Automated Cartilage and Meniscus Segmentation on 7T T2* MRI for Quantitative Assessment of Knee Joint with Minimal Expert Refinement

Eisa Hedayati*, Karsten Knutsen, Abdul Wahed Kajabi, Collin Steinberger, Abhinav Lamba, Asif Abul Hassan, Hassan Ahad, Jutta Ellermann
University of Minnesota, Minneapolis, MN, * (heday015@umn.edu),

Disclosures: All authors have nothing to disclose.

INTRODUCTION: Accurate evaluation of meniscus and cartilage integrity is essential for understanding the onset and progression of knee osteoarthritis (OA), particularly in patients with medial meniscus posterior root tears (MMPRTs). Quantitative magnetic resonance imaging (MRI), including T2* relaxation time mapping, provides noninvasive biomarkers of collagen fiber organization and cartilage health [1]. However, these approaches rely heavily on precise tissue segmentation. Manual segmentation of cartilage and meniscus is labor-intensive, time-consuming, and prone to variability, and the challenge is amplified in T2* images due to limited tissue contrast and close anatomical proximity in the knee joint. Deep learning has enabled automated segmentation of musculoskeletal tissues, but applications to ultra–high-field 7T T2* imaging remain limited. Leveraging automated segmentation could facilitate reproducible and efficient quantitative mapping, expanding the utility of 7T MRI in orthopaedic research. The purpose of this study was to develop and evaluate a deep learning–based auto-segmentation framework for articular cartilage and menisci on 7T T2* MRI. We hypothesize that despite the inherent difficulty of T2* segmentation, automated methods can achieve sufficient accuracy to serve as a strong foundation for quantitative assessment, requiring only minimal refinement by an experienced annotator.

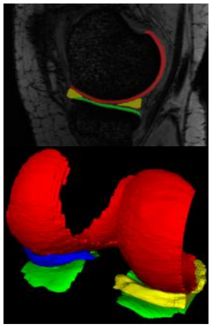

Figure 1. Sample prediction mask of a volunteer.

METHODS: This IRB-approved study included 19 patients with MMPRTs and 8 healthy volunteers. Manual segmentations of femoral and tibial cartilage as well as medial and lateral menisci were performed using ITK-SNAP by experienced annotators. Imaging was conducted on a 7T MRI system using a standardized protocol that included a high-resolution 3D multi-echo gradient recalled echo (GRE) sequence for T2* mapping (TR = 26 ms, seven echo times [3.1–21.4 ms], resolution = 0.4 × 0.4 × 1 mm³, scan time = 5:20 min). The dataset was divided into training, validation, and test sets. The independent test set consisted of 5 patients and 4 volunteers. The training set included 11 patients and 3 volunteers, while the validation set consisted of 3 patients and 1 volunteer. To reduce image size and improve network generalization, MRIs were down-sampled in the sagittal plane by selecting every other pixel, producing four down-sampled versions of each scan. Random cropping was then applied, reducing slices from 600×600 to 300×300 after down-sampling, and further to 256×256. Model inputs were 256×256×32×4 volumes, where 32 contiguous slices were randomly sampled from the ~120-slice stack, and four echoes were included. A batch size of 16 windows was used, with all possible 32-slice windows sampled once per epoch. This combination of down-sampling and window selection served as the primary augmentation strategy. A residual U-Net [1] architecture from MONAI [2] was employed. During validation, each MRI was divided into four sections using the same down-sampling strategy. Each section was zero-padded to 128 slices, split into four windows of 32 slices, and segmented independently. Predictions were merged to reconstruct the full MRI volume. All predictions were further refined using a seeded label re-assignment workflow to correct boundary label bleeding. Segmentation performance was evaluated using the Dice similarity coefficient [4], comparing predicted against manual ground truth masks. Automated and manual segmentation masks were applied to the T2* maps to extract relaxation times from femoral cartilage, tibial cartilage, and the medial and lateral menisci. For each tissue region, paired comparisons between manual and automated measurements were performed. Statistical analyses included paired t-tests, Wilcoxon signed-rank tests, and effect size calculations (Cohen's d). Multiple comparison correction was applied using false discovery rate (FDR) adjustment. The agreement between manual and automated measurements was further assessed with Bland–Altman analysis.

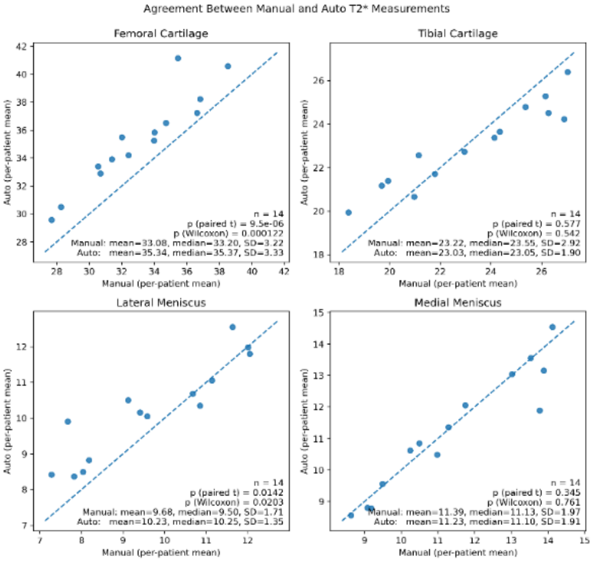

Figure 2. Bland-Altman plot for comparing automatic segmentation vs manual for all 4 different segmentation masks.

RESULTS: The residual U-Net achieved mean Dice scores of 0.73 across all tissues (femoral cartilage 0.80, tibial cartilage 0.78, lateral meniscus 0.64, medial meniscus 0.72). After applying the seeded label re-assignment workflow, overall Dice improved to 0.76, with the largest gain observed in the lateral meniscus (0.71 vs 0.64) and medial meniscus (0.77 vs 0.72). A predicted mask is depicted in Figure 1. Quantitative comparison of T2* values derived from manual versus automated segmentations demonstrated tissue-specific differences. Femoral cartilage values were significantly higher with automated segmentation (35.3 vs 33.1 ms, $p < 0.001$, dz = 1.87). Lateral meniscus T2* values were also modestly but significantly higher with automated masks (10.2 vs 9.7 ms, FDR-adjusted $p < 0.05$, dz = 0.76). No significant differences were observed between manual and automated measurements for tibial cartilage (23.0 vs 23.2 ms, $p > 0.5$) or medial meniscus (11.2 vs 11.4 ms, $p > 0.3$). Bland–Altman analysis confirmed strong agreement overall, with systematic bias observed in femoral cartilage and lateral meniscus (Figure 2).

DISCUSSION: Automated segmentation provided a strong baseline for quantitative 7T T2* analysis, but expert correction remains necessary. Femoral cartilage and lateral meniscus require closer attention, while tibial cartilage and medial meniscus showed good agreement with manual annotations and need less correction. Importantly, the 3D network was trained with only 14 subjects, demonstrating the feasibility of deep learning in very limited, high-quality datasets. These findings support a workflow where automated segmentation accelerates analysis while targeted expert refinement ensures reliability.

CLINICAL RELEVANCE: Automated segmentation on 7T T2* MRI can substantially reduce the burden of manual annotation in quantitative cartilage and meniscus assessment. While targeted correction by an experienced reader remains necessary, particularly for femoral cartilage and lateral meniscus, automated methods provide reliable results for tibial cartilage and medial meniscus. This workflow enables more efficient and scalable evaluation of joint health in patients with meniscal root tears and may facilitate broader adoption of quantitative MRI in orthopaedic research and clinical studies.

REFERENCES: **[1]** Kajabi, Abdul Wahed, et al. "Seven tesla knee MRI T2*-mapping detects intrasubstance meniscus degeneration in patients with posterior root tears." *Radiology Advances* 1.1 (2024): umae005. [2] Siddique, N., et al. "U-net and Variants for Medical Image Segmentation: Review." IEEE Access 9 (2021): 82031-82057. **[3]** Cardoso, M. J., et al. "Monai: Open-Source Deep Learning in Healthcare." arXiv 2211.02701 (2022). **[4]** Dice, L. R. "Ecologic Association Between Species." Ecology 26.3 (1945): 297-302.

ACKNOWLEDGEMENTS: National Institute of Biomedical Imaging and Bioengineering (P41 EB027061), S10 OD025256 and R01EB034575.

ORS 2026 Annual Meeting PaperNo.1211