Retinal artery/vein classification using genetic-search feature selection

Computer Methods and Programs in Biomedicine 161 (2018) 197–207

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Retinal artery/vein classification using genetic-search feature selection

Fan Huang[a], Behdad Dashtbozorg[a], Tao Tan[a,c], Bart M. ter Haar Romeny[b,a,*]

[a] Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Department of Biomedical and Information Engineering, Northeastern University, Shenyang, China
[c] Mammography, ScreenPoint Medical, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 22 November 2017
Revised 9 March 2018
Accepted 17 April 2018

Keywords:
Fundus image
Artery/vein classification
Genetic search feature selection

ABSTRACT

Background and objectives: The automatic classification of retinal blood vessels into artery and vein (A/V) is still a challenging task in retinal image analysis. Recent works on A/V classification mainly focus on the graph analysis of the retinal vasculature, which exploits the connectivity of vessels to improve the classification performance. While they have overlooked the importance of pixel-wise classification to the final classification results. This paper shows that a complicated feature set is efficient for vessel centerline pixels classification.

Methods: We extract enormous amount of features for vessel centerline pixels, and apply a genetic-search based feature selection technique to obtain the optimal feature subset for A/V classification.

Results: The proposed method achieves an accuracy of 90.2%, the sensitivity of 89.6%, the specificity of 91.3% on the INSPIRE dataset. It shows that our method, using only the information of centerline pixels, gives a comparable performance as the techniques which use complicated graph analysis. In addition, the results on the images acquired by different fundus cameras show that our framework is capable for discriminating vessels independent of the imaging device characteristics, image resolution and image quality.

Conclusion: The complicated feature set is essential for A/V classification, especially on the individual vessels where graph-based methods receive limitations. And it could provide a higher entry to the graph-analysis to achieve a better A/V labeling.

1. Introduction

Many systemic diseases including diabetes and hypertension cause blood vessels change (becoming tortuous, narrowing etc.) and even leakage may occur, leading to serious complications like blurry vision and hand/feet tingling and pain [1]. A retinal image provides direct access to vascular abnormalities and enables further quantitative analysis on the retinal vasculature. The study on retinal arteries and veins has received much attention in the field of retinal image analysis, since many artery-vein related biomarkers have been found significantly associated with the progress of diseases. In diabetic retinopathy (DR), the narrowing on arterioles and the widening on venules are observed, which result in a lower arteriolar-to-venular diameter ratio (AVR) of DR patients [2–4]. In hypertension, decrease on generalized arteriolar diameter is associated to the increased blood pressure level [5]. Additional measurements such as tortuosity (generalized vascular curvature) [6,7], vessel branching angle [8,9] and junction exponents [9,10] have received more and more interest. It is important to note that these clinical relevant features behave differently on arteries and veins respectively under pathological conditions. For instance, the arterial vessel wall is more elastic and thinner than the venous wall, thus abnormal arteries are usually more tortuous than veins [11]. Therefore, quantitative biomarkers extracted from arteries and veins separately might reveal more information for diseases progress rather than examining them together.

Due to the fact that high resolution fundus imaging is mostly low cost and fast, retinal screening programs usually produce huge amounts of data for analysis. It is then unrealistic to let human observers manually label the arteries and veins. Therefore, developing a fully automatic artery/vein (A/V) classification system is a prerequisite for automated large-scale retinal image analysis.

* Corresponding author at: Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.
E-mail address: b.m.terhaarromeny@tue.nl (B.M. ter Haar Romeny).

https://doi.org/10.1016/j.cmpb.2018.04.016

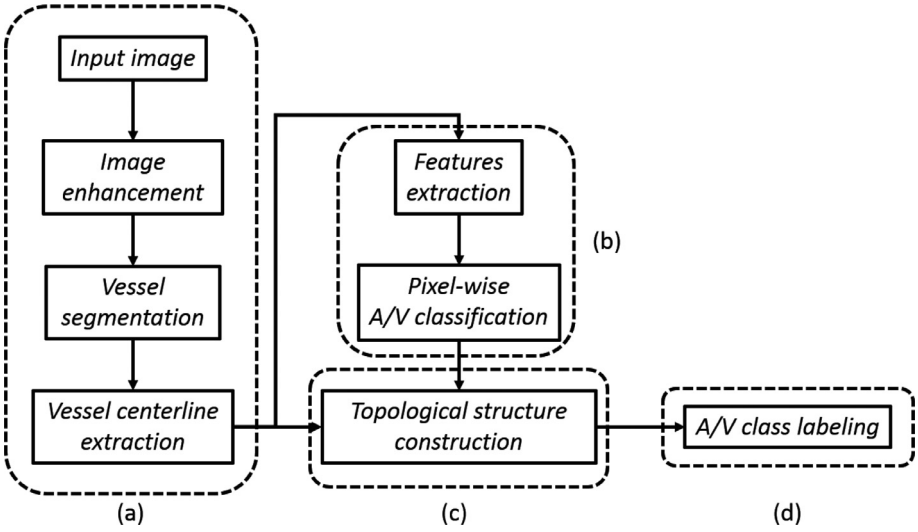


Fig. 1. The general pipeline for an automatic A/V classification program, where (a): image preprocessing, (b): pixel-wise A/V classification, (c): vessel topological structure construction and (d): A/V label assignment.

Several automatic A/V classification systems have been proposed in literature [12–16]. In summary, most of these methods consist of 4 main modules: (1) image pre-processing, (2) pixel-wise A/V probability assignment, (3) vessel topological structure construction and (4) A/V label determination (see Fig. 1). In the pre-processing step, images are enhanced by image processing techniques such as luminosity normalization and histogram equalization which correct for the illumination and background inhomogeneity. Afterwards, the retinal vessels are segmented yielding a vessel binary map. In the pixel-wise A/V classification module, intensity-based features are extracted for all vessel centerline pixels. Using these features, a supervised or unsupervised machine learning technique is exploited to assign a probability (between 0 and 1) to these pixels. After that, a topological structure of the vascular network is built in order to extract the connectivity relation between each individual segment. It improves the result of pixel-wise classification, because arteries only cross veins but not themselves and vessels connected to each other must be of the same type. Finally, the A/V label of each vessel segment, or even a full vessel tree is determined by using both the local and contextual information.

In the last few years, publications on A/V classification mainly focus on constructing the topological structure of vessels using graph theory. Joshi et al. used Dijkstra's graph search algorithm to connect vessel segments as subtrees and clustered them into arterial and venous classes [13]. Dashtbozorg et al. applied graph analysis on individual vessel segments and determined the type of vessels by combining the graph label and the pixel-wise A/V label [14]. Hu et al. constructed the vascular structure by a graph-based and a meta-heuristic algorithm [15]. Estrada et al. incorporated domain-specific features with a topology framework to construct a global likelihood model for A/V classification [17].

Exploiting vessel contextual information for A/V classification is novel, while good pixel-wise classification is also a crucial entry step. Because even if the graph analysis was perfect, an incorrect local A/V probability might still result in wrong A/V label estimation and further affect the corrected labeling of the whole vessel tree. The recent frameworks proposed in literature still use the information extracted by a small amount of features for supervised/un-supervised classification. Joshi et al. extracted only 4 features, Dashtbozorg et al. used 19 features (after feature selection), Niemeijer et al. [12] and Hu et al. [15] used 31 features, Mirsharif et al. [18] used 8 features (after feature selection) and Xu et al. [19] used 21 features for pixel-wise classification. Additionally, the category of features used in these works is limited, where only the local intensity values on multiple color channels (e.g. RGB, HSB and CIExyz) are used.

In this paper, we show that a more complicated feature set is more efficient in the discrimination of artery and vein. We developed a novel framework for pixel-wise A/V classification, which extracts features of different categories for vessel centerline pixels. An advanced feature selection technique, named genetic-search feature selection, is applied to obtain the optimal subset of features for classification. Then this framework was validated on five retinal image databases, including two public datasets and three clinical datasets.

2. Method

Arteries carry oxyhemoglobin which transports oxygen molecules from respiratory organs (e.g. lungs) to the rest of body (e.g. tissue), while veins carry deoxygenated hemoglobin which without the bound oxygen. Oxyhemoglobin is visually brighter and deoxygenated hemoglobin is darker. Therefore, on retinal images arteries are mostly brighter than veins, which makes the pixel intensities of vessels become very important features for discrimination. In the proposed method, we examined many intensity based features such as red, green, blue, hue, saturation, brightness etc. for every vessel centerline pixel. While if we simply feed the color intensities (such as RGB or HSB) of a pixel to a machine learning classifier, the classification results are usually disappointing. It turns out that our brain must take into account more information than considering only the local intensities to determine the vessel type. In addition, the fundus cameras installed in eye clinic are different from each other in the sense of field-of-view, image resolution, imaging flashlight and the embedded post-processing techniques. A well-trained system might work perfectly on the images from one dataset, but it may fail on the ones from others.

In this paper, we developed a framework to address this issue, which extracts a large amount of features for each vessel pixels, followed by a feature selection algorithm. The methodology starts by enhancing the contrast and correcting the luminosity variation of imported image. Afterwards, we apply vessel segmentation technique to obtain the vessel binary map as well as centerlines. For each centerline pixel we extract in total 455 features containing the categories of local intensity, contextual intensity, global intensity and spatial information. Then we use genetic-search

based feature selection technique to search for the subset of features giving the best performance using a supervised classifier. Using the selected features, the classifier assigns an A/V probability to each pixel (pixel-wise classification). Finally, the label of each vessel segment is determined by averaging the A/V probability values assigned to its pixels (segment-wise classification).

2.1. Image preparation

The imported images are preprocessed before the feature extraction step including (a) image luminosity normalization, (b) vessel segmentation and centerline extraction, (c) vessel width measurement and (d) optic-disc center detection.

2.1.1. Image luminosity normalization

Retinal images often suffer from local illumination and background variation. It is mainly due to the non-uniform illumination and the irregular surface of the retina. This problem seriously affects intensity-based A/V separation, because at the central region and the peripheral region of the image, blood vessels might have quite different appearances, where arteries might even be darker than veins. Therefore, we apply two illumination normalization approaches to the images and make it locally homogeneous.

The pixel intensity of a retinal image $f(x, y)$ can be modeled by an illumination-reflection model:

$$f(x, y) = r(x, y)\ l(x, y), \tag{1}$$

where $r(x, y)$ is the reflection property with regard to the absorbed light spectrum of a material, and $l(x, y)$ is the general luminosity around a small local area, which causes the inhomogeneity of pixel intensity. The normalization approach proposed by Foracchia et al. [20] divides the local pixel intensity by the average intensity within its neighbor to cancel the luminosity factor, as described by:

$$N(x, y) = \frac{r(x, y)\ l(x, y)}{\frac{1}{n^2}\sum_i^{n^2} r(x_i, y_i)\ l(x_i, y_i)} \approx \frac{r(x, y)}{\frac{1}{n^2}\sum_i^{n^2} r(x_i, y_i)}. \tag{2}$$

In the above equation, the numerator is the pixel intensity at position (x, y). The denominator is the mean filter applied to the $n \times n$ neighbor around (x, y). Since $l(x, y)$ is the image luminosity caused by the remote light source, we assume $l(x, y)$ within a certain region has little change, so $N(x, y)$ computes the ratio between local reflection and the average reflection inside its $n \times n$ neighbors.

In addition, we use another normalization method, the multi-scale Retinex method (MSR) proposed by Jobson et al. [21], which uses the logarithm transformation to eliminate the term $l(x, y)$. From the reflection model, we have

$$R(x, y) = \log I(x, y) - \log\left[G(x, y, \sigma) * I(x, y)\right], \tag{3}$$

where $I(x, y)$ is the original image intensity at position (x, y), $G(x, y, \sigma)$ is the Gaussian surrounding of (x, y) with scale σ and $*$ denotes the convolution operation. By taking the inverse-logarithm transform on $R(x, y)$, we obtain a luminosity homogeneous image.

2.1.2. Vessel segmentation and centerline extraction

Retinal vessel segmentation is a hot topic in retinal image analysis, and many techniques are developed and proposed in literature including supervised and un-supervised approaches [22–24]. The outcome of a segmentation technique is a vessel binary map, where the foreground is blood vessel and the background is retinal tissue. For artery/vein classification, a preferred vessel segmentation technique should solve two issues: closely parallel vessels and crossing vessels. Because in the region around the optic disc, arteries and veins are often closed and parallel to each other, thus they can be easily segmented as one merged vessel instead of two. In addition, arteries only cross with veins on retinal images. So a segmentation map that preserves the vessel junction points well would benefit further analysis and provide extra information for the classification. The segmentation technique proposed by Zhang et al. [23] applies multi-scale and rotating filters in a position and orientation domain named 'orientation scores'. An orientation score is a 3-D space with 3 axis: the spatial coordinate x, y and the orientation θ, in which vessels with different orientations lay in different planes. The benefit of this construction is that difficult cases like parallel closed vessels are discriminated by the utilization of spatial location and orientation, and vessel crossings are solved because they are disentangled. Rotating derivatives are taken in the directions that are perpendicular to the vessel structures at their corresponding orientation planes, which is similar to the vesselness filtering technique by Frangi et al. [25]. The multi-scale nature of the Gaussian derivative filters ensure that disentangled vessels with various sizes are equally enhanced. Afterwards the 3D structure is projected on the spatial plane by taking the maximum filter response over all orientations yielding a 2D enhanced vessel map. A proper threshold value is applied on the enhanced image to obtain a binary vascular map.

The vessels within the optic disc region are eliminated by the OD mask. An iterative thinning algorithm [26] is used to obtain the centerline of vasculature. Junction points like vessel branchings and crossings are also removed in a manner that pixels connected to each other represent an individual vessel segment.

2.1.3. Vessel width estimation

The caliber of vessels is measured in order to characterize them as small, medium or large vessels. It does not need to be accurate, thus we estimate the width values by a simple distance transform applied on the vessel segmentation map. It calculates the Euclidean distance d of every foreground pixel (blood vessel) to the nearest background pixel (not blood vessel), such that the value $2 \times d - 1$ of every centerline pixel represents the width of the vessel. In some cases, the extracted centerline pixel might not exactly be located at the same position as we found on the distance map. Therefore, we apply a maximum filtering process on the distance map, such that even if the extracted centerline deviates a little bit from the real one, we can still obtain the vessel caliber values. In this study, we estimate the centerline deviation as 2 pixels, therefore we apply a 5×5 maximum filter.

2.1.4. Optic-disc center detection

Detection and parameterization of the optic disc enables the creation of a binary mask which is used to remove the vessels within the optic disc (OD) region as a preprocessing step. In addition, the centralis position is used in calculating some of the spatial features, which are introduced in Section 2.2. In this work, we use the automatic OD detection technique proposed by Dashtbozorg et al. [27]. It uses a new convergence index operator named super-elliptical filter (SEF) to detect semi-elliptical convex shapes in the image. It performs well in localizing both the OD and the fovea. Furthermore, in order to prevent artifacts, such as the OD or the fovea is not clear in the image and the interference of other elliptical shapes like hemorrhages and large junction of vessels, a setup called paired SEF (PSEF) is introduced. It simultaneously localizes the OD and the fovea with two individual SEF filters which are located at a specified distance to each other. The PSEF filter is applied on the normalized green channel image which gives the best contrast. The locations of the OD and the fovea are determined by finding the position giving the maximum paired filter response.

2.2. Features extraction

For each centerline pixel we extract in total 455 features, which reveal information about the spatial location, local intensity,

Table 1
The complete set of features extracted for each centerline pixel.

Category	Description	Index
Local intensities	The pixel intensity of normalized R, G, B, H, Sat., Bri., RG, MSR-R, MSR-G, MSR-RG, Ill.	F1–F11
Vessel width	The vessel width estimated at each pixel.	F12
Spatial coordinate	Euclidean distance to **OD** and **image center**, angle with respect to **OD**.	F13–F15
Circular zone A	The mean, std, min, med and max of the 11 local intensities within a circular region with radius **0.5*vessel width**.	F16–F70
Circular zone B	The mean, std, min, med and max of the 11 local intensities within a circular region with radius **1.0*vessel width**.	F71–F125
Circular zone C	The mean, std, min, med, max of the 11 local intensities within a circular region with radius **2.0*vessel width**.	F126–F180
Centerlines	The mean, std, min, med and max of the 11 local intensities along **every centerline**.	F181–F235
Vessel segments	The mean, std, min, med and max of the 11 local intensities within **every vessel segment**.	F236–F290
All centerlines	The mean, std, min, med and max of the 11 local intensities of **all centerline pixels**.	F291–F345
All vessel segments	The mean, std, min, med and max of the 11 local intensities of **all vessel segment pixels**.	F346–F400
Whole field-of view	The mean, std, min, med and max of the 11 local intensities of **all pixels inside the field-of-view**.	F401–F455

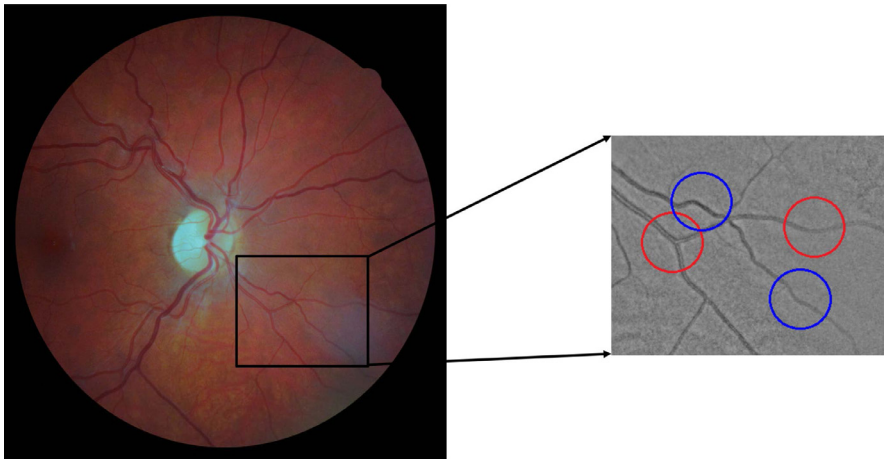


Fig. 2. The discrimination between an artery and a vein is obvious when they are close to the center of the image (right figure: top left), while it becomes difficult at the peripheral part (right figure: bottom right), where the red circles indicate the arteries and the blue circles indicate the veins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

neighborhood intensity and global intensity (as summarized in Table 1). In followed section, we introduce the feature categories that are considered for A/V classification.

2.2.1. Polar coordinates

First, according to the work by Zamperini et al. [28], spatial location has strong discriminative power in A/V classification. Therefore, we extract the polar coordinate for each centerline pixel with respect to the OD center. In addition, we measure the distance from every pixel to the image center. It is motivated by the fact that most fundus cameras focus the imaging light on the center of the image, resulting in that the central region is mostly clearer than the peripheral region giving more reliable information than the latter. Therefore, the spatial location might act as weighting factor to the other intensity-based features and be helpful for improving the classification performance. For instance, Fig. 2 shows an artery and a vein originated from the optic disc (top left) toward the peripheral region (bottom right). In the illumination-normalized red channel, the difference between the artery (red circle) and the vein (blue circle) is obvious when close to the OD, while the discrimination become much more difficult at the peripheral part, even for the same two vessels.

2.2.2. Local intensities

For each centerline pixel, we exploit 11 intensity-based features which are commonly used in literature [14,29]. They include luminosity normalized red (R), green (G), blue (B), hue (H), saturation (Sat.), brightness (Bri.) and the mean squared of red and green (RG), which is computed by $\sqrt{\frac{1}{2}(R^2+G^2)}$. The multi-scale retinex (MSR) values are computed on the R, G and RG channel. Finally, the image illumination term, $G(x, y, \sigma)^*I(x, y)$ with a sufficiently large σ, in Eq. 3 is computed on RG channel as the 11th local intensity feature.

2.2.3. Circular neighbor intensities

After the local intensity features, for all centerline pixels, we extend our feature vector by considering their neighborhood. Similar to the study by Zamperini et al. [28], for each color channel, three circular regions are considered around each centerline pixel with radius 0.5, 1 and 2 times of the vessel width, named zone A, zone B and zone C, respectively (as shown in Fig. 3). In every zone, the mean, standard deviation (std), median (med), minimum (min) and maximum (max) values are computed. Zone A measures the vessel central reflex (if it exists), zone B expands the measurement to the vessel segment and zone C looks at both the vessel and background pixels. Moreover, we compute the five measurements on the pixel intensities along each centerline, in order to keep the classification consistent for every vessel segment. Finally, the automatic vessel segmentation method might not produce perfect segmentation on all vessels, which means the centerlines we obtain are not all exactly at the center of vessels. Therefore, our contextual features consist of the five measurements calculated on the pixels within the whole vessel segment for each vessel.

2.2.4. Global intensities

In order to avoid the effect of color differences between images, we introduce a set of global features to our feature vector, which represents the characteristics of the whole image. The mean, standard deviation, median, minimum and maximum are calculated on the pixels of all vessel centerlines. With the same motivation

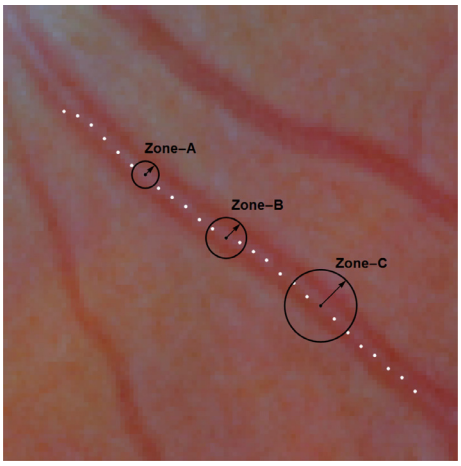


Fig. 3. Three circular regions centered at each centerline pixel with radius 0.5, 1 and 2 times of the vessel width are drawn, named zone A, zone B and zone C respectively. The mean, std, med, min and max of the intensity values within each region are used as features.

discussed above, we also calculate the five measurements on all vessel segments, and within the whole field of view.

2.3. Genetic-search feature selection technique

After the feature extraction step, a novel framework based on evolutionary computation is used to find the subset of features giving the highest pixel-wise A/V classification accuracy. The proposed feature vector contains 455 features, which results in an extremely high dimensional space for feature selection. Traditional feature selection procedures such as the greedy forward/backward selection and the best-first selection can easily stop at any local minima. Therefore, we exploit an advanced evolutionary algorithm which is inspired by the natural selection process in biology [30]. In nature selection, the fitness of individuals, which represents the chance they survive in the environment, is determined by a unique combination of phenotypic traits. Phenotypic traits are encoded by a genetic sequence in chromosomes, and thus they are inheritable. They are propagated from the elder to the next generation via reproduction. After several generations, favorable traits gather together yielding a population with great fitness.

When this concept is exploited for feature selection, the phenotypic traits are the extracted features, and the genetic sequence is a binary vector where the elements 1 or 0 indicate if a feature is used or not. The fitness of each gene is the performance of the classification using the indicating subset of the features. For every generation, we select the parents and obtain their offspring by mating and mutation (as shown in Fig. 4). After expanding the population pool, the fitness of every individual is computed and the survivors are selected. By simulating the evolution, we gradually converge to the global optimum in the feature space, and finally end up with the subset of features giving the desired accuracy.

The evolution is initialized by setting randomized sequence $C_i = \{c_1, c_2, \ldots, c_n\}$ in the population pool where $c_n \in \{0, 1\}$ and n is the length of sequence (the number of feature in use). The initial sequences have different levels which are determined by a level function $l(C_i) = \sum_{j=1}^{n} c_j$. Since the C_i is a binary vector, level $l(C_i)$ represents the number of features in use for each sequence.

At each generation, individuals are paired with each other in the population pool yielding parents (C_i, C_j). Then two variation operators, mating and mutation, are performed on (C_i, C_j) to obtain their offspring. Firstly, we apply mating to expand the pool, which randomly select elements of the two individuals and exchange them. The number of exchanged parts is a percentage, $p_{mate} \in (0, 1)$, of the length of C_i. In addition, during selecting the recombination parts, we use a parameter $p_{pres} \in (0, 1)$ to determine the amount of 1s and 0s to be selected. Therefore, we exchange the information of using and not using certain features during mating. After obtaining the offspring, we compute the fitness, which is the performance of a simple classifier, such as accuracy, sensitivity or specificity, on a given dataset, of every individual. Genes given higher fitness are directly survive (by using a predefined criterion), and the remains move forward to the next step where mutation happen. By mutating, we randomly select several elements of the remained individuals, then switch their values. A predefined value $p_{mute} \in (0, 1)$ determines the number of elements to be mutated. Afterwards, the fitness of mutated genes are recalculated and the survivors are chosen. Finally, two groups of survivor are combined and fed to the next generation. The evolution progress stops after a certain number of generations, or the average fitness in the pool reaches a threshold value. The framework is summarized in Algorithm 1.

Algorithm 1 Genetic-search feature selection.

```
Initilize population pool with random candidates
Evaluate the fitness of each candidate
repeat
    procedure MATING
        Select parents from the pool
        Recombine each parent to get their offspring
        Evaluate the fitness of resulted offspring
        for all offspring do
            if the fitness satisfies the survival criteria then
                Add the individual to the population pool
            else
                procedure MUTATION
                    Mutate the individual
                    Evaluate the mutated offspring
                    if the fitness satisfies the survival criteria then
                        Add the individual to the pool
                    else
                        Eliminate the individual
until termination condition is satisfied.
```

3. Experimental results

3.1. Materials

The proposed framework is validated on the images of five databases including the DRIVE, INSPIRE-AVR, Nidek, Canon and Topcon datasets:

DRIVE dataset is a public dataset provided by Staal et al. [31]. The images are fovea-centered and were acquired by a Canon CR5 non-mydriatic 3CCD camera with a 45 ° field of view (FOV) at resolution of 768 × 584 pixels. The dataset is originally split into a training set and a testing set, each of which contains 20 images. The ground truth of vessel segmentations and A/V labels of the DRIVE images, provided by Hu et al. named RITE dataset [32], are used for validation.

INSPIRE-AVR dataset (referred as INSPIRE) is a public dataset provided by Niemeijer et al. [33]. It contains 40 OD-centered images at resolution 2392 × 2048, where the vessel centerlines and the vessel types are labeled by Dashtbozorg et al. [34].

NIDEK dataset consists of 200 retinal images, where 100 are fovea-centered and 100 are OD-centered images, with size of 3744 × 3744 acquired in the Ophthalmology department of the Academic Hospital Maastricht (AZM) the Maastriche Study [35] in

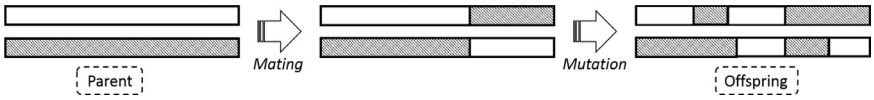


Fig. 4. Offspring are obtained via two steps: mating and mutation.

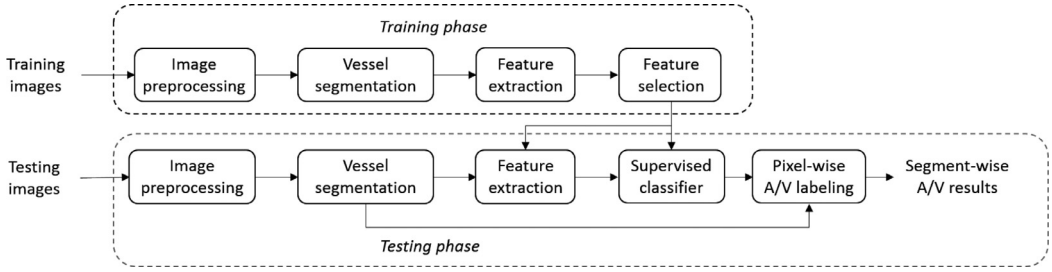


Fig. 5. The proposed framework is validated on 5 databases using the pipeline as shown.

the Netherlands. These images were captured using a NIDEK AFC-230 non-mydriatic auto fundus camera. The blood vessels are segmented by the automatic vessel segmentation described in previous section and the vessel types were labeled by experts using the manual annotation tool in "RHINO" software developed by Dashtbozorg et al. [36].

Canon dataset and **Topcon** dataset both contain 60 fovea-centered and 60 OD-centered retinal images. The images were captured by a Canon Cr-1 Mark II and a Topcon NW300 on 12 healthy subjects, and each subject received 5 acquisitions. The images of Canon camera have size of 3456 × 2304 pixels and the images of Topcon camera have size of 2048 × 1536 pixels. Similar to the NIDEK dataset, the automatic vessel segmentation is used for vessel extraction, the A/V labels are obtained using the manual annotation tool in "RHINO" software.

3.2. Experiment settings

The A/V features are extracted directly for vessel centerline pixels on the original size images. The feature extraction process is implemented in a CUDA parallel programming. Half of the images from each dataset are used for feature selection and classifier training and the rest are for testing. In the training phase, the training images are firstly processed by image normalization and vessel segmentation. Vessel centerlines including large, medium and tiny size of vessels are extracted, and the features are obtained as discussed above. Afterwards, the optimal feature subset is found via the genetic-search feature selection technique. The fitness of each gene is computed as the average accuracy of a Linear Discriminate Analysis (LDA) classifier using a 10-fold cross-validation on the training data. Since we are dealing with a large set of data, we set the stop criteria for all datasets to be 10 generations, and the individuals of 100 features are prior to survived at each generation. When the evolution is terminated, the individual giving the highest fitness is selected and a final LDA classifier is trained using all training data with the optimal features. For the test phase, the same preprocessing steps are applied to the test images. The trained classifier assigns a probability value (between 0 (vein) and 1 (artery)) to each centerline pixel, and a threshold value of 0.5 is used for the A/V label decision. The performance of classification on the centerline pixels, named pixel-wise classification, is evaluated by computing the sensitivity (arteries classified correctly), the specificity (veins classified correctly) and the accuracy (the average of sensitivity and specificity):

$$Sensitivity = \frac{TP}{TP + FN}, \tag{4}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

where *TP*, *TN*, *FP* and *FN* represent the true positive, true negative, false positive and false negative, respectively. In addition, we apply a voting procedure to the A/V probabilities, named segment-wise classification, where the label of each segment is determined as the majority vote of pixels belong to that segment. Then the pixel labels are corrected based on the segment label, and the same performance measurements are calculated (see Fig. 5).

3.3. Results

For each dataset, we apply a 10-fold cross validation on the test images. The results are summarized in Table 2. In this table, we categorize the datasets into fovea-centered images and OD-centered images, where the DRIVE dataset contains only fovea-centered images, the INSPIRE dataset contains only OD centered and the rest contain both types. The results are shown in terms of accuracy (Acc.), sensitivity (Sens.), specificity (Spec.) and area-under-the-curve (AUC). In Table 3, we compare the performance of our framework with the most recent works on A/V classification using the two publicly available datasets: the DRIVE and the INSPIRE. In Table 4, we compare our results on the INSPIRE dataset with the results of greedy-forward feature selection, principal component analysis (PCA) and without using any feature selection techniques.

Figs. 6 and 7 illustrate the sample of A/V classification results for each dataset using the proposed framework. At each row, we show the original retinal image, the A/V ground truth for the vessel centerline, pixel-wise classification and segment-wise classification results. The vessels with red and blue color represent the correctly classified artery and vein, while the yellow color represents the wrongly classified vessels.

4. Discussion

In this paper, we present a framework for automatic pixel-wise artery/vein classification on retinal color images. We have validated the framework on various datasets including publicly available datasets: the DRIVE and the INSPIRE dataset, and three clinical datasets: the NIDEK, the Canon and the Topcon datasets. We divided the dataset into fovea-centered (FOV) and optic disc-centered (OD) groups, where the two types of images are validated separately. Moreover, each group is equally divided into a training

Table 2
The performance of the proposed framework on the DRIVE, INSPIRE, NIDEK, Canon and Topcon datasets.

Dataset	Resolution	Fovea-centered images					Optic disc-centered images				
		Num.	Acc.	Sens.	Spec.	AUC	Num.	Acc.	Sens.	Spec.	AUC
DRIVE	565 × 584	20	72.0%	70.9%	73.8%	0.78	–	–	–	–	–
INSPIRE	2392 × 2048	–	–	–	–	–	20	90.2%	89.6%	91.3%	0.95
NIDEK	3744 × 3744	50	81.1%	81.3%	81.6%	0.89	50	83.6%	83.2%	84.9%	0.91
Canon	3456 × 2304	30	76.8%	77.8%	75.3%	0.84	30	78.3%	79.4%	76.1%	0.85
Topcon	2048 × 1536	30	82.5%	83.5%	81.8%	0.90	30	86.9%	87.6%	86.2%	0.93

Table 3
Result of the proposed framework compared with the most recent A/V classification on the DRIVE and INSPIRE dataset. Bold values indicate the best performance in each column.

Method	Num.[a]	DRIVE				INSPIRE			
	features	Accu.	Sens.	Spec.	AUC	Accu.	Sens.	Spec.	AUC
Proposed framework	100	72.0%	70.9%	73.8%	0.78	**92.0%**	89.6%	**91.3%**	**0.95**
Hu et al. [15]	31	88.0%	–	–	–	–	–	–	–
Estrada et al. [17]	–	**91.7%**	**91.7%**	**91.7%**	–	90.9%	**91.5%**	90.2%	–
Dashtbozorg et al. [14]	30	87.4%	90.0%	84.0%	–	84.9%	–	–	–
Niemeijer et al. [12]	27	–	80.0%	80.0%	**0.88**	–	–	–	0.84

[a] Num. feature: the number of features used for the classification.

Table 4
The comparison between genetic search approach, greedy forward approach, greedy backward approaches, PCA and without feature selection using the INSPIRE dataset. Bold values indicate the best performance in each column.

Dataset	Method	Acc.	Sens.	Spec.	AUC
INSPIRE	Genetic-search	**90.2%**	**89.6%**	**91.3%**	**0.95**
	Greedy-forward	83.0%	81.7%	85.7%	0.87
	PCA	85.6%	85.4%	86.0%	0.86
	No feature selection	85.5%	85.4%	86.5%	0.89

set and a testing set. The framework extracts 455 features for each vessel centerline pixel including information over multi-scales scopes (from local pixel intensity to global luminosity). On the training images, we use a genetic-search based feature selection technique to look for the subset of features giving the highest performance. Afterwards, a LDA classifier is trained with the optimal subset of features and validated on the testing sets.

As we can see from Table 2, the proposed framework achieved an average pixel-wise accuracy of 83% on high resolution retinal images (more than 3 megapixels) including the INSPIRE, NIDEK, Topcon and Canon dataset. However, on the DRIVE dataset where the images have the resolution of only 0.45 Megapixels, we obtain a lower performance compared to other works in A/V classification (as shown in Table 3). This is because when images have low resolution, our feature vector provides limited information for the classification of centerline pixels. For instance, features F16–F180 measure the intensity within three circular regions centered at the centerline pixels, with the radii of 0.5, 1.0 and 1.5 of the corresponding vessel width. On the low resolution images like the DRIVE dataset, the average width of trunk vessels is around 7 pixel, which results in the circular regions have almost the same size (radii of 2, 3.5 and 5 pixels). Therefore, the mentioned features (F16–F180) do not contribute significantly for the discrimination of arteries and veins in low resolution images.

In Table 3, we compare our results with recent works on the INSPIRE dataset. Our pixel-wise approach achieve an average accuracy of 92.0%, sensitivity of 89.6% and specificity of 91.3%, which performs slightly better than the techniques proposed by Estrada et al. [17] and outperform than the framework proposed Dashtbozorg et al. [14], where complicated graph analysis is exploited. Additionally, our AUC value (0.95) is much better than the value (0.84) obtained by Niemeijer et al. [12]. In Table 4, the performance of genetic-search feature selection is compared with greedy-forward feature selection, principal component analysis and without feature selection on the INSPIRE dataset. As we can see, the subset of features selected by genetic-search algorithm provides better result than the other three regular strategies. Note that the performance of the greedy-forward feature selection is even worse than using the full feature set, which is similar to the findings reported by Niemeijer et al. [12].

In Fig. 6 and Fig. 7 we show the examples of pixel-wise and segment-wise classification results for each dataset. First of all, the results demonstrate the good performance of our framework on large and small vessels near the optic disc. These vessels are important in biomarker measurements such as central-retinal-arteriole-equivalent (CRAE), central-retinal-venous-equivalent (CRVE) and arterial-to-venous diameter ratio (AVR). However, the classification of vessels near the OD is a challenge task, since some of them have limited contextual information for doing graph-based analysis. Secondly, the performance of pixel-wise and segment-wise classification are very similar. In general, the voting procedure for deciding vessel segment labels improves the pixel-wise accuracy a lot, because it considers the average probability of all vessel segment pixels instead of using only one pixel. However, our feature vector has already examined information along the vessel centerlines (F181–F235, F291–F345) during the feature extraction. Therefore, in our framework the voting procedure is not necessary as it does not significantly improve the performance of classification, which could speed up the processing.

From the selected features for classifying the vessels in the FOV- and OD-centered images, we note that the contribution of features for both image types is different. The pie charts (inner and outer) shown in Fig. 8a compare the predictive power of multiple feature categories in terms of the FOV- and OD-centered images. Firstly, the best 100 feature subsets giving the highest accuracy on the training data are chosen, and the number of times that each feature is being selected in feature selection process is counted. Features which are selected more than 60 times are considered as the most important ones. The percentage values shown on the pie charts indicate the proportion the feature categories account for in the important-features set. As we can compare the two charts, the category of circular zone C, centerlines, vessel segments, all centerlines and whole field-of-view are almost identical. The category of circular zone A and B account for 16% and 13% for the OD images, while both of them are only 11% for the FOV images. This

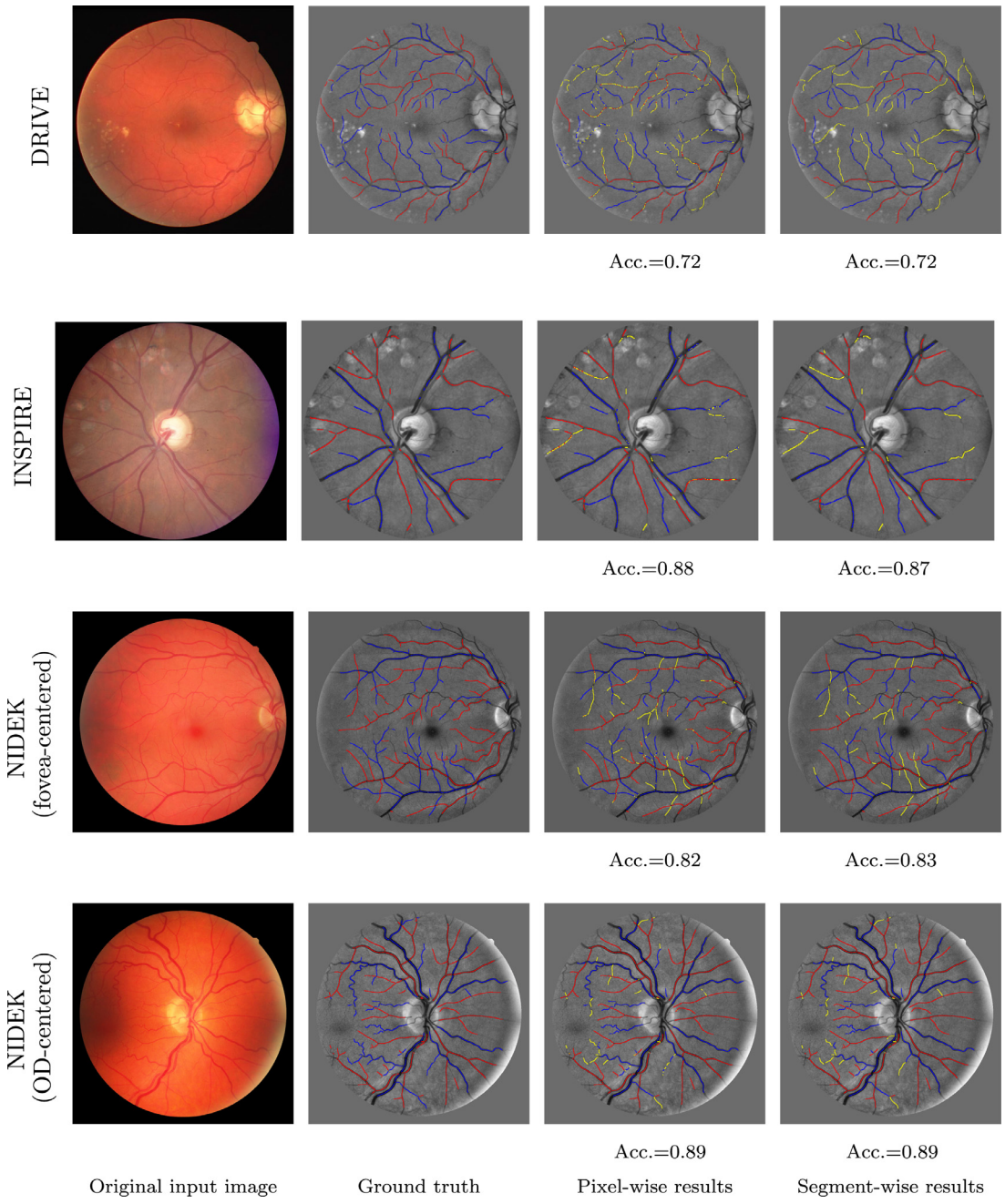


Fig. 6. A/V classification results on the DRIVE, INSPIRE and NIDEK datasets. The figures include: 1st column: the original images; 2nd column: the A/V label of the vessel centerlines; 3rd column: the pixel-wise classification and 4th column: the segment-wise classification. Red: correctly classified arteries; blue: correctly classified veins; yellow: wrongly classified vessels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is because the former contains vessels of mainly large and medium size, while the latter contains vessels of many different sizes. It implies that when classifying large vessels, the information extracted locally (by the small and medium regions) are more discriminative in comparison with the case when many small vessels are involved. In turn, when using only local intensities is not enough to classify small vessels, global information become predictive. As we can see in the charts that the category of all vessel segments for the OD images is 7%, while it increases to 13% for the FOV images. Additionally, spatial coordinate becomes an important category in the FOV images, while its count is less than 60 for the OD images.

We can also compare the characteristic of images acquired by different cameras through the selected feature sets. In Fig. 8b, we show the importance of different feature categories for the A/V classification on the INSPIRE, NIDEK, Canon and Topcon datasets using only the OD-centered images. Similarly, the number of times each feature is being selected by the best 100 subsets is counted. The percentage values shown on the vertical-axis indicate the proportion the features account for in the important-features set. As we can see from the charts, some of the features perform variously on the four datasets. For example, the red-related and brightness-related features for the INSPIRE dataset have a proportion of 14% and 10%, which are much higher than the corresponding values of the other 3 datasets. It implies that these two intensity categories are more descriptive for the images of the INSPIRE than NIDEK, Canon and Topcon. For the NIDEK camera, the green-related and saturation-related features account for 18% and 14%, while these two intensities have lower percentage values on the other datasets.

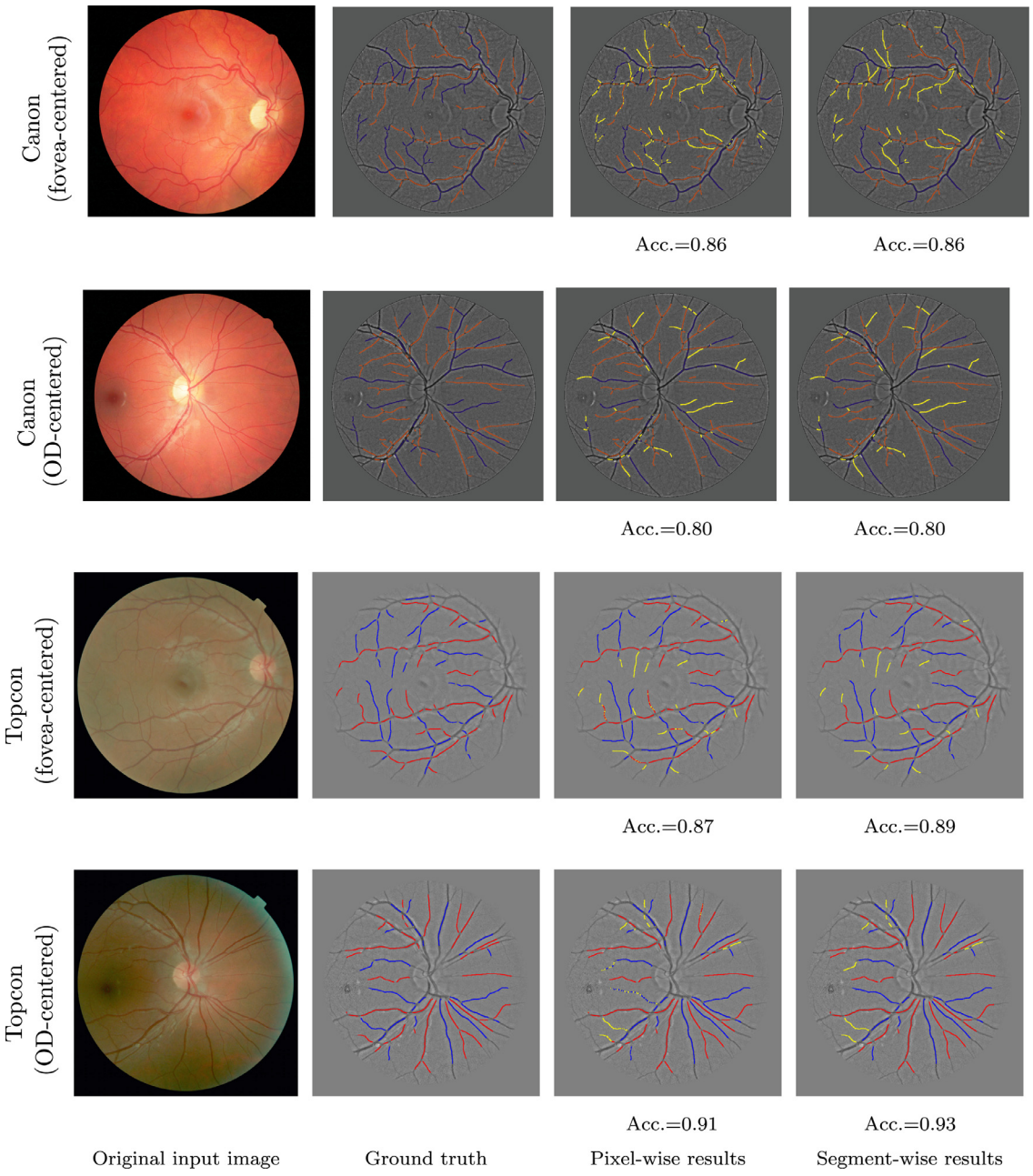
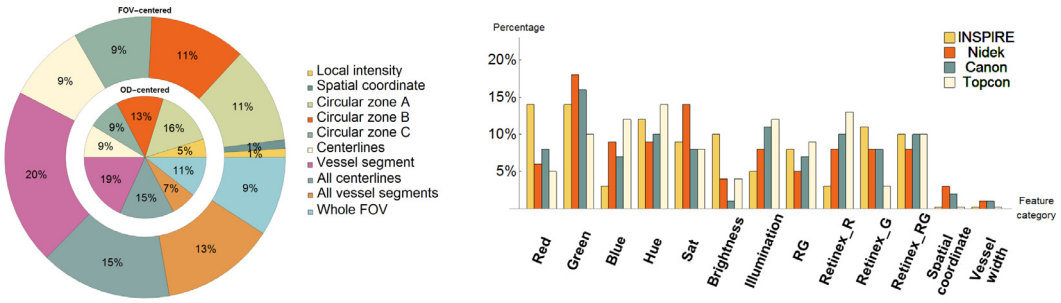


Fig. 7. A/V classification results on the Canon and Topcon datasets. The figures include: 1st column: the original images; 2nd column: the A/V label of the vessel centerlines; 3rd column: the pixel-wise classification and 4th column: the segment-wise classification. Red: correctly classified arteries; blue: correctly classified veins; yellow: wrongly classified vessels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Fovea-centered against OD-centered images (b) INSPIRE, NIDEK, Canon and Topcon datasets

Fig. 8. A comparison of the importance of different feature categories for the A/V classification on the (a) fovea-centered and OD-centered images. (b) INSPIRE, NIDEK, Canon and Topcon datasets.

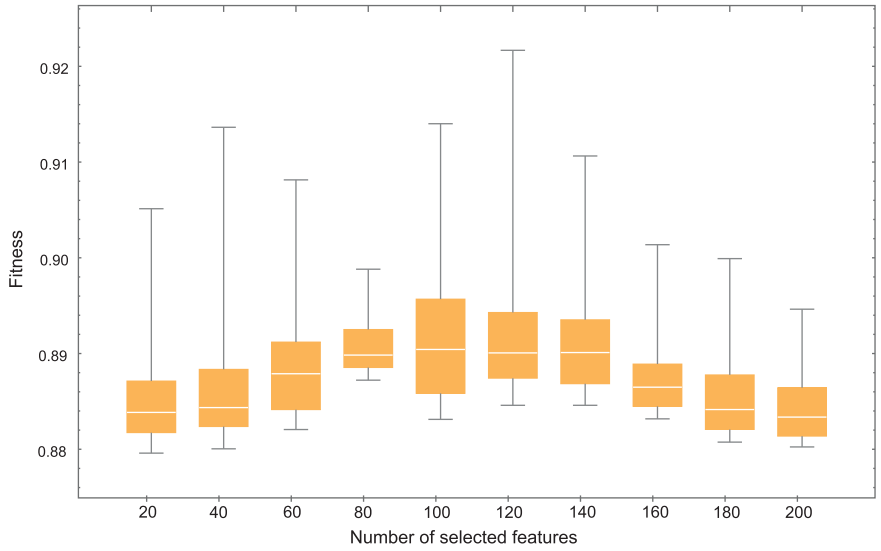


Fig. 9. Box plots of the individual fitnesses on the INSPIRE dataset obtained using various numbers of selected features (*x*-axis: number of feature, *y*-axis: fitness (accuracy) of individuals).

For the Canon dataset, the brightness related features only account for 2%, which is the lowest value among the corresponding values for all cameras. The retinex-green-related features becomes the least descriptive category on the Topcon images with only 3% proportion. These findings imply that our proposed framework is able to capture the special characteristics of the images, and perform a robust automatic pixel-wise artery/vein classification for the images by different fundus cameras.

Additionally, from Fig. 8b, we can learn the features which are the descriptive one among all the datasets. The green-related and hue-related features mostly account for more than 10% among all the datasets. The saturation-related features are also descriptive, where they account for 8% on INSPIRE, Canon and Topcon datasets and the percentage increases to 14% for the NIDEK. Moreover, the retinex-related features give discriminative information for the A/V classification on the 4 datasets. Most of them account for more than 8% on all the datasets, except for the values of the retinex-green on the Topcon and the retinex-red on the INSPIRE are only 3%. It suggests that these intensity features could provide simplified pixel-wise classification on images of multiple cameras with less feature extracted and performance preserved.

The box plots in Fig. 9 show the influence of individual levels (i.e. the number of features selected by the selection techniques) on the classification performance using the INSPIRE dataset. Each plot shows the distribution of accuracy obtained by individuals with level ranging from 20 to 200 with step size 20. As we can see, the performance of classification increases when more features are selected. The best performance 92.2% is found in the population pool of level 120, i.e. 120 features are selected. After that, the performance begins to drop. This reveals that fixing the individuals' level during the genetic search procedure helped improving the searching efficiency if the best individual levels are known/estimated. In this study, as we can see from the figure that the third quantile of level 100 gives better performance that the others, which means better individuals are more likely to be found in the pool of 100, therefore we fix the individual level at 100 in all the experiments.

5. Conclusion

In conclusion, we propose a framework for retinal artery/vein pixel-wise classification. It improves the accuracy by extracting a large set of features for each pixel. Afterwards, a genetic-search feature selection technique is used to select the optimal subset of features for classification. Our experimental result on public datasets shows that its performance on pixel-wise classification is already comparable with recent works using graph-analysis. The improved pixel-wise classification gives a higher entry to the graph-analysis step to achieve better classification. Moreover, it suggests that our framework is capable of blood vessels classification on images by different fundus cameras, as we use a large set of features which covers the characteristics of multiple cameras.

Acknowledgments

The authors would like to thank the University Eye Clinic Maastricht, Maastricht, The Netherlands for providing the retinal images. The work is part of the Hé Programme of Innovation Cooperation, which is financed by the Netherlands Organization for Scientific Research (NWO), dossier No. 629.001.003.

References

[1] American diabetes association (2015). http://www.diabetes.org/diabetes-basics/symptoms/.
[2] T.T. Nguyen, T.Y. Wong, Retinal vascular changes and diabetic retinopathy, Curr. Diab. Rep. 9 (4) (2009) 277–283.
[3] K. Guan, C. Hudson, T. Wong, M. Kisilevsky, R.K. Nrusimhadevara, W.C. Lam, M. Mandelcorn, R.G. Devenyi, J.G. Flanagan, Retinal hemodynamics in early diabetic macular edema, Diabetes 55 (3) (2006) 813–818.
[4] C. Sun, J.J. Wang, D.A. Mackey, T.Y. Wong, Retinal vascular caliber: systemic, environmental, and genetic associations, Surv. Ophthalmol. 54 (1) (2009) 74–95.
[5] A.S. Neubauer, M. Luedtke, C. Haritoglou, S. Priglinger, A. Kampik, Retinal vessel analysis reproducibility in assessing cardiovascular disease, Optom. Vis. Sci. 85 (4) (2008) E247–E254.
[6] C.Y.-l. Cheung, Y. Zheng, W. Hsu, M.L. Lee, Q.P. Lau, P. Mitchell, J.J. Wang, R. Klein, T.Y. Wong, Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors, Ophthalmology 118 (5) (2011) 812–818.
[7] E.J. Bekkers, J. Zhang, R. Duits, B.M. ter Haar Romeny, Curvature based biomarkers for diabetic retinopathy via exponential curve fits in SE(2), in: Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA 2015, held in Conjunction with MICCAI 2015, Iowa Research Online, 2015, pp. 113–120.
[8] B. Wasan, A. Cerutti, S. Ford, R. Marsh, Vascular network changes in the retina with age and hypertension, J. Hypertens. 13 (12) (1995) 1724–1728.
[9] N. Chapman, A. Mohamudally, A. Cerutti, A. Stanton, A.A. Sayer, C. Cooper, D. Barker, A. Rauf, J. Evans, R. Wormald, et al., Retinal vascular network architecture in low-birth-weight men, J. Hypertens. 15 (12) (1997) 1449–1454.
[10] N. Chapman, G. Dell'Omo, M. Sartini, N. Witt, A. Hughes, S. Thom, R. Pedrinelli, Peripheral vascular disease is associated with abnormal arteriolar diameter relationships at bifurcations in the human retina, Clin. Sci. 103 (2) (2002) 111–116.

[11] H.C. Han, Twisted blood vessels: symptoms, etiology and biomechanical mechanisms, J. Vasc. Res. 49 (3) (2012) 185–197.
[12] M. Niemeijer, X. Xu, A.V. Dumitrescu, P. Gupta, B. van Ginneken, J.C. Folk, M.D. Abramoff, Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs, IEEE Trans. Med. Imaging. 30 (11) (2011) 1941–1950.
[13] V.S. Joshi, J.M. Reinhardt, M.K. Garvin, M.D. Abramoff, Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks, PloS ONE 9 (2) (2014) 1–12.
[14] B. Dashtbozorg, A.M. Mendonça, A. Campilho, An automatic graph-based approach for artery/vein classification in retinal images, IEEE Trans. Image Process. 23 (3) (2014) 1073–1083.
[15] Q. Hu, M.D. Abràmoff, M.K. Garvin, Automated construction of arterial and venous trees in retinal images, J. Med. Imaging 2 (4) (2015) 1–6.
[16] S. Vázquez, B. Cancela, N. Barreira, M. Penedo, M. Rodríguez-Blanco, M. Seijo, G. de Tuero, M. Barceló, M. Saez, Improving retinal artery and vein classification by means of a minimal path approach, Mach. Vis. Appl. 24 (5) (2013) 919–930.
[17] R. Estrada, M.J. Allingham, P.S. Mettu, S.W. Cousins, C. Tomasi, S. Farsiu, Retinal artery-vein classification via topology estimation, IEEE Trans. Med. Imaging 34 (12) (2015) 2518–2534.
[18] Q. Mirsharif, F. Tajeripour, H. Pourreza, Automated characterization of blood vessels as arteries and veins in retinal images, Comput. Med. Imaging Graph. 37 (7) (2013) 607–617.
[19] X. Xu, W. Ding, M.D. Abràmoff, R. Cao, An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image, Comput. Methods Progr. Biomed. 141 (2017) 3–9.
[20] M. Foracchia, E. Grisan, A. Ruggeri, Luminosity and contrast normalization in retinal images, Med. Image Anal. 9 (3) (2005) 179–190.
[21] D.J. Jobson, Z.-U. Rahman, G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Trans. Image Process. 6 (7) (1997) 965–976.
[22] G. Azzopardi, N. Strisciuglio, M. Vento, N. Petkov, Trainable COSFIRE filters for vessel delineation with application to retinal images, Med. Image Anal. 19 (1) (2015) 46–57.
[23] J. Zhang, B. Dashtbozorg, E. Bekkers, J.P. Pluim, R. Duits, B.M. ter Haar Romeny, Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores, IEEE Trans. Med. Imaging 35 (12) (2016) 2631–2644.
[24] P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, IEEE Trans. Med. Imaging 35 (11) (2016) 2369–2380.
[25] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering, in: Proceedings of the Medical Image Computing and Computer-Assisted Intervention — MICCAI'98: First International Conference Cambridge, 1496, Springer, MA, USA, 1998, pp. 130–137.
[26] Z. Guo, R.W. Hall, Parallel thinning with two-subiteration algorithms, Commun. ACM 32 (3) (1989) 359–373.
[27] B. Dashtbozorg, J. Zhang, F. Huang, B.M. ter Haar Romeny, Automatic optic disc and fovea detection in retinal images using super-elliptical convergence index filters, in: Image Analysis and Recognition, Lecture Notes in Computer Science, 9730, Springer, 2016, pp. 697–706.
[28] A. Zamperini, A. Giachetti, E. Trucco, K.S. Chin, Effective features for artery-vein classification in digital fundus images, in: Proceedings of the Twenty-Fifth International Symposium on Computer-Based Medical Systems (CBMS), 2012, IEEE, 2012, pp. 1–6.
[29] M. Niemeijer, B. van Ginneken, M.D. Abràmoff, et al., Automatic classification of retinal vessels into arteries and veins, Med. Imaging (2009) 72601F.
[30] A.E. Eiben, J.E. Smith, et al., Introduction to evolutionary computing, 53, Springer, 2003.
[31] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509.
[32] Q. Hu, M.D. Abràmoff, M.K. Garvin, Automated separation of binary overlapping trees in low-contrast color retinal images, in: Proceedings of the Sixteenth International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI, 8150, Springer, Nagoya, Japan, 2013, pp. 436–443.
[33] M. Niemeijer, X. Xu, A. Dumitrescu, P. Gupta, B. van Ginneken, J. Folk, M. Abramoff, Inspire-avr: Iowa normative set for processing images of the retina-artery vein ratio, 2011, http://webeye.ophth.uiowa.edu/component/k2/item/270.
[34] B. Dashtbozorg, A.M. Mendonça, S. Penas, A. Campilho, Retinacad, a system for the assessment of retinal vascular changes, in: Proceedings of the Thirty-Sixth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 6328–6331.
[35] M.T. Schram, S.J. Sep, C.J. van der Kallen, P.C. Dagnelie, A. Koster, N. Schaper, R.M. Henry, C.D. Stehouwer, The maastricht study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities, Eur. J. Epidemiol. 29 (6) (2014) 439–451.
[36] B. Dashtbozorg, J. Zhang, S. Abbasi, F. Huang, B.M. ter Haar Romeny, Retinal health information and notification system (RHINO), in: Proceedings of the SPIE Medical Imaging, International Society for Optics and Photonics, 2017, pp. 1013437(1–6).