Original Article

Published in Medical and Biological Engineering and Computing. Full reference: Franz T, Reddy BD, Human P, Zilla P. A mathematical method for constraint-based cluster analysis towards optimized constrictive diameter smoothing of saphenous vein grafts. Med Biol Eng Comput, 2010, 48(6), 519-29

A Mathematical Method for Constraint-based Cluster Analysis towards Optimized Constrictive Diameter Smoothing of Saphenous Vein Grafts

Thomas Franz[a,*], B. Daya Reddy[b], Paul Human[a], Peter Zilla[a]

[a]Chris Barnard Division of Cardiothoracic Surgery,
[b]Department of Mathematics and Applied Mathematics,
University of Cape Town, Cape Town, South Africa

*Send correspondence to:
Thomas Franz, PhD
Cardiovascular Research Unit
Faculty of Health Sciences
University of Cape Town
Observatory 7935, South Africa
Tel: +27 21 406 6418, Fax: +27 21 448 5935
E-mail: thomas.franz@uct.ac.za

Abstract

This study was concerned with the cluster analysis of saphenous vein graft data to determine a minimum number of diameters, and their values, for the constrictive smoothing of diameter irregularities of a cohort of veins. Mathematical algorithms were developed for data selection, transformation and clustering. Constrictive diameter values were identified with interactive pattern evaluation and subsequently facilitated in decision-tree algorithms for the data clustering. The novel method proved feasible for the analysis of data of 118 veins grafts, identifying the minimum of two diameter classes. The results were compared to outcome of a statistical recursive partitioning analysis of the data set. The method can easily be implemented in computer-based intelligent systems for the analysis of larger data sets using the diameter classes identified as initial cluster structure.

Keywords: saphenous vein; diameter irregularity; constrictive smoothing, mathematical model; pattern recognition;

Nomenclature

Symbol	**Unit**	**Description**
$C_{R,i}$	-	Constriction degree required for complete smoothing of vein i, by reducing the vein's maximum diameter, $D_{max,i}$, to the minimum diameter, $D_{min,i}$, where i = 1 to n
$C_{A,i}$	-	Constriction degree applied to vein i, to reduce the maximum diameter, $D_{max,i}$, to the constricted diameter $d^{(k)}$, where i = 1 to n and k = 1 to l
$C_{A,i,j}$	-	Constriction degree applied to vein i, to reduce the diameter, $D_{i,j}$, at each measurement position, $x_{i,j}$, along the length of the vein to the constriction diameter, $d_P^{(s)}$, proposed with the recursive partitioning, where i = 1 to n and s = 1 to l_P
C_A^{max}	-	Maximum permissible constriction degree applied to a vein to reduce the maximum diameter, $D_{max,i}$, to the constricted diameter $d^{(k)}$, where i = 1 to n and k = 1 to l
C_A	-	Mean of applied constriction degree $C_{A,i}$ for all veins i, with i = 1 to n, across all constriction diameters $d^{(k)}$, with k = 1 to l
$C_A^{(k)}$	-	Mean of applied constriction degree $C_{A,i}$ for all veins z, with z = 1 to n_k, for one constriction diameters $d^{(k)}$, with k = 1 to l
$D_{i,j}$	mm	Outer diameter of vein i measured at luminal pressure associated with post-harvest leak test syringe inflation at measurement position $x_{i,j}$, where i = 1 to n and j = 1 to m_i
$D_{C,i}$	mm	Constriction diameter for vein i, i.e. outer diameter of vein i after constriction, where i = 1 to n
$D_{max,i}$	mm	Maximum value of the outer diameter , $D_{i,j}$, of vein i at luminal pressure associated with post-harvest leak test syringe inflation ,where i = 1 to n
$D_{min,i}$	mm	Minimum value of the outer diameter, $D_{i,j}$, of vein i at luminal pressure associated with post-harvest leak test syringe inflation, where i = 1 to n
$D_{max,z}^{(k)}$	mm	Maximum outer diameter of vein z of sub set k, where z = 1 to n_k and k = 1 to l
$D_{min,z}^{(k)}$	mm	Minimum outer diameter of vein z of sub set k, where z = 1 to n_k and k = 1 to l
d	mm	Constrictive smoothing diameter for all veins i where i = 1 to n
$d^{(k)}$	mm	Constrictive smoothing diameter for a sub set k of veins z where k = 1 to l

		and z = 1 to n_k
$d_{\mathrm{P}}^{(s)}$	mm	Partitions of $D_{\min,i}$, determined with the recursive partitioning method, that represent constrictive smoothing diameters with i = 1 to n and s = 1 to l_{P}
i	-	Identifier of a vein in full set (i = 1 to n)
j	-	Identifier of the measurement position along a vein (j = 1 to m_i)
k	-	Identifier of a vein set (k = 1 to l)
L_i	mm	Harvested length of vein i, where i = 1 to n
l	-	Number of vein sets: l = 1 for the entire set of veins i with i = 1 to n, and l > 1 if it is required to divide the set of n veins into sub sets k of veins
l_{P}	-	Number of partitions $d_{\mathrm{P}}^{(s)}$ determined with the recursive partitioning method
m_i	-	Number of measurement points for the outer diameter $D_{i,j}$ along vein i with the length L_i, where m_i = int (L_i / 20)
n	-	Number of veins in analysis
n_{e}	-	Number of veins excluded from analysis
n_k	-	Number of veins in sub set k where k = 1 to l
$\hat{n}_k$	%	Normalized number of veins in sub set k: $\hat{n}_k = n_k / \sum_{k=1}^{l} n_k$, where k = 1 to l
s	-	Identifier of partitions of $D_{\min,i}$, $d_{\mathrm{P}}^{(s)}$, determined with the recursive partitioning method where s = 1 to l_{P}
$x_{i,j}$	-	Position of $D_{i,j}$ measurement along vein i, where i = 1 to n and j = 1 to m_i
z	-	Identifier of a vein in sub set k where z = 1 to n_k and k = 1 to l

1. Introduction

With the steadily increasing potential to collect and store data, in particular large data sets, the analysis and exploration of such data has received wide attention and stimulated diverse research. Knowledge discovery from data (KDD) is a multidisciplinary field which focuses on the identification of meaningful patterns in data sets. Data mining constitutes a key element of the knowledge discovery process by employing intelligent methods to extract data patterns. Additional steps such as data cleaning, data selection and knowledge presentation are usually necessary for a successful data analysis. Amongst the concepts of data mining are classification and cluster analysis. Classification aims at finding a model or function that describes the data and allows distinguishing data classes. Cluster analysis refers to the grouping of data into clusters, or classes, such that data objects in the same cluster are similar to one another but dissimilar to data objects of other clusters. A main difference between classification and cluster analysis is that the former requires knowledge of class labels a priori whereas the latter does not [1].

Knowledge discovery from large data sets and high-dimensional data often employs intelligent systems such as neural networks [2], fuzzy approaches [3, 4] and genetic algorithms and programming [5, 6] which can form hybrid methods with more analytical knowledge discovery techniques. KDD methods have generally employed intelligent system when dealing with large-scale data [6-8] whereas intelligent systems may not be required for the analysis small and intermediate data sets.

In this paper, we present a simple mathematical method for the clustering of small- to medium-sized sets of single-parameter data with the objective of identifying the minimum number of classes and their labels to satisfy a set of constraints. The feasibility of the proposed method was demonstrated using a data set comprising diameters measured from 118 saphenous veins obtained from 100 patients with the aim of finding the smallest number of diameters, and their values, that allow for the constrictive elimination of diameter irregularities in the maximum number of the vein grafts. This study was motivated by positive effects associated with the constrictive reinforcement of saphenous vein grafts, namely reduction of intimal hyperplasia and endothelial preservation [9, 10]. Despite the developments in vascular tissue engineering, vein grafts are still preferred to prosthetic small diameter vascular grafts [11]. The concept of constrictive elimination of diameter irregularities is the external, i.e. ablumenal, application of a constrictive tubular mesh with constant diameter to the entire length of a vein. Any point along the vein length with a diameter larger than the constriction diameter, i.e. the diameter of the mesh, is reduced to the constriction diameter whereas any point of the vein smaller than the constriction diameter remains smaller. It was, however, not desired to allow for any points of a vein smaller than the constriction diameter, in order to obtain complete smoothing of the vein lumen. Providing a regular luminal diameter along the length of the vein prevents local disturbances of the blood flow and changes in wall shear stress, main factors for the development of focal intimal hyperplasia [12, 13]. Focal intimal hyperplasia typically leads to vein graft stenosis and occlusion constituting serious clinical problems [14]. Flow disturbances along the graft may not only cause focal intimal hyperplasia but contribute to adverse flow fields in the distal anastomoses and consequently favour the progression of intimal hyperplasia in this critical graft region [15]. A further benefit of constriction is the prevention of overstretching the vein graft when exposed to the arterial circulation. By causing non-physiological mechanical stresses in the vein wall, overstretching may lead to intimal hyperplasia and graft failure [16]. Minimizing the number of constriction diameters, i.e. sizes of mesh devices, that are required for a cohort of veins is desirable in the context of translation of this concept and technology into clinical practice. A limited number of mesh devices will be more readily accepted by the medical device industry as well as the clinicians.

2. Methods

2.1 Data Acquisition

The outer diameter of 118 saphenous veins [harvested length: 284 ± 95 mm (mean ± standard deviation), range 100 – 520 mm] of 100 patients undergoing aorto-coronary bypass surgery was recorded every 20 mm along the harvested length using a vernier caliper during post-harvest in situ syringe inflation. The diameter measurements were taken in close succession during the same inflation procedure to ensure constant pressure throughout the measurement process. The inflation pressure was 245 ± 90.4 mmHg. The measurements were assumed representative for systolic arterial pressure of 120 mmHg as saphenous veins, although highly compliant at venous pressure of 30 mmHg, are nearly non-compliant at a pressure of 100 mmHg [17]. The measurement procedure was approved by the institutional review boards of the University of Cape Town and informed consent was obtained from all patients.

2.2 Mathematical Constraint-based Clustering

2.2.1 Data Selection and Transformation

Each vein i, where i = 1 to n, has an outer diameter $D_{i,j}$ at the measurement points $x_{i,j}$, where i = 1 to n and j = 1 to m_i, equally spaced along the entire length, L_i, of the vein. It is desired to find a minimum number k of constrictive smoothing diameters, d_k, where k = 1 to l, to ensure complete diameter smoothing of these veins. The constriction degree required to reduce the maximum diameter of vein i, $D_{\max,i}$, to the minimum diameter, $D_{\min,i}$, is given in percent by:

$$C_{\mathrm{R},i} = \left(\frac{D_{\max,i} - D_{\min,i}}{D_{\max,i}} \right). \tag{1}$$

The constriction degree applied to vein i at the position along the vein with the maximum outer diameter, is defined as

$$C_{\mathrm{A},i} = \left(\frac{D_{\max,i} - D_{\mathrm{C},i}}{D_{\max,i}} \right), \tag{2}$$

where $D_{\mathrm{C},i}$ is the outer diameter of vein i after constriction, i.e. constriction diameter.

<u>Constraint 1</u>: The applied constriction degree, $C_{\mathrm{A},i}$, must be greater than or equal to the to the required constriction degree, $C_{\mathrm{R},i}$, but may not exceed $C_A^{\max}$ = 0.5:

$$C_{\mathrm{A},i} \in \left[C_{\mathrm{R},i}, C_{\mathrm{A}}^{\max} \right]. \tag{3}$$

The maximum constriction degree of 0.5 was found to be feasible and to provide the desired biological response, i.e. mitigation of intimal hyperplasia, in a prior study [9] for femoral vein grafts in the femoral artery position of a non-human primate model. It follows from Eq. (3), with Eqs. (1, 2), that the constriction diameter, $D_{\mathrm{C},i}$, of vein i must be larger than or equal to the maximally constricted largest outer diameter, $(1 - C_{\mathrm{A}}^{\max}) \cdot D_{\max,i}$, and smaller than or equal to the minimum outer diameter, $D_{\min,i}$, of vein i:

$$D_{\mathrm{C},i} \in \left[\left(1 - C_{\mathrm{A}}^{\max}\right) \cdot D_{\max,i}, D_{\min,i} \right] \qquad \text{with } i = 1 \text{ to } n\,. \tag{4}$$

Figure 1 illustrates the outer diameter, $D_{i,j}$, which also represents the maximum individual constrictive smoothing diameter, and the minimum individual constrictive smoothing diameter, $(1 - C_A^{max})D_{i,j}$, at each measurement point $x_{i,j}$ along the length of a non-specific vein i. Veins with $D_{\max,i} > 2\,D_{\min,i}$ were excluded from the analysis since complete smoothing was not possible with the maximum constriction of $C_A^{\max}$ = 0.5. This ensured that the maximally constricted largest diameter was smaller than or equal to the smallest diameter of vein i:

$$\left(1 - C_{\mathrm{A}}^{\max}\right) \cdot D_{\max,i} \leq D_{\min,i} \qquad \text{for } i = 1 \text{ to } n\,. \tag{5}$$

Constraint 2: Any vein i with $D_{\min,i}$ < 3.0 mm was excluded from the analysis following clinical reports of higher occlusion rates in vein grafts that show narrowings of less than 3 mm [18].

The smoothing constriction diameter, d, that accommodates all of n veins, is defined as:

$$d \in \left[\max_i \left(1 - C_{\mathrm{A}}^{\max}\right) \cdot D_{\max,i}\,,\ \min_i D_{\min,i} \right], \qquad \text{with } i = 1 \text{ to } n, \tag{6}$$

with the condition that the largest minimum individual smoothing diameter, $\max_i [(1 - C_{\mathrm{A}}^{\max}) \cdot D_{\max,i}]$, must be smaller than or equal to the smallest minimum diameter, $\min_i D_{\min,i}$:

$$\max_i \left(1 - C_{\mathrm{A}}^{\max}\right) \cdot D_{\max,i} \leq \min_i D_{\min,i}\,, \qquad \text{with } i = 1 \text{ to } n\,. \tag{7}$$

Figure 2 illustrates the range of the individual constrictive smoothing diameter, $D_{\mathrm{C},i}$, for a number of veins, ranked according to minimum and maximum constriction diameter, and the resulting constriction diameter, d, accommodating these veins.

For that case that $\max_i [(1 - C_{\mathrm{A}}^{\max}) D_{\max,i}] > \min_i D_{\min,i}$, a single constrictive smoothing diameter, d, does not exist for the all the n veins, see Eq. (7). The set of veins i with individual constriction diameters $D_{\mathrm{C},i}$, with i = 1 to n, needs to be divided in two or more sub sets k with veins z in each sub set, where k = 1 to l and z = 1 to n_k, such that Eq. (7) is satisfied for each sub set. The constrictive smoothing diameter, $d^{(k)}$, for a sub set of veins is then derived from the largest minimum individual constrictive smoothing diameter and the smallest minimum diameter of the vein sub set k:

$$d^{(k)} \in \left\{ \max_z \left(1 - C_{\mathrm{A}}^{\max}\right) \cdot D_{\max,z}^{(k)}\ ,\ \min_z D_{\min,z}^{(k)} \right\}, \qquad \text{with } k = 1 \text{ to } l \text{ and } z = 1 \text{ to } n_k\,. \tag{8}$$

2.2.2 Data Clustering

Based on the set of individual constrictive smoothing diameters, $D_{\mathrm{C},i}$, with i = 1 to 118, determined in the prior constraint-based data analysis, a number of constrictive smoothing diameters, $d^{(k)}$, with k = 1 to l, with $d^{(1)} < d^{(k)} < d^{(l)}$, were proposed with the aim to accommodate a maximum number of the n veins. The smallest constriction diameter was determined by constraint 2, i.e. $d^{(1)}$ = 3.0 mm, whereas the largest constriction diameter, $d^{(l)}$, was selected such that it fell between the minimum and the maximum of constriction diameter, $D_{C,i}$, of the vein with the largest maximum constriction diameter, $\max_i D_{\min,i}$.

Intermediate constriction diameters $d^{(k)}$, with k = 2 to l-1, were evenly distributed between the minimum and maximum value. The number of constriction diameters, k, was chosen larger than the expected minimum to accommodate the n vein and within the range which was assumed to be acceptable by medical companies and clinicians for the clinical application.

Two alternative decision-tree based clustering algorithms (CA), illustrated in Fig. 3, were utilised to assign a vein i to either the smallest (CA1) or largest (CA2) constriction diameter $d^{(k)}$ possible with the individual constriction diameter $D_{C,i}$ of that vein. It was assumed that all veins i undergoing classification satisfied the conditions formulated in Eqs. (5) and (7),

The CAs utilized the smallest individual constriction diameter (SICD), $(1 - C_A^{\max})D_{\max,i}$, and the largest individual constriction diameter (LICD), $D_{\min,i}$, of a vein i. For small-diameter prioritization (CA1), the search for a suitable constriction diameter $d^{(k)}$ started with the smallest constriction diameter $d^{(1)}$ and proceeded towards increasing constriction diameters as follows.

1. In the case that the LICD of a vein i was smaller than $d^{(1)}$, none of the constriction diameters $d^{(k)}$ with k = 1 to l was suitable for this vein. The vein could not be smoothed and the clustering ceased.
2. If the LICD that was not smaller than $d^{(1)}$, it was checked whether the SICD was smaller than or equal to $d^{(1)}$, in which case constriction diameter $d^{(1)}$ was assigned to the vein and the clustering ceased.
3. If the SICD was larger than $d^{(1)}$ but smaller than or equal to $d^{(2)}$, it was checked whether LICD was larger than or equal to $d^{(2)}$ in which case $d^{(2)}$ was assigned to the vein and the clustering ceased. If LICD was smaller than $d^{(2)}$, none of the $d^{(k)}$ was suitable for the vein, and the clustering ceased.
4. If the clustering was not concluded, the procedure described in step 3 was repeated for k = 2 to l-1, which takes the form in more general terms: If the SICD was larger than $d^{(k-1)}$ but smaller than or equal to $d^{(k)}$, it was checked whether LICD was larger than or equal to $d^{(k)}$ in which case $d^{(k)}$ was assigned to the vein and the clustering ceased. If LICD was smaller than $d^{(k)}$, none of the $d^{(k)}$ was suitable for the vein, and the clustering ceased.
5. If the SICD was larger than $d^{(l-1)}$ but smaller than or equal to $d^{(l)}$, it was checked whether LICD was larger than or equal to $d^{(l)}$ in which case $d^{(l)}$ was assigned to the vein and the

clustering ceased. For the cases that SICD was larger than $d^{(l)}$ and LICD was smaller than $d^{(l)}$, respectively, none of the $d^{(k)}$ was suitable for the vein, and the clustering ceased.

The clustering of a vein *i* with large-diameter prioritization (CA2) commenced with the largest constriction diameter $d^{(l)}$ and proceeded towards decreasing constriction diameters.

1. If the LICD of vein *i* was larger than or equal to the largest constriction diameter $d^{(l)}$, and SICD was larger than $d^{(l)}$, none of the constriction diameters $d^{(k)}$,with k = 1 to l, were suitable for vein *i* and the clustering ceased. If SICD was not larger than $d^{(l)}$, the constriction diameter $d^{(l)}$, was assigned to vein *i*.
2. If the LICD was smaller than the largest constriction diameter $d^{(l)}$, and the SICD was larger than next smaller constriction diameter $d^{(l-1)}$, none of the $d^{(k)}$ was suitable for the vein.
3. If SICD was not larger than $d^{(l-1)}$, it was checked whether the LICD was larger than or equal to $d^{(l-1)}$, in which case $d^{(l-1)}$ was assigned to the vein.
4. If clustering was not concluded, steps 2 and 3 were repeated for $k = l$-2 to 1 in the general form: If the LICD was smaller than $d^{(k+1)}$, and the SICD was larger than $d^{(k)}$, none of the $d^{(k)}$ was suitable for the vein.
5. If SICD was not larger than $d^{(k)}$ and the LICD was larger than or equal to $d^{(k)}$, the constriction diameter case $d^{(k)}$ was assigned to the vein.
6. Finally, if the LICD was not larger than or equal to the smallest constriction diameter $d^{(1)}$, none of the $d^{(k)}$ could be assigned to the vein and the clustering ceased for vein *i*.

The clustering of veins amongst the constriction diameter, $d^{(k)}$,i.e. class, was evaluated using three parameters: 1) Number n_k of veins assigned to each constriction diameter $d^{(k)}$; providing a measure of the distribution of veins amongst, and the utilisation of the constriction diameters, $d^{(k)}$, for different clustering solutions. In particular, n_k = 0 indicated that a constriction diameter, $d^{(k)}$, was not required for the cohort of veins studied. 2) Mean, minimum, and maximum applied constriction degree, $C_A^{(k)}$, for each constriction diameter $d^{(k)}$, providing an indication of the applied constriction degree obtained for the assigned veins per constriction diameter, in particular for assessment of the mean constriction degree obtained with, and the level of dissimilarity between different constriction diameters, $d^{(k)}$, of one clustering solution. 3) Overall applied constriction degree, C_A, across all constriction diameters $d^{(k)}$, with k = 1 to l, presented a comparison between different clustering solutions.

Ranking of values of these criteria was beyond the scope of this study which primarily aimed at developing the clustering method. These criteria's were however assumed to provide beneficial information for further studies.

2.3 Statistical Recursive Partitioning

The data $D_{i,j}$ of the veins i , with i = 1 to n and j = 1 to m_i, described above was additionally examined using a statistical recursive partitioning method, implemented in the statistical software package JMP (version 6.0.3, SAS, Cary, NC), to assess the mathematical method and results. By exhaustively searching all possible partitions and groupings, the method determined those partitions, $d_\text{P}^{(s)}$, of the minimum vein diameter, $D_{\text{min},i}$, with i = 1 to n, for which the difference in the ratio of maximum to minimum vein diameter, $D_{\text{max},i}$ /$D_{\text{min},i}$, between the resulting groups was maximized. The number of partitions, s, where s = 1 to l_P, was increased until further partitions did not substantially add information. The latter was indicated from coefficient of determination, R^2, associated with the number of partitions, s, approaching a plateau, or maximum. The criterion used during visual inspection of the R^2-s graph was that the increase of R^2 for the increase in partition number from s to $s + 1$ was small compared to the increase of R^2 for the increase in the partition number from s - 1 to s, i.e. $R^2(s + 1) - R^2(s) \ll R^2(s) - R^2(s - 1)$. The resulting values for $d_\text{P}^{(s)}$ were subsequently used as constrictive smoothing diameters to cluster the veins in solutions comprising different numbers of partitions observing the constraints established for the mathematical method. The appropriate partition for a vein i was determined by the vein's minimum diameter, $D_{\text{min},i}$, and the applied constriction degree, $C_{\text{A},i}$, (see Eq. 2).

3. Results

3.1 Data Set

The data set obtained from 100 patients (age: 59.7 ± 8.5 years, weight: 80.3 ± 17.2 kg, gender: 64 male, 36 female) comprised 118 veins (data objects) with 1687 measured diameter values, $D_{i,j}$, (data points) i.e. 14.3 ± 4.7 data points per data object. The minimum and maximum outer vein diameter was on average D_{min} = 3.50 ± 0.61 mm and D_{max} = 4.77 ± 0.75 mm with grand minimum and maximum of 2.1 mm and 6.5 mm, respectively. (Average values are given as mean ± SD throughout the paper.)

3.2 Mathematical Data Analysis and Clustering

Of the 118 veins, 117 veins satisfied the condition $D_{\text{max},i} \leq 2D_{\text{min},i}$, with $D_{\text{max},i} < 2D_{\text{min},i}$ for 115 veins and $D_{\text{max},i} = 2D_{\text{min},i}$ for 2 veins. The range of the constriction diameter, $D_{\text{C},i}$, with i = 1 to 117, for the 117 veins is illustrated in Fig. 4. The largest minimum individual constriction diameter was $\max_i (1 - C_{\text{A}}^{\text{max}}) D_{\text{max},i}$ = 3.25 mm and exceeded the smallest maximum individual constriction diameter of $\min_i D_{\text{min},i}$ = 2.10 mm, where i = 1 to 117. Hence a single constriction diameter, d, accommodating all of the 117 veins did not exist.

Four constrictive smoothing diameters were identified, as described in section 2.2.1, for the clustering of the veins: $d^{(1)} = 3.0$ mm, $d^{(2)} = 3.3$ mm, $d^{(3)} = 3.6$ mm and $d^{(4)} = 3.9$ mm. Fourteen veins, with a maximum individual constriction diameter, $D_{\text{min},i}$, smaller than 3.0 mm, were excluded from the data set prior to the cluster analysis.

The clustering of the data of the remaining 103 veins with small-diameter prioritization, using algorithm CA1, assigned n_1 = 95 and n_2 = 8 veins to the constriction diameters $d^{(1)}$ and $d^{(2)}$, respectively, whereas the constriction diameters $d^{(3)}$ and $d^{(4)}$ were not utilized, i.e. $n_3 = n_4 = 0$ (see Table 1). This was a result of algorithm CA1 assigning a vein i to the smallest constriction diameter, $d^{(k)}$, possible for that vein's individual constriction diameter, $D_{\text{C},i}$.

The constriction degree for the constriction diameters $d^{(1)}$ and $d^{(2)}$ was $C_{\text{A}}^{(1)}$ = 0.36 ± 0.09 and $C_{\text{A}}^{(2)}$ = 0.48 ± 0.01, respectively.

The accommodation of all 103 veins with the two smallest of the four constriction diameters when clustering with the data with CA1 indicated that clustering of the data set with large-diameter prioritisation, using CA2, would accommodate all veins with the constriction diameters $d^{(1)}$ and $d^{(2)}$, as well as with more constriction diameters. Algorithm CA2 assigned a vein i to the largest constriction diameter, $d^{(k)}$, possible for that vein's individual constriction diameter, $D_{C,i}$. To facilitate a comparison of clustering results with different numbers of constriction diameters, it was necessary to implement solutions with different maximum constriction diameter. The three solutions studied comprised two ($d^{(1)}$, $d^{(2)}$), three ($d^{(1)}$, $d^{(2)}$,$d^{(3)}$), and four ($d^{(1)}$, $d^{(2)}$, $d^{(3)}$, $d^{(4)}$) constriction diameters.

For the solution with two constriction diameters, $n_1 = 27$ and $n_2 = 76$ veins were assigned to the constriction diameter $d^{(1)} = 3.0$ mm and $d^{(2)} = 3.3$ mm, respectively. Adding the third constriction diameter, $d^{(3)} = 3.6$ mm, resulted in re-assignment of 51 veins from $d^{(2)}$ to $d^{(3)}$ while the number of veins assigned to $d^{(1)}$ remained the same, hence $n_1 = 27$, $n_2 = 25$ and $n_3 = 51$. For four constriction diameters, 33 veins were reassigned from $d^{(3)}$ to $d^{(4)} = 3.9$ mm and assignment of veins to constriction diameters $d^{(1)}$ and $d^{(2)}$ remained unchanged, resulting in a distribution of veins of $n_1 = 27$, $n_2 = 25$, $n_3 = 18$ and $n_4 = 33$.

3.3 Statistical Recursive Partitioning

Figure 5 illustrates the specific relationship of the coefficient of determination, R^2, with the number of partitions, s, for the recursive partitioning of the $D_{i,j}$ data of the 118 veins. The R^2-s curve approached a plateau for $s = 4$ partitions. This indicated the lack of substantial additional information, and benefit, for more than four partitions. The first four partitions proposed for the minimum vein diameter $D_{\min,i}$ were $d_P^{(1)} = 3.5$ mm, $d_P^{(2)} = 2.7$ mm, $d_P^{(3)} = 3.9$ mm and $d_P^{(4)} = 3.3$ mm, in that order. Since a diameter of less than 3.0 mm was considered impractical (constraint 2), the smallest value of $d_P^{(2)} = 2.7$ mm was replaced with $d_P^{(2)} = 3.0$ mm, in conjunction with exclusion of 15 veins with $D_{\min,i} < 3.0$ mm from the analysis. When using a single partition, $d_P^{(1)} = 3.5$ mm, cases of undesired vein distension, indicated by negative values of the applied constriction degree, were observed (see Fig. 6 A). The solutions with more than one partition sufficed for all 103 veins that remained in the analysis without distension and with an applied constriction degree of less than or equal to $C_A^{\max} = 0.5$ (Fig. 6 B-D). Compared to the two-partition solution with $d_P^{(1)} = 3.5$ mm and $d_P^{(2)} = 3.0$ mm (Fig. 6 B), the addition of the third partition $d_P^{(3)} = 3.9$ mm (Fig. 6 C) and fourth partition $d_P^{(4)} = 3.3$ mm (Fig. 6 D) did not appear to affect the extent of constriction. The distribution of veins amongst the partitions, ordered $d_P^{(1)} = 3.5$ mm, $d_P^{(2)} = 3.0$ mm, $d_P^{(3)} = 3.9$

mm and $d_{\mathrm{p}}^{(4)}$ = 3.3 mm, was $\hat{n}_1$ = 41.7% and $\hat{n}_2$ = 58.3 % for the two-partition solution, $\hat{n}_1$ = 22.3 %, $\hat{n}_2$ = 58.3 % and $\hat{n}_3$ = 19.4 % for the three-partition solution, and $\hat{n}_1$ = 22.3 %, $\hat{n}_2$ = 36.0 %, $\hat{n}_3$ = 19.4 % and $\hat{n}_4$ = 22.3 % for the four-partition solution.

4. Discussion

In this study, a simple mathematical method was developed for the constraint-based analysis and supervised clustering of small to medium-sized single parameter data sets. The method was motivated by the search for the minimum number of diameters, and their values, required for the constrictive elimination of diameter irregularities, or smoothing, of saphenous vein grafts of 100 patients. The elimination of diameter irregularities in vein grafts, a cause of local flow disturbances and changes in wall shear stress, promotes the prevention of focal intimal hyperplasia and graft stenosis [12, 13]. For the constrictive smoothing through the ablumenal application of a tubular mesh device to the vein graft [9, 10], a limited number of mesh devices with different diameters is beneficial, if not imperative for the translation of this technology into clinical practice.

The developed method comprises several elements of knowledge discovery from data, namely data selection, transformation, pattern evaluation and cluster analysis [1].

The data selection and transformation process served to extract minimum and maximum data points, $D_{\min,i}$ and $D_{\max,i}$, and to implement constraints on these data points for each vein. Pattern evaluation of the transformed data of all veins provided the basis for supervised identification of the required number, k, of constrictive smoothing diameters, $d^{(k)}$, and their values. Using the identified constrictive smoothing diameters, the clustering was performed employing decision-tree algorithms. Two algorithms were implemented to provide means for different prioritisation of the clustering while using the same constrictive diameter values.

Many clustering algorithms require *a priori* knowledge of the number of clusters [19]. Various methods for the estimation of the cluster number have been suggested in the statistical literature. The supervised identification of class labels, i.e. constrictive smoothing diameters, employed here in conjunction with targeting moderately sized data sets allowed for a simple clustering method without *a priori* knowledge of the number of classes. The method succeeded without the need for computer-based intelligent systems such as genetic algorithms, fuzzy techniques and machine learning which have been employed extensively for clustering of large and multidimensional data sets [3, 7, 20-22]. The proposed method was applied to diameter data obtained from 118 human saphenous vein grafts with a total of 1687 diameter values. The evaluation of the data of each vein and implementation of constraints (limits for constriction and distension and minimum for the vein graft diameter) resulted in exclusion of the data of 15 veins (i.e. 12 % of the entire data set) that were unfit for the objectives of the cluster analysis. For each of the 103 veins that remained in the analysis, the upper and lower threshold for the constriction diameter was determined. Using interactive pattern evaluation, the minimum number of constrictive smoothing diameters required for the 103 veins was found to be two. The subsequent clustering of the data was performed with the minimum of two constriction diameters as well as with three and four constriction diameters to evaluate the two alternative clustering algorithms and the resulting distribution of veins to the different constriction diameters.

The veins were most unevenly clustered when employing clustering algorithm CA1 which prioritised the smaller constriction diameter for veins that qualified for more than one constriction diameter ($\hat{n}_1$: 92.2 %, $\hat{n}_2$: 7.8 %, $\hat{n}_3 = \hat{n}_4 = 0$ %). The most even population of the constriction diameter classes was obtained with four diameters and prioritisation of the largest constriction diameter for veins fitting multiple constriction diameters ($\hat{n}_1$: 26.2 %, $\hat{n}_2$: 24.3 %, $\hat{n}_3$: 17.5 %, $\hat{n}_4$: 32.0 %). The even distribution resulted in a 24.3 % decrease in grand mean constriction of the vein grafts ($C_A = 27.8 \pm 9.5$ %) compared to the uneven two-class distribution favouring low diameter classes ($C_A = 36.7 \pm 8.8$ %). The two- and three-diameter clusters obtained with large-diameter class prioritisation (CA2) provided intermediate distribution patterns and intermediate overall constriction degrees.

The mathematical method was assessed by comparing the constriction diameters and vein clustering with the results of the statistical recursive partitioning of the data of the 118 veins. Both methods indicated that a single constriction diameter was insufficient to accommodate the cohort of veins under the given constraints, whereas two or more diameters sufficed. Good agreement was observed between the mathematically determined constrictive smoothing diameters (3.0, 3.3, 3.6 and 3.9 mm) and those obtained by recursive portioning (3.0, 3.3, 3.5 and 3.9 mm). The distribution of the veins amongst the constrictive diameters differed somewhat between the mathematical and the statistical method. This was however a result of different rules or priorities of the clustering and not a limitation of the novel mathematical method.

The fact that the two larger diameter classes, $d^{(3)}$ and $d^{(4)}$, were not populated when employing small-diameter prioritisation indicated the capacity of the solution to

accommodate vein data with larger diameter values, e.g. when dealing with vein grafts for peripheral grafting in the leg. For peripheral grafting, the harvested length of the vein is generally longer (mean: 686 mm, range: 610 – 740 mm, standard deviations not provided [23]), in conjunction with a flaring vein diameter from the knee towards the thigh [24], compared coronary bypass grafting (mean: 284 ± 95 mm, range: 100 – 520 mm, this study).
While the size of the data set was small enough for the clustering without computer-based intelligent systems, the data source of 100 patients allows for a cautious indication that the constrictive smoothing diameters found extend to larger data sets. This was supported by good agreement of the mathematically and statistically determined constrictive smoothing diameters. The analysis and clustering algorithms may be easily implemented in a programming code and computer-based intelligent systems for the analysis of extensive data sets where the constrictive diameters determined in this study can serve as initial cluster structure.

Acknowledgements

The authors wish to acknowledge Dr L Moodley, Dr W Lichtenberg, Dr J Scherman, Dr H Legodi, Dr G Mphahlele, and Dr I Taunyane of the Chris Barnard Division of Cardiothoracic Surgery, Groote Schuur Hospital, Cape Town, for performing the saphenous vein dimensional measurements. This research was mainly funded through a research collaboration grant by Medtronic Inc. (Minneapolis, MN, USA) to the University of Cape Town and a THRIP grant of the National Research Foundation (South Africa). The salary of TF was fully funded and that of PH was partially funded from the Medtronic research grant.

References

1. Han J, Kamber M: Data mining: Concepts and techniques. 2 ed. The morgan kaufmann series in data managment systems, ed. Gray J. 2006, Amsterdam: Elsevier. 770.
2. Durbin B, Dudoit S, van der Laan MJ: A deletion/substitution/addition algorithm for classification neural networks, with applications to biomedical data. J Stat Plan Inference, 2008; 138:464-488.
3. Hu Y-C, Chen R-S, Tzeng G-H: Finding fuzzy classification rules using data mining techniques. Pattern Recog Lett, 2003; 24:509-519.
4. Paetz J: Supervised neuro-fuzzy clustering for life science applications., in *Biological and medical data analysis*, Maglaveras N, Chouvarda I, Koutkias V, Brause R, Editors. 2006, Springer: Berlin. p. 378-389.
5. Zhang H, Ishikawa M: Bagging using hybrid real-coded genetic algorithm with pruning and its applications to data classification. Int Congr Ser, 2007; 1301:184-187.
6. Zhang Y, Bhattacharyya S: Genetic programming in classifying large-scale data: An ensemble method. Inf Sci, 2004; 163:85-101.
7. Vieira A, Barradas N: A training algorithm for classification of high-dimensional data. Neurocomputing, 2003; 50:461-472.
8. Xu P, Brock GN, Parrish RS: Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. Computational Statistics & Data Analysis, 2009; 53:1674-1687.
9. Zilla P, Human P, Wolf M, Lichtenberg W, Rafiee N, Bezuidenhout D, Samodien N, Schmidt C, Franz T: Constrictive external nitinol meshes inhibit vein graft intimal hyperplasia in nonhuman primates. J Thorac Cardiovasc Surg, 2008; 136:717-725.
10. Zilla P, Wolf M, Rafiee N, Moodley L, Bezuidenhout D, Black M, Human P, Franz T: Utilization of shape memory in external vein-graft meshes allows extreme diameter constriction for suppressing intimal hyperplasia: A non-human primate study. J Vasc Surg, 2009; 49:1532-1542.
11. Sarkar S, Schmitz-Rixen T, Hamilton G, Seifalian A: Achieving the ideal properties for vascular bypass grafts using a tissue engineered approach: A review. Med Biol Eng Comput, 2007; 45:327-336.

12. Yang C, Tang D, Liu SQ: A multi-physics growth model with fluid-structure interactions for blood flow and re-stenosis in rat vein grafts: A growth model for blood flow and re-stenosis in grafts. Computers & Structures, 2003; 81:1041-1058.
13. Liu SQ, Moore MM, Yap C: Prevention of mechanical stretch-induced endothelial and smooth muscle cell injury in experimental vein grafts. J Biomech Eng, 2000; 122:31-38.
14. Mills JL, Bandyk DF, Gahtan V, Esses GE: The origin of infrainguinal vein graft stenosis: A prospective study based on duplex surveillance. J Vasc Surg, 1995; 21:16-22; discussion 22-15.
15. Zhang JM, Chua LP, Ghista DN, Yu SC, Tan YS: Numerical investigation and identification of susceptible sites of atherosclerotic lesion formation in a complete coronary artery bypass model. Med Biol Eng Comput, 2008; 46:689-699.
16. Cacho F, Doblare M, Holzapfel GA: A procedure to simulate coronary artery bypass graft surgery. Med Biol Eng Comput, 2007; 45:819-827.
17. Tai NR, Salacinski HJ, Edwards A, Hamilton G, Seifalian AM: Compliance properties of conduits used in vascular reconstruction. Br J Surg, 2000; 87:1516-1524.
18. Schanzer A, Hevelone N, Owens CD, Belkin M, Bandyk DF, Clowes AW, Moneta GL, Conte MS: Technical factors affecting autogenous vein graft failure: Observations from a large multicenter trial. J Vasc Surg, 2007; 46:1180-1190.
19. Mardia KV, Kent JT, Bibby JM: Multivariate analysis. 1980, San Diego: Academic Press. 521.
20. Armand S, Watelain E, Mercier M, Lensel G, Lepoutre F-X: Identification and classification of toe-walkers based on ankle kinematics, using a data-mining method. Gait Posture, 2006; 23:240-248.
21. Moon H, Ahn H, Kodell RL, Baek S, Lin C-J, Chen JJ: Ensemble methods for classification of patients for personalized medicine with high-dimensional data. Artif Intell Med, 2007; 41:197-207.
22. Nadkarni NM, McIntosh ACS, Cushing JB: A framework to categorize forest structure concepts. For Ecol Manage, 2008; 256:872-882.
23. Papadopoulos NJ, Sherif MF, Albert EN: A fascial canal for the great saphenous vein: Gross and microanatomical observations. J Anat, 1981; 132:321-329.
24. Human P, Franz T, Scherman J, Moodley L, Zilla P: Dimensional analysis of human saphenous vein grafts: Implications for external mesh support. J Thorac Cardiovasc Surg, 2009; 137:1101-1108.

Tables

Table 1. Distribution of veins and constriction degrees for two, three, and four constriction diameters, $d^{(k)}$, obtained with clustering algorithms CA1 and CA2.

	CA1 4 classes	CA2 2 classes	CA2 3 classes	CA2 4 classes
Excluded				
n_e	14	14	14	14
$d^{(1)}$ [mm]	3.0			
n_1	95	27	27	27
$C_A^{(1)}$	0.36±0.09	0.30±0.10	0.30±0.10	0.30±0.10
$C_{A,min}^{(1)}$	0.09	0.09	0.09	0.09
$C_{A,max}^{(1)}$	0.50	0.47	0.47	0.47
$d^{(2)}$ [mm]	3.3			
n_2	8	76	25	25
$C_A^{(2)}$	0.48±0.01	0.34±0.08	0.31±0.09	0.31±0.09
$C_{A,min}^{(2)}$	0.46	0.13	0.13	0.13
$C_{A,max}^{(2)}$	0.49	0.49	0.48	0.48
$d^{(3)}$ [mm]	3.6			
n_3	0	N/A	51	18
$C_A^{(3)}$			0.29±0.08	0.27±0.09
$C_{A,min}^{(3)}$			0.14	0.14
$C_{A,max}^{(3)}$			0.45	0.45
$d^{(4)}$ [mm]	3.9			
n_4	0	N/A	N/A	33
$C_A^{(4)}$				0.24±0.08
$C_{A,min}^{(4)}$				0.09
$C_{A,max}^{(4)}$				0.40
Overall				
n	103	103	103	103
C_A	0.37±0.09	0.33±0.09	0.30±0.09	0.28±0.10
$C_{A,min}$	0.09	0.09	0.09	0.09
$C_{A,max}$	0.50	0.49	0.48	0.48

Figures

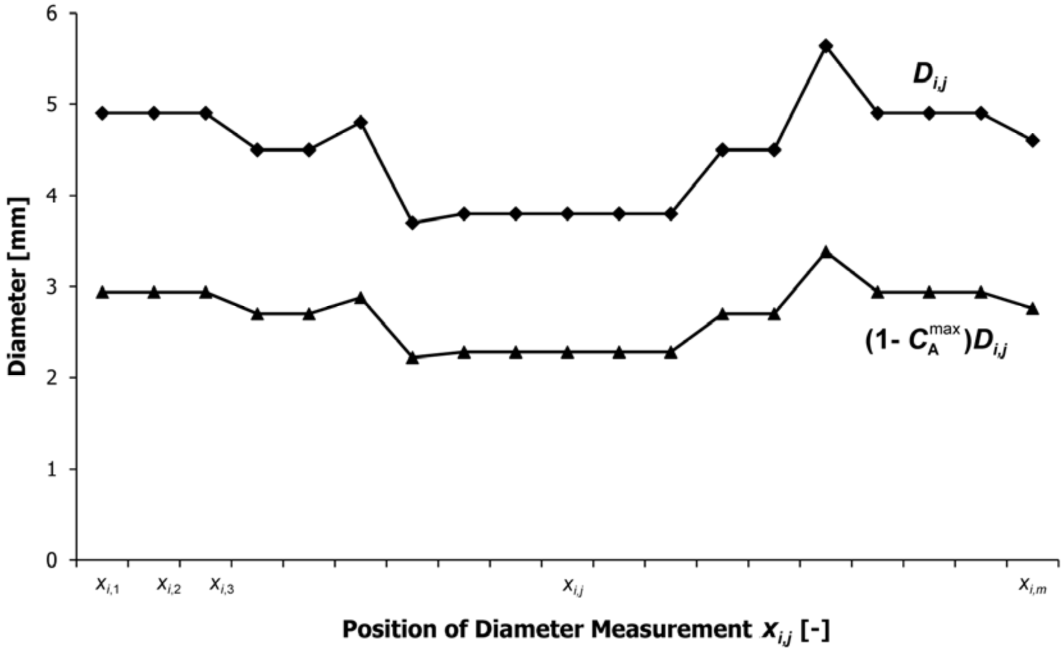

Figure 1. Graph illustrating the outer vein diameter, $D_{i,j}$, (diamonds) and the lower bound of the constrictive smoothing diameter, $(1-C_{\mathrm{A}}^{\max})D_{i,i}$, (triangles) at each measurement point, $x_{i,j}$, along the length of a harvested vein i.

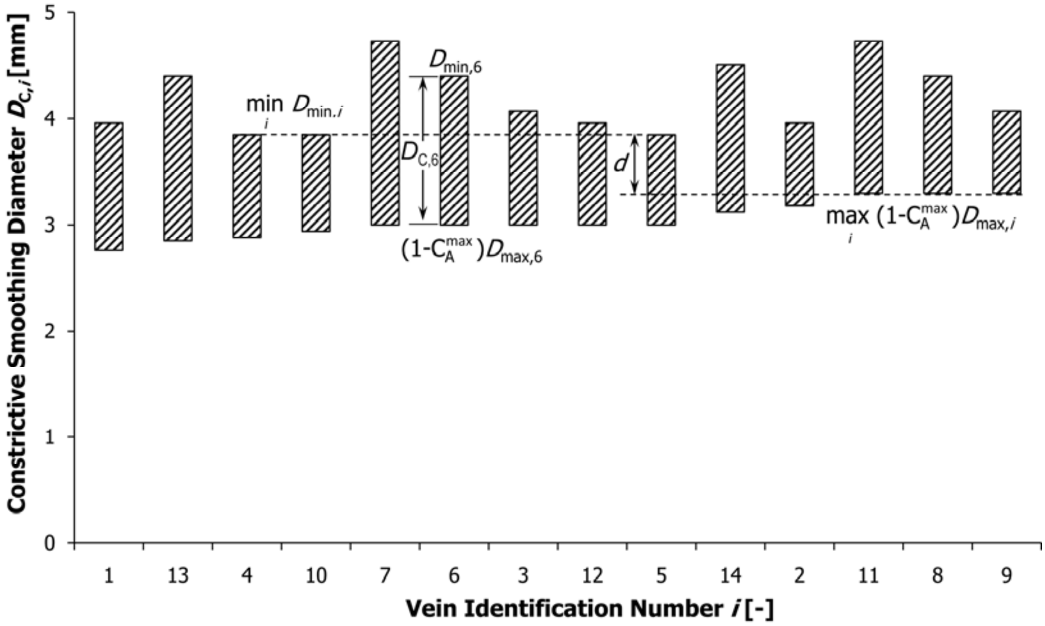

Figure 2. Graph showing individual mesh diameter ranges, $D_{C,i}$, for 14 veins ranked according to the minimum and the maximum individual constriction diameter, $(1-C_A^{max})D_{max,i}$ and $D_{min,i}$. where i = 1 to 14. For vein 6, the smallest and the largest constriction diameter, and its range are indicated. The range of the constrictive smoothing diameter, d, accommodating all 14 veins is determined by the largest minimum individual constriction diameter, $\max_i (1-C_A^{max}) \cdot D_{max,i}$, and the smallest maximum individual constriction diameter, $\min_i D_{min,i}$. (The 14 veins represent an arbitrary selection for demonstration purposes.)

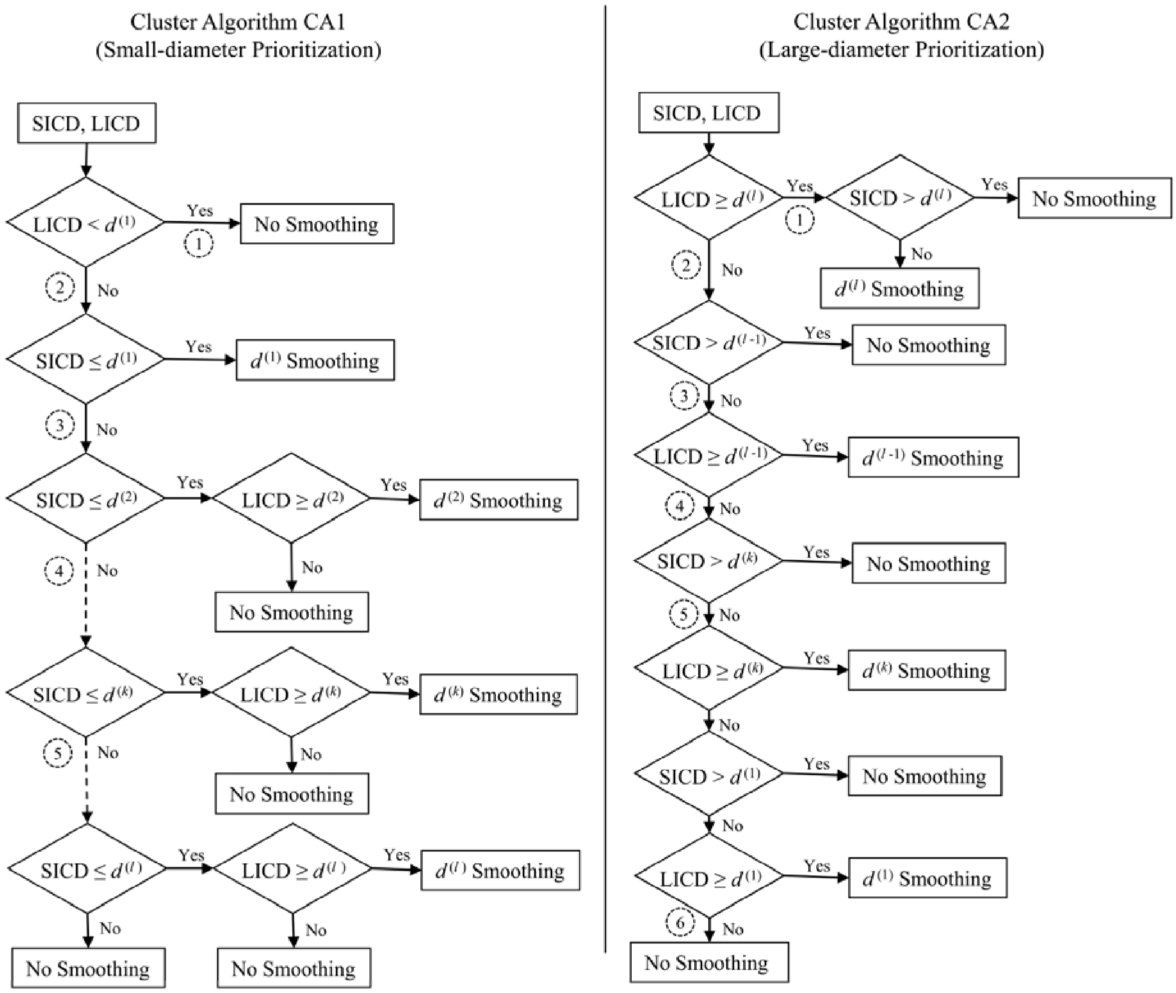

Figure 3. Decision tree charts of the clustering algorithms. Using the smallest individual constriction diameter (SICD), $(1 - C_{\mathrm{A}}^{\max})D_{\max,i}$, and the largest individual constriction diameter (LICD), $D_{\min,i}$, of a vein i, algorithms CA1 and CA2 assigned a vein i to the smallest and the largest constrictive smoothing diameter, $d^{(k)}$, respectively, suitable for that vein, where i = 1 to n and k = 1 to l. Numbers in dashed circles refer to step numbers in the description of the algorithms in section 2.2.2.

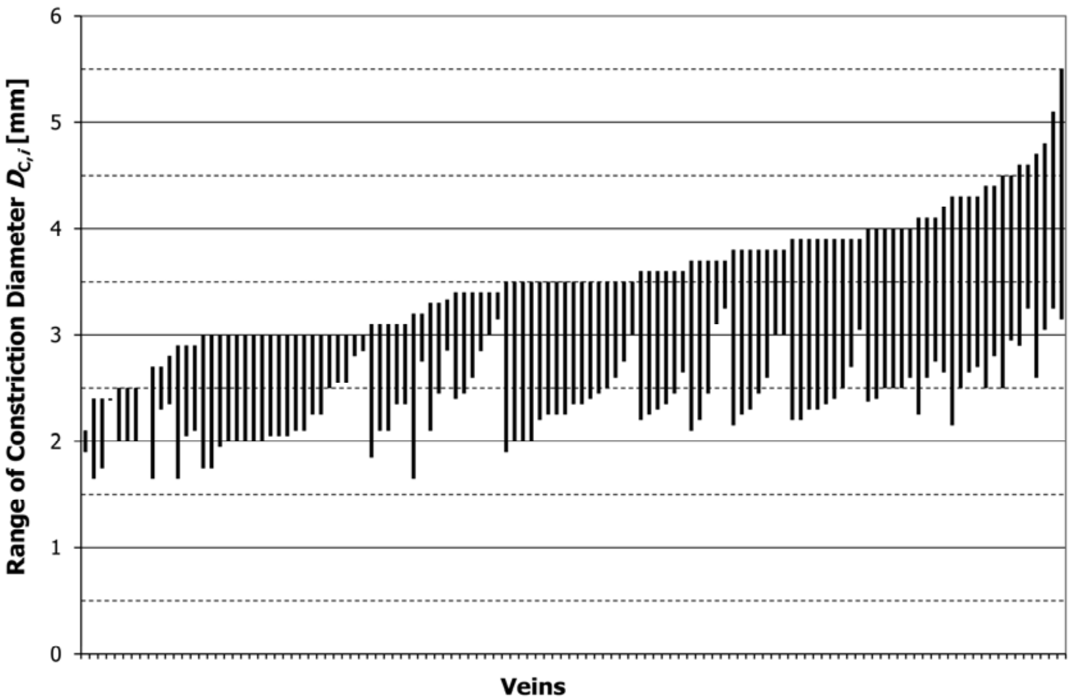

Figure 4. Graph showing the range of the individual constriction diameter, $D_{C,i}$, for the 117 veins that satisfied the condition that the maximum outer diameter is smaller than or equal to twice the minimum outer diameter, $D_{\max,i} \leq 2D_{\min,i}$, based on the maximum applied constriction degree, $C_A^{\max}$= 0.5. The veins i, with i = 1 to 117, are sorted according to maximum individual constriction diameter, $D_{\min,i}$, (first criteria) and then the minimum individual constriction diameter, $(1 - C_A^{\max})D_{\max,i}$, (second criteria). The vein identification numbers are omitted from the horizontal axis for clarity purposes.

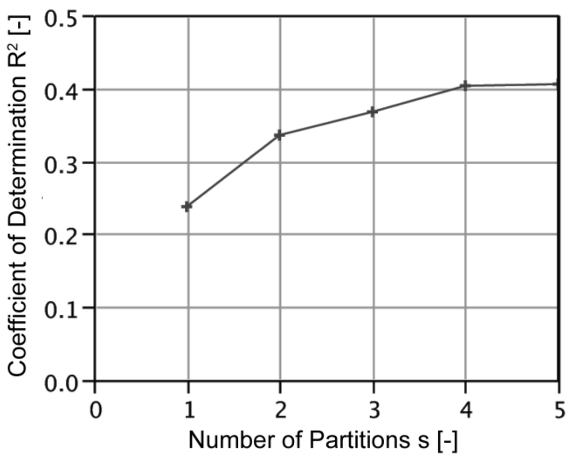

Figure 5. Graph of coefficient of determination, R^2, versus number of partitions, s, for the data set analysed demonstrating the impact of recursive partitioning of the veins into groups, by $D_{\min,i}$, that were most distinctly separated by the ratio of $D_{\max,i}/D_{\min,i}$. After the fourth partition, negligible benefit was derived which was indicated by R^2 approaching a plateau.

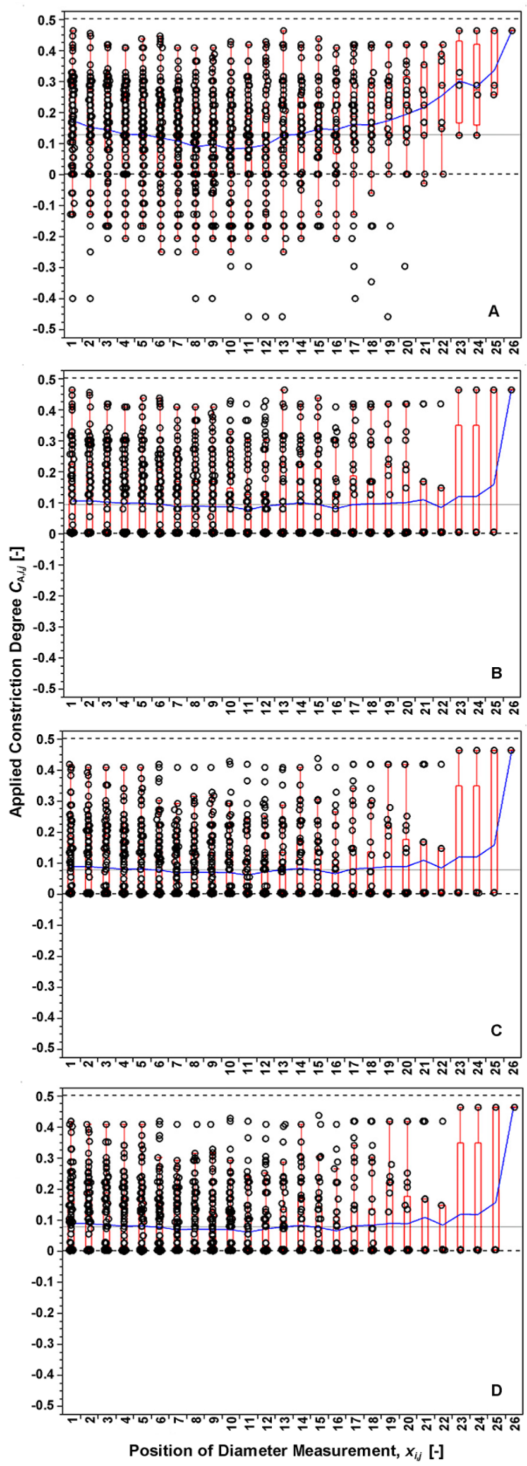

Figure 6. Box and whisker plots of the applied constriction degree, $C_{A,i,j}$, at each diameter measurement position, $x_{i,j}$, along the length of the 103 veins considered in the cluster analysis using the constrictive smoothing diameters (partitions), $d_P^{(s)}$, obtained with recursive partitioning. The results are given for four solutions comprising different numbers of partitions, or constrictive smoothing diameters: A) one partition $d_P^{(s)}$ = 3.5 mm; B) two partitions $d_P^{(1)}$ = 3.5 mm and $d_P^{(2)}$ = 3.0 mm; C) three partition $d_P^{(1)}$ = 3.5 mm, $d_P^{(2)}$ = 3.0 mm

and $d_{\mathrm{P}}^{(3)}$ = 3.9 mm; D) four partition $d_{\mathrm{P}}^{(1)}$ = 3.5 mm, $d_{\mathrm{P}}^{(2)}$ = 3.0 mm, $d_{\mathrm{P}}^{(3)}$ = 3.9 mm and $d_{\mathrm{P}}^{(4)}$ = 3.3 mm. For each solution, the following data is presented at each measurement position, $x_{i,j}$: Applied constriction degree, $C_{\mathrm{A},i,j}$, for each vein i [open circles]; median, lower quartile (i.e. 25$^{\mathrm{th}}$ percentile) and upper quartile (i.e. 75$^{\mathrm{th}}$ percentile) of $C_{\mathrm{A},i,j}$ [open boxes]; lowest value of $C_{\mathrm{A},i,j}$ within 1.5 interquartile range of the lower quartile [bottom whiskers]; and highest value of $C_{\mathrm{A},i,j}$ within 1.5 interquartile range of the upper quartile [upper whiskers]. In addition, the mean of the applied constriction degree $C_{\mathrm{A},i,j}$ plotted along the length of the vein [solid lines].