The final publication is available at

Circumpapillary OCT-Focused Hybrid Learning for Glaucoma Grading Using Tailored Prototypical Neural Networks

Gabriel García[1], Rocío del Amor[1], Adrián Colomer[1], Rafael Verdú-Monedero[2], Juan Morales-Sánchez[2] and Valery Naranjo[1]

[1]*Instituto de Investigación e Innovación en Bioingeniería, Universitat Politècnica de València, 46022, Valencia, Spain*

[2]*Departamento de Tecnologías de la Información y las Comunicaciones, Universidad Politécnica de Cartagena, 30202, Cartagena, Spain*

Abstract

Glaucoma is one of the leading causes of blindness worldwide and Optical Coherence Tomography (OCT) is the quintessential imaging technique for its detection. Unlike most of the state-of-the-art studies focused on glaucoma detection, in this paper, we propose, for the first time, a novel framework for glaucoma grading using raw circumpapillary B-scans. In particular, we set out a new OCT-based hybrid network which combines hand-driven and deep learning algorithms. An OCT-specific descriptor is proposed to extract hand-crafted features related to the retinal nerve fibre layer (RNFL). In parallel, an innovative CNN is developed using skip-connections to include tailored residual and attention modules to refine the automatic features of the latent space. The proposed architecture is used as a backbone to conduct a novel few-shot learning based on static and dynamic prototypical networks. The $k-$shot paradigm is redefined giving rise to a supervised end-to-end system which provides substantial improvements discriminating between healthy, early and advanced glaucoma samples. The training and evaluation processes of the dynamic prototypical network are addressed from two fused databases acquired via Heidelberg Spectralis system. Validation and testing results reach a categorical accuracy of 0.9459 and 0.8788 for glaucoma grading, respectively. Besides, the high performance reported by the proposed model for glaucoma detection deserves a special mention. The findings from the class activation maps are directly in line with the clinicians' opinion since the heatmaps pointed out the RNFL as the most relevant structure for glaucoma diagnosis.

Keywords: Glaucoma Grading, Prototypical Neural Networks, Circumpapillary, Hybrid Learning, Retinal Nerve Fiber Layer, Optical Coherence Tomography

1. Introduction

Glaucoma is a chronic and progressive disease that affects the optic nerve head (ONH) of the retina causing several structural changes and functional damage [1]. Nowadays, this optic neuropathy has become the leading cause of blindness worldwide, according to [2]. Recent studies suggest that the impact of this disease will continue to rise, affecting 111.8 million people in 2040 [3]. Therefore, early diagnosis of glaucoma could be essential for timely treatment in order to prevent irreversible vision loss [2].

Currently, there is no single accurate test to certify glaucoma, the diagnostic procedure includes several time-consuming tests such as pachymetry, tonometry and visual field tests, as well as the examination of different kinds of interpretable retinal images. Specifically, techniques based on image analysis like fundus photography and optical coherence tomography (OCT) have become very important in the context of glaucoma detection. Fundus image is a great cost-effectiveness technique which has reported promising results in the diagnosis of several eye-focused diseases, e.g. diabetic retinopathy [4, 5] and age-related macular degeneration [6, 7]. However, OCT imaging modality [8] is the quintessential technique for glaucomatous damage evaluation [9] since it allows quantifying glaucoma-specific regions such as retinal nerve fibre layer (RNFL) and ganglion cell inner plexiform layer (GCIPL), which are useful biomarkers for the progression of this disease [10]. Additionally, glaucoma is evident in the deterioration of the cell layers around the optic disc, whose infor-

mation could be exploited by the OCT imaging modality since it focuses on the depth axis of the retina to identify structural changes, unlike the 2D projection of the fundus image (see Fig. 1).

Note that, although fundus image modality is cheaper than OCT, it is colour-dependent on the training data set and its interpretation remains subjective [11, 12]. The contrary, OCT is a non-contact and non-invasive technique that provides objective information about the optic nerve head and RNFL structures [13]. It is important to remark that glaucoma detection entails a subjective examination from different experts, whose mismatch ratio is usually high [14]. Consequently, many state-of-the-art studies developed different machine learning algorithms intended to detect glaucoma via fundus image and OCT samples.

2. Related work

According to the recent study [15], which describes a review of deep learning methods for glaucoma detection, spectral domain (SD) OCT has become the most widespread diagnostic tool for analysing glaucomatous pathologies. Besides, the OCT system is routinely used in clinical practice for determining glaucoma severity since it allows emphasising the significance of the structural changes of the retina [16]. There are several clinical studies which claim the potential of the OCT imaging modality in the glaucoma detection paradigm. Medeiros et al. [17] enhanced the importance of the RNFL structure demonstrating that an early pathological degeneration of the retinal cells is associated with a thinning of the RNFL. In [16], authors evidenced the usefulness of the RNFL thickness to determine pathological variations of the retina associated with different glaucomatous stages. Ojima et al. [18] reported that the RNFL thickness has a higher potential for glaucoma diagnosis than a complete macular volume. Moreover, in [19], researchers declared that RNFL features extracted from SD-OCT scans are a powerful indicator for glaucomatous damage evaluation, which is directly in line with the outcomes from our previous work [20].

Inspired by the aforementioned clinical studies, many researchers have developed predictive models to discern between normal and glaucomatous patients via OCT scans analysis. Most of them focused on the optic disc due to the known potential of the circumpapillary OCT images for glaucoma detection [21]. In [22, 23], authors extracted hand-crafted features from the B-scans and combined them with transformations of A-scans and visual field parameters to discriminate

between healthy and glaucoma classes. They used different machine learning classifiers such as support vector machine (SVM), random forest (RF) and k-nearest neighbour (KNN). In [24], an automated framework to estimate the thickness of specific OCT layers was presented in order to monitor retinal abnormalities, similarly to [25, 26], where the proposed methods were intended to quantify retinal layers thicknesses in search of abnormal retinal pathologies. Gao et al [27], made a comparison between the thickness measurements automatically reported by the Topcon OCT system and their end-to-end framework composed of a segmentation stage followed by a feature extraction based on the RNFL thicknesses average. It is important to note that this kind of hand-driven learning methodologies usually requires a previous manual or automatic segmentation of the retinal layers to delimit the regions from which extracting the discriminative features [15]. For that reason, several studies have been proposed in the literature for the sole purpose of addressing this long-standing problem [28–35].

However, segmentation algorithms should entail errors which are transferred to the feature extraction stage [15]. To avoid this shortcoming, deep learning models arise to provide alternative ways of quantifying structural damage since they can learn features from data automatically without reliance on previous segmentation stages or predefined features. In this context, ophthalmology has risen to a forefront in which application of deep learning (DL) algorithms can boost to a better artificial intelligence-based diagnosis in the medical fields. Specifically, glaucoma broadly meets the conditions to aid in the management of the vast amount of information coming from SD-OCT scans, according to the review outlined in [15]. Following the deep learning trend, Thompson et al. [36] demonstrated in a recent study that their segmentation-free DL algorithm (trained with raw SD-OCT B-scans) exceeded the conventional RNFL thickness parameters to directly discern between glaucomatous and healthy eyes. Maetschke et al. [37] also conducted a comparison between hand-driven and data-driven learning strategies applied on raw OCT volumes around the ONH of the retina for glaucoma detection. The main outcomes revealed that the deep learning approach outperformed conventional SD-OCT parameters to discriminate glaucomatous from normal samples. At this point, it should be noted that there are other studies which addressed the binary classification between healthy and glaucomatous cases by applying deep learning algorithms on spectral-domain OCT volumes [38–40], including our previous work [41].

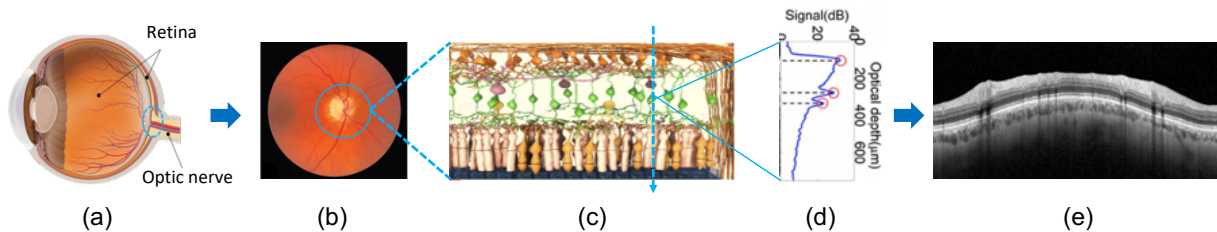However, to the best of the authors' knowledge, the

Figure 1: (a) Eyeball with interesting regions. (b) Fundus image focused on the optic nerve head (ONH). (c) Arrangement of the retinal fiber layers. (d) A-scan corresponding to the depth of retinal cells at a specific point. (e) B-scan representing the fiber layers of the retina with different gray-intensity levels.

deep learning application is not very widespread on circumpapillary OCT images. A recent review of the literature found that most of the deep learning-focused studies using B-scans are conducted in a combination of fundus images to detect glaucoma via RNFL probability maps [42–44]. To fill this gap in the literature, we previously proposed different hand-driven and data-driven learning strategies for glaucoma diagnosis in [20, 41, 45], whose outcomes support the basis of this paper.

2.1. Contribution of this work

Inspired by the high performance reported from our previous circumpapillary-based studies, we have expanded the database of B-scans around the ONH to go deeper into the glaucoma paradigm. In particular, we propose in this paper an innovative framework based on prototypical networks for glaucoma grading using raw circumpapillary images. As far as we know, this work is the first OCT-focused study intended to grade the glaucoma severity by discerning between healthy, early and advanced classes, which adds significant value to the body of knowledge. Most of the previous studies centred on B-scans are designed to discriminate glaucoma from healthy samples [36, 46–50]. Other state-of-the-art studies also pursued the classification of healthy and glaucomatous cases, but using the OCT volumes as an input to their models [37–41]. Additional glaucoma-related studies were addressed from B-scans to accomplish different discrimination tasks such as preperimetric vs perimetric glaucoma [51, 52], progressing vs non-progressing glaucoma [53, 54] and close angle vs open-angle glaucoma [55, 56], among others. Furthermore, there are studies which used a different kind of input data, e.g. visual field tests [57, 58] to discern between healthy and glaucomatous patients. On a wider level, the sub-classification of early and advanced glaucoma has already been conducted in the literature, but throughout fundus image material [59–61].

At this point, it should be noted that a very recent work [62] also proposes a kind of glaucoma grading set up making use of the Armed Forces Institute of Ophthalmology (AFIO) data set [63], which contains OCT scans centred on the ONH. However, it pursues the discrimination between healthy, suspects and glaucomatous samples by computing the distance of different retinal layers of interest previously segmented. In particular, Raja et al. in [62] propose an encoder-decoder architecture to carry out the glaucoma classification. The encoder structure was used to provide a feature map capable of discerning between healthy and glaucoma classes via softmax function; whereas the decoder component was intended to segment the interesting layers of the retina to distinguish between suspect and glaucoma cases from the samples previously predicted as glaucoma. To accomplish this part, the mean of the segmented layers was used as a feature input of an SVM classifier. Unlike AFIO database whose labels correspond to non-successive phases of the disease (healthy, suspect and glaucoma), our database includes different levels of glaucoma severity annotated according to the medical literature [64, 65]. Contrary to Raja et al. [62], we can conduct a learning framework that enables the analysis of glaucoma progression by differentiating between healthy, early and advanced glaucomatous samples. Another essential difference with respect to [62] is that we develop the predictive models from raw grayscale B-scans, whereas Raja et al. [62] made use of pre-processed RGB scans containing manual annotations and highlighted structures.

Furthermore, the proposed work documents additional key contributions concerning the deep learning application in the glaucoma field. For the first time, we raise a glaucoma scenario based on prototypical neural networks (PNN) [66], which have demonstrated a high rate of performance in recent image analysis tasks, such as domain adaptation [67], noisy evaluation [68], text

3

classification [69], etc. Note that prototypical networks are usually formulated as a baseline within the few-shot paradigm [70–73], but in this paper, we exploit the prototypical concept in the *k*-shot methodology to optimize the learning process for glaucoma grading.

Tatham et al. [74] argued that circumpapillary RNFL (cpRNFL) thickness was the best structure to measure glaucoma progression and the most widely used parameter in clinical practice. So, according to this, and inspired by our previous works [20, 45], we outline in this paper a novel OCT-based hybrid backbone as a feature extractor of the prototypical framework. Specifically, from [20], we observed that hand-crafted features could outperform the automatic features extracted by deep learning models trained from scratch. Nevertheless, from the study carried out in [45], we detected that fine-tuned models also improved the model's performance compared with algorithms trained from scratch. For that reason, in this approach, we propose a novel backbone composed of pre-trained deep learning networks (with additional attention modules and residual blocks) in a combination of hand-crafted RNFL-based features, similar to the hybrid methodology that reported the best performance in [20].

In summary, the main contributions of this work are:

- Raw circumpapillary OCT images are used for the first time to measure the glaucoma severity.

- Tailored prototype-based solutions are formulated in a novel framework for glaucoma grading.

- An adapted *k*-shot supervised learning, inspired by the few-shot paradigm, is conducted to exploit the specific-glaucoma knowledge.

- A new OCT-based hybrid backbone is proposed as a feature extractor to combine automatic and hand-crafted information from B-scans.

The rest of the paper is organized as follows. Section 3 outlines the proposed methodology composed of two stages: i) base encoder development and ii) prototype-based learning strategies. Section 4 shows the ablation experiments performed during the validation stage. Section 5 presents the quantitative and qualitative results achieved in the prediction of the test set. A wide discussion about the proposed framework is conducted in Section 6, and a summary of the main conclusions is addressed in Section 7.

3. Methods

3.1. Base encoder development

In this paper, we pay special attention to this section since the performance of the prototypical frameworks largely depends on the representation vectors encoded in the latent space by the feature extractor. The original study of the prototypical neural networks (PNN) [66], as well as others derived from it [67, 70], made use of a 4-layer CNN trained from scratch as an encoder of the feature representation. However, we observed from our previous glaucoma-based works [20, 45] that deep learning models trained from scratch reported the poorest performance in comparison to fine-tuning the models or even extracting hand-crafted features. For that reason, we built on our previous experience to propose in this work a new tailored backbone able to capture the OCT-specific cues for an optimal glaucoma grading. Specifically, we inspired on the OCT hybrid methodology followed in [20], but using pre-trained networks according to [41, 45], to provide a novel base encoder Ψ_ϕ with some novelties that allow improving the previous approaches. The proposed backbone is composed of two learning branches giving rise to a multi-input feature model, as observed in Fig. 2.

3.1.1. Deep learning branch

From the sweep of the pre-trained networks carried out in [45], we selected the VGG16 architecture as the baseline of our deep learning branch Γ_ϕ. In particular, we applied a deep fine-tuning strategy [75] to freeze the weights of the three first convolutional blocks, which were pre-trained with around 14 million of natural images corresponding to the *ImageNet* data set. As a novelty, we propose a glaucoma-specific residual structure which allows propagating the information from the initial VGG layers to the last ones, via convolutional-skip connections. Thereby, the shortcut flows through the gradient of a deeper network to mitigate the problem of vanishing gradients. The proposed residual block aims to optimise the dimensionality of the filters by combining 1×1 convolutions (green boxes) with a customized 3×1 convolutional layer (yellow box). This tailored kernel size allows taking advantage of the domain-specific knowledge of the OCT images to provide local cues for the glaucoma learning process. In this way, the network is encouraged to focus on the OCT vertical axis in order to learn the glaucoma-specific information underlying the contrast differences of the retinal layers. An additional 1×1 convolutional layer was included to reduce the filters' dimension after concatenating the feature maps from the VGG and residual structures. At this
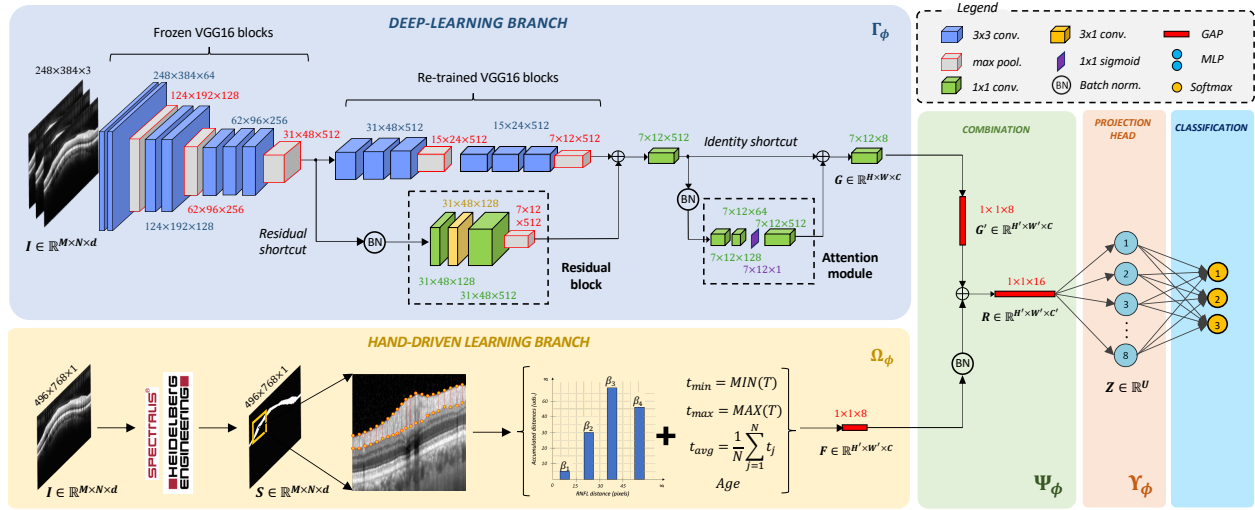
Figure 2: Illustration of the end-to-end backbone proposed as a benchmark to conduct the prototype-based learning strategies. Blue, yellow and green frames correspond to the base encoder network which consists of deep learning and hand-driven learning branches followed by a combination module, respectively. A projection head (red) maps the embedded representations in a lower-dimensional space to maximize the agreement in the classification stage (cyan).

point, we also included an attention module via skip-connections to refine the features in the spatial dimension. The proposed module works as a kind of autoencoder composed of 1×1 convolutions in which the filters are decreased and increased, respectively. At the bottleneck, a single convolutional filter activated by a sigmoid function (purple layer) was included to recalibrate the inputs and forcing the network to learn useful properties from the input representations. The skip-connection allows propagating larger gradients to previous layers throughout an identity shortcut. At the end of the deep learning branch (see Fig. 2), another 1×1 convolutional layer was defined to provide a volume map $G = \{g_1, g_2, ..., g_k, ..., g_C\}$, where $C = 8$ is the number of feature maps g_k, of size $H \times W = 7 \times 12$, which compose the set of the deep learning representations $\Gamma_\phi : I \rightarrow G$.

3.1.2. Hand-driven learning branch

Otherwise, inspired by the clinical study [74], we conducted an additional hand-driven learning branch Ω_ϕ focused on hand-crafted features extracted from the retinal nerve fibre layer (RNFL). Specifically, we made use of an innovative RNFL descriptor, which was proposed in our previous work [20] to include RNFL thickness-based information by computing a bespoke OCT-specific histogram. Unlike [19, 22, 23, 62, 76], among others, where descriptors were based on the RNFL mean, the proposed method allows leveraging the individual information provided by the RNFL thickness

at each point of the B-scan, by considering bags of similar thicknesses values. Thus, let I be a raw circumpapillary image of dimensions $M \times N \times d$, and S its corresponding RNFL mask segmented automatically by the Heidelberg Spectralis OCT system, a vector of thicknesses $T = \{t_1, t_2, ..., t_j, ..., t_N\}$ was computed, where t_j is the thickness value at the position j of the B-scan I, with $j = 1, 2, 3, ..., N$. The proposed histogram-based descriptor is able to quantify the RNFL information into $b = 4$ bags depending on the thickness values. In this way, each bag β_b collects the number of thicknesses t_j whose value is ranged between D_b and D_{b+1}, being $D = [0, 15, 30, 45, \infty]$ a vector of relevant distances optimized on the training images. Additionally, the minimum, maximum and average of the RNFL thickness, besides the age of the patients, were also considered according to the equations formulated in Fig. 2. Finally, the hand-driven learning branch provides a feature vector F consisted of $C = 8$ RNFL-specific features, such that $\Omega_\phi : I \rightarrow F$.

3.1.3. Combination module

Once automatic and hand-crafted features were extracted from their respective branches, a simple combination module was proposed to join the embedded information in a holistic map representation R composed of $C' = 16$ variables per learning instance, as observed in Fig. 2. In particular, the feature volume G extracted from the deep learning branch was mapped to a vector

5

G' by a Global Average Pooling (GAP) layer, according to Equation 1. This operation computes a spatial squeeze from $H \times W$ to $H' \times W'$ that enables the concatenation between the features G and F coming from the different branches. Thereby, the proposed hybrid encoder Ψ_ϕ learns the embedding representations R from the input I as follows: $R = \Psi_\phi(I) = \Gamma_\phi(I) \oplus \Omega_\phi(I)$, where $\oplus$ denotes a concatenation operation.

$$g'_k = \frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} g_k(h, w) \qquad (1)$$

3.1.4. Projection head module

In this paper, we instantiate a projection head network Υ_ϕ that maps the representations R to an embedding vector Z where the classification stage is addressed in a lower-dimensional space. The projection head Υ_ϕ is comprised of a small multi-layer perceptron (MLP) with one hidden layer non-linearly activated by the ReLU function (see Fig. 2). The use of a projection head network is widely used in very recent state-of-the-art techniques, such as contrastive learning [77–79], to maximize the classification agreement. In this paper, we project the representations of the latent space via $Z = \Psi_\phi(R)$ to evidence that the new backbone Ψ_ϕ is better than the previous feature extractors proposed in [20, 41, 45]. According to the aforementioned studies [77–79], the projection head network is then discarded during the prototypical learning stage to measure the distances from the representations $R = \Psi_\phi(I)$ to each prototype. For comparison purposes, a softmax function with three neurons was defined in the last dense layer to contrast a conventional classification approach with the proposed prototypical frameworks.

As a summary of the detailed base encoder backbone, we show a pipeline in Fig. 3 that collects the essential information. Given an input B-scan $I \in \mathbb{R}^{M \times N \times d}$, with $M \times N \times d = 248 \times 384 \times 3$ the dimensions of I, a feature embedded map $R \in \mathbb{R}^{H' \times W' \times C'}$ is provided by the encoder $\Psi_\phi : I \rightarrow R$. Then, the projection head module Υ_ϕ maps R to a metric vector $Z \in \mathbb{R}^U$, $\Upsilon_\phi : R \rightarrow Z$, where $U = 8 < C'$ denotes a lower dimensional space than the latent space R. At the end of the convolutional network, a softmax-activated dense layer is applied to address the classification stage.

3.2. Prototype-based learning strategies

Prototypical networks were born from the idea that there exists an embedding space in which the features from the same class cluster around a single latent group,

a.k.a prototype [66]. In this paper, we conduct an experimental methodology to analyse the performance of two different prototype-based solutions with respect to conventional learning strategies for glaucoma grading. To this end, the traditional approach was defined according to the backbone architecture exposed in Fig. 3, i.e. a base encoder network followed by a projection head module with a softmax-activated function of three neurons corresponding to healthy, early and advanced glaucomatous classes. Concerning the prototype strategies, we present below a comparison between two novel approaches -static and dynamic- that use the embedding space R to determine the glaucoma severity.

3.2.1. Static prototypes

In this paper, we introduce the concept "*static*" to reference the use of rigid prototypes extracted from an encoder network Ψ_ϕ whose weights ϕ were pre-trained in a previous stage through a conventional approach. Thereby, the frozen weights (denoted by Θ) are inferred to the base encoder model to extract the embedding representations R from both training τ and validation ν sets. Note that R^τ is used to obtain the prototypes and R^ν to find the nearest class prototype in the latent space.

A similar procedure based on measuring the similarity of the latent features from different input data has been applied on different state-of-the-art techniques, such as contrastive learning [77–79] or content-based image retrieval (CBIR) [80–82]. In contrastive learning tasks, an encoder network followed by a projection head is trained to differentiate between positive and negative samples, being positive samples the augmented version of a query (or samples with the same label in the case of supervised contrastive [77]), and negative samples the entire remainder of the batch. Contrarily, CBIR studies train convolutional autoencoders and extract the latent space from the encoder structure to find relevant images retrieved from a set of reference that shares similar embedding features with the query image. Previous systems present some similarities with the proposed static prototype approach, since all the methods train an encoder network that is then frozen to face a second classification stage. In the case of contrastive-learning studies, the feature map extracted by the encoder architecture is used to predict the class of the query sample via MLP, k-nearest neighbours (KNN) or inferred prototypes, according to [77]. In contrast, CBIR studies are intended to find the reference set samples that most closely resemble the query image by measuring the similarity of the embedded representations extracted by the encoder structure. The proposed approach differs from the previous ones in a critical point: the encoder net-
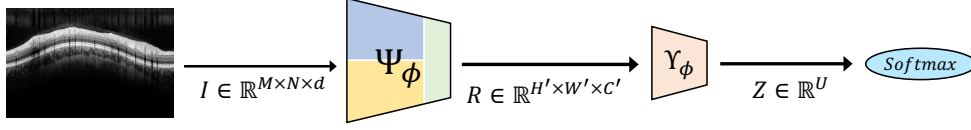
Figure 3: Pipeline showing the backbone architecture composed of the base encoder Ψ_ϕ, the projection head network Υ_ϕ and the softmax function.

work is trained during the first stage for the same objective to be achieved in the second one, i.e. glaucoma grading. Oppositely, contrastive-learning and CBIR-based studies use backbones that were pre-trained for a different task during the first stage.

Below, we detail the training of the proposed approach based on static prototypes, which is composed of two (online and offline) stages, according to Fig. 4.

In the online stage, a conventional classification pipeline was conducted to optimize the weights of the proposed OCT-hybrid backbone by minimizing the categorical cross-entropy loss function $\mathcal{L}(y^\tau, \hat{y}^\tau)$ in each training epoch $e = 1, 2, 3, ..., \epsilon$, as detailed in Algorithm 1. In the offline stage, the weights of the pre-trained encoder Ψ_ϕ were frozen ($\phi \rightarrow \Theta$) and the projection head Υ_ϕ and softmax modules were discarded to avoid the non-linearity of the top model. The embedding representations $E^\tau = \{R_1, R_2, ..., R_i, ..., R_{P^\tau}\}$, with P^τ the number of samples of the minority class in the training set τ, were used to infer the rigid prototypes. In particular, each ρ_c was calculated as the mean of the latent representations $R_i^{\tau_c} = \Psi_\phi(I_i)$, where $I_i \in \tau_c \subset \tau$ denotes the i-sample of the training set τ associated with the class c (see Equation 2).

$$\rho_c = \frac{1}{P^{\tau_c}} \sum_{i=1}^{P^{\tau_c}} R_i^{\tau_c} \qquad (2)$$

During the prediction phase, a matrix of distances $\delta_{c,i}$ was achieved by measuring the Euclidean distance (Equation 3) between each prototype ρ_c and the embedding representation $R_i^\nu = \Psi_\phi(V_i)$, where $V_i \in \nu$ corresponds to the i-scan of the validation subset ν. From here, a probability of belonging to each class was calculated (Equation 4) to determine the predicted class $\hat{y}_i^\nu$, as detailed in Algorithm 1.

$$\delta_{c,i} = \sqrt{(\rho_c - R_i^\nu)^2} \qquad (3)$$

$$p_{i,c} = \frac{\exp(-\delta_{c,i})}{\sum_{c'} \exp(-\delta_{c',i})} \qquad (4)$$

Algorithm 1: Static prototype-based strategy.

Data: Training $\tau = \{(I_1, y_1^\tau), ..., (I_{P^\tau}, y_{P^\tau}^\tau)\}$ and
validation $\nu = \{(V_1, y_1^\nu), ..., (V_{P^\nu}, y_{P^\nu}^\nu)\}$ sets.

Results:
Online stage $\leftarrow$ Trained base encoder Ψ_ϕ ;
Offline stage $\leftarrow$ Inferred prototypes ρ_c ;
Prediction stage $\leftarrow$ Predicted labels $\hat{y}_i^\nu$.

Algorithm:
Online stage:
$\phi \leftarrow$ random;
for $e \leftarrow 1$ **to** ϵ **do**
    for $i \leftarrow 1$ **to** P^τ **do**
        $R_i^\tau \leftarrow \Psi_\phi(I_i)$;
        $Z_i \leftarrow \Upsilon_\phi(R_i^\tau)$;
        $\hat{y}_i^\tau \leftarrow$ softmax(Z_i) ;
    $\mathcal{L}(y^\tau, \hat{y}^\tau) \leftarrow - \sum_i y_i^\tau \, log(\hat{y}_i^\tau)$;
    Update ϕ using $\nabla_\phi \mathcal{L}$;

Offline stage:
for $c \leftarrow 1$ **to** 3 **do**
    $\rho_c \leftarrow \frac{1}{P^\tau} \sum_{i=1}^{P^\tau} \Psi_\Theta(I_i)$;

Prediction phase:
for $i \leftarrow 1$ **to** P^ν **do**
    $R_i^\nu \leftarrow \Psi_\Theta(V_i)$;
    for $c \leftarrow 1$ **to** 3 **do**
        $\delta_{i,c} \leftarrow \sqrt{(\rho_c - R_i^\nu)^2}$;
        $p_{i,c} \leftarrow \frac{\exp(-\delta_{i,c})}{\sum_{c'} \exp(-\delta_{i,c'})}$;
    $\hat{y}_i^\nu \leftarrow argmax(p_{i,c})$

3.2.2. Dynamic prototypes

Inspired by [66], where prototypical neural networks (PNNs) were proposed for few-shot learning, we present in this paper a PNN-based framework for grading glaucoma by exploiting the *k*-shot methodology. The main difference with respect to the previous static approach lies in the online stage since dynamic prototypes are trained in an end-to-end manner, such that prototypes are updated after each epoch e. In this way, the base encoder network can be optimized according to latent distances, instead of a conventional classification top model, as observed in Fig. 5.
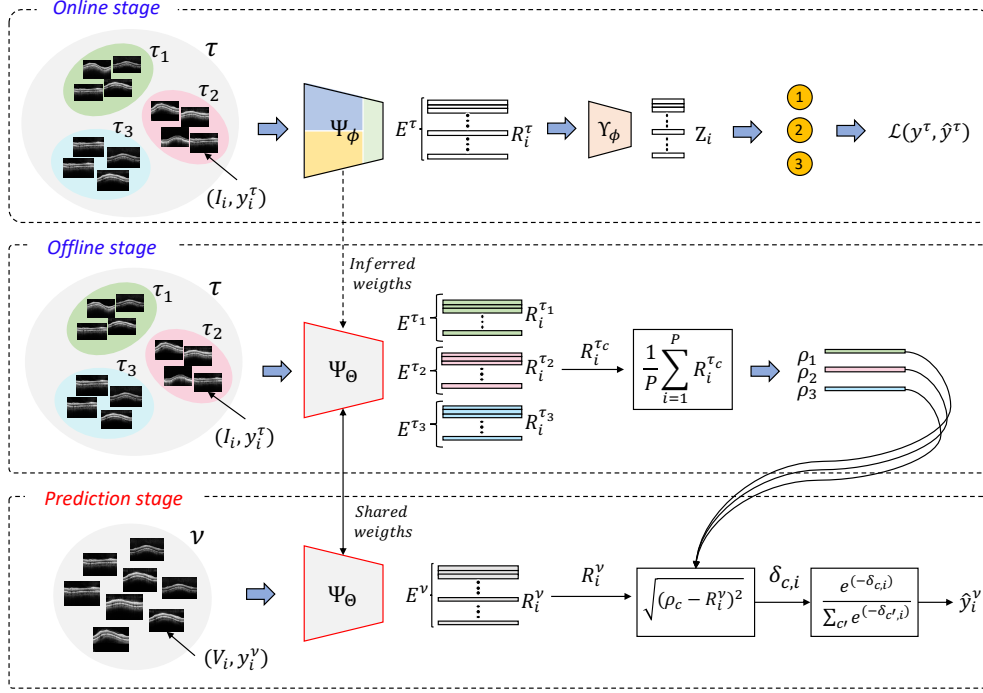
Figure 4: Proposed static prototype-based learning strategy. A conventional approach is conducted in the online stage to optimize the encoder. In the offline stage, the weights of the pre-trained backbone are inferred to extract a prototype ρ_c per class. In the prediction phase, the class of a validation B-scan V_i is determined by measuring the latent distance between each prototype ρ_c and the embedding representations $R_i^\nu = \Psi_\phi(V_i)$.

From here on, we will employ the terminology ($\mathcal{N}$-way, $\mathcal{K}$-shot) used in the literature for few-shot learning, where $\mathcal{K}$ is the number of labelled samples and $\mathcal{N}$ the number of classes in the training set [66]. In the state-of-the-art studies based on prototypical networks [66, 83, 84], a labelled support set $\mathcal{S} = \{(s_1, y_1), ..., (s_i, y_i), ..., (s_\mathcal{K}, y_\mathcal{K})\}$ and an unlabelled query set $\mathcal{Q} = \{q_1, q_2, ..., q_i, ..., q_\mathcal{U}\}$, are considered to train the PNN-based models in a few-shot scenario, being $\mathcal{U}$ the number of unlabelled samples selected from the training set τ. Specifically, $\mathcal{S}$ is used to extract each of $\mathcal{N}$ prototypes ρ_c, whereas $\mathcal{Q}$ is employed to find the nearest class prototype for an embedded query point $R_i^Q = \Psi_\phi(q_i)$. A negative log-probability loss is updated according to the distance metrics from ρ_c and R_i^Q.

The proposed dynamic prototype-based learning strategy differs from the state of the art in multiple ways. Unlike aforementioned studies [66, 83, 84], where authors selected a specific number of $\mathcal{U} = 5, 10$ or 15 unlabelled samples, depending on the database addressed, in this study we made use of $\mathcal{U} = P - \mathcal{K}$ labelled query samples to get the most out of the training set τ. Note that we propose, for the first time, the use of a labelled query set $\mathcal{Q}$, giving rise to a novel Supervised Proto-

typical Neural Network for Glaucoma Grading (SPNN-GG). This results in another difference concerning the literature, since a supervised approach allows including the ground truth label y^Q in the loss function, so learning proceeds by minimizing the categorical cross-entropy, instead of a log-probability function. Also, as a novelty, we move away from the few-shot setting by proposing an optimal $k-$shot scenario in which $\mathcal{K} \in [1, P - \mathcal{U}]$ is optimized as an additional hyper-parameter, unlike in the previous studies where $\mathcal{K} \in [1, 5]$.

The online training of the proposed dynamic prototype-based approach is conducted in a supervised end-to-end way, according to Fig. 5. It is important to remark that the main difference between static and dynamic prototypes lies here since the static approach infers rigid prototypes from a backbone trained by a projection head module, whereas the dynamic strategy updates the prototypes during the training phase by optimizing a backbone based on latent distances. Also note that offline and prediction phases are addressed as in the static approach, i.e. using the entire training set τ to extract the prototypes ρ_c, as detailed in Algorithm 2.

From the training set τ containing P_c^τ samples for each class c, a support set $\mathcal{S} = \{(s_1, y_1^S), ..., (s_\mathcal{K}, y_\mathcal{K}^S)\}$
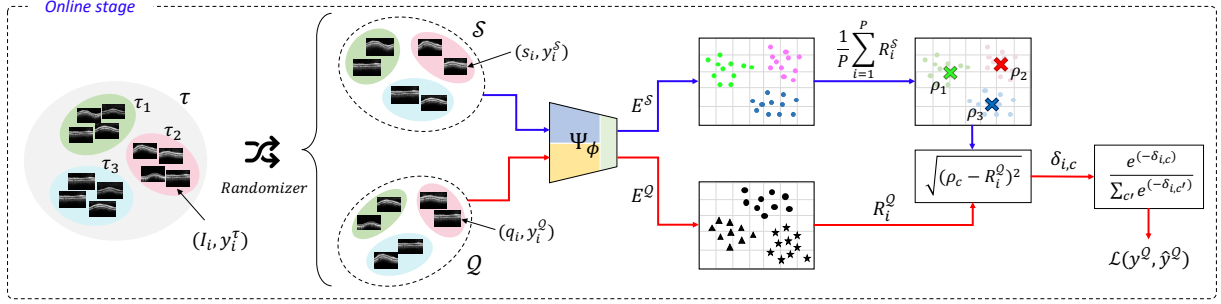
8

Figure 5: Flowchart of the online stage corresponding to the dynamic prototype-based learning strategy. A support set $\mathcal{S}$ and a query set $\mathcal{Q}$ are randomly selected from the training set τ to develop a supervised PNN. The hybrid OCT-based backbone Ψ_ϕ is used to extract the embedding representations from both $\mathcal{S}$, to determine the prototypes ρ_c, and $\mathcal{Q}$, to map the latent representations from the query samples. At the end of the training process, a softmax function is applied to predict the query label based on the latent distances $\delta_{i,c}$.

and a query set $\mathcal{Q} = \{(q_1, y_1^Q), ..., (q_\mathcal{U}, y_\mathcal{U}^Q)\}$, with $\mathcal{U} = P - \mathcal{K}$, are obtained after a randomization process at every epoch e. The hybrid OCT-based backbone Ψ_ϕ was used as the encoder network to extract the embedding representations from the support s_i and query q_i samples. Specifically, latent support set-coming features $E^\mathcal{S} = \Psi_\phi(\mathcal{S}) = \{R_1^\mathcal{S}, ..., R_i^\mathcal{S}, ...R_\mathcal{K}^\mathcal{S}\}$ are used to extract each class-prototype ρ_c as the mean of the embedded representations $R_i^\mathcal{S} = \Psi_\phi(s_i)$. In contrast, query representations $E^\mathcal{Q} = \Psi_\phi(\mathcal{Q}) = \{R_1^Q, ..., R_i^Q, ...R_\mathcal{U}^Q\}$ are mapped in the latent space to find the closest prototype ρ_c, in terms of Euclidean distance. Then, a softmax function is applied to determine the probability of belonging to each class $p(\hat{y}_i^Q = c|R_i^Q)$. During the backward-propagation step, the embedding representations are refined by updating the weights of the base encoder network at every epoch, according to the categorical cross-entropy loss function, denoted by $\mathcal{L}(y^Q, \hat{y}^Q)$. In this way, prototypes are optimized under the hypothesis that each class can be described by just one subspace. Therefore, learning progresses by minimizing the latent distances between $R_i^{Q_c \subset \tau_c}$ and ρ_c, unlike in the static approach where the encoder was updated according to the embedding representations $Z = \Upsilon(\Psi(I_i))$ extracted from the projection network.

4. Ablation experiments

4.1. Data sets

Two private databases coming from different sources were used to develop and evaluate the predictive models for glaucoma grading. Both data sets (*Database 1* and *Database 2*) contain high-resolution SD-OCT scans, which were acquired from healthy, early and advanced-glaucomatous patients with an axial resolution of 3-4

μm using the Heidelberg Spectralis OCT system. This equipment provides circumpapillary B-scans centred on the optical nerve head (ONH) of the retina throughout a super-luminescence diode with an infrared beam of an average wavelength of 870 nm and a bandwidth of 25 nm. In particular, the samples were extracted with a resolution of 496×768 pixels.

Note that subjects with open-angle glaucoma (POAG) were included in the study, whereas patients suffering other ocular disorders, such as closed-angle glaucoma or pseudoexfoliation syndrome were discarded from both databases. Also, patients with media opacity were excluded if the opacity disturbs the B-scan OCT imaging critically. Thus, B-scans with a poor-quality OCT image were discarded from the study. Two different senior ophthalmologists (with more than 25 years of professional experience in clinical ophthalmology) carried out the annotation of the databases. Specifically, each expert manually labelled just one data set, following the European guideline for Glaucoma diagnosis. The examination included several tests such as Goldman applanation tonometry, gonioscopy, slit lamp examination, standard automated perimetry and thickness measurement of specific retinal layers of interest. According to the clinical literature [64, 65], the mean deviation (MD) score plays an essential role in the glaucoma grading scale, such that the severity of the glaucoma depends on the range in which the MD value is found. Conforming to [64], $MD >= -6dB$ is Early; $-6dB > MD \geq -12dB$ is Moderate; $-12dB > MD \geq -20dB$ is Advanced and $MD < -20dB$ is Severe. Since our objective is contributing to the glaucoma grading just from OCT images, without performing additional time-consuming tests, we simplify the glaucoma staging scale by labelling as Advanced those glaucomatous

9

Algorithm 2: Dynamic prototype-based learning strategy.

Data: Training $\tau = \{(I_1, y_1^\tau), ..., (I_{P^\tau}, y_{P^\tau}^\tau)\}$ and
validation $\nu = \{(V_1, y_1^\nu), ..., (V_{P^\nu}, y_{P^\nu}^\nu)\}$ sets.

Results:
Online stage $\leftarrow$ Trained base encoder Ψ_ϕ ;
Offline stage $\leftarrow$ Refined prototypes ρ_c ;
Prediction stage $\leftarrow$ Predicted labels $\hat{y}_i^\nu$.

Online stage:
$\phi \leftarrow$ random;
for $e \leftarrow 1$ **to** ϵ **do**
  for $i \leftarrow 1$ **to** $\mathcal{K}$ **do**
    $(s_i, y_i^S) \leftarrow (I_{random(i)}, y_{random(i)}^\tau)$;
    $R_i^S \leftarrow \Psi_\phi(s_i)$;
  for $i \leftarrow 1$ **to** $\mathcal{U}$ **do**
    $(q_i, y_i^Q) \leftarrow (I_{random(i)}, y_{random(i)}^\tau) \notin \mathcal{S}$;
    $R_i^Q \leftarrow \Psi_\phi(q_i)$;
    for $c \leftarrow 1$ **to** $\mathcal{N}$ **do**
      $\rho_c \leftarrow \frac{1}{\mathcal{K}} \sum_{i=1}^{\mathcal{K}} R_i^S$;
      $\delta_{i,c} \leftarrow \sqrt{(\rho_c - R_i^Q)^2}$;
      $p_{i,c} \leftarrow \frac{\exp(-\delta_{i,c})}{\sum_{c'} \exp(-\delta_{i,c'})}$;
    $\hat{y}_i^Q \leftarrow argmax(p_{i,c})$
  $\mathcal{L}(y^Q, \hat{y}^Q) \leftarrow -\sum_u y_i^Q \, log(\hat{y}_i^Q)$;
  Update ϕ using $\nabla_\phi \mathcal{L}$;

Offline stage:
for $c \leftarrow 1$ **to** N **do**
  $\rho_c \leftarrow \frac{1}{P^\tau} \sum_{i=1}^{P^\tau} \Psi_\Theta(I_i)$;

Prediction phase:
for $i \leftarrow 1$ **to** $\mathbf{P}^\nu$ **do**
  $R_i^\nu \leftarrow \Psi_\Theta(V_i)$;
  for $c \leftarrow 1$ **to** **3** **do**
    $\delta_{i,c} \leftarrow \sqrt{(\rho_c - R_i^\nu)^2}$;
    $p_{i,c} \leftarrow \frac{\exp(-\delta_{i,c})}{\sum_{c'} \exp(-\delta_{i,c'})}$;
  $\hat{y}_i^\nu \leftarrow argmax(p_{i,c})$;

Table 1: Number of patients (pat.) and samples (samp.) in each database grouped by categories, according to the experts' annotation.

	Healthy	Early	Advanced	TOTAL
	(pat./samp.)	*(pat./samp.)*	*(pat./samp.)*	*(pat./samp.)*
Database 1	32 / 41	28 / 35	25 / 31	85 / 107
Database 2	26 / 49	24 / 37	21 / 26	71 / 112
TOTAL	58 / 90	52 / 72	46 / 57	156 / 219

Table 2: Additional information about the age and gender of the patients who compose each database.

	Age		Gender	
	Range	$\mu \pm \sigma$	*Male*	*Female*
Database 1	[19-88]	60,45±16,54	46 (54,12%)	39 (45,88%)
Database 2	[30-90]	64,80±13,93	26 (36,62%)	45 (63,38%)
TOTAL	[19-90]	62,44±15,51	72 (46,15%)	84 (53,85%)

tute of Ophthalmology (AFIO) data set [63] is the only publicly available repository of ONH SD-OCT scans of healthy and glaucomatous subjects. However, the differences between those B-scans and ours make the direct application of our algorithms impossible for multiple reasons: i) B-scans from AFIO data set present manual annotations and highlighted structures of interest. ii) OCT images have been pre-processed showing an RGB colour mode in which cup-to-disk regions appears remarked. iii) AFIO database was acquired using Topcon 3D OCT-1000 machines. iv) The experts' annotations include healthy, glaucoma and suspect labels. Differently, the databases used here contain the grayscale OCT samples extracted in raw by the Heidelberg Spectralis OCT equipment, without any manual annotation or pre-processing. In addition, our database is explicitly labelled into healthy, early and advanced glaucoma classes, according to the visual field-based criteria of the medical literature [64, 65].

Furthermore, another OCT database (OCTID) (with raw ONH B-scans similar to ours) is publicly available in [85]. However, OCTID data set includes Normal (NO), Macular Hole (MH), Age-related Macular Degeneration (AMD), Central Serous Retinopathy (CSR) and Diabetic Retinopathy (DR) image classes, but it does not contemplate the glaucoma class.

Data partitioning. In this paper, we fuse *Database 1* and *Database 2* in order to increase the number of samples from which to develop our machine learning algorithms and to evidence the reliability of the predictive models using two data sets coming from different sources. Making use of the entire fused database, we conducted a patient-level data partitioning procedure to separate training and testing sets. Specifically, $\frac{1}{6}$ of the data was used to test the models, whereas the remainder of the database was employed to train the algorithms.

samples with an $MD < -6dB$. Note that the proposed system is not intended to serve as a definitive glaucoma diagnosis, but as a diagnositc tool which allows guiding the expert's decision through approximate but reliable OCT-based results. Based on the above analyses, all B-scans were classified as healthy, early or advanced. In Tables 1 and 2, we show more information about the specifications of each data set.

It is important to note that, as claimed in [62], there are no public glaucoma-labelled OCT databases that enable an objective comparison with our work. To the best of the authors' knowledge, Armed Forces Insti-

From the training set, we randomly split the data again into training and validation subsets, according to Table 3, to optimise the models' hyper-parameters and monitor the over-fitting.

Table 3: Partition of the circumpapillary B-scans to develop and evaluate the predictive models.

	Healthy	Early	Advanced
Training	60	48	41
Validation	15	12	10
Test	15	12	6

4.2. Backbone selection

Unlike most of the state-of-the-art studies which used as a feature extractor either well-known architectures such as ResNet or VGG [77, 84] or simpler CNNs trained from scratch [66, 67], we pretend to exploit the feature extraction stage to get the most out from the circumpapillary OCT scans. For this reason, we propose a novel OCT-hybrid backbone inspired by ophthalmic clinical studies [74] and our previous glaucoma detection-based experience applying hand-driven [20] and deep learning algorithms [41, 45]. To address an objective comparison with other state-of-the-art studies, we contrast in Tables 4 and 5 the validation results achieved by different architectures trained in a multi-class scenario. Particularly, we compared four approaches, as detailed below.

1. *RNFL features.* A simple MLP was trained using the output of the hand-driven learning branch $\Omega_\phi(I)$ as an input data, similarly to [20].
2. *Fine-tuned VGG16.* This very popular architecture was used (freezing the three first convolutional blocks) as a feature extractor followed by the same MLP classifier as before, according to [45].
3. *Fine-tuned RAGNet.* An expanded version of the previous approach was conducted by including residual and attention modules in the feature extraction architecture. This approach, corresponding to the deep learning branch $\Gamma_\phi(I)$, was introduced in our previous work [41].
4. *OCT-hybrid network.* This approach corresponds to the end-to-end backbone $\Psi_\phi(I)$ exposed in Fig. 2, which proposes a combination of hand-crafted and automatic features before the top model.

The comparison was handled by means of different figures of merit, such as sensitivity (SN), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), F-score (FS) and accuracy (ACC).

Notably, the backbone reporting the best performance during the validation stage was selected as the base encoder network to address the next prototypical-based learning strategy.

Training details. All the contrasting approaches were implemented using Tensorflow 2.3.1 on Python 3.6. Experiments were conducted on a machine with Intel(R) Core(TM) i7-9700 CPU @3.00GHz processor and 16GB RAM. A single NVIDIA GeForce RTX 2080 having cuDNN 7.5 and a CUDA Toolkit 10.1 was used to develop the deep learning algorithms. All models were trained during 200 epochs using a learning rate of 0.0005 with a batch size of 16. Stochastic gradient descent (SGD) optimizer was applied trying to minimize the categorical cross-entropy (CCE) loss function at every epoch. The rest of the architecture hyper-parameters and input dimensions are shown in Fig. 2.

4.3. Prototype-based learning strategies

In this section, we report the validation performance of the static and dynamic prototype-based methods in comparison to the conventional multi-class approach. It is important to note that the comparison was conducted using the proposed OCT-hybrid backbone as a feature extractor for all the scenarios. Following the organization of the previous section, Tables 6 and 7 show the comparison between the three involved learning strategies during the validation phase.

Training details. The same hardware and software systems as before were used to accomplish this section. However, some differences in the dynamic prototypical approach are worth noting. Dimensions of the input image must be downsized to $124 \times 192 \times 3$ to face the GPU memory constraints. Additionally, a decreased learning rate of 0.001 allowed the convergence of the model in 50 training epochs. A batch size of 16 samples was defined to minimize the CEE loss function using the SGD optimizer. Regarding the specific parameters of the dynamic prototype-based strategy, the number of $\mathcal{K}$ shots and the number of $\mathcal{U}$ query samples were determined after an optimization process, according to Fig. 6. Specifically, $\mathcal{K} = 20$ support samples were selected to extract the prototypes ρ_c as the mean of the embedding representations E^S, whereas $\mathcal{U} = 21$ query samples were used to measure the Euclidean distance between the latent features E^Q and each ρ_c. Note that $P = 41$ denotes the number of training samples of the minority class (see Table 1). Also, other statistics and distance metrics were considered during the optimization of the models, as observed in Table 8. The rest of
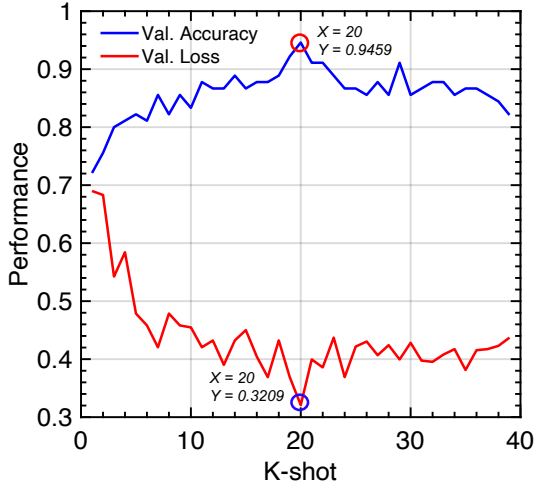
Figure 6: Models' performance using $\mathcal{K}$ samples from each class c to define the prototypes ρ_c and $\mathcal{U} = P - \mathcal{K}$ query samples to measure the embedding distances to minimize the loss function.

Figure 7: Confusion matrix obtained by the dynamic prototype-based approach during the prediction phase.

the hyper-parameters related to the dynamic prototypes is detailed in Section 3.2.2.

5. Prediction results

5.1. Quantitative results

In this section, we show the quantitative results achieved by the three learning strategies conducted during the prediction of the test set. It is worth noting that all the approaches contrasted here were addressed using the proposed OCT-based hybrid backbone as a feature extractor since it reported the best results in the validation phase. As before, we evaluate the models' performance both per class (Table 9) and in terms of average (Table 10), using different figures of merit. Additionally, in Fig. 7, we show the confusion matrix obtained by the best approach, i.e. the dynamic prototype-based model, to evidence the overall behaviour of the proposed method when predicting new samples.

Also, to provide a more comprehensive interpretation of the glaucoma grading scenario, we illustrate in Fig. 8 a 2D map corresponding to the latent space arranged by the dynamic learning. In particular, the prototypes (denoted by asterisks) were calculated from the training and validation sets, whereas spots and crosses make reference to the embedding representations of the well and miss-classified test data, respectively. In addition to this, Euclidean distance-based probabilities reported from the miss-classified samples are also detailed in Fig. 8 to manifest the confidence of the dynamic model when it is wrong in the prediction.
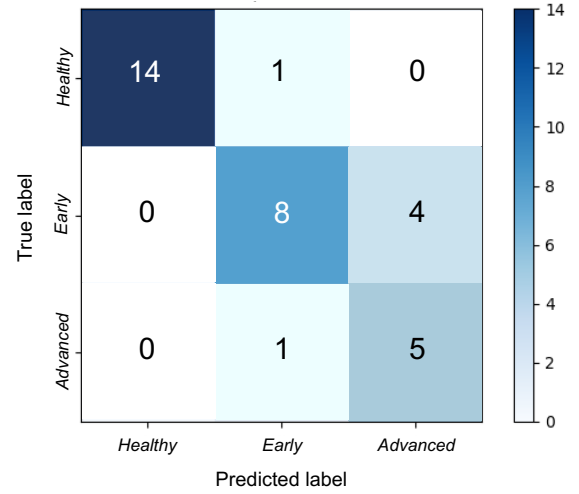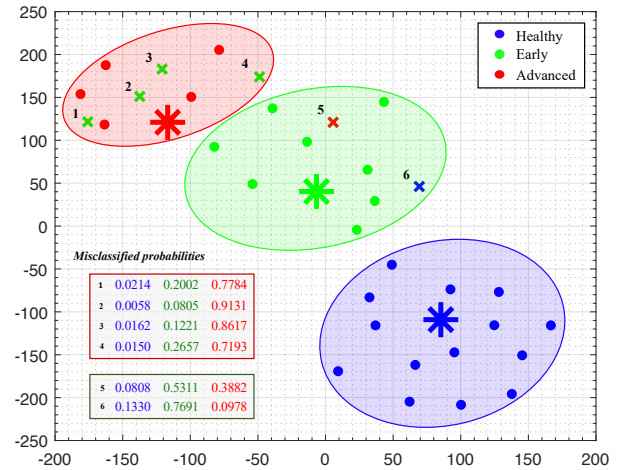

Figure 8: Latent space showing the prototypes and the embedding test features on a 2D map via t-distributed Stochastic Neighbor Embedding (TSNE) tool. Blue, green and red colours denote the ground truth labels of the healthy, early and advanced representations, respectively.

5.2. Qualitative results

Class activation maps (CAMs) [86] were computed to remark the regions in which the proposed dynamic prototypical network paid attention to predict the class of the test samples. The reported heatmaps allow a better understanding of the CNN-extracted features by highlighting the most relevant information of the B-scan for the predictions. These activated maps may provide an additional interpretation of the results for glaucoma grading depending on the patterns highlighted for each

12

Table 4: Backbone selection: Validation results per class achieved from different architectures in a multi-class scenario.

	HEALTHY				EARLY GLAUCOMA				ADVANCED GLAUCOMA			
	RNFL features [20]	*Fine-tuned VGG16 [45]*	*Fine-tuned RAGNet [41]*	*OCT-hybrid network*	*RNFL features [20]*	*Fine-tuned VGG16 [45]*	*Fine-tuned RAGNet [41]*	*OCT-hybrid network*	*RNFL features [20]*	*Fine-tuned VGG16 [45]*	*Fine-tuned RAGNet [41]*	*OCT-hybrid network*
SN	**1**	**1**	0.9333	**1**	0.6667	0.5833	**0.7500**	**0.7500**	0.8000	**0.9000**	**0.9000**	**0.9000**
SP	0.8636	0.8636	**0.9545**	0.9091	0.9200	**0.9600**	0.9200	**0.9600**	**0.9630**	0.9259	0.9259	**0.9630**
PPV	0.8333	0.8333	**0.9333**	0.8824	0.8000	0.8750	0.8182	**0.9000**	0.8889	0.8182	0.8182	**0.9000**
NPV	**1**	**1**	0.9545	**1**	0.8519	0.8276	0.8846	**0.8889**	0.9286	0.9615	0.9615	**0.9630**
FS	0.9091	0.9091	0.9333	**0.9375**	0.7273	0.7000	0.7826	**0.8182**	0.8421	0.8571	0.8571	**0.9000**
ACC	0.9189	0.9189	**0.9459**	**0.9459**	0.8378	0.8378	0.8649	**0.8919**	0.9189	0.9189	0.9189	**0.9459**

Table 5: Backbone selection: Validation results from different multi-class approaches in terms of micro and macro-averages.

	Micro-Average				Macro-Average			
	RNFL features [20]	*Fine-tuned VGG16 [45]*	*Fine-tuned RAGNet [41]*	*OCT-hybrid network*	*RNFL features [20]*	*Fine-tuned VGG16 [45]*	*Fine-tuned RAGNet [41]*	*OCT-hybrid network*
SN	0.8378	0.8378	0.8649	**0.8919**	0.8222	0.8278	0.8611	**0.8833**
SP	0.9189	0.9189	0.9324	**0.9459**	0.9155	0.9165	0.9335	**0.9440**
PPV	0.8378	0.8378	0.8649	**0.8919**	0.8407	0.8422	0.8566	**0.8941**
NPV	0.9189	0.9189	0.9324	**0.9459**	0.9268	0.9297	0.9336	**0.9506**
FS	0.8378	0.8378	0.8649	**0.8919**	0.8262	0.8221	0.8577	**0.8852**
ACC	0.8919	0.8919	0.9099	**0.9279**	0.8919	0.8919	0.9099	**0.9279**

Table 6: Learning strategy: Validation results per class using the proposed OCT-hybrid backbone for glaucoma grading.

	HEALTHY			EARLY GLAUCOMA			ADVANCED GLAUCOMA		
	Conventional multi-class	*Static prototypes*	*Dynamic prototypes*	*Conventional multi-class*	*Static prototypes*	*Dynamic prototypes*	*Conventional multi-class*	*Static prototypes*	*Dynamic prototypes*
SN	**1**	**1**	**1**	0.7500	0.7500	**0.8333**	**0.9000**	0.8000	**0.9000**
SP	0.9091	0.9091	**1**	**0.9600**	0.9200	**0.9600**	**0.9630**	**0.9630**	0.9259
PPV	0.8824	0.8824	**1**	0.9000	0.8182	**0.9091**	**0.9000**	0.8889	0.8182
NPV	**1**	**1**	**1**	0.8889	0.8846	**0.9231**	**0.9630**	0.9286	0.9615
FS	0.9375	0.9375	**1**	0.8182	0.7826	**0.8696**	**0.9000**	0.8421	0.8571
ACC	0.9459	0.9459	**1**	0.8919	0.8649	**0.9189**	**0.9459**	0.9189	0.9189

Table 7: Learning strategy: Validation results in terms of micro and macro-averages using the proposed OCT-hybrid backbone.

	Micro-Average			Macro-Average		
	Conventional multi-class	*Static prototypes*	*Dynamic prototypes*	*Conventional multi-class*	*Static prototypes*	*Dynamic prototypes*
SN	0.8919	0.8649	**0.9189**	0.8833	0.8500	**0.9111**
SP	0.9459	0.9324	**0.9595**	0.9440	0.9307	**0.9620**
PPV	0.8919	0.8649	**0.9189**	0.8941	0.8631	**0.9091**
NPV	0.9459	0.9324	**0.9595**	0.9506	0.9377	**0.9615**
FS	0.8919	0.8649	**0.9189**	0.8852	0.8541	**0.9089**
ACC	0.9279	0.9099	**0.9459**	0.9279	0.9099	**0.9459**

Table 8: Validation accuracy reached by the dynamic prototypical approach using different statistics and distance metrics.

Distance	Euclidean	Cosine	Manhattan	Canberra	**Euclidean**	Cosine	Manhattan	Canberra
Statistic	median	median	median	median	**mean**	mean	mean	mean
Accuracy	0.9279	0.9279	0.9099	0.8559	**0.9459**	0.9099	0.9279	0.8739

class. However, it should be mentioned that CAMs usually do not present a high precision at pixel-level since the heatmaps are created from the last layers of the proposed model, which output a down-sampled map from the input image.

In Fig. 9, we show several examples of CAMs corresponding to correctly and wrongly predicted samples to elucidate the relevant patterns found by the dynamic prototypical network to address the B-scans prediction. Specifically, we expose three well-classified heatmaps for each class to demonstrate the criteria followed by the proposed model to determine the predicted labels. Also, we report in the red frame of Fig. 9 some examples of miss-classified B-scans to visually evidence the reason why the model gets wrong in the prediction. Note that rows and columns in the illustration represent the predictions and labels, respectively.

6. Discussion

It should be noted that OCT B-scans contain information too limited to provide a complex diagnosis based on glaucoma grading. So, additional tests are usually carried out to accomplish a more exhaustive and reliable glaucoma grading diagnosis considering all the stages of the disease. Currently, the OCT systems can provide a result indicating if the B-scan is within normal limits, borderline or outside normal limits, which is equivalent to discern between healthy, suspect and glaucomatous cases. At this point, our system has demonstrated to provide a very high performance achieving an accuracy of 100% and 96, 97% during the validation and testing phases, respectively. However, with the proposed system, we pretend to go beyond the binary classification of healthy vs glaucoma by means of a prototypical framework able to discern early from advanced glaucoma only using OCT samples, which adds significant value to the body of knowledge. Below, we discuss about the different CNN configurations and learning strategies carried out in this work for glaucoma grading.

6.1. About the ablation experiments

Backbone selection. One of the main novelties addressed in this paper lies in the proposed OCT-based hybrid network, which combines hand-crafted and automatic features extracted from the input B-scans. In order to elucidate the superiority of the proposed model with respect to similar approaches, we raise a multi-class glaucoma-grading scenario comparing different methodologies. Note that a direct comparison with other state-of-the-art studies was not possible because

of the lack of public repositories with annotations of different glaucoma severity levels, as claimed in [62]. For that reason, in this paper, we contrasted the proposed model with other own glaucoma-detection methods recently published in [20, 41, 45]. It should be mentioned that all the approaches were adapted to the multi-class environment to provide a reliable comparison during the validation stage, as detailed in Tables 4 and 5. Particularly, from Table 4 we can observe that the fine-tuned RAGNet model [41] introduces slight improvements as compared with the traditional VGG16 architecture and the hand-crafted RNFL features. The use of tailored residual blocks and attention modules allows surpassing the results for most of the metrics by providing more distinctive feature maps. However, the proposed OCT-based hybrid network outperforms the rest of the models in the discrimination of all the classes. Focusing on the healthy column, RNFL features, VGG16 architecture and the proposed hybrid backbone report a more sensible behaviour highlighting for the SN and NPV metrics, whereas RAGNet model showcases more specific results by outstanding for the SP and PPV figures of merit. More global metrics, such as FS and ACC show higher values for RAGNet and OCT-hybrid approaches. Contrarily, the models differ to a greater extent in the discrimination of the glaucoma severity grades. Specifically, the proposed OCT-hybrid network report the best performance distinguishing between early and advanced glaucomatous samples for all the measures. The results detailed in Table 5 further strengthens our confidence in the proposed OCT-hybrid network since it reports the higher values for all figures of merit in terms of micro and macro-average. Table 5 is especially interesting to compare the model's performance since it gives an idea of their overall precision. In such a table, we can observe that hand-driven and basic deep learning approaches present a similar behaviour, whereas more sophisticated CNNs provide substantial performance improvements. Nevertheless, the combination of hand-crafted RNFL-based features and refined CNN architectures yields the best multi-class model for glaucoma grading, achieving an average accuracy of 0.9279.

Prototype-based learning strategy. The best approach reported during the validation phase was used as a feature extractor to address the next stage corresponding to the prototype-based learning strategy. So, according to the previous statements, the proposed OCT-hybrid network Ψ_ϕ was selected as the backbone of the prototypical architectures. At this point, several contributions are proposed in a novel framework for glaucoma grading. The method based on static prototypes

Table 9: Prediction stage: Test results per class reached by the different proposed learning strategies for glaucoma grading.

	HEALTHY			EARLY GLAUCOMA			ADVANCED GLAUCOMA		
	Conventional multi-class	*Static prototypes*	*Dynamic prototypes*	*Conventional multi-class*	*Static prototypes*	*Dynamic prototypes*	*Conventional multi-class*	*Static prototypes*	*Dynamic prototypes*
SN	**0.9333**	**0.9333**	**0.9333**	0.5000	0.4167	**0.6667**	1	0.8333	0.8333
SP	1	1	1	**0.9524**	0.9048	0.9048	0.7778	0.7407	**0.8519**
PPV	1	1	1	**0.8571**	0.7143	0.8000	0.5000	0.4167	**0.5556**
NPV	0.9474	0.9474	0.9474	0.7692	0.7308	0.8210	1	0.9524	0.9583
FS	0.9655	0.9655	0.9655	0.6316	0.5263	**0.7273**	0.6667	0.5556	**0.6667**
ACC	**0.9697**	**0.9697**	**0.9697**	0.7879	0.7273	**0.8182**	0.8182	0.7576	**0.8485**

Table 10: Prediction stage: Test results reached in terms of micro and macro-average by the different proposed learning strategies.

|	Micro-Average			Macro-Average		
	Conventional multi-class	*Static prototypes*	*Dynamic prototypes*	*Conventional multi-class*	*Static prototypes*	*Dynamic prototypes*
SN	0.7879	0.7273	**0.8182**	0.8111	0.7278	**0.8112**
SP	0.8939	0.8636	**0.9091**	0.9101	0.8818	**0.9189**
PPV	0.7879	0.7273	**0.8182**	0.7852	0.7103	**0.7857**
NPV	0.8939	0.8636	**0.9091**	0.9055	0.8768	**0.9106**
FS	0.7879	0.7273	**0.8182**	0.7446	0.6825	**0.7865**
ACC	0.8586	0.8182	**0.8788**	0.8586	0.8182	**0.8788**
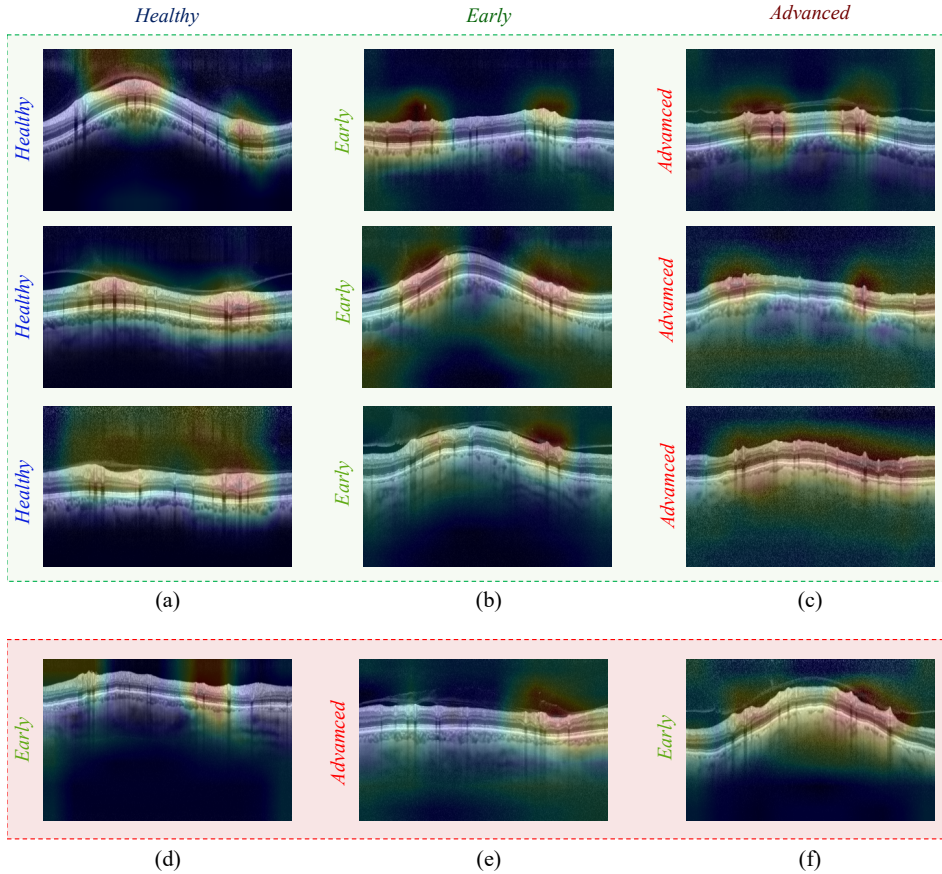
Figure 9: CAMs showing the regions of interest in which the proposed prototypical-dynamic model pays attention to predict each class. Heatmaps over the green background (a-c) correspond to well-classified images, whereas the red frame (d-f) represents the miss-classified samples.

15

was conducted for the first time by inferring the weights ($\phi \rightarrow \Theta$) of class-based prediction networks, instead of auto-encoders, as in the case of CBIR-based studies [80–82], or architectures intended to discern between positive and negative classes, as in the contrastive learning works [77–79]. Notwithstanding, the main novelties related to the prototypical environment were introduced in the dynamic approximation. As detailed in Section 3.2.2, state-of-the-art studies usually made use of $\mathcal{K} \in [1, 5]$ labelled images to extract the prototypes and $\mathcal{U} = 5$, 10 or 15 unlabelled query samples to measure the latent distance to the prototypes during the training of the models. Unlike them, in this paper, the few-shot paradigm was redefined by dealing the $\mathcal{K}$ and $\mathcal{U} = P - \mathcal{K}$ variables as any other hyper-parameter to be optimized in the validation phase. As another novelty, we made use of labelled samples to build the query set $\mathcal{Q}$ in order to make the most of all available information about the B-scans for glaucoma grading. Our hypothesis claimed that the use of more than $\mathcal{K} = 5$ support images to extract the prototypes would provide better results. However, it would be necessary to find the optimal balance between the number of support and query samples, because the data used to extract the prototypes ρ_c is just as important as the data employed to know how well the prototypes work. For that reason, we show in Fig. 6 the categorical accuracy and the loss value achieved by the dynamic prototypical model using $\mathcal{K} \in [1, P]$ samples, being $P = 41$ the total number of the training samples of the minority class (see Table 3). As expected, the model's performance using a few $\mathcal{K}$ samples remains low and it improves as the number of support samples increases. However, the model reaches the best in the middle of the plotting, and then, the curve stabilizes or even worsens. This fact consolidates our hypothesis which holds the importance of a similar splitting of the support and query samples to train the models. In particular, the highest performance is reached using $\mathcal{K} = 20$ shots and $\mathcal{U} = 21$ query scans, which reports a validation accuracy of 0.9459 and a loss value of 0.3209, according to Fig. 6.

In Tables 6 and 7, both static and dynamic prototype-based approaches are compared with a conventional multi-class system based on the OCT-hybrid network trained for glaucoma grading. The classification of healthy samples in Table 6 deserves special attention since the dynamic prototypical model perfectly works for detecting the non-glaucomatous data, achieving the 100% of performance for all figures of merit. The discrimination of the middle class, i.e. early class, makes the outperforming of the dynamic prototypical network evident. However, the conventional multi-class strategy exceeds the results for advanced glaucoma cases in all the metrics. Similar results can be appreciated in Table 7, which provides a comparison of the models' behaviour in terms of micro and macro-average. Specifically, the dynamic prototype reaches the best results with values higher than 0.90 for all the figures of merit. It should be highlighted that the baseline approach outperforms the strategy based on static prototypes, which reveals that end-to-end methodologies provide more robust models, as expected.

Furthermore, we carried out an empirical exploration of different distance metrics and statistical parameters to maximize the classification agreement in the latent space between the prototypes and the embedding query representations. Particularly, statistics based on mean and median operations were considered to extract the prototypes from the support samples, whereas Euclidean, Cosine, Manhattan and Canberra distances were subject to study to determine the class of the query representations according to the closest prototype in the latent space. As detailed in Table 8, the highest validation accuracy was reached using the Euclidean distance and the mean statistic, which reported a validation accuracy of 0.9459.

6.2. About the prediction results

Quantitative results. In this section, we compare the performance of the different proposed learning strategies during the prediction of the test set. As in the validation phase, results per class and in terms of micro and macro-average are reported in Tables 9 and 10, respectively. According to the previous section, the discrimination between healthy and glaucoma classes is successfully accomplished by achieving results higher than 0.93 in the prediction of healthy samples for all figures of merit (see Table 9). It is remarkable that the three contrasted models provide the same effectiveness in distinguishing the healthy class. However, the glaucoma grading results make clear the differences between the models. In particular, the dynamic prototypical network shows better SN and NPV results, whereas the baseline approach out stands for SP and PPV metrics in the early glaucoma detection. The opposite happens in the advanced glaucoma case since the conventional approach reveals a more sensitive behaviour (higher recall and NPV) and the dynamic prototypical model provides more specific results (better specificity and precision). Note that F-score (FS) and Accuracy (ACC) metrics reach higher values using the dynamic prototypical strategy. Additionally, the results in Table 10 are directly in line with those reported in the validation

phase, since the dynamic prototypical method is consolidated as the best model, followed by the conventional multi-class approach and leaving the static prototypical network in the last position. From here, we can conclude that end-to-end trained systems are more compelling and allow providing more reliable and robust models.

From the confusion matrix in Fig. 7, promising results are evidenced for glaucoma detection since the dynamic prototype strategy almost perfectly distinguishes glaucomatous from healthy samples. Only a specific healthy B-scan was wrongly predicted as a sample with early glaucoma. However, the proposed model makes more mistakes predicting the early class since it sometimes confuses early and advanced patterns of the disease. Nonetheless, in the case of the advanced glaucoma class, the prototypical network only disagrees with the expert ophthalmologist once, in which a severe glaucomatous scan was classified as early glaucoma.

In order to visualise the results of the confusion matrix in a more interpretable scenario, we illustrated the latent space environment arranged by the proposed dynamic prototypical network in Fig. 8. The TSNE technique allows showing the distribution of the embedding test representations on a 2D map, as well as the prototypes extracted from the training samples. As appreciated in Fig. 8, three well-differentiated sub-spaces arise on the 2D map locating the healthy and advanced representations in the opposite corners and the prototypes of early and advanced classes very close, as expected. Additionally, early glaucoma-related features are mapped in the middle of the plotting, which matches with the order of the glaucoma severity scale. In addition, the prediction probabilities for the miss-classified samples are detailed in Fig. 8 to demonstrate the coherence of the proposed model using the embedding features for glaucoma grading. For example, the nicknamed representations $1 - 4$ correspond to early glaucomatous samples which were miss-classified as advanced glaucoma items with a level of confidence less than 0.8 in some cases. Regarding the tagged representation 5, the dynamic prototypical network wrongly predicted an early B-scan as an advanced sample, but with serious difficulties since the prediction probabilities are 0.53 and 0.39 for early and advanced classes, respectively. Something similar happens for the nicknamed point 6, which was miss-classified as an early glaucoma sample when it was actually healthy. In spite of this, the embedding representation 6 is located near the decision boundary between early and healthy classes, which manifests the robustness of the model despite being wrong.

Qualitative results. The findings from the Class Activation Maps (CAMs) reported in Fig. 9 keep consistency with the clinical interpretation provided by the ophthalmologists, who claim that a thinning of the RNFL structure is usually associated with glaucomatous patterns, whereas a thickening of the RNFL denotes cues of healthy samples [16, 17]. As appreciated in Fig. 9, the proposed model pays special attention to specific regions of the RNFL depending on the reported prediction, according to the outcomes found in our previous work [45].

In the green frame of Fig. 9, three random examples are illustrated to demonstrate that there is a repeating pattern for each of the classes. Well-classified samples corresponding to the healthy class (Fig. 9. (a)) show heatmaps with highlighted pixels in regions characterised by a thickening of the RNFL. Oppositely, well-classified early and advanced glaucomatous samples (Fig. 9. (b), (c)) provide apparent cues of RNFL deterioration, which is visible in the heatmaps highlighting regions where a thinning of the RNFL is evident. As observed, the degradation level of the RNFL thickness is accentuated the greater the severity of the disease.

The wrongly predicted CAMs can be also appreciated in the red frame of Fig. 9. Particularly, the example showed in Fig. 9 (d) shows a healthy B-scan missclassified as an early glaucoma sample. Nevertheless, the model's decision shows to be coherent with the established patterns, since the B-scan presents several areas in which the RNFL appears slightly thinned. Something similar happens with the remainder of cases of the red frame. Specifically, the example in Fig. 9 (e) is striking since the RNFL seems to be widely deteriorated throughout the entire B-scan, so the model associates the image with an advanced glaucomatous sample. The opposite occurs in Fig. 9 (f), where an advanced glaucoma OCT image is predicted with the early label. In this case, the B-scan shows several regions in which the RNFL thickness agrees more with early than advanced glaucoma patterns. Therefore, from the red frame of Fig. 9, it is possible to elucidate the complexity in the diagnosis of the different grades of glaucoma, even for expert ophthalmologists, who often disagree. It is important to remember at this point that, in this paper, we only made use of raw circumpapillary OCT images to build the predictive models. So, in future research lines, it would be interesting to include additional data in order to provide the model with the necessary information that clinicians take into account to determine the severity of the glaucoma disease.

17

7. Conclusion

In this paper, we have proposed several artificial intelligence solutions for glaucoma grading using raw circumpapillay OCT images. A new base encoder network has been developed improving the multi-class methodologies addressed to date for glaucoma severity detection. The proposed model introduces a residual convolutional block with tailored kernel sizes to get the most out from the B-scans and an attention module able to refine the features of the latent space to maximize the classification agreement. Besides, the encoder network uses, for the first time, a combination of hand-crafted and automatic features giving rise to an OCT-based hybrid model which adds substantial improvements in the final prediction. As a novelty, this architecture was employed as the backbone of a novel framework based on prototypical learning, in which the limits of the few-shot paradigm have been redefined for an optimal glaucoma grading procedure. This innovative approach carried out in an end-to-end manner has surpassed the multiclass baseline and it has reported promising results for both glaucoma detection and glaucoma grading scenarios, achieving testing accuracy of 0.9697 and 0.8788, respectively. In future research lines, the efforts should be focused on improving the discrimination of the different grades of glaucoma severity by including additional information outside the SD-OCT images.

References

[1] R. N. Weinreb, P. T. Khaw, Primary open-angle glaucoma, The Lancet 363 (9422) (2004) 1711–1720.

[2] J. B. Jonas, T. Aung, R. R. Bourne, A. M. Bron, R. Ritch, S. Panda-Jonas, Glaucoma–authors' reply, The Lancet 391 (10122) (2018) 740.

[3] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, C.-Y. Cheng, Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis, Ophthalmology 121 (11) (2014) 2081–2090.

[4] A. García-Floriano, Á. Ferreira-Santiago, O. Camacho-Nieto, C. Yáñez-Márquez, A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images, Computers & Electrical Engineering 75 (2019) 218–229.

[5] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, Jama 316 (22) (2016) 2402–2410.

[6] P. M. Burlina, N. Joshi, M. Pekala, K. D. Pacheco, et al., Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks, JAMA ophthalmology 135 (11) (2017) 1170–1176.

[7] F. Grassmann, J. Mengelkamp, C. Brandl, S. Harsch, M. E. Zimmermann, B. Linkohr, A. Peters, I. M. Heid, C. Palm, B. H. Weber, A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography, Ophthalmology 125 (9) (2018) 1410–1420.

[8] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, et al., Optical coherence tomography, science 254 (5035) (1991) 1178–1181.

[9] I. I. Bussel, G. Wollstein, J. S. Schuman, Oct for glaucoma diagnosis, screening and detection of glaucoma progression, British Journal of Ophthalmology 98 (Suppl 2) (2014) ii15–ii19.

[10] F. A. Medeiros, L. M. Zangwill, L. M. Alencar, C. Bowd, P. A. Sample, R. Susanna, R. N. Weinreb, Detection of glaucoma progression with stratus oct retinal nerve fiber layer, optic nerve head, and macular thickness measurements, Investigative ophthalmology & visual science 50 (12) (2009) 5741–5748.

[11] P. R. Lichter, Variability of expert observers in evaluating the optic disc., Transactions of the American Ophthalmological Society 74 (1976) 532.

[12] R. Varma, W. C. Steinmann, I. U. Scott, Expert agreement in evaluating the optic disc for glaucoma, Ophthalmology 99 (2) (1992) 215–221.

[13] G. J. Jaffe, J. Caprioli, Optical coherence tomography to detect and manage retinal disease and glaucoma, American journal of ophthalmology 137 (1) (2004) 156–169.

[14] G. A. U. National, Glaucoma: diagnosis and management (2017).

[15] A. C. Thompson, A. A. Jammal, F. A. Medeiros, A review of deep learning for screening, diagnosis, and detection of glaucoma progression, Translational Vision Science & Technology 9 (2) (2020) 42–42.

[16] C. K.-S. Leung, M. Yu, R. N. Weinreb, G. Lai, G. Xu, D. S.-C. Lam, Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: patterns of retinal nerve fiber layer progression, Ophthalmology 119 (9) (2012) 1858–1866.

[17] F. A. Medeiros, L. M. Zangwill, C. Bowd, K. Mansouri, R. N. Weinreb, The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change, Investigative ophthalmology & visual science 53 (11) (2012) 6939–6946.

[18] T. Ojima, T. Tanabe, M. Hangai, S. Yu, S. Morishita, N. Yoshimura, Measurement of retinal nerve fiber layer thickness and macular volume for glaucoma detection using optical coherence tomography, Japanese journal of ophthalmology 51 (3) (2007) 197–203.

[19] A. E. Abd El-Naby, H. Y. Abouelkheir, H. T. Al-Sharkawy, T. H. Mokbel, et al., Correlation of retinal nerve fiber layer thickness and perimetric changes in primary open-angle glaucoma, Journal of the Egyptian Ophthalmological Society 111 (1) (2018).

[20] G. García, A. Colomer, V. Naranjo, Analysis of hand-crafted and automatic-learned features for glaucoma detection through raw circumpapillary oct images, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2020, pp. 156–164.

[21] D. C. Hood, A. S. Raza, On improving the use of oct imaging

for detecting glaucomatous damage, British Journal of Ophthalmology 98 (Suppl 2) (2014) ii1–ii9.

[22] D. Bizios, A. Heijl, J. L. Hougaard, B. Bengtsson, Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by stratus oct, Acta ophthalmologica 88 (1) (2010) 44–52.

[23] S. J. Kim, K. J. Cho, S. Oh, Development of machine learning models for diagnosis of glaucoma, PLoS One 12 (5) (2017) 1–16.

[24] G. Ometto, I. Moghul, G. Montesano, A. Hunter, N. Pontikos, P. R. Jones, P. A. Keane, X. Liu, A. K. Denniston, D. P. Crabb, Relayer: A free, online tool for extracting retinal thickness from cross-platform oct images, Translational vision science & technology 8 (3) (2019) 25–25.

[25] M. Kamal Abdellatif, Y. e. a. Abdelmaguid Mohamed Elzankalony, Outer retinal layers' thickness changes in relation to age and choroidal thickness in normal eyes, Journal of ophthalmology 2019 (2019).

[26] T. Hassan, M. U. Akram, M. F. Masood, U. Yasin, Deep structure tensor graph search framework for automated extraction and characterization of retinal layers and fluid pathology in retinal sd-oct scans, Computers in Biology and Medicine 105 (2019) 112–124.

[27] E. Gao, B. Chen, J. Yang, F. Shi, W. Zhu, D. Xiang, et al., Comparison of retinal thickness measurements between the topcon algorithm and a graph-based algorithm in normal and glaucoma eyes, PloS one 10 (6) (2015) 1–13.

[28] S. Niu, Q. Chen, L. de Sisternes, D. L. Rubin, W. Zhang, Q. Liu, Automated retinal layers segmentation in sd-oct images using dual-gradient and spatial correlation smoothness constraint, Computers in biology and medicine 54 (2014) 116–128.

[29] R. Kromer, S. Rahman, F. Filev, M. Klemm, An approach for automated segmentation of retinal layers in peripapillary spectralis sd-oct images using curve regularisation, Insights in Ophthalmology 1 (7) (2017) 1–6.

[30] L. Fang, D. Cunefare, C. Wang, R. H. e. a. Guymer, Automatic segmentation of nine retinal layer boundaries in oct images of non-exudative amd patients using deep learning and graph search, Biomedical optics express 8 (5) (2017) 2732–2744.

[31] W. Duan, Y. Zheng, Y. Ding, S. Hou, Y. Tang, Y. Xu, M. Qin, J. Wu, D. Shen, H. Bi, A generative model for oct retinal layer segmentation by groupwise curve alignment, IEEE Access 6 (2018) 25130–25141.

[32] S. K. Devalla, P. K. Renukanand, B. K. Sreedhar, G. Subramanian, L. Zhang, S. Perera, J.-M. Mari, K. S. Chin, T. A. Tun, N. G. Strouthidis, et al., Drunet: a dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images, Biomedical optics express 9 (7) (2018) 3244–3265.

[33] P. Zang, J. Wang, T. T. e. a. Hormel, Automated segmentation of peripapillary retinal boundaries in oct combining a convolutional neural network and a multi-weights graph search, Biomedical optics express 10 (8) (2019) 4340–4352.

[34] M. Pekala, N. Joshi, T. A. Liu, N. M. Bressler, D. C. DeBuc, P. Burlina, Deep learning based retinal oct segmentation, Computers in Biology and Medicine 114 (2019) 103445.

[35] E. B. Mariottoni, A. A. Jammal, C. N. Urata, S. I. Berchuck, A. C. Thompson, T. Estrela, F. A. Medeiros, Quantification of retinal nerve fibre layer thickness on optical coherence tomography with a deep learning segmentation-free approach, Scientific reports 10 (1) (2020) 1–9.

[36] A. C. Thompson, A. A. Jammal, S. I. Berchuck, E. B. Mariottoni, F. A. Medeiros, Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans, Jama Ophthalmology 138 (4) (2020)

[37] S. Maetschke, B. Antony, H. Ishikawa, G. Wollstein, J. Schuman, R. Garnavi, A feature agnostic approach for glaucoma detection in oct volumes, PloS one 14 (7) (2019).

[38] E. Noury, S. S. Mannil, R. T. Chang, A. R. Ran, C. Y. Cheung, S. S. Thapa, H. L. Rao, S. Dasari, M. Riyazuddin, S. Nagaraj, et al., Detecting glaucoma using 3d convolutional neural network of raw sd-oct optic nerve scans, arXiv preprint arXiv:1910.06302 (2019).

[39] X. Wang, H. Chen, L. Luo, A.-r. Ran, P. P. Chan, C. C. Tham, C. Y. Cheung, P.-A. Heng, Unifying structure analysis and surrogate-driven function regression for glaucoma oct image screening, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 39–47.

[40] X. Wang, H. Chen, A.-R. Ran, L. Luo, P. P. Chan, C. C. Tham, R. T. Chang, S. S. Mannil, C. Y. Cheung, P.-A. Heng, Towards multi-center glaucoma oct image screening with semi-supervised joint structure and function multi-task learning, Medical Image Analysis 63 (2020) 101695.

[41] G. García, A. Colomer, V. Naranjo, Glaucoma detection from raw sd-oct volumes: a novel approach focused on spatial dependencies, Computer Methods and Programs in Biomedicine (2020) 105855.

[42] G. An, K. Omodaka, K. Hashimoto, S. Tsuda, Y. Shiga, N. Takada, et al., Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images, Journal of healthcare engineering 2019 (2019).

[43] K. A. Thakoor, X. Li, E. Tsamis, P. Sajda, D. C. Hood, Enhancing the accuracy of glaucoma detection from oct probability maps using convolutional neural networks, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 2036–2040.

[44] T. Shehryar, M. U. Akram, S. Khalid, S. Nasreen, A. Tariq, A. Perwaiz, A. Shaukat, Improved automated detection of glaucoma by correlating fundus and sd-oct image analysis, International Journal of Imaging Systems and Technology (2020).

[45] G. García, R. d. Amor, A. Colomer, V. Naranjo, Glaucoma detection from raw circumpapillary oct images using fully convolutional neural networks, in: 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 2526–2530. doi:10.1109/ICIP40778.2020.9190916.

[46] R. Asaoka, H. Murata, K. Hirasawa, Y. Fujino, M. Matsuura, A. Miki, T. Kanamoto, et al., Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images, American journal of ophthalmology 198 (2019) 136–145.

[47] C. K. Leung, C. Ye, R. N. Weinreb, M. Yu, G. Lai, D. S. Lam, Impact of age-related change of retinal nerve fiber layer and macular thicknesses on evaluation of glaucoma progression, Ophthalmology 120 (12) (2013) 2485–2492.

[48] S. M. Iverson, W. J. Feuer, W. Shi, D. S. Greenfield, A. I. for Glaucoma Study Group, et al., Frequency of abnormal retinal nerve fibre layer and ganglion cell layer sdoct scans in healthy eyes and glaucoma suspects in a prospective longitudinal study, British Journal of Ophthalmology 98 (7) (2014) 920–925.

[49] F. Naghizadeh, A. Garas, P. Vargha, G. Holló, Detection of early glaucomatous progression with different parameters of the rtvue optical coherence tomograph, Journal of Glaucoma 23 (4) (2014) 195–198.

[50] N. Hammel, A. Belghith, R. N. Weinreb, F. A. Medeiros, N. Mendoza, L. M. Zangwill, Comparing the rates of retinal nerve fiber layer and ganglion cell–inner plexiform layer loss in healthy eyes and in glaucoma eyes, American journal of ophthalmology 178 (2017) 38–50.

[51] J. H. Na, K. R. Sung, S. Baek, Y. J. Kim, M. K. Durbin, H. J. Lee, H. K. Kim, Y. H. Sohn, Detection of glaucoma progression by assessment of segmented macular thickness data obtained using spectral domain optical coherence tomography, Investigative ophthalmology & visual science 53 (7) (2012) 3817–3826.

[52] J. H. Na, K. R. Sung, S. H. Baek, S. T. Kim, K. Shon, J. J. Jung, Rates and patterns of macular and circumpapillary retinal nerve fiber layer thinning in preperimetric and perimetric glaucomatous eyes, Journal of Glaucoma 24 (4) (2015) 278–285.

[53] J. H. Na, K. R. Sung, J. R. Lee, K. S. Lee, S. Baek, H. K. Kim, Y. H. Sohn, Detection of glaucomatous progression by spectral-domain optical coherence tomography, Ophthalmology 120 (7) (2013) 1388–1395.

[54] J. M. Wessel, F. K. Horn, R. P. Tornow, M. Schmid, C. Y. Mardin, F. E. Kruse, A. G. Juenemann, R. Laemmer, Longitudinal analysis of progression in glaucoma using spectral-domain optical coherence tomography, Investigative ophthalmology & visual science 54 (5) (2013) 3613–3620.

[55] B. Y. Xu, M. Chiang, S. Chaudhary, S. Kulkarni, A. A. Pardeshi, R. Varma, Deep learning classifiers for automated detection of gonioscopic angle closure based on anterior segment oct images, American journal of ophthalmology 208 (2019) 273–280.

[56] H. Fu, M. Baskaran, Y. Xu, S. Lin, D. W. K. Wong, J. Liu, T. A. Tun, M. Mahesh, S. A. Perera, T. Aung, A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images, American journal of ophthalmology 203 (2019) 37–45.

[57] F. Li, Z. Wang, G. Qu, D. Song, Y. Yuan, Y. Xu, K. Gao, G. Luo, Z. Xiao, D. S. Lam, et al., Automatic differentiation of glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network, BMC medical imaging 18 (1) (2018) 35.

[58] Ş. S. Kucur, G. Holló, R. Sznitman, A deep learning approach to automatic detection of early glaucoma from visual fields, PloS one 13 (11) (2018) e0206081.

[59] J. M. Ahn, S. Kim, K.-S. Ahn, S.-H. Cho, K. B. Lee, U. S. Kim, A deep learning model for the detection of both advanced and early glaucoma using fundus photography, PloS one 13 (11) (2018) e0207982.

[60] Y. Zhen, L. Wang, H. Liu, J. Zhang, J. Pu, Performance assessment of the deep learning technologies in grading glaucoma severity, arXiv preprint arXiv:1810.13376 (2018).

[61] A. Serener, S. Serte, Transfer learning for early and advanced glaucoma detection with convolutional neural networks, in: 2019 Medical Technologies Congress (TIPTEKNO), IEEE, 2019, pp. 1–4.

[62] H. Raja, T. Hassan, M. U. Akram, N. Werghi, Clinically verified hybrid deep learning system for retinal ganglion cells aware grading of glaucomatous progression, IEEE Transactions on Biomedical Engineering (2020).

[63] H. Raja, M. U. Akram, S. G. Khawaja, M. Arslan, A. Ramzan, N. Nazir, Data on oct and fundus images for the detection of glaucoma, Data in brief (2020) 105342.

[64] R. P. Mills, D. L. Budenz, P. P. Lee, R. J. Noecker, J. G. Walt, L. R. Siegartel, S. J. Evans, J. J. Doyle, Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease, American journal of ophthalmology 141 (1) (2006) 24–30.

[65] R. Susanna Jr, R. M. Vessani, Staging glaucoma patient: why and how?, The open ophthalmology journal 3 (2009) 59.

[66] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: Advances in neural information processing systems, 2017, pp. 4077–4087.

[67] Y. Pan, T. Yao, Y. Li, Y. Wang, C.-W. Ngo, T. Mei, Transferrable prototypical networks for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 2239–2247.

[68] T. Gao, X. Han, Z. Liu, M. Sun, Hybrid attention-based prototypical networks for noisy few-shot relation classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 6407–6414.

[69] S. Sun, Q. Sun, K. Zhou, T. Lv, Hierarchical attention prototypical networks for few-shot text classification, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 476–485.

[70] S. Fort, Gaussian prototypical networks for few-shot learning on omniglot, arXiv preprint arXiv:1708.02735 (2017).

[71] R. Boney, A. Ilin, Semi-supervised few-shot learning with prototypical networks, CoRR abs/1711.10856 (2017).

[72] J. Lu, Z. Cao, K. Wu, G. Zhang, C. Zhang, Boosting few-shot image recognition via domain alignment prototypical networks, in: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), IEEE, 2018, pp. 260–264.

[73] J. Wang, Y. Zhai, Prototypical siamese networks for few-shot learning, in: 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC), IEEE, 2020, pp. 178–181.

[74] A. J. Tatham, F. A. Medeiros, Detecting structural progression in glaucoma with optical coherence tomography, Ophthalmology 124 (12) (2017) S57–S65.

[75] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning?, IEEE transactions on medical imaging 35 (5) (2016) 1299–1312.

[76] M. A. Mayer, J. Hornegger, C. Y. Mardin, R. P. Tornow, Retinal nerve fiber layer segmentation on fd-oct scans of normal subjects and glaucoma patients, Biomedical optics express 1 (5) (2010) 1358–1383.

[77] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, arXiv preprint arXiv:2004.11362 (2020).

[78] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, arXiv preprint arXiv:2002.05709 (2020).

[79] P. H. Le-Khac, G. Healy, A. F. Smeaton, Contrastive representation learning: A framework and review, IEEE Access (2020).

[80] S. Petscharnig, M. Lux, S. Chatzichristofis, Dimensionality reduction for image features using deep learning and autoencoders, in: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, 2017, pp. 1–6.

[81] J. Song, Binary generative adversarial networks for image retrieval, arXiv preprint arXiv:1708.04150 (2017).

[82] M. I. Daoud, A. Saleh, I. Hababeh, R. Alazrai, Content-based image retrieval for breast ultrasound images using convolutional autoencoders: A feasibility study, in: 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), IEEE, 2019, pp. 1–4.

[83] C. Simon, P. Koniusz, M. Harandi, Projective subspace networks for few-shot learning (2018).

[84] C. Simon, P. Koniusz, R. Nock, M. Harandi, Adaptive subspaces for few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4136–4145.

[85] P. Gholami, P. Roy, M. K. Parthasarathy, V. Lakshminarayanan, Octid: Optical coherence tomography image database, Computers & Electrical Engineering 81 (2020) 106532.

[86] B. Zhou, A. Khosla, A. Lapedriza, et al., Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921–2929.