

Quantification in musculoskeletal imaging using computational analysis and machine learning: segmentation and radiomics

Authors:

Meritxell Bach Cuadra, PhD^{1,2,3}, meritxell.bachcuadra@unil.ch

Julien Favre, PhD⁴, julien.favre@chuv.ch

Patrick Omoumi, MD, PhD^{1,4}, Patrick.omoumi@chuv.ch

1. Department of Radiology, Lausanne University Hospital and University of Lausanne (UNIL), Lausanne, Switzerland
2. Centre d'Imagerie BioMédicale (CIBM), Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
3. Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
4. Swiss BioMotion Lab, Department of Musculoskeletal Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Corresponding author:

Patrick Omoumi

Department of Diagnostic and Interventional Radiology

Lausanne University Hospital

Bugnon 46

CH-1011 Lausanne, Switzerland

E-mail: pomoumi@gmail.com

Phone: +41 21 314 45 64

Fax: +41 21 314 44 43

Abstract

Although still limited in clinical practice, quantitative analysis is expected to increase the value of musculoskeletal imaging. Segmentation aims at isolating the tissues and/or regions of interest in the image and is crucial to the extraction of quantitative features, such as size, signal intensity, or image texture. These features may serve to support the diagnosis and monitoring of disease. Radiomics refers to the process of extracting large amounts of features from radiological images and combining them with clinical, biological, genetical, or any other type of complementary data to build diagnostic, prognostic or predictive models.

The advent of machine learning offers promising prospects for automatic segmentation and integration of large amounts of data.

In this paper, we present commonly used segmentation methods and describe the radiomics pipeline, highlighting the challenges to overcome for adoption in clinical practice. We will provide some examples of applications from the musculoskeletal literature.

Keywords

Artificial intelligence, radiomics, musculoskeletal imaging, radiology

Today's practice of musculoskeletal imaging remains mainly based on the subjective, qualitative analysis of imaging data by radiologists. However, there is a general agreement on the fact that quantitative assessment could increase the value of radiological examinations, both in the research and the clinical settings¹. Quantitative information can be retrieved not only from functional or compositional imaging techniques such as perfusion MRI or T2 mapping, which are quantitative by nature, but also for morphological techniques in which the information contained in the image can be quantified through the analysis of size, intensity, or texture for instance.

Regardless of the organ, disease, or imaging modality, the processing pipeline for quantitative image analysis usually involves image segmentation (Figure 1). This step is critical as it isolates the tissue(s) of interest in the image. Once the tissues and/or regions of interest have been identified, quantitative variables also called features, such as size, signal intensity, or image texture, can be extracted. These features may be used directly to support the diagnosis or monitoring of disease. They can also be combined with clinical, biological, or genetical data in decision-support models to assist physicians, either for the diagnosis and prognosis of pathological conditions, or for the prediction of therapeutic responses. The process of extracting large amounts of features from radiological images and combining them with clinical, biological, genetical, or any other type of complementary data to build diagnostic, prognostic or predictive models is referred to as radiomics, and is expected to be core to the development of personalized medicine.

The advent of machine learning methods is expected to improve and facilitate the quantitative image analysis workflow. It notably offers promising prospects for automatic segmentation and integration of large amounts of data, which are critical to adoption in routine practice.

This review paper on image quantification will describe the methods for image segmentation as well as the concept of radiomics, which is becoming a hot topic in the field of imaging thanks to the recent progress made in machine learning. First, we will review some definitions and background information. Second, we will present commonly used segmentation methods, highlighting their strengths and weaknesses. Finally, we will present the concept of radiomics in greater details, emphasizing the radiomics pipeline, as well as the challenges to overcome. We will provide some examples of applications taken from the musculoskeletal literature.

1. Definitions and background information

1.1. Image segmentation

Image segmentation refers to the process of delineating region(s) of the image with a specific semantic meaning, such as bone, cartilage, or muscle. In other words, it corresponds to the fact of identifying groups of pixels or voxels that, together, constitute a tissue of interest, or a portion of it. The tissues or regions of interest are defined by their content (such as their signal intensity, or texture) and their contour.

Segmentation can be done manually, semi-automatically, or automatically.

Manual segmentation is the most basic form of image segmentation, and is considered as the gold standard. However, manual segmentation is usually both time-consuming and subject to

intra- and inter-user variability. To streamline and improve this process, much effort has focused on developing semi-automatic and automatic segmentation methods, which in theory should present higher reproducibility and accuracy.

A variety of segmentation tools already exist as part of computer-aided diagnostic (CAD) software, and some basic segmentation tools are included in most image viewers and Picture Archiving and Communication Systems (PACS). These tools are frequently used in clinical practice, for example, to detect and quantify oncological lesions.

While segmentation tools have long existed, machine learning techniques have opened the doors for new automatic processes. However, whether it is applied to musculoskeletal imaging or medical imaging in general, whether manual, semi-automatic or automatic, machine-learning-based or not, image segmentation remains a challenging task. This is particularly true for the musculoskeletal system due to the complexity and variability of the structures of interest.

Since segmentation is fundamental to almost every quantification task, and may significantly influence the results of feature extraction, it must be properly assessed and validated. Although the details of the validation process are beyond the scope of this review, it is important to keep the following basic notions in mind. The segmentation methods need to be validated notably in terms of repeatability (same results when performed multiple times on the same individuals in the same conditions), reproducibility (same results when performed using different equipment or operators), error (difference with respect to a ground truth, usually through manual segmentation), and efficiency (including both computational and user times)^{2,3}. The error is usually determined using metrics that measure spatial overlap (the most common of which is the Dice similarity coefficient) or boundary mismatch (such as the Hausdorff distance)².

1.2. Radiomics

The term radiomics was originally defined as the extraction of large amount of quantitative features from medical images⁴. The name radiomics is derived from similar terms such as genomics and proteomics, which refer to features of genes or proteins, respectively, and which have been analyzed and used in the modeling of pathological conditions. These “omics” are considered as the building blocks required for precision or personalized medicine, where the aim is to tailor medical treatment to the characteristics of individual patients^{4,5}.

The term radiomics has also been used when referring to the whole process leading to decision support models, including data extraction and subsequent analysis, which may be done in consideration of other information obtained from the patient, such as clinical, biological, or genetical data^{4,6-8}. For the purpose of this paper, we will refer to radiomics as to all the steps of this “radiomics pipeline”. Another important term is “radiogenomics”, which was first used in radiation oncology to refer to the prediction of radiotherapy-induced toxicity. Currently it is most commonly used in reference to the analysis of the links that exist between radiomics (or imaging phenotype) and genomics (genetic profile). Sometimes, it has also been used in reference to the relationship of radiomic features to biological data beyond genomics, such as proteomics and metabolomics^{6,7,9,10}.

Whether visible to the human eye or not, the quantitative features extracted from images are believed to provide information on tissue structure, biology or pathology that are correlated with, or complementary to clinical, biological or genetical data from the patient⁶. In other words, radiomics may provide biomarkers of biological and pathological processes, potentially leading to diagnostic, prognostic, as well as predictive information.

It is to note that the process of radiomics differs from the image analyses typically performed in CAD, which have long been used in imaging¹¹. In particular, the number of features incorporated by radiomics is much larger, up to hundreds or thousands. Second, the field of radiomics is not restricted to diagnostic aids as in CAD, but also includes prognostic and predictive models by integrating non-imaging data.

Radiomics has mostly been developed in oncology, where it has become a hot topic thanks to its promise to provide non-invasive assessments of an entire tumor and its microenvironment^{5,7}. An additional advantage of radiomics is that it is not subject to the sampling errors, opposite to biopsies, especially when dealing with heterogeneous lesions¹².

While radiomics research has mostly focused on oncological applications, the process of radiomics can be applied to any condition that can be explored with imaging. This being said, very limited radiomics and radiogenomics literature is currently available in the other fields, particularly in the musculoskeletal field.

In section 3, we will summarize the principle of the radiomics process, as well as some challenges to overcome in order for radiomics to translate into clinical practice. Moreover, we will present a few applications in the musculoskeletal field. For more in-depth presentation of the radiomics process and potential, the readers can refer to the oncology literature^{6,9-12,12,13}.

2. Segmentation methods

This section will describe the segmentation methods that aim to delineate regions of interest, also referred to as target objects. We will describe the families of methods available for segmenting the musculoskeletal system, independently of the imaging modality or tissue of interest (Figure 2). Although current efforts primarily focus on machine learning methods, we will include other families of segmentation methods in this review. Indeed, all segmentation methods remain interesting depending on the application and complex segmentation problems sometimes require combining different methods.

We will first review the basic methods of segmentation which in their majority rely on simple intensity-based operations, such as thresholding or image filter convolutions (section 2.1.1 and 2.1.2). These methods were the first to be used for medical image segmentation and largely adopted for musculoskeletal imaging. However, since they only use intensity-derived properties, they are moderately robust to image artefacts. The second group of methods corresponds to the model-based segmentation approaches (section 2.2). These methods were adopted to overcome the limitations of basic methods by including additional prior information beyond image intensity. For instance, the target can be set to have a closed and smooth parametrized contour, or the segmentation process can use more complex anatomical priors from statistical shape models or atlases. These mathematically more complex techniques come at the cost of being more computationally expensive. In the third group of methods (section 2.3), we will briefly present the graph-based techniques that have been proposed for musculoskeletal imaging. These methods differ from the previous ones in that they do not analyze the image in terms of voxels but in terms of graphs or networks of nodes and edges. Finally, we will present the methods based on machine learning, including both supervised and unsupervised approaches in Section 2.4. These methods, although already used for many years in segmentation tasks, have recently significantly gained in interest due to the rise of deep learning, which has shown impressively good performance in

various computer-assisted tasks. However, the use of machine learning-based methods is only at its beginning in clinical practice, as their ability to be generalized remains to be proven. It is to note that the methods listed here are commonly used in combination rather than in isolation, as illustrated by some of the examples below. Moreover, many of the methods require pre-processing steps, such as denoising filters or inhomogeneity intensity correction, which are beyond the scope of this review.

For the reader who is interested, sections 2.1 to 2.4 presents these methods in greater detail, along with specific examples taken from the musculoskeletal literature. There is considerable literature dedicated to segmentation methods that can be referred to for more extensive description^{14–20}.

2.1. Basic methods

2.1.1. Thresholding

Image thresholding is an intensity-based method that allows simple segmentation at little computational expenses. It is based on the assumption that the region of interest has distinct intensity values compared to the rest of the image. The threshold can be applied to a pre-defined area of interest or to the entire image. Multiple thresholds, including lower and upper thresholds, can be set.

Thresholding is based on the image intensity only, and does not take into account any spatial or neighborhood information. It is therefore highly affected by noise and other intensity artefacts. Thanks to their simplicity, thresholding methods have been applied to segment a variety of musculoskeletal structures, such as bone, bone marrow lesions, knee effusion, or muscles^{21–25}. In particular, local or global thresholding techniques are often applied to radiography or CT images to segment mineralized structures that present high contrast with surrounding structures^{26–29}.

One limitation of threshold-based segmentation is that threshold values are dependent on the modality, scanner or acquisition parameters.

2.1.2. Edge-detection

Edge-detection methods aim at finding object boundaries based on strong intensity transitions or discontinuities within the image. A large variety of edge-detector filters have been developed in the fields of computer vision and image processing³⁰. These filters are based either on a first-order derivative of the image intensity (eg. Sobel, Prewitt, Roberts,...) or on a second-order derivative (eg. Laplacian of Gaussian (LoG), Canny,...). Figure 3 illustrates the results obtained with different types of edge detector filters applied to a spine MR image. Since the filtering operation is based on intensity values, the detected boundaries are often incomplete or discontinuous (see Figure 3). Moreover, image noise and partial-volume effects, which are particularly strong at the edges, may blur the contours of the anatomical structures. These segmentation errors can be corrected to some extent using post-processing techniques such as mathematical morphological operations or more complex methods (e.g., interpolation or active contours)³¹.

Edge-based techniques have been often applied to segment knee structures with MRI. Lee et al. segmented the femurs and patellae based on a combination of second-order edge-detection filters (LoG) and thresholding methods³². In Carballido-Gamio et al., a semi-automated approach was proposed using the first derivative of the image intensity for

contour extraction of the femoral cartilage, with further refinement by the fitting of a curve on the extracted edge³³. Similarly, segmentation of knee cartilage has been done based on the semi-automated software proposed by Duryea et al., using a customized edge-tracking algorithm as an initial segmentation with further active-contour segmentation (see section 2.2.1) and manual edits of the final contour^{34–36}.

2.1.3. Region-growing methods

Region growing methods typically start from seed points, which propagate to neighboring voxels based on predefined criteria, including homogeneity criteria of the region or edge properties³⁷. These methods are often semi-automated as seed-points are manually identified, with the limitation that the result will depend on the initial seeding. To overcome this limitation, strategies for automatically determining the initial seeds have been developed, such as atlas-based registration methods (see section 2.2.3). Adaptive region-growing methods where the homogeneity criterion is automatically extracted from the image have also been proposed³⁸. Due to their reliance on intensity values, region-growing methods could be affected by “leakage” outside the region of interest³⁸.

Waterton et al. applied region-growing methods to characterize diurnal changes in thickness or volume of the femoral articular cartilage of the knee in asymptomatic young adults³⁹. They used a 3D MRI segmentation of the femoral articular cartilage, with a manual seeding point, and a volume-growing algorithm. Thus, the resulting regions consisted of connected areas with similar pixel intensity values, determined via a threshold manually set by the user or automatically determined from a sample region. Finally, the segmentation was extended to adjacent slices, resulting in a three-dimensional segmentation.

2.1.4. Watershed segmentation

Watershed segmentation is a hybrid method that combines edge-based and region-based approaches (Figure 4)^{40–42}. The name watershed comes from a geographical analogy, where water collects in basins and as the water level increases, small basins merge into larger ones. Catchments basins corresponds to different regions of the image with homogeneous grey level. Watershed lines are ridges that separate these regions or catchment basins. They corresponded to the edges in the image. As a first step, the image is usually pre-processed using an edge-detection algorithm, influencing the end result (Figure 4.B1 and 4.B2, leading to 4.D1 and 4.D2, respectively). The pre-processed image is then treated as a topographic structure, where the intensity value of each image element characterizes its height at this point, forming ridges (watershed lines), valleys (minimums) and slopes (catchment basins) (Figure 4). Watershed strategies often lead to over-segmentation (Figure 4.D1). Merging strategies to fuse the obtained segments are required as additional post-processing steps.

Grau et al. proposed a method based on an improved watershed transform, which instead of relying on the classical gradient calculation prior to the watershed, was based on differences of probabilistic maps and the introduction of prior knowledge in watershed segmentation using a statistical atlas⁴³. Preliminary results on four knees showed accurate cartilage segmentation. Włodarczyk et al. used watershed to segment wrist bone at MRI in 37 subjects⁴⁴. The segmentation procedure sequentially targeted the distal parts of the ulna and radius, the proximal parts of metacarpal bones and carpal bones. At every stage, markers of bones were determined first (using an atlas-based approach) and then a watershed from those markers was applied to find the final segmentation. Watershed segmentation has also been recently applied to segment knee cartilage with ultrasound images⁴⁵.

2.2. Model-based segmentation

The basic methods described in the previous section were the first applied for medical image segmentation for a variety of targets, and modalities. Basic methods are very efficient computationally, and require low level of input information as they solely rely on image intensities. However, as described above, they have intrinsic limitations such as the lack of generalizability, low robustness to image artefacts resulting in discontinuous edges or the need for post-processing. To simplify the highly complex task of segmentation, the use of prior anatomical information has proven to be very efficient. Prior information may be provided in many ways. For instance, by pre-defining the contour of the target object as smooth and closed, either explicitly or implicitly through parametrization (by active contours). Another type of prior knowledge could be a set of predefined rules based on known tissue properties or statistical measures (statistical shape models), or as a set of previous manual expert annotations (atlas-based techniques). Model-based segmentation strategies are the family of methods that first exploited these different types of prior knowledge to further improve and automatize the segmentation task.

2.2.1. Geometric deformable models – Active contours

Active Contour (AC) models perform image segmentation by tracking the evolution of parametric curves called *snakes* ^{46–49}. AC models are mathematically formulated as an optimization problem, where the curve parameters are defined to maximize/minimize a certain criterion, referred to as *energy*. The deformation of the snake is guided by combining internal or external energies, such as a smooth and continuous contour (internal energy) or the fitting of the contour to image properties such as the image gradient or a region homogeneity (external energy). Thus, AC models have prior knowledge on human topology (segmented regions are closed and smooth). These techniques have been widely adopted in 2D segmentation tasks but their extension to 3D segmentation is challenging. Another drawback of snakes is that they do not directly allow simultaneous treatment of multiple contours: one initial closed contour (or snake) has to be defined for each structure of interest to be segmented. This is illustrated in Figure 5 where we have plotted the evolution of multiple snakes for the segmentation of vertebrae on a sagittal MR image. Geodesic active contours (GAC) were introduced to overcome this limitation. GAC models are widely used in medical image segmentation for their flexibility, as they allow the handling of topology variations in the shape of the evolving contour. Thanks to the Chan-Vese formulation, GAC models may split apart and merge together, contrary to snakes ⁵⁰. Moreover, their extension to 3D or higher dimensions is rather straightforward.

AC contours were adopted by Duryea et al. to refine the initial edge-detection of the knee cartilage at MRI, leading to reproducible cartilage thickness and volume measurements ³⁴. AC has also been applied to the segmentation of hip joint cartilage in MRI ^{34,51}. Authors segmented the deep and superficial boundaries of acetabular and femoral cartilage with a semi-automated technique using 2D active contours. Based on an initial contour manually defined by the user, the segmentation was achieved through an optimization process based on an edge-based energy.

2.2.2. Parametric deformable models: statistical shape models

Statistical shape models (SSM), also known as point distribution models, can actually be seen as a statistical version of snakes, where additional information on the possible curve

deformation is obtained from a training set and the energy minimization process restricted based on this information. Specifically, SSM are represented using a set of landmarks that allow using previously segmented objects (the training sample set) to describe their average shape and valid range of shape variation with a few parameters only (Figure 6) ⁵²⁻⁵⁴. In order to segment a new image, the shape model can be matched using different algorithms, the Active Shape Models (ASM) and the Active Appearance Models (AAM) being the most widely used ⁵⁵. The ASM seeks to match a set of model points to the dataset by using image information in small regions around the landmark points. AAMs on the contrary act more globally, by minimizing the difference between the model and the target image based on the appearance of a whole region.

SSMs have been used extensively for the segmentation of the musculoskeletal system. Solloway et al. performed one of the first studies using an ASM to segment the femoral cartilage from 0.5T MR images in both healthy subjects and patients with minor cartilage degeneration ^{53,56}. They showed that the precision of ASM was significantly better than both manual and data-driven methods. Fripp et al. later proposed a new method to automatically segment the bones and extract the bone–cartilage interfaces in the knee using three-dimensional active shape models, initialized with anatomical atlases ^{57,58}. They demonstrated a high overlap agreement (the median Dice similarity coefficient was above 0.9) for all structures of interest in a dataset of 20 healthy subjects. Schmid et al. used SSMs for the segmentation of bone at MRI with small field-of-view images ⁵⁹. They found clinically acceptable errors on a dataset of 86 MRI exams.

2.2.3. “Anatomical” models: Atlas-based segmentation methods

Atlas-based methods rely on anatomical priors in the form of an atlas to be registered to the target volume that has to be segmented ⁶⁰. Thus, the segmentation problem is turned into an image registration problem (Figure 7). An atlas includes a template as well as the information on the object to be segmented, or label. This atlas is then registered to the dataset to be segmented by searching a voxel-to-voxel (or pixel-to-pixel) correspondence in 3D (or 2D) images. Volumetric registration is often done in two steps. Firstly, a global registration (affine or rigid transformation) is performed to obtain an initial alignment at a low computational cost. Secondly, a local registration is applied to address the anatomical variability among individuals. Once a spatial transformation is found between the template and the subject, segmentation is done by propagating the label to the data to be segmented. Labels from multiple atlases, instead of a single atlas, may be used when available ⁶¹.

Kapur et al. used a single atlas-based approach to segment the tibia and the femur from MRI images ⁶². Building on this initial atlas-based segmentation, the authors used a region growing algorithm to refine the tibia and femur segmentations and a Bayesian probabilistic framework (density estimation methods in Section 2.4) to also segment the cartilage. Their algorithm was compared to manual segmentation on seven healthy subjects with low error measures.

Lee et al. have proposed a fully automated method to segment cartilage from MRI using multiple atlases ⁶³. They worked on a dataset of 150 MRI images, where 100 were used as atlases and 50 images were used to test the segmentation algorithm. All atlases were registered to a target image by a non-rigid registration scheme and the best matched atlases were used to propagate the labels. This first segmentation was then refined using a graph cut algorithm (see section 2.3).

2.3. Graph-based methods

Graph-based techniques have been very successful in many computer vision problems such as image denoising, texture synthesis or image segmentation problems. They consist in modeling the image as a graph of nodes (or vertexes) connected by a set of arcs (or edges). Such mathematical representation is very flexible as it generalizes well to different types of data, such as high dimensional datasets (i.e. multi-modal/spectral images or 3D or 4D volumes) or images with complex data representation at each point (e.g. diffusion tensor MRI). Graph theory analysis has been extensively studied in many different computer vision applications, and the existing algorithms and theorems can be easily translated to medical image segmentation.

Graph-based methods may encode global image information, thus allowing the use of contextual information⁶⁴. In practice, the segmentation problem is transformed into a “vertex-labeling” or “graph-partitioning” problem which requires assigning correct labels to each node of the graph according to its properties⁶⁵. While the nodes of the graph most commonly are the image pixels or voxels, other types of vertices can also be used to construct the graph, such as region- or user-specific markers. The edges are then set between the nodes with weights corresponding to a similarity metric between the nodes, for instance based on similarity intensity differences. Like active contour-based methods (see section 2.2.1), graph-partitioning methods aim also at minimizing an energy computed from vertex properties (e.g. intensities) and coherence between neighboring vertices (boundary term).

Shim et al. have developed a semi-automated graph cut program for segmenting knee cartilage on high-spatial-resolution 3T MRI⁶⁶. They used the classical graph cut algorithm with user-defined seeds as hard constraints to narrow the search space for possible segmentations⁶⁵. They evaluated the proposed semi-automated algorithm on MR images of 10 patients with varying severity of osteoarthritis and compared it to manual segmentations from two expert radiologists. They showed that the semi-automated method was significantly faster and more reproducible.

2.4. Machine learning methods

In the last few years, a great body of work has focused to employ machine learning techniques for medical imaging segmentation⁶⁷. Machine learning lies on the boundary of several different disciplines, mainly computer science, statistics, mathematics, and engineering (pattern recognition and signal processing). Machine learning algorithms aim at learning a function (f) that best maps input variables (X) to an output variables (Y). f can then be used to predict (Y), given new examples of input variables (X). f can represent a large variety of approaches such as classification, regression, dimensionality reduction or density estimation. Both unsupervised and supervised learning strategies have been used to estimate the function f . In unsupervised data-driven learning, the image segmentation is solved by exploiting some image features such as latent patterns in the voxel intensity distribution. Such methods are in the family of **generative learning** and include data *clustering* or *density estimation methods*. On the other hand, supervised learning techniques require a training set including labelled samples from which algorithm can learn. This is known as **discriminative learning**. Most common applications exploiting discriminative learning are *classification* and *regression* problems. Voxel-wise classification methods are most commonly used for segmentation.

2.4.1. Generative learning

Generative-based approaches are widely used for image segmentation as they require few user interactions. Among generative models, clustering algorithms partition the feature space of an image into clusters, which correspond to data having a certain level of similarity or distance function. The number of clusters need to be determined. One of the most widely used clustering algorithm is K-means, which, in its original form, uses the Euclidean distance as a distance metric and variance as a measure of cluster scatter. K-means tend to obtain clusters of the same size. Fuzzy C-Means (FCM) is a soft version of K-means that relaxes the constraint that a voxel may belong only to one cluster by using the membership function^{68,69}. Carballido-Gamio *et al.* used a classical intensity-based fuzzy clustering method for trabecular bone segmentation with MRI⁷⁰. The method assigned a partial membership to each voxel based on the distance between image point intensities from the centroid of the cluster which accounted also for partial volume effects. Folkesson *et al.* extended the work of Carballido-Gamio *et al.* by including, in addition to signal intensity, a local bone enhancement feature^{70,71}. This approach allowed the method to account for partial volume effects, noise, and to some extent also for signal intensity inhomogeneity. They showed the performance of their method in terms of reproducibility, correlation with high-resolution peripheral quantitative CT, as well as the ability to detect fractures.

2.4.2. Discriminative learning

Supervised voxel-wise classifiers are based on a training dataset from which each voxel is assigned to a specific class. A challenging aspect of these methods is knowing which is the best set of image features for segmentation, including intensity, texture, or gradients. There is a wide family of classifiers proposed in the literature. However only few have been applied to the musculoskeletal system. The K-nearest neighbor (KNN), Support Vector Machines (SVM) and Random Forests (RF) have been most widely used as they have shown promising results in solving different classification problems in other research fields^{72,73}.

Folkesson *et al.* presented a fully automatic method for the segmentation of knee cartilage with MRI based on a KNN classifier⁷². Their method was evaluated by comparisons to manual segmentations by a radiologist and by examining the inter-scan reproducibility of the volume and area estimates. 139 scans of knees with varying degree of disease were used for training and testing, showing good agreement with manual segmentation, and interscan reproducibility as good as that of a human expert.

2.4.3. Deep learning architectures

Recently, machine learning research has focused on the use of the learning procedure to define a better transformation of raw data into a representation that can effectively accomplish the assigned task. In other words, the feature engineering task is left to the machine learning algorithm. Deep learning is one of the many approaches to machine learning, but contrary to the shallow learning methods reported above, deep learning is focused on learning data representations⁷⁴. It is based on artificial neural network architectures, including many layers of neurons (more than three layers). Each layer learns a specific feature, such as curves or edges in the image. Moreover, deep learning methods are more scalable than other machine learning techniques, meaning that their performance increases significantly with increasing training data.

Deep learning algorithms have been applied to musculoskeletal imaging for segmentation and identification of bone and associated soft tissue abnormalities using various imaging modalities^{19,75}.

Prasoon et al. presented a novel method for voxel classification of knee cartilage MRI, integrating three 2D Convolutional Neural Networks (CNNs)⁷⁶. They tested their method on the tibial cartilage in low-field knee MRI scans. They showed improved results with their method in comparison to a reference machine-learning method in terms of the Dice similarity coefficient⁷².

Recently, Liu et al., presented a fully automated segmentation pipeline to segment knee bone and cartilage, combining a semantic pixel-wise segmentation CNN and 3D deformable modeling to preserve the overall shape and maintain a desirable smooth surface of the structures⁷⁷. They tested their approach in three different datasets of 100, 60 and 100 patients respectively, including different MR sequences and vendors. They showed superior performance to state-of-the-art image processing-based methods⁷⁸.

One of the main limitations of supervised methods is the fact that they depend on the existence of a large quantity of good quality of labeled data, which requires a large amount of resources to be obtained, particularly in the medical image domain. This is particularly important in the deep learning domain in order to avoid overfitting, that is, the inability of the classifier to generalize the segmentation to unseen images.

3. Radiomics

3.1. The radiomics pipeline

There are essentially two types of pipelines for radiomics: the conventional or deep learning approaches. The conventional pipeline consists of multiple steps including image acquisition, image segmentation, extraction and selection of handcrafted features, data integration, analysis and model building (Figure 1). In the deep learning pipeline, all the steps after the acquisition of images can be left to neural networks.

3.1.1. Conventional radiomics

Following is a brief description of each of these steps, including some related challenges and difficulties.

3.1.1.1. Image acquisition

Radiomics can use any type of imaging sources, although most commonly applied to cross-sectional modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and Positron emission tomography (PET).

Because of the major influence of acquisition and image reconstruction parameters on the image aspect (Figure 8), the performance of radiomics feature extraction heavily depends on this step of the process^{79,80}. The consistency of feature extraction needs to be addressed for different acquisition protocols, reconstruction parameters, and different scanners^{10,81}.

Pre-processing of the imaging dataset may be performed prior to segmentation in order to reduce noise and artefacts.

3.1.1.2. Image segmentation

Image segmentation is a fundamental step in the radiomics pipeline and may use any of the methods developed in section 2, alone or in combination.

3.1.1.3. *Feature extraction and selection*

Typical features that are extracted in this process include first-order feature (such as morphological measures (size and shape) and signal intensity), as well as second-order features (which describe the relationships between image pixels or voxels, referred to as texture features, commonly used to assess tumor heterogeneity for instance) ^{11,82,83}. Higher-order features can also be extracted from the images. For example, the use of filters (such as wavelet transforms, or fractal analyses) allow the extraction of patterns from the images ⁶. Virtually an unlimited number of features could be extracted from an image, many of which are redundant, irrelevant, or contribute to over-fitting of the model later used for data integration. Therefore, an important step of the pipeline consists in the selection of a subset of features that will allow appropriate predictions (Figure 9) ⁸⁴. This process, called feature selection, can be done using either supervised or unsupervised techniques. In supervised techniques, features are selected based on their relevance for the outcomes of interest: the features that best classify or predict labeled data are selected. Unsupervised techniques aim mainly at removing redundant features, without considering the outcomes of interest ¹¹. Common approaches include clustering and principal component analysis (these techniques belong to the family of generative machine learning, please refer to section 2.4.1 for further details) ⁸¹. Feature selection may be performed in conjunction with the next step, which is data integration, analysis, and model building.

3.1.1.4. *Data integration, analysis and model building*

In this step, image features are integrated with other types of patients' data in order to build models. Any type of information may be considered, including clinical, genetical, or biological (such as blood biomarkers, or histopathological) information. The models built may have a variety of applications that can be categorized as clustering, classification and time-related analysis (survival analysis). The methods used for the analysis depend on the task at hand, ranging from statistical methods to data-mining and machine-learning approaches. Some commonly methods used for classification include Random Forests, support vector machine (SVM), neural networks, generalized linear models, naive Bayes algorithm and least shrinkage and selection operators (LASSO) ^{11,85}. Methods used for survival analysis include Kaplan-Meier survival analysis, Cox proportional hazards regression models and log-rank tests¹¹. The models built through this process may be used as decision-making support, for example to diagnose a condition, stage a tumor, or predict patient survival or treatment response.

3.1.2. Deep learning-based radiomics

Deep learning methods have the advantage of bypassing the difficulties related to the segmentation and feature selection steps of the conventional radiomics pipeline, potentially increasing reproducibility. They have minimal reliance on user input, and feature extraction and selection happen automatically within a convolutional neural network. In theory, deep learning allows more complex and generalizable architectures to be built ¹⁰. Their performance can further increase with more training samples. Disadvantages include high computational cost and the requirements for large volumes of training data ¹¹.

3.2. Radiomics literature: general considerations and musculoskeletal applications

While many research studies have shown the potential of radiomics to help decision-making, especially in oncology, radiomics is still an emerging field, where translation into clinical practice has been slow. In particular, the scientific quality and quality of reporting were found to be yet insufficient⁸⁶. While guidelines for scientific quality and quality of reporting exist (including the radiomics quality score (RQS), the Transparent Reporting of a multivariable prediction model for Individual Prognosis and Diagnosis (TRIPOD), Image biomarker standardization initiative (IBSI)), few papers comply with these guidelines^{8,87,88}. Areas where improvement is particularly needed include the analysis of feature reproducibility, the validation of models in prospective studies, and the assessment of clinical utility^{3,86}. The use of phantom studies is essential to assess the inter-scanner and inter-vendor variability of features⁸. Validation of a model in terms of reproducibility and performance is a particularly important aspect of the assessment. While most studies use internal cross-validation, prospective studies with independent cohorts are the ideal scenario, although very few studies have so far achieved such a high level of evidence⁸⁶.

The applications of radiomics to musculoskeletal imaging are particularly scarce. Out of 77 scientific papers containing the terms “radiomics” and “radiogenomics” published until December 2018 in high-impact clinical journals (with an impact factor ≥ 7) (n=25) and those published in top imaging journals (Radiology and European Radiology) (n=52), only one pertained to musculoskeletal imaging (focusing on sacral tumors)^{85,86}. It is to note that 70 out of these 77 papers were focusing on oncological applications (mainly lung, brain, digestive and breast cancers)⁸⁶.

Most studies were performed with small sample sizes and lacked proper validation, among other limitations. Therefore, they should be considered as feasibility studies in an emerging field of research.

Opposite to other imaging subspecialties, only a few studies have focused on oncological applications of radiomics in musculoskeletal imaging. Examples include a study by Yin et al. that aimed at developing and validating a classification model for the preoperative differentiation of three sacral tumors (chordoma, giant cell tumor, and metastatic tumor), nicely illustrating the conventional radiomics pipeline. First, image acquisitions included T2-weighted fat-suppressed (T2w FS) and contrast-enhanced T1-weighted (CE T1w) MRI on 120 patients presenting with one of the three tumors. Eighty-three exams were assigned to the training set, and 37 to the validation set. The study was retrospective, but acquisition parameters were standardized. Second, sacral tumors were manually segmented, and interobserver agreement tested. Third, 385 radiomics features were extracted for each examination, including first- and second-order (texture) features. Fourth, feature selection was performed using ANOVA and Pearson correlation statistics, as well as LASSO regression and Random Forest (RF). Fifth, the model was built with the ten most significant features, using RF. Sixth, the model was validated using receiver operating characteristic (ROC) curve and accuracy analysis. In the training set, the best performance was based on joint T2w FS and CE T1w sequences, with an area under the curve (AUC) of 0.773 and an accuracy of 0.711, although the performance was more moderate in the validation set (0.643 and 0.649 for the AUC and accuracy, respectively). Interestingly, the performance of radiomics was superior to the performance of imaging data integrated with clinical data.

Another paper by Zhang et al. aimed at developing a radiomics model for predicting the histopathological grades of soft tissue sarcomas preoperatively, with MRI. Over 1000

features were extracted from T2w FS sequences in 35 MRI exams retrospectively included. Tumors were manually segmented (but the interobserver agreement was not assessed). Feature selection was made using LASSO. Models were built using three machine learning algorithms: RF, k-nearest neighbor and SVM. Five-fold cross-validation of the model was performed. A model using five texture features showed high performance in differentiating low- and high-grade soft tissue sarcomas (AUC of 0.92 and accuracy of 0.88).

Hayano et al. showed in 20 patients with soft tissue sarcoma treated with neoadjuvant bevacizumab and radiotherapy that texture features extracted from non-contrast-enhanced CT were correlated with neoangiogenesis and survival. Another paper by Yi et al. showed that first-order and texture features on CT could be used to assess response to denosumab therapy in patients with giant-cell tumors of bone ⁸⁹.

As with most studies currently available, these papers suffer from common limitations, including small sample size, heterogeneity in image acquisition parameters, and the lack of external validation, impeding their generalizability.

Apart from oncology, machine learning applications for quantification tasks have also gained interest in the fields of osteoporosis and cartilage research.

A recent review paper has focused on the applications of machine learning to osteoporosis ⁹⁰. A PubMed/Medline search for papers on bone/ fracture/ osteoporosis and artificial intelligence/machine learning/deep learning, published from January 2017 to March 2019 was performed. They retrieved fifteen studies, most of which aimed at predicting osteoporotic fractures or at diagnosing osteoporosis. While some studies used radiomics features such as texture analysis from CT or X-rays to predict the risk of fracture, three large cohort studies used machine learning to predict the risk of fracture in osteoporotic patients using different clinical parameters combined with quantitative DXA imaging results ⁹¹⁻⁹⁷. Another paper used high-order radiomics features (fractals) extracted from radiographs to assess response to denosumab treatment in postmenopausal osteoporotic patients ⁹⁸. While these studies illustrate promising applications of machine learning, some limitations need to be addressed before translating into clinical practice. These limitations include small sample sizes for most studies, the lack of external validation, and some issues with the performance metrics used, such as failure to report the rate of false negatives for screening studies ⁹⁰.

In the field of osteoarthritis, cartilage volume and thickness measurements represent natural biomarkers of the disease and have been extensively studied for the last decades, mostly through manual image segmentation. As seen in section 2, a large body of work has aimed at automatizing cartilage segmentation, which has remained a challenging task for years. However, atlas-based methods and deep learning approaches have opened new possibilities to perform this task successfully ^{99,100}. These novel methods could be applied to compositional MRI techniques such as T2 mapping to automatize the extraction of quantitative data. As an example, features such as texture extracted from T2 maps have shown their potential to provide prognostic information on the development of OA ¹⁰¹. Furthermore, different methods have been proposed to analyze the local variations of the quantitative data obtained from cartilage across patients, or longitudinally in the same patients, bringing valuable spatial information on the patterns of changes occurring with the disease ^{102,103}. Although cartilage remains the hallmark of the disease, osteoarthritis is now largely considered as a condition affecting the whole joint, where all intra-articular tissues

including menisci, ligaments, bones, and synovium are involved. The interactions among knee structures, and between these structures and other factors involved in osteoarthritis are not fully understood, and there is a need for a multivariable assessment of imaging-derived quantitative data, integrated with other patient-derived information, including mechanical and genetical information^{104,105}. Analytical tools offered by machine learning provide an excellent opportunity to perform such analyses to advance in our understanding of osteoarthritis.

In conclusion, the use of quantitative analyses has been slow in translating into the clinical practice of musculoskeletal imaging, despite the general agreement that it increases the value of imaging examinations. This is at least partly due to the difficulties posed by image segmentation, which is a crucial step in the processing pipeline for extracting quantitative information from regions of interest in the image. Progress has been slow in automatizing this task, and segmentation remains challenging. However, atlas-based methods and deep learning approaches have recently shown promising results.

The next step of the radiomics pipeline consists of features extraction, leading to virtually an infinite number of features that may be analyzed in association with other non-imaging-derived information to build decision-making models. Musculoskeletal imaging has been lagging behind other subspecialties as to the adoption of radiomics, as shown by the literature. Apart from oncology, imaging of osteoporosis and osteoarthritis will most likely be the first to benefit from these new tools.

Overall, radiomics is still an emerging field, where the models that are generated still need to be validated following high-quality standards, which should lead radiomics to become a viable decision-making tool in the future.

Acknowledgement

The authors would like to thank Jonas Richiardi, Yasser Alemán-Gómez and Hugo Babel for their contribution to the illustrations.

The work is supported by the Centre d'Imagerie BioMédicale (CIBM) of the University of Lausanne (UNIL), the Swiss Federal Institute of Technology Lausanne (EPFL), the University of Geneva (UniGe), the Centre Hospitalier Universitaire Vaudois (CHUV), the Hôpitaux Universitaires de Genève (HUG), and the Leenaards and Jeantet Foundations.

References

1. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, For the Group. A Collaborative Enterprise for Multi-Stakeholder Participation in the Advancement of Quantitative Imaging. *Radiology*. 2011;258(3):906-914. doi:10.1148/radiol.10100799
2. Mansoor A, Bagci U, Foster B, et al. Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. *RadioGraphics*. 2015;35(4):1056-1076. doi:10.1148/rg.2015140232
3. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol*. 2018;102(4):1143-1158. doi:10.1016/j.ijrobp.2018.05.053
4. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol*. 2010;65(7):517-521. doi:10.1016/j.crad.2010.04.005
5. Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. 2017;28(6):1191-1206. doi:10.1093/annonc/mdx034
6. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169
7. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol*. 2017;86:297-307. doi:10.1016/j.ejrad.2016.09.005
8. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141
9. Bodalal Z, Trebeschi S, Beets-Tan R. Radiomics: a critical step towards integrated healthcare. *Insights Imaging*. 2018;9(6):911-914. doi:10.1007/s13244-018-0669-3
10. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. *Abdom Radiol*. 2019;44(6):1960-1984. doi:10.1007/s00261-019-02028-w
11. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Hand-Crafted to Deep Learning-based Cancer Radiomics: Challenges and Opportunities. *IEEE Signal Process Mag*. 2019;36(4):132-160. doi:10.1109/MSP.2019.2900993
12. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5(1):4006. doi:10.1038/ncomms5006
13. Aerts HJWL. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol*. 2016;2(12):1636-1642. doi:10.1001/jamaoncol.2016.2631

14. Ma Z, Tavares JMRS, Jorge RN, Mascarenhas T. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Comput Methods Biomech Biomed Engin.* 2010;13(2):235-246. doi:10.1080/10255840903131878
15. Pedoia V, Majumdar S, Link TM. Segmentation of joint and musculoskeletal tissue in the study of arthritis. *Magn Reson Mater Phys Biol Med.* 2016;29(2):207-221. doi:10.1007/s10334-016-0532-9
16. Arovitola A, Gallo L. Knee bone segmentation from MRI: A classification and literature review. *Biocybern Biomed Eng.* 2016;36(2):437-449. doi:10.1016/j.bbe.2015.12.007
17. Zhang B, Zhang Y, Cheng HD, et al. Computer-Aided Knee Joint Magnetic Resonance Image Segmentation - A Survey. *ArXiv180204894 Cs.* February 2018. <http://arxiv.org/abs/1802.04894>. Accessed August 16, 2019.
18. Ryzhkov MD. Knee Cartilage Segmentation Algorithms: a Critical Literature Review. <http://dspace.library.uu.nl/handle/1874/308831>. Published March 5, 2015. Accessed August 16, 2019.
19. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88. doi:10.1016/j.media.2017.07.005
20. John NW. Segmentation of Radiological Images. In: Neri E, Caramella D, Bartolozzi C, eds. *Image Processing in Radiology.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2008:45-54. doi:10.1007/978-3-540-49830-8_4
21. Dalvi R, Abugharbieh R, Wilson D, Wilson DR. Multi-contrast MR for enhanced bone imaging and segmentation. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf.* 2007;2007:5620-5623. doi:10.1109/IEMBS.2007.4353621
22. Li X, Ma BC, Bolbos RI, et al. Quantitative assessment of bone marrow edema-like lesion and overlying cartilage in knees with osteoarthritis and anterior cruciate ligament tear using MR imaging and spectroscopic imaging at 3 Tesla. *J Magn Reson Imaging.* 2008;28(2):453-461. doi:10.1002/jmri.21437
23. Li W, Abram F, Pelletier J-P, et al. Fully automated system for the quantification of human osteoarthritic knee joint effusion volume using magnetic resonance imaging. *Arthritis Res Ther.* 2010;12(5):R173. doi:10.1186/ar3133
24. Kim S, Lee D, Park S, Oh K-S, Chung SW, Kim Y. Automatic segmentation of supraspinatus from MRI by internal shape fitting and autocorrection. *Comput Methods Programs Biomed.* 2017;140:165-174. doi:10.1016/j.cmpb.2016.12.008
25. Hacihaliloglu I. Ultrasound Imaging and Segmentation of Bone Surfaces: A Review. *Technol Singap World Sci.* 2017;5(2):74-80. doi:10.1142/S2339547817300049
26. Zhang J, Yan C-H, Chui C-K, Ong S-H. Fast segmentation of bone in CT images using 3D adaptive thresholding. *Comput Biol Med.* 2010;40(2):231-236. doi:10.1016/j.combiomed.2009.11.020

27. Jena S, Sahu B, Jagadev AK. Analysis of Medical X-ray Bone Images Using Image Segmentation. In: Jain LC, Patnaik S, Ichalkaranje N, eds. *Intelligent Computing, Communication and Devices*. Advances in Intelligent Systems and Computing. Springer India; 2015:787-794.
28. van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng Phys*. 2018;51:6-16. doi:10.1016/j.medengphy.2017.10.008
29. Silva G, Oliveira L, Pithon M. Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Syst Appl*. 2018;107:15-31. doi:10.1016/j.eswa.2018.04.001
30. Szeliski R. *Computer Vision: Algorithms and Applications*. London: Springer-Verlag; 2011. <https://www.springer.com/gp/book/9781848829343>. Accessed August 16, 2019.
31. Serra J. Introduction to mathematical morphology. *Comput Vis Graph Image Process*. 1986;35(3):283-305. doi:10.1016/0734-189X(86)90002-2
32. Lee J-S, Chung Y-N. Integrating Edge Detection and Thresholding Approaches to Segmenting Femora and Patellae from Magnetic Resonance Images. *Biomed Eng Appl Basis Commun*. 2005;17(01):1-11. doi:10.4015/S1016237205000020
33. carballido. Inter-Subject Comparison of MRI Knee Cartilage Thickness. *Med Image Anal*. 2008;12(2):120-135. doi:10.1016/j.media.2007.08.002
34. Duryea J, Neumann G, Brem MH, et al. Novel fast semi-automated software to segment cartilage for knee MR acquisitions. *Osteoarthritis Cartilage*. 2007;15(5):487-492. doi:10.1016/j.joca.2006.11.002
35. Iranpour-Boroujeni T, Watanabe A, Bashtar R, Yoshioka H, Duryea J. Quantification of cartilage loss in local regions of knee joints using semi-automated segmentation software: analysis of longitudinal data from the Osteoarthritis Initiative (OAI). *Osteoarthritis Cartilage*. 2011;19(3):309-314. doi:10.1016/j.joca.2010.12.002
36. Brem MH, Lang PK, Neumann G, et al. Magnetic resonance image segmentation using semi-automated software for quantification of knee articular cartilage—initial evaluation of a technique for paired scans. *Skeletal Radiol*. 2009;38(5):505-511. doi:10.1007/s00256-009-0658-1
37. Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*. 1994;16(6):641-647. doi:10.1109/34.295913
38. Chang YL, Li X. Adaptive image region-growing. *IEEE Trans Image Process Publ IEEE Signal Process Soc*. 1994;3(6):868-872. doi:10.1109/83.336259
39. Waterton JC, Solloway S, Foster JE, et al. Diurnal variation in the femoral articular cartilage of the knee in young adult humans. *Magn Reson Med*. 2000;43(1):126-132. doi:10.1002/(SICI)1522-2594(200001)43:1<126::AID-MRM15>3.0.CO;2-#

40. Digabel H, LANTUEJOUL C. Iterative algorithms, special issues of practical metallography, vol. 8. 1977.
41. Meyer F. Topographic distance and watershed lines. *Signal Process.* 1994;38(1):113-125. doi:10.1016/0165-1684(94)90060-4
42. Kornilov A, Safonov I. An Overview of Watershed Algorithm Implementations in Open Source Libraries. *J Imaging.* 2018;4:123. doi:10.3390/jimaging4100123
43. Grau V, Mewes AUJ, Alcaniz M, Kikinis R, Warfield SK. Improved watershed transform for medical image segmentation using prior information. *IEEE Trans Med Imaging.* 2004;23(4):447-458. doi:10.1109/TMI.2004.824224
44. Włodarczyk J, Czaplicka K, Tabor Z, Wojciechowski W, Urbanik A. Segmentation of bones in magnetic resonance images of the wrist. *Int J Comput Assist Radiol Surg.* 2015;10(4):419-431. doi:10.1007/s11548-014-1105-x
45. Desai P, Hacihaliloglu I. Knee-Cartilage Segmentation and Thickness Measurement from 2D Ultrasound. *J Imaging.* 2019;5:43. doi:10.3390/jimaging5040043
46. Caselles V, Kimmel R, Sapiro G. Geodesic Active Contours. *Int J Comput Vis.* 1997;22(1):61-79. doi:10.1023/A:1007979827043
47. Malladi R, Sethian JA, Vemuri BC. Shape Modeling with Front Propagation: A Level Set Approach. *Ieee Trans Pattern Anal Mach Intelligence.* 1995;17:158–175.
48. Osher S, Sethian JA. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J Comput Phys.* 1988;79(1):12-49. doi:10.1016/0021-9991(88)90002-2
49. Sethian JA. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science.* 2nd ed. Cambridge, U.K. ; New York: Cambridge University Press; 1999.
50. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process.* 2001;10(2):266-277. doi:10.1109/83.902291
51. Li W, Abram F, Beaudoin G, Berthiaume M, Pelletier J, Martel-Pelletier* J. Human Hip Joint Cartilage: MRI Quantitative Thickness and Volume Measurements Discriminating Acetabulum and Femoral Head. *IEEE Trans Biomed Eng.* 2008;55(12):2731-2740. doi:10.1109/TBME.2008.925679
52. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. *Comput Vis Image Underst.* 1995;61:38-59. doi:10.1006/cviu.1995.1004
53. Cootes TF, Edwards GJ, Taylor CJ. Comparing Active Shape Models with Active Appearance Models. In: *BMVC.* ; 1999. doi:10.5244/c.13.18

54. Davies RH, Twining CJ, Cootes TF, Waterton JC, Taylor CJ. A minimum description length approach to statistical shape modeling. *IEEE Trans Med Imaging*. 2002;21(5):525-537. doi:10.1109/TMI.2002.1009388
55. Heimann T, Meinzer H-P. Statistical shape models for 3D medical image segmentation: A review. *Med Image Anal*. 2009;13(4):543-563. doi:10.1016/j.media.2009.05.004
56. Solloway S, Hutchinson CE, Waterton JC, Taylor CJ. The use of active shape models for making thickness measurements of articular cartilage from MR images. *Magn Reson Med*. 1997;37(6):943-952.
57. Fripp J, Crozier S, Warfield SK, Ourselin S. Automatic Segmentation and Quantitative Analysis of the Articular Cartilages From Magnetic Resonance Images of the Knee. *IEEE Trans Med Imaging*. 2010;29(1):55-64. doi:10.1109/TMI.2009.2024743
58. Fripp J, Crozier S, Warfield SK, Ourselin S. Automatic segmentation of the bone and extraction of the bone-cartilage interface from magnetic resonance images of the knee. *Phys Med Biol*. 2007;52(6):1617-1631. doi:10.1088/0031-9155/52/6/005
59. Schmid J, Kim J, Magnenat-Thalmann N. Robust statistical shape models for MRI bone segmentation in presence of small field of view. *Med Image Anal*. 2011;15(1):155-168. doi:10.1016/j.media.2010.09.001
60. Bach Cuadra M, Duay V, Thiran J-Ph. Atlas-based segmentation. In: *Handbook of Biomedical Imaging*. ; 2015:221-244. doi:10.1007/978-0-387-09749-7_12
61. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. *Med Image Anal*. 2015;24(1):205-219. doi:10.1016/j.media.2015.06.012
62. Kapur T, Beardsley P, Frisken S, Eric W, Grimson L, Wells W. Model Based Segmentation of Clinical Knee MRI. January 1998.
63. Lee J-G, Gumus S, Moon CH, Kwok CK, Bae KT. Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method. *Med Phys*. 2014;41(9). doi:10.1118/1.4893533
64. Chen X, Pan L. A Survey of Graph Cuts/Graph Search Based Medical Image Segmentation. *IEEE Rev Biomed Eng*. 2018;11:112-124. doi:10.1109/RBME.2018.2798701
65. Boykov Y, Veksler O, Zabih R. Fast Approximate Energy Minimization via Graph Cuts. :8.
66. Shim H, Chang S, Tao C, Wang J-H, Kwok CK, Bae KT. Knee cartilage: efficient and reproducible segmentation on high-spatial-resolution MR images with the semiautomated graph-cut algorithm method. *Radiology*. 2009;251(2):548-556. doi:10.1148/radiol.2512081332
67. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer; 2006.

68. Zhang D-Q, Chen S-C. A novel kernelized fuzzy C-means algorithm with application in medical image segmentation. *Artif Intell Med.* 2004;32(1):37-50. doi:10.1016/j.artmed.2004.01.012
69. Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Springer US; 1981. <https://www.springer.com/gp/book/9781475704525>. Accessed August 26, 2019.
70. Carballido-Gamio J, Phan C, Link TM, Majumdar S. Characterization of Trabecular Bone Structure from High-Resolution Magnetic Resonance Images Using Fuzzy Logic. *Magn Reson Imaging.* 2006;24(8):1023-1029. doi:10.1016/j.mri.2006.04.010
71. Folkesson J, Carballido-Gamio J, Eckstein F, Link TM, Majumdar S. Local bone enhancement fuzzy clustering for segmentation of MR trabecular bone images. *Med Phys.* 2010;37(1):295-302. doi:10.1118/1.3264615
72. Folkesson J, Dam EB, Olsen OF, Pettersen PC, Christiansen C. Segmenting Articular Cartilage Automatically Using a Voxel Classification Approach. *IEEE Trans Med Imaging.* 2007;26(1):106-115. doi:10.1109/TMI.2006.886808
73. Zhang K, Lu W, Marziliano P. Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies. *Magn Reson Imaging.* 2013;31(10):1731-1743. doi:10.1016/j.mri.2013.06.005
74. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* Cambridge, Massachusetts: The MIT Press; 2016.
75. Hirschmann A, Cyriac J, Stieltjes B, Kober T, Richiardi J, Omoumi P. Artificial Intelligence in Musculoskeletal Imaging: Review of Current Literature, Challenges, and Trends. *Semin Musculoskelet Radiol.* 2019;23(3):304-311. doi:10.1055/s-0039-1684024
76. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network. In: Salinesi C, Norrie MC, Pastor Ó, eds. *Advanced Information Systems Engineering.* Vol 7908. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013:246-253. doi:10.1007/978-3-642-40763-5_31
77. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med.* 2018;79(4):2379-2391. doi:10.1002/mrm.26841
78. Heimann T, Morrison BJ, Styner MA, Niethammer M, Warfield SK. Segmentation of Knee Images: A Grand Challenge. :8.
79. Omoumi P, Becce F, Ott J, Racine D, Verdun F. Optimization of Radiation Dose and Image Quality in Musculoskeletal CT: Emphasis on Iterative Reconstruction Techniques (Part 1). *Semin Musculoskelet Radiol.* 2015;19(05):415-421. doi:10.1055/s-0035-1569255

80. Omoumi P, Verdun F, Becce F. Optimization of Radiation Dose and Image Quality in Musculoskeletal CT: Emphasis on Iterative Reconstruction Techniques (Part 2). *Semin Musculoskelet Radiol*. 2015;19(05):422-430. doi:10.1055/s-0035-1569254
81. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. 2018;2(1):36. doi:10.1186/s41747-018-0068-z
82. Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D, Müller H. Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Med Image Anal*. 2014;18(1):176-196. doi:10.1016/j.media.2013.10.005
83. Miles KA, Ganeshan B, Hayball MP. CT texture analysis using the filtration-histogram method: what do the measurements mean? *Cancer Imaging*. 2013;13(3):400-406. doi:10.1102/1470-7330.2013.9045
84. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics*. 2017;37(2):505-515. doi:10.1148/rg.2017160130
85. Yin P, Mao N, Zhao C, Wu J, Chen L, Hong N. A Triple-Classification Radiomics Model for the Differentiation of Primary Chordoma, Giant Cell Tumor, and Metastatic Tumor of Sacrum Based on T2-Weighted and Contrast-Enhanced T1-Weighted MRI: A Triple-Classification Radiomics Model for the Differentiation of Primary Chordoma, Giant Cell Tumor, and Metastatic Tumor of Sacrum. *J Magn Reson Imaging*. 2019;49(3):752-759. doi:10.1002/jmri.26238
86. Park JE, Kim D, Kim HS, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol*. July 2019. doi:10.1007/s00330-019-06360-z
87. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63. doi:10.7326/M14-0697
88. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *ArXiv161207003 Cs Eess*. December 2016. <http://arxiv.org/abs/1612.07003>. Accessed September 1, 2019.
89. Yi J, Lee YH, Kim SK, et al. Response evaluation of giant-cell tumor of bone treated by denosumab: Histogram and texture analysis of CT images. *J Orthop Sci*. 2018;23(3):570-577. doi:10.1016/j.jos.2018.01.006
90. Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures: *Curr Opin Rheumatol*. 2019;31(4):368-375. doi:10.1097/BOR.0000000000000607
91. Ho-Le TP, Center JR, Eisman JA, Nguyen TV, Nguyen HT. Prediction of hip fracture in post-menopausal women using artificial neural network approach. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf*. 2017;2017:4207-4210. doi:10.1109/EMBC.2017.8037784

92. Kruse C, Eiken P, Vestergaard P. Clinical fracture risk evaluated by hierarchical agglomerative clustering. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA*. 2017;28(3):819-832. doi:10.1007/s00198-016-3828-8
93. Kruse C, Eiken P, Vestergaard P. Machine Learning Principles Can Improve Hip Fracture Prediction. *Calcif Tissue Int*. 2017;100(4):348-360. doi:10.1007/s00223-017-0238-7
94. Areckal AS, Jayasheelan N, Kamath J, Zawadynski S, Kocher M, David S S. Early diagnosis of osteoporosis using radiogrammetry and texture analysis from hand and wrist radiographs in Indian population. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA*. 2018;29(3):665-673. doi:10.1007/s00198-017-4328-1
95. Singh A, Dutta MK, Jennane R, Lespessailles E. Classification of the trabecular bone structure of osteoporotic patients using machine vision. *Comput Biol Med*. 2017;91:148-158. doi:10.1016/j.combiomed.2017.10.011
96. Valentinitzsch A, Trebeschi S, Kaesmacher J, et al. Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA*. 2019;30(6):1275-1285. doi:10.1007/s00198-019-04910-1
97. Muehlematter UJ, Mannil M, Becker AS, et al. Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning. *Eur Radiol*. 2019;29(5):2207-2217. doi:10.1007/s00330-018-5846-8
98. Dimai HP, Ljuhar R, Ljuhar D, et al. Assessing the effects of long-term osteoporosis treatment by using conventional spine radiographs: results from a pilot study in a sub-cohort of a large randomized controlled trial. *Skeletal Radiol*. 2019;48(7):1023-1032. doi:10.1007/s00256-018-3118-y
99. Carballido-Gamio J, Majumdar S. Atlas-Based Knee Cartilage Assessment. *Magn Reson Med*. 2011;66(2):574-583. doi:10.1002/mrm.22836
100. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging. *Magn Reson Med*. 2018;79(4):2379-2391. doi:10.1002/mrm.26841
101. Ashinsky BG, Bouhrara M, Coletta CE, et al. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *J Orthop Res*. 2017;35(10):2243-2250. doi:10.1002/jor.23519
102. Pedoia V, Li X, Su F, Calixto N, Majumdar S. Fully automatic analysis of the knee articular cartilage $T_{1\rho}$ relaxation time using voxel-based relaxometry: $T_{1\rho}$ VBR for Local Cartilage Evaluation. *J Magn Reson Imaging*. 2016;43(4):970-980. doi:10.1002/jmri.25065

103. Favre J, Erhart-Hledik JC, Blazek K, Fasel B, Gold GE, Andriacchi TP. Anatomically Standardized Maps Reveal Distinct Patterns of Cartilage Thickness With Increasing Severity of Medial Compartment Knee Osteoarthritis: CARTILAGE THICKNESS MAPS WITH OA. *J Orthop Res.* 2017;35(11):2442-2451. doi:10.1002/jor.23548
104. Edd SN, Omoumi P, Andriacchi TP, Jolles BM, Favre J. Modeling knee osteoarthritis pathophysiology using an integrated joint system (IJS): a systematic review of relationships among cartilage thickness, gait mechanics, and subchondral bone mineral density. *Osteoarthritis Cartilage.* 2018;26(11):1425-1437. doi:10.1016/j.joca.2018.06.017
105. Andriacchi TP, Favre J, Erhart-Hledik JC, Chu CR. A systems view of risk factors for knee osteoarthritis reveals insights into the pathogenesis of the disease. *Ann Biomed Eng.* 2015;43(2):376-387. doi:10.1007/s10439-014-1117-2
106. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage.* 2006;31(3):1116-1128. doi:10.1016/j.neuroimage.2006.01.015

Figures and figure legends

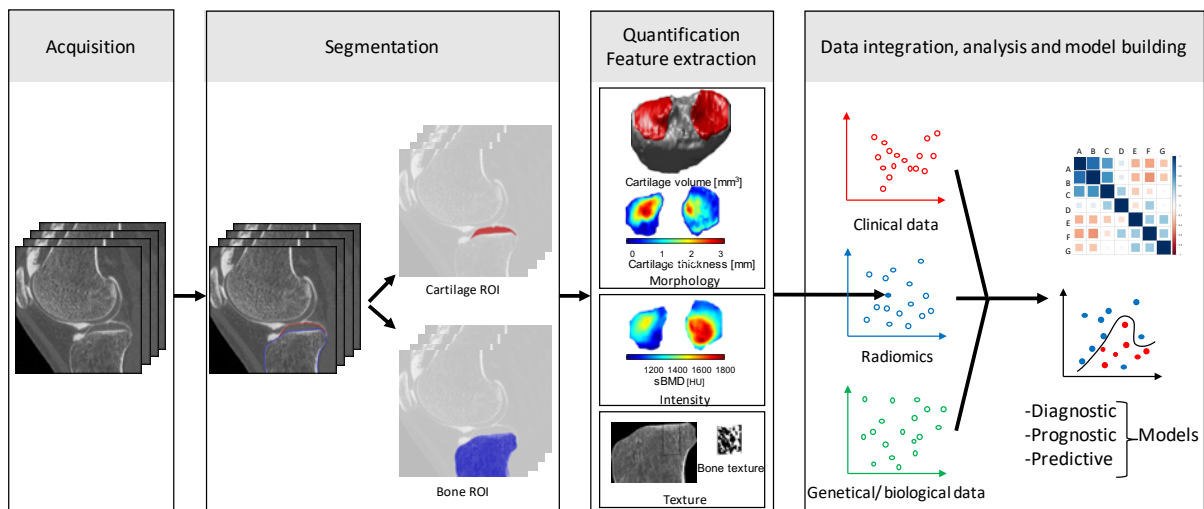


Figure 1: Image quantification analysis workflow and conventional radiomics pipeline. The first step of the process is image acquisition (in this case CT arthrography). Second, images are segmented to isolated regions of interest (ROI) (tibial cartilage and bone in this case). Quantitative information, called radiomics features, are extracted from the ROIs (here morphological features including cartilage volume and thickness, first-order features (signal intensity) and second-order features (bone texture)). The extracted radiomics features are then integrated with other information derived from the patient including clinical, genetical, or biological data. Data analysis is performed to build diagnostic, prognostic or predictive models.

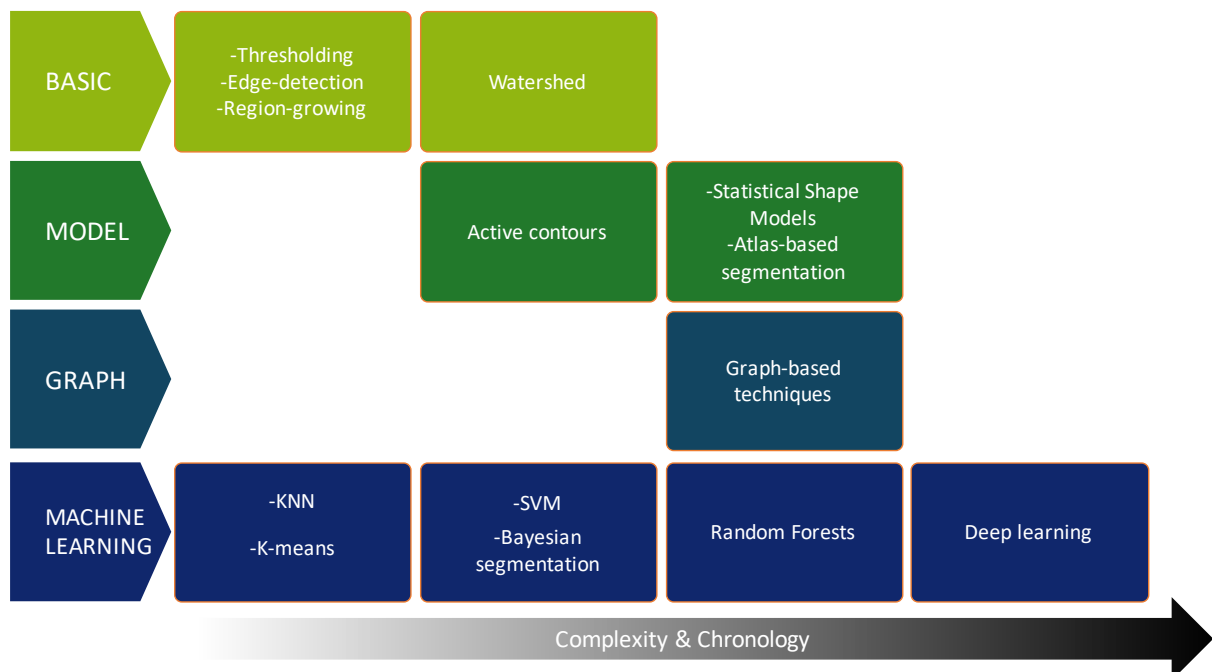


Figure 2: Overview of the different segmentation methods. Methods are grouped into four families (basic, model-based, graph-based and machine learning-based techniques (from

top to bottom rows)). The horizontal axis represents their increasing complexity and/or chronology of adoption by the medial image community.
KNN: K-nearest neighbor, SVM: Support Vector Machines.

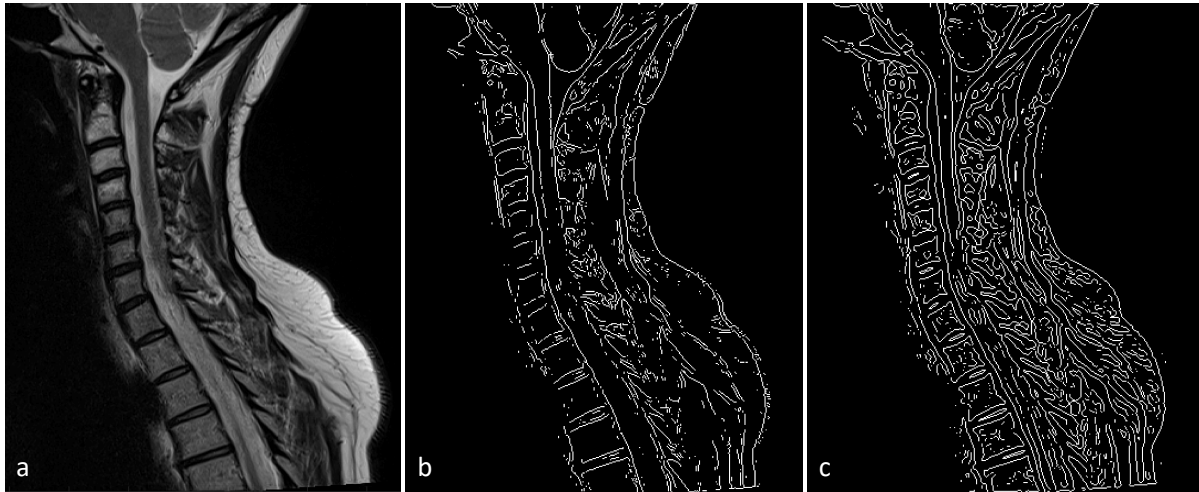


Figure 3: Contour extraction using different edge-detector filters. (a): original T2-weighted sagittal MRI image of the spine. (b): image filtered based on the first derivative of the image intensity (Sobel). (c): image filtered based on the second derivative of the image intensity (Laplacian of Gaussian). Note the differences in edge detection between (b) and (c).

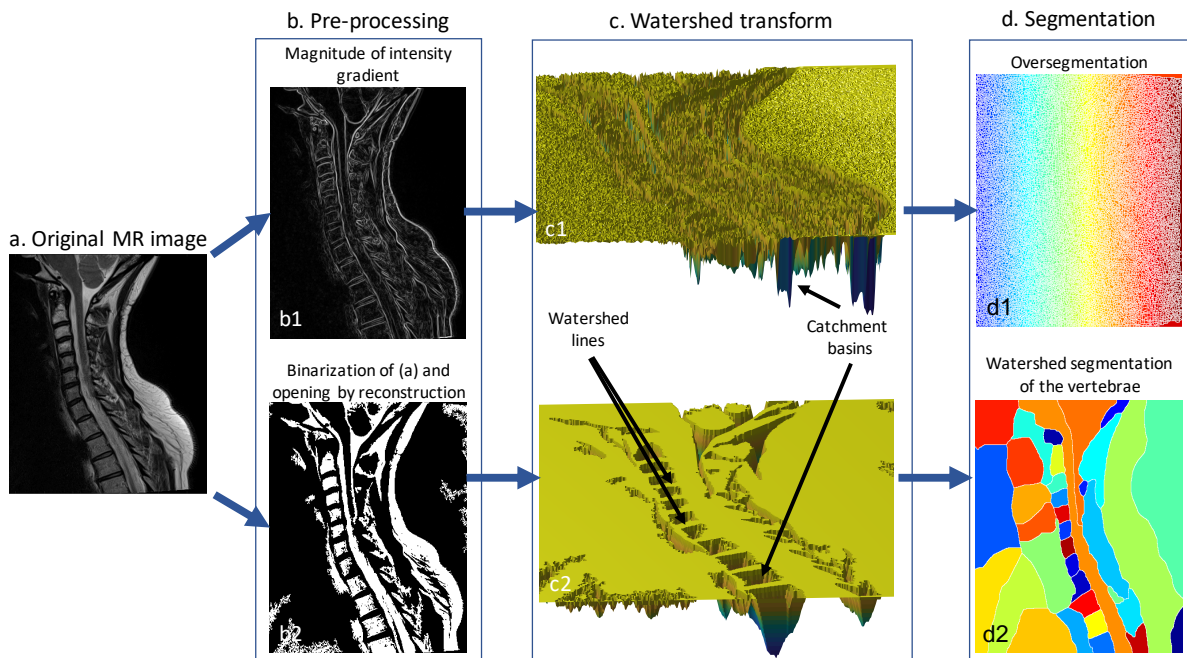


Figure 4: Watershed segmentation for two different pre-processing approaches. A) Original MRI intensity. (b1) Magnitude of intensity image and (c1) watershed transform of (b1) illustrating many regional minima, forming tiny catchment basins, resulting in oversegmentation problem as illustrated in (d1). (b2) Another pre-processing approach based on mathematical morphology to remove regional minima. A distance map of (b2) is then computed and used as input for the watershed transform, as illustrated in (c2). The

final segmentation contains well-distinguished segments for the vertebra, and the non-desired bigger regions can be easily identified and removed with post-processing.

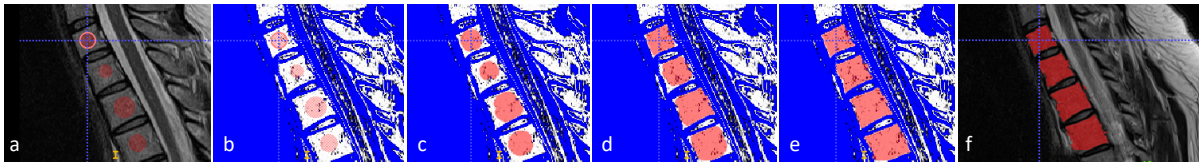


Figure 5: Active contour segmentation¹⁰⁶. (a) Original sagittal T2-weighted MR image of the cervical spine. (b) The image is preprocessed based on region homogeneity. One initial closed contour (or snake) is placed per structure to be segmented. (c), (d) and (e) represent the snake evolution at iterations 46, 92, and 200 respectively. (f) Overlay of the final segmentation on the original MR image.

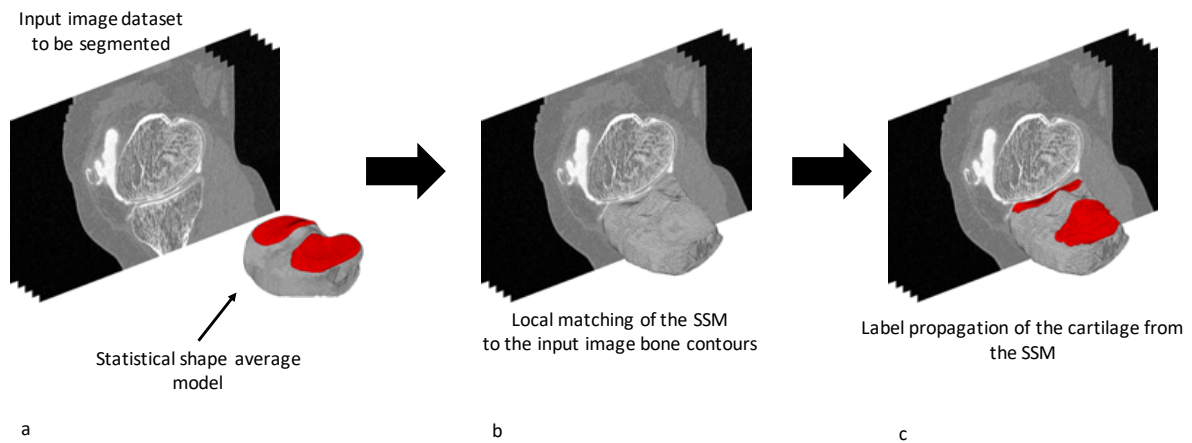


Figure 6: Statistical shape model (SSM) segmentation of the knee bone and cartilage. (a) A statistical shape average model is built from several manually segmented image sets. Initial global positioning of the average shape can be done via point-based registration. (b) Local adaptation of the SSM is performed based on both contours and intensity profiles across bone contours. (c) Finally, label propagation of cartilage is applied.

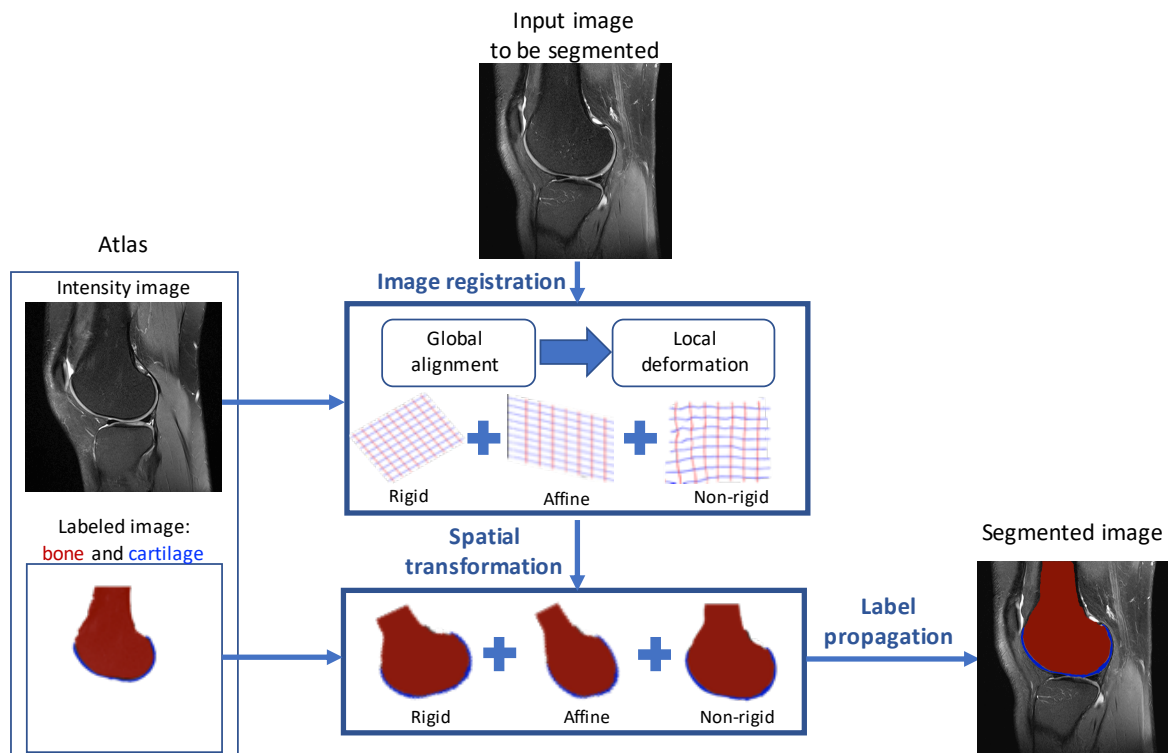


Figure 7: Atlas-based segmentation. Image registration is performed between the atlas image and the input image to be segmented. Point to point spatial correspondence is often done using voxel-wise volume registration between the two gray-level images (atlas and input image). The image registration consists in a combination of global alignment (rigid and affine transformations) and a local non-rigid deformation. This spatial transformation is then applied to the labeled image and the label is propagated to the input image, giving the segmented image.

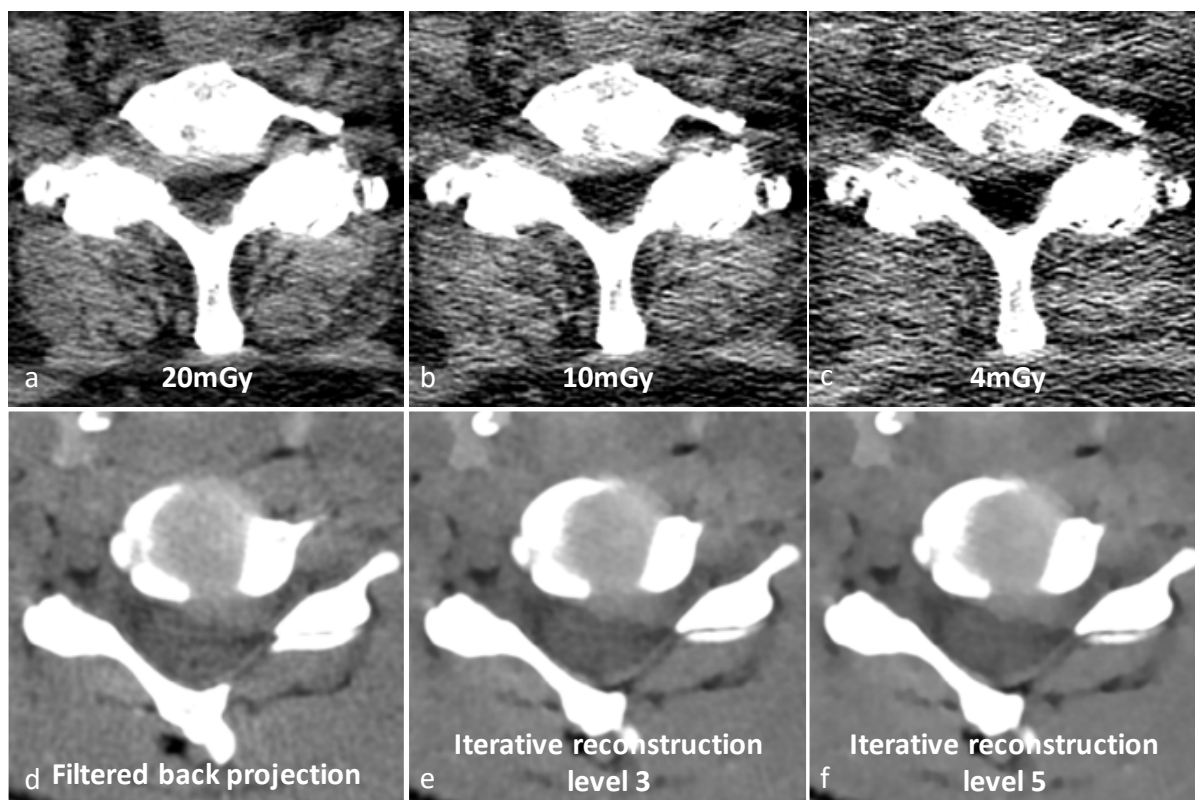


Figure 8: Influence of acquisition and reconstruction parameters on image aspect, hereby potentially affecting segmentation and feature extraction. (a-c) Axial CT of cadaveric spine shows decreasing levels of radiation dose, associated with increased image noise and artifacts ⁷⁹. (d-f) Axial CT of cadaveric spine shows the influence of different reconstructions algorithms on image noise and texture ⁸⁰.

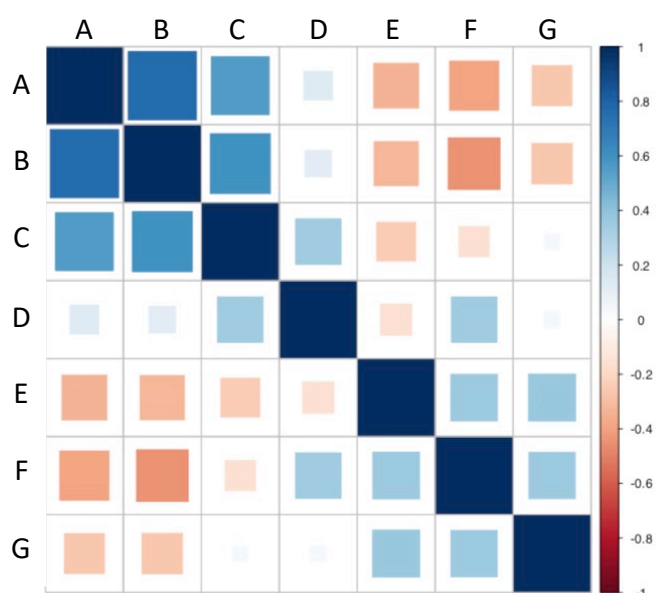


Figure 9: Correlation matrix used for radiomics feature selection. Correlation coefficients are color-coded (according to the scale on right side) and displayed for each pair of radiomics features (A to G). Highly correlated features are redundant and are grouped in (A and B are highly correlated in this example). One feature is chosen to be representative of

each cluster. Note that only seven features were included in this matrix for sake of simplicity, but correlation matrices may contain a much higher number of features.