Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness

Original

Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans / Meiburger, K. M.; Marzola, F.; Zahnd, G.; Faita, F.; Loizou, C. P.; Laine, N.; Carvalho, C.; Steinman, D. A.; Gibello, L.; Bruno, R. M.; Clarenbach, R.; Francesconi, M.; Nicolaides, A. N.; Liebgott, H.; Campilho, A.; Ghotbi, R.; Kyriacou, E.; Navab, N.; Griffin, M.; Panayiotou, A. G.; Gherardini, R.; Varetto, G.; Bianchini, E.; Pattichis, C. S.; Ghiadoni, L.; Rouco, J.; Orkisz, M.; Molinari, F.. - In: COMPUTERS IN BIOLOGY AND MEDICINE. - ISSN 0010-4825. - ELETTRONICO. - 144:(2022), p. 105333. [10.1016/j.compbiomed.2022.105333]

Availability:
This version is available at: 11583/2959327 since: 2022-03-24T11:06:47Z

Publisher:
Elsevier Ltd

Published
DOI:10.1016/j.compbiomed.2022.105333

10 May 2026

Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans

Kristen M. Meiburger, Francesco Marzola, Guillaume Zahnd, Francesco Faita, Christos P. Loizou, Nolann Lainé, Catarina Carvalho, David A. Steinman, Lorenzo Gibello, Rosa Maria Bruno, Ricarda Clarenbach, Martina Francesconi, Andrew N. Nicolaides, Hervé Liebgott, Aurelio Campilho, Reza Ghotbi, Efthyvoulos Kyriacou, Nassir Navab, Maura Griffin, Andrie G. Panayiotou, Rachele Gherardini, Gianfranco Varetto, Elisabetta Bianchini, Constantinos S. Pattichis, Lorenzo Ghiadoni, José Rouco, Maciej Orkisz, Filippo Molinari

PII: S0010-4825(22)00125-1

DOI: https://doi.org/10.1016/j.compbiomed.2022.105333

Reference: CBM 105333

To appear in: *Computers in Biology and Medicine*

Received Date: 18 December 2021

Revised Date: 2 February 2022

Accepted Date: 16 February 2022

Please cite this article as: K.M. Meiburger, F. Marzola, G. Zahnd, F. Faita, C.P. Loizou, N. Lainé, C. Carvalho, D.A. Steinman, L. Gibello, R.M. Bruno, R. Clarenbach, M. Francesconi, A.N. Nicolaides, Hervé. Liebgott, A. Campilho, R. Ghotbi, E. Kyriacou, N. Navab, M. Griffin, A.G. Panayiotou, R. Gherardini, G. Varetto, E. Bianchini, C.S. Pattichis, L. Ghiadoni, José. Rouco, M. Orkisz, F. Molinari, Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans, *Computers in Biology and Medicine* (2022), doi: https://doi.org/10.1016/j.compbiomed.2022.105333.

Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans

Kristen M. Meiburger[a], Francesco Marzola[a], Guillaume Zahnd[b], Francesco Faita[c], Christos P. Loizou[d, q], Nolann Lainé[e], Catarina Carvalho[f], David A. Steinman[g], Lorenzo Gibello[h], Rosa Maria Bruno[i,s], Ricarda Clarenbach[j], Martina Francesconi[c,i], Andrew N. Nicolaides[k], Hervé Liebgott[e], Aurelio Campilho[f,l], Reza Ghotbi[j], Efthyvoulos Kyriacou[d], Nassir Navab[b,m], Maura Griffin[n], Andrie G. Panayiotou[o], Rachele Gherardini[i], Gianfranco Varetto[h], Elisabetta Bianchini[c], Constantinos S. Pattichis[p], Lorenzo Ghiadoni[i], José Rouco[q,r], Maciej Orkisz[e] and Filippo Molinari[a]

[a]Biolab, Department of Electronics and Communications, Politecnico di Torino, Torino, Italy

[b]Computer Aided Medical Procedures, Technische Universität München, München, Germany

[c]Institute of Clinical Physiology, Italian National Research Council, Pisa, Italy

[d]Department of Electrical Engineering and Computer Engineering and Informatics, Faculty of Engineering and Technology, Cyprus University of Technology, 3036 Limassol, Cyprus

[e]Univ Lyon, Université Claude Bernard Lyon 1, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, LYON, France

[f]INESC Technology and Science, Porto, Portugal

[g]Biomedical Simulation Lab, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada

[h]Dipartimento di Scienze Chirurgiche, University of Torino, Torino, Italy

[i]Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

[j]Helios Klinikum München West, München, Germany

[k]Cyprus Cardiovascular Disease Educational Research Trust, Nicosia, Cyprus

[l]FEUP - Faculty of Engineering, University of Porto, Porto, Portugal

[m]Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, MD, USA

[n]The Vascular Screening and Diagnostic Centre, Nicosia, Cyprus

[o]Cyprus International Institute for Environmental and Public Health, Cyprus University of Technology, Limassol, Cyprus

[p]Department of Computer Science, University of Cyprus, Nicosia, Cyprus

[q]Research Center of Information and Communication Technologies, UDC, A Coruña, Spain

[r]Department of Computer Science, University of A Coruña, A Coruña, Spain

[s]INSERM U970, Paris Cardiovascular Research Centre – PARCC and Université de Paris, Paris, France

Corresponding author:

Kristen M. Meiburger

Tel: +39 011 090 4207

Fax: +39 011 090 4217

E-mail: kristen.meiburger@polito.it

Abstract

After publishing an in-depth study that analyzed the ability of computerized methods to assist or replace human experts in obtaining carotid intima-media thickness (CIMT) measurements leading to correct therapeutic decisions, here the same consortium joined to present technical outlooks on computerized CIMT measurement systems and provide considerations for the community regarding the development and comparison of these methods, including considerations to encourage the standardization of computerized CIMT measurements and results presentation. A multi-center database of 500 images was collected, upon which three manual segmentations and seven computerized methods were employed to measure the CIMT, including traditional methods based on dynamic programming, deformable models, the first order absolute moment, anisotropic Gaussian derivative filters and deep learning-based image processing approaches based on Unet convolutional neural networks. An inter- and intra-analyst variability analysis was conducted and segmentation results were analyzed by dividing the database based on carotid morphology, image signal-to-noise ratio, and research center. The computerized methods' obtained CIMT absolute bias results that were comparable with studies in literature and they generally were similar and often better than the observed inter- and intra-analyst variability. Several computerized methods showed promising segmentation results, including one deep learning method (CIMT absolute bias = 106±89 µm vs. 160 ± 140 µm intra-analyst variability) and three other traditional image processing methods (CIMT absolute bias = 139±119 µm, 143±118 µm and 139±136 µm). The entire database used has been made publicly available for the community to facilitate future studies and to encourage an open comparison and technical analysis (http://dx.doi.org/10.17632/m7ndn58sv6.1).

Keywords: intima-media thickness, carotid artery, ultrasound imaging, segmentation, deep learning, open database

1. Introduction

Carotid intima-media thickness (CIMT) is a commonly used marker for atherosclerotic risk assessment. The CIMT is typically measured on B-mode ultrasound images of the common carotid artery (CCA) by delineating the intima-media complex (IMC) [1] (see also Fig. 1), although in some studies carotid Doppler images are also employed to assist CIMT measurement [2]. Increased CIMT values have been associated with future cardiovascular events in high-risk subjects [3], yet studies have also shown negative results on the independent predictive value of CIMT for cardiovascular events [4]. As ultrasound imaging heavily depends on the clinical operator and his/her skills, a common limitation of CIMT measurements is the heterogeneity in technical approaches [3]. Namely, the CIMT value can vary based on the protocol used (e.g., single- vs. multiple-angles acquisition), the measurement location (e.g., distance from bulb), the ultrasound equipment used to acquire the image (e.g., transducer central frequency), and finally the image acquisition setup (e.g., filters, image gain, time-gain compensation, depth). This last factor (i.e., image acquisition setup) has been found to potentially impact the robustness of the CIMT measurements [5] and is in dire need of standardization when using B-mode-based systems [6]. Finally, several large studies focusing on the predictive value of CIMT employed manual measurements by the placement of calipers, which can be subject to inter- and intra-analyst variability and require a large amount of time and effort.

A myriad of computerized methods have been proposed to extract the contours of the IMC and measure the CIMT in longitudinal ultrasound images of the CCA [7] . The problem of segmenting the CCA can be formally described as determining the position of the lumen-intima (LI) and media-adventitia (MA) anatomical interfaces (see also Fig. 1). Usually, CIMT quantification is performed on the far wall of the CCA, as recommended by the Mannheim consensus [8]. A large number of segmentation methods have been extensively described, categorized, and compared in several dedicated reviews [7,9,10]. A brief summary of the major IMC segmentation approaches is presented hereafter. Anatomical interfaces have a continuous profile and are represented by abrupt changes in image intensity, due to the change of acoustic impedance between different tissue types [11]. A variety of edge detection operators have been proposed to analyze and identify the gradient extrema for IMC segmentation [11,12]. An advanced analysis of local image properties was moreover introduced to detect gradient regions (namely, where anatomical interfaces are likely to be located) as a saliency map [13]. Active contours, also referred to as snakes, have been exploited in several different approaches [14,15]. Snakes segmentation is well suited to the IMC geometry due to the clear double-line pattern; however, convergence depends on initialization conditions. Another widely used approach is dynamic programming, which ensures a globally optimal and deterministic solution [16]. Based on graph theory, this approach provides the minimum-cost path running from the left to the right borders of the image which represents the target anatomical interface. In the aim to further increase segmentation robustness, a scheme known as dual-dynamic programming was introduced to extract simultaneously the LI and MA interfaces from ultrasound CCA images [17] and videos [18,19]. The above presented techniques can be described as contour-based, since the focus is to extract the interfaces between the different anatomical regions using various edge operators. Contrarily, region-based methods are devised to extract entire areas based on texture or intensity measurements, using for example adaptive thresholding and morphological operations [17] or fuzzy C-mean clustering [20]. Finally, machine learning approaches have been proposed to segment the layers and quantify the CIMT [21], with a particular focus on deep learning methods in recent years [22,23], which is undergoing an incredible evolution and has recently shown to provide robust performances in several tasks. These systems, known as deep neural networks, extract a high-dimensional representation of the image and use this information to reconstruct a segmentation map of the objects in the image. Deep neural networks perform the extraction, transformation, and interpretation of intrinsic image features via many multi-scale layers of operations (such as, for instance, convolutions), whose parameters are trained in a supervised manner. Deep learning approaches provide a generalization of the segmentation problem, being applied for the segmentation of both the IMC and/or plaques [24,25]. Briefly, some of the recent works on deep learning and CIMT measurement include studies by Jain et al. [26–28] showing the feasibility of deep learning frameworks for segmentation of the IMC and plaques present in the common and internal carotid artery. Biswas et al. published a review on artificial intelligence frameworks for CIMT and plaque area measurements [29] and numerous studies for accurate CIMT measurement and joint wall thickness and plaque burden [24,30]. Deep neural networks benefit from a tremendously fast inference time, they can therefore be also applied for video annotation [31].

A literature review by Saba et al. [32] pointed out the importance of standardizing tools used for measurement, risk stratification and risk assessment, demonstrating the importance of employing full-length CIMT measurements using a large number of equidistant sample points [33].

We recently published a study presenting the comparison between five different computerized CIMT measurement methods on a common large clinical database [34]. In that study, the correlation with clinical parameters and differences in ability to predict cardiovascular events were confronted, where it was found that the CIMT measurements obtained with a skilled analyst’s segmentation and the computerized segmentation were comparable in (i) correlation with clinical parameters, (ii) cardiovascular events prediction through a generalized linear model, and (iii) the Kaplan-Meier hazard ratio, suggesting they can be used interchangeably for CIMT quantification and clinical outcome investigation.

Here we aim to provide, for the first time in literature, a technical outlook on seven different computerized CIMT measurement systems on a unique database, in an attempt to provide considerations (such as the importance of how the CIMT is computed, how the segmentation results should be presented, and how the carotid geometry and image signal-to-noise-ratio influence the segmentation results) for the scientific and medical communities when facing the issue of computerized CIMT measurement systems, both in their development and in their comparison with other methods. The database used in this study, including the ultrasound images used, manual segmentations and computerized segmentations of all analyzed methods, has been made publicly available (http://dx.doi.org/10.17632/m7ndn58sv6.1).

2. Materials and Methods

2.1 Database description

A total of 500 images were included in this study complying with the Declaration of Helsinki. In particular, 400 images were acquired at four different centers (100 per center) using different ultrasound device systems and probes. Table 1 shows the breakdown of the images and the systems/settings used at each center, and Figure 1 shows an example image from each center. The Ethics Committee of each relevant center approved the study and all participants provided written informed consent. The Mannheim consensus guidelines for image acquisition was followed for all participants [8]. Moreover, 100 images were simulated using the Fast And Mechanistic Ultrasound Simulation (FAMUS) software [35,36]. Briefly, this simulator relies on a point source/receive approach and combines the speed of other approximate approaches with the flexibility and realism of mechanistic approaches. More details can be found in Aguilar et al. [35,36]. In particular, the in-silico phantoms were defined using a binary mask where the IMC was manually drawn along the image width. FAMUS was subsequently used to simulate the final B-mode images, thus providing ground truth profiles to be used for comparison between both manual and computerized profiles. The dimensions of the simulated phantoms were 40×1×25 mm^3 in the x, y and z (lateral, elevation and axial) directions using 30,000 scatterers (30 scatterers/mm^3). The simulated array transducer had a central frequency of 7 MHz with 192 active elements, and the B-mode image was reconstructed using 128 scanlines. More details about the simulation parameters can be found in the Supplementary Materials.

The pixel dimension of the images included in the entire dataset presented a mode equal to 0.060 mm/pixel, with a minimum of 0.029 mm/pixel and a maximum of 0.099 mm/pixel. The pixel dimension, also referred to as the calibration factor (CF), of each image is included in the publicly available dataset (http://dx.doi.org/10.17632/m7ndn58sv6.1).

Table 1. Database description

Center	Number of images	Average conversion factor (mm/pixel)	Ultrasound scanner and probe	Settings
Munich	100	0.073	EPIQ 7G (Philips, Amsterdam, The Netherlands) L12-3 12-3 MHz linear array	- Central TGC - Gain 44-48 dB - Dynamic range: 56 dB - No persistence
Pisa [37,38]	100	0.076	MyLab25 (Esaote, Florence, Italy) LA523 4-13 MHz linear array	- Central TGC - Average 96 dB dynamic range - No persistence - No filters
Porto	100	0.062	ATL HDI500 (ATL Ultrasound, Seattle, WA) L12-5 5-12 MHz linear array	- Varying TGC and dynamic range - No persistence - No filters
Torino [39–42]	100	0.064	ATL HDI500 (ATL Ultrasound, Seattle, WA) L12-5 5-12 MHz linear array	- Varying TGC and dynamic range - No persistence - No filters
Toronto (FAMUS)	100	0.045	Linear array with central frequency = 7MHz*	- No TGC applied* - Dynamic range: 45 dB* - No persistence* - No filters*

*For detailed simulation parameters, see Supplementary Materials. References next to Center indicate previous studies where the same dataset either in part or in full was employed.

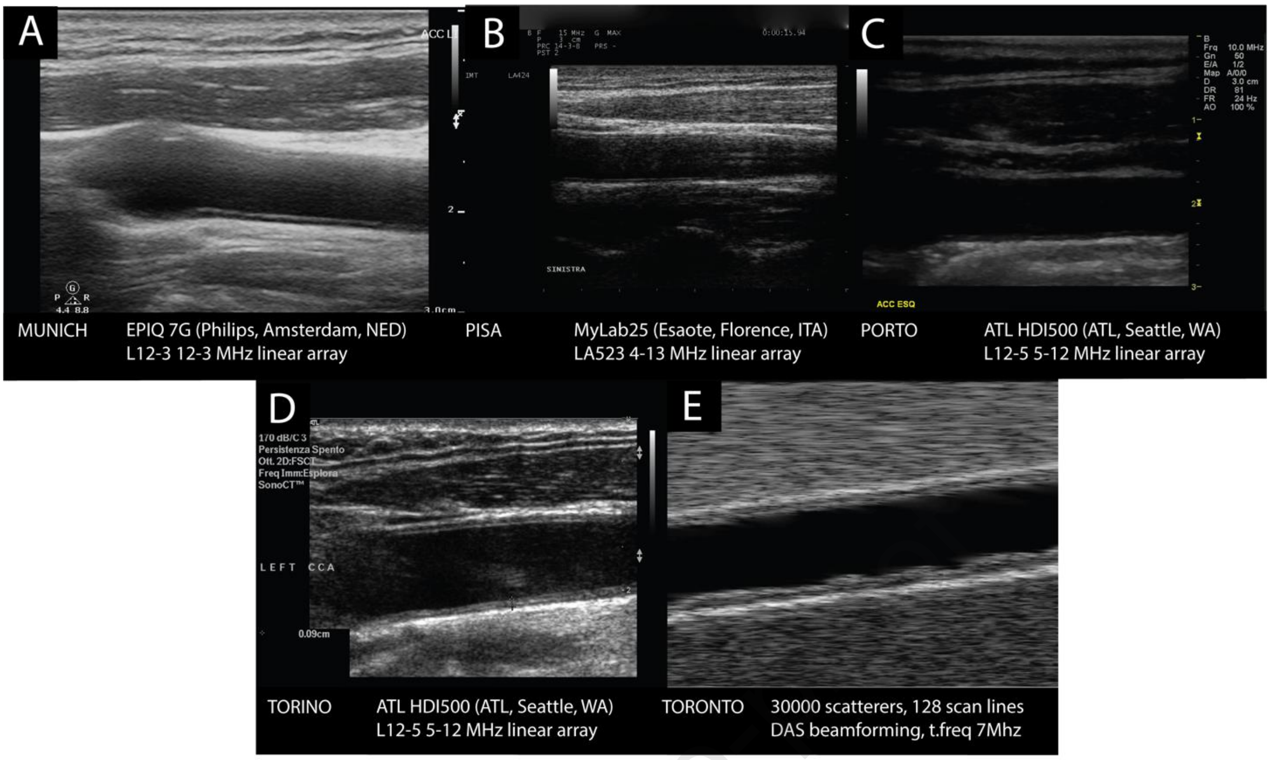


Figure 1. Example of images from each center (A-D) in-vivo images. (E) FAMUS simulated images. (A) Munich, Germany; (B) Pisa, Italy; (C) Porto, Portugal; (D) Torino, Italy; and (E) Toronto, Canada.

2.2. Manual and Automated CIMT measurement methods

For all 500 CCA ultrasound images investigated in this study, both manual and computerized measurement methods were performed, which are described hereafter.

2.2.1. Manual measurements

In order to assess the accuracy of each computerized segmentation method, a manual gold standard reference was generated, despite the lack of absolute ground truth inherent to ultrasound in-vivo data. After determining the full exploitable width of each image in order to (i) exclude regions of poor image quality if present, and to (ii) follow the Mannheim consensus guidelines [8], the contours of the LI and MA anatomical interfaces were subsequently manually traced by two experts. Manual LI and MA tracings were performed on each image by an experienced analyst A1 (L.G. from Torino, >10 years of experience in carotid sonography) and were considered the gold standard. The entire manual segmentation process was performed again by A1 one month later to assess intra-analyst variability (referred to as A1'), and by another expert analyst A2 (G.V. from Torino, > 25 years of experience in carotid sonography). The annotations performed by A1, A1', and A2 were blinded from each other. Manual segmentations were performed with care using a graphical user interface developed specifically for this purpose. More details of the manual segmentation GUI are provided in section 2 of the Supplementary Materials.

2.2.2 Computerized measurements

Seven computerized segmentation methods were developed by the authors, originating from six different research groups. For simplicity, each technique is named by the affiliation institution of the first author, with a subscript that indicates the corresponding country. The methods were employed on all 500 images used in this study. All methods produce the LI and MA tracings; some require user-interaction ($\mathrm{CNR_{IT}}$, $\mathrm{INESCTEC_{PT}}$, $\mathrm{TUM_{DE}}$, $\mathrm{UCY_{CY}}$, $\mathrm{CREATIS_{FR}}$), while others ($\mathrm{POLITO^{Snakes}_{IT}}$, and $\mathrm{POLITO^{UNet}_{IT}}$), are completely automatic. Each method was employed independently, so the manually determined region-of-interest (ROI) in the case of semi-automatic methods could vary. Five of the methods are based on traditional image processing techniques ($\mathrm{CNR_{IT}}$, $\mathrm{INESCTEC_{PT}}$, $\mathrm{TUM_{DE}}$, $\mathrm{UCY_{CY}}$, $\mathrm{POLITO^{Snakes}_{IT}}$), while the remaining two ($\mathrm{CREATIS_{FR}}$, and $\mathrm{POLITO^{UNet}_{IT}}$) are based on deep learning convolutional neural network (CNN) methods[1]. Briefly, the methods are based on the first-order absolute moment [$\mathrm{CNR_{IT}}$] [6,11], anisotropic Gaussian derivative filters [$\mathrm{INESCTEC_{PT}}$] [12,43], dynamic programming [$\mathrm{TUM_{DE}}$] [19,44], snakes [$\mathrm{UCY_{CY}}$] [14], dual snakes [$\mathrm{POLITO^{Snakes}_{IT}}$] [15], a dual-resolution UNet [$\mathrm{CREATIS_{FR}}$] [45], and a standard UNet [$\mathrm{POLITO^{UNet}_{IT}}$]. More details of the methods are described below, and readers are directed to each respective reference for a complete in-depth presentation and validation of the computerized methods.

2.2.2.1. Method by Consiglio Nazionale delle Ricerche (CNR) in Italy **($\mathrm{CNR_{IT}}$)**

---

[1] The $\mathrm{POLITO_{IT}}$ research group developed two algorithms: one based on traditional image processing and the other one based on CNNs. To distinguish between these two, a superscript is used.

The computerized algorithm adopted by the research group from the Consiglio Nazionale delle Ricerche (CNR) in Pisa, Italy [11,46] requires the user to specify a rectangular ROI that includes the entire section of the carotid (i.e., both near wall and far wall) that is to be analyzed, and then includes three automatic stages namely: (i) edge detection, (ii) heuristic search of local maxima, and (iii) identification of boundary regression lines. If an image sequence is analyzed, the user interaction is required only for the first frame and then the software automatically segments all subsequent frames.

For the edge detection stage and the identification of the IMC, the first order absolute moment (FOAM) operator method has been implemented [47] according to the following equation:

$$\mathrm{FOAM}(x,y)=\frac{1}{A_\Theta}\iint_\Theta \left|f(x,y)-f(x-\tau_x,y-\tau_y)\right| d\tau_x d\tau_y$$

where *f(x,y)* is the gray map of an image, and Θ is a circular domain with area A_Θ whose center is the point with coordinates *(x,y)*. The FOAM operator was evaluated on all pixels belonging to a manually placed ROI by the user of the system containing the lumen and the far wall of the CCA. A typical response of the FOAM operator applied to the aforementioned grey-level discontinuities is minimal on pixels belonging to the lumen of the vessel (anechoic), and maximal when the FOAM is evaluated at the level of the LI and MA interfaces [11,46,48].

As a second step, a heuristic search was applied to every column of the FOAM operator output distinctly [11,46]. The search algorithm finds and highlights with a flag all local maxima. The first local maximum greater or equal to the absolute maximum of the FOAM in the column multiplied by a first threshold is searched for; then, a second local maximum, following the first, greater or equal to the first maximum multiplied by a second threshold is located. The first maximum identifies the position of the LI interface, while the second one pinpoints the presence of the media-adventitia interface.

Subsequently, an outlier removal stage was applied to suppress all spurious local maxima and two regression lines were evaluated using first and second maxima (of all columns), respectively, to provide the final segmentation profiles.

2.2.2.2. Method by INESC Technology and Science (INESCTEC) in Portugal (**INESCTEC$_{\mathbf{PT}}$)**

The computerized method proposed by the research group from the INESC Technology and Science (INESCTEC) in Porto, Portugal [12] requires the user to manually click on the top left corner and bottom right corner to determine the ROI in which the algorithm then operates, which uses a robust method for edge extraction in ultrasound images [43] that is based on anisotropic Gaussian derivative filters and non-maximum suppression over the overall artery wall orientation in local regions. The anisotropic filters allow using a wider integration scale along the edges while preserving the edge location precision. The method also performs edge continuation, resulting in the connection of isolated edge points along linear segments, which is a valuable feature for the segmentation of the CCA wall layers. However, as also reported by Rouco et al. [49], this forced continuation, usually results in false edges being detected near convex contours and isolated points. The non-maximum suppression over pooled local orientations allows for a robust estimation of the wall orientation and solves this issue as already proposed in a previous study [43]. The LI and MA boundaries are obtained using a dynamic programming optimization procedure over the edge maps, based on a previously introduced approach [12].

2.2.2.3. Method by Technische Universität München (TUM) in Germany (**TUM$_{\mathbf{DE}}$**)

The main characteristics of the computerized approach proposed by an international team under the leadership of Technische Universität München (TUM) in Germany [19,44] is to extract both anatomical interfaces (namely, LI and MA) in one go, as opposed to one after the other. This is achieved via the introduction of two key ingredients: an original filter bank and a specific dynamic programming scheme. The filter bank consists of shape-adapted filters $\mathcal{H}_\Delta$: $\mathcal{H}_\Delta = \pm {G'}_{\sigma_y} * \mathcal{M}_\Delta$, where $(*)$ is the convolution operator, ${G'}_{\sigma_y}$ is the first derivative of a Gaussian with standard deviation σ_y, $\mathcal{M}_\Delta$ is a pair of Diracs separated by a distance Δ, and the $\pm$ sign is positive for the near wall and negative for the far wall. The filters are applied column-wise, for several values of Δ. The output is a 3D velocity map (x, y, Δ) where the high positive values correspond to the points located exactly mid-way between the two contours to be extracted. The dynamic programming scheme consists in determining the optimal path that describes the IMC skeleton, more specifically: for each x coordinate, the y location of the point located mid-way between the two interfaces, and the local thickness Δ associated. A 3D positive cost map is first directly constructed from the velocity map, then a front propagation along increasing values of x is carried out, penalizing abrupt displacements along the two other directions y and Δ. The 2D optimal path is found via back-tracking, and the two anatomical 1D contours can be immediately deduced.

The method is devised to fully automatically segment image sequences, in which case it can take advantage of local changes from one frame to another to define the region of interest. For still images used in this study, a lightweight interactive initialization process from the previous version of the method [50] was adopted, the main purpose of which is to select, for each image, the largest exploitable region. First, the left and right borders of the region were indicated (2 mouse clicks), then the approximate location and curvature of the far wall was provided (3 clicks), finally the rough IMT value was specified (2 clicks). The method is implemented within the CAROLAB platform developed in-house and is freely available [44].

2.2.2.4. Method by University of Cyprus in Cyprus (**$\mathbf{UCY_{CY}}$**)

The computerized technique proposed by the research group from the University of Cyprus [14] is based on deformable models, also known as snakes. Before running the IMC snakes segmentation algorithm, an IMC initialization procedure was carried out for positioning the initial snake contour as close as possible to the area of interest [14]. At the first stage of the proposed initialization procedure, the user of the system selects using the mouse two points including the ROI where the IMT will be detected. After several automated steps the snake converges to the final LI and MA borders, where the user is allowed to correct or move the snake points before the final snake contours are saved. The Williams & Shah snake segmentation method [51] was used to deform the snake and segment the IMC borders in each image. The snake contour, v(s), adapts itself by a dynamic process that minimizes an energy function:

$$E_{snake}[\upsilon(s)]=E_{int}\big(\mathrm{v(s)}\big)+\mathrm{E}_{image}\big(\mathrm{v(s)}\big)+\mathrm{E}_{ext}\big(v(s)\big)$$

$$= \int_{s} \alpha_s E_{cont}(v(s)) + \beta_s E_{curv}(v(s)) + \gamma_s E_{image}(v(s)) + \mathrm{E}_{ext}(v(s))\, ds$$

Where E_{int}, $E_{image}(v(s))$, $\mathrm{E}_{ext}(v(s))$, $E_{cont}(v(s))$, and $E_{curv}(v(s))$ are the internal, image, external, continuity, and curvature energies of the snake and α_s, β_s and γ_s are the strength, tension, and stiffness parameters, respectively. This method was also proposed and evaluated in two other studies on different datasets: on 1104 ultrasound images of the left and right CCA [52] and in 100 ultrasound images of the CCA [14]. For the Williams & Shah snake, the strength, tension and stiffness parameters were equal to 0.6, 0.4 and 2 respectively. The extracted final snake contours correspond to the LI and MA borders of the IMC.

2.2.2.5. Snakes method by Politecnico di Torino in Italy (**POLITO$^{\mathbf{Snakes}}_{\mathbf{IT}}$**)

The computerized technique based on traditional image processing methods proposed by the research group from the Politecnico di Torino in Turin, Italy is completely automatic and is based on a first-order gaussian filtering and a multi-resolution approach for carotid artery identification and far adventitial wall tracing [39] and then a dual snake formulation for LI and MA identification, similar to the approach presented in [15]. Both the LI and MA snakes are automatically initialized by placing the automatic far adventitial wall segmentation upwards. The dual snakes then simultaneously evolve using the equation as presented in the study by Molinari et al. [15]. The method adopted for the segmentation uses the same internal energy and a mutual distance energy term, but rather differs from the previously published approach because of its external energy term. Specifically, instead of using the first-order absolute moment energy contribution, a phase symmetry contribution was employed [53], based on the source code provided by Peter Kovesi [54]. A description of the energy of the snake to be minimized is displayed in the following equation:

$$\mathrm{E}\left[\mathrm{v}\left(\mathrm{x_i},\mathrm{y_j}\right)\right] = \int_0^{\mathrm{N}} \alpha \mathrm{F_{int}}\left(\mathrm{x_i},\mathrm{y_j}\right) + \beta \mathrm{F_{ext}}\left(\mathrm{x_i},\mathrm{y_j}\right) + \gamma \mathrm{g}$$

where:

$$F_{int}(x_i)=\alpha\left(\frac{1}{2}\left(y(x_{left})+y\left(x_{right}\right)\right)-y(x_i)\right)$$

(xi, yi) are the coordinates of each point on the LI and MA profiles. F_{ext} is calculated based on an external force map derived from the calculation of the symmetries inside the image [48,49]. The last term γg is a downward (for LI) or upward (for MA) force to drive the convergence of the LI and MA snakes. This CPU-based method processes an image in about 2 seconds on a laptop with an i5 4-core CPU. The parameters used for the dual snake model are shown in the Supplementary Materials (Table S3).

2.2.2.6. Method by Université Claude Bernard Lyon in France (**CREATIS$_{\mathbf{FR}}$**)

The method proposed by CREATIS team [45], like several state-of-the-art methods, begins by a very simple user interaction: two mouse clicks defining the left and right limits of the exploitable part of the image. The remainder is fully automatic and comprises two major steps: (i) extraction of a line approximately localizing the IMC of the arterial far wall, and (ii) precise delineation of the LI and MA interfaces along this line. Each of these steps first uses a deep convolutional neural network, actually a dilated UNet, to predict a region, and then seeks (a) smooth path(s) fitting the contour(s) of the region. The first UNet was trained to segment the region extending from the centreline of IMC to the bottom of the image; the upper boundary of the predicted region, approximated by a polynomial curve, is passed to the second step. The second UNet was trained to segment IMC along this curve; the upper and lower contours of the predicted IMC constitute the output of the method. Both UNets use fixed-size overlapping windows distributed between the previously defined left and right limits. The predictions obtained within these windows are combined to produce the final segmentation. The vertical extent of the windows differs between the steps: the first UNet uses windows extending on the whole height of the image, while the height of the windows used by the second UNet is restricted to a few millimetres below and above the curve localizing IMC. This small height is finally upsampled to permit a sub-pixel precision. The code for this method is available at https://github.com/nl3769/caroSegDeep.

2.2.2.7. UNet method by Politecnico di Torino in Italy (**POLITO$^{\mathbf{UNet}}_{\mathbf{IT}}$**)

The deep learning method proposed by the Politecnico di Torino team is based on two steps and is completely automatic, requiring no input from the user: (i) a rule-based algorithm for ROI extraction, and (ii) a UNet [55] for the definition of the IMC segmentation mask. In the first step, the image is cropped to include the whole acquired US image, avoiding the ancillary information about the acquisition. First of all, the rows and columns with only zeros at the borders of the image are removed. Then, the entropy of each row (column) is calculated, and the steepest changes in two consecutive rows (columns) are detected. The image is cropped at those points, and then padding is performed to reach an aspect ratio similar for each image to avoid too much distortion in the following resize. Finally, the image is resized to 480×352 pixels to train the UNet. The same procedure with cropping, padding, and resizing is applied on the mask created on the original image.

A standard UNet with a Dense201 encoder [56] and Channel "Squeeze and Excitation" Blocks [57] was employed. The softmax output was thresholded at 0.5 (0-1 range) and if more than one connected area was detected, only the biggest one was maintained. Finally, the LI and MA profiles were extracted from the semantic map of the image.

The five traditional image processing techniques (CNR$_{IT}$, INESCTEC$_{PT}$, TUM$_{DE}$, UCY$_{CY}$, POLITO$^{Snakes}_{IT}$) were already analyzed and validated on clinical images (n=2176) in our previous study [34], whereas the two methods that rely on deep learning (i.e., CREATIS$_{FR}$ and POLITO$^{UNet}_{IT}$) were not. Hence, the training of the networks included also these clinical images. In particular, the entire dataset presented here (n=500) and the images from the previous study (n=2176) were mixed together and randomly divided into 5 folds, in order to perform a 5-fold cross validation. Hence, each of the 500 images considered for the results analysis belonged to the test subset of one fold and the corresponding segmentation results for these images were put together for the analysis reported here. The division of the dataset for each fold into training, validation, and test sets are included in the publicly available dataset.

A table summarizing the pros and cons of each method briefly presented here, along with the processing time, is provided in the Supplementary Materials (Table S4).

2.3. CIMT computation and common support calculation

The computerized methods produce LI and MA profiles that contain a variable number of points along the image width; some methods (TUM$_{DE}$, POLITO$^{Snakes}_{IT}$, POLITO$^{UNet}_{IT}$, CREATIS$_{FR}$) present one point for each image column, where the profiles are defined, while others have one point every three columns (CNR$_{IT}$) or every 0.1 mm approximately (INESCTEC$_{PT}$). The UCY$_{CY}$ computerized method produces an amount of points that is equivalent to the manual segmentations, thus containing 10-20 points along the image width. In order to standardize CIMT computation and to facilitate the definition of the common support the following procedure was employed. The area of the image where all the methods proposed in this work have properly defined the LI and MA profiles was firstly estimated. Then the LI and MA segmentation profiles were interpolated to contain one point for each image column (within the original width of each profile) using a shape-preserving piecewise cubic interpolation. It should be noted that this interpolation was not done for the TUM$_{DE}$, POLITO$^{Snakes}_{IT}$, POLITO$^{UNet}_{IT}$, CREATIS$_{FR}$ methods which already presented one point for each column.

Using all manual and computerized segmentations (interpolated, if needed as described previously), the final CIMT values were computed using the polyline distance method [33]. This is based on computing the average distance between each point on one profile (i.e., LI or MA) and the distance resulting from the normal projection to the segment of the other profile (i.e., LI or MA, respectively) [33]. A more complete description of the polyline distance is reported in the Supplementary Materials (Figure S3). For the statistical analyses reported in subsections 2.4.1 and 2.4.2, two separate validations were done: first, each method was compared independently with the gold standard reference (i.e., A1) only on the common support that was found between the two considered segmentations. Second, a global common support was determined for all considered methods (i.e., three manual segmentations and seven computerized segmentations). More detailed information on how the common support was determined is reported in the Supplementary Materials.

Finally, the influence of interpolating the LI and MA profiles, the distance metric used to compute the CIMT, and whether the CIMT is computed on the same portion of the image (i.e., if a common support is used or not) was determined. To do so, a comparison was done between the CIMT measurements obtained when: (i) interpolating or not interpolating the profiles, (ii) when using the polyline distance or the Euclidean distance, and (iii) when using a common support or not.

2.4. Statistical analysis

All statistical analyses were done employing R 4.1.2 software, on a laptop with an i5 4-core CPU and 16 GB of RAM. Regression plots and Bland Altman plots were obtained using MATLAB 2020b on the same machine.

2.4.1. CIMT bias and Hausdorff distance and point-wise bias

The signed difference (bias) and absolute bias of the CIMT values against the manual A1 values were computed as follows:

$$\mathrm{CIMT}_{\mathrm{bias}}=\mathrm{CIMT}_{\mathrm{method}}-\mathrm{CIMT}_{\mathrm{A1}} \tag{1}$$

$$\mathrm{CIMT}_{\mathrm{AbsBias}}=|\mathrm{CIMT}_{\mathrm{method}}-\mathrm{CIMT}_{\mathrm{A1}}| \tag{2}$$

where *method* refers to either the reproducibility manual CIMT values (i.e., A2, A1’) or the automatic CIMT measurements (i.e., $\mathrm{UCY}_{\mathrm{CY}}$, $\mathrm{TUM}_{\mathrm{DE}}$, $\mathrm{CNR}_{\mathrm{IT}}$, $\mathrm{INESCTEC}_{\mathrm{PT}}$, $\mathrm{POLITO}^{\mathrm{Snakes}}_{\mathrm{IT}}$, $\mathrm{CREATIS}_{\mathrm{FR}}$, $\mathrm{POLITO}^{\mathrm{UNet}}_{\mathrm{IT}}$). A Wilcoxon paired signed rank sum test was also done to test for statistically significant differences between the CIMT measurements and the reference measure at $p \leq 0.05$ and non-statistical significant differences at $p > 0.05$.

A further statistical analysis was done separating the LI and MA profiles. In particular, the Hausdorff distance [33] between the ground truth LI or MA reference profile and the analyzed method LI or MA profile was computed (referred to as HDM_{LI} and HDM_{MA}, respectively). The Hausdorff distance was considered as it gives an overall idea of the worst-case scenario in terms of distance between two profiles [58–60]. For a visual representation of how the Hausdorff distance is computed, please see the Supplementary Materials (Figure S4). The Hausdorff distance gives an estimate of the largest distance between two profiles but doesn't determine if one profile is more consistently inside or outside of the other. To mitigate this, we analyzed each profile point-by-point and computed the point-wise bias. The results are reported in the Supplementary Material (Table S5).

2.4.2. Analysis of inter and intra-observer variability

As briefly discussed in the Introduction section, a certain level of discrepancy is expected between different manual tracings and measurements. This is demonstrated when two different analysts perform measurements on the same dataset (inter-analyst variability) or the same analyst performs the measurements at two different time points (intra-analyst variability) on the same dataset. Errors are unavoidable when generating gold standard references via manual measurements and no ground truth can be obtained when using in-vivo ultrasound data. It is however insightful to quantify the agreement between different expert analysts, and to then compare it with computerized methods. In order to assess this variability, the difference of CIMT values between A1 and A2 as well as between A1 and A1' was computed. Then, a regression analysis and a Bland-Altman analysis were performed. For comparison purposes, the regression and Bland-Altman analyses were also performed comparing the computerized methods CIMT values with the reference A1 CIMT values.

Moreover, to partially mitigate the issue of not having ground truth data when considering in-vivo ultrasound data, in this study we simulated 100 images using the FAMUS software [35], as described previously. The CIMT values obtained using all three manual and seven computerized techniques were compared with the a-priori known CIMT values using the LI and MA profiles employed to simulate the images, which represent the ground truth boundary profiles. Specifically, a regression and Bland-Altman analysis were performed.

2.4.3. Division of the dataset

In order to provide general guidelines and recommendations, the original database was divided in three different ways: by carotid morphology, image quality, and research center. First, the 500 images were divided by visually inspecting the carotid morphology into five different classes: (i) inclined upward, (ii) inclined downward, (iii) straight, (iv) curved upward or (v) curved downward. Secondly, all images except the 100 FAMUS images, were divided by taking into consideration the signal-to-noise ratio (SNR) of each image. For the i-th image, the region between the LI and MA profiles defined by A1 was considered the one containing the signal. The area for measuring the noise level was defined as the previously defined region shifted upwards towards the lumen by N pixels ($N = 1.5 \times cIMT^{i\text{-}th}_{A1}$). SNR was then calculated as

$$SNR_i = 10log_{10}\frac{P_{Signal}}{P_{Noise}}$$

Three classes were defined:

Low SNR: $SNR_i \leq \mu_{SNR} - \sigma_{SNR}$

Average SNR: $\mu_{SNR} - \sigma_{SNR} < SNR_i < \mu_{SNR} + \sigma_{SNR}$

High SNR: $SNR_i \geq \mu_{SNR} + \sigma_{SNR}$

where SNR_i represents the SNR of the i-th considered image; μ_{SNR} is the average SNR value of all images within the dataset; σ_{SNR} is the standard deviation of all SNR image values. The FAMUS images were excluded from the SNR division to avoid bias in the average SNR estimation. This choice was made since the FAMUS images showed a SNR of 74.66 ± 26.05 dB, while the images acquired with the US devices had an SNR of 33.53 ± 18.55 dB. A two-sample t-test result between the two SNR distributions confirmed the statistically significant difference ($p<0.001$). The division of the dataset in these two ways allows for technical considerations to be made about when and why certain methods may tend to fail segmentation or produce larger bias values. Finally, the database was also divided by center, so as to discuss the segmentation results obtained using different ultrasound devices and imaging settings.

The results obtained by dividing the dataset are presented as box plots, where limits of the boxes correspond to the 1st and 3rd quartile whereas the top and bottom whiskers end at 1.5 times above or below, respectively, the interquartile range (±IQR) which is defined as the difference between the 3rd and 1st quartile. To test if dataset division by morphology, SNR, and center has a significant effect on the CIMT absolute bias, we performed a Kruskal-Wallis (KW) non-parametric test. If not conversely specified, for these tests we considered the CIMT absolute biases from both manual and computerized operators together. Then we compared the median of each subset (e.g., Low, Average, and High SNR for SNR division, or straight, inclined up, inclined down, curved up, curved down for morphology classification) using KW test.

3. Results

3.1. Overall segmentation results

Table 2 shows the CIMT bias results for each manual and computerized segmentation method along with the number of unprocessed images (see second column of Table 2). The results take into consideration the individual common support computed between the ground truth (i.e., A1) and the analyzed method. The four computerized methods that were able to process 100% of the images were the ones based on deep learning (that is, $\text{CREATIS}_{\text{FR}}$ and $\text{POLITO}^{\text{UNET}}_{\text{IT}}$) and those based on deformable models (that is, $\text{POLITO}^{\text{Snakes}}_{\text{IT,}}$ and UCY_{CY}). The $\text{INESCTEC}_{\text{PT}}$ method was unable to process 113 images, whereas the other computerized methods ranged from not processing one (TUM_{DE}) to four (CNR_{IT}) images.

Table 2. Segmentation results (mean±std) for all images investigated (N=500), for all the manual and the automated segmentation methods investigated in this study. The methods are compared to expert analyst 1 (A1) on individual common supports.

Method	Number of unprocessed images	CIMT (µm)	$\text{CIMT}_{\text{bias}}$ (µm)	$\text{CIMT}_{\text{AbsBias}}$ (µm)	HDM_{LI} (µm)	HDM_{MA} (µm)
A1	0	857 ± 255	-	-	-	-
A1’	0	922 ± 285	70 ± 201	160 ± 140	352 ± 140	346 ± 185
A2	0	895 ± 300	46 ± 259	194 ± 177	327 ± 138	338 ± 184
CNR_{IT} [6,11]	4	774 ± 205	-69 ± 169*	139 ± 119	**297 ± 120**	341 ± 144
$\text{INESCTEC}_{\text{PT}}$ [12,43]	113	865 ± 360	31 ± 297*	150 ± 258	874 ± 986	408 ± 545
TUM_{DE} [19,44]	1	750 ± 215	-96 ± 159	143 ± 118	311 ± 148	327 ± 148
UCY_{CY} [14]	0	765 ± 197	-82 ± 177	139 ± 136	309 ± 134	313 ± 126
$\text{POLITO}^{\text{Snakes}}_{\text{IT}}$ [15]	0	849 ± 164	-11 ± 286*	224 ± 178	544 ± 294	536 ± 333
$\text{POLITO}^{\text{UNet}}_{\text{IT}}$	0	1011 ± 213	156 ± 147	178 ± 120	328 ± 189	325 ± 219
$\text{CREATIS}_{\text{FR}}$ [45]	0	861 ± 224	**4 ± 138***	**106 ± 89**	305 ± 197	**289 ± 147**

A1: Manual segmentations of analyst 1; A1’: manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR_{IT}: Method from Consiglio Nazionale delle Ricerche; $\text{INESCTEC}_{\text{PT}}$: INESC Technology and Science; TUM_{DE}: Technische Universität München; UCY_{CY}: University of Cyprus; $\text{POLITO}^{\text{Snakes}}_{\text{IT}}$: Politecnico di Torino (dual snakes); $\text{POLITO}^{\text{UNet}}_{\text{IT}}$: Politecnico di Torino (UNet); $\text{CREATIS}_{\text{FR}}$: CREATIS lab Université de Lyon. CIMT: carotid intima-media thickness. HDM: Hausdorff distance metric. In $\text{CIMT}_{\text{bias}}$, the asterisk (*) defines measurements not statistically significantly different ($p >= 0.05$) using the Wilcoxon rank sum parametric paired test with respect to A1. Bolded values indicate best performance.

In terms of CIMT absolute bias, three computerized methods showed both a lower mean and a lower standard deviation when compared to both the manual intra/inter-analyst results. (A1’: 160 ± 140 µm, A2: 194 ± 177 µm). More specifically, the $\text{CREATIS}_{\text{FR}}$ method showed the lowest absolute bias (106 ± 89 µm), followed by CNR_{IT} (139 ± 119 µm), UCY_{CY} (139 ± 136 µm), and then TUM_{DE} (143 ± 118 µm). The $\text{INESCTEC}_{\text{PT}}$ (150 ± 258 µm) computerized method showed a low mean CIMT bias but a rather high standard deviation, and the bias of the $\text{POLITO}^{\text{UNet}}_{\text{IT}}$ (178 ± 120 µm) computerized method was higher than the intra-observer results but lower than the inter-observer results. Only the $\text{POLITO}^{\text{Snakes}}_{\text{IT}}$ (224 ± 178 µm) computerized method showed both a higher mean and standard deviation than the manual results. ~~The best results, based on the smallest bias, are reported with bold values in Table 2. The asterisk in Table 2 highlights measurements that are not statistically significantly different (p >= 0.05) using the Wilcoxon rank sum parametric paired test with respect to A1.~~ From Table 2, it can be observed how the majority of the computerized methods along with both manual segmentations produce a similar Hausdorff distance when considering the LI and MA boundaries (i.e., HDM_{LI} and HDM_{MA}). Two computerized methods show a higher Hausdorff distance between the computerized and manual A1 LI profiles ($INESCTEC_{PT}$ and $CREATIS_{FR}$) when compared to the Hausdorff distance between the MA profiles. The $INESCTEC_{PT}$ method is much more sensitive to noise in the lumen, showing a HDM_{LI} over twice as high as HDM_{MA}. The $CREATIS_{FR}$ method, on the other hand, shows the lowest HDM_{MA} results, differing by only approximately 16 µm when compared to HDM_{LI}. The CNR_{IT} method shows the lowest HDM_{LI} results overall but has about a 45 µm difference compared to the HDM_{MA} results. Figure 2 provides segmentation results on an example image. More segmentation results are included in the Supplementary Materials, showing the overall common support (Figure S5) and an example FAMUS image (Figure S6). Moreover, the fold results of both deep learning methods are shown in the Supplementary Material.

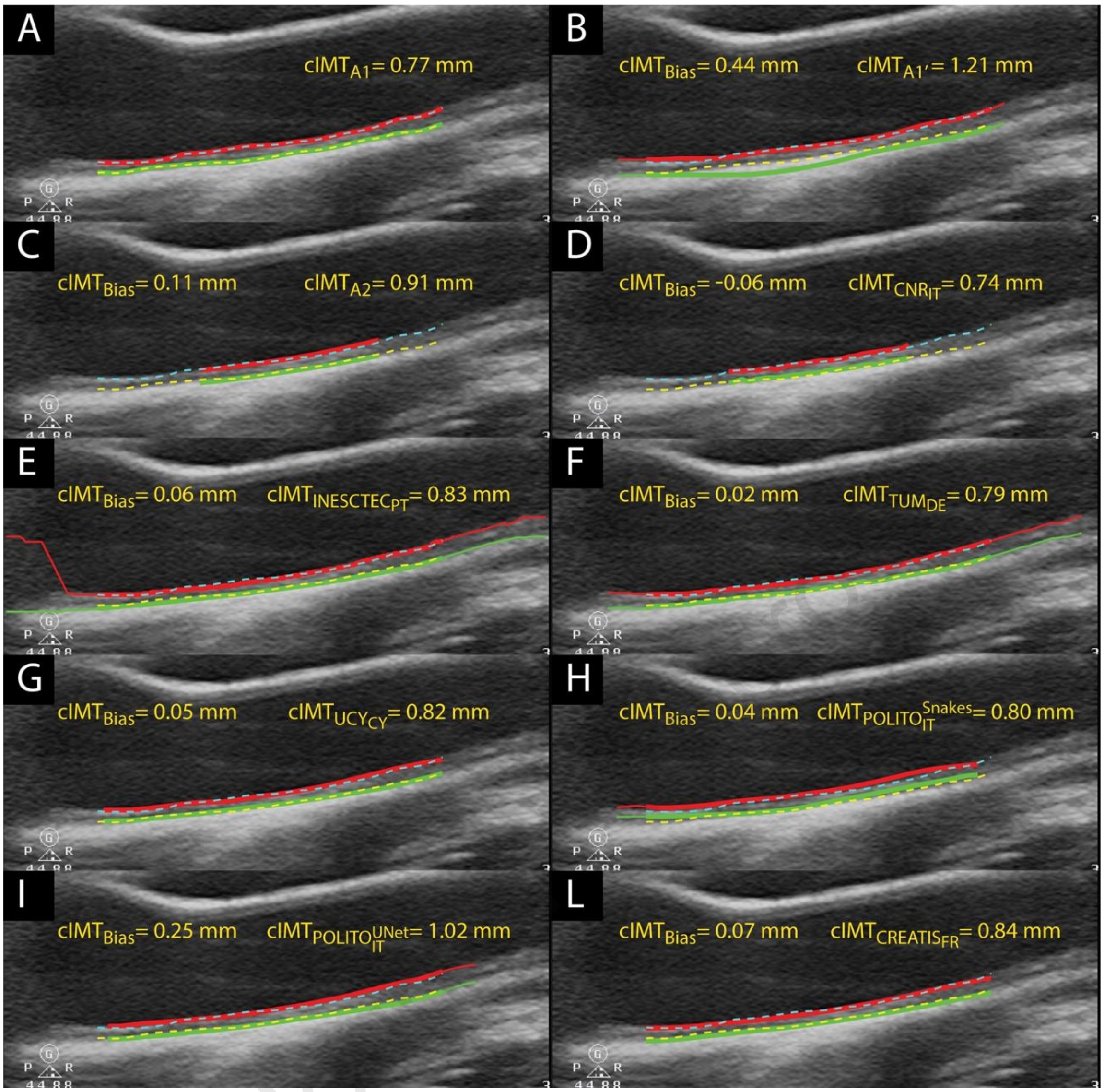


Figure 2. Segmentation results on an example image. In each panel, the dotted blue and yellow profiles represent the manual A1segmentations of the LI and MA boundaries, respectively. The solid red and green profiles represent on the other hand the LI and MA boundaries, respectively, of the compared method (A1', A2, or a computerized method). The red and green profiles are shown in bold where the individual common support is found and where the CIMT bias results are computed (i.e., where both the examined method and A1 provides a segmentation). The CIMT values and CIMT bias results are displayed in yellow on each panel. (A) A1; (B) A1'; (C) A2; (D) CNR$_{IT}$; (E) INESCTEC$_{PT}$; (F) TUM$_{DE}$; (G) UCY$_{CY}$; (H) POLITO$^{Snakes}_{IT}$; (I) POLITO$^{Unet}_{IT}$; (L) CREATIS$_{FR}$.

A1: Manual segmentations of analyst 1; A1': manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon.

Our findings when interpolating or not interpolating the profiles, when using two different distance metrics (i.e., polyline distance (PDM) and Euclidean distance (EDM)), and when using a common support (CS) or not to compute the CIMT are shown in Table 3, where it can be observed how the CIMT absolute bias results can change in an important way depending on the distance metric used, if a common support is used, and if the profiles are interpolated if needed. The $\text{CREATIS}_{\text{FR}}$ method showed the best results for all cases, except for when the Euclidean distance is used to compute the CIMT, there is no profile interpolation and a common support is defined, where the TUM_{DE} and $\text{POLITO}^{\text{Unet}}_{\text{IT}}$ methods tied for the lowest CIMT absolute bias.

Table 3. Segmentation bias results $\text{CIMT}_{\text{AbsBias}}$ (mean±std) between the manual and the automated segmentation methods for all images investigated (N=500). The methods are comparing interpolation, distance metric, and common support with respect to expert analyst 1 (A1).

Method	$\text{CIMT}_{\text{AbsBias}}$ (μm)					
	PDM - interp - CS	PDM – no interp - CS	PDM – interp – no CS	EDM - interp – CS	EDM – no interp - CS	EDM – interp – no CS
A1’	160 ± 140	152 ± 131	159 ± 142	160 ± 140	193 ± 164	159 ± 145
A2	194 ± 177	179 ± 160	194 ± 175	194 ± 178	427 ± 385	194 ± 175
CNR_{IT} [6,11]	139 ± 119	134 ± 116	143 ± 123	139 ± 118	202 ± 146	141 ± 122
$\text{INESCTEC}_{\text{PT}}$ [12,43]	150 ± 258	146 ± 259	213 ± 326	145 ± 245	203 ± 236	202 ± 308
TUM_{DE} [19,44]	143 ± 118	132 ± 117	152 ± 129	142 ± 118	**114 ± 95**	152 ± 129
UCY_{CY} [14]	139 ± 136	140 ± 132	145 ± 139	139 ± 136	145 ± 132	145 ± 138
$\text{POLITO}^{\text{Snakes}}_{\text{IT}}$ [15]	224 ± 178	222 ± 173	213 ± 169	223 ± 177	238 ± 204	212 ± 169
$\text{POLITO}^{\text{UNet}}_{\text{IT}}$	178 ± 120	189 ± 119	175 ± 125	179 ± 120	**114 ± 95**	175 ± 125
$\text{CREATIS}_{\text{FR}}$ [45]	**106 ± 89**	**104 ± 84**	**106 ± 88**	**106 ± 89**	168 ± 135	**105 ± 88**

A1: Manual segmentations of analyst 1; A1’: manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR_{IT}: Method from Consiglio Nazionale delle Ricerche; $\text{INESCTEC}_{\text{PT}}$: INESC Technology and Science; TUM_{DE}: Technische Universität München; UCY_{CY}: University of Cyprus; $\text{POLITO}^{\text{Snakes}}_{\text{IT}}$: Politecnico di Torino (dual snakes); $\text{POLITO}^{\text{UNet}}_{\text{IT}}$: Politecnico di Torino (UNet); $\text{CREATIS}_{\text{FR}}$: CREATIS lab Université de Lyon. CIMT: carotid intima-media thickness. HDM: Hausdorff distance metric. EDM: Euclidean distance metric. Bolded values indicate best performance.

3.2. Inter/intra-observer variability compared to computerized methods

The first column of Figure 3 shows the regression analysis results for inter- and intra-observer variability. The Spearman correlation coefficient, ρ, between A1 and A1’ was $\rho=0.6996$ (see also Fig. 3a), whereas it decreased to $\rho=0.5942$ (see also Fig. 3b), between A1 and A2. The second column of Figure 3 shows the Bland-Altman analysis, where the CIMT bias and 95% confidence intervals can be appreciated. While the average bias between the two measurements (y-axis value of the middle horizontal line) does not exceed 0.07 mm in either cases, the bias was found to be statistically different from zero ($p < 0.05$ – see also Table 2).

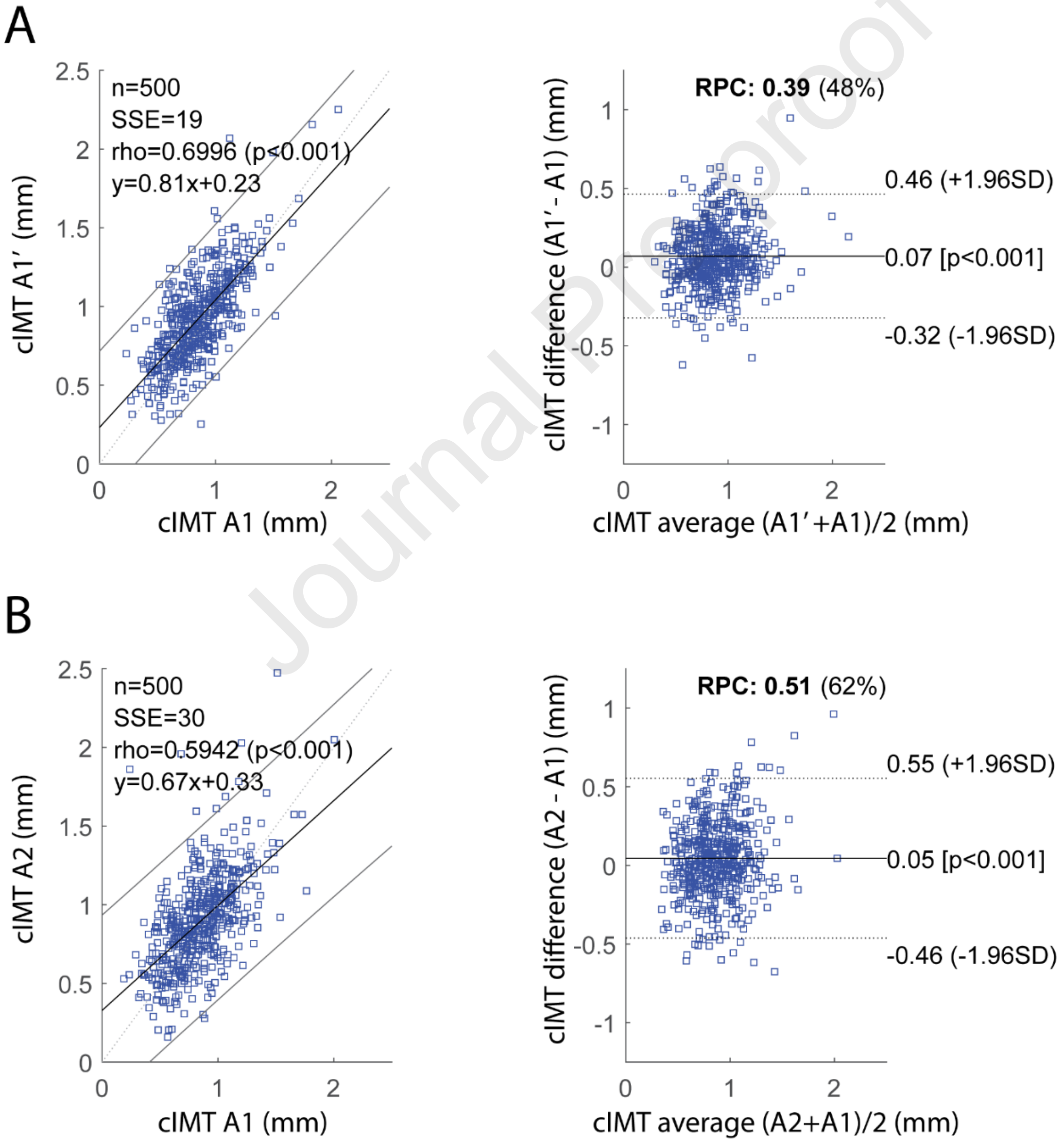


Figure 3. Intra- (top row) and inter- (bottom row) analyst- variability analysis for all manual CIMT segmentation measurements (N=500) performed in this study. First column: correlation analysis. Second column: Bland-Altman analysis.

A1: Manual segmentations of analyst 1; A1': Manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; n: Number of images; SSE: Sum of squared error; rho: Spearman correlation rho value; SD: Standard deviation of the segmentation measurements differences; RPC: Reproducibility coefficient (1.96×SD).

Regression and Bland-Altman analysis for all the computerized methods against A1 CIMT measurements are reported in Fig. 4 and Fig. 5 respectively. Briefly, considering the regression analysis, the computerized methods can be divided into three categories based on the Spearman correlation coefficient, ρ, as follows:

- $\rho \geq 0.8$: CREATIS$_{FR}$ (ρ=0.824) and POLITO$^{UNet}_{IT}$ (ρ=0.813)
- $0.75 < \rho < 0.8$: TUM$_{DE}$ (ρ=0.789), UCY$_{CY}$ (ρ=0.747), and CNR$_{IT}$ (ρ=0.775)
- $\rho \leq 0.75$: INESCTEC$_{PT}$ (ρ=0.617) and POLITO$^{Snakes}_{IT}$ (ρ=0.187).

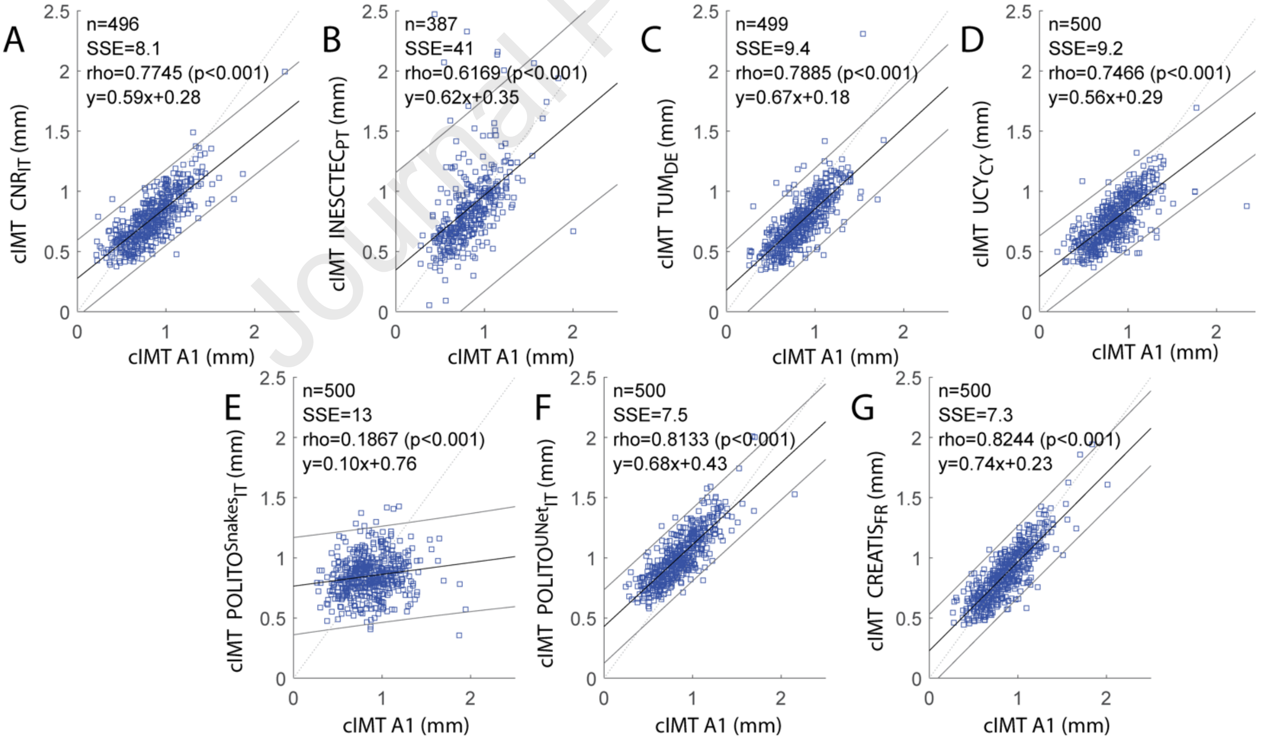


Figure 4. CIMT regression and correlation analysis of the computerized techniques compared to the manual A1 segmentations (n=500).
A1: Manual segmentations of analyst 1; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes);

POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon. n: Number of images; SSE: Sum of squared error; rho: Spearman rho value.

Regarding the Bland-Altman analysis (Fig. 5), we can see how only a few methods showed not statistically significant differences for the CIMT bias (INESCTEC$_{PT}$, POLITO$^{Snakes}_{IT}$ and CREATIS$_{FR}$), but two of these three methods (INESCTEC$_{PT}$ and POLITO$^{Snakes}_{IT}$) are also the ones that present the highest standard deviation, in accordance with Table 2. None of the methods show a visible trend.

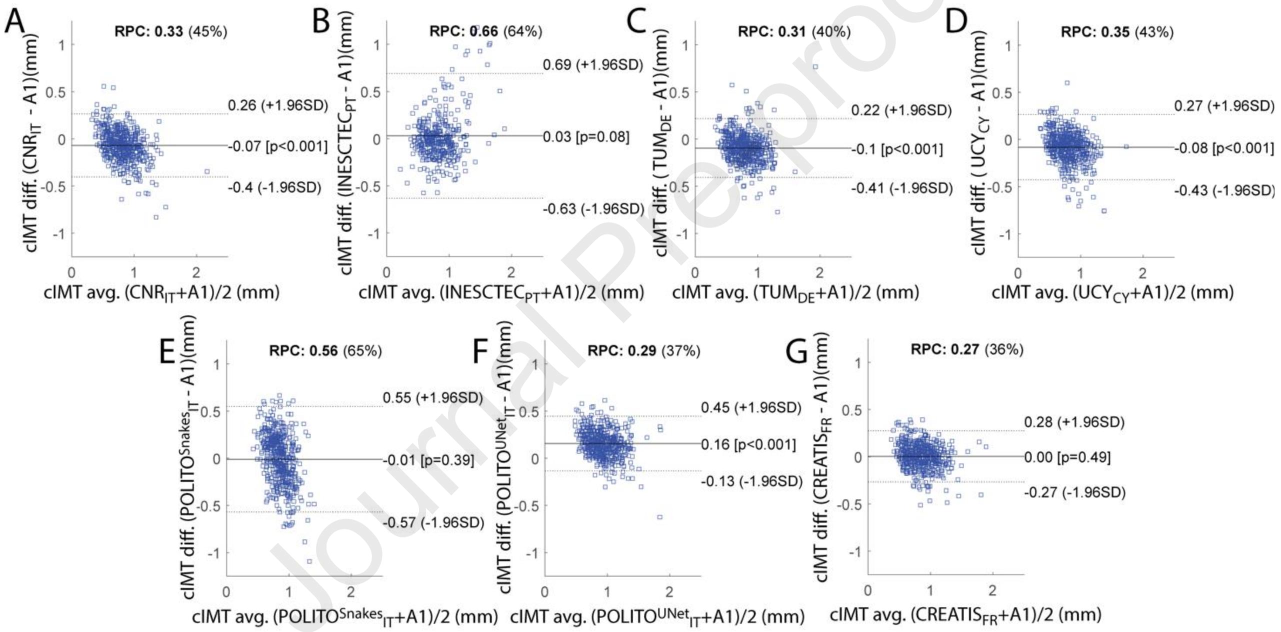


Figure 5. CIMT Bland-Altman analysis of the computerized segmentation techniques compared to the manual A1 segmentations. A1: Manual segmentations of analyst 1; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; SD: Standard deviation of the differences; RPC: Reproducibility coefficient (1.96×SD).

3.3. Division by morphology, SNR, and center

For the in-depth analysis of segmentation results dividing the database by morphology, SNR, and center, the overall common support between all ten considered methods (i.e., three manual and seven computerized) was first found and all results reported refer to those obtained on this overall common support. First of all, 22 images (4.4%) had to be discarded because the common support was found to be less than 1 mm between all methods (see Supplementary Materials). On the remaining 478 images, the LI and MA widths were found to be equal to 8.85 ± 4.09 mm and 9.02 ± 4.03 mm, respectively. Table 4 shows the overall division of the dataset between the various groups that were analyzed. The FAMUS simulated images were not included in this table since they were simulated to be equally distributed in morphology and all with a high SNR.

Table 4. Division of the dataset between groups that were analyzed (center, SNR, morphology)

			Morphology				
Center			**Curved down**	**Curved up**	**Inclined down**	**Inclined up**	**Straight**
Munich	**SNR**	**Average SNR**	4	4	15	25	17
		High SNR	0	0	0	0	0
		Low SNR	2	4	9	12	8
Pisa		**Average SNR**	2	2	13	9	60
		High SNR	0	1	0	0	2
		Low SNR	0	0	1	1	9
Porto		**Average SNR**	6	1	6	18	26
		High SNR	3	1	4	10	24
		Low SNR	0	0	0	1	0
Torino		**Average SNR**	6	3	34	15	29
		High SNR	1	1	4	5	1
		Low SNR	0	0	0	1	0

SNR: Signal to noise ratio

3.3.1. Morphology analysis results

The number of images that were classified by morphology is as follows: straight (n=199, 39.8%), inclined up (n=117, 23.4%), inclined down (n=106, 21.2%), curved up (n=41, 8.2%), curved down (n=37, 7.4%). Figure 6 illustrates an overall box plot of the CIMT absolute bias results for all different methods investigated, when dividing the database by morphology. From the Kruskal-Wallis statistical test, it was found that the inclined-down morphology shows a median CIMT absolute bias that is higher than the other geometries (p<0.001). The complete table with all segmentation results (i.e., CIMT bias, CIMT absolute bias, HDM_{LI}, and HDM_{MA}) can be found in the Supplementary Materials (Table S6).

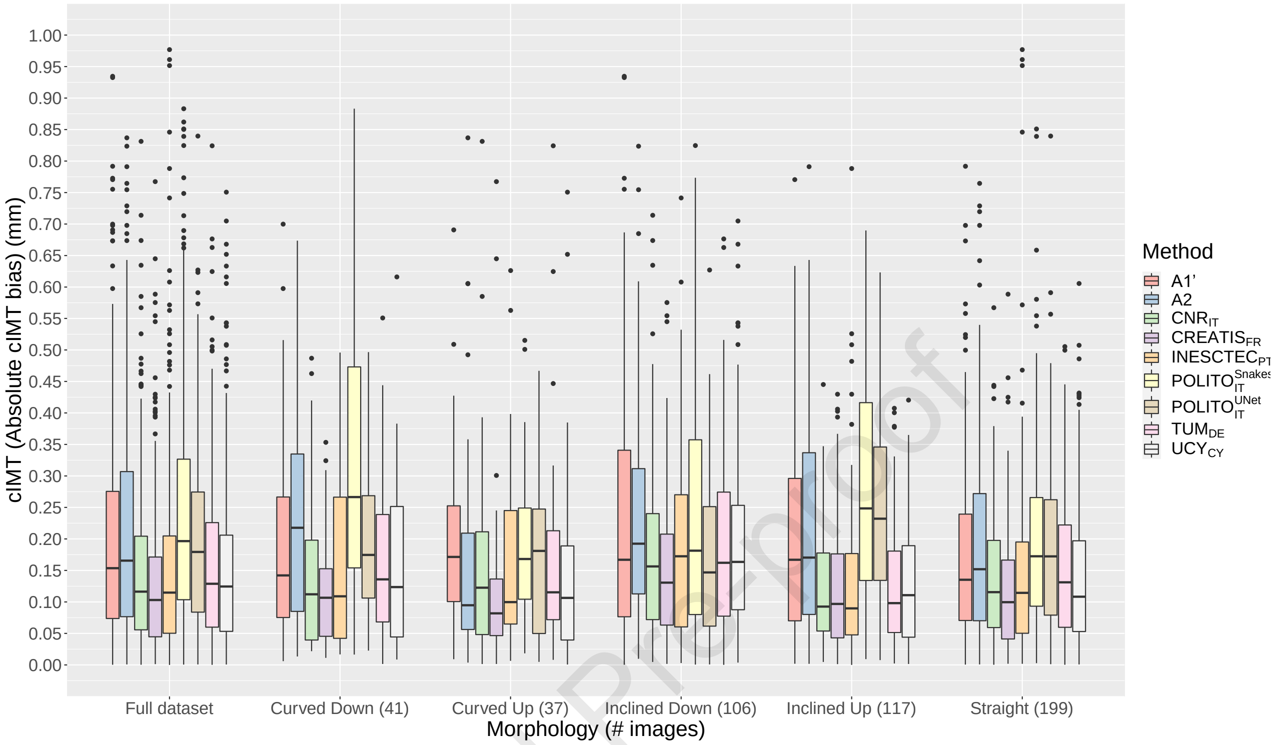


Figure 6. Box plot of the absolute CIMT biases compared to the expert analyst A1 segmentations when dividing the database by carotid morphology. A1': Manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; A1: Manual segmentations of analyst 1; CNR_{IT}: Method from Consiglio Nazionale delle Ricerche; $INESCTEC_{PT}$: INESC Technology and Science; TUM_{DE}: Technische Universität München; UCY_{CY}: University of Cyprus; $POLITO^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); $POLITO^{UNet}_{IT}$: Politecnico di Torino (UNet); $CREATIS_{FR}$: CREATIS lab Université de Lyon.

3.3.2. Signal-to-noise analysis results

The number of images that were classified by SNR were as follows: low SNR (n=48, 12.00%), average SNR (n=295, 73.75%), high SNR (n=57, 14.25%). Figure 7 presents an overall box plot of the CIMT absolute bias results when dividing the database by SNR. According to the Kruskal-Wallis statistical test, both manual and computerized operators showed no statistical difference in CIMT absolute bias between the Low and Average SNR groups (p=0.31), whereas when comparing the CIMT absolute biases in the High SNR class with respect to the rest of the dataset, a significant difference is observed (p<0.001). The complete table with all segmentation results (i.e., CIMT bias, CIMT absolute bias, HDM_{LI}, and HDM_{MA}) is found in the Supplementary Materials (Table S7).

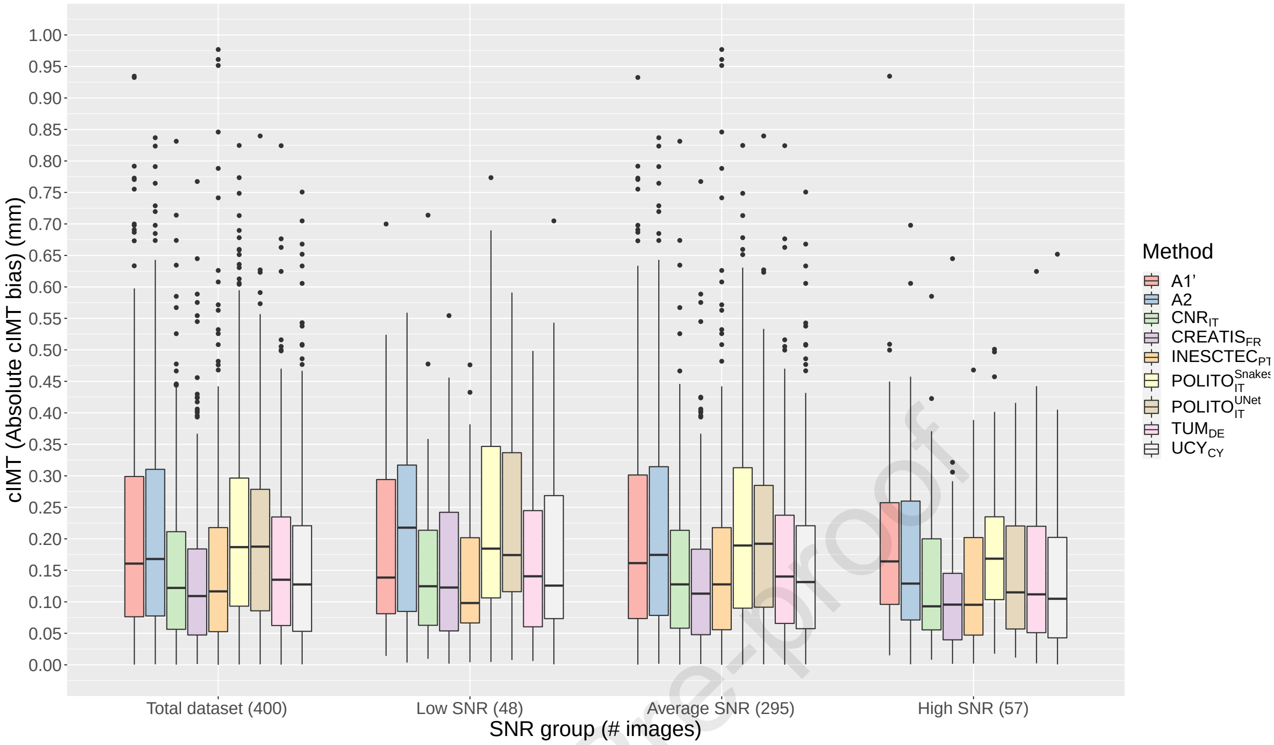


Figure 7. Box plots of the absolute CIMT biases compared to the expert analyst A1 segmentations when dividing the database by signal to noise (SNR). A1: Manual segmentations of analyst 1; A1': manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon.

3.3.3. Acquisition center analysis results

Finally, the dataset was divided by acquisition center. Figure 8 shows an overall box plot of the CIMT absolute bias results when dividing the database in this way. According to the Kruskal-Wallis statistical test, when comparing the CIMT absolute bias for all the segmentation methods, the dataset from the Torino center shows a significant difference ($p=0.039$) with respect to the rest of the dataset. The complete table with all segmentation results (i.e., CIMT bias, CIMT absolute bias, HDM$_{LI}$, and HDM$_{MA}$) is found in the Supplementary Materials (Table S8).

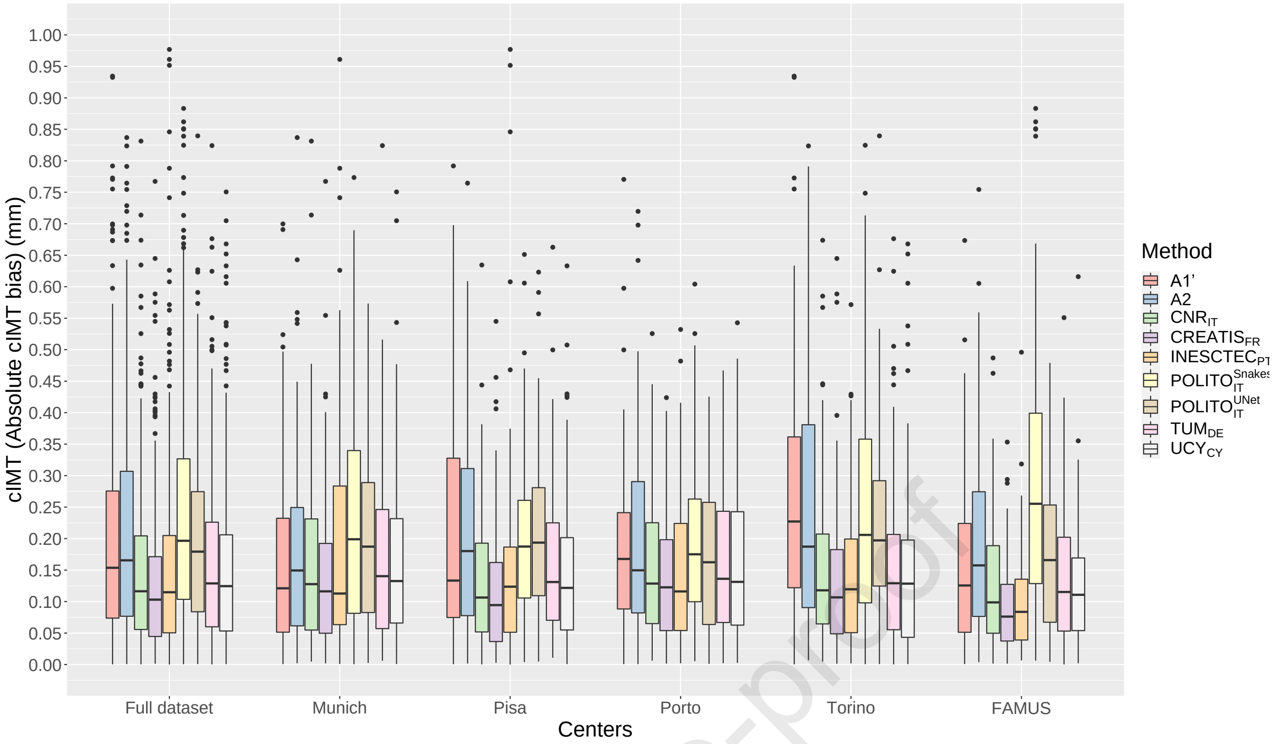


Figure 8. Box plots of the absolute CIMT biases compared to the expert analyst A1 segmentations when dividing the database by centers. A1: Manual segmentations of analyst 1; A1': manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon.

3.4 Simulation and ground truth results

When considering the 100 images that were simulated using the FAMUS software with a-priori known CIMT values, Figures 9 and 10 show the regression and Bland-Altman analysis, respectively. The Spearman correlation coefficients for the manual tracings were equal to 0.721, 0.703, and 0.349 for A1, A1' and A2, respectively. The computerized methods overall showed a higher correlation than the manual methods with Spearman correlation coefficients greater than 0.845 in all cases, with the exception of POLITO$^{Snakes}_{IT}$ ($\rho = 0.453$). The Bland-Altman plots demonstrate how the majority of the computerized methods present much better limits of agreement with the a-priori known CIMT values, showing a smaller standard deviation of the obtained biases.

The complete table with all segmentation results (i.e., CIMT bias, CIMT absolute bias, HDM$_{LI}$, and HDM$_{MA}$) is found in the Supplementary Materials (Table S9).

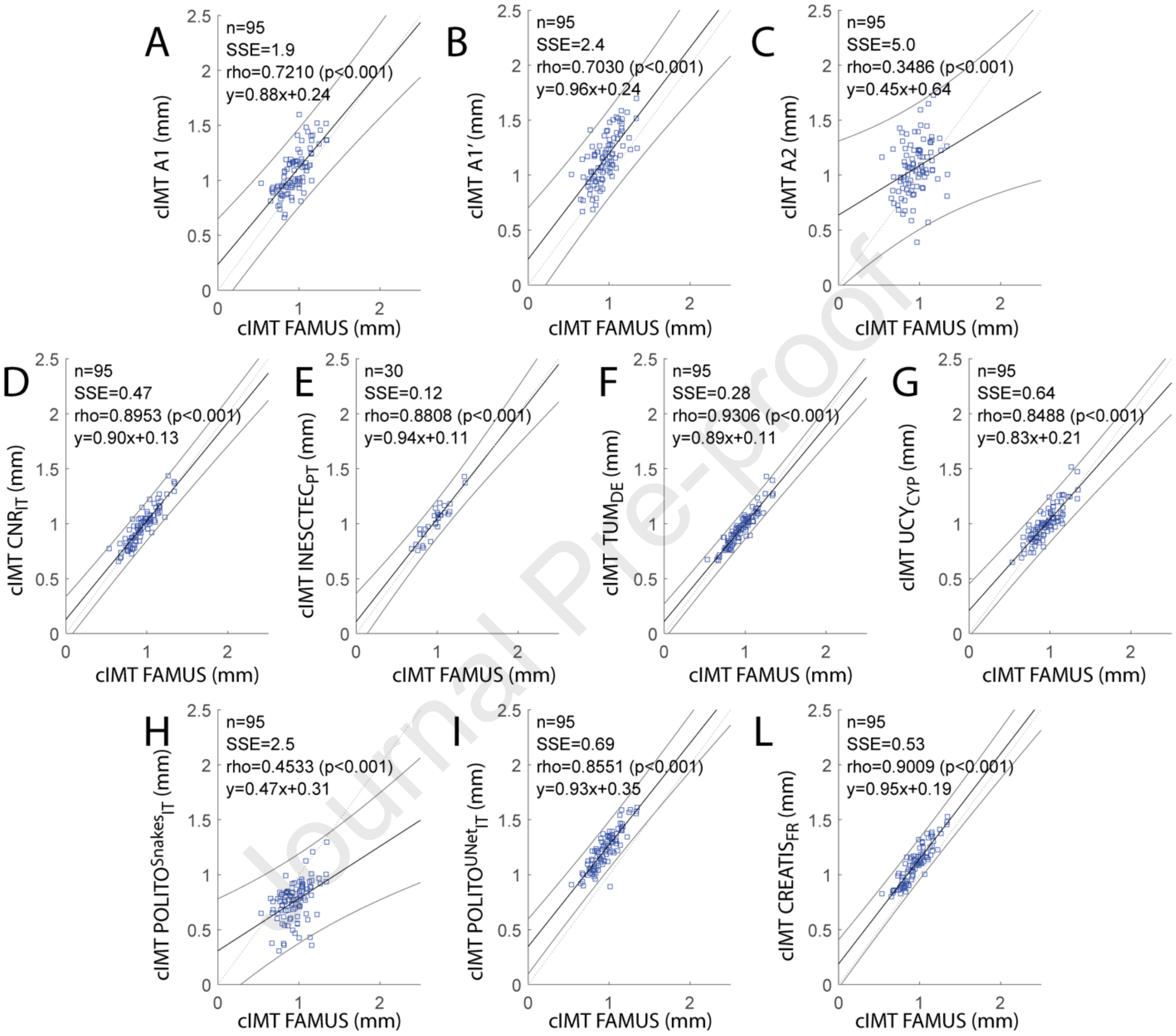


Figure 9. CIMT correlation analysis for all the cIMT segmentation methods investigated in this study compared to the a-priori known FAMUS simulated image values. A1: Manual segmentations of analyst 1; A1': manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; n: Number of images; SSE: Sum of squared error; rho: Spearman rho value.

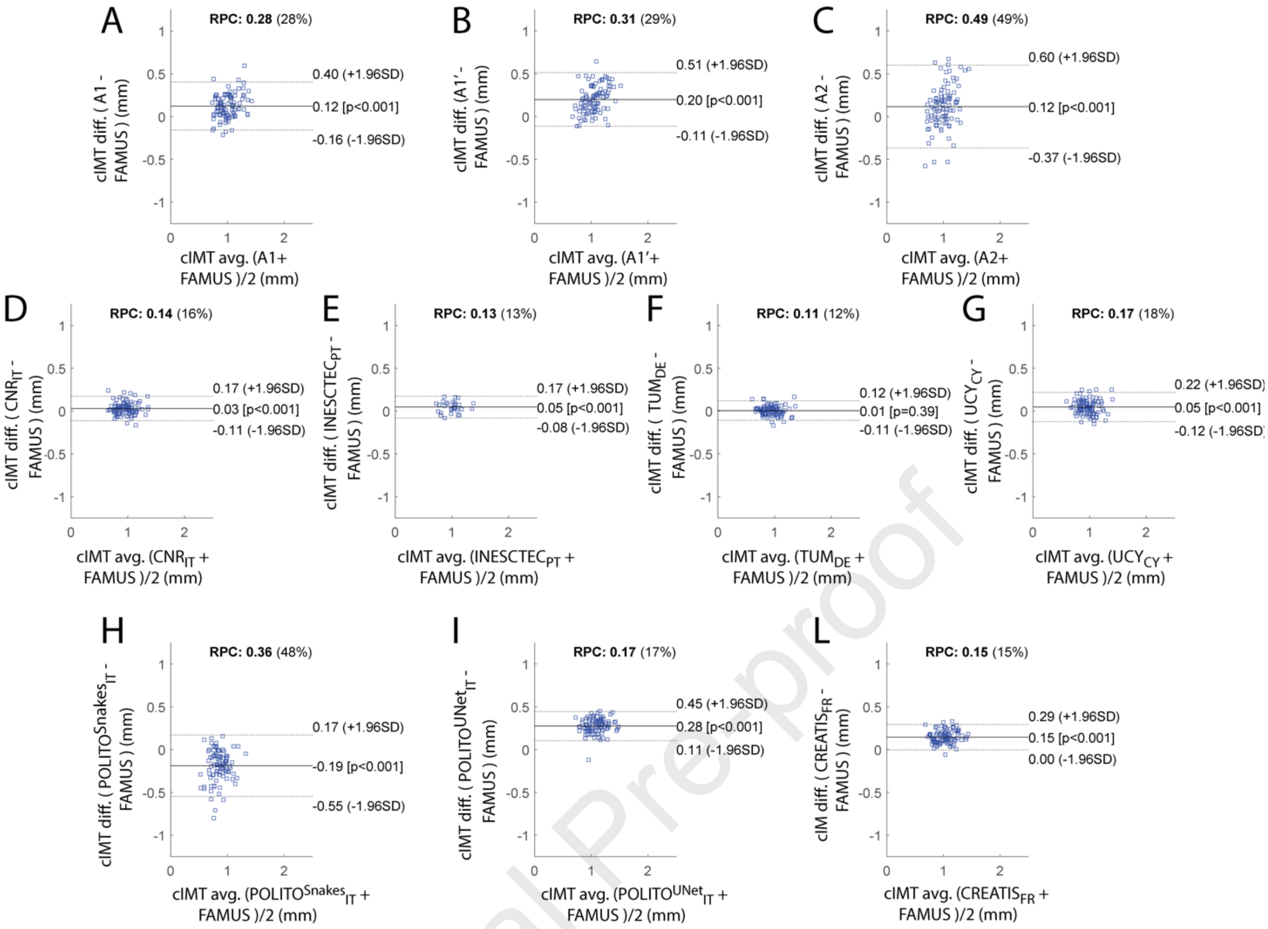


Figure 10. CIMT Bland-Altman analysis of the all methods compared to the FAMUS a-priori known simulated image values. A1: Manual segmentations of analyst 1; A1': manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; SD: standard deviation of the differences; RPC: reproducibility coefficient (1.96×SD).

4. Discussion

4.1 Study findings on analyzed computerized methods

4.1.1. Inter-/Intra- analyst variability compared to computerized methods

This study demonstrated that CIMT measurement variability is lower between computerized segmentations than skilled analysts' manual segmentations. The inter- and intra-analyst reproducibility of manual CIMT measurements often depends on analyst effort and concentration, representing an obstacle especially when considering large databases and cohort studies. In fact, the computerized methods systematically showed comparable and often better CIMT bias results than the inter- and intra-analyst variability. Hence, the results presented here are very promising for computerized methods, showing how they can be an important asset for CIMT measurement and suggesting that the computerized methods measurements are as accurate as the measurement obtained using a gold standard skilled analyst's manual segmentation. Importantly, when considering both the comparison with the manual A1 profiles and with the a-priori known simulated images, the majority of the computerized methods demonstrated better results specifically in terms of robustness when compared to the manual segmentations, showing lower standard deviations of the CIMT bias distribution and hence better limits of agreement. This can be considered the real added value of computerized methods, where a potential fixed bias can be theoretically corrected but the distribution of the bias cannot be corrected.

4.1.2. Computerized methods comparison

When comparing the computerized methods between each other, it can be observed how the deep learning-based $\text{CREATIS}_{\text{FR}}$ method showed promising results on the entire dataset on all morphologies and SNR values, with a slight reduction in performance for the simulated images when compared with the a-priori known values. However, the other methods based on traditional image processing methods, in particular the TUM_{DE}, CNR_{IT}, and UCY_{CY} methods, also demonstrated good performances. In fact, they typically presented slightly higher CIMT bias values when compared to $\text{CREATIS}_{\text{FR}}$, but also showed lower values when compared with the a-priori known values on the simulated images. The $\text{INESCTEC}_{\text{PT}}$ method was unable to process a large number of images and also tended to show a lower robustness when compared to the others, presenting a higher standard deviation of the CIMT bias distribution. The $\text{POLITO}^{\text{Unet}}_{\text{IT}}$ method showed typically robust results but was generally affected by a fixed bias, tending to overestimate the CIMT value and was sensitive to carotid geometry. Finally, the $\text{POLITO}^{\text{Snakes}}_{\text{IT}}$ method showed the highest absolute CIMT value, was sensitive to carotid geometry and had difficulty segmenting the simulated images. When considering the FAMUS simulated images and comparing the a-priori known CIMT results of the simulated images with the manual and computerized methods, a first observation is that the INESCTEC$_{PT}$ method was unable to process 70% of the images, which explains the obviously lower number of points in both Figure 9 and Figure 10 for this method. Interestingly, all computerized methods showed a higher correlation than the manual methods except for the POLITO$^{Snakes}_{IT}$ algorithm. Moreover, by observing the Bland-Altman plots we can see that four computerized methods presented a lower average CIMT bias (solid horizontal line on Bland-Altman plots) than the manual methods: CNR$_{IT}$, INESCTEC$_{PT}$, TUM$_{DE}$, and UCY$_{IT}$. The other three methods showed a rather large fixed bias: the POLITO$^{Unet}_{IT}$ and CREATIS$_{FR}$ methods both tended to overestimate the CIMT, whereas the POLITO$^{Snakes}_{IT}$ method underestimated it. It is interesting how both deep learning methods presented an important bias – this could be due to two factors: first, these methods were trained to mimic the reference annotation from A1, and therefore their results correlate well with A1 who, in turn, is approximately as far away from the a-priori ground truth values as these CNN-based methods; second, this bias could also be partially due to the fact that the simulated images are still rather visually different when compared to the clinical images which were much more represented, hence limiting how much the network can learn about the simulated images during the training phase. Another interesting observation from Figure 10 is how the majority of the computerized methods present much better limits of agreement with the a-priori known CIMT values, showing a smaller standard deviation of the obtained biases (i.e., the dashed horizontal lines enclose a smaller range of bias values). In fact, only the POLITO$^{Snakes}_{IT}$ method shows a 95% confidence interval larger than the A1, A1’ and A2 manual methods.

Moreover, the segmentation errors found in this study, and particularly by the CREATIS$_{FR}$ method (CIMT$_{bias}$ = 4 ± 138 µm, correlation coefficient = 0.82), are comparable with or better than computerized techniques found in literature that have been employed on other databases. More specifically, in the study by del Mar Vila et al. [25], the CIMT estimation showed a correlation coefficient of 0.81, and a CIMT mean bias of 20 µm when employing a DenseNet for segmentation.

Furthermore, in Biswas et al. [30] a CIMT error was found equal to 94 ± 64 μm with a two-stage deep learning technique on 250 images.

4.2. Considerations and recommendations

A main goal of this study is to not only compare the seven computerized methods, but also provide considerations and recommendations for the research community together with an open dataset and segmentation results, in the hopes that they can in the future lead to solid validation guidelines for CIMT computerized methods.

4.2.1. Image acquisition settings, CIMT measurement, and results presentation

4.2.1.1. Image acquisition settings

As can be observed in Figure 1, the images contained in this database vary greatly among each other. This can be due to the different imaging settings and presets and/or to potential filters (e.g., despeckling filters) that were applied on the image before being exported. Figure 8 shows the CIMT absolute bias results of the considered methods compared to the ground truth manual segmentation A1 when dividing the database by centers. From this figure it can be observed that the Torino, Italy center presents the highest absolute CIMT bias for the manual methods, showing how it may present some features that make a manual segmentation more difficult. In fact, comparing the absolute bias for all the operators from the Torino dataset to the rest of the dataset using two classes, shows a significant difference ($p=0.039$). However, it must be noted that with this division interactions between Center and Shape, and Center and SNR also showed a strong significant difference ($p<0.001$). In fact, as will be discussed in a further section, this difference may also be due mainly to the large amount of inclined-down images within the Torino center dataset, as can be seen in Table 4. To further demonstrate this, when considering all the centers (i.e., using 5 different classes (Munich, Pisa, Porto, Torino, FAMUS)), comparing the absolute bias does not show any statistically significant difference ($p=0.21$). Hence, all methods generally present similar performances when dividing the dataset by center. The biggest difference can be observed for the

POLITO$^{\text{Snakes}}_{\text{IT}}$ method that shows a high median absolute CIMT bias for the simulated images. When considering the in-vivo images, it can be therefore concluded that the basic image acquisition settings (e.g., TGC, filters) do not play a significant role in the computerized methods' performance results.

4.2.1.2. CIMT measurement

An often-overlooked issue when comparing methods for CIMT measurement is how the actual measurements are obtained, whether comparing a manual segmentation with a computerized segmentation or comparing two computerized segmentations. Hence, it is important to determine whether the measurements are obtained on the same section of the CCA (i.e., is a common support used), if the segmentation profiles are interpolated, and what distance metric is used to compute the distance between the LI and MA profiles. Our analysis of this issue is reported in Table 3, where we can see that if interpolation is used, there is not a noticeable difference between the polyline distance and the Euclidean distance. On the other hand, if interpolation is not used, the PDM values are comparable between each other whereas the Euclidean distance presents a noticeable difference, especially with the profiles that did not present a point for each column (i.e., A1', A2, CNR, UCY, INESCTEC). When no common support is used, it is possible to see how the absolute CIMT bias values can differ greatly for some methods, and not as much for others. This is mainly due to the fact that some computerized methods try to segment on a portion of the carotid artery as large as possible, whereas others are guided by an operator and so may be limited to only a certain portion. It is not a goal of this study to determine if it is always necessary to interpolate the segmentation profiles or which distance metric is ideal for CIMT measurement [33], but from these results we can underline some important considerations on how to compute the CIMT: it is recommended to measure the CIMT on the same portion of the CCA (i.e., use a common support), and to ensure consistency when comparing CIMT measurements between each other (e.g., a CIMT measurement obtained with the polyline distance with interpolated profiles should not be compared with a CIMT obtained with the Euclidean distance with non-interpolated profiles).

4.2.1.3. CIMT measurement results presentation

Another often overlooked issue is how to present the results of the obtained CIMT values. The CIMT bias is often reported along with a statistical test (typically a paired Student's t-test or paired Wilcoxon test) to determine if there is a statistically significant difference between the gold standard CIMT values (i.e., obtained with a manual segmentation) and the proposed computerized method's CIMT values. What can be observed from Table 2 is how the Wilcoxon test is actually not very representative of the segmentation results, which can be inferred from the definition of the statistical test. In fact, the paired Wilcoxon test determines if the distribution of the difference can be determined as equal to zero, but gives no information on how wide the distribution is and does not take into account the standard deviation of the distribution. Hence, some methods do not produce a statistically significant difference when computing the paired Wilcoxon test (e.g., the $\text{POLITO}^{\text{Snakes}}{}_{\text{IT}}$ method) as they equally overestimate and underestimate the CIMT measurement. This underlines the importance of taking into consideration the standard deviation of the bias distribution. It is therefore recommended to not only compute the CIMT bias but also the absolute CIMT bias, to facilitate a quick interpretation of the segmentation results.

4.2.2. Database division by morphology and SNR

4.2.2.1. Carotid morphology analysis

When dividing the database by morphology (Figure 6), many interesting observations can be done. First of all, when analyzing only the manual segmentations, we can see how in general the images representing the carotid artery as straight across the image are those that present the lowest median CIMT absolute bias, yet it also presents the most outliers shown as black circles above the whiskers. The morphologies that presented the most spread out distributions are the inclined up/down and the curved down morphologies, showing how these morphologies may be the most difficult ones for operators to manually segment. When moving on to observe the box plots also of the computerized methods, it is interesting how at least one (and often more than one) computerized method always showed a lower median CIMT absolute bias than both of the manual segmentations. In particular, the CREATIS$_{FR}$ method showed overall good performances on all of the considered morphologies, presenting the lowest median value in all cases except for the inclined down geometry, where it is slightly higher, yet still very close, to the CNR$_{IT}$ and INESCTEC$_{PT}$ methods. Overall, the inclined-down morphology shows a median CIMT absolute bias that is higher than the other geometries ($p<0.001$), illustrating how the computerized methods tended to all have a more difficult time correctly segmenting these images. In fact, for five out of the seven computerized methods, the inclined-down geometry is the one that shows the highest median absolute CIMT bias. The two exceptions are POLITO$^{Snakes}_{IT}$ and POLITO$^{Unet}_{IT}$ which, unlike the other computerized methods, show a clear difference in performance when dividing the database by geometry ($p<0.001$). In particular, both methods show much higher absolute CIMT biases with inclined-up morphology, whereas the POLITO$^{Snakes}_{IT}$ method also has a more difficult time segmenting images where the carotid artery is presented with a curved down geometry. After a visual analysis, the higher CIMT absolute error in the inclined down images is mainly due to two reasons. First, in a number of images the IMC is clearly visible only on a limited section of the image, and the manual operator extrapolated where the LI and MA boundaries should be located. In these cases, the computerized methods typically underestimated the final IMT. Second, in some cases the manual segmentations that were considered as gold standard were erroneous, and the computerized methods actually followed the visual boundaries in a more accurate manner than the manual segmentation. Still, it can be recommended to include a large proportion of non-straight arteries in the training dataset for deep learning methods. For traditional image processing techniques, it is recommended to allow any potential models to differ from straight ones up to a realistic limit and to be sure to not include any heuristic rules that may be mainly valid for straight arteries within the ultrasound image.

4.2.2.2. Image signal to noise analysis

When dividing the database by signal to noise ratio (Figure 7), a first observation is how the Average SNR group is the one that presents the most outliers, which are defined as absolute CIMT bias values that are more than 1.5×IQR times further away from the 3rd quartile. This is most likely due to the high number of images included in this class. So, while the median CIMT absolute bias is not necessarily the highest for this group, it is the one that presents both the highest number of outliers for manual and for computerized methods. A reduction in SNR does not have a visible impact on the CIMT bias, since the box plots are similar in terms of median and IQR range between the Low and Average SNR groups and the KW test did not show any statistically significant differences ($p=0.31$). In the High SNR group, we find the lowest median CIMT absolute bias for all the computerized methods and A2, with A1' also showing a median comparable to its best one. Moreover, a general reduction in the IQR range can be observed for the High SNR group. Overall, it can be observed that all methods, both manual and computerized, show the largest absolute CIMT bias range in the Low and Average SNR classes (i.e., each method shows the highest bias for an image classified as having a low SNR). Both the absolute bias median and the IQR tend to decrease when the SNR increases and comparing the absolute biases in the High SNR class with respect to the rest of the dataset shows a significant difference ($p<0.001$) supporting this observation. Hence, as can be expected, it is recommended to limit the noise within the lumen as much as possible when acquiring ultrasound carotid artery images that will be analyzed by computerized methods.

4.2.3. Inclusion of in-silico data

The inclusion of the in-silico FAMUS images in this study provided an important analysis, especially considering the deep learning methods. In fact, it underlined the importance of the data that is used as ground truth for training deep neural networks, as the trained model is then "biased" to that data. This can be especially noticed with the CREATIS$_{FR}$ algorithm which provided very good segmentation results when compared to the manual A1 segmentations which were used for training and validating the networks, but showed a higher fixed bias for the in-silico simulated images. As deep learning methods have become and will continue to be the state-of-the-art for medical image analysis, it is therefore recommended to also include some data where the a-priori ground truth is known in order to evaluate this issue.

5. Conclusions

In this study, we presented an open database of 500 carotid ultrasound images acquired at four different centers and including simulated images with a-priori known CIMT values and LI/MA locations. Three manual segmentations and seven different computerized methods were employed to segment the LI and MA boundaries, including two deep-learning-based methods. The segmentation results were analyzed in-depth, through an inter- and intra-analyst variability study, and the comparison between the CIMT values obtained using computerized methods and an expert analyst manual segmentation.

After publishing an in-depth study comparing five different computerized CIMT measurement methods on a common large clinical database where we found that the CIMT measurements obtained with a skilled analyst's segmentation and the computerized segmentations were comparable in numerous clinical aspects, here the same consortium came together to provide a further technical analysis on computerized methods for CIMT measurement. Importantly, the present study provides a first-of-its-kind analysis on technical considerations, analyzing the performances based on morphology, signal to noise ratio, and imaging system used to acquire images. We underline the importance of standardizing computerized CIMT measurements, both when comparing with manual segmentations and between computerized methods. Future aspects will aim for the development of explainable and reliable deep-learning methods, with a particular focus on the inclusion of in-silico data where the a-priori ground truth is known. The entire database used in this study (500 images, division by morphology and SNR, manual tracings and segmentation contours from all compared computerized methods) has been made publicly available for the community. It is the authors intention to facilitate future studies by different research groups by providing a large annotated data set, encouraging an open comparison and technical analysis.

Acknowledgments

This work was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR); LABEX PRIMES provides N. Lainé's doctoral grant.

Conflict of Interest: Dr. Faita and Dr. Bianchini are both co-founders of Quipu Srl, Pisa, Italy, a spin-off company of the Italian National Research Council and the University of Pisa.
Dr. Ghiadoni reports other from Quipu srl, outside the submitted work; Dr. Campilho, Dr. Carvalho and Dr. Rouco report ECHOCAD CAROTID (IGAC software registration 185/2918).

References

[1] Stein JH, Korcarz CE, Hurst RT, Lonn E, Kendall CB, Mohler ER, et al. Use of Carotid Ultrasound to Identify Subclinical Vascular Disease and Evaluate Cardiovascular Disease Risk: A Consensus Statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force Endorsed by the Society for Vascular. J Am Soc Echocardiogr 2008;21:93–111. https://doi.org/10.1016/j.echo.2007.11.011.

[2] Kutbay U, Hardalaç F, Akbulut M, Akaslan Ü, Serhatlıoğlu S. A computer-aided diagnosis system for measuring carotid artery Intima-Media Thickness (IMT) using quaternion vectors. J Med Syst 2016;40:149. https://doi.org/10.1007/s10916-016-0507-4.

[3] Lorenz MW, Gao L, Ziegelbauer K, Norata GD, Empana JP, Schmidtmann I, et al. Predictive value for cardiovascular events of common carotid intima media thickness and its rate of change in individuals at high cardiovascular risk – Results from the PROG-IMT collaboration. PLoS One 2018;13:e0191172. https://doi.org/10.1371/journal.pone.0191172.

[4] Plichart M, Celermajer DS, Zureik M, Helmer C, Jouven X, Ritchie K, et al. Carotid intima-media thickness in plaque-free site, carotid plaques and coronary heart disease risk prediction in older adults. The Three-City Study. Atherosclerosis 2011;219:917–24. https://doi.org/10.1016/j.atherosclerosis.2011.09.024.

[5] Potter K, Reed CJ, Green DJ, Hankey GJ, Arnolda LF. Ultrasound settings significantly alter arterial lumen and wall thickness measurements. Cardiovasc Ultrasound 2008;6:6. https://doi.org/10.1186/1476-7120-6-6.

[6] Bianchini E, Giannarelli C, Maria Bruno R, Armenia S, Landini L, Faita F, et al. Functional and Structural Alterations of Large Arteries: Methodological Issues. Curr Pharm Des 2013;19:2390–400. https://doi.org/10.2174/1381612811319130007.

[7] Meiburger KM, Acharya UR, Molinari F. Automated localization and segmentation techniques for B-mode ultrasound images: A review. Comput Biol Med 2018;92:210–35. https://doi.org/10.1016/j.compbiomed.2017.11.018.

[8] Touboul P, Hennerici M, Meairs S, Adams H. Mannheim carotid intima-media thickness and plaque consensus (2004–2006–2011). Cerebrovasc Dis 2012;34:290–6.

[9] Loizou CP. A review of ultrasound common carotid artery image and video segmentation techniques. Med Biol Eng Comput 2014;52:1073–93. https://doi.org/10.1007/s11517-014-1203-5.

[10] Molinari F, Meiburger KM, Saba L, Acharya UR, Ledda G, Zeng G, et al. Ultrasound IMT measurement on a multi-ethnic and multi-institutional database: Our review and experience using four fully automated and one semi-automated methods. Comput Methods Programs Biomed 2012;108:946–60. https://doi.org/10.1016/j.cmpb.2012.05.008.

[11] Faita F, Gemignani V, Bianchini E, Giannarelli C, Ghiadoni L, Demi M. Real-time Measurement System for Evaluation of the Carotid Intima-Media Thickness With a Robust Edge Operator. J Ultrasound Med 2008;27:1353–61. https://doi.org/10.7863/jum.2008.27.9.1353.

[12] Rocha R, Silva J, Campilho A. Automatic segmentation of carotid B-mode images using fuzzy classification. Med Biol Eng Comput 2012;50:533–45. https://doi.org/10.1007/s11517-012-0883-y.

[13] Molinari F, Rajendra Acharya U, Zeng G, Meiburger KM, Suri JS. Completely automated robust edge snapper for carotid ultrasound IMT measurement on a multi-institutional database of 300 images. Med Biol Eng Comput 2011;49:935–45. https://doi.org/10.1007/s11517-011-0781-8.

[14] Loizou CP, Pattichis CS, Pantziaris M, Tyllis T, Nicolaides A. Snakes based segmentation of the common carotid artery intima media. Med Biol Eng Comput 2007;45:35–49. https://doi.org/10.1007/s11517-006-0140-3.

[15] Molinari F, Meiburger KM, Saba L, Zeng G, Acharya UR, Ledda M, et al. Fully automated dual-snake formulation for carotid intima-media thickness measurement. A new approach. J Ultrasound Med 2012;31:1123–36. https://doi.org/10.7863/jum.2012.31.7.1123.

[16] Rocha R, Campilho A, Silva J, Azevedo E, Santos R. Segmentation of the carotid intima-media region in B-Mode ultrasound images. Image Vis Comput 2010;28:614–25. https://doi.org/10.1016/j.imavis.2009.09.017.

[17] Ilea DE, Duffy C, Kavanagh L, Stanton A, Whelan PF. Fully automated segmentation and tracking of the intima media thickness in ultrasound video sequences of the common carotid artery. IEEE Trans Ultrason Ferroelectr Freq Control 2013;60. https://doi.org/10.1109/TUFFC.2013.2547.

[18] Menchón-Lara R-M, Bastida-Jumilla M-C, Morales-Sánchez J, Sancho-Gómez J-L. Automatic detection of the intima-media thickness in ultrasound images of the common carotid artery using neural networks. Med Biol Eng Comput 2014;52:169–81. https://doi.org/10.1007/s11517-013-1128-4.

[19] Zahnd G, Kapellas K, van Hattem M, van Dijk A, Sérusclat A, Moulin P, et al. A Fully-Automatic Method to Segment the Carotid Artery Layers in Ultrasound Imaging: Application to Quantify the Compression-Decompression Pattern of the Intima-Media Complex During the Cardiac Cycle. Ultrasound Med Biol 2017;43:239–57. https://doi.org/10.1016/J.ULTRASMEDBIO.2016.08.016.

[20] Hassan M, Chaudhry A, Khan A, Iftikhar MA. Robust information gain based fuzzy c-means clustering and classification of carotid artery ultrasound images. Comput Methods Programs Biomed 2014;113:593–609. https://doi.org/10.1016/j.cmpb.2013.10.012.

[21] Menchón-Lara R-M, Sancho-Gómez J-L. Fully automatic segmentation of ultrasound common carotid artery images based on machine learning. Neurocomputing 2015;151:161–7. https://doi.org/10.1016/J.NEUCOM.2014.09.066.

[22] Zhou R, Fenster A, Xia Y, Spence JD, Ding M. Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images. Med Phys 2019;46:3180–93. https://doi.org/10.1002/mp.13581.

[23] Savaş S, Topaloğlu N, Kazcı Ö, Koşar PN. Classification of Carotid Artery Intima Media

Thickness Ultrasound Images with Deep Learning. J Med Syst 2019;43:273. https://doi.org/10.1007/s10916-019-1406-2.

[24] Biswas M, Kuppili V, Araki T, Edla DR, Godia EC, Saba L, et al. Deep learning strategy for accurate carotid intima-media thickness measurement: An ultrasound study on Japanese diabetic cohort. Comput Biol Med 2018;98:100–17. https://doi.org/10.1016/j.compbiomed.2018.05.014.

[25] Vila M del M, Remeseiro B, Grau M, Elosua R, Betriu À, Fernandez-Giraldez E, et al. Semantic segmentation with DenseNets for carotid artery ultrasound plaque segmentation and CIMT estimation. Artif Intell Med 2020;103:101784. https://doi.org/10.1016/j.artmed.2019.101784.

[26] Jain PK, Sharma N, Saba L, Paraskevas KI, Kalra MK, Johri A, et al. Unseen Artificial Intelligence—Deep Learning Paradigm for Segmentation of Low Atherosclerotic Plaque in Carotid Ultrasound: A Multicenter Cardiovascular Study. Diagnostics 2021, Vol 11, Page 2257 2021;11:2257. https://doi.org/10.3390/DIAGNOSTICS11122257.

[27] Jain PK, Sharma N, Giannopoulos AA, Saba L, Nicolaides A, Suri JS. Hybrid deep learning segmentation models for atherosclerotic plaque in internal carotid artery B-mode ultrasound. Comput Biol Med 2021;136:104721. https://doi.org/10.1016/J.COMPBIOMED.2021.104721.

[28] Jain PK, Sharma N, Saba L, Paraskevas KI, Kalra MK, Johri A, et al. Automated deep learning-based paradigm for high-risk plaque detection in B-mode common carotid ultrasound scans: an asymptomatic Japanese cohort study. Int Angiol 2021. https://doi.org/10.23736/S0392-9590.21.04771-4.

[29] Biswas M, Saba L, Omerzu T, Johri AM, Khanna NN, Viskovic K, et al. A Review on Joint Carotid Intima-Media Thickness and Plaque Area Measurement in Ultrasound for Cardiovascular/Stroke Risk Monitoring: Artificial Intelligence Framework. J Digit Imaging 2021;34:581–604. https://doi.org/10.1007/s10278-021-00461-2.

[30] Biswas M, Saba L, Chakrabartty S, Khanna NN, Song H, Suri HS, et al. Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: A screening tool for cardiovascular/stroke risk assessment. Comput Biol Med 2020;123:103847. https://doi.org/10.1016/j.compbiomed.2020.103847.

[31] Zhou Z, Shin J, Feng R, Hurst RT, Kendall CB, Liang J. Integrating Active Learning and Transfer Learning for Carotid Intima-Media Thickness Video Interpretation. J Digit Imaging 2019;32:290–9. https://doi.org/10.1007/s10278-018-0143-2.

[32] Saba L, Jamthikar A, Gupta D, Khanna NN, Viskovic K, Suri HS, et al. Global perspective on carotid intima-media thickness and plaque: should the current measurement guidelines be revisited? Int Angiol 2019;38:451–65. https://doi.org/10.23736/S0392-9590.19.04267-6.

[33] Saba L, Molinari F, Meiburger KM, Piga M, Zeng G, Rajendra Achraya U, et al. What is the correct distance measurement metric when measuring carotid ultrasound intima-media thickness automatically? Int Angiol 2012;31.

[34] Meiburger KM, Zahnd G, Faita F, Loizou CP, Carvalho C, Steinman DA, et al. Carotid Ultrasound Boundary Study (CUBS): An Open Multicenter Analysis of Computerized Intima–Media Thickness Measurement Systems and Their Clinical Impact. Ultrasound Med Biol 2021;47:2442–55. https://doi.org/10.1016/j.ultrasmedbio.2021.03.022.

[35] Aguilar L, Cobbold RSC, Steinman D. Fast and mechanistic ultrasound simulation using a point source/receiver approach. Ultrason Ferroelectr Freq Control IEEE Trans 2013;60:2335–46. https://doi.org/10.1109/TUFFC.2013.6644737.

[36] Aguilar L, Wong J, Steinman DA, Cobbold RSC. FAMUS II: A fast and mechanistic ultrasound simulator using an impulse response approach. IEEE Trans Ultrason Ferroelectr Freq Control 2017;64:362–73. https://doi.org/10.1109/TUFFC.2016.2632706.

[37] Bruno RM, Cartoni G, Stea F, Armenia S, Bianchini E, Buralli S, et al. Carotid and aortic stiffness in essential hypertension and their relation with target organ damage: the CATOD study. J Hypertens 2017;35:310–8. https://doi.org/10.1097/HJH.0000000000001167.

[38] Bruno RM, Stea F, Sicari R, Ghiadoni L, Taddei S, Ungar A, et al. Vascular Function Is Improved After an Environmental Enrichment Program. Hypertension 2018;71:1218–25. https://doi.org/10.1161/HYPERTENSIONAHA.117.10066.

[39] Molinari F, Pattichis CS, Zeng G, Saba L, Acharya UR, Sanfilippo R, et al. Completely automated multiresolution edge snapper-A new technique for an accurate carotid ultrasound IMT measurement: Clinical validation and benchmarking on a multi-institutional database. IEEE Trans Image Process 2012;21:1211–22. https://doi.org/10.1109/TIP.2011.2169270.

[40] Molinari F, Meiburger KM, Saba L, Zeng G, Acharya UR, Ledda M, et al. Fully automated dual-snake formulation for carotid intima-media thickness measurement: A new approach. J Ultrasound Med 2012;31.

[41] Meiburger KM, Molinari F, Acharya UR, Saba L, Rodrigues P, Liboni W, et al. Automated carotid artery intima layer regional 2011;4073. https://doi.org/10.1088/0031-9155/56/13/021.

[42] Molinari F, Meiburger KM, Zeng G, Nicolaides A, Suri JS. CAUDLES-EF: Carotid automated ultrasound double line extraction system using edge flow. J Digit Imaging 2011;24:1059–77. https://doi.org/10.1007/s10278-011-9375-0.

[43] Rouco J, Carvalho C, Domingues A, Azevedo E, Campilho A. A robust anisotropic edge detection method for carotid ultrasound image processing. Procedia Comput Sci 2018;126:723–32. https://doi.org/10.1016/J.PROCS.2018.08.006.

[44] Zahnd G, Orkisz M, Dávila Serrano EE, Vray D. CAROLAB – A platform to analyze carotid ultrasound data. IEEE Int. Ultrason. Symp. (IUS), Glas., 2019, p. 463–6. https://doi.org/10.1109/ULTSYM.2019.8925673.

[45] Lainé N, Zahnd G, Liebgott H, Orkisz M. Carotid artery wall segmentation in ultrasound image sequences using a deep convolutional neural network. ArXiv Prepr ArXiv 2022:https://arxiv.org/abs/2201.12152.

[46] Bianchini E, Bozec E, Gemignani V, Faita F, Giannarelli C, Ghiadoni L, et al. Assessment of Carotid Stiffness and Intima-Media Thickness From Ultrasound Data. J Ultrasound Med

2010;29:1169–75. https://doi.org/10.7863/jum.2010.29.8.1169.

[47] Demi M. Contour Tracking with a Spatio-Temporal Intensity Moment. IEEE Trans Pattern Anal Mach Intell 2016;38:1141–54. https://doi.org/10.1109/TPAMI.2015.2478438.

[48] Faita F, Gemignani V, Bianchini E, Giannarelli C, Demi M. Real-time measurement system for the evaluation of the intima media thickness with a new edge detector. 2006 Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE; 2006, p. 715–8. https://doi.org/10.1109/IEMBS.2006.260450.

[49] Rouco J, Campilho A. Robust common caritod artery lumen detection in B-Mode ultrasound images using local phase symmetry. IEEE Int. Conf. Acoust. Speech Signal Process., 2013, p. 929–33. https://doi.org/10.1109/ICASSP.2013.6637785.

[50] Zahnd G, Orkisz M, Sérusclat A, Moulin P, Vray D. Simultaneous extraction of carotid artery intima-media interfaces in ultrasound images: assessment of wall thickness temporal variation during the cardiac cycle. Int J Comput Assist Radiol Surg 2014;9:645–58. https://doi.org/10.1007/s11548-013-0945-0.

[51] Williams DJ, Shah M. A Fast algorithm for active contours and curvature estimation. CVGIP Image Underst 1992;55:14–26. https://doi.org/10.1016/1049-9660(92)90003-L.

[52] Loizou CP, Nicolaides A, Kyriacou E, Georghiou N, Griffin M, Pattichis CS. A comparison of ultrasound intima-media thickness measurements of the left and right common carotid artery. IEEE J Transl Eng Heal Med 2015;3:1–10. https://doi.org/10.1109/JTEHM.2015.2450735.

[53] Kovesi P. Symmetry and Asymmetry from Local Phase. Tenth Aust. Jt. Conf. Artifi cial Intell., 1997.

[54] Kovesi P. Peter Kovesi n.d. https://www.peterkovesi.com/matlabfns/index.html.

[55] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2015;9351:234–41. https://doi.org/10.1007/978-3-319-24574-4_28.

[56] Huang G, Liu Z, van der Maaten L. Densely Connected Convolutional Networks Gao. IEEE Conf. Comput. Vis. pattern Recognit., 2017, p. 4700–8.

[57] Roy AG, Navab N, Wachinger C. Recalibrating Fully Convolutional Networks With Spatial and Channel “Squeeze and Excitation” Blocks. IEEE Trans Med Imaging 2019;38:540–9. https://doi.org/10.1109/TMI.2018.2867261.

[58] Loizou CP, Pattichis CS, Pantziaris M, Tyllis T, Nicolaides a. Snakes based segmentation of the common carotid artery intima media. Med Biol Eng Comput 2007;45:35–49. https://doi.org/10.1007/s11517-006-0140-3.

[59] Molinari F, Meiburger KM, Zeng G, Nicolaides A, Suri JS. CAUDLES-EF: Carotid automated ultrasound double line extraction system using edge flow. J Digit Imaging 2011;24:1059–77. https://doi.org/10.1007/s10278-011-9375-0.

[60] Molinari F, M. Meiburger K, Zeng G, Acharya UR, Liboni W, Nicolaides A, et al. Carotid artery recognition system: A comparison of three automated paradigms for ultrasound images. Med Phys 2012;39:378. https://doi.org/10.1118/1.3670373.

Highlights:

- Seven computerized and three manual CIMT measurement methods are compared
- Recommendations are given to encourage CIMT measurement standardization
- A technical analysis is given on carotid morphology, SNR, and acquisition settings
- An open dataset with images and manual/computerized segmentations is provided

Conflict of Interest: Dr. Faita and Dr. Bianchini are both co-founders of Quipu Srl, Pisa, Italy, a spin-off company of the Italian National Research Council and the University of Pisa.

Dr. Ghiadoni reports other from Quipu srl, outside the submitted work; Dr. Campilho, Dr. Carvalho and Dr. Rouco report ECHOCAD CAROTID (IGAC software registration 185/2918).

Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans

Supplementary Materials

1. Fast And Mechanistic Ultrasound Simulation (FAMUS) software simulation parameters

The simulation parameters for the images are shown in the table below:

Table S1. FAMUS simulation parameters

Parameter	Value
Velocity of sound (m/s)	1540
Sampling frequency (MHz)	100
Transducer frequency (MHz)	7
Number of elements in x-direction	192
Number of elements in y direction	5
Element height (m)	5×10^{-3}
Size of transducer in lateral direction (m)	31×10^{-3}
F-number (depth/aperture size)	2
Elevation focus (m)	30×10^{-3}
Number of focal zones in transmission	1
Number of focal zones in reception	4
Focal zone in transmission (m)	50×10^{-3}
Focal zone in reception (m)	$[40, 45, 50, 55] \times 10^{-3}$
Number of cycles of initial excitation	2
Apodization window	Hanning
Image width (m)	40×10^{-3}
Number of lines to form B-mode image	128
Number of scatterers	30000
Beamforming method	Delay and Sum

For each image, the artery was divided into seven sections: lumen, intima layer of the distal wall, media layer of the distal wall, adventitia layer of the distal wall, intima-media complex of the proximal wall, adventitia layer of the proximal wall, and everywhere else. The distal wall layers were manually drawn, whereas the radius of the lumen was always equal to 3.5 mm and the proximal wall layers were automatically drawn following the geometry of the artery (i.e., straight, inclined up, inclined down, curved up, curved down). For the simulation, three feature index parameters were defined for each section, as reported in the following table:

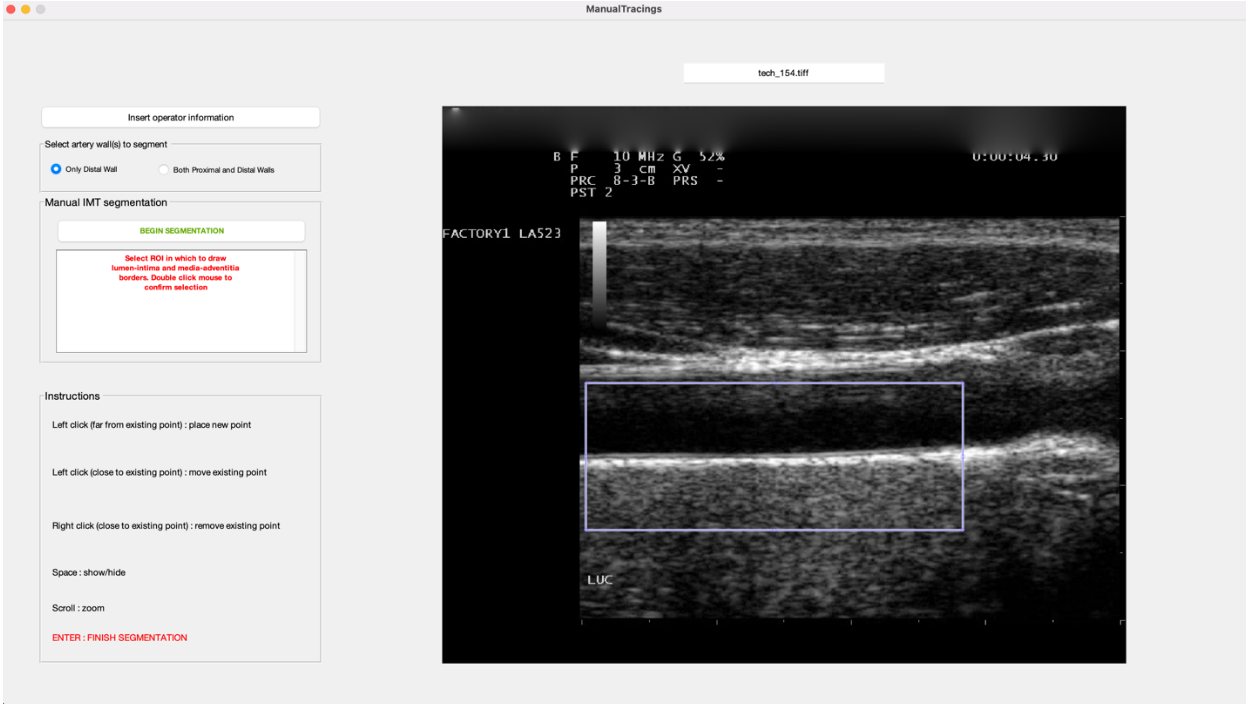
Table S2. Feature indexes for simulation

Image section	Backscatter coefficient	Attenuation (dB/MHz×cm)	Characteristic acoustic impedance (z_0 = rho×c)
Lumen	5×10^{-5}	0.1	1000000
Media layer – distal wall	0.0224	0.22	1400000
Intima layer – distal wall	0.0424	0.22	1700000
Adventitial layer – distal wall	0.0605	0.52	7700000
Intima-media complex – proximal wall	0.0224	0.22	1700000
Adventitial layer – proximal wall	0.0553	0.12	6000000
Everywhere else	0.0505	0.1	5620000

These three parameters are combined, along with distance to transducer surface and view angle to generate a specific amplitude for each scatterer. For more information, this approach is inspired by the study by Pham et al. [1]

2. Manual segmentation Graphical User Interface

The manual segmentations were done independently using the same Graphical User Interface (GUI) that was developed specifically for this purpose. The GUI was developed in MATLAB and was saved as a stand-alone program for the clinicians to easily install and use on their own personal computers. The GUI reads standard image files (e.g. .jpg) and the user is first requested to select a ROI inside of which to trace the lumen-intima and media-adventitia borders (Figure S1).



Select artery wall(s) to segment

● Only Distal Wall ○ Both Proximal and Distal Walls

Select ROI in which to draw lumen-intima and media-adventitia borders. Double click mouse to confirm selection

Instructions

Left click (far from existing point) : place new point

Left click (close to existing point) : move existing point

Right click (close to existing point) : remove existing point

Space : show/hide

Scroll : zoom

ENTER : FINISH SEGMENTATION

Figure S1. Screenshot of graphical user interface and ROI selection.

Subsequently, the user proceeds to manually trace the boundary that is specified in the Instructions panel. The user can add a point by left-clicking far from already placed points or move a point by left-clicking close to the already placed point. The user can also remove a point by right-clicking close to the point that needs to be to removed. If desired, the user can press the space bar to hide the points and profile or zoom in by scrolling. Once the segmentation is done and the user is satisfied with the results, the user presses Enter to save the manually traced coordinates in a .txt file. Figure S2 shows another screenshot of the GUI showing how the points can be manually traced.

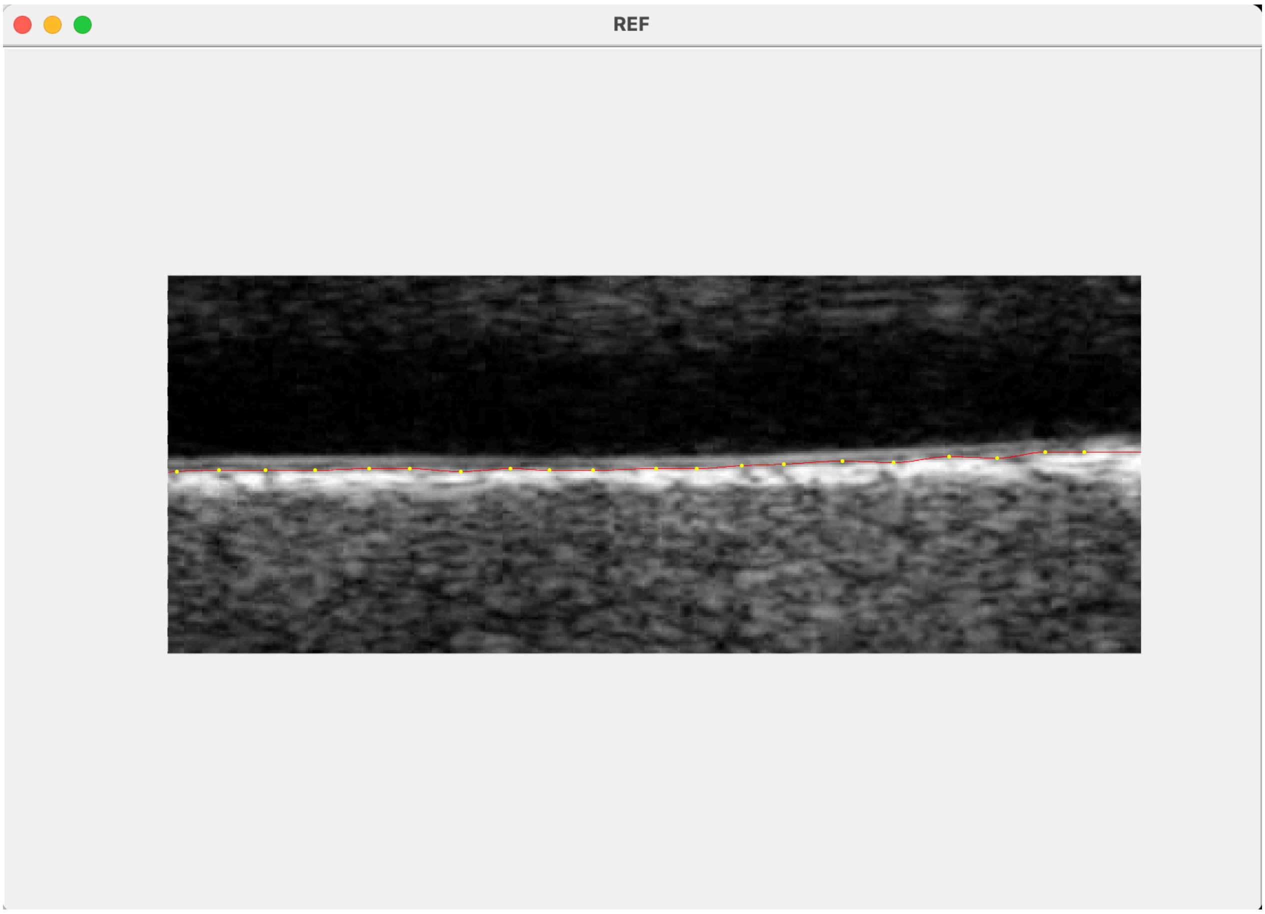


Figure S2. Screenshot of graphical user interface and example segmentation of media-adventitia border.

3. Parameters used for dual snake model (POLITO$^{\text{Snakes}}$$_{\text{IT}}$)

Table S3. Dual snake parameter values

Parameter	Value	Range
α_{LI}	0.1	(0.05 – 0.15)
β_{LI}	-0.5	-(0.25 – 0.75)
γ_{LI}	0.35	(0.175 – 0.525)
α_{MA}	0.5	(0.25 – 0.75)
β_{MA}	0.5	(0.25 – 0.75)
γ_{MA}	-0.25	-(0.125 – 0.375)

4. Computerized methods comparison

Table S4 summarizes the pros and cons of the computerized methods that were employed in the study can be found below. The computational time of each method is also reported.

Table S4. Methods comparison

Method	**Computational time per image**	**Pros**	**Cons**
CNR$_{\mathrm{IT}}$	~0.01 s (i5 4-core CPU)	• Robust to speckle noise thanks to edge detector • Tested in a clinical environment • Validated against commercially available systems	• Requires manual identification of the initial ROI by the operator • Needs calibration factor
INESCTEC$_{\mathrm{PT}}$	~ 1 s (i5 4-core CPU)	• Robust to speckle noise thanks to edge detector • Robust parallel edge location and continuation	• Requires manual identification of the initial ROI by the operator
TUM$_{\mathrm{DE}}$	0.8 ± 0.3 s (2.90 GHz processor and 32 Gb RAM)	• Robustness tested on data acquired with different scanners • Includes a mechanism to control the smoothness of the contours by considering the spatial IMT variations • Is fully-automatic when applied to temporal image sequences	• Is not expected to perform well on images where atherosclerotic plaques are present • Requires manual initialization (seven mouse clicks) when applied to a static image
UCY$_{\mathrm{CY}}$	2.4 ± 1.7 s (i5 4-core CPU)	• Good agreement with manual measurements • Initialization brings snake close to the borders of interests • Human interpretable method – general approach applied successfully in numerous other studies • Accuracy is about 0.95	• Relatively high computational time • Initialization dependent • In 5% of the cases snake converges to a wrong location • Presence of acoustic shadowing together with strong speckle noise hinders the visual and automatic analysis in ultrasound images
POLITO$^{\mathrm{Snakes}}_{\mathrm{IT}}$	~ 2 s (i5 4-core CPU)	• Human interpretable method – general approach applied successfully in other studies • Robust over small areas with very low/high echogenicity thanks to the internal forces and the mutual distance energy • Fully automatic method	• Initialization dependent • Low generalizability over FAMUS dataset • Average to high computational time
CREATIS$_{\mathrm{FR}}$	2.19 ± 0.38 s (NVIDIA T1200 laptop GPU)	• High robustness and accuracy • Incorporation of a mechanism coping with inclined and curved arteries • Interaction limited to two mouse clicks	• Decreased accuracy in the presence of calcified plaques • Inference time increased by the use of overlapping sub-windows • Two separate networks independently trained for distinct successive tasks: far wall detection and actual IMC segmentation
POLITO$^{\mathrm{UNet}}_{\mathrm{IT}}$	0.06 ± 0.02 s (GPU-based, RTX 3070 8GB)	• Fully automatic method • High and stable performance over the entire dataset • Robust on FAMUS dataset • GPU optimized, fast	• Lower interpretability of the segmentation mechanics given the use of a deep learning model

CNR$_{\mathrm{IT}}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{\mathrm{PT}}$: INESC Technology and Science; TUM$_{\mathrm{DE}}$: Technische Universität München; UCY$_{\mathrm{CY}}$: University of Cyprus; POLITO$^{\mathrm{Snakes}}_{\mathrm{IT}}$: Politecnico di Torino (dual snakes); POLITO$^{\mathrm{UNet}}_{\mathrm{IT}}$: Politecnico di Torino (UNet); CREATIS$_{\mathrm{FR}}$: CREATIS lab Université de Lyon .

5. Polyline and Hausdorff distance metric computation

5.1 Polyline distance metric

For the sake of clarity, here we will consider the LI border as boundary A and the MA border as boundary B. A Polyline is a line identified by two successive points on boundary B or A. On the basis of this definition, the perpendicular distance from a boundary point A (B) to the Polyline (i.e., the segment connecting two successive points) on the opposite boundary B (A) can be calculated [18]. In this way, the perpendicular distances for all the points from boundary A onto boundary B and vice-versa, is computed. The Polyline distance is then defined as the average of the computed distances from A to B and B to A. Figure S3 shows an example of Polyline distance computation.

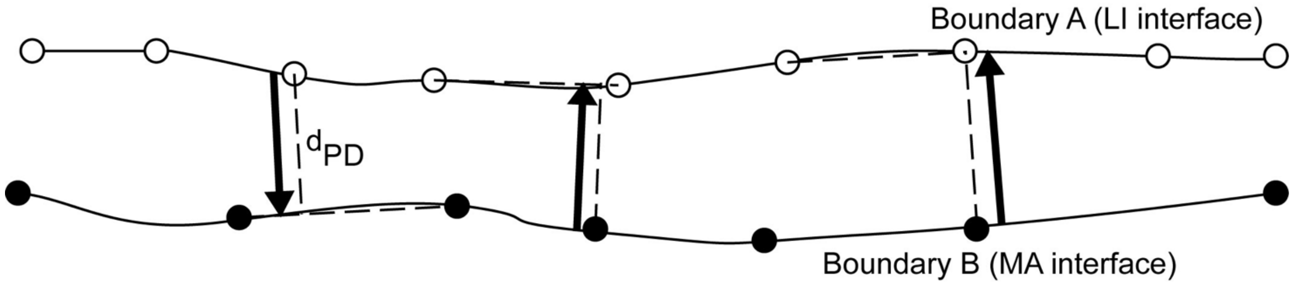


Figure S3. Example of Polyline Distance computation. The distance (dPD) is calculated as the perpendicular distance of one point on a boundary with respect to the closest segment of the other boundary. LI: lumen-intima; MA: media-adventitia.

5.2. Hausdorff distance metric

The Hausdorff distance is defined as the maximum distance of a boundary A to the nearest point in Boundary B. Considering the first point of the boundary B, B_1, the Euclidean distance is computed between B1 and each point on boundary A: $d_{B_1A_1}$, $d_{B_1A_2}$, $d_{B_1A_3}$, … , $d_{B_1A_M}$, where M is the number of points of boundary A and the minimum of these distances is kept, called $dmin_{B_1}$. This process is repeated for every point of boundary B, reaching up to computing the Euclidean distances $d_{B_NA_1}$, … , $d_{B_NA_{M-2}}$, $d_{B_NA_{M-1}}$, $d_{B_NA_M}$ and then obtaining $dmin_{B_N}$ where N is the number of points of boundary B. From all dmin values, the maximum is kept, called $dmax_B$:

$$dmax_B = maximum\{dmin_{B_1}, dmin_{B_2},\ldots,dmin_{B_N}\}$$

The same procedure is done considering boundary A compared to boundary B, hence obtaining $dmax_A$.

The Hausdorff distance metric (HDM) is then computed as:

$$HDM = max\{dmax_B, dmax_A\}$$

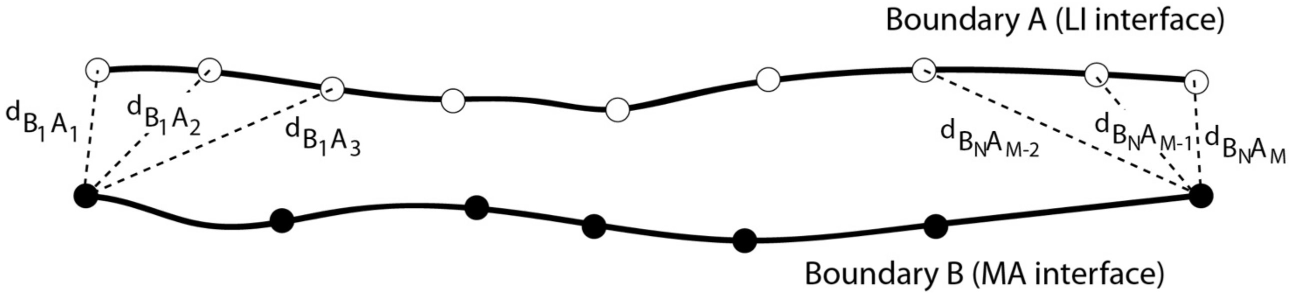


Figure S4. Visual representation of the Hausdorff distance metric.

6. Global common support determination

The global common support is computed by performing the intersection between the individual supports of each methods' profiles. In more detail, for each image, we start from a vector of ones spanning the whole width of the image. For each method, we create a vector of the same width, with ones where the LI and MA profiles are defined and zeros elsewhere. We concatenate vertically all the vectors producing a matrix and we perform a logical AND on its columns. The resulting vector has ones in the columns where all the profiles are defined and zeros elsewhere. If the total length of the common support is higher than 1 mm the vector is retained for subsequent computations; otherwise, the current image is labeled as not having a global common support.

7. Point-wise bias analysis

The Hausdorff distance gives an estimate of the largest distance between two profiles but doesn't determine if one profile is more consistently inside or outside of the other. Hence, here we analyzed each profile point-by-point compared to the expert analyst A1 and compute the point-wise bias. The point-wise bias is considered positive if the computerized profile is outside of the manual profile (i.e., more towards the bottom of the image when considering the MA profile and more towards the top when considering the LI profile), and negative otherwise.

The overall results are shown in Table S5.

Table S5. Point-wise bias computation compared to expert analyst A1

Method	**LI point bias (µm)**	**LIperc$_{out}$ (%)**	**MA point bias (µm)**	**MAperc$_{out}$ (%)**
A1'	46 ± 126	58.9 ± 27.1	23 ± 153	54.5 ± 32.1
A2	11 ± 163	52.3 ± 33.7	35 ± 187	55.2 ± 35.4
CNR$_{IT}$	85 ± 110	72.9 ± 27.7	-145 ± 127	16.5 ± 23.3
INESCTEC$_{PT}$	70 ± 477	38.3 ± 25.5	-21 ± 228	52.9 ± 30.5
TUM$_{DE}$	8 ± 106	52.9 ± 27.2	-101 ± 115	23.2 ± 25.0
UCY$_{CY}$	-7 ± 130	50.8 ± 28.2	-76 ± 126	31.3 ± 27.7
POLITO$^{Snakes}_{IT}$	162 ± 240	78.8 ± 27.1	-177 ± 301	33.1 ± 34.0
POLITO$^{UNet}_{IT}$	75 ± 101	69.7 ± 25.1	80 ± 115	72.6 ± 27.8
CREATIS$_{FR}$	-53 ± 90	32.3 ± 24.6	59 ± 105	69.9 ± 27.5

CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon. LIperc$_{out}$ = percentage of LI points found above the manual A1 LI border. MAperc$_{out}$ = percentage of MA points found below the manual A1 MA border.

8. Example segmentation results

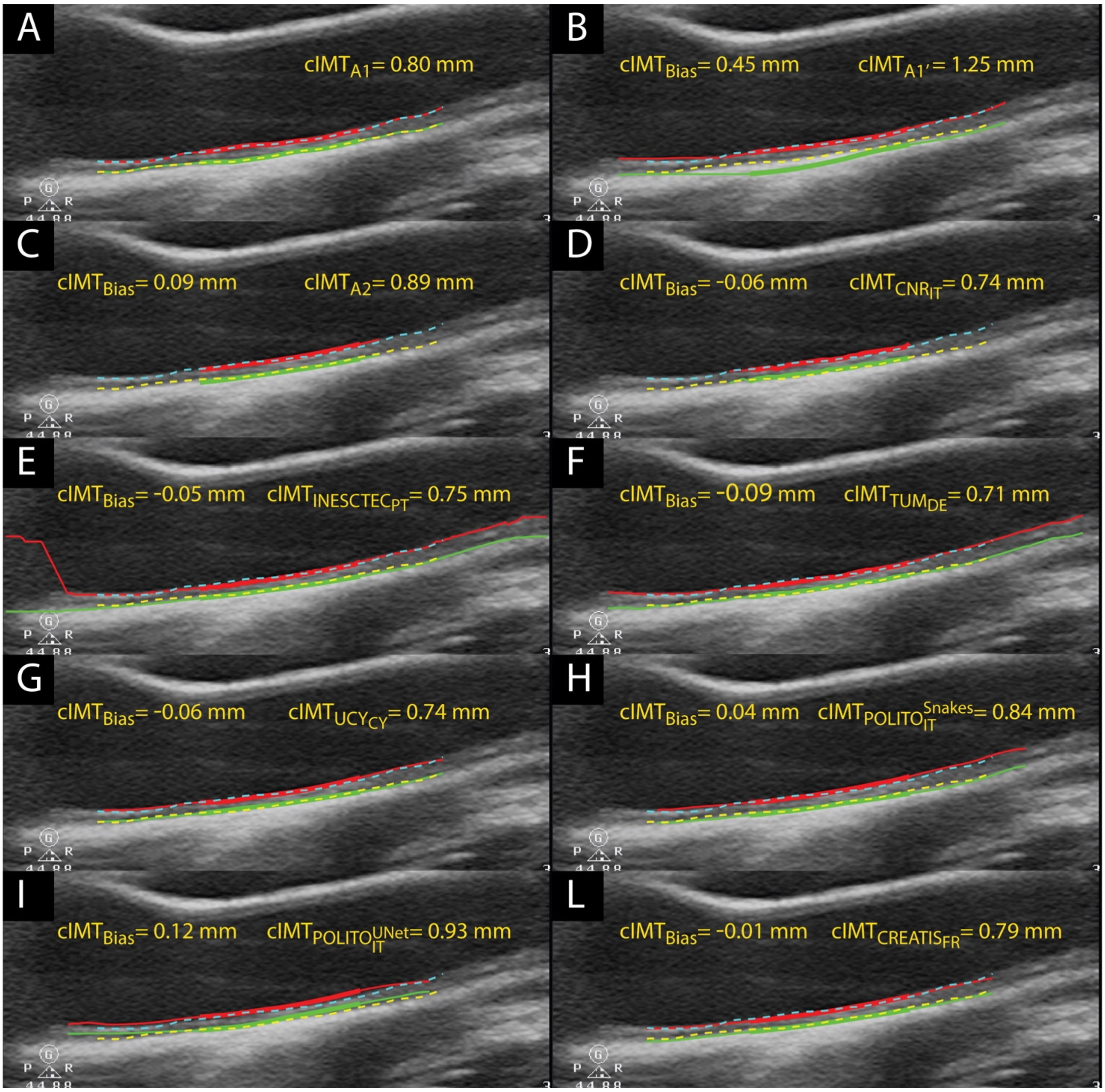


Figure S5. Segmentation results on an example image using an overall common support. In each panel, the dotted blue and yellow profiles represent the manual A1segmentations of the LI and MA boundaries, respectively. The solid red and green profiles represent on the other hand the LI and MA boundaries, respectively, of the compared method (A1', A2, or a computerized method). The red and green profiles are shown in bold where the overall common support is found (i.e., where each method, manual or computerized, provides a segmentation). The CIMT values and CIMT bias results are displayed in yellow on each panel. (A) A1; (B) A1'; (C) A2; (D) CNR$_{IT}$; (E) INESCTEC$_{PT}$; (F) TUM$_{DE}$; (G) UCY$_{CY}$; (H) POLITO$^{Snakes}_{IT}$; (I) POLITO$^{Unet}_{IT}$; (L) CREATIS$_{FR}$. A1: Manual segmentations of analyst 1; A1': manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon.

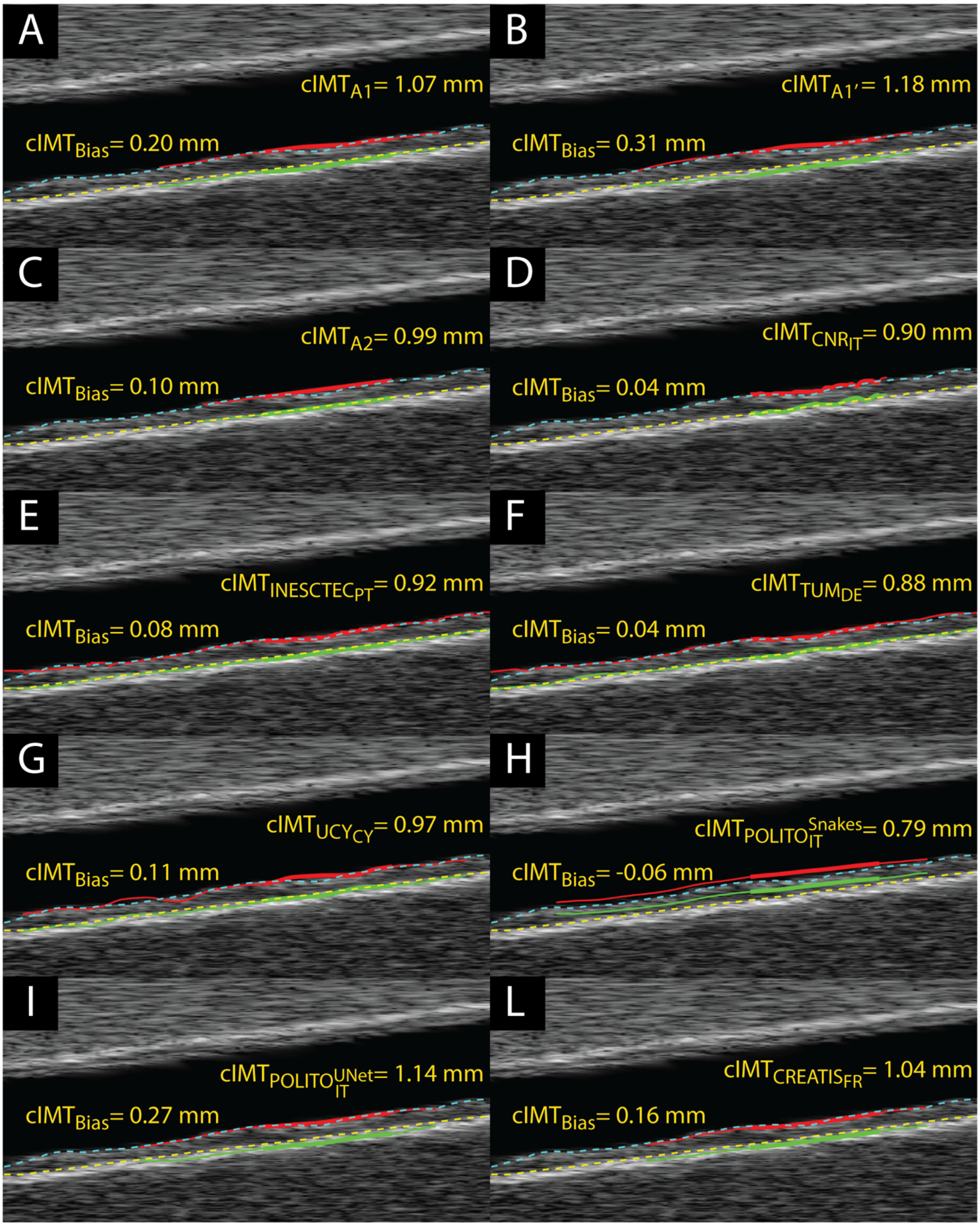


Figure S6. Segmentation results on an example FAMUS image. In each panel, the dotted blue and yellow profiles represent the LI and MA boundaries used for simulation, respectively. The solid red and green profiles represent on the other hand the LI and MA boundaries, respectively, of the compared method (A1, A1’, A2, or a computerized method). The red and green profiles are shown in bold where the overall common support is found (i.e., where each method, manual or computerized, provides a segmentation). The CIMT values and CIMT bias results are displayed in yellow on each panel. (A) A1; (B) A1’; (C) A2; (D) CNR$_{IT}$; (E) INESCTEC$_{PT}$; (F) TUM$_{DE}$; (G) UCY$_{CY}$; (H) POLITO$^{Snakes}_{IT}$; (I) POLITO$^{Unet}_{IT}$; (L) CREATIS$_{FR}$. A1: Manual segmentations of analyst 1; A1’: manual segmentations of analyst 1 traced one month after A1; A2: Manual segmentations of analyst 2; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon .

9. Complete segmentation results when dividing database by morphology

Table S6. Segmentation results compared to expert analyst 1 (A1) dividing by geometry

Geometry	Method	Number of unprocessed images	CIMT$_{bias}$ (µm)	CIMT$_{AbsBias}$ (µm)	HDM$_{LI}$ (µm)	HDM$_{MA}$ (µm)
Inclined upward (117)	A1’	7	70 ± 243	198 ± 156	285 ± 125	276 ± 136
	A2	7	102 ± 261	219 ± 174	259 ± 132	287 ± 148
	CNR$_{IT}$ [2,3]	9	11 ± 155*	124 ± 094	312 ± 124	306 ± 135
	INESCTEC$_{PT}$ [4,5]	27	60 ± 349*	167 ± 311	360 ± 492	254 ± 233
	TUM$_{DE}$ [6,7]	7	-24 ± 159*	127 ± 098	237 ± 127	247 ± 118
	UCY$_{CY}$ [8]	7	6 ± 164*	131 ± 098	238 ± 104	257 ± 114
	POLITO$^{Snakes}_{IT}$ [9]	7	177 ± 279	277 ± 179	454 ± 164	351 ± 230
	POLITO$^{UNet}_{IT}$	7	236 ± 150	241 ± 142	252 ± 112	237 ± 98
	CREATIS$_{FR}$ [10]	7	76 ± 145	125 ± 105	194 ± 90	206 ± 96
Inclined downward (106)	A1’	8	108 ± 294	234 ± 207	279 ± 163	284 ± 157
	A2	8	-22 ± 294*	238 ± 173	265 ± 127	283 ± 142
	CNR$_{IT}$ [2,3]	9	-142 ± 193	184 ± 152	269 ± 97	347 ± 160
	INESCTEC$_{PT}$ [4,5]	32	-52 ± 376	243 ± 290	431 ± 564	254 ± 194
	TUM$_{DE}$ [6,7]	8	-164 ± 183	195 ± 149	235 ± 105	283 ± 135
	UCY$_{CY}$ [8]	8	-147 ± 206	198 ± 157	257 ± 118	284 ± 140

	POLITO$^{Snakes}_{IT}$ [9]	8	15 ± 306*	234 ± 196	432 ± 163	392 ± 230
	POLITO$^{UNet}_{IT}$	8	100 ± 186	166 ± 129	223 ± 114	222 ± 106
	CREATIS$_{FR}$ [10]	8	-69 ± 187	155 ± 125	234 ± 127	200 ± 105
Straight (199)	A1’	3	46 ± 218	171 ± 141	266 ± 113	242 ± 113
	A2	3	61 ± 249	193 ± 169	258 ± 109	258 ± 156
	CNR$_{IT}$ [2,3]	3	-64 ± 158	136 ± 102	269 ± 119	308 ± 124
	INESCTEC$_{PT}$ [4,5]	34	7 ± 338	189 ± 280	439 ± 729	270 ± 449
	TUM$_{DE}$ [6,7]	3	-101 ± 157	150 ± 111	226 ± 100	252 ± 100
	UCY$_{CY}$ [8]	3	-67 ± 168	139 ± 115	228 ± 100	247 ± 103
	POLITO$^{Snakes}_{IT}$ [9]	3	6 ± 257*	204 ± 155	378 ± 170	368 ± 248
	POLITO$^{UNet}_{IT}$	3	162 ± 154	185 ± 125	229 ± 111	218 ± 097
	CREATIS$_{FR}$ [10]	3	10 ± 147*	114 ± 93	200 ± 92	194 ± 087
Curved downward (37)	A1’	0	-7 ± 247*	195 ± 148	234 ± 93	269 ± 136
	A2	0	-89 ± 246	173 ± 195	296 ± 112	282 ± 118
	CNR$_{IT}$ [2,3]	0	-145 ± 177	156 ± 167	272 ± 98	376 ± 186
	INESCTEC$_{PT}$ [4,5]	18	-15 ± 396*	234 ± 315	543 ± 673	327 ± 313
	TUM$_{DE}$ [6,7]	0	-166 ± 174	174 ± 167	214 ± 91	288 ± 143
	UCY$_{CY}$ [8]	0	-131 ± 184	152 ± 166	233 ± 97	278 ± 139
	POLITO$^{Snakes}_{IT}$ [9]	0	-144 ± 277	223 ± 216	395 ± 214	469 ± 244
	POLITO$^{UNet}_{IT}$	0	113 ± 190	178 ± 130	195 ± 90	263 ± 105
	CREATIS$_{FR}$ [10]	0	-34 ± 197*	125 ± 156	211 ± 96	210 ± 116
Curved upward (41)	A1’	4	61 ± 257*	197 ± 173	251 ± 100	254 ± 161
	A2	4	42 ± 286*	236 ± 162	249 ± 131	289 ± 128
	CNR$_{IT}$ [2,3]	4	-112 ± 164	148 ± 131	269 ± 95	349 ± 133
	INESCTEC$_{PT}$ [4,5]	18	-71 ± 213*	162 ± 152	436 ± 548	244 ± 216
	TUM$_{DE}$ [6,7]	4	-143 ± 171	172 ± 141	232 ± 67	275 ± 111
	UCY$_{CY}$ [8]	4	-112 ± 173	156 ± 133	232 ± 78	268 ± 124
	POLITO$^{Snakes}_{IT}$ [9]	4	-252 ± 377	357 ± 276	420 ± 194	664 ± 404
	POLITO$^{UNet}_{IT}$	4	127 ± 179	188 ± 112	209 ± 84	226 ± 088
	CREATIS$_{FR}$ [10]	4	-20 ± 161*	126 ± 101	195 ± 71	193 ± 087

A1: Manual segmentations of analyst 1; A1: Manual segmentations of analyst 1; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di

Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; CIMT: carotid intima-media thickness. HDM: Hausdorff distance metric. The asterisk (*) defines measurements without statistically significant differences ($p > 0.05$) using the Wilcoxon paired test with respect to A1.

10. Complete segmentation results when dividing database by signal to noise ratio (SNR)

Table S7. Segmentation results compared to expert analyst 1 (A1) dividing by SNR

Geometry	Method	Number of unprocessed images	CIMT$_{bias}$ (µm)	CIMT$_{AbsBias}$ (µm)	HDM$_{LI}$ (µm)	HDM$_{MA}$ (µm)
Low SNR (48)	A1’	5	65 ± 241	192 ± 157	281 ± 126	273 ± 124
	A2	5	58 ± 254*	211 ± 149	262 ± 96	253 ± 129
	CNR$_{IT}$ [2,3]	6	-60 ± 209	162 ± 143	278 ± 127	338 ± 156
	INESCTEC$_{PT}$ [4,5]	8	36 ± 486*	211 ± 438	411 ± 575	234 ± 157
	TUM$_{DE}$ [6,7]	5	-75 ± 186	161 ± 118	231 ± 116	251 ± 100
	UCY$_{CY}$ [8]	5	-50 ± 234*	179 ± 157	229 ± 94	268 ± 128
	POLITO$^{Snakes}_{IT}$ [9]	5	30 ± 307*	241 ± 188	354 ± 216	256 ± 162
	POLITO$^{UNet}_{IT}$	5	192 ± 199	229 ± 153	223 ± 112	248 ± 111
	CREATIS$_{FR}$ [10]	5	24 ± 210*	162 ± 132	229 ± 120	218 ± 107
Average SNR (295)	A1’	8	-11 ± 260	202 ± 161	297 ± 134	255 ± 111
	A2	8	-64 ± 220	171 ± 151	282 ± 124	256 ± 93
	CNR$_{IT}$ [2,3]	10	-78 ± 162	136 ± 117	259 ± 104	303 ± 118
	INESCTEC$_{PT}$ [4,5]	42	-79 ± 157	134 ± 113	267 ± 129	215 ± 97
	TUM$_{DE}$ [6,7]	8	-112 ± 159	146 ± 128	228 ± 75	256 ± 106
	UCY$_{CY}$ [8]	8	-89 ± 170	140 ± 130	225 ± 88	245 ± 105
	POLITO$^{Snakes}_{IT}$ [9]	8	95 ± 201	187 ± 118	335 ± 141	270 ± 129
	POLITO$^{UNet}_{IT}$	8	92 ± 154	145 ± 106	231 ± 80	209 ± 85
	CREATIS$_{FR}$ [10]	8	-7 ± 157*	112 ± 108	215 ± 87	210 ± 92
High SNR (57)	A1’	4	262 ± -11*	175 ± 202	129 ± 297	143 ± 255
	A2	4	279 ± -64*	184 ± 171	128 ± 282	161 ± 256
	CNR$_{IT}$ [2,3]	4	186 ± -78	126 ± 136	114 ± 259	146 ± 303
	INESCTEC$_{PT}$ [4,5]	9	361 ± -79	291 ± 134	720 ± 267	405 ± 215
	TUM$_{DE}$ [6,7]	4	187 ± -112	133 ± 146	115 ± 228	125 ± 256
	UCY$_{CY}$ [8]	4	193 ± -89	133 ± 140	111 ± 225	124 ± 245

	POLITO$^{Snakes}_{IT}$ [9]	4	265 ± 95	177 ± 187	174 ± 335	176 ± 270
	POLITO$^{UNet}_{IT}$	4	176 ± 92	134 ± 145	112 ± 231	101 ± 209
	CREATIS$_{FR}$ [10]	4	177 ± -7*	115 ± 112	105 ± 215	99 ± 210

A1: Manual segmentations of analyst 1; A1: Manual segmentations of analyst 1; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; CIMT: carotid intima-media thickness. HDM: Hausdorff distance metric. The asterisk (*) defines measurements without statistically significant differences ($p > 0.05$) using the Wilcoxon paired test with respect to A1.

11. Complete segmentation results when dividing database by center

Table S8. Segmentation results compared to expert analyst 1 (A1) dividing by center

Center	Method	Number of unprocessed images	CIMT$_{bias}$ (µm)	CIMT$_{AbsBias}$ (µm)	HDM$_{LI}$ (µm)	HDM$_{MA}$ (µm)
Munich	A1’	4	24 ± 222*	160 ± 155	245 ± 114	281 ± 150
	A2	4	15 ± 234*	174 ± 156	241 ± 112	240 ± 132
	CNR$_{IT}$ [2,3]	5	-084 ± 193	153 ± 143	270 ± 123	344 ± 186
	INESCTEC$_{PT}$ [4,5]	9	50 ± 403*	231 ± 332	627 ± 957	386 ± 616
	TUM$_{DE}$ [6,7]	4	-111 ± 181	158 ± 138	220 ± 104	260 ± 140
	UCY$_{CY}$ [8]	4	-92 ± 201	164 ± 147	228 ± 110	263 ± 136
	POLITO$^{Snakes}_{IT}$ [9]	4	37 ± 314	234 ± 211	382 ± 234	309 ± 179
	POLITO$^{UNet}_{IT}$	4	155 ± 179	188 ± 137	225 ± 108	233 ± 109
	CREATIS$_{FR}$ [10]	4	-9 ± 187*	137 ± 127	211 ± 114	206 ± 111
Pisa	A1’	2	93 ± 257	202 ± 182	307 ± 146	271 ± 130
	A2	2	84 ± 248	205 ± 161	277 ± 117	286 ± 152
	CNR$_{IT}$ [2,3]	2	-65 ± 162	131 ± 114	286 ± 124	330 ± 139
	INESCTEC$_{PT}$ [4,5]	18	76 ± 457*	208 ± 411	499 ± 663	258 ± 201
	TUM$_{DE}$ [6,7]	2	-110 ± 166	156 ± 121	239 ± 114	271 ± 118
	UCY$_{CY}$ [8]	2	-70 ± 179	145 ± 124	235 ± 113	258 ± 120
	POLITO$^{Snakes}_{IT}$ [9]	2	123 ± 201	192 ± 134	404 ± 164	285 ± 166
	POLITO$^{UNet}_{IT}$	2	188 ± 158	203 ± 132	243 ± 113	236 ± 100

	CREATIS$_{FR}$ [10]	2	-10 ± 160*	118 ± 108	219 ± 115	194 ± 92
Porto	A1’	5	-20 ± 218*	173 ± 133	279 ± 107	242 ± 134
	A2	5	-20 ± 252*	193 ± 162	275 ± 140	254 ± 136
	CNR$_{IT}$ [2,3]	6	-73 ± 177	147 ± 120	284 ± 131	295 ± 136
	INESCTEC$_{PT}$ [4,5]	5	-76 ± 176	145 ± 123	298 ± 157	211 ± 106
	TUM$_{DE}$ [6,7]	5	-101 ± 182	160 ± 128	255 ± 126	251 ± 117
	UCY$_{CY}$ [8]	5	-85 ± 188	156 ± 132	256 ± 119	253 ± 121
	POLITO$^{Snakes}_{IT}$ [9]	5	105 ± 214	189 ± 140	350 ± 167	268 ± 149
	POLITO$^{UNet}_{IT}$	5	119 ± 170	163 ± 122	247 ± 114	208 ± 100
	CREATIS$_{FR}$ [10]	5	16 ± 165*	130 ± 103	221 ± 104	215 ± 107
Torino	A1’	6	132 ± 294	247 ± 201	295 ± 171	303 ± 159
	A2	6	129 ± 307	246 ± 220	265 ± 154	331 ± 193
	CNR$_{IT}$ [2,3]	7	-50 ± 192	144 ± 134	269 ± 130	312 ± 156
	INESCTEC$_{PT}$ [4,5]	27	-16 ± 167*	108 ± 127	322 ± 356	230 ± 143
	TUM$_{DE}$ [6,7]	6	-82 ± 192	150 ± 141	232 ± 124	256 ± 137
	UCY$_{CY}$ [8]	6	-53 ± 196	143 ± 143	227 ± 114	264 ± 143
	POLITO$^{Snakes}_{IT}$ [9]	6	121 ± 282	232 ± 194	428 ± 182	350 ± 205
	POLITO$^{UNet}_{IT}$	6	182 ± 193	205 ± 156	247 ± 130	233 ± 122
	CREATIS$_{FR}$ [10]	6	-5 ± 186*	133 ± 129	204 ± 111	208 ± 112
Toronto (FAMUS)	A1’	5	78 ± 178	146 ± 126	219 ± 104	211 ± 119
	A2	5	-6 ± 247*	186 ± 160	252 ± 117	259 ± 139
	CNR$_{IT}$ [2,3]	5	-92 ± 126	116 ± 098	286 ± 125	338 ± 151
	INESCTEC$_{PT}$ [4,5]	70	-54 ± 081*	32 ± 76	241 ± 120	164 ± 088
	TUM$_{DE}$ [6,7]	5	-117 ± 125	130 ± 106	203 ± 087	272 ± 118
	UCY$_{CY}$ [8]	5	-74 ± 133	117 ± 95	238 ± 104	265 ± 119
	POLITO$^{Snakes}_{IT}$ [9]	5	-311 ± 250	301 ± 243	494 ± 174	792 ± 280
	POLITO$^{UNet}_{IT}$	5	154 ± 143	166 ± 120	182 ± 106	226 ± 108
	CREATIS$_{FR}$ [10]	5	25 ± 110	88 ± 70	175 ± 077	174 ± 081

A1: Manual segmentations of analyst 1; A1: Manual segmentations of analyst 1; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; CIMT: carotid intima-media thickness. HDM: Hausdorff distance metric. The asterisk (*) defines measurements without statistically significant differences ($p > 0.05$) using the Wilcoxon paired test with respect to A1.

12. Complete segmentation results in FAMUS dataset using ground truth profiles

Table S9. Segmentation results compared to ground truth in FAMUS dataset

Center	Method	N° unprocessed images	CIMT$_{bias}$ (µm)	CIMT$_{AbsBias}$ (µm)	HDM$_{LI}$ (µm)	HDM$_{MA}$ (µm)
FAMUS	A1’	5	122 ± 143	149 ± 115	236 ± 86	256 ± 108
	A2	5	200 ± 160	212 ± 143	225 ± 79	307 ± 142
	CNR$_{IT}$ [2,3]	5	116 ± 248	201 ± 176	297 ± 111	304 ± 167
	INESCTEC$_{PT}$ [4,5]	5	31 ± 73	59 ± 53	242 ± 110	213 ± 80
	TUM$_{DE}$ [6,7]	70	47 ± 65	66 ± 44	213 ± 85	182 ± 87
	UCY$_{CY}$ [8]	5	5 ± 58*	44 ± 37	171 ± 63	166 ± 62
	POLITO$^{Snakes}_{IT}$ [9]	5	48 ± 087	79 ± 060	197 ± 76	177 ± 82
	POLITO$^{UNet}_{IT}$	5	-189 ± 183	205 ± 165	453 ± 143	620 ± 213
	CREATIS$_{FR}$ [10]	5	276 ± 86	279 ± 78	177 ± 92	360 ± 101
	A1’	5	147 ± 75	148 ± 73	192 ± 73	298 ± 92

A1: Manual segmentations of analyst 1; A1: Manual segmentations of analyst 1; CNR$_{IT}$: Method from Consiglio Nazionale delle Ricerche; INESCTEC$_{PT}$: INESC Technology and Science; TUM$_{DE}$: Technische Universität München; UCY$_{CY}$: University of Cyprus; POLITO$^{Snakes}_{IT}$: Politecnico di Torino (dual snakes); POLITO$^{UNet}_{IT}$: Politecnico di Torino (UNet); CREATIS$_{FR}$: CREATIS lab Université de Lyon; CIMT: carotid intima-media thickness. HDM: Hausdorff distance metric. The asterisk (*) defines measurements without statistically significant differences ($p > 0.05$) using the Wilcoxon paired test with respect to A1.

6. Segmentation results divided by cross-validation fold for Deep Learning methods

Table S10 shows the cIMT Bias with respect to A1 for both the deep learning based methods dividing the results between each fold in the cross validation process. In this table, the results are calculated on the Test set.

Table S10. Segmentation results divided by cross-validation fold (test set)

Fold	cIMT Bias (µm)		cIMT Abs. Bias (µm)		HDM LI (µm)		HDM MA (µm)	
	CREATIS$_{\text{FR}}$ [10]	POLITO$^{\text{UNet}}_{\text{IT}}$	CREATIS$_{\text{FR}}$ [10]	POLITO$^{\text{UNet}}_{\text{IT}}$	CREATIS$_{\text{FR}}$ [10]	POLITO$^{\text{UNet}}_{\text{IT}}$	CREATIS$_{\text{FR}}$ [10]	POLITO$^{\text{UNet}}_{\text{IT}}$
0	-8 ± 169*	152 ± 169	128 ± 110	190 ± 125	205 ± 90	219 ± 107	193 ± 105	222 ± 105
1	0 ± 138*	181 ± 151	111 ± 81	201 ± 122	198 ± 105	234 ± 106	178 ± 75	208 ± 94
2	2 ± 183*	145 ± 173	133 ± 125	188 ± 124	213 ± 117	229 ± 112	211 ± 106	229 ± 106
3	20 ± 176*	186 ± 191	132 ± 117	217 ± 155	205 ± 91	238 ± 124	214 ± 96	252 ± 104
4	2 ± 169*	135 ± 165	128 ± 110	172 ± 126	209 ± 93	224 ± 96	198 ± 90	223 ± 85

POLITO$^{\text{UNet}}_{\text{IT}}$: Politecnico di Torino (UNet); CREATIS$_{\text{FR}}$: CREATIS lab Université de Lyon; the asterisk (*) defines measurements without statistically significant differences ($p > 0.05$) using the t-test on cIMT Bias.

7. Segmentation results in validation sets for each fold in deep learning methods

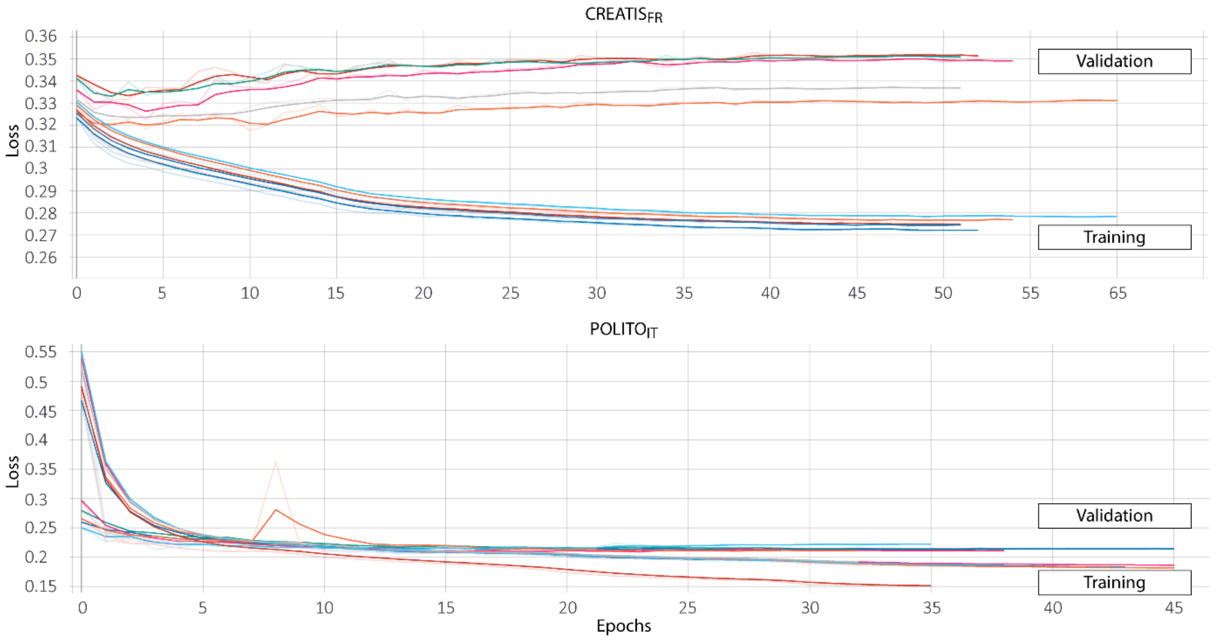


Figure S7. Evolution of the losses of the deep learning methods during the learning process: CREATIS$_{\text{FR}}$ (top) and POLITO$^{\text{Unet}}_{\text{IT}}$ (bottom).

Figure S7 shows the evolution of the losses of the deep learning methods (CREATIS$_{FR}$ and POLITO$^{Unet}{}_{IT}$) during the learning process. In the top part of the figure (CREATIS$_{FR}$ method), we note an overfitting which is justified by the use of transfer learning. The CREATIS team has pre-trained their models on another annotated database. During transfer learning, the learning rate was reduced to 0.0001, which resulted in overfitting. Therefore, only the models minimizing the loss in each fold on the validation subset (namely between the first and tenth epoch) were saved and used for the subsequent testing. Figure S7 (bottom) shows the losses of the POLITO$_{UNet}{}^{IT}$ method. This model was trained starting from the weights trained on ImageNet, showing a much higher loss in the first epochs. For 4 out of 5 folds training and validation perfromances started to diverge from epoch 20.

REFERENCES

1. Pham, A.; Lundgren, B.; Stage, B. Shadow effects in simulated ultrasound images derived from computed tomography images using a focused beam tracing model. *J. Acoust. Soc. Am.* **2012**, *132*, 487, doi:10.1121/1.4726031.
2. Bianchini, E.; Giannarelli, C.; Maria Bruno, R.; Armenia, S.; Landini, L.; Faita, F.; Gemignani, V.; Taddei, S.; Ghiadoni, L. Functional and Structural Alterations of Large Arteries: Methodological Issues. *Curr. Pharm. Des.* **2013**, *19*, 2390–2400.
3. Faita, F.; Gemignani, V.; Bianchini, E.; Giannarelli, C.; Ghiadoni, L.; Demi, M. Real-time Measurement System for Evaluation of the Carotid Intima-Media Thickness With a Robust Edge Operator. *J. Ultrasound Med.* **2008**, *27*, 1353–1361, doi:10.7863/jum.2008.27.9.1353.
4. Rouco, J.; Carvalho, C.; Domingues, A.; Azevedo, E.; Campilho, A. A robust anisotropic edge detection method for carotid ultrasound image processing. *Procedia Comput. Sci.* **2018**, *126*, 723–732, doi:10.1016/J.PROCS.2018.08.006.
5. Rocha, R.; Silva, J.; Campilho, A. Automatic segmentation of carotid B-mode images using fuzzy classification. *Med. Biol. Eng. Comput.* **2012**, *50*, 533–545, doi:10.1007/s11517-012-0883-y.
6. Zahnd, G.; Kapellas, K.; van Hattem, M.; van Dijk, A.; Sérusclat, A.; Moulin, P.; van der Lugt, A.; Skilton, M.; Orkisz, M. A Fully-Automatic Method to Segment the Carotid Artery Layers in Ultrasound Imaging: Application to Quantify the Compression-Decompression Pattern of the Intima-Media Complex During the Cardiac Cycle. *Ultrasound Med. Biol.*

2017, *43*, 239–257, doi:10.1016/J.ULTRASMEDBIO.2016.08.016.

7. Zahnd, G.; Orkisz, M.; Dávila Serrano, E.E.; Vray, D. CAROLAB – A platform to analyze carotid ultrasound data. In Proceedings of the IEEE International Ultrasonics Symposium (IUS), Glasgow (Scotland); 2019; pp. 463–466.
8. Loizou, C.P.; Pattichis, C.S.; Pantziaris, M.; Tyllis, T.; Nicolaides, A. Snakes based segmentation of the common carotid artery intima media. *Med. Biol. Eng. Comput.* **2007**, *45*, 35–49, doi:10.1007/s11517-006-0140-3.
9. Molinari, F.; Meiburger, K.M.; Saba, L.; Zeng, G.; Acharya, U.R.; Ledda, M.; Nicolaides, A.; Suri, J.S. Fully automated dual-snake formulation for carotid intima-media thickness measurement. A new approach. *J. Ultrasound Med.* **2012**, *31*, 1123–36, doi:10.7863/jum.2012.31.7.1123.
10. Lainé, N.; Zahnd, G.; Liebgott, H.; Orkisz, M. Carotid artery wall segmentation in ultrasound image sequences using a deep convolutional neural network. In Proceedings of the International Symposium on Biomedical Imaging; 2022; p. Submitted.