

Article

Choroidal Neovascularization Screening on OCT-Angiography Choriocapillaris Images by Convolutional Neural Networks

Kawther Taibouni ¹, Alexandra Miere ² , Abdourahmane Samake ¹, Eric Souied ², Eric Petit ¹
and Yasmina Chenoune ^{1,3,*}

¹ Laboratory of Images, Signals and Intelligent Systems (LISSI, EA N° 3956), University Paris-Est Créteil, Vitry sur Seine, 94400 Paris, France; kawther.taibouni@univ-paris-est.fr (K.T.); abdou.samake0908@gmail.com (A.S.); petit@u-pec.fr (E.P.)

² Centre Hospitalier Intercommunal de Créteil, Department of Ophthalmology, 40, Avenue de Verdun, Créteil, 94010 Paris, France; alexandra.miere@chicreteil.fr (A.M.); Eric.Souied@chicreteil.fr (E.S.)

³ ESME Sudria Research Lab., 34 Rue de Fleurus, 75006 Paris, France

* Correspondence: yasmina.chenoune@esme.fr

Abstract: Choroidal Neovascularization (CNV) is the advanced stage of Age-related Macular Degeneration (AMD), which is the leading cause of irreversible visual loss for elder people in developed countries. Optical Coherence Tomography Angiography (OCTA) is a recent non-invasive imaging technique widely used nowadays in diagnosis and follow-up of CNV. In this study, an automatic screening of CNV based on deep learning is performed using OCTA choriocapillaris images. CNV eyes (advanced wet AMD) are diagnosed among healthy eyes (no AMD) and eyes with drusen (intermediate AMD). An OCTA dataset of 1396 images is used to train and evaluate the model. A pre-trained convolutional neural network (CNN) is fine-tuned and validated on 80% of the dataset while the remaining 20% is used independently for predictions. The model can accurately detect CNV on the test set with an accuracy of 89.74%, precision of 0.96 and 0.99 area under the curve of the receiver operating characteristic. A good overall classification accuracy of 88.46% is obtained on a balanced test set. Detailed analysis of misclassified images shows that they are also considered ambiguous images for expert clinicians. This novel CNN-based application is truly a breakthrough to assist clinicians in the challenging task of screening for neovascular complications.

Keywords: age-related macular degeneration; choroidal neovascularization; convolutional neural networks; image classification; optical coherence tomography angiography



Citation: Taibouni, K.; Miere, A.; Samake, A.; Souied, E.; Petit, E.; Chenoune, Y. Choroidal Neovascularization Screening on OCT-Angiography Choriocapillaris Images by Convolutional Neural Networks. *Appl. Sci.* **2021**, *11*, 9313. <https://doi.org/10.3390/app11199313>

Academic Editors: Leonardo Rundo, Carmelo Militello and Andrea Tangherloni

Received: 15 September 2021

Accepted: 2 October 2021

Published: 8 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Age-related Macular Degeneration (AMD) is the leading cause of irreversible blindness in the elderly population of developed countries. AMD is characterized by changes in the Retinal Pigment Epithelium (RPE), Bruch's Membrane (BM), or Choriocapillaris (CC) complex [1]. There are several staging systems for AMD, but the most widely used is the AREDS (Age Related Eye Disease Study) classification, distinguishing between early, intermediate and late AMD (see Figure 1) [2]. Early and intermediate AMD are characterized by the presence of drusen and pigmentary changes. Late AMD consists of wet AMD, characterized by choroidal neovascularization (CNV), and dry AMD, characterized by geographic atrophy (GA) in the macular area [3,4]. While both dry and wet AMD are visually threatening, in the particular case of wet AMD, CNV progression can result in rapidly deteriorating visual acuity, leading to scarring and irreversible visual loss [4,5]. Moreover, as hallmarks of early and intermediate AMD, drusen precede the progression to late AMD [6,7]. Hence, distinguishing between the early, intermediate, and late AMD plays a key role in both follow-up and treatment decisions, in order to preserve the visual prognosis.

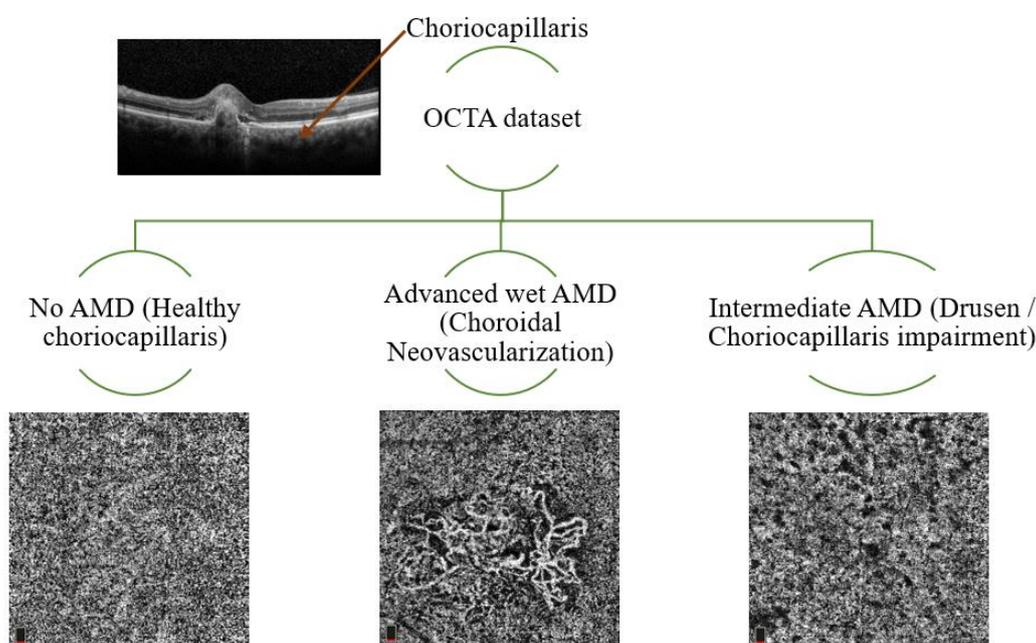


Figure 1. Classification of three types of OCTA choriocapillaris images (no AMD, advanced wet AMD, intermediate AMD). On the top-left: the OCTA cross-sectional B-scan showing retinal layers.

Novel imaging techniques, such as optical coherence tomography angiography (OCTA), contribute to the accurate diagnosis of early, intermediate and late AMD in a depth-resolved and non-invasive manner [8–10]. Besides the accurate detection of CNV or GA in the CC segmentation [11], OCTA has demonstrated CC alterations, i.e., flow deficits (FD) and signal attenuation associated with drusen, in early and intermediate AMD [12]. Moreover, recent literature has shown that choriocapillaris FD predict drusen enlargement, therefore being a significant risk factor for late forms of AMD [7]. Therefore, earlier diagnosis could ensure better follow-up of patients at high risk of conversion to advanced wet AMD. A computer-aided diagnosis (CAD) tool in ophthalmic clinical routine could be of significant assistance for clinicians at daily diagnosis and during follow-up.

The detection of CNV in the context of wet AMD on OCTA images is very challenging due to OCTA various artifacts [8,9]. Many recent papers dealt with the contributions of artificial intelligence (AI) and deep learning (DL) approaches in ophthalmology [13–15]. However, only few works have addressed CNV screening on OCTA images using DL methods [16]. Most of the existing works focused on the diagnosis of AMD using OCT or Color Fundus Photography. Moreover, published works on OCTA images involved other retinal diseases such as Diabetic Retinopathy (DR) [17].

In 2017, Rasti et al. [18] proposed a CAD system based on a multi-scale convolutional mixture of expert model to identify accurately dry AMD and diabetic macular edema using OCT images. Two different macular OCT datasets of 4142 and 3247 B-scans were used for the training step. A very good classification score of 99.85% was derived from the receiver operating characteristic curve (ROC-AUC).

In 2018, Burlina et al. [19] used a set of 67,401 color fundus images of AMD patients to estimate 5-year risk of progression to advanced stages of AMD by DL techniques. Two AMD severity scales (4-step and 9-step) were considered, and a human versus machine comparison was carried out. This study achieved weighted k scores of 0.77 for the 4-step and 0.74 for the 9-step AMD severity scales. The same year, Grassmann et al. [20] exploited a database of 120,656 color fundus images, manually graded in 13 AMD severity levels, to train several CNN architectures (AlexNet, GoogLeNet, VGG, Inception-V3, ResNet and I-ResNet-V2). A very good-weighted k of 92% was obtained. Govindaiah et al. [19] have also shown that deep CNN could be efficient to grade color fundus images in four classes: no AMD, early AMD, intermediate AMD, and advanced AMD. The study included

a comparison between training using the transfer learning approach and training from scratch. The obtained accuracies were 78.1% for transfer learning and 83% without transfer learning.

More recently, Russakoff et al. [21] developed a DL architecture so called “AMDnet” on OCT images to predict the conversion from early/intermediate AMD to advanced wet AMD. The study included 71 patients with confirmed early/intermediate AMD that were imaged with OCT three times over 2 years. Results showed a ROC-AUC of 0.89 at the B-scan levels and 0.91 for volumes. Hwang et al. [22] used 35,900 labeled OCT images from AMD patients to train three types of Convolutional Neural Networks (CNNs), VGG19, InceptionV3 and ResNet50, to perform AMD diagnosis. The authors developed an AI and cloud-based telemedicine interaction tool dedicated to diagnosis and therapeutic of AMD. The image discrimination rates obtained by expert clinicians (92.73% and 91.90%) and provided by the AI-based platform (above 90%) were almost the same.

Further works on OCT imaging, such as the recent study of Romo-Bucheli et al. [23], proposed a treatment predictive model using a densely connected neural network (DenseNet) and a recurrent neural network (RNN) on longitudinal OCT scans for neovascular AMD patients (281 patients for training and 69 for tests). The CNN model achieved 0.85 AUC in detecting patients with low treatment requirements and 0.81 AUC for patients with high treatment requirements.

In what concerns the use of OCTA in DL, Le et al. [17] tested the feasibility of using DL for DR detection from OCTA including 77 patients and 20 control subjects. The authors applied transfer learning on a VGG16 network for robust OCTA classification. The obtained results showed an accuracy of 87.27% in differentiating healthy, no DR and DR eyes. In the same period, Wang et al. [16] developed an algorithm based on two CNNs to classify input OCTA images (using structural volumes and enface retinal angiograms) as CNV or Non-CNV and then segment the CNV membrane when present. The proposed neural network included a cutoff threshold for CNV area to overcome the residual artifacts limitation that could be confounded with CNV. CNV binary classification ROC-AUC was 0.997.

In this work, we aim to fill the gap of CNV screening on OCTA images using DL by promoting a novel application of CNNs on OCTA images using the choriocapillaris slab. The main contribution of this paper is the deep learning-based solution to classify AMD on two major forms: advanced wet AMD (CNV) and intermediate AMD (drusen/pigmentary changes) including a healthy control group (no AMD) using choriocapillaris OCTA images. A second contribution is the adaptation of a pre-trained VGG19 model on non-medical ImageNet dataset to medical domain using an adapted densely connected classifier on our limited OCTA data. Additionally, class activation mapping is used to interpret the CNN prediction on choriocapillaris OCTA images, which is a promising DL application for CAD systems in retinal clinical routine.

2. Materials and Methods

2.1. Dataset and Study Population

Data from patients with AMD is collected from the Ophthalmology Department of Intercommunal Hospital Center of Créteil, France, between September 2014 and July 2019. A database of 1396 choriocapillaris OCTA images of size 304×304 with a pixel size of $9.87 \times 9.87 \mu\text{m}$ is built from 787 eyes related to 508 patients (mean age 70.67 ± 17.74 years). All patients underwent a 3×3 mm OCTA examination (AngioVue, Optovue, Fremont, CA, USA). The choriocapillaris slab is extracted and there are no excluded images due to motion or projection artifacts.

A retina specialist (A.M.) classified the OCTA images into three classes (391 with no AMD images from healthy eyes of 156 subjects, 457 images with CNV from 187 AMD patients, and 548 images with intermediate AMD from 274 patients). Multiple images per patient are included in this database. On one hand, follow-up images acquired at different dates are considered in this study as they show notable and significant changes in the CNV progression or in the number and size of drusen. On the other hand, both eyes are

considered for some of the patients when both eyes' examinations are available. This study is performed in accordance with the Declaration of Helsinki and current French legislation and with approval of our local ethics committee.

2.2. CNV Screening on OCTA Images

Our goal is to discriminate from OCTA images the three predefined classes: No AMD (healthy CC), advanced wet AMD (CNV) and intermediate AMD (drusen or CC impairment) (see Figure 1). OCTA imaging allows physicians to visualize blood vessels in the individual layers of the retina and choroid without dye injection. Thus, CNV, drusen and impairment within the CC (pigmentary changes) can be clearly identified on OCTA images. The healthy choriocapillaris appears on OCTA images as a grainy texture with bright and dark spots corresponding to blood flow and flow deficits, respectively [7]. Drusen and CC impairment are characterized by black nonflow areas of different sizes related to flow deficits that can appear anywhere on the OCTA image surrounded by the grainy texture of the choriocapillaris. Regarding CNV, neovascular membranes harbor the aspect of a vascular branching, surrounded by the grainy texture of the choriocapillaris.

Nevertheless, these images are corrupted by speckle noise due to the physical principles of OCT, in addition to the image acquisition process and artifacts [24]. Moreover, included CNV lesions could have different sizes and locations with irregular shapes of neovascular membranes, thus small ones may be confused with the grainy texture of the choriocapillaris. This makes CNV detection on OCTA images a very challenging task. Figure 1 illustrates the OCTA choriocapillaris images of the three classes used for the classification in this work.

2.3. CNN Architecture and Transfer Learning

As depicted on Figure 2, our methodological approach consists of two parts: the VGG19 deep network [25] that provides the features extraction process on the OCTA image and a personalized densely connected network that represents the classification part.

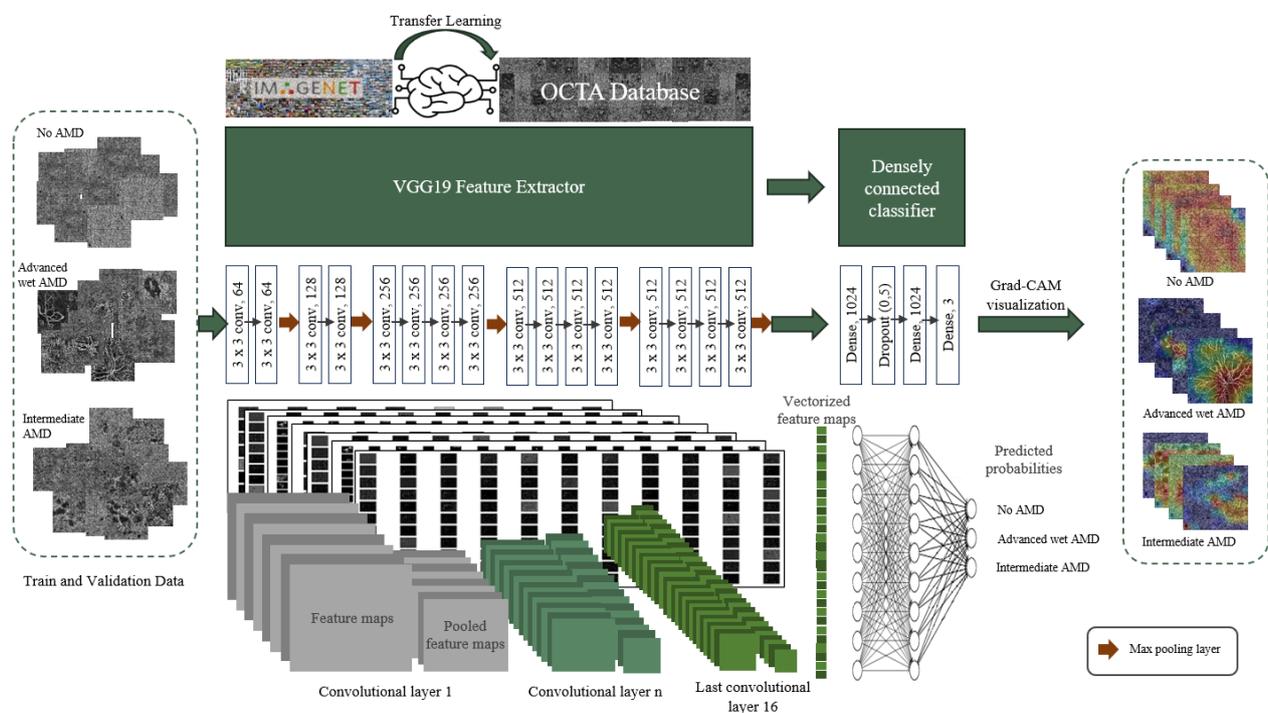


Figure 2. Our proposed modified VGG19 architecture. OCTA images are mapped through the VGG19 feature extractor to build discriminant 9×9 feature maps that are used by the densely connected classifier to generate predicted probabilities and Grad-CAM heatmaps to visualize feature attribution for each class: No AMD (healthy choriocapillaris), advanced wet AMD (CNV), intermediate AMD (drusen/choriocapillaris impairment).

VGG19 is a competition-winning model of the *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) [26], that has a sequential pipeline architecture consisting of 16 convolutional and 3 fully connected trainable layers including five max-pooling layers. In this work, only the VGG19 convolutional part is used as a feature extractor where convolution layers include 3×3 convolution filters. A rectified linear unit (ReLU) activation function and a downsampling 2×2 max-pooling operation follow each convolutional stack. This convolutional part provides vectorized feature maps used as input to the densely connected classifier.

The densely connected classifier contains three fully connected layers, layers 1 and 2 are composed of 1024 nodes each. The last one consists of three nodes that provide the classification result into three types of OCTA images. A regularization dropout layer is included after the first dense layer to overcome overfitting the model by randomly dropping out 50% of the activations at that layer. ReLU activation function is used on the two first dense layers, whereas a softmax activation function is used on the last one.

According to the huge number of learnable VGG19 parameters (144 million) and the limited amount of OCTA data in our training dataset, transfer learning from non-medical data is applied in our approach [27]. Therefore, learned knowledge from the ImageNet dataset [26,28] is transferred to the model and adapted to our application by fine-tuning the convolutional part using OCTA images. The densely connected layers are trained from scratch on our OCTA data to classify OCTA images. Finally, feature maps from the last convolutional layer are mapped through the densely connected classifier to generate Gradient-weighted Class Activation Mapping (Grad-CAM) visualization [29]. The Grad-CAM produces a localization map that highlights the image's important features to the CNN for class predictions.

Additionally, to assess the impact of our approach (transfer learning on modified VGG19 model) on CNV detection accuracy, we trained the original VGG19 model independently from scratch with random initialization on our OCTA data.

2.4. Implementation Details

The dataset is divided into two independent subsets for training and testing. From the whole dataset, 80% (1115 images) is dedicated to fine-tuning, training, and validation. Subsequently, this first partition is further separated into 80% (892 images) for train and 20% (223 images) for validation. The remaining 20% (281 images) of the whole dataset is used for the performance evaluation and tests.

The whole network is trained end-to-end on 100 epochs for transfer learning and 200 epochs for training original VGG19. Stochastic Gradient Descent (SGD) optimization algorithm [30] and categorical cross entropy loss function are used. The learning-rate is set to 10^{-5} for transfer learning and to 10^{-4} for training original VGG19. Data augmentation is applied during training to reduce overfitting. Only random zoom is used in transfer learning to generate 16 OCTA images at each batch while rotation, horizontal and vertical flip are used in addition in training from scratch the original VGG19 model to generate 8 OCTA images at each batch.

The pipeline is implemented in Python with the Keras-TensorFlow library [31,32]. Training and testing are performed on a NVIDIA Corporation GP104GL [Quadro P4000] Graphics Processing Unit.

2.5. Performance Evaluation

Performance is evaluated using an independent balanced test set (78 images for each class) and there are no excluded images due to motion and projection artifacts or to image quality. The CNN prediction output is compared to the ground truth set by the expert reader (A.M.). Four statistical metrics are used to report classification performance [33]:

First, the accuracy that generally describes how the model performs across all classes. It is obtained as the ratio between the number of correct predictions to the total number of predictions:

$$\text{Accuracy} = \frac{\text{True}_{\text{positive}} + \text{True}_{\text{negative}}}{\text{True}_{\text{positive}} + \text{True}_{\text{negative}} + \text{False}_{\text{positive}} + \text{False}_{\text{negative}}}$$

Then the precision, that is calculated as the ratio between the number of positive samples correctly classified to the total number of samples classified as positive. The precision measures the model's accuracy in classifying a sample as positive:

$$\text{Precision} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{positive}}}$$

The recall is calculated as the ratio between the number of positive samples correctly classified as Positive to the total number of positive samples. The recall measures the model's ability to detect positive samples. The higher the recall, the more positive samples detected:

$$\text{Recall} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{negative}}}$$

Finally, the F1-score is a way of combining the precision and recall:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Confusion matrix and area under the curve (AUC) of the receiver operating characteristic (ROC) and precision-recall (PRC) curves (ROC-AUC and PRC-AUC) are supplied. Additionally, multiple class activation maps are generated to analyze the feature attribution and understand the CNN predictions.

3. Results

CNV screening evaluation, reported in Table 1 and Figure 3, shows that CNV detection on OCTA images achieves the best performance with a precision of 96%, recall of 90%, F1-score of 0.93 and an accuracy of 89.74%. In addition, ROC-AUC and PRC-AUC are 0.99 each. No AMD (healthy CC) OCTA images are also well classified by the proposed VGG19 modified model with a very good accuracy of 94.87% and F1-score of 0.90 (precision 0.85, recall 0.95). Regarding intermediate AMD (drusen/CC impairment) class, the images are, in some cases, confused with no AMD class. Sixty-three images from the intermediate AMD test dataset are correctly classified while 12 images are predicted as no AMD images and 3 as CNV images. This is summarized by the confusion matrix in Table 2.

Table 1. CNV screening performance of the modified VGG19. Accuracy, precision, recall, F1-score, area under the curve (AUC) for precision-recall (PRC) and receiver operation characteristic (ROC) curves for the three classes of OCTA images (HCC: healthy choriocapillaris, DCCI: drusen/choriocapillaris impairment, CNV: choroidal neovascularization).

	Accuracy (%)	Precision	Recall	F1-Score	PRC-AUC	ROC-AUC
HCC	94.87	0.85	0.95	0.90	0.97	0.99
DCCI	80.77	0.85	0.81	0.83	0.94	0.97
CNV	89.74	0.96	0.90	0.93	0.99	0.99

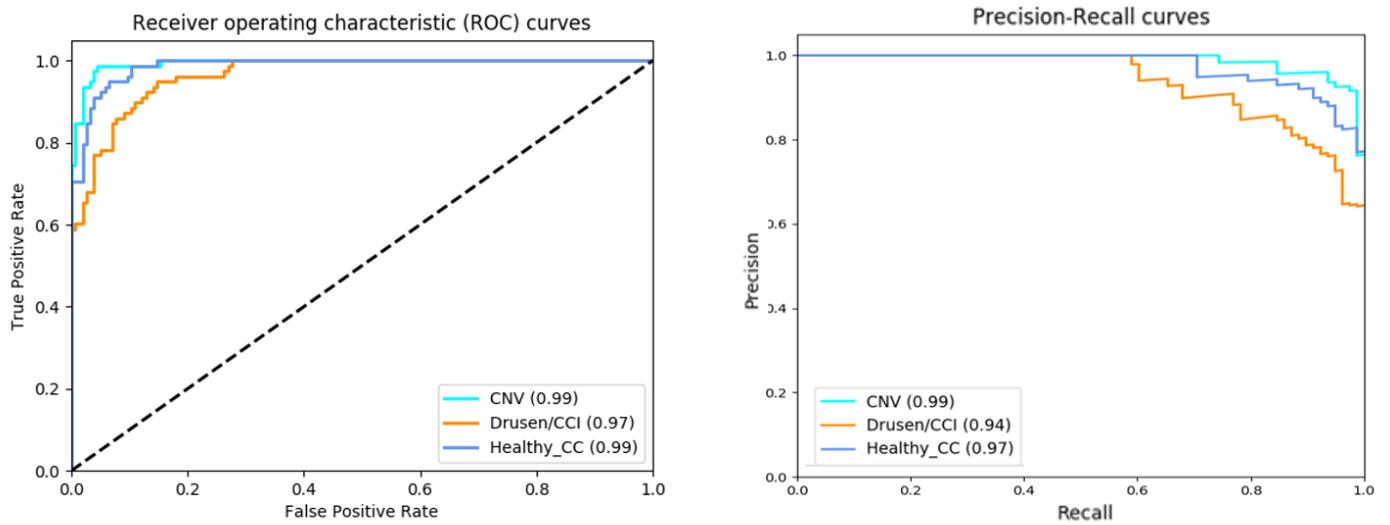


Figure 3. Receiver operating characteristic (ROC) and precision-recall (PRC) curves for the three classes: advanced wet AMD (CNV), intermediate AMD (drusen/CCI: choriocapillaris impairment), no AMD (healthy choriocapillaris). Area under the curve (AUC) values for each class.

Table 2. Confusion matrix of the modified VGG19 prediction on a balanced test dataset of 234 OCTA images (78 images for each class). CNV: choroidal neovascularization, DCCI: drusen/choriocapillaris impairment, HCC: healthy choriocapillaris.

Ground Truth Class	Predicted Class		
	CNV	DCCI	HCC
CNV	70	7	1
DCCI	3	63	12
HCC	0	4	74

The overall classification accuracy using transfer learning on our modified VGG19 is 88.46% with a loss of 0.089, while the overall classification accuracy using original VGG19 is 83.76% with a loss of 0.37. The proposed approach on the modified VGG19 achieved better performance than original VGG19 for CNV detection. This is reported in classification reports (Tables 1 and 3) and confusion matrices (Tables 2 and 4) where CNV screening accuracy is 74.36% for original VGG19 against 89.74% for our proposed approach. In addition, PRC-AUC and ROC-AUC are 0.99 each using our modified VGG19, while PRC-AUC is 0.95 and ROC-AUC is 0.97 using original VGG19.

Table 3. CNV screening performance of the original VGG19. Accuracy, precision, recall, F1-score, area under the curve (AUC) for precision-recall (PRC) and receiver operation characteristic (ROC) curves for the three classes of OCTA images (HCC: healthy choriocapillaris, DCCI: drusen/choriocapillaris impairment, CNV: choroidal neovascularization).

	Accuracy (%)	Precision	Recall	F1-Score	PRC-AUC	ROC-AUC
HCC	97.44	0.84	0.97	0.90	0.96	0.99
DCCI	79.49	0.75	0.79	0.77	0.83	0.93
CNV	74.36	0.97	0.74	0.84	0.95	0.97

Table 4. Confusion matrix of the original VGG19 prediction on a balanced test dataset of 234 OCTA images (78 images for each class). CNV: choroidal neovascularization, DCCI: drusen/choriocapillaris impairment, HCC: healthy choriocapillaris.

Ground Truth Class	Predicted Class		
	CNV	DCCI	HCC
CNV	58	19	1
DCCI	2	62	14
HCC	0	2	76

Figures 4 and 5 display Grad-CAM visualizations for correct predictions and incorrect predictions of the three classes, respectively, (no AMD, intermediate AMD, and advanced wet AMD). Grad-CAM heatmaps are superimposed on original OCTA choriocapillaris images with warm colors (red, orange, and yellow) for discriminant features and cold colors (blue, cyan, and green) for non-discriminant features.

In Figure 4, expected discriminant features are correctly highlighted by Grad-CAM heatmaps for each class: grainy texture throughout the whole OCTA image for no AMD images (Figure 4A,B), flow deficits/nonflow areas for intermediate AMD images (Figure 4C,D), and high flow vascular networks (CNV) for advanced wet AMD images (Figure 4E,F,G,H). This is further supported by the CNN predicted probabilities for each image. Regarding no AMD images (Figure 4A,B), the CNN predicted high probabilities were 0.94 and 0.97, respectively, and were attributed to no AMD class. Predicted probabilities for intermediate AMD images (Figure 4C,D) were 0.70 and 0.99, respectively, and were attributed to the correct class. Finally, regarding advanced wet AMD images (Figure 4E,F,G,H) predicted probabilities were 0.99 for images Figure 4E,G,H and 0.63 for image Figure 4F, correctly attributed to CNV.

On the other hand, non-discriminant CNV features are highlighted by Grad-CAM heatmaps in Figure 5I,J, including the flow deficits/nonflow areas or regions in Figure 5I and grainy texture in Figure 5J.

The CNN predicted probabilities reinforce this observation where image Figure 5I is predicted as intermediate AMD with 0.84 probability and image Figure 5J as no AMD image with 0.90 probability. Regarding the Figure 5K, non-discriminant drusen features are highlighted by Grad-CAM heatmap, grainy texture is highlighted as discriminant features showing the CNN prediction as no AMD with 0.59 of probability against 0.41 for intermediate AMD. Conversely, Figure 5L represents a healthy CC (no AMD) image predicted as intermediate AMD with 0.58 of probability against 0.42 for no AMD. The Grad-CAM heatmap supports these probabilities by highlighting only drusen features as discriminant rather than those from grainy texture of no AMD images.

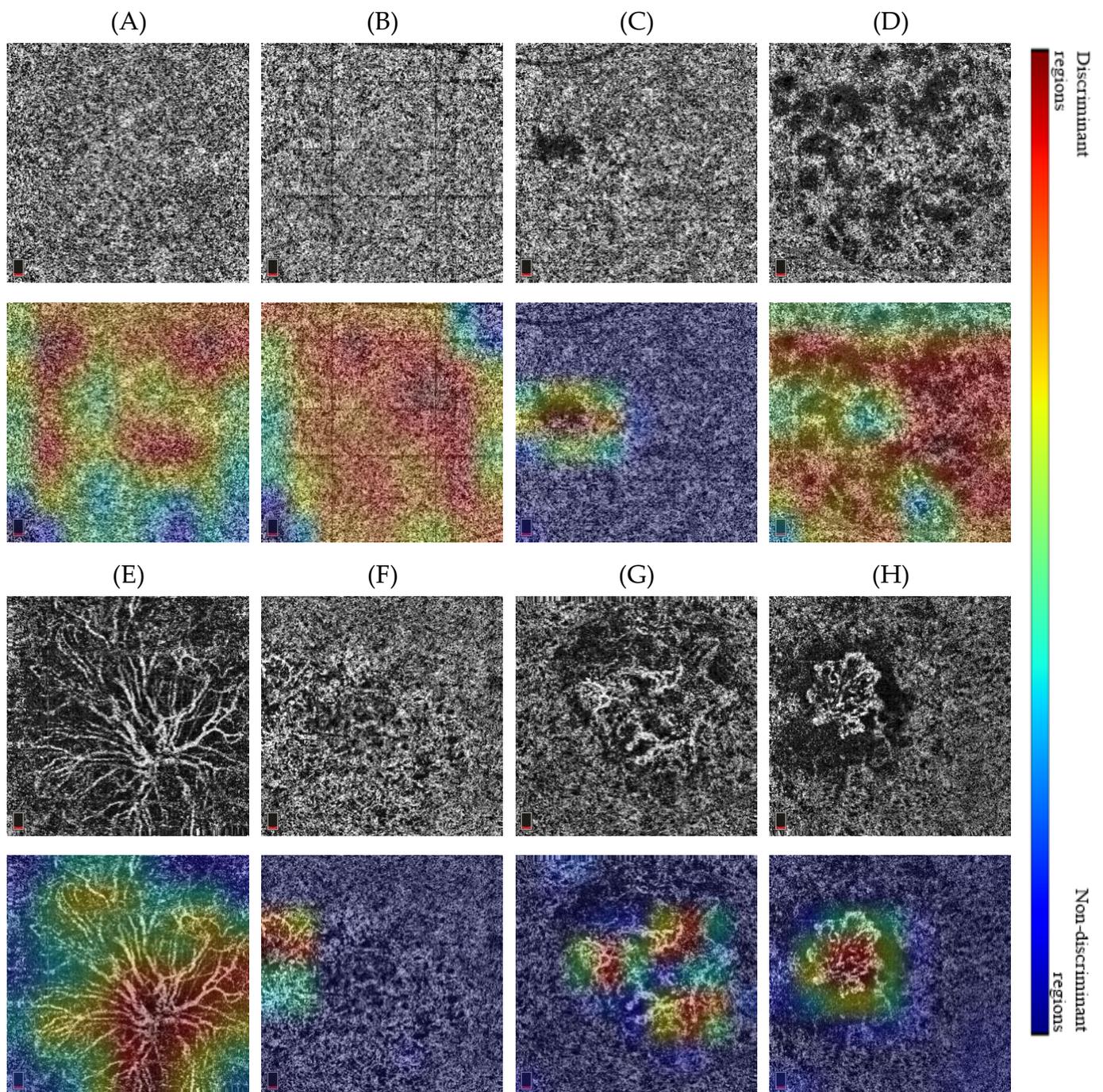


Figure 4. Grad-CAM visualizations for correct predictions by the proposed VGG19 modified model. Below each OCTA image, the corresponding Grad-CAM visualization of the CNN prediction. (A,B) Healthy choriocapillaris—no AMD. (C,D) Drusen/Choriocapillaris impairment—intermediate AMD. (E–H) Choroidal Neovascularization—advanced wet AMD.

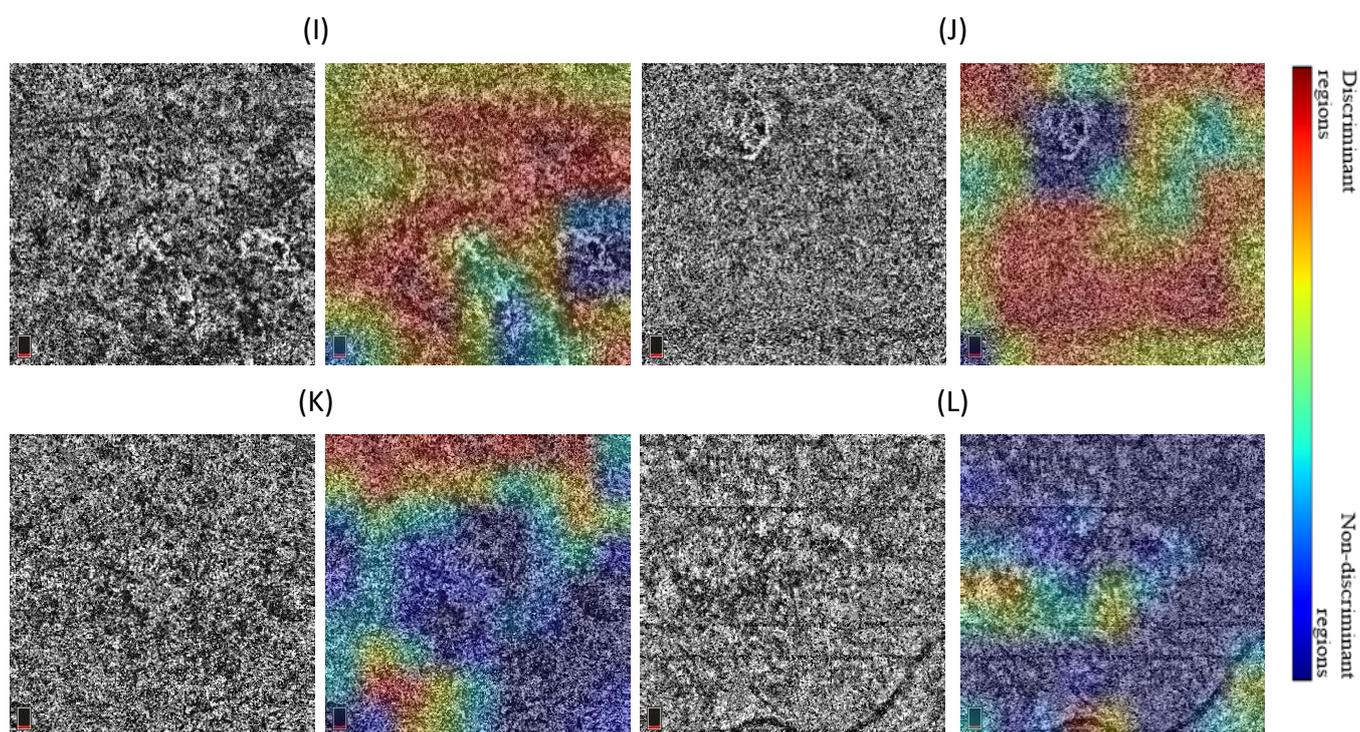


Figure 5. Grad-CAM visualizations for incorrect predictions by the proposed VGG19 modified model. To the right of each OCTA image, the corresponding Grad-CAM visualization of the CNN prediction. (I,J) advanced wet AMD (CNV) images, (K) intermediate AMD image and (L) No AMD (healthy CC) image.

4. Discussion and Conclusions

CNV screening on OCT-angiography choriocapillaris images is challenging due to the high variance in neovascular membranes sizes, shapes, and locations. Additionally, the speckle noise and the image acquisition process makes the detection task even more difficult, particularly to differentiate CNV or early drusen from the grainy texture of the choriocapillaris. Our proposed VGG19 modified model achieved very good CNV detection performance with an accuracy of 89.74% and F1-score of 0.93 (as reported in Table 1 [0.96 of precision, 0.90 of recall]). The ROC-AUC and PRC-AUC were 0.99 each (Table 1 and Figure 3). The confusion matrix in Table 2 shows that the CNN could accurately detect the CNV lesion on 70 images out of 78 independent test images. Seven CNV images are confused with intermediate AMD images, only one CNV image is misclassified as healthy CC, and 3 images of intermediate AMD are misclassified as CNV.

Concerning the no AMD (healthy CC) classification performance by our model, the statistical analysis reported in Table 1 and Figure 3 shows that it is the second-best classified class after CNV. Precision and recall are 0.85 and 0.95, respectively, with a ROC-AUC of 0.99, a PRC-AUC of 0.97 and an accuracy of 94.87%. The confusion matrix in Table 2 indicates that 74 no AMD images are correctly classified, only 4 images are confused with intermediate AMD images and no healthy CC image is misclassified as CNV.

Finally, for intermediate AMD detection, statistics show a F1-score of 0.83 (precision of 0.85 and recall of 0.81), an accuracy of 80.77%, a ROC-AUC of 0.97 and a PRC-AUC of 0.94 (reported in Table 1 and Figure 3). These results are better explained by the confusion matrix attributions in Table 2, where 63 of the 78 intermediate AMD images are correctly classified, while 12 images are predicted as no AMD and 3 images confused with CNV.

Furthermore, the CNN predicted probabilities demonstrate, on one hand, the CNN certainty when predicting correct classes (probability higher than 0.90 for the images in Figure 4A,B,D,E,G,H), and on the other hand, the CNN uncertainty for more difficult cases, such as images C and F of Figure 4. Image C illustrates a flow impairment clustered only on one region of the image and surrounded by the grainy texture of the CC on the whole

image. Therefore, the CNN predicted probabilities for image C are 0.70 as intermediate AMD and 0.29 as no AMD. Regarding image F, the CNV membrane is covered by different drusen surrounded by the grainy texture of the CC. Consequently, the CNN predicted probabilities are 0.63 for CNV and 0.36 for intermediate AMD. This shows the CNN's high ability to discriminate wet AMD eyes from healthy and intermediate AMD eyes and to show uncertainty for ambiguous and confusing cases.

The Grad-CAM visualization (see Figure 4) provides a better understanding of these results. No matter the shape, size, and location of the CNV on images of Figure 4E–H; the CNN's predicted high probability is attributed to the correct region with high discriminative CNV features. Conversely, the CNN's low probability is attributed to the non-discriminative regions. Similarly, regardless of the presence of artifacts on images of Figure 4B,C; the CNN feature attribution is correctly highlighted by Grad-CAM heatmaps. This proves that the CNN prediction is based on relevant regions of the OCTA choriocapillaris images related to the three classes: no AMD, intermediate AMD, and advanced wet AMD, that are also considered by the expert reader to detect CNV on OCTA images.

The most difficult step of this classification problem is, on one hand, the detection of some confusing CNV where a small, indefinite vascular shape is visible and for which very small amount of OCTA images are available in our training dataset. On the other hand, the discrimination of early small drusen from healthy choriocapillaris remains problematic. These difficulties are addressed in Figure 5 that illustrates two misclassified CNV images (Figure 5I,J), a misclassified intermediate AMD image (Figure 5K) and a misclassified no AMD image (Figure 5L).

Image I is misclassified as intermediate AMD with 0.84 of probability and image J is the only CNV image of our test dataset misclassified as no AMD with 0.90 of probability. Feature attribution visualization for these two misclassified images illustrated by Grad-CAM heatmaps in Figure 5 helps understanding the CNN prediction for these cases. The Grad-CAM heatmap of image Figure 5I shows that discriminant features are those of drusen by only black nonflow regions which explains the CNN predicted high probability attributed to intermediate AMD class rather than CNV class. Only one small CNV is visible in Figure 5I and is considered as non-discriminant feature. The CNN prediction is thus based on the most present features on the image. Such images are also ambiguous for clinicians and additional imaging modalities are needed to establish a clear diagnosis.

The Grad-CAM heatmap of image Figure 5J explains again the CNN misclassification. Only grainy regions present throughout the whole image are considered as discriminant features by the CNN. The tiny CNV membrane visible on this image is considered as non-discriminant as it is hidden in the grainy texture. This explains the CNN's prediction with high probability (0.90) for image Figure 5J, attributed to no AMD class.

To overcome these classification errors, we should supply our training dataset with more ambiguous OCTA choriocapillaris images such as images I and J of Figure 5 to train the model to detect tiny CNV membranes when they are hidden and confused with drusen or CC grainy texture.

Drusen or significant flow impairment on OCTA choriocapillaris images appear as nonflow areas/flow deficits that are generally surrounded by the grainy texture of CC. When these areas are small and less important than the CC texture, the OCTA image is considered as ambiguous and confusing even by clinicians. This is the case of the 12 OCTA images misclassified by the CNN and predicted as no AMD images. Early small drusen are hardly visible on OCTA images and can be considered as flow deficits related to OCTA image acquisition process. Thus, the early small drusen manually classified as intermediate AMD images are, in some cases, predicted as no AMD by the CNN. These cases are illustrated by typical images in Figure 5K,L.

The above analysis is supported by the predicted probabilities for images K and L of Figure 5. Figure 5K represents an intermediate AMD image predicted as no AMD by the CNN with close probabilities for both classes (0.59 for no AMD against 0.41 for intermediate AMD). Similarly, Figure 5L is a no AMD image predicted as intermediate

AMD by the CNN with close probabilities (0.58 for intermediate AMD against 0.42 for no AMD). This last result reveals again the CNN's ability to show uncertainty in cases of confusing features. To resolve this ambiguity, it would be relevant to classify these images as uncertain images for which the clinician should use additional information from other imaging modalities or patient history to decide and make a diagnosis.

Despite these few misclassification errors, the CNN showed a great ability to screen and detect CNV on OCTA choriocapillaris images. This main finding is achieved through the transfer learning approach that is used to train the proposed VGG19 modified model to overcome the limitation of the small amount of training data. Fine-tuning the proposed modified VGG19 improved the overall classification accuracy compared to that obtained from training the original VGG19 from random initialization. The overall accuracy increased from 83.76% using original VGG19 to 88.46% using transfer learning on the modified VGG19 although it was applied from non-medical data.

This study is one of the few works dealing with CNV screening on OCTA data using only images at the choriocapillaris slab. Obtained results revealed a promising application of CAD systems to diagnose CNV on OCTA choriocapillaris images in clinical routine using DL-based methods. In order to produce more reliable results to clinicians and to help them quantify the CNN uncertainty, we aim to measure the CNN prediction uncertainty in further works to identify how much a CNN could be trusted in diagnosis [34,35] and to avoid using images not suitable for diagnosis when high uncertainty is detected [35].

Further studies on this topic will focus on data augmentation, as well as including more CNV images to work on larger datasets and different imaging modalities to improve classification performance for ambiguous cases.

Author Contributions: Conceptualization, K.T., A.M., E.S., E.P. and Y.C.; methodology, K.T.; validation, E.S., E.P. and Y.C.; investigation, K.T., A.M. and A.S.; resources, A.M., A.S. and E.S.; writing—original draft preparation, K.T.; writing—review and editing, K.T., A.M., E.S., E.P. and Y.C.; supervision, E.S., E.P. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and current French legislation and with approval of our local ethics committee.

Informed Consent Statement: Patient consent was waived due to the retrospective nature of the study.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author.

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

References

1. Biesemeier, A.; Taubitz, T.; Julien, S.; Yoeruek, E.; Schraermeyer, U. Choriocapillaris breakdown precedes retinal degeneration in age-related macular degeneration. *Neurobiol. Aging* **2014**, *35*, 2562–2573. [[CrossRef](#)] [[PubMed](#)]
2. Age-Related Eye Disease Study Research Group A Randomized, Placebo-Controlled, Clinical Trial of High-Dose Supplementation With Vitamins C and E, Beta Carotene, and Zinc for Age-Related Macular Degeneration and Vision Loss. *Arch. Ophthalmol.* **2001**, *119*, 1417–1436. [[CrossRef](#)]
3. Colijn, J.M.; Buitendijk, G.H.S.; Prokofyeva, E.; Alves, D.; Cachulo, M.L.; Khawaja, A.P.; Cougnard-Gregoire, A.; Merle, B.M.J.; Korb, C.; Erke, M.G.; et al. Prevalence of age-related macular degeneration in Europe: The past and the future. *Ophthalmology* **2017**, *124*, 1753–1763. [[CrossRef](#)] [[PubMed](#)]
4. Abramoff, M.D.; Garvin, M.K.; Sonka, M. Retinal Imaging and Image Analysis. *IEEE Rev. Biomed. Eng.* **2010**, *3*, 169–208. [[CrossRef](#)] [[PubMed](#)]
5. De Jong, P.T. Age-related macular degeneration. *N. Engl. J. Med.* **2006**, *355*, 1474–1485. [[CrossRef](#)]
6. Lipez, A.; Miller, L.; Kovacs, I.; Czako, C.; Csipo, T.; Baffi, J.; Csiszar, A.; Tarantini, S.; Ungvari, Z.; Yabluchanskiy, A.; et al. Microvascular contributions to age-related macular degeneration (AMD): From mechanisms of choriocapillaris aging to novel interventions. *GeroScience* **2019**, *41*, 813–845. [[CrossRef](#)] [[PubMed](#)]

7. Nassisi, M.; Tepelus, T.; Nittala, M.G.; Sadda, S.R. Choriocapillaris flow impairment predicts the development and enlargement of drusen. *Graefes Arch. Clin. Exp. Ophthalmol.* **2019**, *257*, 2079–2085. [[CrossRef](#)]
8. de Carlo, T.E.; Romano, A.; Waheed, N.K.; Duker, J.S. A review of optical coherence tomography angiography (OCTA). *Int. J. Retin. Vitro.* **2015**, *1*, 1–15. [[CrossRef](#)]
9. Spaide, R.F.; Fujimoto, J.G.; Waheed, N.K. Image artifacts in optical coherence angiography. *Retina* **2015**, *35*, 2163. [[CrossRef](#)]
10. Ma, J.; Desai, R.; Nesper, P.; Gill, M.; Fawzi, A.; Skondra, D. Optical Coherence Tomographic Angiography Imaging in Age-Related Macular Degeneration. *Ophthalmol. Eye Dis.* **2017**, *9*, 1179172116686075. [[CrossRef](#)]
11. Miere, A.; Butori, P.; Cohen, S.Y.; Semoun, O.; Capuano, V.; Jung, C.; Souied, E.H. Vascular Remodeling of Choroidal Neovascularization after Anti-Vascular Endothelial Growth Factor Therapy Visualized on Optical Coherence Tomography Angiography. *Retina* **2019**, *39*, 548–557. [[CrossRef](#)] [[PubMed](#)]
12. Alten, F.; Lauermann, J.L.; Clemens, C.R.; Heiduschka, P.; Eter, N. Signal reduction in choriocapillaris and segmentation errors in spectral domain OCT angiography caused by soft drusen. *Graefes Arch. Clin. Exp. Ophthalmol.* **2017**, *255*, 2347–2355. [[CrossRef](#)]
13. Schmidt-Erfurth, U.; Sadeghipour, A.; Gerendas, B.S.; Waldstein, S.M.; Bogunović, H. Artificial intelligence in retina. *Prog. Retin. Eye Res.* **2018**, *67*, 1–29. [[CrossRef](#)] [[PubMed](#)]
14. Ting, D.S.W.; Pasquale, L.R.; Peng, L.; Campbell, J.P.; Lee, A.Y.; Raman, R.; Tan, G.S.W.; Schmetterer, L.; Keane, P.A.; Wong, T.Y. Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **2018**, *103*, 167–175. [[CrossRef](#)] [[PubMed](#)]
15. Sarhan, M.H.; Nasser, M.A.; Zapp, D.; Maier, M.; Lohmann, C.P.; Navab, N.; Eslami, A. Machine Learning Techniques for Ophthalmic Data Processing: A Review. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3338–3350. [[CrossRef](#)]
16. Wang, J.; Hormel, T.T.; Gao, L.; Zang, P.; Guo, Y.; Wang, X.; Bailey, S.T.; Jia, Y. Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning. *Biomed. Opt. Express* **2020**, *11*, 927–944. [[CrossRef](#)] [[PubMed](#)]
17. Le, D.; Alam, M.; Yao, C.K.; Lim, J.I.; Hsieh, Y.-T.; Chan, R.V.P.; Toslak, D.; Yao, X. Transfer Learning for Automated OCTA Detection of Diabetic Retinopathy. *Transl. Vis. Sci. Technol.* **2020**, *9*, 35. [[CrossRef](#)]
18. Rasti, R.; Rabbani, H.; Mehridehnavi, A.; Hajizadeh, F. Macular OCT Classification Using a Multi-Scale Convolutional Neural Network Ensemble. *IEEE Trans. Med. Imaging* **2018**, *37*, 1024–1034. [[CrossRef](#)]
19. Burlina, P.M.; Joshi, N.; Pacheco, K.D.; Freund, D.E.; Kong, J.; Bressler, N.M. Use of Deep Learning for Detailed Severity Characterization and Estimation of 5-Year Risk Among Patients With Age-Related Macular Degeneration. *JAMA Ophthalmol.* **2018**, *136*, 1359–1366. [[CrossRef](#)]
20. Grassmann, F.; Mengelkamp, J.; Brandl, C.; Harsch, S.; Zimmermann, M.E.; Linkohr, B.; Peters, A.; Heid, I.M.; Palm, C.; Weber, B.H. A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. *Ophthalmology* **2018**, *125*, 1410–1420. [[CrossRef](#)]
21. Russakoff, D.B.; Lamin, A.; Oakley, J.D.; Dubis, A.M.; Sivaprasad, S. Deep Learning for Prediction of AMD Progression: A Pilot Study. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, 712–722. [[CrossRef](#)]
22. Hwang, D.-K.; Hsu, C.-C.; Chang, K.-J.; Chao, D.; Sun, C.-H.; Jheng, Y.-C.; Yarmishyn, A.A.; Wu, J.-C.; Tsai, C.-Y.; Wang, M.-L.; et al. Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics* **2019**, *9*, 232–245. [[CrossRef](#)]
23. Romo-Bucheli, D.; Erfurth, U.S.; Bogunovic, H. End-to-End Deep Learning Model for Predicting Treatment Requirements in Neovascular AMD From Longitudinal Retinal OCT Imaging. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3456–3465. [[CrossRef](#)]
24. Baghaie, A.; Yu, Z.; D’Souza, R.M. State-of-the-art in retinal optical coherence tomography image analysis. *Quant. Imaging Med. Surg.* **2015**, *5*, 603–617. [[CrossRef](#)]
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
26. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
27. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
28. Morid, M.A.; Borjali, A.; Del Fiol, G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Comput. Biol. Med.* **2021**, *128*, 104115. [[CrossRef](#)] [[PubMed](#)]
29. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [[CrossRef](#)]
30. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
31. Chollet, F. Keras. 2015. Available online: <https://keras.io/> (accessed on 11 May 2021).
32. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, C.S.; Davis, A.; Dean, J.; Devin, M.; et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Prepr.* **2016**, arXiv:1603.04467.
33. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 1.
34. Araújo, T.; Aresta, G.; Mendonça, L.; Penas, S.; Maia, C.; Carneiro, Â.; Mendonça, A.M.; Campilho, A. DR|GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. *Med. Image Anal.* **2020**, *63*, 101715. [[CrossRef](#)]
35. Laves, M.-H.; Ihler, S.; Ortmaier, T. Uncertainty Quantification in Computer-Aided Diagnosis: Make Your Model say “I don’t know” for Ambiguous Cases. *arXiv* **2019**, arXiv:1908.00792.