

# OCTA Multilayer and Multisector Peripapillary Microvascular Modeling for Diagnosing and Staging of Glaucoma

Danilo Andrade De Jesus<sup>1,\*</sup>, Luisa Sánchez Brea<sup>1,\*</sup>, João Barbosa Breda<sup>2-4</sup>, Ella Fokkinga<sup>1,5</sup>, Vera Ederveen<sup>1,5</sup>, Noor Borren<sup>1,5</sup>, Amerens Bekkers<sup>1,5</sup>, Michael Pircher<sup>6</sup>, Ingeborg Stalmans<sup>2,7</sup>, Stefan Klein<sup>1</sup>, and Theo van Walsum<sup>1</sup>

<sup>1</sup> Biomedical Imaging Group Rotterdam, Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands

<sup>2</sup> Research Group of Ophthalmology, Department of Neurosciences, KU Leuven, Leuven, Belgium

<sup>3</sup> Cardiovascular R&D Center, Faculty of Medicine of the University of Porto, Porto, Portugal

<sup>4</sup> Ophthalmology Department, Centro Hospitalar Sao João, Porto, Portugal

<sup>5</sup> Delft University of Technology, Delft, The Netherlands

<sup>6</sup> Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria

<sup>7</sup> Ophthalmology Department, University Hospitals Leuven, Leuven, Belgium

**Correspondence:** Danilo Andrade De Jesus, Dr. Molewaterplein 40, Na 25.12, 3015 GD Rotterdam, The Netherlands. e-mail: [d.andradedejesus@erasmusmc.nl](mailto:d.andradedejesus@erasmusmc.nl)

**Received:** October 25, 2019

**Accepted:** October 1, 2020

**Published:** November 5, 2020

**Keywords:** OCT angiography; microvascular density; glaucoma; machine learning

**Citation:** Andrade De Jesus D, Sánchez Brea L, Barbosa Breda J, Fokkinga E, Ederveen V, Borren N, Bekkers A, Pircher M, Stalmans I, Klein S, van Walsum T. OCTA multilayer and multisector peripapillary microvascular modeling for diagnosing and staging of glaucoma. *Trans Vis Sci Tech.* 2020;9(2):58, <https://doi.org/10.1167/tvst.9.2.58>

**Purpose:** To develop and assess an automatic procedure for classifying and staging glaucomatous vascular damage based on optical coherence tomography angiography (OCTA) imaging.

**Methods:** OCTA scans (Zeiss Cirrus 5000 HD-OCT) from a random eye of 39 healthy subjects and 82 glaucoma patients were used to develop a new classification algorithm based on multilayer and multisector information. The averaged circumpapillary retinal nerve fiber layer (RNFL) thickness was also collected. Three models, support vector machine (SVM), random forest (RF), and gradient boosting (xGB), were developed and optimized for classifying between healthy and glaucoma patients, primary open-angle glaucoma (POAG) and normal-tension glaucoma (NTG), and glaucoma severity groups.

**Results:** All the models, the SVM (area under the receiver operating characteristic [AUROC]  $0.89 \pm 0.06$ ), the RF (AUROC  $0.86 \pm 0.06$ ), and the xGB (AUROC  $0.85 \pm 0.07$ ), with 26, 22, and 29 vascular features obtained after feature selection, respectively, presented a similar performance to the RNFL thickness (AUROC  $0.85 \pm 0.06$ ) in classifying healthy and glaucoma patients. The superficial vascular plexus was the most informative layer with the infero temporal sector as the most discriminative region of interest. No significant differentiation was obtained in discriminating the POAG from the NTG group. The xGB model, after feature selection, presented the best performance in classifying the severity groups (AUROC  $0.76 \pm 0.06$ ), outperforming the RNFL (AUROC  $0.67 \pm 0.06$ ).

**Conclusions:** OCTA multilayer and multisector information has similar performance to RNFL for glaucoma diagnosis, but it has an added value for glaucoma severity classification, showing promising results for staging glaucoma progression.

**Translational Relevance:** OCTA, in its current stage, has the potential to be used in clinical practice as a complementary imaging technique in glaucoma management.

## Introduction

Glaucoma is the leading cause of irreversible blindness, affecting over 80 million people worldwide.<sup>1</sup> It is a chronic progressive optic neuropathy characterized by the thinning of the peripapillary retinal nerve fiber

layer (RNFL) and optic disc cupping as a result of axonal and retinal ganglion cell loss.<sup>2,3</sup> Glaucoma is an irreversible but preventable disease, which requires a management strategy involving risk stratification. Risk assessment is meant to properly allocate intensive treatment and monitoring to those who are more vulnerable while avoiding overburdening both patient

and health care system in cases when the disease is less likely to cause loss of visual function. However, the limited understanding of the disease's pathophysiology hampers effective risk stratification. Glaucoma assessment is currently based on a set of risk factors, previously identified in epidemiologic studies (e.g., RNFL thickness and intraocular pressure [IOP]), but with limited predictability for the individual patient.<sup>4-6</sup> Therefore, there is a need to identify additional risk factors/features that may contribute to improve this clinical decision process.

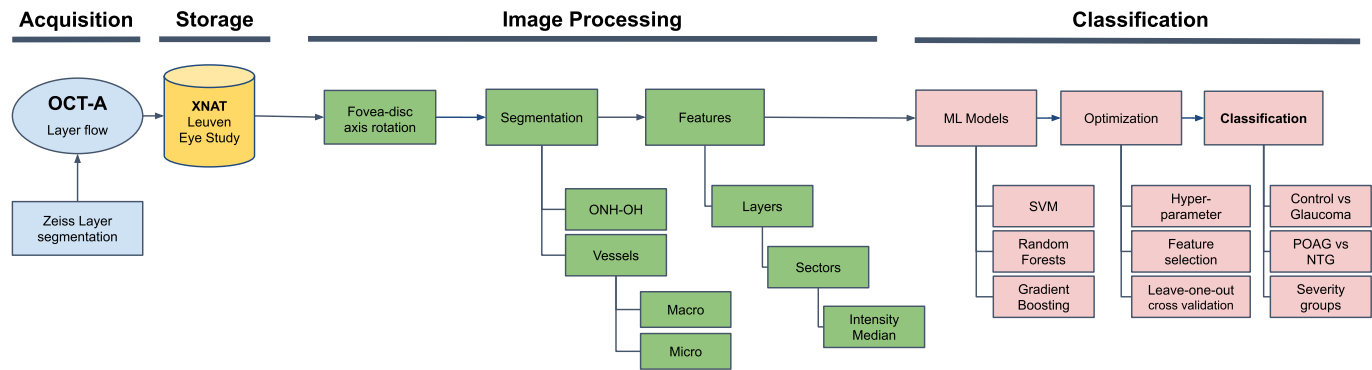
Differences in vascular parameters have been reported between glaucoma patients and healthy individuals, at ocular and systemic levels.<sup>7,8</sup> A number of techniques, such as fluorescein angiography,<sup>9</sup> laser speckle flowgraphy,<sup>10</sup> laser Doppler flowmetry,<sup>11</sup> and color Doppler imaging,<sup>12,13</sup> have been used for the evaluation of ocular and retinal blood perfusion. With the recent introduction of optical coherence tomography angiography (OCTA), standard OCT imaging devices are now capable of analyzing retinal vascular flow and to link it to a number of ocular diseases.<sup>10,14</sup> The application of OCTA to glaucoma has contributed to a more comprehensive assessment of the vascular supply role in the disease modulation. Significantly lower vessel density and blood flow index in the peripapillary area,<sup>15-19</sup> optic disc,<sup>15-24</sup> and macular area,<sup>15,17,18,23-25</sup> have been observed in glaucoma eyes in comparison with normal eyes. For all these areas, the diagnostic abilities of the imaged features increased with the glaucoma severity.<sup>23-26</sup> Moreover, it has been previously reported,<sup>27</sup> that vascular parameters increase the ability to discriminate between types of glaucoma (primary open-angle glaucoma [POAG] and normal-tension glaucoma [NTG]), as parameters linked to vascular dysfunction are more prominent in NTG patients.

Although three-dimensional OCTA information is generated from OCT imaging, most studies have only investigated the vascular density in the superficial layers (above the inner plexiform layer). Only a few studies considered the choroidal layers, including the choriocapillaris. Kiyota et al.,<sup>28</sup> found significantly lower OCTA parameters in the superficial choroid (0–70  $\mu\text{m}$  below Bruch's membrane, including the choriocapillaris) of glaucoma eyes in comparison to healthy controls but not in the deep choroid (70–140  $\mu\text{m}$  below Bruch's membrane). Two other studies reported no significant differences in choroidal results.<sup>29,30</sup> In addition to the analysis of all information available in a given layer, studies have been conducted discriminating different sectors. Andrade De Jesus et al.,<sup>19</sup> showed that the inferior and superior sectors of the peripapillary superficial layer present the most severe

vascular damage in glaucoma individuals compared to healthy controls. A similar study was performed by Lommatzsch et al.,<sup>31</sup> who observed that macular vessel density in both superficial and deep retinal vascular plexus in glaucomatous eyes was significantly lower than in healthy eyes, with the largest reduction found in the inferior macular sector. Rao et al.,<sup>23</sup> showed that the area under the receiver operating characteristic (AUROC) of the averaged peripapillary vessel density was significantly larger than the AUROC of the average inside the optic disc or at the macula, with the inferotemporal sector as its best discriminator.

Despite the increasing number of studies on OCTA features for glaucoma diagnosis, the data are frequently conflicting and/or arising from small-scale studies. Therefore, the current results have not allowed researchers to reach conclusions on the added value of a vascular analysis in clinical practice. The vascular glaucomatous damage in the retinal layers, choriocapillaris, and choroid, and the advantage of the vascular parameters compared to structural parameters such as the RNFL thickness or the intraocular pressure, need additional research to determine their contribution to the risk assessment and staging of glaucoma disease. Hence, the aim of this study is to further contribute to the understanding of the vascular role of glaucoma in the retinal layers, choriocapillaris, and choroid, looking at the information available at each layer and also at specific sectors within a layer. To that end, this article describes the design and optimization of classification models based on different layers and sectors. The models are compared with the aim to infer whether the information from multiple regions of interest (ROIs) has an added value for both the diagnostic accuracy and the discrimination between types of glaucoma (NTG and POAG). In summary, this article was designed to further contribute to the understanding of the following points:

- How good is a model based on multilayer and multisector (MLS) OCTA data at classifying between glaucoma patients and healthy subjects?
- Is there an added value of using OCTA imaging features in comparison to RNFL thickness?
- Can OCTA imaging features be used to discriminate between POAG and NTG?
- Can OCTA imaging be used to discriminate between different glaucoma severity levels?
- Which OCTA imaging features have the highest discriminant power in diagnosing and staging glaucoma disease?



**Figure 1.** Study pipeline structured into four parts (blue: acquisition; yellow: storage; green: image processing; red: classification). ONH-OH, optic nerve head optically hollow area; ML, machine learning; SVM, support vector machine; POAG, primary open-angle glaucoma; NTG, normal-tension glaucoma.

## Methods and Materials

The study pipeline was structured into four parts (acquisition, storage, image processing, and classification) as shown in Figure 1. In the first part, OCTA data were acquired by glaucoma specialists (J.B.B. and I.S.). Next, the data were exported to the Extensible Neuroimaging Archive Toolkit (XNAT) platform. Then, the imaged data were processed for masking out the spurious information and defining the different ROIs, in order to compute the respective microvascular density features. Lastly, two machine learning classification models were trained with the features obtained in the previous step. Feature selection was performed in order to gain insight into which features are the most important for risk stratification and disease staging.

### Image Acquisition

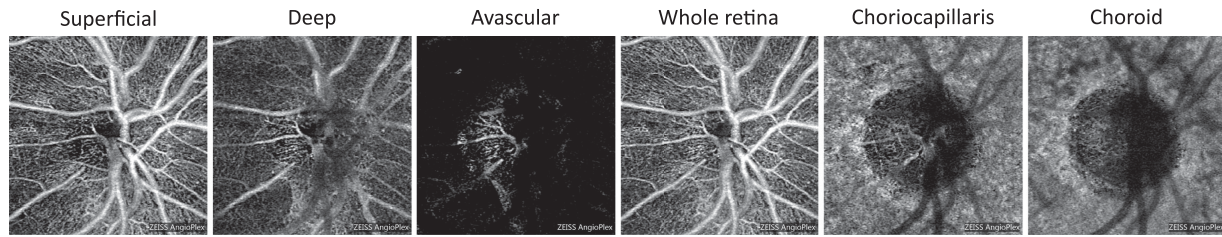
An OCTA data set selected from the Leuven Eye Study cohort,<sup>12</sup> was used in this study. Thirty-nine healthy subjects (aged  $63 \pm 13$  years) and 82 glaucoma patients (aged  $69 \pm 10$  years with an average visual field mean deviation [VF MD] of  $-8.1 \pm 6.7$  dB) were included, as shown in Table 1. The OCTA data in the Leuven Eye Study cohort consist of, for each subject, a  $3 \times 3$ -mm optic disc centered angiography scan acquired (via undilated pupil) using the Cirrus 5000 HD OCT (Carl Zeiss, Dublin, CA, USA; 10.0 software version). In this cohort, healthy individuals were recruited from those accompanying glaucoma patients. The exclusion criteria included blood relatives, those with a family history of glaucoma, rim notching or thinning, optic disc structural changes such as asymmetrical cup/disc ratio, disc hemorrhage, or an IOP above 21 mm Hg. Patients with glaucoma were

defined as having characteristic optic disc damage and visual field loss as defined elsewhere.<sup>32</sup> Glaucoma patients were excluded if they had a history of ocular trauma or any other eye disease (including high ametropias, defined as hyperopia higher than 4 diopters and myopia higher than 6 diopters). Additionally, patients with diabetes mellitus were excluded, since it is a known confounder in vascular-related research. Patients with open-angle glaucoma were stratified based on their maximal recorded untreated IOP as having POAG ( $> 21$  mm Hg) or NTG if equal or below that threshold. The eye with greater glaucomatous damage was chosen to be included in the study whenever both eyes were considered eligible. Six layers (superficial and deep vascular plexus, avascular, whole retina, choriocapillaris, and choroid) segmented by the manufacturer's software were exported (see Fig. 2). Although the avascular layer should not contain flow information, the glaucoma progression or the performance of the device's layer segmentation software may lead to microvasculature imaged in this layer, as seen in Figure 2. Therefore, the avascular layer was also included in order to maximize the information retrieved from the OCTA scan. Images with signal strength index below the suggested inclusion value provided by the manufacturer, 6 out of 10,<sup>33</sup> were excluded from the study. Images with severe movement artifacts or visible floaters were also excluded. In addition to the angiography scans, the averaged circumpapillary RNFL thicknesses were exported from the device. The glaucoma group was further divided by severity. Patients with a VF MD higher than  $-6$  dB were considered mild (37 subjects), between  $-6$  and  $-12$  dB were considered moderate (26 subjects), and those with a VF MD worse than  $-12$  dB were considered severe (19 subjects). Visual fields were obtained using the Humphrey Field Analyzer (HFA)

**Table 1.** Demographics and Characteristics of the Data Set for the Control and Glaucoma Groups, Type of Glaucoma, and Glaucoma Severity

	Control		Type of Glaucoma				Glaucoma Severity				p-Value
	39	82	Glaucoma	p Value	POAG	NTG	p-Value	Mild	Moderate	Severe	
Number of eyes	39	82			38	44		37	26	19	
Age (years)	62.2 ± 13.9	65.9 ± 8.9	0.10	0.10	64.4 ± 9.9	67.3 ± 7.8	0.20	66.3 ± 8.6	66.5 ± 7.7	64.6 ± 11.1	0.74
Gender (male/female)	24/15	41/41			18/20	23/21		13/24	17/9	11/8	
Eye (OD/OS)	20/19	47/35			17/21	30/14		19/18	16/10	12/7	
VF MD (dB)	0.38 ± 0.78	-7.8 ± 6.4	<0.001	<0.001	-8.6 ± 7.4	-7.8 ± 6.4	0.49	-2.4 ± 2.4	-8.6 ± 1.7	-17.1 ± 4.1	<0.001
RNFL (µm)	88.9 ± 6.4	67.1 ± 12.2	<0.001	<0.001	65.6 ± 12.2	68.4 ± 12.0	0.20	73.6 ± 11.8	63.3 ± 10.7	59.7 ± 7.2	<0.001
								73.6 ± 11.8	63.3 ± 10.7	59.7 ± 7.2	0.002
								73.6 ± 11.8	63.3 ± 10.7	59.7 ± 7.2	<0.001
S-IT (au)	72.8 ± 17.2	33.1 ± 20.6	<0.001	<0.001	33.5 ± 21.3	34.1 ± 19.7	0.8	47.1 ± 19.1	26.8 ± 16.1	17.5 ± 9.4	<0.001
								47.1 ± 19.1	26.8 ± 16.1	17.5 ± 9.4	<0.001
								47.1 ± 19.1	26.8 ± 16.1	17.5 ± 9.4	<0.001
								26.8 ± 16.1	26.8 ± 16.1	17.5 ± 9.4	0.034

OD, right eye; OS, left eye; Age, VF MD, RNFL thickness, and S-IT are presented as mean ± standard deviation; S-IT, superficial inferotemporal microvascular density; au, arbitrary units.



**Figure 2.** Optic disc centered peripapillary OCTA coronal projections of a healthy individual. From left to right: superficial, deep, avascular, whole retina, choriocapillaris, and choroid layers.

(Zeiss, Oberkochen, Germany) or the Octopus (Interzeag, Schlieren, Switzerland) perimeters on the same day as the OCT and OCTA examinations. The visual field programs were the 24-2 SITA standard program (HFA) or the G1 dynamic strategy (Octopus). The VF MD was extracted through a software that can extract data from both devices (Peridata; PeriData Software GmbH, Germany, version 3.5.4). Unreliable VFs (false positive >20%, false negative >40%, or fixation loss >30%) were excluded from analysis. These cutoffs are set by default by the manufacturer and widely used in previously published studies. The study adhered to the tenets of the Declaration of Helsinki, and it was approved by the Institutional Review Board of the University Hospitals Leuven. An informed consent was signed by all the participants prior to the study evaluation.

## Storage and Database

All the acquisitions were exported via Hyper-Text Transfer Protocol Secure (HTTPS) from a picture archiving and communication system (PACS) at University Hospitals Leuven to the Extensible Neuroimaging Archive Toolkit (XNAT,<sup>34</sup>) hosted within the Dutch national research infrastructure (Trait/Health-RI, <https://trait.health-ri.nl/trait-tools/xnat>). XNAT is an open-source platform for imaging-based research and clinical investigations, which manages access to different data sets compartmentalized into separate projects. All data were anonymized before being transferred to the XNAT.

## Image Processing

Python version 3.7, in combination with OpenCV,<sup>35</sup> and Scikit-image,<sup>36</sup> was used for the implementation of all the developed algorithms. The image processing was structured into three parts: fovea-disc axis correction, segmentation, and feature computation.

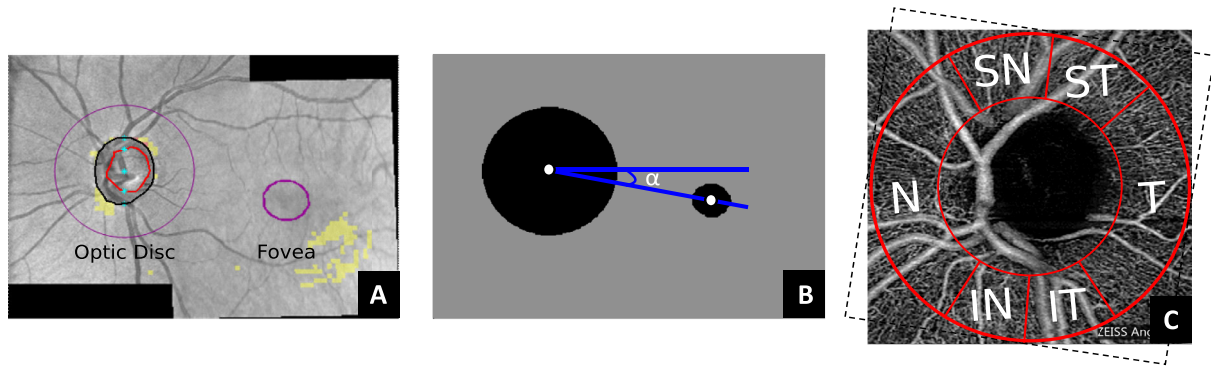
## Fovea-Disc Axis Correction

First, the OCTA images were rotationally corrected to a common reference. This was applied to ensure that the same area was being compared during the sectorial analysis, as there are differences in the head position during OCTA acquisition. In this work, the Panomap images provided by the device (see Fig. 3A) were used to correct the angle of the fovea-disc axis to zero degrees. The Panomap provides colored circular delineations of the fovea and optic disc, which were used to estimate the angle between the centroids of both circles (Fig. 3B). The mask used to extract the ROIs in the OCTA image (explained in the following subsections) was rotated according to the estimated fovea-disc axis angle ( $\alpha$ ) and taking into account if the eye was a left (OS) or right (OD) eye (Fig. 3C).

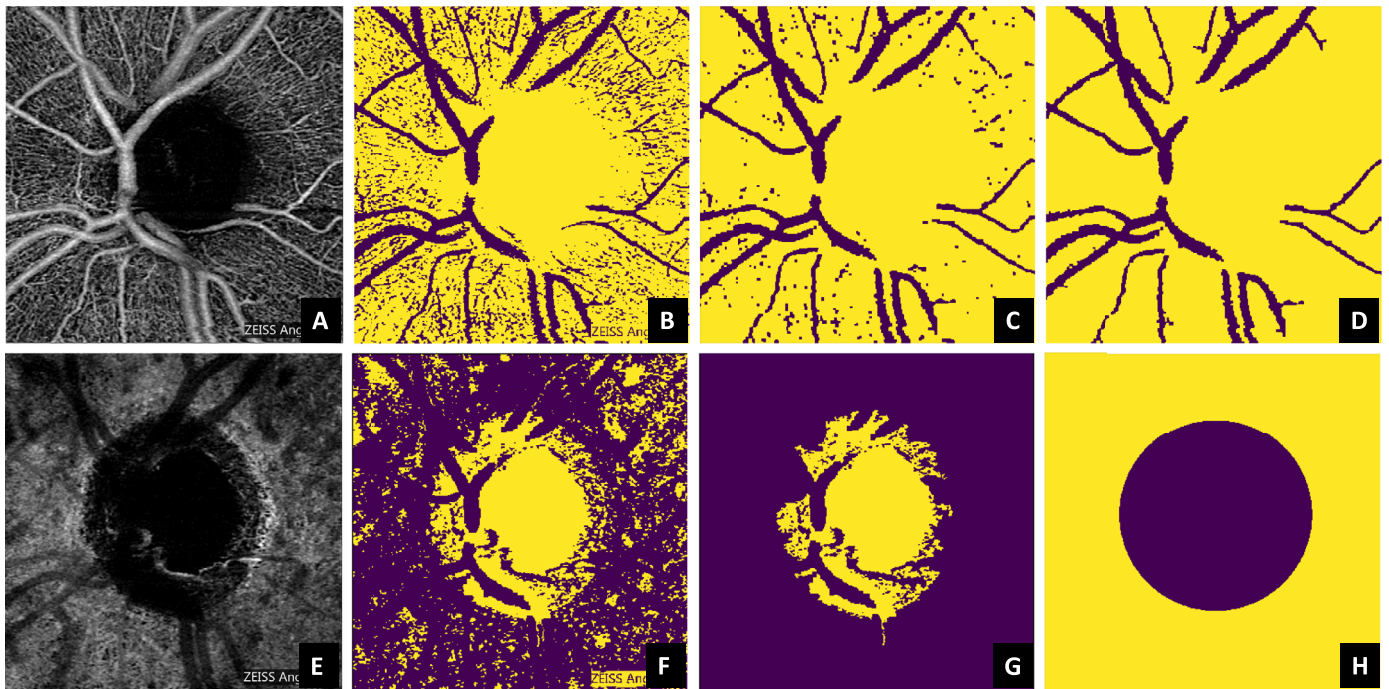
## Segmentation

Two main structures of interest were segmented in order to be masked out during the feature computation: the macrovasculature and the optic nerve head optically hollow area (ONH-OH). The superficial layer (Fig. 4A) was chosen as an anchor reference for obtaining a macrovascular mask for all the layers, due its higher contrast in comparison to the other ones. For achieving an approximate estimate of the macrovasculature, a binary image (Fig. 4B) was obtained from the superficial OCTA layer, based on a threshold set by the upper 88th percentile of the image intensity histogram. The selection of the percentile was done empirically, and it may depend on the data set. The definition of a relative threshold (in terms of a percentile), rather than an absolute value, makes the method more robust to the variations of intensity expected among individuals. A morphologic opening followed by a closing was then applied to denoise the binary image (Fig. 4C). Next, all connected components with an area less than 250 pixels were removed, ensuring that the remaining area was macrovasculature/background (Fig. 4D).

Regarding the ONH-OH, the choroid was used as anchor reference instead of the superficial layer. This is because, in some cases (Fig. 2), the center of the



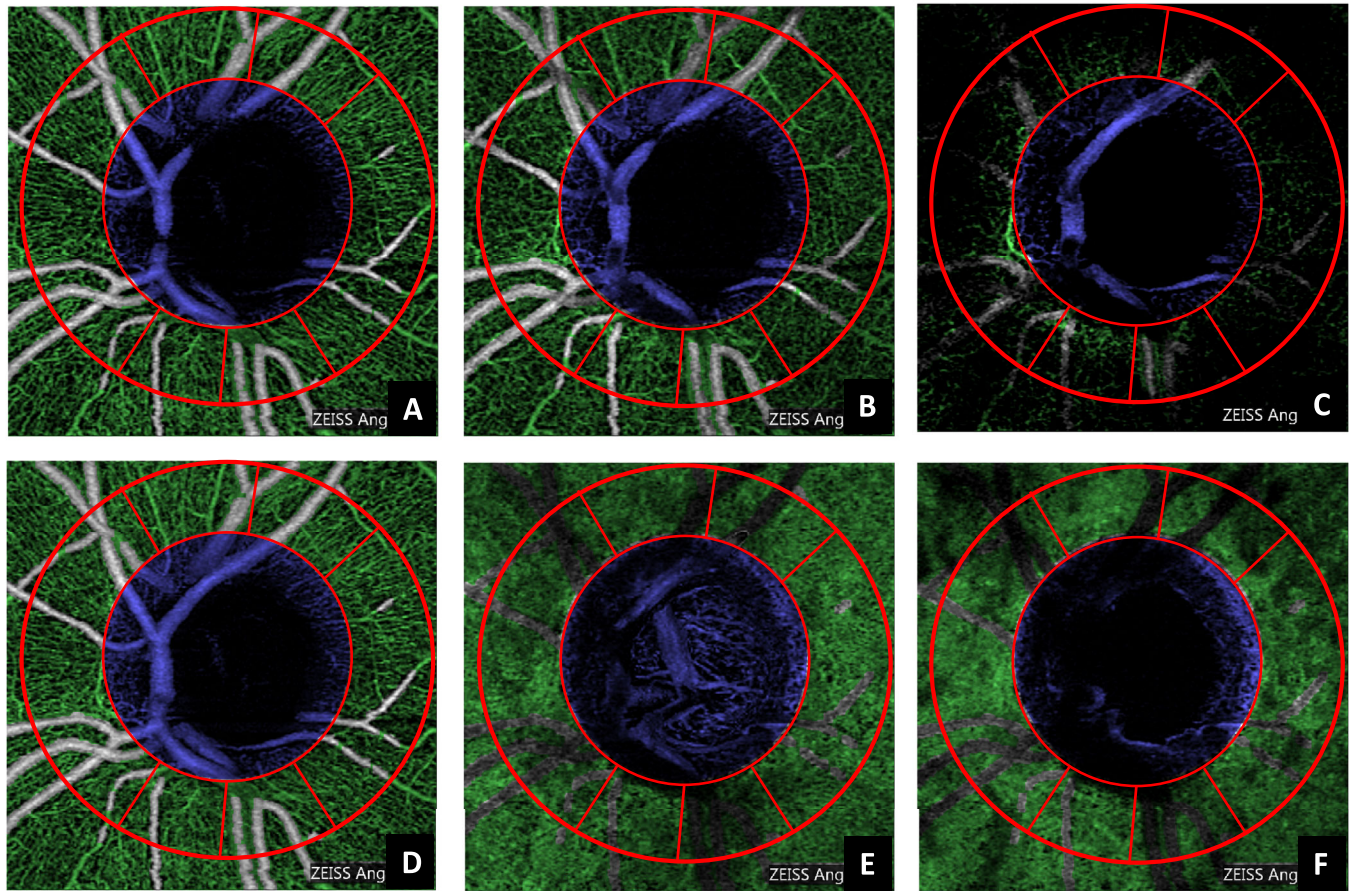
**Figure 3.** Example of a Panomap image used to estimate the fovea-disc axis angle (A). The binarization around the optic disc and the fovea, based on the predelineated areas by the device, is followed by the detection of the circles' centroids and the estimation of the rotation angle ( $\alpha$ ) (B). The superimposed mask on the OCTA image at the optic disc (C) is rotated according to  $\alpha$ . The Garway-Heath sectors nasal (N), inferonasal (IN), inferotemporal (IT), temporal (T), superotemporal (ST), and superonasal (SN) are delineated between the red lines.



**Figure 4.** Segmentation of the microvasculature (*upper row*) and the optic nerve head optically hollow area (*lower row*) based on the superficial vascular plexus and the choroid, respectively. The steps from the original image until the respective mask are denoted from the *left* to the *right* and succinctly explained in [Image Processing](#).

ONH-OH is partially covered by macrovasculature in the superficial layer, while the choroid offers a more robust alternative, as it provides a higher contrast between the ONH-OH and the vascularized areas (Fig. 4E). A binary image (Fig. 4F) was obtained from the choroidal OCTA image based on the lower 40th percentile of the equalized histogram. This percentile was empirically determined based on the data set characteristics. All connected components except the one with the largest area (estimated with Grana's

algorithm for eight-way connectivity,<sup>37</sup>) were removed from the image (Fig. 4G). Its location and respective area were used to estimate the centroid and the radius of a circle, which was used as a mask for the optically hollow area (Fig. 4H). Even if the dimension of this region is anatomically dependent and may vary with the disease progression, the shape remains similar, allowing this generalization. Figure 5 illustrates the automatic segmentation applied to all the layers considered in this study.



**Figure 5.** Automatic segmentation of the microvasculature (*green*) and the optic nerve head optically hollow area (*blue*) applied to all the layers included in this study (A, superficial vascular plexus; B, deep vascular plexus; C, avascular; D, whole retina; E, choriocapillaris; F, choroid). The *red lines* denote the Garway-Heath sectors already rotated according to the fovea-disc axis correction.

## Features

Once the vasculature and the optically hollow regions had been determined, the Garway-Heath map,<sup>38</sup> was used to set the ROIs according to seven sectors: superotemporal (ST), superonasal (SN), nasal (N), inferonasal (IN), inferotemporal (IT), temporal (T), and the circumpapillary (CP) — all sectors included. A circular mask centered at the ONH-OH was created, ensuring that the dimension of the sectors was the largest possible, while keeping the radius constant for all of them. The last step of the image processing was to retrieve the microvascular density measurements (features) from the ROIs based on the microvascular intensity median (VIM) within each ROI. The median was chosen instead of the mean because it is more robust to outliers in the sample set, and it is more applicable to not normally distributed data. Hence, 42 features were obtained based on the information retrieved from six layers and seven sectors

at the peripapillary region. For comparison purposes, the RNFL measure from the device was obtained for each patient. Additionally, the VIM measure for the IT sector of the superficial layer (S-IT) was used as a single-feature-based reference method, since it has been reported as the best glaucoma discriminator in the peripapillary region.<sup>23</sup>

## Classification

Python 3.7, in combination with Scikit-learn,<sup>39</sup> and Numpy,<sup>40</sup> was used for the classification step. Pandas,<sup>41</sup> was used for data manipulation, and Matplotlib<sup>42</sup> for obtaining the graphs and images for the results. The classification was organized into two parts: definition and training of the machine learning models, and statistical analysis and feature study, as described below.

### Machine Learning Models

Three classification models were used for the experiments: a support vector machine (SVM),<sup>43</sup> a random forest (RF),<sup>44</sup> and a gradient boosting classifier (xGB).<sup>45</sup> They were trained for three classification tasks: (1) differentiating glaucoma patients from controls, (2) distinguishing POAG and NTG types of glaucoma, and (3) identifying glaucoma severity levels (healthy, mild, moderate, and severe).

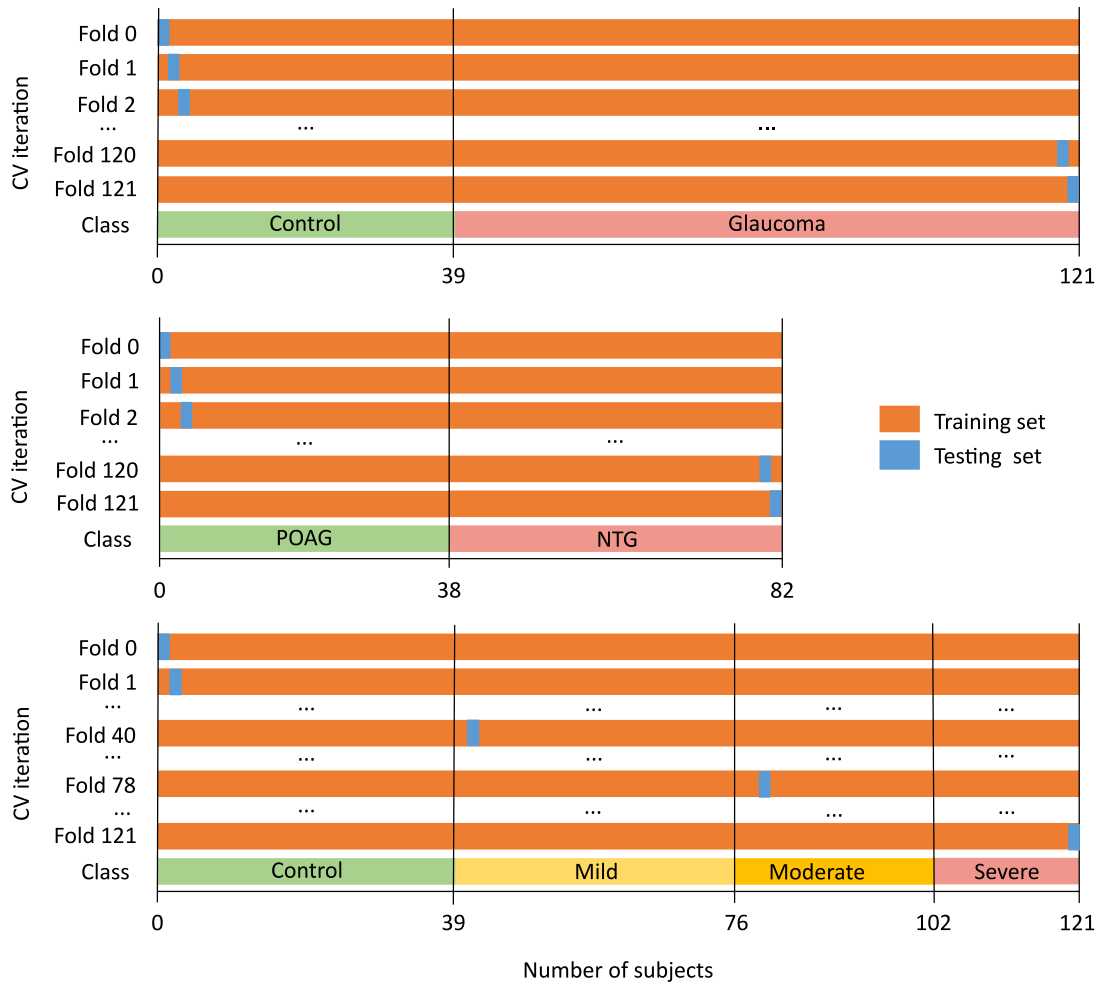
SVM was chosen because it typically provides robust classification performance even in case of small data sets and high-dimensional feature spaces. It is a supervised machine learning method that creates a hyperplane that separates the data points into two classes (negative and positive). The hyperplane is based on the maximal margin between classification points. The margin can be defined as the largest distance between nearest data points (the support vectors) and the hyperplane. Ideally, data points that are far apart are separable for classification, and that explains why the maximal margin is searched. The second model, RF, is a supervised machine learning model that is commonly used and relatively fast. It is an ensemble method that aggregates the individual predictions of decision trees into a combined prediction. RF creates several trees from subsets of the data, which are randomly selected. Each node in the tree represents a decision boundary that takes a certain amount of feature values into account. The features are also randomly divided as subsets for each tree. The obtained (random) trees are then merged to form the final class decision. The third model, xGB, is another supervised model that, like RF, consists of a set of decision trees. There are two main differences between these two models. First, RF builds each tree independently while gradient boosting uses the previous trees as a basis to build the subsequent ones. This additive model works in a forward stage wise manner, introducing a weak learner to improve the shortcomings of existing weak learners. Second, the RF combines results at the end of the process, while gradient boosting combines results along the way. For all models, one-versus-all strategy,<sup>46</sup> was used for differentiating the glaucoma severity levels. One-versus-all involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives.

Leave-one-out cross-validation was used to split the data in 121 folds. This approach leaves one subject out of the training data; that is, if there are  $n$  data points in the original sample, then  $n - 1$  samples are used to train the model and 1 point is used as the test set. Figure 6 summarizes the cross-validation scheme for each of the three classification tasks.

For each fold, hyperparameter optimization was applied on the training set. The hyperparameters of both classifiers were adjusted through grid search using stratified cross-validation,<sup>47</sup> (5-folds) to prevent overfitting. Hyperparameters are parameters with specified values per classification model. They influence the accuracy, complexity, and computational time of the model. The following parameters were considered for the SVM: kernel type (linear or radial basis function); penalty parameter  $C$  of the error term  $\in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1, 10, 10^2, 10^3\}$ ; gamma coefficient for the radial basis function kernel  $\in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ . The grid search for the RF model included the following parameters: number of trees in the forest  $\in \{10, 25, 50\}$ ; function to measure the quality of a split (Gini index or entropy); whether bootstrap samples are used when building trees (yes, no); percentage (%) of the minimum number of samples required to be at a leaf node  $\in \{10, 30, 50\}$ ; percentage (%) of the minimum number of samples required to split  $\in \{10, 30, 50\}$ ; maximum depth of the tree  $\in \{5, 10, 15, unlimited\}$ . The following parameters were included in the xGB grid search: minimum sum of instance weight (Hessian) needed in a child  $\in \{1, 5, 10\}$ ; step size shrinkage used in update to prevent overfitting (learning rate)  $\in \{0.01, 0.1, 0.25\}$ ; minimum loss reduction required to make a further partition on a leaf node of the tree  $\in \{0.1, 1, 3\}$ ; subsample ratio of the training instances  $\in \{0.6, 0.8, 1\}$ ; subsample ratio of columns when constructing each tree  $\in \{0.6, 0.8, 1\}$ ; percentage (%) of the minimum number of samples required to be at a leaf node  $\in \{10, 30, 50\}$ ; maximum depth of the tree  $\in \{5, 10, 15, unlimited\}$ .

In order to study the contribution of different features to the model and to improve the diagnostic accuracy, feature selection and hyperparameter tuning were applied on the training set. Ideally, including more features in a model should imply that a better outcome is achieved, as more information is available. However, this is not always the case, as some features may be redundant or irrelevant to discriminate glaucoma. These features may decrease the accuracy and increase training time. Hence, a feature selection algorithm based on univariate analysis of the features was applied. The method consisted of applying a statistical test to compare the discriminating power of each feature and hence retain the subset of the  $k$  features that presented the lowest  $P$  values. For the binary classification, the Mann-Whitney test's  $P$  value was computed between each feature and the class labels. The Mann-Whitney test was replaced by the Kruskal-Wallis test for the multiclass classification (healthy, mild, moderate, and severe). For each possible value





**Figure 6.** Leave-one-out cross-validation for differentiating glaucoma patients from controls (*first row*), distinguishing POAG and NTG types of glaucoma (*second row*), and identifying glaucoma severity levels (*third row*).

of  $k$ , the cross-validation accuracy estimate within the training set was used as a criterion to be maximized during the hyperparameter optimization. The optimal value of  $k$  was determined on the test set by selecting the value that maximized the average accuracy over the 121 outer cross-validation folds.

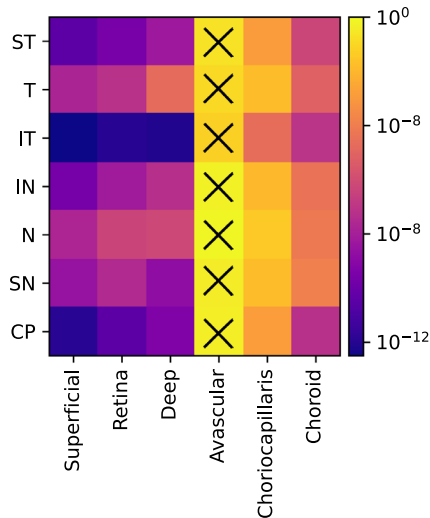
### Statistical Analysis and Feature Study

The normal distribution of the data on each studied group was assessed with the Kolmogorov-Smirnov test. The Mann-Whitney test was used to compare the statistical differences for all features between control and glaucoma and between POAG and NTG subjects. The severity levels of glaucoma were compared with the one-way analysis permutation test for all features followed by a post hoc analysis using the pairwise permutation test with false discovery rate adjustment. The AUROC, accuracy, sensitivity, and specificity of the classification between control and glaucoma subjects, as well as types of glaucoma, were computed

for each model. The 95% confidence interval of each metric was obtained with stratified bootstrap resampling.<sup>48</sup> The comparison between the ROC curves from the binary classification models was performed with the DeLong test,<sup>49,50</sup> whereas the comparison between contingency tables for the multiclass models was performed with the McNemar-Bowker test.<sup>51,52</sup> For all tests,  $P < 0.05$  was used to declare significance. The multiclass microaveraged AUROC and accuracy were used to assess the performance of the multiclass models (i.e., to classify between the four glaucoma severity levels: healthy, mild, moderate, and severe).<sup>39,53</sup>

## Results

The demographics and characteristics of all subjects included in the data set are listed in [Table 1](#). No statistically significant age differences were observed between



**Figure 7.** Mann-Whitney test,  $p$  value, for each vascular feature between control and glaucoma subjects. The X symbol denotes the groups that did not present a statistically significant difference ( $P > 0.05$ ).

the healthy control and the glaucoma groups ( $P = 0.1$ ), type of glaucoma ( $P = 0.2$ ), or severity groups ( $P = 0.74$ ). A statistically significant difference was observed for the VF MD between the control and glaucoma groups ( $P < 0.001$ ) and between the severity groups ( $P < 0.001$ ) but not between the POAG and NTG groups ( $P = 0.49$ ). The RNFL thickness was significantly different between the control and glaucoma groups but not between POAG and NTG groups ( $P = 0.2$ ). A significant difference was also observed between the three severity groups ( $P < 0.001$ ) for the RNFL. However, the post hoc analysis revealed that no significant differences existed between the moderate and severe groups ( $P = 0.23$ ). The inferotemporal microvascular density of the superficial layer was not significant to differentiate between POAG and NTG ( $P = 0.77$ ) but presented a statistically significant difference between control and glaucoma groups ( $P < 0.001$ ) and among severity groups ( $P < 0.001$ ), including between moderate and severe glaucoma ( $P = 0.034$ ).

## Glaucoma Classification

The Garway-Heath sectorial analysis of the six layers of the peripapillary region resulted in 42 microvascular features. Figure 7 shows the logarithm of the Mann-Whitney test,  $P$  value, between control and glaucoma subjects for each feature, color-coded by significance level. All features except the avascular layer for all sectors presented a statistically significant difference ( $P < 0.05$ ) between glaucoma and controls. The IT sector had the lowest  $P$  values, in agreement with

what has been reported in the literature. The results of the classification (AUROC, accuracy, sensitivity, and specificity) between glaucoma and healthy individuals before and after feature selection (according to the method described in Classification) are summarized in Table 2. For both models, a reduction of the number of features was observed (42 to 26, 22, and 29 for SVM, RF, and xGB, respectively). Figure 8 shows the accuracy as a function of the selected number of features ( $k$ ). The selected features based on the highest accuracy observed in Figure 8 are listed in Table 3. Although the optimal number of features differed between models, an increase of the accuracy can be observed in the three models when increasing the number of features. Since the MLS SVM after feature selection presented the highest score in AUROC and accuracy, it was considered the best model for glaucoma detection.

Further analysis of the ROC curves showed a statistically significant difference ( $P = 0.02$ ) between one single feature (S-IT SVM) and the model after feature selection (MLS SVM). However, no statistically significant difference was observed ( $P = 0.24$ ) between the best structural model (RNFL SVM) and the best vascular model (MLS SVM).

## Type of Glaucoma

The results of the classification between the two types of glaucoma, POAG and NTG, are shown in Table 4. The accuracy was maximized at selecting 4, 6, and 38 features for the MLS SVM, MLS RF, and MLS xGB model, respectively. However, in none of the cases, neither before nor after feature selection, a high discrimination ability was observed. The structural model (RNFL RF) presented the highest sensitivity, although, as shown in Table 1, no statistically significant difference existed between both groups (POAG versus NTG,  $P = 0.20$ ).

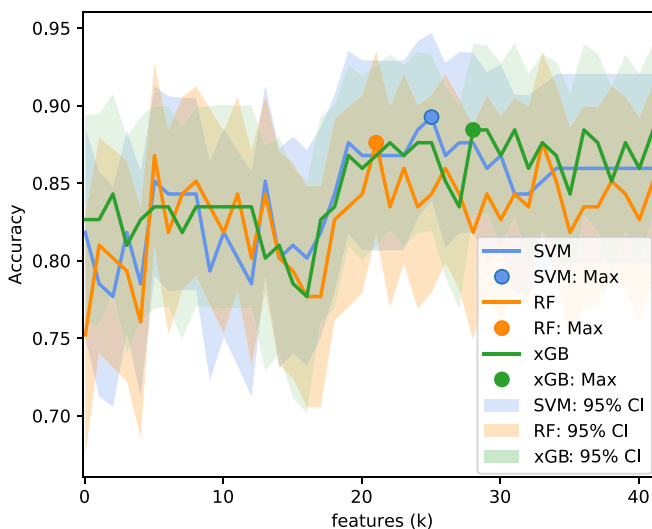
## Glaucoma Severity

Figure 9 shows the logarithm of the permutation test for one-way analysis,  $\log_{10}(P$  value), between glaucoma severity groups, including the healthy controls for each feature, color-coded by significance level. All features except the avascular layer (for all sectors) and the choriocapillaris (SN sector) presented a statistically significant difference ( $P < 0.05$ ) between groups. The inferotemporal (at the superficial, retina, and deep layers) and circumpapillary sectors (at the superficial layer) were the most discriminant for severity classification.

**Table 2.** AUROC, Accuracy, Sensitivity, and Specificity ( $\pm$  Confidence Interval at 95%) for the Vascular and Structural Parameters before and after Feature Selection for the Classification between Glaucoma and Healthy Individuals

Parameter	Model	No. of Features	Before Feature Selection			
			AUROC	Accuracy	Sensitivity	Specificity
Vascular	S-IT SVM	1	0.81 $\pm$ 0.07	0.82 $\pm$ 0.07	0.83 $\pm$ 0.08	0.79 $\pm$ 0.13
	S-IT RF	1	0.76 $\pm$ 0.08	0.75 $\pm$ 0.08	0.73 $\pm$ 0.10	0.79 $\pm$ 0.13
	S-IT xGB	1	0.81 $\pm$ 0.08	0.83 $\pm$ 0.07	0.87 $\pm$ 0.07	0.74 $\pm$ 0.14
	MLS SVM	42	0.87 $\pm$ 0.06	0.86 $\pm$ 0.06	0.84 $\pm$ 0.08	0.90 $\pm$ 0.09
	MLS RF	42	0.80 $\pm$ 0.08	0.81 $\pm$ 0.07	0.83 $\pm$ 0.08	0.77 $\pm$ 0.13
	MLS xGB	42	0.84 $\pm$ 0.07	0.86 $\pm$ 0.06	0.90 $\pm$ 0.06	0.77 $\pm$ 0.13
Structural	RNFL SVM	1	0.85 $\pm$ 0.06	0.82 $\pm$ 0.07	0.76 $\pm$ 0.09	<b>0.95 <math>\pm</math> 0.07</b>
	RNFL RF	1	0.82 $\pm$ 0.07	0.80 $\pm$ 0.07	0.77 $\pm$ 0.09	0.87 $\pm$ 0.11
	RNFL xGB	1	0.82 $\pm$ 0.07	0.82 $\pm$ 0.07	0.82 $\pm$ 0.08	0.82 $\pm$ 0.12
			After Feature Selection			
Vascular	MLS SVM	26	<b>0.89 <math>\pm</math> 0.06</b>	<b>0.90 <math>\pm</math> 0.05</b>	0.91 $\pm$ 0.06	0.87 $\pm$ 0.10
	MLS RF	22	0.86 $\pm$ 0.06	0.87 $\pm$ 0.06	0.88 $\pm$ 0.07	0.85 $\pm$ 0.11
	MLS xGB	29	0.85 $\pm$ 0.07	0.88 $\pm$ 0.06	<b>0.93 <math>\pm</math> 0.05</b>	0.77 $\pm$ 0.13

The highest mean value per metric is highlighted in bold.



**Figure 8.** Accuracy for the selected features for each subset size  $k$ . The same procedure was applied to all models, SVM (blue), RF (orange), and xGB (green), for classifying between the healthy control and glaucoma groups.

Table 5 summarizes the AUROC, accuracy, sensitivity, and specificity for all multiclass models before and after feature selection. The accuracy as a function of the number of features retrieved from the feature selection step, as well as the respective ranked features, is shown in Figure 10 and Table 6, respectively. In interpreting these results, please note that the multi-

class accuracy measure for this four-class classification scenario would be 0.25 in case of complete random guessing; a value of 0.5 to 0.6 therefore indicates good classification accuracy. Among the compared models, the MLS xGB after feature selection was the most discriminant for the multiclass comparison. The accuracy slightly improved after feature selection, although no statistically significant difference was observed according to the McNemar-Bowker test ( $P = 0.57$ ).

Figure 11 shows the confusion matrices for the best structural (RNFL xGB) and vascular (MSL xGB after feature selection) models. In both cases, the advanced stages of glaucoma were challenging to classify, with the RNFL xGB misclassifying all the severe cases. It can be observed from the confusion matrices that the RNFL xGB performed better for the healthy controls. However, the results of Glaucoma Classification showed that there was no statistical difference between the RNFL and the best vascular density model on classifying healthy and glaucoma subjects. The vascular model had a better performance discriminating between different glaucoma severity levels: the confusion matrices show that it was more likely to classify the right class but also that mistakes were more likely to happen in consecutive classes (i.e., mistake severe for moderate rather than mistake severe for mild). Moreover, a statistically significant difference was observed between models based on the

**Table 3.** List of Features for the SVM, RF, and xGB Models Used to Train the Final Multilayer and Multisector Models for Classifying Control and Glaucoma Subjects

SVM				RF				xGB			
Layer	Sector	$\log_{10}(p)$	Folds	Layer	Sector	$\log_{10}(p)$	Folds	Layer	Sector	$\log_{10}(p)$	Folds
Superficial	IT	-12.6	121	Superficial	IT	-12.6	121	Superficial	IT	-12.6	121
Deep	IT	-12.0	121	Deep	IT	-12.0	121	Deep	IT	-12.0	121
Retina	IT	-11.9	121	Retina	IT	-11.9	121	Retina	IT	-11.9	121
Superficial	CP	-11.8	121	Superficial	CP	-11.8	121	Superficial	CP	-11.8	121
Superficial	ST	-10.3	121	Superficial	ST	-10.3	121	Superficial	ST	-10.3	121
Retina	CP	-10.3	121	Retina	CP	-10.3	121	Retina	CP	-10.3	121
Superficial	IN	-9.5	121	Superficial	IN	-9.5	121	Superficial	IN	-9.5	121
Retina	ST	-9.4	121	Retina	ST	-9.4	121	Retina	ST	-9.4	121
Deep	CP	-9.2	121	Deep	CP	-9.2	121	Deep	CP	-9.2	121
Deep	SN	-8.7	121	Deep	SN	-8.7	121	Deep	SN	-8.7	121
Superficial	SN	-8.4	121	Superficial	SN	-8.4	121	Superficial	SN	-8.4	121
Deep	ST	-8.2	121	Deep	ST	-8.2	121	Deep	ST	-8.2	121
Retina	IN	-8.1	121	Retina	IN	-8.1	121	Retina	IN	-8.1	121
Superficial	T	-7.7	121	Superficial	T	-7.7	121	Superficial	T	-7.7	121
Superficial	N	-7.6	121	Superficial	N	-7.6	121	Superficial	N	-7.6	121
Retina	SN	-7.5	121	Retina	SN	-7.5	121	Retina	SN	-7.5	121
Deep	IN	-7.2	121	Deep	IN	-7.2	121	Deep	IN	-7.2	121
Choroid	CP	-7.2	121	Choroid	CP	-7.2	121	Choroid	CP	-7.2	121
Retina	T	-7.1	121	Retina	T	-7.1	121	Retina	T	-7.1	121
Choroid	IT	-7.0	121	Choroid	IT	-7.0	121	Choroid	IT	-7.0	121
Retina	N	-6.3	121	Retina	N	-6.3	120	Retina	N	-6.3	121
Choroid	ST	-6.2	121	Choroid	ST	-6.3	93	Choroid	ST	-6.2	121
Deep	N	-6.1	121	Deep	N	-6.3	29	Deep	N	-6.1	121
Choroid	T	-5.0	121					Choroid	T	-5.0	121
Deep	T	-4.7	116					Deep	T	-4.7	121
Choriocapillaris	IT	-4.6	110					Choriocapillaris	IT	-4.6	121
Choroid	IN	-4.6	15					Choroid	IN	-4.4	121
Choroid	SN	-4.3	1					Choroid	N	-4.1	121
								Choroid	SN	-3.9	121

The selected features are ranked from the top to the bottom with the  $\log_{10}(p)$  as the averaged logarithm value over the folds in which the feature was selected.

McNemar-Bowker test ( $P = 0.01$ ), confirming that, overall, the vascular information outperformed the structural information at discriminating different glaucoma severity groups.

## Discussion

A number of approaches and techniques have been published over the past two decades to infer and study vascular glaucomatous damage.<sup>26</sup> OCTA is a new submodality that has emerged from OCT, which measures changes in backscattered signal intensity or amplitude in order to differentiate areas of blood flow

from areas of static tissue. OCTA requires a very high sampling density in order to achieve the resolution needed to detect the tiny capillaries found in the retina. Its current limitations include inability to show leakage, proclivity for image artifacts due to patient movement/blinking,<sup>54</sup> and a relatively small field of view (e.g., 3 x 3 mm, 4.5 x 4.5 mm, and 6 x 6 mm). Hence, studies incident on glaucoma diagnosis and progression have been focused on specific regions of interest, namely the macula and the optic disc, including the peripapillary region. Richter et al.,<sup>55</sup> reported that, although the superficial vascular plexus micro-circulation in both macular and peripapillary regions is significantly reduced in glaucoma patients, global

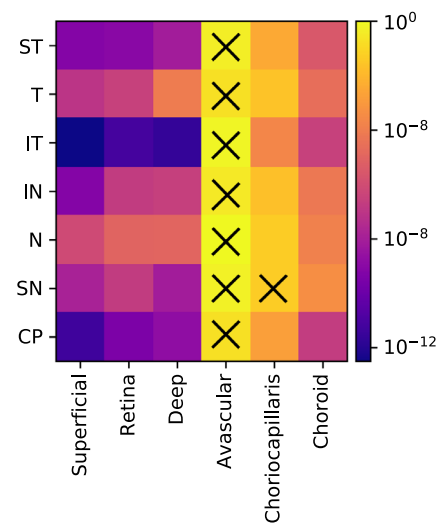
**Table 4.** AUROC, Accuracy, Sensitivity, and Specificity ( $\pm$  Confidence Interval at 95%) for the Vascular and Structural Parameters before and after Feature Selection for Classifying POAG and NTG Patients

Parameter	Model	No. of Features	Before Feature Selection			
			AUROC	Accuracy	Sensitivity	Specificity
Vascular	S-IT SVM	1	0.11 $\pm$ 0.07	0.11 $\pm$ 0.07	0.14 $\pm$ 0.10	0.08 $\pm$ 0.09
	S-IT RF	1	0.13 $\pm$ 0.07	0.12 $\pm$ 0.07	0.00 $\pm$ 0.00	0.26 $\pm$ 0.14
	S-IT xGB	1	0.32 $\pm$ 0.08	0.34 $\pm$ 0.11	0.59 $\pm$ 0.15	0.05 $\pm$ 0.07
	MLS SVM	42	0.49 $\pm$ 0.11	0.49 $\pm$ 0.11	0.43 $\pm$ 0.14	0.55 $\pm$ 0.16
	MLS RF	42	0.45 $\pm$ 0.10	0.45 $\pm$ 0.10	0.45 $\pm$ 0.14	0.45 $\pm$ 0.16
	MLS xGB	42	0.44 $\pm$ 0.11	0.45 $\pm$ 0.11	0.55 $\pm$ 0.15	0.34 $\pm$ 0.16
Structural	RNFL SVM	1	<b>0.61 <math>\pm</math> 0.11</b>	0.61 $\pm$ 0.11	0.59 $\pm$ 0.14	0.63 $\pm$ 0.15
	RNFL RF	1	<b>0.61 <math>\pm</math> 0.10</b>	<b>0.62 <math>\pm</math> 0.11</b>	<b>0.77 <math>\pm</math> 0.13</b>	0.45 $\pm$ 0.16
	RNFL xGB	1	0.49 $\pm$ 0.11	0.49 $\pm$ 0.11	0.52 $\pm$ 0.15	0.45 $\pm$ 0.16
After Feature Selection						
Vascular	MLS SVM	4	0.57 $\pm$ 0.10	0.56 $\pm$ 0.11	0.45 $\pm$ 0.15	<b>0.68 <math>\pm</math> 0.15</b>
	MLS RF	6	0.51 $\pm$ 0.11	0.51 $\pm$ 0.11	0.55 $\pm$ 0.15	0.47 $\pm$ 0.16
	MLS xGB	38	0.46 $\pm$ 0.11	0.46 $\pm$ 0.11	0.50 $\pm$ 0.15	0.42 $\pm$ 0.16

The highest mean value per metric is highlighted in bold.

peripapillary perfusion parameters outperform global macular perfusion parameters. The same observations were reported by Triolo et al.,<sup>56</sup> and Rao et al.,<sup>23</sup> which led to the conclusion that the peripapillary region should provide the highest discriminant power to classify glaucoma. Although the OCTA data, in their raw form, are provided in three dimensions, the two-dimensional data created from projections between specific depths are often used. Coronal projections of different layers (superficial vascular plexus,<sup>31,57,58</sup> deep vascular plexus,<sup>31,59,60</sup> whole retina,<sup>21,61</sup> choriocapillaris,<sup>61</sup> and choroid<sup>61,62</sup>) have been studied individually. However, to the best of our knowledge, no one has reported whether including information from multiple layers at the peripapillary region has an added value for the glaucoma classification.

Three classification models (SVM, RF, and xGB) were used in this work. All models are relatively fast and appropriate for small data sets and high-dimensional feature spaces. However, there are a number of other suitable options that could have been used in this work, such as KNN, or decision trees.<sup>63</sup> Besides conventional machine learning models, deep learning approaches can be used in classification tasks and have become the state of the art in several medical imaging applications. The main drawback when training a deep learning model is that it requires large amounts of data, which were not available in this study. Nevertheless, a convolutional neural network (CNN) model (VGG16,<sup>64</sup>) was explored to compare with the proposed methods in all three classification



**Figure 9.** Permutation test,  $p$  value, for each vascular feature between severity groups (healthy controls, mild, moderate, and severe glaucoma). The X symbol denotes the groups that did not present a statistically significant difference ( $P > 0.05$ ).

tasks. Pretrained weights and data augmentation were used to minimize the issues derived from the small data set size. Unfortunately, the CNN was not able to outperform the traditional classifiers in any of the cases. The implementation details and results of this approach are detailed in the Appendix.

In **Glaucoma Classification**, we show that combining the information from multiple layers and multiple sectors at the peripapillary region has an added

**Table 5.** AUROC and Accuracy ( $\pm$  Confidence Interval at 95%) for the Vascular and Structural Parameters before and after Feature Selection for Classifying Healthy Controls and Different Glaucoma Severity Levels (Mild, Moderate, and Severe)

Parameter	Model	No. of Features	Before Feature Selection			
			AUROC	Accuracy	Sensitivity	Specificity
Vascular	S-IT SVM	1	0.72 $\pm$ 0.06	0.58 $\pm$ 0.09	0.58 $\pm$ 0.09	0.86 $\pm$ 0.04
	S-IT RF	1	0.70 $\pm$ 0.06	0.55 $\pm$ 0.09	0.55 $\pm$ 0.09	0.84 $\pm$ 0.04
	S-IT xGB	1	0.72 $\pm$ 0.06	0.58 $\pm$ 0.09	0.58 $\pm$ 0.09	0.85 $\pm$ 0.04
	MLS SVM	42	0.67 $\pm$ 0.06	0.51 $\pm$ 0.09	0.51 $\pm$ 0.09	0.83 $\pm$ 0.04
	MLS RF	42	0.72 $\pm$ 0.06	0.60 $\pm$ 0.09	0.60 $\pm$ 0.09	0.85 $\pm$ 0.04
	MLS xGB	42	0.72 $\pm$ 0.06	0.59 $\pm$ 0.09	0.59 $\pm$ 0.09	0.85 $\pm$ 0.04
Structural	RNFL SVM	1	0.60 $\pm$ 0.06	0.40 $\pm$ 0.09	0.40 $\pm$ 0.09	0.79 $\pm$ 0.04
	RNFL RF	1	0.59 $\pm$ 0.06	0.39 $\pm$ 0.09	0.39 $\pm$ 0.09	0.80 $\pm$ 0.03
	RNFL xGB	1	0.67 $\pm$ 0.06	0.53 $\pm$ 0.09	0.53 $\pm$ 0.09	0.82 $\pm$ 0.04
			After Feature Selection			
Vascular	MLS SVM	21	0.74 $\pm$ 0.06	0.60 $\pm$ 0.09	0.60 $\pm$ 0.09	<b>0.87 <math>\pm</math> 0.04</b>
	MLS RF	24	0.69 $\pm$ 0.06	0.55 $\pm$ 0.09	0.55 $\pm$ 0.09	0.84 $\pm$ 0.04
	MLS xGB	39	<b>0.76 <math>\pm</math> 0.06</b>	<b>0.64 <math>\pm</math> 0.08</b>	<b>0.64 <math>\pm</math> 0.08</b>	<b>0.87 <math>\pm</math> 0.04</b>

The highest mean value per metric is highlighted in bold.

value for glaucoma classification, in comparison to using only the most discriminant region (superficial layer, inferotemporal sector). The comparison of healthy individuals with glaucoma patients showed that it is possible to achieve a classification accuracy using vascular information similar to that obtained with structural information (statistically nonsignificant differences). The MLS SVM obtained the best results among the compared models to differentiate healthy from glaucoma individuals, although the differences with the RNFL were not significant. The differences between using the most discriminant feature for microvascular analysis, the inferotemporal sector in the superficial layer, and using the feature selection subset were significant, showing that there is an added value of using different layers and sectors for glaucoma classification. However, it is difficult to conclude whether these differences are from the pathology itself or a consequence of the imaging artifacts. In general, deeper layers have not been considered in glaucoma analysis due to the difficulty in explaining the physical meaning of the imaged content. Deeper layers are influenced by light propagation through the overlying vessels, which cause the projection of intensity and phase fluctuations into deeper layers. These are known as shadowing artifacts. Although latest algorithms provided by the manufacture tend to minimize these influences, residual artifacts remain. From the univariate analysis, it is also possible to observe that the superficial is the most infor-

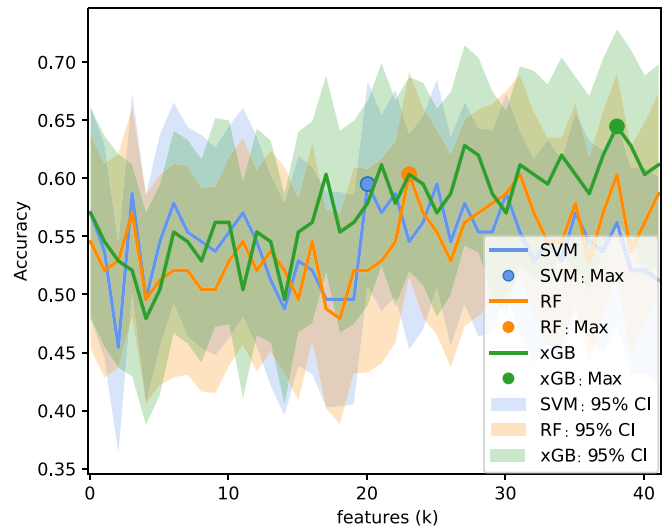
mative layer for discriminating glaucoma, in agreement with what has been reported in the literature. Moreover, looking at the sectors, it can be confirmed that the inferotemporal sector is the most affected by the disease, following also what has already been published about typical locations for optic nerve damage.<sup>65</sup> Overall, the obtained results regarding the most informative feature and accuracy of the models are in agreement with the latest analysis of the peripapillary region.<sup>23,31,66</sup>

Only a few studies have analyzed the differences between POAG and NTG. Bojikian et al.,<sup>67</sup> observed that the perfusion detected on the optic disc was significantly reduced in POAG and NTG groups compared to normal controls, but no difference was seen between POAG and NTG groups with similar levels of VF MD damage. In this study, an attempt to differentiate both groups was also made. However, the results show that, regardless of the type of parameters used (structural or vascular), the model (SVM, RF, or xGB), and the application or not of feature selection, no differentiation could be made in this data set. This can be due to the data set not being sufficiently large or representative to obtain conclusions, but it can also point to the impossibility of making a distinction between POAG and NTG based on a single, static OCTA analysis. Since many studies have shown that NTG patients are more prone to vascular dysregulation, it is possible that these differences are

clearer in a dynamic (rather than static) examination of ONH-OH/retinal vasculature, comparing OCTA data obtained through several acquisitions in the same patient.

The glaucoma group was subdivided into three groups according to the severity (mild, moderate, and severe). The MLS xGB model achieved the best outcomes for accuracy in the multiclass analysis. From the multiclass confusion matrices in Figure 11, it was possible to observe that the best vascular model classifies the severity level correctly more often than the structural model, especially in the intermediate levels. The poorer performance of the structural parameters compared to the vascular analysis may be explained by the floor effect in the RNFL as the disease progresses. Structural parameters measured by OCT reach a base level beyond which little change was seen with increasing severity of glaucoma.<sup>68</sup> The average value generally lies between 50% and 70% of the RNFL in normal eyes.<sup>69-71</sup>

Despite the results observed in this study, further research needs to be done in order to reduce the image variability due to the image acquisition and post-processing. Since there is not a clear boundary that can be applied to differentiate micro- from macrovasculature in OCTA imaging, different image-processing approaches may lead to different microvascular interpretations within the same data. It is a subjective definition that may be interpreted differently depending on the image operator. A popular method used to extract the microvascular density is to generate a binarization from the OCTA image, based on thresholding techniques.<sup>72</sup> The ratio of white or black pixels over a specific area is used to estimate the microvascular dropout. In general, the threshold is chosen based on an empirical analysis using general-purpose image-processing programs such as ImageJ.<sup>73</sup> These binarization approaches lead to results that are valid for a specific method, which ultimately hinder comparisons between different studies, and preclude the development of a robust classification model. In this study, it is assumed that the OCTA images exported from a single device are comparable between subjects and, hence, the median serves as a robust measure to estimate the microvascular density within a ROI, provided that the microvasculature is properly segmented. The separation of micro- from macrovasculature is another source of variability between studies. In some studies, the macrovasculature is segmented and extracted from the region of interest. Other authors have opted for estimating the vascular density based on all the information presented on the OCTA image. Since the macrovasculature is not expected to be affected by glaucoma and is a subject-dependent anatomic feature,



**Figure 10.** Accuracy for the selected features for each subset size  $k$ . The same procedure was applied to all models, SVM (blue), RF (orange), and xGB (green), for classifying the severity groups.

an analysis including the macrovasculature based on image pixel intensity is not desirable, as it may bias the results.<sup>74</sup> Similarly, the size of the ONH-OH is subject dependent and does not provide relevant information. Therefore, it is desirable to segment and exclude these areas from the ROI before the microvascular density estimation is done. Deep learning segmentation approaches could eventually reduce the current variability given their proven efficiency in vascular segmentation in fundus photography and scanning laser ophthalmoscopy imaging.<sup>75</sup> All of these adjustments should be routinely done to guarantee more accurate (and comparable) results when analyzing the role of vascular parameters in glaucoma with OCTA.

Although this study hints that microvascular density outperforms the RNFL thickness in multiclass classification of glaucoma severity, further work on larger cohorts is needed in order to validate this hypothesis. As seen in Figures 8 and 10, the accuracy as a function of the number of features (based on the entire data set) was fluctuating substantially, and there was no clear and robust maximum in the graph. Thus, the accuracies reported after feature selection might be a bit optimistic. Future studies should also evaluate the predictive value of RNFL when using the same Garway-Heath sectors corrected for the fovea-disc axis. In addition, the prediction ability of vascular and structural information using longitudinal data should be analyzed, as well as the design of new models combining multisectorial data from both types of parameters.

**Table 6.** List of Features for the SVM and RF Models Used to Train the Final Multilayer and Multisector Models for Classifying the Severity Groups Including the Healthy Controls

SVM				RF				xGB			
Layer	Sector	$\log_{10}(p)$	Folds	Layer	Sector	$\log_{10}(p)$	Folds	Layer	Sector	$\log_{10}(p)$	Folds
Superficial	IT	-14.9	121	Superficial	IT	-14.9	121	Superficial	IT	-14.9	121
Retina	IT	-14.0	121	Retina	IT	-14.0	121	Retina	IT	-14.0	121
Deep	IT	-13.9	121	Deep	IT	-13.9	121	Deep	IT	-13.9	121
Superficial	CP	-13.7	121	Superficial	CP	-13.7	121	Superficial	CP	-13.7	121
Superficial	IN	-12.5	121	Superficial	IN	-12.5	121	Superficial	IN	-12.5	121
Superficial	ST	-11.7	121	Superficial	ST	-11.7	121	Superficial	ST	-11.7	121
Retina	CP	-11.6	121	Retina	CP	-11.6	121	Retina	CP	-11.6	121
Retina	IN	-11.0	121	Retina	IN	-11.0	121	Retina	IN	-11.0	121
Deep	CP	-10.9	121	Deep	CP	-10.9	121	Deep	CP	-10.9	121
Deep	IN	-10.7	121	Deep	IN	-10.7	121	Deep	IN	-10.7	121
Retina	ST	-10.1	121	Retina	ST	-10.1	121	Retina	ST	-10.1	121
Deep	ST	-10.1	121	Deep	ST	-10.1	121	Deep	ST	-10.1	121
Deep	SN	-9.7	121	Deep	SN	-9.7	121	Deep	SN	-9.7	121
Superficial	SN	-9.6	121	Superficial	SN	-9.6	121	Superficial	SN	-9.6	121
Superficial	N	-8.8	121	Superficial	N	-8.8	121	Superficial	N	-8.8	121
Superficial	T	-8.5	121	Superficial	T	-8.5	121	Superficial	T	-8.5	121
Retina	SN	-8.3	121	Retina	SN	-8.3	121	Retina	SN	-8.3	121
Deep	N	-7.1	121	Deep	N	-7.1	121	Deep	N	-7.1	121
Retina	N	-7.1	121	Retina	N	-7.1	121	Retina	N	-7.1	121
Retina	T	-7.1	121	Retina	T	-7.1	121	Retina	T	-7.1	121
Choroid	IT	-6.3	120	Choroid	IT	-6.3	121	Choroid	IT	-6.3	121
Choroid	CP	-6.3	1	Choroid	CP	-5.9	121	Choroid	CP	-5.9	121
				Choriocapillaris	IT	-5.0	121	Choriocapillaris	IT	-5.0	121
				Choroid	ST	-4.9	121	Choroid	ST	-4.9	121
								Deep	T	-4.3	121
								Choroid	T	-4.2	121
								Choroid	IN	-3.1	121
								Choroid	N	-3.0	121
								Choroid	SN	-3.0	121
								Choriocapillaris	CP	-2.1	121
								Choriocapillaris	ST	-1.9	121
								Choriocapillaris	T	-1.7	121
								Choriocapillaris	IN	-1.2	121
								Choriocapillaris	SN	-1.2	121
								Choriocapillaris	N	-0.9	121
								Avascular	IT	-0.9	121
								Avascular	CP	-0.5	120
								Avascular	IN	-0.4	117
								Avascular	T	-0.3	115
								Avascular	ST	-0.3	10
								Avascular	SN	-0.3	1

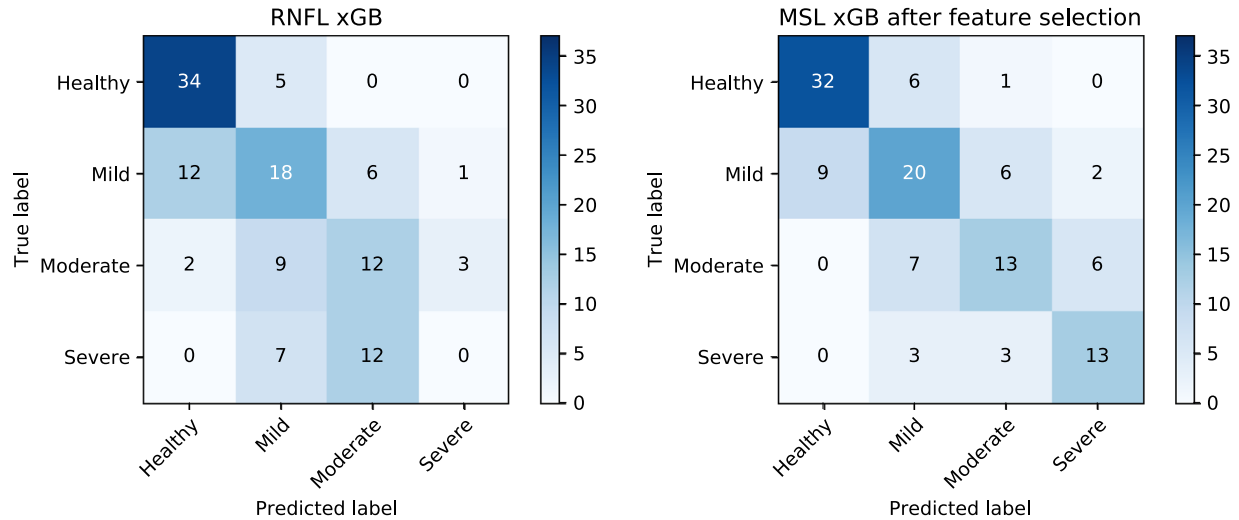
The selected features are ranked from the top to the bottom with the  $\log_{10}(p)$  as the averaged logarithm value over the folds in which the feature was selected.

## Conclusion

In order to ensure individualized health care for glaucoma patients, taking into account the disease

and severity, it is necessary to develop models and software that are able to integrate the increasingly complex web of risk factors and parameters associated with the disease. Hence, studies in glaucoma research must expand beyond the established risk factors and





**Figure 11.** Confusion matrices for each severity group according to the best structural (RNFL xGB model, *left*) and the best vascular model (MSL xGB model after feature selection, *right*).

explore the contribution of novel technologies. In this sense, the inclusion of image-based parameters, such as measurements derived from OCT and OCTA, is particularly relevant, since these have high reproducibility and are largely operator independent. In line with this research goal, this study assesses the effect of combining vascular information of different retinal and choroidal layers to improve the prediction and staging of glaucoma disease. The results show that, although the OCTA superficial vascular plexus is the most informative at discriminating healthy subjects from glaucoma patients, and also between glaucoma severity levels, there is an added value of including multilayer and multisector information in a classification model, instead of restricting the information to the most discriminative region. Moreover, a combination of multilayer and multisector microvascular information seems to yield a higher discriminative power than the circumpapillary RNFL thickness to discriminate between severity levels of glaucoma. Nevertheless, further studies must be done to better understand and validate the role of deeper layers such as choriocapillaris and choroid in glaucomatous vascular damage, as well as to reduce the bias due to image acquisition and postprocessing. In addition, multisectorial RNFL data and larger cohorts should also be considered in future research.

## Acknowledgments

Supported by the Horizon 2020 research and innovation program (grant 780989: multi-modal,

multi-scale retinal imaging project). The results in this study were partially presented and discussed at ARVO Imaging Eye Conference, Vancouver, Canada.<sup>66</sup>

Disclosure: **D. Andrade De Jesus**, None; **L.S. Brea**, None; **J.B. Breda**, None; **E. Fokkinga**, None; **V. Ederveen**, None; **N. Borren**, None; **A. Bekkers**, None; **M. Pircher**, None; **I. Stalmans**, None; **S. Klein**, None; **T. van Walsum**, None

\* DADJ and LSB contributed equally to this work.

## References

1. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081–2090.
2. Güngör SG, Ahmet A. Are all retinal nerve fiber layer defects on optic coherence tomography glaucomatous? *Turkish Ophthalmol*. 2017;47:267.
3. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311:1901–1911.
4. Coleman AL, Miglior S. Risk factors for glaucoma onset and progression. *Surv Ophthalmol*. 2008;53:S3–S10.
5. Cook C, Foster P. Epidemiology of glaucoma: what's new? *Can Ophthalmol*. 2012;47:223–226.
6. Jesus DA, Majewska M, Krzyzanowska-Berkowska P, Iskander DR. Influence of eye

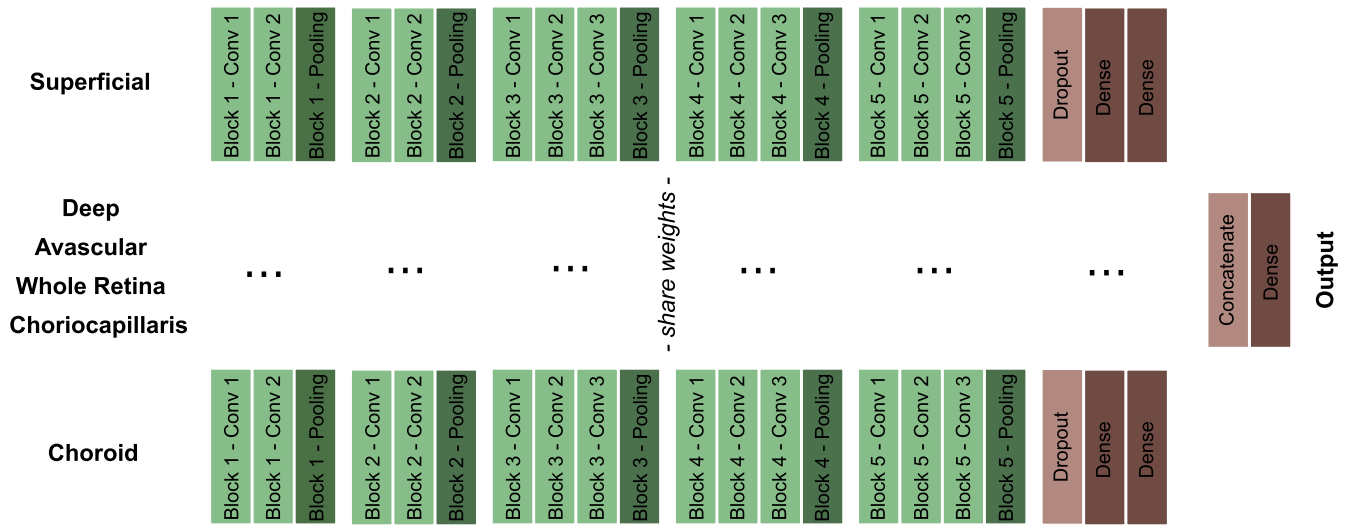
- biometrics and corneal micro-structure on non-contact tonometry. *PLoS One*. 2017;12:e0177180.
7. Cherecheanu AP, Garhofer G, Schmidl D, Werkmeister R, Schmetterer L. Ocular perfusion pressure and ocular blood flow in glaucoma. *Curr Opin Pharmacol*. 2013;13:36–42.
  8. Barbosa-Breda J, Abegão-Pinto L, Van Keer K, et al. Heterogeneity in arterial hypertension and ocular perfusion pressure definitions: towards a consensus on blood pressure-related parameters for glaucoma studies. *Acta Ophthalmol*. 2019;97:e487–e492.
  9. Spaide RF, Klancnik JM, Cooney MJ. Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography. *JAMA Ophthalmol*. 2015;133:45–50.
  10. Sugiyama T, Araie M, Riva CE, Schmetterer L, Orgul S. Use of laser speckle flowgraphy in ocular blood flow research. *Acta Ophthalmol*. 2010;88:723–729.
  11. Michelson G, Schmauss B, Langhans M, Harazny J, Groh M. Principle, validity, and reliability of scanning laser doppler flowmetry. *J Glaucoma*. 1996;5:99–105.
  12. Abegão Pinto L, Willekens K, Van Keer K, et al. Ocular blood flow in glaucoma—the Leuven Eye Study. *Acta Ophthalmol*. 2016;94:592–598.
  13. Stalmans I, Vandewalle E, Anderson DR, et al. Use of colour doppler imaging in ocular blood flow research. *Acta Ophthalmol*. 2011;89:e609–e630.
  14. Fang P, Lindner M, Steinberg J, et al. Clinical applications of OCT angiography. *Ophthalmologe*. 2016;113:14–22.
  15. Bekkers A, Borren N, Ederveen V, et al. Microvascular damage assessed by optical coherence tomography angiography for glaucoma diagnosis: a systematic review of the most discriminative regions. *Acta Ophthalmol*. 98:537–558, <https://onlinelibrary.wiley.com/doi/full/10.1111/aos.14392>.
  16. Shin JW, Lee J, Kwon J, Choi J, Kook MS. Regional vascular density–visual field sensitivity relationship in glaucoma according to disease severity. *Br J Ophthalmol*. 2017;101:1666–1672.
  17. Venugopal JP, Rao HL, Weinreb RN, et al. Repeatability of vessel density measurements of optical coherence tomography angiography in normal and glaucoma eyes. *Br j Ophthalmol*. 2018;102:352–357.
  18. Takusagawa HL, Liu L, Ma KN, et al. Projection-resolved optical coherence tomography angiography of macular retinal circulation in glaucoma. *Ophthalmology*. 2017;124:1589–1599.
  19. Jesus DA, Breda JB, Van Keer K, Sousa AR, Pinto LA, Stalmans. Quantitative automated circum-pillary microvascular density measurements: a new angio OCT-based methodology. *Eye*. 2019;33:320.
  20. Chihara E, Dimitrova G, Amano H, Chihara T. Discriminatory power of superficial vessel density and prelaminar vascular flow index in eyes with glaucoma and ocular hypertension and normal eyes. *Invest Ophthalmol Vis Sci*. 2017;58:690–697.
  21. Akil H, Huang AS, Francis BA, Sadda SR, Chopra V. Retinal vessel density from optical coherence tomography angiography to differentiate early glaucoma, pre-perimetric glaucoma and normal eyes. *PLoS One*. 2017;12:e0170476.
  22. Yarmohammadi A, Zangwill LM, Diniz-Filho A, et al. Optical coherence tomography angiography vessel density in healthy, glaucoma suspect, and glaucoma eyes. *Invest Ophthalmol Vis Sci*. 2016;57:OCT451–OCT459.
  23. Rao HL, Pradhan ZS, Weinreb RN, et al. Regional comparisons of optical coherence tomography angiography vessel density in primary open-angle glaucoma. *Am J Ophthalmol*. 2016;171:75–83.
  24. Wang X, Jiang C, Ko T, et al. Correlation between optic disc perfusion and glaucomatous severity in patients with open-angle glaucoma: an optical coherence tomography angiography study. *Graefes Arch Clin Exp Ophthalmol*. 2015;253:1557–1564.
  25. Rao HL, Pradhan ZS, Weinreb RN, et al. A comparison of the diagnostic ability of vessel density and structural measurements of optical coherence tomography in primary open angle glaucoma. *PLoS One*. 2017;12:e0173930.
  26. Van Melkebeke L, Barbosa-Breda J, Huygens M, Stalmans I. Optical coherence tomography angiography in glaucoma: a review. *Ophthalmic Res*. 2018;60:139–151.
  27. Barbosa-Breda J, Van Keer K, Abegão-Pinto L, et al. Improved discrimination between normal-tension and primary open-angle glaucoma with advanced vascular examinations—the Leuven Eye Study. *Acta Ophthalmol*. 2019;97:e50–e56.
  28. Kiyota N, Kunikata H, Shiga Y, Omodaka K, Nakazawa T. Relationship between laser speckle flowgraphy and optical coherence tomography angiography measurements of ocular microcirculation. *Graefes Arch Clin Exp Ophthalmol*. 2017;255:1633–1642.
  29. Jia Y, Wei E, Wang X, et al. Optical coherence tomography angiography of optic disc perfusion in glaucoma. *Ophthalmology*. 2014;121:1322–1332.
  30. Lee EJ, Kim S, Hwang S, Han JC, Kee C. Microvascular compromise develops following

- nerve fiber layer damage in normal-tension glaucoma without choroidal vasculature involvement. *J Glaucoma*. 2017;26:216–222.
31. Lommatzsch C, Rothaus K, Koch J, Heinz C, Grisanti S. OCTA vessel density changes in the macular zone in glaucomatous eyes. *Graefes Arch Clin Exp Ophthalmol*. 2018;256:1499–1508.
  32. Jampel HD. Glaucoma care update target pressure in glaucoma therapy. *J Glaucoma*. 1997;6:133–138.
  33. Spaide RF, Fujimoto JG, Waheed NK. Image artifacts in optical coherence angiography. *Retina*. 2015;35:2163.
  34. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The extensible neuroimaging archive toolkit. *Neuroinformatics*. 2007;5:11–33.
  35. Bradski G. The OpenCV Library. *Dr. dobb's journal software tools*. 2000;25:120–125, <https://www.proquest.com/docview/202684726?accountid=13598>.
  36. Van der Walt S, Schönberger JL, Nunez-Iglesias J, et al. Scikit-image: image processing in python. *Peer J*. 2014;2:e453.
  37. Grana C, Borghesani D, Cucchiara R. Optimized block-based connected components labeling with decision trees. *IEEE Trans Image Process*. 2010;19:1596–1609.
  38. Garway-Heath DF, Poinosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107:1809–1815.
  39. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Machine Learning Res*. 2011;12:2825–2830.
  40. Van Der Walt S, Colbert SC, Varoquaux G. The numpy array: a structure for efficient numerical computation. *Comput Sci Eng*. 2011;13:22.
  41. McKinney W. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX: SciPy; 2010:51–56, <https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>.
  42. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng*. 2007;9:90.
  43. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. 1999;9:293–300.
  44. Liaw A, Wiener M. Classification and regression by random forest. *R News*. 2002;2:18–22.
  45. Chen T, He T, Benesty M, Khotilovich V, Tang Y. Xgboost: extreme gradient boosting. R package version 0.4-2 1–4. 2015, <http://cran.fhcr.org/web/packages/xgboost/vignettes/xgboost.pdf>.
  46. Vapnik V. *The Nature of Statistical Learning Theory*. Springer Science & Business Media. New York: Springer-Verlag; 2013:93–122.
  47. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification *Technical Report, Department of Computer Science and Information Engineering, University of National Taiwan, Taipei*. 2003:1–12, <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
  48. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci*. 1996;11:189–212.
  49. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
  50. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett*. 2014;21:1389–1393.
  51. Bowker AH. A test for symmetry in contingency tables. *J Am Stat Assoc*. 1948;43:572–574.
  52. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput*. 1998;10:1895–1923.
  53. Haghghi S, Jasemi M, Hessabi S, Zolanvari A. PyCM: multiclass confusion matrix library in python. *J Open Source Softw*. 2018;3:729.
  54. Sánchez Brea L, Andrade De Jesus D, Shirazi MF, Pircher M, Walsum T van, Klein S. Review on retrospective procedures to correct retinal motion artefacts in OCT imaging. *Appl Sci*. 2019;9:2700.
  55. Richter GM, Chang R, Situ B, et al. Diagnostic performance of macular versus peripapillary vessel parameters by optical coherence tomography angiography for glaucoma. *Transl Vis Sci Technol*. 2018;7:21–21.
  56. Triolo G, Rabiolo A, Shemonski ND, et al. Optical coherence tomography angiography macular and peripapillary vessel perfusion density in healthy subjects, glaucoma suspects, and glaucoma patients. *Investig Ophthalmol Vis Sci*. 2017;58:5713–5722.
  57. Chen HS-L, Liu C-H, Wu W-C, Tseng H-J, Lee Y-S. Optical coherence tomography angiography of the superficial microvasculature in the macular and peripapillary areas in glaucomatous and healthy eyes. *Invest Ophthalmol Vis Sci*. 2017;58:3637–3645.
  58. Rolle T, Dallorto L, Tavassoli M, Nuzzi R. Diagnostic ability and discriminant values of OCT-angiography parameters in early glaucoma diagnosis. *Ophthalmic Res*. 2019;61:143–152.
  59. Kromer R, Glusa P, Framme C, Pielen A, Junker B. Optical coherence tomography angiography analysis of macular flow density in glaucoma. *Acta Ophthalmol*. 2019;97:e199–e206.
  60. Kurysheva NI, Maslova EV, Zolnikova IV, Fomin AV, Lagutin MB. A comparative study of

- structural, functional and circulatory parameters in glaucoma diagnostics. *PLoS One*. 2018;13:e0201599.
61. Yip VC, Wong HT, Yong VK, et al. Optical coherence tomography angiography of optic disc and macula vessel density in glaucoma and healthy eyes. *J Glaucoma*. 2019;28:80–87.
  62. Kim SB, Lee EJ, Han JC, Kee C. Comparison of peripapillary vessel density between preperimetric and perimetric glaucoma evaluated by OCT-angiography. *PLoS One*. 2017;12:e0184297.
  63. Olson RS, Cava WL, Mustahsan Z, Varik A, Moore JH. Data-driven advice for applying machine learning to bioinformatics problems. *Biocomputing*. 2018;23:192–203, [https://doi.org/10.1142/9789813235533\\_0018](https://doi.org/10.1142/9789813235533_0018).
  64. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015. ICLR CoRR, abs/1409.1556. <https://www.robots.ox.ac.uk/~vgg/publications/2015/Simonyan15/simonyan15.pdf>.
  65. Mok KH, Lee VW-H, So KF. Retinal nerve fiber loss pattern in high-tension glaucoma by optical coherence tomography. *J. Glaucoma*. 2003;12:255–259.
  66. Andrade De Jesus D, Sanchez L, Breda J, et al. The superficial peripapillary vascular layer is the most informative for glaucoma diagnosis: a multi-layer study based on OCT angiography. In *The Association for Research in Vision and Ophthalmology Imaging Eye Conference*. Vol. 60. Association for Research in Vision and Ophthalmology; 2019, PB052, <https://iovs.arvojournals.org/article.aspx?articleid=2748135>.
  67. Bojikian KD, Chen CL, Wen JC, et al. Optic disc perfusion in primary open angle and normal tension glaucoma eyes using optical coherence tomography-based microangiography. *PLoS One*. 2016;11:e0154691.
  68. Rao HL, Pradhan ZS, Weinreb RN, et al. Vessel density and structural measurements of optical coherence tomography in primary angle closure and primary angle closure glaucoma. *Am Ophthalmol*. 2017;177:106–115.
  69. Kanamori A, Nakamura M, Tomioka M, Kawaka Y, Yamada Y, Negi A. Structure–function relationship among three types of spectral-domain optical coherent tomography instruments in measuring parapapillary retinal nerve fibre layer thickness. *Acta Ophthalmol*. 2013;91:e196–e202.
  70. Ueda K, Kanamori A, Akashi A, Kawaka Y, Yamada Y, Nakamura M. Difference in correspondence between visual field defect and inner macular layer thickness measured using three types of spectral-domain OCT instruments. *Jpn Ophthalmol*. 2015;59:55–64.
  71. Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci*. 2012;53:6939–6946.
  72. Mehta N, Liu K, Alibhai AY, et al. Impact of binarization thresholding and brightness/contrast adjustment methodology on optical coherence tomography angiography image quantification. *Am Ophthalmol*. 2019;205:54–65.
  73. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9:671.
  74. Holló G. Influence of removing the large retinal vessels–related effect on peripapillary vessel density progression analysis in glaucoma. *J Glaucoma*. 2018;27:e137–e139.
  75. Brea LS, Andrade De Jesus D, Klein S, van Walsum T. *Proceedings of the Third Conference on Medical Imaging with Deep Learning, PMLR 121*. 2020:709–720.
  76. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
  77. Torrey L, Shavlik J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA: IGI Global; 2010:242–264.
  78. Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient object localization using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE; 2015:648–656.
  79. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami: IEEE; 2009:248–255.
  80. Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning. In *12th fUSENIXg Symposium on Operating Systems Design and Implementation (fOSDIg 16)*. Savannah, GA, USA: Usenix; 2016:265–283.
  81. Gulli A, Pal S. *Deep Learning With Keras*. Birmingham, UK: Packt Publishing Ltd.; 2017.
  82. Zeiler MD. Adadelata: an adaptive learning rate method. arXiv preprint arXiv:1212.5701. 2012.

## Appendix

In this section, we summarize the obtained results for the classification of glaucoma disease (healthy,



**Figure A1.** VGG16 architecture modified for each of the six OCTA images. The weights of the convolutional blocks (in green) were frozen, while the dense layers were trained.

glaucoma), type of glaucoma (POAG, NTG), and glaucoma severity (healthy, mild, moderate, and severe) using deep learning (DL).<sup>76</sup> Among the different possibilities, a transfer-learning approach based on a CNN classification model was applied. In transfer learning, a new task is learned through the transfer of knowledge from a task that has already been learned.<sup>77</sup> The task used to train the model does not need to be related to the objective task. Such a technique reduces the amount of data required and, in many cases, increases the accuracy as compared to models built from scratch.

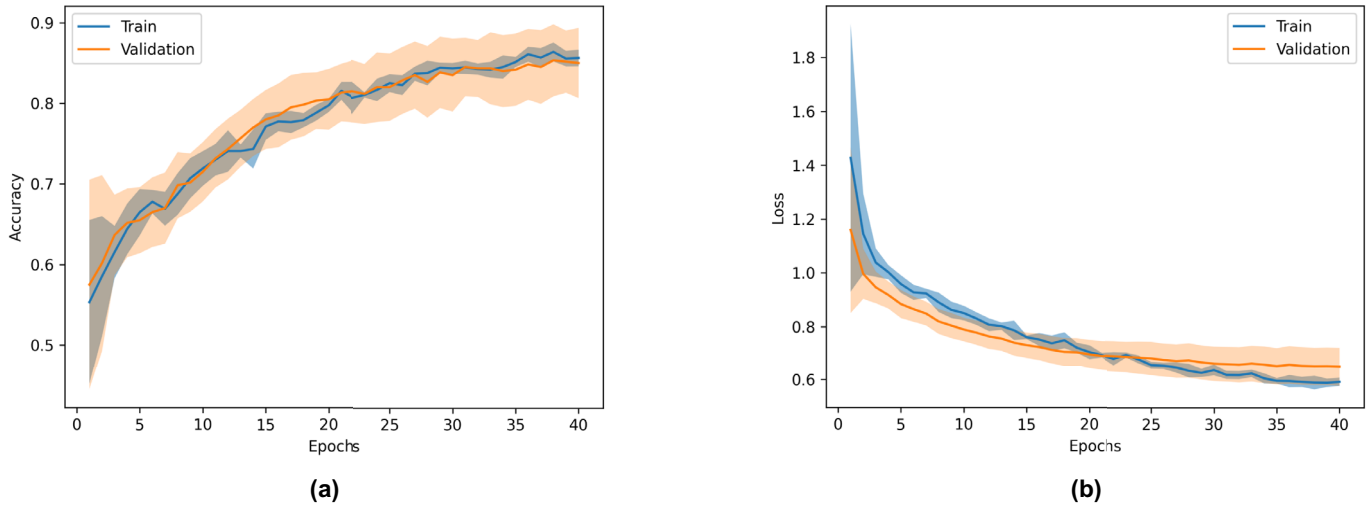
## Methods

A state-of-the-art architecture for classification problems, VGG16,<sup>64</sup> was used as basis for the model. All OCTA images were resized to match the VGG16's expected input ( $224 \times 224 \times 3$ ). Since the current data set consists of six vascular layers per subject (see Fig. 2), a VGG16 model was implemented per layer. The weights of the six models were shared, as otherwise the number of trainable parameters would be too high, potentially leading to overfitting. Regarding each individual VGG16, the convolutional blocks were kept as in the original architecture, but the last block, which consists of three fully connected layers, was altered to reduce the model complexity and to prevent overfitting (see Fig. A1). A spatial dropout<sup>78</sup> layer of 0.25 was added to the block, and the number of neurons of the two fully connected layers was reduced from 4096 to 2048. A concatenation layer was used to merge the output of all six models, and the number of neurons

of the output layer was changed from 1000 to 2 and 4 for the binary and multiclass classification, respectively. The network was initialized with the ImageNet<sup>79</sup> weights, and only the fully connected layers were retrained. Tensorflow<sup>80</sup> and Keras<sup>81</sup> libraries were used for the implementation of the models.

To prevent the model from overfitting the training set and to improve its generalization capabilities, data augmentation was performed, increasing the number of samples in the data set by a factor of 5. The following ranges of transformation and normalization operations were applied to the data: rotation (up to 10 degrees), height and width shift (ranging up to 10%), horizontal flip, zoom, and brightness (ranging from 80% to 120%).

Experiments with various hyperparameter values were conducted to maximize the performance of the CNN model for classifying healthy controls against glaucoma subjects. The optimizer was the stochastic gradient descent that adapts learning rates based on a moving window of gradient updates (Adadelta).<sup>82</sup> The binary cross-entropy was chosen as loss function, and the training ran for 40 epochs with a batch size of 16. The hyperparameters chosen for this first classification task were also used for training the models to classify POAG versus NTG and glaucoma severity levels. For the latter, the categorical cross-entropy loss function was used. Given the constraints in computing power, stratified cross-validation was used to split the data in five folds, instead of the leave-one-out pipeline used in the classical methods. The number of images for each class was the same for each fold.



**Figure A2.** Mean and confidence interval at 95% for the accuracy (a) and loss (b) of the model trained to classify healthy controls versus glaucoma subjects.

**Table A1.** AUROC, Accuracy, Sensitivity, and Specificity ( $\pm$  Confidence Interval at 95%) for the Glaucoma Classification, Type of Glaucoma, and Glaucoma Severity Analysis Based on the Transfer Learning Approach

Analysis	AUROC	Accuracy	Sensitivity	Specificity
Glaucoma classification ( <i>healthy, glaucoma</i> )	$0.83 \pm 0.05$	$0.87 \pm 0.03$	$0.73 \pm 0.10$	$0.94 \pm 0.03$
Type of glaucoma ( <i>POAG, NTG</i> )	$0.65 \pm 0.09$	$0.65 \pm 0.09$	$0.65 \pm 0.19$	$0.66 \pm 0.13$
Glaucoma severity ( <i>healthy, mild, moderate, severe</i> )	$0.63 \pm 0.03$	$0.47 \pm 0.05$	$0.47 \pm 0.05$	$0.82 \pm 0.02$

## Results and Discussion

Figure A2 shows the mean and confidence interval at 95% for the accuracy (a) and loss (b) of the model trained to classify healthy controls versus glaucoma subjects. The AUROC, accuracy, sensitivity, and specificity for the three models are summarized in Table A1.

The results show that the presented transfer learning approach does not outperform the best conventional models presented earlier in this work. For the healthy control versus glaucoma subjects, the specificity was better than the best model presented in Table 2. However, the AUROC, the accuracy, and the sensitivity presented lower values. The model trained to classify the type of glaucoma did present slightly better results for the AUROC, accuracy, and specificity than the best model in Table 4. However, no significant statistical difference was observed between the ROC curves ( $P = 0.46$ ). Moreover, the results are still underwhelming, and it is not possible to claim that OCTA data can be

used to differentiate between POAG and NTG subjects. Lastly, the model trained to differentiate between glaucoma severity levels presented lower values for all metrics when compared to the best model in Table 5. This may be explained by the low number of images of the highest severity level, as depicted in Table 1.

## Conclusion

The application of the CNN model presented in this Appendix did not show added benefits when compared to the best models presented in Tables 2, 4, and 5. Although we can not claim that other deep learning approaches could not give better results, it is certain that in-depth research would be needed to significantly improve the performance of the current classification given the size of the OCTA data set. Nevertheless, larger data sets would likely benefit from the performance of deep learning-based models.