# pyKNEEr: An image analysis workflow for open and reproducible research on femoral knee cartilage

Serena Bonaretti[1,2], Garry E. Gold[1], and Gary S. Beaupre[2,3]

[1]Department of Radiology, Stanford University, Stanford, CA, USA
[2]Musculoskeletal Research Laboratory, VA Palo Alto Health Care System, CA, USA
[3]Department of Bioengineering, Stanford University, Stanford, CA, USA

**Abstract**

Transparent research in musculoskeletal imaging is fundamental to reliably investigate diseases such as knee osteoarthritis (OA), a chronic disease impairing femoral knee cartilage. To study cartilage degeneration, researchers have developed algorithms to segment femoral knee cartilage from (MR) images and to measure cartilage morphology and relaxometry. The majority of these algorithms are not publicly available or require advanced programming skills to be compiled and run. However, to accelerate discoveries and findings, it is crucial to have open and reproducible workflows. We present `pyKNEEr`, a framework for open and reproducible research on femoral knee cartilage from MR images. `pyKNEEr` is written in `python`, uses Jupyter notebook as a user interface, and is available on GitHub with a GNU GPLv3 license. It is composed of three modules: 1) image preprocessing to standardize spatial and intensity characteristics, 2) femoral knee cartilage segmentation for intersubject, multimodal, and longitudinal acquisitions, and 3) analysis of cartilage morphology and relaxometry. Each module contains one or more Jupyter notebooks with narrative, code, visualizations, and dependencies to reproduce computational environments. `pyKNEEr` facilitates transparent image-based research of femoral knee cartilage because of its ease of installation and use, and its versatility for publication and sharing among researchers. Finally, thanks to its modular structure, `pyKNEEr` favors code extension and algorithm comparison. We tested our reproducible workflows with experiments that also constitute an example of transparent research with `pyKNEEr`. We provide links to executed notebooks and executable environments for immediate reproducibility of our findings.

*Keywords*: reproducible research, open science, knee cartilage, segmentation, medical image analysis, interactive paper

# 1 Introduction

Open science and computational reproducibility are recent movements in the scientific community that aim to promote and encourage transparent research. They are supported by national and international funding agencies, such as the United States National Institutes of Health (NIH) [9] and the European Commission [10]. Open science refers to the free availability of data, software, and methods developed by researchers with the aim to share knowledge and tools [75]. Computational reproducibility is the ability of researchers to duplicate the results of a previous study, using the same data, software, and methods used by the original authors [5]. Openness and reproducibility are essential to researchers to assess the accuracy of scientific claims [55], build on the work of other scientists with confidence and efficiency (i.e. without "reinventing the wheel") [54], and collaborate to improve and expand robust scientific workflows to accelerate scientific discoveries [53, 14, 42]. Historically, research data, tools, and processes were rarely openly available because of limited storage and computational power [42]. Nowadays, there are several opportunities to conduct transparent research: data repositories (e.g. Zenodo and FigShare), code repositories (e.g. GitHub, GitLab, and Bitbucket), and platforms for open science (e.g. The European Open Science Cloud and Open Science Framework). In addition, there exist computational notebooks that combine narrative text, code, and visualization of results (e.g. Jupyter notebook and R markdown), allowing researchers to create workflows that are computationally transparent and well documented [54]. Finally, it is possible to recreate executable environments from repositories to run notebooks directly in a browser and thus make code immediately reproducible (e.g. Binder).

In the evolution of research practice, the structure of scientific papers, intended as vehicles to communicate methods and results to peers, is changing. In 1992, Claerbout was among the first to envision interactive publications: "[...] an author attaches to every figure caption a pushbutton or a name tag usable to recalculate the figure from all its data, parameters, and programs. This provides a concrete definition of reproducibility in computationally oriented research" [8]. Following this vision, papers are transforming from static to interactive. They will progressively integrate data and code repositories, metadata files describing data characteristics (e.g. origin, selection criteria, etc.), and computational notebooks used to compute results and create graphs and tables [19, 21] for more transparent research.

Transparency in image-based research is crucial to provide meaningful and reliable answers to medical and biological questions [50]. In the musculoskeletal field, quantitative analysis from magnetic resonance (MR) imaging has assumed an increasingly important role in investigating osteoarthritis (OA) [22]. OA is the most common joint disease worldwide, affecting about 2 in 10 women and 1 in 10 men over 60 years of age [76]. It causes structural changes and loss of articular cartilage, with consequent pain, stiffness, and limitation of daily activities [45]. OA of the knee is one of the main forms of OA, affecting 1/3 of the adults with OA [39] and accounting for 83% of the total OA economic burden [25]. To investigate knee OA, scientists have developed algorithms to preprocess MR images, segment femoral knee cartilage, and extract quantitative measurements of morphology, such as thickness [15] and volume [56], and relaxation times, such as $T_{1\rho}$ and $T_2$ [32].

In the image analysis pipeline, segmentation constitute a major challenge. Researchers still tend to segment femoral knee cartilage manually or semi-automatically, using commercial or in-house software, in a tedious and non-reproducible manner [41, 35]. However, there exist several algorithms that researchers have developed to automatically segment knee cartilage. In the literature and in published reviews [24, 48, 79], we have found 29 relevant publications that propose new algorithms to segment femoral knee cartilage. These algorithms are based on different principles, namely active contours, atlas-based, graph-based, machine and deep learning, and hybrid combinations, and were developed by various research groups worldwide (Fig. 1). Of these, only the implementations by Wang et al. [69] and by Shan et al. [59] are open-source and hosted in public repositories (see Wang's repository and Shan's repository). These two implementations, however, have some limitations: in the first case, documentations of code and usage are not extensive, while in the second case the code is written in C++ and requires advanced programming skills to be compiled and run. Other communities, such as neuroimaging, largely benefit from robust, open-source, and easy-to-use software to segment and analyze images (e.g. ANTs [3], FreeSurfer [17],

Nipype [20]). Thanks to these open-access tools, researchers do not need to re-implement algorithms for basic processing and can focus on further statistical analyses [65, 29, 13]. To accelerate discoveries and findings, it is fundamentally important to have not only open-source tools, but also workflows that are computationally reproducible, and thus enhance scientific rigor and transparency [14].

In this paper, we present `pyKNEEr`, an automatic workflow to preprocess, segment, and analyze femoral knee cartilage from MR images specifically designed for open and reproducible research. `pyKNEEr` is written in python with Jupyter notebooks as graphical user-interface, is shared on GitHub, and has a documentation website. In addition, we provide an example of transparent research with `pyKNEEr` through our validation study, implemented using images from the Osteoarthitis Initiative (OAI) [49] as well as in-house images. Finally, to be compliant with recommendations for interactive publications, throughout the paper we provide links to data files and repositories (🛢), software repositories (), specific code and Jupyter notebooks (</>), executable environments (), metafiles and web documentation (🗎), and websites [37].

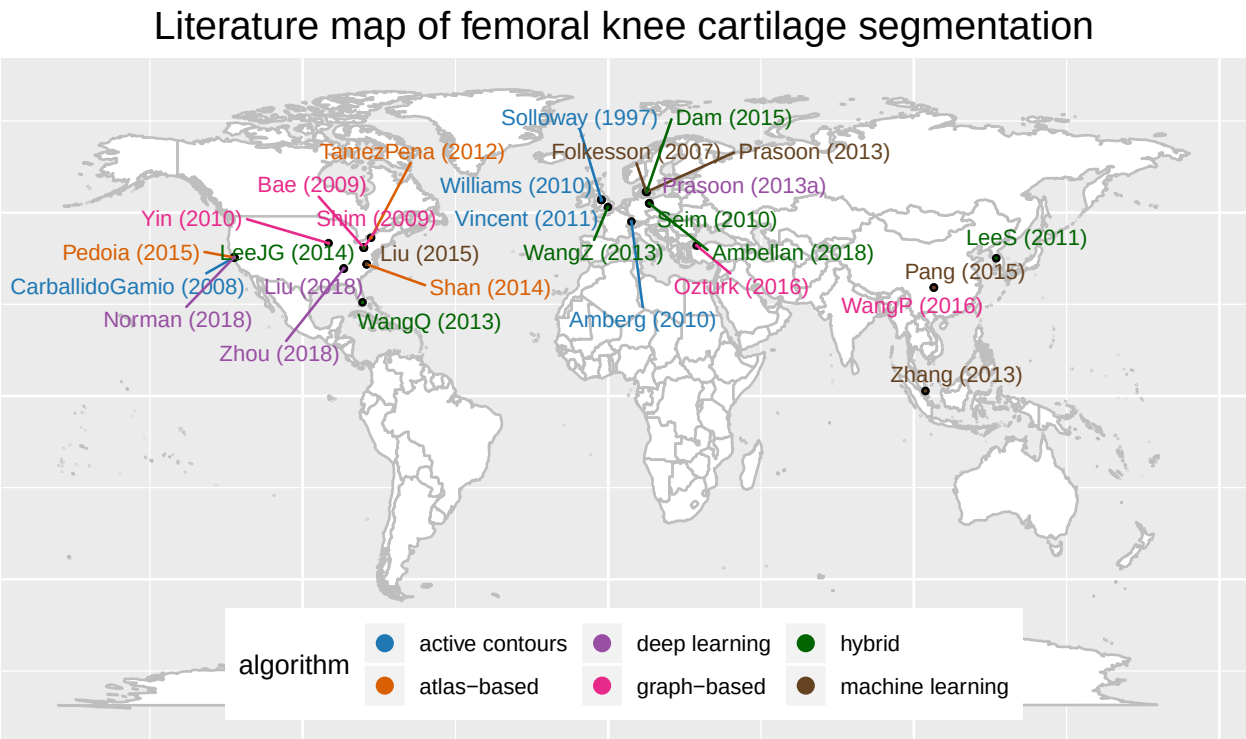



Figure 1: The visualization shows name of first author, year of publication, affiliation of last author, and segmentation method for 29 relevant publications on femoral knee cartilage segmentation from 1997 to 2018. Publications by segmentation method and in alphabetical order are: Active contours: Amberg(2010)[2], Carballido-Gamio(2008)[6], Solloway(1997)[62], Vincent(2011)[67], Williams(2010)[74]; Atlas-based: Pedoia(2015)[47], Shan(2014)[59], Tamez-Pena(2012)[64]; Deep-learning: Liu(2018)[33], Norman(2018)[43], Prasoon(2013a)[52], Zhou(2018)[81]; Graph-based: Bae(2009)[4], Ozturk(2016)[44], Shim(2009)[60], WangP(2016)[68], Yin(2010)[78]; Hybrid: Ambellan(2018)[1], Dam(2015)[11], LeeJG(2014)[30], LeeS(2011)[31], Seim(2010)[57], WangQ(2013)[69], WangZ(2013)[70]; Machine learning: Folkesson(2007)[18], Liu(2015)[34], Pang(2015)[46], Prasoon(2013)[51], Zhang(2013)[80]. This graph and graphs in Fig. 4 and Fig. 5 were made in Jupyter notebook using `ggplot2` [71], an R package based on the grammar of graphics [72]. 🛢 </> 

## 2 Characteristics and structure of pyKNEEr

The main characteristics of pyKNEEr are embedded in its name: *py* is for *python*, to indicate openness, *KNEE* is for femoral knee cartilage, and *r* is for *reproducibility*.

### 2.1 Openness: python, file formats, code reuse, and GitHub

Characteristics and structure of pyKNEEr are based on recommendations for open scientific software in the literature, such as usage of open language and file formats, code reuse, and licensed distribution [53, 26, 54]. We wrote pyKNEEr in the open language python, using open-access libraries to perform computations, such as NumPy for linear algebra, pandas for data analysis, matplotlib for visualizations, SimpleITK for medical image processing and analysis [36], and itkwidgets for 3D rendering. We used widespread open-source formats for input and output data, such as text files (.txt) for input image lists, metafile (.mha) for images, and tabular files (.csv) for tables. To favor our code reuse, we organized pyKNEEr in three modules: 1) image preprocessing; 2) femoral knee cartilage segmentation; and 3) morphology and relaxometry analysis (Fig. 3). Modularity will allow us and other researchers to test, enhance, and expand the code by simply modifying, substituting, or adding Jupyter notebooks. At the same time, we reused open-source code developed by other scientists, such as preprocessing algorithms developped by Shan et al. [59] and elastix for atlas-based segmentation [28]. Finally, we released pyKNEEr on GitHub with a GNU GPLv3 license, which requires openness of derivative work. For citation, we assigned pyKNEEr a digital object identifier (DOI), obtained through the merge of the GitHub release to Zenodo (Table 1).

| | Repository | Metadata / Documentation | Software / Language | License | DOI | Citation |
|---|---|---|---|---|---|---|
| Software | | | | | | |
| Used | | | | | | |
| Preprocessing | Bitbucket | Wiki | C++, ITK | Apache | https://doi.org/10.1016/j.media.2014.05.008* | [59]* |
| elastix 4.8 | GitHub | Github Wiki | C++, ITK | Apache | https://doi.org/10.1109/TMI.2009.2035616* | [28]* |
| Developed | | | | | | |
| pyKNEEr | GitHub | Website | python, Jupyter notebook | GNU GPLv3 | https://doi.org/10.5281/zenodo.2574172 | Bonaretti S. et al. "pyKNEER" (v0.0.1). Zenodo. 2019. 10.5281/zenodo.2574172 |
| Data | | | | | | |
| Original | OAI | Website | - | Data user agreement | https://doi.org/10.1016/j.joca.2008.06.016* | [49]* |
| Derived (results) | Zenodo | Jupyter notebook | - | CC-BY-NC-SA | https://doi.org/10.5281/zenodo.2530609 | Bonaretti S. et al. Dataset used in (Bonaretti et al. 2019). Zenodo. 2019. 10.5281/zenodo.2530609 |

Table 1: Openness and reproducibility of pyKNEEr code and experimental data. DOIs and citations of used software and original data refers to their corresponding publication (*).

### 2.2 Reproducibility: Jupyter notebooks with computational narratives and dependencies

We designed pyKNEEr as a tool to perform and support computational reproducible research, using principles recommended in the literature [55, 54]. For each module of the framework, we used one of more Jupyter notebooks as a user-interface, because of their versatility in combining code, text, and visualization, and because they can be easily shared among researchers, regardless of operating systems.

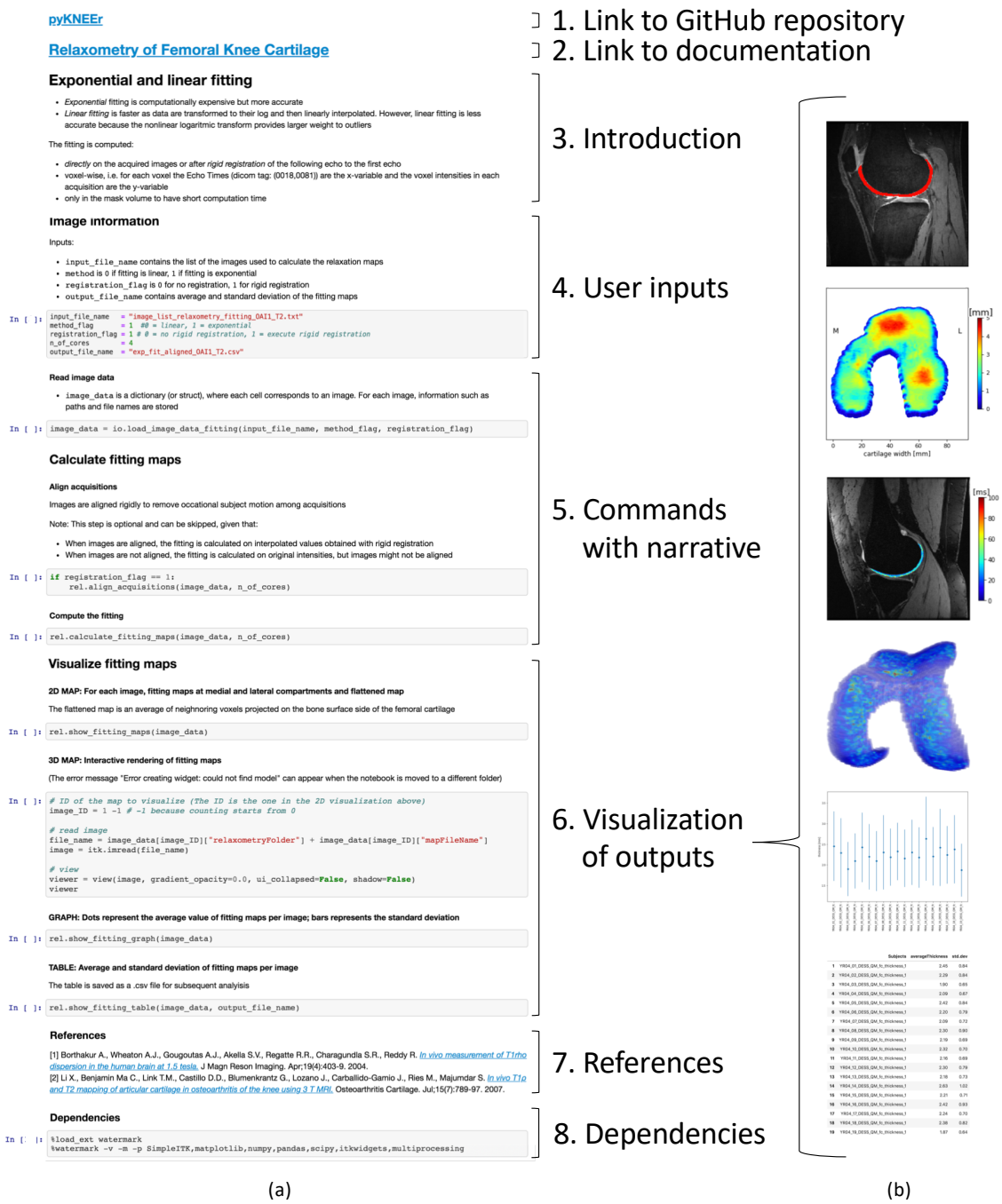



Figure 2: User-interface of modules in pyKNEEr: (a) Structure of Jupyter notebooks and (b) qualitative and quantitative visualization of outputs (from top: cartilage segmentation on image slice, flattened map of cartilage thickness, relaxation map on image slice, 3D relaxation map, and plot and table with average and standard deviation of thickness values).

Across pyKNEEr modules, we used the same notebook structure for consistent computational narratives (Fig. 2). Each notebook contains:

1. Link to the GitHub repository: The repository contains code and additional material, such as source files of documentation and publication;
2. Link to documentation: Each notebook is associated with a webpage containing instructions on how to create input text files, run notebooks, and evaluate outputs. Explanations include step-by-

step instructions for a demo dataset, provided to the user to become familiar with the software. Single webpages are part of a documentation website, comprehensive of installation instructions and frequently asked questions. We created the website using sphinx, the python documentation generator;

3. Introduction: Brief explanation of the algorithms in the notebook;
4. User inputs: The input of each notebook is a text file (`.txt`) with folder paths and file names of images to process or analyze. Additional inputs are variables to customize the process, such as number of cores and algorithm options;
5. Commands with narrative: Titles and subtitles define narrative units of computations, and additional texts provide information about command usage and outcome interpretation. Commands in the notebook call functions in python files associated with that notebook (e.g. in the preprocessing module, the notebook `preprocessing.ipynb` calls the python file `preprocessing_for_nb.py`). In turn, associated python files call functions in core files (e.g. the python file `preprocessing_for_nb.py` calls `sitk_functions.py`, containing image handling functions);
6. Visualization of outputs: Qualitative visualizations include sagittal slices with superimposed cartilage mask or relaxometry map, 2D thickness maps, and 3D relaxometry maps, to allow users a preliminary evaluation of outputs. Quantitative visualizations include scatter plots and tables with numerical values and descriptive statistics (Fig. 2b), which are also saved in `.csv` files to allow researcher subsequent analysis;
7. References: List of main references used in notebook algorithms;
8. Dependencies: Code dependencies (i.e., version of python and its packages) to allow researchers to recreate the current computational environment and thus reproduce findings. To print dependencies, we used the python package `watermark`.

## 2.3 Algorithms in `pyKNEEr`

`pyKNEEr` contains specific algorithms to preprocess, segment, and analyze femoral knee cartilage from MR images. Wherever possible, we implemented multiprocessing using multiple computer cores to optimize computation effort.

### Image preprocessing

Spatial and intensity preprocessing provide standardized high quality images to the segmentation algorithm [16]. In spatial preprocessing, we transform images to right-anterior-inferior (RAI) orientation, we flip right knees (when present) to the left laterality, and we set image origin to the origin of the cartesian system (0,0,0). In intensity preprocessing, we correct image intensities for the inhomogeneities of the static magnetic field ($B_0$) [61], we rescale intensities to a common range [0 - 100], and we enhance cartilage edges with edge-preserving smoothing using curvature flow [58] (Fig. 3-I). Implementation of intensity preprocessing is a translation of the open access code by Shan et al. [59] ( ) from C++ to python.

### Femoral knee cartilage segmentation

Three steps comprise femoral cartilage segmentation: 1) finding a reference image; 2) segmenting femoral cartilage; and 3) evaluating segmentation quality (Fig. 3-II). Finding reference image and evaluating segmentation quality are possible only when ground truth segmentations are available.

*Finding a reference image*. We propose a convergence study to find the reference image, i.e. a segmented image used as a template, or atlas, for the segmentation. First, we randomly select an image as a reference image. Then, we register all images of the dataset to the reference using rigid, similarity, and spline transformations, as explained in the following paragraph. Next, we average the vector fields that result from the registrations. Finally, we choose the image whose vector field is the closest to the average vector

field as the new reference for the following iteration. We repeat until two consecutive iterations converge to the same reference image or after a fixed number of iterations. It is possible to execute the search for the reference several times using different images as the initial reference to confirm the selection of the reference image. This algorithm requires femur masks because the comparison among vector fields and their average is calculated in the femur volume, as cartilage volume is limited.

*Atlas-based segmentation*. Initially, we register a moving image (i.e. any image of the dataset) to a reference image, by transforming first the femur and then the cartilage. Then, we invert the transformation describing the registration. Finally, we apply the inverted transformation to the cartilage mask of the reference image to obtain the cartilage mask of the moving image. The images to segment can be new subjects (intersubject), images of the same subject acquired at different time points (longitudinal), or images of the same subject acquired with different protocols (multimodal). To segment intersubject images, we use rigid, similarity, and spline registration, to segment longitudinal images only rigid and spline registration, and to segment multimodal images only rigid registration. We perform image registration and mask warping with `elastix` and `transformix`, respectively [28, 47], using a multiresolution approach with smoothing image pyramid, random coordinate sampler, adaptive stochastic gradient descent optimizer, and B-spline interpolators [47]. Detailed parameters are in the code repository (🐙).

*Evaluating segmentation quality*. We quantitatively evaluate quality of segmentation using the Dice Similarity Coefficient (DSC), a measure of the overlap between a newly segmented mask and the corresponding ground truth segmentation [12]. The Dice Similarity Coefficient is calculated as:

$$DSC = \frac{2|NM \cap GT|}{|NM| + |GT|}$$

where NM is the newly segmented mask, and GT is the ground truth.

**Morphology and relaxometry analysis**

In `pyKNEEr`, cartilage analysis includes morphology and relaxometry (Fig. 3-III).

*Cartilage morphology*. Morphology quantifications are cartilage thickness and cartilage volume. To calculate cartilage thickness, first we extract contours from each slice of the cartilage mask as a point cloud. Then, we separate the subchondral side of the cartilage from the articular side, we interpolate each cartilage side to a cylinder that we unroll to flatten cartilage [41], and we calculate thickness between the two cartilage sides using a nearest neighbor algorithm in the 3D space [6, 38]. Finally, we associate thicknesses to the subchondral point cloud to visualize them as a 2D map. We compute cartilage volume as the number of voxels of the mask multiplied by the voxel volume.

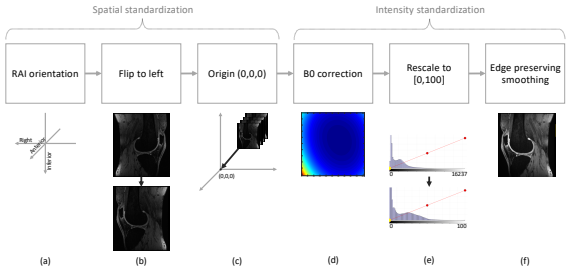
*Cartilage relaxometry*. We implemented two algorithms to calculate relaxometry maps: Exponential or linear fitting and Extended Phase Graph (EPG) modeling. We use exponential or linear fitting to compute $T_{1\rho}$ maps from $T_{1\rho}$-weighted images and $T_2$ maps from $T_2$-weighted images. We calculate exponential fitting by solving a mono-exponential equation voxel-wise using a Levenberg-Marquardt fitting algorithm [32]:

$$S(T_a) = K \cdot exp(-T_a/T_b)$$

where: for $T_{1\rho}$-weighted images, $T_a$ is time of spin-lock (TSL) and $T_b$ is $T_{1\rho}$; for $T_2$-weighted images, $T_a$ is echo time (TE) and $T_b$ is $T_2$; and $K$ is a constant. We compute linear fitting by transforming the images to their logarithm and then linearly interpolating voxel-by-voxel. Linear fitting is not recommended when signal-to-noise ratio is high because the logarithm transformation alters the normality of noise distribution, but it is fast and computationally inexpensive [7]. Before calculating exponential or linear fitting, the user has the option to register the images with lowest TE or TSL to the image with the highest TE or TSL to correct for image motion during acquisition [66]. We use EPG to calculate $T_2$ maps from DESS acquisition. The implementation in `pyKNEEr` is the one proposed by Sveinsson et al [63], which is based on a linear approximation of the relationship between the two DESS signals.

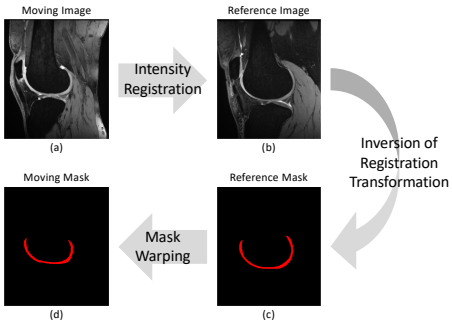
**MAIN ALGORITHMS IN pyKNEEr**

**I. Preprocessing**
`preprocessing.ipybn` </> 🗎



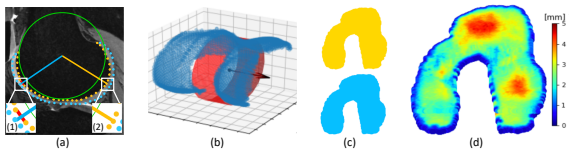

Spatial standardization: (a) image set to RAI orientation (right-anterior-inferior with respect to the subject's coordinate system), (b) right knee (when present) flipped to left, (c) image origin set to (0,0,0). Intensity standardization: (d) correction of $B_0$ inhomogeneity, (e) intensity rescaling to [0,100], and (f) edge preserving smoothing.

**II. Segmentation**
`segmentation.ipybn` </> 🗎
(`find_reference.ipybn` </>)
(`segmentation_quality.ipybn` </>)



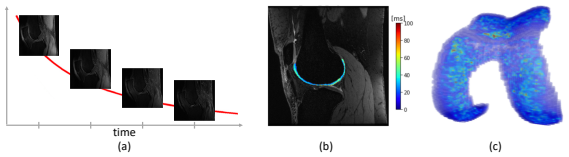

Atlas-based segmentation. We register a moving image (a) to a reference image (b). We invert the computed transformation and we apply it to the mask of the reference image (c) to obtain the mask of the moving image (d).

**III. Analysis**
`morphology.ipybn` </> 🗎




Calculating cartilage thickness. (a) To separate cartilage sides, we interpolate a circle (green) in the cartilage and we connect its center to cartilage points through rays. When a ray (blue) intersects a segment (red) connecting two consecutive points, the point belongs to articular cartilage (1), otherwise to subchondral cartilage (2). (b) Interpolation to a cylinder to flatten the cartilage. (c) Visualization of subchondral (yellow) and articular (blue) surfaces. (d) 2D map of cartilage thickness.

`relaxometry_fitting.ipybn` </> 🗎
(`relaxometry_EPG.ipybn` </>)




Calculating relaxometry maps using exponential fitting. (a) Voxel-wise interpolation of images at subsequent TE or TSL. (b) Relaxation map superimposed on one slice of the image. (c) Relaxation map in 3D.

Figure 3: Main algorithms in pyKNEEr modules: I. Image preprocessing; II. Femoral cartilage segmentation; and III. Analysis of morphology and relaxometry. Left: Links to Jupyter notebooks (</>) and documentation (🗎). In parenthesis, notebooks in the module not depicted here. Middle: Graphic summary of algorithms. Right: Algorithm descriptions.

# 3 Open and reproducible research with pyKNEEr: Our validation study

We validated pyKNEEr with experiments that also constitute an example of open and reproducible research with pyKNEEr.

## 3.1 Image data

We used three datasets that we named OAI1, OAI2, and inHouse (Table 2). OAI1 contained 19 Double-Echo in Steady-State (DESS) images and $T_2$-weighted ($T_2$-w) spin-echo images acquired at year 4 of the OAI study. Ground truth segmentations were created using an atlas-based method (Qmetrics Technologies, Rochester, NY, USA) [64] for a previous study [23]. OAI2 consisted of 88 DESS images acquired at baseline and at 1-year followup. Ground truth segmentations were computed using an active appearance model (imorphics, Manchester, UK) [67]. Finally, inHouse contained 4 images acquired at Stanford University using DESS and CubeQuant protocols. For clarity in the following, OAI1 will be split in OAI1-DESS and OAI1-T2, OAI2 in OAI2-BL (baseline) and OAI2-FU (followup), and inHouse in inHouse-DESS and inHouse-CQ (CubeQuant). Details of the acquisition parameters are in Table 2-I.

| Dataset | OAI1-DESS | OAI1-$T_2$ | | OAI2-BL | OAI2-FU | inHouse-DESS | | inHouse-CQ | |
|---|---|---|---|---|---|---|---|---|---|
| Number of subjects | 19 | 19 | | 88 | 88 | 4 | | 4 | |
| **I. Acquisition parameters** | | | | | | | | | |
| Acquisition protocol | DESS | $T_2$-w | | DESS | | DESS | | CubeQuant | |
| Acquisition plane | sagittal | sagittal | | sagittal | | sagittal | | sagittal | |
| Number of images in series | 2 (1 available)$^\diamond$ | 7 | | 2 (1 available)$^\diamond$ | | 2 | | 4 | |
| In-plane spacing [mm] | 0.3646 x 0.3646 (0.4270 x 0.4270)$^\star$ | 0.3125 x 0.3125 (0.4296 x 0.4296)$^\star$ | | 0.3646 x 0.3646 | | 0.3125 x 0.3125 | | 0.3125 x 0.3125 | |
| Slice thickness [mm] | 0.7 (0.75)$^\star$ | 3 (3.5)$^\star$ | | 0.7 | | 1.5 | | 3 | |
| Echo time (TE) [ms] | 4.7 | 10, 20, 30, 40, 50, 60, 70 | | 4.7 | | 42.52 | | - | |
| Spin-lock time (TSL) [ms] | - | - | | - | | - | | 1, 10, 30, 60 | |
| Repetition time (TR) [ms] | 16.32 | 2700 (2900)* | | 16.32 | | 25 | | 1302 | |
| Flip angle [°] | 25 | 180 | | 25 | | 30 | | 90 | |
| **II. Ground truth segmentation** | | | | | | | | | |
| Method | atlas-based | | | active models | | - | - | - | - |
| Anatomy | femur, femoral cartilage | | | femoral cartilage | | - | - | - | - |
| Type | mask | | | contour | | - | - | - | - |
| **III. Experimental results** | | | | | | | | | |
| Image number in series | 1 | 1 | 2-7 | 1 | 1 | 1 | 2 | 1 | 2-4 |
| Preprocessing | | | | | | | | | |
| Spatial standardization | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Intensity standardization | ● | ● | - | ● | - | ● | - | ● | - |
| Segmentation | | | | | | | | | |
| Find reference | 4, 8, 10, 13, 16 | - | - | - | - | - | - | - | - |
| Intersubject | ● | - | - | ● | - | ● | - | - | - |
| Longitudinal | - | - | - | - | ● | - | - | - | - |
| Multimodal | - | ● | - | - | - | - | - | ● | - |
| Segmentation quality | | | | | | | | | |
| Dice coefficient | ● | ● | - | ● | ● | - | - | - | - |
| Analysis | | | | | | | | | |
| Morphology | ●○ | ●○ | - | ●○ | ●○ | ● | - | ● | - |
| Relaxation | - | ●○ | | - | - | ● | | ● | |

Table 2: Datasets used to evaluate `pyKNEEr`. **I. Acquisition parameters**: Parameters of the protocols used to acquire the images. Images of OAI1-DESS, OAI2-BL, and OAI2-FU were acquired with the same DESS protocol, consisting of 2 echos, although only their average was available ($\diamond$). Images of one subject of the dataset OAI1 had different slice spacing and thickness ($\star$). Data queries to obtain acquisition parameters are in a Jupyter notebook (</>). The original identification numbers (IDs) of the OAI images are in a Jupyter notebook used as a metafile (). **II. Ground truth segmentation**: The datasets OAI1 and OAI2 have ground truth segmentations. They differ for computational method, segmented anatomy, and label type. **III. Experimental results**: Details of the steps in `pyKNEEr` for each dataset. Full circle (●) indicates processing of the dataset, while empty circle (○) indicates processing of ground truth segmentations. The numbers in "Find reference" indicated the ID of the seed images used in the convergence study. Links are to the executed notebooks on GitHub.

## 3.2 Results

We preprocessed, segmented, and analyzed all the dataset using different options in `pyKNEEr`, according to dataset characteristics and availability of ground truth segmentation () (Table 2-III).

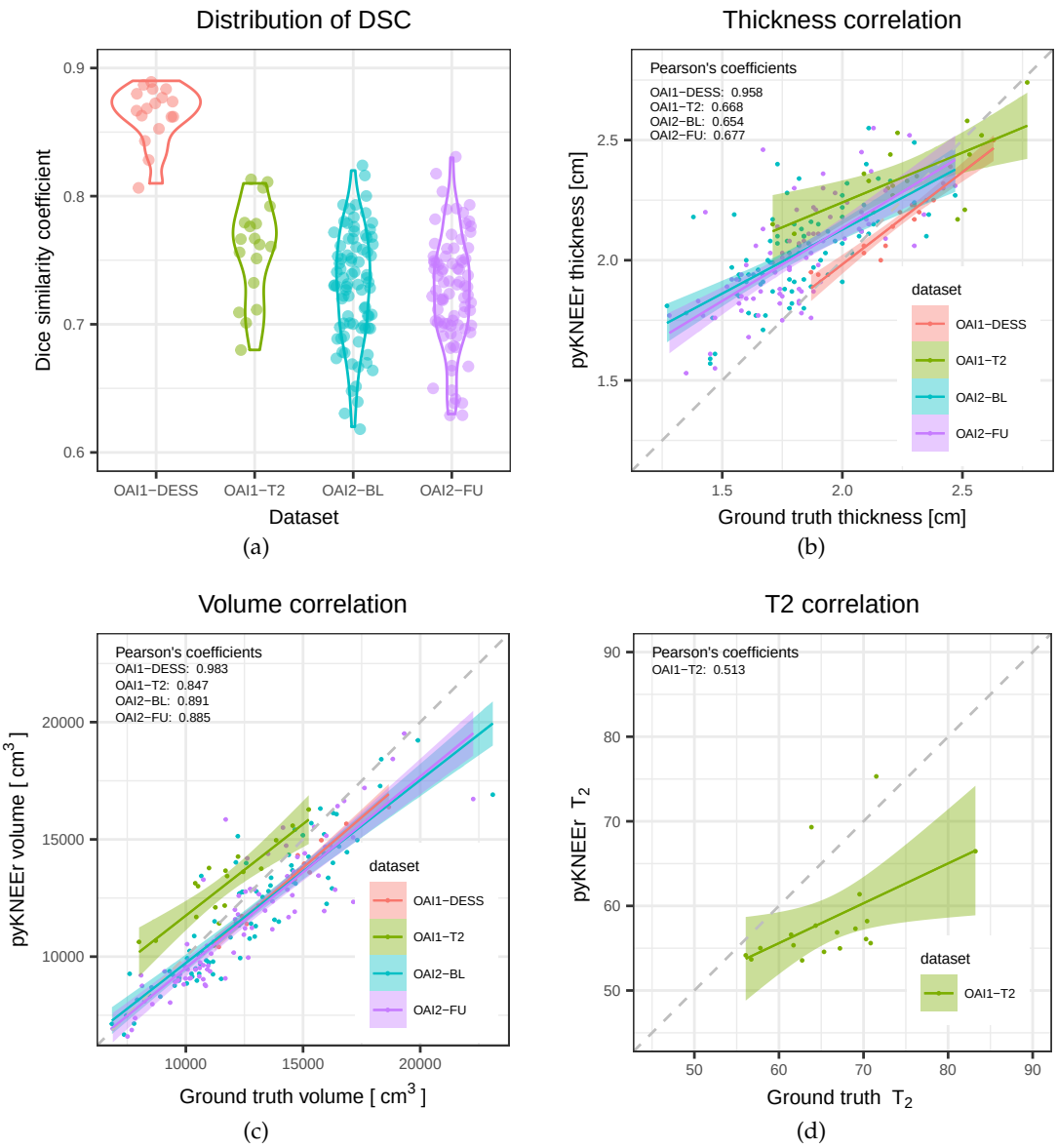



Figure 4: Results for the datasets OAI1-DESS (red), OAI1-T2 (green), OAI2-BL (cyan), and OAI2-FU (purple). (a) Violin plots describing the distribution of the DSC within each dataset. The dots represent DSC values spread around the y-axis to avoid visualization overlap. (b-d) Correlation between measurements derived from ground truth segmentations and `pyKNEEr`'s segmentations, i.e. cartilage thickness (b), cartilage volume (c), and $T_2$ maps (d).

*Preprocessing*. We executed spatial preprocessing for all images of the datasets and intensity preprocessing only for the images directly involved in segmentation.

*Finding reference*. We selected the reference mask from the dataset OAI1-DESS because of the availability of ground truth segmentations of the femur. We picked 5 images as initial reference for our parallel investigation using a python random function (random seed = 4 ⟨/⟩). For all the studies, we found the reference as the subject whose vector field distance to the average vector field was the minimum (subject ID = 9).

*Segmenting intersubject, longitudinal, and multimodal images*. We segmented images from OAI1-DESS, OAI2-BL, and inHouse-DESS as new subjects. Segmentation failure were 1 for OAI1-DESS (ID = 6 ⟨/⟩), 3 for OAI2-BL (IDs = 6,24,31 ⟨/⟩), and none inHouse-DESS (⟨/⟩). We excluded the failed registrations from the following analysis of segmentation quality, cartilage morphology, and cartilage relaxometry. We

segmented the first acquisition of OAI1-T2 images (</>) and inHouse-CQ images (</>) using the multimodal option in pyKNEER, and OAI2-FU images (</>) using the longitudinal option.

*Segmentation quality*. We evaluated segmentation quality for the datasets OAI1 and OAI2 because they had ground truth segmentations of femoral cartilage. The Dice similarity coefficients were $0.86 \pm 0.02$ (mean $\pm$ standard deviation) for OAI1-DESS, $0.76 \pm 0.04$ for OAI1-T2, $0.73 \pm 0.04$ for OAI2-BL, and $0.72 \pm 0.04$ for OAI2-FU (Fig. 4a) (</>).

*Morphology*. We calculated cartilage thickness and volume for all datasets, including ground truth segmentations. We computed correlations of cartilage thickness calculated from pyKNEEr's segmentation and ground truth segmentation, and we found that Pearson coefficients were 0.958 for OAI1-DESS, 0.668 for OAI1-T2, 0.654 for OAI2-BL, and 0.667 for OAI2-FU (Fig 4b). Similarly, we computed correlations for cartilage volume, and we found that Pearson coefficients were 0.983 for OAI1-DESS, 0.847 for OAI1-T2, 0.891 for OAI2-BL, and 0.885 for OAI2-FU (Fig 4c) (</>).

*Relaxometry*. Before calculating relaxometry maps for OAI1-T2, we rigidly registered the images with shortest TE to the image with longest TE. Similarly, before calculating $T_{1\rho}$ maps for inHouse-CQ, we rigidly registered the images with shortest TSL to the image with longest TSL. Then, we calculated $T_2$ maps for OAI-T2 images extracting values in pyKNEEr's masks (</>) and ground truth masks (</>), and we compared them, obtaining a Pearson's coefficient of 0.513 (</>). Finally, we computed relaxometry maps using exponential fitting for inHouse-CQ (</>) and EPG modeling for inHouse-DESS (</>) to show feasibility of the methods.

# 4 Discussion

To test possible reproducible workflows with pyKNEEr, we ran experiments with three different datasets. Image preprocessing was successful in all cases, while image segmentation failed in 4 cases. Average DSC were 0.81 for dataset OAI1 and 0.73 for dataset OAI2, which are in the range of published values (Fig. 5). Discrepancies of DSC between OAI1 and OAI2 can be due to the different characteristics of ground truth segmentations. OAI1 ground truth segmentations were created using an atlas-based method with $DSC = 0.88$ [64] (see "TamezPena (2012)" in Fig. 5), whereas OAI2 ground truth segmentations were created using an active appearance model with $DSC = 0.78$ [67] (see "Vincent (2011)" in Fig. 5). In addition, to calculate DSC we transformed OAI2 ground truth segmentations from contours to volumetric masks, potentially adding discretization error. Quality of segmentation had a direct impact on morphology and relaxometry analysis. Pearson's coefficient was higher for cartilage volume than cartilage thickness, suggesting higher preservation of volume, and it was low for $T_2$ relaxation times, suggesting higher dependency on segmentation quality for intensity-based measurements. Finally, regression lines show that measurements from pyKNEEr segmentation overestimated small thicknesses and underestimated large volumes and $T_2$ values (Fig. 4). Despite its modest performance, we implemented atlas-based segmentation because it has the advantage to provide byproducts for further analysis. Image correspondences established during the registration step can be used for intersubject and longitudinal comparison of cartilage thickness and relaxation times, and voxel-based morphometry and relaxometry [47].

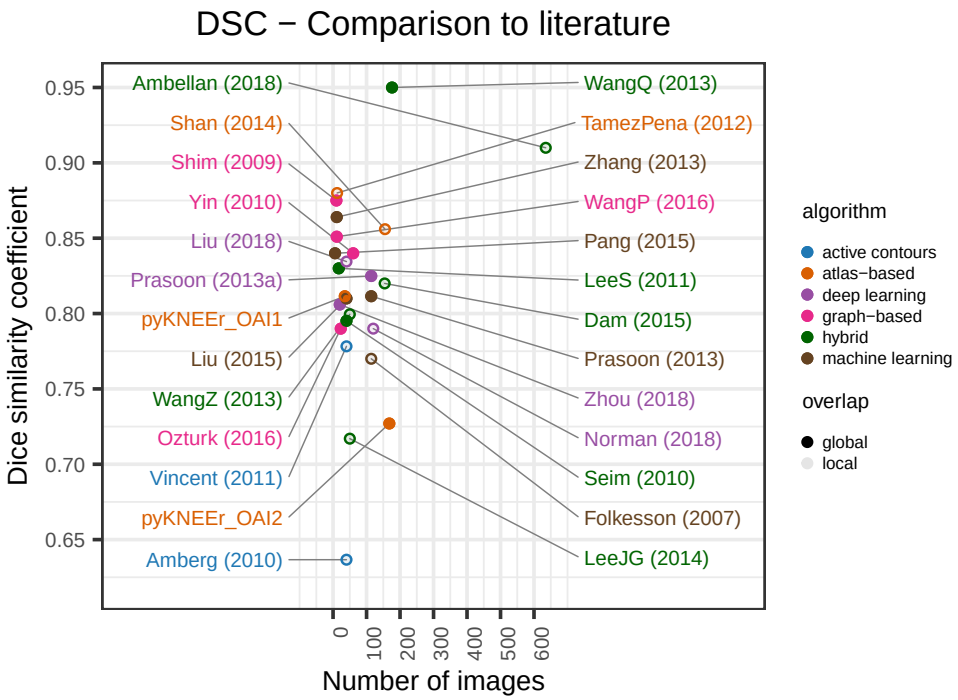



Figure 5: Performances of the segmentation module of `pyKNEEr`, compared with 24 studies in literature that report it. Full dots represents studies where DSCs were calculated on the whole mask, whereas empty dots represent studies where DSCs were calculated in specific parts of the cartilage, e.g. the weight-bearing area [24].

We designed `pyKNEEr` to facilitate transparent research on femoral knee analysis from MR images. Traditionally, medical image analysis workflows are in ITK, VTK, and Qt, requiring advanced computational skills in C++ to build, run, and extend code. We wrote `pyKNEEr` in `python` because of its ease of use, compatibility with various operating systems, and extensive computing support through packages and open code. As a consequence, `pyKNEEr` can be easily installed as a package in the `python` environment and does not require advanced programming skills. In addition, we used Jupyter notebooks as a user-interface because of their ease of use, literate computing approach [40], versatility for publications, and sharing among researchers. In `pyKNEEr`, Jupyter notebooks can be simply downloaded from our GitHub repository ( ) to a local folder. Researchers have to input an image list and optionally set a few variables, and after automatic execution of the notebook, they directly obtain visualizations, graphs, and tables for further analysis. In addition, researchers can link the executed notebook directly to papers (similarly to Table 2) and thus create an interactive publication with reproducible analysis. In the medical image analysis community, other examples of combined use of `python` and Jupyter notebooks are mainly for educational and general research purpose (e.g SimpleITK notebooks [77]), while usage of `python` as a programming language is rapidly gaining popularity in neuroimaging (e.g. Nipype [20]).

Several extensions of `pyKNEEr` could be imagined, thanks to the modularity of its structure. In the segmentation module, the current notebook implementing atlas-based segmentation (`segmentation.ipynb`) could be substituted by notebooks with hybrid machine or deep learning algorithms, which can provide higher DSC [69, 1] (Fig.5). In the morphology module (`morphology.ipynb`), the code structure already includes a flag (`thickness_algo`) to integrate additional algorithms for cartilage thickness, such as surface normal vectors, local thickness, and potential field lines [38]. Finally, new notebooks could be added to the workflow to segment and analyze more knee tissues, such as tibial cartilage, patellar cartilage, and the menisci. Extensions will require a limited amount of effort because of the popularity and ease of `python`, the free availability of a large number of programming packages, and the flexibility of Jupyter notebooks [77]. In addition, standardized file format and computational environment will facilitate comparison of

findings and performances of new algorithms.

In conclusion, we have presented `pyKNEEr`, an image analysis workflow for open and reproducible research on femoral knee cartilage. We validated `pyKNEEr` with three experiments, where we tested preprocessing, segmentation, and analysis. Through our validation test, we presented a possible modality of conducting open and reproducible research with `pyKNEEr`. Finally, in our paper we provide links to executed notebooks and executable environments for computational reproducibility of our results and analysis.

# References


[1] Ambellan, F., Tack, A., Ehlke, M., and Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Medical Image Analysis* (2018).

[2] Amberg, M., Luthi, M., and Vetter, T. Fully automated segmentation of the knee using local deformation-model fitting. In *MICCAI 2010 Workshop Medical Image Analysis for the Clinic—A Grand Challenge (SKI10)* (2010), pp. 251–260.

[3] Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage 54*, 3 (2011), 2033–2044.

[4] Bae, K. T., Shim, H., Tao, C., Chang, S., Wang, J. H., Boudreau, R., and Kwoh, C. K. Intra- and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method. *Osteoarthritis and Cartilage 17*, 12 (2009), 1589–1597.

[5] Bollen, K., Cacioppo, J. T., Kaplan, R., Krosnick, J., and Olds, J. L. Social, behavioral, and economic sciences perspectives on robust and reliable science. Tech. rep., 2015.

[6] Carballido-Gamio, J., Bauer, J. S., Stahl, R., Lee, K. Y., Krause, S., Link, T. M., and Majumdar, S. Inter-subject comparison of MRI knee cartilage thickness. *Medical Image Analysis 12*, 2 (2008), 120–135.

[7] Chen, W. Errors in quantitative T1rho imaging and the correction methods. *Quantitative imaging in medicine and surgery 5*, 4 (2015), 583–91.

[8] Claerbout, J. F., and Karrenbach, M. Electronic documents give reproducible research a new meaning. *SEG Technical Program Expanded Abstracts 1992 11*, 1 (1992), 601–604.

[9] Collins, F. S., and Tabak, L. A. NIH plans to enhance reproducibility. *Nature 505*, 7485 (2014), 612–613.

[10] Commission, T. E. Commission recommendations of 17 July 2012 on access to and preservation of scientific information. Tech. rep., 2012.

[11] Dam, E. B., Lillholm, M., Marques, J., and Nielsen, M. Automatic Segmentation of High- and Low-Field Knee MRIs Using Knee Image Quantification with Data from the Osteoarthritis Initiative. *Journal of Medical Imaging 2*, 2 (2015), 1–13.

[12] Dice, L. Measures of the amount of ecologic association between species. *Ecology* (1945), 297–302.

[13] Doehrmann, O., Ghosh, S. S., Polli, F. E., Reynolds, G. O., Horn, F., Keshavan, A., Triantafyllou, C., Saygin, Z. M., Whitfield-Gabrieli, S., Hofmann, S. G., Pollack, M., and Gabrieli, J. D. Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *Archives of General Psychiatry 70*, 1 (2013), 87–97.

[14] Donoho, D. L., Maleki, A., Rahman, I. U., Shahram, M., and Stodden, V. Reproducible research in computational harmonic analysis. *Comput. Sci. Eng. 11*, 1 (2009), 8–18.

[15] Eckstein, F., Boudreau, R., Wang, Z., Hannon, M. J., Duryea, J., Wirth, W., Cotofana, S., Guermazi, A., Roemer, F., Nevitt, M., John, M. R., Ladel, C., Sharma, L., Hunter, D. J., and Kwoh, C. K. Comparison of radiographic joint space width and magnetic resonance imaging for prediction of knee replacement: A longitudinal case-control study from the Osteoarthritis Initiative. *European Radiology 26*, 6 (2016), 1942–1951.

[16] Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Ayse, I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., Gorgolewski, K. J., Isik, A. I., Erramuzpe Aliaga, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., and Gorgolewski, K. J. FMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods 16*, January (2019), 1–20.

[17] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A. M. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron 33*, 3 (2002), 341–355.

[18] Folkesson, J., Dam, E. B., Olsen, O. F., Pettersen, P. C., and Christiansen, C. Segmenting articular cartilage automatically using a voxel classification approach. *IEEE Transactions on Medical Imaging 26*, 1 (2007), 106–115.

[19] Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J.-h. H., Pierce, S. A., Pope, A., Tzeng, M. W., Villamizar, S. R. S., and Yu, X. Toward the geoscience paper of the future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science 3*, 10 (2016), 388–415.

[20] Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., and Ghosh, S. S. Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics 5*, August (2011).

[21] Gundersen, O. E., Gil, Y., and Aha, D. W. On reproducible AI: Towards reproducible research, open science, and digital scholarship in AI publications. *AI Magazine 39*, 3 (2017), 56–68.

[22] Hafezi-Nejad, N., Demehri, S., Guermazi, A., and Carrino, J. A. Osteoarthritis year in review 2017: updates on imaging advancements. *Osteoarthritis and Cartilage 26*, 3 (2018), 341–349.

[23] Halilaj, E., Hastie, T. J., Gold, G. E., and Delp, S. L. Physical activity is associated with changes in knee cartilage microstructure. *Osteoarthritis and Cartilage 26*, 6 (2018), 770–774.

[24] Heimann, T., and Morrison, B. Segmentation of knee images: A grand challenge. *Proc. Medical Image Analysis for the Clinic: A Grand Challenge. Bejing, China* (2010), 207–214.

[25] Hunter, D. J., Schofield, D., and Callander, E. The individual and socioeconomic impact of osteoarthritis. *Nature Reviews Rheumatology 10*, 7 (2014), 437–441.

[26] Jiménez, R. C., Kuzak, M., Alhamdoosh, M., Barker, M., Batut, B., Borg, M., Capella-Gutierrez, S., Chue Hong, N., Cook, M., Corpas, M., Flannery, M., Garcia, L., Gelpí, J. L., Gladman, S., Goble, C., González Ferreiro, M., Gonzalez-Beltran, A., Griffin, P. C., Grüning, B., Hagberg, J., Holub, P., Hooft, R., Ison, J., Katz, D. S., Leskošek, B., López Gómez, F., Oliveira, L. J., Mellor, D., Mosbergen, R., Mulder, N., Perez-Riverol, Y., Pergl, R., Pichler, H., Pope, B., Sanz, F., Schneider, M. V., Stodden, V., Suchecki, R., Svobodová Vařeková, R., Talvik, H.-A., Todorov, I., Treloar, A., Tyagi, S., van Gompel, M., Vaughan, D., Via, A., Wang, X., Watson-Haigh, N. S.,

and Crouch, S. Four simple recommendations to encourage best practices in research software. *F1000Research 6* (2017), 876.

[27] Jupyter, P., Bussonnier, M., Forde, J., Freeman, J., Granger, B., Head, T., Holdgraf, C., Kelley, K., Nalvarte, G., Osheroff, A., Pacer, M., Panda, Y., Perez, F., Ragan-Kelley, B., and Willing, C. Binder 2.0 - Reproducible, interactive, sharable environments for science at scale. In *Proceedings of the 17th Python in Science Conference* (2018), no. Scipy, pp. 113–120.

[28] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. elastix: A Toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging 29*, 1 (2010), 196–205.

[29] Lawson, G. M., Duda, J. T., Avants, B. B., Wu, J., and Farah, M. J. Associations between children's socioeconomic status and prefrontal cortical thickness. *Developmental Science 16*, 5 (2013), 641–652.

[30] Lee, J.-G., Gumus, S., Moon, C. H., Kwoh, C. K., and Bae, K. T. Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. *Medical Physics 41*, 9 (2014), 092303.

[31] Lee, S., Park, S. H., Shim, H., Yun, I. D., and Lee, S. U. Optimization of local shape and appearance probabilities for segmentation of knee cartilage in 3-D MR images. *Computer Vision and Image Understanding 115*, 12 (2011), 1710–1720.

[32] Li, X., Benjamin Ma, C., Link, T. M., Castillo, D. D., Blumenkrantz, G., Lozano, J., Carballido-Gamio, J., Ries, M., and Majumdar, S. In vivo T1$\rho$ and T2 mapping of articular cartilage in osteoarthritis of the knee using 3 T MRI. *Osteoarthritis and Cartilage*, 15 (2007), 789–797.

[33] Liu, F., Zhou, Z., Jang, H., McMillan, A., and Kijowski, R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magnetic Resonance in Medicine 79* (2018), 2379–2391.

[34] Liu, Q., Wang, Q., Zhang, L., Gao, Y., and Shen, D. Multi-atlas context forests for knee MR image segmentation. In *International Workshop on Machine Learning in Medical Imaging* (2015), no. June 2016, pp. 186–193.

[35] Liukkonen, M. K., Mononen, M. E., Tanska, P., Saarakkala, S., Nieminen, M. T., and Korhonen, R. K. Application of a semi-automatic cartilage segmentation method for biomechanical modeling of the knee joint. *Computer Methods in Biomechanics and Biomedical Engineering 20*, 13 (2017), 1453–1463.

[36] Lowekamp, B. C., Chen, D. T., Ibáñez, L., and Blezek, D. The Design of SimpleITK. *Frontiers in Neuroinformatics 7*, December (2013), 1–14.

[37] Luger, R., Foreman-Mackey, E. A. D., Fleming, D. P., Lustig-Yaeger, J., and Deitrick, R. STARRY: Analytic occultation light curves. *arXiv:1810.06559v1 [astro-ph.IM]* (2018).

[38] Maier, J., Black, M., Bonaretti, S., Bier, B., Eskofier, B., Choi, J. H., Levenston, M., Gold, G., Fahrig, R., and Maier, A. Comparison of different approaches for measuring tibial cartilage thickness. *Journal of integrative bioinformatics 14*, 2 (2017), 1–10.

[39] Martel-Pelletier, J., Barr, A. J., Cicuttini, F. M., Conaghan, P. G., Cooper, C., Goldring, M. B., Goldring, S. R., Jones, G., Teichtahl, A. J., and Pelletier, J. P. Osteoarthritis. *Nature Reviews Disease Primers 2* (2016), 1–18.

[40] Millman, K. J., and Pérez, F. Developing Open Source Practices. In *Implementing Reproducible Research*, V. Stodden, F. Leisch, and R. D. Peng, Eds. Taylor & Francis, 2014, pp. 1–29.

[41] Monu, U. D., Jordan, C. D., Samuelson, B. L., Hargreaves, B. A., Gold, G. E., and McWalter, E. J. Cluster analysis of quantitative MRI T2 and T1rho relaxation times of cartilage identifies differences between healthy and ACL-injured individuals at 3T. *Osteoarthritis and Cartilage 25*, 4 (2017), 513–520.

[42] Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, N., Simonsohn, U., and Wagenmakers, E.-j. A manifesto for reproducible science. *Nature Publishing Group 1*, January (2017), 1–9.

[43] Norman, B., Pedoia, V., and Majumdar, S. Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology 288*, 1 (2018), 177–185.

[44] Öztürk, C. N., and Albayrak, S. Automatic segmentation of cartilage in high-field magnetic resonance images of the knee joint with an improved voxel-classification-driven region-growing algorithm using vicinity-correlated subsampling. *Computers in Biology and Medicine 72* (2016), 90–107.

[45] Palazzo, C., Ravaud, J.-F., Papelard, A., Ravaud, P., and Poiraudeau, S. The burden of musculoskeletal conditions. *PLoS ONE 9*, 3 (2014), e90633.

[46] Pang, J., Li, P. Y., Qiu, M., Chen, W., and Qiao, L. Automatic articular cartilage segmentation based on pattern recognition from knee MRI images. *Journal of Digital Imaging 28*, 6 (2015), 695–703.

[47] Pedoia, V., Li, X., Su, F., Calixto, N., and Majumdar, S. Fully automatic analysis of the knee articular cartilage T 1$\rho$ relaxation time using voxel-based relaxometry. *Journal of Magnetic Resonance Imaging 43* (2015), 970–980.

[48] Pedoia, V., Majumdar, S., and Link, T. M. Segmentation of joint and musculoskeletal tissue in the study of arthritis. *Magnetic Resonance Materials in Physics, Biology and Medicine* (2016).

[49] Peterfy, C. G., Schneider, E., and Nevitt, M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis and Cartilage 16*, 12 (2008), 1433–1441.

[50] Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J. B., Vul, E., and Yarkoni, T. Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience 18*, 2 (2017), 115–126.

[51] Prasoon, A., Igel, C., Loog, M., Lauze, F., Dam, E. B., and Nielsen, M. Femoral cartilage segmentation in knee MRI scans using two stage voxel classification. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (2013), 5469–5472.

[52] Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., and Nielsen, M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* (2013), no. 09, pp. 246–253.

[53] Prlić, A., and Procter, J. B. Ten simple rules for the open development of scientific software. *PLoS Computational Biology 8*, 12 (2012), 8–10.

[54] Rule, A., Birmingham, A., Zuniga, C., Altintas, I., Huang, S.-C., Knight, R., Moshiri, N., Nguyen, M. H., Rosenthal, S. B., Pérez, F., and Rose, P. W. Ten simple rules for reproducible research in Jupyter notebooks. *arXiv:1810.08055* (2018).

[55] Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. Ten simple rules for reproducible computational research. *PLoS Computational Biology 9*, 10 (2013), 1–4.

[56] Schaefer, L. F., Sury, M., Yin, M., Jamieson, S., Donnell, I., Smith, S. E., Lynch, J. A., Nevitt, M. C., and Duryea, J. Quantitative measurement of medial femoral knee cartilage volume – analysis of the OA Biomarkers Consortium FNIH Study cohort. *Osteoarthritis and Cartilage 25*, 7 (2017), 1107–1113.

[57] Seim, H., Kainmueller, D., Lamecker, H., Bindernagel, M., Malinowski, J., and Zachow, S. Model-based auto-segmentation of knee bones and cartilage in MRI data. In *Proc. Medical Image Analysis for the Clinic: A Grand Challenge. Bejing, China* (2010), pp. 215–223.

[58] Sethian, J. *Level set methods and fast marching methods*. Cambridge Press, 1999.

[59] Shan, L., Zach, C., Charles, C., and Niethammer, M. Automatic atlas-based three-label cartilage segmentation from MR knee images. *Medical Image Analysis 18*, 7 (2014), 1233–1246.

[60] Shim, H., Chang, S., Tao, C., Wang, J.-H., Kwoh, C. K., and Bae, K. T. Knee cartilage: Efficient and reproducible segmentation on high-spatial-resolution MR images with the semiautomated graph-cut algorithm method. *Radiology 251*, 2 (2009), 548–556.

[61] Sled, J., Zijdenbos, A., and Evans, A. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging 17*, 1 (1998), 87–97.

[62] Solloway, S., Hutchinson, C. E., Waterton, J. C., and Taylor, C. J. The use of active shape models for making thickness measurements of articular cartilage from MR images. *Magnetic resonance in medicine 37*, 6 (1997), 943–952.

[63] Sveinsson, B., Chaudhari, A. S., Gold, G. E., and Hargreaves, B. A. A simple analytic method for estimating T2 in the knee from DESS. *Magnetic Resonance Imaging 38* (2016), 63–70.

[64] Tamez-Peña, J. G., Farber, J., González, P. C., Schreyer, E., Schneider, E., and Totterman, S. Unsupervised segmentation and quantification of anatomical knee features: Data from the osteoarthritis initiative. *IEEE Transactions on Biomedical Engineering 59*, 4 (2012), 1177–1186.

[65] Van Erp, T. G., Hibar, D. P., Rasmussen, J. M., Glahn, D. C., Pearlson, G. D., Andreassen, O. A., Agartz, I., Westlye, L. T., Haukvik, U. K., Dale, A. M., Melle, I., Hartberg, C. B., Gruber, O., Kraemer, B., Zilles, D., Donohoe, G., Kelly, S., McDonald, C., Morris, D. W., Cannon, D. M., Corvin, A., Machielsen, M. W., Koenders, L., De Haan, L., Veltman, D. J., Satterthwaite, T. D., Wolf, D. H., Gur, R. C., Gur, R. E., Potkin, S. G., Mathalon, D. H., Mueller, B. A., Preda, A., Macciardi, F., Ehrlich, S., Walton, E., Hass, J., Calhoun, V. D., Bockholt, H. J., Sponheim, S. R., Shoemaker, J. M., Van Haren, N. E., Pol, H. E., Ophoff, R. A., Kahn, R. S., Roiz-Santiaez, R., Crespo-Facorro, B., Wang, L., Alpert, K. I., Jönsson, E. G., Dimitrova, R., Bois, C., Whalley, H. C., McIntosh, A. M., Lawrie, S. M., Hashimoto, R., Thompson, P. M., and Turner, J. A. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Molecular Psychiatry 21*, 4 (2016), 547–553.

[66] van Tiel, J., Kotek, G., Reijman, M., Bos, P. K., Bron, E. E., Klein, S., Nasserinejad, K., van Osch, G. J. V. M., Verhaar, J. A. N., Krestin, G. P., Weinans, H., and Oei, E. H. G. Is T1$\rho$ mapping an alternative to delayed gadolinium-enhanced MR imaging of cartilage in the assessment of sulphated glycosaminoglycan content in human osteoarthritic knees? An in vivo validation study. *Radiology 279*, 2 (2016), 523–531.

[67] Vincent, G., Wolstenholme, C., Scott, I., and Bowes, M. Fully automatic segmentation of the knee joint using active appearance models. *MICCAI 2010 Workshop Medical Image Analysis for the Clinic—A Grand Challenge (SKI10)* (2011).

[68] Wang, P., He, X., Li, Y., Zhu, X., Chen, W., and Qiu, M. Automatic knee cartilage segmentation using multi-feature support vector machine and elastic region growing for magnetic resonance images. *Journal of Medical Imaging and Health Informatics 6*, 4 (2016), 948–956.

[69] Wang, Q., Wu, D., Lu, L., Liu, M., Boyer, K. L., and Zhou, S. K. Semantic context forests for learning-based knee cartilage segmentation in 3D MR images. In *Medical Computer Vision. Large Data in Medical Imaging Lecture Notes in Computer Science* (Newyork, 2013), Springer, Ed., pp. 105–115.

[70] Wang, Z., Donoghue, C., and Rueckert, D. Patch-based segmentation without registration: Application to knee MRI. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2013), vol. 8184 LNCS, pp. 98–105.

[71] Wickham, H. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York, 2016.

[72] Wilkinson, L. *The grammar of graphics (statistics and computing)*. Springer-Verlag, Berlin, Heidelberg, 2005.

[73] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data 3* (mar 2016), 160018.

[74] Williams, T. G., Holmes, A. P., Waterton, J. C., MacIewicz, R. A., Hutchinson, C. E., Moots, R. J., Nash, A. F. P., and Taylor, C. J. Anatomically corresponded regional analysis of cartilage in asymptomatic and osteoarthritic knees by statistical shape modelling of the bone. *IEEE Transactions on Medical Imaging 29*, 8 (2010), 1541–1559.

[75] Woelfle, M., Olliaro, P., and Todd, M. H. Open science is a research accelerator. *Nature Chemistry 3*, 10 (2011), 745–748.

[76] Woolf, A. D., and Pfleger, B. Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization 81*, 9 (2003), 646–656.

[77] Yaniv, Z., Lowekamp, B. C., Johnson, H. J., and Beare, R. SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research. *Journal of Digital Imaging 31*, 3 (2018), 290–303.

[78] Yin, Y., Zhang, X., Williams, R., Wu, X., Anderson, D., and Sonka, M. LOGISMOS—Layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint. *IEEE Trans Med Imaging 29*, 12 (2010), 2023–2037.

[79] Zhang, B., Zhang, Y., Cheng, H. D., Xian, M., Gai, S., Cheng, O., and Huang, K. Computer-aided knee joint magnetic resonance image segmentation - A survey. *biorxiv=11802.04894v1* (2018).

[80] Zhang, K., Lu, W., and Marziliano, P. Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies. *Magnetic Resonance Imaging 31*, 10 (2013), 1731–1743.

[81] Zhou, Z., Zhao, G., Kijowski, R., and Liu, F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magnetic Resonance in Medicine 80*, 6 (2018), 2759–2770.

## Acknowledgements

This work was supported by Merit Review Award Number I01 RX001811 from the United States (U.S.) Department of Veterans Affairs Rehabilitation R&D (Rehab RD) Service and by NIHEB002524 and NIHAR062068. The funding sources did not play a role in data analysis, manuscript preparation, or the decision to publish this work.

We would like to thank Uche Monu for her insights on cartilage segmentation, Bragi Sveinsson for sharing his code on EPG modeling, Valentina Mazzoli for her insight on relaxometry, Christof Seiler for software engineering support, Amy Silder, Julie Kolesar, and Scott Uhlrich for beta testing, and all the researchers that make their code open-source, on whose effort we could create `pyKNEEr`.

## Conflict of interest

We do not have any conflict of interest.