



Full Length Article

## DeepOCT: An explainable deep learning architecture to analyze macular edema on OCT images



Gokhan Altan <sup>a,\*</sup>

<sup>a</sup> Iskenderun Technical University, Central Campus, Block A, Computer Engineering Dept., İskenderun, Hatay, Turkey

### ARTICLE INFO

*Article history:*

Received 8 March 2021

Revised 13 November 2021

Accepted 23 December 2021

Available online 24 January 2022

*Keywords:*

Deep learning

Convolutional neural networks

Optical coherence tomography

Macular edema

DeepOCT

### ABSTRACT

Macular edema (ME) is one of the most common retinal diseases that occur as a result of the detachment of the retinal layers on the macula. This study provides computer-aided identification of ME for even small pathologies on OCT using the advantages of Deep Learning. The study aims to identify ME on OCT images using a lightweight explainable Convolutional neural networks (CNN) architecture by composing significant feature activation maps and reducing the trainable parameters. A CNN is an effective Deep Learning algorithm, which consists of feature learning and classification stages. The proposed model, DeepOCT, focuses on reaching high classification performances as well as popular pre-trained architectures using less feature learning and shallow dense layers in addition to visualizing the most responsible regions and pathology on feature activation maps. The DeepOCT encapsulates the block-matching and 3D filtering (BM3D) algorithm, flattening the retinal layers to avoid the effects arising from different macula positions, and excluding non-retinal layers by cropping. DeepOCT identified OCT with ME with the rates of 99.20%, 100%, and 98.40% for accuracy, sensitivity, and specificity, respectively. The DeepOCT provides a standardized analysis, a lightweight architecture by reducing the number of trainable parameters, and high classification performances for both large- and small-scale datasets. It can analyze medical images at different levels with simple feature learning, whereas the literature uses complicated pre-trained feature learning architectures.

© 2021 Karabuk University. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Optical Coherence Tomography (OCT) is a non-invasive ophthalmic technique that enables controlling the retinal conditions. The pattern of the retina is comprised of cross-sectional layers. The distinctive layers' map, thickness, and contiguity values state the type and severity of the ophthalmic diseases. The OCT images are commonly employed to help identify ophthalmological diseases by visualizing the ocular pathologies and formal degeneration in retinal layers [1]. The OCT images are frequently utilized to estimate the performance of deep learning (DL) algorithms due to providing facilities to control the tissue at microscopic resolution [2,3].

The number of researches on the computerized analysis of OCT has increased in popularity depending on the developments in the medical image processing techniques and DL algorithms. One of the most frequent ophthalmic diseases is macular edema in literature. The macula is the pigmented tissue near the centre of the

retina. Macular edema is described with the detachment of retinal layers and the occurrence of fluid-filled pathologies in macula owing to many conditions, including diabetes, age-related macular degeneration, genetic disorders, injuries, and more [3–5]. Conventional image processing techniques focused on the identification of diabetic macular edema (DME) using hand-crafted features [6,7]. In recent years, a majority of the studies focused on adapting the efficacy of DL algorithms into the identification of pathologies and diagnosis of DME on OCT images. Transfer learning, feature learning, and pruning the pre-trained Convolutional Neural Networks (CNN) architectures are the common approaches for many research fields to handle novel DL architectures. The studies utilized various DL architectures, including VGG-16 [8–12], VGG-19 [13], ResNET50 [13,14], ImageNET-InceptionV3 [15–17], and AlexNET [16–18] to detect DME on OCT images.

The recent developments on pre-trained DL architectures have reached high levels that are difficult to outperform. The focus of novel papers is composing lightweight architectures presenting similar capabilities as well as deeper models besides ensuring the explainability with visual supports for clinical validation. The importance of the proposed method is composing a lightweight CNN architecture with high DME identification performances using

\* Corresponding author.

E-mail address: [gokhan.altan@iste.edu.tr](mailto:gokhan.altan@iste.edu.tr)

† Peer review under responsibility of Karabuk University.

an acceptable number of trainable parameters compared to popular pre-trained architectures in addition to visualizing the most responsible retinal activation maps with clinical relevance.

Computerized analysis of the OCT images is an important step to provide a more detailed analysis for ophthalmologists. This study aims at identifying macular edema on OCT images using a lightweight explainable CNN architecture by composing significant feature activation maps and reducing the trainable parameters. The paper experimented with a variety of CNN structures on both large- and small-scale OCT databases to classify DME and non-DME subjects. The exact identification of DME with small pathologies is a time-consuming detailed examination for even skilled ophthalmologists. The proposed DeepOCT architecture performs identification of small edemas with feature mapping, assigns a standardized analysis with preprocessing, and attains an optimized lightweight DL architecture by reduced feature learning and fine-tuned supervised model on OCT images.

Herein, block-matching and 3D filtering (BM3D), flattening of retinal layers and cropping were applied to subtract non-retina layers. Various CNN architectures on large- and small-scale datasets were experimented with to specify a high-performance CNN structure for DME identification. The main contributions of this proposal can be encapsulated as:

1. The proposal is a lightweight architecture with much less number of the parameters required to train the network compared to popular pre-trained architectures
2. A standardized aligning position for OCT scans is provided by curvature correction of retinal layers analysis and excluding non-retinal sections
3. The architecture reached promising DME identification performance rates by modelling from scratch. Hence, it composed initialization weights for transfer learning to adapt knowledge into the analysis of related medical images.
4. The DeepOCT is separately implemented on small- and large-scale datasets To see the effectiveness of the proposed architecture. It determined the robustness and generalization impact of the same CNN architecture on OCT scans with different specifications.

The remaining of the paper is organized as follows: A detailed related works are presented with a detailed comparison to highlight the superiority of the proposal in 2. The database specifications, preprocessing, the proposed DeepOCT architecture, and experimental setup in training are explained in 3. 4 presents the experimental results for proposals. A complete comparison with the state-of-the-art method is explained in 5. Finally, the proposal is concluded in 6.

## 2. Related Works

Several researchers have put effort into computer-aided analysis for identifying retinal diseases with the recent advancements in deep learning. Significant disadvantages of existing works are the explainability and interpretability of the architectures, limited clinical relevance, the complicated CNN architectures, and their dependency on data specifications. This section explains the main differences of the proposal and further clarification the main contributions concerning related works.

Training a CNN model is a time-consuming and big data dependent process for obtaining a better generalization. Transfer learning enables overcoming these limitations by recounting the knowledge of trained architectures. Transfer learning is only convenient to be adapted if the initial and target problems are related enough for the feature learning of the architecture to be relevant.

However, it has been commonly adapted for the OCT images for DME identification from conventional object detection models [14,9,10,16–18,11,19]. DeepOCT was trained from scratch without any weight factor optimization on pre-trained architectures. Hence, the generated feature activation maps present low-, mid-, and high-level with regard to retinal evaluation.

The capabilities of DL algorithms have taken classification and detection techniques to a new level. That is why advanced image processing techniques have equipped feature learning. Composing deeper CNN models provides the opportunity for more detailed analysis with sequential convolutions besides an increase in the number of parameters required to train the network and time-complexity as a disadvantage. The recent main focus of the researchers is performing high classification achievements as well as deeper and combined models using lightweight architectures with a reduced number of trainable parameters. Whereas popular pre-trained architectures have a variety of classification parameters through 25 M (ResNet-50) [14], 60 M (AlexNet) [19,18], 138 M (VGG-16) [9–11], DeepOCT is an acceptable lightweight architecture compared to other architectures with 7.9 M parameters.

Many related works reported high classification performances for accuracy, specificity, sensitivity, the area under curve ratio, and more using various techniques. The most crucial disadvantage of the proposals is that it is a black box in terms of estimating the responsibility of learned features and how much they affect the model [14,13,20,11,9,10,16–18,21,12,2]. Even if the model achieved high performance, it is more valuable for clinical experiments to estimate the most responsible regions for the learning procedure rather than predicting the category. In recent years, explainable AI techniques have been used to support medical decision support mechanisms and for assessing the clinical relevance of the output [8,19]. The DeepOCT has the advantages of modeling an explainable lightweight architecture related to medical images from scratch.

Due to the intensive speckle noise in the OCT, analyzing the denoised OCT images has always been a prerequisite for many types of research on segmentation and retinal pathology localization. In contrast to analysis of raw OCT scans [9,12,10,18,8,19], various filtering techniques, which are hard to perform a complete comparison, were applied in different sequences with different parameters for even the same procedures. Extracting the retinal boundaries can be challenging for severe DME pathologies with different curvature and background specifications. Therefore, the papers followed various procedures, including median filtering [14,9], BM3D [14,16], sparsity-based block matching [14,21], and Surrogate Image Generation [21] for extraction of ophthalmological layers. OCT-specific non-automatic cropping, which is a time-consuming procedure for large-scale databases, is utilized for extracting the relevant regions for DME [13,20,10,16,17]. Since the CNN architectures have the ability to learn shape-variants of the OCT scans, a flattened retinal layer for DME is more effective to force learning the pathology in the retinal layer [20]. The proposal performs a standardized preprocessing by applying BM3D for noise suppression and retinal edge detail preservation, flattening for curvature correction on OCT, binarization for uppermost and lowermost boundary detection, and automatic cropping for excluding non-ophthalmologic patterns in OCT.

Though a limited number of researchers have handled the disadvantages of transfer learning by training their own CNN architectures on OCT scans, none of the related works has nonetheless reached a satisfying level that would overcome all these disadvantages.

In [14], fine-tuning on the popular CNN models is used to improve the identification accuracy of DME on preprocessed OCT images. They experimented with various removals of deep layers on GoogLeNet, ResNet, and DenseNet architectures. Their proposal

reduced the number of parameters to 8.5 M in the CNN model about 99% classification performances. However, it is not clear to understand the clinical validity using the achievements without the correlation between the feature activation maps and disease pathology.

In [13], the efficiency of multiple machine learning algorithms and CNN was compared using the wavelet transform as unsupervised learning to define the spatial-frequency domain. They reported the supremacy of the CNN for the diagnosis of macular pathology. However, they trained their proposal using a limited number of OCT scans which is not suitable for the generalization of DME pathology.

In [20], a multi-scale CNN model with a new cost function was proposed for macular edema pathology on OCT. They compared the efficiency of their CNN model in training time and discrimination ability with existing CNN architectures and designated the superiorities of their CNN model. However, it has no visual evaluation to ensure which sections were learned by the proposed CNN architecture.

In [12], a ribcage network for dense layers was suggested by concatenating the hand-crafted features (Gabor filters and scale-invariant feature transform) with a convolutional block at the middle layers. Their proposal was trained on the ZhangLab database and achieved a significant reduction in training time and higher retinal disease identification performance than VGG-16, DenseNet, and Xception architectures on OCT images. However, the popular architectures initialized with the weights on ImageNet, which is unrelated knowledge for OCT. Moreover, their proposal includes a series of time-consuming procedures, including various hand-crafted techniques for feature integration in the discrimination network. Moreover, using random crop in the data augmentation risks feeding the network with non-related pathologies for small DMEs in the ZhangLab.

In [21], a surrogate feature generation was predicted by applying an optimized non-orthogonal wavelet filter to OCT images. They proposed a lightweight CNN model with four convolutional layers. The proposal reported the profitability of denoising and mask extraction to identify the retinal diseases using CNN models. However, a small number of OCT scans were utilized in training. This case is prone to a low generalization of diversity on unsupervised data instead of complex representation. Moreover, visual evaluation is not possible for the correlation of learned data and pathology.

Lastly, it is essential for the proposal to highlight the evaluation of datasets concerning the scales and variety of usage. A majority of the literature trained CNN architectures using a limited number of OCT scans. However, the small-scale dataset is not suitable for training deep learning models with high generalization even if using data augmentation. Although not sufficient, this effect is minimized by using cross-validation [13,20,9,21,10,16–18]. Held-out procedure for splitting training and testing sets is used for various CNN architectures on the large-scale database (ZhangLab) [14,12,18,8,19]. Whereas the related works commonly focused on combining the available databases or just one of them, this study implements separate training on each OCT database for the same architecture stratifying by databases. Thus, it determines the generalization impact of the same CNN architecture on OCT scans with different specifications.

### 3. Materials and Methods

#### 3.1. Database

OCT images allow checking the condition of the retinal layers during the clinical examination as a consequence of presenting the different scans of the macula in volumes. Therefore, different

scans can be obtained from the patient to assess the presence of pathology, edema, and degeneration-based ophthalmic disorders. The size and contrast of the pathology may vary depending on the contiguity of scans.

The OCT from non-DME subject presents normal foveal structural features with a subsidence of the foveal pit and extrusion of plexiform layers in particular. The OCT from the patients with DME lacks extrusion of plexiform layers, causing retinal detachment and a foveal pit with shallow indent for even various sizes for different macular severities.

There are different local, private, and open-access clinical databases on OCT images. The most commonly used databases are Zhanglab [19], Singapore Eye Research Institute (SERI) [22], and Duke datasets. ZhangLab is a comprehensive database among these databases by the number of subjects and scope of retinal diseases. In contrast, SERI and DUKE Datasets have limitations on the number of OCT images and subjects. In this study, ZhangLab and SERI datasets were used to analyze how the proposed DL-based macular edema classification models are effective on various specifications. ZhangLab dataset consists of 62,489 OCT images (11,349 DME and 51,140 normal) with different resolutions. The SERI dataset was collected from 32 OCT volumes (16 DME and 16 normal). It has 128 B-scan OCT images for each subject with a resolution of 512x1024 pixels.

Dataset is normally divided into three subsets (training, validation, and testing) in the training of the DL algorithms. The training set is utilized for training the model, validation is used for hyper-parameter optimization to evaluate the training using loss at the end of each epoch, and testing is used for final evaluation. 10% of the training set was split as a validation set to prevent the proposed CNN models from overfitting in training.

The datasets were stratified by patients instead of stratifying by OCT images for validation of proposals. Thus, none of the OCT images from a patient is used in both training and test set. Test group of ZhangLab is comprised of 500 images (250 OCT with DME and 250 normal OCT) as validation set. The rest of the ZhangLab database was used in training the proposed models on DME identification. The SERI dataset is separated using an 8-fold cross-validation algorithm to avoid overfitting. The test groups composed using 8-fold cross-validation and reserved discrete folds provide an appreciation performance measurement on the proposed DL models using independent feature sets for training and testing.

The proposed DL models were trained with held-out training and testing sets on ZhangLab to attain a precise performance comparison. Due to the limited data, 8-fold cross-validation was carried out on SERI.

#### 3.2. Preprocessing

Due to noise in raw OCT images, they may not have the essential sufficiency for a complete analysis of retinal edema or pathology. Therefore, the retinal layers and edema pathology can be designated using enhancing techniques by surpassing the unclear edges and eliminating non-ophthalmological background surfaces. The OCT images have speckle noise with a low signal-to-noise ratio. The preprocessing steps of the proposal are (1) denoising the OCT images for noise suppression and retinal edge detail preservation, (2) flattening for curvature correction on the shape variety of OCT scans, (3) binarization for detection of uppermost and lowermost boundary in OCT, (4) cropping at the obtained boundaries to excluding non-ophthalmologic patterns, and (5) resizing the image to feed the DL architectures. In the first step, the BM3D algorithm was applied to the OCT images to reduce noise [23]. The BM3D algorithm has higher superiority and effectiveness than conventional filtering methods on OCT images [14–16]. It is

comprised of designating a group of similar patches from non-local regions in OCT images and extracting 3-D wavelet representation of patch group, respectively [23].

Sigma parameter of BM3D is the value of the noise in the matching step to provide noise suppression and edge detail preservation in the OCT. Using higher sigma values resulted in nesting layers in the retina and sharp edges. Small sigma values had deficiencies in the restoration of retinal details. The BM3D parameters are 45 and 35 for sigma for the SERI dataset and ZhangLab dataset, respectively. In this way, the attenuation of noises from OCT images and more apparent edges were generated to attain the retinal layers. Hence, the morphological details in the retinal layers and pathologies become more prominent [24,25]. The position of the retina varies depending on the scans in OCT images. Using a flattening algorithm is necessary to get the retinal layers in proportion to a unified aligning position to minimize morphological variations. Therefore, the anisotropic diffusion strategy in [11] was adapted as the second step of preprocessing. Anisotropic diffusion strategy consists of intra-region smoothing, aligning the individual retinal edges, and performing curvature correction of retina [26]. The parameters of the anisotropic diffusion are applied at one iteration using the default conduction coefficient (kappa) value of 50 as a function of gradient and standard deviation of 0.1 for Gaussian blurring. Each of the OCT images was horizontally cropped using the positioned retinal layers' uppermost and lowermost locations that were detected by binarization of the images using an adaptive threshold value of 44. The threshold of 44 is a manually determined average value to have absolute retinal layers after the binarization of random ten OCT scans with DME and nonDME. In this manner, the non-ophthalmologic patterns in OCT images were excluded. At the last step, each OCT image was resized to 224x224 pixels for introducing a comparable architecture with the popular pre-trained networks (VGG-16, VGG-19, ResNet, EfficientNet, and more). 1 depicts the OCT images and preprocessing stage in a sequel manner for DME and normal.

### 3.3. Convolutional Neural Networks

The CNN architectures consist of feature learning with sequentially combining a variety of convolutional layers (CONV), pooling layers, and supervised learning stages with multiple fully-connected layers (FC) [9,13,20,27,28]. The most prevalent pre-trained CNN architectures are AlexNet [16–19], VGGNet [10–13,20], GoogLeNet [14], ResNet [14,17,20], and DenseNet [12]. In this study, a novel lightweight CNN architecture, DeepOCT, was proposed to identify macular edema in OCT images. At the feature learning stage, the proposed models had three CONVs (descending and ascending order of 32, 64, and 96 feature maps) in which the size of convolution filters varied at a range of  $3 \times 3 \sim 13 \times 13$ . Adding noise after each CONV is a popular method in training to improve the robustness of the CNN models for real-valued inputs, regularization progress, and detract overfitting. Gaussian noise ( $\mu=0.35, \sigma=0.01$ ) was added to emulate the noise effect to the pre-processed OCT images for each CONV. The CNN architecture consisted of a max-pooling ( $2 \times 2$ ) layer to identify dominant pixels of feature maps. Rectified linear unit (ReLU) activation function is commonly utilized to rule out the negative pixels of convolution operation after each CONV. At the classification stage, the classifier was composed of an input layer in which the feature map of the last CONV is flattened, one-or-two FCs, and an output layer with a softmax activation layer (DME-non-DME). The neuron variety at each FC was empirically iterated at a range of 100~600 neurons. The proposed CNN models were trained using a limited number of FCs and neurons at each FC.

In consequence of proposing a CNN architecture that includes not only feature learning but also supervised learning with DNNs,

the experimental setup performs a brute-force (exhaustive) search with a variety of convolution filters and dense layer optimization. The reason for the preference is on the ground of the no free lunch theorem. In other words, each feature map may have the possibility of performing high generalized predictions for different dense layer specifications. Hence, the training assumption is attempting the different feature activation maps by a variety of CONV layers to achieve the highest classification performance by learning the most significant characteristics. The presented achievements belong to the best model combinations in the experiments. Especially, sequential convolutional layers on low-level characteristics at the first steps evolve to reach mid- and high-level features.

The proposals were trained from scratch without any weight factor optimization on pre-trained architectures. Furthermore, owing to many classification parameters, including dropout factor, learning rate, and more in supervised learning, the proposed architecture used fixed parameters as a dropout factor of 0.5 for each FC, 50 epochs, and a batch size of 50. Adaptive Moment Estimation (Adam) algorithm with an initial learning rate of 0.001 and epsilon constant of 1e-07 for numerical stability on stochastic gradient descent is selected as the optimizer of DL. No early stopping was utilized in training. The softmax function is set as the output function and binary cross-entropy loss function is adapted to the proposed DeepOCT. The sigmoid function is chosen as the activation function of the FCs.

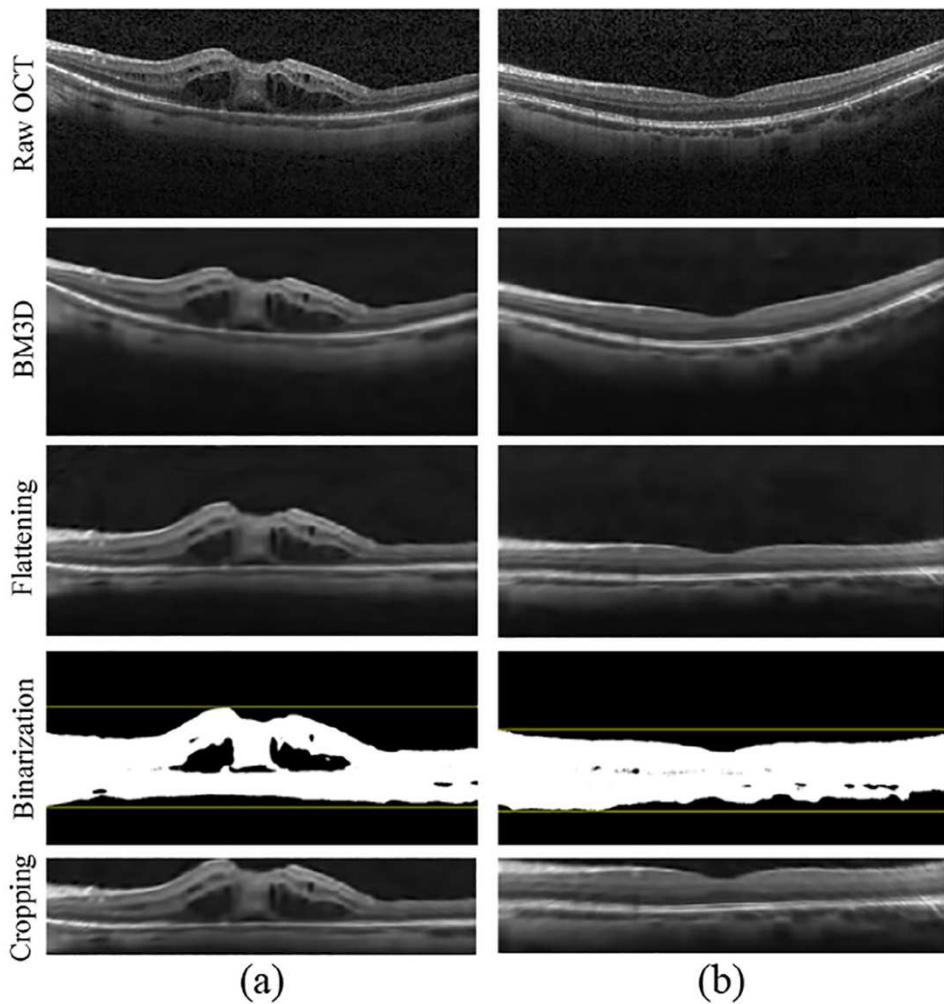
Feature gradient activation map was used to generate a visualization map by the sum of pooling layers to visualize the result of proposed DeepOCT architecture [29,30]. It enables preserving the spatial location information of the targets for each corresponding class in the CNN-based models. The heat map on OCT presents the degree of significance depending on the CNN architecture to determine which regions are more responsible for diagnosing DME. Thus, DeepOCT responds to validating the predictions for actual pathology regions on the OCT as an explainable model.

The DL analysis on the separation of OCT images with DME and non-DME concentrated on the same image options, CONV layers with the same arguments, and a fixed range of classification parameters. The supervised stage of the CNN models was performed at a limited variety of neuron numbers, dense layers, iterations, and other classification parameters for getting an optimized DL model.

The statistical test characteristics are calculated using the contingency table distribution of the predicted and actual labels to designate the classification performance of the models in machine learning. Therefore, overall accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were considered in the comparison of the CNN models [31]. BDPV (Inference and Design for Predictive Values in Diagnostic Tests) package in R was applied to obtain these characteristics. Moreover, the DeLong method was used to compare the performances of proposed CNN models at 95% confidence interval (pROC package in R). It calculates the variance of the area under the curve (AUC) via the quantile function of the normal distribution [32]. 1 presents the highest achievements for various CNN models, separately. Since training CNN models is very time-consuming and requires advanced hardware (GPU), the classification performances on DME identification, which were achieved using the common pre-trained CNN models and fine-tuning, were directly compared with the literature. The efficiency of the proposed CNN models for various CONV filters and dense layers has been experimented with two datasets.

## 4. Experimental Results

The DeepOCT model was trained on different datasets (ZhangLab and SERI) separately. 2 presents the highest classification



**Fig. 1.** Preprocessing stage for OCT images with DME (a) and non-DME (b). It includes BM3D filtering to obtain more clear edges among retinal layers, flattening to handle a standardized positioning of OCT, binarization to get the uppermost and lowermost position of retinal layers, and cropping to exclude non-retinal layers from OCT, respectively.

**Table 1**

The classification performances (%) of the proposed CNN architectures on DME identification.

Model Structure	Dataset	Accuracy (CI)	Sensitivity (CI)	Specificity (CI)	NPV (CI)	PPV (CI)
CONV(12)@96 CONV(5) @64 CONV (13)@32 FC1(370)	ZhangLab	94.90 (91.22–95.06)	92.60 (89.93–94.11)	97.20 (96.23–98.56)	92.93 (89.86–94.33)	97.06 (96.13–98.42)
	SERI	81.20 ± 1.26 (80.79–82.48)	90.53 ± 1.42 (89.80–93.17)	71.88 ± 1.03 (71.63–73.42)	88.36 ± 0.94 (88.29–90.04)	76.30 ± 0.97 (75.83–76.93)
CONV(12)@96 CONV(7) @64 CONV(6) @32 FC1(240) - FC2(520)	ZhangLab	94.80 (94.34–95.04)	99.60 (99.43–99.81)	90.00 (88.79–90.35)	99.56 (98.55–99.68)	90.88 (89.72–91.62)
	SERI	91.60 ± 2.41 (86.32–92.75)	93.31 ± 2.69 (88.79–94.62)	89.89 ± 3.10 (88.99–91.83)	93.07 ± 1.72 (91.91–94.32)	90.23 ± 0.96 (90.01–91.23)
CONV(4) @32 CONV(5) @64 CONV (13)@96 FC1(230)	ZhangLab	92.40 (91.84–93.11)	88.60 (87.79–89.21)	96.20 (94.81–96.79)	89.41 (88.87–91.21)	95.89 (95.13–96.23)
	SERI	93.41 ± 0.87 (91.77–96.80)	94.97 ± 0.79 (92.34–95.85)	91.85 ± 0.86 (91.11–92.43)	94.81 ± 0.52 (94.72–94.94)	92.09 ± 1.03 (90.79–93.17)
CONV(4) @32 CONV(12)@64 (DeepOCT) CONV(13)@96 FC1 (140) - FC2(470)	ZhangLab	99.20 (98.94–99.96)	100 (99.64–100)	98.40 (95.48–99.12)	100 (99.87–100)	98.43 (96.75–99.43)
	SERI	99.12 ± 0.26 (97.27–99.41)	98.58 ± 0.21 (98.44–99.02)	99.66 ± 0.34 (98.89–99.87)	99.65 ± 0.25 (99.51–99.99)	98.60 ± 0.32 (96.59–99.24)

CI: 95% confidence intervals (Lower–Upper bound), CONV: Convolution layer, FC: Fully connected layer, \*\*\*each CONV was followed by max pooling (2 × 2) and RELU.

results for four CNN architectures at the experimented range of parameters. The proposed method worked very well on both datasets from different institutions with a variety of specifications. The highest classification accuracy rate was achieved via 2 FCs with 140 at the 1<sup>st</sup> FC and 470 neurons for 2<sup>nd</sup> FC using three CONVs feature maps.

The proposed DeepOCT model (see 2) has achieved an AUC value of 0.9828 for DME identification on OCT images. The most responsible features are designated using the feature activation map on the OCT images for DME. Even though existing many retinal layers, the activation map has designated the macular region for tested OCT images. Preprocessing on the retinal layer has

**Table 2**

Comparison of related works on identification of DME using different CNN architectures.

Related Works	CNN Architecture	Parameters	Accuracy	Sensitivity	Specificity	Database
Awais et al. [9]	VGG-16	138 M	87.50	93.50	81.00	SERI
Perdomo et al. [10]	VGG-16	138 M	93.75	93.75	93.75	SERI
Chan et al. [16]	AlexNET + PCA + SVM	60 M	96.07	97.66	94.48	SERI
Chan et al. [17]	AlexNET + SVM	60 M	96.88	93.75	100	SERI
This study	<b>DeepOCT</b>	<b>7.9 M</b>	<b>99.12</b>	<b>98.58</b>	<b>99.66</b>	SERI
Kaymak et al. [18]	AlexNET	60 M	99.80	100	99.60	ZhangLab
Li et al. [11]	VGG-16	138 M	98.80	98.80	98.80	ZhangLab
Kermany et al. [19]	AlexNet	60 M	98.20	96.80	99.60	ZhangLab
This study	<b>DeepOCT</b>	<b>7.9 M</b>	<b>99.20</b>	<b>100</b>	<b>98.40</b>	ZhangLab

LoO-CV: Leave-one-Out Cross Validation, VGG: Visual Geometry Group, PCA: Principal component analysis, SVM: Support vector machines.

contributed positively to the performance of the CNN models for feature dimensionality. This finding proves that eliminating non-retinal regions in medical images has improved the capabilities of detecting the macular abnormalities on OCT.

In terms of visualization performance of the DeepOCT architecture on DME identification, the most responsible area is at the foveal pit for the normal OCT images (see 3.a). The most responsible areas are macular edema areas for not only extensive macular pathology but also small ones (see 3.b-c). The DeepOCT has a sufficient capability to identify the small macular edemas using the advantages of feature activation maps and feature learning.

## 5. Discussion

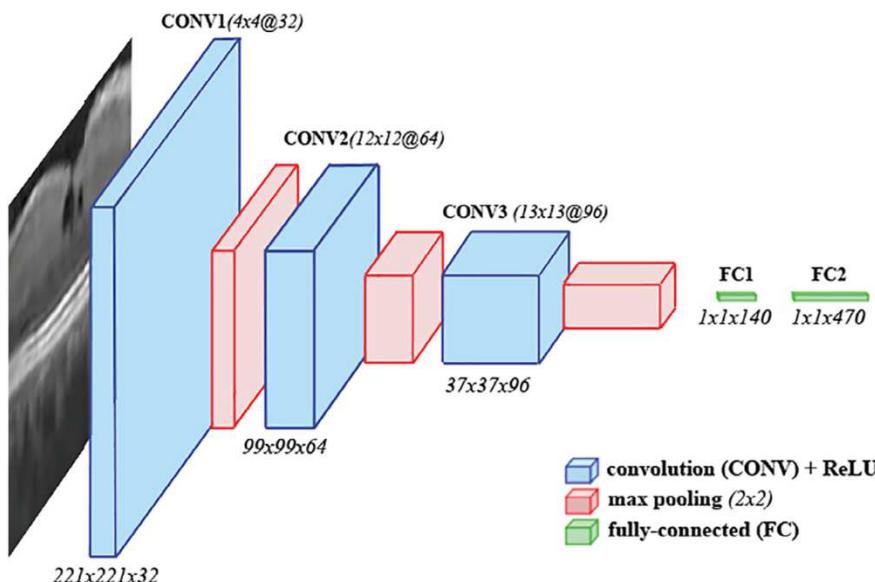
This study's main finding is that the lightweight CNN model also has capabilities to identify the DME on OCT images just as complicated CNN models, including AlexNET, GoogleNET, VGGNet, and more. Despite the fact that a CNN often needs a large amount of data to propose practical applications, the lightweight CNN models have high enough generalization performance for SERI datasets and ZhangLab on the classification of the DME and non-DME. Herein, excluding non-retinal layers from OCT by BM3D, flattening, and cropping provided learning the significant characteristic features of the convolved representations. Raw OCT results in the curse of feature dimensionality, learning non-

ophthalmological sections, and indirectly decelerating the training for even a tiny number of OCT images.

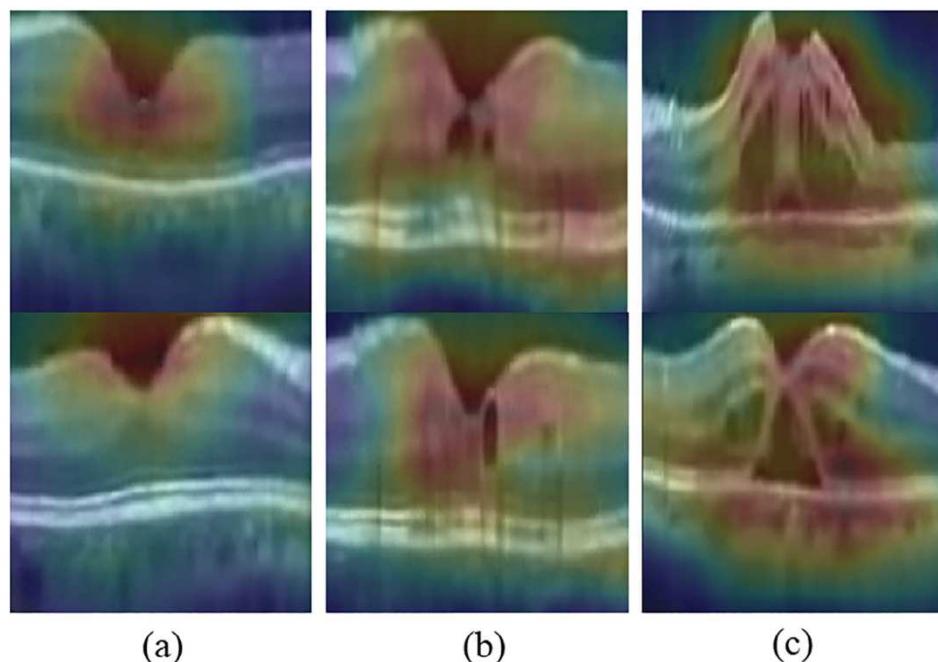
The researchers focused on a variety of preprocessing, feature extraction, and multiple machine learning algorithms to identify abnormalities and different pathologies on retinal layers. However, it could be inadequate to identify DME as similar detachments on retinal layers may cause similar symptoms of other retinal diseases on OCT images.

Sidibe et al. intended to prove the capability of the components of the Gaussian mixture model on DUKE and SERI datasets. Their method reached the rates of 80% and 93% on DUKE; 100% and 80% on SERI for sensitivity and specificity, respectively [6]. Furthermore, Lemaître et al. applied multiple machine learning algorithms to the texture features and reported the efficiency of support vector machines on SERI with separation performance rates of 93.33%, 86.67%, and 100% for accuracy, sensitivity, and specificity, respectively [7].

In recent years, most of the studies applied the pre-trained CNN models to identify pathologies and DME on OCT images. Lee et al. used VGG-16 on the Heidelberg Spectralis database that is comprised of 2.6 million OCT images and separated age-related macular edema from normal with an accuracy rate of 87.63%, a sensitivity rate of 84.63%, and a specificity rate of 91.54% [8]. Rasti et al. used wavelet-based CNN on DUKE and likewise the Heidelberg Spectralis database. They achieved the rates of 98.67% and 98.22% on Heidelberg; 99.33% and 99.11% on DUKE for precision and sensitivity, respectively [13]. Moreover, they proposed a



**Fig. 2.** The architecture of the DeepOCT including the convolutional layer (CONV) with kernel specifications, max-pooling layers, fully-connected layers (FC) with neurons, and the dimensionalities of the feature maps after each CONV.



**Fig. 3.** Visualization of feature activation maps through DeepOCT for normal OCT (a), DME with small macular edema on OCT (b), and DME with extensive macular edema. The red regions indicate the most responsible feature activation maps for DeepOCT.

multi-scale CNN model and reported a precision rate of 99.39% and a sensitivity rate of 97.78% on DUKE [20]. Rong et al. proposed a CNN structure on DUKE using surrogate image generation. They separated DME from non-DME with classification rates of 95.09%, 96.39%, and 93.60% for accuracy, sensitivity, and specificity [21]. Ji et al. performed DME identification using many ImageNet (Inception D2, V2, E3) CNN architectures. They reported the highest classification accuracy rate of 100% using the Inception V3 model [15]. Furthermore, they fine-tuned the same architecture on a local Beijing dataset consisting of 1680 OCT images. They separated three eye diseases with an accuracy rate of 98.86%, a sensitivity rate of 98.30%, and a specificity rate of 99.15%. Whereas these studies focused on identifying DME on the DUKE and local datasets, recent studies commonly analyzed ZhangLab and SERI datasets using efficient pre-trained CNN architectures. Despite using the same databases, a complete comparison is improbable for SERI because none of the papers shared in which subjects were incorporated into training or testing folds. However, the most related papers in terms of database and DL algorithm are compared in 2.

It is possible to improve the identification performance of DME using lightweight architecture and reduce the computational burden on training. The proposed DeepOCT is a functional architecture in terms of classifying the OCT with DME and non-DME. Due to the advanced capabilities of computerized analysis on OCT images, DeepOCT detects small macular edemas that are hard to identify even other pathologies on retinal layers during the retinal angiography. The clarity of the OCT may be different even for the same subject depending on the specifications of the medical device, scan type, pose of the eye, and more. The proposed DeepOCT model implements a standardized approach by flattening the retinal layers, excluding non-retinal layers, focusing on the most diagnostic and deterministic areas for the ophthalmic diseases, and transferring the low- and high-level features among CONVs. The DeepOCT architecture achieves high accuracy rates of 99.12% and 99.20% for DME identification on SERI and ZhangLab datasets. The main benefits of the model are identifying even small macular edema in the retinal layers, decreasing the dependency of the experts, examin-

ing retinal layers with a standardized method, and proposing an optimized lightweight CNN model by comparison to popular CNN architectures.

Although the AI models reach high-enough capabilities for many fields, it is still a major issue of possessing black-box architectures with no explanations for the predicted outputs. This case leads to suspicions on the knowledge of how an AI-based system reached the decision, what to do for an optimum high generalization capability, how to understand prediction straightness and limitations, what the AI model learned, and more. It is a big necessity to reveal the extensions that are trustable by clinicians instead of focusing on nothing but output in the medical decision support system. Even if the prediction is true, the certainty in the visualization of learned features and precise detection of pathological patterns in medical images are the most important parameters for the clinical relevance of the AI systems. By focusing on this issue, the concept of explainable AI has been introduced as a novel approach to enhance intelligent algorithms more interpretable and comprehensible using feature activation maps, visualization techniques on outputs, and rule-based hybrid learning models. The correlation between the feature activation maps and disease symptoms were also assessed by an ophthalmologist. The learned feature activation maps (red regions in 3) had a high responsibility for identification of pathological regions for DME and foveal pit for non-DME in the ophthalmological assessment.

The main benefits of the model are identifying even small macular edema in the retinal layers, decreasing the dependency of the experts, examining retinal layers with a standardized method, and proposing an optimized lightweight CNN model compared to popular CNN architectures.

## 6. Conclusion

The main superior qualitative of the proposed model is estimating DME on OCT images using an optimized pruned CNN architecture. The DeepOCT has the advantages of DL completely

as pre-trained CNN (GoogleNet, AlexNet, VGGNet, DenseNet, and more) in terms of high generalization capability via a simplistic feature learning and FCs. The DeepOCT reached a high generalization for DME identification on SERI dataset with a limited number of OCT. Besides, it was experienced that using large filters at first CONV and decreasing the filter size CONV by CONV affected the performance unlikely. On the contrary, using small filters at CONV1 and increasing the filter size layer by layer improved the classification performance for DeepOCT. Therefore, the low-level features generated in the first CONV are transferred to the subsequent CONVs to extract middle- and high-level features.

Flattening and cropping of OCT scans in preprocessing may be appraised as an estimated data loss that negatively impacts machine learning algorithms. However, excluding less responsible non-retinal regions from OCT images has enhanced visualization capability for macular edema besides providing a standardized analysis. The learned feature activation maps can be supported with regression concept vectors [33] to enhance the explainability.

Medical image processing depends on high-capacity devices due to the complexity of the DL models and needs a high energy supply to predict the diseases. Moreover, it results in delayed responses and a longtime analysis. The proposed lightweight DL model is an energy-efficient model with a low number of classification parameters to identify DME on OCT. It provides saving battery life with low energy consumption, high accurate predictions using energy-efficient electronics, efficient deployability in mobile devices, low-cost embedded systems, and IoT devices in medicine. The main scripts, available feature activation maps on random OCT scans for visual validation, and the weights of DeepOCT will be fully available at [34] for training, testing, and easy adaptability to transfer learning.

The weakest aspect of the DeepOCT is the variety in medical devices. Each medical device has its specifications, including noise reduction technologies, lateral resolutions, view of angles, and more. Therefore, although the DeepOCT has high DME identification performances, it still needs to be tested with various OCT scans from different medical devices before settling as a diagnostic tool. Nevertheless, due to the ability of the DeepOCT for the identification of small macular edemas, it has potential clinical relevance for DME to be an alternative decision support tool for ophthalmologists.

## Data Availability Statement

The data underlying this article are available in Mendeley Data Repository, at <https://doi.org/10.17632/rscbjbr9sj.2>.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The author would like to thank Dr. Abdullah BEYOGLU for his ophthalmological assessments in the validation and evaluation of the proposal. The author also thanks to the editors and anonymous reviewers for providing insightful suggestions and comments to improve the quality of research paper.

## References

- [1] J.S. Schuman, C.A. Puliafito, J.G. Fujimoto, J.S. Duker, Optical coherence tomography of ocular diseases, Slack New Jersey (2004), <https://doi.org/10.1097/01.opx.0000431611.44976.20>.
- [2] J. Wu, Y. Zhang, J. Wang, J. Zhao, D. Ding, N. Chen, L. Wang, X. Chen, C. Jiang, X. Zou, X. Liu, H. Xiao, Y. Tian, Z. Shang, K. Wang, X. Li, G. Yang, J. Fan, AttenNet: Deep Attention Based Retinal Disease Classification in OCT Images, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, doi: 10.1007/978-3-030-37734-2\_46, 2020, pp. 565 – 576. doi: 10.1007/978-3-030-37734-2\_46.
- [3] M. Badar, M. Haris, A. Fatima, Application of deep learning for retinal image analysis: A review (2020). doi: 10.1016/j.cosrev.2019.100203..
- [4] C.A. Puliafito, M.R. Hee, C.P. Lin, E. Reichel, J.S. Schuman, J.S. Duker, J.A. Izatt, E. A. Swanson, J.G. Fujimoto, Imaging of Macular Diseases with Optical Coherence Tomography, Ophthalmology 102. doi: 10.1016/S0161-6420(95)31032-9..
- [5] K. Alsaih, M.Z. Yusoff, T.B. Tang, I. Faye, F. Mériadeau, Deep learning architectures analysis for age-related macular degeneration segmentation on optical coherence tomography scans, Computer Methods and Programs in Biomedicine 195 (2020), <https://doi.org/10.1016/j.cmpb.2020.105566> 105566.
- [6] D. Sidié, S. Sankar, G. Lemaitre, M. Rastgoo, J. Massich, C.Y. Cheung, G.S. Tan, D. Milea, E. Lamoureux, T.Y. Wong, F. Mériadeau, An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images, Computer Methods and Programs in Biomedicine 139 (2017) 109–117, <https://doi.org/10.1016/j.cmpb.2016.11.001>.
- [7] G. Lemaitre, M. Rastgoo, J. Massich, C.Y. Cheung, T.Y. Wong, E. Lamoureux, D. Milea, F. Mériadeau, D. Sidié, Classification of SD-OCT Volumes Using Local Binary Patterns: Experimental Validation for DME Detection, Journal of Ophthalmology (2016), <https://doi.org/10.1155/2016/3298606>.
- [8] C.S. Lee, D.M. Baughman, A.Y. Lee, Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images, Kidney International Reports 1 (2017) 322–327. arXiv:1612.04891, doi: 10.1016/j.kiret.2016.12.009..
- [9] M. Awais, H. Muller, T.B. Tang, F. Mériadeau, Classification of SD-OCT images using a Deep learning approach, in: Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017 IEEEExplore, Kuching, Malaysia, 2017, pp. 489–492, <https://doi.org/10.1109/ICSIPIA.2017.8120661>.
- [10] O. Perdomo, S. Otalora, F.A. Gonzalez, F. Mériadeau, H. Muller, OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes, in: Proceedings - International Symposium on Biomedical Imaging, IEEEExplore, Washington, DC, USA, 2018, pp. 1423–1426, <https://doi.org/10.1109/ISBL.2018.8363839>.
- [11] F. Li, H. Chen, Z. Liu, X. Zhang, Z. Wu, Fully automated detection of retinal disorders by image-based deep learning, Graefe's Archive for Clinical and Experimental Ophthalmology 257 (2019) 495–505, <https://doi.org/10.1007/s00417-018-04224-8>.
- [12] X. Li, L. Shen, M. Shen, C.S. Qiu, Integrating Handcrafted and Deep Features for Optical Coherence Tomography Based Retinal Disease Classification, IEEE Access 7 (2019) 33771–33777, <https://doi.org/10.1109/ACCESS.2019.2891975>.
- [13] R. Rasti, A. Mehridehnavi, H. Rabbani, F. Hajizadeh, Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier, Journal of Biomedical Optics 23 (2018) 1–10, <https://doi.org/10.1117/1.jbo.23.3.035005>.
- [14] Q. Ji, J. Huang, W. He, Y. Sun, Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images, Algorithms 12 (2019) 51, <https://doi.org/10.3390/a12030051>.
- [15] Q. Ji, W. He, J. Huang, Y. Sun, Efficient deep learning-based automated pathology identification in retinal optical coherence tomography images, Algorithms 11 (2018) 88, <https://doi.org/10.3390/a11060088>.
- [16] G.C. Chan, A. Muhammad, S.A. Shah, T.B. Tang, C.K. Lu, F. Mériadeau, Transfer learning for Diabetic Macular Edema (DME) detection on Optical Coherence Tomography (OCT) images, in: Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017 IEEEExplore, Kuching, Malaysia, 2017, pp. 493–496, <https://doi.org/10.1109/ICSIPIA.2017.8120662>.
- [17] G.C. Chan, S.A. Shah, T.B. Tang, C.K. Lu, H. Muller, F. Mériadeau, Deep Features and Data Reduction for Classification of SD-OCT Images: Application to Diabetic Macular Edema, in: International Conference on Intelligent and Advanced System, IEEE, Kuala Lumpur, Malaysia, 2018, pp. 1–4, <https://doi.org/10.1109/ICIAS.2018.8540579>.
- [18] S. Kaymak, A. Serener, 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), IEEE, Cluj-Napoca, Romania, 2018, pp. 265–269, <https://doi.org/10.1109/ICCP.2018.8516635>.
- [19] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M.K. Prasadha, J. Pei, M. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V.A. Huu, C. Wen, E.D. Zhang, C.L. Zhang, O. Li, X. Wang, M.A. Singer, X. Sun, J. Xu, A. Taftreshi, M.A. Lewis, H. Xia, K. Zhang, Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning, Cell 172 (2018) 1122–1133. doi: 10.1016/j.cell.2018.02.010..
- [20] R. Rasti, H. Rabbani, A. Mehridehnavi, F. Hajizadeh, Macular OCT Classification Using a Multi-Scale Convolutional Neural Network Ensemble, IEEE Transactions on Medical Imaging 37 (2018) 1024–1034, <https://doi.org/10.1109/TMI.2017.2780115>.
- [21] Y. Rong, D. Xiang, W. Zhu, K. Yu, F. Shi, Z. Fan, X. Chen, Surrogate-assisted retinal OCT image classification based on convolutional neural networks, IEEE

- Journal of Biomedical and Health Informatics 23 (2019) 253–263, <https://doi.org/10.1109/JBHI.2018.2795545>.
- [22] G. Lemaitre, J. Massich, M. Rastgoo, F. Meriaudeau, Seri Dataset (2019). URL: <http://i2cvb.github.io/#ret>.
- [23] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, IEEE Transactions on Image Processing 16 (2007) 2080–2095, <https://doi.org/10.1109/TIP.2007.901238>.
- [24] B. Chong, Y.K. Zhu, Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter, Optics Communications 291 (2013) 461–469, <https://doi.org/10.1016/j.optcom.2012.10.053>. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0030401812012199>.
- [25] S. Huang, C. Tang, M. Xu, Y. Qiu, Z. Lei, BM3D-based total variation algorithm for speckle removal with structure-preserving in OCT images, Applied Optics 58 (23) (2019) 6233, <https://doi.org/10.1364/AO.58.006233>. URL: <https://www.osapublishing.org/abstract.cfm?URI=ao-58-23-6233>.
- [26] P. Perona, J. Malik, Scale-Space and Edge Detection Using Anisotropic Diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence doi: 10.1109/34.56205.
- [27] B.S. Gerendas, H. Bogunovic, A. Sadeghipour, T. Schlegl, G. Langs, S.M. Waldstein, U. Schmidt-Erfurth, Computational image analysis for prognosis determination in DME, Vision Research 139 (2017) 204–210, <https://doi.org/10.1016/j.visres.2017.03.008>.
- [28] D.X. Zhou, Theory of deep convolutional neural networks: Downsampling, Neural Networks 124 (2020) 319–327, <https://doi.org/10.1016/j.neunet.2020.01.018>.
- [29] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: Proceedings of the IEEE International Conference on Computer Vision, IEEE, Venice, Italy, 2017, pp. 618–626, <https://doi.org/10.1109/ICCV.2017.74>. arXiv:1610.02391.
- [30] M. Menikdiwela, C. Nguyen, H. Li, M. Shaw, CNN-based small object detection and visualization with feature activation mapping, in: International Conference Image and Vision Computing New Zealand, IEEE, Christchurch, New Zealand, 2018, pp. 1–5, <https://doi.org/10.1109/IVCNZ.2017.8402455>.
- [31] G. Altan, Y. Kutlu, N. Allahverdi, Deep Learning on Computerized Analysis of Chronic Obstructive Pulmonary Disease, IEEE Journal of Biomedical and Health Informatics 24 (5) (2020) 1344–1350, <https://doi.org/10.1109/JBHI.2019.2931395>. URL: <https://ieeexplore.ieee.org/document/8777195>.
- [32] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach, Biometrics 44 (1988) 837–845, <https://doi.org/10.2307/2531595>.
- [33] M. Graziani, V. Andreadczyk, H. Müller, Regression Concept Vectors for Bidirectional Explanations in Histopathology, in: S.D. et Al. (Ed.), Lecture Notes in Computer Science book series (LNCS, volume 11038), Springer, Cham, 2018, pp. 124–132. doi: 10.1007/978-3-030-02628-8\_14. URL: [http://link.springer.com/10.1007/978-3-030-02628-8\\_14](http://link.springer.com/10.1007/978-3-030-02628-8_14).
- [34] DeepOCT architecture: Scripts and pre-trained weights for transfer learning (2021). URL: [https://github.com/galtan-PhD/DeepOCT-Architecture/..](https://github.com/galtan-PhD/DeepOCT-Architecture/)