Segmentation of Deep Inferior Epigastric Artery Perforators for Breast Reconstruction using Deep Learning in Computed Tomography Angiography: A Quantitative and Qualitative Evaluation

Master Thesis Technical Medicine

Iris Moes

June 2023

Universiteit Leiden

TU Delft — Delft University of Technology

Erasmus Universiteit Rotterdam

Erasmus MC — Universitair Medisch Centrum Rotterdam

This page is initially left blank

Segmentation of Deep Inferior Epigastric Artery Perforators for Breast Reconstruction using Deep Learning in Computed Tomography Angiography:
A Quantitative and Qualitative Evaluation

I. (Iris) Moes

Student number : 4470990

June 12th, 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)

Dept. of Plastic Reconstructive Surgery and Dept. of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
September 2022 – June 2023

Supervisors:

Dr. E.M.L. (Eveline) Corten
Dr. Ir. T. (Theo) van Walsum
R. (Ruisheng) Su
A.(Abdullah) Thabit

Thesis committee members:

Dr. ir. T. (Theo) van Walsum, Erasmus MC (chair)
Dr. E.M.L. (Eveline) Corten, Erasmus MC
Dr. W.A. (Wyanne) Noortman, Leiden University Medical Center
R. (Ruisheng) Su, Erasmus MC

An electronic version of this thesis is available at http://repository.tudelft.nl/

Universiteit Leiden

TUDelft — Delft University of Technology

ERASMUS UNIVERSITEIT ROTTERDAM

This page is initially left blank

Preface and acknowledgments

It was the blue cart that was connected to my brother which was my first introduction to medical technology. Not until later did I find out it was a heart pump, which was keeping him alive. As a child, I always had an interest in becoming a doctor. However, this introduction to medical technology must have planted a seed that started to grow as I got older because I realized that a career as a doctor would not be a good fit for me. Instead, I developed an ambition for mathematics, physics, and mostly formulas, equations, and numbers. Yet, the human body and the medical field still held my interest. During my orientation for studies, my mom showed me information about the relatively new study Clinical Technology. Without any doubt, I was excited and knew that this program would meet both my technical and medical desires.

The choice of our internships in the Master program of Technical Medicine provided me with the opportunity to develop myself as Clinical Technologist. During my final internship at the Department of Cardiothoracic Surgery, I was introduced to the world of 3D models, medical image segmentation, analyzing numbers, and writing an article. Aspects that I enjoyed and wanted to learn more about.

Tessa, I appreciate you introducing me to Eveline, whose project aligns perfectly with my interests, as well as providing me with the opportunity to learn more about Plastic Reconstructive surgery, a field that has always intrigued me. Eveline, I am grateful for the chance to embark on my own clinical journey during my thesis. Theo, thank you for your technical guidance and support. Your expertise and feedback have helped me develop my skills. Abdullah and Ruisheng, thank you for being part of our team, providing your time and effort to answer all my questions, and inspiring me with your work. Finally, Wyanne, it is an honor to have you in my thesis committee. We met during the master intervisions, where your personality and experience as a Clinical Technologist were thoroughly valuable.

I am deeply grateful to my family and closest friends for their unwavering support during this exceptionally difficult year. Their belief in my capabilities and unwavering faith in my personality has been invaluable. I would like to extend a special appreciation to my parents and younger brother. My dad, who was always there to listen and to support me. And my mom, who was, and I know still is a constant source of love and encouragement. I hope I can make them proud of the person I will become.

Now, after several years of studying, I am nearing the end of my journey as a student of Technical Medicine. While I remain unsure if this field is my ideal future after all, and if my true passion lies more in the technical or clinical aspects of medicine, I am confident that with everything I have experienced and learned, this degree and background will unlock numerous opportunities for me.

<div align="right">

Iris Moes
June, 2023

</div>

Table of contents

Abstract

Introduction: Breast reconstruction after a mastectomy is crucial to improve a patient's quality of life. The Deep Inferior Epigastric artery Perforator (DIEP) flap procedure is considered the golden standard for breast reconstruction. Three-dimensional (3D) visualization methods have shown promise in providing a better understanding of Computed Tomography Angiography (CTA) information and potentially reducing operative time. However, for these methods manual segmentation of the perforators is needed, which is time-consuming and prone to variability. Automated segmentation using Deep learning (DL) can potentially overcome these limitations and provide accurate and efficient segmentation of the DIEP flap perforators. This study aimed to evaluate the application of DL for automated perforator segmentation in DIEP flap breast reconstruction, improving the efficiency and objectivity of DIEP flap perforator segmentation.

Materials and methods: The dataset comprised 25 CTA scans for training and validation, and 5 CTA scans for testing. DL was employed for automated segmentation, and quantitative evaluation included metrics such as Dice coefficient score, recall, precision, surface distance, and centerline overlap. The qualitative evaluation involved grading the clinical acceptability of segmentations by four experienced plastic reconstructive surgeons.

Results: On the training set, a Dice score of 0.58 ($\pm$0.08) and a true positive centerline overlap of 0.66 ($\pm$0.10) were achieved for perforator segmentation. The DL model successfully segmented the intramuscular main branch, but some perforators were missed in the subcutaneous fat tissue. Combined grading by all surgeons showed no statistical difference between manual and automated segmentations and both segmentations were evaluated as clinically acceptable.

Conclusion: Automated DL segmentation holds promise for enhancing the efficiency and objectivity of identifying DIEP flap perforators in CTA images, providing an alternative to manual segmentation. Nonetheless, further research is needed to refine the automated segmentation results and to validate the generalizability and clinical applicability of the DL segmentation approach in larger patient cohorts and different clinical settings.

List of abbreviations

2D	Two-dimensional
3D	Three-dimensional
ASSD	Average Symmetric Surface Distance
CTA	Computed Tomography Angiogram
DIEP	Deep Inferior Epigastric artery Perforator
DL	Deep Learning
DSC	Dice similarity coefficient
DTUNet	Dual branch Topology-aware U-Net
FN	False Negatives
FP	False Positives
FPCR	False Positive Centerline Rate
nnU-Net	no new U-Net
RA	Rectus Abdominis
SIEA	Superficial Inferior Epigastric Artery
TN	True Negatives
TP	True Positives
TPCR	True Positive Centerline Rate

1 Introduction

Breast cancer is the most prevalent cancer worldwide and the second leading cause of cancer-related death in women ([1](#)). Although breast-conservative treatment has increased due to early detection and improved treatments, a mastectomy, involving removal of all breast tissue, may still be required or preferred ([2](#)). Additionally, the number of prophylactic mastectomies in women with a genetic higher risk for breast cancer has increased over time ([3](#)). A mastectomy can be traumatic and is associated with unaesthetic results, and high-quality breast reconstruction is crucial for the quality of life of patients ([4](#), [5](#)). In addition to improving body image, successful breast reconstruction can also result in better psychological patient well-being.

Following mastectomy, the Deep Inferior Epigastric artery Perforator (DIEP) flap procedure is considered the gold standard for breast reconstruction ([6](#), [7](#)). In DIEP flap breast reconstruction, skin, and subcutaneous tissue, together with one or more perforating vessels from the epigastric vessels, are taken from the abdomen. Preserving as much muscular tissue as possible, the flap perforators and corresponding veins are dissected through the rectus abdominis (RA) muscle ([8](#)). Subsequently, the subcutaneous tissue, skin, veins, and perforators are transplanted to the recipient's breast to recreate a naturally appearing breast (see Figure 1). The wide variety in the vascular anatomy of the lower anterior abdominal wall between patients ([6](#), [9](#), [10](#)) complicates the procedure and can increase the operative time. To optimize the time to harvest the DIEP flap, preoperative imaging can be used to visualize and map the perforators ([11](#), [12](#)). It has demonstrated to reduce operative time, surgeon stress, and postoperative complications in DIEP flap breast reconstruction ([13-15](#)).
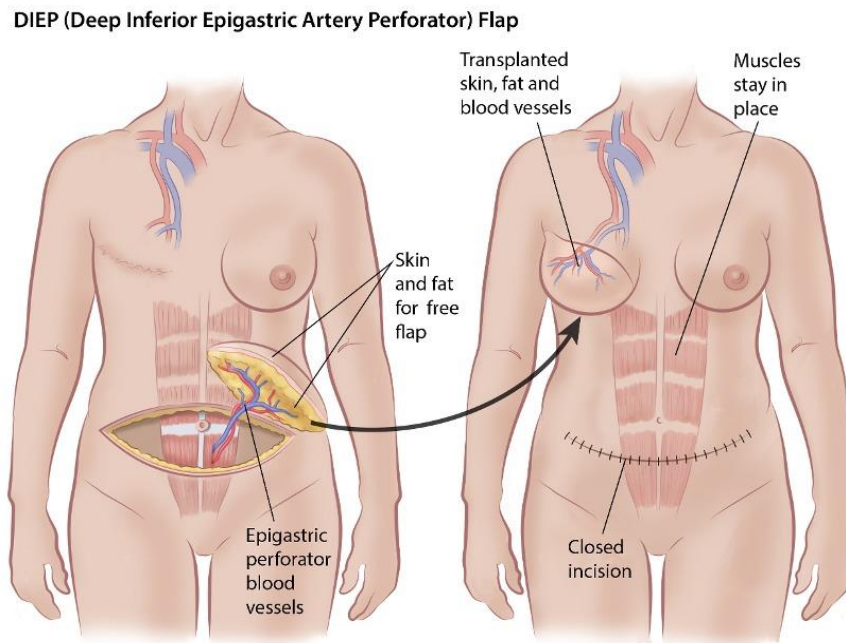

Figure 1. Deep Inferior Epigastric artery Perforator (DIEP) flap procedure. In this surgery, the skin, and subcutaneous fat tissue, together with one or more perforating vessels from the epigastric vessels, are taken from the abdomen and transplanted to the recipient's breast. The muscle is preserved during surgery. Image from ([16](#)).

Preoperative visualization of the abdominal vascular anatomy with Computed Tomography Angiography (CTA) is the gold standard for the identification of perforators for DIEP flap breast reconstruction ([17-19](#)). The current visualization, however, is limited by display in two-dimension (2D) only. Three-dimensional (3D) visualization has been suggested and investigated for DIEP flap breast reconstruction. Compared to identification solely on 2D CTA scans, 3D visualization provides a better understanding of the CTA information and reduces the operative time spent on the dissection of the perforators ([20](#)).

For DIEP flap breast reconstruction, different 3D visualization methods have been proposed to assist in the identification of the perforators and intraoperative mapping. These methods include the projection of the perforators on the abdomen, virtual reality, augmented reality, and y, and creating a 3D-printed model of the abdominal wall's vascular anatomy ([20-25](#)). However, the previous studies exploring these 3D visualization methods relied on manual segmentation of the flap perforators, which is known to be time-consuming and subjective.

Fully automated methods offer the potential for faster and more objective perforator segmentation in DIEP flap breast reconstruction. However, to date, there is no automated method for the segmentation of the DIEP flap perforator vessels available. Semi-automated methods using supervised identification with computer-aided detection to support the perforator identification have been described previously ([26-28](#)). In these studies, a vessel centerline tracking method was used, resulting in the reduction of the time required for perforator annotation. However, user input was still needed by manually marking perforators, and the reported outcomes did not include the most commonly used performance evaluation metrics for 3D medical imaging segmentation ([29](#)). These metrics are relevant for the assessment of segmentation results, reproducibility, and comparability of the outcomes ([30](#)). Deep Learning (DL) could provide a fully automated segmentation method and is already widely used in research for several complex medical imaging tasks ([31](#)), including (micro)vessel segmentation ([32](#), [33](#)). Nonetheless, the small caliber of the DIEP flap perforator vessels ($\sim 1 - 3.5$ mm) ([34](#)) and the low contrast to the surrounding tissue provide a challenge for the direct application of existing DL vessel segmentation methods.

Therefore, this study aimed to evaluate the application of DL for automated perforator segmentation in DIEP flap breast reconstruction, to improve the efficiency and objectivity of DIEP flap perforator segmentation.

2 Materials and methods

2.1 Dataset

In this study, we included 30 anonymized abdominal CTA scans acquired for preoperative planning of unilateral or bilateral DIEP flap breast reconstruction in the Erasmus MC between 2020 and 2022. The selection of scans within this timeframe was based on the availability of adequate contrast. We included the first 30 scans that met this criterion, excluding two scans with insufficient contrast uptake. The average age of 26 of the 30 patients was 48 (± 8.31) years. For 4 patients, the age was unknown.

Out of the 30 scans, two scans were acquired with a Siemens Naeotom Alpha, with a slice thickness of 0.6 mm, and the rest of the scans were obtained on a Siemens Somatom Drive, Edge Plus, Somatom Force or Somatom Definition Edge, reconstructed with a slice thickness of 0.75 mm. The peak kilo voltage of the generator ranged from 70 to 120 kVp, the average X-ray tube current was 378 (± 172.60) mA, and the spiral pitch factor ranged from 0.60 to 0.95. The collimation width used was 0.40 mm for the photon counting CTA scans and 0.60 mm for the other scans. For all patients, Iomeron was used as contrast agent.

2.2 Manual cropping

Prior to manual annotation and model training, the scans were manually cropped using 3D slicer[1] (version 5.0.3) ([35]). The CTA scans were loaded as DICOM into 3D Slicer. The images were cropped to obtain the same input sizes and to only include the region of interest for the DIEP segmentation. The cropping area used had a region of interest with dimensions of 512 voxels (x; width) x 220 voxels (y; depth) x 700 voxels (z; height). The placement of the region of interest was determined based on the location of the most medial aspect of the umbilicus. The center of the upper axis of the cropping region was aligned 3 mm superior to this determined location, as generally, perforators above this area are not used as flap perforators. This alignment was discussed with two plastic reconstruction surgeons with over 15 years of experience in DIEP flap breast reconstruction. The cropped scans were stored in NIfTI file format. More details on the manual cropping can be found in Appendix A, section 3.

2.3 Manual data annotation

The segmentation process involved annotating the following structures on the cropped scans: the rectus abdominis (RA) muscle, subcutaneous fat tissue, skin tissue, and the DIEP vessels. This annotation was performed semi-automatically using 3D slicer. Segmentation was supported with thresholding and region-growing-based algorithms. For the DIEP vessels, the draw tube function was used, which creates a tube along manually placed markers (see Figure 2). A comprehensive protocol, including all steps and software requirements for the manual annotation process, can be found in Appendix A, section 4.

All annotations were performed by one Technical Medicine Master student. Subsequently, a random subset of annotations (n = 6) was independently reviewed by two plastic reconstruction surgeons to ensure accuracy. Their observations and evaluations were combined and thoroughly discussed until a consensus was reached regarding the annotated structures. Adjustments, if

---

[1] http://www.slicer.org

required, were made to the segmentations accordingly. The segmented RA muscle, fat tissue, and skin tissue were then exported as a single binary mask that contained the information of all three segmentations. In cases where adjustments were necessary, the segmentations were modified accordingly. The segmented RA muscle, fat tissue, and skin tissue were exported as one binary mask, containing the information of the three segmentations, and stored in NIfTI file format. The DIEP segmentation was exported and stored in the same format.
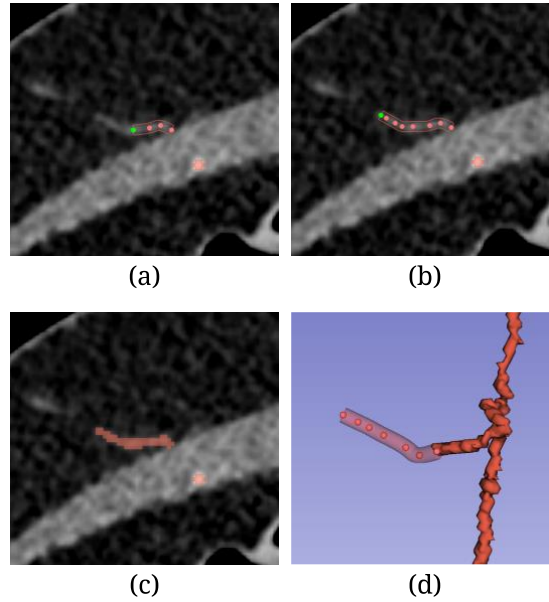

Figure 2. Manual segmentation of deep inferior epigastric artery perforators on a CTA scan. (a-c) Axial view of CTA scan where the tube wire function was used to perform the segmentation of the perforator vessels, (d) 3D view tube wire segmentation.

2.4 Deep learning models

In this study, we evaluated two DL models: no new U-Net (nnU-Net) (version 1) ([36](#)) and Dual branch Topology-aware U-Net (DTUNet) ([37](#)). nnU-Net was evaluated in various experiments, and a pre-trained DTUNet was only used and evaluated for inference.

2.4.1 nnU-Net

nnU-Net is an open-source network, that can be trained for various imaging segmentation tasks. It is a self-adapted framework and includes preprocessing of the images. The network architecture itself, training, and post-processing of the images. nnU-Net is based on the U-Net architecture ([38](#)) and is designed to handle various 2D and 3D imaging techniques. The wide applicability of nnU-Net was shown by applying the model to 53 segmentation tasks, including a diversity of imaging types, structures, and properties.

The preprocessing, framework, training, and post-processing steps for nnU-Net used in our study include:

- Preprocessing: No adaptations were made to the original nnU-Net preprocessing pipeline. nnU-Net automatically determines the preprocessing parameters based on the data fingerprint. The determined intensity normalization for CT scans includes percentile clipping and z-score normalization with the global foreground mean and standard deviation. Our input data was anisotropic, which means that nnU-Net used third-order

spline interpolation for resampling in the in-plane dimensions and nearest-neighbor interpolation for resampling in the out-of-plane dimension. For the corresponding annotations, the same resampling strategy is performed, yet only nearest neighbor interpolation is used to ensure the labels remain binary.

- Framework: In our experiments, the 3D full-resolution U-Net and an ensemble of the 3D full-resolution and 3D full-resolution cascade U-Net networks were used.

- Training: 5-fold cross-validation was used for training. For the tissue segmentations, 1000 epochs, and for DIEP segmentation, 1500 epochs were used. No additional adaptations were made to the nnU-Net training strategy. A wide range of data-augmentation methods were used, including rotations, scaling, mirroring, brightness, contrast, Gaussian noise, Gaussian blur, simulation of low resolution, and gamma correction, leading to an increased amount of training data. Moreover, the Stochastic Gradient Descent optimizer and cross-entropy loss function were used. nnU-Net uses a learning rate schedule starting with a high learning rate that decays during learning. The initial learning rate starts at 1e-3 and is reduced by a factor of 10 when the validation loss plateaus. The minimum learning rate is 1e-6.

- Post-processing: nnU-Net post-processing steps include a combination of connected component analysis, hole filling, and morphological operations. However, in our results, no differences were observed when comparing the raw and post-processed predictions, indicating that the post-processing steps did not affect the results.

For training and evaluating nnU-Net, the dataset was randomly divided into a train and test set of 25 and 5 scans, respectively. The RA muscle, fat, and skin segmentations were used separately from the DIEP segmentations.

2.4.2 DTUNet

DTUNet is a network developed for automated brain vessel extraction from CTA images. Brain vessel segmentation poses the challenge of segmenting small blood vessels. Considering the similar challenges encountered in DIEP vessel segmentation makes the application of DTUNet interesting for our case.

- Preprocessing: The preprocessing steps consisted of clipping the voxel intensities between -100 and 1500 HU and normalization to a range of 0-1. Furthermore, the intensities of the pixels in the subcutaneous fat tissue were scaled to the same median intensity of the RA muscle tissue in the DIEP scan (see Figure 3b). This scaling process was performed because the contrast between the vessels and subcutaneous fat tissue did not present the same low contrast as observed in brain vessels. Subsequently, the images were masked using the manually annotated RA muscle and fat tissue masks to only provide results for the region of interest (see Figure 3c).

- Framework: The DTUNet architecture consists of a dual-branch 3D U-Net with one branch for vessel centerline, and one branch for lumen segmentation. In our study, we only extracted the results from the lumen segmentation branch. The 3D-UNet consisted of a classic encoder architecture with a depth of four layers, with the addition of a fusion block to concatenate the two branches, using PyTorch. The network was trained using a topology-aware loss function, consisting of the Dice loss for the lumen segmentation.

- Post-processing: DTUNet was trained on brain vessel CTA images to predict all vessels within these input images. Since there were no ground truth annotations of all vessels in the CTA scans, additional post-processing steps were performed to remove small structures in the lateral parts of the fat tissue and small segments (<10 pixels) to eliminate noise. Furthermore, the Superficial Inferior Epigastric Artery (SIEA) vessels were manually removed.
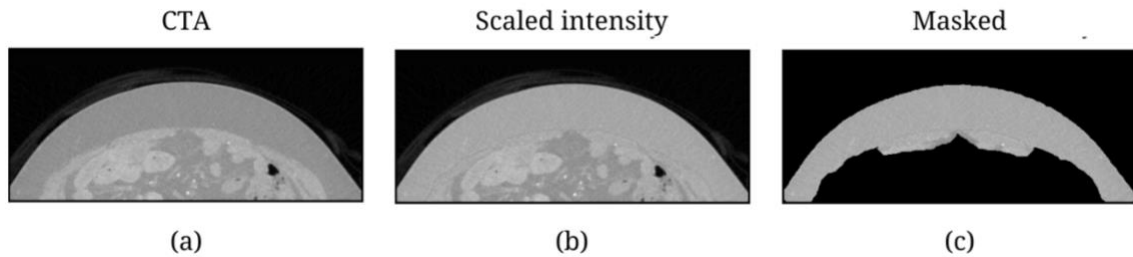

Figure 3. Intensity scaling of subcutaneous fat tissue to the median intensity value of rectus abdominis muscle in an abdominal CTA. Scans viewed in the axial plane. (a) Original CTA scan; (b) Scan after scaling intensity in subcutaneous fat tissue; (c) Scan masked to subcutaneous fat, muscle tissue, and the dilated region around muscle tissue.

2.5 Quantitative evaluation

The main evaluation metrics used in medical image segmentation include overlap-based metrics and distance-based metrics. This section outlines the metrics we used to evaluate the performance of DL models for DIEP segmentation.

2.5.1 Overlap-based metrics

The Dice similarity coefficient (DSC), recall, and precision, are commonly used overlap-based metrics for the evaluation of the performance of DL models for medical image segmentation ([29]). These measures are bounded between 0 and 1, with a score of 1 indicating a perfect alignment between the automated and manual segmentations.

The DSC ([39]) measures the overlap between the automated and the manual segmentation (ground truth), and is defined using overlap partitions, including the true positive (TP), true negative (TN), false positive (FP), and false negative (FN):

$$DSC = \frac{2 \; x \; TP}{2 \; (TP+FP+FN)} \qquad (1)$$

In addition to the DSC, recall, and precision were evaluated. Recall, or sensitivity, measures the proportion of voxels successfully detected by the automatic segmentation, while precision measures the accuracy of positive predictions. The recall was included as an evaluation metric since it is desired to miss as fewest perforators as possible, thus aiming for a high recall. Precision was evaluated since it represents the proportion of automated predicted vessels that are true positives, which gives us information on whether the automated predicted perforators are actual true positives. The recall and precision are defined as:

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

2.5.2 Distance-based metrics

Distance-based metrics, such as the Average Symmetric Surface Distance (ASSD), are recommended for evaluating the performance of DL models for small segments ([29]). In this study, the ASSD was included as metric for quantitative evaluation of the automated segmentation performance. The ASSD measures the average of all distances at every point on their contour ([40]), where a distance of 0 mm would refer to a perfect segmentation.

2.5.3 Distance-overlap derived metrics

Most widely used distance-based metrics only look at maximum or average values of the surface distances. These outcomes, however, can be highly affected by outliers and consist of only a single averaged or maximum value. For visualization of the DIEP vessels, small differences between a manual and automated segmented vessel may still be quantified as true vessel. It is therefore relevant to have insight into the distribution of the differences between the manual and the automated segmentation. It was discussed with two surgeons, that for our application, a 1.0 mm difference between the manual and the automated segmentation centerlines was still acceptable (true positive).

We defined two metrics that represent this accepted difference in distance: the True Positive Centerline Rate (TPCR), computed by dividing the number of TP centerline points by the total of ground truth centerline points; and the False Positive Centerline rate (FPCR), defined as the ratio between FP points and the total number of ground truth centerline points. These metrics give a more valuable insight into the performance of the model for our application since it provides information on the actual ratio of the prediction that is suitable for clinical use.

For the TPCR and FPCR metrics we first extracted the centerlines of both the manual and the automated DIEP segmentation. Secondly, for each point on the centerline of the automated prediction the Euclidian distance to the closest point on the manual centerline was computed, and vice versa. When the computed distance was less than 1.0 mm the point was labeled as TP, while a distance equal to or greater than 1.0 mm was counted as. Both the TPCR and FPCR are bounded between 0 and 1, where a score of 1 indicates perfect performance of the automated DL model in terms of centerline overlap alignment. More detail on the computation and definition of these metrics is provided in Appendix B.

2.5.4 Metric computation

The metrics were computed using masked labels. Masking of the DIEP segmentations was performed to exclude the area where the perforator branches from the femoral artery (see Figure 4). The manual and automated DIEP segmentations were masked using the manually annotated RA muscle and fat tissue. To still include the perforator main branch outside of the RA muscle tissue, the RA muscle mask was dilated using binary dilation from the SciPy ([41]) library, implemented in Python (version 3.9.16), with five iterations. The metrics were computed using the masked labels since the large caliber of the area where the DIEP branches from the femoral artery could highly affect the outcomes, and only the small caliber of the DIEP vessels is of our interest.

All metrics were computed with the left and right perforators separated. This means that the automated prediction and corresponding manual annotation were sliced in half. Metrics were computed on the separated vessels to obtain comparable results for all experiments.
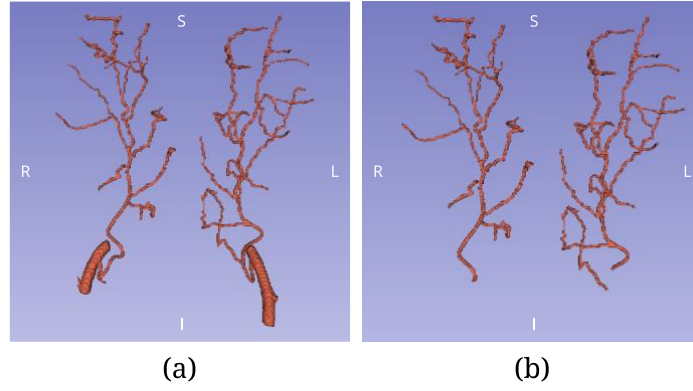
<div align="center">(a)   (b)</div>

Figure 4. Visualization of masked segmentation of the deep inferior epigastric artery perforator labels that were used in the quantitative metric evaluation. (a) original manual annotation; (b) masked manual annotation.

2.6 Qualitative evaluation

When evaluating the performance of DL segmentation models, quantitative metrics are commonly used to measure the differences between manual and automated segmentation However, these metrics alone may not provide a comprehensive assessment of the clinical acceptability of the automated segmentations. To ensure the clinical utility of the automated DIEP segmentations, it is essential to obtain qualitative feedback from clinical experts. Therefore, in addition to quantitative metrics, the clinical acceptability of the automated DL segmentations of the DIEP vessels was evaluated.

Four plastic reconstructive surgeons from the Erasmus MC, each with over 15 years of experience, participated in the qualitative evaluation of the DIEP segmentations. These surgeons were independent of the manual annotations and model training process. For the evaluation, a grading scale was used to record the acceptability of the presented segmentation by blinded surgeons. The surgeons were asked to assign a grade to each DIEP segmentation using the grading rates provided in Table 2. To minimize bias in the rating process, the type of segmentation (manual or automated) was not disclosed during the rating. A total of 15 CTA scans were included in this clinical evaluation.

The manual and automated segmentations of each subject were graded in a randomized order. The CTA scans and segmentation masks of the DIEP were visualized in axial view on a 2D screen in 3D Slicer, where the original scan was shown parallel to the scan with segmentation (Figure 5). Segmentations for the same subject were graded at least one week apart to reduce the bias of seeing the same images twice.

First, the mean grading scores of the manual and automated segmentations for each surgeon were compared to assess whether there was a difference in grading of the two segmentation methods by one surgeon. Subsequently, the mean scores of all surgeons for both the manual and automated segmentations were compared. Finally, the inter-rater agreement of the grading scores for each subject was assessed.

Table 2. Grading scheme used for clinical acceptability evaluation of deep inferior epigastric artery perforator segmentations in computed tomography angiogram.

| Grade	Definition
0	Not acceptable, manual (re)drawing of the entire structure is required
1	Acceptable, but corrections necessary, perforator(s) missing suitable as ideal flap perforators.
2	Acceptable, only minor corrections necessary. Perforator(s) missing, but not suitable or relevant for ideal flap perforator selection.
3	Accepted, no corrections necessary for clinical use.
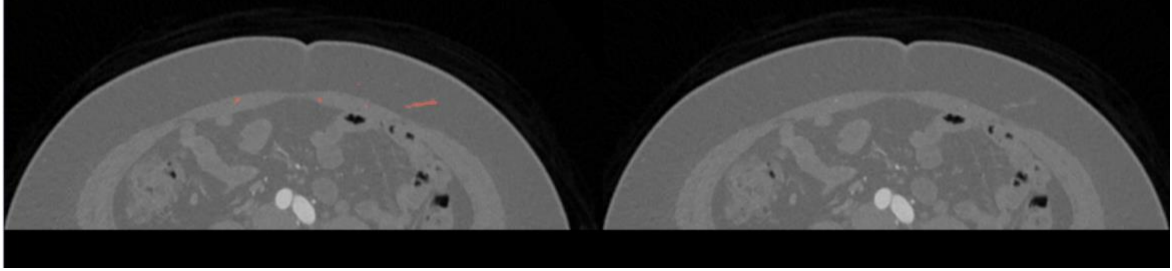
Figure 5. Parallel view in 3D Slicer that was shown to the surgeons during grading the acceptability of manual and automated segmentation of the deep inferior epigastric artery perforators (red) on CTA.

2.7 Statistical analysis

For the quantitative metrics, the mean and standard deviation were reported. The mean was computed as the micro average, defined as:

$$Metric_{mean\ value} = \frac{1}{n} * (M_{v1} + M_{v2} + \cdots M_{vn}) \qquad (4)$$

Where n represents the number of validation or test cases in a set, and M_{v1} represents the metric for one validation or test subject.

Statistical analysis was performed using Python (version 3.9.13) with functions from the SciPy library ([41](#)). A p-value lower than 0.05 was considered statistically significant. For statistical analysis of the qualitative evaluation, the paired t-test or Mann-Whitney U test was used depending on the characteristics of the data. Specifically, the paired t-test was utilized when analyzing paired data with equal dataset sizes, while the Mann-Whitney U test was used for comparisons involving unequally sized datasets. The decision to use the Mann-Whitney U test was motivated by the small dataset sizes and the observed lack of normality in the data. Alternatively, if the dataset sizes were larger or the data met the assumptions of normality, an independent t-test would have been used instead.

For the clinical acceptability ratings, the gradings of the manual and automated segmentation for each surgeon and the gradings for the two methods of all surgeons combined were compared using the Wilcoxon signed-rank test. This is a non-parametric statistical test that is used to compare two paired samples.

For assessment of the inter-rater agreement between the clinical experts, the Fleiss' Kappa was used. This measures the extent of agreement between the ratings of the surgeons occurring by change, where values closer to 1.00 indicate a higher agreement among multiple raters ([42]). Generally, values above 0.80 indicate excellent or almost perfect agreement, values between 0.60 and 0.79 indicate substantial agreement, and values below 0.60 indicate moderate, fair, or poor agreement ([43]). However, given the participation of only four surgeons in the qualitative evaluation, the Fleiss' Kappa can yield small values due to minor disagreements among the surgeons. To complement the assessment of the inter-rater agreement, the percentage agreement between the surgeons was also examined. This provides a measure of the proportion of segmentations where the surgeons' ratings agreed.

3 Automated segmentation experiments

3.1 nnU-Net RA muscle, subcutaneous fat tissue, and skin tissue segmentation

nnU-Net was trained and evaluated with quantitative metrics for automated segmentation of the RA muscle, subcutaneous fat tissue, and skin tissue using the 25 scans in the train set. The 3D full-resolution and an ensemble with the 3D full-resolution and the 3D full-resolution cascade model were both trained and evaluated. We quantitatively evaluated the automated segmentations using the DSC score, recall, and precision. Moreover, we qualitatively evaluated the results by visualizing the predictions of the RA muscle. We focused more on the visualization of the RA muscle since for use in clinical practice, visualization of the segmented perforators and the muscle is most relevant compared to the other tissue structures. Based on the quantitative measures and visualization, as well as the ease of use and training, the best model was evaluated with the test set.

3.2 DIEP vessel segmentation experiments

3.2.1 nnU-Net

Different training set compositions were used to evaluate the performance of nnU-Net for DIEP segmentation. Training with different training set sizes was performed to evaluate the performance of nnU-Net using the framework's original settings for our dataset. Moreover, we evaluated the influence of training when splitting the scans in half, since the input scan size influences the patch size used when training with nnU-Net. It was evaluated what influence this would have on the performance.

The training compositions used in these nnU-Net experiments included:

- Small train set: Training with 15 of the 25 scans in the training set using the 3D full-resolution model. This experiment was performed to evaluate the influence of dataset size on the model performance.
- Full train data: The network was trained with all train data (n = 25) using the 3D full-resolution model.
- Left and right perforators separated: Scans were divided into two halves, resulting in separate input data for both the right and left DIEP vessels. The 3D full-resolution model was again used for training.
- Ensemble of networks: An ensemble of the 3D full-resolution network and the 3D full-resolution cascade network was used on the training set composition with the perforators separated. It was assessed whether using this ensemble would improve the segmentation performance.

The different experiments on training with different training compositions are summarized in Table 1.

Table 1. Training set compositions investigated in training and testing nnU-Net for deep inferior epigastric artery perforator segmentation on CTA scans. Each training set composition was trained for 1500 epochs, without any additional adaptions to the nnU-Net default configurations and settings. 3D fullres = 3D full-resolution U-Net model; 3D cascade = 3D full-resolution cascade model.

Training set composition	nnU-Net framework	Image size (x, y, z)	Patch size (x, y, z)	# train
Small train set	3D fullres	512, 220, 700	160, 64, 224	15
Full train data set	3D fullres	512, 220, 700	160, 64, 224	25
Left and right perforators separated	3D fullres	256, 220, 700	96 , 96, 256	50
	Ensemble(3D fullres, 3D cascade)	256, 220, 700	96 , 96, 256	50

3.2.2 DTUNet

Inference with the pre-trained DTUNet was performed on the small training set composition (n = 15) used in the previous experiments. It was assessed whether this performance was suitable for our application based on quantitative and visual evaluation.

3.3 DIEP vessel segmentation evaluation with outer-cross validation

Because of our small dataset (n = 30), outer-cross validation was performed to evaluate whether the best-performing model would still achieve the same performance when using different train and test splits. In outer-cross validation, the dataset is divided into multiple non-overlapping train and test sets. For each iteration, a different split is used, ensuring that all samples are included in the test set at least once. Consistency and similarity in the results obtained from different outer-cross validation splits would indicate that the model's performance is not highly dependent on the specific initial train-test split. By employing outer-cross validation and observing consistent results across iterations, we aimed to strengthen the validity and reliability of our findings.

4 Results

4.1 RA muscle, subcutaneous fat tissue, and skin tissue segmentation

Segmentation of the RA muscle, subcutaneous fat tissue, and skin tissue resulted in a mean DSC score of respectively 0.92 (± 0.02), 0.96 (± 0.06), and 0.85 (± 0.09) using the 3D full-resolution network. For the ensemble of the 3D full-resolution and 3D full-resolution cascade network, mean DSC scores of 0.92 (± 0.02), 0.97 (± 0.04), and 0.85 (± 0.09) were measured. All the quantitative metrics for the tissue segmentations are presented in Table 3.

While both the full-resolution and the ensemble of networks resulted in the same mean DSC score, a significant difference (p = 0.002) for the RA muscle, favoring the ensemble model, was observed. Nonetheless, when visually comparing the predicted segmentations of the two network configurations, no clinically relevant differences were identified. A visual comparison of the predicted segmentations for a subset of CTA scans can be found in Appendix C, Figure C1. Since using one model for prediction is computationally less expensive and quicker, we evaluated the 3D full-resolution model with the test set.

Using the 3D full-resolution U-Net network on the test set resulted in a mean DSC score for the RA muscle, fat tissue, and skin tissue of 0.90 (± 0.03), 0.97 (± 0.02), and 0.86 (± 0.04) respectively. The DSC, recall, and precision are listed in Table 3. The results of the statistical analysis can be found in Appendix C, Table C1. Visualization of the automated predicted tissue segmentations of the test set can be found in Appendix C, Figure C2.

Table 3. nnU-Net performance for automated rectus abdominis muscle, subcutaneous fat tissue, and skin tissue segmentation. The mean value and standard deviation of the metrics are presented. The models used consisted of the 3D full-resolution U-Net and an ensemble of the 3D full-resolution and 3D full-resolution cascade U-Net networks. The test set was evaluated using the 3D full network. * Indicates that there was a significant difference (p < 0.05) between the 3D full-resolution network and ensemble, or between the full-resolution network cross-validation and test outcomes. DSC = Dice similarity coefficient; RA = rectus abdominis.

	RA Muscle			Fat tissue			Skin tissue		
	DSC	*Recall*	*Precision*	*DSC*	*Recall*	*Precision*	*DSC*	*Recall*	*Precision*
3D fullres	0.92 ± 0.02	0.93 ± 0.05	0.92 ± 0.06	0.96 ± 0.04	0.96 ± 0.06	0.97 ± 0.03	0.85 ± 0.09	0.87 ± 0.10	0.87 ± 0.13
Ensemble	0.92* ± 0.02	0.93 ± 0.05	0.92 ± 0.05	0.97 ± 0.04	0.96 ± 0.06	0.97 ± 0.02	0.85 ± 0.09	0.87* ± 0.10	0.87 ± 0.13
Test	0.90 ± 0.03	0.93 ± 0.05	0.89 ± 0.07	0.97* ± 0.02	0.97* ± 0.04	0.97* ± 0.03	0.86 ± 0.04	0.88 ± 0.07	0.85 ± 0.08

4.2 DIEP vessel segmentation experiments

4.2.1 nnU-Net

With the small train set, an average DSC score of 0.55 (± 0.08), and TPCR of 0.61 (± 0.11) were obtained for automated DIEP segmentation. Using the full train data set and the 3D full-resolution model on the separated perforator set, both resulted in a DSC score of 0.58 (± 0.08), with TPCR values of 0.63 (0.11) and 0.66 (± 0.10) respectively. The ensemble of the 3D full-resolution and 3D cascade model resulted in a DSC score of 0.60 (± 0.08), and TPCR of 0.66 (± 0.10). All the quantitative metrics for DIEP segmentation, obtained through cross-validation using nnU-Net with different training set compositions and networks, are summarized in Table 4. The loss curves for the experiments are provided in Appendix C, Figure C3.

Figure 6 shows the box plots of the results, with statistically significant differences indicated. For all but one, a statistically significant difference for the DSC was found. For the TPCR, a statistically significant difference was found between the full train set composition and both the 3D full-resolution (p = 0.01) and the ensemble of two U-Net networks (p = 0.02). All the outcomes of the statistical analysis can be found in Appendix C, Table C2.

Although the DSC scores were comparable, the use of the separated train set composition is slightly more supported based on the results and significant differences. When comparing the quantitative results of the ensemble of the 3D full-resolution and 3D full-resolution cascade networks with the results of solely the 3D full-resolution model, the ensemble resulted in a statistically significant higher DSC (p = 0.02) and precision (p = 0.03). However, upon the visual comparison of the cross-validation predictions between the 3D full-resolution network and the ensemble of networks, no clinically relevant differences were observed (see Figure 7).

Table 4. Quantitative metrics for different experiments with nnU-Net for automated deep inferior epigastric artery perforator segmentation on CTA. The mean value and standard deviation of the metrics are presented. ASSD = average symmetric surface distance; DSC = Dice similarity coefficient; FPCR = false positive centerline rate; TPCR = true positive centerline rate.

| Training set composition	Network	DSC	Recall	Precision	ASSD	TPCR	FPCR
Small train data set	3d fullres	0.55 ± 0.08	0.46 ± 0.11	0.70 ± 0.10	2.00 ± 1.11	0.61 ± 0.11	0.08 ± 0.04
Full train data set	3d fullres	0.58 ± 0.08	0.50 ± 0.12	0.72 ± 0.12	1.89 ± 0.88	0.63 ± 0.11	0.10 ± 0.06
Left and right perforators separated	3d fullres	0.58 ± 0.08	0.52 ± 0.10	0.71 ± 0.09	1.75 ± 0.87	0.66 ± 0.10	0.11 ± 0.05
	Ensemble	0.60 ± 0.08	0.52 ± 0.10	0.71 ± 0.09	1.75 ± 0.87	0.66 ± 0.10	0.10 ± 0.05
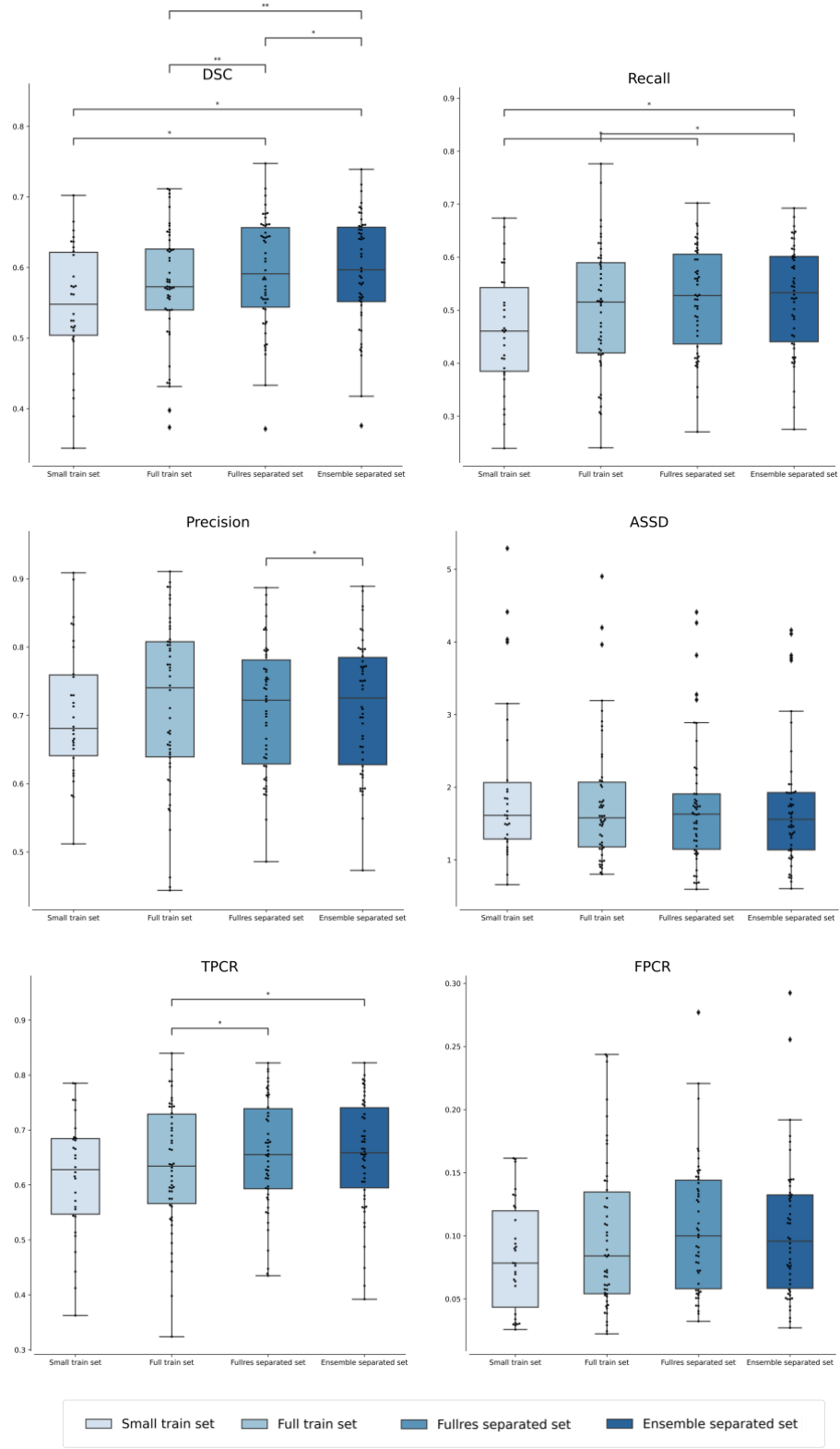
Figure 6. Boxplots of the quantitative metrics in the nnU-Net experiments for deep inferior epigastric artery perforator segmentation when training different training set compositions. The small (n = 15) and full (n = 25) train sets were trained with the 3D full-resolution network, and 5-fold cross-validation was used for training. The full training set with the left and right perforators separated was trained with the 3D full-resolution and the 3D full-resolution cascade model. ASSD = average symmetric surface distance (in mm); DSC = Dice similarity coefficient; FPCR = false positive centerline rate; TPCR = true positive centerline rate; * 1.00e-02 < p <= 5.00e-02; **: 1.00e-03 < p <= 1.00e-02.
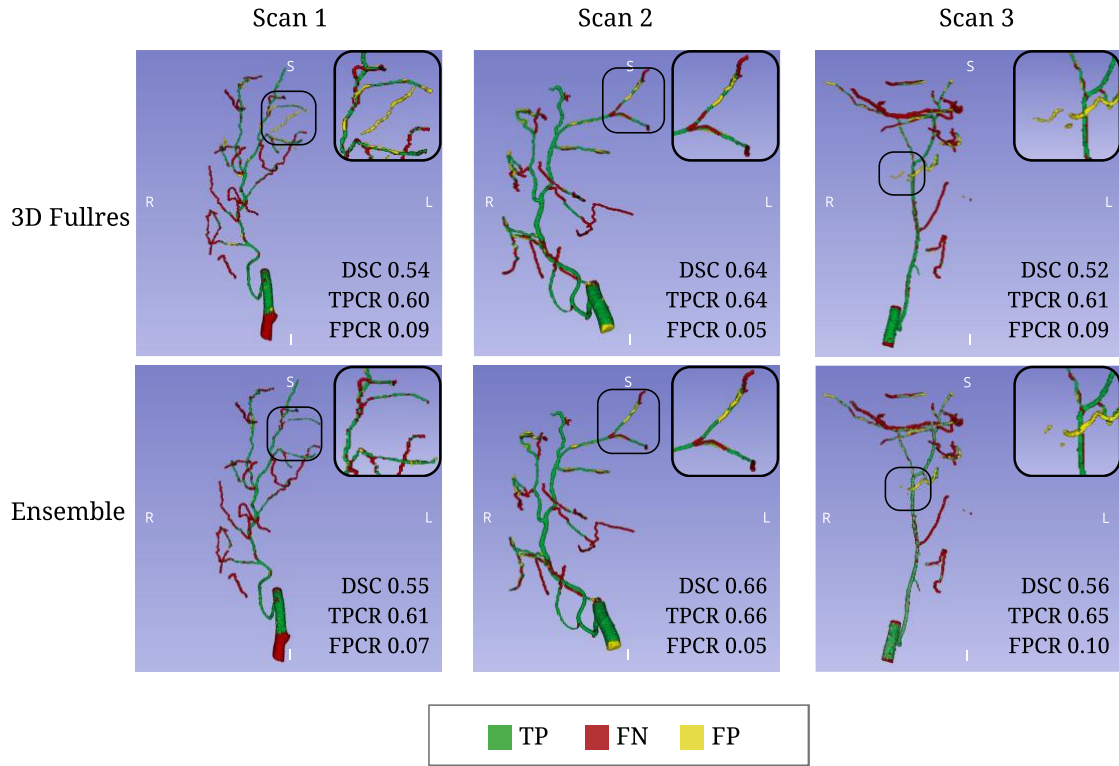
Figure 7. Visualization of automated segmentation of the deep inferior epigastric artery perforators for three CTA scans using nnU-Net with different networks. The first row displays the manual and automated predicted segmentation with the 3D full-resolution U-Net network, while the second row represents the ensemble of networks. No clinically relevant differences were observed when visually comparing the cross-validation predictions of the networks. DSC = Dice similarity coefficient; FPCR = false positive centerline rate; TPCR = true positive centerline rate; TP = true positives; FN = false negatives; FP = false positives.

4.2.2 DTUNet

After the manual post-processing steps, the average DSC between the manual and predicted segmentation was found to be 0.35 (± 0.11). The average TPCR was 0.38 (± 0.14). In addition, the DTUNet recall, precision, ASSD, and FPCR were 0.25 (± 0.11), 0.66 (± 0.17), 5.21 (± 2.84), and 0.10 (± 0.08) respectively. Figure 8 illustrates the automated segmentation results obtained by DTUNet before and after manual post-processing for the subject with the lowest and the subject with the highest TPCR.

Before post-processing   After post-processing

DSC 0.55
TPCR 0.68
FPCR 0.12
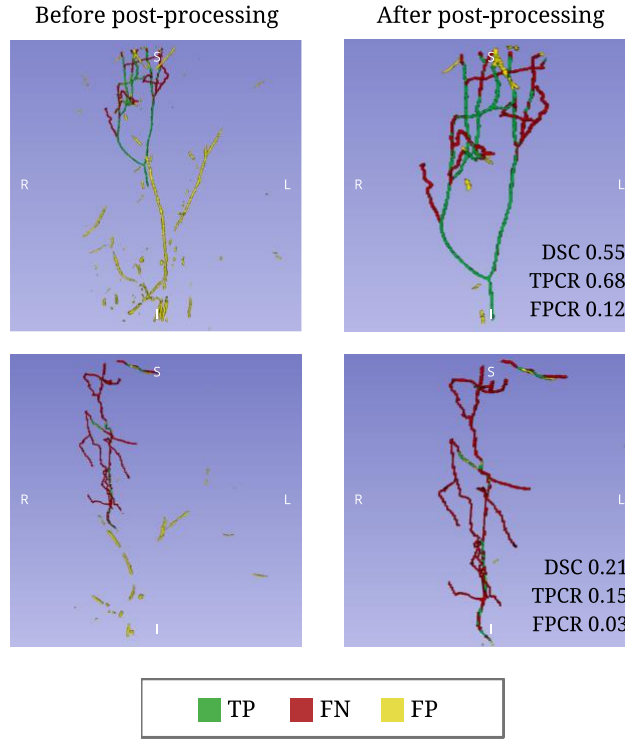
DSC 0.21
TPCR 0.15
FPCR 0.03

TP   FN   FP

Figure 8. Visual result of comparison of automated predictions with the pre-trained DTUNet and the manual segmentation before and after post-processing. DSC = Dice similarity coefficient; FPCR = false positive centerline rate; TPCR = true positive centerline rate; TP = true positives; FN = false negatives; FP = false positives.

4.3 DIEP vessel segmentation evaluation with outer-cross validation

The nnU-Net was used for outer-cross validation, as it demonstrated superior performance compared to DTUNet for our specific application. Since the 3D full-resolution U-Net model showed no visual differences when compared to the ensemble of two U-Net models, and using one model is computationally less expensive, we evaluated the 3D full-resolution model in the DIEP outer-cross validation.

Table 5 summarizes the quantitative metrics for the different train-test splits. Overall, the results across all folds demonstrated comparable performance, which validates the findings in our experiments. However, one split showed a statistically significant difference for the TPCR between the training and test set (p = 0.01) of the cross-validation results, and for one split, the precision of the test set was significantly different (p = 0.00) compared to the training set and the test set of the initial train-test split used in the experiments (p = 0.01). Detailed results of the statistical analysis can be found in Appendix C, Table C3. The average prediction time for one scan across the six repeated train-test splits was 231.17 (± 48.41) seconds.

During visual observation, we noticed that the automated segmentation successfully captured the main intramuscular DIEP branch, but it had difficulty identifying vessels in the subcutaneous fat. Interestingly, the 3D full-resolution model predicted some perforators that were overlooked in the manual segmentation. Figure 9 provides a visual representation of a test subject where the automated segmentation missed perforators, as well as another test subject where a perforator was overlooked in the manual segmentation.

A visualization of the manual and automated segmentations of the DIEP, along with the RA muscle, fat tissue, and skin tissue for one CTA scan, is provided in Appendix C, Figure C5.

Table 5. Results of nnU-Net using an ensemble of the 3D full-resolution and 3D full-resolution cascade U-Net network for segmentation of the deep inferior epigastric artery perforators on CTA. The first train set split corresponds to the initial train test split used in our experiments. ASSD = Average Symmetric Surface Distance; CV = cross-validation; DSC = Dice similarity coefficient; FPCR = false positive count; TPCR = True positive centerline rate; * = significantly different compared to the same split's CV results; ÿ = significantly different compared to the #1 test results.

| Train test split		DSC	Recall	Precision	ASSD	TPCR	FPCR
#1	CV	0.58 ± 0.08	0.52 ± 0.10	0.71 ± 0.09	1.75 ± 0.87	66% ± 10%	0.11 ± 0.05
	Test	0.55 ± 0.05	0.47 ± 0.09	0.70 ± 0.09	1.52 ± 0.25	63% ± 9%	0.08 ± 0.05
#2	CV	0.58 ± 0.07	0.51 ± 0.09	0.70 ± 0.10	1.67 ± 0.69	65% ± 10%	0.11 ± 0.05
	Test	0.58 ± 0.10	0.49 ± 0.14	0.74 ± 0.07	2.22 ± 1.31	64% ± 11%	0.11 ± 0.04
#3	CV	0.59 ± 0.07	0.53 ± 0.09	0.71 ± 0.10	1.76 ± 1.04	67% ± 9%	0.10 ± 0.05
	Test	0.55 ± 0.10	0.47 ± 0.12	0.67 ± 0.08	2.22 ± 1.06	57% ± 10%*	0.09 ± 0.04
#4	CV	0.58 ± 0.08	0.50 ± 0.10	0.70 ± 0.10	1.79 ± 0.87	64% ± 10%	0.10 ± 0.05
	Test	0.60 ± 0.06	0.52 ± 0.09	0.75 ± 0.10	1.83 ± 0.97	66% ± 7%	0.09 ± 0.04
#5	CV	0.59 ± 0.08	0.51 ± 0.11	0.71 ± 0.08	1.86 ± 0.90	64% ± 10%	0.11 ± 0.05
	Test	0.59 ± 0.05	0.57 ± 0.10	0.65 ± 0.16	1.40 ± 0.46	70% ± 11%	0.13 ± 0.07
#6	CV	0.58 ± 0.08	0.50 ± 0.10	0.69 ± 010	1.77 ± 0.74	64% ± 10%	0.10 ± 0.05
	Test	0.59 ± 0.04	0.48 ± 0.07	0.81 ± 0.06*ÿ	1.56 ± 0.46	65% ± 7%	0.09 ± 0.04
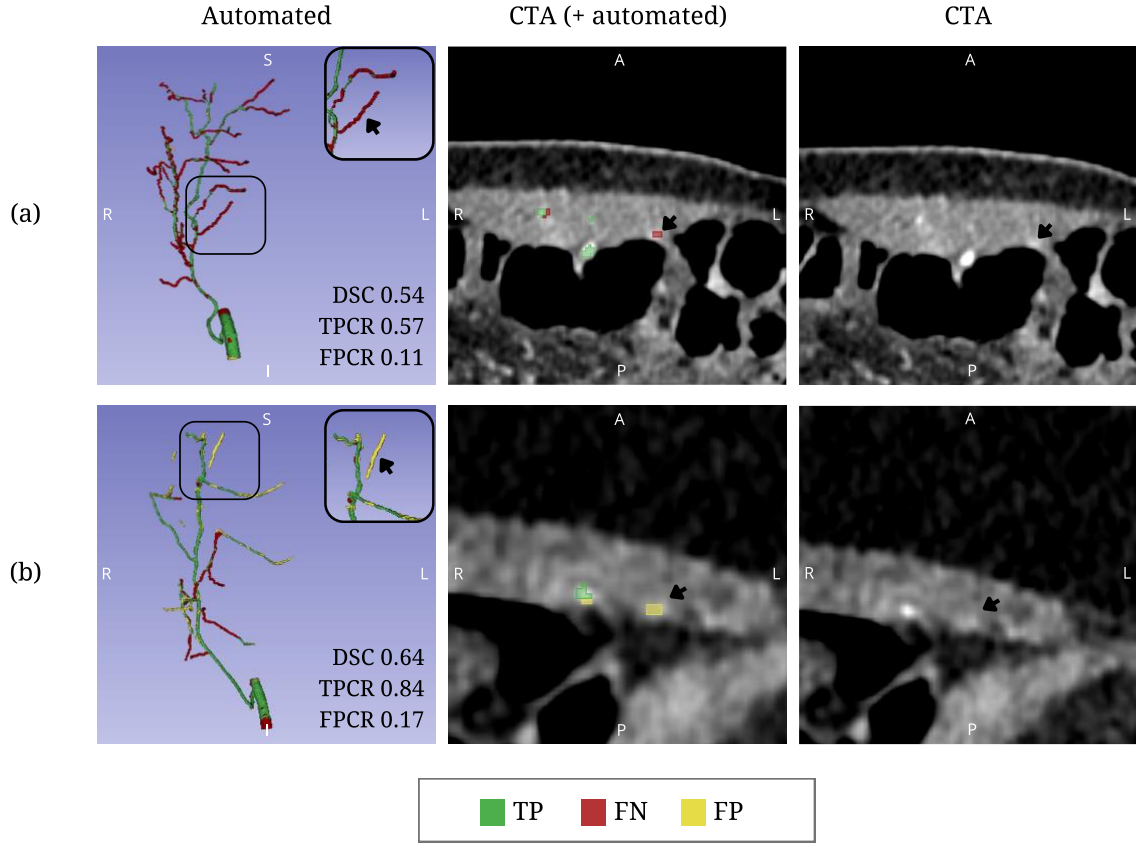
Figure 9. Visualization of automated segmentation outcomes for two test cases. The axial slices of the CTA scans are presented using a CT-abdomen window leveling in 3D Slicer. (a) Visualization of a false negative automated perforator; (b) Visualization of a false positive automated perforator that was missed in the manual annotation. DSC = Dice similarity coefficient; FPCR = false positive centerline rate; TPCR = true positive centerline rate; TP = true positives; FN = false negatives; FP = false positives.

4.4 Qualitative evaluation

The mean grades for the manual and automated segmentations were respectively 2.52 (± 0.62) and 2.37 (± 0.73). All except one rated both segmentations above 2.00. However, one surgeon gave a rating above 2.00 for the manual segmentation but not for the automated segmentation (1.67 ± 0.70). There were no statistical differences observed between the two segmentations for individual surgeons or in the combined result. Table 6 provides the mean grading scores for each individual surgeon, as well as the overall gradings and accompanying statistical results. Figure 10 visually presents the scores attributed by each surgeon separately. Appendix C, Table C4, and Figure C4 offer a detailed comparison of the results obtained from the evaluation of each of the 15 CTA scans individually.

The inter-rater agreement, as measured by Fleiss' Kappa, was 0.06 for manual segmentations, 0.02 for automated segmentations, and 0.04 when combining both segmentations. This indicated a very low level of agreement among the surgeons. The percentage agreements were respectively 13%, 7%, and 7%, for manual, automated, and combined gradings. The results for pair-wise percentage agreement are provided in Appendix C, Table C5.

Table 6. Mean acceptability score and the corresponding p-value (Wilcoxon signed-rank test) for surgeon qualitatively evaluating manual and automated deep inferior epigastric artery vessel segmentations. The mean scores are presented with ± one standard deviation. Four plastic reconstructive surgeons graded manual and automated segmentations of 15 CTA scans. A 4-point scale was used for grading: 0 = not acceptable; 1 = Acceptable, but corrections necessary; 2 = Acceptable, only minor corrections necessary; 3 = Accepted, no corrections necessary for clinical use.

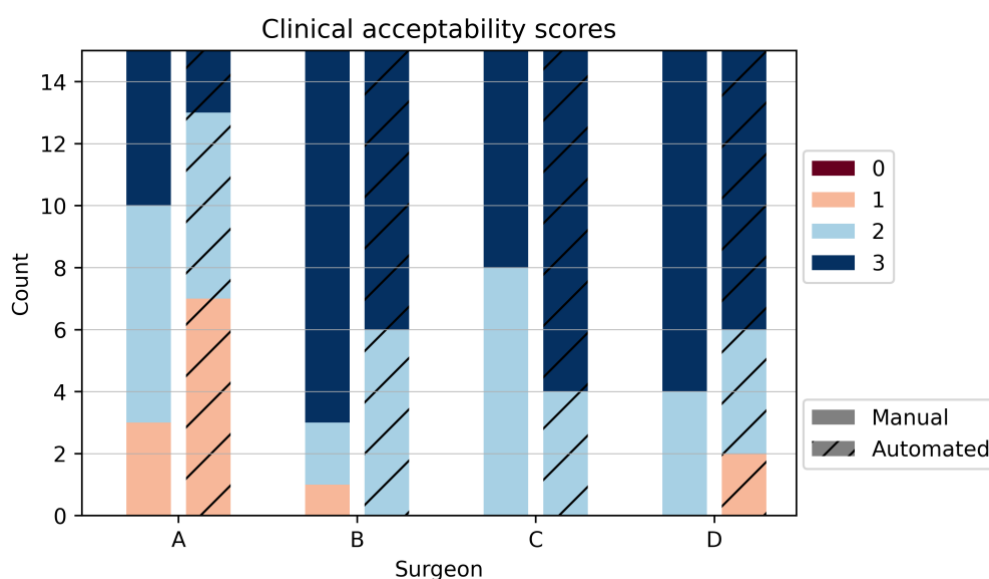| Surgeon	Mean acceptability score		p-value (Wilcoxon test)
	Manual	*Automated*	
A	2.13 ± 0.72	1.67 ± 0.70	0.08
B	2.73 ± 0.72	2.60 ± 0.49	0.32
C	2.47 ± 0.50	2.73 ± 0.44	0.10
D	2.73 ± 0.44	2.47 ± 0.72	0.21
All	2.52 ± 0.62	2.37 ± 0.73	0.13

Figure 10. Distribution of clinical acceptability scores assigned by each surgeon for manual and automated segmentations. The segmentations of 15 CTA scans were graded using a 4-point scale: 0 = not acceptable; 1 = Acceptable, but corrections necessary; 2 = Acceptable, only minor corrections necessary; 3 = Accepted, no corrections necessary for clinical use. Note that grade 0 was not assigned by any of the surgeons.

5 Discussion

The presented study aimed to evaluate the utility of DL segmentation for automated DIEP vessel segmentation in breast reconstruction surgery. 3D visualization of the perforators has shown to provide a better understanding of the anatomy and could reduce the harvest time of the DIEP flap. These 3D visualization methods, however, involve manual segmentation of the perforators, which is time-consuming, subjective, and prone to inter- and intra-observer variability. Automated segmentation using DL networks has the potential to overcome these limitations and provide more accurate and efficient segmentation of the DIEP vessels.

5.1 Quantitative results

Using automated DL for segmentation of the RA muscle, subcutaneous fat tissue, and skin tissue, resulted in high DSC scores for the test set. This means that there was a substantial overlap and agreement between the automated and manual segmentations, demonstrating the effectiveness of the DL model in accurately delineating these structures. These results, however, were not unexpected since the manual segmentation of the structures was already supported by intensity thresholding algorithms and the contours of the structures are distinguishable from their background. Nonetheless, the use of nnU-Net for the segmentations offers the advantage of providing an objective and fully automated result. For skin tissue, however, a simpler method such as volume rendering would suffice and yield faster results. Kreher et al. (2022) presented a U-Net network for muscle segmentation on CT scans evaluated using 130 scans and reported a DSC score of 0.86 (± 0.12) for segmentation of the RA muscle ([44]). Islam et al. (2022) reported a DSC score of 0.94 (± 0.04) for RA muscle segmentation when training with 1070 scans ([45]). In our study, we obtained a DSC score of 0.92 (± 0.02) on our training set. This indicates that using nnU-Net for RA muscle segmentation using a limited amount of data already shows comparable results with other studies.

The automated DIEP vessel segmentation on the initial test set resulted in a mean DSC of 0.58 (± 0.08) and a TPCR of 63% (± 9%), indicating moderate agreement between the automated and manual segmentations. However, a relatively low recall score of 0.47 (± 0.09) indicated that the model missed some perforators, which was evident during the visualization of the automated segmentations. Conversely, the precision score of 0.70 (± 0.09) highlighted that the majority of the predicted vessels were true positives rather than false positives. Further supporting this finding, the low FPCR of 0.08 (± 0.05) indicated that the segmentation model accurately identified DIEP vessels in the images.

Studies that have used deep learning for intracranial blood vessel segmentation on CTA scans have reported DSC scores of 0.91 and 0.87 on their respective datasets ([33], [46]), and the original study of DTUNet reported an ASSD of 0.84 (± 0.07) mm ([37]). While both intracranial blood vessel segmentation and DIEP vessel segmentation share the challenge of segmenting small vessels, the former has seen more advanced developments, leading to superior quantitative performance. Nevertheless, as the first study to explore DL for automated DIEP vessel segmentation, our obtained DSC score of 0.58 shows promise as a foundation. With further refinement and substantial expansion of the dataset, there is potential for improved performance to be achieved.

5.2 Qualitative results

As shown in Table 6, the gradings assigned by all surgeons for both manual and automated segmentations reveal no statistically significant difference between the two methods, and both segmentations were evaluated as clinically acceptable. However, the inter-rater agreement (0.04) and percentage agreement (7%) when combining all gradings indicate a low level of agreement among the surgeons. It is important to note that the limited number of participating surgeons in the evaluation may contribute to these low values. A more comprehensive assessment of the inter-rater agreement would require the involvement of a larger number of surgeons or segmentations.

Although the segmentations in our study were evaluated as clinically acceptable, it is worth noting that surgeons may have individual preferences regarding visualization. In our study, we only presented the axial view of the original CTA scans to the surgeons. However, some surgeons may be accustomed to assessing the DIEP anatomy using the maximum intensity projection scan or by viewing the scan from different planes. This variation in preferences could potentially impact the surgeons' grading. Another noteworthy observation from the evaluation was the potential influence of scan quality on the grading assigned by surgeons. We observed a bias towards assigning lower grades to both the manual and automated segmentations for CTA scans with lower image quality. This suggests that image quality plays a role in the surgeons' perception and assessment of the segmentations.

5.3 Future directions for clinical use

Although the presented study shows promising results for automated DL segmentation of DIEP vessels in breast reconstructive surgery, further improvement and development is necessary for clinical acceptability and use in clinical practice. Based on our quantitative and qualitative results, different recommendations can be made.

First, transfer learning and training could be used to further improve the performance of automated DL segmentation for DIEP vessels. In this study, transfer learning was not experimented with, but it could help improve the performance of the segmentation network by leveraging pre-trained models. Second, future studies could investigate the possibility of automatically calculating the perforator caliber, orientation, and distance of the segmented vessels to the umbilicus, both horizontally and vertically. This information could help surgeons in determining the most suitable vessel for flap harvesting and improve surgical outcomes.

Finally, it is worth investigating how the clinical acceptability of automated segmentations in 3D can be investigated. The set-up for our qualitative evaluation in 2D was based on previous studies that examined the clinical acceptability of automated segmentations for clinical applications using qualitative evaluation ([47-49]). However, acceptability in 2D may not accurately reflect the acceptability of segmentations in 3D. In our study, we chose not to evaluate the segmentations in 3D, as the differences between manual and automated segmentation would have been too apparent, and due to the potential impact of the visual-spatial ability of a surgeon on the qualitative grading when visualizing the results in 3D ([50-52]). Visual-spatial ability refers to the ability to perceive, analyze, and mentally manipulate visual information and understand how objects and spaces are arranged in relation to each other ([53], [54]). To date, no study has evaluated the clinical acceptability of automated segmentation for 3D visualization, which could be of interest for future research.

5.4 Limitations

The methods and experiments used in this study were subject to several limitations that may have impacted the results. First, a limited amount of annotated data was used for both training and testing, which could have affected the model's performance and may have contributed to a weaker statistical power. While there were no significant differences between the experiment with the small and full datasets, the addition of only 10 scans for training already showed less overfitting when evaluating the loss curves. Therefore, a larger annotated dataset could have potentially improved the performance of the model. Secondly, there was a wide range of scan quality which could not only have influenced the quantitative but also the qualitative results.

Furthermore, the manual segmentations of the perforators may have been inadequate. Although the learning curve was corrected by checking earlier annotations, the results showed that some perforators were still missed in the manual segmentations (see Figure 9). Another limitation of the manual annotations includes that only the DIEP vessels were annotated. Since the slices also contained other similar vessels, it would have been easier for the model to have also included the SIEA vessels. Finally, this study did not include external validation to evaluate the performance on an independent dataset, which could have provided more robust validation of the model's generalization to new datasets.

6 Conclusion

This study investigated the application of DL for automated perforator segmentation in DIEP flap breast reconstruction. Overall, our findings suggest that automated DL segmentation holds the potential to enhance the efficiency and objectivity of identifying DIEP vessels in CTA images, offering an alternative to manual segmentation. Nonetheless, further research is needed to refine the automated segmentation results and to validate the generalizability and clinical applicability of the DL segmentation approach in larger patient cohorts and different clinical settings.

Ethics statement

For this study, the Medical Research Involving Humans Subject Act (WMO) does not apply. Approval for the use of the CTA scans in our study investigating DIEP vessel segmentation and visualization for preoperative imaging was obtained from the medical ethical review committee. for preoperative imaging. For the clinical acceptability evaluation, the participants gave their informed written consent in participating in the experiment.

References

1.   Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.

2.   McGuire KP, Santillan AA, Kaur P, Meade T, Parbhoo J, Mathias M, et al. Are mastectomies on the rise? A 13-year trend analysis of the selection of mastectomy versus breast conservation therapy in 5865 patients. Ann Surg Oncol. 2009;16(10):2682-90.

3.   Carbine NE, Lostumbo L, Wallace J, Ko H. Risk-reducing mastectomy for the prevention of primary breast cancer. Cochrane Database Syst Rev. 2018;4(4):CD002748.

4.   Shiraishi M, Sowa Y, Tsuge I, Kodama T, Inafuku N, Morimoto N. Long-Term Patient Satisfaction and Quality of Life Following Breast Reconstruction Using the BREAST-Q: A Prospective Cohort Study. Front Oncol. 2022;12:815498.

5.   Nano MT, Gill PG, Kollias J, Bochner MA, Malycha P, Winefield HR. Psychological impact and cosmetic outcome of surgical breast cancer strategies. ANZ J Surg. 2005;75(11):940-7.

6.   Doval AF, Lamelas AM, Daly LT, Tobias AM, Lin SJ, Singhal D, et al. Deep Inferior Epigastric Artery Perforator Flap Breast Reconstruction in Women With Previous Abdominal Incisions: A Comparison of Complication Rates. Ann Plast Surg. 2018;81(5):560-4.

7.   Healy C, Allen RJ, Sr. The evolution of perforator flap breast reconstruction: twenty years after the first DIEP flap. J Reconstr Microsurg. 2014;30(2):121-5.

8.   Seidenstuecker K, van Waes C, Munder BI, Claes KE, Witzel C, Roche N, et al. DIEP flap for safe definitive autologous breast reconstruction. Breast. 2016;26:59-66.

9.   El-Mrakby HH, Milner RH. The vascular anatomy of the lower anterior abdominal wall: a microdissection study on the deep inferior epigastric vessels and the perforator branches. Plast Reconstr Surg. 2002;109(2):539-43; discussion 44-7.

10.   Ireton JE, Lakhiani C, Saint-Cyr M. Vascular anatomy of the deep inferior epigastric artery perforator flap: a systematic review. Plast Reconstr Surg. 2014;134(5):810e-21e.

11.   Smit JM, Klein S, Werker PM. An overview of methods for vascular mapping in the planning of free flaps. J Plast Reconstr Aesthet Surg. 2010;63(9):e674-82.

12.   Pratt GF, Rozen WM, Chubb D, Ashton MW, Alonso-Burgos A, Whitaker IS. Preoperative imaging for perforator flaps in reconstructive surgery: a systematic review of the evidence for current techniques. Ann Plast Surg. 2012;69(1):3-9.

13.   Teunis T, Heerma van Voss MR, Kon M, van Maurik JF. CT-angiography prior to DIEP flap breast reconstruction: a systematic review and meta-analysis. Microsurgery. 2013;33(6):496-502.

14.   Masia J, Kosutic D, Clavero JA, Larranaga J, Vives L, Pons G. Preoperative computed tomographic angiogram for deep inferior epigastric artery perforator flap breast reconstruction. J Reconstr Microsurg. 2010;26(1):21-8.

15.   Wade RG, Watford J, Wormald JCR, Bramhall RJ, Figus A. Perforator mapping reduces the operative time of DIEP flap breast reconstruction: A systematic review and meta-analysis of preoperative ultrasound, computed tomography and magnetic resonance angiography. J Plast Reconstr Aesthet Surg. 2018;71(4):468-77.

16.   DIEP flap breast reconstruction [Image]. Canadian Breast Cancer Network; [cited 2023 March 29]. Available from: https://www.cbcn.ca/en/autologous-reconstruction.

17.   Rozen WM, Ashton MW, Stella DL, Phillips TJ, Taylor GI. The accuracy of computed tomographic angiography for mapping the perforators of the DIEA: a cadaveric study. Plast Reconstr Surg. 2008;122(2):363-9.

18.   Rozen WM, Phillips TJ, Ashton MW, Stella DL, Gibson RN, Taylor GI. Preoperative imaging for DIEA perforator flaps: a comparative study of computed tomographic angiography and Doppler ultrasound. Plast Reconstr Surg. 2008;121(1):9-16.

19.   Cina A, Barone-Adesi L, Rinaldi P, Cipriani A, Salgarello M, Masetti R, et al. Planning deep inferior epigastric perforator flaps for breast reconstruction: a comparison between multidetector computed tomography and magnetic resonance angiography. Eur Radiol. 2013;23(8):2333-43.

20.   Hummelink S, Hoogeveen YL, Schultze Kool LJ, Ulrich DJO. A New and Innovative Method of Preoperatively Planning and Projecting Vascular Anatomy in DIEP Flap Breast Reconstruction: A Randomized Controlled Trial. Plast Reconstr Surg. 2019;143(6):1151e-8e.

21.   Wesselius TS, Meulstee JW, Luijten G, Xi T, Maal TJJ, Ulrich DJO. Holographic Augmented Reality for DIEP Flap Harvest. Plast Reconstr Surg. 2021;147(1):25e-9e.

22.   Fitoussi A, Tacher V, Pigneur F, Heranney J, Sawan D, Dao TH, et al. Augmented reality-assisted deep inferior epigastric artery perforator flap harvesting. J Plast Reconstr Aesthet Surg. 2021;74(8):1931-71.

23.   Jablonka EM, Wu RT, Mittermiller PA, Gifford K, Momeni A. 3-DIEPrinting: 3D-printed Models to Assist the Intramuscular Dissection in Abdominally Based Microsurgical Breast Reconstruction. Plast Reconstr Surg Glob Open. 2019;7(4):e2222.

24.   Chae MP, Hunter-Smith DJ, Chung RD, Smith JA, Rozen WM. 3D-printed, patient-specific DIEP flap templates for preoperative planning in breast reconstruction: a prospective case series. Gland Surg. 2021;10(7):2192-9.

25.   Freidin D, Singolda R, Tejman-Yarden S, Parmat I, Liran A, Ofir H, et al. Using Virtual Reality for Deep Inferior Epigastric Perforator Flap Preoperative Planning. Plast Reconstr Surg Glob Open. 2023;11(1):e4773.

26.   Araujo RJ, Garrido V, Baracas CA, Vasconcelos MA, Mavioso C, Anacleto JC, et al. Computer aided detection of deep inferior epigastric perforators in computed tomography angiography scans. Comput Med Imaging Graph. 2019;77:101648.

27.   Mavioso C, Araujo RJ, Oliveira HP, Anacleto JC, Vasconcelos MA, Pinto D, et al. Automatic detection of perforators for microsurgical reconstruction. Breast. 2020;50:19-24.

28.   Lange CJ, Thimmappa ND, Boddu SR, Dutruel SP, Pei M, Farooq Z, et al. Automating Perforator Flap MRA and CTA Reporting. J Digit Imaging. 2017;30(3):350-7.

29.   Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 2015;15:29.

30.   Muller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes. 2022;15(1):210.

31.   Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017;19:221-48.

32.   Zhou T, Tan T, Pan X, Tang H, Li J. Fully automatic deep learning trained on limited data for carotid artery segmentation from large image volumes. Quant Imaging Med Surg. 2021;11(1):67-83.

33.   Tetteh G, Efremov V, Forkert ND, Schneider M, Kirschke J, Weber B, et al. DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes. Front Neurosci. 2020;14:592352.

34.   Ayhan S, Oktar SO, Tuncer S, Yucel C, Kandal S, Demirtas Y. Correlation between vessel diameters of superficial and deep inferior epigastric systems: Doppler ultrasound assessment. J Plast Reconstr Aesthet Surg. 2009;62(9):1140-7.

35.   Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323-41.

36.   Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203-11.

37.   Liu S, Su R, Su J, Xin J, Wu J, van Zwam W, et al. A topology-aware framework for brain vessel centerline extraction from CTA images. Submitted (2023).

38.   Ronneberger O, Fischer P, Brox T, editors. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015 2015//; Cham: Springer International Publishing.

39.   Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology. 1945;26(3):297-302.

40.   Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. J Med Imaging (Bellingham). 2018;5(1):015006.

41.   Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-72.

42.   Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin. 1971;76:378-82.

43.   McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.

44.   Kreher R, Hinnerichs M, Preim B, Saalfeld S, Surov A. Deep-learning-based Segmentation of Skeletal Muscle Mass in Routine Abdominal CT Scans. In Vivo. 2022;36(4):1807-11.

45.   Islam S, Kanavati F, Arain Z, Da Costa OF, Crum W, Aboagye EO, et al. Fully automated deep-learning section-based muscle segmentation from CT images for sarcopenia assessment. Clin Radiol. 2022;77(5):e363-e71.

46.   Nazir A, Cheema M, Sheng B, Li H, Li P, Yang P, et al. OFF-eNET: An Optimally Fused Fully End-to-End Network for Automatic Dense Volumetric 3D Intracranial Blood Vessels Segmentation. IEEE Transactions on Image Processing. 2020;PP.

47.   Rauseo E, Omer M, Amir-Khalili A, Sojoudi A, Le TT, Cook SA, et al. A Systematic Quality Scoring Analysis to Assess Automated Cardiovascular Magnetic Resonance Segmentation Algorithms. Front Cardiovasc Med. 2021;8:816985.

48.   Garrett Fernandes M, Bussink J, Stam B, Wijsman R, Schinagl DAX, Monshouwer R, et al. Deep learning model for automatic contouring of cardiovascular substructures on radiotherapy planning CT images: Dosimetric validation and reader study based clinical acceptability testing. Radiotherapy and Oncology. 2021;165:52-9.

49.   Schreier J, Genghi A, Laaksonen H, Morgas T, Haas B. Clinical evaluation of a full-image deep segmentation algorithm for the male pelvis on cone-beam CT and CT. Radiotherapy and Oncology. 2020;145:1-6.

50.   Hegarty M, Keehner M, Cohen C, Montello D. The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals. 2010.

51.   Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED, Cusimano MD. Effect of visual-spatial ability on learning of spatially-complex surgical skills. Lancet. 2002;359(9302):230-1.

52.   Hedman L, Ström P, Andersson P, Kjellin A, Wredmark T, Felländer-Tsai L. High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. Surg Endosc. 2006;20(8):1275-80.

53.   Grieve J. Neuropsychology for Occupational Therapists; Assessment of Perception and Cognition Second Edition: Wiley-Blackwell; 1999.

54.   Goldstein EB, Cacciamani L. Sensation and perception: Cengage Learning; 2021.

Appendices

Appendix A - Manual segmentation protocol

1 Requirements

- Abdominal computed tomography angiography (CTA) scan (slice thickness of 0.6 – 0.75 mm)
- 3D Slicer (version 5.0.3) (http://www.slicer.org/) with FastMarching extension [a]

[a] Can be downloaded from the 3D Slicer manager.

2 Import data from XNAT

1. Login to XNAT and open the project-ID "diap" (MEC-2020-0504).
2. Select patient.
3. Select the CT-scan.
4. Select "Download images" and select the 0.75 CTA DICOM file of the abdomen.
5. Click download.

3 Cropping images in 3D Slicer

3.1 Load scan data

1. Open 3D Slicer
2. Click "Load DICOM Data"

3. Click "Import DICOM files"

4. Select the folder with the DICOM file.
5. Select scan and click "Load."

3.2 Crop volume in 3D Slicer

1. Switch to the "Markups" module and create a "Line" markup.
2. Position the first point (L1) at the height of the umbilicus using the different viewing plane (see Figure A1)
3. Position the second point (L2) 3000mm superior to L-1 in the sagittal viewing plane (see Figure A1)
   a. The exact position of the markup can be changed in the Markup module.
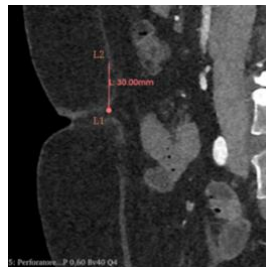

Figure A1. Sagittal view of mark-ups for cropping the image. L1 = first marker point; L2 = second marker point; L = length between L1 and L2

4. Switch to "Crop Volume" module

This module can be found with "Find module". Search for "Crop Volume" and press "Switch to module".

5. Create a new input ROI (Region of Interest) named "R". Select "Create new AnnotationROI as...". The input ROI must be named "R", else there will be an error in the next step.
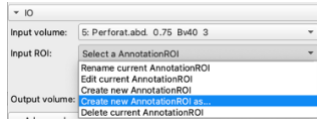
6. Open the "Python Interactor"(small blue and yellow symbol on the far right on the top toolbar). Copy and paste the following code in the interactor:

```python
# specify desired dimensions in voxels
radiusIJK = [512, 220, 700]
# assign the first volume node in the data module to a variable
volumeNode = slicer.util.getNode('vtkMRMLScalarVolumeNode1')
# assign roi node to a variable
roiNode = slicer.util.getNode('R')
# get the size of the volume voxels
spacing = volumeNode.GetSpacing()
# convert the number of pixels to measurements in mm
radiusRAS = [(radiusIJK[0]*spacing[0])/2, (radiusIJK[1]*spacing[1])/2,
(radiusIJK[2]*spacing[2])/2]
# change the size of the ROI
roiNode.SetRadiusXYZ(radiusRAS)
```

(This code sets the ROI in the 3D Slicer panels to a volume of 512x220x700 voxels.)

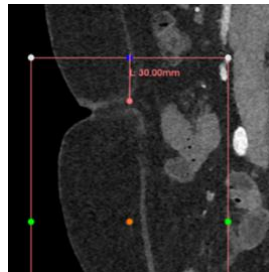7. Manually adjust the center superior ROI position to align it with the L-2 markup (see Figure A2).

Figure A2. Sagittal view of aligning the region of interest (ROI) for cropping the image. Orange dot = center of ROI; White dots = edges ROI; green dots = anterior and posterior middle points of ROI; Blue dot = middle of superior side of the ROI; L1 = first marker point; L2 = second marker point; L = length between L1 and L2

8. Align the ROI to cover the entire anterior abdominal skin in all axial slices.
   a. Do not change the dimensions of the ROI. If this occurs, then run the previous code again.
9. In the Advanced settings, set the "Fill value" to -1000 (HU air) and select interpolated cropping (keep interpolator on linear).
10. Press "Apply".

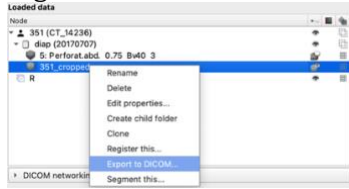## 3.3 Export cropped scan

3.3.1 Export to DICOM

1. After cropping the volume, switch to the "Add DICOM Data" module.

2. In the "loaded data" panel, right-click on the cropped volume and rename the file to "patientID_cropped".

3. Right click on the cropped volume and select "Export to DICOM…".

4. Change the "Seriesdescription" to "Cropped".
5. Export the file to the desired folder (use the default, "Scalar Volume" export type).

3.3.2 Export to NiFTI (.nii.gz)
To train the nnU-Net, the cropped scan volumes need to be saved in .nii.gz format!

1. Switch to the "Data" module.
2. Right-click on the cropped scan volume and select "Export to file…".
3. Name the file "diep_patientID_0000".
   a. For use with nnU-Net, the patientID needs to be in an XXX-digit format (e.g., 001, 002, etc.).

Example file name: DIEP_001_0000.nii.gz.

4. Select the desired folder to save the cropped scan.
5. Press "Save".

Note: for reproducibility, notate the ROI coordinates, these can be found in "MRML node information" under "Node information" in the "Data" module.

4 Segmentations

4.1 Import data

1. Repeat all steps in 3.1 (Load scan data) but select and import the cropped DICOM file.
   Or continue in 3D slicer after the previous step (cropping image), delete the original volume, and only work on the cropped scan.

4.2 Tissue annotations in 3D Slicer

4.2.1 Rectus Abdominis annotation
1. Open the "Segmentation editor" module.
2. Make sure that the master volume is set to the cropped scan.
3. Add a new segment and rename the segment to "Muscle."
4. Select "Paint" and activate "Sphere brush."
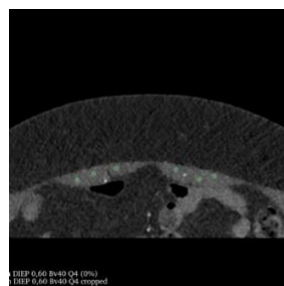5. Paint several spheres inside the abdominal muscle tissue (see Figure A3)

Figure A3. Axial view of spheres (green) painted inside the abdominal muscle tissue.

6. Select "Fast Marching" and set the "Maximum volume" to 6.00%.
7. Click "Initialize" and check the segmentation in the 3D view window.
    a. Click "Show 3D"
    b. Center the view in the 3D window ⊕.
8. Adjust the "Segment volume" to reduce artifacts.
9. Press "Apply".
10. Adjust the segmentation manually using "Scissors" and "Erase" to remove remaining artifacts.
11. Select "Smoothing", select "Closing" (kernel size 3.00 mm), and press "Apply".
12. Check the slices and adjust the segmentation manually to remove artifacts and fill the remaining holes.
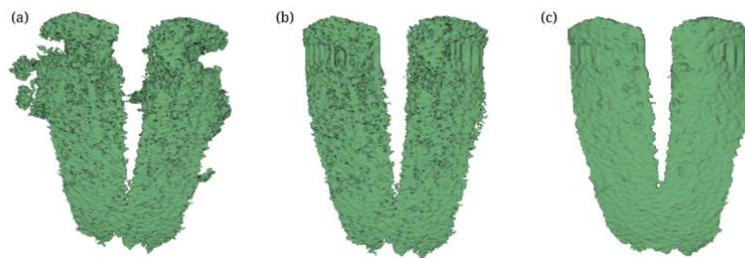

Figure A4: Anterior view of (a) rectus abdominis tissue after "Fast marching" at 6.00% Maximum volume and 85% Segment volume; (b) result after manually removing artifacts and manually filling gaps with "Paint" function; (c) final segmentation result after smoothing closing operator and filling remaining holes.

4.2.2 Fat tissue annotation
The steps for the fat tissue annotation are similar to the muscle tissue annotation (section 4.2.1).

1. Add a new segment and rename the segment to "Fat."
2. In "Masking" select editable area "Outside other segments" and "Modify other segments" on "Allow overlap".
3. Select "Paint" and activate "Sphere brush."
4. Paint several spheres inside the anterior abdominal fat tissue. Make sure to place the spheres distributed in the entire region (see Figure B5).
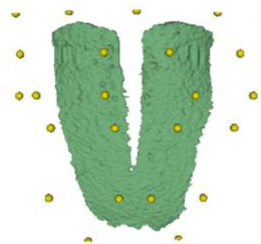

Figure A5. 3D view of painted spheres (yellow) for the fat tissue segmentation. Green = rectus abdominal muscle tissue segmentation

5. Select "Fast Marching" and set the "Maximum volume" to 40.00% (if needed, adjust to a higher or lower value based on the size of the fat region).
6. Click "Initialize" and check the segmentation in the 3D view window.
7. If necessary, adjust the "Segment volume" to reduce artifacts.
8. Press "Apply".
9. Adjust the segmentation using "Scissors" and "Erase" to remove the remaining artifacts.

10. Select "Smoothing", select "Closing" (kernel size 3.00 mm), and press "Apply". Adjust the kernel size if holes are not filled sufficiently.

11. Check the slices and adjust the segmentation manually to remove artifacts and fill the remaining holes. In Figure 6 the results before and after this step are visualized.
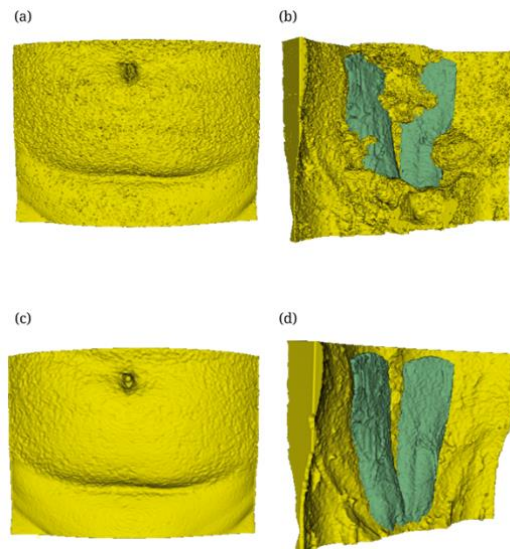

Figure A6. (a) Anterior view of lower anterior abdomen fat tissue after "Fast marching" at 40.00% Maximum volume and 90% Segment volume (b) posterior side view of the fat tissue; (b) anterior view of the result after manually removing artifacts and manually filling gaps with "Paint" function, and the smoothing closing operator; (d) posterior side view of the final segmentation result. Yellow = fat tissue segmentation, green = rectus abdominis segmentation

4.2.3 Skin tissue annotation

1. Add a new segment and rename the segment to "Skin."

2. Check that in "Masking" the selected editable area is still "Outside other segments" and that "Modify other segments" is set to "Allow overlap"!

3. Click on the "Threshold" effect and set the threshold range to intensities that the skin region is within these intensities (see Figure A7). The fat tissue layer can also be in-between these intensities.

Figure A7. Threshold effect for skin tissue segmentation. Red = threshold area; green = rectus abdominis muscle segmentation

4. Click "Apply".

5. In "Island", select "Keep selected island" and click on one of the pixels containing the skin layer.

    a. If this does not result in sole the skin layer, the segmented pixels behind the fat tissue could still be connected. Adjust the parts where the segmentation connects with artifacts using the eraser and/or scissors, and repeat "Keep selected island".

6. Select "Smoothing", select "Closing" (kernel size 3.00 mm), and press "Apply". Adjust the kernel size if holes are not filled sufficiently.

7. Adjust the segmentation using "Scissors" and "Erase" to remove remaining artifacts. A result of a skin segmentation is visualized in Figure A8.
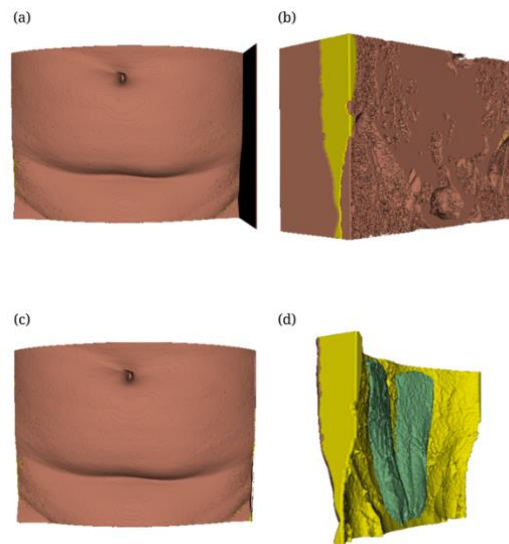

Figure A8. (a) Anterior view of skin tissue segmentation after thresholding; (b) posterior side view of a result after thresholding; (c) anterior view of a result after "Keep selected island"; (d) posterior view of skin tissue segmentation after "Keep selected island", removing artifacts and filling holes. Red = skin tissue segmentation; green = rectus abdominis segmentation; yellow = fat tissue segmentation

4.2.4 Perforator annotation

1. in "Masking" set the selected editable area to "Everywhere" and keep "Modify other segments" on "Allow overlap"!
2. In "Threshold" set the threshold range to intensities that cover the femoral artery and press "Use for masking" (see Figure A9)
3. Create a new segment named "DIEP"!
4. Using the "Paint" effect, manually fill in parts of the femoral artery to obtain segment parts of the Femoral artery inferior and superior to the DIEP branch (see Figure A9)
5. Create a new segment (use the default name or change the name is desired)
6. Select "Draw tube" .
7. Using control points with the "Draw tube" effect, follow the DIEP vessel and place marker points (see Figure A10). Press "Apply" when done with a vessel segment (keep the interpolation at "Cardial spline").
   a. Adjust the threshold in the masking setting when segmenting the vessel.
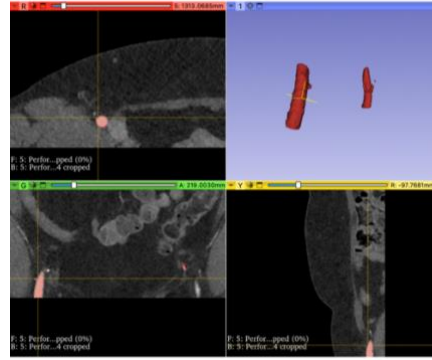   b. Adjust the radius of the tube if needed.

Figure A9. Screenshot of 3D Slicer scene for manual deep inferior epigastric artery segmentation. In the views the axial, coronal, and sagittal computed tomography angiograms respectively. In the upper right view, the 3D view of the segment.
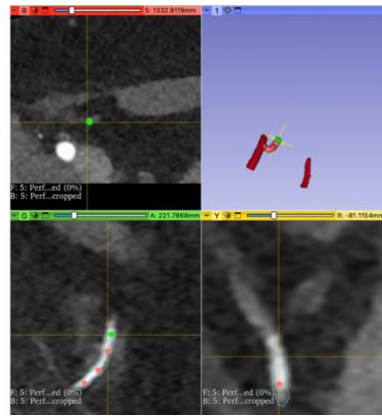

Figure A10. Screenshot of 3D Slicer scene for manual deep inferior epigastric artery segmentation using the "Draw tube" effect to segment the vessel. In the views the axial, coronal, and sagittal computed tomography angiograms respectively. In the upper right view, the 3D view of the marker points. The red dots are marker points; The blue lines indicate the tube path.

8. Create a new segment and repeat the previous steps until the whole vessel is segmented (see Figure A11a for a result).
    a. Using "Draw tube" again in a segment will replace the previous tube, therefore a new segmented needs to be created for each annotated vessel part.
    b. For each vessel part, adjust the threshold and/or radius if needed.
9. When done for all parts, go to "Logical operators"
10. In the segmentation editor, select the DIEP segment.
11. In "Logical operators", set the operation to "Add" and select a vessel segment.
12. Press "Apply".
13. Repeat this step until all vessel segments are added to the DIEP segment (see Figure A11b for a result).
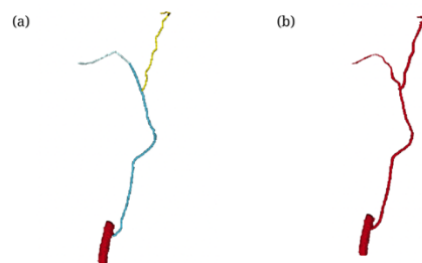

Figure A11. Anterior 3D view of right deep inferior epigastric artery perforator (DIEP) segmentation with 3D slicer. (a) Result of different vessel parts annotated using "Draw tubes". In red the femoral part. In blue and yellow colors, the annotated vessel parts; (b) results of the vessel after adding up the vessel parts.

14. Delete the vessel branch segments.
15. Check the DIEP segmentation.
16. Add vessel branches by repeating the previous steps. Figure A12 shows the results of the DIEP segmentation for one subject.
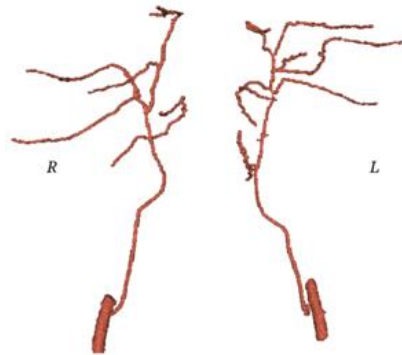

Figure A12. Anterior 3D view of a result of deep inferior epigastric artery perforator (DIEP) segmentation with 3D slicer.

4.3 Export data

4.3.1 Tissues

1. Go to the "Segmentation" module.
2. Make only the muscle, fat, and skin tissue visible.
3. In Export/Import models and label maps select "Export to new label map" as "output node"
4. In the advanced settings, set "Exported segments" to "Visible" and the "Reference volume" to the cropped scan.
5. Click "Export".
6. Go to the "Data" module.
7. Right-click on the created binary label map and select "Export to file…".
8. Change the filename to "TASKID_patientID.nii.gz" (TASKID for tissues used in this study is "tiss")
9. Change the desired folder for the segmentation to be saved in.
10. Select the file format .nii.gz.
11. Click "Export".

To export the tissue masks separately, only make the desired tissue segmentation visible, and repeat the same steps.

4.3.2 Perforators

1. Go to the "Segmentation" module.
2. Make only the DIEP vessel segmentation visible.
3. In Export/Import models and label maps select "Export to new label map" as "output node"
4. In the advanced settings, set "Exported segments" to "Visible" and the "Reference volume" to the cropped scan.
5. Click "Export".
6. Go to the "Data" module.
7. Right-click on the created binary label map and select "Export to file…"

8. Change the filename to "TASKID_patientID.nii.gz" (TASKID for DIEP vessels used in this study is "diep").
9. Change the desired folder for the segmentation to be saved in.
10. Select the file format .nii.gz.
11. Click "Export". The results of a DIEP and tissue segmentation are shown in Figure A13.
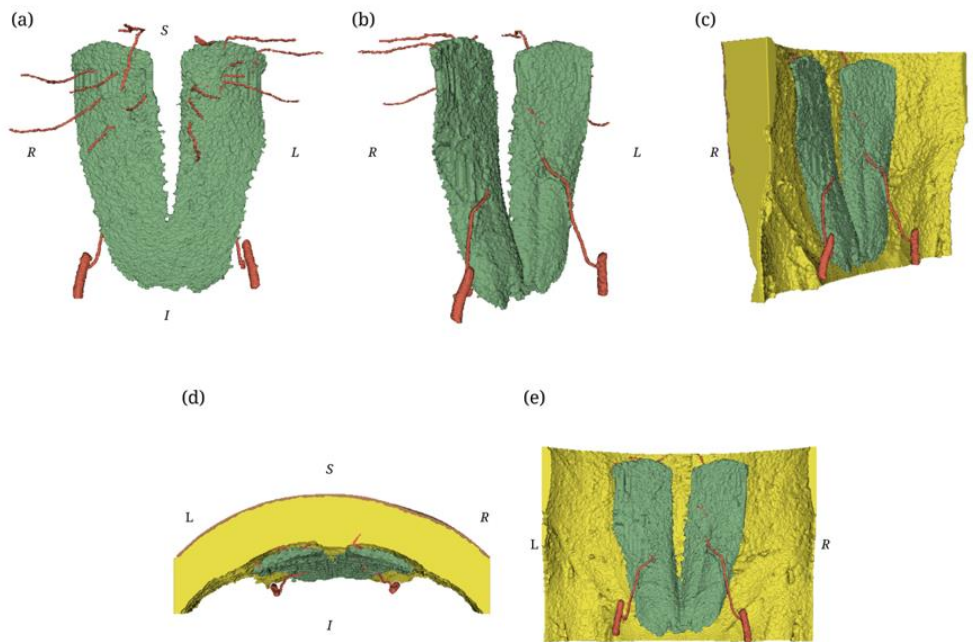
Figure A13. 3D view of the result of deep inferior epigastric artery perforator (DIEP), rectus abdominal muscle, anterior abdominal fat, and anterior skin layer segmentation with 3D slicer. Surface smoothing is turned off. (a) Anterior view of DIEP and muscle; (b) right side view of DIEP and muscle; (c) right side view of all segmentations; (d) superior view of all segmentations; (e) posterior view of all segmentations. Red = DIEP; green = rectus abdominal muscle; yellow = fat tissue; brown = skin tissue.

Appendix B – Centerline distance-based metrics

For the evaluation of the model performance, the distances between the centerlines of the predicted and manual segmentation were evaluated. It was discussed with two surgeons, that for our application, 1.0 mm of difference was still accepted (true positive). Two metrics were defined for this quantitative evaluation: True Positive Centerline Rate (TPCR) (Equation 1) and False Positive Centerline Rate (FPCR) (Equation 2).

$$TPCR = \frac{TP}{\#\ manual\ centerline\ points}\ (1) \qquad FPCR = \frac{FP}{\#\ manual\ centerline\ points}\ (2)$$

A centerline point was marked true positive (TP) when the distance from the ground truth centerline to the closest centerline point of the prediction was smaller than 1.0 mm, and a centerline point was marked false positive (FP) when the distance from the prediction centerline to the closest centerline point of the ground truth equaled or was larger than 1.0 mm. This computation and definitions are visualized in Figures B1 and B2. In these figures, the false negative (FN) and FP computed for the distances from automated to manual centerline points are also visualized, however, these were not reported or used in our defined metrics. The code used to compute the distances between the centerline points is provided on the next page.
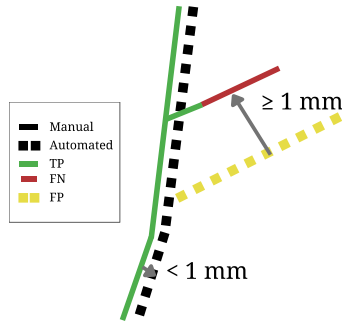
Figure B1: Visualization of metric definition. Distances < 1 mm were still considered a true positive, whereas points with distances ≥ 1 mm were either false negative or false positive. Automated = automated segmentation; FP = false positives; Manual = manual segmentation; TP = true positives.
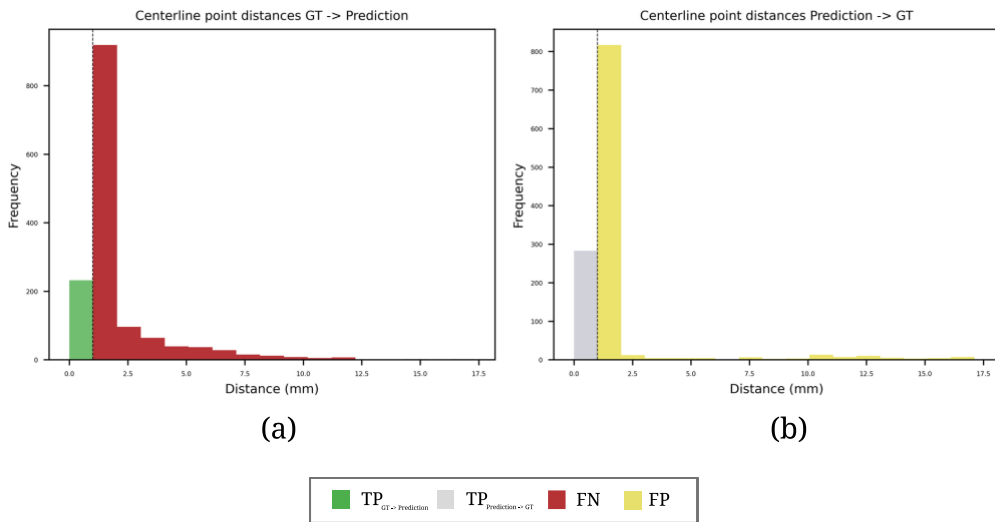
Figure B2: (a) Histogram of distances computed from the manual to the automated segmentation centerline points; (b) Histogram of distances computed from the automated to the manual segmentation centerline points. Automated = automated segmentation; FP = false positives; GT = ground truth; Manual = manual segmentation; TP = true positives; TP = true positives.

```python
def centerline_distances(path_gt, path_pred):
    """
    This function first extracts the centerlines of two binary masks (.nii.gz) format.
    Secondly, for each non-zero point in one binary array, it computes the distance (in pixels)
    to the closest non-zero point in the oposite binary array, and vice versa.

    Parameters
    ----------
    path_gt : str
        Path to folder with ground truth .nii.gz files.
    path_pred : str
        Path to folder with predicted .nii.gz files.

    Returns
    -------
    dis_pred_gt : list
        List with distances between the points on the centerline of the prediction
        and the closest point on the centerline of the centerline of the ground truth.
    dis_gt_pred : list
        List with distances between the points on the centerline of the ground truth
        and the closest point on the centerline of the centerline of the prediction.
    """

    # load data
    gt_data = nib.load(path_gt).get_fdata()
    pred_data = nib.load(path_pred).get_fdata()
    # skeletonize gt and prediction
    skeleton3d_gt = skeletonize_3d(gt_data)
    skeleton3d_pred = skeletonize_3d(pred_data)
    # voxel dimensions
    sampling_gt = nib.load(path_gt).header.get_zooms()
    sampling_pred = nib.load(path_pred).header.get_zooms()

    # Compute the Euclidean distance transform of the binary arrays
    dist_transform_gt = distance_transform_edt(~skeleton3d_gt, sampling = sampling_gt)
    dist_transform_pred = distance_transform_edt(~skeleton3d_pred, sampling = sampling_pred)

    # Initialize lists to store the distances
    dis_pred_gt = []
    dis_gt_pred = []

    # For each non-zero point in the first binary array,
    # find the nearest non-zero point in the distance transform of the second binary array
    for i,j,k in np.argwhere(skeleton3d_pred):
        local_dist_transform = dist_transform_gt[i, j, k]
        closest_point = np.unravel_index(np.argmin(local_dist_transform), local_dist_transform.shape)
        closest_point = np.array(closest_point) - 1
        distance = local_dist_transform[tuple(closest_point)]
        dis_pred_gt.append(distance)

    # For each non-zero point in the first binary array,
    # find the nearest non-zero point in the distance transform of the second binary array
    for i,j,k in np.argwhere(skeleton3d_gt):
        local_dist_transform = dist_transform_pred[i, j, k]
        closest_point = np.unravel_index(np.argmin(local_dist_transform), local_dist_transform.shape)
        closest_point = np.array(closest_point) - 1
        distance = local_dist_transform[tuple(closest_point)]
        dis_gt_pred.append(distance)

    return dis_pred_gt, dis_gt_pred
```

Appendix C - Supplementary tables and figures

Table C1. Statistical comparison of two models' cross-validation results and the cross-validation and test outcomes. Fullres refers to the 3D full-resolution network, and ensemble to the ensemble of the 3D full-resolution and the 3D full-resolution cascade U-Net networks. DSC = Dice similarity coefficient, RA = Rectus Abdominis muscle; * = significantly different (p < 0.05).

		DSC	Recall	Precision	Statistical test
RA	Fullres vs ensemble	0.00*	0.06	0.18	Paired t-test
	Fullres vs test	0.28	0.83	0.11	Mann-Whitney U
Fat tissue	Fullres vs ensemble	0.54	0.93	0.33	Paired t-test
	Fullres vs test	0.00*	0.01*	0.01*	Mann-Whitney U
Skin tissue	Fullres vs ensemble	0.11	0.02*	0.94	Paired t-test
	Fullres vs test	0.91	1.00	0.79	Mann-Whitney U

Table C2. Statistical comparison between cross validation results of nnU-Net experiments with different training set compositions. Fullres refers to the 3D full-resolution network, and ensemble to the ensemble of the 3D full-resolution and the 3D full-resolution cascade U-Net networks. ASSD = average symmetric surface distance; DSC = Dice similarity coefficient; FPCR = false positive count; TPCR = true positive centerline rate; * = significantly different p < 0.05).

	DSC	Recall	Precision	ASSD	TPCR	FPCR	Statistical test
Small set vs. full set	0.14	0.11	0.53	0.51	0.51	0.45	Mann-Whitney U
Small set vs. separated fullres	0.03*	0.02*	0.78	0.48	0.14	0.06	Mann-Whitney U
Small set vs. separated ensemble	0.01*	0.01*	0.67	0.36	0.12	0.14	Mann-Whitney U
Full set vs. separated fullres	0.00*	0.07	0.34	0.62	0.01*	0.30	Paired t-test
Full set vs. separated ensemble	0.00*	0.03	0.56	0.64	0.02*	0.54	Paired t-test
Separated fullres vs separated ensemble	0.02*	0.09	0.03*	0.96	0.97	0.17	Paired t-test

Table C3. Statistical comparison between repeated train-test splits and the initial train-test split. ASSD = average symmetric surface distance; CV = cross-validation; DSC = Dice similarity coefficient; FPCR = false positive centerline rate; TPCR = true positive centerline rate; * = significantly different (p < 0.05).

Train-test split		DSC	Recall	Precision	ASSD	TPCR	FPCR	Statistical test
#1	Test vs. CV	0.12	0.13	0.96	0.71	0.40	0.08	Mann-Whitney U
#2	Test vs. CV	0.71	0.49	0.24	0.27	0.70	0.96	Mann-Whitney U
	CV vs. CV #1	0.17	0.20	0.61	0.53	0.58	0.93	Paired t-test
	Test vs. test #1	0.56	0.72	0.26	0.16	0.93	0.18	Paired t-test
#3	Test vs. CV	0.17	0.21	0.16	0.19	0.01*	0.60	Mann-Whitney U
	CV vs. CV #1	0.81	0.74	0.87	0.95	0.59	0.71	Paired t-test
	Test vs. test #1	0.82	0.97	0.30	0.12	0.12	0.49	Paired t-test
#4	Test vs. CV	0.36	0.53	0.25	0.96	0.43	0.45	Mann-Whitney U
	CV vs. CV #1	0.09	0.10	0.70	0.79	0.23	0.40	Paired t-test
	Test vs. test #1	0.19	0.34	0.09	0.44	0.48	0.64	Paired t-test
#5	Test vs. CV	0.73	0.13	0.07	0.12	0.07	0.29	Mann-Whitney U
	CV vs. CV #1	0.75	0.70	0.98	0.47	0.40	0.85	Paired t-test
	Test vs. test #1	0.20	0.09	0.27	0.54	0.25	0.18	Paired t-test
#6	Test vs. CV	0.43	0.41	0.00*	0.55	0.99	0.48	Mann-Whitney U
	CV vs. CV #1	0.12	0.26	0.27	0.92	0.31	0.75	Paired t-test
	Test vs. test #1	0.20	0.96	0.01*	0.85	0.80	0.53	Paired t-test

Table C4. Results for qualitative evaluation of manual and automated DIEP segmentation for each CTA scan. A total of four plastic reconstructive surgeons (A-D) graded the manual and automated segmentation of 15 CTA scans using a 4-point grading scale. 0 = not acceptable; 1 = Acceptable, but corrections necessary, perforator(s) missing suitable as ideal flap perforators; 2 = Acceptable, only minor corrections necessary. Perforator(s) missing, but not suitable or relevant for ideal flap perforator selection; 3 = Accepted, no corrections necessary for clinical use. All represents the mean and standard deviation of the gradings of all surgeons combined. A = automated segmentation; M = manual segmentation.

Scan	Surgeon A		Surgeon B		Surgeon C		Surgeon D		All	
	M	A	M	A	M	A	M	A	M	A
1	3	1	3	3	2	3	3	3	2.75 ± 0.43	2.50 ± 0.87
2	2	1	3	3	3	3	3	1	2.75 ± 0.43	2.00 ± 1.00
3	2	2	3	3	3	3	3	3	2.75 ± 0.43	2.75 ± 0.43
4	3	3	3	3	3	2	3	2	3.00 ± 0.00	2.50 ± 0.50
5	2	1	3	2	2	3	3	3	2.50 ± 0.50	2.25 ± 0.83
6	3	1	2	2	2	3	2	3	2.25 ± 0.43	2.25 ± 0.83
7	2	2	2	2	2	3	2	3	2.00 ± 0.00	2.50 ± 0.50
8	3	1	3	3	3	3	2	1	2.75 ± 0.43	2.00 ± 1.00
9	3	2	3	3	3	3	3	3	3.00 ± 0.00	2.75 ± 0.43
10	1	1	3	2	2	2	3	3	2.25 ± 0.83	2.00 ± 0.71
11	1	2	3	3	2	3	3	2	2.25 ± 0.83	2.50 ± 0.50
12	2	2	3	2	2	2	3	2	2.50 ± 0.43	2.00 ± 0.00
13	1	1	1	2	3	3	2	2	1.75 ± 0.83	2.00 ± 0.71
14	2	3	3	3	3	3	3	3	2.75 ± 0.43	3.00 ± 0.00
15	2	2	3	3	2	2	3	3	2.50 ± 0.50	2.50 ±0.50

Table C5. Pair-wise percentage agreements between surgeons for manual and automated deep inferior epigastric artery perforator segmentation. A total of four surgeons (A-D), graded a total of 15 CTA scans using a 4-point grading scale. All refers to the percentage agreement for gradings of both the manual and automated segmentations combined.

Pair	Method		
	Manual	Automated	All
A, B	20%	27%	10%
A, C	20%	20%	10%
A, D	20%	27%	10%
B, C	13%	7%	7%
B, D	13%	13%	7%
C, D	13%	7%	7%

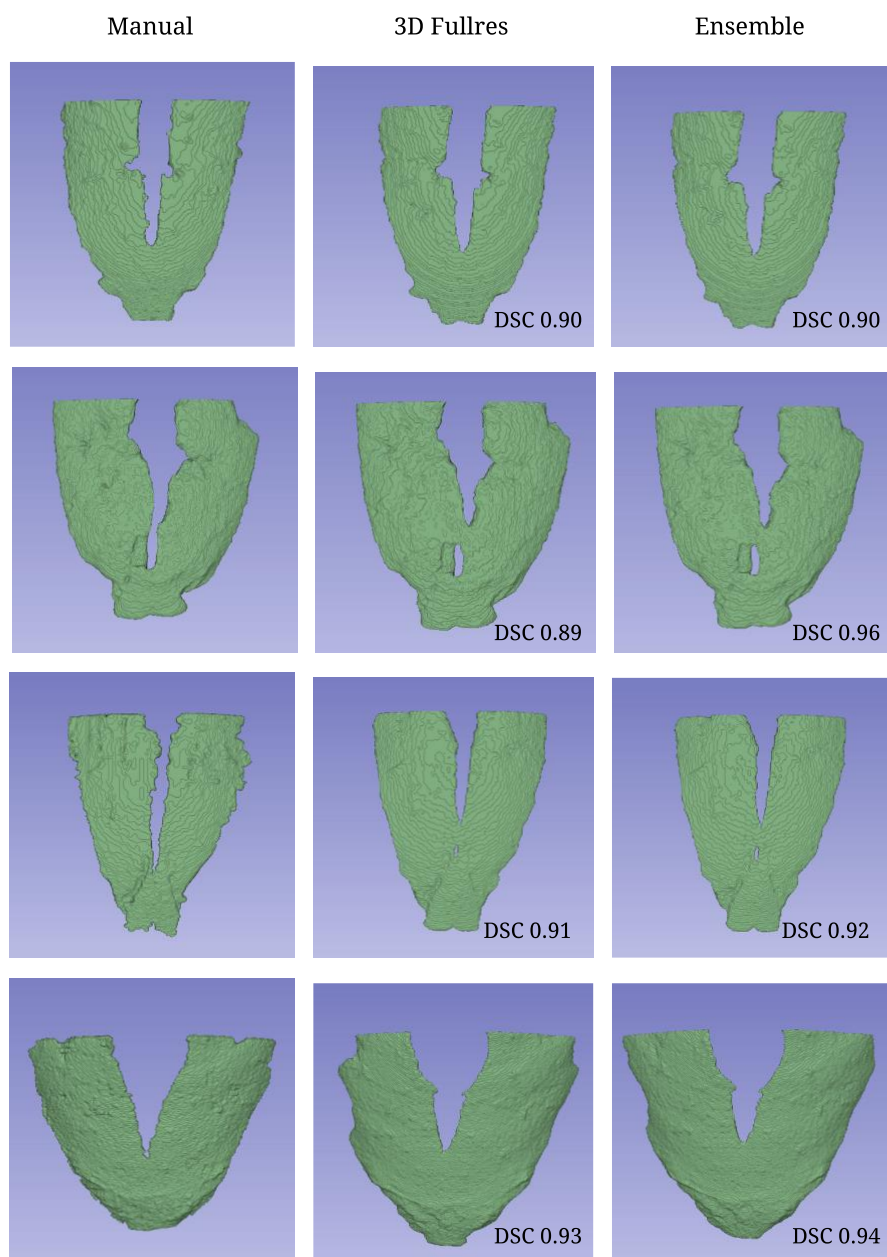| Manual	3D Fullres	Ensemble
	DSC 0.90	DSC 0.90
	DSC 0.89	DSC 0.96
	DSC 0.91	DSC 0.92
	DSC 0.93	DSC 0.94

Figure C1. Visualization of cross-validation segmentation results of nnU-Net for automated segmentation of the rectus abdominis (RA) muscle. Predictions with the 3D full-resolution network and predictions when using an ensemble of the 3D full-resolution (3D Fullres) and 3D full-resolution cascade (Ensemble) network are shown. The DSC for the RA muscle segmentations is presented. Automated = automated segmentation; DSC = Dice similarity coefficient.
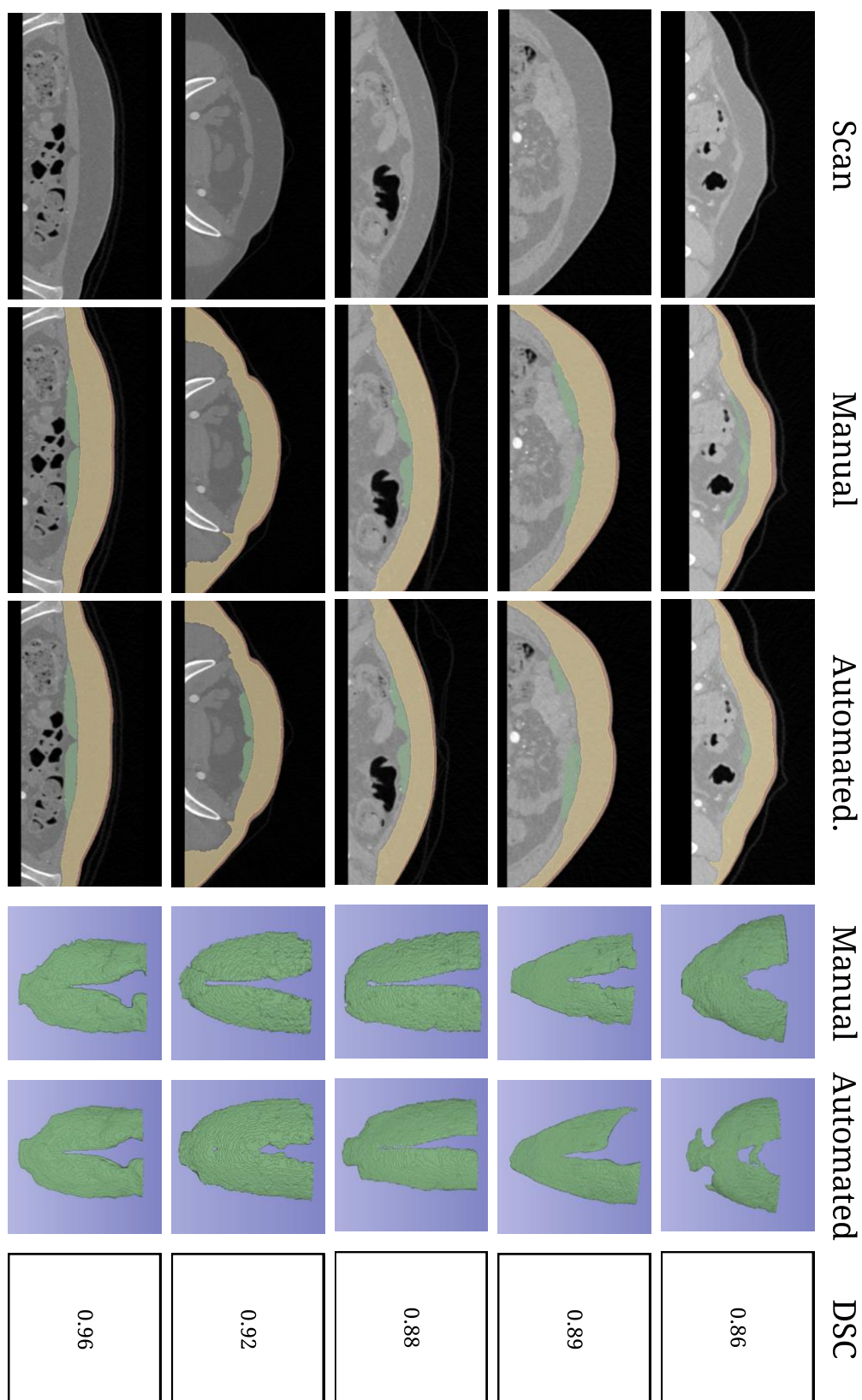
Figure C2. Visualization of segmentation results of nnU-Net for automated segmentation of the rectus abdominis (RA) muscle (green), subcutaneous fat tissue (yellow), and skin tissue (red). The scans are axial views of CTA scans. In the 3D models, the ground truth and predicted RA muscle segmentations are shown in the anterior view. The DSC scores for the RA muscle segmentations are presented. Automated = automated segmentation; DSC = Dice similarity coefficient; Manual = manual segmentation.
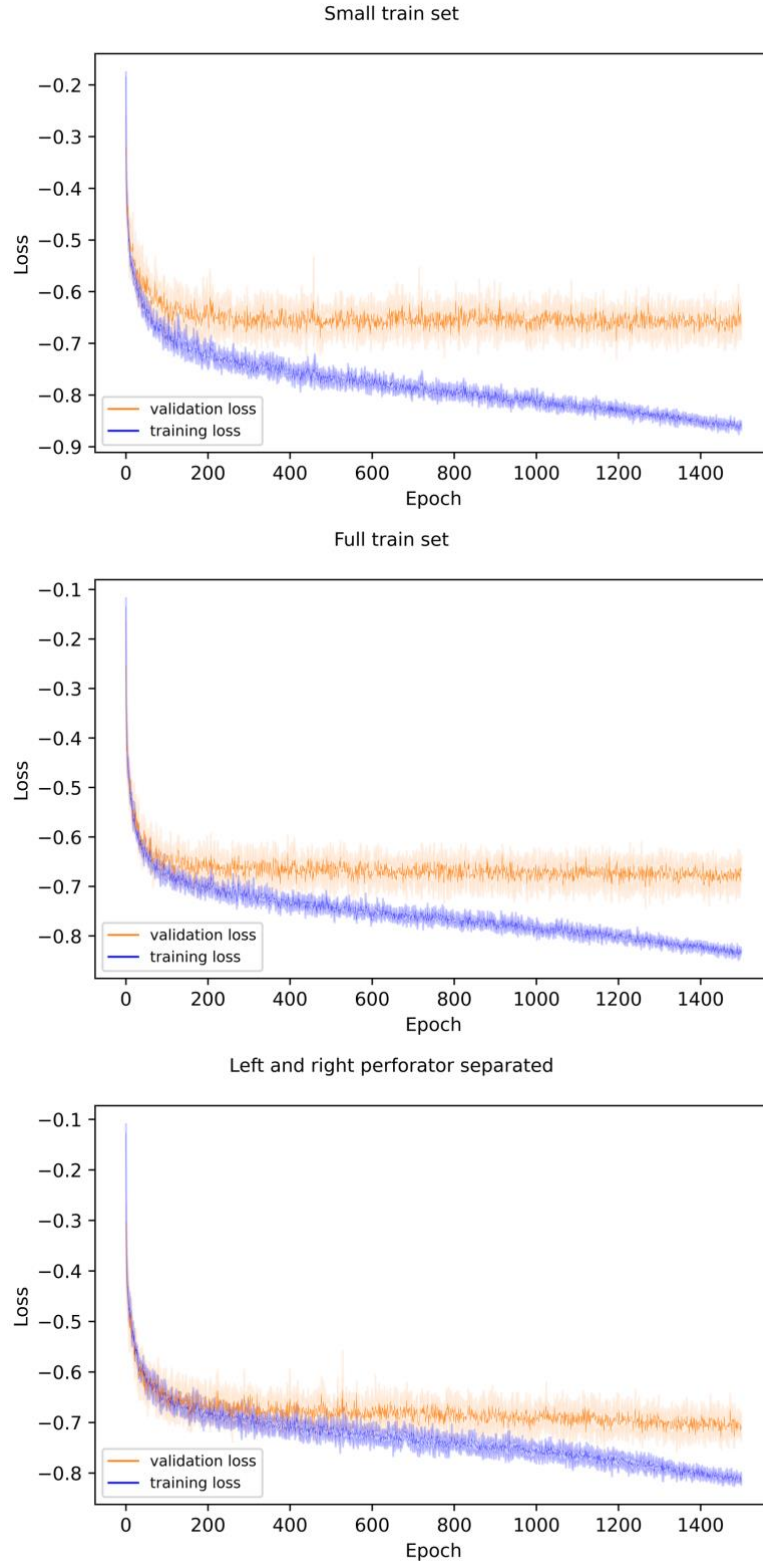
Figure C3. Loss curves of experiments using nnU-Net and the 3D full-resolution U-Net model for segmentation of deep inferior epigastric artery perforators on CTA. The validation and loss mean values of 5-folds are plotted with one standard deviation range.
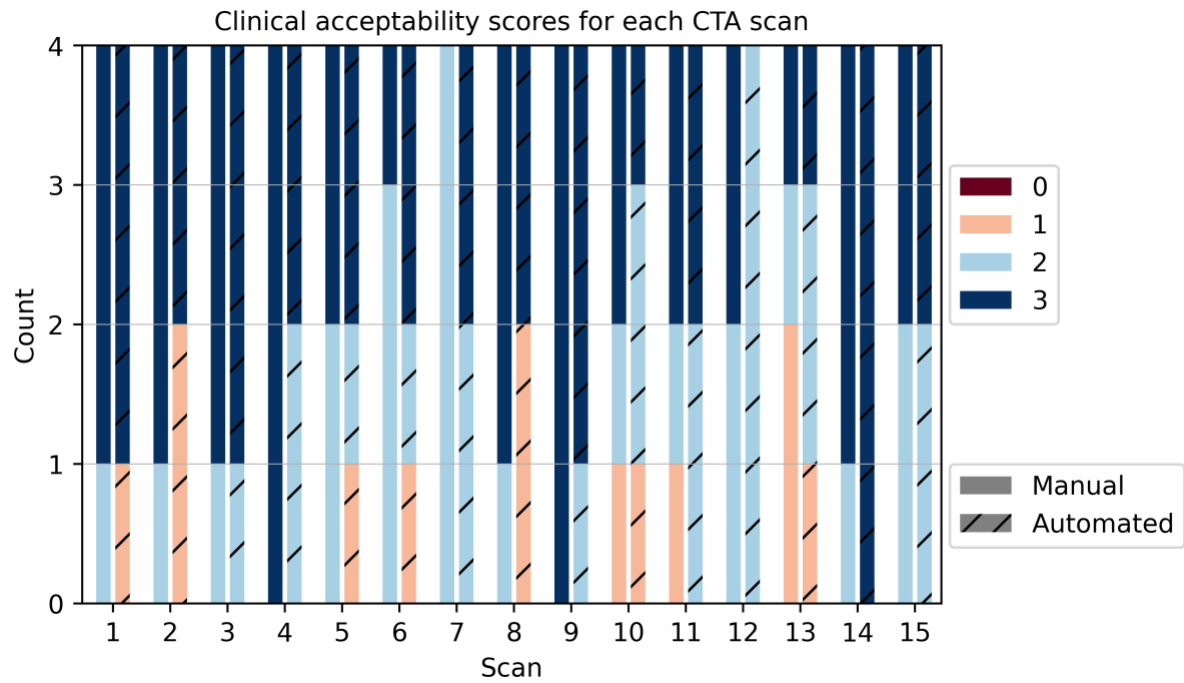
Figure C4. Distribution of clinical acceptability scores for manual and automated segmentations across the CTA scans. The scores were assigned by four surgeons to the segmentations of 15 CTA scans using a 4-point grading scale: 0 = not acceptable; 1 = Acceptable, but corrections necessary, perforator(s) missing suitable as ideal flap perforators; 2 = Acceptable, only minor corrections necessary. Perforator(s) missing, but not suitable or relevant for ideal flap perforator selection; 3 = Accepted, no corrections necessary for clinical use. None of the surgeons assigned a score of 0 to any segmentation.
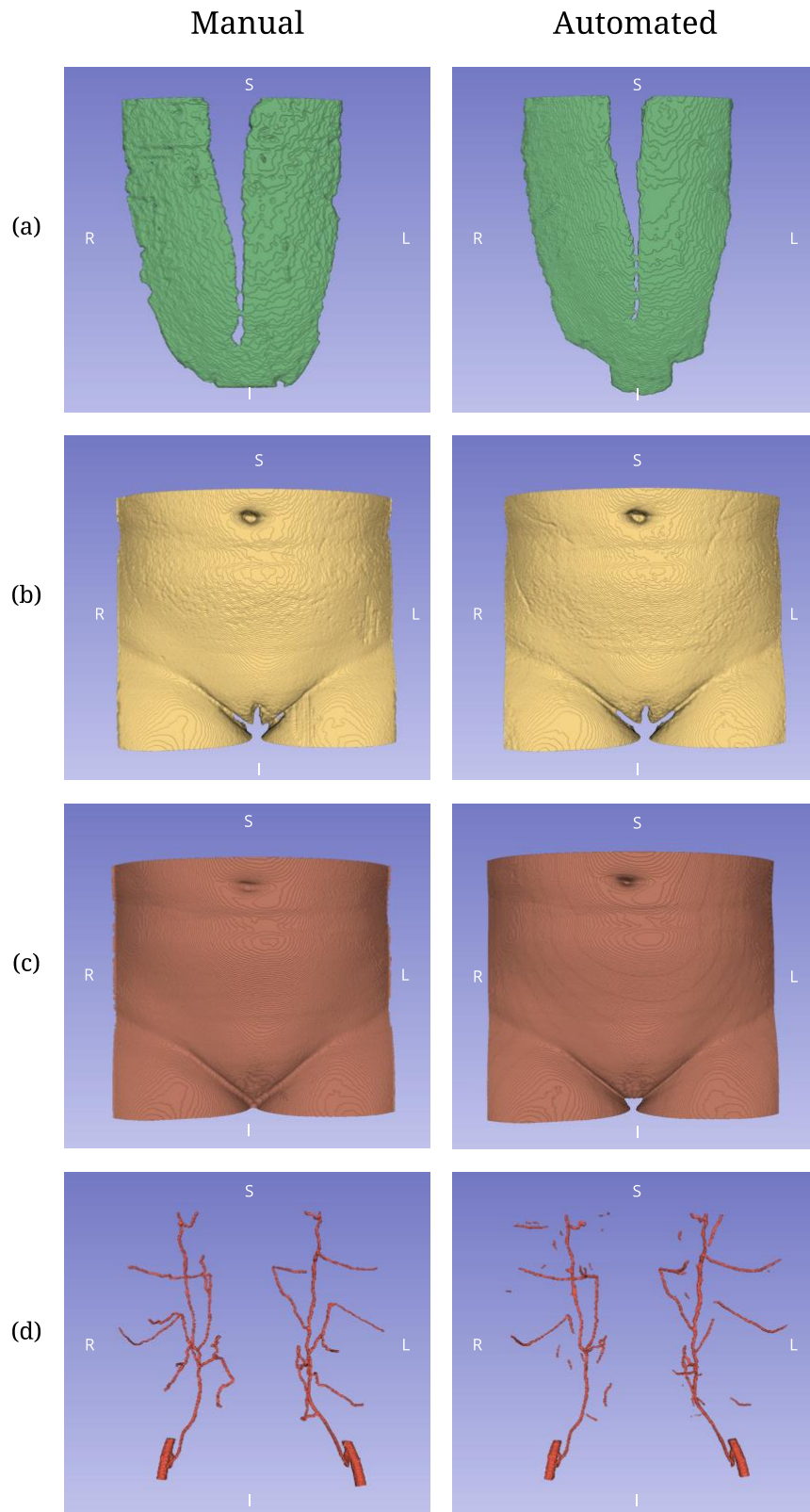
Figure C5. Visualization of the manual and automated segmentations of the (a) Rectus Abdominis muscle, (b), fat tissue, (c) skin tissue, and (d) deep inferior epigastric artery perforators for the CTA scan of one patient.