Corneal Nerve Tortuosity Grading via Ordered Weighted Averaging-based Feature Extraction

Pan Su[1,2], Tianhua Chen[3], Jianyang Xie[1], Yalin Zheng[4], Hong Qi[5], Davide Borroni[6], Yitian Zhao[1], Jiang Liu[7]

1.Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, 315300, China

2.School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

3.School of Computing and Engineering, University of Huddersfield, Huddersfield, HD1 3DH, United Kingdom

4.Department of Eye and Vision Science, University of Liverpool, Liverpool, L69 3BX, United Kingdom

5.Department of Ophthalmology, Peking University Third Hospital, Beijing, 100191, China

6.St. Paul's Eye Unit, Royal Liverpool University Hospital, Liverpool, L69 3BX, United Kingdom

7.Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China

Version typeset September 28, 2020

Yitian Zhao (yitian.zhao@nimte.ac.cn) and Jiang Liu (liuj@sustech.edu.cn)

Abstract

Purpose: Tortuosity of corneal nerve fibers acquired by *in vivo* Confocal Microscopy (IVCM) are closely correlated to numerous diseases. While tortuosity assessment has conventionally been conducted through labor-intensive manual evaluation, this warrants an automated and objective tortuosity assessment of curvilinear structures. This paper proposes a method that extracts image-level features for corneal nerve tortuosity grading.

Methods: For an IVCM image, all corneal nerve fibers are first segmented and then their tortuosity are calculated by morphological measures. The Ordered Weighted Averaging (OWA) approach, and the k-Nearest-Neighbor guided Dependent Ordered Weighted Averaging (kNNDOWA) approach are proposed to aggregate the tortuosity values and form a set of extracted features. This is followed by running the wrapper method, a supervised feature selection, with an aim to identify the most informative attributes for tortuosity grading.

Results: Validated on a public and an in-house benchmark data sets, experimental results demonstrate superiority of the proposed method over the conventional averaging and length-weighted averaging methods with performance gain in accuracy (15.44% and 14.34%, respectively).

Conclusions: The simultaneous use of multiple aggregation operators could extract

image-level features that lead to more stable and robust results compared with that using average and length-weighted average. The OWA method could facilitate the explanation of derived aggregation behavior through stress functions. The kNNDOWA method could mitigate the effects of outliers in the image-level feature extraction.

The table of contents is for drafting and refereeing purposes only. Note that all links to references, tables and figures can be clicked on and returned to calling point using cmd[on a Mac using Preview or some equivalent on PCs (see View - go to on whatever reader).

Contents

I. Introduction

The *in vivo* Confocal Microscopy (IVCM) is a non-invasive technique to imaging the corneal nerves, particularly, for the examination of the subbasal nerve plexus[1]. Since the IVCM was successfully applied to corneal nerve imaging in 2001[2], a number of studies[3,4,5] have shown that numerous properties of corneal nerve, such as nerve fiber branching, density, length, and tortuosity, are related to both eye conditions and systemic diseases. As the *tortuosity* can be used to interpret the degeneration and subsequent regeneration of nerves, which leads to active neural growth[5], substantial attention has been paid to tortuosity among other morphological properties of nerve fiber. For example, important correlation has been identified between the tortuosity of nerve and severity of diabetic neuropathy[6], which is one of the most common and serious long-term complications of diabetes[7]. In addition, tortuosity has also been associated with various ocular diseases, such as retinopathy of prematurity[8], herpes simplex keratitis[9], and fungal keratitis[10]. In order to reveal the correlation between the degrees of fiber tortuosity and associated medical conditions, the tortuosity levels of nerve fibers could be labeled in a rough band of 3-5 grades,[1,3,11] and could also be labeled using an interval of real numbers with refined resolution of 0.1 or even 0.01[12]. However, such empirical assessment is subjective. It may lead to substantial inter and intra-observer variability, and also making it susceptible to human errors[3,12]. With the ever increasing collection of high resolution IVCM images, the inefficient labor-intensive approach necessities an automated tortuosity assessment method.

Once the corneal nerves are traced in IVCM images, each nerve fiber can be represented as pixels which form a curvilinear structure and its tortuosity can then be measured. The definition and measurement of tortuosity has been extensively studied on medical[13,14] and other forms of images in the literature[15,16]. A number of measures have been defined with respect to different criteria, such as the length-based[17,18,19], the angle-based[12,13,15,20], and the curvature-based measures[14,21]. It is worth noting that most of these existing measures are designed for quantifying specific anatomical structures (such as retinal vessel[17], intracerebral vasculature[18], and corneal nerve[22]). As such, there is no universal agreement as to which standard or measure to apply for when quantifying the tortuosity of nerve fibers.

Many of the existing methods in the literature focus on defining and calculating the tortuosity of individual curvilinear structures, i.e., the fiber-level tortuosity. However, in working towards the automated grading of IVCM images with respect to the tortuosity of corneal nerve fibers, a step that has substantial influence on the quality of grading is extracting image-level tortuosity from fiber-level tortuosity. In the literature, this is conducted through the simple average of fiber-level tortuosity degrees or the weighted average by fiber lengths in many existing automated methods[11,23,24]. However, as the nerve fibers of varying lengths could exhibit considerably different tortuosity characteristics, this approach could lead to misclassification of nerve fibers, particularly those that consist of only a handful of highly twisted nerves among many other flat ones, which are empirically labeled highly tortuous by ophthalmologists[1]. Furthermore, when multiple measures are simultaneously utilized, the averaging methods for extracting image-level features can be different for different fiber-level tortuosity measures[25]. Although a number of researches have pointed out the importance of image-level feature extraction to tortuosity assessment[1,3,25], to our best knowledge, so far there is no such a pipeline that can choose the aggregation methods automatically rather than empirically calibrated.

In order to address the issues that may result from existing approaches, a module which enables automated aggregation of tortuosity on individual fibers is proposed and added to the conventional pipeline in this paper, whereby both experts-defined and data-driven weighting vectors are employed in the aggregation. To be more specific, an image-level feature extraction method based on the Ordered Weighted Averaging (OWA) and k-Nearest-Neighbor guided Dependent OWA (kNNDOWA) is proposed. For each fiber-level tortuosity measure, the tortuosity degrees of all nerve fibers in an image are aggregated by the OWA with a set of stress functions that aims to enhance the diversity and interpretability of extracted image-level features. Furthermore, the kNNDOWA is also employed to learn the weight of each nerve fiber by using an unsupervised approach. These initially generated features are refined by supervised feature selection techniques and the selected features are then fed into classifiers to perform the corneal nerve tortuosity grading. The proposed method (named as Mixed OWA and Feature Selection, MOWAFS) is verified on both a public and an in-house data sets. The in-house collection includes 300 images of the corneal subbasal nerve plexus obtained through a scanning laser confocal microscope in normal and pathologic subjects. Experimental results demonstrate the superior performance of the proposed methods over conventional approaches with aggregation operators including the averaging, maximum, and length-weighted averaging. It is worth noticing that there exists end-to-end models such as the use of convolutional neural network, which enables to directly generate predictions from images through a "black-box" model[26]. However, its application in grading the tortuosity of nerve fibers is not preferred so far, owing to the very limited labeled data available as well as the requirement for the extraction of meaningful knowledge to clinicians.

II. Materials and Methods

II.A. Materials

Two IVCM image data sets are employed in this paper. Apart from being tested on the public[1] Corneal Nerve Tortuosity data set[27] (indicated as PUB hereinafter), which consists of 30 images labeled into 3 grades as low, mid, and high, the proposed approach is also applied to a recently collected in-house data set with a larger collection. In order to validate the effectiveness of MOWAFS in clinical practice, the in-house data set (indicated as OWN hereinafter) comprises 300 images which are randomly selected from the IVCM library of Peking University Third Hospital. All images were taken in normal and pathological subjects with a Heidelberg Retina Tomograph HRT-III combined with Rostock Cornea Module. No preference on disease, age, or corneal location was set over the selection of images. However, images containing abnormal structures such as noticeable langerhans cells and obvious neuromas, were excluded to avoid biases in the automated segmentation of fiber nerves. The images are acquired in the view field of $400 \times 400 \mu m^2$ and are stored with the resolution of 384×384 pixels.

The images are graded into 4 levels of tortuosity based on the Laura protocol[2] by an experienced ophthalmologist. The original protocol categorizes the IVCM images into 5 grades with respect to their tortuosity: Grade 0, the nerve fibers appear almost straight; Grade 1, the nerve fibers are slightly tortuous; Grade 2, the nerve fibers appear moderately tortuous; there are frequent changes in the direction of the fiber, although these are of small amplitude; Grade 3, the nerve fibers are quite tortuous and the amplitude changes in the fiber direction can be quite severe; Grade 4, the nerve fibers appear very tortuous, showing abrupt and frequent changes in the nerve fiber direction. Since it is practically difficult

[1]Available at: http://bioimlab.dei.unipd.it/

for the clinician to discriminate the images of Grade 0 over those of Grade 1, the original Laura protocols of Grade 0 and 1 are merged. Therefore, the employed grading scales of the in-house data set are from Grade 1 to Grade 4 with images distributed in corresponding grades being 41, 173, 66, and 20, respectively.

As the first step of the automated tortuosity grading pipeline, segmentation is required to locate the nerves in IVCM images. The nerve fibers shown on IVCM images from both the PUB and OWN data sets are segmented by a recently proposed deep learning based algorithm named CS-NET[28]. In addition, for the PUB data set, the nerve fibers were also manually segmented by an ophthalmologist who traced the centerlines of all visible nerves. Depending on whether the images are segmented automatically or manually (indicated as -auto and -man, respectively), these result in three data sets (PUB-auto, OWN-auto, and PUB-man) in total for subsequent validation. Figure 1 shows examples of original in-house images as well as corresponding images segmented automatically. It can be seen from Figure 1F and 1G that artifacts such as small dendritic cells exist in some IVCM images. These artifacts may be regarded as dots or very short curves by the selected segmentation method. Therefore, a simple post-process is employed to delete the segments which are shorter than 10 pixels following the running of CS-NET.

II.B. Framework of Automated Tortuosity Grading

The conventional pipeline of automated corneal nerve tortuosity grading consists of nerve fiber segmentation, fiber-level and image-level feature extraction, and tortuosity classification (as shown in Figure 2). Given an IVCM image $Img_p \in \mathbb{U}$, the nerve fibers $Seg_1, Seg_2, \cdots, Seg_m$ on Img_p are first located by image segmentation techniques, which can be performed either through manual annotations or an automated algorithm. Then, M measures $F^1, F^2, \cdots, F^M$, which follow different criteria and standards, are calculated on each nerve segment $Seg_i, i = 1, 2, \cdots, m$ to characterize its degrees of tortuosity by real numbers. The collected tortuosity degrees of Seg_i with regard to the M measures are represented by $(f_i^1, f_i^2, \cdots, f_i^M)$, where large values indicate high tortuosity.

An aggregation operator $Agg : \mathbb{R}^m \to \mathbb{R}$ is then applied on the tortuosity degrees of all nerve fibers with respect to a certain measure, which generates image-level feature $(a_p^1, a_p^2, \cdots, a_p^M)$ for image Img_p. As mentioned before, the averaging and length-weighted averaging of all or selected segmented nerve fibers are commonly employed as the Agg. Formally, given the length of Seg_i denoted as l_i and the tortuosity degree of Seg_i with respect to measure F^I is f_i^I, $I = 1, 2, \cdots, M$, the image-level tortuosity can be calculated as length-weighted average[3]:

$$Agg^{\text{length}}(f_1^I, f_2^I, \cdots, f_m^I) = \frac{\sum_{i=1}^{m} l_i f_i^I}{\sum_{i=1}^{m} l_i} \tag{1}$$

or simply as arithmetic average:

$$Agg^{\text{average}}(f_1^I, f_2^I, \cdots, f_m^I) = \frac{\sum_{i=1}^{m} f_i^I}{m}. \tag{2}$$

In doing so, the M fiber-level tortuosity measures $F^1, F^2, \cdots, F^M$ are transformed into M image-level features $A^1, A^2, \cdots, A^M$, respectively. Finally, a classifier is trained to assign each image with one of N tortuosity grade labels $g_1, g_2, \cdots, g_N$. A feature selection algorithm can be applied on the image-level features to select the most discriminative ones for the tortuosity classification optionally[22,29].

II.C. Fiber-level Feature Extraction

The tortuosity has been estimated using various criteria, which are derived from corresponding geometric measurements such as length, angle, and curvature. Since there is no universal measure that can capture the characteristics of all types of tortuosity, multiple measures are employed simultaneously to evaluate the fiber-level tortuosity. The following introduces how various measures are calculated and utilized in this paper.

As the IVCM images are stored as pixels, the discrete approximation of geometric quantities is employed to measure the degree of tortuosity of a nerve fiber. Formally, given the centreline of nerve fiber Seg_i described by the ordered set of pixels $[(x_j, y_j)|j = 1, 2, \cdots, n]$, amongst which (x_1, y_1) and (x_n, y_n) represent the two ends of the centreline, the chord length L_x and curve length L_c, are defined as

$$L_x = \sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2}$$

and

$$L_c = \sum_{j=2}^{n} \sqrt{\Delta x_j^2 + \Delta y_j^2},$$

respectively, where $\Delta x_j = x_j - x_{j-1}, j = 2, 3, \cdots, n$. A simple and widely used measure of curvilinear structure tortuosity, i.e., the Arc Length over Chord Length Ratio, is then defined as the ratio between curve length and the chord length of Seg_i[30]:

$$\tau_L(Seg_i) = L_c/L_x. \tag{3}$$

A number of tortuosity measures are defined base on the concept of curvature K, which is a metric for indicating directional change of investigated curve. For each point (x_j, y_j) in Seg_i, the curvature $K(j)$ is defined as

$$K(j) = \frac{\Delta^2 x_j \Delta y_j - \Delta x_j \Delta^2 y_j}{(\Delta x_j^2 + \Delta y_j^2)^{3/2}},$$

where $\Delta^2 x_j = \Delta x_j - \Delta x_{j-1}, j = 3, 4, \cdots, n$. In[14], τ_C: the sum of absolute $K(j)$ and τ_{SC}: the sum of squared $K(j)$ over the whole segment Seg_i are employed as measures of the nerve directional variability. In[23], the maximum of absolute $K(j)$ over a corneal nerve fiber (indicated as τ_{MC}) is also used as a measure of tortuosity. The τ_C, τ_{SC}, and τ_{MC} are defined as follows:

$$\tau_C(Seg_i) = \sum_{j=3}^{n} |K(j)|, \tag{4}$$

$$\tau_{SC}(Seg_i) = \sum_{j=3}^{n} K(j)^2, \tag{5}$$

$$\tau_{MC}(Seg_i) = \max_{\{j=3,\cdots,n\}} |K(j)|. \tag{6}$$

An alternative curvature-based tortuosity measure has been proposed in[31], where the derivative of the curvature is used to quantify the directional change of a line. Similar to the τ_{SC}, the tortuosity level can be defined as the sum of the squared derivative of curvature:

$$\tau_{DCI}(Seg_i) = \frac{1}{L_c} \sum_{j=4}^{n} (K(j) - K(j-1))^2, \tag{7}$$

where the L_c is the curve length and $K(j)$ is the curvature of point (x_j, y_j) in Seg_i as described in the previous text.

It may be assumed that the greater the number of curvature sign is changed along a curve, the more tortuous the curve is. Therefore, several tortuosity measures are defined base on the concept of inflection points (also known as twists). The number of inflection points n' equals to the number of changes in sign of the curvature $K(j)$ for planar curves [17]. Since the tortuosity measure τ_L may not distinguish between smoothly curved structures and structures that make abrupt changes in direction, a new tortuosity based on τ_L, i.e., the Inflection Count Metric (ICM) is proposed in[18]:

$$\tau_{ICM}(Seg_i) = (n' + 1) \cdot \tau_L(Seg_i). \tag{8}$$

Moreover, if a turn curve $s_{j'}, j' = 1, 2, \cdots, n'+1$ is defined as the portion of a nerve segment Seg_i located between two consecutive twists (or the portion between one end of Seg_i to its nearest twist), it can be assumed that the greater the amplitude (maximum distance of the curve from the underlying chord) of a turn curve, the greater the tortuosity associated with it[1]. Then, the tortuosity of the nerve segment Seg_i is calculated as:

$$\tau(Seg_i) = \frac{n'}{n' + 1} \frac{1}{L_c} \sum_{j'=1}^{n'+1} (\tau_L(s_{j'}) - 1). \tag{9}$$

An angle-based tortuosity measure termed Slope Chain Code (SCC) is proposed in[15], where a curve is traced by a chain, which is essentially a sequence of fixed-length straight lines placed along the curve. The corresponding slope angle between such two adjacent straight line segments is employed to estimate the curvature of the point at which the end of a line segment and the original curve intersect. As the original curve is approximated by a sequence of constant-length segments in SCC, the selection of length will not only decide the number of sampling points in SCC calculation, but also affect the resultant tortuosity degrees. Therefore, since it is difficult to decide the length of line segments in SCC for corneal nerve fibers in this paper, the constant-length line segments in SCC are replaced by straight line segments between two points which achieve the local maximum and local minimum of K adjacently. Figure 3 illustrates the calculation of slope angle α at the local maximum point.

Given that the total number of points which achieve local maximal K and adjacent to two local minimal K points is n'', the slope angles at such points are $\alpha_{j''}, j'' = 1, 2, \cdots, n''$, the tortuosity measure τ_{MS} is defined as

$$\tau_{MS}(Seg_i) = \frac{1}{n''} \sum_{j''=1}^{n''} \alpha_{j''}. \tag{10}$$

As summarized in Table ??, eight geometric measures designed for tortuosity evaluation of curvilinear structures are employed for fiber-level feature extraction on IVCM images. Since there is no universal agreement as to which measure to apply for when quantifying the tortuosity of nerve fibers, this paper empirically uses a comprehensive range of measures, which are defined following different geometric standards. It is worth noticing that other fiber-level tortuosity measures can also be employed in the MOWAFS framework.

Each measure is deemed to be a mapping from nerve fibers to real-valued numbers where high values represent high tortuosity. Once the tortuosity of m traced nerve fibers in an IVCM image are obtained via such a fiber-level measure (the value of m may vary in different IVCM images), the next step aims to aggregate the tortuosity values of the m fiber segments to form an image-level tortuosity value. This is nontrivial as IVCM images usually contain a variable number of corneal nerve fibers which could show considerably different tortuosity characteristics[3]. The following subsections present the proposed method which can effectively extract image-level features for the tortuosity grading.

II.D. Image-level Feature Extraction

It turns out that[1] the averaged fiber-level tortuosity could be rather crude and cannot provide an accurate estimation of image-level tortuosity, particularly those images that consist of only a handful of highly twisted nerves among many other flat ones, but are empirically labeled highly tortuous by ophthalmologists. With the existing averaging method, high tortuosity values from a small amount of nerves are averaged out in comparison with low tortuosity values from the majority, which leads to low estimation of tortuosity at the image level. In order to solve this problem, the OWA based image-level feature extraction method is proposed, which aims to flexibly adjust the contributions made by different nerve fibers through tuning weights. As an alternative to OWA, in order to promote nerve fibers that are deemed more reliable and demote ones that are likely to be outliers, the kNNDOWA is also employed in the image-level feature extraction, by considering the similarity of individual fibers with regard to its nearest neighbors. The flowchart of the proposed method named as Mixed OWA and Feature Selection (MOWAFS) is shown in Figure 4.

II.D.1. OWA-based Feature Extraction

In case where multiple arguments are required to aggregate in order to produce a more robust outcome[32,33], the simple average, maximum, and minimum are among the popular aggregation operators. Apart from these conventional operators, the family of OWA operators is an alternative and more general type of operator for aggregation. The distinguishing aspect of OWA is the reordering step in which the input values are rearranged in descending/ascending order before they are integrated into a single aggregated one[34].

Formally, a mapping $Agg^{\text{owa}} : \mathbb{R}^m \to \mathbb{R}$ is called an OWA operator if

$$Agg^{\text{owa}}(f_1, f_2, \cdots, f_m) = \sum_{i=1}^{m} w_i f_{\pi(i)} \tag{11}$$

where $f_{\pi(i)}$ is a permutation of f_i, which satisfies that $f_{\pi(i)}$ is the i-th largest of the f_i, and $w_i \in [0, 1]$ is a collection of weights that satisfies $\sum_{i=1}^{m} w_i = 1$, $i = 1, 2, \cdots, m$.

The weights of an OWA operator are hereafter denoted as a weighting vector $W = (w_1, w_2, \cdots, w_m)$, in which w_i is associated with the i-th largest input values. Different from weighted averaging, the ordering of inputs gives OWA a nonlinear feature. Different choices of the weighting vector W can lead to different aggregation results. For example, the classical averaging is an special case of OWA by setting $w_i = 1/m$. The maximum operator can be formed by OWA with $w_1 = 1$ and $w_i = 0$ for $i \neq 1$, and the minimum can be formed by $w_m = 1$ and $w_i = 0$ for $i \neq m$. An important feature of the OWA operator is that it provides an output value between the maximum and the minimum of the arguments. A straightforward way of applying OWA to the image-level feature extraction is by defining a feature A^I based on the fiber-level tortuosity measure F^I, as its value of Img_p is $a_p^I = Agg^{\text{owa}}(f_1^I, f_2^I, \cdots, f_m^I)$, $I = 1, 2, \cdots, M$.

The conventional aggregation operators are inflexible in the utilization of expert perceptions to control the aggregation behavior. In OWA, a simple mechanism named stress function has been introduced for deriving weights with explicit andness/orness and attaining interpretability. Let the stress function[35] $h : [0, 1] \to \mathbb{R}^+$ be a non-negative function on the unit interval. The OWA weighting vector $W = (w_1, \cdots, w_i, \cdots, w_m)$ can then be defined as

$$w_i = \frac{h(\frac{i}{m})}{\sum_{i=1}^{m} h(\frac{i}{m})}, \tag{12}$$

such a function h is termed a stress function for OWA[35].

Since the number of segmented nerve fibers m varies in different IVCM images, the number of input values and the number of weights are different when OWA is applied to aggregate the fiber-level features on different images. Therefore, in the OWA-based image-level feature extraction, a stress function can be predefined with the corresponding weighting vector derived based on the number of nerve fibers on each IVCM image. The OWA weighting vector obtained with the associated stress function can be directly used to explain the overall aggregation behavior. That is, the values from a stress function $h(z)$ on the left side of $[0, 1]$ reflect weights associated with the larger inputs, i.e., nerve fibers with higher tortuosity degrees, whereas the values associated with the right side of the unit interval reflect the weights associated with smaller inputs, i.e., nerve fibers with lower tortuosity degrees.

Stress functions of different shapes can be used to impose constraints over the distribution of weights in OWA and hence resulting in different andness/orneess of the aggregation. Andness suggests that the aggregated result is influenced by smaller input values and the aggregation operator behaves similarly to conjunction, while orness indicates that the aggregated result is influenced by greater input values and the aggregation operator behaves similarly to disjunction. Figure 5 shows the examples of stress functions which define the behavior of OWA operators.

An indicator which can be adopted to quantify the andness/orness of an OWA aggregation operator is the Attitudinal Character (A-C)[35]. In particular, the attitudinal character of an OWA operator can be calculated from the stress function h as

$$\text{A-C}(h) = \int_0^1 \frac{\int_0^t h(z)\mathrm{d}z}{\int_0^1 h(z)\mathrm{d}z}\mathrm{d}t. \tag{13}$$

The value of attitudinal character gives an idea that an aggregation operator behaves similarly to conjunction/andness (influenced by smaller argument values) if A-C is closer to 0 or disjunction/orness (influenced by larger values) if A-C is closed to 1. It can easily be shown that the attitudinal character of the minimum, average, and maximum are 0, 0.5, and 1, respectively.

It is worth noticing that the attitudinal character can also be calculated from the weights by

$$\text{A-C}(W) = \frac{1}{m-1}\sum_{i=1}^{m}((m-i)w_i) \tag{14}$$

and A-C$(W) \rightarrow$ A-C(h) as $m \rightarrow \infty$. Since the number of nerve fibers m varies with different IVCM images and a stress function is used to derive weighting vectors for all images in a data set, the most accurate way to calculate A-C(h) is using Eqn. (13) directly. However, in practical experimentations and applications, it is more convenient to calculate an approximation of A-C(h) by using A-C(W) and Eqn. (12) while setting the value of m to a very large integer.

Instead of using the conventional minimum, average, and maximum, the OWA operator is able to generate aggregated results in between the minimum and maximum. More specifically in this paper, each of the $M(M = 8)$ tortuosity measures will form an input to twenty-one OWA operators whose A-C values are distributed in $[0, 1]$. The linear stress function $h(z) = 1 + u(z - 1)$ is employed to derive the weighting vectors for OWA-based image-level feature extraction. By setting $u = 0.2, 0.4, \cdots, 1.0, \frac{1}{0.8}, \frac{1}{0.6} \cdots, \frac{1}{0.2}$, nine weighting vectors are generated. For each of the generated weighting vector W, $1 - W$ is also employed. Including the features extracted by using minimum, average and maximum, a total number of M by 21 image-level features can be generated based on the M tortuosity measures for each IVCM image.

II.D.2. *k*NNDOWA-based Feature Extraction

While aggregating the tortuosity at nerve fiber level, instead of simply adopting a fixed set of weighting vectors, whose generation is independent of the specific input values, this paper also explores a data-driven generation of weights by considering the reliability of each individual inputs with respect to its neighbors. When automated segmentation algorithms are employed in the pipeline of tortuosity grading, inaccurate segmentation may generate outlier values in the tortuosity measurement. A typical weighting vector given by a certain OWA operator may suffer from assigning largest weights to outlier arguments (e.g., maximum and minimum), resulting in biased or even false results. Therefore, an unsupervised learning mechanism is also adopted to differentiate between nerve fibers that are deemed more reliable and those that are likely to be outliers by considering the interplay between their tortuosity values.

A type of OWA operator named Dependent OWA has been introduced in the literature[36], in which the values of inputs are used to determine the weights in the aggregation in order to produce reliable aggregated outcomes. In particular, the k-Nearest-Neighbor guided Dependent OWA (kNNDOWA)[37] determines the reliability of input values (i.e., the tortuosity degrees) by its nearest neighbors. This modeling of reliability helps differentiate amongst a set of nerve fibers in an IVCM image such that a certain tortuosity degree of a nerve fiber which is similar to those of other nerve fibers is deemed reliable and can be assigned a higher weight. In contrast, a tortuosity degree that is different from its neighbors is assigned a lower weight. Formally, the reliability of an input value f_i, $i = 1, 2, \cdots, m$ in kNNDOWA is defined as

$$R_i^k = 1 - \frac{\sum_{t=1}^{k} |f_i - n_t^{f_i}|/k}{\max_{i', i'' \in \{1,2,\cdots,m\}} |f_{i'} - f_{i''}|}, \tag{15}$$

where $n_t^{f_i}$ is the value of t-th nearest neighbor ($t = 1, 2, \cdots, k$) of the input tortuosity f_i, and the absolute difference between two tortuosity degrees is used to perform neighbor-searching. In this paper, the k in kNNDOWA is set to the round number of $m/3$ for each image, which accounts one third of nerve fibers in an IVCM image while calculating the reliability and subsequently the weight for the tortuosity of each nerve fiber.

Having obtained the reliability degrees of all tortuosity values as per Eqn. (15), they can then be normalized to generate the weighing vector for nerve fibers. Given the reliability R_i^k of each tortuosity degree f_i, the corresponding kNNDOWA operator $Agg^{\text{knn-dowa}} : \mathbb{R}^m \to \mathbb{R}$ can be defined by

$$Agg^{\text{knn-dowa}}(f_1, f_2, \cdots, f_m) = \frac{\sum_{i=1}^{m} R_i^k f_i}{\sum_{i=1}^{m} R_i^k}. \tag{16}$$

For each fiber-level tortuosity measure F^I, an image-level feature A^I can be defined as the feature value of Img_p, i.e., $a_p^I = Agg^{\text{knn-dowa}}(f_1^I, f_2^I, \cdots, f_m^I), I = 1, 2, \cdots, M$. It is worth noticing that by using the kNNDOWA based feature extraction, the number of extracted feature remains the same as the number of selected tortuosity measures.

II.D.3. Supervised Feature Selection based on Wrapper

Once the tortuosity measurements of multiple nerve fibers are aggregated by OWA and kNNDOWA with respect to all predefined measures as per the above procedure, a supervised feature selection is performed to identify the most informative features utilizing the manual grading as ground truth. This also comes with two more advantages: First, the removal of redundant features improves computational efficiency for the subsequent operations. Second, it simplifies the resultant model, making it easier to interpret by clinicians.

Although a wide range of methods have been established for feature selection, the Wrapper feature selection[38] is employed for its being highly effective at retaining or improving the accuracy of classification. Moreover, combing with a forward stepwise searching scheme, the Wrapper feature selection also retains original feature semantics and enables to explore the feature selected at each iteration, which can be helpful for clinicians to decompose the rationale against domain expertise. In particular, the feature selection algorithm employed here uses classification accuracy to select a subset of features through a process which starts off with an empty feature subset. In each iteration, a most influential feature that obtains the biggest gain in the classification accuracy is added to the feature subset. This is iterated until the accuracy does not increase by adding remaining features. It is worth noticing that other feature selection algorithms[39] can also be employed in this MOWAFS to select an effective subset of the extracted image-level features.

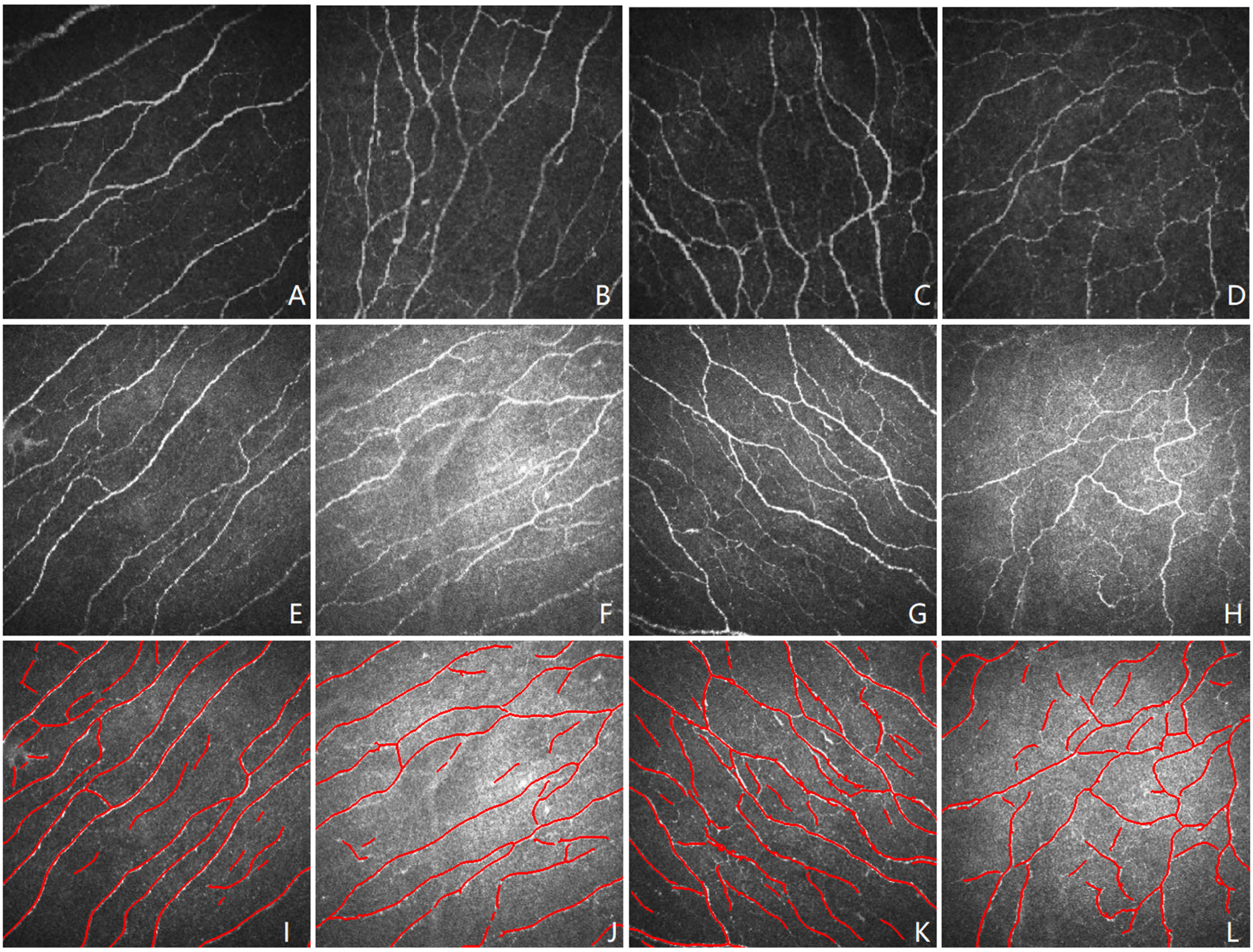

Figure 1: A-H are examples of corneal nerve images with different tortuosity levels of the OWN dataset (columns from left to right: Grades 1 to 4) and I-L result from the automated fiber tracing of E-H, respectively.

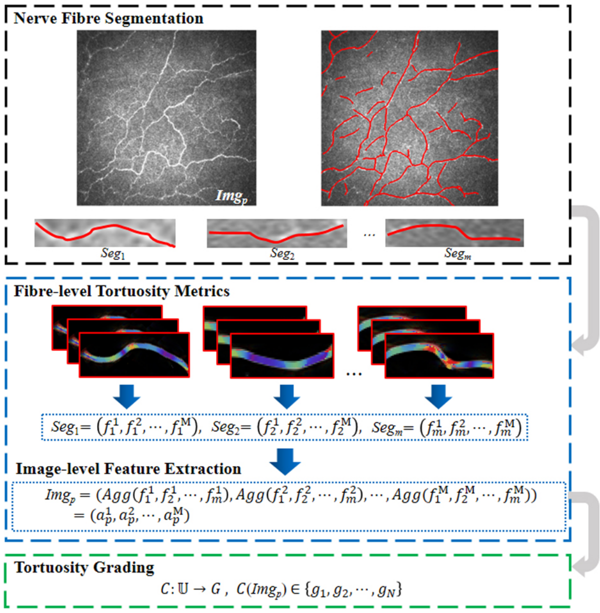

Figure 2: The framework of automated corneal nerve tortuosity grading with an explicit feature extraction

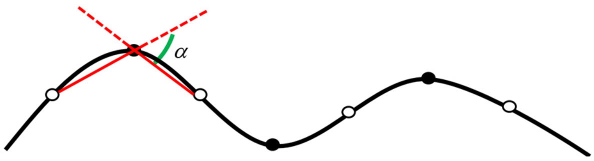

Figure 3: An example for the calculation of straight line segments (solid red lines) and the slope angle α. The dots and circles indicate the points of local maximal and minimal K, respectively.

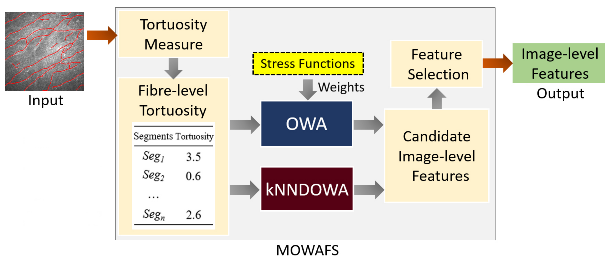

Figure 4: The flowchart of the MOWAFS method, in which the input is IVCM images with segmented nerve fibers and output is a set of image-level features for tortuosity analysis. Two types of aggregations, the OWA and kNNDOWA, are jointly used in this method.

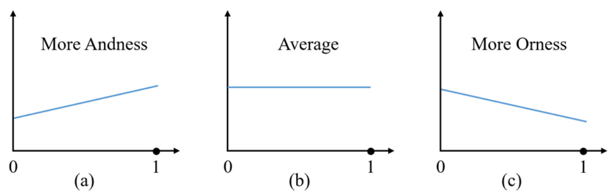

Figure 5: Examples of linear stress functions. The higher values from a stress function on the right/left side of $[0, 1]$ reflect higher weights associated with nerve fibers with lower/higher tortuosity degrees in the OWA aggregation.

III. Results

Once the nerve fibers are segmented in IVCM images by the deep learning-based algorithm CS-NET[28], tortuosity values of each single nerve fiber can be calculated using those measures which are summarized in Table **??** and detailed in Section II.C.. For each of the eight tortuosity measures, the OWA based image-level features defined by Eqn. (11) with twenty-one weighting vectors (see Section II.D.1. for details), and the kNNDOWA based image-level feature defined by Eqn. (16) are extracted.

Three classic classification algorithms are employed to evaluate the performance of feature subsets in Wrapper, i.e., the Support Vector Machines with radical basic function kernels [40], k Nearest Neighbors[41], and C.45 Decision Tree[42] (denoted in the following as SVM, NN, and DT, respectively). The Weka[43] implementations of the three classification models are employed in the experiment. In addition to the proposed MOWAFS, the features extracted by using kNNDOWA independently and the OWA extracted features with Wrapper feature selection (OWA-FS) are also tested. Since the number of extracted image-level features equals the number of employed fiber-level tortuosity measures by using conventional aggregation methods, the maximum size of selected feature subsets in Wrapper is set to M, (i.e., 8) in this experiment for fair comparison.

Extracted image-level features now form input to the classification models for the overall tortuosity grading. The classification accuracy results across the PUB-man, PUB-auto, and OWA-auto data sets are summarized in Table 1, where the row represents the conventional methods: (Average, Maximum, LenA: length-weighted averaging) and the new methods (OWA-FS, kNNDOWA, MOWAFS) used for extracting features. For the PUB data set, the original labels of "High, Medium, and Low" are employed as the ground truth of the classification task, and for the OWN data set, the manual labels of "Grade1-4" are employed as the ground truth. The accuracy is calculated as the ratio of correctly classified images over all images in the data set. Following the standard performance assessment protocol employed in[3], the weighted accuracy (wAcc), the sensitivity (wSe), the specificity (wSp), positive predicted value (wPpv) and negative predictive value (wNpv), which are defined as follows, are also employed to evaluate the performance based on the SVM and the results are reported in Table 2.

$$\text{wAcc} = \sum_{c=1}^{N} r_c \frac{\text{TP}_c + \text{TN}_c}{\text{TP}_c + \text{TN}_c + \text{FP}_c + \text{FN}_c}, \tag{17}$$

$$\text{wSe} = \sum_{c=1}^{N} r_c \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \tag{18}$$

$$\text{wSp} = \sum_{c=1}^{N} r_c \frac{\text{TN}_c}{\text{TN}_c + \text{FP}_c}, \tag{19}$$

$$\text{wPpv} = \sum_{c=1}^{N} r_c \frac{\text{TP}}{\text{TP}_c + \text{FP}_c}, \tag{20}$$

$$\text{wNpv} = \sum_{c=1}^{N} r_c \frac{\text{TN}_c}{\text{TN}_c + \text{FN}_c}, \tag{21}$$

where TP_c, TN_c, FP_c, and FN_c are the true positives, true negatives, false positives, and false negatives, respectively, for the c-th grade ($c = 1, 2, \cdots, N$). N denotes the number of total tortuosity grades, i.e., $N = 3$ for PUB and 4 for OWN. r_c represents the percentage of images whose grade is g_c in a data set.

Each value in Tables 1 and 2 is calculated by averaging 10 random runs of 10-fold cross validation, with the best performance for each performance criterion highlighted in boldface. To validate the statistical significance of the experimental results, the paired t-test is carried out between the LenA and MOWAFS. The differences of all such paired results are statistically significant with p-values are smaller than 0.05.

Table 1: Summary of classification accuracy (%)

	PUB-man			PUB-auto			OWN-auto		
	SVM	NN	DT	SVM	NN	DT	SVM	NN	DT
Average	73.00	75.00	75.00	64.33	64.00	55.00	60.57	64.60	64.57
Maximum	66.34	57.33	67.00	43.66	44.33	58.67	51.69	55.57	55.60
LenA	71.01	73.33	51.67	75.67	76.67	75.00	59.31	61.37	61.97
OWA-FS	88.66	90.73	**90.00**	78.67	83.79	84.32	66.25	69.22	69.02
kNNDOWA	73.00	75.33	65.33	75.00	69.67	66.00	66.12	64.80	62.97
MOWAFS	**89.01**	**91.19**	**90.00**	**82.33**	**85.54**	**86.32**	**69.87**	**69.86**	**70.94**

As clearly reflected in Table 1, performances computed on top of features extracted by the proposed method dominate those calculated using the conventional averaging and

Table 2: SVM classification performance on image-level features (%)

		wAcc	wSe	wSp	wPpv	wNpv
PUB-man	Average	82.00	73.00	86.50	74.69	87.02
	Maximum	77.56	66.33	83.17	70.48	83.45
	LenA	80.67	71.00	85.50	69.43	86.61
	OWA-FS	92.44	88.67	94.33	88.60	94.49
	kNNDOWA	82.00	73.00	86.50	72.75	86.78
	MOWAFS	**92.67**	**89.00**	**94.50**	**88.98**	**94.67**
PUB-auto	Average	76.22	64.33	82.17	63.97	82.67
	Maximum	62.44	43.67	71.83	43.02	72.17
	LenA	83.78	75.67	87.83	75.19	88.61
	OWA-FS	85.78	78.67	89.33	**85.30**	91.93
	kNNDOWA	83.33	75.00	87.50	77.70	88.76
	MOWAFS	**88.22**	**82.33**	**91.17**	84.91	**92.50**
OWN-auto	Average	73.71	62.08	58.13	51.05	80.79
	Maximum	67.79	56.93	42.99	36.48	59.51
	LenA	72.87	61.47	57.61	50.27	78.74
	OWA-FS	77.50	68.72	62.76	**63.57**	87.10
	kNNDOWA	77.41	66.64	65.32	56.12	84.20
	MOWAFS	**79.91**	**72.28**	**67.54**	62.90	**87.24**

length-weighted averaging methods, regardless of the selection of classifier and testing data. Notably, the averaged performance using the proposed MOWAFS has significantly improved upon the Average and LenA methods by 15.44% and 14.34%, respectively.

Table 2 shows that the OWA based method (OWA-FS) outperforms the conventional methods (Average, Maximum, and LenA) over all performance criteria, which indicates that weighting nerve fibers with respect to tortuosity measures instead of nerve length can extract image-level features of higher quality, while the independent use of kNNDOWA based method is not as good as the OWA based ones. However, the joint use of OWA based and kNNDOWA based features, i.e., the MOWAFS, also outperforms the conventional aggregations and the OWA-FS over most of the performance criteria for the tested data sets. This also demonstrates it is beneficial to utilize the reliability of each fiber's tortuosity degree, which leads the generation of data-driven weights in extracting image-level features.

Although the experimental results show high accuracy for the proposed methods, the difference between results of the PUB data set and OWN data set is apparent. One possible explanation is that the OWN data set is labeled by only one clinician, thus making it potentially suffer from high intra-observer variability. Since the MOWAFS is based on a supervised feature selection mechanism, the quality of training data is crucial to the performance of the resultant model. The subjectivity embedded in the protocols of manual corneal nerves tortuosity analysis directly influences the accuracy of labeled data, which forms a big challenge to building an accurate automated system for corneal nerve tortuosity grading. The proposed MOWAFS provides a computational way to characterize the clinicians' perception of how the tortuosity of multiple nerve fibers in an IVCM image is aggregated. By using the proposed method, accuracy improvement on both the PUB and OWN data sets validates that the modeling and optimization of the fiber-level tortuosity aggregation is a significant step in building an effective automated corneal nerve tortuosity grading system based on IVCM images.

IV. Discussion

IV.A. Analysis on Correlations of Image-level Features

To evaluate the correlation between the extracted features and the ground truth provided by ophthalmologists, the Spearman's rank correlation coefficient r_s is employed, which falls in the range of $[-1, +1]$ with the $+/-$ sign indicating the positive/negtive correlation between the two ranks. The resultant values of r_s between image-level features extracted by different approaches and the ground truth are shown in Figure 6, where the Y-axis is r_s and the X-axis represents the attitudinal character of a weighting vector generated by OWA, which indicates the aggregation behavior. The results of OWA-based features are represented as curves marked by circles. The results of kNNDOWA-based and length-weighted features (LenA) are represented as solid and dash straight lines, respectively.

From an overall perspective, regardless of the data sets or the fiber-level tortuosity measures used, significantly different correlations between the ground truth and the extracted image-level features may be obtained depending on the choice of a particular aggregation method. This clearly demonstrates the significance of image-level feature extraction to the overall automated evaluation of corneal nerve tortuosity. To examine more closely, all the highest r_s values are achieved using features generated by OWA and kNNDOWA based methods for the OWN-auto data set. For the PUB-auto data set, the OWA and kNNDOWA based features also result in higher or at least comparable correlation values that are obtainable by the conventional length-weighted features with only one exception at F^5. With two exceptions at F^2 and F^7 out of all eight metrics, similar results are achieved on the PUB-man scenario. Instead of favoring any particular choice of aggregation, the proposed methods are able to generate a variety of features, some of which are clearly more correlated with the ground truth than the conventional length-weighted method.

The OWA operator can generally be characterized by the attitudinal character A-C with the overall aggregation showing more andness if A-C is closer to 0 or more orness if A-C is closer to 1. With regard to Figure 6, the correlations resulted from using extreme A-C values (i.e., close to 1 or 0), are generally not as high as those using mid-range values. Another observation is that the highest values of r_s do not result from using aggregator with A-C$(W) = 0.5$ (i.e., the conventional average operator), which generally lie in the range of

$[0.2, 0.4]$ or $[0.6, 0.8]$. As such this demonstrates OWA operators with appropriately selected attitudinal character can be more effective than the classical aggregation operators such as minimum, maximum, and average for extracting image-level features.

Another interesting point to note is that correlations on the in-house data set are generally lower that those achieved on the public one. The gap is possibly attributed to a significantly higher number of instances embedded in the in-house collection. For manual grading involving subjective bias, it is naturally more difficult for experts to reach higher consensus on a data set with more instances, hence more challenging to model the manual grading process. This in turn calls for data-driven methods to select features of the most indicative, which then forms input to powerful and interpretable classifiers[44] to advance the tortuosity evaluation.

IV.B. Analysis on Selected Features

The proposed method generates a set of aggregation operators whose weighting vectors are predefined or learned from the input values, thereby possibly resulting in image-level features being redundant or even misleading in the classification. The Wrapper based feature selection is employed to select a subset of those features of the most informative to the tortuosity evaluation. Figure 7 demonstrates the iterative generations of the algorithm on three data sets, where the X-axis indicates the number of iterations and the Y-axis indicates the classification accuracy. Each data point is labeled with the choice of fiber-level tortuosity measure and the underlying attitudinal character with respect to the OWA based image-level feature. It is worth noticing that, as more instances and higher complexity are contained in the OWN-auto than those in the PUB-auto and OWN-auto data sets, both the SVM and DT classifiers select more than eight features before the accuracy stop increasing in the Wrapper.

By using a greedy searching scheme initialized with an empty feature subset, it is not surprising that the accuracy increases with the increment of features selected for inclusion. While iteratively adding features with the Wrapper algorithm, the choice of classifier may affect the evaluation of feature subsets. Nevertheless, all the three classifiers select the features generated by kNNDOWA on the PUB-auto and OWA-auto data sets in the first iteration, which indicates that the weighting vector learned from data can be more informative than those pre-defined ones in this experiment. The main reason is that the automated segmentation of nerve fibers is not as accurate as the manual tracing and the resultant tortuosity values of nerve fibers in an image may contain noises. By using the kNNDOWA with k set to a high number (e.g. in this experiment, k is one third of the number of segmented nerve fibers in an image), the noises in the tortuosity values are less weighted. For different classification algorithms, the selected features in the subsequent iterations do not remain as similar as the initial iteration. As the accuracy is evaluated on the feature subset, individual features with high r_s coefficient does not necessarily indicate a better choice for the feature subset as a whole. Instead, individual features with small correlations (e.g., the $F^1(0.00)$ in the PUB-man data set) may be selected, as their inclusion may contribute more to the increase of classification accuracy for the underlying feature subset.

It can be seen from Figure 7 that all three tested classification algorithms tend to select fewer features on the PUB-man data set. The faster convergence and higher resultant accuracy of the Wapper algorithm on the PUB-man data set reveal the potential flaw of conventional tortuosity grading pipeline whereby the quality of tortuosity evaluation is dependent on the quality of segmentation of nerve fibers.

What is reflected in the experiment is that despite the tortuosity degrees of individual nerve fibers are known, the aggregation over all nerve fibers is crucial to the final performance of the automated tortuosity grading pipeline. The experimental results demonstrate the effectiveness of the proposed MOWAFS to perform image-level feature extraction based on both the pre-defined and data-driven weighting vectors. This also suggests that the simultaneous use of multiple and diverse aggregation operators could lead to more stable and robust results compared with those using individual feature extraction method.

IV.C. Limitations of MOWAFS

Although the MOWAFS can substantially increase the accuracy using geometric measures to predict subjective tortuosity grading, the limitations of the proposed method are also worth discussing. First of all, the limited availability of corneal nerve images, particularly the lack of universally accepted tortuosity grading labels, restricts the experimental validation of any automated method including the MOWAFS. It also limits the development of machine learning models which requires large training examples, such as the deep learning models.

To our best knowledge, the MOWAFS is the first one which focuses on the formal computational modeling of clinicians' perception in fiber-level tortuosity aggregation amongst the state-of-the-art automated pipelines of corneal nerve tortuosity analysis. However, it can be expected that the joint use of MOWAFS with other techniques such as the ensemble of multiple scaled images, multiple tortuosity measures, and multiple segmentation algorithms may produce a more robust automated system for corneal nerve analysis than the individual use of MOWAFS.

From the perspective of machine learning models, the hypothesis space of MOWAFS can be further extended. The proposed method assumes pre-defined stress functions or aggregation weights in OWA. A complete data-driven modeling of clinicians' perception in fiber-level tortuosity aggregation should also include the learning of stress functions from labeled data. The implementation of such a model requires experts from both machine learning and ophthalmology to takes efforts to investigate the types and parameters of learnable stress functions, and also develop proper learning algorithms for optimizing their values.

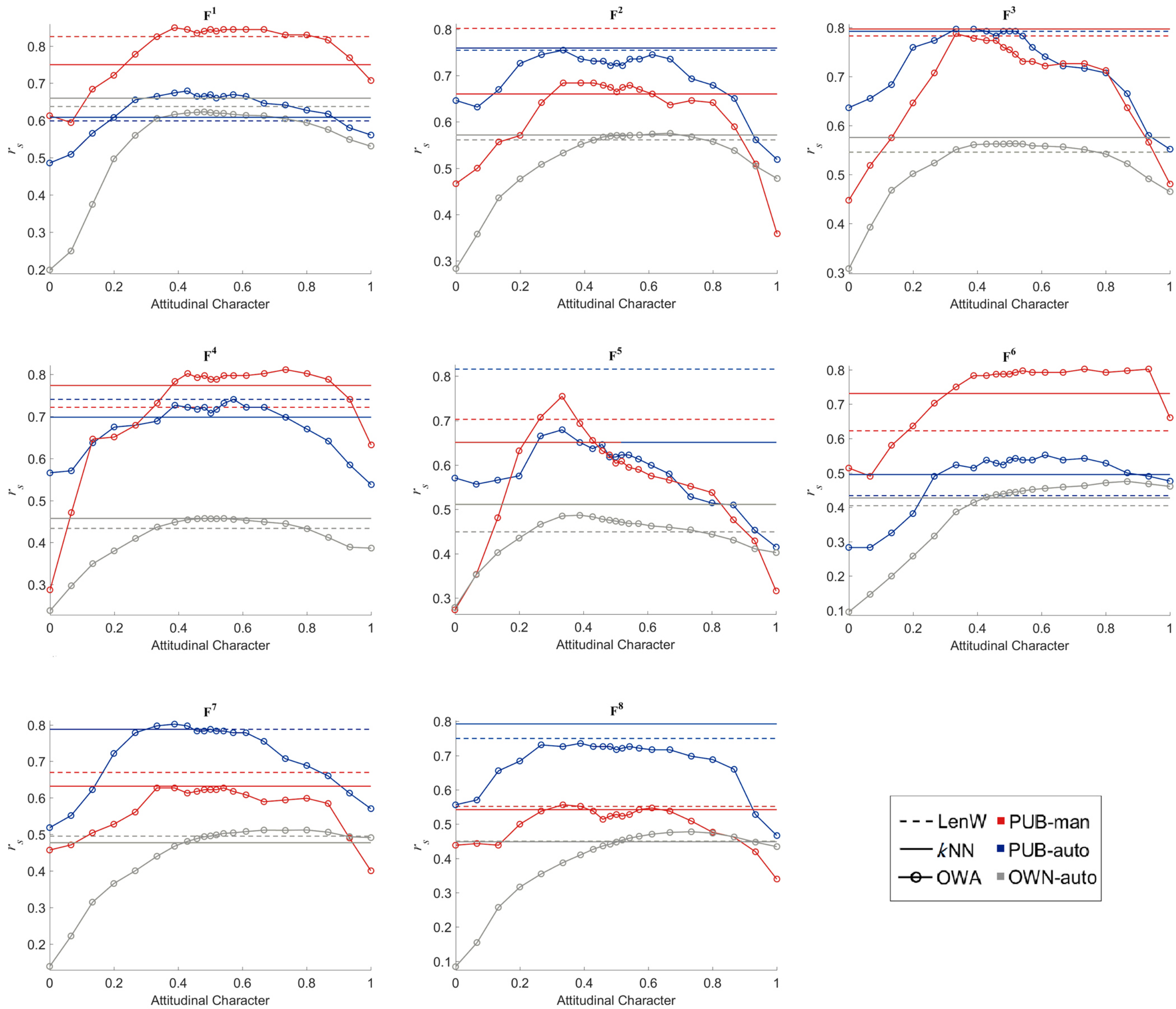

Figure 6: r_s coefficients between ground truth and extracted image-level features based on each tortuosity measure. The circles or solid lines above the dot line indicate the r_s values of OWA or kNNDOWA based features are higher than those based on the length-weighted averaging.

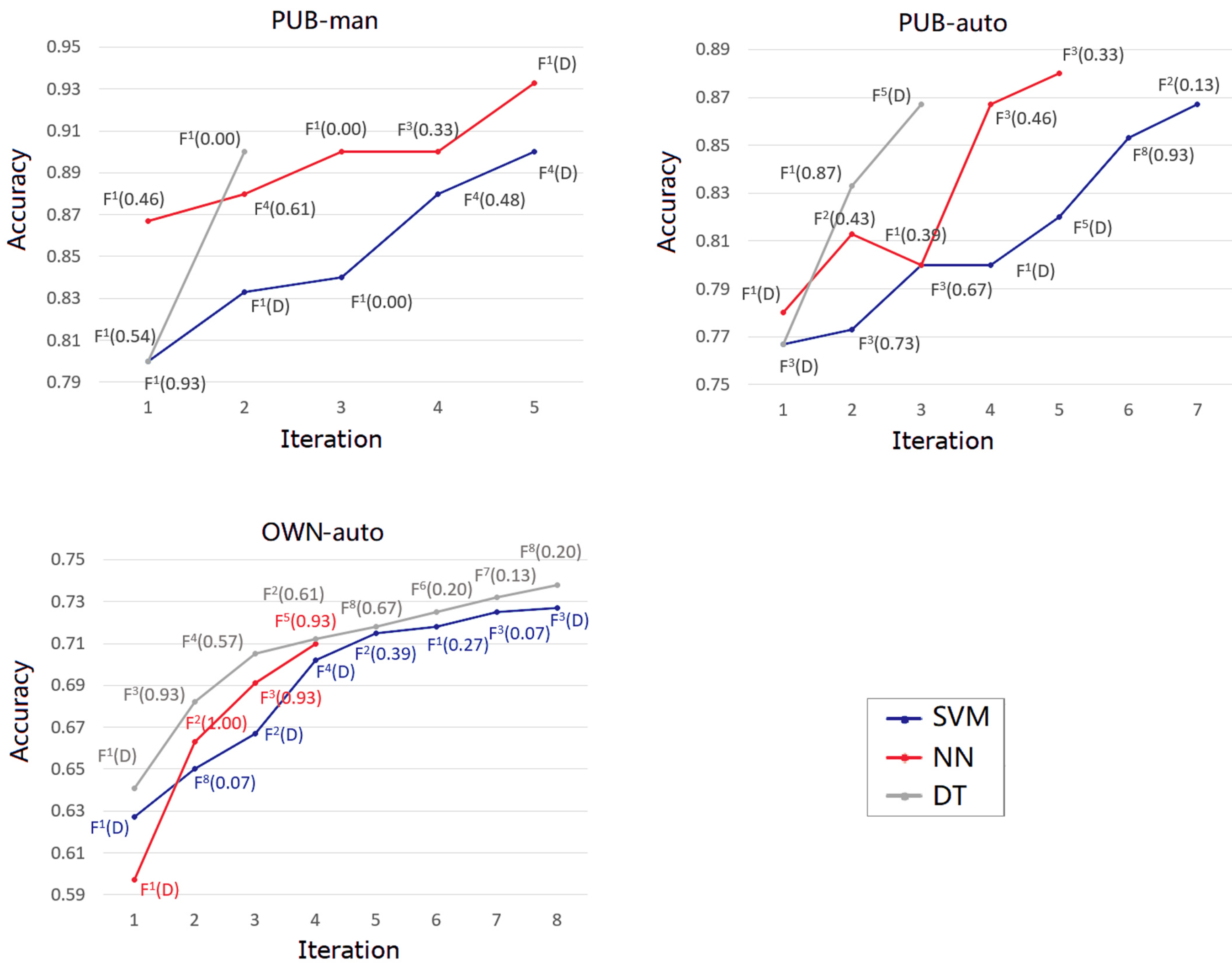

Figure 7: Results of the wrapper-based feature selection. The labels of data points indicate fiber-level tortuosity measure and the underlying attitudinal character value (features generated by kNNDOWA are indicated as 'D'). For example, $F^1(0.00)$ indicates the feature generated from the tortuosity measure F^1 with an OWA operator whose attitudinal character value is zero.

V. Conclusions

Owing to the significance of the corneal nerve in support of the examination and diagnosis for a number of diseases, this paper presents a transparent framework with novel image-level feature extraction for the tortuosity grading of corneal nerve fibers, whereby an image-level feature extraction approach is proposed based on two types of aggregation methods and feature selection. Supported with statistical tests, experimental studies on two real-world data sets demonstrate the effectiveness of the proposed method, in comparison with the conventional length-weighted averaging approach.

Whist promising, this research also opens up an avenue for significant further investigation of applying OWA and alternative fuzzy methods[45,46] for interpretable medical image processing. For instance, it would be potentially more effective to develop a method which supports the aggregation of fiber-level tortuosity with adaptive stress functions or weighting vectors in a supervised manner. With the present work focusing on the reliability of each corneal nerve fiber, it would be interesting to alternatively investigate the reliability of tortuosity measures and regions of IVCM images for tortuosity grading. Finally, the proposed feature extraction methods could be naturally extended to cope with a broader range of medical imaging tasks[47].

VI. Acknowledgment

This work was supported by National Natural Science Foundation of China (61906181), Project funded by China Postdoctoral Science Foundation (2019M652156), Ningbo "2025 S&T Megaprojects" (2019B10033, 2019B10061), and the Key Research and Development Program of Zhejiang Province (2020C030360).

Conflict of Interest

The authors have no conflicts to disclose.

References

[1] F. Scarpa, X. Zheng, Y. Ohashi, and A. Ruggeri. Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. Investigative ophthalmology visual science, 52 9:6404–8, 2011.

[2] Laura Oliveira-Soto and Nathan Efron. Morphology of corneal nerves using confocal microscopy. Cornea, 20 4:374–84, 2001.

[3] R. Annunziata, A. Kheirkhah, S. Aggarwal, P. Hamrah, and E. Trucco. A fully automated tortuosity quantification system with application to corneal nerve fibres in confocal microscopy images. Medical Image Analysis, 32:216–232, 2016.

[4] K. Edwards, N. Pritchard, D. Vagenas, A. W. Russell, R. A. Malik, and N. Efron. Standardizing corneal nerve fibre length for nerve tortuosity increases its association with measures of diabetic neuropathy. Diabetic medicine : a journal of the British Diabetic Association, 31 10:1205–9, 2014.

[5] J. Kim and M. Markoulli. Automatic analysis of corneal nerves imaged using in vivo confocal microscopy. Clinical experimental optometry, 101 2:147–161, 2018.

[6] Panagiotis A. Kallinikos, Michael Berhanu, C. H. O'Donnell, Andrew J. M. Boulton, Nathan Efron, and Rayaz A. Malik. Corneal nerve tortuosity in diabetic patients with neuropathy. Investigative ophthalmology & visual science, 45 2:418–22, 2004.

[7] N. Pritchard, K.Edwards, A. W. Russell, B. A Perkins, R. A. Malik, and N. Efron. Corneal confocal microscopy predicts 4-year incident peripheral neuropathy in type 1 diabetes. Diabetes care, 38 4:671–5, 2015.

[8] Conor Heneghan, John Flynn, Michael O'Keefe, and Mark Cahill. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. Medical image analysis, 6 4:407–29, 2002.

[9] Hamrah Pedram, Cruzat Andrea, Mohammad H Dastjerdi, Zheng Lixin, Bashar M Shahatit, Hasan A Bayhan, Dana Reza, and Pavan Langston Deborah. Corneal sensation and subbasal nerve alterations in patients with herpes simplex keratitis: an in vivo confocal microscopy study. Ophthalmology, 117(10):1930–1936, 2010.

[10] K Kurbanyan, LM Hoesl, WA Schrems, and P Hamrah. Corneal nerve alterations in acute acanthamoeba and fungal keratitis: an in vivo confocal microscopy study. Eye, 26 (1):126, 2012.

[11] R. Annunziata, A. Kheirkhah, S. Aggarwal, B.M. Cavalcanti, P. Hamrah, and E. Trucco. Tortuosity classification of corneal nerves images using a multiple-scale-multiple-window approach. In Proceedings of the Ophthalmic Medical Image Analysis, pages 113–120, 2014.

[12] Philip Mehrgardt, Seid Miad Zandavi, Simon K. Poon, Juno Kim, Maria Markoulli, and Matloob Khushi. U-net segmented adjacent angle detection (usaad) for automatic analysis of corneal nerve structures. Data, 5(2):37, Apr 2020. ISSN 2306-5729. doi: 10.3390/data5020037. URL http://dx.doi.org/10.3390/data5020037.

[13] O. Smedby, N Högman, S. Nilsson, U. Erikson, A. G. Olsson, and G. Walldius. Two-dimensional tortuosity of the superficial femoral artery in early atherosclerosis. Journal of vascular research, 30 4:181–91, 1993.

[14] W. E. Hart, M. H. Goldbaum, P. Kube, and M. Nelson. Measurement and classification of retinal vascular tortuosity. International journal of medical informatics, 53 2-3:239–52, 1999.

[15] E. Bribiesca. A measure of tortuosity based on chain coding. Pattern Recognition, 46 (3):716–724, 2013.

[16] Montserrat Alvarado-Gonzalez, Wendy Aguilar, Edgar Garduño, Carlos Velarde, Ernesto Bribiesca, and Verónica Medina-Bañuelos. Mirror symmetry detection in curves represented by means of the slope chain code. Pattern Recognition, 87:67–79, 2019.

[17] E. Grisan, M. Foracchia, and A. Ruggeri. A novel method for the automatic grading of retinal vessel tortuosity. IEEE Transactions on Medical Imaging, 27:310–319, 2008.

[18] E. Bullitt, G. Gerig, S. M. Pizer, W. Lin, and S. R. Aylward. Measuring tortuosity of the intracerebral vasculature from mra images. IEEE Transactions on Medical Imaging, 22:1163–1171, 2003.

[19] D. Bracher. Changes in peripapillary tortuosity of the central retinal arteries in newborns. Graefe's Archive for Clinical and Experimental Ophthalmology, 218:211–217, 1982.

[20] K. G. Goh, W. Hsu, M. L. Lee, and H. Wang. Adris: An automatic diabetic retinal image screening system. Studies in Fuzziness & Soft Computing, 60:181–210, 2001.

[21] Y. Zhao, J. Zhang, E. Pereira, Y. Zheng, P. Su, J. Xie, Y. Zhao, Y. Shi, H. Qi, J. Liu, and Y. Liu. Automated tortuosity analysis of nerve fibers in corneal confocal microscopy. IEEE Transactions on Medical Imaging, pages 1–1, 2020.

[22] F. Scarpa and A. Ruggeri. Development of clinically based corneal nerves tortuosity indexes. In Fetal, Infant and Ophthalmic Medical Image Analysis - International Workshop, FIFI 2017, and 4th International Workshop, OMIA 2017, Held in Conjunction with MICCAI 2017, Proceedings, pages 219–226, 2017.

[23] Roberto Annunziata, Ahmad Kheirkhah, Shruti Aggarwal, Bernardo M Cavalcanti, Pedram Hamrah, and Emanuele Trucco. Two-dimensional plane for multi-scale quantification of corneal subbasal nerve tortuosity. Investigative ophthalmology & visual science, 57(3):1132–1139, 2016.

[24] Lucía Ramos, Jorge Novo, José Rouco, Stephanie Romeo, María D Álvarez, and Marcos Ortega. Retinal vascular tortuosity assessment: inter-intra expert analysis and correlation with computational measurements. BMC medical research methodology, 18(1): 1–11, 2018.

[25] Neil Lagali, Enea Poletti, Dipika V Patel, Charles NJ McGhee, Pedram Hamrah, Ahmad Kheirkhah, Mitra Tavakoli, Ioannis N Petropoulos, Rayaz A Malik, Tor Paaske Utheim, et al. Focused tortuosity definitions based on expert clinical assessment of corneal subbasal nerves. Investigative ophthalmology & visual science, 56(9):5102–5109, 2015.

[26] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. Medical image analysis, 42:60–88, 2017.

[27] Scarpa Fabio, Zheng Xiaodong, Ohashi Yuichi, and Ruggeri Alfredo. Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. Investigative Ophthalmology & Visual Science, 52(9):6404, 2011.

[28] Lei Mou, Yitian Zhao, Li Chen, Jun Cheng, Zaiwang Gu, Huaying Hao, Hong Qi, Yalin Zheng, Alejandro Frangi, and Jiang Liu. Cs-net: Channel and spatial attention network for curvilinear structure segmentation. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Proceedings, pages 721–730, 2019. ISBN 978-3-030-32239-7.

[29] K Narasimhan and K Vijayarekha. Automatic grading of images based on retinal vessel tortuosity analysis. Indian Journal of Science and Technology, 8(29):1, 2015.

[30] W. Lotmar, A. Freiburghaus, and D. Bracher. Measurement of vessel tortuosity on fundus photographs. Albrecht Von Graefes Archiv Fr Klinische Und Experimentelle Ophthalmologie, 211(1):49–57, 1979.

[31] M. Patasius, V. Marozas, A. Lukosevicius, and D. Jegelevicius. Evaluation of tortuosity of eye blood vessels using the integral of square of derivative of curvature. In Proceedings Eur. Med. Biol. Eng. Conf. (EMBEC05), page 1589, 2005.

[32] Pan Su, Changjing Shang, Tianhua Chen, and Qiang Shen. Exploiting data reliability and fuzzy clustering for journal ranking. IEEE Transactions on Fuzzy Systems, 25(5): 1306–1319, 2017.

[33] Pan Su, Qiang Shen, Tianhua Chen, and Changjing Shang. Ordered weighted aggregation of fuzzy similarity relations and its application to detecting water treatment plant malfunction. Engineering Applications of Artificial Intelligence, 66:17–29, 2017.

[34] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. Systems, Man and Cybernetics, IEEE Transactions on, 18(1):183–190, 1988.

[35] R.R. Yager. Using stress functions to obtain owa operators. Fuzzy Systems, IEEE Transactions on, 15(6):1122–1129, 2007.

36 Zeshui Xu. Dependent owa operators. In Modeling Decisions for Artificial Intelligence, pages 172–178. Springer, 2006.

37 T. Boongoen and Q. Shen. Nearest-neighbor guided evaluation of data reliability and its applications. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 40(6):1622–1633, 2010.

38 Ron Kohavi and George H. John. Wrappers for feature subset selection. Artificial Intelligence, 97(1-2):273–324, 1997.

39 Qiang Shen, Ren Diao, and Pan Su. Feature selection ensemble. In Andrei Voronkov, editor, Alan Turing Centenary, pages 289–306, 2012.

40 Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol., 2(3), May 2011. ISSN 2157-6904. doi: 10.1145/1961189.1961199. URL https://doi.org/10.1145/1961189.1961199.

41 David W. Aha, Dennis F. Kibler, and Marc K. Albert. Instance-based learning algorithms. Machine Learning, 6(1):37–66, 1991.

42 J Ross Quinlan. C4.5 : programs for machine learning. Morgan Kaufmann Publishers Inc., 1992.

43 Mark A. Hall Eibe Frank and Ian H. Witten. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Fourth Edition. Morgan Kaufmann, 2016.

44 Tianhua Chen, Changjing Shang, Pan Su, and Qiang Shen. Induction of accurate and interpretable fuzzy rules from preliminary crisp representation. Knowledge-Based Systems, 146:152–166, 2018. doi: 10.1016/j.knosys.2018.02.003. URL https://doi.org/10.1016/j.knosys.2018.02.003.

45 Tianhua Chen, Qiang Shen, Pan Su, and Changjing Shang. Fuzzy rule weight modification with particle swarm optimisation. Soft Computing, 20(8):2923–2937, 2016. doi: 10.1007/s00500-015-1922-z. URL https://doi.org/10.1007/s00500-015-1922-z.

[46] T. Chen, C. Shang, J. Yang, F. Li, and Q. Shen. A new approach for transformation-based fuzzy rule interpolation. IEEE Transactions on Fuzzy Systems, pages 1–1, 2019. doi: 10.1109/TFUZZ.2019.2949767.

[47] Y. Zhao, J. Xie, H. Zhang, Y. Zheng, Y. Zhao, H. Qi, Y. Zhao, P. Su, J. Liu, and Y. Liu. Retinal vascular network topology reconstruction and artery/vein classification via dominant set clustering. IEEE Transactions on Medical Imaging, 39(2):341–356, 2020.