



OPEN ACCESS

EDITED BY

Xin Gao,
King Abdullah University of Science and
Technology, Saudi Arabia

REVIEWED BY

Chi-Wen Lung,
Asia University, Taiwan
Ashley Weaver,
Wake Forest University, United States
Diana Madrid Fuentes,
Wake Forest University, United States in
collaboration with reviewer AW

*CORRESPONDENCE

Sokratis Makrogiannis,
smakrogiannis@desu.edu

SPECIALTY SECTION

This article was submitted to
Computational Physiology and
Medicine,
a section of the journal
Frontiers in Physiology

RECEIVED 23 May 2022

ACCEPTED 26 September 2022

PUBLISHED 14 October 2022

CITATION

Makrogiannis S, Okorie A, Di Iorio A,
Bandinelli S and Ferrucci L (2022), Multi-
atlas segmentation and quantification of
muscle, bone and subcutaneous
adipose tissue in the lower leg using
peripheral quantitative
computed tomography.
Front. Physiol. 13:951368.
doi: 10.3389/fphys.2022.951368

COPYRIGHT

© 2022 Makrogiannis, Okorie, Di Iorio,
Bandinelli and Ferrucci. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Multi-atlas segmentation and quantification of muscle, bone and subcutaneous adipose tissue in the lower leg using peripheral quantitative computed tomography

Sokratis Makrogiannis^{1*}, Azubuike Okorie¹, Angelo Di Iorio²,
Stefania Bandinelli³ and Luigi Ferrucci⁴

¹Math Imaging and Visual Computing Lab, Division of Physics, Engineering, Mathematics and Computer Science, Delaware State University, Dover, DE, United States, ²Antalgic Mini-invasive and Rehab-Outpatients Unit, Department of Innovative Technologies in Medicine & Dentistry, University "G.d'Annunzio", Chieti-Pescara, Italy, ³Geriatric Unit, Local Health Unit Tuscany Centre, Florence, Italy, ⁴National Institute on Aging, National Institutes of Health, Baltimore, MD, United States

Accurate and reproducible tissue identification is essential for understanding structural and functional changes that may occur naturally with aging, or because of a chronic disease, or in response to intervention therapies. Peripheral quantitative computed tomography (pQCT) is regularly employed for body composition studies, especially for the structural and material properties of the bone. Furthermore, pQCT acquisition requires low radiation dose and the scanner is compact and portable. However, pQCT scans have limited spatial resolution and moderate SNR. pQCT image quality is frequently degraded by involuntary subject movement during image acquisition. These limitations may often compromise the accuracy of tissue quantification, and emphasize the need for automated and robust quantification methods. We propose a tissue identification and quantification methodology that addresses image quality limitations and artifacts, with increased interest in subject movement. We introduce a multi-atlas image segmentation (MAIS) framework for semantic segmentation of hard and soft tissues in pQCT scans at multiple levels of the lower leg. We describe the stages of statistical atlas generation, deformable registration and multi-tissue classifier fusion. We evaluated the performance of our methodology using multiple deformable registration approaches against reference tissue masks. We also evaluated the performance of conventional model-based segmentation against the same reference data to facilitate comparisons. We studied the effect of subject movement on tissue segmentation quality. We also applied the top performing method to a larger out-of-sample dataset and report the quantification results. The results show that multi-atlas image segmentation with diffeomorphic deformation and probabilistic label fusion produces very good quality over all tissues, even for scans with significant quality degradation. The application of our technique to the larger dataset reveals trends of age-related body composition changes that are consistent with the literature.

Because of its robustness to subject motion artifacts, our MAIS methodology enables analysis of larger number of scans than conventional state-of-the-art methods. Automated analysis of both soft and hard tissues in pQCT is another contribution of this work.

KEYWORDS

tissue segmentation, tissue quantification, multi-atlas techniques, subject movement, clinical application, pQCT

1 Introduction

Accurate segmentation of tissues using medical imaging is key for the quantification of changes in the structure and composition of tissues, which may result from diseases, aging, and other risk factors related to the tissue(s) in question (Cordova et al., 2014; Töpfer et al., 2015; Owen et al., 2019; Rodrigues and Pinheiro 2019; Wong and Manske 2020). Recent advances of artificial intelligence (AI) in the field of medical imaging have also drawn the interest of researchers in the application of computer techniques in the area of bone and muscle imaging (Burns et al., 2020). In clinical studies, segmentation of bone, muscle, and adipose tissue can be used for computing objective measures and descriptors of body composition and for exploring the causes and effects of differences of these descriptors between subject groups (Lauretani et al., 2008; Makrogiannis et al., 2018).

In the past 2 decades, peripheral quantitative computed tomography (pQCT) and high resolution peripheral quantitative computed tomography (HR-pQCT) have emerged as essential technologies for segmentation and quantification of bone, muscle and adipose tissue properties at the diaphyseal regions of the limbs. Segmentation of hard and soft tissues in pQCT and HR-pQCT imaging has been used to assess the effects of type-2 diabetes mellitus (T2DM) (Starr et al., 2018), osteoporosis (Simon et al., 2022) and osteoarthritis (Chen et al., 2018), to establish measures for characterizing sex-, ethnic-, site-, and age-related outcomes (Gabel et al., 2018), to study the effect of exercise on the muscle and fat cross-sectional areas (Rowe et al., 2019), and in studies of aging and age-related diseases (Chow et al., 2022; Liu et al., 2022). A challenge in pQCT-based segmentation is subject movement and the associated motion artifacts. Subject movement, subtle or obvious, occurs frequently in standard pQCT and HR-pQCT scans (Wong 2016). It may degrade the image quality and affect the assessment of bone and muscle properties (Pialat et al., 2012; Chan et al., 2018). pQCT motion has been assessed by visual inspection followed by a pass or fail decision. Usual criteria are the presence of discontinuities and streaks at the cortical bone and changes in intensity of trabecular bone. Motion streaks originating from the cortical bone extend into the muscle. Quantitative evaluation methods have been proposed for pQCT (Blew et al., 2014) and HR-pQCT (Pauchard et al., 2011; Sode et al., 2011). Thresholding (Blew et al., 2014) and

watershed segmentation techniques (Wong 2016) have been employed to identify and assess motion streaks in the muscle. Motion artifact correction and automatic segmentation are desirable.

However, to the best of our knowledge, there is no previous report in the literature on pQCT segmentation techniques explicitly addressing subject movement and the limited contrast-to-noise ratio that are characteristic of this modality. In this work we propose to address this gap by developing a multi-atlas image segmentation (MAIS) framework (Rohlfing et al., 2001, 2005; Shen and Hammer, 2002; Langerak et al., 2010; Sotiras et al., 2013; Iglesias and Sabuncu 2015) for identification of soft and hard tissues in pQCT scans of the lower leg. The MAIS framework includes the stages of statistical atlas generation, linear and non-linear registration, and label fusion for tissue segmentation. In these stages we use pQCT images of the lower leg at 4%, 38%, and 66% of the tibial length. We validated segmentation performance against manual reference masks using the Dice Similarity Coefficient (DSC). We evaluated the performance of multiple atlas-based tissue segmentation techniques and an established model-based tissue segmentation technique. We expect that the segmentation performance of multi-atlas based methods is largely unaffected by motion. We have also applied our framework to a larger out-of-sample dataset and reported our results on age-related tissue composition changes. Figure 1 summarizes the main components of the proposed framework and related experiments.

2 Our methodology

2.1 Atlas-based tissue segmentation

We formulate the problem of atlas-based segmentation next. Given a subject S , an atlas A and its atlas label map SA , we aim to produce the segmentation of S by warping the atlas to the spatial domain of the subject. This stage is called image registration, and is followed by pixel-wise assignment of tissue labels from the warped segmentation atlas to the subject (Figure 2). Since a single atlas is used for segmentation, we refer to this method as *single atlas image segmentation (SAIS)*.

In multi-atlas image segmentation (MAIS), multiple atlases A_i , $i = 1, \dots, N$ and corresponding segmented atlases SA_i , $i = 1,$

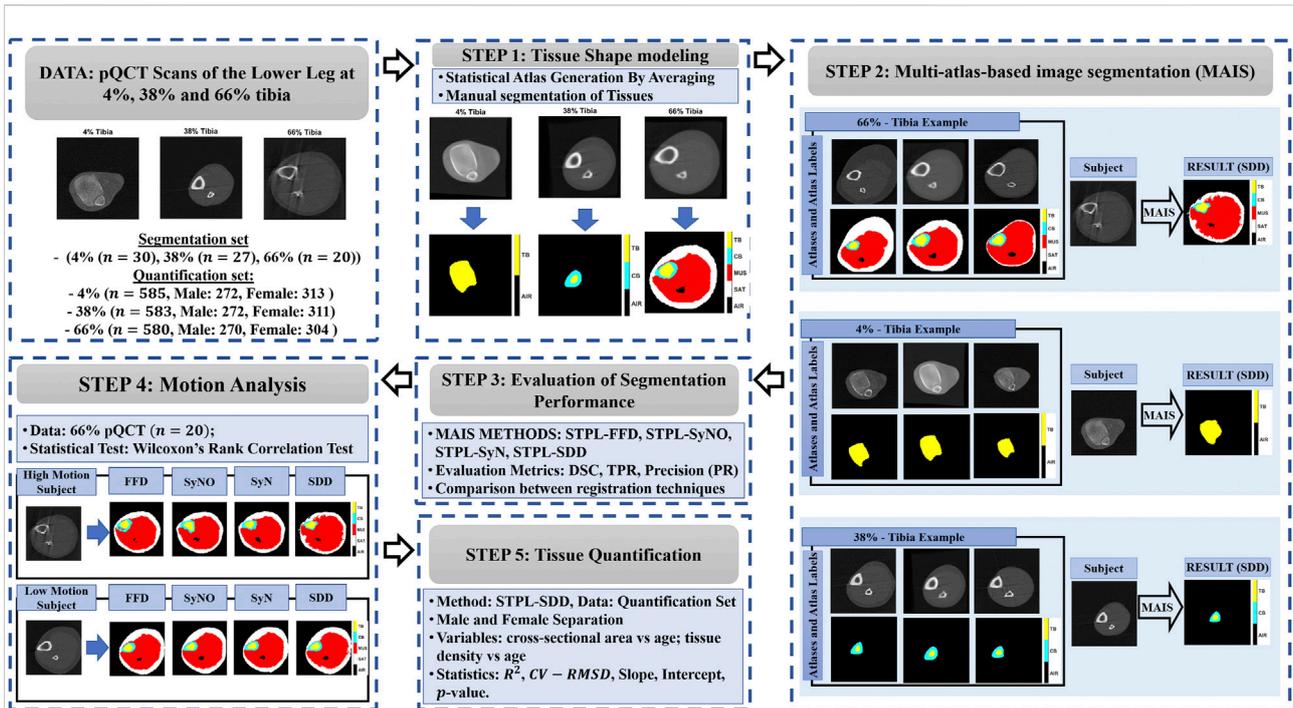


FIGURE 1

Main components of the proposed framework. STAPLE (STPL), free-form deformation (FFD), symmetric diffeomorphic demons (SDD), symmetric normalization (SyN), symmetric normalization - only (SyNO), dice similarity coefficient (DSC), true positive rate (TPR), precision (PR) squared-Spearman's correlation coefficient (R^2), and coefficient of variation - root mean squared difference (CV-RMSD).

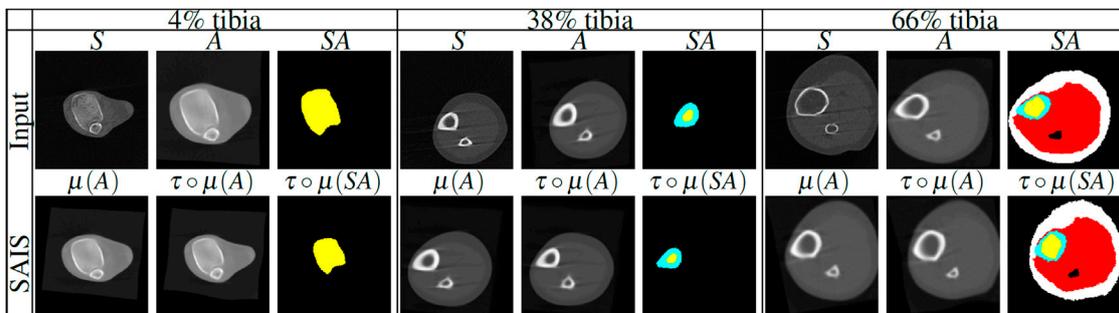


FIGURE 2

Illustration of single atlas based image segmentation (SAIS) stages of lower leg scans at 4%, 38% and 66% tibia. First row: subject (S), statistical atlas (A) and segmented atlas label map (SA) for each tibial length that are used as inputs for segmentation. Second row: linear registration, nonlinear registration and label propagation output. The tissue labels are color-coded as follows: trabecular bone (yellow), cortical bone (cyan), muscle (red), and SAT (white). μ denotes linear transformation, $\tau \circ \mu$ denotes the composition of linear (μ) with nonlinear (τ) transformations.

... , N are used to produce a segmentation of the subject S. The main stages of a multi-atlas-based segmentation algorithm are registration, label propagation and label fusion. Figure 3 shows the main stages of MAIS. In the following subsections, we elaborate on our method based on these stages. We will use A when referring to each atlas $A_i, i \in \{1, \dots, N\}$.

2.2 Image registration

Image registration is a key stage of atlas-based segmentation as described in the previous section. The goal of registration is to align the spatial domain of a subject with that of an atlas. In other words, we wish to find the optimal

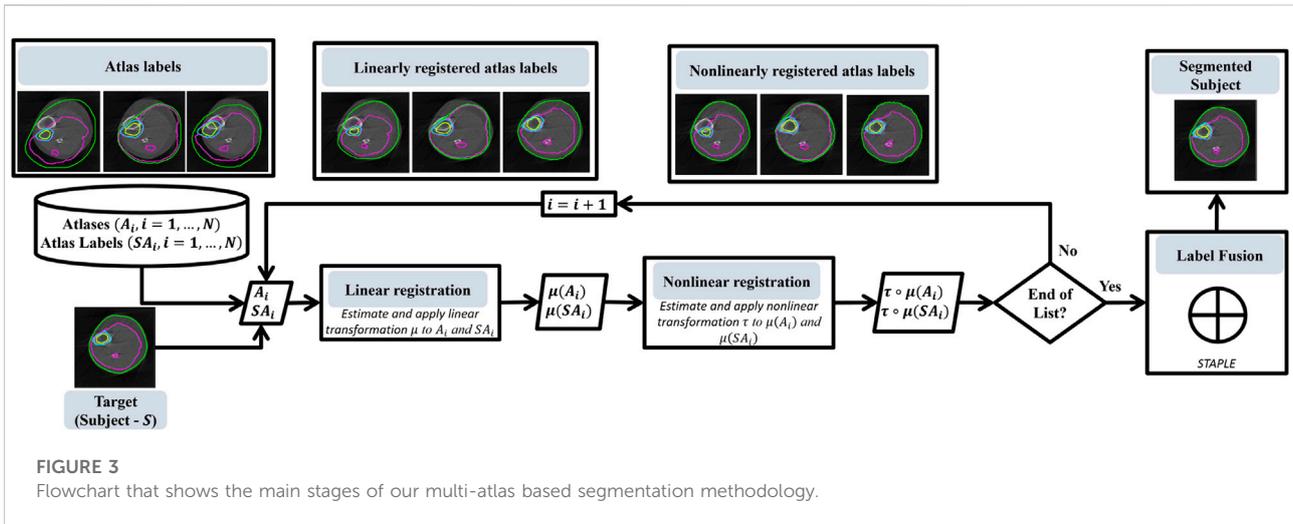


TABLE 1 Tissues to quantify in each anatomical site.

Anatomical site	Tissues to quantify
4% tibia	Trabecular Bone (TB)
38% tibia	Cortical Bone (CB), Trabecular Bone
66% tibia	Cortical Bone, Trabecular Bone, Subcutaneous Adipose Tissue (SAT), Muscle

deformation τ^* in the set $\{\tau, \tau : (A, \Omega_A) \rightarrow (S, \Omega_S)\}$ of all transformations from the spatial domain of the source image (or atlas) (A, Ω_A) to the subject space (S, Ω_S) , that minimizes the following energy functional (Sotiras et al., 2013; Vercauteren et al., 2007),

$$\mathcal{E}(\tau) := \lambda_\zeta \zeta(S, A \circ \tau) + \lambda_\rho \rho(\tau), \quad (1)$$

where ζ is the similarity term, ρ is the regularization term, λ_ζ and λ_ρ are the weights of the similarity and regularization terms, respectively. Therefore,

$$\tau^* = \arg \min_{\tau} \mathcal{E}(\tau). \quad (2)$$

To ensure accurate registration we applied both linear and nonlinear registrations. Linear registration is used to capture the rigid displacement of the subject while nonlinear registration is used to capture the local deformation of the anatomical structures of the subject. Anatomical structures are the tissue types, i.e., the trabecular bone, cortical bone, muscle, and subcutaneous adipose tissue as shown in Table 1. Figure 2 shows examples of rigid displacements of atlas A denoted by $\mu(A)$, and local deformations $\tau \circ \mu$ produced by nonlinear registration $\tau \circ \mu$. Our techniques for linear and nonlinear registration are described below.

2.2.1 Linear registration

Given a subject S and for each atlas A , we estimate the parameters of a linear transformation μ from atlas space (A, Ω_A) to the subject space (S, Ω_S) that defines the rigid motion between the atlas and the subject. Our linear transformation is modeled using affine transformations. We utilize the Mattes' mutual information similarity measure (Mattes et al., 2003) given in Eq. 3,

$$\zeta(S, A \circ \mu) = - \sum_{\iota} \sum_{\kappa} p(\iota, \kappa | \mu) \log \frac{p(\iota, \kappa | \mu)}{p_A(\iota | \mu) p_S(\kappa)} \quad (3)$$

and a regular step gradient descent optimizer, to estimate the parameters of the affine transformation μ , where p , is the joint probability distribution of subject and the atlas, p_A and p_S are the marginal probability distributions of the atlas and the subject respectively, $\iota = 1, \dots, n_A$ and $\kappa = 1, \dots, n_S$ are the indices of the histogram bins for the source and target image. The image $\mu(A)$ of the atlas A lives in the subject space and approximates the rigid motion between the subject and the atlas (Figure 2).

2.2.2 Nonlinear registration

In this stage, the goal is to correct the local deformations between the subject and the atlas. Hence we seek the parameters of deformation $\tau : (\mu(A), \Omega_S) \rightarrow (S, \Omega_S)$ from $\mu(A) \subset S$ onto S . We introduce the use of three nonlinear deformable registration techniques, free-form deformation (FFD), symmetric diffeomorphic demons (SDD), and symmetric image normalization (SyN, SyNO) for our multi-atlas-based registration. We discuss the nonlinear methods for our multi-atlas-based image segmentation below.

2.2.2.1 Free-form deformations

Free-form deformation was originally proposed by Rueckert et al., 1999 and was applied to automated registration of breast

MRI scans. It uses a spline interpolation kernel to compute the deformation values between the control points that produces a locally controlled, globally smooth transformation.

Given a 2D spatial domain $\Omega = \{\mathbf{x} = (x, y) | 0 \leq x < X, 0 \leq y < Y\}$ of an image, let Φ denote an $n_x \times n_y$ mesh of control points $\phi_{i,j} = (i\delta, j\delta)$ with uniform spacing δ . Ruckert et al., 1999 proposed a method that seeks the optimal FFD (displacement field τ) written as the 2-D tensor product of 1-D cubic B-Splines

$$\tau(\mathbf{x}) = \sum_{l=0}^2 \sum_{m=0}^2 B_l(u)B_m(v)\phi_{i+l,j+m} \tag{4}$$

that optimizes the energy functional in Eq. 1. In Eq. 4, $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$, and B_l represents the l th basis function of the B-spline. The regularization term ρ is given by the bending energy of a thin-plate of metal, which controls the smoothness of the transformation, defined by

$$\rho(\tau) = \frac{1}{|\Omega|} \iint_{\Omega} \left(\frac{\partial \tau}{\partial x} + \frac{\partial \tau}{\partial y} \right)^2 d\Omega, \tag{5}$$

where $|\Omega|$ is the cardinality of the image spatial domain. We employed Matte’s mutual information (Mattes et al., 2003) defined in Eq. 3 as the similarity measure ζ . We used $\lambda_\zeta = -1$ and $\lambda_\rho = 0.01$ in the energy function of Eq. 1. The optimization process is based on updating control points *via* the gradient of the cost function. We employed Limited-memory Broyden–Fletcher–Goldfarb–Shannon optimization to find the energy minimum of Eq. 1. We embedded this method in a hierarchical multi-scale structure to be able to capture a wide range of deformations. This structure contains a coarse and fine scale for optimization. At the coarse scale, the optimizer can capture extensive deformations, but the solution may have limited precision. So we use the coarse solution to initialize the optimization at the fine scale and find a more precise deformation field.

2.2.2.2 Symmetric diffeomorphic demons

We utilize a variant of the Demons algorithm that is optimized in the log domain as proposed in Vercauteren et al., 2008. This is a variational method that seeks to minimize the following energy functional:

$$\mathcal{E}(c, \tau) = \frac{1}{\lambda_\zeta^2} \frac{1}{2|\Omega_p|} \sum_{\mathbf{x} \in \Omega_p} |S(\mathbf{x}) - A(\tau(\mathbf{x}))|^2 + \frac{1}{\lambda_h^2} \|\log(\tau^{-1} \circ c)\|^2 + \frac{1}{\lambda_\rho^2} \|\nabla \log(\tau)\|^2 \tag{6}$$

where the variable c was introduced to approximate the error in the correspondence between image pixels, λ_h accounts for spatial uncertainty on the correspondences, and Ω_p is the region of overlap between S and $A \circ \tau$.

In the update step, and under the assumption that the current transformation τ is expressible as an exponential of smooth vector fields \mathbf{v} , i.e., $\tau = \exp(\mathbf{v})$, the Baker-Campbell Hausdorff (BCH) approximations are used to seek a smooth velocity field $Z(\mathbf{v}, \epsilon \mathbf{u})$, such that $\exp(Z(\mathbf{v}, \epsilon \mathbf{u})) \approx \tau \circ \exp(\epsilon \mathbf{u})$, where ϵ is a weight parameter. Then \mathbf{u} is given by

$$\mathbf{u}(\mathbf{x}) = - \left(\frac{(S(\mathbf{x}) - A \circ \tau(\mathbf{x}))}{(\|J^p\|^2 + (\lambda_\zeta^2(\mathbf{x})/\lambda_h^2))} \right) J^{\mathbf{x}\tau}, \tag{7}$$

and J is the Jacobian matrix. In the log-domain, the inverse of a spatial transform τ^{-1} , parametrized by $\tau = \exp(\mathbf{v})$, can be obtained efficiently by backward computation $\tau^{-1} = \exp(-\mathbf{v})$. A symmetric transformation can be obtained from a nonsymmetric one by making the global energy symmetric, i.e.,

$$\tau_{opt} = \arg \min_{\tau} (\mathcal{E}_{sym} := \mathcal{E}(S, A \circ \tau) + \mathcal{E}(A, S \circ \tau^{-1})). \tag{8}$$

The minimization of the energy functional in Eq. 8 can be formulated as a constrained equation using two diffeomorphisms

$$[\tau_{opt}, \tau_{opt}^{-1}] = \arg \min_{[\tau, \tau^{-1}]} \mathcal{E}_{sym}. \tag{9}$$

2.2.2.3 Symmetric image normalization

This method uses diffeomorphisms as the transformation model (Avants et al., 2008). SyN performs the normalization by minimizing the energy functional defined in Eq. 10. SyN searches for a symmetric diffeomorphic spatiotemporal mapping, $\tau \in Diff_0 := \{\text{the space of diffeomorphic mappings with homogeneous boundary conditions}\}$ that minimizes the energy functional in the optimization problem defined in Eq. 10.

$$\mathcal{E}_{sym}(S, A) = \inf_{\tau_1} \inf_{\tau_2} \int_{t=0}^1 \{ \|v_1(\tau_1(\mathbf{x}, t), t)\|_L^2 + \|v_2(\tau_2(\mathbf{x}, t), t)\|_L^2 \} dt + \int_{\Omega} \zeta(S(\tau_2(0.5), A(\tau_1(0.5))) d\Omega$$

subject to each $\tau \in Diff_0$ the solution of: $d\tau(\mathbf{x}, t)/dt = v_i(\tau(\mathbf{x}, t), t)$
with $\tau_i(\mathbf{x}, 0) = \mathbf{I}$, and $\tau_i^{-1}(\tau_i) = \mathbf{I}$, $\tau_i(\tau_i^{-1}) = \mathbf{I}$ (10)

The first integral in Eq. 10 corresponds to the regularization term that is induced by a functional norm $\|\cdot\|_L$ through a linear differential operator $L = a\nabla^2 + b\mathbf{I}$, with constants a and b , and \mathbf{I} is the identity mapping. The second integral corresponds to the similarity term between the reference and input image, where ζ is Matte’s mutual information defined in Eq. 3, Ω is the common spatial domain of the images, $v(\mathbf{x}, t)$ is the velocity field, and t is the time. The optimization process performs gradient descent to update the deformation field and a fixed point method calculates the inverse transformation. The velocity fields v_i are computed iteratively, and they update the deformation τ_i , $i = 1, 2$. The deformable registration stage is preceded by rigid and affine transformation steps to address global misalignments as described in.

2.3 Shape modeling - statistical atlas generation

We generate the statistical atlases by iteratively averaging subject scans that are mapped onto a common reference space. We first choose one subject to serve as the reference scan. Then, we linearly register all subjects to the selected reference and compute the group average. In the second iteration, we use as reference the average image produced by linear group registration. Next, we map all subjects to the new computed template using nonlinear registration, and compute the average. In the remaining iterations we only apply nonlinear mappings to update the average image. We repeat the above steps until the final iteration $i = K$. This process converged to an atlas template within $K = 5$ iterations on our data.

The above steps generate sequences of transformations $\tau_i^{(n)}$, where $n = 0, 1, 2, \dots, N$ is the n th image and $i = 0, 1, \dots, K$ is the i th iteration. Rohlfsing et al., 2001 showed that in each iteration i , the transformation $\tau_i^{(n+1)}$ and the preceding transformation $\tau_i^{(n)}$ differ only by a small amount of deformation. Finally, a human operator labeled the trabecular bone, cortical bone, muscle, and SAT using the MIPAV software suite (McAuliffe et al., 2001) by manual selection of control points and spline interpolation. In the first row of Figure 2, we display our computed atlas image and the atlas label map for 4%, 38% and 66% of the tibia length, respectively. In the first row, third, sixth and ninth columns of Figure 2, we display the color-coded atlas labels, for each tibial site under consideration.

2.4 Multi-atlas based tissue segmentation

2.4.1 Label propagation

Label propagation is the process of assigning labels from the warped atlas labels to the reference space. We use the linear (μ) and nonlinear (τ) transformations between S and A that we found in the registration stage, to map labels from the atlas to the subject space. Label propagation is achieved by nearest neighbor interpolation after warping the atlas label to the subject domain via the estimated deformation $\tau := \tau \circ \mu$. The segmentation map is produced by $\tau(SA)$.

2.4.2 Label fusion

This is a key stage of MAIS. Here, we combine all the propagated atlases to obtain a final segmentation (Iglesias and Sabuncu 2015). Various methods have been proposed for this stage including best atlas selection, a selective and iterative method for performance level estimation (SIMPLE), joint label fusion, majority voting, weighted majority voting, and simultaneous truth and performance level estimation (STAPLE) algorithm (Warfield et al., 2004;

Langerak et al., 2010; Wang et al., 2013; Iglesias and Sabuncu 2015). In this work, we utilized STAPLE for fusing segmentation results by individual atlases. We utilize STAPLE for label fusion because it has performed very well over a range of applications Cardoso et al., 2013; Weston et al., 2019.

STAPLE can be formulated using probabilistic classification terms. Given K segmentations (classifications) of a subject S having N pixels, let $e_k(\mathbf{x})$ be the decision of classifier k at voxel \mathbf{x} . If the (unknown) ground truth label for voxel \mathbf{x} is l , we say that $\mathbf{x} \in C_l$. The performance of classifier k is determined by two parameters p (sensitivity) and q (specificity), referring to the fractions of true positives and true negatives among the classified voxels, that maximizes the complete log likelihood function, $(p, q) = \arg \max_{p,q} \ln f(D, T|p, q)$, where $T = \cup_{l=1}^n C_l$ is the true segmentation, also called missing or hidden data, and $D = [e_k(\mathbf{x})]$ is an $N \times K$ decision matrix.

For each classifier k , and each class C_b , the parameters p and q are modeled independently as the following conditional probabilities: $p_k = \Pr(e_k(\mathbf{x}) = l | \mathbf{x} \in C_l)$ and $q_k = \Pr(e_k(\mathbf{x}) \neq l | \mathbf{x} \notin C_l)$. The process of estimation of p and q is achieved by the Expectation-Maximization (EM) algorithm.

The final segmentation \hat{S} at voxel \mathbf{x} is computed by $E(\mathbf{x}) = \arg \max_x P(x \in C_i | e_1(\mathbf{x}), \dots, e_K(\mathbf{x}))$. The probability $p(x \in C_i | e)$ follows from the classifier's decisions and their performance parameters using Bayes' rule (Rohlfsing et al., 2005).

3 Data description and performance evaluation measures

3.1 Overview of dataset

We used pQCT data obtained from the InCHIANTI clinical study to evaluate the performance of the methods. InCHIANTI is a longitudinal study of risk factors for mobility disability performed in a representative sample of the middle aged and older populations living in Tuscany, Italy (Ferrucci et al., 2000; Lauretani et al., 2008; Makrogiannis et al., 2018). The validation dataset is randomly sampled from the original InCHIANTI baseline database as in Makrogiannis et al., 2018. It consists of pQCT scans of the lower leg acquired at the 4%, 38% and 66% tibial length of the lower leg. Our randomly sampled dataset contains a total of 77 samples, that is 30 samples at 4%, 27 samples at 38% and 20 samples at 66% tibial length. The pQCT scans of the right lower leg were acquired using a XCT 2000 scanner (Stratec Medizintechnik GmbH, Pforzheim, Germany). The slice thickness of each scan is 2.1 mm and the in-plane slice is 0.5 mm. At each tibial location, we used three templates for MAIS. One of these three samples was used to initialize statistical atlas generation. We cross-validated the performance of each method on the testing samples that

remained after removing the three templates. We focused on specific tissue(s) of interest at each tibial location: trabecular bone at 4%; cortical and trabecular bone at 38%; subcutaneous adipose tissue (SAT), muscle, cortical bone and trabecular bone at 66% as shown in [Table 1](#).

3.2 Motion artifacts

Unique characteristics of the test data that made it suitable for our analysis and clinically relevant are: 1) the strong representation of older persons who are likely to experience walking difficulties, 2) the different locations of the tibial length that we considered for identification of six different tissue types ([Table 1](#)), 3) the incidence of motion artifacts in the dataset making it complicated for segmentation. Motion artifacts occur as a result of the voluntary and involuntary movement of the subject during image acquisition.

Frequent criteria for motion assessment are discontinuities and streaks in the cortical bone, and blurring and shifting of trabecular bone. A clinical specialist used the above visual inspection criteria and a grading system from 1 to 5 (1: no artifacts) to classify the subject motion artifacts as described in [Wong 2016](#). This effect was more evident at the 66% tibial length with more than 50% of the samples having significant motion artifacts corresponding to grades 4 and 5.

3.3 Validation

We performed quantitative analysis of the performance of the atlas-based tissue identification schemes by calculating the Dice Similarity Coefficient (DSC), sensitivity or True Positive Rate (TPR), and Precision (PR) between segmentation results and their ground truths T . We cross-validated the performance metrics by excluding the two subjects we used as templates in MAIS, and the subject we used as template for statistical atlas generation. We made comparisons across single- and multi-atlas based image segmentation methods with respect to registration algorithms. We compared the results obtained by multi-atlas based techniques with those obtained by the automated tissue identification and quantification (TIDAQ) method ([Makrogiannis et al., 2018](#)), to emphasize the performance of MAIS techniques. We chose TIDAQ for our comparisons because it is a model-based tissue segmentation method that has produced good results for images of good quality, but may be challenged by images that have moderate to significant motion artifacts. We performed nonparametric Wilcoxon rank sum tests to examine the effects of motion artifacts on tissue identification using pQCT scans of 66% tibia. The sample size is $n = 20$.

TABLE 2 Summary of the unlabeled pQCT scans of the lower leg from the InCHIANTI Study.

Site (%)	Gender	N	Age (yrs)				
			Min	Max	Med	Ave	Std
4	Male	272	28	104	74	87.3	14.2
	Female	313	26	104	74	87.4	15.0
	Total	585	26	104	73	87.4	14.6
38	Male	272	28	104	73	69.2	14.5
	Female	311	26	104	74	69.4	14.7
	Total	583	26	104	74	69.3	14.6
66	Male	270	28	104	73	69.2	14.3
	Female	304	26	104	73	69.1	14.8
	Total	580	26	104	73	69.1	14.5

3.4 Tissue quantification

We provide an extension of our analysis to non-labeled pQCT scans of the lower leg from the InCHIANTI dataset to quantify body composition changes caused by aging. Our aim is to show the reliability of the MAIS-technique using SDD-STPL for tissue quantification on an extended clinical dataset and evaluate the agreement of our results with clinical observations. We decided to quantify the baseline InCHIANTI dataset that includes a total of 2,425 pQCT scans. We applied the following procedures to prepare the data for analysis. With the help of TIDAQ software, we sorted the scans according to anatomical sites, selected scans at 4%, 38%, and 66% of tibial length and removed scans with different orientations. We then separated the remaining scans according to their gender (male and females). After data preparation, we had a total of 1748 scans for quantification that may be grouped as follows: 585 scans (males: 272, and females: 313) at 4%; 583 scans (males: 272, females: 311) at 38%; and 580 scans (Males: 270, females: 304) at 66%. A summary of our quantification dataset, including gender and age distribution, is provided in [Table 2](#). To characterize the effect of aging on body composition, we calculated the cross-sectional area (CSA) and density of the trabecular bones at 4% tibia, cortical bones at 38% tibia, and muscle and subcutaneous adipose tissue (SAT) at 66% tibia. To obtain these measures, we ran our SDD-STPL technique on all datasets to automatically identify these tissues and computed the total CSA and the average density for each tissue type. We decided to perform analyses of both genders jointly, as well as separate gender-conditional analyses.

We used scatter plots and regression analyses of quantification results to study the changes in body composition (response variable) with respect to age (predictor variable). We utilized the following statistical measures to analyze the relationship between the two variables: the square of the correlation coefficient (R^2), the coefficient of variation of

TABLE 3 4% Trabecular Bone Segmentation Performance (mean \pm standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Method	DSC	TPR	PR
STAT-FFD	0.947 \pm 0.045	0.96 \pm 0.021	0.939 \pm 0.084
STPL-FFD	0.936 \pm 0.056	0.971 \pm 0.018	0.911 \pm 0.103
STAT-SDD	0.959 \pm 0.019	0.939 \pm 0.037	0.981 \pm 0.015
STPL-SDD	0.972 \pm 0.009	0.965 \pm 0.021	0.98 \pm 0.014
STAT-SyNO	0.906 \pm 0.072	0.922 \pm 0.081	0.9 \pm 0.103
STPL-SyNO	0.914 \pm 0.068	0.936 \pm 0.07	0.902 \pm 0.104
STAT-SyN	0.947 \pm 0.05	0.972 \pm 0.017	0.93 \pm 0.122
STPL-SyN	0.947 \pm 0.054	0.975 \pm 0.02	0.928 \pm 0.128
TIDAQ	0.941 \pm 0.022	0.913 \pm 0.038	0.972 \pm 0.024

the root mean squared difference (*CV-RMSD*) (Eq. 11) between the reference (y) and the predicted (\hat{y}) measurements, the slope and intercept of the regression line, and the p - values.

$$CV - RMSD = \frac{1}{\mu} \sqrt{\frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{N}}; \quad \mu = \frac{\sum_{k=1}^N y_k}{N} \quad (11)$$

4 Experiments and results

Here we evaluate the performances of the four MAIS techniques and an automated model-based tissue quantification method (TIDAQ) (Makrogiannis et al., 2018). We then explore the effect of subject movement artifacts on tissue segmentation performance. Finally, we expand our analysis to quantify the complete InCHIANTI baseline dataset.

Our aim is to support or reject the hypotheses that 1) MAIS techniques, in general, improve the segmentation performance of SAIS, 2) STAPLE on SDD mappings produces better segmentation quality than the other methods, 3) MAIS is more resilient to subject movement than model-based segmentation.

We analyzed the performance of the deformable methods by validating the SAIS results of statistical atlases against reference standards, T . Reference standard is a tissue label map that was generated manually by a clinical specialist. We evaluated single atlas segmentation performances of STAT-FFD, STAT-SDD, STAT-SyN, and STAT-SyNO, where ‘STAT’ represents the statistical atlas. The MAIS techniques we developed and evaluated in this framework are STPL-FFD, STPL-SDD, STPL-SyN, and STPL-SyNO. ‘SyNO’ denotes an ‘Symmetric Normalization with our own linear registration’ and ‘STPL’ denotes ‘STAPLE.’

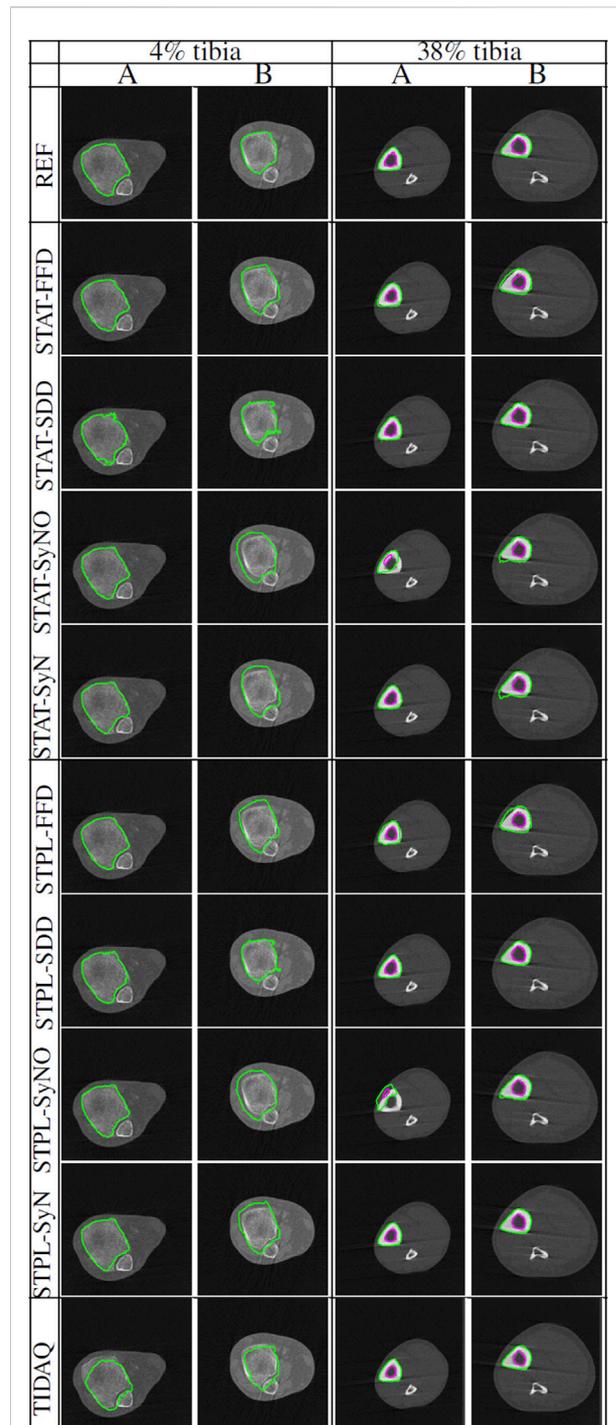


FIGURE 4

Comparisons of tissue segmentation by MAIS and model-based methods at 4% tibia and 38% tibia. Two subjects (subject (A) and subject (B)) are selected at each tibial site to demonstrate the performance of the compared methods. At 4% the trabecular bone is delineated by the green contour, and at 38% the cortical bone and the trabecular bone are delineated by the green and magenta contours, respectively.

TABLE 4 38% Cortical Bone Segmentation Performance (mean \pm standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Method	DSC	TPR	PR
STAT-FFD	0.809 \pm 0.126	0.933 \pm 0.064	0.721 \pm 0.125
STPL-FFD	0.781 \pm 0.136	0.889 \pm 0.135	0.706 \pm 0.153
STAT-SDD	0.932 \pm 0.022	0.965 \pm 0.025	0.902 \pm 0.032
STPL-SDD	0.947 \pm 0.021	0.949 \pm 0.036	0.947 \pm 0.023
STAT-SyNO	0.78 \pm 0.174	0.891 \pm 0.173	0.704 \pm 0.184
STPL-SyNO	0.76 \pm 0.215	0.841 \pm 0.233	0.704 \pm 0.216
STAT-SyN	0.871 \pm 0.07	0.983 \pm 0.026	0.787 \pm 0.104
STPL-SyN	0.873 \pm 0.112	0.96 \pm 0.079	0.807 \pm 0.141
TIDAQ	0.897 \pm 0.15	0.954 \pm 0.156	0.847 \pm 0.147

TABLE 5 38% Trabecular Bone Segmentation Performance (mean \pm standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Method	DSC	TPR	PR
STAT-FFD	0.823 \pm 0.084	0.774 \pm 0.145	0.917 \pm 0.121
STPL-FFD	0.797 \pm 0.131	0.759 \pm 0.177	0.891 \pm 0.155
STAT-SDD	0.92 \pm 0.027	0.868 \pm 0.052	0.981 \pm 0.019
STPL-SDD	0.939 \pm 0.019	0.91 \pm 0.041	0.972 \pm 0.026
STAT-SyNO	0.703 \pm 0.294	0.627 \pm 0.282	0.831 \pm 0.309
STPL-SyNO	0.687 \pm 0.343	0.624 \pm 0.338	0.807 \pm 0.363
STAT-SyN	0.849 \pm 0.079	0.764 \pm 0.081	0.97 \pm 0.109
STPL-SyN	0.877 \pm 0.08	0.811 \pm 0.11	0.972 \pm 0.078
TIDAQ	0.825 \pm 0.167	0.726 \pm 0.156	0.96 \pm 0.187

4.1 4% tibia segmentation

Table 3 displays the summarized performance measures obtained by the single-atlas image segmentation using the statistical atlas and multi-atlas image segmentation methods, and TIDAQ for identification of the trabecular bone (TB) over the test-set. In Figure 4, column 1 we show examples of multi-atlas segmentations of the trabecular bone at 4% tibia by the compared methods, delineated by the green contours.

A comparison of single-atlas and multi-atlas segmentation results in Table 3 shows that the use of multiple atlases improves segmentation quality. With respect to multi-atlas image segmentation techniques, we examined the performance quality of the methods and the effect of label fusion on the improvement of the results. The range and mean \pm standard deviation of DSC values produced by STPL-SDD are [0.944, 0.992] and 0.972 ± 0.009 , by STPL-FFD are [0.739, 0.991] and 0.936 ± 0.056 , by STPL-SyN are [0.743, 0.990] and 0.947 ± 0.054 , and by STPL-SyNO are [0.731, 0.991] are 0.914 ± 0.068 , respectively, as can be seen in Table 3. In summary, we observe that STPL-SDD outperformed the other multi-atlas techniques.

The range of DSC values produced by TIDAQ for trabecular bone identification is [0.897, 0.967] with mean \pm standard deviation of 0.941 ± 0.022 . The results in Table 3 indicate that TIDAQ outperformed STPL-FFD, and STPL-SyNO registration techniques. On the other hand, TIDAQ was less accurate than STPL-SDD, and STPL-SyN.

4.2 38% tibia segmentation

Tables 4 and 5 contain the performance measures produced by these experiments. Figure 4 displays examples of tissue delineations, on two of our test subjects.

Overall, quantitative results of the methods reported in Tables 4 and 5, show that SDD produced better segmentation

quality than the other deformable models, followed by SyN. Low DSC values are produced by SyN, FFD, and SyNO models (in decreasing order) for both cortical and trabecular bone in some subjects. SyNO missed the trabecular bone in few subjects producing zero DSC. We also noticed that all deformable methods produced DSC means that are greater than 75% for cortical and trabecular bone except SyNO in the trabecular bone. We infer that the range of DSC values for SDD is more compact than the other deformable models. The minimum value of DSC produced by SDD is greater than 85% for cortical bone, and about 82% for trabecular bone. The DSC values produced by SyN are fairly compact in the identification of bone compartments, and their mean values are greater than 85% with standard deviations about 10%.

In Tables 4 and 5, we observe that FFD results for both single- and multi-atlas image segmentation techniques show a wider DSC dispersion than SDD, the former producing values lower than 60% for cortical bone and about 45% for trabecular bone.

SyNO results show wider DSC spread than all of the other deformable models. SyNO produced the least mean DSC of about 78% for identification of cortical bone and about 68% for trabecular bone. This error is usually caused by linear registration failures propagated to the symmetric normalization stage.

TIDAQ performed very well in the identification of trabecular bone, and much better in the identification of cortical bone. Overall, the identification accuracy is promising with DSC mean \pm standard deviation of $89.7 \pm 2.2\%$ for cortical bone and $82.5 \pm 16.7\%$ for trabecular bone. Despite the good performance of TIDAQ, STPL-SDD outperformed it in the identification of cortical bone, while STPL-SDD and STPL-SyN outperformed it in the identification of trabecular bone. This shows that multi-atlas image segmentation techniques have the potential to produce higher tissue identification accuracy than TIDAQ.

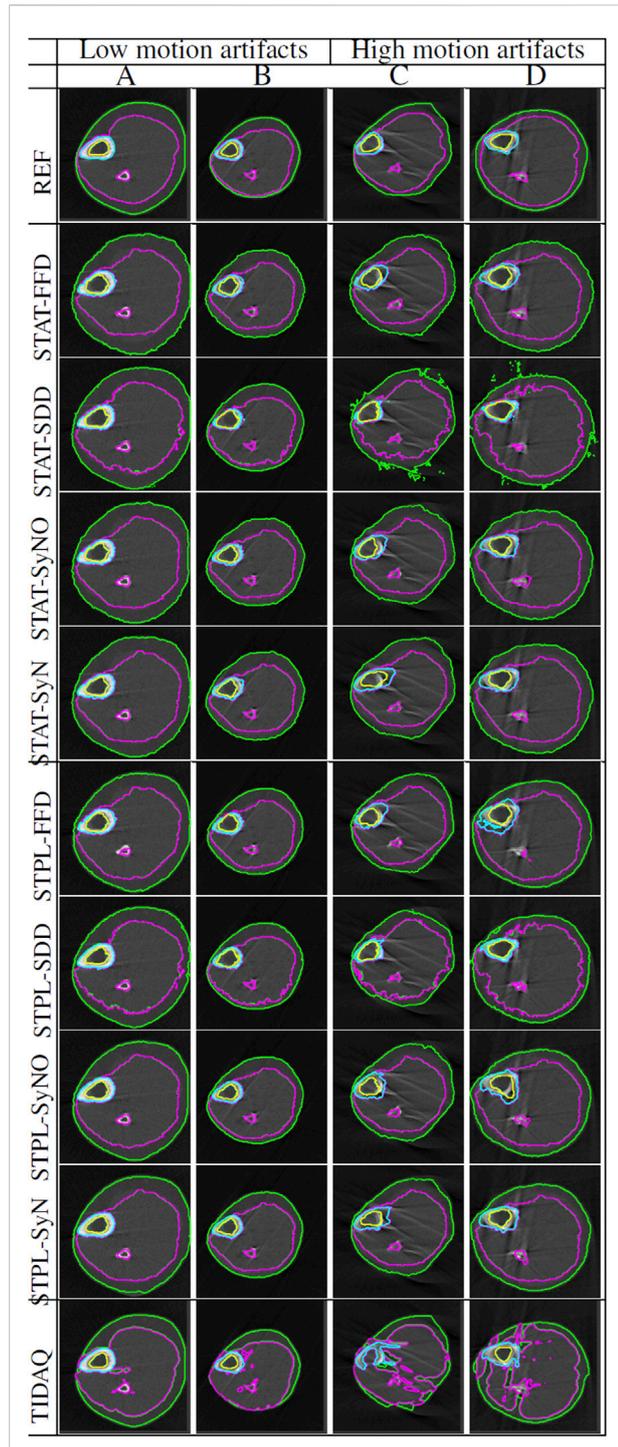


FIGURE 5
 Segmentation comparisons of scans at 66% tibial length with low and high levels of artifacts caused by subject motion. Scans (A,B) have low motion artifacts, while scans (C,D) have the high motion artifacts. The delineation of the subcutaneous adipose tissue (SAT) is represented by the green contour, muscle by magenta color, cortical bone by cyan color, and trabecular bone by yellow color.

TABLE 6 66% SAT Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.682 ± 0.158	0.953 ± 0.051	0.562 ± 0.2
STPL-FFD	0.73 ± 0.16	0.912 ± 0.085	0.649 ± 0.227
STAT-SDD	0.708 ± 0.151	0.939 ± 0.042	0.594 ± 0.189
STPL-SDD	0.771 ± 0.144	0.917 ± 0.057	0.692 ± 0.202
STAT-SyNO	0.684 ± 0.152	0.938 ± 0.065	0.572 ± 0.2
STPL-SyNO	0.729 ± 0.161	0.892 ± 0.093	0.659 ± 0.229
STAT-SyN	0.677 ± 0.147	0.935 ± 0.079	0.565 ± 0.197
STPL-SyN	0.734 ± 0.157	0.889 ± 0.078	0.664 ± 0.225
TIDAQ	0.746 ± 0.228	0.845 ± 0.14	0.702 ± 0.267

TABLE 7 66% Muscle Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.893 ± 0.032	0.837 ± 0.067	0.962 ± 0.037
STPL-FFD	0.902 ± 0.034	0.857 ± 0.074	0.959 ± 0.037
STAT-SDD	0.922 ± 0.029	0.891 ± 0.057	0.958 ± 0.032
STPL-SDD	0.938 ± 0.028	0.914 ± 0.057	0.966 ± 0.029
STAT-SyNO	0.894 ± 0.038	0.847 ± 0.073	0.953 ± 0.054
STPL-SyNO	0.911 ± 0.038	0.874 ± 0.078	0.959 ± 0.048
STAT-SyN	0.893 ± 0.035	0.844 ± 0.077	0.957 ± 0.051
STPL-SyN	0.913 ± 0.04	0.875 ± 0.084	0.962 ± 0.045
TIDAQ	0.904 ± 0.085	0.861 ± 0.131	0.963 ± 0.025

4.3 66% tibia segmentation

Figure 5 displays segmentation results produced by the compared approaches on two scans with low motion degradation and two scans with high motion degradation. The mean ± standard deviation values of DSC, TPR and PR of each tissue over all testing samples are given in Tables 6–9. Overall, SDD exhibited better performance across the three metrics than the other deformable models for the identification of all tissues, followed by SyN and TIDAQ. In addition, SDD produced DSC, true positive rate, and precision values of lower dispersion (expressed by smaller standard deviation) than the other methods. STPL-SDD yielded the top DSC performance for SAT, muscle, and trabecular bone segmentation. The DSC minimum values for this method were about 62.7% for SAT, 89% for muscle, 75% for cortical bone, and 86.2% for trabecular bone. All MAIS techniques produced mean DSC greater than 89%, mean TPR greater than 84%, and mean PR greater than 95% in muscle segmentation. Conversely, all tested methods yielded

TABLE 8 66% Cortical Bone Segmentation Performance (mean \pm standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.665 \pm 0.152	0.668 \pm 0.121	0.677 \pm 0.196
STPL-FFD	0.668 \pm 0.132	0.696 \pm 0.11	0.666 \pm 0.19
STAT-SDD	0.681 \pm 0.094	0.55 \pm 0.112	0.911 \pm 0.059
STPL-SDD	0.829 \pm 0.083	0.776 \pm 0.118	0.898 \pm 0.064
STAT-SyNO	0.721 \pm 0.156	0.708 \pm 0.145	0.754 \pm 0.201
STPL-SyNO	0.765 \pm 0.194	0.792 \pm 0.181	0.761 \pm 0.225
STAT-SyN	0.718 \pm 0.147	0.719 \pm 0.084	0.737 \pm 0.21
STPL-SyN	0.788 \pm 0.152	0.827 \pm 0.112	0.769 \pm 0.194
TIDAQ	0.851 \pm 0.115	0.987 \pm 0.032	0.76 \pm 0.157

TABLE 9 66% Trabecular Bone Segmentation Performance (mean \pm standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.825 \pm 0.087	0.902 \pm 0.124	0.783 \pm 0.124
STPL-FFD	0.827 \pm 0.085	0.911 \pm 0.123	0.783 \pm 0.138
STAT-SDD	0.86 \pm 0.049	0.983 \pm 0.023	0.769 \pm 0.08
STPL-SDD	0.913 \pm 0.045	0.968 \pm 0.037	0.869 \pm 0.088
STAT-SyNO	0.843 \pm 0.179	0.91 \pm 0.203	0.803 \pm 0.142
STPL-SyNO	0.868 \pm 0.167	0.913 \pm 0.191	0.838 \pm 0.14
STAT-SyN	0.856 \pm 0.116	0.946 \pm 0.073	0.791 \pm 0.149
STPL-SyN	0.895 \pm 0.105	0.939 \pm 0.077	0.862 \pm 0.133
TIDAQ	0.871 \pm 0.221	0.816 \pm 0.216	0.994 \pm 0.026

mean DSC less than 80%, and a top PR of 70.2% in SAT segmentation.

In SAT delineation, FFD and SyNO models compete with each other in the identification of SAT, but SyNO outperforms free-form deformation in the identification of other tissues by at least 10% accuracy. Except for cortical bone, we observe that MAIS outperformed TIDAQ.

4.4 Effects of subject motion on tissue identification at 66% tibia

In this experiment, we studied the effect of artifacts on segmentation performance. The motion was assessed using 5-level visual grading as described in Wong (2016). We separated the samples into low level of motion determined by grades 1–3, and high level of motion with grades 4 and 5, and compared the performance of all methods. As stated above, 12 out of 20 pQCT scans at 66% tibia contain high to severe motion artifacts (grades 4 and 5).

The next step is to explore the differences in segmentation performance between the two groups. Table 10 summarizes the segmentation performance measured by DSC, true positive rate (TPR) and precision (PR) for each tissue type and each method. This table also contains the relative differences of the performance measures. To estimate the statistical significance of the differences in segmentation performance, we applied nonparametric Wilcoxon rank sum tests between the two groups and we report the p -values.

Considering SAT, we observe consistent decrease of average DSC and PR with increasing motion artifacts for all methods. We observe that TIDAQ shows the highest decrease in all performance measures. The Wilcoxon tests indicate statistically significant performance differences for TIDAQ in DSC and precision, and for STPL-SDD in DSC and true positive rate.

4.5 Tissue composition assessment

We applied our STPL-SDD method to the baseline InCHIANTI dataset that we described in Section 3 and summarized in Table 2. We then analyzed the quantification results of cross-sectional areas and average densities for each tissue to explore changes in its composition as a function of age. The scatter plots and regression results in Figures 6, 7 lead to the following observations. Trabecular bone density decreases with age at similar rates for males and females. The CSA of cortical bone decreases with age at similar rates for males and females. Cortical bone density decreases with age for males and females, and the rate of decrease is higher for females. Muscle CSA decreases with age for males and females, and the rate of decrease is higher for males. Muscle density decreases with age at similar rates for males and females. Our analysis does not reveal significant changes with age for SAT CSA, SAT density, and trabecular bone CSA.

In addition, the statistical results reported in Tables 11–13 show that there is increased correlation of cortical bone density with age, and trabecular bone density with age, especially for females, relative to the other tissues. In addition, there is noticeable correlation between muscle CSA and age for males. The CV-RMSD values show decreased variation mostly for cortical density and muscle density for each gender. We also observe that the gender-conditional analyses produce lower variability than joint analyses of males and females as we expected.

4.6 Method implementation and execution time

We developed the programs of the proposed methodologies in C++, Python 3.7, and used the ITK library. We implemented

TABLE 10 Effect of motion artifacts on segmentation performance.

Dice similarity coefficient (DSC)

TISSUE	66%-SAT				66%-MUSCLE				66%-CB				66%-TB			
	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>
STPL-FFD	0.799	0.703	-11.9%	0.208	0.919	0.896	-2.5%	0.143	0.691	0.660	-4.5%	0.849	0.785	0.843	7.4%	0.173
STPL-SDD	0.866	0.734	-15.2%	0.046	0.964	0.928	-3.8%	0.004	0.876	0.810	-7.5%	0.059	0.908	0.914	0.7%	0.849
STPL-SyN	0.794	0.711	-10.5%	0.246	0.926	0.908	-2.0%	0.503	0.792	0.786	-0.8%	0.173	0.845	0.914	8.2%	0.453
STPL-SyNO	0.814	0.697	-14.4%	0.173	0.936	0.901	-3.7%	0.095	0.813	0.747	-8.1%	0.143	0.900	0.856	-4.9%	0.775
TIDAQ	0.922	0.678	-26.4%	0.014	0.978	0.876	-10.5%	4.7·10 ⁻⁴	0.956	0.811	-15.2%	2.3·10 ⁻⁴	0.960	0.837	-12.8%	0.035

TRUE POSITIVE RATE (TPR)

TISSUE	66%-SAT				66%-MUSCLE				66%-CB				66%-TB			
	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>
STPL-FFD	0.914	0.912	-0.3%	0.703	0.882	0.847	-4.0%	0.443	0.695	0.697	0.2%	0.633	0.952	0.895	-6.0%	0.633
STPL-SDD	0.958	0.901	-5.9%	0.007	0.950	0.901	-5.1%	0.117	0.819	0.760	-7.2%	0.336	0.966	0.969	0.3%	0.387
STPL-SyN	0.886	0.890	0.5%	0.775	0.894	0.868	-2.9%	0.775	0.822	0.829	0.8%	0.849	0.925	0.944	2.1%	1.000
STPL-SyNO	0.903	0.888	-1.6%	0.703	0.911	0.860	-5.7%	0.246	0.843	0.772	-8.4%	0.387	0.980	0.887	-9.4%	0.143
TIDAQ	0.911	0.820	-10.0%	0.336	0.978	0.817	-16.4%	2.3·10 ⁻⁴	1.000	0.982	-1.8%	0.185	0.925	0.774	-16.3%	0.035

PRECISION (PR)

TISSUE	66%-SAT				66%-MUSCLE				66%-CB				66%-TB			
	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>	Low	High	Diff%	<i>p</i>
STPL-FFD	0.732	0.617	-15.7%	0.503	0.964	0.957	-0.6%	0.924	0.719	0.645	-10.3%	0.387	0.672	0.826	23.0%	0.026
STPL-SDD	0.797	0.651	-18.3%	0.289	0.980	0.960	-2.0%	0.173	0.958	0.875	-8.7%	0.002	0.868	0.870	0.3%	0.775
STPL-SyN	0.747	0.631	-15.5%	0.443	0.966	0.961	-0.6%	0.775	0.794	0.759	-4.4%	0.117	0.791	0.890	12.6%	0.443
STPL-SyNO	0.766	0.617	-19.4%	0.246	0.967	0.956	-1.1%	1.000	0.814	0.740	-9.1%	0.117	0.838	0.839	0.2%	0.566
TIDAQ	0.933	0.613	-34.3%	0.010	0.979	0.957	-2.2%	0.117	0.915	0.700	-23.5%	2.3·10 ⁻⁴	1.000	0.992	-0.8%	1.000

the symmetric normalizations, -SyN and -SyNO (originally -SyNOnly) of ANTs on the ANTsPY python library (antspyx version 0.2.5) with default parameters to generate the segmentation results corresponding to these nonlinear registration algorithms. The TIDAQ backend is implemented in C++ and uses ITK, while the user interface is a Java plugin. We executed our experiments on a system with Linux CentOS 7, 2 x Intel(R) Xeon(R) CPU E5-2,690 v4 2.60 GHz, and 128 GB RAM. We computed the execution time of SAIS and MAIS, with respect to the deformable registration algorithms, for all subjects at the different tibia location. We calculated the mean \pm standard deviation of the execution time over all subjects in the segmentation set and report the values in Table 14. We

observe that the execution time of STAT-SDD is about 27 s for 4% tibia, and about 32 s for 38% and 66% tibia, while the execution time of STPL-SDD is about 102 s for 4%, 118 s for 38%, and 148 s for 66%.

5 Discussion

5.1 Method comparisons, tissue separability and technical characteristics

Tissue delineation is an important, but challenging task in medical image analysis. The accuracy of tissue delineation

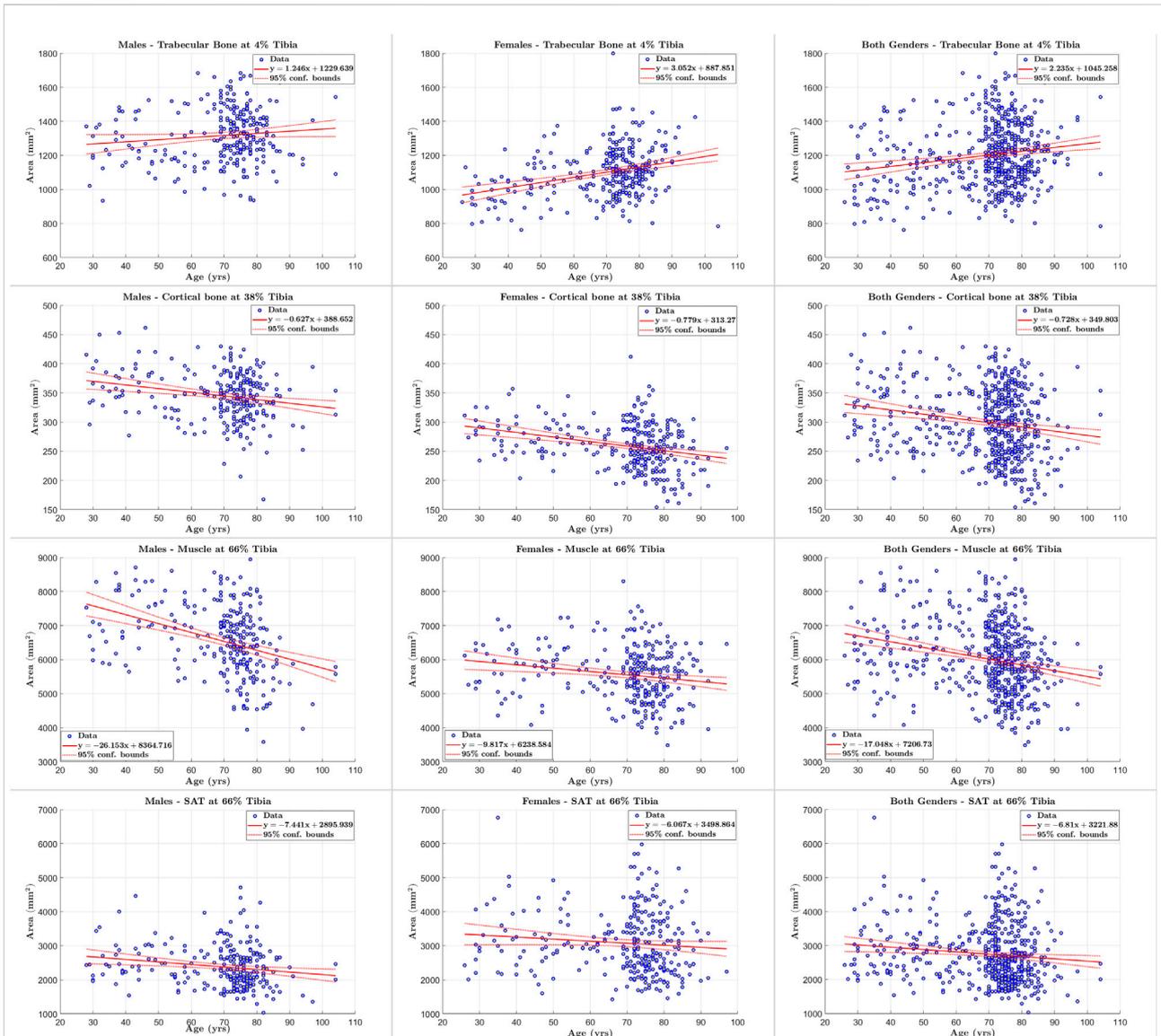


FIGURE 6 Scatter plots of tissue of interest’s cross sectional area versus age for males, females, and both genders at 4%, 38% and 66% tibia.

depends on many factors, including the intrinsic characteristics of the technique, image modality, artifacts or noise, and the number of tissues to be identified in each scan. Peripheral quantitative computed tomography imaging quality is significantly affected by artifacts caused by subject movement. Overall, the proposed multi-atlas image segmentation techniques address the aforementioned factors. Furthermore, the multi-atlas symmetric diffeomorphic demons technique proved to be more robust to reduced image quality than the other methods, followed by symmetric image normalization.

Visual inspection of tissue densities in the image at multiple tibial sites that are displayed in Figure 8, second row, shows that

the distributions of different tissues significantly overlap with one another. At 4% tibia, we observe a clear overlap between the distribution of all leg tissues and the trabecular bone. An optimized thresholding technique based on tissue densities, for example, may not separate the distributions accurately, because of high false positives and false negatives. At 38% tibia, although it appears that there is a valley between the cortical and trabecular bones, yet there is still significant overlap between the distributions of the trabecular bone tissues and all leg tissues. At 66% tibia, we note the extensive overlap among the distributions of SAT, muscle, and trabecular bone. These distributions illustrate the difficulties that would be

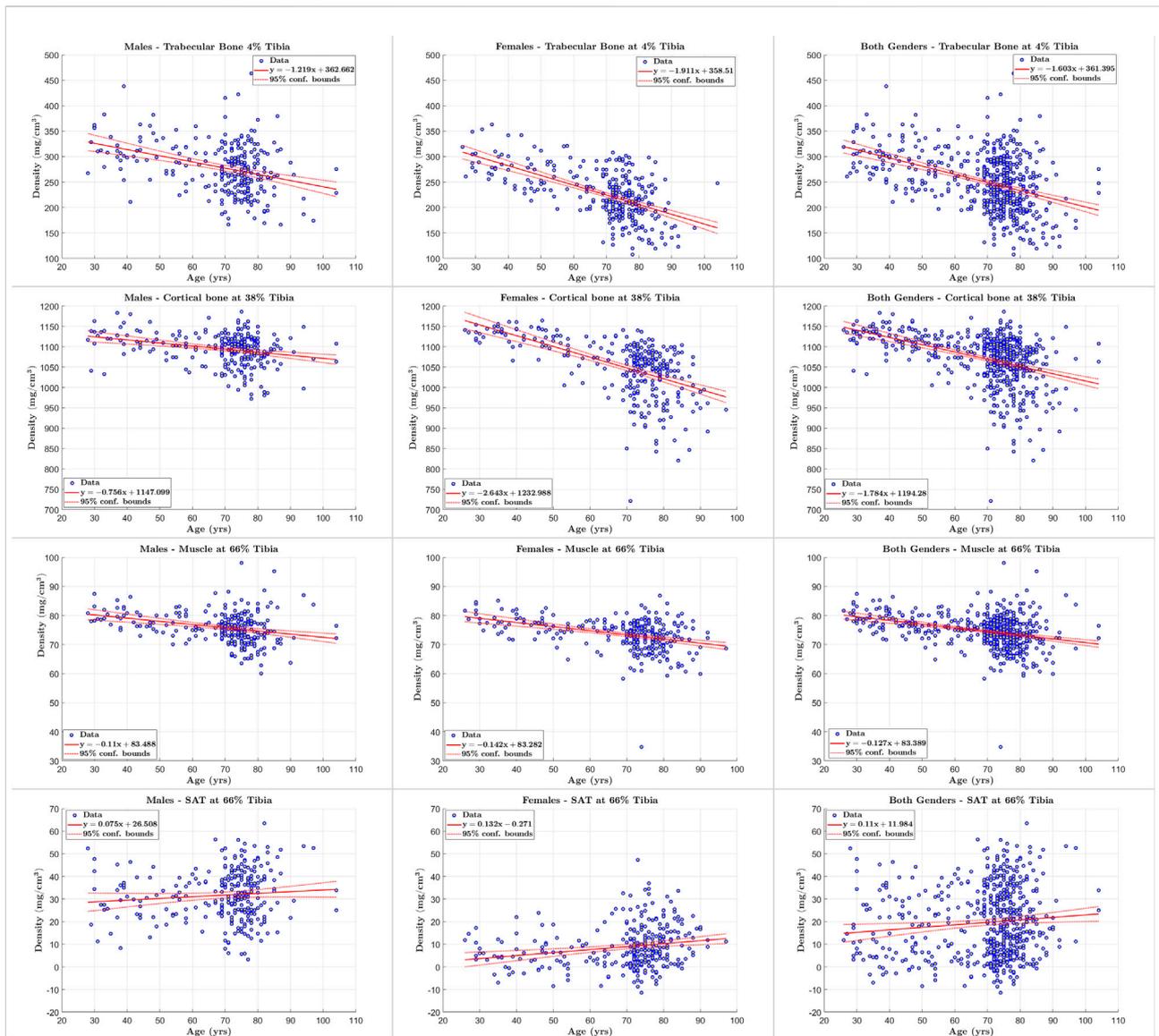


FIGURE 7
Scatter plots of tissue of interest’s density versus age for males, females, and both genders at 4%, 38% and 66% tibia.

encountered by segmentation techniques that rely on density alone. Yet, the proposed MAIS technique STPL-SDD, yields high segmentation accuracy (DSC >90%), in almost all tissue delineations at different sites, except for 66%-SAT and 66%-cortical bone.

Specific properties of these methods that improved the tissue delineation accuracy are their 1) symmetric nature, and 2) diffeomorphism. Concerning symmetry, the method in question provides equal treatment to both fixed and moving images. In addition, the interactive force between the two images can produce accurate registration of one part of the image to the other, and *vice versa* (Rogelj and Kovačić, 2006). On the other

hand, diffeomorphism affords the algorithms the ability to handle both large and small deformations (Sotiras et al., 2013). It is important to note that large deformations are a result of large strains or rotations, which are caused by subject movement. Thus, symmetric diffeomorphic demons and symmetric normalization are robust to local image artifacts or large image deformations that are difficult to register.

The segmentation results shown in Figure 4, at 4% and 38% tibial length, correspond to subjects corresponding to variable levels of delineation challenges, caused by either the condition of the subject (in the 4% examples) or subject motion (in the 38% examples). At both anatomical sites, subject ‘B’ presents

TABLE 11 Statistical measures of the relationships between tissue properties and age for males and females.

Measurement	Site	R ²	CV-RMSD	Slope	Intercept	p-value
Trabecular CSA	4	0.031	3.646	2.235	1,045.3	< 10 ⁻⁴
Trabecular density	4	0.181	4.799	-1.603	361.4	< 10 ⁻⁴
Cortical CSA	38	0.033	4.661	-0.728	349.8	< 10 ⁻⁴
Cortical density	38	0.178	1.260	-1.784	1,194.3	< 10 ⁻⁴
SAT CSA	66	0.013	7.361	-6.810	3,221.9	0.0054
SAT density	66	0.011	18.583	0.110	12	0.0123
Muscle CSA	66	0.058	3.967	-17.048	7,206.7	< 10 ⁻⁴
Muscle density	66	0.113	1.663	-0.127	83.4	< 10 ⁻⁴

TABLE 12 Statistical measures of the relationships between tissue properties and age for females.

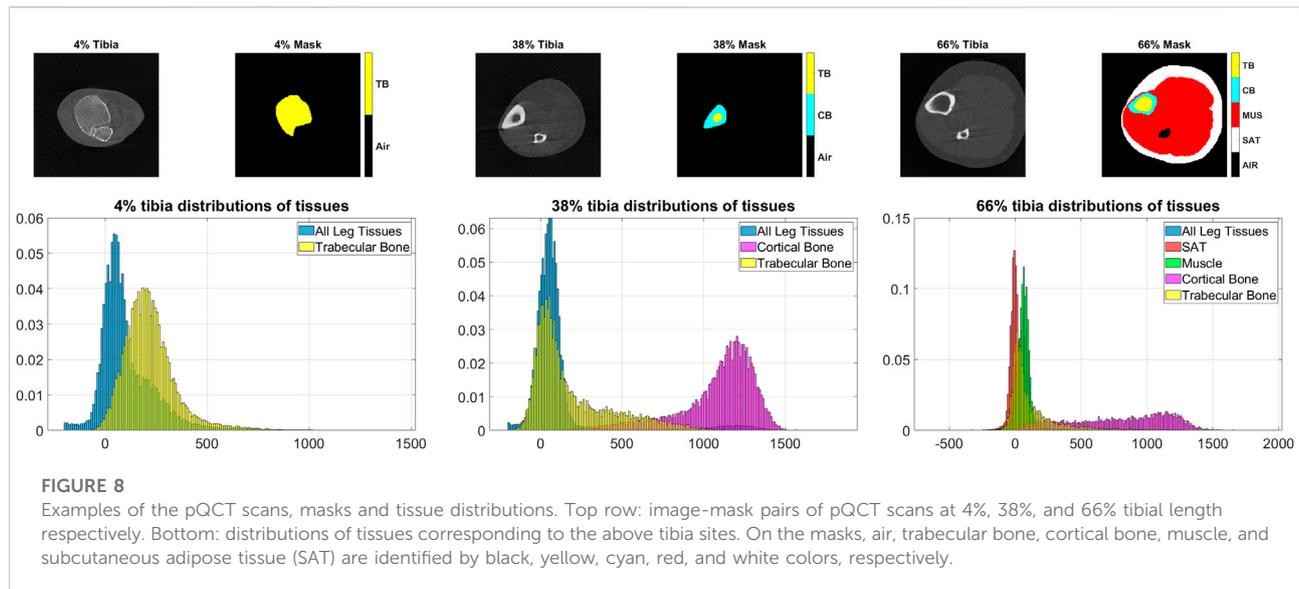
Measurement	Site	R ²	CV-RMSD	Slope	Intercept	p-value
Trabecular CSA	4	0.102	2.159	3.052	887.9	< 10 ⁻⁴
Trabecular density	4	0.348	3.025	-1.911	358.5	< 10 ⁻⁴
Cortical CSA	38	0.086	2.528	-0.779	313.3	< 10 ⁻⁴
Cortical density	38	0.307	0.977	-2.643	1,233	< 10 ⁻⁴
SAT CSA	66	0.010	5.122	-6.067	3,498.9	0.0870
SAT density	66	0.044	17.826	0.132	-0.3	0.0002
Muscle CSA	66	0.034	2.428	-9.817	6,238.6	0.0013
Muscle density	66	0.150	1.185	-0.142	83.3	< 10 ⁻⁴

TABLE 13 Statistical measures of the relationships between tissue properties and age for males.

Measurement	Site	R ²	CV-RMSD	Slope	Intercept	p-value
Trabecular CSA	4	0.013	1.950	1.246	1,229.6	0.0611
Trabecular density	4	0.127	2.704	-1.219	362.7	< 10 ⁻⁴
Cortical CSA	38	0.048	1.937	-0.627	388.7	0.0003
Cortical density	38	0.081	0.554	-0.756	1,147.1	< 10 ⁻⁴
SAT CSA	66	0.032	4.019	-7.441	2,895.9	0.0032
SAT density	66	0.009	5.721	0.075	26.5	0.1141
Muscle CSA	66	0.135	2.362	-26.153	8,364.7	< 10 ⁻⁴
Muscle density	66	0.087	1.098	-0.110	83.5	< 10 ⁻⁴

TABLE 14 Mean ± standard deviation (in seconds) of the execution times of SAIS and MAIS with respect to nonlinear registration models.

Tibia site	Method	FFD	SDD	SyN	SyNO
4%	SAIS	56.47 ± 3.34	27.01 ± 0.81	5.55 ± 0.30	37.25 ± 2.91
	MAIS	191.45 ± 9.54	101.82 ± 1.72	41.27 ± 0.99	136.48 ± 5.61
38%	SAIS	55.49 ± 7.00	31.15 ± 1.46	5.67 ± 0.33	37.16 ± 1.90
	MAIS	189.31 ± 18.62	117.18 ± 3.97	41.88 ± 1.34	140.61 ± 15.20
66%	SAIS	55.47 ± 3.96	31.48 ± 1.29	4.42 ± 4.42	36.38 ± 1.30
	MAIS	209.66 ± 31.41	147.18 ± 36.05	52.51 ± 13.25	151.13 ± 25.72



more segmentation challenges than subject 'A'. Visual inspection of subject 'B' at 4% tibia in Figure 4 shows that STPL-FFD and STPL-SyNO could not delineate the trabecular bone accurately when compared to STPL-SDD and STPL-SyN. Similarly, under 38%-tibia scans in Figure 4, we observe that due to the higher presence of streaks in subject 'B' than in subject 'A', STPL-FFD and STPL-SyNO produced lower tissue delineation accuracy than STPL-SDD and STPL-SyN. This observation provides insight into the improved delineation accuracy produced by symmetric diffeomorphic demons and symmetric normalization.

5.2 Effects of subject motion on tissue identification at 66% tibia

The effect of motion artifacts is less pronounced on muscle identification in terms of the relative differences. We observe statistically significant differences in DSC for STPL-SDD and TIDAQ. TIDAQ also showed statistically significant decrease of muscle true positive rate.

In the cortical bone, most performance changes are not statistically significant. TIDAQ produced p -values smaller than 0.05 for DSC and precision, and STPL-SDD for precision only. In the trabecular bone, STPL-FFD produced p -values smaller than 0.05 for precision, and TIDAQ for DSC and true positive rate. DSC and precision values of the trabecular bone produced by STPL-FFD and STPL-SyN increase from low to high motion group, because of segmentation errors in subjects of the low motion group.

We observe that TIDAQ is significantly affected by motion artifacts, as average DSC clearly decreases from the low to high

subject motion group. MAIS techniques and especially STPL-SyN are more resilient to subject motion. STPL-SDD still produces the highest DSC and precision values in the high motion group overall.

5.3 Tissue composition assessment

The results in Figures 6, 7 and Tables 11–13 are consistent with findings of clinical studies of aging that used semi-manual quantification workflows (Makrogiannis et al., 2018; Ferrucci et al., 2000). These results indicate that our automated methodology can help to increase the throughput of sophisticated cross-sectional and longitudinal analyses of tissue properties. We also note that the proposed methodology enables the analysis of both hard and soft tissues in pQCT. This is a desirable and innovative feature as pQCT has been mostly restricted to quantification of bone in the past (Gabel et al., 2018; Wong and Manske, 2020). Our methodology opens the door for efficient exploration of muscle properties in the lower leg using pQCT. On the other hand, a greater number of reference segmentation masks may be needed to improve the statistical power of performance evaluations.

6 Conclusion

We introduced multi-atlas segmentation methods for soft and hard tissue segmentation in the lower leg using pQCT data. Our results indicate that the MAIS technique, STPL-SDD, produced more accurate tissue delineation as measured by DSC than all compared methods. STPL-SyN is largely resilient

to subject motion artifacts and noise. The results of our experiments indicate that our methodology can analyze data with degradations caused by subject motion that conventional methods cannot analyze. Future directions of this work include extending this framework to 3D imaging data, and using the segmentation and quantification results for disease prognosis and diagnosis.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the National Institute on Research and Care of the Elderly (INRCA) Institutional Review Board (IRB00001456). The patients/participants provided their written informed consent to participate in this study.

Author contributions

(I) Conception and design: SM, LF; (II) Administrative support: SM, LF; (III) Provision of study materials or patients: SB, LF; (IV) Collection and assembly of data: SM, AO, AD, SB, LF; (V) Data

analysis and interpretation: AO, SM, AD; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Funding

This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) (award #SC3GM113754) and by the Army Research Office under grant #W911NF2010095. We acknowledge the support by Delaware CTR-ACCEL (NIH #U54GM104941) and the State of Delaware.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi:10.1016/j.media.2007.06.004
- Blew, R. M., Lee, V. R., Farr, J. N., Schiferl, D. J., and Going, S. B. (2014). Standardizing evaluation of pqct image quality in the presence of subject movement: Qualitative versus quantitative assessment. *Calcif. Tissue Int.* 94, 202–211. doi:10.1007/s00223-013-9803-x
- Burns, J. E., Yao, J., and Summers, R. M. (2020). Artificial intelligence in musculoskeletal imaging: A paradigm shift. *J. Bone Min. Res.* 35, 28–35. doi:10.1002/jbmr.3849
- Cardoso, M. J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., et al. (2013). Steps: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcellation. *Med. Image Anal.* 17, 671–684. doi:10.1016/j.media.2013.02.006
- Chan, A. C., Adachi, J. D., Papaioannou, A., and Wong, A. K. O. (2018). Investigating the effects of motion streaks on pqct-derived leg muscle density and its association with fractures. *J. Clin. Densitom.* 21, 130–139. doi:10.1016/j.jocd.2016.12.001
- Chen, Y., Hu, Y., Yu, Y. E., Zhang, X., Watts, T., Zhou, B., et al. (2018). Subchondral trabecular rod loss and plate thickening in the development of osteoarthritis. *J. Bone Min. Res.* 33, 316–327. doi:10.1002/jbmr.3313
- Chow, S. K. H., van Mourik, M., Hung, V. W. Y., Zhang, N., Li, M. M. C., Wong, R. M. Y., et al. (2022). Hr-pqct for the evaluation of muscle quality and intramuscular fat infiltration in ageing skeletal muscle. *J. Pers. Med.* 12, 1016. doi:10.3390/jpm12061016
- Cordova, J. S., Schreiber, E., Hadjipanayis, C. G., Guo, Y., Shu, H. K. G., Shim, H., et al. (2014). Quantitative tumor segmentation for evaluation of extent of glioblastoma resection to facilitate multisite clinical trials. *Transl. Oncol.* 7, 40–47. doi:10.1593/tlo.13835
- Ferrucci, L., Bandinelli, S., Benvenuti, E., Di Iorio, A., Macchi, C., Harris, T. B., et al. (2000). Subsystems contributing to the decline in ability to walk: Bridging the gap between epidemiology and geriatric practice in the inchiati study. *J. Am. Geriatr. Soc.* 48, 1618–1625. doi:10.1111/j.1532-5415.2000.tb03873.x
- Gabel, L., Macdonald, H. M., Nettlefold, L. A., and McKay, H. A. (2018). Sex-ethnic-and age-specific centile curves for pqct-and hr-pqct-derived measures of bone structure and strength in adolescents and young adults. *J. Bone Min. Res.* 33, 987–1000. doi:10.1002/jbmr.3399
- Iglesias, J. E., and Sabuncu, M. R. (2015). Multi-atlas segmentation of biomedical images: A survey. *Med. Image Anal.* 24, 205–219. doi:10.1016/j.media.2015.06.012
- Langerak, T. R., van der Heide, U. A., Kotte, A. N., Viergever, M. A., Van Vulpen, M., Pluim, J. P., et al. (2010). Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans. Med. Imaging* 29, 2000–2008. doi:10.1109/TMI.2010.2057442
- Lauretani, F., Bandinelli, S., Griswold, M. E., Maggio, M., Semba, R., Guralnik, J. M., et al. (2008). Longitudinal changes in bmd and bone geometry in a population-based study. *J. Bone Min. Res.* 23, 400–408. doi:10.1359/jbmr.071103
- Liu, S., Anwari, V., Yazdankhah, N., Tam, K., Whyte, R., Ha, E., et al. (2022). Bone-muscle-fat interactions and perfusive evidence of inflammation in non-overweight postmenopausal women with early knee oa. *Osteoarthr. Imaging* 2, 100030. doi:10.1016/j.ostima.2022.100030

- Makrogiannis, S., Boukari, F., and Ferrucci, L. (2018). Automated skeletal tissue quantification in the lower leg using peripheral quantitative computed tomography. *Physiol. Meas.* 39, 035011. doi:10.1088/1361-6579/aaaf5
- Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging* 22, 120–128. doi:10.1109/TMI.2003.809072
- McAuliffe, M. J., Lalonde, F. M., McGarry, D., Gandler, W., Csaky, K., and Trus, B. L. (2001). Medical image processing, analysis and visualization in clinical research. IEEE symposium on computer-based medical systems CBMS. 26–27 July 2001. Bethesda, MD, USA. IEEE Computer Society, 381–386.
- Owen, P. J., Hart, N. H., Latella, C., Hendy, A. M., Lamon, S., and Rantalainen, T. (2019). Identifying and assessing inter-muscular fat at the distal diaphyseal femur measured by peripheral quantitative computed tomography (pqct). *J. Clin. Densitom.* 24, 106–111. doi:10.1016/j.jocd.2019.11.001
- Pauchard, Y., Ayres, F. J., and Boyd, S. K. (2011). Automated quantification of three-dimensional subject motion to monitor image quality in high-resolution peripheral quantitative computed tomography. *Phys. Med. Biol.* 56, 6523–6543. doi:10.1088/0031-9155/56/20/001
- Pialat, J., Burghardt, A., Sode, M., Link, T., and Majumdar, S. (2012). Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. *Bone* 50, 111–118. doi:10.1016/j.bone.2011.10.003
- Rodrigues, R., and Pinheiro, A. M. G. Segmentation of skeletal muscle in thigh dixon mri based on texture analysis. arXiv preprint arXiv:1904.04747 (2019).
- Rogelj, P., and Kovačič, S. (2006). Symmetric image registration. *Med. Image Anal.* 10, 484–493. doi:10.1016/j.media.2005.03.003
- Rohlfing, T., Brandt, R., Maurer, C. R., and Menzel, R. (2001). Bee brains, b-splines and computational democracy: Generating an average shape atlas. Proceedings IEEE workshop on mathematical methods in biomedical image analysis (MMBIA 2001). 09–10 December 2001. Kauai, HI, USA. IEEE, 187–194.
- Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, J., and Calvin, R. (2005). *Quo vadis, atlas-based segmentation? The handbook of medical image analysis-volume iii: Registration models.*
- Rowe, G. S., Blazeovich, A. J., and Haff, G. G. (2019). pqct-and ultrasound-based muscle and fat estimate errors after resistance exercise. *Med. Sci. Sports Exerc.* 51, 1022–1031. doi:10.1249/MSS.0000000000001873
- Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging* 18, 712–721. doi:10.1109/42.796284
- Shen, Dinggang, and Hammer, Davatzikos C. (2002). Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21, 1421–1439. doi:10.1109/TMI.2002.803111
- Simon, A., Schäfer, H. S., Schmidt, F. N., Stürznickel, J., Amling, M., and Rolvien, T. (2022). Compartment-specific effects of muscle strength on bone microarchitecture in women at high risk of osteoporosis. *J. Cachexia. Sarcopenia and Muscle* 13, 2310–2321. doi:10.1002/jcsm.13044
- Sode, M., Burghardt, A. J., Pialat, J. B., Link, T. M., and Majumdar, S. (2011). Quantitative characterization of subject motion in hr-pqct images of the distal radius and tibia. *Bone* 48, 1291–1297. doi:10.1016/j.bone.2011.03.755
- Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* 32, 1153–1190. doi:10.1109/TMI.2013.2265603
- Starr, J. F., Bandeira, L. C., Agarwal, S., Shah, A. M., Nishiyama, K. K., Hu, Y., et al. (2018). Robust trabecular microstructure in type 2 diabetes revealed by individual trabecula segmentation analysis of hr-pqct images. *J. Bone Min. Res.* 33, 1665–1675. doi:10.1002/jbmr.3465
- Töpfer, D., Gerner, B., Finzel, S., Kraus, S., Museyko, O., Schett, G., et al. (2015). Automated three-dimensional registration of high-resolution peripheral quantitative computed tomography data to quantify size and shape changes of arthritic bone erosions. *Rheumatology* 54, 2171–2180. doi:10.1093/rheumatology/kev256
- Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2007). Non-parametric diffeomorphic image registration with the demons algorithm. International conference on medical image computing and computer-assisted intervention. Springer, 319–326.
- Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008). Symmetric log-domain diffeomorphic registration: A demons-based approach. International conference on medical image computing and computer-assisted intervention. Springer, 754–761.
- Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., and Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 611–623. doi:10.1109/TPAMI.2012.143
- Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23, 903–921. doi:10.1109/TMI.2004.828354
- Weston, A. D., Korfiatis, P., Kline, T. L., Philbrick, K. A., Kostandy, P., Sakinis, T., et al. (2019). Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* 290, 669–679. doi:10.1148/radiol.2018181432
- Wong, A. K. (2016). A comparison of peripheral imaging technologies for bone and muscle quantification: A technical review of image acquisition. *J. Musculoskelet. Neuronal Interact.* 16, 265–282.
- Wong, A. K. O., and Manske, S. L. (2020). A comparison of peripheral imaging technologies for bone and muscle quantification: A review of segmentation techniques. *J. Clin. Densitom.* 23, 92–107. doi:10.1016/j.jocd.2018.04.001