by image-based deep learning. Graefe's Arch Clin Exp Ophthalmol 257(3):495–505

33. Lin M, Chen Q, Yan S (2013) Network in network, arXiv:1312.4400

34. Ling CX, Sheng VS (2010) Cost-sensitive learning. Encyclopedia of machine learning, pp 231–235

35. Liu Y-H, Liu C-L, Tseng S-M (2018) Deep discriminative features learning and sampling for imbalanced data problem. In: 2018 IEEE International Conference on Data Mining (ICDM), IEEE, pp 1146–1151

36. Najeeb S, Sharmile N, Khan MS, Sahin I, Islam MT, Bhuiyan MIH (2018) Classification of retinal diseases from oct scans using convolutional neural networks. In: 2018 10th International Conference on Electrical and Computer Engineering (ICECE), IEEE, pp 465–468

37. Nanni L, Ghidoni S, Brahnam S (2017) Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recogn 71:158–172

38. Powers DM (2020) Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation

39. Rasti R, Mehridehnavi A, Rabbani H, Hajizadeh F (2018) Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier. J Biomed Opt 23(3):035005

40. Rasti R, Rabbani H, Mehridehnavi A, Hajizadeh F (2017) Macular oct classification using a multi-scale convolutional neural network ensemble. IEEE Trans Med Imaging 37(4):1024–1034

41. Rong Y, Xiang D, Zhu W, Yu K, Shi F, Fan Z, Chen X (2018) Surrogate-assisted retinal oct image classification based on convolutional neural networks. IEEE J Biomed Health Inform 23(1):253–263

42. Sanyal A, Kumar P, Kar P, Chawla S, Sebastiani F (2018) Optimizing non-decomposable measures with deep networks. Mach Learn 107(8-10):1597–1620

43. Settles B (2012) Active learning. morgan claypool, Synthesis Lectures on AI and ML

44. Srinivasan PP, Kim LA, Mettu PS, Cousins SW, Comer GM, Izatt JA, Farsiu S (2014) Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. Biomed Opt Express 5(10):3568–3577

45. Sunija A, Kar S, Gayathri S, Gopi VP, Palanisamy P (2021) Octnet: A lightweight cnn for retinal disease classification from optical coherence tomography images. Comput Methods Programs Biomed 200:105877

46. ŢăLu S-D (2013) Optical coherence tomography in the diagnosis and monitoring of retinal diseases, ISRN biomedical imaging

47. Thomas A, Harikrishnan P, Krishna AK, Palanisamy P, Gopi VP (2021) A novel multiscale convolutional neural network based age-related macular degeneration detection using oct images. Biomed Signal Process Control 67:102538

48. Venhuizen FG, van Ginneken B, Bloemen B, van Grinsven MJ, Philipsen R, Hoyng C, Theelen T, Sánchez CI (2015) Automated age-related macular degeneration classification in oct using unsupervised feature learning. In: Medical imaging 2015: Computer-aided diagnosis, vol 9414, International society for optics and photonics, p 94141I

49. Wu J, Zhang Y, Wang J, Zhao J, Ding D, Chen N, Wang L, Chen X, Jiang C, Zou X et al (2020) Attennet: deep attention based retinal disease classification in oct images. In: International conference on multimedia modeling, Springer, pp 565–576

50. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.