IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 February 2023, accepted 6 March 2023, date of publication 16 March 2023, date of current version 21 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3258408

RESEARCH ARTICLE

UR-CarA-Net: A Cascaded Framework With Uncertainty Regularization for Automated Segmentation of Carotid Arteries on Black Blood MR Images

ELIZAVETA LAVROVA[1,2], ZOHAIB SALAHUDDIN[2], HENRY C. WOODRUFF[2,3], MOHAMED KASSEM[3,4], ROBIN CAMARASA[5], ANJA G. VAN DER KOLK[6,7], PAUL J. NEDERKOORN[8], DANIEL BOS[5,9], JEROEN HENDRIKSE[6], M. ELINE KOOI[3,4], AND PHILIPPE LAMBIN[2,3]

[1]GIGA Cyclotron Research Centre In Vivo Imaging, University of Liège, 4000 Liège, Belgium
[2]The D-Laboratory, Department of Precision Medicine, GROW—School for Oncology and Reproduction, Maastricht University, 6200 MD Maastricht, The Netherlands
[3]Department of Radiology and Nuclear Medicine, Maastricht University Medical Centre+, 6229 HX Maastricht, The Netherlands
[4]CARIM School for Cardiovascular Diseases, Maastricht University, 6200 MD Maastricht, The Netherlands
[5]Department of Radiology and Nuclear Medicine, Erasmus University Medical Center Rotterdam, 3015 GD Rotterdam, The Netherlands
[6]Department of Radiology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands
[7]Department of Medical Imaging, Radboudumc, 6525 GA Nijmegen, The Netherlands
[8]Department of Neurology, Amsterdam UMC, 1007 MB Amsterdam, The Netherlands
[9]Department of Epidemiology, Erasmus University Medical Center Rotterdam, 3015 GD Rotterdam, The Netherlands

Corresponding author: Elizaveta Lavrova (e.lavrova@maastrichtuniversity.nl)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program CHAIMELEON under Grant 952172, in part by EuCanImage under Grant 952103, in part by the TRANSCAN Joint Transnational Call 2016 (JTC2016 CLEARLY) under Grant UM 2017-8295, and in part by the Innovative Medicines Initiative (IMI)-Optimal Treatment for Patients with Solid Tumours in Europe Through Artificial intelligence (OPTIMA) under Grant 101034347. The work of Elizaveta Lavrova was supported by the Liege-Maastricht Imaging Valley Grant.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Ethical Committees of Amsterdam Medical Center, Erasmus Medical Center, Maastricht University Medical Center+, and University Medical Center Utrecht.

ABSTRACT We present a fully automated method for carotid artery (CA) outer wall segmentation in black blood MRI using partially annotated data and compare it to the state-of-the-art reference model. Our model was trained and tested on multicentric data of patients (106 and 23 patients, respectively) with a carotid plaque and was validated on different MR sequences (24 patients) as well as data that were acquired with MRI systems of a different vendor (34 patients). A 3D nnU-Net was trained on pre-contrast T1w turbo spin echo (TSE) MR images. A CA centerline sliding window approach was chosen to refine the nnU-Net segmentation using an additionally trained 2D U-Net to increase agreement with manual annotations. To improve segmentation performance in areas with semantically and visually challenging voxels, Monte-Carlo dropout was used. To increase generalizability, data were augmented with intensity transformations. Our method achieves state-of-the-art results yielding a Dice similarity coefficient (DSC) of 91.7% (interquartile range (IQR) 3.3%) and volumetric intraclass correlation (ICC) with ground truth of 0.90 on the development domain data and a DSC of 91.1% (IQR 7.2%) and volumetric ICC with ground truth of 0.83 on the external domain data outperforming top-ranked methods for open-source CA segmentation. The uncertainty-based approach increases the interpretability of the proposed method by providing an uncertainty map together with the segmentation.

INDEX TERMS Auto-segmentation, carotid MRI, vessel segmentation, U-Net, uncertainty regularization.

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

I. INTRODUCTION

Stroke is the second leading cause of death and a leading cause of disability in adults worldwide [1]. Carotid artery

(CA) atherosclerosis is one of the major causes of stroke as it can lead to the formation of an embolus from an atherosclerotic plaque or hypoperfusion due to narrowing of the CA lumen. Current risk assessment and treatment decision strategies for patients with carotid artery stenosis due to the presence of a carotid plaque who suffered a recent transient ischemic attack (TIA) or stroke are based on the degree of CA stenosis [2]. However, recent studies have shown that plaque morphology and composition can improve stroke prediction [3], [4], [5], [6].

Modern medical imaging techniques, such as ultrasound (US), computed tomography angiography (CTA), and magnetic resonance imaging (MRI) help shed light on CA plaque characteristics in a non- or minimally-invasive manner. MRI, especially when combining multiple contrast weightings with a reference black blood T1-weighted scan, can provide extensive information about the CA, plaque morphology, and even plaque subcomponents in 3D without the adverse effects of radiation dose on the patient [5], [7]. One of the challenges of MRI is the fact that manual CA plaque characterization is time-consuming and subjective [8]. Therefore, there is a demand to automate this process.

In the last decades, several research groups reported on various methods to enable automated plaque characterization on multi-contrast MRI by segmenting plaque components. Computer vision approaches such as shape fitting, active contours, and level sets, in combination with simple machine learning methods such as classification and clustering, were attempted early on [8], [9], [10], [11], [12], [13], [15], [16], [17], [18], [19], [20], [21], and [14]. In more recent years, convolutional neural networks (CNN), including U-Net, have gained increasing attention [22], [23], [24], [25], [26], [27], [28]. Whereas in early works, mostly area or volume differences were used as segmentation evaluation metrics, [10] was one of the first to report traditional segmentation scores such as Dice Score Coefficient and Hausdorff distance.

When analyzing the many techniques that have attempted to characterize CA plaque and its subcomponents, it becomes clear that the levels of automation are widely disparate. Even though plaque characterization methods are automated, they require CA localization first, and this step is performed manually. In some studies, readers need to delineate the outer wall and lumen on every MRI slice [8], [14], [15], [16], [17], [22]. In other studies, plaque characterization needs manually pre-cropped regions containing the CA or lumen seed points in every slice as an input [9], [23], [26], [27]. Sometimes even additional delineation of a muscle region is needed for intensity re-scaling [9], [15], [16]. While there are some studies where the CA was located using the lumen seed points in the distal slices alone, or in different CA branches [11], [12], [13], [18], [19], user interaction is still necessary, and no studies show robustness to seed positioning. A couple of publications report on automated CA localization, but the detection area is limited to manually selected slices [10]. These approaches are time-consuming and introduce inter- and intra-reader variability. Moreover, with a multi-contrast approach, different MRI scans should be co-registered, which is at the moment performed in a semi-automated manner. Therefore, to increase the speed, robustness, reliability, and reproducibility of advanced quantitative CA analysis, while decreasing cost and clinical burden, automated detection and segmentation of the CA on black blood MRI sequences remains an unmet clinical need.

There are a number of challenges related to CA segmentation. First of all, currently popular deep learning-based semantic segmentation approaches show robust performance but require large amounts of fully annotated training data [29]. In most datasets, only the symptomatic CA is delineated. Moreover, in most cases, only the internal CA is delineated meaning that in addition to the external branch of the symptomatic CA and the entire asymptomatic CA, vertebral arteries are also not delineated. Even though those arteries are not of high clinical interest, for a deep learning network they are essentially the same objects as the CA. Other neck arteries are also present on the slice, and if their external walls and lumens are clear on the image, being unlabelled, they can confuse a segmentation model. Second, multi-contrast MRI often experiences a domain shift caused by different acquisition protocols [30]. Therefore, models trained on the data from some particular scanner and acquisition protocol might not be performing well on the data from slightly different acquisition settings. Third, MRI is expressed in arbitrary units [31]. Therefore, segmentation by thresholding characteristic physical density units, such as Hounsfield units, is problematic. Fourth, ground-truth segmentation is usually performed on the reference MR sequence, and the other black blood sequences are rigidly coregistered in a semi-automated manner, which might cause co-registration issues related to patient movement. Finally, Most modern deep learning applications do not provide uncertainty estimations of the segmentations and retail "black boxes" regarding the interpretability of the outputs. Nevertheless, at the moment, the demand for interpretable methods is growing [32]. At the moment, there are several studies aimed to perform the whole CA segmentation [27], [28], [33], but they all used the data from the same domain for both training and evaluation of their models.

To address these issues, we aimed to develop an automated method for common and internal CA detection on black blood MRI which would be robust against image quality and acquisition protocol variations and could be trained on partially segmented data, as well as providing uncertainty estimates of the segmentations generated. For this, we trained a nnU-Net [34], known for its high semantic segmentation performance in similar tasks [35], on lateral halves of T1-weighted (T1w) scans containing CA outer wall contours. After an nnU-Net model was trained to perform semantic segmentation of CA on 3D MRI scans, an additional U-Net was trained on 2D patches, containing the CA contour, to refine

the output contours in anatomically challenging slices such as bifurcation areas and increase the generalizability of the method on the external data. To refine segmentation in visually challenging voxels, we incorporated an uncertainty component into the loss function using Monte-Carlo dropout during training. Within this study, we hypothesized that the proposed workflow would be able to segment CA on black blood MRI robustly regardless of black blood sequence, contrast media, or MRI vendor. Our secondary hypothesis was that these nnU-Net refinements will improve the segmentation. We refer to the refined segmentation network as ''U-CarA-Net'' (short for ''U-Net-based carotid artery segmentation''), and the network with added uncertainty regularization as ''UR-CarA-Net'' (short for ''U-Net-based carotid artery segmentation with regularization'').

We tested U-CarA-Net and UR-CarA-Net on the training data as well as on data from an external center. Additionally, we tested the method on the regular T2w TSE scans and post-contrast T1w TSE scans. We performed an ablation study comparing the segmentations produced by the backbone nnU-Net, U-CarA-Net, and UR-CarA-Net. Our contribution in this work lies in the implementation of a fully automated model for the segmentation of CA in two clinically relevant MRI sequences which can be trained on partially labeled data as only the symptomatic CA was delineated - and the addition of an uncertainty estimate of the segmentations. The code is open-source and available at GitHub: https://github.com/lavrovaliz/ur-cara-net [36].

II. METHODS

A. METHOD OVERVIEW

The main steps of the proposed method are:

1) Splitting the image of the neck through the central medial plane and keeping only the halves containing ground-truth CA segmentation, keeping right side halves and reflecting left sides to preserve the anatomy;
2) Training a 3D nnU-Net on lateral halves of T1w TSE images, containing ground-truth CA contours, for CA segmentation;
3) Training the U-CarA-Net on 2-channel 2D patches localized along the CA centerline localized by the nnU-Net;
4) Training the UR-CarA-Net on 2-channel 2D patches localized along the CA centerline using U-CarA-Net weights as initial weights.

The pipeline is illustrated in Figure 1. A detailed description of the steps is presented in the next sections.

B. MRI SCAN SPLIT

As ground-truth segmentations were available only for the symptomatic side, we split all the scans with the central median plane to deal with the missing contra-lateral segmentation. For training and validation, we selected the halves containing the ground-truth segmentations. To preserve the anatomy for the nnU-Net model, we keep the right side halves untouched and the left side halves are reflected relative to the median plane assuming lateral symmetry of the neck is not disturbed significantly with the minor deviations in the positioning of the patients.

C. CA SEGMENTATION WITH nnU-NET

To perform the initial CA segmentation, we trained a 3D full-resolution nnU-Net [34] on T1w TSE data. The method described performs data pre-processing and adapts its training and data augmentation parameters according to the properties of the training data. We used the standard settings from the original implementation including a combination of Dice similarity coefficient (DSC) and cross-entropy as a loss function, stochastic gradient descent with Nesterov momentum as an optimizer, poly learning rate schedule with an initial learning rate of 0.01, and training on 1000 epochs.

D. SEGMENTATION REFINEMENT

Using the nnU-Net trained to segment CA in black blood T1w TSE carotid scans, we were able to localize the CA centerline and get the CA contour information. To refine the contours, given there could be cases of slight co-registration errors in the data from the different sequences, we trained a 2D U-Net [37] to segment the CA on square MRI slice patches, moving along the CA centerline.

For every slice, the patch center was placed into the center mass of the nnU-Net outcome. As a 2-channel input for the U-CarA-Net, we used corresponding patches of the T1w slice and the nnU-Net softmax output. The nnU-Net softmax output gives information about the reference segmentation as well as uncertainty information by labeling voxels with values between 0 and 1.

After the U-CarA-Net is trained, we re-trained it in the following manner to improve its performance, calling it UR-CarA-Net. Besides utilizing a 2-channel input, we incorporated uncertainty information into the loss function as suggested in [38] to improve probability calibration in the 'difficult' region. We generated the uncertainty estimate using Monte-Carlo dropout during training. We trained the model with dropout and after every epoch, we sampled outcomes obtained on the validation data from the current epoch and previous N-1 epochs. Then we calculated the value of the loss function composed of the weighted segmentation loss and uncertainty regularization component, where segmentation loss had DSC and cross-entropy components.

In the following, $\mathcal{P} \subset \mathrm{R}^2$ denotes the set pixels of a patch, $\theta \in \Omega$ are the parameters of the model, y_p represents the ''ground-truth'' label of the pixel $p \in \mathcal{P}$, $y_{p,\theta}$ the segmentation prediction for the pixel $p \in \mathcal{P}$ with a model with parameters θ, and $\overline{y}_p = \int_{\theta\in\Omega} \mathbb{P}(y_p|\theta)\mathbb{P}(\theta|\mathcal{D})d\theta$ is the Bayesian average estimate for the pixel $p \in \mathcal{P}$ approximated drawing sets of weights $\Theta \in \Omega$ using Monte-Carlo dropout

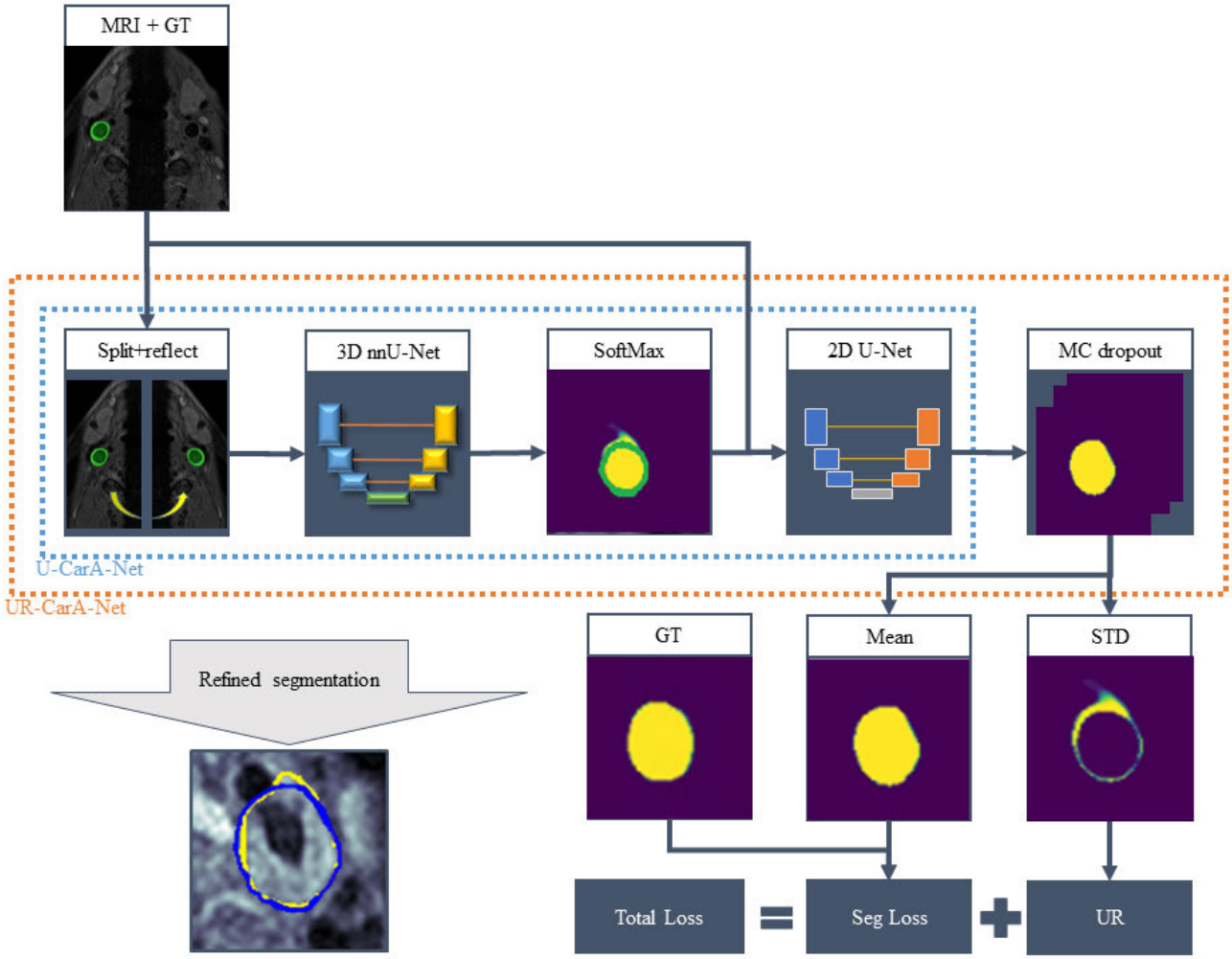


FIGURE 1. Method overview; MRI - magnetic resonance imaging, GT - ground-truth segmentation, SoftMax - nnU-Net SoftMax output, MC dropout - Monte-Carlo dropout, STD - standard deviation, Seg Loss - segmentation loss, UR - uncertainty regularization; all the models have binary outputs: CA and non-CA, models outputs are presented in purple-yellow segmentation map; the presented contours have several colors: green - ground truth, yellow - nnU-Net, blue - UR-CarA-Net.

with $\mathcal{N}_\Theta$ times sampling on the training dataset $\mathcal{D}$:

$$\mathcal{L}_{TOTAL} = \underbrace{w_{DSC}\mathcal{L}_{DSC} + w_{BCE}tanh(\frac{\gamma_{BCE}\mathcal{L}_{BCE}}{2}))}_{\text{Segmentation}} + \underbrace{w_{UNC}tanh(\frac{\gamma_{UNC}\mathcal{R}_{UNC}}{2}))}_{\text{Uncertainty}} \quad (1)$$

where $tanh : x \rightarrow \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is an hyperbolic tangent, $\mathcal{L}_{TOTAL}$ is the final loss function, the segmentation loss contains weighted DSC loss $\mathcal{L}_{DSC}$ and cross-entropy loss $\mathcal{L}_{BCE}$, $\mathcal{R}_{UNC}$ is our proposed uncertainty regularization component, the corresponding weights are w_{DSC}, w_{BCE}, and w_{UNC}, and scaling factors are γ_{BCE} and γ_{UNC}. The loss functions and regularization component are calculated as the following:

$$\mathcal{L}_{DSC} = 1 - \frac{s + 2\sum_{p\in\mathcal{P}} y_p \overline{y}_p}{s + \sum_{p\in\mathcal{P}} y_p + \sum_{p\in\mathcal{P}} \overline{y}_p}, \quad (2)$$

$$\mathcal{L}_{BCE} = -\sum_{p\in\mathcal{P}} y_p log(\overline{y}_p) + (1 - y_p) log(1 - \overline{y}_p)), \quad (3)$$

$$\mathcal{R}_{UNC} = \sum_{p\in\mathcal{P}} \frac{1}{n_\Theta} \sum_{\theta\in\Theta} (\overline{y}_p - y_{p,\theta})^2. \quad (4)$$

where s is a smoothing factor.

We implemented hyperbolic tangent as a scaling function to cross-entropy and uncertainty components since they have ranges of values different from each other and DSC. Both components have 0 as a minimal possible value, and the highest value depends on the image size in the extreme case. We used hyperbolic tangent to scale the values of these components in the range of [0, 1].

A new loss was implemented for training the U-CarA-Net model since we assumed that the learning curve plateau was reached. As we intended to improve the performance of a model trained to solve a particular task, we assumed that the U-CarA-Net weights were close to the optimal point

in the parameter hyperspace. Therefore, for UR-CarA-Net, we decreased the learning rate to find a solution close to the U-CarA-Net weights.

III. EXPERIMENTAL SET-UP

A. IMAGING DATA

This model-building procedure was registered at OSF.io (10.17605/OSF.IO/VPT2B). We used imaging data acquired within the PARISK study (clinical trials.gov NCT01208025) in Amsterdam Medical Center (center 1), Erasmus Medical Center (center 2), Maastricht University Medical Center+ (center 3), and University Medical Center Utrecht (center 4) [2]. PARISK is a large prospective multicenter study to improve recurrent stroke risk stratification based on multi-modality carotid imaging in symptomatic patients with mild to moderate CA stenosis. Inclusion criteria were a transient ischemic attack (TIA), amaurosis fugax, or minor stroke (modified Rankin scale $\leq$3) of the CA territory, CA NASCET stenosis $<$70% of the ipsilateral internal CA detected on Doppler US or CTA, and no indication for a revascularization procedure. Exclusion criteria were a probable cardiac source of embolism, a clotting disorder, severe comorbidity, standard contraindications for MRI,. Written informed consent was obtained from all patients before enrolment. MRI was performed on 3T whole-body scanners. Centers 1, 3, and 4 used an Achieva TX scanner (Philips Healthcare, Best, The Netherlands) with an eight-channel phased-array coil (Shanghai Chenguang Medical Technologies Co., Shanghai, China). Center 2 used a Discovery MR 750 system (GE Healthcare, Milwaukee, MI, USA) with a four-channel phased-array coil with an angulated setup (Machnet B.V., Roden, Netherlands). Apart from the difference in hardware and MRI protocols, the main difference is in acquired and reconstructed voxel sizes. More information is available in [2].

For this study, we used pre-contrast T1w TSE MRI. Additionally, we validated the models on T2w and contrast-enhanced T1w (T1w CE) TSE MRI. The scans were acquired with the same reconstructed slice thickness of 2 mm and contained up to 15 slices, centered on the CA bifurcation. The acquisition plane for all the protocols was transversal. The MRI protocols were described previously [2] and are summarized in Table 1. We selected patients for whom the described sequences were available. Therefore, our data contained 13, 115, 25, and 34 patients from centers 1, 2, 3, and 4, respectively.

B. IMAGE PRE-PROCESSING

The slices from centers 1, 3, and 4 were cropped to a 512 pixels $\times$ 512 pixels matrix by removing 8 border pixels from each side. The slices from center 2 were resampled with cubic interpolation to an in-plane 0.3 mm $\times$ 0.3 mm pixel size and reshaped to a 512 $\times$ 512 matrix. The reshaping was performed by padding. Intensity normalization was performed by subtracting the minimum intensity and dividing it by the intensity range for every slice.

TABLE 1. MRI scan parameters.

Sequence	T1w QIR TSE	T1w DIR FSE	T2w TSE
Center	1, 3, 4	2	1, 3, 4
TR (ms)	800	1 RR	4800
TE (ms)	10	5.2	49
Acquired voxel size (mm$\times$mm)	0.62$\times$0.67	0.55$\times$0.71	0.62$\times$0.63
Reconstructed voxel size (mm$\times$mm)	0.30$\times$0.30	0.55$\times$0.63	0.30$\times$0.30
Acquisition matrix (pixels$\times$pixels)	260$\times$240	256$\times$224	260$\times$252
Reconstruction matrix (pixels$\times$pixels)	528$\times$528	256$\times$256	528$\times$528

We used the following abbreviations: TR — repetition time, TE — echo time, RR — R wave to R wave interval (1 heartbeat), QIR — quadruple inversion recovery, TSE — turbo spin echo, DIR — double inversion recovery, FSE — fast spin echo.

C. EXPERIMENTS

We trained and evaluated three CA detection and segmentation models: the baseline nnU-Net and U-CarA-Net with and without uncertainty component in the loss function. All the models were trained, validated, and tested on the same data, for both U-CarA-Net models the same data augmentation transformations were applied to the same slices.

The patients from centers 1, 3, and 4 were split in a center-stratified manner into training, validation, and test sets in the proportions of 0.70, 0.15, and 0.15, respectively. Training and validation sets were used while training the models. Data from the test set were used to evaluate the performance scores. Data from center 2 were used for external testing. The summary of the resulting data split can be seen in Table 2.

All the data we had, we could separate into 4 domains. The pre-contrast T1w TSE data from centers 1, 3, and 4 were split into training, validation, and test sets and represented the development domain (referred to as DD). The scans from the training and validation sets were used to train the CA detection models and tune hyperparameters. The models were evaluated on the data from 3 domains different from the development domain:

1) post-contrast T1w TSE MRI from the patients of centers 1, 3, and 4 test sets (referred to as D1),
2) T2w MRI from the patients of centers 1, 2, and 3 test set (referred to as D2),
3) T1w FSE from center 2 (referred to as external domain, ED).

To investigate the generalizability of the models trained on the DD, they were evaluated on D1, D2, and ED to assess the center-specific impact of a different vendor, and different acquisition and reconstruction protocols. A summary of the data domains we used in this study is presented in Figure 2.

D. EVALUATION

To evaluate the results, we used common segmentation metrics such as the DSC and Hausdorff distance (HD). These scores are traditionally reported in segmentation studies and

TABLE 2. Patients split into training, validation, and test sets.

Center	1	2	3	4	Total
Total (N)	13	34	115	25	187
Training (N)	9	0	80	17	106
Validation (N)	2	0	17	4	23
Test (N)	2	0	18	4	23
External test (N)	0	34	0	0	34

give an understanding of the overlap and distance between the segmented areas. Additionally, we used problem-specific metrics. As for our task, it is important to preserve the centerline of the segmented area, so we used the center line Dice similarity coefficient (clDSC) to evaluate the models [39]. As we are dealing with uncertain reference annotations in D1 and D2, we utilized the normalized surface distance (NSD) [39], [40] as an uncertainty-aware score with a tolerance of 2 voxels as this is the average contouring disagreement we calculated from the manual annotations. Even though we suggest a fully automated method, we still assume an expert interaction in the cases where the method fails. To assess a value of a possible expert correction, we evaluate the relative added path length (relAPL), which is the length of the contour that has to be drawn while editing a segmentation [41] reported relative to the ground-truth contour length.

Finally, we computed the clinical metric used in cardiovascular clinical practice, the root-mean-square error between the ground truth and automatically segmented volumes (VRMSE). To measure the agreement between ground-truth and automated segmentations, we calculated the intraclass correlation (ICC) and performed Bland-Altman analysis for the CA volumes. For the non-normal distributed scores, we reported median values and interquartile ranges (IQR). For the differences in the scores, statistical significance was assessed using the two-sided Wilcoxon test.

To better gauge the performance of our model in different conditions, we investigated the influence of the slice location and image quality on 2D segmentation scores. As previously described evaluation scores assess segmentation quality in the whole 3D scan, to compare the segmentation performance in different anatomical parts of the CA, we calculated 2D DSCs in common and internal CA. We were dealing with multiple data domains originating from contrast media presence, different acquisition protocols, and equipment, resulting in different levels of noise and intensity bias, which, from the digital imaging point of view, resulted in different image contrast. Therefore, we compared 3D DSCs obtained for D1, D2, and ED scans, having the contrast within and beyond DD contrast values. As a contrast evaluation metric, we chose Michelson contrast, which characterizes areas with non-uniform textures and is used in medical imaging [42].

To evaluate the effects of adding uncertainty information, we compared the baseline model (nnU-Net) with U-CarA-Net and UR-CarA-Net. All the models were trained on the whole training and validation datasets with only the symptomatic CA labeled. All the pipelines used the same trained nnU-Net model, both U-CarA-Net and UR-CarA-Net models had the same architectures. Evaluations were performed on the test data.

E. IMPLEMENTATION DETAILS

We used the same architecture of U-Net as a backbone model for both U-CarA-Net and UR-CarA-Net. The input shape was $64 \times 64 \times 2$ since we had 2 channels and a square patch size of 64 pixels $\times$ 64 pixels. A patch size of 64×64 was selected since 64 is the smallest power of two which exceeds the root square of the maximum CA area in training data slices. Convolutional layers (Conv2D) with a kernel size of 3 pixels $\times$ 3 pixels were followed by batch normalization for faster training and to reduce overfitting. The first Conv2D consisted of 16 filters. In the contracting path, a number of filters were duplicated in every other Conv2D, resulting in 256 filters in the bottleneck Conv2D layer. In the expanding path, the number of filters was halved in every Conv2D. All the activation layers after batch normalization layers were exponential linear units (ELUs) for simplicity and generalizability, and to avoid the vanishing gradient and dying node problems, except for sigmoid activation in the last conv2D for pixel-wise CA probability prediction for a non-linearly separable problem. Activation layers were followed by 2D max-pooling layers downsampling the input with a 2×2 window and a stride equal to the pool size selecting the max value from the window. Every max-pooling layer was followed by a dropout layer to prevent overfitting and enable uncertainty regularization component in UR-CarA-Net. Every transposed Conv2D (Conv2DTranspose) had a kernel size of 3 pixels $\times$ 3 pixels, strides of 2 pixels $\times$ 2 pixels, and the same number of filters as an upcoming Conv2D. The initial learning rate of the Adam optimizer was reduced by a factor of 0.1 after 3 epochs of non-improvement of the loss function.

We trained the U-CarA-Net for 100 epochs with a batch size of 64. The dropout rate was 0.05. The initial rate of the Adam optimizer was 0.001, it was reduced by the factor of 0.1 while the learning curve was on a plateau for 3 epochs, and the lowest bound for the learning rate was 10-5. We trained UR-CarA-Net using U-CarA-Net weights as the initial weights. The dropout rate was 0.1 to increase its influence on the uncertainty regularization component. We used the same batch size and learning rate-reducing strategy as for U-CarA-Net. However, the initial learning rate of the Adam optimizer was set to 10^{-9}, and the lower bound for the learning rate was 10^{-15}. Small values of the learning rate were used since the baseline model (U-CarA-Net) was already trained to solve the task, hence, only refinement was needed. Therefore, it was important not to move far in the hyperparameter space. Weights for the loss functions components and scaling function factor were set to $w_{DSC} = 0.5$, $w_{BCE} = 0.5$, $w_{UNC} = 1$, $\gamma_{BCE} = 1$, and $\gamma_{UNC} = 1$. These values were estimated empirically while experimenting with different weighting factors using weights from [38] as a reference.

To avoid overfitting and increase the robustness of the model to different hardware, and reconstruction and

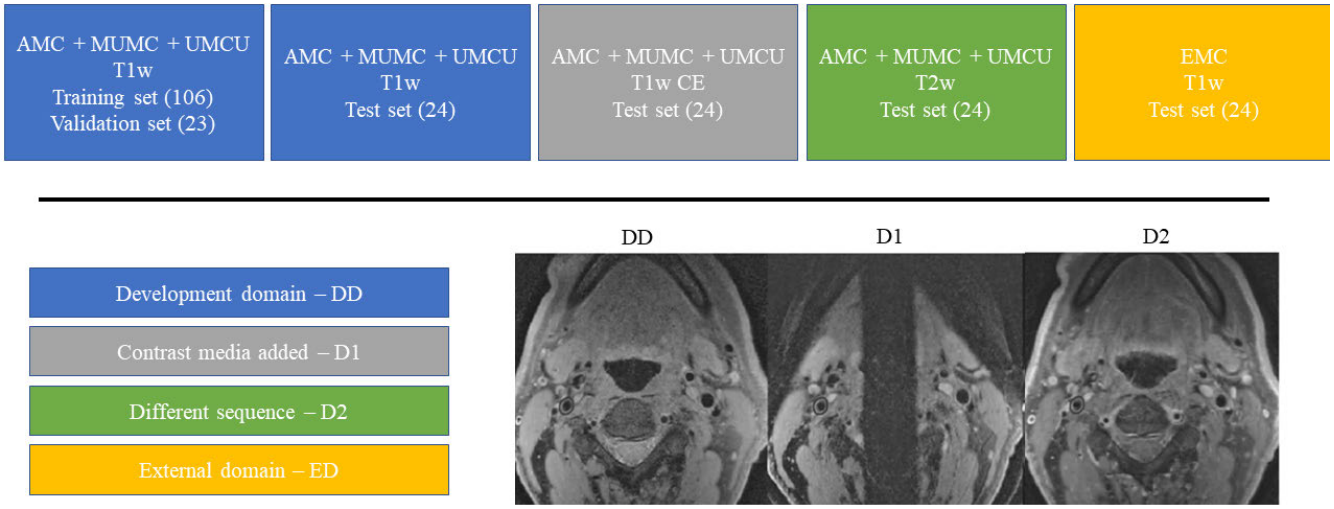


FIGURE 2. Data domain description.

TABLE 3. Comparison of carotid artery (CA) segmentation performance of the baseline (nnU-Net), improved (U-CarA-Net), and proposed (UR-CarA-Net) methods calculated on development domain (DD), first domain (D1), second domain (D2), and external domain (ED).

Domain	DSC (%)	HD (mm)	clDSC (%)	NSD (%)	relAPL	VRMSE (ml)	ICC
nnU-Net							
DD	91.6 (4.0)	2.5 (5.3)	98.4 (5.6)	99.3 (3.1)	0.28 (0.12)	0.22	0.94
D1	89.6 (12.2)	3.6 (9.6)	99.2 (27.5)	98.7 (16.2)	0.34 (0.15)	0.72	0.55
D2	89.9 (4.9)	2.6 (7.0)	100.0 (8.1)	99.6 (2.2)	0.32 (0.09)	0.30	0.83
ED	88.7 (11.6)	3.0 (3.1)	98.9 (7.2)	97.5 (11.7)	0.47 (0.21)	1.26	0.53
U-CarA-Net							
DD	91.9 (5.0)	2.3 (1.3)	100.0 (1.7)	99.5 (2.8)	0.30 (0.11)	0.28	0.88
D1	88.6 (13.2)	3.5 (6.1)	97.8 (19.0)	98.9 (16.7)	0.35 (0.19)	0.44	0.74
D2	89.7 (5.0)	2.5 (4.8)	100.0 (7.8)	99.4 (6.4)	0.33 (0.06)	0.28	0.85
ED	88.3**** (8.5)	2.9 (3.0)	98.7 (9.8)	98.2 (6.0)**	0.44 (0.19)****	0.82	0.77
UR-CarA-Net							
DD	91.7 (3.3)	2.2 (2.3)	99.6 (4.5)	99.3 (2.6)	0.28 (0.07)	0.24	0.91
D1	89.2 (11.4)	3.6 (6.3)	98.8 (26.7)	98.9 (15.7)	0.34 (0.14)	0.45	0.74
D2	89.8 (5.1)	2.5 (5.9)	100.0 (9.4)	99.5 (3.6)	0.31 (0.07)	0.27	0.86
ED	91.1 (7.2)****	2.9 (2.6)	99.2 (7.2)	98.4 (3.9)**	0.39 (0.18)****	0.68	0.83

Comparison of carotid artery (CA) segmentation performance of the baseline (nnU-Net), improved (U-CarA-Net), and proposed (UR-CarA-Net) methods calculated on development domain (DD), first domain (D1), second domain (D2), and external domain (ED). The values mentioned are median (IQR). We used the following abbreviations: DSC — Dice similarity coefficient, HD — Hausdorff distance, clCSC — centerline Dice Similarity Coefficient, NSD — normalized surface distance, VRMSE — volume root-mean-square error, ICC — volume intraclass correlation. The best scores are in bold. For DSC, JSC, clDSC, NSD, and ICC the higher the value the better whereas for HD, relAPL, and VRMSE the lower the value the better. Segmentation scores between the proposed and baseline and proposed and improved methods were compared on ED: p-values were calculated using a two-sided Wilcoxon test and referred as follows: "ns" refers to the p value in the range $5 \times 10^{-2} < p \leq 1$, "*" refers to the p value in the range $1 \times 10^{-2} < p \leq 5 \times 10^{-2}$, "**" refers to the p value in the range $1 \times 10^{-3} < p \leq 1 \times 10^{-2}$, "***" refers to the p value in the range $1 \times 10^{-4} < p \leq 1 \times 10^{-3}$, "****" refers to the p value in the range $p \leq 1 \times 10^{-4}$.

acquisition protocols, data augmentation was applied to all the training slices. The transformations had a probability of 0.5 each and mostly simulated differences in intensity distributions. The following transformations were performed: horizontal and vertical flips, blurring with a kernel size ranging from 3 to 7 pixels, Gaussian noise with 0 mean and variance in the range from 10 to 50, brightness and contrast variation by 25%, and gamma transformation with gamma ranging from 0.8 to 1.2. Every transformation parameter was a random number from the uniform distribution of the corresponding range. From every original slice used for data augmentation, 10 augmented slices were generated.

We utilized Keras 2.2.4 with a TensorFlow 1.14.0 backend. All the training and testing were performed on one NVIDIA GeForce RTX 2080 Ti.

IV. RESULTS

The resulting nnU-Net configuration can be found in the debug file in the project GitHub repository [36].

As can be seen from Table 3, the overlap-based metrics, such as DSC, clDSC, and NSD, are relatively high for all the models and data domains. The highest medium DSC is 91.9% for the U-CarA-Net in DD, the lowest is 88.3% for the U-CarA-Net in ED. The clDSC is above 97% for all the

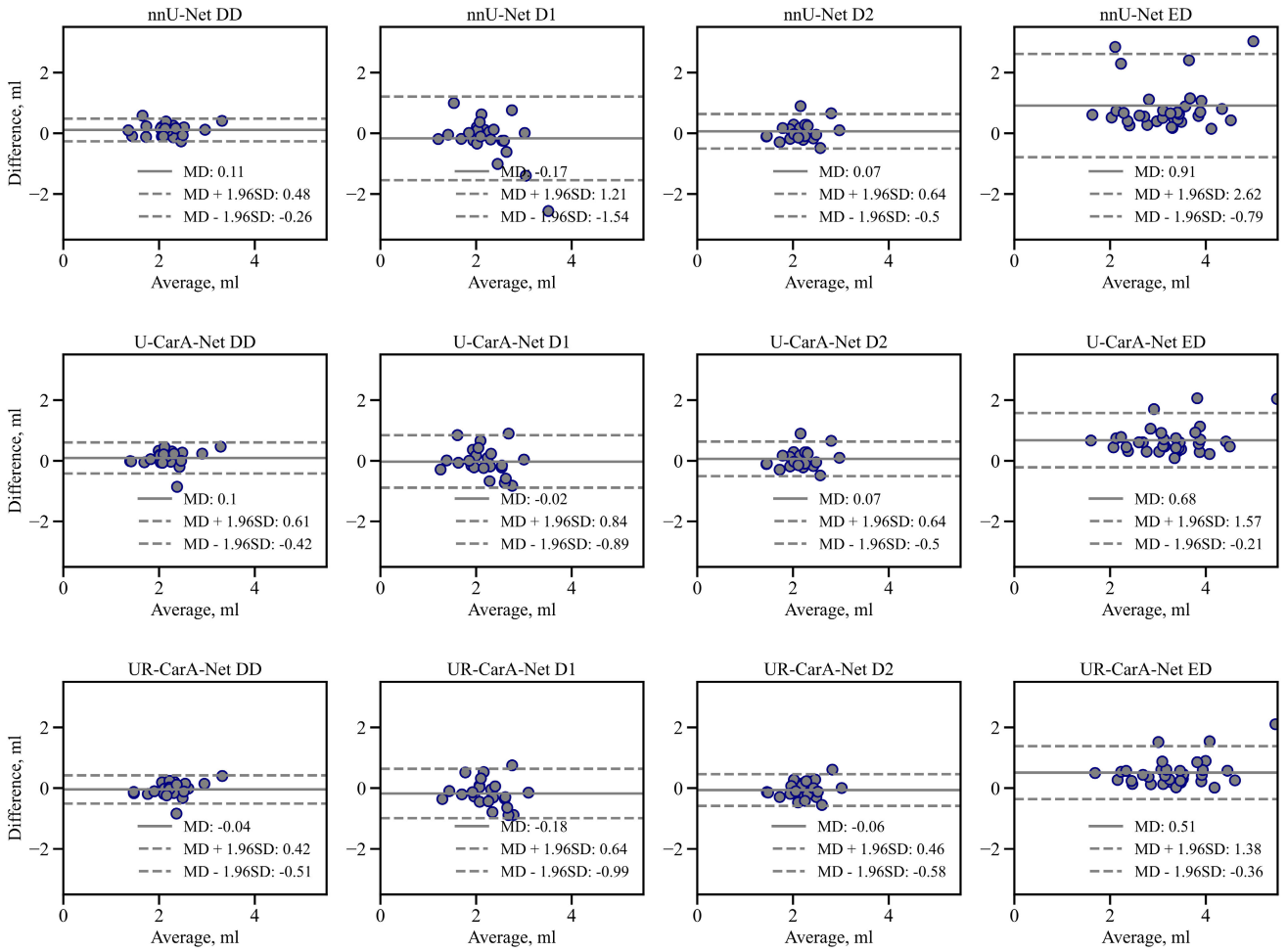


FIGURE 3. Bland-Altman plots for the models obtained on DD, D1, D2, and ED. The solid line represents the mean difference. The dashed lines represent +1.96 standard deviations (top) and −1.96 standard deviations (bottom) from the mean.

models and all the data domains, even yielding a median value of 100.0% for nnU-Net in D2, the U-CarA-Net in DD and D2, and the UR-CarA-Net in D2. NSD scores are significantly higher than the corresponding DSC, exceeding 97% for all the models and data domains.

The best performance among the data domains is achieved on the DD. The common segmentation metrics (DSC, clDSC, and NSD) yielded by the baseline model were relatively high in D1, D2, and ED. Nevertheless, clinically relevant volume-related scores, such as VRMSE, showed insufficient segmentation in these domains, especially in ED yielding a VRMSE of 1.26 compared to a VRMSE of 0.22 in DD. Low values of ICC between ground-truth and auto-segmented contours in D1 and ED show a poor agreement with the manual segmentation. Contour refinement with the U-CarA-Net improved segmentation scores for the DD and distance-based metric HD for the D1 and D2. This resulted in the improvement of VRMSE and ICC yielding better agreement with the manual segmentation. Finally, contour refinement with the UR-CarA-Net improved distance-based HD and relAPL for all the data domains. It resulted in a lower VRMSE in D2 and ED and in higher ICC for D1, D2, and ED. Moreover, for ED, compared to nnU-Net, ICC increased by 0.30 and VRMSE decreased by a factor of 2. Additionally, the UR-CarA-Net application results in IQR drop for HD, NSD, and relAPL in all the domains, as well as DSC and clDSC for DD, D1, and ED. For the ED, all the metrics obtained with the UR-CarA-Net show the best performance of this model and the lowest IQRs. Nevertheless, for the DD data, the best volume-based scores are still yielded with nnU-Net. The most challenging data domain, even for the UR-CarA-Net, was post-contrast MRI, where we could achieve an ICC of 0.74.

Even though we noticed an improvement in the distance- and volume-based metrics while using the proposed model, overlap-based metrics have high values which are close to each other. Also a statistical comparison of the interval-based estimations of the metrics is shown. We compared the proposed method (UR-CarA-Net) with the baseline (nnU-Net) and improved (U-CarA-Net) ones on the ED. As we can

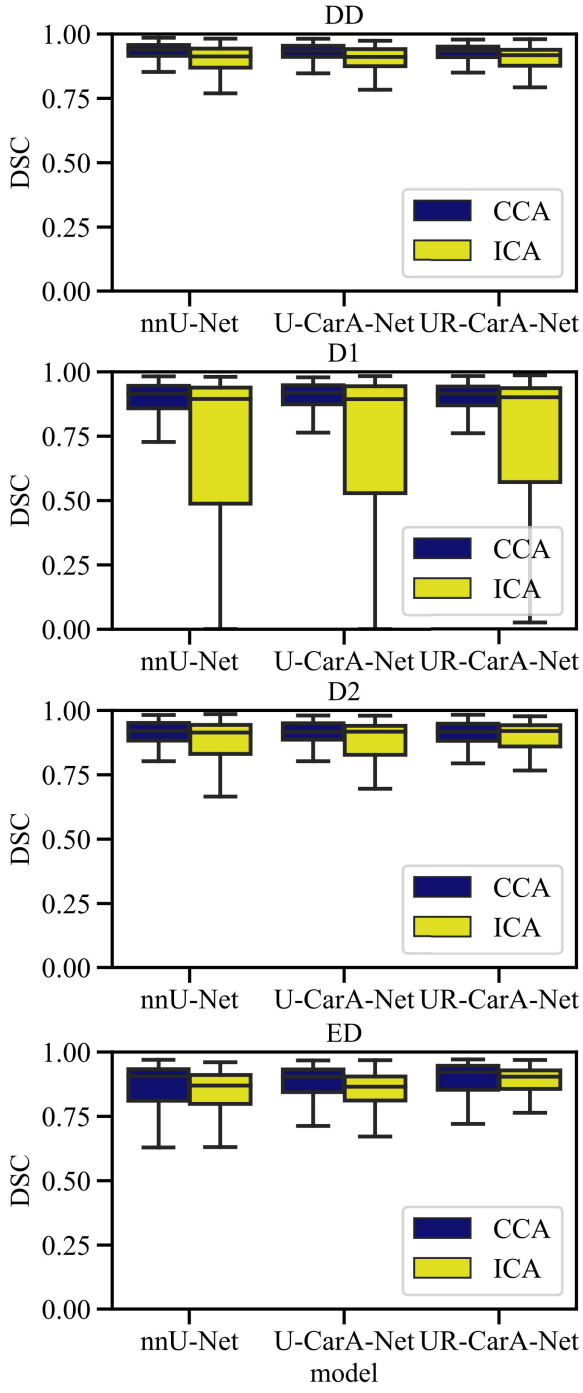


FIGURE 4. Segmentation performance with regards to carotid artery (CA) anatomy. Segmentation performance is measured in 2D Dice similarity coefficient (DSC). CA anatomy is defined as common CA (CCA) or internal CA (ICA). Data were presented as box plots, where boxes are representing the interquartile range (IQR), extending from Q1 to Q3 and centered on the median value. Upper whiskers represent the highest data point that is less than $Q3 + 1.5 \times IQR$. Lower whiskers represent the smallest data point that is greater than $Q3 - 1.5 \times IQR$.

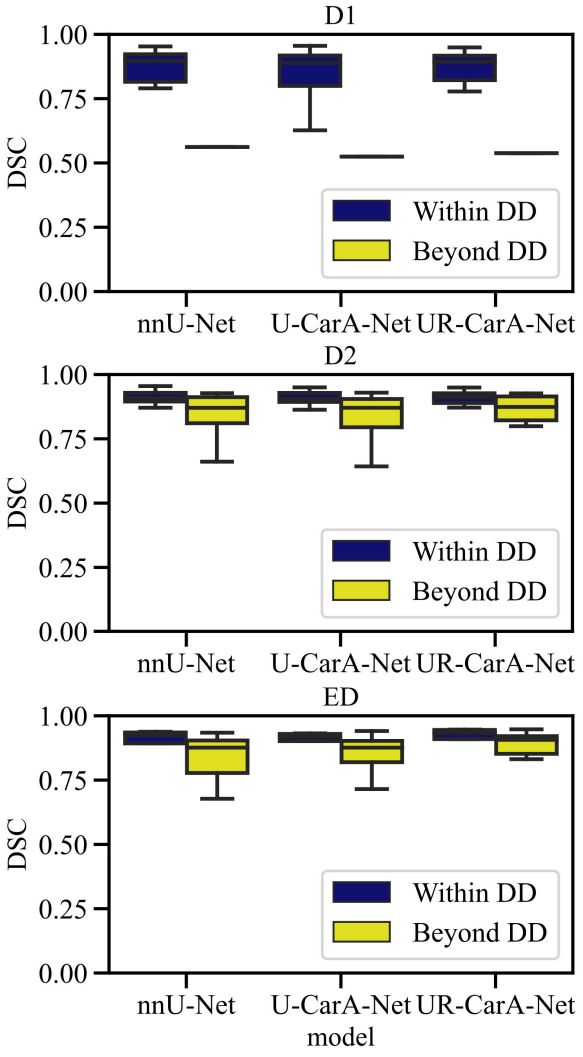


FIGURE 5. Segmentation performance with regards to the image quality corresponding or not to image quality in development domain (DD) data. Segmentation performance is measured in 3D Dice similarity coefficient (DSC). Image quality is defined in terms of Michelson contrast lying within and beyond DD values. Data were presented as box plots, where boxes are representing the interquartile range (IQR), extending from Q1 to Q3 and centered on the median value. Upper whiskers represent the highest data point that is less than $Q3 + 1.5 \times IQR$. Lower whiskers represent the smallest data point that is greater than $Q3 - 1.5 \times IQR$.

see from results table, UR-CarA-Net significantly improved DSC, NSD, and relAPL.

Bland-Altman plots in Figure 3 show, that even though all the models are yielding absolute bias values close to 0 ml, the proposed method could decrease the absolute systematic bias in DD by 0.07 ml, in D2 by 0.01 ml, and in ED by 0.40 ml. Nevertheless, for DD and D1 the lowest bias values were yielded with U-CarA-Net. We can also see that the implementation of U-CarA-Net and UR-CarA-Net increased the segmented volumes in DD, D2, and ED.

Figure 4 shows that 2D DSCs calculated in CCA are higher than in ICA. The UR-CarA-Net application decreases the range of DSC deviation, especially in ICA. Nevertheless, the lowest 2D DSC values and the widest DSC IQR are observed in D1 ICA, where U-CarA-Net and UR-CarA-Net do not improve the segmentation performance.

According to Figure 5, segmentation performance drops when the Michelson contrast is not within the range presented

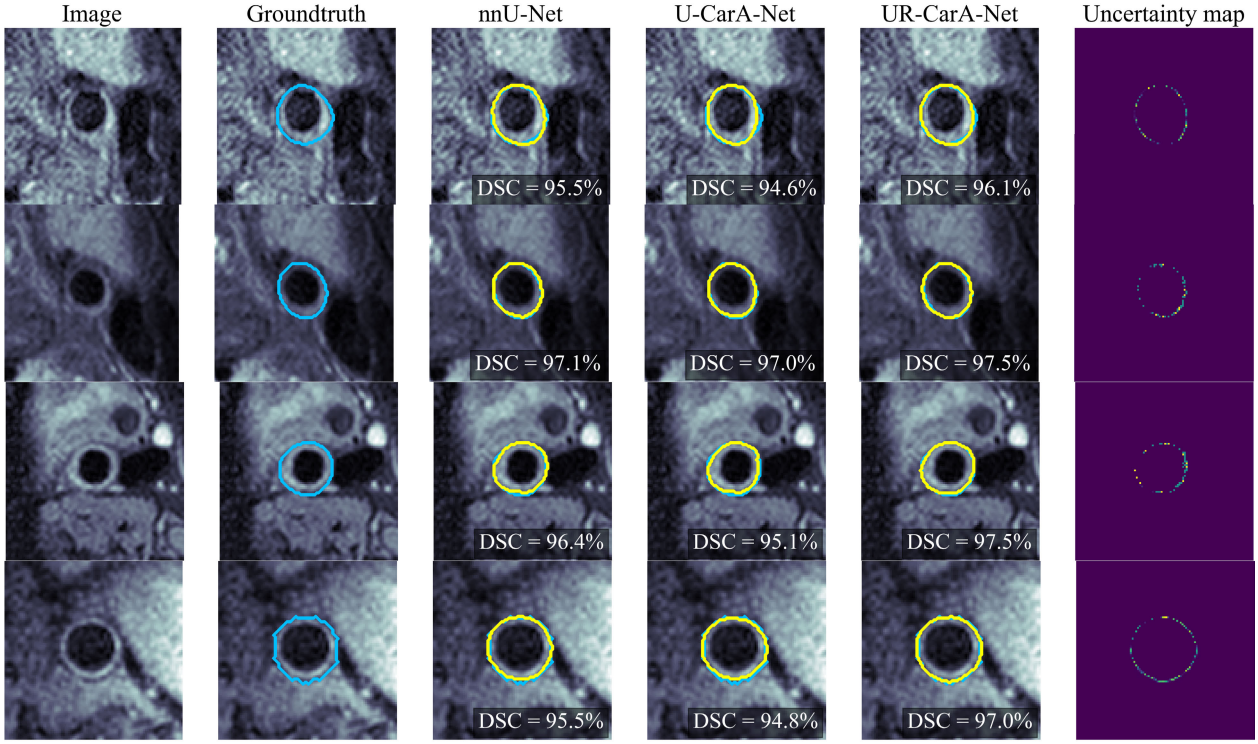


FIGURE 6. Segmentation of the 'simple' slices: blue contour — ground-truth, yellow contour — segmentation model, DSC — 2D Dice similarity coefficient for 2D contour, GT — ground-truth segmentation; first line — T1w TSE from the development domain, second line — post-contrast T1w TSE, third line — T2w, fourth line — external domain T1w FSE.

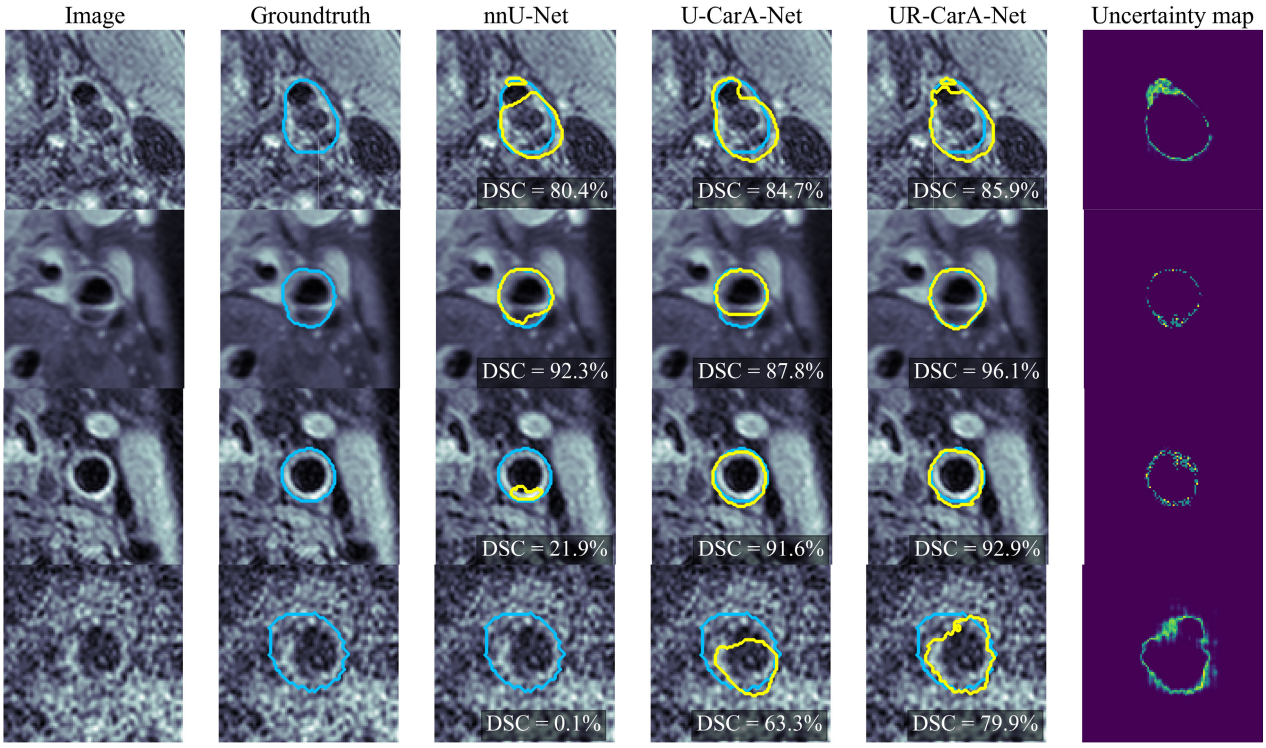


FIGURE 7. Segmentation of the 'challenging' slices: blue contour — ground-truth, yellow contour — segmentation model, DSC — 2D Dice similarity coefficient for 2D contour in percentage; first line — T1w TSE from the development domain, second line — post-contrast T1w TSE, third line — T2w, fourth line — external domain T1w FSE.

in the DD. In the D1, DSC is decreased by more than 20%, but in the D2 and ED, UR-CarA-Net application reduces the DSC variance for the scans with Michelson contrast values outside of the DD values.

Figure 6 shows segmentation results in the 'simple' cases. These slices are taken from the different patients and data domains. On these slices, the CA is clearly visible, it has a sufficient area, and the anatomy is simple, without any bifurcations and other CA branches present. The DSC values calculated for the same slice but with different models are similar.

Figure 7 shows segmentation results in 'challenging' cases. The slices are taken from the different patients with 2 representations from each data domain.

The segmentation challenges were caused by complicated anatomy (DD, D1), low image quality (ED), or different intensity distribution (D2). In these examples, the nnU-Net contours do not cover the whole CA area, especially in the bifurcation area or in the CA with stenosis. This results in the non-smooth contours not covering the whole CA area. In the DD case, the ICA was not a part of the contour. In the D2 case, the CA had only some pixels segmented; in the case of ED, it was not segmented. Implementation of the U-CarA-Net and UR-CarA-Net increased the segmented areas and DSCs. For case A, even though the highest DSC was yielded with UR-CarA-Net, the model still did not segment the whole CA area, and the contour was not smooth. However, the uncertainty map for this slice has high values in the pixels, classified as false negatives, showing that additional attention is needed. Finally, the UR-CarA-Net segmentation in the case ED overlaps the CA area. Still, it has a curved shape, which is reflected in the corresponding uncertainty map, having many bright pixels along the predicted segmentation edges. Nevertheless, DSC increased from 0.0% to 79.9%.

V. DISCUSSION AND CONCLUSION

In this work, we presented a method for CA segmentation on BB MRI. The robustness of Our proposed method is more robust for different MRI protocols and acquisition equipment. Even though we did not observe any significant improvement testing our model on the data from the development domain, volumetric ICC between ground-truth and automated segmentations improved by 0.19, 0.03, and 0.30 in the contrast-enhanced, T2w, and different center MRI, respectively. Moreover, in the external domain data, DSC, NSD, and relAPL improved significantly. Even though the proposed method brought segmentation performance scores in the external domain more in alignment with the ones in the development domain, in the contrast-enhanced and T2w MRI, the volumetric scores were lower. This can be explained by the fact that according to the protocol, the ground-truth contours are drawn on the T1w TSE data. Therefore, they were drawn on development domain data and then projected on contrast-enhanced and T2w MRI. Due to the minor patient movement, some slight misregistrations were possible. Additionally, in the T1w CE scans, contrast accumulates in the highly vascularised outer vessel wall (adventitial layer), which can cause a bias in the quantification of the wall volumes. High clDSC and NSD scores were observed, which did not change much between the models. Therefore, CA centerlines and main areas segmented by the nnU-Net did not change much, and contours were refined at the edges resulting in the difference between the segmented volumes, which proves our secondary hypothesis.

In comparison to similar published work, [27] used a 3D U-Net to segment CA outer wall and lumen on the PARISK data excluding EMC scans. The authors applied Monte-Carlo dropout as well to assess the segmentation uncertainty, yielding DSCs of 76.4% for the vessel wall and 88.5% for the lumen. The authors limited the scans by a $128 \times 128 \times 16$ bounding box placed in the center of the ground-truth segmentations. Alblas et al. [28] and Zhu et al. [33] applied 3D U-Net for CA localization and 2D CNN for CA segmentation method similar to ours that employs the sequential application of two deep learning models for 1) 3D CA centerline detection and 2) 2D CA wall contouring. They achieved a median DSC of 81.3% for the vessel wall on the open-source dataset acquired with the same MR protocol. Zhu et al. [33] achieved DSCs of 89.68% and 80.29% for the lumen and wall segmentation, respectively. They combined deep learning and graph-based approaches and applied them to the multi-sequence MRI acquired with the same protocols. DSCs reported in the current study, obtained on the PARISK data, exceed the values reported in the literature. But for a fair comparison, the same dataset should be used.

Although several studies to develop CA segmentation methods have been previously performed, we believe our work still stands out as a fully automated pipeline. Moreover, to our knowledge, it is the first study performing validation in multi-domain and multi-center data. We showed that the implementation of our method improves the segmentation performance on the external data domain to the level of the performance achieved on the development data domain. To improve our segmentation performance, we introduced three additional steps to the high-performing baseline model (nnU-Net): 1) carotid MRI lateral split to deal with one-sided annotations, 2) 2D contour refinement in a patch put into the localized CA area using the original image and nnU-Net softmax output, 3) introduction of an uncertainty component to the loss function using Monte-Carlo dropout. The reproducibility of this study is guaranteed by using carotid MRI data from a national multicenter PARISK study which is a highly recognized dataset. As all the patients had mild to moderate CA stenosis, the models were trained on the target patient cohort data, so they are robust to the pathological CA shape variations. We separately applied the method to the different BB MR sequences from the multi-contrast dataset instead of merging all these data into different channels which makes our method less demanding for the input data. Finally, UR-CarA-Net does not utilize the uncertainty information only in the training process but also enables uncertainty map generation by activating the dropout layers during the inference. Therefore, scans with high uncertainty scores can be reviewed separately.

There are some limitations of the presented approach. First of all, it is highly dependent on the baseline model (nnU-Net)

used for the initial CA segmentation. After nnU-Net, U/UR-CarA-Net scans the patches along the primarily segmented CA. Therefore, if nnU-Net does not segment the vessel or segments the wrong object in the scan, this error cannot be corrected by our method. Moreover, as presented in Figure 4, whereas CCA segmentation DSCs are relatively high, there might be errors in ICA segmentation. Nevertheless, Figure 4 also shows that the proposed method improves the ICA segmentation. However, these errors can be alleviated by increasing the dataset size. The second limitation is related to the blind lateral split of the MR scans into the right and left sides. If the patient is not positioned perfectly in the scanner, it results in a significant disturbance of the anatomy. Nevertheless, we rely on the carotid MRI data corresponding to the acquisition protocols where a patient is carefully centered in the scan. The third limitation is in the application of the 2D approach in contour refinement. We selected this approach for several reasons: 1) by using a 2D input, the training data size is increasing, 2) scanning 3D scans with a cubic 3D U-Net window will require more computational capacity, 3) reconstructed z-axis resolution is much lower than in-plane resolution (2 mm vs 0.30 mm $\times$ 0.30 mm). Finally, it is necessary to be careful in selecting the BB MR sequences since the enhancement of the outer layer of the vessel wall can cause a bias in volumetric measurements.

As the proposed method is the first step in CA plaque characterization, in future work, we aim to build a model for cardiovascular event prediction in patients with CA stenosis to be able to stratify patients based on their risk of stroke. Additionally, the results show that it is possible to use our method for BB MR sequence automated co-registration by segmenting CA on the scans of the same patient and minimizing the distance between the segmentations obtained on the different sequences. The achieved state-of-the-art segmentation metrics of our method together with its interpretability due to the uncertainty maps generation means the approach can be used as an initial step in CA plaque analysis. It can be followed by automated plaque components segmentation with one of the existing methods and handcrafted or deep radiomics applications for clinical outcomes prediction.

To conclude, in this first externally validated multi-center fully automated CA segmentation study, our model showed good segmentation performance (DSC of 91.7% (IQR 3.3%) on the development domain scans and 91.1% (IQR 7.2%) on the external domain scans), as well as an agreement with the manual segmentation (volume ICC of 0.91 on the development domain scans and 0.83 on the external domain scans). Its application is also feasible for the other BB MRI, obtained with other equipment or sequences. The suggested approach can be used for the other tasks on partially labeled data. The code is available on GitHub [36].

VI. AUTHOR CONTRIBUTION

Elizaveta Lavrova, Henry C. Woodruff, M. Eline Kooi, and Philippe Lambin conceived the idea of the paper. Mohamed Kassem, M. Eline Kooi, Anja G. Van Der Kolk, Philippe Lambin, Daniel Bos, and Jeroen Hendrikse participated in the data acquisition. Elizaveta Lavrova and Zohaib Salahuddin contributed to the development of the methodology. Elizaveta Lavrova implemented the analysis. All the authors contributed to the writing of the manuscript. Henry C. Woodruff, Philippe Lambin, and M. Eline Kooi supervised the work.

VII. CONFLICT OF INTEREST STATEMENT

Philippe Lambin reports, within and outside the submitted work, grants/sponsored research agreements from Radiomics SA, ptTheragnostic/DNAmito, Health Innovation Ventures. He received an advisor/presenter fee and/or reimbursement of travel costs/consultancy fee and/or in kind manpower contribution from Radiomics SA, BHV, Merck, Varian, Elekta, ptTheragnostic, BMS and Convert pharmaceuticals. Dr. Lambin has minority shares in the company Radiomics SA, Convert Pharmaceuticals, MedC2, and LivingMed Biotech, he is the co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Radiomics SA, one issued patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patented inventions (softwares) licensed to ptTheragnostic/DNAmito, Radiomics SA, and Health Innovation Ventures, and three non-issued, non-licensed patents on deep learning radiomics and LSRT (N2024482, N2024889, N2024889). He confirms that none of the above entities or funding were involved in the preparation of this paper. Henry Woodruff has minority shares in the company Radiomics.

REFERENCES

[1] E. J. Benjamin et al., "Heart disease and stroke statistics—2019 update: A report from the American Heart Association," *Circulation*, vol. 139, no. 10, pp. e56–e528, Mar. 2019.

[2] M. T. B. Truijman, M. E. Kooi, A. C. van Dijk, A. A. J. de Rotte, A. G. van der Kolk, M. I. Liem, F. H. B. M. Schreuder, E. Boersma, W. H. Mess, R. J. van Oostenbrugge, P. J. Koudstaal, L. J. Kappelle, P. J. Nederkoorn, A. J. Nederveen, J. Hendrikse, A. F. W. van der Steen, M. J. A. P. Daemen, and A. van der Lugt, "Plaque at RISK (PARISK): Prospective multicenter study to improve diagnosis of high-risk carotid plaques," *Int. J. Stroke*, vol. 9, no. 6, pp. 747–754, Aug. 2014.

[3] K. P. H. Nies, L. J. M. Smits, M. Kassem, P. J. Nederkoorn, R. J. V. Oostenbrugge, and M. E. Kooi, "Emerging role of carotid MRI for personalized ischemic stroke risk prediction in patients with carotid artery stenosis," *Frontiers Neurol.*, vol. 12, Aug. 2021, Art. no. 718438.

[4] A. C. van Dijk, M. T. B. Truijman, B. Hussain, T. Zadi, G. Saiedie, A. A. J. de Rotte, M. I. Liem, A. F. W. van der Steen, M. J. A. P. Daemen, P. J. Koudstaal, P. J. Nederkoorn, J. Hendrikse, M. E. Kooi, and A. van der Lugt, "Intraplaque hemorrhage and the plaque surface in carotid atherosclerosis: The plaque at RISK study (PARISK)," *Amer. J. Neuroradiol.*, vol. 36, no. 11, pp. 2127–2133, Nov. 2015.

[5] L. Saba, C. Yuan, T. S. Hatsukami, N. Balu, Y. Qiao, J. K. DeMarco, T. Saam, A. R. Moody, D. Li, C. C. Matouk, M. H. Johnson, H. R. Jäger, M. Mossa-Basha, M. E. Kooi, Z. Fan, D. Saloner, M. Wintermark, D. J. Mikulis, and B. A. Wasserman, "Carotid artery wall imaging: Perspective and guidelines from the ASNR vessel wall imaging study group and expert consensus recommendations of the American society of neuroradiology," *Amer. J. Neuroradiol.*, vol. 39, no. 2, pp. E9–E31, Feb. 2018.

[6] D. H. van Dam-Nolen et al., "Carotid plaque characteristics predict recurrent ischemic stroke and TIA: The PARISK (plaque at risk) study," *Cardiovascular Imag.*, vol. 15, no. 10, pp. 1715–1726, 2022.

[7] A. Saxena, E. Y. K. Ng, and S. T. Lim, "Imaging modalities to diagnose carotid artery stenosis: Progress and prospect," *BioMed. Eng. OnLine*, vol. 18, May 2019, Art. no. 66.

[8] A. van Engelen, A. C. van Dijk, M. T. B. Truijman, R. Van't Klooster, A. van Opbroek, A. van der Lugt, W. J. Niessen, M. E. Kooi, and M. de Bruijne, "Multi-center MRI carotid plaque component segmentation using feature normalization and transfer learning," *IEEE Trans. Med. Imag.*, vol. 34, no. 6, pp. 1294–1305, Jun. 2015.

[9] I. M. Adame, R. J. van der Geest, B. A. Wasserman, M. A. Mohamed, J. H. C. Reiber, and B. P. F. Lelieveldt, "Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 16, no. 5, pp. 227–234, Apr. 2004.

[10] D. S. Jodas, A. S. Pereira, and J. M. R. S. Tavares, "Lumen segmentation in magnetic resonance images of the carotid artery," *Comput. Biol. Med.*, vol. 79, pp. 233–242, Dec. 2016.

[11] A. M. Arias-Lorza, J. Petersen, A. van Engelen, M. Selwaness, A. van der Lugt, W. J. Niessen, and M. de Bruijne, "Carotid artery wall segmentation in multispectral MRI by coupled optimal surface graph cuts," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 901–911, Mar. 2016.

[12] S. Gao, R. van't Klooster, P. H. Kitslaar, B. F. Coolen, A. M. van den Berg, L. P. Smits, R. Shahzad, D. P. Shamonin, P. J. H. de Koning, A. J. Nederveen, and R. J. van der Geest, "Learning-based automated segmentation of the carotid artery vessel wall in dual-sequence MRI using subdivision surface fitting," *Med. Phys.*, vol. 44, no. 10, pp. 5244–5259, Oct. 2017.

[13] A. M. Arias-Lorza, D. Bos, A. van der Lugt, and M. de Bruijne, "Cooperative carotid artery centerline extraction in MRI," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0197180.

[14] R. Zhang, Q. Zhang, A. Ji, P. Lv, J. Zhang, C. Fu, and J. Lin, "Identification of high-risk carotid plaque with MRI-based radiomics and machine learning," *Eur. Radiol.*, vol. 31, no. 5, pp. 3116–3126, May 2021.

[15] J. M. A. Hofman, W. J. Branderhorst, H. M. M. T. Eikelder, V. C. Cappendijk, S. Heeneman, M. E. Kooi, P. A. J. Hilbers, and B. M. T. H. Romeny, "Quantification of atherosclerotic plaque components using in vivo MRI and supervised classifiers," *Magn. Reson. Med.*, vol. 55, no. 4, pp. 790–799, Apr. 2006.

[16] R. van't Klooster, O. Naggara, R. Marsico, J. H. C. Reiber, J.-F. Meder, R. J. van der Geest, E. Touzé, and C. Oppenheim, "Automated versus manual in vivo segmentation of carotid plaque MRI," *Amer. J. Neuroradiol.*, vol. 33, no. 8, pp. 1621–1627, Sep. 2012.

[17] F. Liu, D. Xu, M. S. Ferguson, B. Chu, T. Saam, N. Takaya, T. S. Hatsukami, C. Yuan, and W. S. Kerwin, "Automated in vivo segmentation of carotid plaque MRI with morphology-enhanced probability maps," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 55, no. 3, pp. 659–668, 2006.

[18] W. Kerwin, D. Xu, F. Liu, T. Saam, H. Underhill, N. Takaya, B. Chu, T. Hatsukami, and C. Yuan, "Magnetic resonance imaging of carotid atherosclerosis," *Topics Magn. Reson. Imag.*, vol. 18, no. 5, pp. 371–378, 2007.

[19] H. Tang, T. van Walsum, R. S. van Onkelen, R. Hameeteman, S. Klein, M. Schaap, F. L. Tori, Q. J. A. van den Bouwhuijsen, J. C. M. Witteman, A. van der Lugt, L. J. van Vliet, and W. J. Niessen, "Semiautomatic carotid lumen segmentation for quantification of lumen geometry in multispectral MRI," *Med. Image Anal.*, vol. 16, no. 6, pp. 1202–1215, Aug. 2012.

[20] A. van Engelen, W. J. Niessen, S. Klein, H. C. Groen, H. J. M. Verhagen, J. J. Wentzel, A. van der Lugt, and M. de Bruijne, "Atherosclerotic plaque component segmentation in combined carotid MRI and CTA data incorporating class label uncertainty," *PLoS ONE*, vol. 9, no. 4, Apr. 2014, Art. no. e94840.

[21] H. Tang, M. Selwaness, R. Hameeteman, A. van Dijk, A. van der Lugt, J. C. Witteman, W. J. Niessen, L. J. van Vliet, and T. van Walsum, "Semi-automatic MRI segmentation and volume quantification of intraplaque hemorrhage," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 1, pp. 67–74, Jan. 2015.

[22] Y. Dong, Y. Pan, X. Zhao, R. Li, C. Yuan, and W. Xu, "Identifying carotid plaque composition in MRI with convolutional neural networks," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, May 2017, pp. 1–8.

[23] J. Wu, J. Xin, X. Yang, J. Sun, D. Xu, N. Zheng, and C. Yuan, "Deep morphology aided diagnosis network for segmentation of carotid artery vessel wall and diagnosis of carotid atherosclerosis on black-blood vessel wall MRI," *Med. Phys.*, vol. 46, no. 12, pp. 5544–5561, 2019.

[24] L. Chen, J. Sun, G. Canton, N. Balu, D. S. Hippe, X. Zhao, R. Li, T. S. Hatsukami, J.-N. Hwang, and C. Yuan, "Automated artery localization and vessel wall segmentation using tracklet refinement and polar conversion," *IEEE Access*, vol. 8, pp. 217603–217614, 2020.

[25] L. Chen, H. Zhao, H. Jiang, N. Balu, D. B. Geleri, B. Chu, H. Watase, X. Zhao, R. Li, J. Xu, T. S. Hatsukami, D. Xu, J.-N. Hwang, and C. Yuan, "Domain adaptive and fully automated carotid artery atherosclerotic lesion detection using an artificial intelligence approach (LATTE) on 3D MRI," *Magn. Reson. Med.*, vol. 86, no. 3, pp. 1662–1673, Sep. 2021.

[26] D. D. Samber, S. Ramachandran, A. Sahota, S. Naidu, A. Pruzan, Z. A. Fayad, and V. Mani, "Segmentation of carotid arterial walls using neural networks," *World J. Radiol.*, vol. 12, no. 1, pp. 1–9, Jan. 2020.

[27] R. Camarasa, D. Bos, J. Hendrikse, P. Nederkoorn, E. Kooi, A. van der Lugt, and M. de Bruijne, "Quantitative comparison of Monte–Carlo dropout uncertainty measures for multi-class segmentation," in *Proc. Int. Workshop Uncertainty Safe Utilization Mach. Learn. Med. Imag.*, 2020, pp. 32–41.

[28] D. Alblas, C. Brune, and J. M. Wolterink, "Deep-learning-based carotid artery vessel wall segmentation in black-blood MRI using anatomical priors," *Proc. SPIE*, vol. 12032, Apr. 2021, Art. no. 120320Y.

[29] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: Methodological failures and recommendations for the future," *NPJ Digit. Med.*, vol. 5, no. 1, p. 48, Apr. 2022.

[30] E. Kondrateva, M. Pominova, E. Popova, M. Sharaev, A. Bernstein, and E. Burnaev, "Domain shift in computer vision models for MRI data analysis: An overview," *Proc. SPIE*, vol. 11605, Jan. 2020, Art. no. 116050H.

[31] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage, Clin.*, vol. 6, pp. 9–19, Jan. 2014.

[32] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021.

[33] C. Zhu, X. Wang, S. Chen, Z. Teng, C. Bai, X. Huang, M. Xia, Z. Shao, Z. Gu, and P. Sun, "Complex carotid artery segmentation in multi-contrast MR sequences by improved optimal surface graph cuts based on flow line learning," *Med. Biol. Eng. Comput.*, vol. 60, no. 9, pp. 2693–2706, Sep. 2022.

[34] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[35] S. Hu, Z. Liao, and Y. Xia, "Label propagation for 3D carotid vessel wall segmentation and atherosclerosis diagnosis," Aug. 2022, *arXiv:2208.13337*.

[36] L. Lavrova, "Lavrovaliz/ur-cara-net: 0.1," Zenodo, Version 0.1, Mar. 2023, doi: 10.5281/zenodo.7741155.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

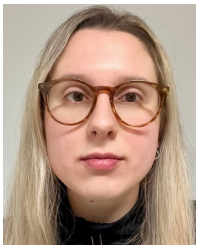
[38] T. W. Arega, S. Bricq, F. Meriaudeau, "Leveraging uncertainty estimates to improve segmentation performance in cardiac MR," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis*, C. H. Sudre, R. Licandro, C. Baumgartner, A. Melbourne, A. Dalca, J. Hutter, R. Tanno, E. A. Turk, K. Van Leemput, B. T. Jordina, W. M. Wells, and C. Macgowan, Eds. Cham, Switzerland: Springer, 2021, pp. 24–33.

[39] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. W. Pluim, U. Bauer, and B. H. Menze, "clDice—A novel topology-preserving loss function for tubular structure segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16560–16569.

[40] A. Reinke et al., "Common limitations of image processing metrics: A picture story," Apr. 2021, *arXiv:2104.05642*.

[41] F. Vaassen, C. Hazelaar, A. Vaniqui, M. Gooding, B. van der Heyden, R. Canters, and W. van Elmpt, "Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy," *Phys. Imag. Radiat. Oncol.*, vol. 13, pp. 1–6, Jan. 2020.

[42] P. Fält, J. Hyttinen, L. Fauch, A. Riepponen, A. Kullaa, and M. Hauta-Kasari, "Spectral image enhancement for the visualization of dental lesions," in *Proc. 8th Int. Conf. Image Signal Process.*, 2018, pp. 490–498.



ELIZAVETA LAVROVA was born in Moscow, Russia, in 1992. She received the B.S. and M.S. degrees in biomedical engineering from Bauman Moscow State Technical University, Moscow, in 2014 and 2016, respectively.

From 2013 to 2014, she was with the Computed Tomography Department, GE Healthcare, Russia. In 2016, she was with the Technical Evangelism Department, Microsoft, Russia. From 2016 to 2018, she was a Research Assistant with Bauman Moscow State Technical University. Since 2019, she has been a Ph.D. Researcher with Liege University, Liege, Belgium; and Maastricht University, Maastricht, The Netherlands. Her research interest includes the application of machine learning and deep learning methods for medical imaging tasks.

ZOHAIB SALAHUDDIN was born in Lahore, Pakistan, in 1995. He received the bachelor's degree in electrical engineering from the University of Engineering and Technology, Lahore, in 2017, and the joint master's degree in medical imaging and applications from the University of Burgundy, University of Cassino, and University of Girona, in 2020. He is currently pursuing the Ph.D. with the Department of Precision Medicine, Maastricht University. He received an Erasmus Plus Scholarship to pursue his joint master's degree. He completed his master's thesis at Philips, Hamburg. He had presented his work on Coronary Artery Centerline Extraction at International Symposium on Biomedical Imaging in 2021.

He was with Siemens, Lahore, from 2017 to 2018. His research work focuses on the interpretability of artificial intelligence networks for medical image analysis, uncertainty estimation, and standardization of reporting guidelines for AI tools for medical image analysis.

HENRY C. WOODRUFF received the M.S. degree from Max Planck Institute for Radio Astronomy, Bonn, Germany, in 2003, and the Ph.D. degree in astrophysics from Sydney University, Sydney, Australia, in 2009.

From 2011 to 2014, he was a Postdoctoral Researcher with the Hunter Cancer Research Alliance, Newcastle, Australia; and the Central Medical School, University of Sydney, Australia, from 2014 to 2017. From 2017 to 2020, he was a Senior Postdoctoral Research Fellow with Maastricht University, The Netherlands. Since 2020, he has been an Assistant Professor with the Department of Precision Medicine and the Head of the D-Lab. His research interest includes the application of advanced machine learning and image processing methods on vast amounts of medical data to improve the lives of patients, with a focus on decision support systems and precision medicine.

MOHAMED KASSEM received the M.D. degree from Aleppo University, Syria, in 2003, and the M.S. degree in biomedical sciences from Maastricht University, in 2019.

From 2012 to 2014, he was a Medical Coordinator of the Qatar Red Crescent Society. Since 2019, he has been a Ph.D. Researcher with Maastricht University.

ROBIN CAMARASA was born in Toulouse, France, in 1996. He received the Master of Science degree in computer and data sciences from Saint-Etienne University, France, in 2020.

Since 2020, he has been a Ph.D. Researcher with the Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands. His current research interest includes the interpretability and uncertainty of machine learning models in medical image analysis.

ANJA G. VAN DER KOLK was born in Wageningen, The Netherlands, in 1984. She received the med school degree from Utrecht University, the Netherlands, in 2008, the Ph.D. degree in ultra-high-field MRI (7T) from Utrecht University/UMC Utrecht, the Netherlands. She completed the radiology residency from Gelre Ziekenhuizen Apeldoorn, the Netherlands, in 2020.

From 2020 to 2021, she was a Neuroradiologist with the Antoni van Leeuwenhoek Hospital/Netherlands Cancer Center, Amsterdam. Since 2021, she has been a Neuroradiologist and a Clinician-Scientist with the Rabdoudumc in Nijmegen, the Netherlands, combining her clinical duties 50/50 with (clinically driven) research. She has more than 50 publications in international peer-reviewed journals and has coauthored with researchers from several groups in the USA. Her research interests include ultra-high-field (7T and higher) MRI, metabolic imaging of brain tumors, (intracranial) vessel wall imaging, and neuroinflammation.

Dr. Kolk is an elected member of the ISMRM Annual Meeting Program Committee, an elected Vice Chair of the ISMRM Education Committee (which she will continue for another three years as the Chair), and an elected member of the editorial board of *European Radiology*. She was a recipient of the (Dutch) Frederik Philips Award for the Best Thesis on Clinical Radiologic and Interventional Techniques in 2014 and the Female Scientist Innovations in Imaging Award in 2018.

PAUL J. NEDERKOORN has been a Neurologist with Academic Medical Center, Amsterdam, The Netherlands, since 2009. His research interest includes carotid endarterectomy or optimized medical therapy in patients with symptomatic carotid artery stenosis: a cost-effectiveness analysis.

DANIEL BOS was born in Rotterdam, the Netherlands, in 1985. He received the med school degree from Erasmus University Rotterdam, in 2009, and the Ph.D. degree in epidemiology and radiology from Erasmus University Rotterdam, in 2013. He received his formal registration as a Clinical Epidemiologist from the Dutch Society for Epidemiology and the National Committee for Medical-Biological Research Netherlands (SMBWO), in 2019.

He currently holds a position as an Associate Professor of epidemiology and radiology and nuclear medicine with Erasmus MC, where he is the Vice Chair of the Department of Epidemiology. Additionally, he holds a position as a Professor with the Department of Cardiovascular Sciences, KU Leuven (Belgium), and an Adjunct Associate Professor with the Department of Epidemiology, Harvard School of Public Health (USA). He has more than 100 publications in international peer-reviewed journals. His research is focused on the interface of neurovascular and neurodegenerative disease, with a strong emphasis on the use of imaging.

Prof. Bos is a member of the Dutch Society for Epidemiology, the European Stroke Organisation, and the European Atherosclerosis Society. He is an Associate Editor of *The British Journal of Radiology* and the *European Stroke Journal*. He was a recipient of the (Dutch) Lourens Penning Prize for Best Paper in Neuroradiology and the Van Leersum Prize of the Royal Netherlands Academy of Arts and Sciences.

JEROEN HENDRIKSE received the Ph.D. degree from Utrecht University.

From 2004 to 2015, he was a Radiologist with the Radiology Department, University Medical Center Utrecht, where he was the Head of the Radiology Department, from 2017 to 2022. Since 2017, he has been a Professor of neuroradiology with the University Medical Center Utrecht. Since 2020, he has been the Chairperson of the Image and Oncology Division, University Medical Center Utrecht.

M. ELINE KOOI was born in Purmerend, the Netherlands, in 1971. She received the master's degree in physics and the Ph.D. degree in molecular physics from the University of Amsterdam, in 1995 and 1999, respectively. She was trained in medical physics, specializing in radiology from Maastricht UMC.

Since 2004, she has been a registered Medical Physicist with MUMC. She is a Professor of medical physics, especially vascular imaging with the Department of Radiology and Nuclear Medicine, CARIM, School for Cardiovascular Diseases, Maastricht University Medical Centre. Her research is focused on sophisticated multi-modality cardiovascular imaging. Her group is at the international forefront of developing and evaluating imaging techniques to study plaque vulnerability. She has focused on cutting-edge imaging of the hallmarks of high-risk plaques. She was the first to demonstrate that plaque inflammation can be visualized in humans. Her group has the unique position of multi-modality imaging of the vessel wall (PET/MRI, ultrasound, and CT) in different vascular beds (carotid and coronary arteries and aorta). She has performed numerous MR Spectroscopy studies on lipid metabolism, also in the heart. Recently, she has expanded her research toward cardiac MRI (perfusion MRI, T1 mapping, and strain). She is a Principal Investigator of a multi-center trial to determine the clinical value of plaque imaging (PARISK). She coordinated plaque imaging within the European Carotid Surgery Trial-2 (ECST-2). She has published more than 140 scientific publications in international peer-reviewed journals.

Prof. Kooi is a member of the International Society of Magnetic Resonance Imaging and the Netherlands Society of Medical Physics and of the NWO Vici committee.

PHILIPPE LAMBIN was born in Belgium, in 1961. He received the Medical degree from the Catholic University of Louvain, Belgium, in 1986, the degree in clinical oncology from Institut Gustave Roussy, Villejuif, France, in 1990, the degree in radiation oncologist from the Catholic University of Louvain, in 1991, and the Ph.D. degree from Université Paris-Sud, France, in 1997.

Since 2000, he has been a Professor of radiation oncology with the University of Maastricht, The Netherlands. From 2000 to 2015, he was a Medical Director of the MAASTRO Clinic. Since 2017, he has been the Head of the Department of Precision Medicine, University of Maastricht. From 2000 to 2006, he was a Visiting Professor with the Department of Radiation Oncology, University Hospital Gasthuisberg, Leuven, Belgium. In 2007, he was a Visiting Professor with the University of Philadelphia. From 2010 to 2017, he was a Visiting Professor with the University of Toronto. In 2016, he was a Visiting Professor with the University of California, San-Francisco (UCSF) School of Medicine. In 2018, he was a Visiting Professor with the Memorial Sloan Kettering Cancer Center, New York. In 2019, he was a Visiting Professor with UCSF, San Francisco. His main research interests include directed toward translational research in tumour biology with a specific focus on tumour hypoxia, functional imaging, and lung, head, and neck cancer.

Prof. Lambin has a number of awards, including the Breur Award from the European Society for Therapeutic Radiology and Oncology in 2009, the Prof. Raymond Bush Award in 2010, the European Society for Therapeutic Radiology and Oncology Lecturer Award in 2012, and the Nederland's Respiratory Society Science Award 2016.