AMERICAN COLLEGE of SPORTS MEDICINE
LEADING THE WAY

. . . Published ahead of Print

DeepACSA: Automatic Segmentation of Cross-sectional Area in Ultrasound Images of Lower Limb Muscles Using Deep Learning

Paul Ritsche[1], Philipp Wirth[2], Neil J. Cronin[3], Fabio Sarto[4], Marco V. Narici[4], Oliver Faude[1], and Martino V. Franchi[4]

[1]Department of Sport, Exercise & Health, University of Basel, SWITZERLAND; [2]Lightly AG, Zurich, SWITZERLAND; [3]Neuromuscular Research Centre, Faculty of Sport and Health Sciences, University of Jyvaskyla, FINLAND; [4]Department of Biomedical Sciences, University of Padova, ITALY

Accepted for Publication: 14 July 2022

Medicine & Science in Sports & Exercise, Publish Ahead of Print
DOI: 10.1249/MSS.0000000000003010

DeepACSA: Automatic Segmentation of Cross-sectional Area in Ultrasound Images of Lower Limb Muscles Using Deep Learning

Paul Ritsche[1], Philipp Wirth[2], Neil J. Cronin[3], Fabio Sarto[4], Marco V. Narici[4], Oliver Faude[1], and Martino V. Franchi[4]

[1]Department of Sport, Exercise & Health, University of Basel, SWITZERLAND; [2]Lightly AG, Zurich, SWITZERLAND; [3]Neuromuscular Research Centre, Faculty of Sport and Health Sciences, University of Jyvaskyla, FINLAND; [4]Department of Biomedical Sciences, University of Padova, ITALY

Oliver Faude and Martino V. Franchi share last authorship

Address for Correspondence:

Martino Franchi, Department of Biomedical Sciences, University of Padova, Via Francesco Marzolo 5, 35131 Padova, Italy; Phone: +39 (0) 49 827 5309; E-mail: martino.franchi@unipd.it

Conflict of Interest and Funding Source:

This investigation was not supported by any funding source. All authors report no competing interests. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by the American College of Sports Medicine.

ABSTRACT

Purpose: Muscle anatomical cross-sectional area (ACSA) can be assessed using ultrasound and images are usually evaluated manually. Here, we present DeepACSA, a deep learning approach to automatically segment ACSA in panoramic ultrasound images of the human rectus femoris (RF), vastus lateralis (VL), gastrocnemius medialis (GM) and lateralis (GL) muscles. **Methods:** We trained three muscle-specific convolutional neural networks (CNNs) using 1772 ultrasound images from 153 participants (age=38.2 years, range: 13-78). Images were acquired in 10% increments from 30 to 70% of femur length for RF and VL and at 30 and 50% of muscle length for GM and GL. During training, CNN performance was evaluated using intersection-over-union scores. We compared the performance of DeepACSA to manual analysis and a semi-automated algorithm using an unseen test set. **Results:** Comparing DeepACSA analysis of the RF to manual analysis with erroneous predictions removed (3.3%) resulted in intra-class correlation (ICC) of 0.989 (95% CI 0.983;0.992), mean difference of 0.20 cm^2 (0.10;0.30) and standard error of the differences (SEM) of 0.33 cm^2 (0.26,0.41). For the VL, ICC was 0.97 (0.96,0.968), mean difference was 0.85 cm^2 (-0.4,1.31) and SEM was 0.92 cm^2 (0.73,1.09) following removal of erroneous predictions (7.7%). After removal of erroneous predictions (12.3%), GM/GL muscles demonstrated an ICC of 0.98 (0.96,0.99), a mean difference of 0.43 cm^2 (0.21,0.65) and a SEM of 0.41 cm^2 (0.29,0.51). Analysis duration was 4.0s standard deviation (SD) ± 0.43 for analysis of one image in our test set using DeepACSA. **Conclusions:** DeepACSA provides fast and objective segmentation of lower limb panoramic ultrasound images comparable to manual segmentation. Inaccurate model predictions occurred predominantly on low-quality images, highlighting the importance of high image quality for accurate prediction.

Key Words: IMAGE SEGMENTATION, MUSCLE, ULTRASONOGRAPHY, U-NET, DEEP NEURAL NETWORKS

INTRODUCTION

The anatomical cross-sectional area of a muscle (ACSA) represents a two-dimensional index of muscle size acquired in the transversal plane. (1) Muscle ACSA is important for clinical and scientific practice; it is related to the capacity of a muscle to generate force and, as a consequence, to locomotor performance. (2,3) Recent investigations have shown that muscle ACSA represents a useful parameter in the diagnosis and classification of several muscular disorders as well as the potential monitoring of disease progression. (4–7) Lower limb muscle ACSA is often assessed in the diagnosis of sarcopenia and is strongly associated with frailty. (4–6) Moreover, measuring lower limb muscle mass or ACSA is necessary to assess the extent of induced muscle loss in cachectic, dystrophic, and even intensive care unit patients. (7–9) Muscle mass is also a predictor of hospital stay duration in patients with moderate to severe COVID-19. (10) Thus, whole muscle ACSA represents a crucial variable when monitoring muscle decline in pathological settings, in response to disuse and ageing, and when investigating muscle adaptations to training in rehabilitation and return to sport scenarios.

Muscle ACSA can be assessed using several techniques such as Magnetic Resonance Imaging (MRI) and computer tomography, or with ultrasound imaging. Due to technical advances, the ability to perform measurements at the bedside, and drastically lower costs compared to other imaging modalities, the use of ultrasound has increased in many research and clinical settings. (4,11,12) Because of the shape and size of several lower limb muscles, conventional static B-mode ultrasound is often unsuitable to assess whole muscle ACSA. (1) The main reason for this is the limited field of view of most transducers. (1,11) Hence, panoramic ultrasound, which is used to obtain a panoramic image of the whole muscle, is employed to

circumvent the limitations imposed by the field of view of most commercially available transducers. (1) Brightness mode panoramic ultrasound inherits the benefit that it enables the user to image structures extending the field-of-view of the transducer. Thus no stitching of several non-panoramic images as well as manual extrapolation is required for complete ACSA assessment in ultrasonography images. (1,13) Yet, a limitation of panoramic ultrasonography is that the algorithms used for image calculation by the manufacturers are usually unknown. The technical principles are similar to those of non-panoramic ultrasonography, requiring the same methodological aspects to be considered. (1,13,14) However, panoramic ultrasonography is a more challenging technique, as it requires steady and consistent probe movement in the same imaging plane. (1,15) Nonetheless, panoramic ultrasound has previously shown good comparability to MRI and excellent inter-session and inter-rater repeatability. (16,17)

Once ultrasound is implemented as a routine part of ACSA monitoring, large volumes of image data could be collected, necessitating efficient and reliable analysis methods. To date, ultrasound images of muscle ACSA are mostly evaluated manually. However, manual analysis is subjective, laborious, and requires great experience. (18,19) Semi-automated and automated algorithms have been developed to accelerate the ACSA segmentation process and reduce the subjectivity of manual segmentation in lower limb muscle ultrasound images. (20,21) These approaches used sophisticated image processing steps to localize features and objects within an image. However, because of this, their generalisability is limited, and they present inconsistent results depending on image properties.

Making use of deep learning might be advantageous compared to non-trainable image processing alone, especially since convolutional neural networks (CNNs) have shown great promise identifying features in medical ultrasound and MRI scans. (12,19) U-net structured CNNs (22) have inter alia been successfully applied to brain MRI segmentation(23), carotid plaque segmentation in ultrasound images (24), as well as lower limb muscle segmentation in ultrasound and MRI images. (25–28) Previously, Chen et al. (27) and Marzola et al. (25,26) presented automatic approaches to segment the ACSA of upper and lower limb muscles, but only in conventional static brightness mode ultrasound images, and not panoramic images.

To our knowledge, the semi-automated ACSAuto program (20) is the only open-access tool available for analysing panoramic lower limb ultrasound images. Thus, it should be determined whether CNN approaches could be used to fully automate the process of ACSA analysis from panoramic ultrasound images. This could increase objectivity and reduce the time and effort needed for the analysis process. (18,25) Although manual ACSA segmentation of ultrasound images is the current gold standard, using automatic approaches might increase the evaluation quality.

In this study we present DeepACSA, a python package incorporating U-net based CNNs to automatically segment the ACSA of lower limb muscles in panoramic ultrasound images. DeepACSA is open source and includes a custom graphical user interface (GUI) to allow straightforward use and implementation. The muscles of interest are the m. rectus femoris (RF), m. vastus lateralis (VL), mm. gastrocnemius medialis (GM) and lateralis (GL). Here we describe

the DeepACSA package and compare it to manual image analysis and a non-trainable image processing script (ACSAuto) (20).

METHODS

Image acquisition and data

The data used in this study contains images from 143 participants of different age groups (adolescent males: n = 50, 16.4 years (13 to 17), adults: n = 83 (25 females, 58 males), 26.9 years (18 to 40), elderly males: n=10, 71.5 years (65 to 78)). We used 602 ultrasound images of the RF, 634 ultrasound images of the VL and 298 ultrasound images of the GM and GL (9 to 36 images per participant) acquired in previous and ongoing studies which received ethical approval from the responsible committees (Ethics Committee of North-Western and Central Switzerland; Local ethics Committee of the Department of Biomedical Sciences, University of Padova) and complied with the Declaration of Helsinki. Participants signed an informed written consent prior to the start of the study after receiving all relevant study information. Images were anonymized and randomly divided into a training set for the models as well as an external test set, making sure that no image of the same muscle region of one participant appeared in both sets. To increase image variability, we used brightness mode panoramic ultrasonography from three devices (ACUSON Juniper, linear-array 54 mm probe, 12L3, Acuson 12L3, SIEMENS Healthineers, Erlangen, Germany; Aixplorer Ultimate, linear-array 38mm probe, Superline SL10-2, SuperSonic Imagine, Aix-en-Provence, France; Mylab 70, linear array 47mm probe, Esaote Biomedica, Genova, Italy) collected by three different experienced operators to assess the ACSA of RF, VL, GM and GL. We acquired scans at rest and a guide was mounted to the leg to control the transversal path. Because of regional differences in muscle size and shape, we

acquired images of the RF and VL at 10% increments from 30 to 70% of the distance between the lateral femur condyle and the trochanter major. Participants laid in a supine position with their legs extended and feet on the bed. We acquired images of the GM and GL at 30 and 50% of muscle length. Participants laid in a prone position with their legs extended and feet on the bed. Scans of the RF and VL muscles were either taken separately or cropped from whole quadriceps scans to increase image variability. We included images of all acquired regions in our training and validation sets. Ultrasound device settings differed between devices and study protocols, and image depth, brightness and contrast settings were always chosen to ensure best visibility of the muscle. Images included in the training and validation sets were manually labelled by an experienced investigator (PR) prior to model training (Fig. 1). Manual segmentation consisted of digitizing the ACSA of each muscle using the polygon tool in FIJI (29). We randomly included 30 images from the proximal (70% of femur length), mid (50% of femur length) and distal (30% of femur length) regions in the test set for the RF and VL. For the GM and GL, 37 images from the mid (50% of muscle length) and 28 images from the proximal (30% of muscle length) region were randomly included in the test set. Therefore, the test set consisted of 90 images for the RF and VL and 65 images for the GM and GL. The ratio of images in the test and training sets for the RF was 0.18 (90:512), for the VL 0.17 (90:544), and for the GM and GL 0.27 (65:233). No cropped images were included in the test set because we only had these images available from one device. We measured ACSA in all images of the test set using the DeepACSA and ACSAuto programs. Because GL segmentation is not supported in ACSAuto, we only analyzed the GM test images with ACSAuto.

Network architecture and training

We performed sematic segmentation with two classes (ACSA and background) on pre-processed ultrasound images. The image-label pairs (Fig. 1) served as input for the models. We compared the performance of the original U-net architecture to models using a pretrained VGG16 encoder path and a U-net decoder path. The U-net models we employed consisted of a contracting and an expanding path. (18,22) The exact architecture of the U-net model can be found elsewhere. (18,22) The encoder (left) path uses convolution and max pooling layers to generate a 1024-dimensional abstract representation of the input image. Subsequently, the decoder (right) path uses deconvolutional layers to generate a pixelwise class prediction from the abstract image representation of the input. Features between layers of the same resolution are shared so that information about small-scale objects is not lost (hence the U-shape). The model outputs a pixelwise binary label, thus, every pixel of an image is predicted to belong to one of two possible classes. Shortly, and incontrast to the U-Net encoder, the VGG16 encoder uses a series of convolutional and may pooling layers to generate a 512-dimensional abstract representation of the input image (Fig. 2). (30) Detailed description of the VGG16 encoder path architecture can be found elsewhere. (30) We adopted U-net and VGG16 models due to their demonstrated success for image segmentation in several fields (18,25–28,30) and their ability to work well with small datasets containing a few hundred images (22). Prior to model training, images and masks in the training sets of all muscles were augmented using height and width shift, rotation and horizontal flipping (see shared code for details). We imported, normalized to a scale between 0 and 1, and resized the images to 256 x 256 pixels for training. Training the CNNs with 512 x 512-pixel sized images yielded inferior intersection-over-union (IoU) and accuracy scores and was computationally more expensive. By convention, we applied a random 90/10 %

training/validation data split within the training data set. (18) We trained three separate models, one for RF, one for VL and one for GM and GL combined with the same U-net model architecture.(18,22) A RTX3060 GPU (NVIDIA, Santa Clara, USA) was used for model training with a defined maximum of 50 epochs and a batch size of one. We used the Adam optimizer for stochastic gradient descent to update network weights with an initial learning rate of 10^{-5}. (18,22) The learning rate was reduced by a factor of 0.1 following ten epochs of loss stagnation. We used binary cross-entropy as a loss function because the segmentation task only represents two classes. To reduce the risk of overfitting, we implemented early stopping when the training error decreased, and the test error reached a plateau or increased. During training, model performance was evaluated using the IoU measure to test the overlap between manually segmented muscle area and labels predicted by the respective model (Table 1). IoU is defined as

$$IoU(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \qquad [1]$$

where $0 \leq \text{IoU}(A, B) \leq 1$, A describes the ground truth and B the model predictions. The IoU is maximal if the ground truth A and model predictions B are the same and minimal if A and B have no overlap.

The output of each CNN is a binary mask image that contains the predicted muscle area. The output is post-processed so that holes within the predicted muscle are filled and, in case of multiple predicted structures, only the largest structure is kept. The DeepACSA code is written in Python using a Keras API to interface with Tensorflow for model training. The code and trained models from this project are openly available at https://github.com/PaulRitsche/DeepACSA.git.

For more information about program use we refer the reader to an instructional online video (Link: https://youtu.be/It9CqVSNc9M).

Analysis metrics

All statistical analyses were performed in R software (31) (Base, BlandAltmanLeh, irr, readxl and rstudioapi packages). The analysis script is shared as supplemental digital content (see Supplemental Digital Content 1, http://links.lww.com/MSS/C684). We compared DeepACSA and ACSAuto measurements to manual measurements. For this purpose, we calculated consecutive-pairwise intra-class correlations (ICC) and standard error of the differences between methods (SEM) with 95% compatibility intervals (CI). We used Bland–Altman analysis (32) to test the agreement between two analysis methods and set the limits of agreement to ± 1.96 standard deviations. We computed the standardized mean bias according to Hopkins (33), with 0.1, 0.3, 0.6, 1.0 and 2.0 representing small, moderate, large, very large and extremely large errors respectively.

RESULTS

The training of the deep neural networks required 23 to 38 epochs, and IoU and loss metrics are shown in Table 1. Performance of the model employing a pretrained VGG16 encoder path and a U-net decoder path outperformed the original U-net architecture slightly, based on IoU and accuracy measures. Thus, we selected this model architecture for comparison to manual analysis. An example of an input image and model prediction following post-processing is shown in Figure 2. ICCs, SEMs, mean bias, standardized mean bias with 95% CI as well as percentage values for all muscles across all regions comparing DeepACSA and ACSAuto to manual

measurements are shown in Table 2 and Fig. 3. We additionally report comparability statistics after removing incorrect predictions (RF: n = 3 (3.3%), VL: n = 7 (7.7%), GM & GL: n = 8 (12.3%)) by the models (Table 2). Predictions were removed following visual inspections by an experienced investigator (PR) based on obvious deviations from the expected ACSA of the respective muscle (Fig. 4). All images wrongly segmented by the models can be found in the supplemental digital content (see Supplemental Digital Content 2, http://links.lww.com/MSS/C685). Comparing DeepACSA analysis of all muscles including the wrong predictions to manual analysis resulted in ICCs between 0.91 and 0.99, mean differences of 0.24 to 0.85 cm^2 and SEMs ranging from 0.46 to 1.53 cm^2 (Table 2). Calculated standardized mean biases were small for all muscles ranging from 0.05 to 0.14 (Table 2). The comparison of ACSAuto with manual analysis of all images and muscles resulted in ICCs between 0.97 and 0.99, mean differences of -0.13 to 0.24 cm^2 and SEMs ranging from 0.23 to 0.66 cm^2. Standardized mean biases were small for all muscles ranging from -0.02 to 0.05 (Table 2). DeepACSA segmentation resulted in differences between muscle regions for the VL. At 30% of femur length, the ICC was 0.75 (95% CI: 0.53;0.87), the mean difference was 1.66 cm^2 (0.47;2.85) and the SEM was 2.24 cm^2 (1.05;3.41). Standardized mean bias was moderate (0.34). In contrast, at 50% of femur length, the ICC was 0.95 (0.89;0.97), the mean difference was 0.72 cm^2 (0.16;1.27) and the SEM was 1.05 cm^2 (0.73;1.27). Standardized mean bias was small (0.14). At 70% of femur length, the ICC was 0.96 (0.93;0.98), the mean difference was 0.18 cm^2 (-0.15;0.51) and the SEM was 0.63 cm^2 (0.29;0.96). Similar to the 50% of femur length site, the observed standardized mean bias was small (0.05). Regional data for RF, GM & GL can be found in the supplemental digital content (see Tables S1-S3, Supplemental Digital Content 3, Comparison of manual analysis and ACSAuto and DeepACSA at 30%, 50%, and 70% femur

length, http://links.lww.com/MSS/C686). Removing incorrect model predictions resulted in increased ICCs and decreased mean differences, SEMs, and standardized mean bias for all muscles (Table 2). In Bland-Altman analysis, concerns regarding heteroscedasticity were only raised when comparing manual to DeepACSA RF analysis. With increasing muscle size, it seems that the variance around the mean difference increases.

Additionally, we compared the efficiency of DeepACSA and ACSAuto in ten images. Analysis duration was 4.0s standard deviation (SD) ± 0.43 for scaling, segmenting and saving the analysis results of one image in our test set using DeepACSA. Compared to that, analysis duration for one image using ACSAuto (whole workflow) was 40.3s ± SD 6.9.

DISCUSSION

Here we present DeepACSA, a deep learning approach for automatic analysis of muscle ACSA in panoramic brightness mode ultrasound images. The predictions of our trained CNNs were comparable to manual segmentation for all the lower limb muscles investigated. To our knowledge, DeepACSA represents the first openly accessible program able to automatically segment the muscle area of rectus femoris, vastus lateralis, gastrocnemius medialis, and gastrocnemius lateralis in panoramic ultrasound images.

Comparing the DeepACSA-based analysis of all muscles, the predictions of GM & GL ACSA were found to have the highest agreement with manual analysis. The worst agreement was observed for the RF. Yet, only the VL displayed regional differences. Most incorrect VL predictions occurred at 30% of femur length. Here the aponeuroses of the VL are more difficult

to scan and in some cases cannot be distinguished from surrounding tissue. Thus, the image quality decreases. Lower pixel contrast and image quality make prediction more difficult and might lead to deletion or selection of the wrong area during post-processing.

All our CNNs demonstrated minimal loss and high IoU scores. Although high IoU and minimal loss values in the training set might indicate overfitting, similar values were observed for the test set. Thus, our models can generalize to unseen data. However, we used ultrasound images of all three devices in our training, validation and test sets leading to similar image characteristics in all data sets. To test whether our CNNs can generalize to unseen image data from other devices, images from a fourth device should solely be added to the test set. Based on our results, visual inspection of the model output subsequent to automatic muscle ACSA segmentation is still necessary, and the removal of erroneous predictions resulted in increased comparability for all muscles. However, we believe that adding more training data might help to eliminate these model failures. Erroneous predictions in the test set mostly occurred when image quality was low, i.e., when tissue contrast was low or pixel values were homogeneous. Therefore, image quality is still an important factor for correct muscle ACSA segmentation using deep learning. Images for which muscle ACSA was predicted incorrectly should be analyzed using ASCAuto (20) or manually.

Given that DeepACSA is comparable to manual evaluation, the program might be implementable in a clinical setting as well. In contrast to MRI or computer tomography, ultrasound can be used quickly at the bedside of patients and practitioners do not necessarily need active cooperation from patients. (4) This might be especially important for patients in the

intensive care unit or with severe muscular disorders. (9,34) DeepACSA analysis is comparable to manual evaluation, more time efficient, and automates the segmentation process. This is important for clinical practice as it decreases the effort needed to analyze ultrasound image data. In fact, reducing the subjectivity of the analysis might increase the overall validity of clinical results. Moreover, our trained CNNs could be used to correctly predict the ACSA of other muscles, in patients with muscular diseases or an ageing population (for example when screening for the presence of sarcopenia). However, because of differences in image properties, it might be necessary to retrain our models including labeled images from the respective muscles or clinical populations in the training dataset. In the future, we aim to release more models that allow data from additional muscles and clinical populations to be analysed using the DeepACSA package.

In contrast to the algorithms presented by Chen et al. (27) and Marzola et al. (25,26), who used a supervised learning approach with U-net structured CNNs and achieved good precision and recall rates, DeepACSA can segment muscles in panoramic ultrasound images. To our knowledge, DeepACSA is the only program that includes images of several lower limb muscles from various muscle regions, different operators, and several ultrasound devices. Whereas Marzola et al. (25) included images from the tibialis anterior, GM and biceps brachii at rest, Chen et al. (27) included images of the RF during contraction. Acquiring images during contraction could limit the generalizability of the model to images acquired at rest. Furthermore, while the algorithms proposed by Marzola et al. (25,26) seem to distinguish between clinical and healthy populations based on echogenicity values, segmenting ACSA in muscles that exceed the field of view of the ultrasound transducer may have limited meaningfulness.

It is important to note that the predictions of our trained CNNs for all muscles showed lower agreement with manual analysis when compared to ACSAuto (DeepACSA: 95% limits of agreement 5.1 to -2.1 cm^2, ACSAuto: 1.7 to -1.9 cm^2). This is due to the manual correction of the suggested outline during ACSAuto analysis. Yet, the feature detection filters used for image segmentation in ACSAuto are highly dependent on image properties. Furthermore, ACSAuto is somewhat subjective because the user must validate the proposed muscle outline. In addition, not correcting the suggested outline led to large measurement errors compared to manual analysis and thus unusable results. (20) By employing trainable CNNs, DeepACSA is more robust to variation in ultrasound image pixel characteristics because the whole image texture is considered, and more complex features are computed. Additionally, DeepACSA reduces user bias in the analysis process, as no user input is required during image segmentation.

This investigation has some limitations. First, we performed no cross-validation or hyper parameter tuning for our models. (19) Additionally, comparing more different deep neural networks with various architectures could be beneficial to determine the ideal model for this task. However, the U-net architecture we employed demonstrated good results when segmenting muscle aponeuroses and fascicles in sagittal plane ultrasound images of the VL and GM. (18) Although our training set consisted of data from different age groups, we only included images from adolescent and young healthy participants and three devices in our test sets because of limited available data from elderly people. Thus, we cannot yet generalize our results to older people, clinical populations or other devices. (25,26) Finally, DeepACSA is currently able to automatically evaluate single muscle images but not videos.

CONCLUSIONS

DeepACSA segmentation of panoramic ultrasound images from rectus femoris, vastus lateralis, gastrocnemius medialis and lateralis muscles yielded comparable results to manual segmentation. Therefore, our trained convolutional neural networks can automatically segment lower limb muscles in panoramic ultrasound images. DeepACSA objectifies and accelerates the evaluation process of panoramic ACSA ultrasound images, allowing large datasets to be evaluated quickly, and representing a valuable tool that can be implemented in clinical settings. However, our results demonstrated that visual inspection of the output subsequent to automatic muscle ACSA segmentation may still be necessary to optimize predictions and avoid misclassifications. In the future, the segmentation performance of different model architectures should be compared whilst using a larger, more variable training dataset.

Declaration of Interest

This investigation was not supported by any funding source. All authors report no competing interests. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by the American College of Sports Medicine.

Data Sharing Statement

All data used in this study is publicly available. The DeepACSA package is available on Github at https://github.com/PaulRitsche/DeepACSA.git. Our trained models are available at https://doi.org/10.5281/zenodo.5799068. Anonymized panoramic ultrasound images are available at https://doi.org/10.5281/zenodo.5799204. The instructional video can be accessed using this link https://youtu.be/It9CqVSNc9M.

Role of Funding Source

This investigation was not supported by any funding source.

REFERENCES

1. Franchi MV, Raiteri BJ, Longo S, Sinha S, Narici MV, Csapo R. Muscle architecture assessment: strengths, shortcomings and new frontiers of in vivo imaging techniques. *Ultrasound Med Biol.* 2018;44(12):2492–504.
2. Evangelidis PE, Massey GJ, Pain MTG, Folland JP. Strength and size relationships of the quadriceps and hamstrings with special reference to reciprocal muscle balance. *Eur J Appl Physiol.* 2016;116(3):593–600.
3. Ritsche P, Bernhard T, Roth R, et al. M. Biceps femoris long head architecture and sprint ability in youth soccer players. *Int J Sports Physiol Perform.* 2021;16(11):1616-24.
4. Ticinesi A, Narici MV, Lauretani F, et al. Assessing sarcopenia with vastus lateralis muscle ultrasound: an operative protocol. *Aging Clin Exp Res.* 2018;30(12):1437–43.
5. Chianca V, Albano D, Messina C, et al. Sarcopenia: imaging assessment and clinical application. *Abdom Radiol.* 2021. doi: 10.1007/s00261-021-03294-3.
6. Mueller N, Murthy S, Tainter CR, et al. Can Sarcopenia quantified by ultrasound of the rectus femoris muscle predict adverse outcome of surgical intensive care unit patients as well as frailty? A prospective, observational cohort study. *Ann Surg.* 2016;264(6):1116–24.
7. Anderson LJ, Chong N, Migula D, et al. Muscle mass, not radiodensity, predicts physical function in cancer patients with or without cachexia. *Oncotarget*. 2020;11(20):1911–21.
8. Jacques MF, Onambele-Pearson GL, Reeves ND, Stebbings GK, Smith J, Morse CI. Relationships between muscle size, strength, and physical activity in adults with muscular dystrophy: muscle strength, size and physical activity in muscular dystrophy. *J Cachexia Sarcopenia Muscle.* 2018;9(6):1042–52.

9. Puthucheary ZA, Rawal J, McPhail M, et al. Acute Skeletal muscle wasting in critical illness. *JAMA*. 2013;310(15):1591-600.

10. Gil S, Jacob Filho W, Shinjo SK, et al. Muscle strength and muscle mass as predictors of hospital length of stay in patients with moderate to severe COVID-19: a prospective observational study. *J Cachexia Sarcopenia Muscle.* 2021;12(6):1871-8.

11. Sarto F, Spörri J, Fitze DP, Quinlan JI, Narici MV, Franchi MV. Implementing ultrasound imaging for the assessment of muscle and tendon properties in elite sports: practical aspects, methodological considerations and future directions. *Sports Med.* 2021;51(6):1151–70.

12. Shin Y, Yang J, Lee YH, Kim S. Artificial intelligence in musculoskeletal ultrasound imaging. *Ultrasonography*. 2021;40(1):30–44.

13. Weng L, Tirumalai AP, Lowery CM, et al. US extended-field-of-view imaging technology. *Radiology*. 1997;203(3):877–80.

14. Aldrich JE. Basic physics of ultrasound imaging: *Crit Care Med.* 2007;35(5 Suppl):S131–7.

15. Franchi MV, Fitze DP, Raiteri BJ, Hahn D, Spörri J. Ultrasound-derived biceps femoris long head fascicle length: extrapolation pitfalls. *Med Sci Sports Exerc.* 2020;52(1):233–43.

16. Noorkoiv M, Nosaka K, Blazevich AJ. Assessment of quadriceps muscle cross-sectional area by ultrasound extended-field-of-view imaging. *Eur J Appl Physiol.* 2010;109(4):631–9.

17. Bolsterlee B, Gandevia SC, Herbert RD. Ultrasound imaging of the human medial gastrocnemius muscle: how to orient the transducer so that muscle fascicles lie in the image plane. *J Biomech.* 2016;49(7):1002–8.

18. Cronin NJ, Finni T, Seynnes O. Fully automated analysis of muscle architecture from B-mode ultrasound images with deep learning. *ArXiv*. 2020;arXiv:2009.04790.

19. Greenspan H, van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging.* 2016;35(5):1153–9.

20. Ritsche P, Wirth P, Franchi MV, Faude O. ACSAuto-semi-automatic assessment of human vastus lateralis and rectus femoris cross-sectional area in ultrasound images. *Sci Rep.* 2021;11(1):13042.

21. Salvi M, Caresio C, Meiburger KM, De Santi B, Molinari F, Minetto MA. Transverse muscle ultrasound analysis (TRAMA): robust and accurate segmentation of muscle cross-sectional area. *Ultrasound Med Biol.* 2019;45(3):672–83.

22. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. ArXiv. 2015;arXiv:1505.04597

23. Lee B, Yamanakkanavar N, Choi JY. Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture. *PLoS One.* 2020;15(8):e0236493.

24. Meshram NH, Mitchell CC, Wilbrand S, Dempsey RJ, Varghese T. Deep learning for carotid plaque segmentation using a dilated U-Net architecture. *Ultrason Imaging.* 2020;42(4–5):221–30.

25. Marzola F, van Alfen N, Doorduin J, Meiburger KM. Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment. *Comput Biol Med.* 2021;135:104623.

26. Marzola F, Alfen N van, Salvi M, Santi BD, Doorduin J, Meiburger KM. Automatic segmentation of ultrasound images of gastrocnemius medialis with different echogenicity

levels using convolutional neural networks. *Annu Int Conf IEEE Eng Med Biol Soc.* 2020;2020:2113-6.

27. Chen X, Xie C, Chen Z, Li Q. Automatic tracking of muscle cross-sectional area using convolutional neural networks with ultrasound. *J Ultrasound Med.* 2019;38(11):2901–8.

28. Rohm M, Markmann M, Forsting J, Rehmann R, Froeling M, Schlaffke L. 3D Automated segmentation of lower leg muscles using machine learning on a heterogeneous dataset. *Diagnostics (Basel)*. 2021;11(10):1747.

29. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9(7):676–82.

30. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv. 2015;arXiv:1409.1556

31. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. 2019. Available from: https://www.R-project.org/

32. Bland MJ, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;(20):307–10.

33. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc.* 2009;41(1):3–13.

34. Zaidman CM, van Alfen N. Ultrasound in the assessment of myopathic disorders. *J Clin Neurophysiol.* 2016;33(2):103–11.

FIGURE LEGENDS

Figure 1. Manual segmentation for convolutional neural network training. a) Original extended-field-of-view ultrasound image of m. vastus lateralis, b) manually segmented binary mask of muscle in a).

Figure 2. DeepACSA workflow. a) Original ultrasound image of the m. rectus femoris (RF) at 50% of femur length that serves as input for the model. b) Detailed U-net CNN architecture (modified from Ronneberger et al. (19) and Cronin et al. (15),). Multi-channel feature maps are represented by the blue boxes with number of channels displayed on top of the respective box. Copied feature maps from the convolutional (left) side that are concatenated with the ones from the expanding (right) side are represented by the white boxes. The different operations are marked by the arrows. c) Model prediction of muscle area following post-processing (shown as a binary image).

Figure 3. Bland-Altman plots of all muscles plotting the difference between manual and DeepACSA with incorrect predictions removed (rm), manual and DeepACSA as well as manual and ACSAuto area segmentation measurements against the mean of both measures. Dotted and solid lines illustrate 95% limits of agreement and bias. M. rectus femoris (RF) and m. vastus lateralis (VL), mm. gastrocnemius medialis (GM), and lateralis (GL).

Figure 4. Examples of incorrect predictions and obvious deviations from the expected ACSA of the respective muscles. a) m. vastus lateralis, b) m. rectus femoris, c) m. gastrocnemius medialis, and d) m. gastrocnemius lateralis.

SUPPLEMENTAL DIGITAL CONTENT

SDC 1: Comparability_analysis_DeepACSA.txt

Analysis script

SDC 2: Supplemental_Digital_File2.pdf

SDC 3: Supplemental Digital File 3.pdf

Table S1 - Comparison of manual analysis and ACSAuto and DeepACSA at 30% femur length

Table S2 - Comparison of manual analysis and ACSAuto and DeepACSA at 50% femur length

Table S3 - Comparison of manual analysis and ACSAuto and DeepACSA at 70% femur length

Figure 1

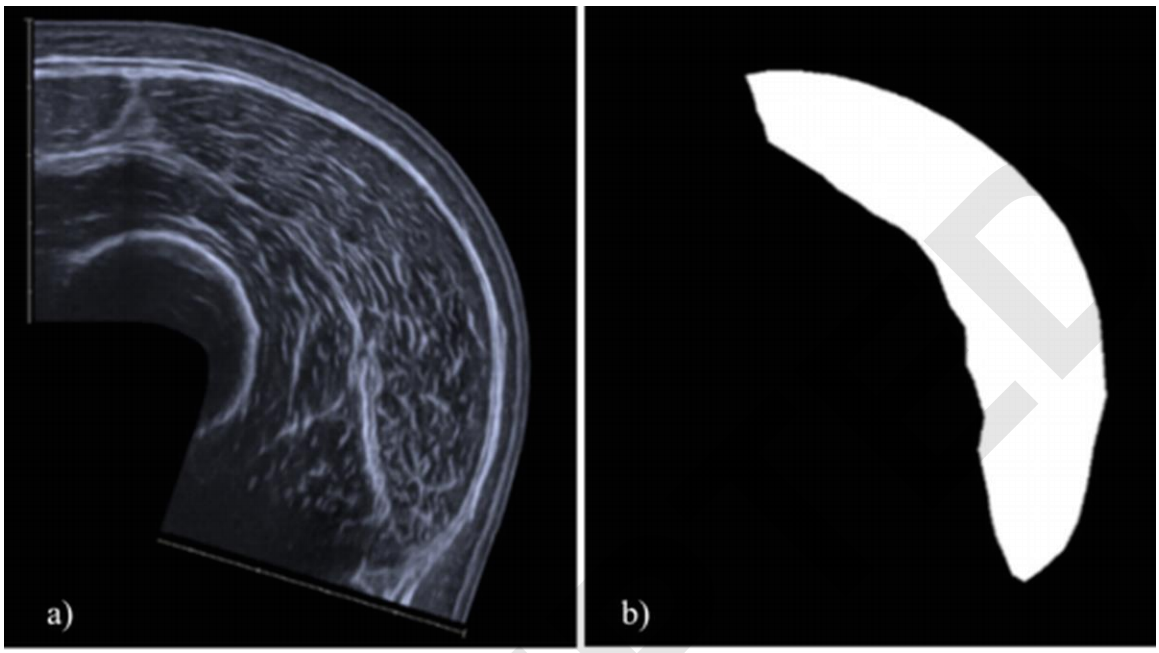


Figure 2

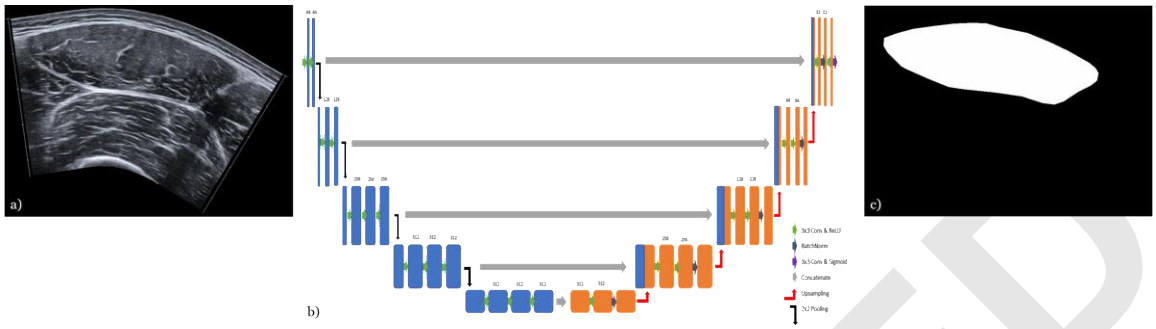


Figure 3

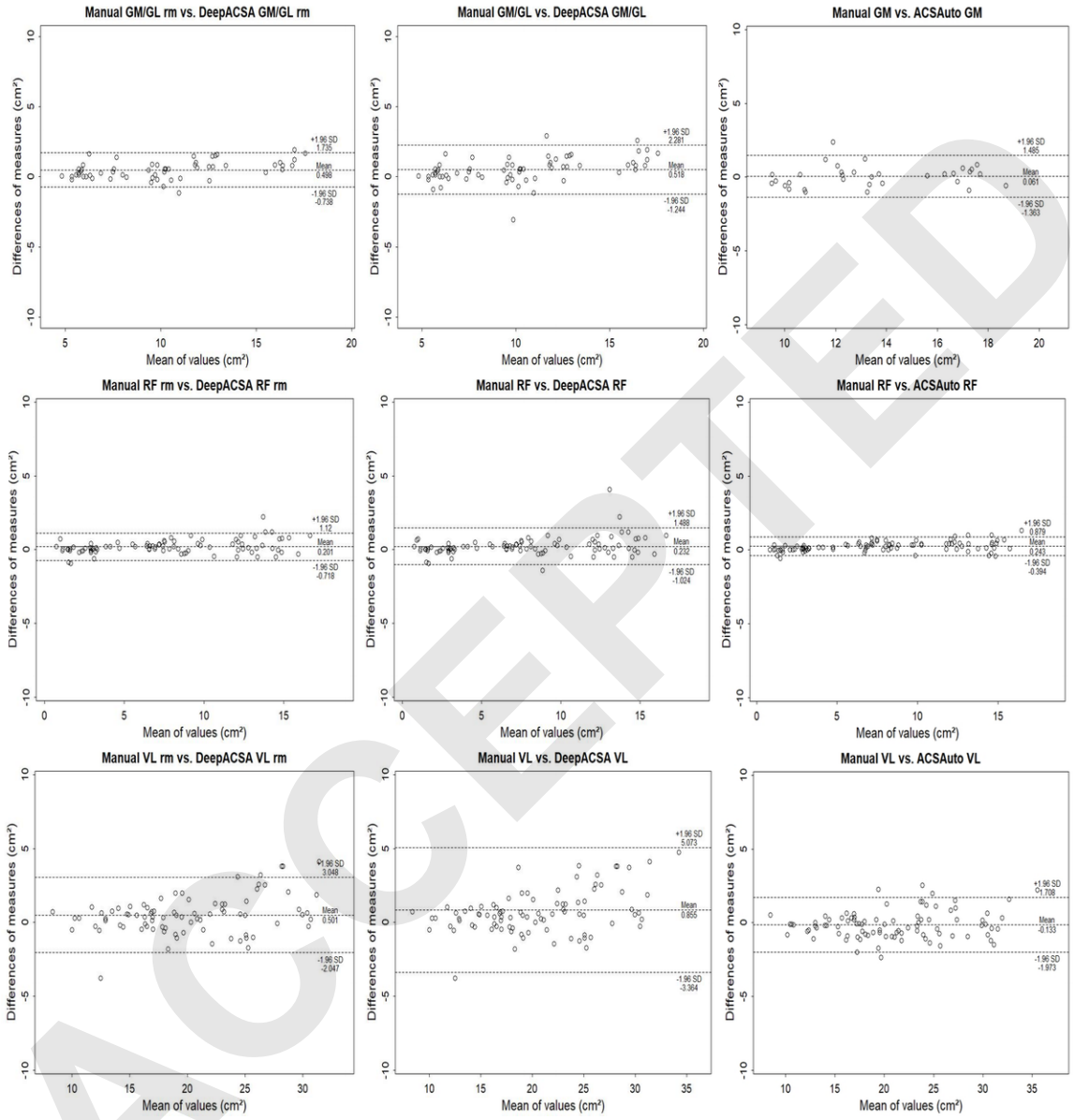


Figure 4

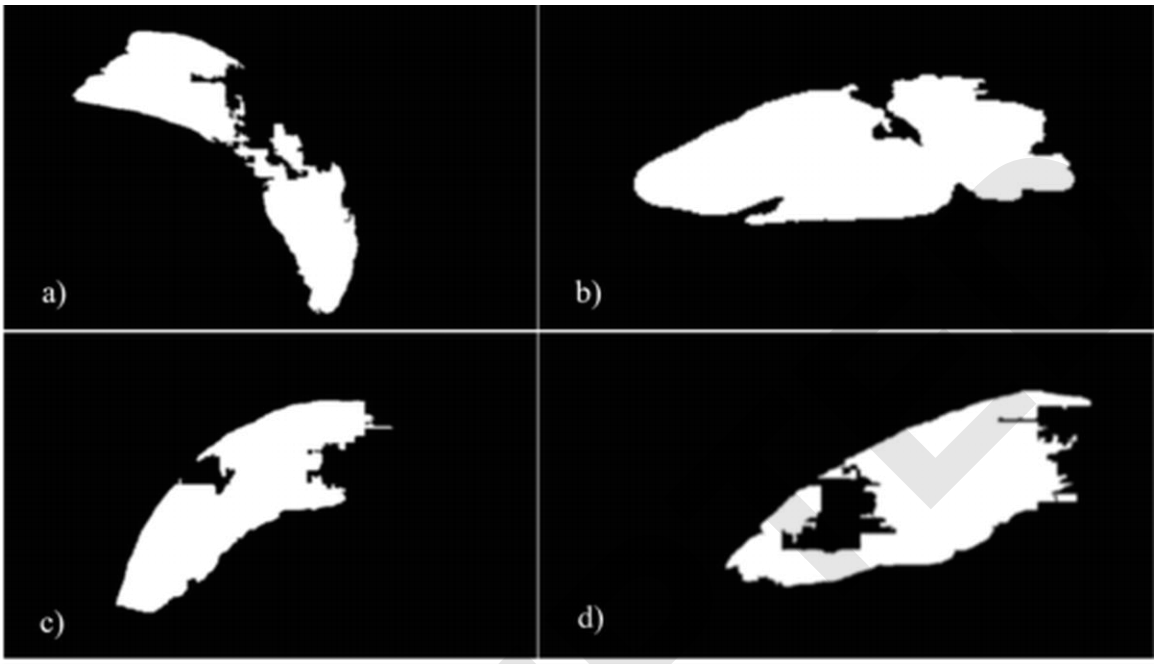


Table 1: Final intersection over union and loss functions for the two different model architectures, original U-net and pretrained VGG16 encoder U-net.

Metric	RF		VL		GM & GL	
	Train	Valid	Train	Valid	Train	Valid
Original U-net						
IoU	0.997	0.996	0.994	0.992	0.997	0.996
Accuracy	0.986	0.984	0.975	0.971	0.975	0.973
Pretrained VGG16 encoder U-net						
IoU	0.998	0.997	0.997	0.996	0.997	0.996
Accuracy	0.996	0.995	0.995	0.992	0.996	0.994

Intersection over union (IoU) was calculated for the model prediction and the manually segmented ground truth in the training and the randomly generated test set using the training/validation split. M. rectus femoris (RF), m. vastus lateralis (VL), mm. gastrocnemius medialis (GM), and lateralis (GL).

Table 2: Comparison of manual analysis and ACSAuto and DeepACSA

Mode	ICC	MD cm²	SEM cm²	SEM %	Stmd
RFM vs. Auto	0.996 (0.994;0.998)	0.24 (0.17;0.31)	0.23 (0.19;0.27)	5.10	0.05
RFM vs. Deep	0.989 (0.983;0.992)	0.23 (0.10;0.37)	0.45 (0.29;0.63)	11.1	0.05
RFM vs. DeepRm	0.989 (0.983;0.992)	0.20 (0.10;0.30)	0.33 (0.26;0.41)	9.49	0.04
VLM vs. Auto	0.988 (0.982;0.992)	-0.13 (-0.33;0.06)	0.66 (0.55;0.77)	3.13	-0.02
VLM vs. Deep	0.926 (0.889;0.95)	0.85 (-0.4;1.31)	1.52 (0.88;2.20)	8.66	0.13
VLM vs. DeepRm	0.972 (0.957;0.982)	0.50 (0.21;0.79)	0.92 (0.73;1.09)	4.48	0.08
GM&GLM vs. Auto	0.97 (0.94;0.985)	0.06 (-0.2;0.32)	0.51 (0.35;0.67)	4.19	0.02
GM&GLM vs. Deep	0.960 (0.925;0.979)	0.53 (0.25;0.82)	0.59 (0.42;0.75)	5.69	0.14
GM&GLM vs. DeepRM	0.978 (0.955;0.989)	0.43 (0.21;0.65)	0.41 (0.29;0.51)	3.89	0.12

RF = rectus femoris, M = Manual analysis, Auto = ACSAuto analysis, Deep = DeepACSA analysis, DeepRM = DeepACSA analysis with wrong predictions removed, VL = vastus lateralis, GM & GL = gastrocnemii, rm = wrong prediction removed, ICC = Intra-class correlation (95% compatibility interval), MD = mean difference (95% compatibility interval), SEM = Standard error of the differences (95% compatibility interval), and Stmd = standardized mean difference.

```
#'Title: Comparability_analysis_DeepACSA
#'Project: DeepACSA
#'Author: Paul Ritsche
#'Last edited: 22.10.2021

rm(list = ls())
library(rstudioapi)
library(BlandAltmanLeh)
library(readxl)
library(irr)

##########################
#####Define Functions#####
##########################

get_file_list <- function(root){
  #'  Retrieves all *.xlsx files in root directory and subfolders.
  #'  Files should contain one measurement parameter each.
  #'  Args: root = Path to directory containgy .xlsx files (String)
  #'  Returns: Vector containing all  files in root and subfolders.

  file_list <- list.files(root, pattern ='*.xlsx')
  df_list <- lapply(file_list, read_excel)
  return(df_list)

}

calc_comp <- function(root, comp_index, filename){
  #'  Calculates several comparability values including ICC, MD, SEM and StMD
  #'  for comparability analysis.
  #'  Args: root = Path to directory containgy .xlsx files (String)
  #'        comp_index = Column to be compared to (Integer)
  #'        filename = Name of file (String)
  #'  Returns: CSV-file with above mentioned values.

  # Get files in root
```

```
df_list <- get_file_list(root)
# Loop trough files
for (data_frame in df_list){

  # Convert tibble to data.frame
  data_frame <- as.data.frame(data_frame)
  data_frame <- na.exclude(data_frame)
  View(data_frame)

  # Loop trough columns of data_frame
  for(i in 2:ncol(data_frame)) {

    # Get variable names
    var_names <- colnames(data_frame)
    var_names <-  paste(var_names[comp_index], var_names[i], sep = " vs. ")

    #Scatterplot of a vs. b
    plot(data_frame[,comp_index], data_frame[,i],
        xlab = colnames(data_frame[comp_index]),
        ylab = colnames(data_frame[i]))

    #ICC calculation
    icc_raw <- icc(cbind(data_frame[,comp_index],data_frame[,i]),
                model="oneway",
                type="agreement",
                unit="single") #Model=only subjects randomly chosen, type=differences in judges rating, unit =
average of several measurements
    ICC <- icc_raw$value
    ICC_L <- icc_raw$lbound
    ICC_U <- icc_raw$ubound

    #Mean difference
    ttest <- t.test(data_frame[,comp_index], data_frame[,i], paired = TRUE)
    MD <- ttest$estimate
    MD_CI <- ttest$conf.int
    diff_log <- (log(data_frame[,comp_index]) - log(data_frame[,i]))
```

```
MDP <- ((100*(exp(mean(diff_log) / 100))) - 100) * 100

#SEM of difference calculation
diff <- (data_frame[,comp_index]-data_frame[,i])
CV <- sd(diff)/sqrt(2)
CV_log <- sd(diff_log) / sqrt(2)
CVP <- ((100*(exp(CV_log / 100))) - 100) * 100

#Bootstrapping CI of SEM
BSW <- NULL #Empty Object for bootstrap values
for (l in 1:10000) {
  BSdiff <- sample(diff, size = length(diff),
                   replace = TRUE)
  BSW <- c(BSW, sd(BSdiff)/sqrt(2)) #Save bootstrap values
}

CV_CI<- quantile(BSW, probs = c(0.025, 0.975), na.rm = TRUE) #Get quantiles = CI

#Mean bias standardized to
Mean_bias <- bland.altman.stats(data_frame[,comp_index],
                                data_frame[,i])$mean.diffs
LLOA <- bland.altman.stats(data_frame[,comp_index],
                           data_frame[,i])$lower.limit
ULOA <- bland.altman.stats(data_frame[,comp_index],
                           data_frame[,i])$upper.limit
Stmd <- Mean_bias / sd(data_frame[,comp_index]) #Standardized mean difference

#Creating dataframe to convert to CSV
comp_vals <- data.frame("Comparison" = var_names,
                        "ICC" = ICC,
                        "ICC_L" = ICC_L,
                        "ICC_U" = ICC_U,
                        "MD" = MD,
                        "MD_CI_L" = MD_CI[1],
                        "MD_CI_U" = MD_CI[2],
                        "MDP" = MDP,
```

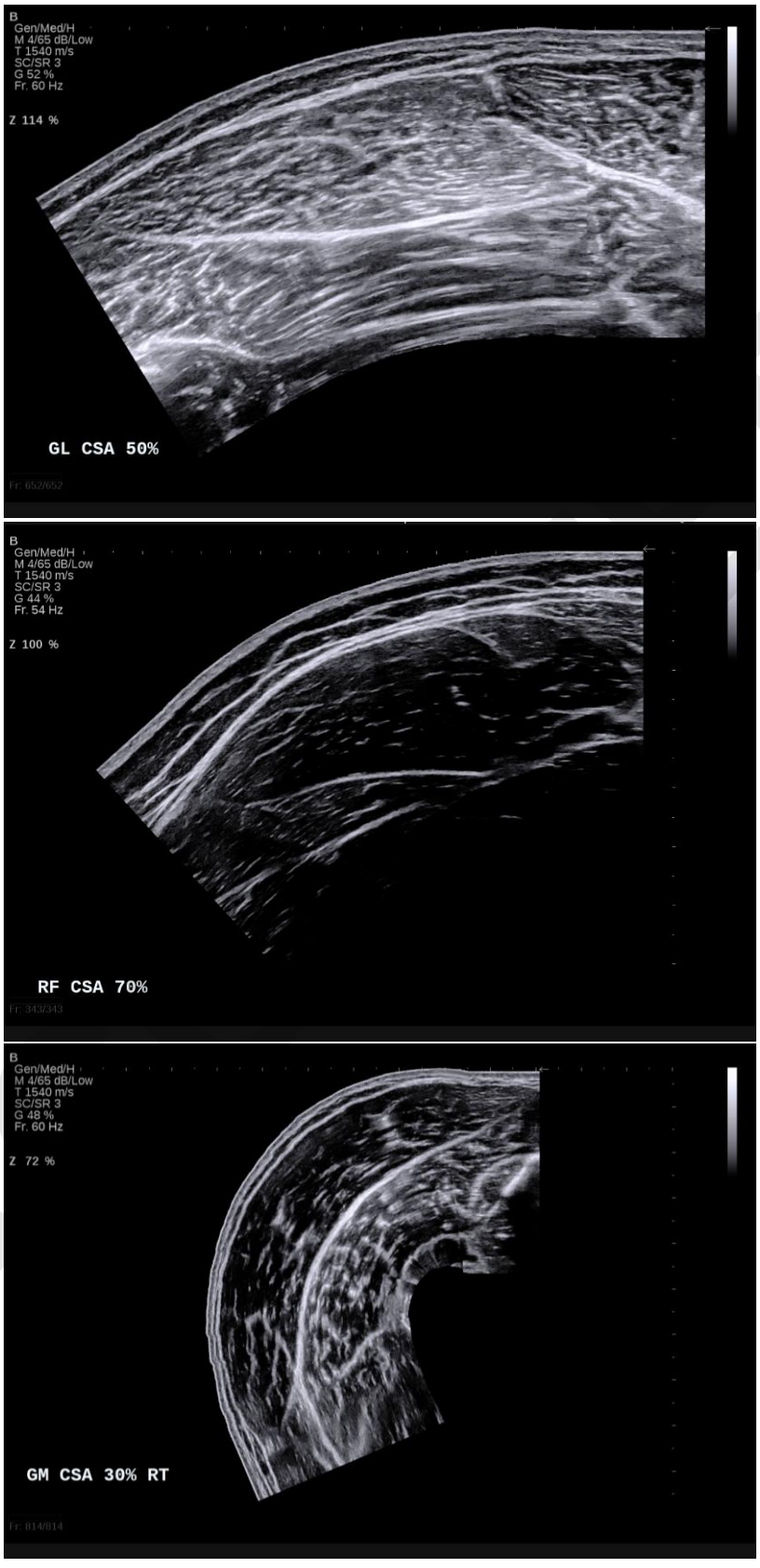
```
                    "CV" = CV,
                    "CV_CI_L" = CV_CI[1],
                    "CV_CI_U" = CV_CI[2],
                    "CVP" = CVP,
                    "Mean_bias" = Mean_bias,
                    "LLOA" = LLOA,
                    "ULOA" = ULOA,
                    "Stmd" = Stmd
    )
    write.table(comp_vals, file = filename,
                append=TRUE,
                sep=",",
                col.names=TRUE,
                row.names=FALSE)
   }
  }
}



calc_BaPlots <- function(root, comp_index){
  #'  Calculates Bland-Altman Plots for comparability analysis.
  #'  Args: root = Path to directory containgy .xlsx files (String)
  #'        comp_index = Column to be compared to (Integer)
  #'  Returns: .JPG file with different Plots.

  # Get files in root
  df_list <- get_file_list(root)
  # Loop trough files
  for (data_frame in df_list){

    # Convert tibble to data.frame
    data_frame <-  as.data.frame(data_frame)
    names <- colnames(data_frame)
    name <- paste(names[3], ".jpg", sep = "")

    # Define image settigns
```

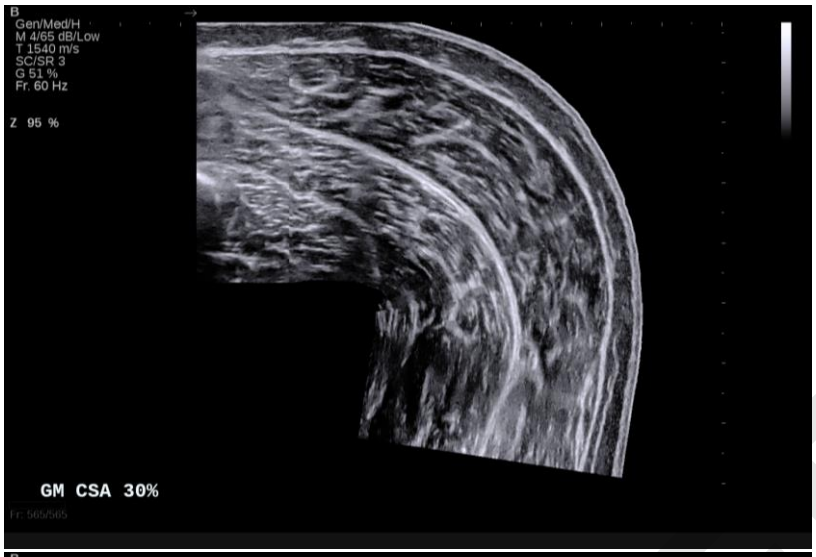
```
jpeg(name,
    width = 1640, height = 640, units="px",
    quality = 100, pointsize = 15, bg = "transparent",
    res = 100)

oldpar <- par()
par(mar = c(3,3,2,1), cex.axis = 1, cex.lab = 1.2,
   mgp = c(1.75, 0.5, 0), tcl = -0.2, mfcol = c(1,2))

# Loop trough columns of data_frame
for(i in 3:ncol(data_frame)) {

 # Get variable names
 var_names <- colnames(data_frame)
 var_names <-  paste(var_names[comp_index], var_names[i], sep = " vs. ")

 ##Manual Quadriceps RF
 a <- bland.altman.stats(data_frame[,comp_index],
                         data_frame[,i])
 b <- a$lines

 #Creating bland-altman Plots
 plot(a$means, a$diffs,
     panel.last = abline(h = a$lines, lty = 2),
     ylab = "Differences of measures (cm²)",
     xlab = "Mean of values (cm²)",
     ylim = c(-9.5, 9.5),
     xlim = c(min(a$means) - 0.5, max(a$means) + 2),
     main = var_names)

 segments(1, a$mean.diffs, 3 , a$mean.diffs)
 text(max(a$means) + 1, b[3] + 0.4, round(b[3], 3), cex = 0.7)
 text(max(a$means) + 1, b[3] + 1,"+1.96 SD", cex = 0.7)
 text(max(a$means) + 1, b[2] - 0.3, round(b[2], 3), cex = 0.7)
 text(max(a$means) + 1, b[2] + 0.37,"Mean", cex = 0.7)
 text(max(a$means) + 1, b[1] - 1, round(b[1], 3), cex = 0.7)
```

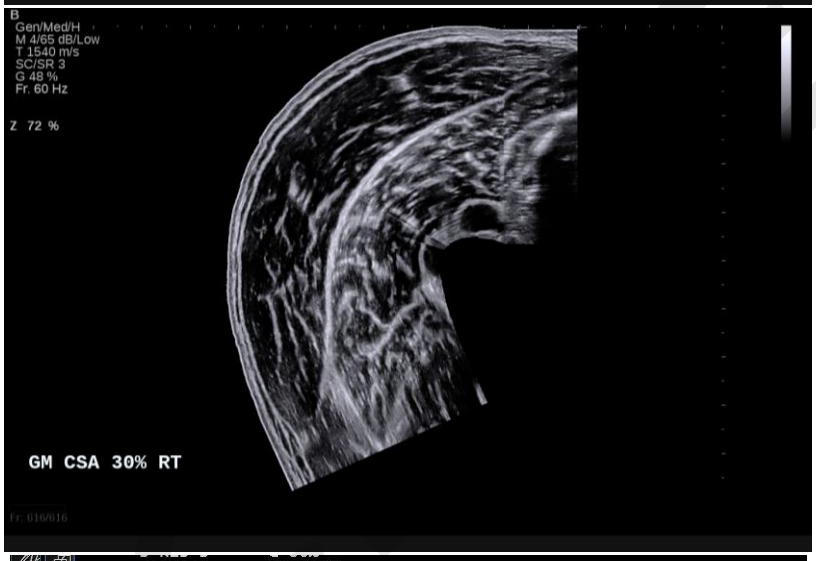
```
    text(max(a$means) + 1, b[1] - 0.4,"-1.96 SD", cex = 0.7)
   }
   dev.off()
  }
}

##################
#####Analysis#####
##################

#####Set root Directory#####
setwd(dirname(getActiveDocumentContext()$path)) # Gets file location
working_dir <-  getwd() # Sets location to working directory

#####Comparability analysis#####
calc_comp(working_dir, 2, "comparability_pre-post.csv")

#####Bland-Altman Plots#####
calc_BaPlots(working_dir, 2)
```

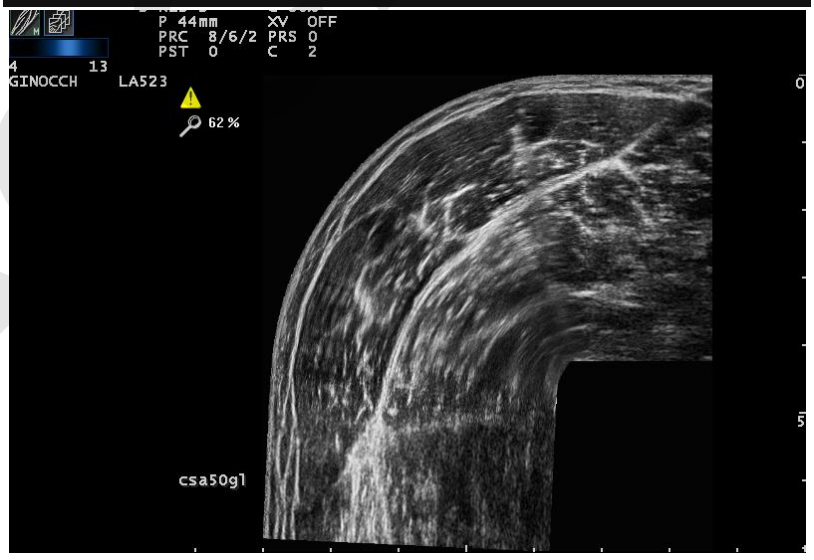
Supplementary file containing all images with erroneous predictions using DeepACSA.

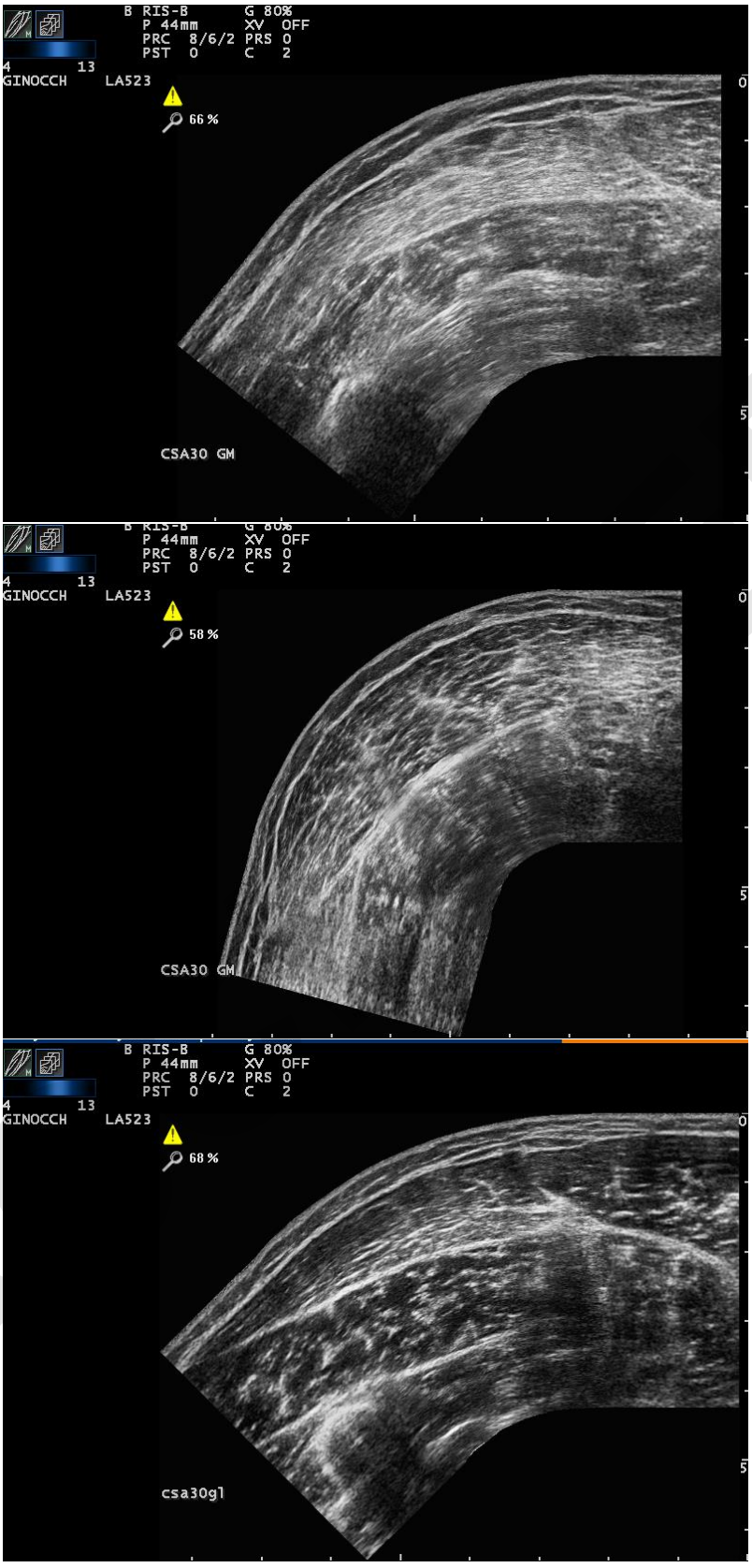
Firstly, the eight images of the gastrocnemius medialis and lateralis are presented. Then, the seven images of the vastus lateralis are presented. Lastly, the two images of the rectus femoris are presented.

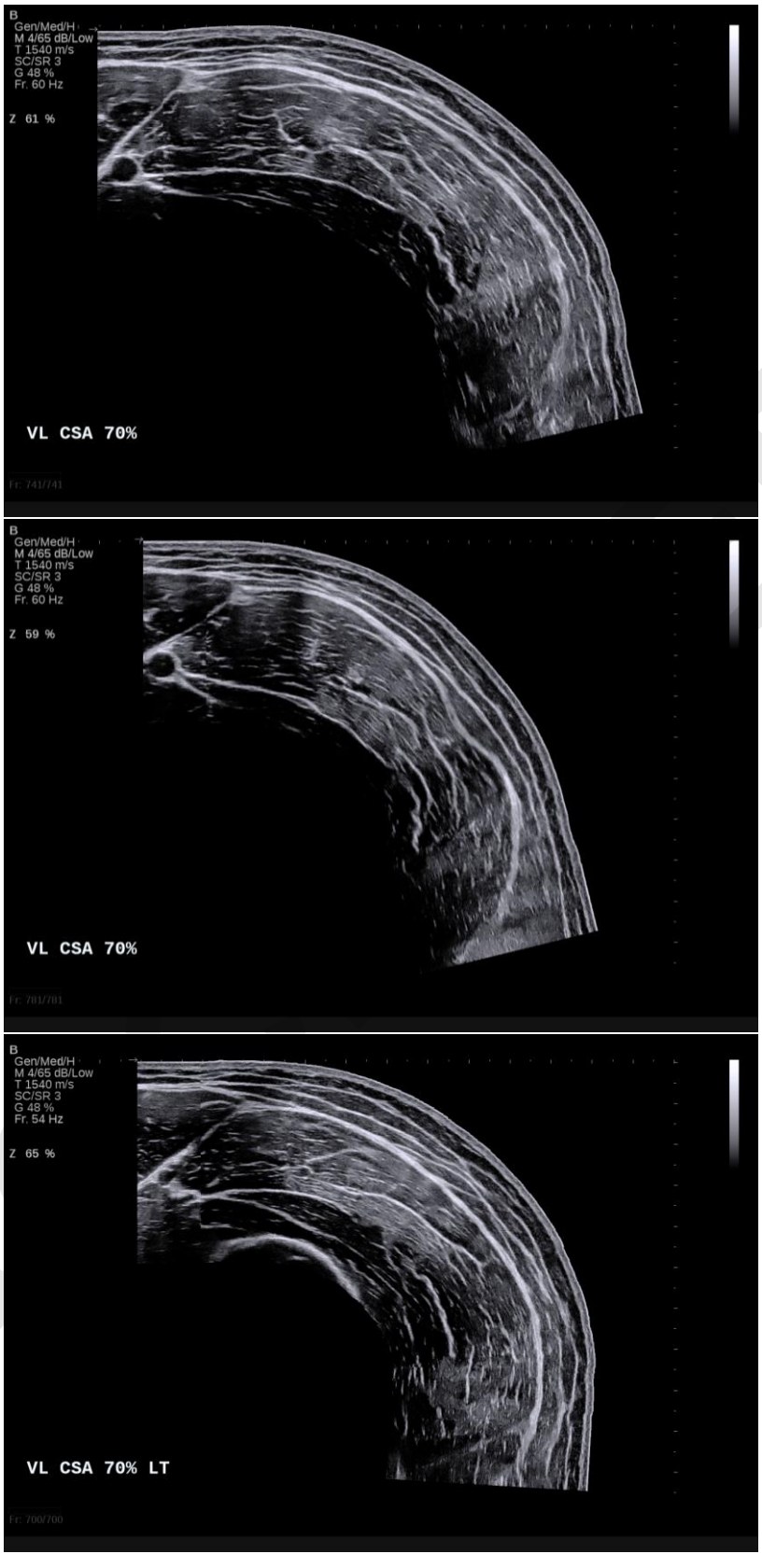


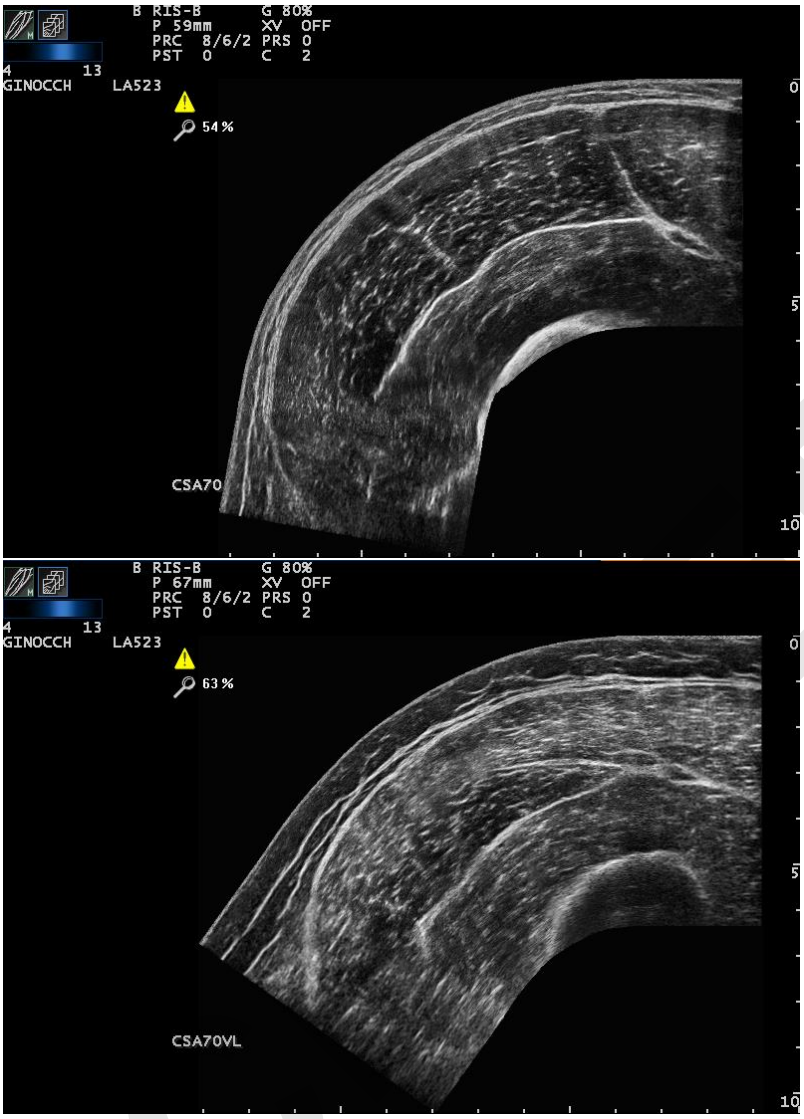


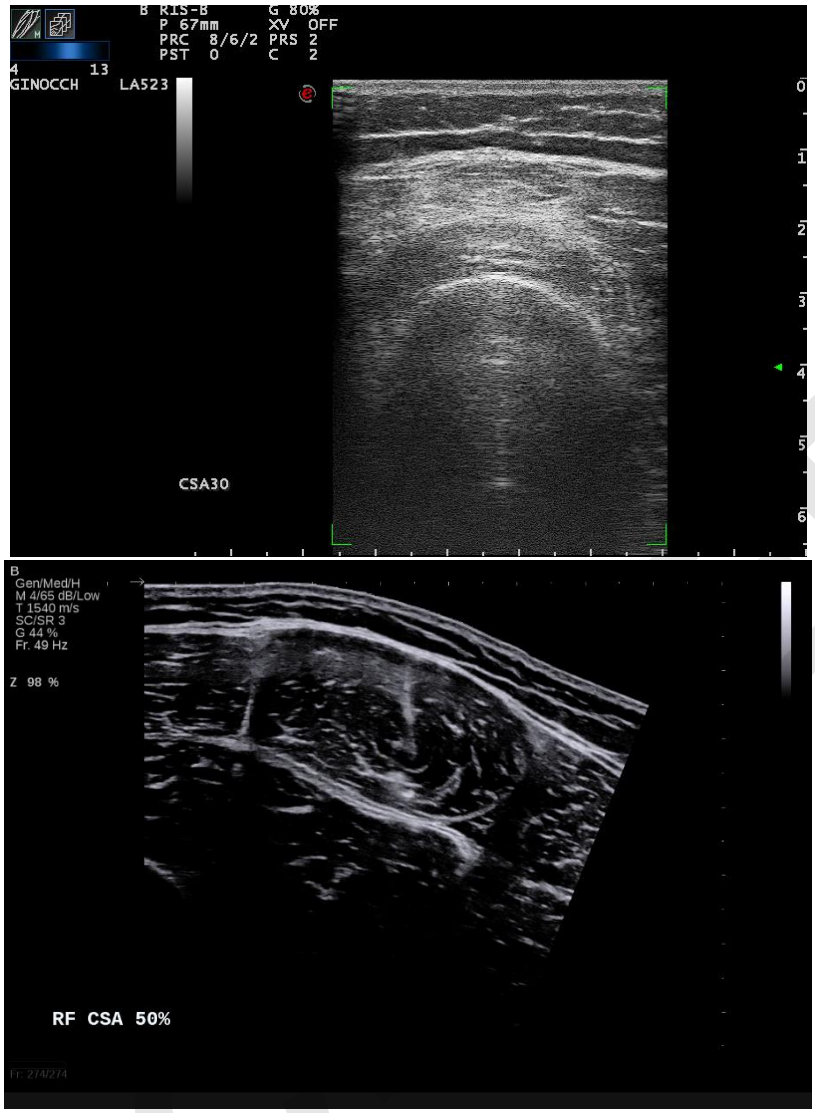












Supplementary Table 1: Comparison of manual analysis and ACSAuto and DeepACSA at 30% femur length

ACSA 30%

Muscle	Mode	ICC	MD	MDP	CV	CVP	Mean_bias	Stmd
RF	Hand vs. ACSAuto	0.983 (0.966;0.992)	0.06 (-0.01;0.13)	1.56	0.13 (0.08;0.18)	8.26	0.06 (-0.32;0.43)	0.05
RF	Hand vs. DeepACSA	0.941 (0.881;0.971)	0.03 (-0.11;0.18)	9.21	0.27 (0.16;037)	38.01	0.03 (-0.72;0.79)	0.03
VL	Hand vs. ACSAuto	0.981 (0.961;0.991)	-0.268 (-0.506; -0.03)	1.66	0.451 (0.333;0.554)	2.911883	-0.268 (-1.517;0.981)	-0.08
VL	Hand vs. Deep256	0.964 (0.923;0.983)	0.18 (-0.15;0.52)	1.13	0.63 (0.29;0.97)	4.94	0.18 (-1.57;1.94)	0.05
GM/GL	Hand vs. ACSAuto	0.961 (0.903;0.985)	-0.11 (-0.49;0.26)	1.34	0.54 (0.27;0.79	4.62	-0.11 (-1.64;1.41)	-0.03
GM/GL	Hand vs. DeepSingleL	0.962 (0.919;0.982)	0.58 (0.30;0.87)	4.17	0.69 (0.39;0.97)	6.83	0.49 (-1.42;2.41)	0.12

RF = rectus femoris, VL = vastus lateralis, GM&GL = gastrocnemii, rm = wrong prediction removed, ICC = Intra-class correlation, CI = compatibility interval, MD = mean difference, SEM = Standard error of the differences, and Stmd = standardized mean difference.

Supplementary Table 2: Comparison of manual analysis and ACSAuto and DeepACSA at 50% femur length

ACSA 50%

Muscle	Mode	ICC	MD	MDP	CV	CVP	Mean_bias	Stmd
RF	Hand vs. ACSAuto	0.975 (0.948;0.988)	0.32 (0.24;0.41)	4.26	0.16 (0.12;0.19)	2.18	0.32 (-0.12;0.77)	0.18
RF	Hand vs. Deep256	0.973 (0.946;0.987)	0.13 (-0.02;0.28)	2.12	0.29 (0.16;0.41)	3.61	0.13 (-0.66;0.92)	0.07
VL	Hand vs. ACSAuto	0.986 (0.97;0.993)	-0.05 (-0.379;0.279)	-0.55	0.623 (0.452;0.743)	2.67	-0.05 (-1.777;1.677)	-0.01
VL	Hand vs. Deep256	0.947 (0.892;0.974)	0.71 (0.17;1.27)	2.64	1.04 (0.73;1.27)	4.17	0.72 (-2.18;3.61)	0.14
GM/GL	Hand vs. ACSAuto	0.973 (0.921;0.991)	0.3 (-0.05;0.65)	2.47	0.42 (0.24;0.54)	3.03	0.3 (-0.88;1.47)	0.07
GM/GL	Hand vs. DeepSingleL	0.962 (0.939;0.977)	0.51 (0.29;0.74)	4.36	0.63 (0.46;0.80)	6.15	0.51 (-1.24;2.28)	0.13

RF = rectus femoris, VL = vastus lateralis, GM&GL = gastrocnemii, rm = wrong prediction removed, ICC = Intra-class correlation, CI = compatibility interval, MD = mean difference, SEM = Standard error of the differences, and Stmd = standardized mean difference.

Supplementary Table 3: Supplementary table 3: Comparison of manual analysis and ACSAuto and DeepACSA at 70% femur length

ACSA 70%

Muscle	Mode	ICC	MD	MDP	CV	CVP	Mean_bias	Stmd
RF	Hand vs. ACSAuto	0.961 (0.921;0.981)	0.35 (0.19;0.51)	2.66	0.3 (0.22;0.36)	2.25	0.35 (-0.49;1.18)	0.18
RF	Hand vs. Deep256	0.862 (0.732;0.931)	0.55 (0.21;0.88)	4.31	0.64 (0.33;0.92)	4.88	0.54 (-1.22;2.32)	0.28
VL	Hand vs. ACSAuto	0.968 (0.935;0.985)	-0.08 (-0.537;0.376)	-0.45	0.865 (0.63;1.037)	3.732241	-0.08 (-2.477;2.317)	-0.02
VL	Hand vs. Deep256	0.745 (0.533;0.869)	0.1.66 (0.47;2.84)	7.98	2.24 (1.05;3.42)	13.23	1.66 (-4.55;7.88)	0.34

RF = rectus femoris, VL = vastus lateralis, GM&GL = gastrocnemii, rm = wrong prediction removed, ICC = Intra-class correlation, CI = compatibility interval, MD = mean difference, SEM = Standard error of the differences, and Stmd = standardized mean difference.