



OPEN

Development of an artificial intelligence powered software for automated analysis of skeletal muscle ultrasonography

Zoe Calulo Rivera^{1,14}, Felipe González-Seguel^{2,3,4,14}, Arimitsu Horikawa-Strakovsky^{2,5,6}, Catherine Granger¹, Aarti Sarwal⁷, Sanjay Dhar⁸, George Ntoumenopoulos⁹, Jin Chen¹⁰, V. K. Cody Bumgardner^{6,11}, Selina M. Parry^{1,15}, Kirby P. Mayer^{2,3,15}✉ & Yuan Wen^{2,6,11,12,13,15}✉

Muscle ultrasound has high utility in clinical practice and research; however, the main challenges are the training and time required for manual analysis to achieve objective quantification of muscle size and quality. We aimed to develop and validate a software tool powered by artificial intelligence (AI) by measuring its consistency and comparability of expert manual analysis quantifying lower limb muscle ultrasound images. Quadriceps complex (QC) and tibialis anterior (TA) muscle images of healthy, intensive care unit, and/or lung cancer participants were captured with portable devices. Manual analyses of muscle size and quality were performed by experienced physiotherapists taking approximately 24 h to analyze all 180 images, while automated analyses were performed using a custom-built deep-learning model (MyoVision-US), taking 247 s (saving time = 99.8%). Consistency between the manual and automated analyses was good to excellent for all QC (ICC = 0.85–0.99) and TA (ICC = 0.93–0.99) measurements, even for critically ill (ICC = 0.91–0.98) and lung cancer (ICC = 0.85–0.99) images. The comparability of MyoVision-US was moderate to strong for QC (adj. R^2 = 0.56–0.94) and TA parameters (adj. R^2 = 0.81–0.97). The application of AI automating lower limb muscle ultrasound analyses showed excellent consistency and strong comparability compared with human analysis across healthy, acute, and chronic population.

Keywords Skeletal muscle, Ultrasound, Automation software, Artificial intelligence, Medical image analysis

Abbreviations

AI Artificial intelligence
CSA Cross-sectional area
EI Echo intensity

¹Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Melbourne, Australia. ²Center for Muscle Biology, University of Kentucky, Lexington, KY, USA. ³Department of Physical Therapy, College of Health Sciences, University of Kentucky, 900 S. Limestone St. CTW, Rm: 204D, Lexington, KY 40536, USA. ⁴School of Physical Therapy, Faculty of Medicine, Clínica Alemana Universidad del Desarrollo, Santiago, Chile. ⁵Math, Science, and Technology Center, Paul Laurence Dunbar High School, Lexington, KY, USA. ⁶Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, USA. ⁷Department of Neurology, Wake Forest Baptist Medical Center, Winston Salem, NC, USA. ⁸Division of Pulmonary, Critical Care & Sleep Medicine, Department of Internal Medicine, College of Medicine, University of Kentucky, Lexington, KY, USA. ⁹Department of Physiotherapy, St. Vincent's Hospital, Sydney, Australia. ¹⁰Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ¹¹Department of Pathology and Laboratory Medicine, College of Medicine, University of Kentucky, 760 Press Ave. HKRB, Rm: 364, Lexington, KY 40508, USA. ¹²Division of Biomedical Informatics, Department of Internal Medicine, College of Medicine, University of Kentucky, Lexington, KY, USA. ¹³Department of Physiology, College of Medicine, University of Kentucky, Lexington, KY, USA. ¹⁴Zoe Calulo Rivera, Felipe González-Seguel, and Arimitsu Horikawa-Strakovsky contributed equally as co-first authors. ¹⁵Selina M. Parry, Kirby P. Mayer and Yuan Wen contributed equally as co-senior authors. ✉email: kpmaye2@uky.edu; ywen2@uky.edu

ICC	Intraclass correlation coefficient
ICU	Intensive care unit
QC	Quadriceps complex
RF	Rectus femoris
TA	Tibialis anterior
VI	Vastus intermedius

Peripheral muscle dysfunction can encompass reductions in muscle mass, strength, endurance, and fatigability¹. Muscle ultrasound imaging enables non-invasive bedside evaluation of skeletal muscle quantity (i.e., muscle thickness, anatomical cross-sectional area [CSA]), quality (echo intensity [EI]) and architectural properties (i.e., pennation angle, fascicle length, contraction potential)^{1,2}. It has become increasingly popular as an assessment tool to evaluate longitudinal changes in muscle characteristics of healthy and diverse populations such as individuals with critical illness and lung cancer^{1,3–6}. Both lung cancer⁷ and critical illness⁸ contribute to a hypercatabolic state driven by systemic inflammation, oxidative stress, and hormonal imbalances, including insulin resistance and cortisol dysregulation. Current literature is focused on muscle ultrasound evaluation of the quadriceps and tibialis anterior muscle groups due to its size and importance in sarcopenia diagnosis and detection of muscle wasting^{9–13}. Muscle ultrasound imaging has demonstrated good to excellent inter- and intra-rater reliability and strong criterion validity compared to histochemistry analyses on muscle biopsy in older adults¹⁴, individuals with myopathic disease¹⁵, and intensive care unit (ICU) survivors 6–12 months after discharge¹⁶.

The benefits of muscle ultrasound support the utilization of this tool in clinical and research practice. However, the ability of parameters defined from muscle ultrasound to predict meaningful outcomes is equivocal^{4,17}. Limited predictive validity may partially be explained by sonographers using a subjective scale at the bedside (ie. the Heckmatt approach for qualitative evaluation of EI)¹⁸ or the need for a trained expert to manually analyze ultrasound images for objective parameters^{2,19,20}. The objective analysis of muscle parameters from ultrasound images is manual operator-dependent, requiring sustained human engagement that is time and labor-intensive, increasing the potential for human biases. Computational programs capable of mimicking human decision-making behavior are broadly known as artificial intelligence (AI)²¹. Machine learning is a branch of AI and involves the process of building algorithms to learn patterns or rules based on data²². Automated methods of analyzing ultrasound images have been investigated to address the limitations created by manual data interpretation. In the last decade, the use of AI and machine learning in medical research, particularly medical image analysis, has grown in popularity, as these models have been shown to improve efficiency and lower the error rate of clinical tasks²². Recent studies by Cronin et al. and Katakis et al. have shown that machine and deep learning can improve existing automated systems and produce comparable results to manual analysis, indicating that these methods are both acceptable and reliable^{23–25}. In 2017, Caresio et al. developed and validated a fully automatic method, named MUSA (Muscle UltraSound Analysis) for measuring muscle thickness on longitudinal ultrasound images only of healthy individuals²⁶. Previous studies have demonstrated variable performance of automated segmentation approaches across different image qualities^{27,28}. We believe that both high-quality data acquisition and robust algorithm development are complementary goals rather than competing priorities. While striving for optimal image acquisition remains essential, clinical realities—particularly in critical care settings—often present challenges that result in variable image quality. Our approach acknowledges this reality by developing algorithms that maintain reasonable performance across the spectrum of image quality encountered in clinical practice. Rather than suggesting that algorithms should compensate for poor imaging technique, we aim to create tools that work effectively within real-world constraints while still emphasizing the importance of standardized acquisition protocols. The goal is to develop AI models that can generalize across reasonable variations in image quality, increasing their clinical utility without diminishing the importance of proper ultrasound technique. This dual approach of improving both image acquisition standards and algorithm robustness will likely yield the best outcomes for patients and researchers.

MyoVision is a freely available automated image analysis software that was originally developed for the quantification of muscle cross-sections from biopsies in animal models and human studies^{29,30}. It is a reliable and accurate program that automates the analyses of immunohistochemistry^{29,30}. The use of MyoVision for muscle biopsies and the need for automated software for ultrasonography led to the development of MyoVision-US to provide a comprehensive automated quantification of lower limb muscle without human supervision. Therefore, this study aimed to develop and validate the AI-powered MyoVision-US by measuring its consistency and comparability of expert manual analysis of muscle thickness, CSA, and EI of ultrasound images across healthy, acute, and chronic illness cohorts.

Methods

Study design

A prospective study for the development and validation of MyoVision-US was conducted in accordance with prior work^{29,30}, and data was compared to manual analysis of muscle ultrasound images. The reporting of clinimetric properties was performed in accordance with the CONensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines³¹. This study received approval from the University of Kentucky Medical Expedited Internal Review Board (IRB #75185) on February 18, 2022 with Data Use Agreement of the University of Kentucky for image sharing across sites. Informed consent was waived by the University of Kentucky Medical Expedited Internal Review Board due to the study design and deidentified approach with ultrasonography images. All methods were performed in accordance with approved institutional regulations.

Musculoskeletal ultrasound image dataset

A diverse dataset of muscle ultrasound images was developed from active (ACTRN12618001733268; ACTRN12617001283369; NCT03141762) and completed studies in United States, Australia (Melbourne), Australia (Sydney), and Chile^{4,9,20,32–35}. All of these studies followed standardized measurement protocols for image acquisition, including knobology, probe type, compression and tilting technique, and muscle and bone visualization. Prior to dataset transfer all personal or patient identifiers were removed from images or data files. A test dataset was curated to assess the performance and generalizability of the automated MyoVision-US software. A blinded researcher randomly selected 90 ultrasound images of the quadriceps complex (QC) including 30 from adults with critical illness in the ICU or during the recovery phase, 30 from adults with lung cancer with outpatient scans, and 30 volunteer healthy adult individuals. Additionally, 90 tibialis anterior (TA) ultrasound images were also randomly extracted (not necessarily from the same individuals with QC images) including 60 selected from adults in the ICU and 30 selected from the volunteer healthy adults. We did not include TA images of lung cancer individuals due to lack of availability across study dataset. The three adult populations were selected to provide a spectrum of muscle health to enhance generalizability during the testing ranging from healthy, acute, and chronic illness states. We focused on the QC and TA muscles because they are commonly assessed in clinical practice for several reasons: (1) they are readily accessible for ultrasound imaging, (2) they have established clinical relevance in sarcopenia diagnosis and frailty assessment, (3) they show significant changes during critical illness and cancer-related muscle wasting, and (4) previous research has demonstrated their predictive value for functional outcomes and mortality in critical illness and those with chronic diseases^{9–13}. These muscles also represent both upper and lower leg compartments, providing a more comprehensive evaluation of peripheral muscle status. Image selection included those that allowed standardized pixel analysis.

Manual image quantification

Manual image analysis was performed by three independent physiotherapists (SMP, KPM, and FGS) with at least seven years of muscle ultrasound practical experience^{2,4,9,16,20,32,34–38} using freely available NIH Image J software (version 1.53t)³⁹. SMP has 14 years of theoretical and practical ultrasound imaging expertise, including training of other clinicians in the use of muscle ultrasound in clinical and research practice internationally at conferences and workshops. KPM has over 9 years of experience utilizing muscle ultrasound in research and practice, including leading educational sessions in the United States. FGS has over 8 years of experience utilizing muscle ultrasound in research and practice and has led Chilean educational courses on muscle ultrasound. All have co-authored published papers by using similar ultrasound protocols in both clinical and educational contexts related to muscle ultrasound imaging^{2,4,9,16,20,32,34–38}.

Each experienced physiotherapist evaluated a third of the total ultrasound images, where all muscle ultrasound parameters (muscle thickness, anatomical CSA, and EI) were evaluated in triplicate using the mean of the three repeated values used in the final analyses. All muscle thickness parameters were measured in centimeters. Rectus femoris (RF) thickness was measured between the two transverse fascial planes of the muscle belly utilizing the center of the femur for consistency. Vastus intermedius (VI) thickness was measured between the uppermost part of the femur bone to the superficial fascia of the VI muscle utilizing the center of the femur for consistency. The thickness of QC was measured between the uppermost part of the bone echo of the femur and the superficial fascia and includes both the RF and VI muscles. TA thickness was measured from the superficial fascia to the uppermost part of the bone echo of the tibia. Cross-sectional area was measured in square centimeters using the free-form mode to trace the inside border of the epimysium for both RF and TA muscles. Echo intensity was quantified based on the grey scale of the pixels of the image ranging from 0 (black) to 255 (white) for RF and TA muscles using the greyscale of the pixels calculated using the free-form trace outline method described to determine CSA.

MyoVision-US software implementation

As shown in Fig. 1, the MyoVision-US workflow consists of three steps. The first step is the prediction by a DeepLabV3⁴⁰ model, followed by the second step of post-processing, and the last step involves the calculation of muscle image parameters.

Step 1 Two semantic segmentation models were trained to automatically segment QC ultrasound images and TA ultrasound images respectively. To obtain the best performance, ablation studies were performed to find the optimal architecture, loss function, and training settings. For model architectures, DeepLabV3 with a ResNet50⁴¹ backbone, UNet⁴² with a ResNet50 backbone, and SegFormer⁴³ with an MIT-B2 and an MIT-B3 backbone were compared. For loss functions, binary-cross entropy loss, Jaccard loss, Tversky loss⁴⁴, Lovasz loss⁴⁵, and focal loss⁴⁶ were tested. For general training settings, epoch count and strong color augmentations were investigated. After such studies, 2 DeepLabV3 models with a ResNet50 backbone were deemed optimal, one for QC ultrasound images and one for TA ultrasound images. DeepLabV3 is a deep, fully convolutional neural network architecture for semantic segmentation, which utilizes a backbone to initially identify and extract useful features from the input image and cascades of atrous convolutions to recognize and segment objects at multiple scales. The ResNet50 model is a residual network architecture that introduced skip connections and is commonly used for computer vision and image classification tasks. The dataset used to train the models consists of 94 QC images and 243 TA images representing a mixed group of ICU, lung cancer, and healthy individuals (this set is separate from the test dataset described in previous sections). These images were labeled by a technician and verified by experienced clinicians with polygon traces of the maximum visible CSA of the RF, VI, and femur on QC images and the TA on TA images. These datasets were split into independent training and validation sets with approximately 20% of the images in the validation set (76/18 and 195/48 for QC and TA respectively). Images were acquired using various portable ultrasound devices across the different

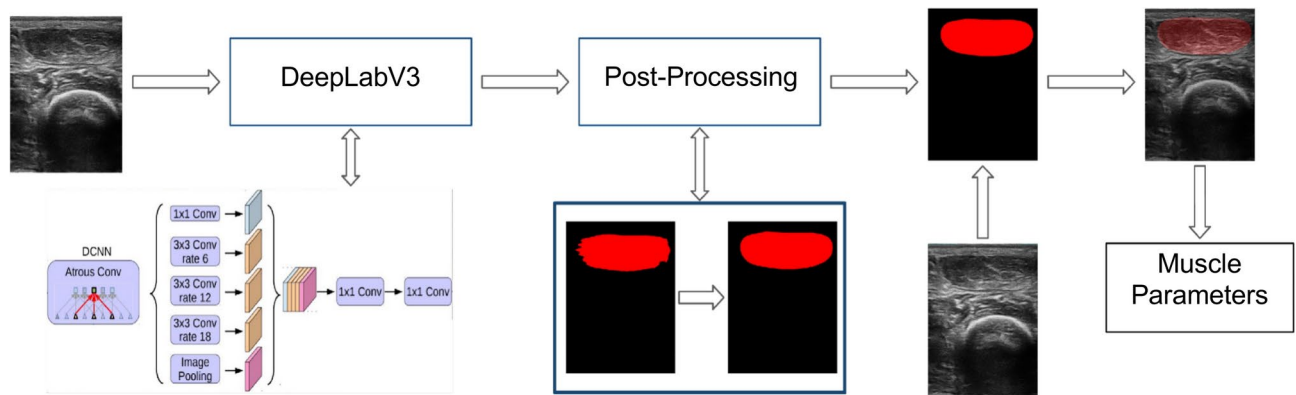


Fig. 1. Diagram of MyoVision-US workflow, with 3 stages: DeepLabV3 model, post-processing, and value calculation. The analysis of the rectus femoris muscle in an ultrasound image of the quadriceps complex is specifically shown.

study sites, including Mindray M7 (Mindray, Shenzhen, China), SonoSite M-Turbo and X-Porte (FUJIFILM SonoSite, Inc., Bothell, WA, USA), and GE LOGIQ E (GE Healthcare, Chicago, IL, USA). All devices used linear array transducers with frequencies between 6–13 MHz, and standardized protocols for image acquisition were followed at all sites to minimize device-specific variations. Images and labels were normalized and standardized to stabilize training while balancing memory usage. Spatial and strong color data augmentation was also used and was performed using Albumentations⁴⁷, a python library used for image augmentation in computer vision tasks. The ResNet50 backbone was initialized with pretrained ImageNet⁴⁸ weights. The whole model was then trained on either the QC or TA datasets for 40 epochs using a cosine annealing learning rate scheduler⁴⁹ with linear warm up, a maximum learning rate of $5e-4$, an AdamW⁵⁰ optimizer with a weight decay of $1e-6$, and a binary cross-entropy loss function. For additional training details, please refer to the supplementary training script provided at: https://github.com/Arimitsu06/myovision-us_training_scripts. The models were evaluated using the Dice coefficient⁵¹ and Intersection over Union (IoU), which are both measures of similarity between predicted and ground truth labels ranging from 0 being no spatial agreement to 1 being complete agreement (in this case, these measures were used to compare model predicted and expert labelled maximum visible CSA of muscles and bone). The trained DeepLabV3 models of QC and TA achieved Dice coefficients of 0.90 (0.88, 0.86, and 0.96 for RF, VI, and femur respectively) and 0.95 and mIoUs of 0.88 (0.87, 0.83, and 0.95 for RF, VI and femur respectively) and 0.90 on the validation sets.

Step 2 A post-processing scheme was used to minimize noise and artifacts in the model's prediction similar to Katakis et al.²⁵. Post-processing starts with the extraction of contours (boundaries of all distinct polygons) in the output prediction (mask) from the model. Then, the contour with the largest internal area was kept and morphological operations, opening and closing, were performed to remove noise by filling in gaps and removing protrusions. In addition, a cubic spline was fit along the contour of the mask to further smooth it.

Step 3 After post-processing, calculations of specific values were made after resizing to the size of the original image with nearest-neighbor interpolation. Muscle thickness was calculated by finding the number of pixels predicted to be the target muscle at either the midpoint of the predicted femur mask or the midpoint of the predicted mask of the target muscle (i.e. RF or TA). QC thickness was found by finding the distance, in pixels, at the midpoint of the predicted femur mask between the top of the predicted mask of the femur and the top of the predicted mask of the RF. Cross-sectional area was calculated by using the number of pixels in the predicted mask of the target muscle. Area and thickness measurements were converted to centimeters using the pixel scale for each image calculated from the corresponding depth parameter. Mean and standard deviation of EI was calculated by averaging or finding the standard deviation of the grayscale value of all pixels within the predicted mask of the target muscle in the original image.

Statistical analysis

Descriptive statistics, histogram plots, and Shapiro–Wilk were performed for all muscle parameters (thickness, anatomical CSA, EI) for both QC and TA muscles. Descriptive statistics are provided for each group (healthy, ICU, and lung cancer) for generalizability. Intraclass correlation coefficients (ICC) were calculated based on the average of measures comparing the software to the manual analysis (95% confidence interval). A two-way random effects test of the average measures was used to examine interrater consistency. The ICC values are scaled from 0 to 1, where an ICC between 0.75 and 0.90 is considered to be good and higher than 0.90 is considered to be excellent reliability⁵². Mean bias error and coefficient of variation were calculated to measure systematic and random errors, respectively. To determine the comparability of MyoVision-US (independent variable) and the manual analysis (dependent variable), separate and adjusted linear regression analyses were performed for each muscle parameter. Standard error of measurement (SEM), or standard errors of the coefficients, were reported to evaluate the precision of the regression estimates. The data were analyzed using IBM SPSS Statistics (version 29). For MyoVision-US software implementation we used Python (coding language) version 3.9.21, Cuda (GPU

coding language) version 12.5, Pytorch 2.6.0, Torchvision 0.21.0, Segmentation_Models_Pytorch 0.4.1.dev0, Albumentations 2.0.2, and Opencv-Python 4.11.0.86.

Results

Ninety QC and 90 TA images independent of the images used in the training or validation set of the AI model training were randomly selected to be analyzed by MyoVision-US and compared to manual analyses performed by expert physiotherapists with ultrasound expertise. Manual analysis took experts collectively approximately 24 h to analyze all 180 images with each image requiring roughly 8 min. MyoVision-US took 247 s to analyze all 180 images (163 s for QC images and 84 s for TA), saving roughly 99.8% of the time used for manual analysis. Three of the 90 QC images (3.3%) and five of the 90 TA images (3.4%) were not analyzed in manual analysis due to poor image quality in the ICU group related to marked muscular or subcutaneous adiposity/edema, otherwise, there were no missing data. Descriptive statistics for the ultrasound images analyzed by MyoVision-US are provided in Table 1. Representative images for RF (Fig. 2A) and TA (Fig. 2B) demonstrate the software’s recognition of the muscles of interest for both healthy and ICU patients. MyoVision-US automated analyses of muscle thickness, CSA, and EI indicate that ICU and/or lung cancer was associated with lower values for RF (Fig. 2C) and TA (Fig. 2D) compared to healthy individuals. Direct comparisons between manual and automated analysis of RF (Fig. 2E) and TA (Fig. 2F) demonstrate high degrees of consistency for thickness, CSA, and EI. Bland–Altman plots indicate that the two methods of analysis are highly consistent for both the RF (Fig. 2G) and the TA (Fig. 2H).

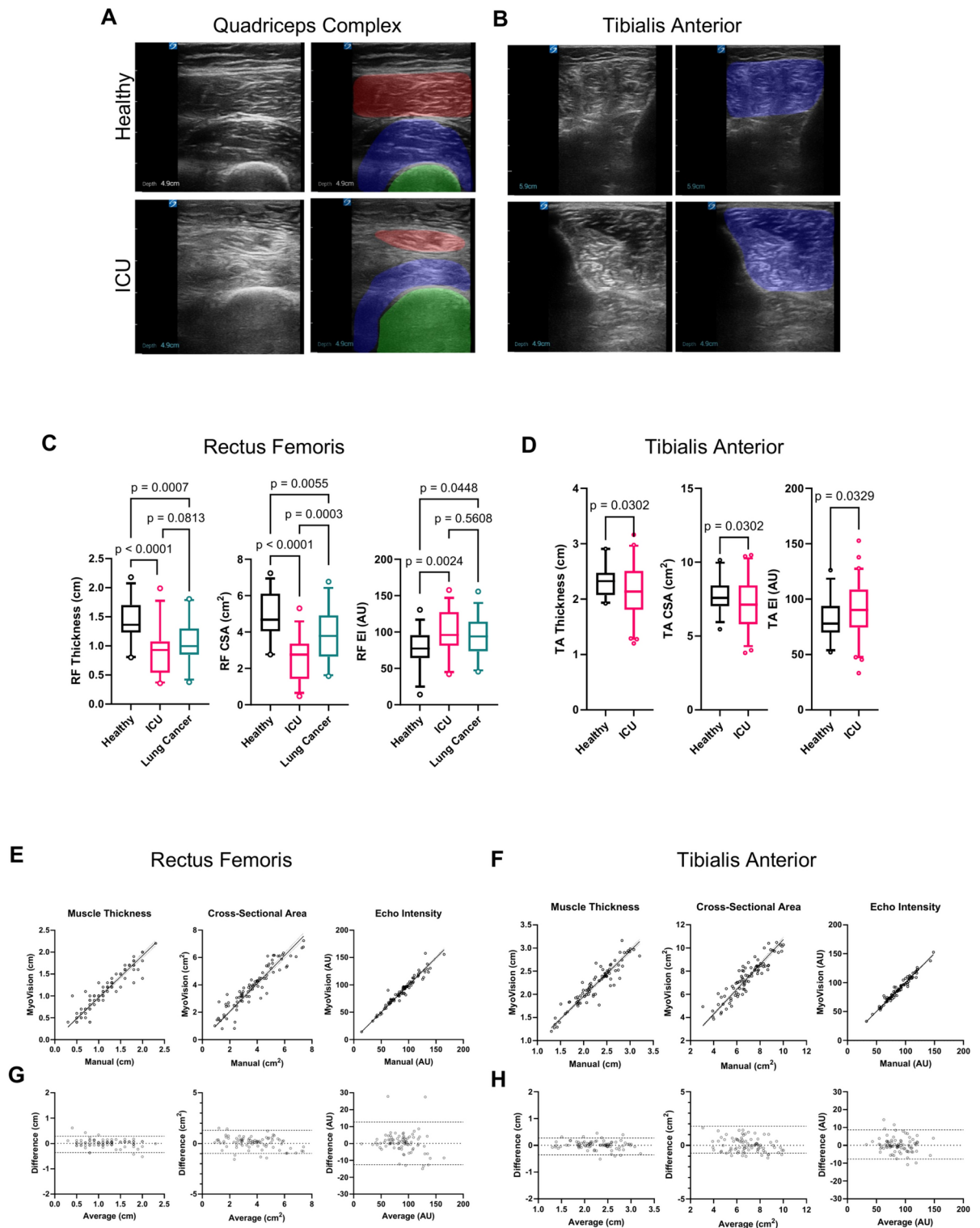
The consistency between MyoVision-US and manual analyses for QC images, including examinations of VI and RF for the entire cohort (n = 87), was excellent with ICC ranging from 0.92–0.99 (Table 2). Values were not attenuated when examining each group separately, showing good to excellent ICC values for ICU (ICC = 0.91–0.98) and lung cancer (ICC = 0.85–0.99) images. MyoVision-US demonstrates excellent agreement with manual analysis having an adjusted R² values ranging from 0.76–0.94 in the images of the entire cohort. These values were not considerably lower for ICU (adjusted R² = 0.68–0.91) and lung cancer (adjusted R² = 0.55–0.99) images.

The consistency between MyoVision-US and manual analyses for TA images of the entire cohort (n = 85) was excellent with ICC ranging from 0.96–0.99 (Table 2). Values were not reduced when examining each sub-cohort separately, showing excellent ICC values for ICU (ICC = 0.96–0.99) and healthy (ICC = 0.94–0.99) individuals. MyoVision-US demonstrates excellent agreement to the manual analysis with adjusted R² values ranging from 0.85–0.96 in the images of the entire cohort. These values were not considerably attenuated for ICU (adj. R² = 0.84–0.97), where the comparability of MyoVision-US for EI was slightly higher for ICU images (adj. R² = 0.97) compared to healthy (adj. R² = 0.95) images.

Coefficient of variation (CV%) was calculated between manual and MyoVision-US measurements to quantify the relative variability between methods. The CV% values (ranging from 18 to 48%) represent the standard deviation of the differences between methods divided by the mean of all measurements, expressed as a percentage. These values indicate the degree of random error between measurement methods, with lower values representing better agreement. The highest variability was observed in the ICU cohort (CV% = 48% for RF CSA), suggesting that critically ill patients’ muscle characteristics present the greatest challenge for consistent measurements between methods. Regardless of muscle and population group, the comparability by muscle parameter was excellent for EI (adj. R²: 0.94–0.96), good to excellent for thickness (adj R² = 0.76–0.99), and moderate to good for CSA (adj R² = 0.85–0.87) values (Table 2). To facilitate improved utilization of the automated analysis, we developed a simple user interface that allows the upload of an ultrasound image to be

	RF mT (cm)	VI mT (cm)	QC mT (cm)	RF CSA (cm ²)	RF EI (a.u.)	TA mT (cm)	TA CSA (cm ²)	TA EI (a.u.)
Overall								
Count	90	90	90	90	90	90	90	90
Mean	1.1	1.2	2.6	3.8	90.7	2.12	7.4	88.0
±SD	0.4	0.4	0.9	1.6	27.6	0.4	1.7	22.2
Healthy								
Count	30	30	30	30	30	30	30	30
Mean	1.4	1.5	3.3	4.9	77.5	2.4	7.9	81.4
±SD	0.3	0.4	0.6	1.2	24.8	0.3	1.2	18.5
ICU								
Count	30	30	30	30	30	60	60	60
Mean	0.9	1.0	2.0	2.5	100.8	2.1	7.1	91.3
±SD	0.4	0.3	0.8	1.2	28.0	0.5	1.8	23.2
Lung cancer								
Count	30	30	30	30	30			
Mean	1.1	1.1	2.4	3.9	93.8			
±SD	0.4	0.3	0.7	1.4	25.2			

Table 1. Descriptive statistics of muscle parameters. RF = rectus femoris, QC = quadriceps complex (rectus femoris + vastus intermedius), TA = tibialis anterior, mT = muscle thickness, CSA = anatomical cross-sectional area, EI = echo intensity, cm = centimeters, a.u. = arbitrary unit, SD = standard deviation.



analyzed (Fig. 3). On the left-hand side, the original image can be seen with the accompanying segmentation by the AI model on the right-hand side. We demonstrate the multiclass segmentation capabilities of the models under development and include all anatomical features within the same image, such as the skin, adipose tissue, muscles, and bone. The model and MyoVision-US software (version 1.0) are available online for demonstration purposes through https://huggingface.co/spaces/ari10/MyoVision-US_Demo.

In ablation studies with performances on the validation set, DeepLabV3 with a ResNet-50 backbone performed the best for QC images with an increase of 0.007 Dice coefficient and 0.006 mIoU over the next best model architecture while SegFormer with an MIT-B2 backbone performed the best for TA images with an increase of 0.001 Dice coefficient and 0.002 mIoU over the next best model architecture as shown in Table 3. Strong color augmentations, binary-cross entropy loss, and a training time of 40 epochs were found to perform best across QC and TA images as well (Table 3). However, while SegFormer performs better than DeepLabV3 on

Fig. 2. Comparison of automated analysis by MyoVision-US and manual analysis of ultrasound images of quadriceps complex and tibialis anterior. **A–B.** Ultrasound images with and without the free-form trace labeled shadow from MyoVision-US analysis of ultrasound images of the quadriceps complex and tibialis anterior from healthy individuals and patients in the intensive care unit. In ultrasound images of the quadriceps complex, the rectus femoris (red), vastus intermedius (blue), and femur (green) are shown, and in ultrasound images of the tibialis anterior, the tibialis anterior (blue) is shown. **C–D.** Boxplots of rectus femoris muscle thickness, cross-sectional area, and echo intensity values of ultrasound images from healthy individuals (black), patients in the intensive care unit (magenta), and patients with lung cancer (green). **E–F.** Scatter plot with regression line plotting values for automated analysis by MyoVision-US against manual analysis of ultrasound images for muscle thickness, cross-sectional area, and echo intensity of rectus femoris (left) and tibialis anterior (right). **G–H.** Bland–Altman plots showing absolute differences for all measures comparing the manual and automated methods. Dashed lines indicate 95% confidence intervals. All correlations had p values < 0.0001 . ICU = intensive care unit, RF = rectus femoris, TA = tibialis anterior, EI = echo intensity, CSA = anatomical cross-sectional area, cm = centimeters.

the validation set, it is somewhat similar to the training dataset due to being subsets of the same dataset picked out for training the model. As SegFormer is a transformer-based architecture and transformers tend to require a relatively larger number of images to generalize than convolutional neural networks like DeepLabV3, SegFormer and DeepLabV3's performances on the more diverse test dataset for TA images are compared in Table 4. Our suspicions on the generalizability of SegFormer are confirmed as SegFormer performs worse than DeepLabV3 on the more varied TA images in the test dataset, and thus, DeepLabV3 with a ResNet-50 backbone is utilized for TA images in MyoVision-US.

Discussion

MyoVision-US was developed as an AI-driven software for muscle ultrasound image analysis, presenting substantial time savings compared to manual analysis by experts. MyoVision-US and manual analysis had moderate to excellent consistency for QC, RF, VI, and TA ultrasound images of healthy individuals as well as patients with acute and chronic diseases. Furthermore, MyoVision-US demonstrated strong agreement when compared directly to the manual analyses by experts, even for challenging images from patients in the ICU and with lung cancer.

The current description of MyoVision-US implementation represents a very superficial application of the newest AI technologies. Previous models and architecture requiring thousands of training images are being overtaken by newer methodologies requiring a tenth that number for training with improved results^{22,23,25,27,28,53}. Our present work tests the newest implementations, provides the scripts and approach as a resource to the community, and a user interface that allows non-computational scientists and clinicians to readily apply our models to real-world applications without any underlying technical expertise in the field of AI. Additional development will allow for not only multiclass segmentation from the same image (i.e., delineating all anatomical features, including muscle, bone, skin, adipose, and fascia) but also allow for real-time integration into ultrasound devices. AI-powered precision guidance systems will facilitate efficient and accurate ultrasound image acquisition and open the doors to greater utilization of bedside ultrasounds, which is demanding from an operator training perspective. The generalizability of such AI algorithms for ultrasonography is highly promising for multiple organ systems in addition to and in conjunction with muscle applications⁵⁴.

Using a deep learning image segmentation model known for performance, this study presents a relevant opportunity to improve the analysis of muscle ultrasound images with clinical and research applications. While previous studies have shown excellent reliability for muscle parameters using deep learning approaches in ultrasound images^{27,28,55}, the present study adds findings with excellent inter-rater consistency for populations with challenging muscle characteristics due to acute and chronic illness. This allowed the spectrum of muscle parameters to be increased, covering high and low values of thickness, CSA, and muscle EI. Muscle quality loss in acute and chronic diseases challenges the quantification of muscle parameters necessary for decision making in clinical practice⁹. In the present study, MyoVision-US demonstrated excellent consistency and comparability compared to expert manual analysis. In some challenging images with poor tissue delineation due to edema or adiposity, manual analysts were unable to confidently identify muscle boundaries, while the AI model, having been trained on a diverse dataset including similar difficult cases, was still able to propose segmentations. Such instances may be useful in populations with severe loss of muscle health (i.e., critical illness and lung cancer)³. However, this does not necessarily mean these AI segmentations are more accurate than what an expert could achieve with additional time or imaging options. Rather, the AI offers consistency and reproducibility in approach. We recommend that particularly challenging images still receive expert verification, especially when clinical decisions depend on these measurements. Therefore, like many automation tools, MyoVision-US has the potential to enhance scientific rigor by reducing human bias while simultaneously reducing the time required to analyze muscle ultrasound images. We reported that there is a significantly higher muscle EI in grayscale units in two different disease states. Marzola et al. described findings that were consistent with ours and further supports the robustness of our models and approach⁵³. Marzola et al. further used a z-score approach that utilizes deviations in EI to classify healthy from diseased muscle tissues. This approach, although potentially very interesting, is not representative of standard of care or a defined diagnostic criterion currently. Research remains ongoing into the nature of muscle changes in various diseases, and EI does not represent a uniform response for muscle in disease.

	RF mT (cm)	VI mT (cm)	QC mT (cm)	RF CSA (cm ²)	RF EI (a.u.)	TA mT (cm)	TA CSA (cm ²)	TA EI (a.u.)
Overall								
Mean difference	0.1	0.2	0.1	0.5	4.0	0.12	0.65	2.94
SD	0.1	0.2	0.2	0.4	5.0	0.12	0.51	2.98
CV (%) ¹	39	34	34	43	30	18	21	25
Mean bias error ²	0.04	0.00	−0.10	−0.16	−0.07	−0.04	−0.60	−1.13
ICC (95% CI) ³	0.97 (0.95–0.98)	0.92 (0.88–0.95)	0.98 (0.97–0.99)	0.97 (0.95–0.98)	0.99 (0.98–0.99)	0.96 (0.95–0.98)	0.958 (0.94–0.97)	0.991 (0.98–0.995)
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Adjusted R ⁴	0.89	0.76	0.93	0.87	0.94	0.87	0.85	0.96
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SEM	0.2	0.2	0.2	0.6	6.3	0.16	0.64	4.16
Healthy								
Mean difference	0.1	0.2	0.1	0.4	1.8	0.11	0.51	3.34
SD	0.0	0.2	0.2	0.3	2.0	0.12	0.37	3.34
CV (%) ¹	24	24	20	25	32	13	15	23
Mean bias error ²	0.00	0.11	−0.13	−0.25	−0.82	−0.03	−0.45	−2.66
ICC (95% CI) ³	0.99 (0.98–0.99)	0.91 (0.81–0.96)	0.99 (0.99–1.00)	0.96 (0.92–0.98)	0.99 (0.99–1.00)	0.94 (0.87–0.971)	0.96 (0.92–0.98)	0.986 (0.969–0.993)
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Adjusted R ⁴	0.97	0.67	0.99	0.85	0.99	0.80	0.87	0.95
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SEM	0.1	0.2	0.1	0.5	2.5	0.14	0.42	4.28
ICU								
Mean difference	0.1	0.2	0.2	0.6	5.5	0.12	0.72	2.74
SD	0.1	0.3	0.4	0.5	5.8	0.12	0.56	2.78
CV (%) ¹	45	24	38	48	28	20	23	25
Mean bias error ²	0.03	−0.11	−0.07	0.11	0.63	−0.05	−0.67	−0.34
ICC (95% CI) ³	0.95 (0.88–0.98)	0.92 (0.82–0.96)	0.93 (0.86–0.97)	0.91 (0.79–0.96)	0.98 (0.95–0.99)	0.97 (0.95–0.98)	0.954 (0.92–0.97)	0.99 (0.987–0.996)
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Adjusted R ⁴	0.80	0.75	0.76	0.68	0.91	0.89	0.84	0.97
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SEM	0.2	0.2	0.4	0.7	8.1	0.16	0.72	3.91
Lung cancer								
Mean difference	0.1	0.1	0.1	0.4	5.0			
SD	0.1	0.2	0.1	0.3	5.7			
CV (%) ¹	36	28	28	37	27			
Mean bias error ²	0.07	0.00	−0.09	−0.31	0.09			
ICC (95% CI) ³	0.95 (0.90–0.98)	0.85 (0.68–0.93)	0.996 (0.99–0.998)	0.98 (0.96–0.99)	0.98 (0.95–0.99)			
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001			
Adjusted R ⁴	0.83	0.55	0.99	0.93	0.92			
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	<0.001			
SEM	0.2	0.2	0.1	0.4	7.1			

Table 2. Consistency and comparability of MyoVision-US with manual ultrasound analyses. RF = rectus femoris. QC = quadriceps complex (rectus femoris + vastus intermedius). TA = tibialis anterior. mT = muscle thickness. CSA = anatomical cross-sectional area. EI = echo intensity. cm = centimeters. a.u. = arbitrary unit. SD = standard deviation, CV = coefficient of variation, ICC = intraclass correlation coefficient, SEM = standard error of measurement (standard error of the estimate). ¹CV% is the coefficient of variations for the MyoVision-US values. ²Mean bias error is the average of differences between Manual and MyoVision-US values. ³Two-way random effect ICC model for consistency between Manual and MyoVision-US values. ⁴Adjusted R-squared is linear regression for the MyoVision-US based output predicting manual values.

This study has potential limitations that need to be addressed. First, the sample size could be viewed as a limitation according to the guidelines to radiology research on artificial intelligence⁵⁶. However, the inclusion of images from multiple sites encompassing images of healthy individuals and patients with critical illness and lung cancer allowed a wide range of muscle parameters to be covered. Second, the findings of this study could only be generalized to adult healthy individuals or adult patients with critical illness and lung cancer. However, studies on populations with similar muscle conditions could also benefit from automating muscle image analysis. Third, although this study did not describe the accuracy result of each iteration during the development of the MyoVision-US, images were randomly selected and blinded for manual/automated analysis

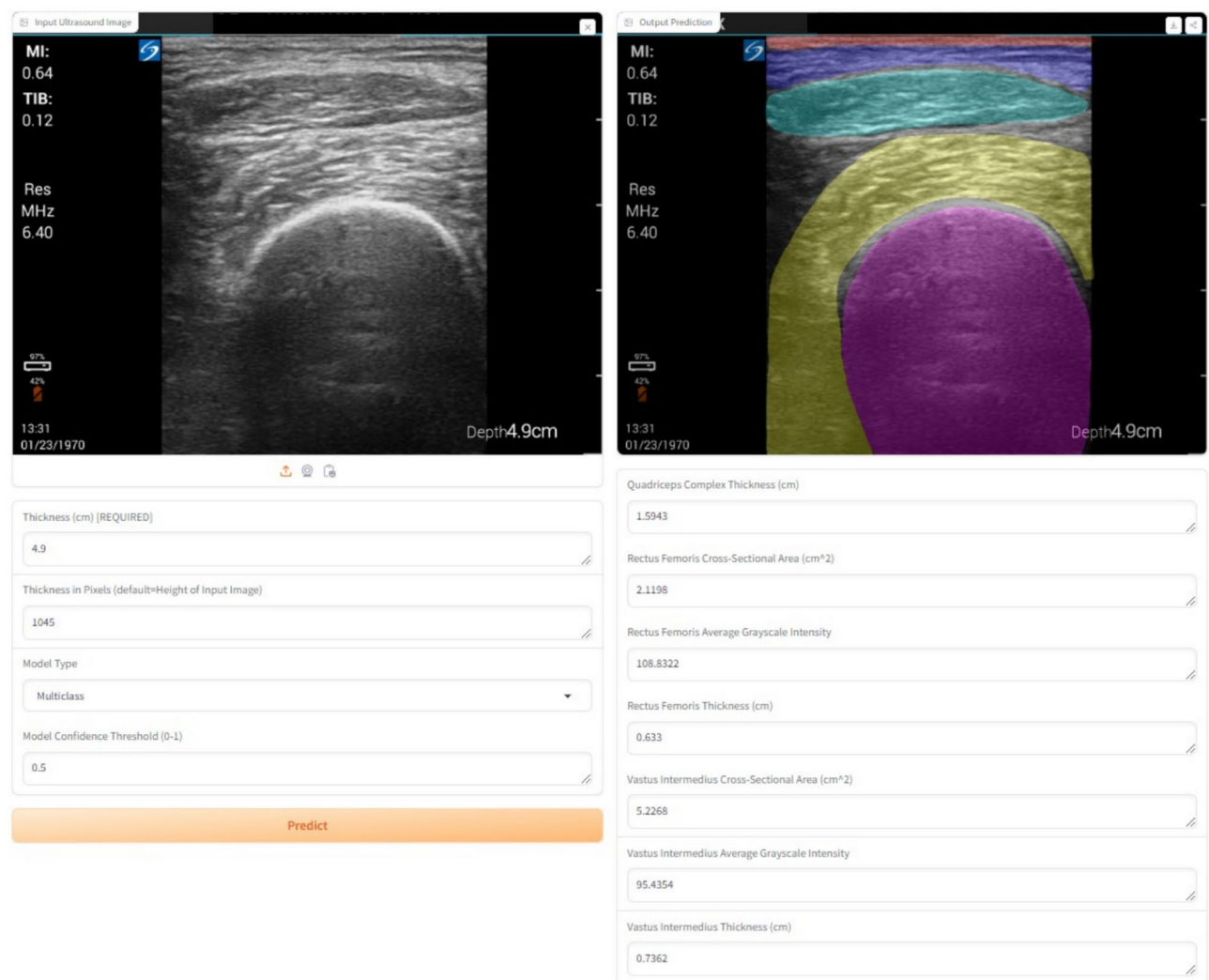


Fig. 3. MyoVision-US graphical user interface hosted on HuggingFace, a commonly used platform for building, training, and deploying machine learning models, with results of multiclass analysis and measurements shown. Skin (red), adipose tissue (blue), rectus femoris (yellow), vastus intermedius (aqua), and femur (pink) are segmented and shown.

Epochs	Strong color augmentations	Backbone architecture	Decoder architecture	Parameters (Millions)	Loss function	QC Dice	QC mIoU	TA Dice	TA mIoU
40	Yes	ResNet-50	DeepLabV3	39.6	BCE	0.900	0.881	0.949	0.904
40	Yes	ResNet-50	UNet	32.5	BCE	0.893	0.875	0.944	0.895
40	Yes	MIT-B2	SegFormer	24.7	BCE	0.883	0.866	0.956	0.916
40	Yes	MIT-B3	SegFormer	44.6	BCE	0.880	0.860	0.955	0.914
20	Yes	ResNet-50*	DeepLabV3*	39.6*	BCE	0.881	0.861	0.946	0.898
80	Yes	ResNet-50*	DeepLabV3*	39.6*	BCE	0.884	0.865	0.943	0.893
40	No	ResNet-50*	DeepLabV3*	39.6*	BCE	0.897	0.876	0.950	0.905
40	Yes	ResNet-50*	DeepLabV3*	39.6*	Focal	0.883	0.866	0.935	0.878
40	Yes	ResNet-50*	DeepLabV3*	39.6*	Tversky	0.851	0.830	0.942	0.892
40	Yes	ResNet-50*	DeepLabV3*	39.6*	Jaccard	0.867	0.849	0.950	0.905
40	Yes	ResNet-50*	DeepLabV3*	39.6*	Lovasz	0.846	0.824	0.953	0.910
40	Yes	ResNet-50*	DeepLabV3*	39.6*	Dice	0.891	0.873	0.950	0.905

Table 3. Dice and mIoU scores of various ablations on the MyoVision-US AI model on the holdout validation set. QC = quadriceps complex (mean of specific measure, e.g. Dice, over rectus femoris, vastus intermedius, and femur). TA = tibialis anterior. mIoU = mean Intersection over Union. BCE = binary cross-entropy. *Indicates the backbone architecture and decoder architecture used for TA ablations were MIT-B2 and SegFormer while QC ablations were performed with ResNet-50 and DeepLabV3.

Backbone architecture	Decoder architecture	TA CSA ICC*	TA mT ICC*	TA EI ICC*	TA CSA adjusted R ²	TA mT adjusted R ²	TA EI adjusted R ²
ResNet-50	DeepLabV3	0.96 (0.94–0.97)	0.96 (0.95–0.98)	0.99 (0.98–0.995)	0.85	0.87	0.96
MIT-B2	SegFormer	0.93 (0.89–0.95)	0.95 (0.93–0.97)	0.99 (0.98–0.99)	0.75	0.83	0.95

Table 4. Performance comparison between SegFormer and DeepLabV3 on the TA test dataset. TA = tibialis anterior. CSA = anatomical cross-sectional area. mT = muscle thickness. EI = echo intensity. ICC = intraclass correlation coefficient. *Two-way random effect model for consistency.

showing excellent consistency with manual analysis. Fourth, the findings of this study are limited to QC and TA muscles, therefore, future reports should explore accuracy in other peripheral and respiratory muscles. Lastly, each experienced physiotherapist manually quantified a subsection of the ultrasound image set, which could suggest risk of operator bias. However, all three experts have been trained and practiced with similar ultrasound protocols for more than 8 years. While MyoVision-US can operate without direct human supervision during the analysis process, we recommend visual confirmation of the segmentation results as a quality control measure, particularly for clinical applications. In our experience, human verification can improve final measurements by identifying cases where anatomical landmarks were incorrectly segmented. For research applications with large datasets, automated analysis without supervision may be appropriate with the understanding that a small percentage (approximately 3–5% in our dataset) of images may require manual correction.

When considering the clinical relevance of measurement errors between manual and automated analyses, it is important to contextualize the magnitude of these differences. The mean differences between methods were 0.1–0.2 cm for muscle thickness, 0.5–0.65 cm² for CSA, and 3–4 a.u. for EI. Based on previously reported interrater differences in these parameters (~0.1 cm for thickness, ~1.0 cm² for CSA, and 5–8 a.u. for EI)³², the errors observed are on par with or below clinically meaningful thresholds. However, in patients with critical illness where changes may be more subtle and rapid, even these small differences could potentially impact clinical decision-making. Therefore, we recommend visual verification of automated measurements in high-stakes clinical applications until further validation studies are conducted.

The bedside measurement of muscle parameters has been progressively replaced by post-hoc analysis using software such as NIH Image J mainly to systematize parameters of muscle quantity and quality. However, these analyzes take time and are not exempt from human error. Therefore, the findings of this study become relevant for the interpretation of muscle quality and quantity parameters if adequate image acquisition is performed.

In conclusion, the application of AI to automate muscle ultrasound analyses showed improvements in speed with strong consistency and comparability compared with manual human analysis evaluating muscle thickness, CSA, and EI of QC and TA muscles. This study introduces a method to automate the analysis of lower limb muscle ultrasound images, which may be useful for future research and clinical applications in acute and chronic settings.

Data availability

Data is available through reasonable request to the corresponding authors.

Received: 23 August 2024; Accepted: 21 April 2025

Published online: 29 April 2025

References

- Parry, S. M., Burtin, C., Denehy, L., Puthucherry, Z. A. & Bear, D. Ultrasound evaluation of quadriceps muscle dysfunction in respiratory disease. *Cardiopul. Phys. Therapy J.* **30**, 15–23. <https://doi.org/10.1097/cpt.000000000000102> (2019).
- Mourtzakis, M., Parry, S., Connolly, B. & Puthucherry, Z. Skeletal muscle ultrasound in critical care: A tool in need of translation. *Ann. Am. Thorac. Soc.* **14**, 1495–1503. <https://doi.org/10.1513/AnnalsATS.201612-967PS> (2017).
- Fazzini, B. et al. The rate and assessment of muscle wasting during critical illness: A systematic review and meta-analysis. *Crit. Care* **27**, 2. <https://doi.org/10.1186/s13054-022-04253-0> (2023).
- Mayer, K. P. et al. Acute skeletal muscle wasting and dysfunction predict physical disability at hospital discharge in patients with critical illness. *Crit. Care* **24**, 637. <https://doi.org/10.1186/s13054-020-03355-x> (2020).
- Alqahtani, J. S. et al. Diagnostic and clinical values of non-cardiac ultrasound in COPD: A systematic review. *BMJ Open Respir. Res.* <https://doi.org/10.1136/bmjresp-2020-000717> (2020).
- Okura, K. et al. Objective physical activity level is associated with rectus femoris muscle echo-intensity in patients with chronic obstructive pulmonary disease. *Clin. Respir. J.* **16**, 572–580. <https://doi.org/10.1111/crj.13528> (2022).
- Rausch, V., Sala, V., Penna, F., Porporato, P. E. & Ghigo, A. Understanding the common mechanisms of heart and skeletal muscle wasting in cancer cachexia. *Oncogenesis* **10**, 1. <https://doi.org/10.1038/s41389-020-00288-6> (2021).
- Batt, J., Herridge, M. & dos Santos, C. Mechanism of ICU-acquired weakness: Skeletal muscle loss in critical illness. *Intensive Care Med.* **43**, 1844–1846. <https://doi.org/10.1007/s00134-017-4758-4> (2017).
- Parry, S. M. et al. Ultrasonography in the intensive care setting can be used to detect changes in the quality and quantity of muscle and is related to muscle strength and function. *J. Crit. Care* **30**(1151), e1159–e1114. <https://doi.org/10.1016/j.jccr.2015.05.024> (2015).
- Pardo, E. et al. Reliability of ultrasound measurements of quadriceps muscle thickness in critically ill patients. *BMC Anesthesiol* **18**, 205. <https://doi.org/10.1186/s12871-018-0647-9> (2018).
- Selva Raj, I., Bird, S. R. & Shield, A. J. Ultrasound Measurements of Skeletal Muscle Architecture Are Associated with Strength and Functional Capacity in Older Adults. *Ultrasound Med. Biol.* **43**, 586–594. <https://doi.org/10.1016/j.ultrasmedbio.2016.11.013> (2017).
- Oranchuk, D. J., Bodkin, S. G., Boncella, K. L. & Harris-Love, M. O. Exploring the associations between skeletal muscle echogenicity and physical function in aging adults: A systematic review with meta-analyses. *J. Sport Health Sci.* <https://doi.org/10.1016/j.jshs.2024.05.005> (2024).

13. Leigheb, M. et al. Sarcopenia diagnosis: reliability of the ultrasound assessment of the tibialis anterior muscle as an alternative evaluation tool. *Diagnostics* **11**, 2158 (2021).
14. Nijholt, W., Scafoglieri, A., Jager-Wittenaar, H., Hobbelen, J. S. M. & van der Schans, C. P. The reliability and validity of ultrasound to quantify muscles in older adults: A systematic review. *J. Cachexia Sarcopenia Muscle* **8**, 702–712. <https://doi.org/10.1002/jcsm.12210> (2017).
15. Paramalingam, S. et al. Muscle B mode ultrasound and shear-wave elastography in idiopathic inflammatory myopathies (SWIM): Criterion validation against MRI and muscle biopsy findings in an incident patient cohort. *BMC Rheumatol* **6**, 47. <https://doi.org/10.1186/s41927-022-00276-w> (2022).
16. Mayer, K. P. et al. Construct and criterion validity of muscle ultrasonography for assessment of skeletal muscle in patients recovering from COVID-19. *Front. Physiol.* **14**, 1231538. <https://doi.org/10.3389/fphys.2023.1231538> (2023).
17. Casey, P. et al. The current use of ultrasound to measure skeletal muscle and its ability to predict clinical outcomes: A systematic review. *J. Cachexia Sarcopenia Muscle* **13**, 2298–2309. <https://doi.org/10.1002/jcsm.13041> (2022).
18. Heckmatt, J. Z., Leeman, S. & Dubowitz, V. Ultrasound imaging in the diagnosis of muscle disease. *J. Pediatr.* **101**, 656–660. [https://doi.org/10.1016/s0022-3476\(82\)80286-2](https://doi.org/10.1016/s0022-3476(82)80286-2) (1982).
19. Whittaker, J. L. et al. Imaging with ultrasound in physical therapy: What is the PT's scope of practice? A competency-based educational model and training recommendations. *Br. J. Sports Med.* **53**, 1447–1453. <https://doi.org/10.1136/bjsports-2018-100193> (2019).
20. Gonzalez-Seguel, F. et al. Evaluating a muscle ultrasound education program: Theoretical knowledge, hands-on skills, reliability, and satisfaction of critical care physiotherapists. *Arch. Rehabil. Res. Clin. Transl.* **3**, 100142. <https://doi.org/10.1016/j.arct.2021.100142> (2021).
21. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444. <https://doi.org/10.1038/nature14539> (2015).
22. Konnaris, M. A. et al. Computational pathology for musculoskeletal conditions using machine learning: Advances, trends, and challenges. *Arthritis. Res. Ther.* **24**, 68. <https://doi.org/10.1186/s13075-021-02716-3> (2022).
23. Katakis, S. et al. Automatic Extraction of Muscle Parameters with Attention UNet in Ultrasonography. *Sensors (Basel)* **22**, <https://doi.org/10.3390/s22145230> (2022).
24. Ritsche, P. et al. Fully automated analysis of muscle architecture from B-mode ultrasound images with DL_Track_US. *Ultrasound Med. Biol.* **50**, 258–267. <https://doi.org/10.1016/j.ultrasmedbio.2023.10.011> (2024).
25. Katakis, S. et al. Muscle cross-sectional area segmentation in transverse ultrasound images using vision transformers. *Diagnostics* **13**, 217 (2023).
26. Caresio, C., Salvi, M., Molinari, F., Meiburger, K. M. & Minetto, M. A. Fully automated muscle ultrasound analysis (MUSA): Robust and accurate muscle thickness measurement. *Ultrasound Med. Biol.* **43**, 195–205. <https://doi.org/10.1016/j.ultrasmedbio.2016.08.032> (2017).
27. Tayfur, B. et al. Automatic segmentation of quadriceps femoris cross-sectional area in ultrasound images: Development and validation of convolutional neural networks in people with anterior cruciate ligament injury and surgery. *Ultrasound Med. Biol.* **51**, 364–372. <https://doi.org/10.1016/j.ultrasmedbio.2024.11.004> (2025).
28. Ritsche, P. et al. DeepACSA: Automatic segmentation of cross-sectional area in ultrasound images of lower limb muscles using deep learning. *Med. Sci. Sports Exerc.* **54**, 2188–2195. <https://doi.org/10.1249/mss.0000000000003010> (2022).
29. Wen, Y. et al. MyoVision: software for automated high-content analysis of skeletal muscle immunohistochemistry. *J. Appl. Physiol. (Bethesda, Md. 1985)* **124**, 40–51. <https://doi.org/10.1152/japplphysiol.00762.2017> (2018).
30. Viggars, M. R., Wen, Y., Peterson, C. A. & Jarvis, J. C. Automated cross-sectional analysis of trained, severely atrophied, and recovering rat skeletal muscles using MyoVision 2.0. *J. Appl. Physiol. (Bethesda, Md. 1985)* **132**, 593–610. <https://doi.org/10.1152/japplphysiol.00491.2021> (2022).
31. Mokkink, L. B. et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J. Clin. Epidemiol.* **63**, 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006> (2010).
32. Mayer, K. P. et al. Interrater reliability of muscle ultrasonography image acquisition by physical therapists in patients who have or who survived critical illness. *Phys. Ther.* **100**, 1701–1711. <https://doi.org/10.1093/ptj/pzaa068> (2020).
33. Mayer, K. P. et al. Physical function measured prior to lung transplantation is associated with posttransplant patient outcomes. *Transplant Proc.* **53**, 288–295. <https://doi.org/10.1016/j.transproceed.2020.07.022> (2021).
34. Sarwal, A. et al. Interobserver reliability of quantitative muscle sonographic analysis in the critically ill population. *J. Ultrasound Med. Biol.* **34**, 1191–1200. <https://doi.org/10.7863/ultra.34.7.1191> (2015).
35. Ntoumenopoulos, G., Parry, S. M. & Neindre, A. L. Impact of an intensive education programme of diagnostic lung and lower limb ultrasound on physiotherapist knowledge: A pilot study. *Australas J. Ultrasound Med.* **21**, 104–114. <https://doi.org/10.1002/ajum.12089> (2018).
36. Lambell, K. J. et al. Comparison of ultrasound-derived muscle thickness with computed tomography muscle cross-sectional area on admission to the intensive care unit: A pilot cross-sectional study. *JPEN J. Parenter. Enteral. Nutr.* **45**, 136–145. <https://doi.org/10.1002/jpen.1822> (2021).
37. El-Ansary, D. et al. Architectural anatomy of the quadriceps and the relationship with muscle strength: An observational study utilising real-time ultrasound in healthy adults. *J. Anat.* **239**, 847–855. <https://doi.org/10.1111/joa.13497> (2021).
38. Mayer, K. P. et al. Muscle power is related to physical function in patients surviving acute respiratory failure: A prospective observational study. *Am. J. Med. Sci.* **361**, 310–318. <https://doi.org/10.1016/j.amjms.2020.09.018> (2021).
39. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675. <https://doi.org/10.1038/nmeth.2089> (2012).
40. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint [arXiv:1706.05587](https://arxiv.org/abs/1706.05587) (2017).
41. He, K., Zhang, X., Ren, S. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.
42. Ronneberger, O., Fischer, P. & Brox, T. 234–241 (Springer International Publishing).
43. Xie, E. et al. in *Proceedings of the 35th International Conference on Neural Information Processing Systems* Article 924 (Curran Associates Inc., 2021).
44. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. 379–387 (Springer International Publishing).
45. Berman, M., Triki, A. R. & Blaschko, M. B. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4413–4421.
46. Lin, T. Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 318–327. <https://doi.org/10.1109/TPAMI.2018.2858826> (2020).
47. Buslaev, A. et al. Albumentations: Fast and flexible image augmentations. *Information* **11**, 125 (2020).
48. Deng, J. et al. in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.
49. Loshchilov, I. & Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint [arXiv:1608.03983](https://arxiv.org/abs/1608.03983) (2016).
50. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. arXiv preprint [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) (2017).
51. Zou, K. H. et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* **11**, 178–189. [https://doi.org/10.1016/s1076-6332\(03\)00671-8](https://doi.org/10.1016/s1076-6332(03)00671-8) (2004).
52. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012> (2016).

53. Marzola, F., van Alfen, N., Doorduyn, J. & Meiburger, K. M. Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment. *Comput. Biol. Med.* **135**, 104623. <https://doi.org/10.1016/j.combiomed.2021.104623> (2021).
54. Levy, B. E. et al. Artificial intelligence evaluation of focused assessment with sonography in trauma. *J. Trauma Acute. Care Surg.* **95**, 706–712. <https://doi.org/10.1097/TA.0000000000004021> (2023).
55. Nhat, P. T. H. et al. Clinical evaluation of AI-assisted muscle ultrasound for monitoring muscle wasting in ICU patients. *Sci. Rep.* **14**, 14798. <https://doi.org/10.1038/s41598-024-64564-w> (2024).
56. Bluemke, D. A. et al. Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers—from the radiology editorial board. *Radiology* **294**, 487–489. <https://doi.org/10.1148/radiol.2019192515> (2020).

Author contributions

AHS and YW designed and developed the MyoVision-US software. ZCR, FGS, AHS, SMP, KPM, and YW performed data analysis, drafted the figures and manuscript. KPM, YW and SD generated training datasets with annotations and labels. FGS, SMP, and KPM generated the validation datasets by measuring muscle regions and parameters. ZCR, FGS, CG, AS, SD, GN, SMP, and KPM participated in image acquisition and critically revised the manuscript. JC and VKCB provided technical and AI expertise. SMP, KPM, and YW were responsible for project and study design; curation of images for software design and analysis; and coordination of data collection and analysis, and overseeing the drafting of the figures and manuscript. All authors critically revised and approved the final manuscript.

Funding

The project was supported by the AIMS—Artificial intelligence in Medicine Alliance of the University of Kentucky and by the NIH National Center for Advancing Translational Sciences through grant number UL1TR001998 to YW, KPM, and SD. KPM and YW were supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institute of Health K23-AR079583 and R00-AR081367, respectively. SMP is supported by the Al and Val Rosenstrauss Fellowship from the Rebecca L. Cooper Foundation.

Declarations

Competing interests

YW is the founder of MyoAnalytics LLC. The remaining authors have no conflicts of interest.

Ethical approval

University of Kentucky Medical Expedited Internal Review Board (IRB) #75185 approved this study in February 18, 2022.

Consistent publication

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Additional information

Correspondence and requests for materials should be addressed to K.P.M. or Y.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025