Deep learning-based segmentation of knee MRI for fully automatic subregional morphological assessment of cartilage tissues: Data from the Osteoarthritis Initiative

Egor Panfilov1 | Aleksei Tiulpin1,2,3 | Miika T. Niemen1,2 | Simo Saarakkala1 | Victor Casula1

1Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland
2Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland
3Ailean Technologies Oy, Oulu, Finland

Correspondence
Egor Panfilov, Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Aapistie 5A, 90220 Oulu, Finland.
Email: egor.panfilov@oulu.fi

Funding information
Oulun Yliopisto; Sigrid Juséliuksen Säätiö; Jane ja Aatos Erkon Säätiö; KAUTE-Säätiö

Abstract
Morphological changes in knee cartilage subregions are valuable imaging-based biomarkers for understanding progression of osteoarthritis, and they are typically detected from magnetic resonance imaging (MRI). So far, accurate segmentation of cartilage has been done manually. Deep learning approaches show high promise in automating the task; however, they lack clinically relevant evaluation. We introduce a fully automatic method for segmentation and subregional assessment of articular cartilage, and evaluate its predictive power in context of radiographic osteoarthritis progression. Two data sets of 3D double-echo steady-state (DESS) MRI derived from the Osteoarthritis Initiative were used: first, \( n = 88 \); second, \( n = 600 \), 0-/12-/24-month visits. Our method performed deep learning-based segmentation of knee cartilage tissues, their subregional division via multi-atlas registration, and extraction of subregional volume and thickness. The segmentation model was developed and assessed on the first data set. Subsequently, on the second data set, the morphological measurements from our and the prior methods were analyzed in correlation and agreement, and, eventually, by their discriminative power of radiographic osteoarthritis progression over 12 and 24 months, retrospectively. The segmentation model showed very high correlation (\( r > 0.934 \)) and agreement (mean difference < 116 mm³) in volumetric measurements with the reference segmentations. Comparison of our and manual segmentation methods yielded \( r = 0.845–0.973 \) and mean differences = 262–501 mm³ for weight-bearing cartilage volume, and \( r = 0.770–0.962 \) and mean differences = 0.513–1.138 mm for subregional cartilage thickness. With regard to osteoarthritis progression, our method found most of the significant associations identified using the manual segmentation method, for both 12- and 24-month subregional cartilage changes. The method may be effectively applied in osteoarthritis progression studies to extract cartilage-related imaging biomarkers.

KEYWORDS
cartilage, deep learning, morphology, osteoarthritis, segmentation
1 | INTRODUCTION

Articular cartilage is often studied in vivo using magnetic resonance imaging (MRI) to understand osteoarthritis-related soft tissue changes in knee joints. Several MRI protocols, particularly, based on 3D double-echo steady-state (DESS) sequence, have been shown to provide a good contrast between the cartilage and the surrounding tissues. However, manual segmentation of cartilage is prone to resegmentation errors, particularly, due to inter-reader variability.1,2 Furthermore, manual delineation is time-demanding, which makes it impractical in large-scale longitudinal studies, such as osteoarthritis progression research. Thus, the development of automatic methods for cartilage segmentation that are fast, accurate, and consistent is crucial.3,4

Numerous studies have been performed to automate knee cartilage segmentation by applying image processing, conventional machine learning algorithms,5 and registration-based methods,6 yet with a limited success. Recently, methods based on deep learning (DL) have been evaluated in this task, showing unprecedented performance.7–9 However, the vast majority of DL-based studies performed validation against the manual segmentations at a tissue level only and without relation to the evolution of cartilage changes caused by osteoarthritis.

Earlier works have shown associations between the changes of cartilage morphology, quantified via total volume and thickness statistics, and the progression of osteoarthritis.10,11 Importantly, the morphological cartilage changes vary in relation to the particular knee anatomy and are heterogeneous,12 such that cartilage thinning in one subregion may be accompanied by swelling in another. Moreover, even the healthy cartilage tissues have complex geometry and non-uniform thickness profiles. Assessment of the cartilage segmentation accuracy on a tissue level does not account for the relative importance of the subregional errors and, therefore, has limited clinical value for understanding the disease.

Multiple studies have been conducted to identify the cartilage subregions mostly affected by osteoarthritis.13,14 Such subregions have been defined not only with respect to certain anatomical landmarks, but also considering the load dynamics within the joint. Several approaches have been proposed, with different levels of subregional granularity.15,16 To date, the most established and validated method for subregional cartilage assessment is still semi-automatic,16–19 and relies on the segmentation performed manually due to accuracy concerns.

In this study, we combined a DL-based cartilage segmentation with a nonrigid registration-based subregional division. Using this pipeline, we analyzed the subregional performance of the DL-based cartilage segmentation and investigated whether it could effectively substitute the manual delineation process in osteoarthritis progression research.

2 | MATERIALS AND METHODS

2.1 | Data

We performed a retrospective analysis using two data sets drawn from the Osteoarthritis Initiative (OAI; https://nda.nih.gov/oai), a multi-center prospective cohort study database (Table 1). Ethical approval and informed consent for all participants were obtained by OAI. The first data set (IMO) included 88 subjects from the OAI Progression cohort: sagittal 3D DESS knee MR scans from baseline and 12-month follow-up visits. The data set had manual annotations for femoral, tibial, patellar cartilage, and menisci produced by IMorphics Ltd.20 The second data set (FBC) included 600 subjects from the FNIH Osteoarthritis Biomarkers Consortium data,19 nested case–control study in OAI. Sagittal 3D DESS MR scans were obtained from the baseline, 12-month, and 24-month visits. Subjects overlapping with the training subset of IMO (n = 13), and subjects with missing assessments for any of the visits (n = 20) were excluded from FBC during the analysis. The data set comprised of four groups of subjects: controls (GC), radiographic progressors (GR), pain progressors (GP), radiographic and pain progressors (GRP). The demographics of the final data set and the groups are summarized in Table 1. We refer to Eckstein et al.19 for the detailed definition of the progressor groups. In short, radiographic progression was defined as the loss of minimum joint space width of 0.7 mm or more from baseline to 24, 36, or 48 months. Pain progression was defined as a persistent increase of 9 or more points in normalized 0–100 pain score (quantified using Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC]) from baseline to 24, 36, 48, or 60 months. All images were acquired with 3T Siemens MAGNETOM Trio scanners and quadrature transmit–receive knee coils (USA Instruments). Sagittal 3D DESS sequence was used (160 slices; voxel size: 0.37 × 0.37 × 0.7 mm, matrix: 384 × 384, field of view (FOV): 140 mm; repetition time (TR): 16.3 ms, echo time (TE): 4.7 ms, flip angle: 25°).

IMO data included the reference annotations for femoral, tibial, patellar cartilage tissues, and menisci.

2.2 | Segmentation

The proposed method performed segmentation of the cartilage tissues, their subsequent subregional division, and quantification of subregional morphological properties (Figure 1). For segmentation, we employed a DL-based approach. Here, IMO data were split subject-wise into training (n = 70, KL1: 4/48/78/8 knees) and test (n = 18, KL1: 0/11/21/4 knees) subsets, maintaining the similar distribution of radiographic osteoarthritis severity, measured by the Kellgren–Lawrence (KL) system. The training subset was used in 5-fold cross-validation scheme to train five segmentation models, further used in a single ensemble. The models were trained to segment femoral, tibial, and patellar cartilage tissues, as well as menisci in 2D, from sagittal slices. During the evaluation phase, continuous model-wise outputs of softmax activation layers were averaged over the ensemble and the dominant class was identified for each voxel. The architecture was based on UNet21 with an ImageNet-pretrained VGG19 encoder,22 and modified to have six levels in the decoder. Further regularization was done using mixup,23 which was previously shown to improve segmentation robustness in the task.8 The segmentation quality was quantified on the test subset using the...
volumetric (Dice score coefficient [DSC]) and surface-based metrics (average symmetric surface distance [ASSD], Hausdorff distance [HD]). Additionally, these metrics were computed for each subregion (see below) to investigate the segmentation quality in relation to anatomical locations.

### 2.3 Subregional division

The cartilage tissues were then divided into the subregions using a multi-atlas registration. We adopted two sets of atlases (Figure 1). The first multi-atlas ($k=5$) was constructed according to Wirth and Eckstein.16 Specifically, the femoral cartilage was split into trochlear, central, and posterior sections, and the central sections were further divided into external, central, and internal subregions, total of 10. Tibial cartilage plates were divided into five subregions each: central, anterior, posterior, internal, and external. The second multi-atlas ($k=5$) adapted the simpler model introduced by Hafezi-Nejad et al.,24 where femoral cartilage was divided into medial and lateral compartments by the trochlear groove, and other tissues were kept intact. Five scans from the training set of IMO were manually annotated with each delineation scheme. The scans were chosen to represent a diverse set of subjects of different age, sex, body mass index (BMI), and osteoarthritis severity, (respectively, 61.8 [12.0] years; M/F, 3/2, 30.0 [4.5] kg/m²; KL1-3, 2/2/1), avoiding the cases with severe cartilage degradation. Subsequently, the division was done by elastic registration of the segmented scan to the multi-atlas, remapping of the tissue masks based on the proximity to the atlas' subregions, and majority voting over the multi-atlas.

### 2.4 Morphological measurements

We performed the assessment both at tissue and subregional levels. The absolute volume was estimated using numerical integration over the mask voxels. The average thickness was estimated in two steps. First, the thickness map was computed from the tissue-level masks using our own implementation of the local thickness algorithm. Ablation study of the thickness measure is provided in Supporting Information S1. Second, the average thickness was computed from the voxels corresponding to a particular tissue or a subregion.

### 2.5 Comparison of the methods

We used the morphological measurements from FBC data produced by two independent groups. Chondrometrics GmbH provided

---

**TABLE 1** Summary of the OAI iMorphics and OAI FNIH Osteoarthritis Biomarkers Consortium data sets

|                          | OAI iMorphics | OAI FNIH Biomarkers Consortium (after exclusion) |
|--------------------------|---------------|--------------------------------------------------|
|                          | Full          | Full.No progression | Pain progression | JSL progression | Pain and JSL progression |
| **n**                   | 88            | 567                | 189             | 99             | 100                       |
| Male                    | 45            | 227                | 63              | 33             | 55                        |
| Female                  | 43            | 340                | 126             | 66             | 45                        |
| **Age (years)**         | 61.2 (10.0), [45, 78] | 61.6 (8.9), [45, 79] | 61.5 (9.1), [45, 79] | 59.2 (8.8), [45, 78] | 63.1 (8.5), [45, 79] |
| Male                    | 62.0 (10.9), [45, 78] | 61.5 (9.2), [45, 79] | 61.0 (9.9), [45, 79] | 58.9 (8.7), [45, 76] | 63.0 (8.4), [45, 76] |
| Female                  | 60.4 (9.0), [46, 78] | 61.7 (8.8), [46, 79] | 61.7 (8.8), [46, 79] | 59.4 (8.9), [46, 78] | 63.2 (8.6), [46, 78] |
| **BMI (kg/m²)**         | 31.1 (4.6), [21.9, 48.7] | 30.7 (4.8), [18.6, 46.7] | 30.4 (4.8), [18.6, 43.9] | 31.1 (5.0), [20.0, 46.7] | 30.6 (4.6), [20.6, 44.6] |
| Male                    | 30.5 (3.9), [23.2, 41.3] | 30.1 (3.8), [20.8, 42.5] | 29.8 (3.6), [20.8, 40.6] | 30.1 (3.8), [22.6, 40.6] | 30.2 (3.9), [23.2, 38.5] |
| Female                  | 31.6 (5.3), [21.9, 48.7] | 31.0 (5.3), [18.6, 46.7] | 30.7 (5.3), [18.6, 43.9] | 31.6 (5.5), [20.0, 46.7] | 31.0 (5.4), [20.6, 44.6] |
| **KL (1-4)**            | 2, 31, 52, 3 | 73, 290, 204, 0 | 23, 110, 56, 0 | 13, 59, 27, 0 | 14, 45, 41, 0 |
| Male                    | 2, 14, 26, 3 | 27, 94, 106, 0 | 5, 33, 25, 0 | 4, 17, 12, 0 | 8, 18, 29, 0 |
| Female                  | 0, 17, 26, 0 | 46, 196, 98, 0 | 18, 77, 31, 0 | 9, 42, 15, 0 | 6, 27, 12, 0 |

Note: OAI FNIH data set subject groups, defined by relation to pain and radiographic progression, are also shown independently. Age and BMI format is average (SD), [min, max]. KL grades are measured at the baseline.

Abbreviations: BMI, body mass index; JSL, joint space length; KL, Kellgren–Lawrence grading score of radiographic osteoarthritis progression (1—doubtful, 4—severe osteoarthritis); OAI, Osteoarthritis Initiative.
volumetric and thickness measurements based on the delineation scheme of Wirth and Eckstein. Cartilage segmentation in the method of Chondrometrics was done manually, followed by an automatic subregional splitting and assessment. Biomediq A/S provided fully automatic volumetric measurements obtained using registration-based approach and the atlas of Hafezi-Nejad et al. We compared the measurements of the methods using correlation and Bland-Altman analyses. For correlation, the $r$ values were read as in Hirvasniemi et al.: $0.00–0.19$, very weak; $0.20–0.39$, weak; $0.40–0.59$, moderate; $0.60–0.79$, strong; and $0.80–1.00$, very strong. However, since the ground truth cartilage volume and thickness in OAI are not known, we also assessed the discriminative power of the methods in the scope of osteoarthritis progression. Four groups of subjects (GC, GR, GP, GRP) were studied in two scenarios: osteoarthritis progression over 12 and 24 months. Within each scenario, we performed two comparisons: GC versus GR, GC and GP versus GR and GRP. Here, single differential morphological features over the considered period, for example, volume change over 24 months, were extracted for each subject and then used to discriminate between the groups. The discriminative power was measured by odds ratio (OR) using SciPy (version 1.4.1) and scikit-learn (version 0.22.0) software. Finally, the significant associations ($p < 0.05$) were identified and compared between the methods.

3 | RESULTS

3.1 | Tissue segmentation

The developed segmentation model showed volumetric and surface-based accuracy comparable to the previously published state-of-the-art methods. Segmentation quality metrics are presented in Table 2. Tissue-wise, on the test subset of IMO data, the model achieved the following DSCs (mean and standard deviation): $0.890 (0.019)$ for femoral cartilage, $0.897 (0.021)$ for lateral and $0.837 (0.062)$ for medial. Exact comparison of our method to the previously published DL-based methods is challenging due to the differences in the data set splits and the validation approaches. A recent study by Desai et al. in a comprehensive evaluation of the available methods on a standardized split of IMO data set ($n = 74/14$ for training/testing, respectively) showed that the DL-based models achieve, generally, similar performance in terms of DSC ($0.81–0.90$), ASSD ($0.18–0.40$ mm), and other segmentation metrics. In comparison to Ambellan et al. where the method was trained and evaluated in a 2-fold cross-validation scheme ($n = 44/44$ for training/testing, respectively) independently for baseline/12-month follow-ups, our
| Tissue/ subregion | DSC | ASSD (mm) | HD (mm) | Average volume | Average local thickness |
|-------------------|-----|-----------|---------|-----------------|------------------------|
|                   |     |           |         | Reference (mm³) | Predicted (mm³) | r       | AMD (mm³) |
| Femoral           | 0.910 | 0.137 | 5.936 | 15430 | 15408 | 0.952 | -22.631 |
|                   | (0.019) | (0.057) | (2.808) | (2992) | (2971) |
| tLF               | 0.907 | 0.154 | 4.584 | 3169 | 3093 | 0.943 | -76.382 |
|                   | (0.047) | (0.163) | (2.917) | (769) | (773) |
| ecLF              | 0.909 | 0.114 | 2.882 | 489 | 493 | 0.957 | 4.270 |
|                   | (0.032) | (0.115) | (1.644) | (136) | (122) |
| ccLF              | 0.953 | 0.043 | 0.911 | 800 | 815 | 0.994 | 14.218 |
|                   | (0.011) | (0.012) | (0.196) | (221) | (234) |
| icLF              | 0.915 | 0.096 | 2.082 | 807 | 819 | 0.964 | 12.135 |
|                   | (0.024) | (0.059) | (1.137) | (227) | (237) |
| pLF               | 0.905 | 0.133 | 3.758 | 2405 | 2496 | 0.903 | 91.202 |
|                   | (0.033) | (0.076) | (2.061) | (603) | (587) |
| tMF               | 0.913 | 0.131 | 4.021 | 3450 | 3397 | 0.948 | -52.557 |
|                   | (0.023) | (0.062) | (1.652) | (758) | (753) |
| ecMF              | 0.876 | 0.122 | 2.569 | 468 | 473 | 0.928 | 4.867 |
|                   | (0.055) | (0.113) | (1.391) | (131) | (122) |
| ccMF              | 0.862 | 0.176 | 2.211 | 488 | 508 | 0.987 | 19.441 |
|                   | (0.129) | (0.456) | (2.131) | (276) | (282) |
| icMF              | 0.901 | 0.105 | 2.323 | 751 | 754 | 0.971 | 2.830 |
|                   | (0.037) | (0.055) | (1.286) | (244) | (240) |
| pMF               | 0.893 | 0.144 | 3.728 | 2599 | 2556 | 0.880 | -42.653 |
|                   | (0.033) | (0.077) | (1.581) | (600) | (532) |
| Tibial lateral    | 0.921 | 0.135 | 3.595 | 2816 | 2823 | 0.948 | 7.379 |
|                   | (0.014) | (0.050) | (1.286) | (588) | (550) |
| aLT               | 0.925 | 0.070 | 1.835 | 461 | 454 | 0.978 | -6.867 |
|                   | (0.017) | (0.035) | (0.708) | (154) | (154) |
| eLT               | 0.914 | 0.087 | 1.861 | 362 | 364 | 0.943 | 1.830 |
|                   | (0.030) | (0.052) | (0.712) | (104) | (100) |
| cLT               | 0.972 | 0.032 | 0.827 | 672 | 679 | 0.997 | 7.211 |
|                   | (0.007) | (0.008) | (0.150) | (205) | (206) |
| iLT               | 0.894 | 0.175 | 3.308 | 741 | 763 | 0.898 | 21.917 |
|                   | (0.034) | (0.113) | (2.707) | (202) | (142) |
| pLT               | 0.885 | 0.145 | 2.783 | 577 | 561 | 0.944 | -16.396 |
|                   | (0.044) | (0.106) | (1.468) | (219) | (178) |
| Tibial medial     | 0.871 | 0.238 | 4.427 | 2433 | 2439 | 0.959 | 6.074 |
|                   | (0.058) | (0.301) | (2.675) | (737) | (693) |
| aMT               | 0.846 | 0.384 | 3.299 | 607 | 621 | 0.969 | 14.073 |
|                   | (0.156) | (1.112) | (3.036) | (275) | (253) |
| eMT               | 0.784 | 0.482 | 3.858 | 322 | 338 | 0.843 | 16.143 |
|                   | (0.207) | (1.092) | (4.047) | (138) | (111) |
| cMT               | 0.911 | 0.106 | 1.942 | 444 | 451 | 0.971 | 6.916 |
|                   | (0.089) | (0.119) | (1.608) | (187) | (169) |
| iMT               | 0.864 | 0.197 | 2.646 | 479 | 449 | 0.830 | -29.950 |
|                   | (0.057) | (0.119) | (1.100) | (169) | (155) |
| pMT               | 0.871 | 0.143 | 2.986 | 578 | 577 | 0.920 | -1.225 |
|                   | (0.041) | (0.061) | (1.162) | (169) | (151) |
| Patellar          | 0.870 | 0.208 | 4.908 | 3010 | 2975 | 0.934 | -34.848 |
|                   | (0.049) | (0.100) | (3.858) | (871) | (815) |

(Continues)
results suggested smaller ASSDs: 0.137 (0.057) versus 0.19 (0.08)/0.20 (0.09) for femoral cartilage, 0.135 (0.050) versus 0.17 (0.06)/0.18 (0.06) for lateral and 0.238 (0.301) versus 0.26 (0.23)/0.28 (0.22) for medial tibial cartilage. Wirth et al.\textsuperscript{30} have recently studied their approach using a distinct healthy reference cohort from OAI (n = 71/21 for training and testing, respectively). Their method yielded larger DSCs and smaller ASSDs for medial tibial cartilage—0.91 (0.02) versus 0.871 (0.058) and 0.13 (0.03) mm versus 0.238 (0.301) mm. However, for lateral tibial cartilage, the DSCs were comparable, while ASSDs were larger with their method—0.92 (0.02) versus 0.921 (0.014) and 0.17 (0.04) mm versus 0.135 (0.050) mm, respectively. Tibial cartilage HDs were comparable across methods before, but smaller with theirs after post-processing. Similar to Wirth et al.,\textsuperscript{30} our segmentations were less accurate for the tissues in the medial compartment compared with lateral, particularly, in tibial, central femoral, and combined femorotibial cartilage regions. Our proposed method showed very high correlation of volumetric measurements for all tissues ($r > 0.934$) and the absolute mean differences (AMDs) were smaller than 116 mm$^3$ (Table 2). Compared with Wirth et al.,\textsuperscript{30} our automatic segmentations resulted in lower correlation with manual for tibial cartilage in volume ($r > 0.948$ vs. 0.98) but similar in cartilage thickness ($r > 0.935$ vs. 0.93).

### 3.2 Subregional division and morphological assessment

The results of the segmentation analysis are summarized in Table 2. Visual inspection of multiple divided segmentation masks confirmed that the proposed division method produces the delineation in accordance with the original protocols of Wirth et al.\textsuperscript{16} and Hafezi-Nejad et al.\textsuperscript{24} In terms of DSC and ASSD, the least accurate segmentation of femoral cartilage was observed in the trochlear and the posterior subregions (DSC = 0.893–0.913; ASSD = 0.131–0.154 mm) and, importantly, in the medial central weight-bearing subregion (DSC = 0.862 [SD, 0.129]; ASSD = 0.176 [SD, 0.456] mm). For tibial cartilage, larger errors were exposed in the lateral internal and posterior subregions (DSC = 0.885–0.894; ASSD = 0.145–0.175 mm), and in the medial anterior and external subregions (DSC = 0.784–0.846; ASSD = 0.384–0.482 mm). The segmentation metrics were additionally analyzed at different KL-grades (Supporting Information S2). The segmentations were notably less accurate toward KL4 in the subregions of medial tibial cartilage and weight-bearing subregions of medial femoral cartilage (DSC lower by 0.027–0.268, ASSD higher by 0.016–0.369 mm, compared with KL2).

For other tissues and subregions a similar trend was observed, while the differences were rather minor (DSC lower by −0.006 to 0.061, ASSD higher by −0.038 to 0.165 mm). HDs were also higher for severe (KL > 2) cases, primarily in the subregions of medial femoral and tibial cartilage tissues (increase in HD by 0.805–4.520 mm at KL4 compared with KL2). Interestingly, comparison of subregional volumetric measurements showed reduced correlation and larger AMDs for the trochlear and the posterior areas of femoral cartilage, both in the medial and the lateral compartments ($r = 0.880–0.943$; AMD = −76.382 to 91.102 mm$^3$). For tibial cartilage, the lowest $r$ and the largest AMDs were found in internal, external, and posterior subregions ($r = 0.830–0.944$; AMD = −29.950 to 21.917 mm$^3$). From the volumetric measurements, it can also be concluded that the division method attributed more tibial tissue to the central and internal subregions. Average local thickness showed similar regularities as the volumetric measurements, with the largest discrepancy in the posterior femoral ($r = 0.787–0.841$; AMD = 0.068–0.127), and the posterior and internal tibial subregions ($r = 0.876–0.955$; AMD = 0.021–0.086 mm). For patellar cartilage and menisci, the correlation for the volumetric and thickness measurements was also very high ($r = 0.893–0.967$), while the AMDs were among the largest across all the tissues (AMD = −115.339 to 20.480 mm$^3$ for volume and 0.131–0.192 for local thickness, respectively).

### 3.3 Comparison of morphological measurements and discriminative powers

The entire developed method, comprising the segmentation, subregional division, and morphological assessment modules, was eventually run on the FBC data set, and subsequently compared with the method of Chondrometrics\textsuperscript{19} (Table 3). The volumetric estimates showed very high correlation ($r > 0.950$) for tibial cartilage, high correlation ($r > 0.845$) for weight-bearing femoral cartilage, and positive mean differences in relation to the measurements of Chondrometrics. For average local thickness, the correlation was from high to very high ($r$ from 0.770 to 0.962), being larger in the

| Tissue/subregion | DSC | ASSD (mm) | HD (mm) | Average volume | Predicted | AMD (mm) |
|------------------|-----|-----------|---------|----------------|-----------|----------|
| | | | | Reference (mm$^3$) | Predicted (mm$^3$) | r | | | Reference (mm) | Predicted (mm) | r | AMD (mm) |
| Meniscus lateral | 0.897 | 0.169 | 3.537 | 2494 | 2515 | 0.967 | 20.48 | 3.541 (0.358) | 3.733 (0.355) | 0.933 | 0.192 |
| (0.021) | (0.077) | (1.708) | (637) | (551) | | | | | | | |
| Meniscus medial | 0.837 | 0.338 | 5.531 | 2698 | 2582 | 0.961 | −115.339 | 3.605 (0.625) | 3.745 (0.533) | 0.893 | 0.140 |
| (0.062) | (0.208) | (2.966) | (1043) | (917) | | | | | | | |

Abbreviations: AMD, absolute mean difference; ASSD, average symmetric surface distance; DSC, Dice score coefficient; HD, Hausdorff distance; OAI, Osteoarthritis Initiative.
Table 3: Agreement and correlation between our method and Chondrometrics measurements on FNIH Osteoarthritis Biomarkers Consortium data set

| Feature          | r    | AMD (mm²) |
|------------------|------|-----------|
| cLF.VC           | 0.888 | 266.031 (±263.058, 795.119) |
| LT.VC            | 0.973 | 500.817 (150.723, 850.912) |
| cMF.VC           | 0.845 | 262.110 (±226.545, 750.764) |
| MT.VC            | 0.950 | 454.490 (59.357, 849.622) |
| LFTC.ThC         | 0.939 | 1.816 (1.277, 2.354) |
| cLFTC.ThC        | 0.953 | 1.450 (0.857, 2.043) |
| cLF.ThC          | rTAB: 0.912 | rTAB: 0.727 (0.430, 1.024) |
| cAB: 0.938       | cAB: 0.706 (0.454, 0.958) |
| ecLF.ThC         | 0.867 | 0.513 (0.172, 0.854) |
| ccLF.ThC         | 0.903 | 0.677 (0.296, 1.059) |
| icLF.ThC         | 0.820 | 0.697 (0.265, 1.129) |
| LT.ThC           | rTAB: 0.944 | rTAB: 1.089 (0.718, 1.459) |
| cAB: 0.956       | cAB: 1.070 (0.715, 1.424) |
| aLT.ThC          | 0.829 | 1.138 (0.607, 1.668) |
| eLT.ThC          | 0.876 | 0.731 (0.389, 1.073) |
| cLT.ThC          | 0.962 | 0.773 (0.387, 1.158) |
| iLT.ThC          | 0.890 | 1.106 (0.563, 1.650) |
| pLT.ThC          | 0.789 | 1.109 (0.539, 1.679) |
| MFTC.ThC         | 0.882 | 1.604 (0.974, 2.234) |
| cMFTC.ThC        | 0.904 | 1.378 (0.601, 2.156) |
| cMF.ThC          | rTAB: 0.859 | rTAB: 0.745 (0.337, 1.152) |
| cAB: 0.907       | cAB: 0.692 (0.376, 1.008) |
| ecMF.ThC         | 0.770 | 0.706 (0.276, 1.135) |
| ccMF.ThC         | 0.899 | 0.637 (0.137, 1.138) |
| icMF.ThC         | 0.845 | 0.678 (0.210, 1.146) |
| MT.ThC           | rTAB: 0.893 | rTAB: 0.859 (0.546, 1.172) |
| cAB: 0.920       | cAB: 0.833 (0.546, 1.120) |
| aMT.ThC          | 0.890 | 0.851 (0.480, 1.221) |
| eMT.ThC          | 0.804 | 0.695 (0.249, 1.142) |
| cMT.ThC          | 0.881 | 0.741 (0.280, 1.202) |
| iMT.ThC          | 0.778 | 0.804 (0.206, 1.403) |
| pMT.ThC          | 0.852 | 0.915 (0.586, 1.244) |

Note: Absolute mean difference (AMD) is shown as mean (95% confidence interval). For correlation, \( p < 0.0001 \) for all features. Features are abbreviated as \( XYZ.Q \). Here, \( X = (a) \) anterior, \( (p)osterior, \( (c)entral, \( (e)xternal, \) or \( (i)nternal; \( Y = (l)ateral, \( (m)edial, \) or whole tissue; \( Z = (f)emoral, \( (t)ibial, \) or \( (p)atellar cartilage, \) or \( (m)eniscus; \( Q = \) volume of cartilage \( VC \)."

Table 4: Agreement and correlation between our method and Biomediq measurements on OAI FNIH Osteoarthritis Biomarkers Consortium data set

| Feature | r    | AMD (mm²) |
|---------|------|-----------|
| F.VC    | 0.942 | 1601.889 (±548.936, 3752.714) |
| LF.VC   | 0.965 | 812.997 (±183.671, 1809.664) |
| MF.VC   | 0.867 | 788.893 (±841.581, 2419.367) |
| LT.VC   | 0.960 | 197.714 (±222.853, 618.281) |
| MT.VC   | 0.966 | 198.743 (±125.168, 522.654) |
| P.VC    | 0.809 | 237.197 (±1028.190, 1502.584) |
| LM.VC   | 0.816 | 41.667 (±693.430, 776.765) |
| MM.VC   | 0.854 | 157.084 (±986.446, 672.277) |

Note: Absolute mean difference (AMD) is shown as mean (95% confidence interval). For correlation, \( p < 0.0001 \) for all features. Features are abbreviated as \( XYZ.Q \). Here, \( X = (a) \) anterior, \( (p)osterior, \( (c)entral, \( (e)xternal, \) or \( (i)nternal; \( Y = (l)ateral, \( (m)edial, \) or whole tissue; \( Z = (f)emoral, \( (t)ibial, \) or \( (p)atellar cartilage, \) or \( (m)eniscus; \( Q = \) volume of cartilage \( VC \)."

central subregions. The mean differences were 0.513–0.706 mm, 0.695–1.138 mm for the femoral and tibial cartilage subregions, respectively.

Comparison of our method to Biomediq\(^\text{24}\) is summarized in Table 4 and highlights the differences between the DL-based and atlas-based segmentation methods. The measurements showed very high correlation in lateral femoral and both tibial cartilage tissues, and high correlation for patellar and medial femoral cartilages and menisci. In terms of AMD, our method tends to relatively overestimate the cartilage volume for all tissues, except for menisci. For combined femorotibial regions, the overestimation was larger than in the individual subregions.

Finally, we studied the discriminative power of the extracted morphological features. Based on the measurements produced by our and Chondrometrics’ methods, the comparison revealed similar findings, both in the medial and lateral compartments (Figures 2 and 3). With regard to the volumetric features, our method showed significant association both for medial weight-bearing femoral and whole tibial cartilages over 24 months, but not over 12 months. With Chondrometrics, the associations were significant for medial weight-bearing femoral subregion in all comparisons, while being inconsistent for medial tibial cartilage. Thickness-wise, the significant associations were largely similar across the methods. The largest ORs were observed in the medial compartment for central weight-bearing femoral subregion, external and central tibial subregions, and combined femorotibial regions. Our approach showed additional significance for the whole medial cartilage and its anterior subregion in all cases except for GC versus GR case over 12 months. The Chondrometrics measurements over the total subchondral bone area (ThCtAB) yielded, generally, larger ORs than the ones obtained with our method. However, their assessments for cartilaginous area thickness (ThCcAB) for whole tibial cartilage showed similar or lower ORs compared with ours. In the lateral compartment (Figure 3), our method provided significant associations for external weight-bearing femoral and anterior tibial subregional thickness change over 24 months. Other associations were rather sparse and did not show a consistent behavior, with both methods. Here, all ORs were ≤1.919.

Relative performance of our and Chondrometrics methods was
additionally analyzed over subject subgroups with specific KL-grades at baseline (Supporting Information S3). Interestingly, for the central subregions of medial femoral and tibial cartilage, our method yielded larger ORs for subjects at KL1 and smaller at KL3, while the values with Chondrometrics were, generally, at the same level. Notably, their method was more specific to volumetric changes in central medial femoral cartilage at all KLs and to thickness changes in external medial tibial subregion at KL4.

Considering Biomediq, in GC versus GR case their method showed significant ORs in femoral cartilage, its lateral subregion, lateral tibial cartilage, and medial meniscus (Figure 4). Over the 24 months, additional significant association with radiographic osteoarthritis progression was found for medial, but no longer in the whole femoral cartilage. Our method showed significant association for lateral femoral cartilage and lateral meniscus over 12 months, medial femoral, medial tibial, and medial meniscus over 24 months. For “GC and GP” versus “GR and GRP” case, Biomediq data yielded similar findings and ORs as previously. Our method, however, showed additional associations in whole femoral cartilage over 12 months, and lateral femoral and lateral meniscus over 24 months. Interestingly, our method showed consistent behavior, somewhat reflecting the findings from the subregion analysis, while the method of Biomediq returned higher relative ORs for the lateral compartments of femoral and tibial cartilage, and medial meniscus. Numerical values of the ORs for all the aforementioned comparisons are available in Supporting Information S4.

FIGURE 2   Odds ratios obtained using the morphological measurements of our and Chondrometrics methods. The figure includes the effects for the features from the medial joint compartment only. (A,C) correspond to 12-month difference, (B,D) correspond to 24-month difference. In (A,B) control and radiographic progression groups are considered, in (C,D) control and pain-progression and radiographic and radiographic and pain progression groups are considered. Significant associations ($p < 0.05$) are shown with stars. Features are abbreviated as xYz.Q. Here, x—(a)nterior, (e)xternal, (c)entral, (i)nternal, or (p)osterior; Y—(M)edial; Z—(F)emoral or (T)ibial cartilage. MFTC—combined femorotibial region comprising cMF and MT. cMFTC—combined weight-bearing femorotibial region comprising cMF and cMT. Q—volume of cartilage (VC) or thickness of cartilage (ThC). Thickness measurements are computed two ways: cAB—over cartilaginous areas of subchondral bone; tAB—over total area of subchondral bone [Color figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

In this study, we presented a fully automatic method for subregional segmentation and morphological assessment of knee cartilage tissues. We validated and analyzed the performance of the method against the reference manual segmentations. Subsequently, we compared our method to two previously published solutions, one of which is an established semi-automatic system based on manual cartilage delineation with subsequent automatic subregional assessment. Our method achieved comparable discriminative power of radiographic osteoarthritis progression over 12 and 24 months as the semi-automatic system. The results show that the proposed method can already be used to automate the cartilage segmentation and subregional assessment in osteoarthritis research.

Our study brings an important novelty to knee MRI research. For the first time in this domain, a DL-based method for segmentation has been combined with a registration-based solution for subregional division of the tissues. DL provides fast (few minutes per scan) and accurate segmentation, while being demanding in terms of annotated data for method development. Registration, in turn, has limited accuracy, but can be inexpensively reconfigured for a different subregional mapping. In the knee joint, the most used cartilage subregions are defined from a small set of anatomically distinct landmarks, otherwise typically being ambiguous. Registration-based approach can easily incorporate the landmarks into process, exploit
the anatomical similarities between a scan and a multi-atlas, thus, closer mimicking the procedure followed by a human annotator.

Even though our method relies on the recent scientific advances, it has space for further improvements. The method yielded lower accuracy in posterior and trochlear cartilage subregions, which are typically of more complex curvature than the weight-bearing areas and are exposed to more apparent partial volume effect. 3D convolutional neural networks could be used to potentially improve segmentation in such areas, providing more accurate cartilage delineation with higher inter-slice consistency. However, it has been recently shown that 2D approaches may be equally as accurate in the problem on the same rather small data set,29 which necessitates a larger and more diverse sample size to produce further improvements. Concerning the registration module, we did not touch on the optimal ways of creating the multi-atlas, which, if done right, can positively affect the accuracy of subregional splitting. It is important to note, however, that the creation of an optimal multi-atlas is an NP-hard problem and for example, OAI would require enormous amount of time and computation.31,32 In this study, we simplified this process by selecting a diverse set of subjects with different age, sex, BMI, and KL-grades to account for the major variability factors in the cohort. Using more advanced voting strategies over the multi-atlas may provide further improvements.33 Finally, recently introduced pre-trained DL-based registration methods can be alternatively used to improve the accuracy of subregional division.34 Accuracy-wise, our method showed reduced performance in the medial compartment of femoral and tibial cartilage tissues as compared with the lateral one. This difference could be potentially attributed to the more complex morphology of the degraded cartilage, noisier reference segmentations in such areas, and higher contribution of the partial volume effect. Analysis of the segmentation model and the complete pipeline at different KL-grades also indicated that our method is less accurate toward severe osteoarthritic cases. Improving the method performance in such cases, particularly, for KL4 almost absent in the considered data sets, would be a valuable direction for further work.

Despite the methodological novelty, and good performance of the proposed method, our work has still some limitations. In general, the validation of algorithms for segmentation and morphological assessment of the cartilage tissues from MRI is a long-standing problem. Previous studies suggested multiple approaches for validation of measurements in vivo, including computed tomography arthrography, stereophotogrammetry, radiography, ultrasonography.35–40 Reference manual segmentation still remains the most accurate and widely used technique, despite the accompanying intra- and inter-user variability and MR contrast-related bias.29,30 In this study, we extensively used manual segmentation for validation. However, meaningful comparison of the methods when no ground truth is available remains an open problem. From side-by-side analysis of the considered methods, it can be concluded that our approach tended to relatively over-segment the cartilage tissues on FBC data, particularly, when compared with

![Figure 3](wileyonlinelibrary.com)
Chondrometrics. However, when assessed on the reference segmentation masks on IMO data, our approach showed small AMD, both in the volumetric and thickness measurements. This finding suggests a presence of a systematic bias between the reference segmentations from IMO data and the manual segmentations of Chondrometrics. Subsequently, a discrepancy of approximately 2–3 voxels was observed between our and Chondrometrics’ measurements of subregional thicknesses. Here, the differences in segmentations were presumably further amplified by the post-processing done solely in Chondrometrics’ method (e.g. meshing) and the different thickness measures used. For instance, Maier et al.\textsuperscript{41} showed that local thickness used in our method may relatively overestimate cartilage thickness compared with distance transform. We believe that already by using manual segmentation masks of Chondrometrics, our method can be trained to produce less biased measurements and achieve better agreement. A more thorough comparison to the method of Chondrometrics is prohibitively complicated by its proprietary nature. Another important difference is in the compared morphological features. Namely, the thickness measurements by Chondrometrics were computed over the total area of subchondral bone, taking into computation also the denuded bone regions. Very sparse measurements were provided specifically over the cartilaginous areas, which were the main target of our method. Presumably, this has negatively affected the agreement between the methods and the final ORs. This is also suggested by the higher similarity of morphological measurements and ORs between our and Chondrometrics methods with features computed over cartilaginous areas. Nonetheless, our approach can be extended by incorporating bone segmentation into the pipeline, to produce similar thickness estimates.

In our study, we used average subregional thickness as the key morphological feature. It is necessary to mention the work of Buck et al.,\textsuperscript{18} where more complex thickness statistics were shown to yield larger effect sizes, at least, in osteoarthritis treatment efficacy studies. Next, even though our study sheds light on the properties of DL-based segmentation, it does not touch on the important aspects of resegmentation errors\textsuperscript{30} and errors due to scanner drifts,\textsuperscript{42} both of which can contribute to segmentation inaccuracies. Finally, adapting our approach to new MRI sequences and acquisition settings in a resource-efficient way remains an open question. It has been recently shown that minor changes in MRI protocol may degrade the performance of DL-based segmentation, necessitating model retraining or using semi-supervised training schemes.\textsuperscript{8} The registration step may require creating new subregional atlases, which is, however, less demanding when provided the cartilage tissue segmentations. Since

**FIGURE 4** Odds ratios obtained using the volumetric measurements of our and Biomediq methods. (A,C) correspond to 12-month difference, (B,D) correspond to 24-month difference. In (A,B) control and radiographic progression groups are considered, in (C,D) control and pain-progression and radiographic and radiographic and pain progression groups are considered. Significant associations ($p < 0.05$) are shown with stars. Features are abbreviated as Y.Z.Q. Here, Y—(L)ateral, (M)edial, or whole tissue; Z—(F)emoral, (T)ibial, or (P)atellar cartilage, or (M)eniscus; Q—volume of cartilage (VC) [Color figure can be viewed at wileyonlinelibrary.com]
the studies on new MRI protocols often include conventional sequences, correspondence of exam scans can be exploited to facilitate the adaptation of our method. We leave these topics to further work.

Our study makes an important step toward automating the morphological analysis of knee cartilage tissues. We introduced a fully automatic method for segmentation and sub-regional cartilage assessment from 3D DESS MR images. By studying the methodological properties of our approach and comparing it to the prior automatic and established manual methods, we provide new insights on the maturity of automatic segmentation methods for osteoarthritis progression studies. The source code of our approach including the local thickness implementation is publicly available at https://github.com/MiPT-Oulu/SubregionalCartilageAnalysis.

ACKNOWLEDGMENTS

The authors would like to acknowledge the following funding sources: strategic funding of University of Oulu (Infotech Oulu), Sigrid Juselius Foundation, KAUTE Foundation, and Jane and Aatos Erkko Foundation, Finland. Additionally, the authors gratefully acknowledge Santeri Rytky for his insight and scientific discussion. The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Egor Panfilov: Conceptualization; methodology; software; formal analysis; investigation; writing—original draft; writing—review and editing; visualization; final approval; integrity of the work. Aleksei Tiulpin: Conceptualization; methodology; data curation; writing—original draft; writing—review and editing; visualization; supervision; funding acquisition; final approval. Milka T. Nieminen: Conceptualization; writing—review and editing; project administration; funding acquisition; final approval. Simo Saarakkala: Conceptualization; writing—review and editing; project administration; funding acquisition; final approval. Victor Casula: Conceptualization; methodology; validation; writing—original draft; supervision; project administration; final approval.

ORCID

Egor Panfilov https://orcid.org/0000-0002-2500-6375
Simo Saarakkala https://orcid.org/0000-0003-2850-5484

REFERENCES

1. Stammberger T, Eckstein F, Michaelis M, Englemier KH, Reiser M. Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. Magn Reson Imaging. 1999;17(7):1033-1042. https://doi.org/10.1016/s0730-725x(99)00040-5
2. Schneider E, Nevitt M, McCulloch C, et al. Equivalence and precision of knee cartilage morphometry between different segmentation teams, cartilage regions, and MR acquisitions. Osteoarthritis Cartilage. 2012;20(8):869-879. https://doi.org/10.1016/j.joca.2012.04.005
3. Hunter DJ, Niu J, Zhang Y, et al. Change in cartilage morphometry: a sample of the progression cohort of the Osteoarthritis Initiative. Ann Rheum Dis. 2009;68(3):349-356. https://doi.org/10.1136/ard.2007.082107
4. Wluka AE, Forbes A, Wang Y, Hanna F, Jones G, Cicutti FM. Knee cartilage loss in symptomatic knee osteoarthritis over 4.5 years. Arthritis Res Ther. 2006;8(4):R90. https://doi.org/10.1186/ar1962
5. Ryzhkov MD Knee Cartilage Segmentation Algorithms: A Critical Literature Review. [thesis]. Utrecht University; Utrecht University, 2015. http://dspace.library.uu.nl/handle/1874/308831
6. Dam EB, Lilholm M, Marques J, Nielsen M. Automatic segmentation of high-and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. J Med Imaging. 2015;2(2):024001. https://doi.org/10.1117/1.JMI.2.2.024001
7. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MRI imaging data to determine relaxometry and morphometry. Radiology. 2018;288(1):177-185. https://doi.org/10.1148/radiol.2018172322
8. Panfilov E, Tiulpin A, Klein S, Nieminen MT, Saarakkala S. Improving robustness of deep learning based knee MRI segmentation: mixup and adversarial domain adaptation. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 0-0). 2019:13:549-556. https://doi.org/10.1109/ICCVW.2019.00057
9. Ambellan F, Tack A, Ehike M, Zachow S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the Osteoarthritis Initiative. Med Image Anal. 2019;52:109-118. https://doi.org/10.1016/j.media.2018.11.009
10. Wirth W, Hunter DJ, Nevitt MC, et al. Predictive and concurrent validity of cartilage thickness change as a marker of knee osteoarthritis progression: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage. 2017;25(12):2063-2071. https://doi.org/10.1016/j.joca.2017.08.005
11. Eckstein F, Glaser C. Measuring cartilage morphology with quantitative magnetic resonance imaging. In Seminars in musculoskeletal radiology. Semin Musculoskelet Radiol. 2004;8:329-353. https://doi. org/10.1055/s-2004-861579
12. Calvo E, Palacios I, Delgado E, et al. High-resolution MRI detects cartilage swelling at the early stages of experimental osteoarthritis. Osteoarthritis Cartilage. 2001;9(5):463-472. https://doi.org/10.1053/joca.2001.0413
13. Qazi AA, Dam EB, Loog M, Nielsen M, Lauze F, Christiansen C. A variational method for automatic localization of the most pathological ROI in the knee cartilage. In Medical Imaging 2008: Image Processing (Vol. 6914, p. 69140T); 2008. International Society for Optics and Photonics. https://doi.org/10.1117/12.769682
14. Jørgensen DR, Dam EB, Lilholm M. Predicting knee cartilage loss using adaptive partitioning of cartilage thickness maps. Comput Biol Med. 2013;43(8):1045-1052. https://doi.org/10.1016/j.compbiomed.2013.05.012
15. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis
26. Hirvasniemi J, Thevenot J, Multanen J, et al. Association between knee osteoarthritis biomarkers consortium. Arthritis Rheum. 2004;50(3):811-816. https://doi.org/10.1016/j.art20191

28. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.

32. Wei K, Iyer, B, Blimes J. Deep submodularity in Data Subset Selection and Active Learning. In International Conference on Machine Learning: June 2015; PMLR. 1954-1963.

34. Ding Z, Han, X, Niethammer M. VoteNet+: an improved deep learning label fusion method for multi-atlas segmentation. arXiv. 2019:111: 00582-56. https://doi.org/10.1105/IS85479.2020.9098493

36. Cohen ZA, McCarthy DM, Kwak SD, et al. Knee cartilage topography, thickness, and contact areas from MRI: in-vitro calibration and in-vivo measurements. Osteoarthritis Cartilage. 1999;7(1):95-109. https://doi.org/10.1053/joca.1998.0165

38. Eckstein F, Adam C, Sittek H, et al. Non-invasive determination of cartilage thickness throughout joint surfaces using magnetic resonance imaging, J Biomech. 1997;30(3):285-289. https://doi.org/10.1007/s00247-9971146-3

40. Yoon CH, Kim HS, Ju JH, Jee WH, Park SH, Kim HY. Validity of the sonographic longitudinal sagittal image for assessment of the cartilage thickness in the knee osteoarthritis. Clin Rheumatol. 2008;27(12):1507-1516. https://doi.org/10.1007/s10067-008-0956-3

42. Maier J, Black M, Bonaretti S, et al. Comparison of different approaches for measuring tibial cartilage thickness. J Integr Bioinform. 2017;14(2):20170015. https://doi.org/10.1515/jib-2017-0015

44. Desai AD, Caliva F, Iriondo C, Mortazi A, Jambawalikar S, Bagci U, IWOAI Segmentation Challenge Writing Group. The international workshop on osteoarthritis imaging knee MRI segmentation challenge: a multi-institute evaluation and analysis framework on a standardized dataset. Radiol Artif Intell. 2021;3:e200078. https://doi.org/10.1148/ryai2021200078

46. Wirth W, Eckstein F, Kemnitz J, et al. Accuracy and longitudinal reproducibility of quantitative femorotibial cartilage measures derived from automated U-Net-based segmentation of two different MRI contrasts: data from the osteoarthritis initiative healthy reference cohort. Mogn Reson Mater Phys Biol Med. 2020;34:337-354. https://doi.org/10.1007/s10334-020-00889-7

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.