Musclesense: a trained, artificial neural network for the anatomical segmentation of lower limb magnetic resonance images in neuromuscular diseases

Baris Kanber^{1,2,3}, Jasper M Morrow³, Uros Klickovic³, Stephen Wastling^{3,4}, Sachit Shah^{3,4}, Pietro Fratta³, Amy R McDowell⁵, Matt G Hall^{5,6}, Chris A Clark⁵, Francesco Muntoni⁵, Mary M Reilly³, Michael G Hanna³, Daniel C Alexander^{2,7}, Tarek Yousry^{2,3,4}, John S Thornton^{2,3,4}

¹ Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK

² National Institute for Health Research Biomedical Research Centre at UCL & UCLH, London UK

³ UCL Queen Square Institute of Neurology, University College London, London, UK.

⁴ The Lysholm Department of Neuroradiology, University College London Hospitals NHS Foundation Trust, London, UK

⁵ UCL GOSH Institute of Child Health, University College London, London, UK

⁶ National Physical Laboratory, Teddington, UK

⁷ Centre for Medical Image Computing, Department of Computer Science, University College London, London, UK

Corresponding author: Baris Kanber, PhD, UCL Centre for Medical Image Computing, 90 High Holborn, London WC1V 6LJ, UK. E-mail: b.kanber@ucl.ac.uk. Telephone: +44 (0)20 7679 2000.

Keywords: Musclesense; MRI; lower limb; muscle; segmentation; neuromuscular diseases

Introduction

Neuromuscular diseases affect more than 17 million children and adults worldwide causing either premature death or lifelong disability¹. Recent scientific advances have resulted in life-changing gene therapies in conditions such as spinal muscular atrophy^{2,3} and Duchenne muscular dystrophy⁴. There are many ongoing therapy trials: in muscular dystrophies alone, there are currently 56 interventional studies actively recruiting⁵. However, assessing the efficacy of potential therapies in clinical trials, and monitoring treatment response in routine clinical practice, requires

¹ Deenen, J. C. W. et al. The Epidemiology of Neuromuscular Disorders: A Comprehensive Overview of the Literature. *J. Neuromuscul. Dis.* 2, 73–85 (2015).

² Finkel, R. S. et al. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N. Engl. J. Med.* 377, 1723–1732 (2017).

³ Mendell, J. R. et al. Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* 377, 1713–1722 (2017).

⁴ Voit, T. et al. Safety and efficacy of drisapersen for the treatment of Duchenne muscular dystrophy (DEMAND II): an exploratory, randomised, placebo-controlled phase 2 study. *Lancet Neurol.* 13, 987–996 (2014).

⁵ ClinicalTrials.gov. Available at: <https://clinicaltrials.gov/>. (Accessed: 7th September 2019)

reliable and sensitive biomarkers. One of the most responsive outcome measures longitudinally has been fat quantification with 3-point Dixon magnetic resonance imaging (MRI), which can accurately track intramuscular fat accumulation, and correlates strongly with disease severity and disability⁶. However, the current technology requires time-consuming manual segmentation of MRI scans, limiting the complexity of practically feasible analyses, and their use in the clinical setting: approximately 15 minutes per scan is required for an expert to draw accurate regions-of-interest on muscles bilaterally on a single slice at thigh and calf levels. Due to this limitation, data from a single axial slice has provided the primary outcome in many published studies even though data is usually acquired of multiple slices, and in more recent studies of entire muscle volumes. Finally, manual muscle segmentation is necessarily subjective and so is liable to inter-rater bias, reducing outcome measure responsiveness and study power.

Artificial-intelligence (AI)-enabled, automated segmentation offers a solution to these problems with orders-of-magnitude increases in speed, enabling more anatomically complete, complex analyses, and eliminating the inter-observer bias. Recent developments in AI, and increased computational power, provide a timely opportunity to introduce automated segmentation for clinical MRI biomarker extraction. In this article, we introduce Musclesense, a trained, artificial neural network that anatomically segments muscles from lower limb MRI images in neuromuscular diseases.

Methods

Musclesense employs a two-dimensional, multi-channel, deep, convolutional neural network in a U-net configuration⁶. It has been trained and validated with calf- and thigh-level, 3 point Dixon gradient-echo MRI data (Siemens 3T Prisma, TR = 125.0 ms, TE = 3.45/ 4.60/ 5.75 ms, NSA = 4, 11 slices, FOV= 180 mm, voxel size = $0.6 \times 0.6 \times 10.0 \text{ mm}^3$) from amyotrophic lateral sclerosis (ALS, n=21), spinal bulbar muscular atrophy (KD, n=21), inclusion body myositis (IBM, n=20), Charcot-Marie-Tooth disease (CMT, n=20), and hypokalemic periodic paralysis (HYPOPP, n=12) patients, as well as 72 adult, control subjects. We used 571 MRI acquisitions in total: 310 of these were at the calf-level, while 261 were at the thigh-level. All datasets were manually labelled by experienced observers, identifying the muscle compartment boundaries using the open-source ITK-

⁶ Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, et al., editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 Cham: Springer International Publishing; 2015. p. 234–41. (Lecture Notes in Computer Science).

SNAP software⁷, on single, mid-level slices for the calf-level and thigh-level images. The manual segmentation at the thigh included all muscles at mid-thigh level except adductor longus, whilst at calf level included all muscles. Separated fat and water images were obtained with the conventional algorithm due to Glover and Schneider (1991)⁸.

A separate dataset of calf-, and thigh level MRI images from four paediatric Duchenne muscular dystrophy (DMD) patients, four paediatric control subjects and a healthy adult volunteer, obtained at a different centre, using a markedly different protocol (T2-weighted fat-saturated turbo-spin-echo sequence obtained on a Siemens 3T Prisma using a 15-channel knee coil, TR = 3500 ms, TE = 73 ms, NSA = 1, 4 slices, FOV= 160 mm, voxel size = $0.36 \times 0.36 \times 3.0$ mm³) was used for secondary validation. Separated fat and water images, in this dataset, were produced by Siemens proprietary software at the scanner.

Two editions of Musclesense were developed. The first edition was optimised for the 3-point Dixon sequence images of the main dataset and was trained on the fat, water, TE=3.45, and 5.75 ms Dixon MR images normalised to zero mean and unit standard deviation. The second, simplified edition of Musclesense was tailored for centres that use MRI acquisition protocols different to ours (including that of the paediatric cohort) and was trained only on the normalised fat images. In both cases, training data was augmented with random horizontal and vertical shifts (range $\pm 50\%$ for both), random scaling (range 70%-130%), horizontal, and vertical flips.

We measured segmentation performance as the spatial overlap between the manually and computationally segmented regions employing the Dice score coefficient (DSC) in a nested cross-validation setting. The available data was split (5-folds) into datasets D1 (80%) and D2 (20%), avoiding the selection of the same patient's data for both datasets. Nested within this, datasets D1 were further split (5-folds) into datasets D₃ (80%) and D₄ (20%). In each of the resulting 25 iterations, the networks were trained on datasets D₃, validating on D₄, and testing on D₂. For the paediatric cohort, for which no manual segmentations were available, the assessment was qualitative by means of visual inspection.

Results

⁷ Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31:1116–28.

⁸ Glover GH, Schneider E. Three-point Dixon technique for true water/fat decomposition with B0 inhomogeneity correction. *Magn Reson Med* 1991;18:371–83.

Network training took approximately 3.5 hours per fold at 320x160 voxel resolution on an NVIDIA GeForce GTX 1080 Ti GPU with 10481 MB of memory (NVIDIA, Santa Clara, CA, USA). Test-time speed of inference was approximately 30 milliseconds per MRI slice.

For the adult cohort images, Musclesense performed well in discriminating between muscle and non-muscle regions, even in images showing marked intra-muscular fat accumulation. The mean DSC, quantifying the degree of spatial overlap between the manual and Musclesense segmentations was 0.950 (standard deviation: 0.025) for the calf, and 0.942 (standard deviation: 0.041) for the thigh. Examples of fat and water images are shown in Figure 1, along with the anatomical segmentations obtained with Musclesense. In the secondary validation dataset, which had a markedly different acquisition sequence, different participant age-distribution, and a different method of fat/water map extraction, the obtained segmentations were of acceptable quality on visual inspection (Figure 2), albeit with poorer separation of intermuscular boundaries as might be expected due to the use of the only fat images for inference. It is also noted that that in the paediatric cohort, which included one adult volunteer, only one limb was scanned per acquisition.

In the adult cohort, 8 subjects had repeat scanning 2-weeks apart, and test-retest differences in calculated, mean muscle fat fractions were 0.65% (range: 0.05-1.84%), and 1.30% (range: 0.04-4.88%) for the calf, and the thigh, respectively (Table 1).

Figure 1 - Examples of fat (first column), and water (second column) images from the first dataset, along with the anatomical segmentations obtained with Musclesense (third column).

Figure 2 - Examples of fat (first column), and water (second column) images from the second dataset, along with the anatomical segmentations obtained with Musclesense (third column).

Calf imaging from an 8-year-old, male control subject

Calf imaging from an 8-year-old, male DMD patient

Calf imaging from an 8-year-old, male DMD patient

Calf imaging from a 40-year-old, female, healthy volunteer

Table 1 - Test-retest values of the mean, calf, and thigh fat fractions measured using the Musclesense segmentations obtained on scans of the same, four subjects, two weeks apart. Case types are inclusion body myositis (IBM), and control subject (CS).

Age (gender)	Case type	Test, mean, calf fat fraction (%)	Retest, mean, calf fat fraction (%)	Test, mean, thigh fat fraction (%)	Retest, mean, thigh fat fraction (%)
28 (male)	CS	3.09	3.04	1.85	1.89
27 (female)	CS	1.57	1.09	1.22	1.16
59 (male)	IBM	51.79	49.95	37.06	41.95
75 (male)	IBM	9.50	9.26	28.51	28.29

Summary

We present Musclesense, a trained, artificial neural network for the anatomical segmentation of MRI images in neuromuscular diseases. The network displayed value in segmenting muscles in the presence of a broad range of disease severity across our adult cohort data, as well as a promising example of generalisation to a different subject cohort with markedly different acquisition parameters. We believe it can help alleviate the need for tedious manual segmentation and enable the rapid assessment and monitoring of disease severity and treatment response in the clinical setting. More complicated analyses, including segmenting individual muscle compartments, as well as 3D segmentation using all MRI slices acquired in a patient visit, that are not practicable using the current, manual methods, may also be enabled. This will potentially lead to imaging biomarkers of greater sensitivity, in part by reducing sensitivity to single slice positioning errors. We continue the development of Musclesense to incorporate such capabilities, and this is expected to increase the reproducibility of the method beyond that reported in Table 1. Further refinement with larger training data sets is also expected to reduce the occasional regional misclassification errors seen with the current Musclesense versions. As with any AI-based approach, we recommend a manual quality control step, where errors may be identified and corrected.

The second, simplified edition of Musclesense, tailored for lower limb fat images obtained using generic MRI acquisition sequences is made freely available to the research community on [NiftyWeb](http://niftyweb.cs.ucl.ac.uk/) (<http://niftyweb.cs.ucl.ac.uk/>). The first, professional edition of Musclesense is also freely available, upon request from the corresponding author, for centres which are able to use our optimised MRI sequence and method of fat/water map extraction or equivalent.

Compliance with Ethical Standards

This study was approved by the joint University College London/University College London Hospitals NHS Foundation Trust (REC reference number 08/HO715/117), and Bromley Research Ethics

Committees (REC reference number 12/LO/0442). All subjects gave informed, written consent before participating in the study.

Information Sharing Statement

The second, simplified edition of Musclesense, tailored for lower limb fat images obtained using generic MRI acquisition sequences is made freely available to the research community on [NiftyWeb](http://niftyweb.cs.ucl.ac.uk/) (<http://niftyweb.cs.ucl.ac.uk/>). The first, professional edition of Musclesense is also freely available, upon request from the corresponding author, for centres which are able to use our optimised MRI sequence and method of fat/water map extraction or equivalent. The datasets used during development are not publicly available.

Acknowledgements

We are grateful to the National Institute for Health Research Biomedical Research Centre at University College London Hospitals and University College London, and the Wellcome Trust for funding our study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.