IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 21, 2020, accepted October 1, 2020, date of publication October 6, 2020, date of current version October 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028960

Data Augmentation for Improving Proliferative Diabetic Retinopathy Detection in Eye Fundus Images

TERESA ARAÚJO[1,2], GUILHERME ARESTA[1,2], LUÍS MENDONÇA[3], SUSANA PENAS[4,5], CAROLINA MAIA[4], ÂNGELA CARNEIRO[4,5], ANA MARIA MENDONÇA[1,2], (Senior Member, IEEE), AND AURÉLIO CAMPILHO[1,2], (Senior Member, IEEE)

[1]INESC TEC–Institute for Systems and Computer Engineering, Technology, and Science, 4200-465 Porto, Portugal
[2]Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal
[3]Department of Ophthalmology, Hospital de Braga, 4710-243 Braga, Portugal
[4]Department of Ophthalmology, Centro Hospitalar São João, 4200-319 Porto, Portugal
[5]Department of Surgery and Physiology, Faculty of Medicine, University of Porto, 4200-465 Porto, Portugal

Corresponding author: Teresa Araújo (tfaraujo@inesctec.pt)

This work was supported in part by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization–COMPETE 2020 Programme, and in part by the National Funds through the Fundação para a Ciência e a Tecnologia (FCT) under Project CMUP-ERI/TIC/0028/2014. The work of Teresa Araújo was supported by the FCT Grant under Contract SFRH/BD/122365/2016. The work of Guilherme Aresta was supported by the FCT Grant under Contract SFRH/BD/120435/2016.

ABSTRACT Proliferative diabetic retinopathy (PDR) is an advanced diabetic retinopathy stage, characterized by neovascularization, which leads to ocular complications and severe vision loss. However, the available DR-labeled retinal image datasets have a small representation of images of the severest DR grades, and thus there is lack of PDR cases for training DR grading models. Additionally, the criteria for labelling these images in the publicly available datasets is not always clear, with some images which do not show typical PDR lesions being labeled as PDR due to the presence of photo-coagulation treatment and laser marks. This problem, together with the datasets' high class imbalance, leads to a limited variability of the samples, which the typical data augmentation and class balancing cannot fully mitigate. We propose a heuristic-based data augmentation scheme based on the synthesis of neovessel (NV)-like structures that compensates for the lack of PDR cases in DR-labeled datasets. The proposed neovessel generation algorithm relies on the general knowledge of common location and shape of these structures. NVs are generated and introduced in pre-existent retinal images which can then be used for enlarging deep neural networks' training sets. The data augmentation scheme was tested on multiple datasets, and allows to improve the model's capacity to detect NVs.

INDEX TERMS Data augmentation, deep learning, neovascularization, proliferative diabetic retinopathy.

I. INTRODUCTION

Diabetic retinopathy (DR) is a complication of diabetes and it constitutes a leading cause of blindness worldwide [1], [2]. In 2030, 552 million people are expected to suffer from diabetes [3]. The majority of visual loss cases can be prevented with early detection and adequate treatment [4], [5]. The earliest diabetes-related changes in the retina are often imperceptible and have minimum impact in the vision [6], and thus approximately one third of the diabetic patients have DR but show no symptoms, leading to the progression of the disease untreated. Contrastingly, proliferative diabetic retinopathy (PDR) is an advanced stage of DR, leading to ocular complications and severe vision loss in many developed countries [7]. PDR is characterized by the appearance and growth of newly formed vessels, neovessels (NV), on the retina or optic disc (OD) that extend along the retinal surface or into the vitreous cavity, increasing significantly the risk of vision loss [7], [8]. There is also a strong correlation between PDR and uncontrolled systemic disease [7], [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Berdakh Abibullaev.

Regular check-ups via DR screening programs are essential for detecting the disease as early as possible and determining the adequate treatment. Specialists search for abnormalities in the retina and classify the severity of the disease according to the findings. However, the diagnosis process is prone to errors due to the large number of patients to be observed, poor image acquisition and variety of lesions to analyse. Computer-aided diagnosis (CAD) systems can improve the DR screening pipeline both reducing the burden [10] and providing a second opinion to the ophthalmologists, reducing diagnosis' subjectivity [11], [12].

Deep learning has recently allowed for CAD systems to achieve near-human performance in DR detection. DR grading, *i.e.* staging of the pathology according to its severity, is a more complex task since it requires the identification and integration of different lesions. Generaly, DR is classified accordingly to the international DR scale as mild non-proliferative DR (NPDR) (R1), moderate NPDR (R2), severe NPDR (R3) or PDR (R4) [1], ordered according to their risk of progression [13]. Despite the task's complexity, several works have shown promising performance [14]–[16]. Namely, Araújo *et al.* [16] proposed a model (DR|Graduate) that not only produces a grade but also an uncertainty associated with the prediction and an attention map that explains the decision. However, the proposed model tends to fail to detect R4 graded images. One of the problems with the DR|Graduate learning phase is the association of photo-coagulation treatment and laser marks to R4. This follows the training set labels, since training images which do not show R4 signs, such as NVs and pre-retinal hemorrhages, are still labeled as R4 due to the presence of photo-coagulation marks. The R4 detection problem is further exacerbated by the high class imbalance in the datasets, since there is a limited variability of the samples, which the data augmentation and class balancing scheme were not able to mitigate.

A possible solution to attenuate this R4-labeling and the class balancing problems is to augment the training data by generating new synthetic images from a given grade, *i.e.*, with specific lesions. Several studies have focused on the generation of artificial eye fundus images [17]–[19], commonly taking advantage of generative adversarial networks (GANs) to generate either an image or a portion of an image containing the structures of interest. The majority of the works use segmented anatomical structures, namely the vasculature, to guide the generation. For instance, [18] has proposed an end-to-end adversarial retinal image synthesis that allows to generate artificial retinal images that are anatomically consistent, which is trained on pairs of retinal images and vessel tree segmentations. However, the method only generates healthy images, of size 256×256 pixels, which is low given the common resolution of acquired images. Although the generated vessels have an overall consistent appearance, they often present interruptions and width inconsistencies along the vessels. Indeed, the quality of the generated images may not be enough when considering the need to distinguish small structures such as DR lesions.

Other researchers have addressed the non-healthy retinal image generation. Pujitha and Sivaswamy [20] proposes a GAN-based method to generate images with hemorrhages. The method requires vessel tree and lesion binary annotations for training the model. A new retinal image can be synthesized by providing the model the vessel tree and lesion (hemorrhage) masks. In the work of [19] the authors incorporated style transfer into the framework. They examined pathological cases, using as style reference images with DR and other pathologies. However, the results were not considered acceptable, with retinal images not being properly replicated.

Despite producing overall plausible retinal images, these systems present several drawbacks. For instance, they require training data with pixel-wise annotations, which are costly to obtain. Also, the training set variability determines the generation capability of the model [18]. Further, the high computational requirements often hinder the quality of the generated images, particularly providing low resolution images which do not much the higher resolutions in the retinographs. Additionally, as referred, these works often rely on the retinal vasculature for image generation and for evaluating its plausibility. However, the synthesized images do not present clinically accepted vessel networks [21]. Also, to our knowledge the generation of pathological images from a given DR grade, with a given lesion type, is not yet achieved [21]. The work of [20] suggested that it was possible to generate images with different number of hemorrhages, but it did not explore further its capability of generating images with other lesions types.

We propose an heuristic-based data augmentation scheme based on NV-like structures generation that compensate for the lack of PDR cases in DR-labeled datasets. The proposed neovessel synthesis algorithm relies on the general knowledge of common location and shape of these structures. NVs are generated and introduced in pre-existent retinal images which can then be used for enlarging deep neural networks' training sets. Training with this type of data augmentation allows to increase detection of real NVs in independent test sets.

II. METHODS

The NV-based data augmentation starts with the generation of prototype NVs using a semi-random graph-based approach. The resulting graphs are processed having into account the type of vessel and the location insertion, as summarized in Fig. 1. The processing includes moving the graph to the pixel-space, adjusting vessel width according to the branching of the structure, rotation and color matching to nearby vessels. Each step of the algorithm is detailed in the remaining of the section.

A. INSERTION LOCATION

NVs occur anywhere in the retina, but they are most common posteriorly, within 45 degrees of the OD [7]. They are particularly common on the OD itself. In the Diabetic

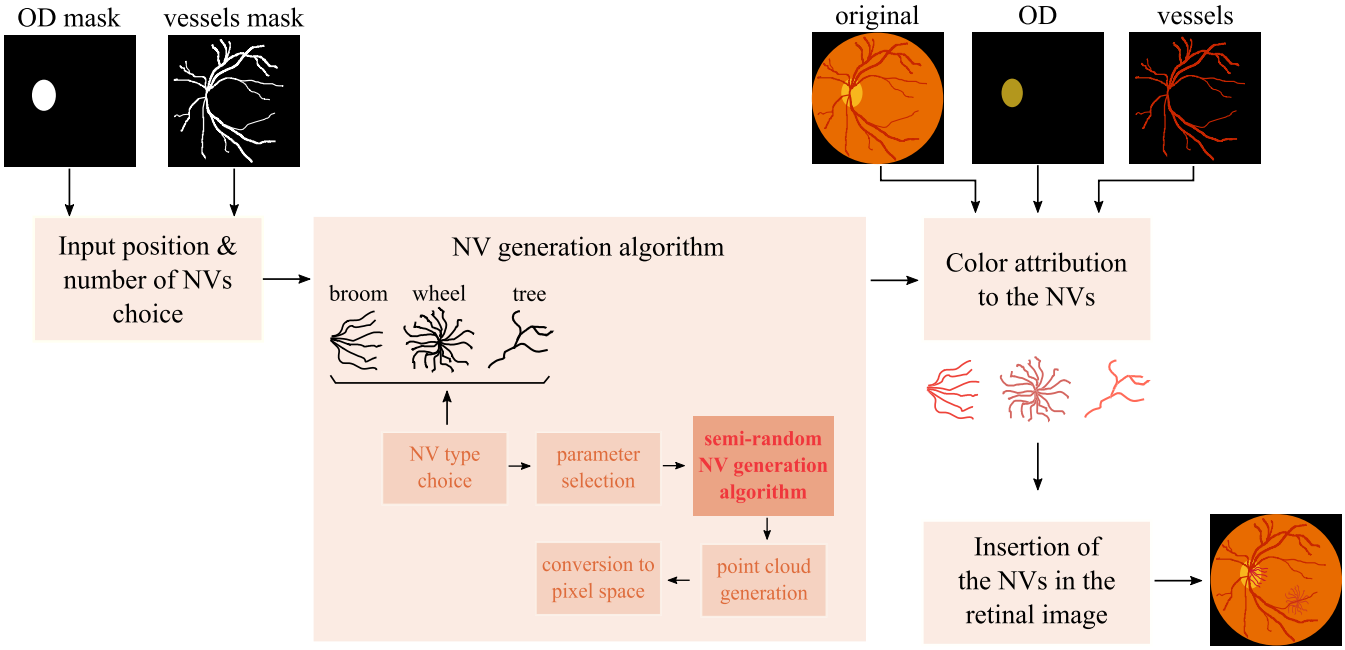


FIGURE 1. Pipeline of the proposed neovessel generation algorithm and insertion in the retinal image.

Retinopathy study (DRS) [7], out of 1377 PDR eyes 15% had NV only on the OD or less than 1 OD diameter from the OD (NV on disc), 40% had NV only outside this zone (NV elsewhere), 45% had NV in both zones. NVs often lie over veins and appear to drain into them [7]. Sometimes NVs grow for several OD diameters across the retina without forming proeminent networks, commonly arising from the OD. In these cases, they are only distinguishable from normal vessels due to their capability of crossing both arterioles and veins [7].

The region of interest (ROI) and location of where the generated NV is inserted is based on the statistical behavior of the pathology. Following the DRS statistics, the probability of inserting generated NVs only in the OD or surrounding region was set to 0.15, only elsewhere 0.40, and on both sites 0.45. The area of the ROI was set to 25% of the image size to properly capture the characteristics of the local vasculature.

1) VESSEL SEGMENTATION

In order to ensure a more natural introduction, NVs were only inserted on pixels located on vessels. This goes in line with the fact that NV often lie over retinal veins. In order to determine which image pixels belong to vessels, a rough vessel segmentation was performed based on Otsu thresholding on the image's green channel, preceded by the application of a bottom hat operator. We call this the vessel segmented image M_{vessels}. To perform A/V segmentation would be a much more complex task and it is out the scope of this study. A more sophisticated deep learning vessel segmentation method based on a U-Net was also explored in initial tests, but the improvement in the vessel segmentation did not lead to an improvement in the performance of the DR grading algorithm, and thus the simpler approach was chosen.

2) OPTIC DISC SEGMENTATION

For segmenting the OD the UOLO framework was used, proposed by Araújo *et al.* [22] for the simultaneous detection and segmentation of structures of interest in medical images. It consists of an object segmentation module which intermediate abstract representations are processed and used as input for object detection. UOLO was validated for OD segmentation on different public datasets. The binary image with the segmented OD is herein referred to as M_{OD}.

B. QUANTITY

The number of introduced NVs is determined by randomly selecting a probability threshold value. Initially the probability of inserting a NV is set to one, and is updated after each insertion (decreased by a factor of $\times 0.65$). The generation-insertion process continues until this probability reaches the pre-determined probability threshold and no further NVs are introduced. This update rule was selected since it allows a non deterministic choice of the number of neovessels to be inserted. The probability decay parameter value was chosen experimentally in order to allow the insertion of a plausible number of NVs, accordingly to the usual appearance of true PDR images with NVs. The most common number of inserted neovessels per image is 3 (average of 3 ± 2 NVs/image). This value seems adequate considering that 1) there is some redundancy, and having images with more than one NV may be better for training the network since it promotes activations on different regions and vessel types on the same image, which we believe that leads the network to learn more generic

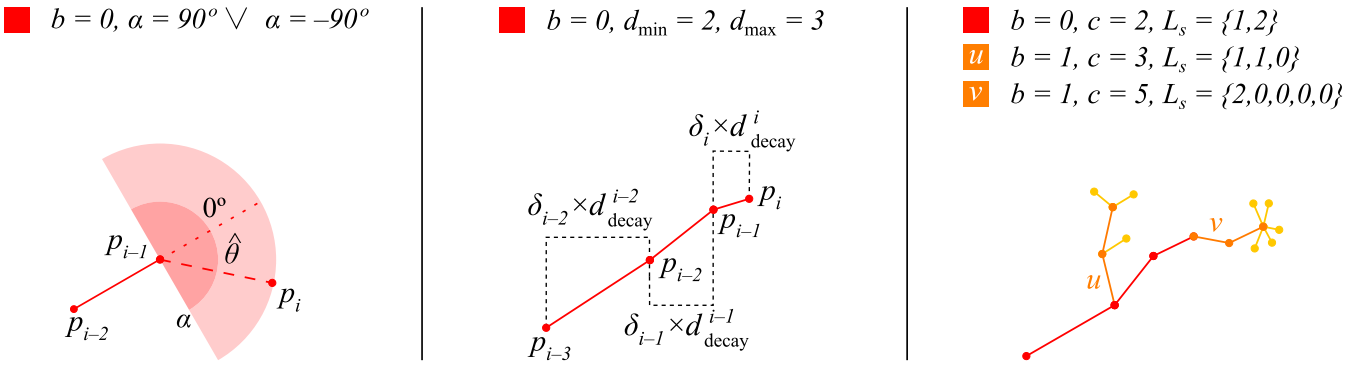


FIGURE 2. Semi-random neovessel generation algorithm's parameters example. ***b*** is the current branch offspring level, ***p*** are the points composing the branch, α is the angular range limit to where the current branch can grow, $d_{\min}$ and $d_{\max}$ are the minimum and maximum distance between two sequential points, δ is the randomly selected distance, ***c*** is the number of offspring branches and L_s are the number of points between branching offspring.

representations of the NVs, and 2) having a very high number of NVs would make the image non realistic.

C. SEMI-RANDOM NEOVESSEL GENERATION

The NV generation algorithm assumes that NVs are composed of branches, similarly to trees. On a tree, there is a main branch (trunk) from which several other branches can spurt. The amount, location, orientation and growth of child branches is dependent on the type of tree (or NV). By their turn, in a recursive process, the child branches can generate offspring. Generically, the algorithm operates by iteratively generating points on the Cartesian space on a parent branch. At each iteration, there is the possibility that the point belongs to a new child branch, affecting the position and distance of all sibling points thereafter. Once a growth limit has been achieved, the child branch ends and the growth of the parent branch continues.

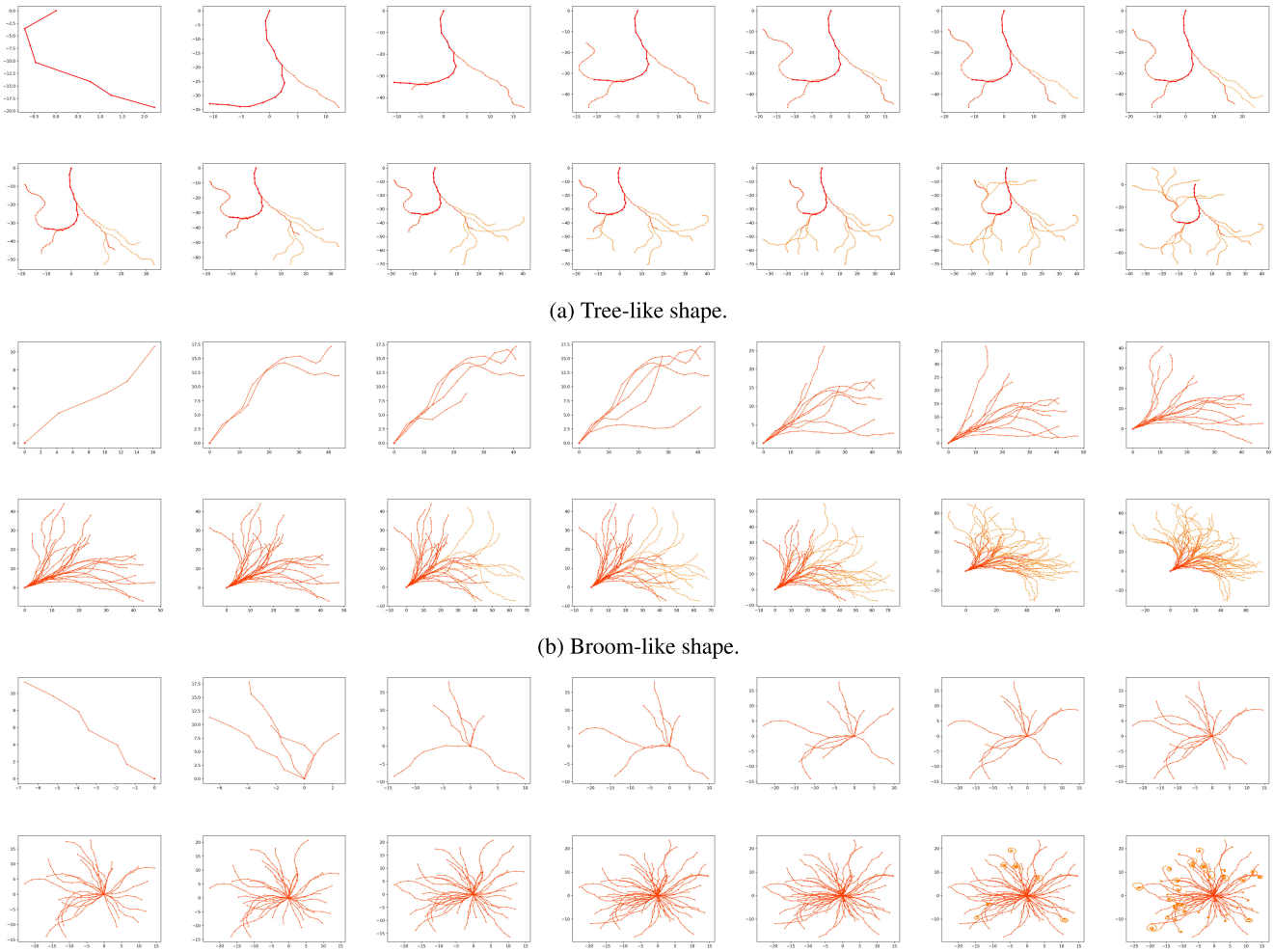
Formally, let $b \leq b_{\max}$ define the current branch offspring level, composed by an iteratively growing set of points $P_b = \{p_0, \ldots, p_i, \ldots, p_c\}$ in the Cartesian x,y space. Each offspring level $b\,(c\,, L_s\,, \alpha, p_\alpha, D)$ has a set of properties that affect the behaviour of the branch: 1) c, with $C_b = \{c_{\min}, \ldots, c_{\max}\} \in \mathbb{N} \xleftarrow{R} c$, is the number of child branches that will offspring from the current branch of level b, randomly selected from a set of integers C_b between $c_{\min}$ and $c_{\max}$; 2) L_s, which is the number of points added between two offspring branches, with $|L_s| = c$; 3) α, the angular range limit where p_i can occur in relation to p_{i-1}; 4) p_α, which controls the probability of changing the sign of α; 5) $D = \{d_{\min}\,, d_{\max}\,, d_{\text{decay}}\}$ is a set of distance related parameters, where $d_{\min}$ and $d_{\max}$ are the minimum and maximum distances between two sequential points, respectively, and d_{decay} is a iteration-wise distance growth factor; Each point $p\,(\theta, \delta, s)_{b\,,i}$ has a behaviour inherited by the branch level b and the previous point $p_{b\,,i}$: 1) $\theta_{b,i} = \theta_{b,i-1} + \hat{\theta}$, where $\hat{\theta} \in [\alpha_b\,, 0] \vee \hat{\theta} \in [0, \alpha_b]$, corresponding to the absolute angle in relation to the origin of the referential where p_i will be located; 2) $[d_{\min}\,, d_{\max}] \xleftarrow{R} \delta \times d_{\text{decay}}^{i-1}$ corresponding to the distance between $p_{b\,,i}$ and $p_{b\,,i-1}$ and; 3) s, true when $i \in L_s$, is a boolean indicating if the point will originate a new child branch. Branch growth can be done using different update rules. In this work, points p can result from a linear coordinate update ($x_i = x_{i-1} + cos(\theta) \times \delta$, $y_i = y_{i-1} + sin(\theta) \times \delta$) or from a logarithm spiral-like curve update ($x = ae^{k\theta} \cos\theta \times \delta$, $y = ae^{k\theta} \sin\theta \times \delta$, where a and k are pre-defined constants). Due to the recursive nature of the method, the maximum number of branching levels b, b_{max} has to be limited. Otherwise, the vessels would grow indefinitely. An exemplification of some of the main parameters is shown in Fig. 2. Illustrative step-wise examples of NVs generation are shown in Fig. 3. Changing the hyperparameters that control the range of possible orientations, distances, number of points and number of branches allows to obtain different NVs types.

The generated NV structures are iteratively converted to the pixel space starting from the lowest branch level b. First, a randomly selected scale factor (between 2 and 6) is applied to the generated NVs in order to increase the range of possible sizes of these structures. Then, for each b, the coordinates of the points are rounded to the nearest integer. The resulting structure is dilated using as structuring element a disk of radius $\propto 1/b$ and smoothed with a Gaussian filter. The insertion of the pixel represented NV, NV_{im}, on the eye fundus image is detailed in Section II-D.

1) SHAPE

New vessels often resemble carriage wheels, radiating from the center of the structure to a circular boundary. They may also be irregularly shaped and not present this radiating pattern. Initially, NVs may be barely visible, but later their caliber is commonly 1/8-1/4 that of a major retinal vein at the disc margin, and occasionally be as large as these veins [7].

Despite NVs may present a high variety of shapes, we considered three different main types of shapes: tree-like, wheel-like and broom-like. The first shape resembles a simple tree, with relatively light branching. The broom shape has a more dense appearance and grows mainly in a given angular range. The wheel shape differs from the broom since it can grow in the full 2π *rad* angular range and its branches may end in curly structures. Different sets of parameter values were



(a) Tree-like shape.

(b) Broom-like shape.

(c) Wheel-like shape.

FIGURE 3. Examples of step-wise generation of artificial neovessels for the different considered shapes.

TABLE 1. Ranges and values of the semi-random generation algorithm parameters for each of the considered neovessel shapes. All shapes have three levels ($b \in \{0, 1, 2\}$). The coordinate update rule is always linear, except for the wheel-like shape, for which the spiral update rule is used in the last level ($b = 2$)). Also, $p_\alpha = 0.5$, except for the wheel at $b = 2$ where $p_\alpha = 0$ to promote curling. $d_{decay} = 0.95$, except for the wheel where $d_{decay} = 1$.

NV type	parameter - level	c *#child branches*	Ls *#points between branches*	α *angular range limit*	d *distance between points*
tree	$b = 0$	$[3, 4]$ relatively low # branches; each level has >childs than the previous	$\in \{4, 5, 6\}$	$\pi/6$ grow in a more or less specific direction (not too spread)	$[3, 5]$ bigger branches in the beginning than in the end (for the same Ls)
	$b = 1$	$[4, 5]$	$\in \{4, 5, 6\}$	$\pi/6$	$[1.2, 3]$
	$b = 2$	1	$\in \{4, 5, 6\}$	$\pi/6$	$[1.2, 3]$
broom	$b = 0$	$[15, 25]$ number of branches in the first level for high density	$\in \{0\}$ in the first level all branches come from the same point	$\pi/7$ grow in a more or less specific direction (not too spread)	-
	$b = 1$	$[2, 3]$	$\in \{4, 5, 6\}$	$\pi/7$	$[2, 7]$
	$b = 2$	1	$\in \{4, 5, 6\}$	$\pi/7$	$[2, 5]$
wheel	$b = 0$	$[50, 60]$ high number of branches in the first level for high density appearance	$\in \{0\}$	2π branches may spread in all directions in the beginning (dense)	-
	$b = 1$	$[1, 2]$	$\in \{4, 5, 6\}$	$\pi/6$	$[2, 2.5]$
	$b = 2$	1	$\in \{75, 100\}$ large # of points in the last level to allow a complete spiral	$\pi/5$	$[1.5, 2]$

selected to associate with each distinct NV shape. These three main shapes are shown in Fig. 3. Table 1 presents a summary of the values/ranges of the most relevant parameters for each of the considered NV shapes, along with a brief explanation of some of the choices. For instance, in the case of the wheel-like shape, the parameter c of the first level ($b = 0$), which

(a) Close up of the region of interest.

(b) Vessel segmentation M_{vessels}

(c) Color matrix C.

(d) Color matrix C after Gaussian smoothing.

(e) Inserted NV without the distance transform ponderation.

(f) Inserted NV with the distance transform ponderation.

FIGURE 4. Color attribution to a generated neovessel (NV).

controls the number of offspring branches, must be very large in order to produce a dense structure, and thus c_{min} and c_{max} must high. Also, the α parameter, which controls the angular range limit for branch growth, must be very large (around 2π) in order to allow the growth in all directions. Contrarily to the wheel-shape, in the tree and broom-like shapes the angular range for the first level must be reduced in order to force a more narrow angular growth, in a specific direction. The broom shape differs from the tree one mainly due to the setting of the c parameter, which must be higher for the broom in the beginning of the growth to allow a more dense structure. Also, in the first level of the broom shape (as well as in the wheel shape) all the branches should grow from the same point, i.e., L_s is has just one index, whereas in the tree they may arise from different locations.

2) ORIENTATION

For the broom-like shape, the orientation of the NV in the image has special importance since, for instance, it may seem unnatural to have a NV orientated perpendicular to the existing vasculature. A rough indicator of the orientation of the main vessels in a given image region can be determined based on the OD position, with the definition of quadrants centered on the OD. A broom NV is thus inserted with an orientation close to that of the vessels in the insertion quadrant ($\pi/4$, $3\pi/4$, $-3\pi/4$, or $-\pi/4$), with a tolerance of 0.17 rad (randomly determined for each insertion).

D. COLOR ASSIGNMENT AND NEOVESSEL INSERTION

The color of the generated NVs must be coherent with the ROI's vasculature to allow a realistic insertion. This color is determined based on the assumption that 1) vessels' colors are not constant and change according to multiple factors, including vessel type (artery or vein), proximity to the OD or fovea, thickness, external illumination and position on the eye; 2) the artificial NVs should have a color similar to the local vasculature, namely the nearest vessels. Because of this, the color of the NV is based on the color of the vessels from M_{vessels} near the insertion region, instead of using only the color of the vessel from which the NV will branch. Further, since the vessel segmentation is not perfect, larger vessels, expected to be more representative of the overall vasculature, are given more weight on the final NV color.

Vessel color is attributed via a color matrix C that is computed based on the colors of the local vasculature exemplified in Fig. 4a. Specifically, each pixel from C, $C(i,j)$, results from the weighted average of the median RGB values of each of the objects from M_{vessels} (Fig. 4b). For instance, for the red channel, $C(i,j)_{\text{red}} = \sum_{v=0}^{N} r_{v,AP} \times DT_v(i,j) \times c_{v,\text{red}}$, where v is the index of each of the N vessels segmented in

M_{vessels} (i.e., each connected component), $c_{v,\text{red}}$ is the median red color of vessel v, $r_{v,AP}$ is the area/perimeter ratio and $DT_v(i,j) = DT'_v(i,j) / \sum_{v=0}^{N} DT'_v(i,j)$, where $DT'_v \in [0\,,\,1]$ is the complement of the distance transform of $\overline{M_{\text{vessels},v}}$. The area/perimeter ratio constitutes an approximation of vessel thickness, given that the vessel shape can be approximated by a rectangle with length c and thickness, l, with $A_v = c \times l$ and $P_v = 2(c + l)$, $r_{AP,v}$ should be higher for thicker vessels (if $c \gg l \rightarrow r_{AP,v} \propto l$). Due to the high color difference that exists between the OD and non-OD regions, the vessels from each of these regions are handled separately. The color matrices are then combined (Fig. 4c) and C is smoothed with a Gaussian filter to reduce the influence of sharp color transitions (Fig. 4d).

Each pixel of NV_{im} is attributed the color of the corresponding position in the color matrix C. To ensure that NV_{im} is properly blended with the image and that branches fade with increasing distance to the insertion point and neighbor vessels, NV_{im} is weighted by the complement of the normalized distance transform of $\overline{M_{\text{vessels}}}$ (see Fig. 4e and 4h). The insertion of the NV is then achieved via $\text{ROI}_{\text{with NV}} = (1 - NV_{\text{im}}) \times \text{ROI} + NV_{\text{im}} \times C$.

E. DR GRADING

1) METHOD - DR|Graduate

DR grading was obtained using the methodology proposed by Araújo *et al.* [16]. DR|Graduate is a deep learning-based DR grading computer-aided diagnosis system that supports its decision by providing an interpretable explanation and an estimation of the uncertainty of that prediction, which gives the ophthalmologist a measure of how much that decision can be trusted. The method was designed having into account the ordinal nature of the DR grading problem. It consists mainly of a Gaussian-sampling approach built upon a Multiple Instance Learning (MIL) framework which allows to infer an image grade together with a prediction uncertainty and an explanation map, being trained only with image-wise labels. The produced explanation maps highlight the regions of the image related with each DR grade. The network is composed of several convolutional-batch normalization blocks interleaved with max-pooling layers. For each input image I, DR|Graduate predicts a diabetic retinopathy (DR) grade $\hat{y}_g$ and a grade uncertainty u. Let M be the output of the last layer of DR|Graduate's backbone after assessing the input image I. The lesion map L is computed by applying a $1 \times 1 \times 1$ convolution with a linear activation over M. L is a map where each element L_{ij} indicates the presence of a lesion of grade $\lfloor L_{ij} \rceil$[†] on a patch of size dependent on the model's receptive field in the input. Following the MIL approach, the grade of an image is initially estimated as $\hat{y}_r = max(L)$. The model's output probabilities ($p_c, c \in \{0 \ldots 4\}$) are computed from $\hat{y}_r$ based on a Gaussian distribution centered on $\hat{y}_r$ with variance equal to the uncertainty u, with u learned during training.

DR|Graduate assumes that grading is a discrete ordinal classification problem, making the goal the minimization of the distance between the target and the predicted classes. It also assumes that there is a different lesion type per grade, and thus the class can be predicted following the standard MIL assumption, meaning that the most severe lesion type will be responsible for the image grade.

The model was trained by optimizing a loss based on the cross-entropy, and the class imbalance problem was addressed by augmentation of the least represented classes via online horizontal and vertical flips, rotations, brightness adjustments and contrast normalization. Training started with all the 5 classes equiprobable and evolved to a more skewed distribution, which is closer to the real world distribution. Images were cropped around the FOV and resized to the input size of the network, 640× 640 pixels. The model was trained for 240 epochs with the Kaggle DR detection training set, using a batch size of 30, the Adam optimizer, and a learning rate of 3×10^{-4}. DR|Graduate was validated in several different datasets. The model showed a robust performance for DR grading. However, as referred in Section I, it usually fails to detect R4/PDR labeled images.

2) NEWLY PROPOSED DATA AUGMENTATION

DR|Graduate was herein trained with the original Kaggle DR detection images for grades R0-R3, and the R4 class was augmented using the images with artificial NVs. Fig. 5 shows examples of Kaggle images originally from grades R0-R3 with the generated NVs. Examples of newly generated images were presented to an ophthalmologist, who accessed their appearance and confirmed that the generated NVs resemble true neovascularizations. The original R4 images were kept, and the same number of newly generated images was added. NVs were inserted in images equitably pooled from the set of R0-R3 images. No insertion was performed on the pre-existant R4 images since this could impair the detection of other PDR signs. The reason behind this R4 training set construction was to have a significant number of images with generated NVs, but still maintain the information regarding the original data in order to not degrade the performance in the Kaggle DR detection and other datasets which may present the same labeling scheme.

The training scheme was kept the same as in [16], in order to assess only the influence of the NV augmentation in the results. Note that the original data augmentation scheme, used in [16] (flips, rotations, brightness adjustments and contrast normalization), were still performed in the new training, together with the newly proposed NV augmentation.

III. RESULTS

A. DATASETS

Kaggle DR detection: This dataset* has a large number of high resolution color eye fundus images (approximately 35 000 in the training set, 55 000 in the test set). Images

[†] $\lfloor\,\rceil$ rounding to the nearest integer, $\lfloor\,\rfloor$ rounding down, $\lceil\,\rceil$ rounding up.

*https://www.kaggle.com/c/diabetic-retinopathy-detection

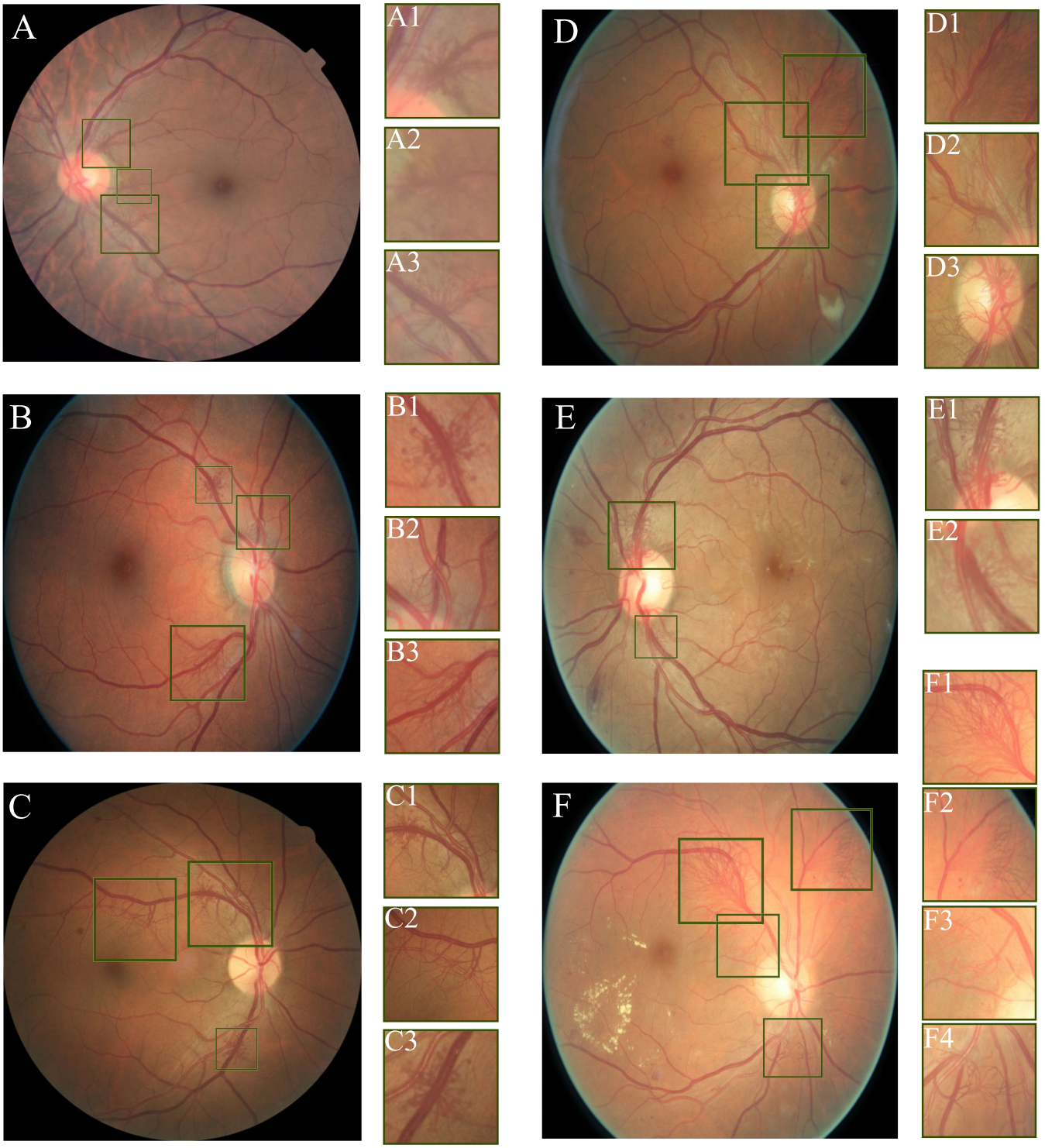
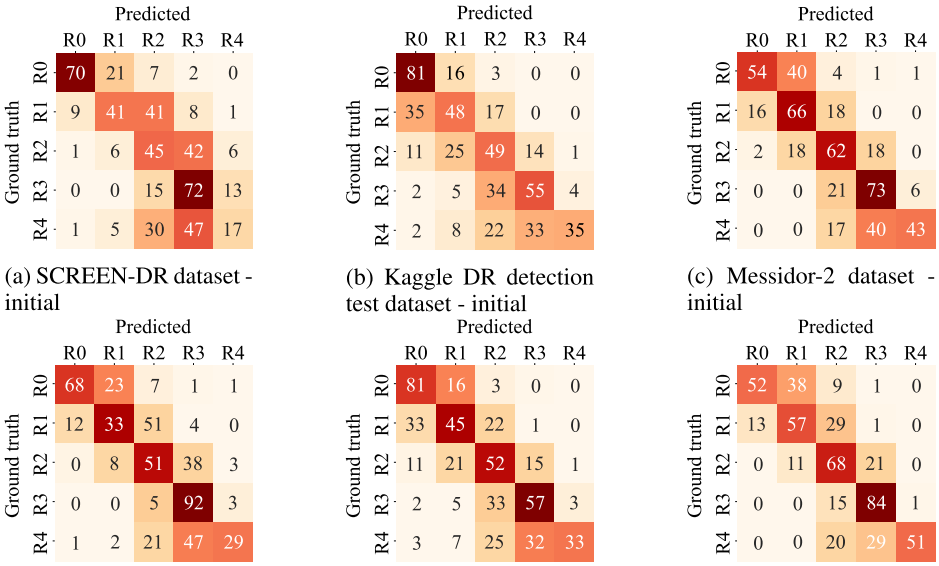


FIGURE 5. Examples of images with synthesized artificial neovessels (NVs). Close-ups of some of these NVs are shown next to the images, along with the NV type. Original DR labels (from the Kaggle DR detection dataset): A: R0; B: R0; C: R1; D: R2; E: R3; F: R3.

were provided by EyePACS,[†] and were acquired under a large variety of conditions, using different types of cameras. Images are labeled by a single clinician with the respective DR grade (5 severity levels). The distribution of the images per grade in this dataset is {73.5, 7.0, 15.0, 2.5, 2.0}% for {R0, R1, R2, R3, R4} grades, respectively. It is a real-world dataset, presenting noise in image and labels, and the images may present bad quality (contain artifacts,

[†]http://www.eyepacs.com



(a) SCREEN-DR dataset - initial

(b) Kaggle DR detection test dataset - initial

(c) Messidor-2 dataset - initial

(d) SCREEN-DR dataset - new

(e) Kaggle DR detection test dataset - new

(f) Messidor-2 dataset - new

FIGURE 6. Confusion matrices of the new and initial (DR|Graduate's) predictions *vs* the ground truth DR grades in the SCREEN-DR, Kaggle DR detection test and Messidor-2 datasets. Matrixes were normalized by dividing each original value by the sum of the values from it's row (i.e., number of images from that grade in the ground truth), and the shown values are percentages.

be out of focus and/or with inadequate exposition). This dataset aims at promoting the development of robust and reliable algorithms that perform well on a large variety of images.

Messidor-2: Messidor-2 [23], [24][‡] has 1748 images from 874 subjects, with one fovea-centered image per eye. Images were acquired in 2010, at: Hôpital Lariboisière (Paris, France), Brest University Hospital (Brest, France), and Saint-Etienne University Hospital (St Etienne, France). The ground truth used in this work is an adjudicated consensus of three specialists [14].[§] A total of 1744 images were used, as the remaining were adjudicated as ungradable. The class distribution is $\{58.3, 15.5, 19.9, 4.3, 2.0\}\%$ for $\{R0, R1, R2, R3, R4\}$ grades, respectively.

SCREEN-DR: This private dataset consists of a subset of retinal images from a Portuguese DR screening program, managed by the Portuguese North Health Administration (ARSN). The dataset has 966 images, labeled by a retinal specialist in 5 severity levels. For 348 of these images, pixel-wise annotations of the following lesions are available: microaneurisms (MAs), hemorrhages (HEMs), cotton-wool spots (CWSs), intra-retinal microvascular abnormalities (IRMAs), hard exudates (EXs), neovessels (NVs), pre-retinal hemorrhages (PHEMs) and pre-retinal fibrosis (PFIB). The class distribution in the SCREEN-DR dataset is $\{43.0, 17.6, 23.4, 6.2, 9.8\}\%$ for $\{R0, R1, R2, R3, R4\}$ grades, respectively.

[‡]http://www.adcis.net/en/third-party/messidor2/
[§]https://www.kaggle.com/google-brain/messidor2-dr-grades

TABLE 2. Quadratic weighted kappa (κ) for the new and initial (DR|Graduate's) DR grading in the SCREEN-DR, Kaggle DR detection test and Messidor-2 datasets.

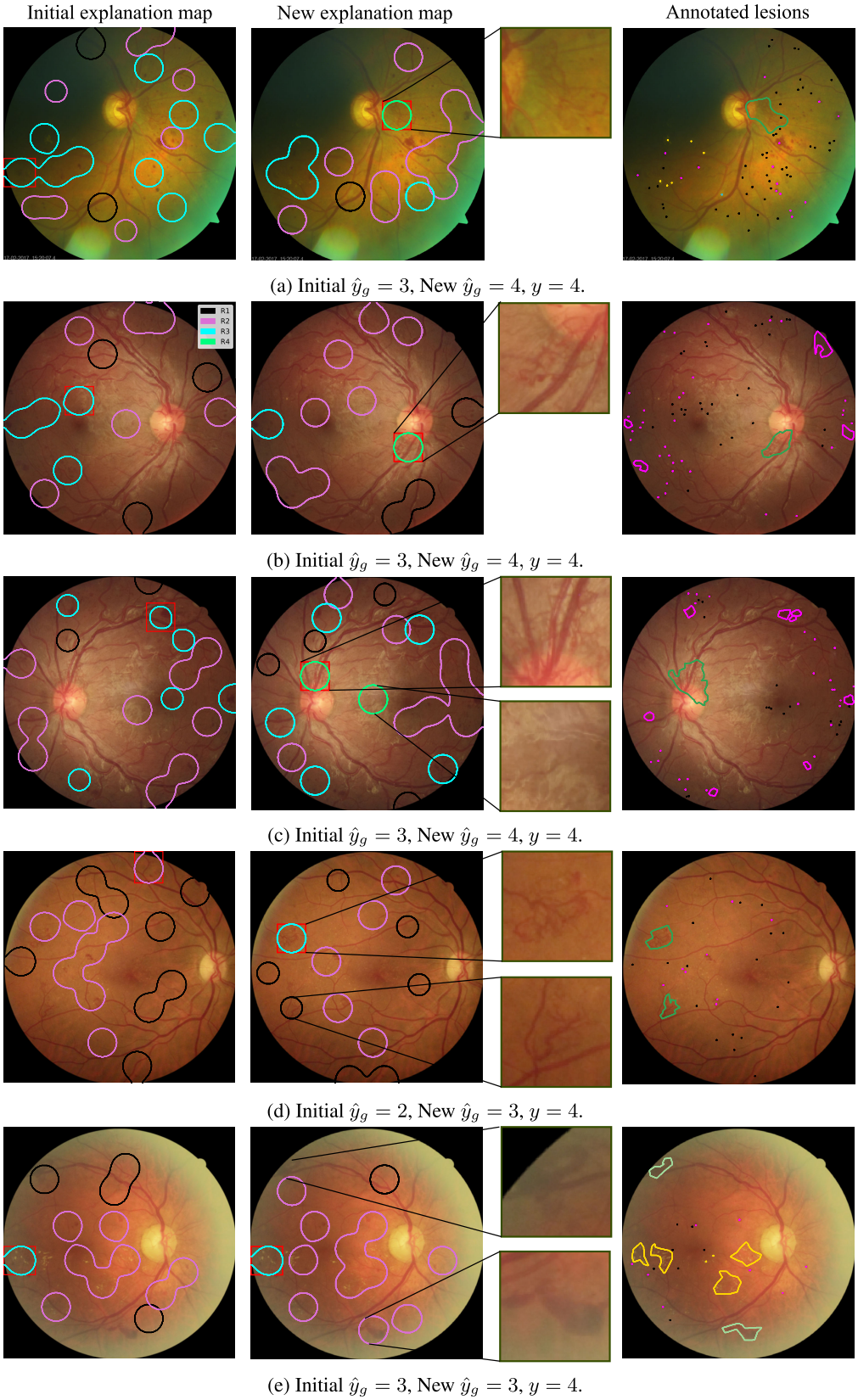
Dataset	SCREEN-DR	Kaggle test	Messidor-2
κ - initial	0.74	0.74	0.71
κ - new	0.78	0.74	0.70

B. RESULTS AND DISCUSSION

The quadratic-weighted Cohen's kappa (κ) values of the new and of the initial DR|Graduate predictions for three different datasets are shown in Table 2. Additionally, in order to perform a grade-wise evaluation, confusion matrices of the initial and new predictions were computed, as shown in Fig. 6. These are normalized versions of the confusion matrices, obtained by dividing each value of the original matrix by the sum of the values of each row.

As shown, the κ values remain approximately the same in the public datasets (Kaggle and Messidor-2), and improved from 0.71 to 0.74 in the SCREEN-DR dataset. By inspecting the confusion matrixes one can see that the grade-wise in the Kaggle dataset the performance remained approximately the same. The fact that the performance in this dataset did not degrade suggests that the inclusion of the original R4 images together with the newly generated allowed the model to be able to classify in a similar manner the original R4-labeled images. For the Messidor-2, despite the κ value having remained the same, one can see that the performance on the R4 class, both in this dataset and in the SCREEN-DR, has improved, with more R4-labeled images being correctly detected.

Examples of the new and of the initial DR|Graduate's attention maps *vs* the specialist's annotations on SCREEN-DR



(a) Initial $\hat{y}_g = 3$, New $\hat{y}_g = 4$, $y = 4$.

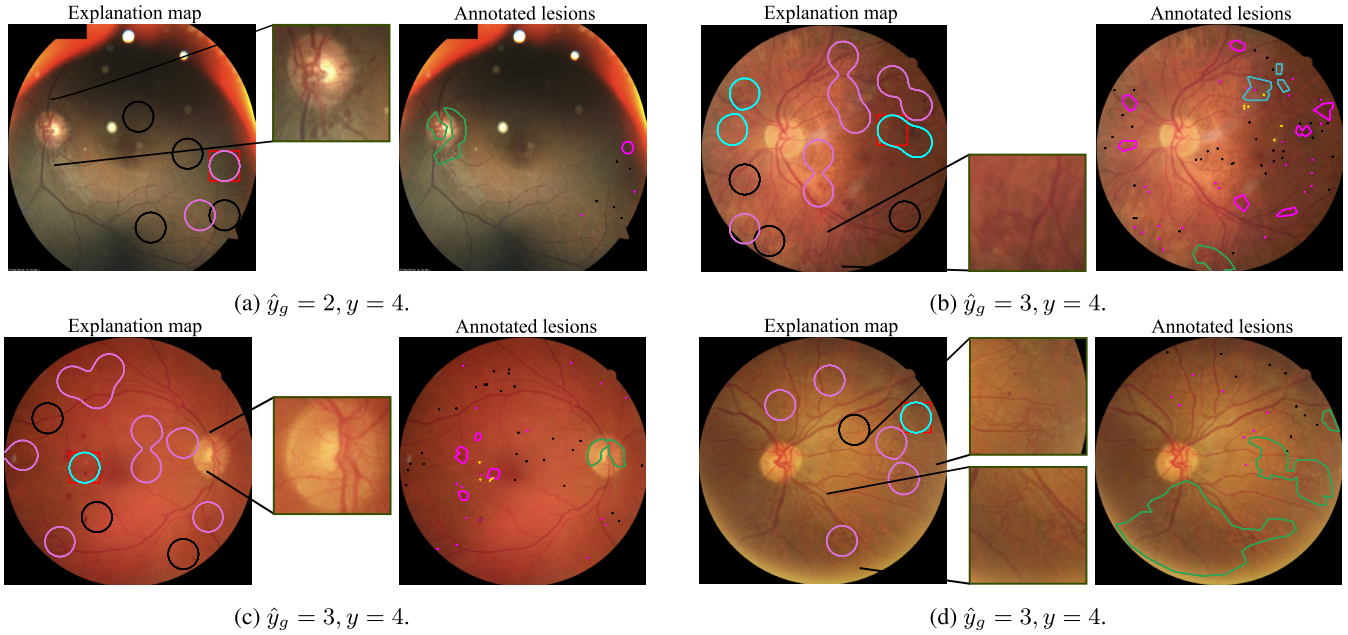
(b) Initial $\hat{y}_g = 3$, New $\hat{y}_g = 4$, $y = 4$.

(c) Initial $\hat{y}_g = 3$, New $\hat{y}_g = 4$, $y = 4$.

(d) Initial $\hat{y}_g = 2$, New $\hat{y}_g = 3$, $y = 4$.

(e) Initial $\hat{y}_g = 3$, New $\hat{y}_g = 3$, $y = 4$.

FIGURE 7. Maps predicted by our method and DR|Graduate along with the ophthalmologist's annotated lesions in the SCREEN-DR dataset. Curves correspond to the contours of the explanation maps (threshold=0.3), and the red square indicates the region of most relevance for diagnosis (corresponding to the maximum in the network's output activation map). Explanation map: ■ R1, ■ R2, ■ R3, ■ R4; ground truth map: ■ MA, ■ HEM, ■ EX, ■ CWS, ■ IrMA, ■ NV, ■ PHEM, ■ PFIB.



(a) $\hat{y}_g = 2, y = 4$. (b) $\hat{y}_g = 3, y = 4$.

(c) $\hat{y}_g = 3, y = 4$. (d) $\hat{y}_g = 3, y = 4$.

FIGURE 8. Maps predicted by our method along with the ophthalmologist's annotated lesions in the SCREEN-DR dataset in cases in which the model mispredicts R4-labeled images. Curves correspond to the contours of the explanation maps (threshold=0.3), and the red square indicates the region of most relevance for diagnosis (corresponding to the maximum in the network's output activation map). Explanation map: ■ R1, ■ R2, ■ R3, ■ R4; ground truth map: ■ MA, ■ HEM, ■ EX, ■ CWS, ■ IrMA, ■ NV, ■ PHEM, ■ PFIB.

images are shown in Fig. 7. In each predicted map, the region with maximum activation is identified by a surrounding square. The model is now capable of detecting NVs in eye fundus images, as shown in Fig. 7a,7b and 7c, which were initially not being detected by DR|Graduate [16]. However, in some cases other structures are counfounded with NVs and mistakenly detected as PDR signs (Fig. 7c). Additionally, NVs are sometimes misdetected as other lesions, such as IRMAs (Fig. 7d), characteristic of R3. To note that DR|Graduate originally was confounding the NV in this image with an HEM rather than an IRMA. The difficulty for specialists to distinguish between neovessels and IRMAs is widely referred in the literature, being specially true for early NVs [7].

One should note that PDR images, which do not present NVs but rather other PDR signs (PHEMs and PFIBs), may still be misclassified since these lesions are not contemplated in the proposed data augmentation scheme. However, one can see in Fig. 7e that despite not being able to identify the PHEMs in the image, the method identified one of them as being an hemorrhage, which is understandable due to the similarity HEMs and these small PHEMs.

A few NVs are still missed, as the cases shown in Fig. 8. Some of these vessels may present an unusual shape and appearance (Fig. 8a and 8b). NVs that are too slight were missed in certain cases (Fig. 8c), despite being contemplated in the NV generation (Fig. 5)-A2). Very large NVs, such as the ones shown from Fig. 8d, may be too large for the network to detect given its receptive field. The NV attributed colors at times may not be very realistic, which mainly happens due to the presence light artifacts, given rise to color aberrations, as shown in Fig. 5-F2 and F3.

IV. CONCLUSION

We proposed a data augmentation scheme to compensate for the lack of PDR cases in DR-labeled datasets. It builds upon a heuristic-based algorithm for the generation of neovessel-like structures which relies on the general knowledge of common location and shape of these structures. The synthesized NVs can be introduced in pre-existent retinal images which can then be used for enlarging the datasets for training deep neural networks. The insertion of the NVs in the retinal images has in account the color coherence with the neighbour vasculature to ensure a realistic insertion. Experiments were performed to assess the influence of this data augmentation scheme in the training of a previous proposed model for DR grading, DR|Graduate. Results have shown that the PDR detection performance has improved, with NVs that were being missed by the model now being highlighted in the explanation maps. However, part of the PDR images are still not detected since they do not present NVs but rather pre-retinal fibrosis or pre-retinal hemorrhages, which were poorly learned by the model. NVs which present an unusual shape or that are too slight are still being missed by the model, likely due to its lack of representation in the generated dataset. This study shows the potential of introducing NVs in retinal images for improving the detection of these proliferative DR signs, thus allowing to improve the performance of computer-aided DR grading methods and easing their clinical application.

REFERENCES

[1] *Diabetic Retinopathy Guidelines*, The Royal College of Ophthalmologists, London, U.K., Dec. 2012.

[2] S. Wild, "Estimates for the year 2000 and projections for 2030," *World Health*, vol. 27, no. 5, pp. 1047–1053, 2004.

[3] T. Scully, "Diabetes in numbers," *Nature*, vol. 485, no. 7398, pp. S2–S3, May 2012.

[4] S. Garg and R. M. Davis, "Diabetic retinopathy screening update," *Clin. Diabetes*, vol. 27, no. 4, pp. 140–145, Sep. 2009.

[5] L. Wu, P. Fernandez-Loaiza, J. Sauma, E. Hernandez-Bogantes, and M. Masis, "Classification of diabetic retinopathy and diabetic macular edema.," *World J. Diabetes*, vol. 4, no. 6, pp. 290–294, 2013.

[6] U. R. Acharya, E. Y. K. Ng, J.-H. Tan, S. V. Sree, and K.-H. Ng, "An integrated index for the identification of diabetic retinopathy stages using texture parameters," *J. Med. Syst.*, vol. 36, no. 3, pp. 2011–2020, Jun. 2012.

[7] J. K. Sun, P. S. Silva, J. D. Cavallerano, B. A. Blodi, M. D. Davis, and L. M. Aiello, "Retinal vascular disease," in *Ryan's Retina*, vol. 2, A. P. Schachat and S. R. Sadda, Eds. Amsterdam, The Netherlands: Elsevier, 2018, pp. 1091–1128.

[8] S. R. Group, "Four risk factors for severe visual loss in diabetic retinopathy: The third report from the diabetic retinopathy study," *Arch. Ophthalmol.*, vol. 97, no. 4, pp. 654–655, 4 1979.

[9] L. P. Aiello, M. T. Cahill, and J. S. Wong, "Systemic considerations in the management of diabetic retinopathy," *Amer. J. Ophthalmol.*, vol. 132, no. 5, pp. 760–776, Nov. 2001.

[10] G. S. Scotland, P. McNamee, S. Philip, A. D. Fleming, K. A. Goatman, G. J. Prescott, S. Fonseca, P. F. Sharp, and J. A. Olson, "Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland.," *The Brit. J. ophthalmol.*, vol. 91, no. 11, pp. 1518–1523, 2007.

[11] M. D. A. Moff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Opthalmol. Vis. Sci.*, vol. 57, no. 13, p. 5200, Oct. 2016.

[12] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," *Med. Image Anal.*, vol. 39, pp. 178–193, Jul. 2017.

[13] J. F. Arévalo, A. F. Lasave, D. G. Zeballos, and S. Bonafonte-Royo, "Diabetic Retinopathy," in *Retinal Choroidal Manifestations Selries Systemic Diseases*. New York, NY, USA: Springer, 2013, pp. 387–416.

[14] J. Krause, "Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy," *Ophthalmology*, vol. 125, no. 8, pp. 1264–1272, Aug. 2018.

[15] J. de la Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," 2017, *arXiv:1712.08107*. [Online]. Available: http://arxiv.org/abs/1712.08107

[16] T. Araújo, G. Aresta, L. Mendonça, S. Penas, C. Maia, Â. Carneiro, A. M. Mendonça, and A. Campilho, "DR|GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101715.

[17] E. Menti, L. Bonaldi, L. Ballerini, A. Ruggeri, and E. Trucco, "Simulation and Synthesis in Medical Imaging," in *Simulation and Synthesis in Medical Imaging* (Lecture Notes in Computer Science), vol. 9968, S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, Eds. Athens, GA, USA: Springer, 2016, pp. 167–176.

[18] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abramoff, A. M. Mendonca, and A. Campilho, "End-to-End adversarial retinal image synthesis," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 781–791, Mar. 2018.

[19] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Med. Image Anal.*, vol. 49, pp. 14–26, Oct. 2018.

[20] A. K. Pujitha and J. Sivaswamy, "Retinal image synthesis for CAD development," in *Proc. 15th Int. Conf. Image Anal. Recognit.*, vol. 10882, Jun. 2018, pp. 6211–6213.

[21] V. Bellemo, P. Burlina, L. Yong, T. Y. Wong, and D. S. W. Ting, "Generative adversarial networks (GANs) for retinal fundus image synthesis," in *Proc. Conf. Artif. Intell. Lecture Notes Bioinformat*, vol. 11367, 2019, pp. 289–302.

[22] T. Araújo, G. Aresta, A. Galdran, P. Costa, A. M. Mendonça, and A. Campilho, "UOLO—Automatic object detection and segmentation in biomedical images," in *Proc. Conf. Artif. Intell. Lecture Notes Bioinformat.*, vol. 11045, 2018, pp. 165–173.

[23] E. Decencière, "Feedback on a publicly distributed image database: The messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

[24] M. D. Abràmoff, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophthalmol.*, vol. 131, no. 3, pp. 351–357, 2013.

TERESA ARAÚJO received the M.Sc. degree in bioengineering engineering and biomedical engineering from the Faculty of the Engineering, University of Porto (FEUP), Porto, Portugal, in 2016, where she is currently pursuing the Ph.D. degree in electrical and computer engineering. Since 2014, she has collaborated as a Researcher with the Biomedical Imaging Laboratory, Center for Biomedical Engineering Research (C-BER), Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto. Her research interest includes the development of computer-aided diagnosis systems for diabetic retinopathy grading in eye fundus images.

GUILHERME ARESTA received the M.Sc. degree in bioengineering engineering and biomedical engineering from the Faculty of the Engineering, University of Porto (FEUP), Porto, Portugal, in 2016, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. He has been a Researcher with the Biomedical Imaging Laboratory, Center for Biomedical Engineering Research (C-BER), Institute for Systems and Computer Engineering, Technology, and Science (INESC-TEC), Porto, since 2014. His research interest includes computer-aided diagnosis systems for lung cancer assessment in chest computed tomographies.

LUÍS MENDONÇA received the M.D. degree from the Medical School, University of Porto, in 2005.

He became a Board Certified Ophthalmologist in 2010 after completing his residency at the Department of Ophthalmology, Centro Hospitalar São João, Porto, Portugal. From 2010 to 2011, he completed a Retinal Research Fellowship at the LuEsther T. Mertz Retinal Research Center, Manhattan Eye, Ear, and Throat Hospital, Department of Ophthalmology, New York, NY, USA. Since 2011, he has been a Surgical and a Medical Retina Specialist with the Department of Ophthalmology, Hospital de Braga, Braga, Portugal.

SUSANA PENAS is currently a Retina Specialist and a Consultant of Ophthalmology with the Department of Ophthalmology, Centro Hospitalar Universitário de São João, Porto, Portugal. She is also an Assistant Teacher of Ophthalmology with the Department of Surgery and Physiology, Faculty of Medicine, University of Porto. She is also a Principal Investigator and a Co-Investigator in many clinical trials related to retinal and choroidal pathology. Her main areas of research are medical retina, retinal and choroidal vascular pathology, angiogenesis, and retina imaging. She has or has been a member of the European Society of Retina Specialists, the American Academy of Ophthalmology, the Association for Research in Vision and Ophthalmology, the European Association for Vision and Eye Research, and the European Vitreo-Retinal Society. She has several publications in international indexed journals, and is a Reviewer for indexed journals.

CAROLINA MAIA is currently a Retina Specialist and a Consultant of Ophthalmology with the Department of Ophthalmology, Centro Hospitalar Universitário de São João, Porto, Portugal. Her areas of research include medical retina, retina imaging, and diabetic retinopathy.

ÂNGELA CARNEIRO is currently a Retina Specialist and a Consultant of Ophthalmology wth the Department of Ophthalmology, Centro Hospitalar Universitário de São João, Porto, Portugal, and a Professor with the Department of Surgery and Physiology, Faculty of Medicine, University of Porto. Her main research interests include age-related macular degeneration (AMD), angiogenesis, diabetic retinopathy, and retina imaging. She has been the Principal Investigator or a Co-Investigator in many clinical trials related to areas, such as AMD, choroidal neovascularization, and diabetic retinopathy. She is a member of the European Society of Retina Specialists, the American Academy of Ophthalmology, and the Association for Research in Vision and Ophthalmology. She has authored more than 50 original articles published in journals, such as *JAMA Ophthalmology*, *Retina*, *British Journal of Ophthalmology*, *Experimental Eye Research*, *Journal of Cellular Biochemistry*, *Acta Ophthalmologica*, *Ophthalmologica*, *Ophthalmic Research*, *Oxidative Medicine and Cellular Longevity*, and *European Journal of Ophthalmology*. She is a Reviewer for several indexed journals.

ANA MARIA MENDONÇA (Senior Member, IEEE) received the Ph.D. degree from the University of Porto (FEUP). She was a Researcher with the Institute for Biomedical Engineering (INEB) until 2014, but since 2015, she has been a Senior Researcher with INESC TEC. She was the Director of the Master in Biomedical Engineering from 2009 to 2014. She is currently an Associate Professor with the Department of Electrical and Computer Engineering (DEEC), Faculty of Engineering, University of Porto (FEUP). She is also a member of the Scientific Committee of the Doctoral Program in Biomedical Engineering. She has been collaborating as a Researcher and also as responsible in several research projects, mostly dedicated to the development of image analysis and classification methodologies aiming at extracting essential information from medical images in order to support the diagnosis process. Her past work has been mostly devoted to three main areas: retinal pathologies, lung diseases, and genetic disorders, but ongoing work is mainly focused on the development of Computer-Aided Diagnosis systems in Ophthalmology and Radiology. She has coauthored more than 100 papers in international journals and conferences, which attracted more than 2350 citations, according to Google scholar. Her current research interests include the areas of biomedical engineering, medical image analysis, computer vision, and machine learning.

AURÉLIO CAMPILHO (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Faculty of Engineering, University of Porto (FEUP), in 1985. He was an Assistant and an Associate Professor of Electrical and Computer Engineering with FEUP, where he was a Professor of Electrical and Computer Engineering, from 2003 to 2020. He was a Co-Founder of the Biomedical Engineering Institute (INEB), and from 1994 to 2000, he was the INEB President and a Research Coordinator. He promoted the creation of the first ever MSc and later PhD degrees in Biomedical Engineering in Portugal. He was one of the promoters of the M.Sc. on Bioengineering degree at FEUP and ICBAS (Instituto de Ciências Biomédicas Abel Salazar) from University of Porto. Since January 2014, he has been a Coordinator of the Center for Biomedical Engineering Research (C-BER), Institute for Systems and Computer Engineering, Technology, and Science (INESC TEC). Since April 2020, he has been a Jubilee Professor of Electrical and Computer Engineering with the University of Porto, and a Senior Researcher with INESC TEC. He has authored one book (with two editions), co-edited 20 books, and published more than 250 papers in international journals and conferences. His current research interests include the areas of biomedical engineering, medical image analysis, image processing, computer vision, and machine learning. He is a Fellow of the European Alliance for Medical and Biological Engineering and Science (EAMBES). He has served as an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING and *Journal of Machine Vision and Applications*. He is the General Chair of the series of International Conferences on Image Analysis and Recognition (ICIAR). He has served as an Organizer of several special issues and conferences.