Comprehensive fully-automatic multi-depth grading of the clinical types of macular neovascularization in OCTA images

Plácido L. Vidal[1,2] · Joaquim de Moura[1,2] · Pablo Almuiña[4,5] · María Isabel Fernández[3,4,5] · Marcos Ortega[1,2] · Jorge Novo[1,2]

Abstract

Optical Coherence Tomography Angiography or OCTA represents one of the main means of diagnosis of Age-related Macular Degeneration (AMD), the leading cause of blindness in developed countries. This eye disease is characterized by Macular Neovascularization (MNV), the formation of vessels that tear through the retinal tissues. Four types of MNV can be distinguished, each representing different levels of severity. Both the aggressiveness of the treatment and the recovery of the patient rely on an early detection and correct diagnosis of the stage of the disease. In this work, we propose the first fully-automatic grading methodology that considers all the four clinical types of MNV at the three most relevant OCTA scanning depths for the diagnosis of AMD. We perform both a comprehensive ablation study on the contribution of said depths and an analysis of the attention maps of the network in collaboration with experts of the domain. Our proposal aims to ease the diagnosis burden and decrease the influence of subjectivity on it, offering a explainable grading through the visualization of the attention of the expert models. Our grading proposal achieved satisfactory results with an AUC of 0.9224 ± 0.0381. Additionally, the qualitative analysis performed in collaboration with experts revealed the relevance of the avascular plexus in the grading of all three types of MNV (despite not being directly involved in some of them). Thus, our proposal is not only able to robustly detect MNV in complex scenarios, but also aided to discover previously unconsidered relationships between plexuses.

Keywords Optical coherence tomography angiography · Computer-aided diagnosis · Age-related macular degeneration · Multi-depth analysis · Qualitative analysis · Attention maps

1 Introduction

Thanks to the advances in sciences and, especially, in medical sciences, we are amidst an epoch of an ageing population [8, 23]. Given this unprecedented situation, we are facing a surge in age-related pathologies, promoting research on methodologies focused on the analysis of said afflictions.

Among the main cases of age-related pathologies, we must highlight those that arise due to consumption habits in developed countries, such as Diabetic Macular Edema or DME [6, 29]. This pathology manifests itself in the degeneration of the blood vessels that irrigate the retina, causing fluid leakages that deform and destroy its natural histology. DME is usually studied by means of Optical Coherence Tomography or OCT, which generates a cross-sectional representation of

the retinal layers [2]. Due to the relevance of the pathology, methodologies have been designed for its study from different perspectives [16], such as looking for a classification of the different fluid accumulations in the work of Barua et al. [4] and Wu et al. [37], segmentation of the accumulations in the work of Wu et al. [36] and Rahil et al. [21], or generation of probability maps of the subtypes in the work of Vidal et al. [33, 34].

However, in this work, we will focus on Age-related Macular Degeneration or AMD, which constitutes the leading cause of blindness in developed countries and is not only affected by consumption habits but is a direct consequence of the increase of life expectancy. Of this pathology, its neovascular exudative variant (nAMD) is the most noteworthy, since it represents a severe type of late AMD [5].

The presence of macular neovascularization (MNV) is the defining characteristic of the nAMD. The degeneration of the macular tissues can stimulate the development of new vessels. These new vessels may leak and bleed (clinically

✉ Joaquim de Moura
joaquim.demoura@udc.es

Extended author information available on the last page of the article

termed exudation), disrupting the normal architecture of the retinal layers. This, ultimately produces a fibrotic disciform scar. Patients with nAMD describe a rapid onset decrease of vision, metamorphopsia (or distortion of objects in the field of vision) and paracentral scotomas (or sections in the field of vision with complete or severe vision loss) [15].

Currently, the clinical literature considers four different types of MNV [27], being Type 1 the most common. It represents a growth of the vessels from the choriocapillaris into the sub-retinal pigmented epithelium (RPE) space with sparse leakages. Type 2 refers to the proliferation of new vessels from the choroid into the subretinal space. With this type of MNV, the vessels traverse the sub-RPE space. Type 3 MNV, occurs when the vascular proliferation starts from the deep capillary plexus in the retina, extending towards the outer retina. Characteristic to this type of lesion, the blood flow within the vascular proliferation is supplied by the retinal vessels, instead of the choriocapillaris. Scattered intraretinal haemorrhages and intraretinal fluid are present, with cystoid macular edemas clearly visible in other imaging modalities. The final stage of this type is the formation of a retinal-choroidal anastomosis (or connection/opening between two diverging structures) [3]. Finally, a fourth type is considered, when the patterns present a mixture between the Type 1 and Type 2 MNV (neovascularization in the sub-RPE region and subretinal compartments). This would represent an intermediate state between Type 1 and 2, albeit still being a mixture between the two for its defining features.

Recently, the development of Optical Coherence Tomography Angiography (OCTA) presented a new explicit way of visualizing the vascular flow (and, consequently, pathologies that leave trace or directly involve vascular structures). This medical imaging modality, unlike the aforementioned OCT modality used in the works of Barua et al. [4], Vidal et al. [33, 34] or Wu et al. [36, 37], generates a representation of the structures with vascular flow by means of comparing the decorrelation signal (differences in the backscattered OCT signal intensity or amplitude) between sequential cross-sectional scans taken at the same location. This results in a representation of the vascular blood flow map [11, 27]. That is, unlike OCT (and despite the name similarities), OCTA allows for the study of only the vascular structures and not the retinal tissues, absent in the generated images. Moreover, this medical imaging modality does not the need invasive contrasts [9] unlike other techniques with similar purpose, with the additional advantage of generating these images faster than other techniques with comparable objectives.

The integrity of this vascular flow map has shown to be correlated with several diseases, as well as the direct visual acuity of the patients [7, 10, 24]. These images are commonly analyzed at different depths, divided into four (and, sometimes, five) main regions: the Superficial Capillary Plexus (SCP), Deep Capillary Plexus (DCP), Avascular Plexus (AP) and Choriocapillaris or CP. In Fig. 1, we present an Optical Coherence Tomography image showing a cross-sectional representation of the retinal layers depths from where the OCTA images are extracted [12, 22]. In this figure, additionally from the different labeled depths, for future reference, we indicate also the boundary layers that establish their limits. The vascular flow of these depths can be seen in Fig. 2, where included a few representative examples of OCTA images for both each type of MNV and normal OCTA images.

Fig. 1 Cross-sectional retinal image (OCT). Over it, the relevant depths with their boundary layers from where OCTA images are generated
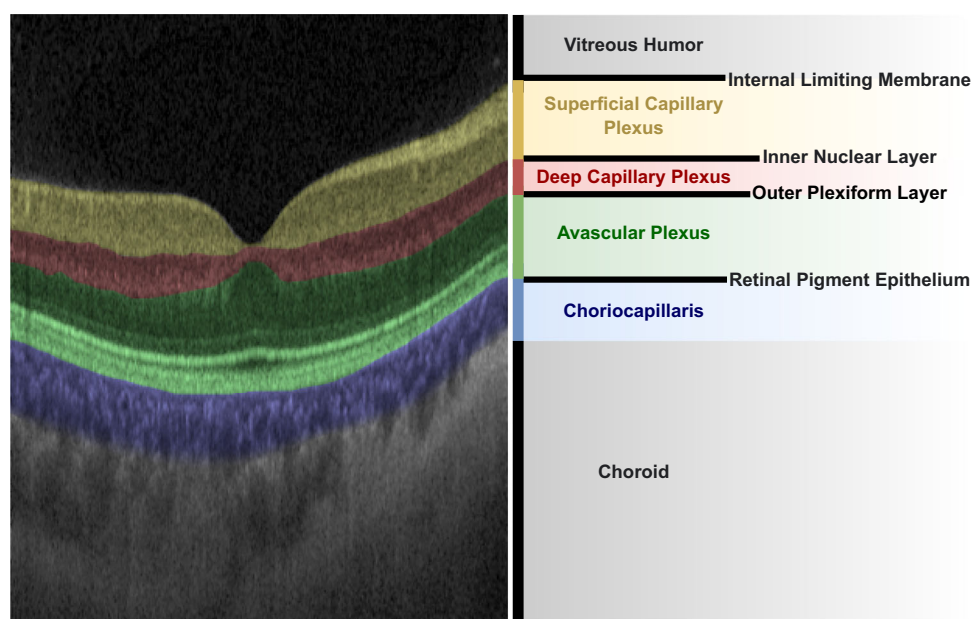
Fig. 2 Representative examples of OCTA images from both normal and each of the types of MNV. The images are presented at the three depths and two area sizes considered in this work

Area	Normal	Type 1	Type 2	Type 3	Mixed
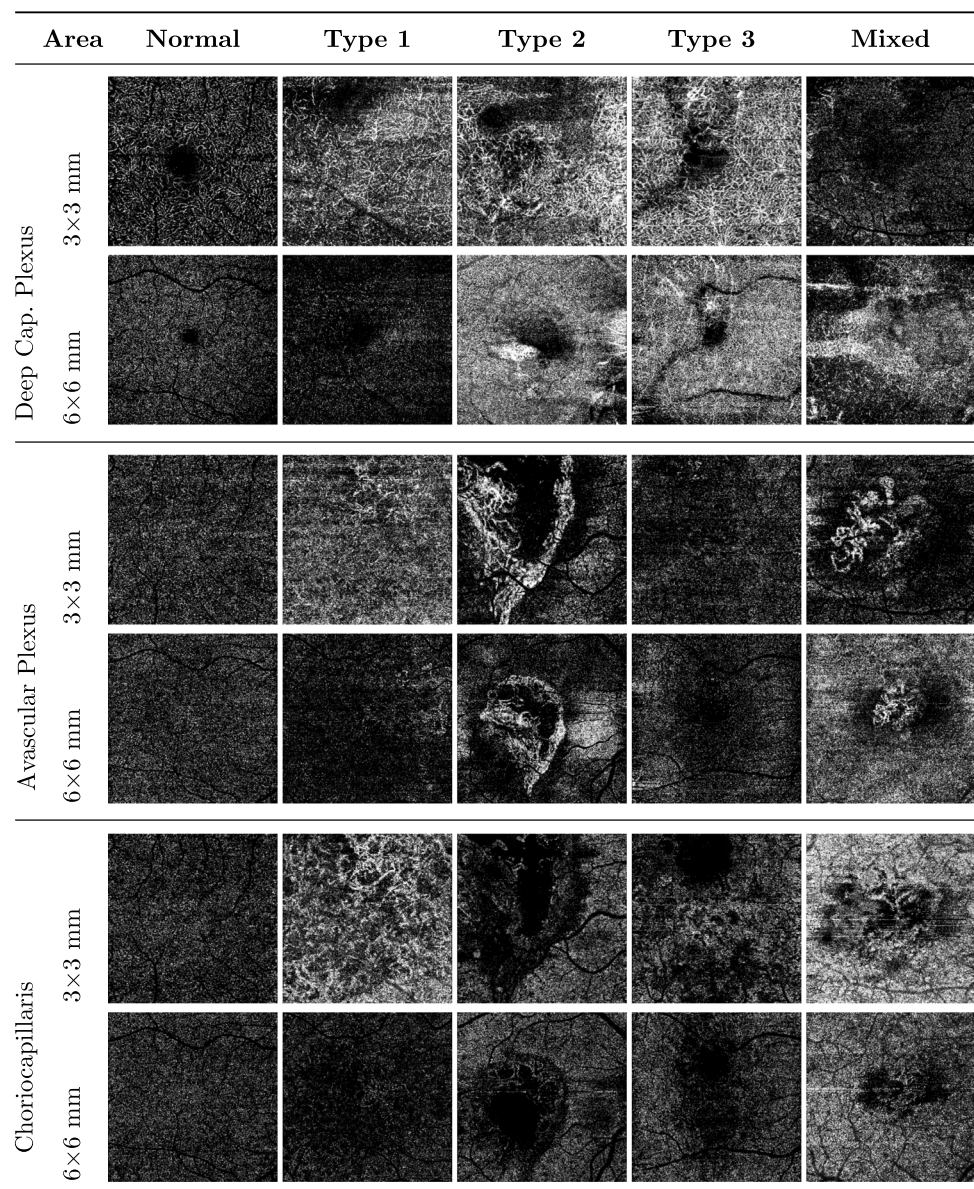
Due to the relevance of the aforementioned AMD, some works have focused on its classification. However, due to the novelty of the medical imaging modality, scarcity of samples and difficulty of its diagnosis in some stages of the disease, works exclusively focused on the detection of AMD-related MNV are not prominent and even less so that focus in obtaining further understanding on the results and/or studied pathology). Some of the proposals, such as the work of Alfahaid et al. [1], center its analysis to determine the presence or absence of AMD. This particular work is based on the analysis of texture descriptors through local binary patterns. In this same line, the works Thakoor et al. [31, 32]

classify this presence or absence of neovascular AMD and non-neovascular AMD on all the considered depths through deep learning strategies. Other works focus on the extraction of a defined segmentation of the MNV, like the work of Liu et al. [17] with a combination of pre-processing filters, saliency map strategies and post-processing stages. Another example is the work of Xue et al. [38], with a thresholding strategy followed by a CLIQUE clustering algorithm. Finally, works like the proposal of Wang et al. [35], Patel et al. [20], and Jin et al. [14] offer a multi-task approach, performing a binary classification on the presence or absence of neovascularization as well as obtaining a segmentation.

1.1 Research gaps

With the above in mind, we can summarize the research gaps we have found in the literature with the following ideas:

- Due to the novelty of OCTA, the features that define the pathology in this medical imaging modality at different depths are ill-defined.
- Most of the research is focused on Type 1 MNV, or in analyzing the presence or absence of the general pathology.
- Only a few retinal depths are used or studied in the literature, severely hindering the potential of the proposed methodologies and potentially offering biased results.

This means that, while there is a clinical definition of each type of MNV, the impact and features that each type presents in the OCTA medical imaging modality remains not well studied. This is mostly because the lack of samples and a properly labeled datasets which, in consequence, limits what these research works can study. In particular, most of the works we found explore only the most common type of MNV, Type 1, that is mostly limited (per definition in the clinical literature) to a very narrow depth in the Choriocapillaris layer. Also, as presented, some works also consider a binary classification on the presence or absence of the pathology altogether, reducing the impact of the two aforementioned factors.

1.2 Our proposal

To approach these research gaps, we propose a fully-automatic methodology with the objective of obtaining a grading on the severity of MNV in all the four types/stages that are considered in the clinical literature only based on OCTA images. With this methodology, we perform an in-depth ablation study of all the different depths relevant for the diagnosis of the pathology: DCP, AP and CP presented

in Fig. 2. We assess the relevance of each depth for each type of the pathology, as well as the contributions of their combinations to the detection and diagnosis of MNV. In Fig. 3, we present all the combinations and surfaces analyzed in this work. The same way, the dataset is composed by images at different stages of treatment of the disease, ensuring that the model is able to perform a detection even in the most borderline scenarios.

Finally, as our proposal aims to be useful for a coherent study on the features that define the target pathology (and to use as much samples as we have available to us from clinical practice) we also include samples with severe artifacts. In Fig. 4, we present a collection of these OCTA images from our dataset where the vasculature and fundus at different analyzed depths are affected by these artifacts. As the reader can appreciate in these examples, these images represent a challenging scenario, a hindrance that has to be taken into account when analyzing the affliction, as the information contained in the images is severely distorted and may lead to false diagnosis.

In summary, the main contributions of our proposal are:

- First work to perform a comprehensive ablation study with different complexity levels of CNNs to analyze the contribution of the relevant OCTA depths and scanning region sizes.
- Fully-automatic grading methodology for the four clinical types of AMD-associated MNV in OCTA imaging.
- Complete qualitative analysis of the results through explainable artificial intelligence strategies.
- Proposal tested with patients at different stages of treatment, including images with complex artifacts from the capturing process.

This manuscript is divided as follows: in Section 2 "Dataset", we present the information and protocols followed to capture the images from the patients, device configuration and other pertinent data related to the acquisition of

Fig. 3 Depths, surfaces and their combinations analyzed with the proposed methodology
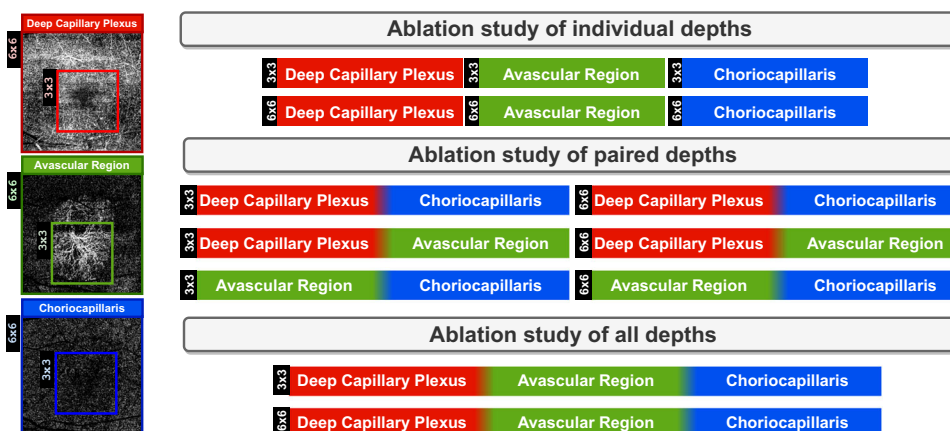
Fig. 4 Representative examples of OCTA images with capture artifacts taken from all the considered depths and areas

Area	Deep Cap. Plexus	Avascular Plexus	Choriocapillaris
3×3 mm			
6×6 mm			
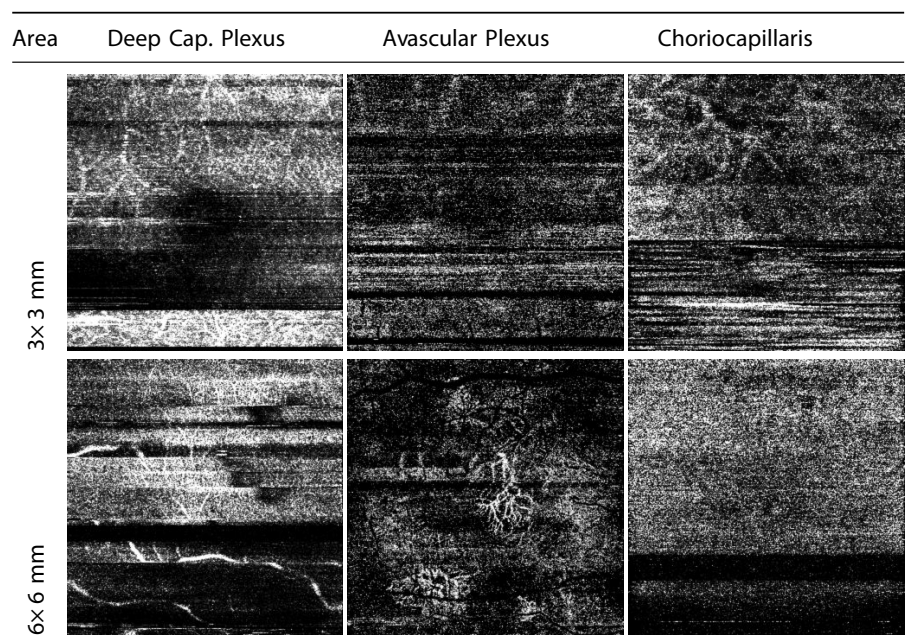
the images. Then, in Section 3 "Methodology", we proceed to explain our proposal and the experiments/analysis that compose it. The results are shown and discussed in Section 4 "Results and discussion". Finally, we provide some final notes and possible future works in Section 5, "Conclusions".

2 Dataset

For the development of this work, we used a dataset composed by 939 OCTA samples generated with a SS-DRI-Triton-OCTA device (Topcon Corp Inc, Tokyo, Japan). This capture device has an A-scan rate of 100,000 scans per second, using a light source with a wavelength of 1 μm allowing a deeper penetration into tissue. This allows better axial resolution and improved detection sensitivity of microvasculature. All these images were taken centered in fovea, and 464 from these images were taken using a 3×3 mm scan pattern and 475 from 6×6 mm scan pattern. The division of samples per MNV class in this work is presented in Table 1.

The images were labeled by a team of expert clinicians, conducting a prospective study of naive neovascular AMD, treated with Ranibizumab in a "Treat and Extend" pattern with a follow up of 12 months. The OCTA images were

Table 1 Number of OCTA samples (three image depths per sample) for each type of MNV

	¬ MNV	Type 1	Type 2	Type 3	Mix 1&2
3 × 3	99	103	42	142	78
6 × 6	98	107	43	143	84

analysed using IMAGEnet 6 and the OCTARA algorithm [28]. All the eyes were studied under pharmacological mydriasis. Two retina specialists assessed OCTA for abnormal MNV flows using both en face and B-Scan flow images with head-to-head comparisons with structural OCT to eliminate misidentification affected by artifacts. All the visits (from the baseline to the final stages) were included in the dataset, independently of the artifacts and the treatment/severity stage.

3 Methodology

Our proposed work is divided into two stages. The first stage represents the main proposal, a methodology focused on the grading of the target pathology, as well as the consequent ablation study on the relevance of each depth (as shown in Fig. 3 and further explained in Section 3.1). In the second stage, on the other hand, we focus in the comprehensive qualitative analysis performed in collaboration with experts of the domain. This analysis, as mentioned, is made by means of the independent and joint analysis of the attention maps at different depths. To do so, we use the models trained in the first stage, so we can also further compare the performance obtained with the real behavior of the system (and, thus, ensure its validity and robustness). Through this analysis, we can better understand both the behavior of the models and the pathology itself (Section 3.2).

3.1 Fully-automatic grading

First of all, we will train the models to perform the grading on an input based on individual depths. This allows us to assess

the contribution of each depth to determine the pathological (or normal) class of the samples. This way, we can find alternative ways of circumnavigating the difficulties that arose from the complex artifacts present in most of the images. The grading is done such as the samples are classified into normal (that is, no MNV is present in the input image), the four different levels of MNV (including the type representing a mixture of Type 1 and Type 2 MVN). While these combinations are explained in Sections 3.1.1 to 3.1.3, the precise configuration and strategy used to train the models of these analysis are explained in Sections 3.1.4 to 3.1.6.

3.1.1 Individual depth ablation study

The first proposal performs an study on the relevance of each considered depth individually. While (by their clinical definition explained the introduction of this manuscript) some types of MNV are limited to one of the considered depths (such as Type 1, limited to the sub-RPE space), others (such as Type 3) transverse multiple regions. However, the structural changes could perfectly affect other depths in a way that could be detected by machine learning strategies. This way, in this first approach we study each of the different depths independently from the others. We want to find which depths are able (to a certain extent, we are not looking for a perfect grading, but more of a significant tendency to be followed in subsequent analysis stages) to distinguish types of MNV that, in theory and by definition, do not leave a trace in said depths. Also, for the types that extend along several layers, the ones that are more significant. Moreover, the experiments will be repeated for the two considered surfaces in this work: 3×3 and 6×6 mm. This way, we can also assess the relevance of peripheral information of the models, as well as study the impact of macro and micro structures (as, in bigger areas, the small structures are degraded by the resolution of the images). Additionally, these models will also be the main analyzed in the posterior qualitative analysis, as they allow to explore the unbiased contribution towards the classification of each class. This way, we could further examine if features present in layers that are not deemed clinically relevant for the disease actually contribute with previously unknown features. Thus, in total, in this analysis we perform six experiments, or two per depth.

3.1.2 Paired depth ablation study

In this second approach, we study all the possible pairs of depths that are relevant for our work. In this case, we want to complement the results of the first proposal. We want to find if, for each type of MNV, the information of each individual layer can be complemented with its neighboring one (or opposite) to improve the resulting grading. While in the first approach the models would only analyze the unique struc-

tures present in the individual layers (or the patterns that leave a trace from another depths), by analyzing other layer the model can have an added information about the structures. For example, the model can assess if a particular unique structural pattern of the retinal/subretinal vessels is part of the normal nature of a patient or might be a pathological formation surging from structures at other depths. The same way as in the previous analysis, we test both 3×3 mm and 6×6 mm, thus also resulting in a total of six experiments for this analysis.

3.1.3 Complete multi-depth ablation study

Finally, we study the potential combination of all the considered depths towards the grading of the pathology. This allows us to compare the results with the other studies to, in effect, perform a complete ablation study. In this analysis, we are able to assess how the system performs with all the available information, and the degradation of the results when the components of each previous iterations are not present. Additionally, we can assess that these layers, instead of providing information, only increment the apparent information noise (for example, by requiring more complex models to obtain similar grading results). Like in the previous analysis, to allow for a full comparison of the results, we also test both 3×3 mm and 6×6 mm, thus resulting in two experiments in this case.

3.1.4 Model training

As we want to perform a coherent study on the performance of our proposal, we study different configurations of a proven model (presented in Section 3.1.5). To train and validate these models we used random repetitions at patient level for each of the depth configurations contemplated in the aforementioned ablation studies and surfaces. That is, no samples from the same patient are present both in training and test, but each repetition distributes the subjects at random between training, validation and test sets with a given proportion each. By using this strategy, we mitigate the effects of the dataset imbalance present in some of the classes and for some subjects, as a cross-validation strategy would result in some folds to have significantly less samples than other iterations in some classes and return biased results. Then, due to the low number of repetitions of a cross-validation and its strict repartition of the folds, each of these biased experiments would highly impact the final results. By performing a high number of random repetitions, these folds contribute less to the final metrics and we get closer to the reality. Nonetheless, for informative and disclosure purposes, we also include the experiments where we train and evaluate the performance of all the stages of our methodology by means of a 10-fold crossvalidation in the Appendix.

The training of each repetition is done by means of an early-stopping strategy. That is, the training of the model ends whenever no improvements are achieved in the validation loss for a given number of epochs. The final weights of the model are the ones which achieved the minimum validation loss through all the epochs. This way, this strategy allows to preserve the model with the best generalisation capabilities while also preventing further overfitting (which would happen in a manually-set number of epochs). Finally, the learning rate of the optimizer is modulated by a scheduler, where the learning rate would decrease if the model reached stagnation. This allows for the training to self-regulate, decreasing the step-sizes of the gradient descent the closer the model is to a minimum. We follow these adaptive strategies to minimize the impact of the initial configuration in the results of the ablation and posterior study performed. The full pseucodode of the training is presented in Algorithm 1.

3.1.5 Model architecture

The CNN network architecture chosen is the DenseNet [13]. As benefits, this network has shown to be a robust solution used in pathologies with similar characteristics to the one studied, as well as in medical imaging modalities with similar properties. Furthermore, as this architecture consists of dense blocks connected by skip-connections, mitigates the possible effects of gradient vanishing in large networks or overfitting in smaller configurations. This resilience allows to minimize the effect of the configuration on the results that were obtained in the conducted ablation study, being the main factor affecting its outcome and the complexity of the problem in relation to the chosen network. Thus, this network architecture represents a robust proven baseline to evaluate both the contribution to the clinical domain of the different factors considered in the proposed ablation study and a wide range of network complexities needed for the classification of the pathology.

In particular, we tested the configurations shown in Table 2: the DenseNet configurations 121, 161, 169 and 201. We chose these configurations because, while the DenseNet 121 configuration is more reliable in simpler tasks, it obtains lesser results in more complex scenarios (albeit able to generalize more). On the other hand, the DenseNet 201 configuration is quite reliable on very complex scenarios, but (despite all the safeguards and measures of the network to prevent it) tends to overfitting when the task at hand has a simpler solution. This way, we can study not only the relevance of the layers, but also compare the models to assess the information noise and complexity of the features needed to solve the task.

The potential drawback of using this architecture is the concatenation of feature maps from each layer with the

---

Algorithm 1 Pseudocode of the training strategy.

Input: $numRepetitions$, the number of times to repeat the training process.
Input: $maxPatience$, the number of epochs without loss improvement to stop training
Input: $maxStagnation$, the number of epochs without loss improvement to reduce the learning rate.
Input: $lrFactor$, the learning rate reduction factor when maxStagnation is reached.
Input: $modelConfigurations$, each network architecture to consider.
Output: $trainMetrics, validationMetrics, testMetrics, model$ for each model configuration, surface and considered combination of depths.
for each surface $\in (3x3, 6x6)$ **do**
    for each combination of 1, 2 and 3 elements of (DCP, AP, CP) **do**
        $repetitions \leftarrow numRepetitions$ random distributions of the patients.
        for each $repetition \in repetitions$ **do**
            divide $repetition$ into $trainSet, validationSet, testSet \mid testSet \cap validationSet \cap trainSet = \emptyset$
        end for
        for each model $\in modelConfigurations$ **do**
            for each $trainSet, validationSet, testSet \in repetitions$ **do**
                $stagnation \leftarrow 0$
                $patience \leftarrow 0$
                do
                    $model, trainMetrics \leftarrow train(model, trainSet, learningRate)$
                    $validationMetrics = model(validationSet)$
                    $loss \leftarrow CrossEntropyLoss(validationMetrics)$
                    if $loss < minimumLoss$ **then**
                        $minimumLoss \leftarrow loss$
                        $stagnation \leftarrow 0$
                        $patience \leftarrow 0$
                    else
                      $stagnation \leftarrow stagnation + 1$
                      $patience \leftarrow patience + 1$
                      if $stagnation >= maxStagnation$ **then**
                        $learningRate \leftarrow learningRate * lrFactor$
                        $stagnation \leftarrow 0$
                    end if
                  end if
                while $stagnation < maxPatience$
                $testMetrics \leftarrow test(model, testSet)$
             end for
        end for
    end for
end for

---

previous layer, leading to duplicated information. As the number of layers in the network grows along the different configurations, this duplication of information can result in an unnecessary increase in the number of model parameters, leading to greater computational and memory requirements during training. While we can use proven strategies to reduce this effect (such as pruning and dropout), using these techniques would require further analysis of the behavior of the network to prevent collateral effects during the analysis (specially in larger datasets).

Table 2 Basic structure of the DenseNet-based network configurations used in this work, where λ represents the size of the sides of the input sample

Layers	Convolution	Pooling	Dense block	Transition layer		Dense block	Transition layer		Dense block	Transition layer		Dense block	Classification layer
Output	$\frac{\lambda}{2} \times \frac{\lambda}{2}$	$\frac{\lambda}{4} \times \frac{\lambda}{4}$	$\frac{\lambda}{4} \times \frac{\lambda}{4}$	$\frac{\lambda}{4} \times \frac{\lambda}{4}$	$\frac{\lambda}{8} \times \frac{\lambda}{8}$	$\frac{\lambda}{8} \times \frac{\lambda}{8}$	$\frac{\lambda}{8} \times \frac{\lambda}{8}$	$\frac{\lambda}{16} \times \frac{\lambda}{16}$	$\frac{\lambda}{16} \times \frac{\lambda}{16}$	$\frac{\lambda}{16} \times \frac{\lambda}{16}$	$\frac{\lambda}{32} \times \frac{\lambda}{32}$	$\frac{\lambda}{32} \times \frac{\lambda}{32}$	5×1
DenseNet 121	7×7 Convolution Stride 2	3×3 Maxpool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 6$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 12$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 24$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 16$	Fully-connected softmax 5 classes
DenseNet 161	7×7 Convolution Stride 2	3×3 Maxpool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 6$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 12$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 36$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 24$	Fully-connected softmax 5 classes
DenseNet 169	7×7 Convolution Stride 2	3×3 Maxpool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 6$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 12$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 32$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 32$	Fully-connected softmax 5 classes
DenseNet 201	7×7 Convolution Stride 2	3×3 Maxpool Stride 2	$\begin{bmatrix} \times1\ conv \\ \times3\ conv \end{bmatrix} \times 6$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 12$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 48$	1×1 Convolution	2×2 Average pool Stride 2	$\begin{bmatrix} 1\times1\ conv \\ 3\times3\ conv \end{bmatrix} \times 32$	Fully-connected softmax 5 classes

3.1.6 Training configuration

First of all, regarding the dataset, the images are rescaled to a size of 224 × 224 as no diminishing results were observed at this resolution configuration and the load on the model was lesser. Additionally, to prevent overfitting and increase the information extracted from our dataset, as data augmentation strategy the images were randomly rotated a random angle between 90 and -90 degrees with nearest neighbors interpolation. Finally, they were randomly flipped horizontally and/or vertically with a probability of 0.5 each.

For the training, we employed the Stochastic Gradient Descent or SGD with Nesterov momentum of 0.9 [30]. The training of the models was performed with a patience of 35 epochs for the early-stopping strategy. The scheduler, on the other hand, reduces to a 75% of its value the learning rate every 10 epochs without improvements in the loss. As optimizer, we used the CrossEntropyLoss [18, 19], with an initial learning rate of 0.005 (obtained by preliminary tests as an stable starting point with acceptable training times). Finally, the experiments are repeated 25 times, with random distributions at patient-level between the train, validation and test sets (that is, no images from the same patient are allowed into both the evaluation and training sets). The dataset was divided into 60% for training, 25% for validation and 15% for testing in each repetition.

3.1.7 Model evaluation

To evaluate the performance of the models, we employed the metrics shown in Equations 1 to 5, where TP are the True Positives, TN the True Negatives, FP the False Positives and FN the False Negatives:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \ \text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{AUC} = \int_{x=0}^{1} Precision(FPR^{-1}(x))dx, \ \text{FPR} = 1 - Recall \tag{3}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{5}$$

These measures are the *accuracy* (or percentage of correctly classified samples), *precision* (or percentage of positives returned by the system that are actually positive), *recall* (or percentage of true positives returned by the system from the total), *Area Under the ROC Curve* or AUC (or probability of the system to assign a higher score to a positive sample than a negative sample), F_1 *Score* (or the harmonic mean of the precision and recall), and *Matthews correlation coefficient* or MCC (the correlation between the true labels and the labels returned by the system). This last score is between -1 and 1,

as it is based on the correlation coefficient of Pearson. These metrics were chosen as they offer different points of view of the results but, as we are dealing with a multiclass problem, the average-per-class result will be presented in the cases the metric does not inherently support it by its definition (such as in the case of MCC).

3.2 Qualitative analysis of the attention

After the training of the different grading models and statistical analysis of the results (where we trained a total of 14 models at different scanning depths), we will study the attention maps of the networks to further understand the results in the scenarios that offer the most interesting study cases. That is, we will analyze in collaboration with experts what structures the networks are looking for at each depth to solve this challenging scenario and the situations where the models achieve outstanding performance for combinations of MNV types and depths where it should not be possible. To do so, we will explore the gradient-weighted class activation mappings describing the regions the network focused its attention [25, 26]. Then, we consult the most representative and relevant ones with the experts to assess what features and structures match with the ones returned by these attention maps.

Our proposal extracts the attention of the network by using the partially retained spatial information of the last convolutional layers and the higher-level features developed in them. It uses the gradient information flowing into this last convolutional layer to assign the importance values to each neuron. These activations reveal which parts of the original image contributed more to the final grading. Usually, these network attention maps are employed by methodologies to further ensure that the model is not taking advantage of other artifacts and elements of the image to improve its results such as noise and artifacts product of the capture device that may reveal a subgroup of patients, clinical devices that suggest symptoms of related pathologies to the expert system (or risk of), or interface information left by the capture device that gives clues related to the nature of the patient and pathology. In our case, the images do not contain any information pertinent to the patient. All the samples were taken with the same device and configuration, and both healthy and pathological images contain different levels of artifacts to ensure that the system can only assess the results based on true biomarkers.

In this work, we explore the situations where, in the previous analyses, we noticed odd behaviors of the models given the expertise of the clinicians. As mentioned, we will mainly focus mainly on the models based on individual depths so we can perform an unbiased per-layer study at this stage that complements and expands what we explored in the first stage. For example, good testing metrics for types of MNV with

theoretical no presence at the studied depth and vice versa. This way, in this work we not only propose a fully-automatic grading solution. In conjunction with the previous per-depth analysis we are further increasing the understanding of the target pathology and proving the robustness and validity of our work.

4 Results and discussion

In Sections 4.1 to 4.6 we will proceed to present and offer a brief discussion on the results obtained for each experiment. Each section offers the results of its analogous section from the explained methodology (including both quantitative and qualitative analyses). For the sake of allowing a better per-MNV-type analysis, the results of each experiment are presented as a normalized confusion matrix. The best global results for each scenario are presented in Section 4.4 with the metrics contemplated in the methodology.

4.1 Individual depth ablation results

In this first section of the results, we present the analysis of the metrics for each depth independently of the rest. That is, the networks were trained by using only one of the considered depths at the same time. This way, in Fig. 5 we present the results by using only the DCP depth. In Fig. 6 the results only using images taken at AP depth. Finally, in Fig. 7, images taken from the CP depth.

The first thing to note is that, as expected, most of the studied types of MNV obtain sensibly bad results at the DCP depth (Fig. 5) except from the particular case of Type 3 MNV. This is understandable, as this type of MNV is the one that extends through the retinal layers and prone to a vascular anastomosis along them. On the other hand, seems that regardless of the severity and type of the pathology, the DCP depth is as good as the rest of considered depths to determine wether a patient presents normal retinal patterns or any type of MNV. This indicates that, despite the clinical affliction be exclusively focused on a given depth or depths, the DCP still manifests artifacts or deformations that reveal the pathological nature present in the other depths. That is, despite not being able to determine wether the MNV is focused on proximal depths to the CP or already has traversed the retinal layers towards the DCP, still the DCP vascular structures seem to be altered in a way that allow to predict wether the samples are pathological or not in a global manner. Finally, we have to remark also how the methodology is slightly better at determining Type 2 MNV at the wider area (6×6 mm) than the smaller area (3×3 mm). From a machine learning standpoint, this could be due to the removal of smaller structures from the
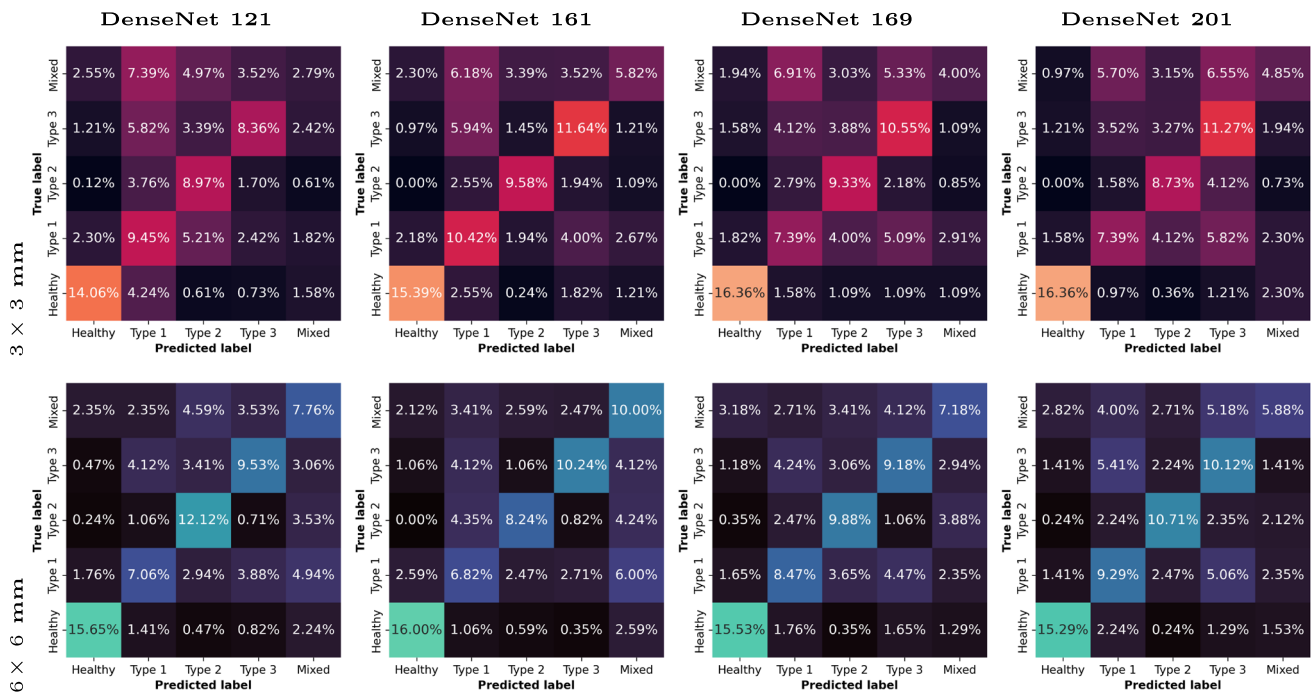
Fig. 5 Normalised confusion matrices for the models trained only using the OCTA images at DCP depth

resolution change. Both areas are covered in an image of the same size, so the 6×6 mm OCTA samples covers more area in an image with the same number of pixels than 3×3 mm. Thus, smaller patterns are lost. This could, theoretically, help

the network to focus on the macro structures that are formed and less on the removed smaller patterns.

Regarding the results from the models trained exclusively with images at the AP depth (Fig. 6), we can see how, even compared to the CP depth (Fig. 7), it obtained the best over-
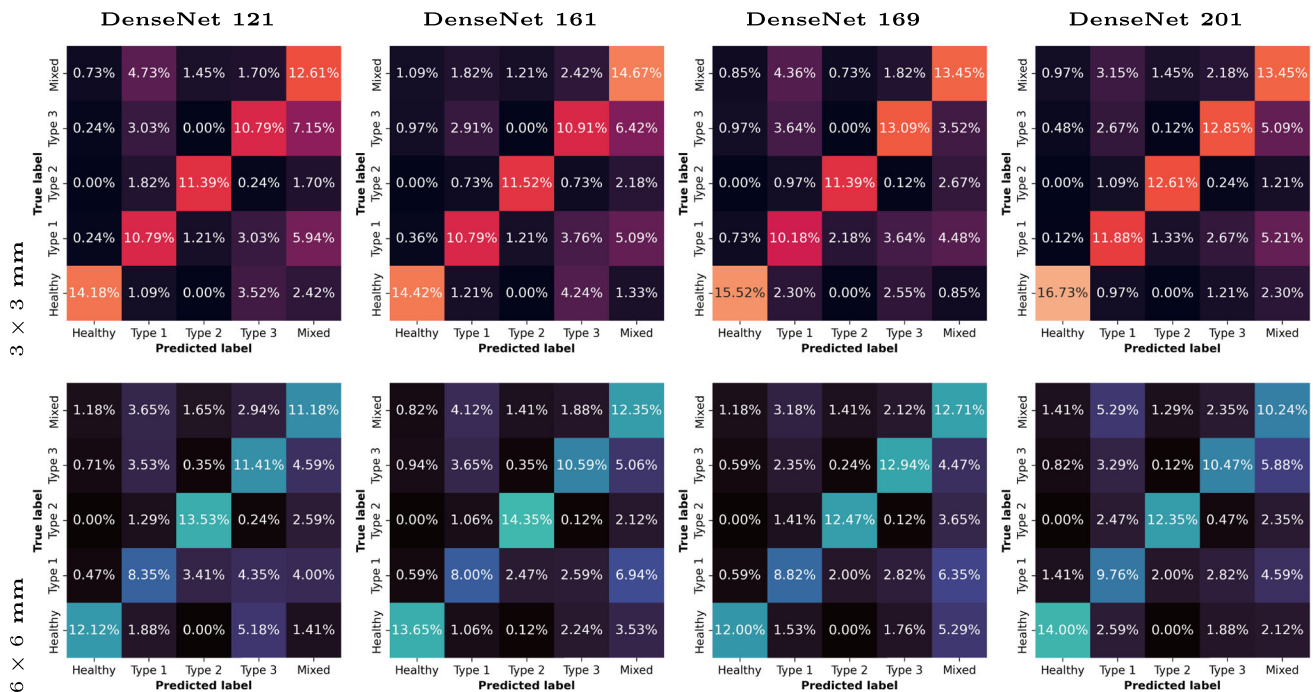

Fig. 6 Normalised confusion matrices for the models trained only using the OCTA images at AP depth
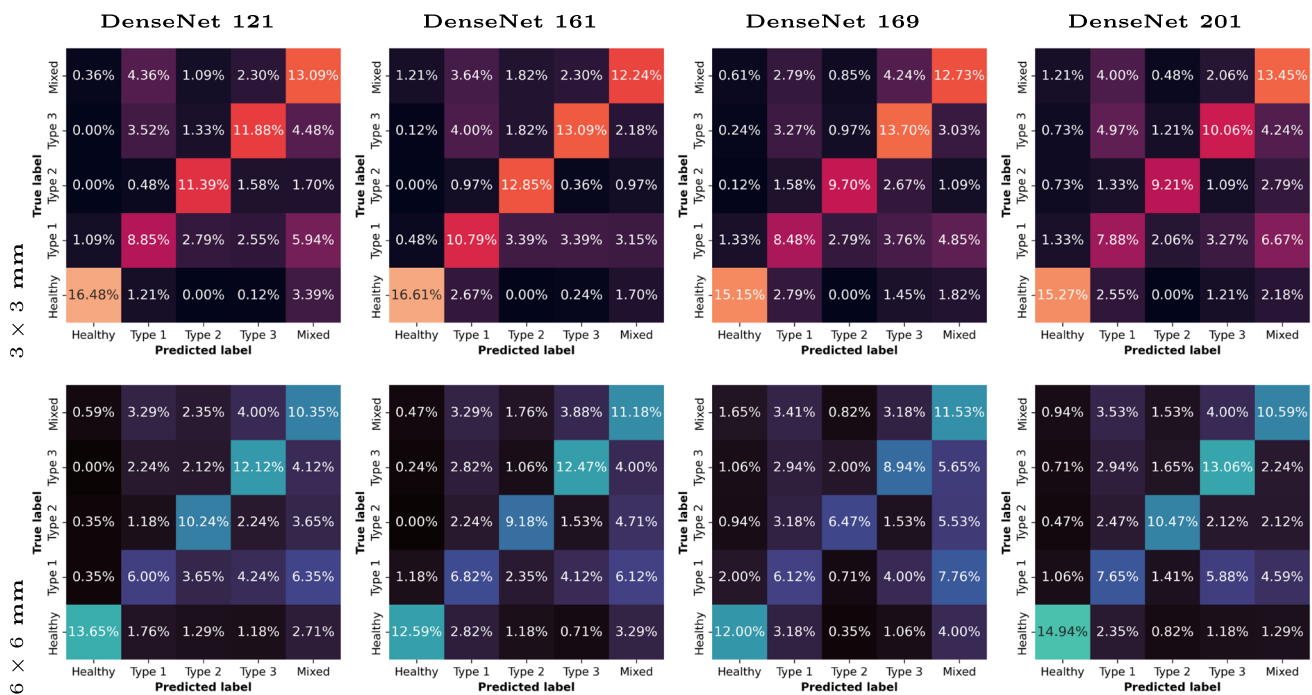
Fig. 7 Normalised confusion matrices for the models trained only using the OCTA images at CP depth

all results to determine both the presence and type of MNV. This is probably due to the fact that at this depth, as its name implies, there is a limited presence of vascular activity and, thus, less noise to infer features related to new vascular formations. Moreover, Type 2 MNV and, by extension, Type 3, are also characterised by the presence of new vessels that transverse this region. This justifies the good results regarding Type 2 and Type 3. However, the good results of Type 1 indicate that, despite this type of MNV only limited to the sub-RPE space (thus, with no presence whatsoever in the AP), microstructures product of MNV can be seen too in Type 1 MNV at this depth. This is confirmed by the fact that these good results are achieved exclusively in 3×3 mm OCTA images, while 6×6 mm OCTA images (that lost the small structures due to the image resolution as mentioned in the previous analysis) obtain understandable significantly worse results for Type 1 MNV.

Finally, as expected, the models trained exclusively with images at the CP depth (Fig. 7) present the best results for distinguishing normal from pathological cases. As MNV cases have its source at this level, its understandable that the most early indicators are present at this depth. However, this also indicates that we can infer a certain degree of severity from this layer despite being a common factor in all the MNV types. Also, we see how it obtains better results the higher the type of MNV (probably due to the fact that, the longer the reach of the retinal vessels, the thicker the base structures that grow from this depth need to be). Also, this depth seems to be also key for detecting cases with mixed patterns of Type 1 and

Type 2, indicating that this MNV already presents particular structural alterations at this depth that suggest the mixture between the types.

4.2 Paired depths ablation results

Regarding the paired analysis, we first will study the results obtained at the combined depths of the AP and the DCP (Fig. 8). In this case, we see that, as happened in the first individual analysis, these two layers mostly favor the Type 2 and 3 MNV, as the main difference between these two types is the presence or absence of MNV in the DCP layers. As no information from the CP is given (and, as mentioned, this is the source point for MNV), Type 1 MNV returns underwhelming results at these depths being confused in several times with the mixed MNV class.

Secondly, we study the results combining OCTA images at DCP depth and from the CP (Fig. 9). This case is particularly interesting, as is at this combination where the best results for detecting Type 1 MNV are found (and only in the 3×3 mm OCTA images). This is probably because the model is taking advantage of microstructures present in the innermost studied depth of the retina (DCP) that aid it to discern between Type 1 and Type 2 MNV. It seems that Type 2 MNV leaves traces in the DCP depth that aid the model to properly distinguish them from the Type 1. This is further confirmed by the good results also obtained in the mixed type MNV class.

Finally, in the particular case where the the model received as input OCTA images from both the CP and the AP depths
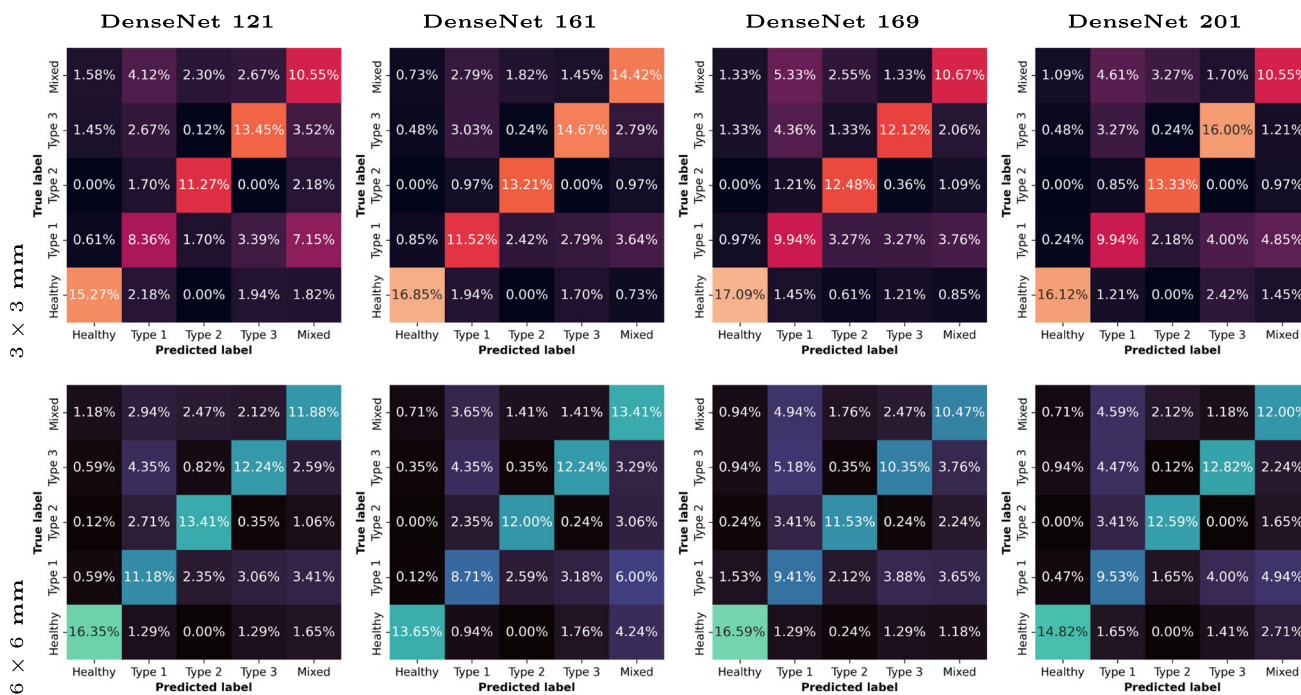
Fig. 8 Normalised confusion matrices for the models trained only using the OCTA images at both DCP and AP depths

(Fig. 10), as expected, the results are the best from the three combinations of this analysis. As seen in the analysis at individual depths, at this depths the most significant features are present: the CP where the MNV grows from, and the AP where features leaked to this layer aid to denote changes in Type 1 MNV (and also present in the definition for both Types 2 and 3).

4.3 Complete multi-depth ablation results

Finally, in Fig. 11 we present the results for the model that combines as input all the three depths. As expected, from all the options, this is the one that obtains best results (as it contains all the information available). However, the Type 1 MNV is still the class which suffers the most in the final
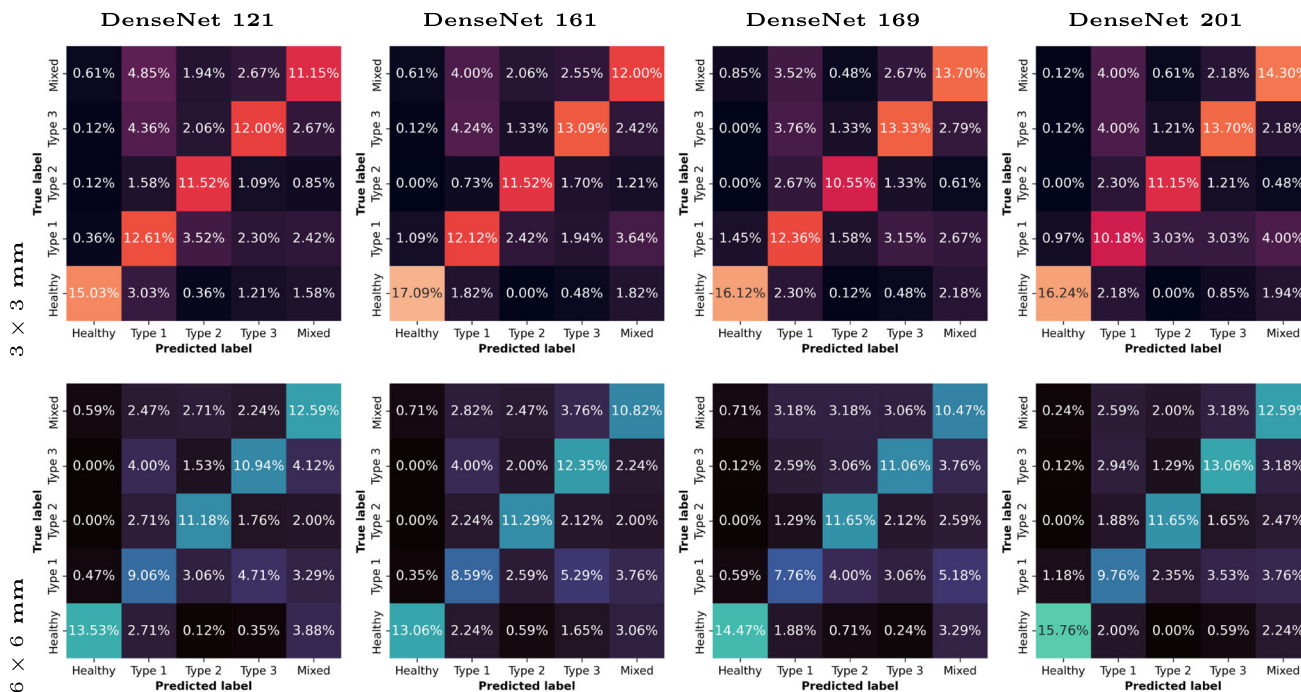
Fig. 9 Normalised confusion matrices for the models trained only using the OCTA images at both DCP and CP depths
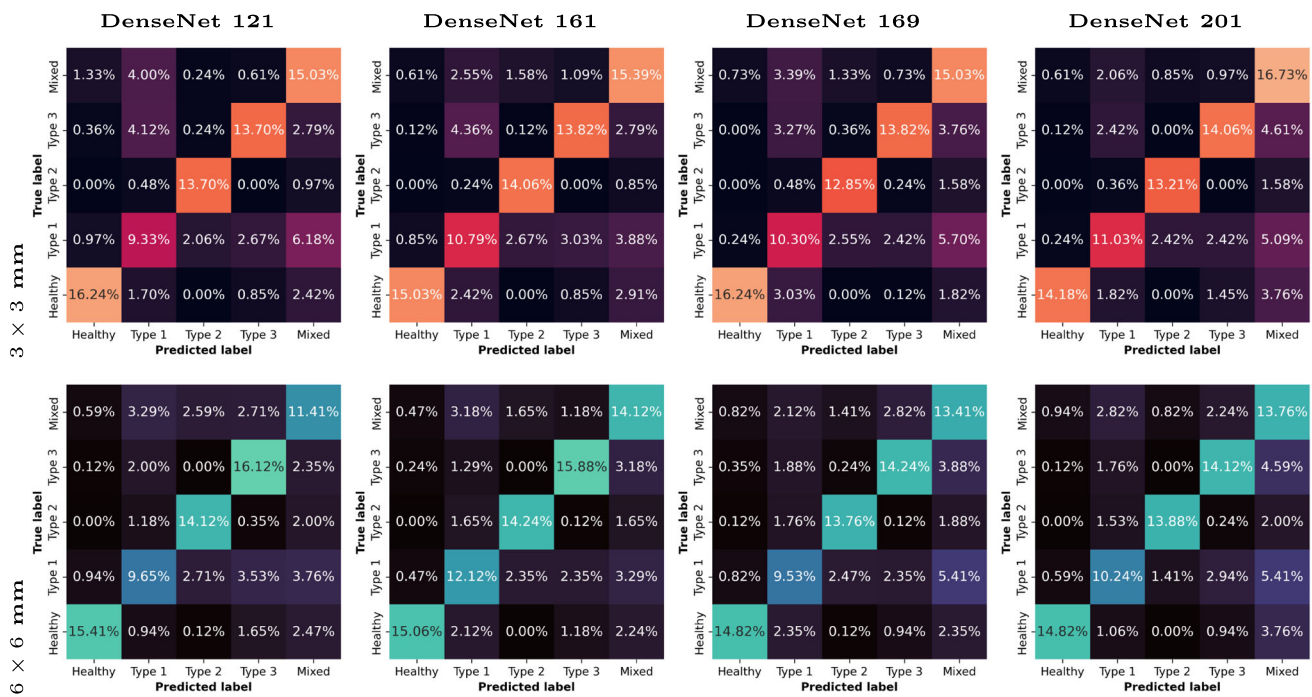
Fig. 10 Normalised confusion matrices for the models trained only using the OCTA images at both AP and CP depths

results. Additionally, the results shown that, when all the depths are available, the system is able to better assess the class of the input in the 3×3 mm OCTAs (being the exception Type 2 MNV that appears to be confused more with Type 1 and the mixed type).

However, all the results are obtained best with the most complex DenseNet configuration, and still significantly close to the combination of both the AP and CP layers in the previous paired approach with simpler models. Thus, we can infer that the addition of the DCP layer does not add significant
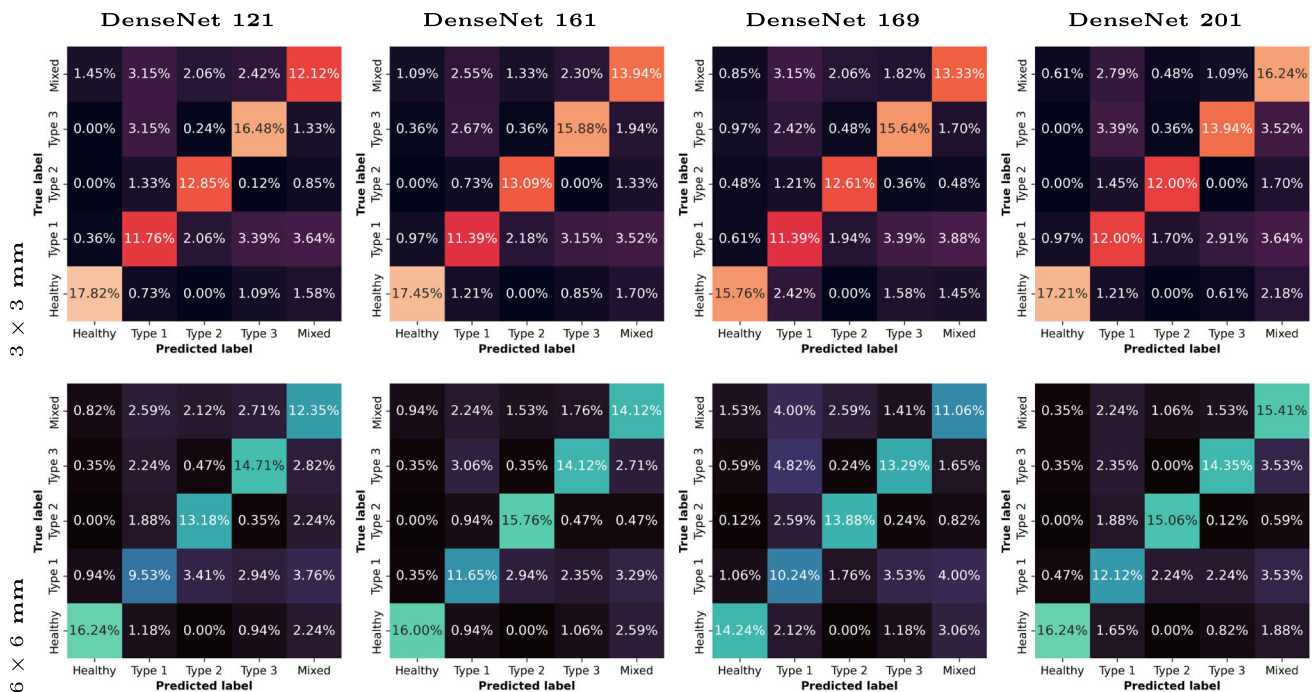

Fig. 11 Normalised confusion matrices for the models trained only using the OCTA images at the three considered depths

information to the detection, and mostly forces the model to use more lower-lever features to be able to remove the added noise from said depth. Nonetheless, these scenarios will be further explored in the following results of the qualitative analysis (Section 4.6), which will help us to confirm if this layer actually contributes and in which manner.

4.4 Global results

The best metrics for each type of analysis are included in Table 3. In this table, the reader can see how our work has achieved more than satisfactory results despite the challenging scenario presented. Our dataset, as shown in the introduction of this work, presents severe artifacts that seemingly completely obfuscates any relevant features present at some depths. This is shown particularly by the improvement of the results as more information is provided to the model. This is not only because the new points of view, but also thanks to the redundancy of information present in the rest of layers layers. Moreover, as shown in the previous confusion matrices, even in completely opposite depths the model is able to infer features to aid in the grading process.

These results confirm what we suggested in previous sections. The performance of the model that uses exclusively the information from the AP depth with an outstanding 0.8983 ± 0.0645 of AUC, proving the usefulness of this individual layer in the proposed grading task. Adding extra depth information adds only an increase of 0.03 of performance in all metrics, well inside the standard deviation of the AP-only model. Also, it is worth mentioning that only in the single depth analysis the favored depth was the 3×3 mm, while the other two experiments favored the 6×6 mm. As mentioned, we can infer that this is caused by the removal of small patterns and noise due to the same resolution of the images while capturing a bigger region, but this will be further studied in the following qualitative analysis stage made in collaboration with the expert clinicians. When we are only considering one of the depths, the model takes advantage of the most information it can from a single bidimensional projection of a region, and maintaining these features is critical to predict the structures at other depths to improve the grading performance. However, when including information about the adjacent layers, these small patterns become significantly less relevant (as this information is present in these extra layers). Thus, their removal greatly simplifies the problem and allows the model to focus on the inter-depth relationship and filter the intra-depth patterns.

As mentioned, this phenomenon seen in the AP depth and its unexpected performance is going to be studied further in the qualitative analysis of these models, where we will infer what structures present in the images allow for this performance despite (in theory) some of the types not considering

to a certain extent this depth (such as Type 1 and the Mixed type).

4.5 Analysis of the misclassifications

In this section we analyze the samples that the proposed methodology misclassified. One of the main causes of misclassifications are the severe artifacts present in the dataset (specially in the tests where we only consider one of the depths at the same time). In particular, in Fig. 12, we present a random assortment of samples for each individual ablation test, network and surface area analyzed. In them, we can note how in all these scenarios the normal structures of the retina are severely distorted. As we will present in following sections, the methodology is able to classify samples up to a certain level of disruption, and these scenarios are extreme. Moreover, this effect is diminished the more depths are considered (even considering that the artifacts tend to affect to similar regions along all depths).

When considering multiple (or all) depths, these artifacts become a lesser concern. Our proposal seems to be able to take advantage from the redundancy of information as well as continuity between depths to compensate for the data degradation. This way, most of the impact now falls on the extreme scenarios of each category where the number of samples is not sufficient. This is reflected in that most of the misclassifications belong to particular patients, instead of a randomized set like in the previous set. In Fig. 13, we present a random selection of misclassified samples per subject. As the reader can see, these represent scenarios for the different types of AMD that might be underrepresented in the dataset (note that, unlike in Fig. 12, all three depths belong to the same patient in each column).

4.6 Qualitative analysis of the attention results

Finally, in this last stage we analyzed some representative cases that would better illustrate the results attained by our grading proposal. In particular, we wanted to further understand why the AP of the retina was significant for a type of MNV that should not be present at that depth. In Figs. 14 and 15 we present representative examples at the relevant depths that explain this scenario.

In these two figures, we can see how, for the networks that attained the best results using exclusively OCTAs taken at the depth of the AP (in both cases, with the DenseNet 201, as shown in Section 4.1), the networks attention is focused on vascular structures present at AP depth. These structures, despite being at AP depth, are actually under the RPE layer, as they deformed the retinal layers but did not cross through. Thus, despite this MNV actually being technically Type 1, due to the deformation of the external retinal layers, its presence is shown at the depth of the AP. Additionally, this MNV

Table 3 Best results for each of the experiments based on the correlation between the results and the true classes (MCC)

Architecture	Surface	Depths	AUC	Accuracy	Precision	Recall	F1 Score	MCC
DenseNet 201	3x3	AP	0.8973 ± 0.0654	0.6752 ± 0.1456	0.7315 ± 0.1128	0.6841 ± 0.1468	0.6794 ± 0.1477	0.6035 ± 0.1798
DenseNet 161	6x6	AP & CP	0.9126 ± 0.0495	0.7141 ± 0.1269	0.7561 ± 0.1074	0.7168 ± 0.1270	0.7122 ± 0.1310	0.6529 ± 0.1520
DenseNet 201	6x6	DCP & AP & CP	0.9224 ± 0.0381	0.7318 ± 0.0825	0.7743 ± 0.0689	0.7352 ± 0.0811	0.7327 ± 0.0803	0.6739 ± 0.1012

AP: Avascular Plexus, CP: Choriocapillaris, DCP: Deep Capillary Plexus

Fig. 12 Random misclassified examples that present severe artifacts. All the presented samples are random, not necessarily belonging to the same patient or visit per row or column

	DenseNet 121		DenseNet 161		DenseNet 169		DenseNet 201	
	3x3	6x6	3x3	6x6	3x3	6x6	3x3	6x6
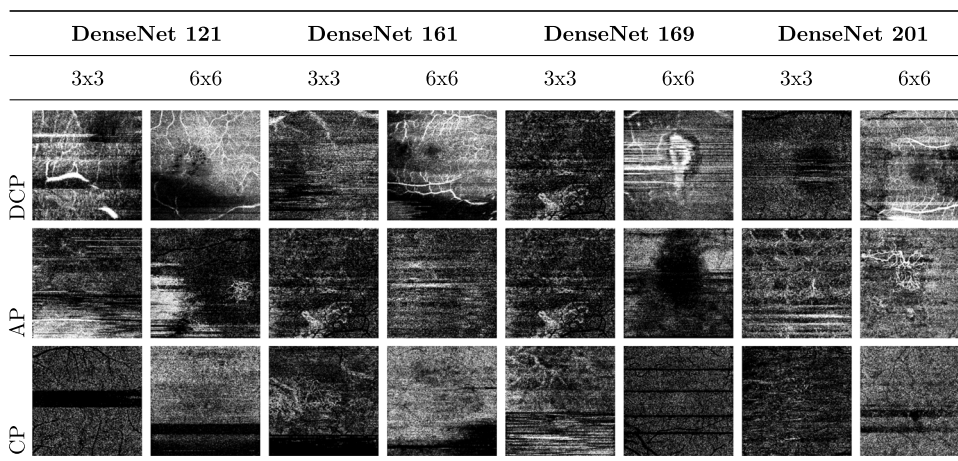
grows from the Bruch's membrane (the innermost layer of the choroid). Thus, this MNV is imperceptible mostly at this depth.

In both figures we also include the attention maps of the network that obtained the best results at the depth of the CP. Again, in both cases this corresponds to the DenseNet 161. As we can see (and, in contrast with the same depth of the DenseNet 201), there are actually some patterns present at this depth that hint the presence of Type 1 MNV despite not being clearly visible. In the case of the patient presented in Fig. 15, this is mostly focused on a particular vascular structure on the rightmost side of the image that could hint what has actually surfaced in innermost layers. On the other hand, in the patient of Fig. 14, no particular notable vascular structures have appeared on said membrane that could lead to believe that there is MNV; but the network has been able to infer its presence by the overall sparsity of the capillar structures at this depth.

In Fig. 16, we include a view from the capture device that confirms our explanation. In the cross-sectional view of these retinal layers (labeled by expert clinicians in the capture device) we can see the RPE layer marked in green. Under it, the vascular flow of the lesion is colored in red,

and the CP flow in purple. As the reader can see, there is a clear disturbance in the RPE layer that pushes over the CP line, facilitating said vascular artifacts to appear at depths they should not. Additionally, as shown by the depth maps in these images from both patients, this is prominent along all the retina, explaining the behavior shown in the attention maps.

The same way, in Fig. 17 we present a case with Type 2 MNV. In this particular case, the MNV should be visible (by definition, as mentioned in the introduction of this work) in both the AP and CP depths. In that figure, we present these two relevant depths with the network (DenseNet161) and surface analyzed size that returned the best results in the individual analysis (6×6 mm for the AP and 3×3 mm for the CP). As we can see in this case, the attention map is more concentrated in both scenarios. The MNV present at AP is clearly visible and attracts the attention of the network. This is the same scenario at CP level, where also the network is clearly focused on the region underneath the CP that presents a darkened pattern underneath the MNV at the AP depth. Moreover, despite attaining better results at different surface regions, both attention maps are consistent, showing that both cases reached the same approximate

Fig. 13 Random misclassified 6x6 samples from the DenseNet 201 network experiment trained with all three depths. Each column represents a random misclassified visit from the subject with the given ID in the dataset
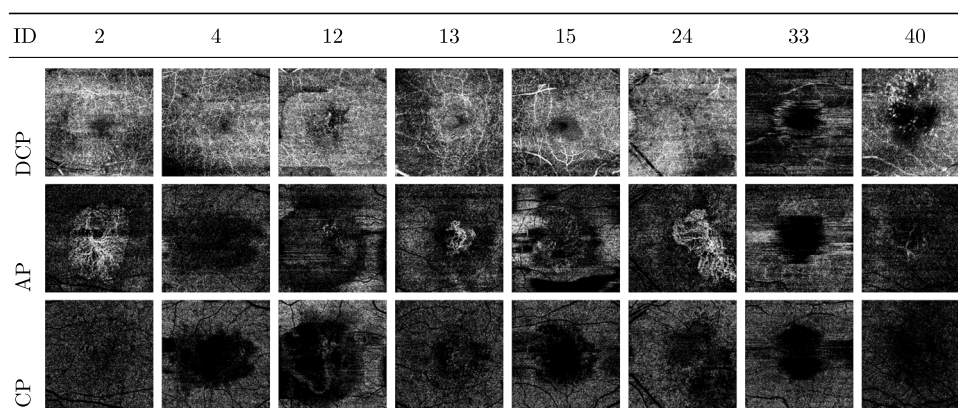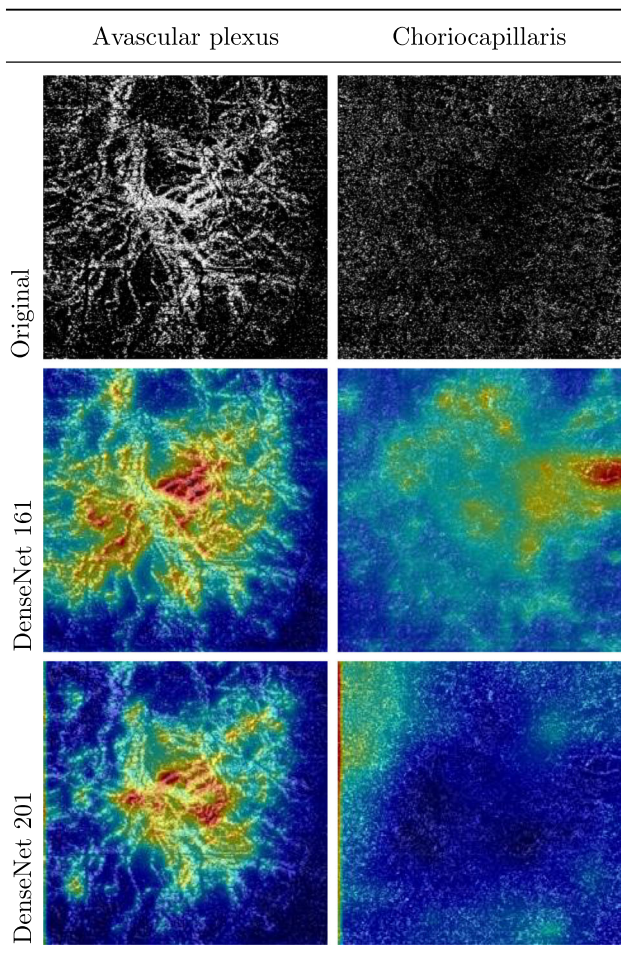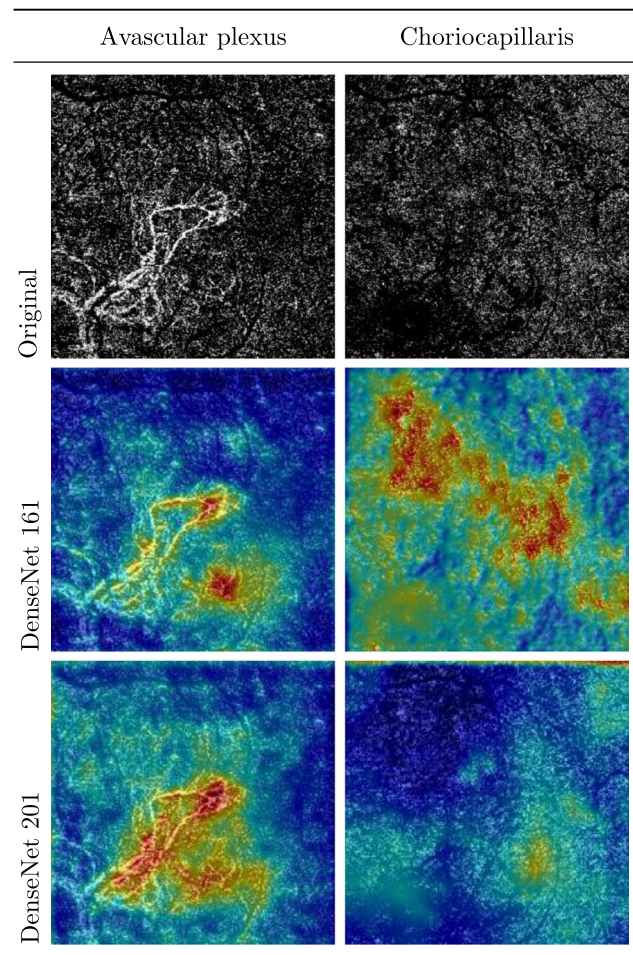

ID	2	4	12	13	15	24	33	40

Fig. 14 Attention maps generated for the Type 1 MNV generated for the images with a surface of 3×3 mm

Fig. 15 Attention maps generated for the Type 1 MNV generated for the images with a surface of 3×3 mm

conclusion on the relevance of the structures present in the OCTAs.

Finally, in Fig. 18 we present a patient with clear Type 3 MNV. In this scenario, we also include the DCP layer as, per definition of this type of MNV, the MNV patterns should also be noticeable in an OCTA at this depth (additionally, the models obtained satisfactory results identifying Type 3 MNV in the individual analysis when only using this depth). In this figure, as we did in previous scenarios, we included the cases of the individual analysis where the networks obtained the best performance at that depths and surface sizes. In this case, the DenseNet 161 for the analysis at DCP depth and the DenseNet 169 for both the AP and CP depths.

In the case of the CP depth, we see the same scenario presented in previous attention map analysis. The networks that perform better are the ones that measure the darkened surface present at the CP level, better so than the ones that consider the formation of pathological structures (such as the ones shown at the same depth with the DenseNet 161). This is a similar situation with what happens at the AP, where

also the model that outperforms the rest is the one focusing on particular darkened patches versus the models that focus on underlying vascular patterns. On the other hand, at DCP level the results show that the vascular formations are actually more relevant than the darkened regions, where no network has focused on them, but more on the lattice patterns with high lacunarity at this depth.

5 Conclusions

In this work, we presented a fully-automatic grading methodology, able to distinguish the four clinical stages of MNV in OCTA images. We performed an in-depth study of all the relevant layers for the disease, taking advantage of models with a range of complexities to understand their contribution to the issue. Furthermore, we studied this already challenging domain with a dataset comprised by images with severe capturing artifacts and with different stages of treatment, which would hinder the diagnosis process. Finally, we presented

Fig. 16 Images from the capture device interface for the patients of Figs. 14 (left) and 15 (right). Top-left corner of each image: layer depth. Top-right corner of each image: the OCTA view. Bottom panel: cross-sectional OCT scan
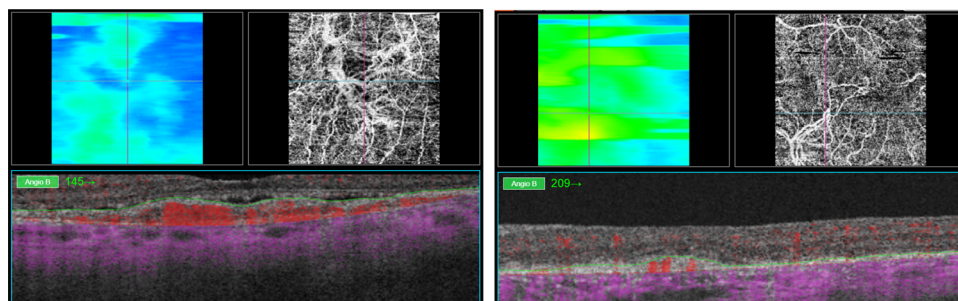
an in-depth qualitative study of the results in collaboration with expert clinicians, analyzing the attention maps of the trained networks to understand the behavior of the proposed methodology and improve the understanding of the target pathology.

We obtained more than satisfactory results even with the most limited approach. By only using the AP layer (which, in theory, is not relevant for some of the stages of the disease) we were able to obtain an 0.8937 ± 0.0654 AUC. The same way, with the complete model considering all the layers, the results improved to an 0.9224 ± 0.0381 AUC. In the qualitative anal-

ysis of the attention maps, we discovered that, both in the AP and the CP layers, vascular formations were present despite (in theory) not being defined by the type. In collaboration with the expert clinicians and the provided maps, we concluded that these vascular structures were under the proper layers for the given type (as they did not cross the limiting membranes), albeit the deformation they caused propagated to upper levels, allowing for its detection at the considered depths despite not being present in the layer. This same scenario occurred with the DCP layer, being able to determine the binary presence of the pathology despite being affected by definition only in latter stages of it. This way, our proposal is the first that considers the full grading of the MNV in OCTA at the three most relevant depths (allowing us to reveal the previously unconsidered line of research in the usage of layers outside the explicit definition of MNV types) and performs a complete qualitative analysis in collaboration with expert clinicians by means of explainable artificial intelligence strategies.

As future works, it would be interesting to consider also the different treatment stages available in the dataset. Some of the patients respond well to the treatment, while others show minimal to no response. A methodology able to predict or that takes into consideration this factor could help to better assess the features that contribute the most to each stage and type of MNV, minimizing (and exploring) the influence and bias of the pharmacological/surgical treatment. Additionally, we would like to further explore in the qualitative analysis the change in detected patterns when considering more than individual independent layers (as well as the implication of said changes with the experts in the domain).

Regarding the learning strategy and explainability techniques, it would be interesting to apply image classification strategies that would inherently include and favor explainability. As an example, we can consider prototype-based learning approaches such as ProtoPNet, which performs the classification based on a set of representative training prototypes. In this way, the classification is both justified and explained with selected explicit examples from the train set that can help to understand the classification process without the requirement of an extensive analysis of samples by
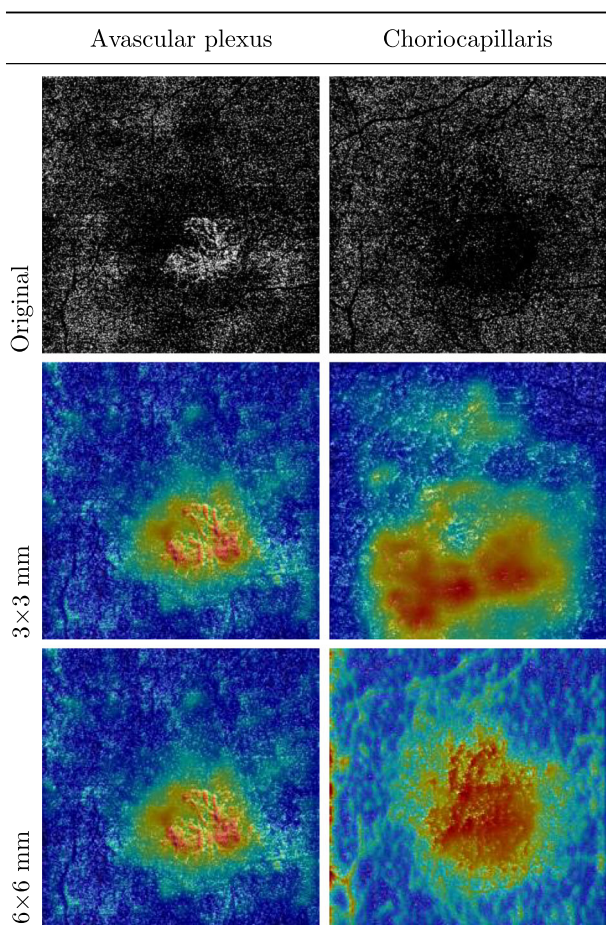

Fig. 17 Attention maps generated for the Type 2 MNV with the DenseNet 161

Fig. 18 Attention maps generated for the Type 3 MNV generated for the images with a surface of 3×3 mm
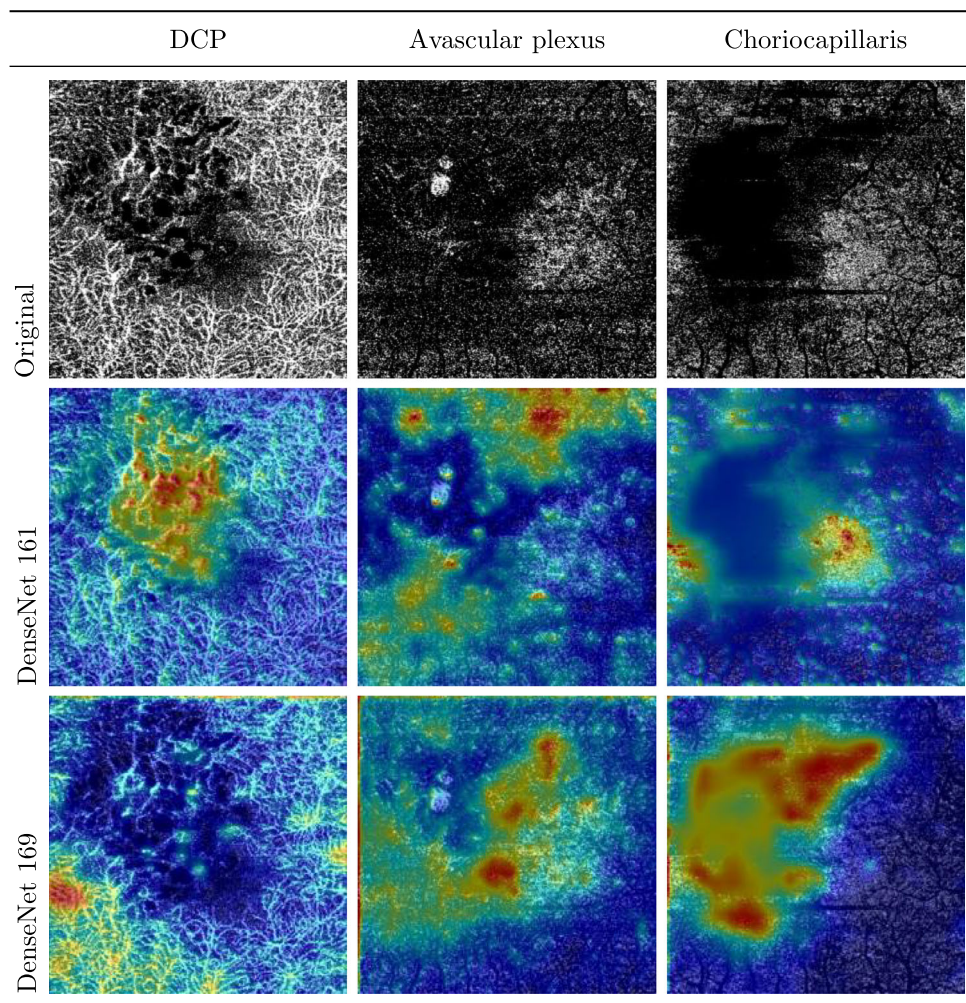
an expert in the domain. Moreover, this paradigm can help to filter problems related to the methodology itself, allowing for a more robust analysis of the results. Finally, it would also be interesting to combine this prototype based strategy with three-dimensional vision transformers or ViT. The retinal vascular tree expands itself along all the three axis of the studied cube. As shown, this is relevant specially in scenarios where artifacts severely distort the generated image, diminishing the impact of these artifacts the more layers were considered during training. Thus, these machine learning strategies that intrinsically consider both the spatial continuity and integrate attention mechanisms could greatly benefit our proposal and its explainability. This would also help with the aforementioned limitations of the network architecture used in this work, which, due to its high number of connections, overfitting may occur in datasets of significant size, being detrimental for future works with a higher number of samples.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval The protocols followed in this work were conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Investigation of Galicia (Comité Autonómico de Ética de la Investigación de Galicia) (2017/614) the 23rd of January, 2018.

Conflict of interest The authors have no financial or proprietary interests in any material discussed in this article.

References

1. Alfahaid A, Morris T (2018) An automated age-related macular degeneration classification based on local texture features in optical coherence tomography angiography. In: Communications in Computer and Information Science, Springer International Publishing, pp 189–200. https://doi.org/10.1007/978-3-319-95921-4_19

2. Aumann S, Donner S, Fischer J, Müller F (2019) Optical coherence tomography (OCT): Principle and technical realization. In: High Resolution Imaging in Microscopy and Ophthalmology, Springer International Publishing, pp 59–85. https://doi.org/10.1007/978-3-030-16638-0_3

3. Bansal R, Hemanth V, Mulkutkar S, Singh R, Gupta V, Dogra MR, Gupta A (2017) OCT angiography demonstrates retinal angiomatous proliferation and chorioretinal anastomosis of type 3 neovascularization. International Ophthalmology 38(5):2149–2151. https://doi.org/10.1007/s10792-017-0664-6

4. Barua PD, Chan WY, Dogan S, Baygin M, Tuncer T, Ciaccio EJ, Islam N, Cheong KH, Shahid ZS, Acharya UR (2021) Multilevel deep feature generation framework for automated detection of retinal abnormalities using OCT images. Entropy 23(12):1651. https://doi.org/10.3390/e23121651

5. Borrelli E, Bandello F, Souied EH, Barresi C, Miere A, Querques L, Sacconi R, Querques G (2022) Neovascular age-related macular degeneration: advancement in retinal imaging builds a bridge between histopathology and clinical findings. Graefe's Archive for Clinical and Experimental Ophthalmology. https://doi.org/10.1007/s00417-022-05577-x

6. Cabrera AP, Wolinsky EL, Mankad RN, Monickaraj F, Das A (2022) Pathophysiology of diabetic macular edema. In: Diabetic Macular Edema, Springer Nature Singapore, pp 7–25. https://doi.org/10.1007/978-981-19-7307-9_2

7. DaCosta J, Bhatia D, Talks J (2019) The use of optical coherence tomography angiography and optical coherence tomography to predict visual acuity in diabetic retinopathy. Eye 34(5):942–947. https://doi.org/10.1038/s41433-019-0606-9

8. de Jong EK, Geerlings MJ, den Hollander AI (2020) Chapter 10 - age-related macular degeneration. In: Gao XR (ed) Genetics and Genomics of Eye Disease, Academic Press, pp 155–180, 10.1016/B978-0-12-816222-4.00010-1, https://www.sciencedirect.com/science/article/pii/B9780128162224000101

9. De Oliveira PRC, Berger AR, Chow DR (2017) Optical coherence tomography angiography in chorioretinal disorders. Can J Ophthalmol 52(1):125–136. https://doi.org/10.1016/j.jcjo.2016.07.015'www.sciencedirect.com/science/article/pii/S0008418216300692'

10. Díaz M, Díez-Sotelo M, Gómez-Ulla F, Novo J, Penedo MFG, Ortega M (2019) Automatic visual acuity estimation by means of computational vascularity biomarkers using oct angiographies. MDPI AG 19:4732. https://doi.org/10.3390/s19214732

11. Greig EC, Duker JS, Waheed NK (2020) A practical guide to optical coherence tomography angiography interpretation. International Journal of Retina and Vitreous 6(1). https://doi.org/10.1186/s40942-020-00262-9

12. Guo Y, Hormel TT, Pi S, Wei X, Gao M, Morrison JC, Jia Y (2021) An end-to-end network for segmenting the vasculature of three retinal capillary plexuses from oct angiographic volumes. Biomed Opt Express 12(8):4889–4900. https://doi.org/10.1364/BOE.431888'opg.optica.org/boe/abstract.cfm?URI=boe-12-8-4889'

13. Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. https://arxiv.org/abs/1608.06993

14. Jin K, Yan Y, Chen M, Wang J, Pan X, Liu X, Liu M, Lou L, Wang Y, Ye J (2022) Multimodal deep learning with feature level fusion for identification of choroidal neovascularization activity in age-related macular degeneration. Acta Ophthalmologica 100(2):e512–e520. https://doi.org/10.1111/aos.14928'onlinelibrary.wiley.com/doi/abs/10.1111/aos.14928'

15. Karpecki PM (2015) Kanski's clinical ophthalmology. Optometry and Vision Science 92(10):e386. https://doi.org/10.1097/opx.0000000000000737

16. Li HY, Wang DX, Dong L, Wei WB (2022) Deep learning algorithms for detection of diabetic macular edema in OCT images: A systematic review and meta-analysis. Eur J Ophthalmol 33(1):278–290. https://doi.org/10.1177/11206721221094786

17. Liu L, Gao SS, Bailey ST, Huang D, Li D, Jia Y (2015) Automated choroidal neovascularization detection algorithm for optical coherence tomography angiography. Biomed Optics Express 6(9):3564. https://doi.org/10.1364/boe.6.003564

18. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, Curran Associates, Inc., pp 8024–8035, http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

19. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2022) Cross-entropy loss. https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html

20. Patel R, Wang J, Campbell JP, Kiang L, Lauer A, Flaxel C, Hwang T, Lujan B, Huang D, Bailey ST, Jia Y (2018) Classification of choroidal neovascularization using projection-resolved optical coherence tomographic angiography. Investigative Opthalmology

& Visual Science 59(10):4285. https://doi.org/10.1167/iovs.18-24624

21. Rahil M, Anoop BN, Girish GN, Kothari AR, Koolagudi SG, Rajan J (2023) A deep ensemble learning-based cnn architecture for multiclass retinal fluid segmentation in oct images. IEEE Access 11:17241–17251. https://doi.org/10.1109/ACCESS.2023.3244922

22. Rocholz R, Corvi F, Weichsel J, Schmidt S, Staurenghi G (2019) OCT Angiography (OCTA) in Retinal Diagnostics, Springer International Publishing, Cham, pp 135–160. https://doi.org/10.1007/978-3-030-16638-0_6

23. Rudnicka E, Napierała P, Podfigurna A, Meczekalski B, Smolarczyk R, Grymowicz M (2020) The world health organization (WHO) approach to healthy ageing. Maturitas 139:6–11. https://doi.org/10.1016/j.maturitas.2020.05.018

24. Samara WA, Shahlaee A, Adam MK, Khan MA, Chiang A, Maguire JI, Hsu J, Ho AC (2017) Quantification of diabetic macular ischemia using optical coherence tomography angiography and its relationship with visual acuity. Ophthalmology 124(2):235–244. https://doi.org/10.1016/j.ophtha.2016.10.008

25. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2019) Grad-CAM: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision 128(2):336–359. https://doi.org/10.1007/s11263-019-01228-7

26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2019b) Grad-CAM: Visual explanations from deep networks via gradient-based localization. https://doi.org/10.1007/s11263-019-01228-7

27. Spaide RF, Jaffe GJ, Sarraf D, Freund KB, Sadda SR, Staurenghi G, Waheed NK, Chakravarthy U, Rosenfeld PJ, Holz FG, Souied EH, Cohen SY, Querques G, Ohno-Matsui K, Boyer D, Gaudric A, Blodi B, Baumal CR, Li X, Coscas GJ, Brucker A, Singerman L, Luthert P, Schmitz-Valckenberg S, Schmidt-Erfurth U, Grossniklaus HE, Wilson DJ, Guymer R, Yannuzzi LA, Chew EY, Csaky K, Monés JM, Pauleikhoff D, Tadayoni R, Fujimoto J (2020) Consensus nomenclature for reporting neovascular age-related macular degeneration data. Ophthalmology 127(5):616–636. https://doi.org/10.1016/j.ophtha.2019.11.004

28. Stanga PE, Tsamis E, Papayannis A, Stringa F, Cole T, Jalil A (2016) Swept-source optical coherence tomography angio™ (topcon corp, japan): Technology review. In: Developments in Ophthalmology, S. Karger AG, pp 13–17. https://doi.org/10.1159/000442771

29. Stewart M, Browning D, Lee C (2018) Diabetic macular edema: Evidence-based management. Indian Journal of Ophthalmology 66(12):1736. https://doi.org/10.4103/ijo.ijo_1240_18

30. Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning. In: Dasgupta S, McAllester D (eds) Proceedings of the 30th International Conference on Machine Learning, PMLR, Atlanta, Georgia, USA, Proceedings of Machine Learning Research, vol 28, pp 1139–1147, https://proceedings.mlr.press/v28/sutskever13.html

31. Thakoor K, Bordbar D, Yao J, Moussa O, Chen R, Sajda P (2021) Hybrid 3d–2d deep learning for detection of neovascularagerelated macular degeneration using optical coherence tomography b-scans and angiography volumes. In, (2021) IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE. https://doi.org/10.1109/isbi48211.2021.9434111

32. Thakoor KA, Yao J, Bordbar D, Moussa O, Lin W, Sajda P, Chen RWS (2022) A multimodal deep learning system to distinguish late stages of AMD and to compare expert vs. AI ocular biomarkers. Scientific Reports 12(1). https://doi.org/10.1038/s41598-022-06273-w

33. Vidal P, de Moura J, Novo J, Ortega M (2023) Multivendor fully automatic uncertainty management approaches for the intuitive representation of DME fluid accumulations in OCT images. Medical and Biological Engineering and Computing. https://doi.org/10.1007/s11517-022-02765-z

34. Vidal PL, de Moura J, Díaz M, Novo J, Ortega M (2020) Diabetic macular edema characterization and visualization using optical coherence tomography images. Applied Sciences 10(21):7718. https://doi.org/10.3390/app10217718

35. Wang J, Hormel TT, Gao L, Zang P, Guo Y, Wang X, Bailey ST, Jia Y (2020) Automated diagnosis and segmentation of choroidal neovascularization in oct angiography using deep learning. Biomed Opt Express 11(2):927–944. https://doi.org/10.1364/BOE.379977'opg.optica.org/boe/abstract.cfm?URI=boe-11-2-927'

36. Wu J, Zhang Y, Xiao Z, Zhang F, Geng L (2022) Automated segmentation of diabetic macular edema in OCT b-scan images based on RCU-net. International Journal of Imaging Systems and Technology 33(1):299–311. https://doi.org/10.1002/ima.22788

37. Wu Q, Zhang B, Hu Y, Liu B, Cao D, Yang D, Peng Q, Zhong P, Zeng X, Xiao Y, Li C, Fang Y, Feng S, Huang M, Cai H, Yang X, Yu H (2021) Detection of morphologic patterns of Diabetic Macular Edema using a deep learning approach based on Optical Coherence Tomography images. Retina 41(5):1110–1117. https://doi.org/10.1097/iae.0000000000002992

38. Xue J, Yan S, Wang Y, Liu T, Qi F, Zhang H, Qiu C, Qu J, Liu X, Li D (2019) Unsupervised segmentation of choroidal neovascularization for optical coherence tomography angiography by grid tissue-like membrane systems. IEEE Access 7:143058–143066. https://doi.org/10.1109/access.2019.2943186

Authors and Affiliations

Plácido L. Vidal[1,2] · Joaquim de Moura[1,2] · Pablo Almuiña[4,5] · María Isabel Fernández[3,4,5] · Marcos Ortega[1,2] · Jorge Novo[1,2]

Plácido L. Vidal
placido.francisco.lizancos.vidal@udc.es

Pablo Almuiña
pablo.almuina@gmail.com

María Isabel Fernández
maribelfernandez@institutogomez-ulla.es

Marcos Ortega
mortega@udc.es

Jorge Novo
jnovo@udc.es

[1]   Centro de investigación CITIC, Universidade da Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain

[2]   Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, Xubias de Arriba, 84, 15006 A Coruña, Spain

[3]   Instituto Oftalmológico Gómez-Ulla, Santiago de Compostela, Galicia, Spain

[4]   Department of Ophthalmology, Complejo Hospitalario Universitario de Santiago de Compostela, Santiago de Compostela, Galicia, Spain

[5]   Department of Surgery and Medico-Surgical Specialties, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain