



DADP: Dynamic abnormality detection and progression for longitudinal knee magnetic resonance images from the Osteoarthritis Initiative

Chao Huang^a, Zhenlin Xu^b, Zhengyang Shen^b, Tianyou Luo^c, Tengfei Li^{d,g}, Daniel Nissman^d, Amanda Nelson^e, Yvonne Golightly^{e,f}, Marc Niethammer^{b,g,**}, Hongtu Zhu^{b,c,g,*}

^a Department of Statistics, Florida State University, Tallahassee, 32304, USA

^b Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, 27599, USA

^c Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, 27599, USA

^d Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, 27599, USA

^e Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, 27599, USA

^f Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, 27599, USA

^g Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, 27599, USA

ARTICLE INFO

Article history:

Received 18 May 2021

Revised 27 November 2021

Accepted 27 December 2021

Available online 1 January 2022

MSC:

41A05

41A10

65D05

65D17

Keywords:

Osteoarthritis

Abnormal region detection

Dynamic functional mixed effect model

Dynamic conditional random field model

ABSTRACT

Osteoarthritis (OA) is the most common disabling joint disease. Magnetic resonance (MR) imaging has been commonly used to assess knee joint degeneration due to its distinct advantage in detecting morphologic cartilage changes. Although several statistical methods over conventional radiography have been developed to perform quantitative cartilage analyses, little work has been done capturing the development and progression of cartilage lesions (or abnormal regions) and how they naturally progress. There are two major challenges, including (i) the lack of building spatial-temporal correspondences and correlations in cartilage thickness and (ii) the spatio-temporal heterogeneity in abnormal regions. The goal of this work is to propose a dynamic abnormality detection and progression (DADP) framework for quantitative cartilage analysis, while addressing the two challenges. First, spatial correspondences are established on flattened 2D cartilage thickness maps extracted from 3D knee MR images both across time within each subject and across all subjects. Second, a dynamic functional mixed effects model (DFMEM) is proposed to quantify abnormality progression across time points and subjects, while accounting for the spatio-temporal heterogeneity. We systematically evaluate our DADP using simulations and real data from the Osteoarthritis Initiative (OAI). Our results show that DADP not only effectively detects subject-specific dynamic abnormal regions, but also provides population-level statistical disease mapping and subgroup analysis.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Osteoarthritis (OA) is a common, chronic disease affecting all joint tissues and characterized by cartilage loss. Over 50 million people have doctor-diagnosed arthritis (Hootman et al., 2016), the majority of which is OA, and even this is likely an underestimate (Jafarzadeh and Felson, 2017), leading to health expenditures totaling over \$27 billion annually (Gupta et al., 2005). OA is commonly associated with symptoms including swelling, pain, discom-

fort, locking, and other problems in mobility, which can be severe (Felson et al., 2000). The disease of OA is also characterized by the progressive loss of joint cartilage, which is believed to be the main factor in OA. Different medical imaging modalities have been employed to investigate and quantitatively measure the knee joint cartilage, such as radiography, Magnetic resonance (MR) imaging, ultrasound (US), Computed Tomography (CT), and optical coherence tomography (OCT) (Braun and Gold, 2012; Kijowski et al., 2020). Among all these available modalities, MR imaging is a non-invasive modality with high spatial resolution. Although MR imaging cannot capture the bone architecture as well as CT, it is useful for capturing some characteristics of the joint, including cartilage morphology and biochemical composition (Guermazi et al., 2003; Roemer et al., 2020). Furthermore, MR imaging is very sensitive

* Corresponding author.

** Co-corresponding author.

E-mail addresses: mn@cs.unc.edu (M. Niethammer), htzhu@email.unc.edu (H. Zhu).

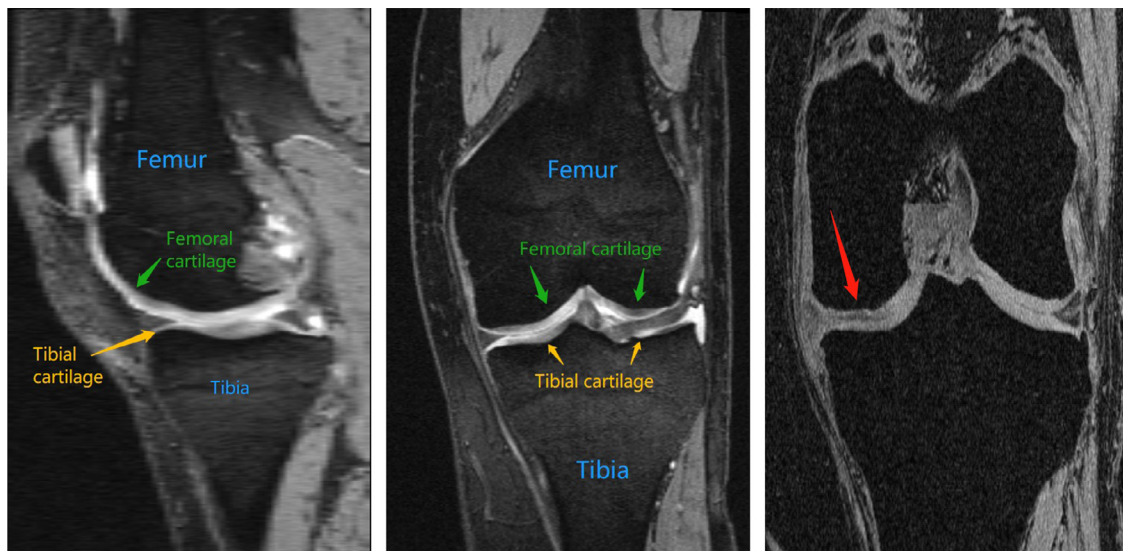


Fig. 1. MR images of the human knee illustrating cartilage loss. (left) A sagittal slice of a healthy knee. Bones are annotated in blue, femoral cartilage in green and tibial cartilage in orange; (middle) A coronal slice of the same healthy knee; (right) A coronal slice of an OA knee with cartilage loss indicated by the red arrow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to full or partial-thickness changes of articular cartilage over time (Huang et al., 2015; Karim et al., 2021). Therefore, MR imaging has been commonly used to detect cartilage morphological changes and understand OA progression (Raynauld, 2003; Cicuttini et al., 2005; Conaghan et al., 2011). Fig. 1 shows MR images of the human knee illustrating cartilage loss.

Several large-scale longitudinal multicenter MR imaging databases, such as the Multicenter Osteoarthritis Study (MOST) (Segal et al., 2013) and the Osteoarthritis Initiative (OAI) (Eckstein et al., 2012), have been acquired to understand the progression of cartilage change and associated risk factors (Felson et al., 2013). Although quantitative MR image-based biomarkers, such as the volume of the articular cartilage (Eckstein et al., 2006; Schaefer et al., 2017), shape information (An et al., 2016; Bowes et al., 2021), and T_2 relaxometry patterns (Pedioia et al., 2017; 2019), show potential for OA diagnosis and treatment planning, little work has been done in assessing the development of cartilage lesions (or abnormal regions), such as soft tissue swelling or small cartilage or bone fragments, and their natural progression, due to several challenges, which will be detailed below. From a clinical perspective, it is also of great interest to delineate the location and progression of abnormal cartilage regions for refining patient treatment.

There are at least two major challenges in existing abnormal region detection approaches. First, spatio-temporal correspondences and correlations in cartilage thickness are required for statistical localized analysis. It is difficult to establish such correspondences due to the small cartilage volume in relation to the remainder of the knee (Huang et al., 2015; Pedioia et al., 2017; 2019; Roach et al., 2021). Second, abnormal regions across both subjects and time points exhibit spatio-temporal heterogeneity. For the spatial heterogeneity, the cartilage loss pattern is only consistent within a small set of the entire population of subjects. For the temporal heterogeneity, over time cartilage damage and cartilage loss have been found from longitudinal OA studies (Crema et al., 2013; Glyn-Jones et al., 2015; Chen et al., 2017). Thus, treating all the subjects across different time points equally and applying standard statistical methods (e.g., voxel-based analysis and subregion-based analysis (Wirth and Eckstein, 2008)) would be problematic.

To address the two challenges above, a Gaussian Hidden Markov Model (GHMM) was proposed in Huang et al. (2015) to simulta-

neously model the progression and abnormality of cartilage thickness. However, there are some drawbacks in GHMM. First, GHMM ignores the temporal heterogeneity and assumes that the subject-specific abnormal region pattern is time-invariant, which would be sub-optimal in practice, especially in characterizing and predicting individual OA progression. Second, for normal control subjects, GHMM applies a voxel-wise linear mixed model and treats all voxels as independent units, which ignores two key functional features of imaging data, including spatial smoothness and spatial correlation. Third, GHMM does not provide the population level statistical disease mapping and the detection of some potential OA-subgroups based on the cartilage thickness abnormality (Liu et al., 2021).

The aim of this work is to propose a Dynamic Abnormality Detection and Progression (DADP) framework. First, longitudinal knee MR images are processed to derive flattened 2D cartilage thickness maps. Second, the subject-specific dynamic abnormal regions are detected based on the thickness maps. Compared to GHMM and other existing approaches, the main contributions of our DADP framework are summarized as follows:

1. **A deep learning based image analysis pipeline** is proposed in DADP to extract 2D cartilage thickness maps from 3D knee MR images, in a common coordinate system, establishing correspondences in space and time;
2. **A dynamic functional mixed effects model** is proposed in DADP to (i) establish the relationship between the cartilage thickness map and a set of covariates of interest, (ii) capture the spatial smoothness and spatio-temporal correlations within the thickness map through functional data analysis tools (Yuan et al., 2014), and (iii) detect the spatio-temporal heterogeneity in subject-specific dynamic abnormal regions through a dynamic conditional random field model (Sutton et al., 2007);
3. **A series of formal statistical inference procedures** are established in DADP to (i) test whether there is any significant effect of some covariates of interest on the cartilage thickness, (ii) detect population level OA risk on the thickness map through the statistical disease mapping approach, and (iii) conduct subgroup analysis through a non-negative matrix factorization method (Ding et al., 2005);
4. **A powerful and user-friendly package** is developed in DADP for the entire OAI MR image dataset. Moreover, our package

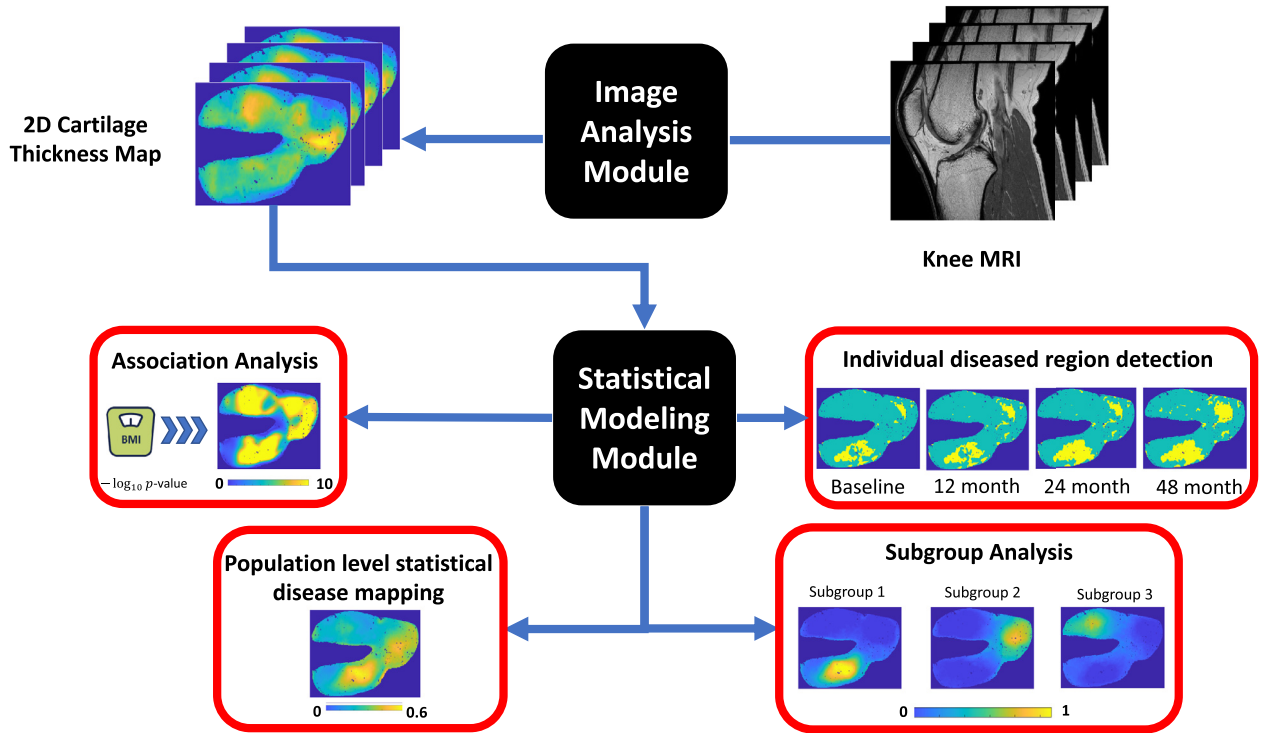


Fig. 2. The DADP framework includes (i) an image analysis module and (ii) a statistical modeling module. *Image analysis module* extracts 2D cartilage thickness maps in a common coordinate system from the 3D knee MR images through a series of image analysis steps. *Statistical modeling module* includes (a) association analysis between cartilage thickness and covariates of interest, (b) individual abnormal regions detection, and (c) multi-level statistical disease mapping.

can be straightforwardly applied for analyzing other knee MR image data at scale. The package will be released to the public through GitHub.

2. DADP

Our DADP includes two main components: (i) an image analysis module and (ii) a statistical modeling module. Specifically, in the image analysis module, the 3D knee MR images are pre-processed through a series of image analysis steps to derive flattened 2D cartilage thickness maps in a common coordinate system. Through this common coordinate system, we obtain spatial correspondence across subjects and/or time points. Then, in the statistical modeling module, a dynamic functional mixed effects model is established based on the 2D cartilage thickness maps to characterize the spatio-temporal correlations within the cartilage thickness map, address the spatio-temporal heterogeneity of abnormal regions, and quantify local OA risk at both population-level and subgroup-level. The workflow of our DADP is summarized in Fig. 2. Next, we introduce the detailed steps in the image analysis module and the estimation and inference procedures in the statistical modeling module.

2.1. Image analysis module

The aim of this image analysis module is to extract spatially aligned 2D thickness maps of femoral/tibial cartilage from the 3D MR images. The whole pipeline consists of four image preprocessing & analysis steps, including ① segmentation & meshing, ② 3D thickness map computation, ③ registration, and ④ thickness map projection. Fig. 3 shows the overall preprocessing pipeline.

First, the femoral cartilage (FC) and tibial cartilage (TC) are segmented from 3D MR images with a U-Net-like network (Ronneberger et al., 2015; Xu et al., 2018), where the model is trained with two parallel output layers. Each layer predicts

a binary segmentation for FC or TC, respectively. Compared to some existing methods for knee cartilage segmentation (e.g., graph model based approach (Yin et al., 2010), 2D U-Net network (Norman et al., 2018), volumetric convolution neural network (Raj et al., 2018), and multi-agent learning (Tan et al., 2019)), our proposed cascaded model of simple U-Nets can match the performance of a complex U-Net, while providing better efficiency in terms of using fewer parameters and requiring less memory (Xu et al., 2018). Based on the resulting (soft) segmentations, we build triangle meshes for each cartilage using marching cubes (Lorensen and Cline, 1987). **Second**, we compute the cartilage thickness for every vertex on the triangle meshes. The thickness value at a vertex on the cartilage surface mesh is approximated by the distance between the vertex's position and its closest point on the opposite side of the surface. Therefore, we divide the cartilage surface mesh into two opposite parts. Some local smoothing is applied to improve incorrect divisions caused by mesh artifacts and imperfect segmentations. **Third**, we align the cartilage meshes to an atlas space using a deep registration network in order to align each MR image to an atlas image. Here the atlas is built via an unbiased atlas-building approach. Specifically, we consider 60 affine-aligned images with segmentations. The initial atlas is simply the average image of these aligned affine images. We then iterate B-Spline registrations (using NiftyReg and NMI as similarity measures) to the current atlas image 10 times. After each round of registrations, the atlas image is updated as the average of the set of deformed images given the current registration transformations. In this way, the atlas gets successively sharper. After the 10 registration rounds, we warp the segmentations given the current transformations from image space to the atlas image and obtain the constructed atlas and a probabilistic segmentation. Compared to some widely-used public registration methods (e.g., SyN (Avants et al., 2008; 2014), Demons (Vercauteren et al., 2008; 2009), NiftyReg (Modat et al., 2010; 2014), and VoxelMorph variant (Dalca et al., 2018)), our proposed registration ap-

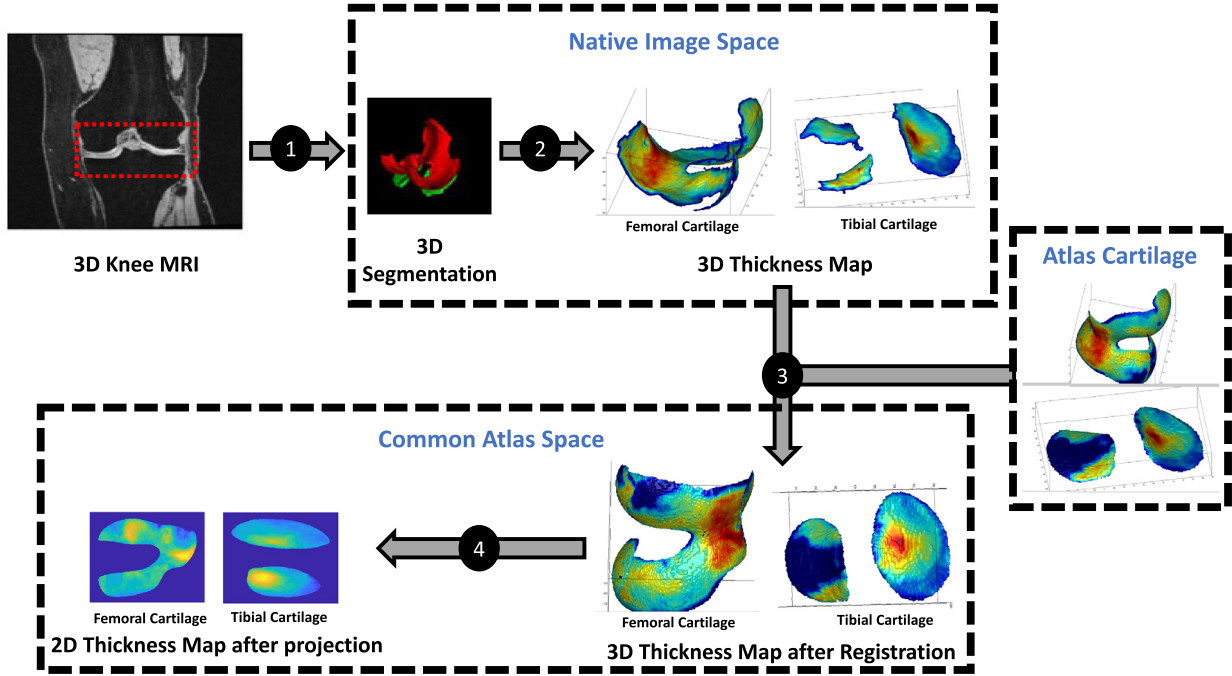


Fig. 3. Image analysis module in DADP: ① segmentation & meshing; ② 3D thickness map computation; ③ registration; and ④ thickness map projection.

proach can achieve comparable and sometimes better performance for longitudinal registration with a dramatically reduced computation time and with excellent deformation regularity and symmetry (Shen et al., 2019). Specifically, the registration involves two phases: an affine registration as pre-alignment followed by a non-parametric registration to capture local deformations. For the non-parametric registration, we use a vector momentum-parameterized stationary velocity field (vSVF) approach (Niethammer et al., 2019). In particular, we use its deep learning version (Shen et al., 2019). We then transfer the cartilage thicknesses to the atlas cartilage mesh. **Fourth**, we project the 3D atlas points onto the 2D plane and interpolate thickness values to obtain spatially aligned flattened 2D cartilage thickness maps for the FC and TC respectively.

2.2. Statistical modeling module

2.2.1. Notations & assumptions

Suppose that we obtain longitudinal imaging data and clinical variables for n unrelated subjects. Let n_i be the total number of observations for the i -th subject for $i = 1, \dots, n$. Our images consist of preprocessed 2D cartilage thickness maps calculated from the OAI study, which are registered to a common template $S \subset \mathbb{R}^2$ with n_v grid points $\{s_1, \dots, s_{n_v}\}$. For the i -th subject at the j -th time point, we observe (i) 2D imaging data, denoted by $y_{i,j}(s)$, indicating the cartilage thickness at pixel s ; (ii) a p_x -dimensional covariate vector of interest, denoted by $\mathbf{x}_{i,j}$, including the intercept, time-independent covariates (e.g., gender), and time-dependent ones (e.g., age and body mass index (BMI)); and (iii) the Kellgren-Lawrence grade (KLG) (Kellgren and Lawrence, 1957) denoted as $g_{i,j}$. The KLG score corresponds to 5 diagnostic outcomes, starting from normal stage ($g_{i,j} = 0$), early stage ($g_{i,j} = 1$), mild OA ($g_{i,j} = 2$), moderate OA ($g_{i,j} = 3$), to severe OA ($g_{i,j} = 4$). Without loss of generality, we assume that the first n^* subjects are OA patients with at least one observation being at OA stage, whereas the rest $n - n^*$ subjects are normal controls with all observations being at normal stage. In addition, for the i -th subject at the j -th time point, we assume that S can be decomposed into the union

of normal regions $\mathcal{R}_{i,j}^0$ and abnormal regions $\mathcal{R}_{i,j}^1$, that is

$$S = \mathcal{R}_{i,j}^0 \cup \mathcal{R}_{i,j}^1 \quad \text{and} \quad \mathcal{R}_{i,j}^0 \cap \mathcal{R}_{i,j}^1 = \emptyset. \quad (1)$$

Here we also assume that: (i) observations at the normal stage are perfectly healthy and do not exhibit any abnormal regions, i.e., $\mathcal{R}_{i,j}^1 = \emptyset$; (ii) for OA patients, the shape, size, and location of abnormal regions $\mathcal{R}_{i,j}^1$ may vary across subjects and time points; and (iii) the abnormal regions satisfy an inclusion property such that $\mathcal{R}_{i,j}^1 \subset \mathcal{R}_{i,j+1}^1$ for $j = 1, \dots, n_i - 1$. To further illustrate these assumptions, Fig. 4 shows three different abnormal region patterns. In this example, patients at the normal stage do not have any abnormal regions, while the shape, size, and location of abnormal regions are different across subjects and time points. In particular, once an abnormal region is present it does not disappear and cannot shrink at later time points.

2.2.2. DFMEM: Dynamic functional mixed effects model

We propose a dynamic functional mixed effects model (DFMEM) to simultaneously detect individual abnormal regions and capture the relationship between cartilage thickness and covariates of interest. The path diagram of DFMEM is presented in Fig. 5. Specifically, DFMEM is defined as

$$y_{i,j}(s) = \mathbf{x}_{i,j}^T \boldsymbol{\beta}(s) + b_{i,j}(s) \mathbf{w}_{i,j}^T \boldsymbol{\alpha} + \mathbf{z}_{i,j}^T \boldsymbol{\gamma}_i(s) + \eta_{i,j}(s) + \epsilon_{i,j}(s) \quad (2)$$

for $j = 1, \dots, n_i$ and $i = 1, \dots, n$. Here $\boldsymbol{\beta}(s)$ is a p_x -dimensional vector representing the fixed effect at pixel s , $\mathbf{w}_{i,j}$ is a p_w -dimensional vector of covariates (commonly a subset of $\mathbf{x}_{i,j}$) associated with the effect $\boldsymbol{\alpha}$ occurring within the abnormal regions (i.e., $\{b_{i,j}(s), s \in S\}$, $b_{i,j}(s) = 0$ if $s \in \mathcal{R}_{i,j}^0$ and $b_{i,j}(s) = 1$ otherwise), and $\mathbf{z}_{i,j}$ is a p_z -dimensional vector of covariates (commonly a subset of $\mathbf{x}_{i,j}$) associated with the random effect $\boldsymbol{\gamma}_i(s)$. In addition, $\eta_{i,j}(s)$ is a stochastic individual function and $\epsilon_{i,j}(s)$ is a measurement error. Furthermore, we assume that $\boldsymbol{\gamma}_i(s)$, $\{\eta_{i,j}(s)\}$, and $\{\epsilon_{i,j}(s)\}$ are independent copies of $\mathbb{GP}(\mathbf{0}, \Sigma_\gamma(s, s'))$, $\mathbb{GP}(0, \sigma_\eta^2(s, s'))$, and $\mathbb{N}(0, \sigma_\epsilon^2(s))$, respectively, where $\mathbb{GP}(\mu(s), \Sigma(s, s'))$ denotes a Gaussian process with mean function $\mu(s)$ and covariance function

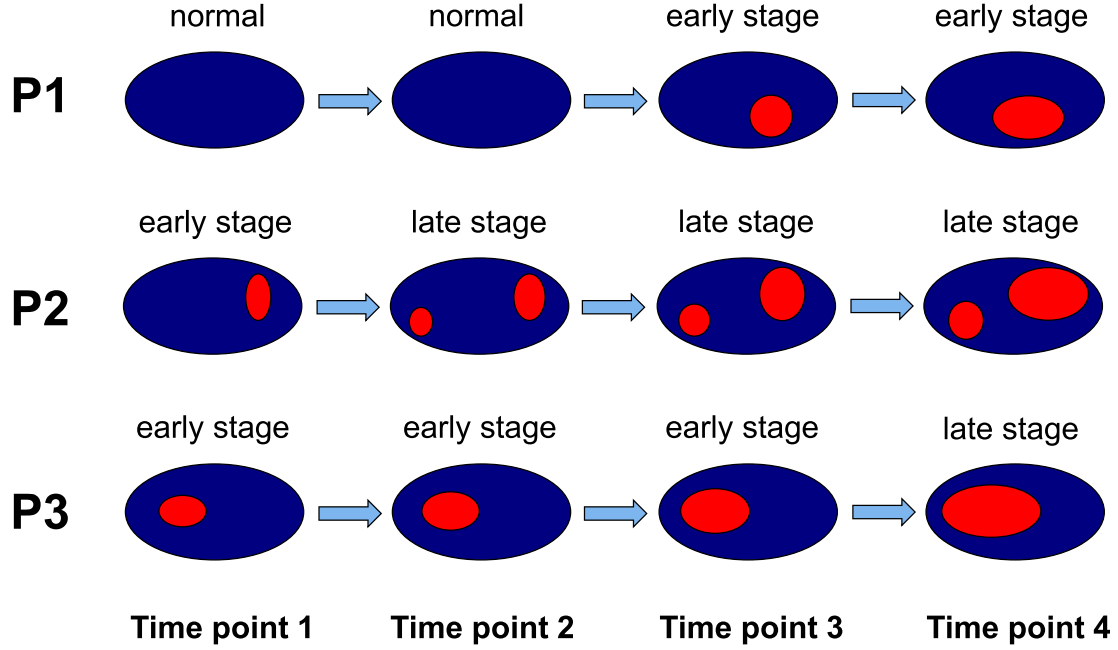


Fig. 4. Three different abnormal region patterns. **P1:** Normal stage for the first two time points and transition to an early stage for the follow-up two time points. One abnormal region (red) occurs at the third time point and becomes larger at the fourth time point; **P2:** Early stage at baseline and transition to late stage for the follow-up three time points. The abnormal regions (red) grow in size and number; **P3:** Early stage at the first three time points and transition to late stage at the fourth time point. One abnormal region (red) grows over time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

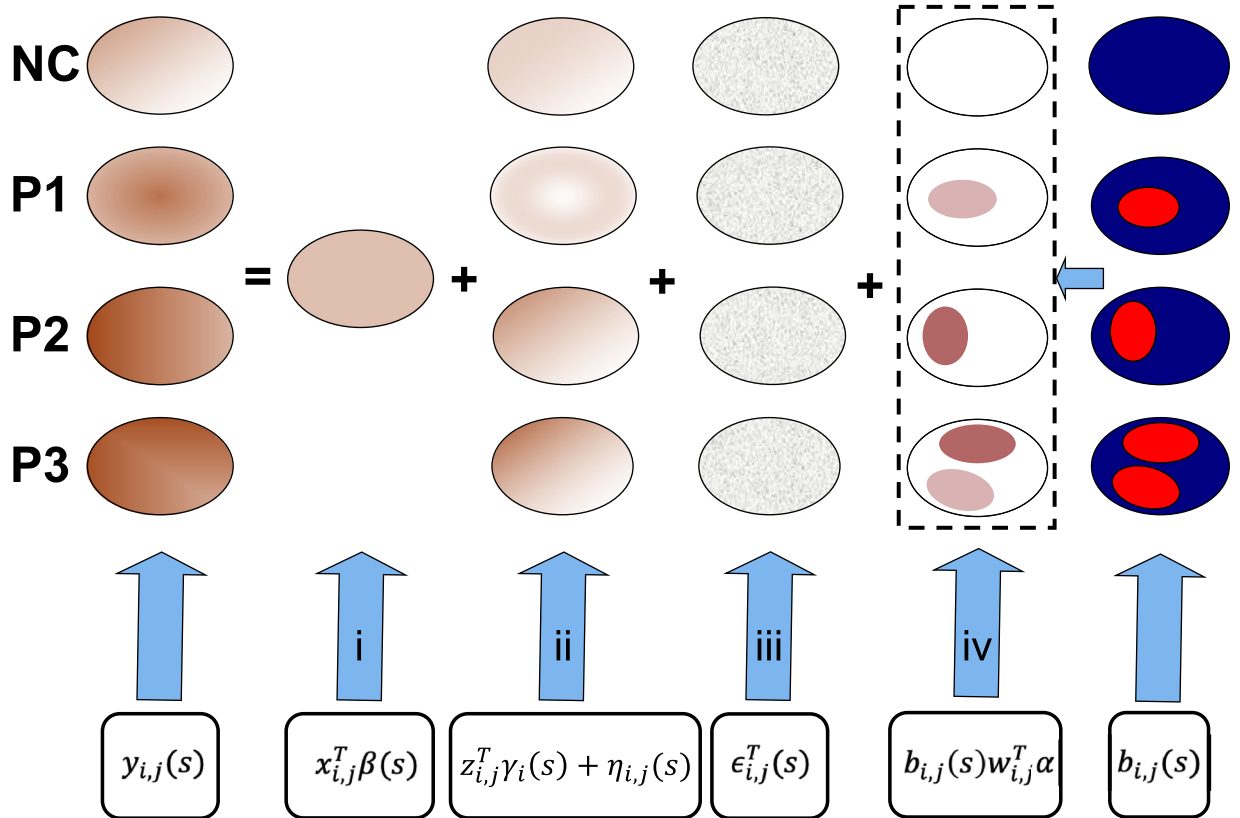


Fig. 5. Path diagram of DFMEM. The change of cartilage thickness at pixel s includes four parts in the proposed model, including (i) the common dynamic change $\mathbf{x}_{i,j}^T \boldsymbol{\beta}(s)$ in both normal controls and OA subjects, (ii) the change caused by random effects $\{\mathbf{z}_{i,j}^T \boldsymbol{\gamma}_i(s), \eta_{i,j}(s)\}$, (iii) measurement errors $\epsilon_{i,j}^T(s)$, and (iv) the change $\mathbf{w}_{i,j}^T \boldsymbol{\alpha}$ caused by dynamic abnormal regions (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\Sigma(s, s')$. Therefore, given $\mathbf{b}_i(s) = (b_{i,1}(s), \dots, b_{i,n_i}(s))^\top$, the covariance structure for $\mathbf{y}_i(s) = (y_{i,1}(s), \dots, y_{i,n_i}(s))^\top$ can be written as

$$\Sigma_{y_i}(s, s') = \mathbf{Z}_i^\top \Sigma_\gamma(s, s') \mathbf{Z}_i + (\sigma_\eta^2(s, s') + \sigma_\epsilon^2(s) \mathbf{1}(s = s')) \mathbf{I}_{n_i}, \quad (3)$$

where $\mathbf{Z}_i = (\mathbf{z}_{i,1}, \dots, \mathbf{z}_{i,n_i})$, $\mathbf{1}(\cdot)$ is an indicator function, and $\mathbf{I}_{n_i}$ is an $n_i \times n_i$ identity matrix. Detailed derivations of (3) can be found in [Appendix A](#). Model (2) consists of four key components: (i) a fixed effect component $\mathbf{x}_{i,j}^\top \boldsymbol{\beta}(s)$ to characterize the fixed effect between cartilage thickness and covariates of interest; (ii) a subject-specific random effect component including $\mathbf{z}_{i,j}^\top \boldsymbol{\gamma}(s)$ to characterize the within-subject spatio-temporal correlation structure and $\eta_{i,j}(s)$ to capture the spatial correlation structure within $\mathcal{S}$; (iii) measurement errors $\epsilon_{i,j}(s)$; and (iv) a disease related random effect component $b_{i,j}(s) \mathbf{w}_{i,j}^\top \boldsymbol{\alpha}$ to characterize the disease related dynamic effect within abnormal regions.

Next, we propose a dynamic conditional random field model (DCRFM) ([Wang and Ji, 2005](#); [Wang et al., 2006](#); [Sutton et al., 2007](#); [Yin et al., 2009](#); [Deufemia et al., 2014](#)) to describe the subject-specific latent variables $\{\mathbf{b}_i(s), s \in \mathcal{S}\}$. Some assumptions are introduced here. First, it is assumed that $\mathbf{b}_i(s)$, $\boldsymbol{\gamma}_i(s)$, $\{\eta_{i,j}(s)\}$, and $\{\epsilon_{i,j}(s)\}$ are mutually independent. Second, it is assumed that the pattern of $b_{i,j}(s)$ at pixel s for the i -th OA patient at the j -th time point is affected by three possible factors: (i) **population factor** $b_0(\mathcal{D}_s) = \{b_0(s), s \in \mathcal{D}_s\}$: patterns of pixels in $\mathcal{D}_s$, including both pixel s and pixels in its neighborhood, at the population level, (ii) **spatial factor** $b_{i,j}(\mathcal{N}_s) = \{b_{i,j}(s), s \in \mathcal{N}_s\}$: patterns of pixels in $\mathcal{N}_s$, the neighborhood of pixel s , for the i -th OA patient at the j -th time point, and (iii) **temporal factor** $b_{i,j-1}(\mathcal{M}_s) = \{b_{i,j-1}(s), s \in \mathcal{M}_s\}$: patterns of pixels in $\mathcal{M}_s$, including both pixel s and pixels in its neighborhood, for the i -th OA patient at the previous time point. Specifically, the pattern of $b_{i,j}(s)$ at baseline is assumed to be affected by the population and spatial factors, whereas that at follow-up visits is assumed to be affected by all the three factors. Further illustration of the three factors is presented in [Fig. 6](#).

In order to model the population level pattern of $b_0(s)$, a Potts model ([Green and Richardson, 2002](#)) is considered here:

$$p_0(b_0(s) | \tau) \propto \exp\{-\tau U(b_0(s), b_0(\mathcal{N}_s))\}, \quad (4)$$

where $U(b_0(s), b_0(\mathcal{N}_s)) = \sum_{s' \in \mathcal{N}_s} \mathbf{1}(b_0(s) \neq b_0(s'))$ and the turning parameter τ is introduced to encourages spatial smoothness in homogeneous regions. Then, given $b_0(\mathcal{D}_s)$ and $\mathbf{b}_i(\mathcal{N}_s)$, $\mathbf{b}_i(s)$ is assumed to follow a DCRFM with the conditional probability mass function, $p(\mathbf{b}_i(s) | \mathbf{b}_i(\mathcal{N}_s), b_0(\mathcal{D}_s), \tau, \kappa, \rho)$ given by

$$p_1(b_{i,1}(s) | b_{i,1}(\mathcal{N}_s), b_0(\mathcal{D}_s), \tau, \kappa) \times \prod_{j=2}^{n_i} p_2(b_{i,j}(s) | b_{i,j-1}(\mathcal{M}_s), b_{i,j}(\mathcal{N}_s), b_0(\mathcal{D}_s), \tau, \kappa, \rho), \quad (5)$$

where $p_1(b_{i,1}(s) | b_{i,1}(\mathcal{N}_s), b_0(\mathcal{D}_s), \tau, \kappa)$ is the conditional probability mass function of $b_{i,j}(s)$ at baseline and proportional to

$$\exp\{-\tau U(b_{i,1}(s), b_{i,1}(\mathcal{N}_s)) - \kappa U(b_{i,1}(s), b_0(\mathcal{D}_s))\}, \quad (6)$$

in which κ is introduced to control the population factor affecting the subject-specific abnormal region pattern at pixel s . In addition, $p_2(b_{i,j}(s) | b_{i,j-1}(\mathcal{M}_s), b_{i,j}(\mathcal{N}_s), b_0(\mathcal{D}_s), \tau, \kappa, \rho)$ is the conditional probability mass function of $b_{i,j}(s)$ at follow-up visits. Specifically, in order to make sure that the inclusion property of abnormal region is satisfied, $b_{i,j}(s) = 1$ with probability 1 when $b_{i,j-1}(s) = 1$. Otherwise, the conditional probability mass function is proportional to

$$\exp\{-\tau U(b_{i,j}(s), b_{i,j}(\mathcal{N}_s)) - \kappa U(b_{i,j}(s), b_0(\mathcal{D}_s)) - \rho U(b_{i,j}(s), b_{i,j-1}(\mathcal{M}_s))\}. \quad (7)$$

where ρ influences the strength of temporal dependencies. It should be noted that, $\mathcal{D}_s$, $\mathcal{N}_s$, and $\mathcal{M}_s$ may have different sizes. For

simplicity, throughout this work we consider $\mathcal{N}_s$ is the set of the closest 8 neighbors of pixel s , and $\mathcal{D}_s = \mathcal{M}_s = \mathcal{N}_s \cup \{s\}$.

2.2.3. Estimation procedure

Our next task is to detect the individual abnormal regions $\{\mathbf{b}_i(s), s \in \mathcal{S}\}_{i=1}^n$, and estimate all unknown parameters decomposed into three parts: (i) $\boldsymbol{\theta}$ including $\boldsymbol{\beta}(s)$, $\Sigma_\gamma(s, s')$, $\sigma_\eta^2(s, s')$, and $\sigma_\epsilon^2(s)$, (ii) $\boldsymbol{\alpha}$, and (iii) τ, κ, ρ . The whole estimation procedure can be treated as an iterative approach, in which either the detected abnormal regions or estimated parameters in one of the three parts are updated with others fixed. Specifically, at the $(t+1)$ -th iteration, the detected abnormal regions $\{\hat{\mathbf{b}}_i^{(t)}(s), s \in \mathcal{S}\}$ and current estimates $\hat{\Theta}^{(t)} \doteq \{\hat{\boldsymbol{\theta}}^{(t)}, \hat{\boldsymbol{\alpha}}^{(t)}, \hat{\tau}^{(t)}, \hat{\kappa}^{(t)}, \hat{\rho}^{(t)}\}$ are updated until certain criterion is satisfied. The overall estimation procedure is summarized in [Algorithm 1](#). More detailed updating procedure and ini-

Algorithm 1: Estimation procedure in statistical modeling module.

Data: cartilage thickness map $\{y_{i,j}(s), s \in \mathcal{S}\}$, covariate vector $\mathbf{x}_{i,j}$, and KLG $g_{i,j}$, $j = 1, \dots, n_i$, $i = 1, \dots, n$

Result: detected individual dynamic abnormal regions $\{\hat{\mathbf{b}}_i(s), s \in \mathcal{S}\}_{i=1}^n$ and estimated parameters $\hat{\Theta}$

Initialization: $\{\hat{\mathbf{b}}_i^{(0)}(s), s \in \mathcal{S}\}_{i=1}^n$, $\hat{\Theta}^{(0)}$, and $t = 0$;

while stopping criterion not satisfied **do**

Update $\hat{\boldsymbol{\theta}}^{(t)}$ via functional data analysis tools;

Update $\hat{\boldsymbol{\alpha}}^{(t)}$ via a weighted least squares approach;

Update $\{\hat{\mathbf{b}}_i^{(t)}(s), s \in \mathcal{S}\}_{i=1}^n$ via the MRF-MAP approach;

Update $\{\hat{\tau}^{(t)}, \hat{\kappa}^{(t)}, \hat{\rho}^{(t)}\}$ via a pseudo-likelihood approach;

Set $t = t + 1$;

end

Output: $\hat{\mathbf{b}}_i(s) = \hat{\mathbf{b}}_i^{(t)}(s), s \in \mathcal{S}, i = 1, \dots, n$, and $\hat{\Theta} = \hat{\Theta}^{(t)}$.

tialization procedure can be found in [Appendix B](#) and [Appendix D](#), respectively.

2.2.4. Inference procedure

After obtaining all the estimated parameters and detected abnormal regions, we carry out a series of formal statistical inference procedures consisting of three different statistical tools: (i) hypothesis testing on parameters of interest including $\boldsymbol{\beta}(s)$ and $\boldsymbol{\alpha}$; (ii) statistical disease mapping; and (iii) subgroup analysis.

Hypothesis testing. In practice, we are interested in investigating (i) whether there is any significant effect associated with some covariates of interest on the cartilage thickness, and (ii) whether there is any significant effect associated with covariates of interest on the cartilage thickness within the abnormal regions. Both of these two scientific questions can be written as hypothesis testing problems in the following general forms respectively:

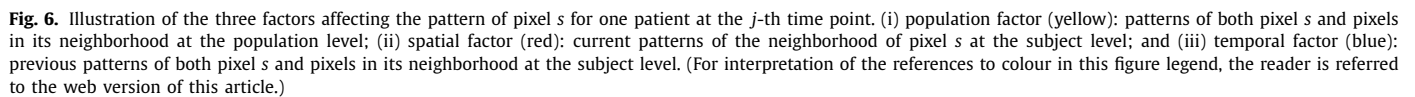
$$\mathbb{H}_0(s) : \mathbf{C}_\beta \boldsymbol{\beta}(s) = 0 \text{ v.s. } \mathbb{H}_1(s) : \mathbf{C}_\beta \boldsymbol{\beta}(s) \neq 0, \quad (8)$$

$$\mathbb{H}_0 : \mathbf{C}_\alpha \boldsymbol{\alpha} = 0 \text{ v.s. } \mathbb{H}_1 : \mathbf{C}_\alpha \boldsymbol{\alpha} \neq 0, \quad (9)$$

where $\mathbf{C}_\beta$ is a $1 \times p_x$ vector and $\mathbf{C}_\alpha$ is a $1 \times p_w$ vector. For testing problem (8), a sequence of Wald tests can be constructed as follows:

$$T_\beta(s) = \mathbf{C}_\beta \hat{\boldsymbol{\beta}}(s) \left[\mathbf{C}_\beta \hat{\Sigma}_\beta(s) \mathbf{C}_\beta^\top \right]^{-1} \hat{\boldsymbol{\beta}}^\top(s) \mathbf{C}_\beta^\top, \quad s \in \mathcal{S}, \quad (10)$$

where $\hat{\Sigma}_\beta(s) = [\sum_{i=1}^n \mathbf{X}_i \{ \mathbf{Z}_i^\top \hat{\Sigma}_\gamma(s, s) \mathbf{Z}_i + \hat{\sigma}_\eta^2(s, s) \mathbf{I}_{n_i} \}^{-1} \mathbf{X}_i^\top]^{-1}$ is the estimated covariance matrix of $\hat{\boldsymbol{\beta}}(s)$. Given the estimate $\hat{\boldsymbol{\alpha}}$ and the


$$T_\alpha = \mathbf{C}_\alpha \hat{\boldsymbol{\alpha}} \left[\mathbf{C}_\alpha \hat{\Sigma}_\alpha \mathbf{C}_\alpha^\top \right]^{-1} \hat{\boldsymbol{\alpha}}^\top \mathbf{C}_\alpha^\top, \quad (12)$$

Subgroup analysis. Besides the abnormal region pattern at both the individual and the population levels, one is also interested

in utilizing the abnormal region based imaging markers to identify subgroups with spatially distinct cartilage changes over time. Specifically, standard subgroup outcomes for OA progression (e.g., KLG) have low reliability, are difficult to interpret, and respond poorly to change (Schiphof et al., 2011; Felson et al., 2011; Guermazi et al., 2012). In comparison, our detected individual dynamic abnormal regions are localized cartilage thickness based features and highly sensitive to the OA progression (Balamoody et al., 2010; Wirth et al., 2013). Therefore, the abnormal region based subgroup analysis has advantages in providing improved subgroup measures, which is important in specific OA prevention and clinical treatment management. In order to conduct this clustering analysis, we consider the non-negative matrix factorization (NMF) method which has an inherent clustering property (Ding et al., 2005). Specifically, we first construct a $n_v \times n^*$ matrix Ω , whose i -th column is defined as $\Omega_i = \sum_{j=1}^{n_i} \omega_{i,j} \mathbf{1}(g_{i,j} > 0) / \sum_{j=1}^{n_i} \mathbf{1}(g_{i,j} > 0)$. Here $\omega_{i,j}$ is a $n_v \times 1$ vector, whose l -th element is the estimated probability $\mathbb{P}(\mathbf{b}_{i,j}(s_l) = 1)$. By adopting the NMF method, the approximation of matrix Ω is achieved by minimizing the following loss function:

$$\|\Omega - MC\|_F, \text{ subject to } 0 \leq M \leq 1, 0 \leq C, \quad (14)$$

where $\|\cdot\|_F$ is the Frobenius matrix norm, M is an $n_v \times k$ matrix, C is a $k \times n^*$ matrix, and k is the number of clusters. The computed matrix C gives the cluster membership, i.e., the i -th patient belongs to the l -th cluster if the element $C_{l,i} > C_{l',i}$ for all $l' \neq l$, while the computed matrix M gives the subgroup level statistical disease mapping, i.e., the l -th column in M represents the disease mapping of the l -th cluster. In addition, the consensus matrix (Brunet et al., 2004; Kim and Park, 2007) is adopted here to determine the number of clusters for both simulation studies and real data analysis. The consensus matrix Λ is defined as the average connectivity matrix whose entry $\Lambda_{i,i'}$ reflects the probability that subjects i and i' belong to the same cluster. To measure the dispersion of a consensus matrix, the dispersion coefficient is defined as $\rho_\Lambda = \sum_{i=1}^{n^*} \sum_{i'=1}^{n^*} 4(\Lambda_{i,i'} - 0.5)^2$. After obtaining ρ_Λ values for various k , we can determine the number of clusters from the maximal ρ_Λ .

3. Experiments and results

3.1. OAI Data description

OAI Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI). OAI is a collaborative informatics system created by the National Institute of Mental Health and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to provide a worldwide resource to quicken the pace of biomarker identification, scientific investigation and OA drug development. Dataset identifier(s): [NIMH Data Archive Collection ID(s) or NIMH Data Archive Digital Object Identifier (DOI)]. If the Research Project involves collaboration with Submitters or NIH staff (as indicated in the DUC), then Recipient will acknowledge Submitters or NIH staff as co-authors, if appropriate, on any presentation, disclosure, or publication.

We consider the longitudinal MR image data from the OAI study, which comprises normal controls and individuals with, or at high risk of, knee OA incidence or progression. The OAI study contains a total of 4796 men and women at ages 45–79 enrolled between 2004 and 2006 (Eckstein et al., 2012). Those subjects consist of a progression sub-cohort (1,389), an incidence sub-cohort (3,285), and a normal control unexposed reference sub-cohort (122) (Peterfy et al., 2008; Nevitt et al., 2006). Further details about OAI can be found on <http://www.oai.ucsf.edu/>.

For the MR images, four 3.0 Tesla Siemens Trio MRI scanners were dedicated to imaging the OAI participants at baseline, 12, 24,

Table 1

Demographic (gender, BMI, and age) and KLG information of patients at baseline and overall visits.

	Left knee		Right knee	
	Baseline	Overall	Baseline	Overall
Gender (F/M)	2,517/1,826	-	2,522/1,817	-
BMI (kg/m ²)	28.6 ± 4.8	28.3 ± 4.8	28.6 ± 4.8	28.3 ± 4.8
Age (years)	61.1 ± 9.2	63.2 ± 9.2	61.1 ± 9.2	63.3 ± 9.2
KLG=0	1729	8339	1627	8223
KLG=1	759	3640	762	3914
KLG=2	1111	4697	1214	5404
KLG=3	603	2491	587	2889
KLG=4	141	748	149	799
Total	4343	19,915	4339	21,229

36, 48, 72 and 96 months. All images have uniform dimension of $384 \times 384 \times 160$ and resolution of $0.36 \times 0.36 \times 0.7 \text{ mm}^3$ per voxel. After preprocessing, we have 19,915 left knee FC/TC thickness maps from 4,343 subjects and 21,229 right knee FC/TC thickness maps from 4,339 subjects. The demographic and KLG information of the OAI dataset used here is presented in Table 1.

3.2. Simulation studies

We examine the finite sample performance of our DADP framework. Here the simulation data was generated based on the pre-processed 2D thickness maps of normal controls in the OAI study. First, we fitted the following model to the left knee FC thickness maps from normal controls,

$$y_{i,j}(s) = \beta_1(s) + \beta_2(s) \times \text{gender} + \beta_3(s) \times \text{BMI} \\ + \beta_4(s) \times \log(\text{Age}) + \gamma_1(s) + \gamma_2(s) \times \text{BMI} \\ + \gamma_3(s) \times \log(\text{Age}) + \eta_{i,j}(s) + \epsilon_{i,j}(s), \quad (15)$$

where the assumptions on random effects $\{\gamma_l(s)\}_{l=1}^3$, individual stochastic term $\eta_{i,j}(s)$, and measurement error $\epsilon_{i,j}(s)$ are the same as those in model (2). Here both BMI and logarithm transformed age information are normalized through the Z-transformation. Then, the obtained parameter estimators of $\{\beta(s), \Sigma_\gamma(s, s'), \sigma_\eta^2(s, s'), \sigma_\epsilon^2(s)\}$ were used as the true values in model (2). In addition, the true values of elements in α were independently generated from $U(-0.3, 0)$ and $\mathbf{w}_{i,j} = (1, \text{normalized BMI, normalized } \log(\text{Age}))^\top$.

On the other hand, the individual abnormal regions were pre-specified according to three types of abnormal region patterns based on the statistical disease mapping learned from the real data (see Fig. 7). Specifically, we first selected two seed points with high probability in the population level statistical disease mapping. For each observation with abnormal region pattern 1, one abnormal region is randomly generated with its center around seed point 1; for each observation with abnormal region pattern 2, one abnormal region is randomly generated with its center around seed point 2; and for each observation with abnormal region pattern 3, two abnormal regions are randomly generated with their centers around two seed points respectively. In addition, the sizes of the abnormal region at baseline and follow-up visits vary across subjects. Finally, we generated 624 thickness maps from 160 subjects with 100 normal controls and 60 diseased patients equally divided into 3 groups according to their abnormal region patterns with 3 or 4 observations for each subject, and the average signal-to-noise ratio (SNR) is 11.44.

First, we fit DFMEM in our DADP with the simulated data and investigate the estimation accuracy of all the functional coefficients $\beta_l(s)$, $l = 1, 2, 3, 4$. The mean absolute error (MAE), defined as $\sum_{k=1}^{n_v} |\hat{\beta}_l(s_k) - \beta_l(s_k)| / n_v$, is calculated and reported in Table 2. For comparison, we consider the spatial random effects model

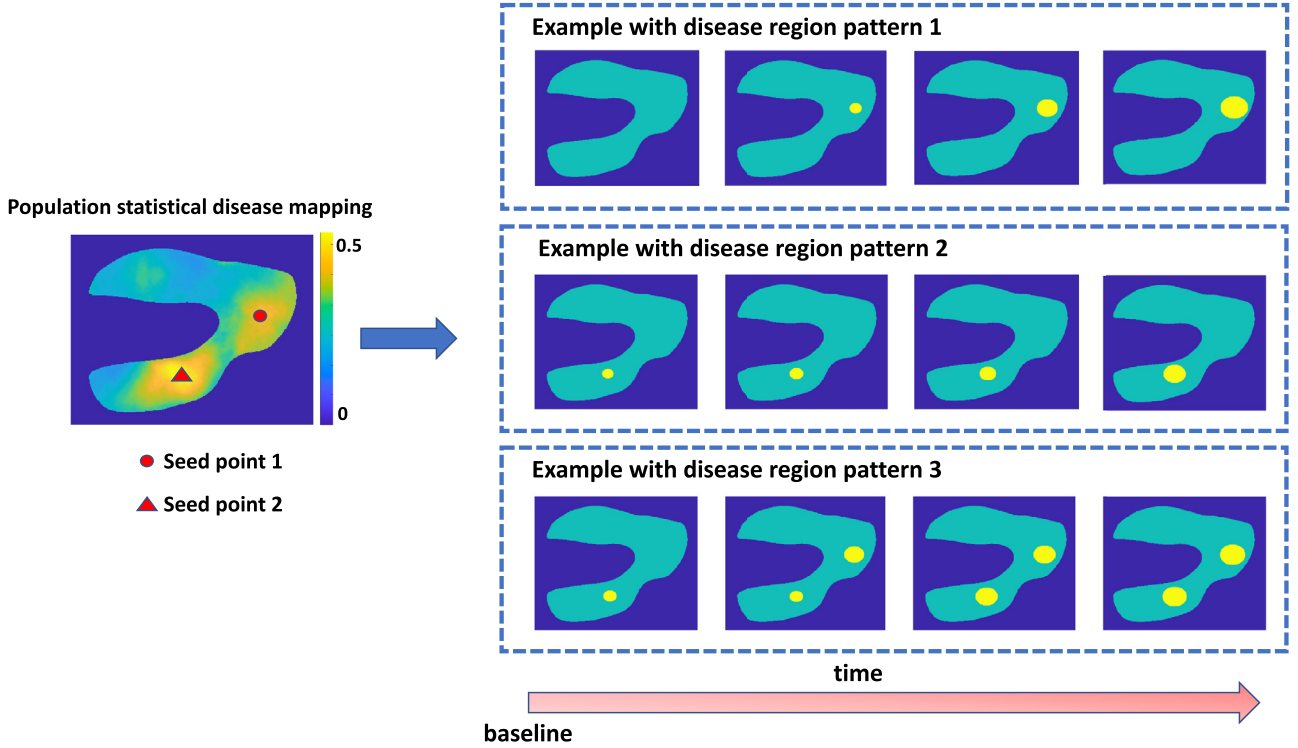


Fig. 7. Illustrations of three predefined abnormal region patterns (background is annotated in dark blue; normal region in light blue; abnormal region in yellow). Two seed points are pre-selected according to the population level statistical disease mapping learned from the left knee FC 2D thickness maps. Pattern 1: one abnormal region is randomly generated around seed point 1; Pattern 2: one abnormal region is randomly generated around seed point 2; Pattern 3: two abnormal regions are randomly generated from two seed points respectively. Examples are presented for the three predefined abnormal region patterns. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Estimation accuracy of functional coefficients $\{\beta_l(s)\}_{l=1}^4$ for DADP and SREM in terms of MAE, defined as $\sum_{k=1}^{n_v} |\hat{\beta}_l(s_k) - \beta_l(s_k)|/n_v$.

Method	$\beta_1(s)$	$\beta_2(s)$	$\beta_3(s)$	$\beta_4(s)$
SREM	0.0452	0.0682	0.0323	0.0322
DADP	0.0134	0.0079	0.0048	0.0051

(SREM, Zhu et al. (2007); Huang et al. (2015)) integrated with DCRFM, and the corresponding MAEs are reported in Table 2 as well. The key difference between SREM and our DFMEM is that the smoothness of both the thickness maps and functional coefficients are considered in DFMEM rather than SREM. According to Table 2, the estimation accuracy based on our DADP outperforms that based on SREM in terms of MAE, showing the power of functional data analysis tools used in our DADP.

Second, we apply our DADP to detect the abnormal regions for each subject at each time point. For three randomly selected abnormal subjects, the detection results are presented in Fig. 8. It can be seen that the abnormal regions for subjects at each time point can be successfully detected, while the detection performance at follow-up visits is better than that at baseline. The possible reason is that the abnormal regions for subjects at baseline are small and the additional effect in the abnormal region is relatively weak. We also apply the *K-means* clustering method, SREM+K (SREM integrated with the *K-means*), and GHMM Huang et al. (2015) to the simulated data. Specifically, the *K-means* clustering method is the initialization procedure in the statistical modeling module of DADP, and GHMM is an integration of SREM with a Potts model. Compared to the alternative approaches, our DADP has three unique features including (i) the smoothness property of thickness maps,

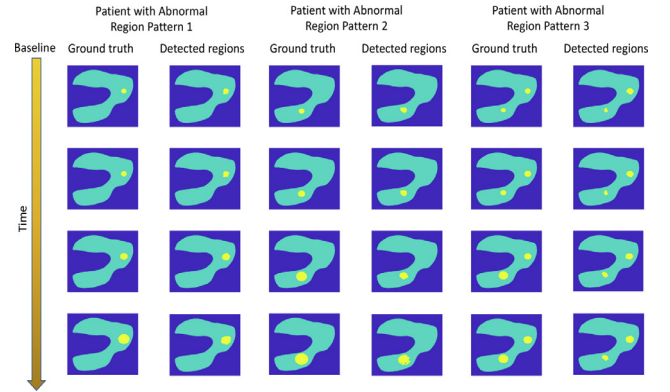


Fig. 8. Detected abnormal regions for four randomly selected subjects including three patients with different abnormal region patterns (background in dark blue; normal region in light blue; abnormal region in yellow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(ii) the dynamic pattern of individual abnormal regions, and (iii) the connection between the individual abnormal region pattern and population level one. For comparison, some metrics, including adjusted Rand index (Hubert and Arabie, 1985), Dice similarity coefficient (Carass et al., 2020), Sensitivity, and Specificity, are considered to evaluate the detection performance of all the four methods for each thickness map, and the mean performance metrics across all thickness maps are reported in Table 3.

It can be found that, our DADP outperforms the other three methods in terms of all the performance metrics, indicating that all three features in DADP enhance its performance in terms of dynamic abnormal region detection.

Table 3

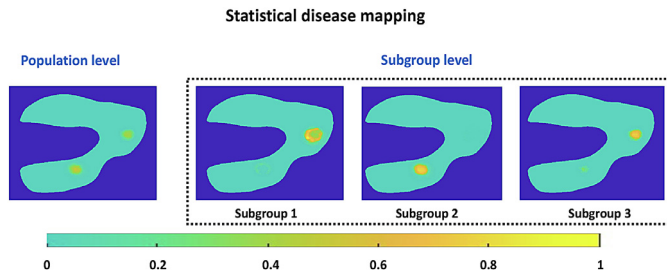
Detection performance for DADP and three other methods (K-means, GHMM, and SREM+K) in terms of adjusted Rand index (ARI), Dice similarity coefficient (DSC), Sensitivity (SEN), and Specificity (SPE). Their corresponding standard errors are presented in parentheses.

	ARI	DSC	SEN	SPE
K-means	0.46 (0.24)	0.47 (0.19)	0.69 (0.29)	0.70 (0.22)
SREM+K	0.61 (0.19)	0.64 (0.14)	0.77 (0.16)	0.75 (0.15)
GHMM	0.75 (0.22)	0.77 (0.16)	0.88 (0.12)	0.86 (0.10)
DADP	0.80 (0.18)	0.81 (0.10)	0.98 (0.09)	0.95 (0.01)

Table 4

Detection performance for DADP with model misspecification and three other methods (K-means, GHMM, and SREM+K) in terms of adjusted Rand index (ARI). Their corresponding standard errors are presented in parentheses.

	K-means	SREM+K	GHMM	DADP
	0.45 (0.28)	0.58 (0.18)	0.71 (0.24)	0.77 (0.20)

**Fig. 9.** Statistical disease mapping at both population level and subgroup level.**Table 5**

Subgroup analysis based on DADP: relationship between clustering results and abnormal region patterns.

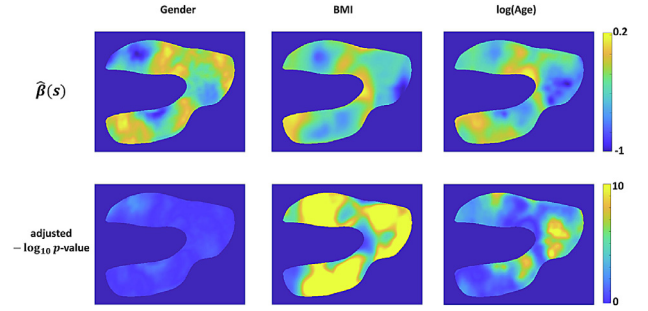
	Subgroup 1	Subgroup 2	Subgroup 3
Pattern 1	11	0	6
Pattern 2	0	24	0
Pattern 3	7	3	9

Next, we investigate the influences of segmentation accuracy on the performance of dynamic abnormal region detection. To mimic the inaccurate segmentation results, we generated the thickness maps from a misspecified model, where the disease related random effect component was constructed as $b_{i,j}(s)[\mathbf{w}_{i,j}^T \boldsymbol{\alpha} + \delta_{i,j}(s)]$ and $\delta_{i,j}(s) \stackrel{\text{iid}}{\sim} N(0, 0.01)$. Here $\delta_{i,j}(s)$ is introduced to capture the potential voxel-wise perturbation caused by the inaccurate segmentation results. The detection performance of all the four methods in terms of ARI is reported in Table 4.

It can be found that, although the ARI of our DADP decreases compared to that in Table 3 (i.e., the “perfect segmentation” scenario), our DADP still achieves reasonable detection performance and outperforms the other three methods when the model is misspecified, which means that our DADP is not sensitive to the accuracy of cartilage segmentation results.

Finally, the statistical disease mapping at the population level is presented in Fig. 9. It can be found that the regions around these two selected seed points are much more likely to be affected by the disease, which is in agreement with the simulation mechanism of abnormal regions. In addition, based on the detected abnormal regions, we conduct the subgroup analysis and present the clustering results in Table 5.

It can be found that subjects with abnormal region pattern 2 are successfully identified by subgroup 2, whereas all patients with abnormal region pattern 1 are clustered into the other two sub-

**Fig. 10.** Left knee FC: estimates of functional coefficients associated to the covariates including gender, normalized BMI, and normalized log(Age) (top); corresponding FDR adjusted $-\log_{10} p$ -value maps (bottom).**Table 6**

Inference on coefficient functions $\beta(s)$: global test statistics and p -values (in parentheses) of three covariates including gender, BMI, and log(Age).

	Gender	BMI (kg/m ²)	log(Age) (years)
Left knee FC	1.02 (0.45)	31.06 (2.3E-3)	11.43 (2.1E-2)
Left knee TC	2.13 (0.17)	20.01 (8.9E-3)	30.12 (9.8E-4)
Right knee FC	0.85 (0.66)	28.41 (3.6E-3)	13.07 (1.0E-2)
Right knee TC	1.69 (0.33)	20.59 (6.2E-3)	31.21 (1.0E-3)

groups. However, the subjects with abnormal region pattern 3 is not well detected due to two reasons. First, pattern 3 has some overlaps with other two patterns in terms of the design mechanism as shown in Fig. 7. Second, although the mechanisms for the generation of the three patterns are not exactly the same, the simulated abnormal region with pattern 3 may be similar to that with other patterns at baseline or some early time points. For example, the simulated disease region with pattern 3 is almost the same as that with pattern 2 at baseline in Fig. 7. Moreover, the corresponding statistical disease mapping at the subgroup level is also presented in Fig. 9, which successfully recovers the predefined abnormal region patterns. Thus, our DADP provides reasonable subgroup analysis in terms of the abnormal region patterns.

3.3. Real data analysis

First, we consider DFMEM (2) with $\mathbf{x}_{i,j} = (1, \text{gender, normalized BMI, normalized log(Age)})^T$, $\mathbf{z}_{i,j} = (1, \text{normalized BMI, normalized log(Age)})^T$, and $\mathbf{w}_{i,j} = \mathbf{z}_{i,j}$. Then, we fit the proposed model on the thickness maps extracted from the left knee FC/TC and the right knee FC/TC, respectively. For the left knee FC, the estimates of coefficient functions associated to the covariates including gender, normalized BMI and normalized log(age) are presented in Fig. 10.

In order to test how different covariates locally affect the cartilage thickness, the pixel-wise Wald test statistics in (10) are calculated, and the FDR adjusted $-\log_{10} p$ -values across all pixels are shown in Fig. 10. For the other three cartilages, the estimates of coefficient functions and corresponding FDR adjusted $-\log_{10} (p)$ -values can be found in Fig. G.14, G.15 and G.16 from Appendix G. According to the adjusted p -value maps, compared to the gender effect, both BMI and age effects are more significant on both medial and lateral regions in all four different cartilages, which is consistent with findings in the existing literature (Cai et al., 2019). Meanwhile, the global effects of these covariates on thickness maps are also investigated. The global test statistics in (11) along with the p -values are calculated and reported in Table 6.

According to the global p -values, the overall BMI effect is found to be more significant than the age effect on both left knee FC and right knee FC, while the overall age effect is more significant than the BMI effect on both left knee TC and right knee TC. In addition

Table 7

Inference on coefficients α : estimates associated to three covariates including gender, BMI, and log(Age); and p -values (in parentheses).

Cartilage	BMI (kg/m ²)	log(Age) (years)
Left knee FC	-0.15 (0.044)	-0.24 (0.012)
Left knee TC	-0.15 (0.046)	-0.26 (0.009)
Right knee FC	-0.18 (0.032)	-0.30 (0.004)
Right knee TC	-0.17 (0.038)	-0.27 (0.009)

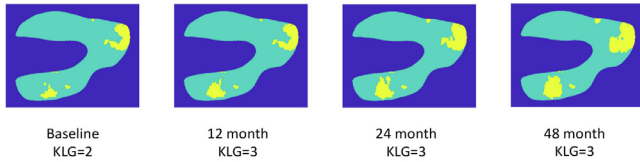
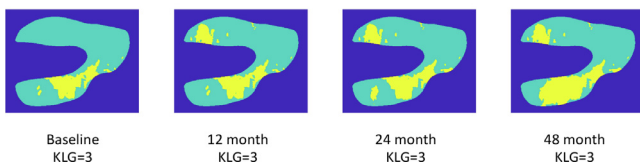
Patient 1 (gender: female, baseline age: 57 years, baseline BMI: 38.6 kg/m²)Patient 2 (gender: male, baseline age: 74 years, baseline BMI: 31.1 kg/m²)

Fig. 11. Left knee FC: detected abnormal regions for two randomly selected patients with observations at baseline, 12 months, 24 months, and 48 months: one female (baseline age: 57 years, baseline BMI: 38.6 kg/m², KLG score at baseline is 2, and 3 at follow-up visits) and one male (baseline age: 74 years, baseline BMI: 31.1 kg/m², KLG score is 2 at all four time points).

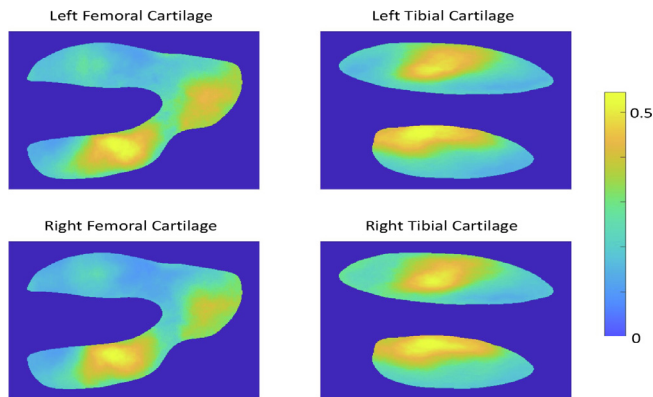


Fig. 12. Population level statistical disease mapping for different cartilages: left knee FC (top left), left knee TC (top right), right knee FC (bottom left), and right knee TC (bottom right).

to the coefficient functions, the inference results of α are presented in Table 7. Both BMI and age effects are found significant on the abnormal regions for all the four cartilages.

Next, the detected abnormal regions on left knee FC for randomly selected two patients are plotted in Fig. 11. Detection results on other cartilages can be found in Fig. H.17, H.18 and H.19 from Appendix H. The corresponding baseline age, baseline BMI, KLG scores, and time points are also presented. It can be found that the abnormal regions vary across subjects and time points and the inclusion assumption is satisfied in the detection results. Besides the detected individual abnormal regions, the population level statistical disease mapping for different cartilages are calculated and reported in Fig. 12. It can be found that, regions with high probability affected by OA are similar between the left knee and the right knee. This symmetry has also been discovered in the existing studies (Metcalf et al., 2012).

Table 8

Subgroup analysis based on the detected individual abnormal regions: clustering results are summarized according to the KLG scores. Three subgroups are detected for left knee FC and right knee FC, while two subgroups are detected for left knee TC and right knee TC.

Cartilage	Subgroup	KLG			
		1	2	3	4
Left knee FC	1	484	3931	713	361
	2	256	326	1449	286
	3	2777	342	179	168
Left knee TC	1	2833	2351	514	405
	2	766	2266	1879	401
Right knee FC	1	597	4594	1220	334
	2	207	441	700	188
	3	3387	415	692	181
Right knee TC	1	2981	2808	701	405
	2	1003	2524	2109	354

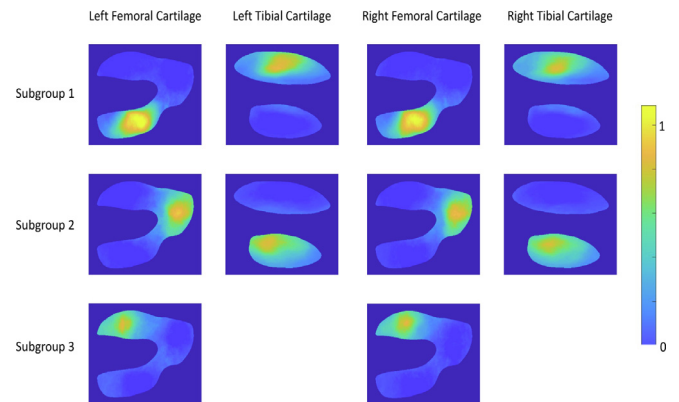


Fig. 13. Subgroup level statistical disease mapping for different cartilages: left knee FC (top left), left knee TC (top right), right knee FC (bottom left), and right knee TC (bottom right).

Finally, the subgroup analysis is conducted based on the detected individual abnormal regions. The clustering results are reported in Table 8, in which the KLG information is reported as well to build up the relationship between clustering results and clinical outcomes. According to the clustering results, 78.9% observations at early stage (KLG score 1) are clustered into subgroup 3, 85.5% observations at mild stage (KLG score 2) are clustered into subgroup 1, and 89.0% observations at moderate and severe stages (KLG score 3 or 4) are clustered into subgroups 1 and 2. Similar clustering results can be found on the right knee FC. Specifically, 80.8% observations at early stage (KLG score 1) are clustered into subgroup 3, 84.3% observations at mild stage (KLG score 2) are clustered into subgroup 1, and 73.7% observations at moderate and severe stages (KLG score 3 or 4) are clustered into subgroups 1 and 2. Therefore, the individual abnormal regions on both left knee FC and right knee FC can be treated as useful biomarkers in identifying OA patients at early stage.

Moreover, the statistical disease maps at the subgroup level are presented in Fig. 13. It can be found that, for all the different cartilages, each subgroup has its unique statistical disease map, which may be treated as a new imaging biomarker for OA early prevention and treatment development.

4. Conclusions & discussions

In this work, a dynamic abnormal region detection framework, including an image analysis module and a statistical modeling module, has been proposed for quantitative cartilage analysis. In the image analysis module, the spatial correspondence has been

well established on 2D cartilage thickness maps extracted from 3D knee MR images through a series of image analysis steps. In the statistical modeling module, a dynamic functional mixed effects model has been proposed. Specifically, the relationship between cartilage thicknesses and covariates of interest has been characterized through functional regression analysis tools, the spatial smoothness and spatial-temporal correlation within the thickness map has been established through functional data analysis tools, and the spatial-temporal heterogeneity in abnormal regions has been detected through dynamic conditional random field models. In addition to the subject-specific dynamic abnormal regions, the population-level statistical disease mapping has been derived, and subgroup analysis has also been conducted through the non-negative matrix factorization method. The proposed framework has been assessed through both simulation studies and real data analysis from the OAI study. It shows that our method outperforms other existing approaches in detecting the subject-specific dynamic abnormal regions and in providing population-level statistical disease mapping and subgroup analysis. All the results derived from our method bring new insights in OA prediction and clinical practice.

Several important issues need to be addressed in our future research. First, the estimation procedure in our DADP can be treated as an iterative approach, and the performance strongly depends on the initialization of abnormal regions and parameter estimation. “Bad” initial values could make the whole algorithm converge to some local maxima or even collapse. To address this issue, a Bayesian framework (Jin et al., 2016; Guo et al., 2020) can be considered here instead, where the estimated parameters can be derived from the posterior distribution via introducing proper prior information, and the latent abnormal regions can be detected via the Monte Carlo Markov Chain (MCMC) sampling method.

Second, in order to detect the subject-specific dynamic abnormal regions, a DCRFM is adopted in this manuscript. Although it works well in our simulated dataset, the simple one-layer latent structure in DCRFM would be inefficient once the spatio-temporal heterogeneity is more complicated within the imaging dataset. To address this issue, the idea of attention-based deep multiple instance learning (Ilse et al., 2018) can be considered and extended to longitudinal datasets. Specifically, both the individual abnormal regions and statistical disease mapping at the subgroup level can be learned through the estimated attention weights.

Third, our current subgroup analysis is conducted via applying the NMF method on the average probability that each pixel belongs to the abnormal region across time, where the dynamic abnormality pattern could not be perfectly characterized. It is of great interest to develop some cartilage thickness progression based features that can integrate the abnormality information to better capture the dynamic abnormality pattern.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Chao Huang: Formal analysis, Methodology, Writing – original draft. **Zhenlin Xu:** Data curation, Writing – original draft. **Zhengyang Shen:** Data curation, Writing – original draft. **Tianyou Luo:** Data curation, Validation. **Tengfei Li:** Data curation, Validation. **Daniel Nissman:** Writing – review & editing. **Amanda Nelson:** Writing – review & editing. **Yvonne Golightly:** Writing – review & editing. **Marc Niethammer:** Supervision, Writing – review & editing. **Hongtu Zhu:** Supervision, Writing – review & editing.

Acknowledgments

The work was supported by the [National Institutes of Health](#) (NIH) grants ([1R01AR072013](#), [MH086633](#), [MH116527](#)) and the [National Science Foundation](#) (NSF) grant ([DMS-1953087](#)). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

Appendix A. Covariance function $\Sigma_{y_i}(s, s')$

Define $\eta_i(s) = (\eta_{i,1}(s), \dots, \eta_{i,n_i}(s))^T$ and $\epsilon_i(s) = (\epsilon_{i,1}(s), \dots, \epsilon_{i,n_i}(s))^T$. Then the covariance function $\Sigma_{y_i}(s, s')$ given $\mathbf{b}_i(s)$ can be derived as

$$\mathbf{Z}_i^T \mathbf{Cov}[\mathbf{y}_i(s), \mathbf{y}_i(s')] \mathbf{Z}_i + \mathbf{Cov}[\eta_i(s), \eta_i(s')] + \mathbf{Cov}[\epsilon_i(s), \epsilon_i(s')]. \quad (\text{A.1})$$

According to the assumptions on $\mathbf{y}_i(s)$, $\eta_{i,j}(s)$, and $\epsilon_{i,j}(s)$, we have $\mathbf{Cov}[\mathbf{y}_i(s), \mathbf{y}_i(s')] = \Sigma_{\gamma}(s, s')$, $\mathbf{Cov}[\eta_i(s), \eta_i(s')] = \sigma_{\eta}^2(s, s') \mathbf{I}_{n_i}$, and $\mathbf{Cov}[\epsilon_i(s), \epsilon_i(s')] = \sigma_{\epsilon}^2(s) \mathbf{1}(s = s') \mathbf{I}_{n_i}$. Therefore, the covariance structure of $\mathbf{y}_i(s)$ given $\mathbf{b}_i(s)$ in (3) holds.

Appendix B. Key steps in estimation procedure

Update $\hat{\theta}^{(t)}$. Given the current detected abnormal regions and estimated parameters in part (ii), model (2) can be written as

$$\tilde{\mathbf{y}}_{i,j}^a(s) = \mathbf{x}_{i,j}^T \boldsymbol{\beta}(s) + \mathbf{z}_{i,j}^T \mathbf{y}_i(s) + \eta_{i,j}(s) + \epsilon_{i,j}(s), \quad (\text{B.1})$$

where $\tilde{\mathbf{y}}_{i,j}^a(s) = y_{i,j}(s) - \hat{\mathbf{b}}_{i,j}^{(t)}(s) \mathbf{w}_{i,j}^T \hat{\boldsymbol{\alpha}}^{(t)}$. Then, model (B.1) can be treated as a spatiotemporal (or functional) linear mixed effects model (Bernal-Rusiel et al., 2013; Yuan et al., 2014; Hyun et al., 2016; Zhu et al., 2019). The updating procedure for θ can be summarized in three key steps: (i) derive an initial estimate of $\boldsymbol{\beta}(s)$ by the local linear regression method (Fan and Gijbels, 1996), while ignoring the temporal structure among repeated observations from one single subject; (ii) derive the updated estimates, $\hat{\Sigma}_{\gamma}^{(t+1)}(s, s')$, $\hat{\sigma}_{\eta}^{2(t+1)}(s, s')$, and $\hat{\sigma}_{\epsilon}^{2(t+1)}(s, s')$, respectively, by the local constant method (Ramsay and Silverman, 2006) given the updated estimate $\hat{\boldsymbol{\beta}}^{(t+1)}(s)$; and (iii) refine the estimated $\hat{\boldsymbol{\beta}}^{(t+1)}(s)$ by incorporating the temporal-spatial structure obtained from the estimated covariance functions. The detailed updating procedure can be found in Appendix C.

Update $\hat{\boldsymbol{\alpha}}^{(t)}$. Let $\mathbf{X}_i = (\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,n_i})$ and $\hat{W}_i(s_k) = W_i \text{diag}(\hat{\mathbf{b}}_i^{(t)}(s_k))$, where $W_i = (\mathbf{w}_{i,1}, \dots, \mathbf{w}_{i,n_i})$ and $\text{diag}(\hat{\mathbf{b}}_i^{(t)}(s_k))$ returns a diagonal matrix with $\hat{\mathbf{b}}_i^{(t)}(s_k)s'$ on the main diagonal. Then, the estimate of $\boldsymbol{\alpha}$ can be updated via minimizing the weighted least squares function

$$\sum_{i=1}^n \sum_{k=1}^{n_v} \|\mathbf{y}_i(s_k) - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}^{(t+1)}(s_k) - \hat{W}_i^T(s_k) \boldsymbol{\alpha}\|_{\hat{\Sigma}_{i,s_k}}^2, \quad (\text{B.2})$$

where $\hat{\Sigma}_{i,s_k} = \mathbf{Z}_i^T \hat{\Sigma}_{\gamma}^{(t+1)}(s_k, s_k) \mathbf{Z}_i + \hat{\sigma}_{\eta}^{2(t+1)}(s_k, s_k)$ and the norm operator $\|\mathbf{a}\|_{\hat{\Sigma}_{i,s_k}}$ is defined as $\sqrt{\mathbf{a}^T \hat{\Sigma}_{i,s_k}^{-1} \mathbf{a}}$.

Update $\hat{\mathbf{b}}_i^{(t)}$. Here the detected abnormal region pattern at both individual level and population level are updated based on the MRF-MAP approach, which is an efficient method in latent variable recovering and adopted in many practical applications such as image segmentation (Zhang et al., 2001; Nie et al., 2009; Ahmadvand et al., 2017). Specifically, the population level abnormal region pattern $\hat{\mathbf{b}}_0^{(t)}(s)$ is updated via minimizing the following loss function

$$\kappa \sum_{g_{i,j} > 0} U(b_0(s), \hat{\mathbf{b}}_{i,j}^{(t)}(\mathcal{D}_s)) + \tau n^* U(b_0(s), \hat{\mathbf{b}}_0^{(t)}(\mathcal{N}_s)), \quad (\text{B.3})$$

while the individual abnormal region pattern is updated via minimizing the following objective function

$$\mathcal{E}_i(s) - \ln p(\mathbf{b}_i(s) | \hat{\mathbf{b}}_i^{(t)}(\mathcal{N}_s), \hat{\mathbf{b}}_0^{(t)}(\mathcal{D}_s), \hat{\tau}^{(t)}, \hat{\kappa}^{(t)}, \hat{\rho}^{(t)}), \quad (\text{B.4})$$

where $\mathcal{E}_i(s) = \frac{1}{2} \|\mathbf{y}_i(s) - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}^{(t+1)}(s) - \mathbf{W}_i^\top \text{diag}(\mathbf{b}_i(s)) \hat{\boldsymbol{\alpha}}^{(t+1)}\|_{\hat{\Sigma}_{i,s}}^2$ and $\hat{\Sigma}_{i,s} = \hat{\Sigma}_{i,s} + \hat{\sigma}_\epsilon^2(s) \mathbf{I}_{n_i}$. Furthermore, the iterated conditional modes (ICM) algorithm (Besag, 1986; Zhang et al., 2001), which uses a greedy iterative strategy for optimization, is adopted to obtain the optimal solutions to (B.3) and (B.4).

Update $\hat{\tau}^{(t)}$, $\hat{\kappa}^{(t)}$, and $\hat{\rho}^{(t)}$. Since τ , κ , and ρ are not the primary parameters of interest, we use an approximate but computationally efficient method based on a pseudo-likelihood function (Huang et al., 2015). A key advantage of using the pseudo-likelihood function is that it does not involve the intractable partition function. The pseudo-likelihood at the $(t+1)$ -th iteration, denoted as $PL(\kappa, \tau, \rho)$, is a simple product of the conditional likelihood

$$\prod_{i=1}^{n^*} \prod_{s \in \partial \mathcal{S}} \frac{p(\hat{\mathbf{b}}_i^{(t+1)}(s) | \hat{\mathbf{b}}_i^{(t+1)}(\mathcal{N}_s), \hat{\mathbf{b}}_0^{(t+1)}(\mathcal{D}_s), \tau, \kappa, \rho)}{\sum_{\mathbf{b}_i(s)} p(\mathbf{b}_i(s) | \hat{\mathbf{b}}_i^{(t+1)}(\mathcal{N}_s), \hat{\mathbf{b}}_0^{(t+1)}(\mathcal{D}_s), \tau, \kappa, \rho)}, \quad (\text{B.5})$$

where $\partial \mathcal{S}$ denotes the set of points at the boundaries of $\mathcal{S}$. Thus, the estimates $\{\hat{\tau}^{(t+1)}, \hat{\kappa}^{(t+1)}, \hat{\rho}^{(t+1)}\}$ can be obtained by maximizing the log-pseudo-likelihood function $\ln PL(\tau, \kappa, \rho)$ through certain numerical algorithm, e.g., Newton-Raphson method.

Appendix C. Key steps in updating the estimate of θ

Step I. We first ignore the temporal structure among repeated observations from one single subject and then derive an initial estimate of $\boldsymbol{\beta}(s)$. Specifically, based on the local linear regression method (Fan and Gijbels, 1996), the initial estimate is obtained via minimizing the following weighted least squares function

$$\sum_{i,j,k} \left(\tilde{y}_{i,j}^a(s_k) - \mathbf{x}_{i,j}^\top (\boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s)(s_k - s)) \right)^2 K_H(s_k - s), \quad (\text{C.1})$$

where $K_H(s) = |H|^{-1} K(H^{-1}s)$, $K(s)$ is the Gaussian kernel function, and H is the positive definite bandwidth matrix.

Step II. Given the updated estimate $\hat{\boldsymbol{\beta}}^{(t+1)}(s)$ from (C.1), we first calculate the residuals

$$\hat{\mathbf{v}}_i(s) = \mathbf{y}_i(s) - \mathbf{X}_i^\top \hat{\boldsymbol{\beta}}^{(t+1)}(s), \quad i = 1, \dots, n. \quad (\text{C.2})$$

Then, let $\sigma_0^2(s, s') \doteq \sigma_\eta^2(s, s') + \sigma_\epsilon^2(s) \mathbf{1}(s = s')$, and the estimates $\hat{\Sigma}_\gamma(s, s')$ and $\hat{\sigma}_0^2(s, s')$ can be derived by minimizing the following least squares function:

$$\sum_{i=1}^n \|\hat{\mathbf{v}}_i(s) \hat{\mathbf{v}}_i^\top(s') - \mathbf{Z}_i^\top \Sigma_\gamma(s, s') \mathbf{Z}_i - \sigma_0^2(s, s') \mathbf{I}_{n_i}\|_F^2. \quad (\text{C.3})$$

Next, by adopting the local constant method (Ramsay and Silverman, 2006), the updated estimates $\hat{\Sigma}_\gamma^{(t+1)}(s, s')$ and $\hat{\sigma}_\eta^{2(t+1)}(s, s')$ can be, respectively, derived by minimizing the following weighted least squares functions,

$$\sum_{k,k'}^{n_\nu} \|\hat{\Sigma}_\gamma(s_k, s_{k'}) - \Sigma_\gamma(s, s')\|_F^2 K_H(s_k - s) K_H(s_{k'} - s'), \quad (\text{C.4})$$

$$\sum_{k,k'}^{n_\nu} \|\hat{\sigma}_0^2(s_k, s_{k'}) - \sigma_\eta^2(s, s')\|_F^2 K_H(s_k - s) K_H(s_{k'} - s'), \quad (\text{C.5})$$

and the estimate $\sigma_\epsilon^{2(t+1)}(s)$ can be consequently written as

$$\sigma_\epsilon^{2(t+1)}(s) = (\hat{\sigma}_0^2(s, s) - \hat{\sigma}_\eta^{2(t+1)}(s, s))_+, \quad (\text{C.6})$$

where the operator $(a)_+$ is defined as $a \cdot \mathbf{1}(a > 0)$.

Step III. Let $\tilde{\mathbf{y}}_i^a(s_k) = (\tilde{y}_{i,1}^a(s_k), \dots, \tilde{y}_{i,n_i}^a(s_k))^\top$ and $\mathbf{X}_i = (\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,n_i})$. Given the estimated covariance functions, we refine the estimate $\hat{\boldsymbol{\beta}}^{(t+1)}(s)$ by minimizing the updated weighted least squares function:

$$\sum_{i=1}^n \sum_{k=1}^{n_\nu} \|\tilde{\mathbf{y}}_i^a(s_k) - \mathbf{X}_i^\top [\boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s)(s_k - s)]\|_{\hat{\Sigma}_{i,s_k}}^2 K_H(s_k - s), \quad (\text{C.7})$$

where the norm operator $\|\mathbf{a}\|_{\hat{\Sigma}_{i,s_k}}$ is defined as $\sqrt{\mathbf{a}^\top \hat{\Sigma}_{i,s_k}^{-1} \mathbf{a}}$,

and $\hat{\Sigma}_{i,s_k} = \mathbf{Z}_i^\top \hat{\Sigma}_\gamma^{(t+1)}(s_k, s_k) \mathbf{Z}_i + \hat{\sigma}_\eta^{2(t+1)}(s_k, s_k)$. In particular, the bandwidth matrices in (C.1), (C.4), (C.5), and (C.7) can be selected by using a leave-one-curve-out cross-validation method (Zhang and Chen, 2007; Zhu et al., 2019).

Appendix D. Initialization procedure

The initialization of abnormal regions and parameter estimation plays a critical role in our iterative estimation procedure. Here, an initial estimation of θ is first derived based on observations at the normal stage. Then, given the estimated $\hat{\theta}^{(0)}$, an initial detection of abnormal regions $\{\hat{\mathbf{b}}_i^{(0)}(s), s \in \mathcal{S}\}_{i=1}^n$ is determined based on the residual maps through a conditional clustering analysis approach. Specifically, the abnormal regions for the i -th OA patient at the disease baseline are initialized via clustering analysis, such as *K-means*, on the residual map, whereas the abnormal regions at follow-up visits are initialized within the detected abnormal regions at the previous time point. Furthermore, the initial estimate of the population level abnormal region pattern is derived from the specified abnormal regions at disease baseline according to the majority rule. In addition to the initialization of abnormal regions, the initial estimator of α can be derived via minimizing the loss function in (B.2). Finally, for the three tuning parameters $\{\tau, \kappa, \rho\}$, we set all of them to 0.5 as initial values, which have been found reasonable according to existing studies (Ahmadvand et al., 2017).

Appendix E. Wild bootstrap procedure for approximating the null distribution of T_β

1. Fit model (2) under the null hypothesis with the original data and retain the fitted values $\hat{y}_{i,j}(s) = \mathbf{x}_{i,j}^\top \hat{\boldsymbol{\beta}}(s) + \hat{b}_{i,j}(s) \mathbf{w}_{i,j}^\top \hat{\boldsymbol{\alpha}}$, individual functions $\hat{f}_{i,j}(s) = \mathbf{z}_{i,j}^\top \hat{\boldsymbol{\gamma}}(s) + \hat{\eta}_{i,j}(s)$, and residuals $\hat{\epsilon}_{i,j}(s) = y_{i,j}(s) - \hat{y}_{i,j}(s) - \hat{f}_{i,j}(s)$;
2. Create synthetic response variables $y_{i,j}^{(l)}(s) = \hat{y}_{i,j}(s) + \zeta_i^{(l)} \hat{f}_{i,j}(s) + \psi_{i,j}^{(l)}(s) \hat{\epsilon}_{i,j}(s)$, where $\zeta_i^{(l)}$ and $\{\psi_{i,j}^{(l)}(s), s \in \mathcal{S}\}$ are random variables following standard normal distribution;
3. Given $\{\hat{b}_{i,j}(s), s \in \mathcal{S}\}$, refit the model using the synthetic response variables and calculate the test statistic $T_\beta^{(l)}$;
4. Aggregate the results of Steps 2 and 3 over $l = 1, \dots, K$ ($K = 500$) to obtain $\{T_\beta^{(l)}\}_{l=1}^K$ and calculate the approximated p -value $p = K^{-1} \sum_{l=1}^K \mathbf{1}(T_\beta^{(l)} > T_\beta)$.

Appendix F. Wild bootstrap method in estimating the covariance of $\hat{\alpha}$

1. Fit model (2) under the null hypothesis with the original data and retain the fitted values $\hat{y}_{i,j}(s) = \mathbf{x}_{i,j}^\top \hat{\boldsymbol{\beta}}(s) + \hat{b}_{i,j}(s) \mathbf{w}_{i,j}^\top \hat{\boldsymbol{\alpha}}$,

- individual functions $\hat{f}_{i,j}(s) = \mathbf{z}_{i,j}^T \hat{\boldsymbol{\gamma}}_i(s) + \hat{\eta}_{i,j}(s)$, and residuals $\hat{\epsilon}_{i,j}(s) = y_{i,j}(s) - \hat{y}_{i,j}(s) - \hat{f}_{i,j}(s)$;
2. Create synthetic response variables $y_{i,j}^{(l)}(s) = \hat{y}_{i,j}(s) + \zeta_i^{(l)} \hat{f}_{i,j}(s) + \psi_{i,j}^{(l)}(s) \hat{\epsilon}_{i,j}(s)$, where $\zeta_i^{(l)}$ and $\{\psi_{i,j}^{(l)}(s), s \in S\}$ are random variables following standard normal distribution;
 3. Given $\{\hat{b}_{i,j}(s), s \in S\}$, refit the model using the synthetic response variables and derive the estimate $\hat{\alpha}^{(l)}$;
 4. Aggregate the results of Steps 2 and 3 over $l = 1, \dots, K$ ($K = 500$) to obtain $\{\hat{\alpha}^{(l)}\}_{l=1}^K$, where the empirical covariance is calculated to represent the covariance matrix of $\hat{\alpha}$.

Appendix G. Inference on coefficient functions for left knee TC and right knee FC/TC

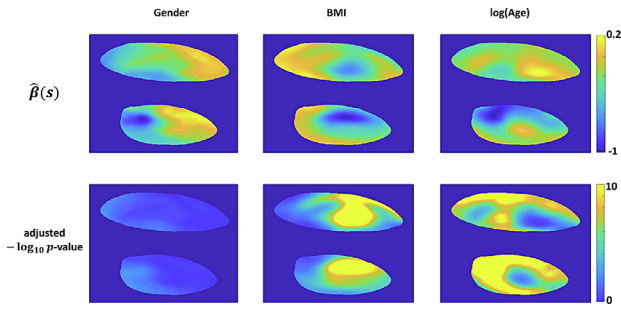


Fig. G.14. Left knee TC: estimates of functional coefficients associated to the covariates including gender, normalized BMI, and normalized log(age) (top); and their corresponding FDR adjusted $-\log_{10} p$ -value maps (bottom).

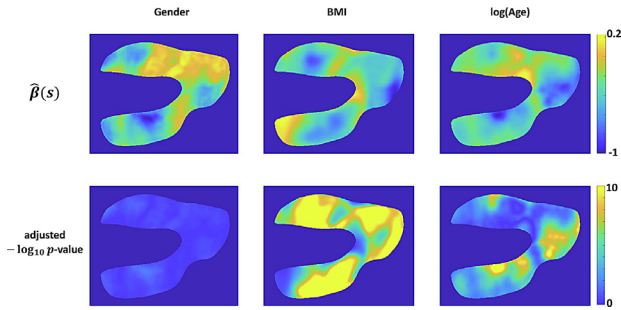


Fig. G.15. Right knee FC: estimates of functional coefficients associated to the covariates including gender, normalized BMI, and normalized log(age) (top); and their corresponding FDR adjusted $-\log_{10} p$ -value maps (bottom).

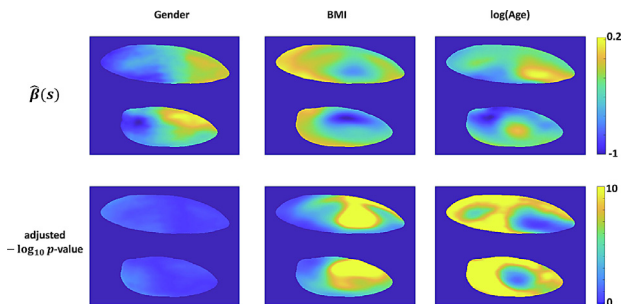


Fig. G.16. Right knee TC: estimates of functional coefficients associated to the covariates including gender, normalized BMI, and normalized log(age) (top); and their corresponding FDR adjusted $-\log_{10} p$ -value maps (bottom).

Appendix H. Detected abnormal regions for left knee TC and right knee FC/TC

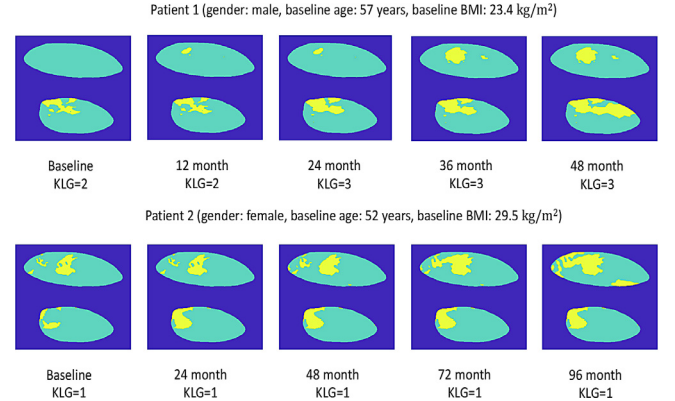


Fig. H.17. Left knee TC: detected abnormal regions for two randomly selected patients with observations at five time points: one male (baseline age: 57 years, baseline BMI: 23.4 kg/m², KLG score is 2 at baseline and 12 months, and 3 at 24 months, 36 months, and 48 months) and one female (baseline age: 52 years, baseline BMI: 29.5 kg/m², KLG score is 1 at all five time points, i.e., baseline, 24 months, 48 months, 72 months, and 96 months).

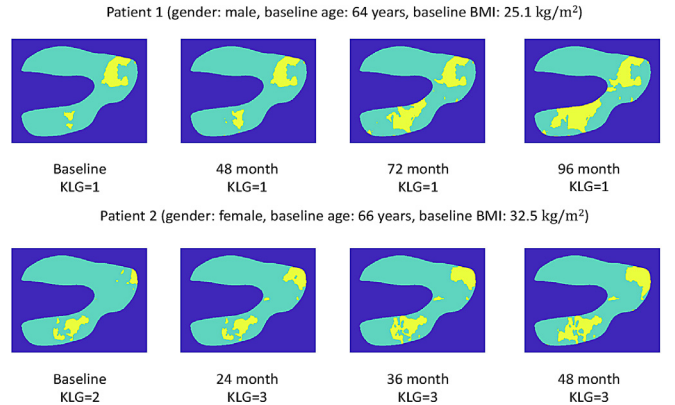


Fig. H.18. Right knee FC: detected abnormal regions for two randomly selected patients with observations at four time points: one male (baseline age: 64 years, baseline BMI: 25.1 kg/m², KLG score is 1 at all the time points including baseline, 48 months, 72 months, and 96 months) and one female (baseline age: 66 years, baseline BMI: 32.5 kg/m², KLG score is 2 at baseline, and 3 at 24 months, 36 months, and 48 months).

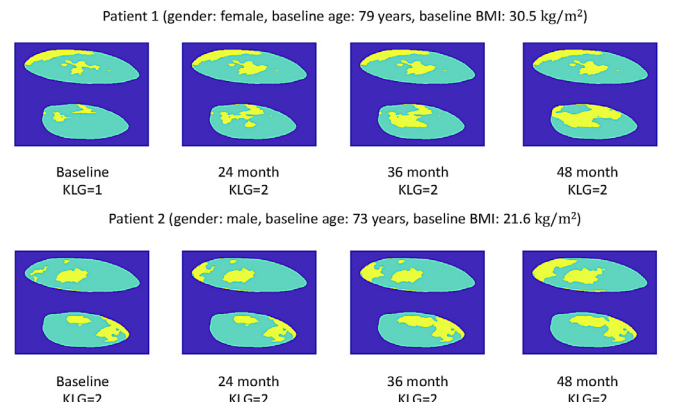


Fig. H.19. Right knee TC: detected abnormal regions for two randomly selected patients with observations at four time points including baseline, 24 months, 36 months, and 48 months: one female (baseline age: 79 years, baseline BMI:

30.5 kg/m², KLG score is 1 at baseline, and 2 at follow-up visits) and one male (baseline age: 73 years, baseline BMI: 21.6 kg/m², KLG score is 2 at all four time points).

References

- Ahmadvand, A., Yousefi, S., Manzuri Shalmani, M.T., 2017. A novel markov random field model based on region adjacency graph for t1 magnetic resonance imaging brain segmentation. *Int. J. Imaging Syst. Technol.* 27 (1), 78–88.
- An, H., Marron, J.S., Schwartz, T.A., Renner, J.B., Liu, F., Lynch, J.A., Lane, N.E., Jordan, J.M., Nelson, A.E., 2016. Novel statistical methodology reveals that hip shape is associated with incident radiographic hip osteoarthritis among african american women. *Osteoarthritis Cartil.* 24 (4), 640–646.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12 (1), 26–41.
- Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C., 2014. The insight toolkit image registration framework. *Front. Neuroinform.* 8, 44.
- Balamoody, S., Williams, T.G., Waterton, J.C., Bowes, M., Hodgson, R., Taylor, C.J., Hutchinson, C.E., 2010. Comparison of 3t MR scanners in regional cartilage-thickness analysis in osteoarthritis: a cross-sectional multicenter, multivendor study. *Arthritis Res. Ther.* 12 (5), 1–9.
- Bernal-Rusiel, J.L., Reuter, M., Greve, D.N., Fischl, B., Sabuncu, M.R., 2013. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage* 81, 358–370.
- Besag, J., 1986. On the statistical analysis of dirty pictures. *J. R. Stat. Soc. Ser. B Methodol.* 48 (3), 259–302.
- Bowes, M.A., Kacena, K., Alabas, O.A., Brett, A.D., Dube, B., Bodick, N., Conaghan, P.G., 2021. Machine-learning, MRI bone shape and important clinical outcomes in osteoarthritis: data from the osteoarthritis initiative. *Ann. Rheum. Dis.* 80 (4), 502–508.
- Braun, H.J., Gold, G.E., 2012. Diagnosis of osteoarthritis: imaging. *Bone* 51 (2), 278–288.
- Brunet, J.-P., Tamayo, P., Golub, T.R., Mesirov, J.P., 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U.S.A.* 101 (12), 4164–4169.
- Cai, G., Jiang, M., Cicuttini, F., Jones, G., 2019. Association of age, sex and BMI with the rate of change in tibial cartilage volume: a 10.7-year longitudinal cohort study. *Arthritis Res. Ther.* 21 (1), 273.
- Carass, A., Roy, S., Gherman, A., Reinhold, J.C., Jesson, A., Arbel, T., Maier, O., Handels, H., Ghafoorian, M., Platel, B., et al., 2020. Evaluating white matter lesion segmentations with refined sørensen-dice analysis. *Sci. Rep.* 10 (1), 1–19.
- Chen, D.I., Shen, J., Zhao, W., Wang, T., Han, L., Hamilton, J.L., Im, H.-J., 2017. Osteoarthritis: toward a comprehensive understanding of pathological mechanism. *Bone Res.* 5 (1), 1–13.
- Cicuttini, F., Hankin, J., Jones, G., Wluka, A., 2005. Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. *Osteoarthritis Cartil.* 13 (8), 722–727.
- Conaghan, P.G., Hunter, D.J., Maillefert, J.-F., Reichmann, W.M., Losina, E., 2011. Summary and recommendations of the OARSI FDA osteoarthritis assessment of structural change working group. *Osteoarthritis Cartil.* 19 (5), 606–610.
- Crema, M.D., Felson, D.T., Roemer, F.W., Wang, K., Marra, M.D., Nevitt, M.C., Lynch, J.A., Torner, J., Lewis, C.E., Guermazi, A., 2013. Prevalent cartilage damage and cartilage loss over time are associated with incident bone marrow lesions in the tibiofemoral compartments: the MOST study. *Osteoarthritis Cartil.* 21 (2), 306–313.
- Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2018. Unsupervised learning for fast probabilistic diffeomorphic registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 729–738.
- Deufemia, V., Risi, M., Tortora, G., 2014. Sketched symbol recognition using latent-dynamic conditional random fields and distance-based clustering. *Pattern Recognit.* 47 (3), 1159–1171.
- Ding, C., He, X., Simon, H.D., 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In: *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM, pp. 606–610.
- Eckstein, F., Cicuttini, F., Raynaud, J.-P., Waterton, J.C., Peterfy, C., 2006. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. *Osteoarthritis Cartil.* 14, 46–75.
- Eckstein, F., Wirth, W., Nevitt, M.C., 2012. Recent advances in osteoarthritis imaging—the osteoarthritis initiative. *Nat. Rev. Rheumatol.* 8 (10), 622.
- Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications. Chapman and Hall, London.
- Felson, D.T., Lawrence, R.C., Dieppe, P.A., Hirsch, R., Helmick, C.G., Jordan, J.M., King-ton, R.S., Lane, N.E., Nevitt, M.C., Zhang, Y., et al., 2000. Osteoarthritis: new insights. part 1: the disease and its risk factors. *Ann. Intern. Med.* 133 (8), 635–646.
- Felson, D.T., Niu, J., Guermazi, A., Sack, B., Aliabadi, P., 2011. Defining radiographic incidence and progression of knee osteoarthritis: suggested modifications of the kellgren and lawrence scale. *Ann. Rheum. Dis.* 70 (11), 1884–1886.
- Felson, D.T., Niu, J., Yang, T., Torner, J., Lewis, C.E., Aliabadi, P., Sack, B., Sharma, L., Guermazi, A., Goggins, J., et al., 2013. Physical activity, alignment and knee osteoarthritis: data from MOST and the OAI. *Osteoarthritis Cartil.* 21 (6), 789–795.
- Glyn-Jones, S., Palmer, A., Agricola, R., Price, A.J., Vincent, T.L., Weinans, H., Carr, A.J., 2015. Osteoarthritis. *Lancet* 386 (9991), 376–387.
- Green, P.J., Richardson, S., 2002. Hidden markov models and disease mapping. *J. Am. Stat. Assoc.* 97 (460), 1055–1070.
- Guermazi, A., Hunter, D.J., Li, L., Benichou, O., Eckstein, F., Kwoh, C.K., Nevitt, M., Hayashi, D., 2012. Different thresholds for detecting osteophytes and joint space narrowing exist between the site investigators and the centralized reader in a multicenter knee osteoarthritis study—data from the osteoarthritis initiative. *Skelet. Radiol.* 41 (2), 179–186.
- Guermazi, A., Zaim, S., Taouli, B., Mieux, Y., Peterfy, C.G., Genant, H.K., 2003. Mr findings in knee osteoarthritis. *Eur. Radiol.* 13 (6), 1370–1386.
- Guo, C., Kang, J., Johnson, T.D., 2020. A spatial bayesian latent factor model for image-on-image regression. *Biometrics*.
- Gupta, S., Hawker, G.A., Laporte, A., Croxford, R., Coyte, P.C., 2005. The economic burden of disabling hip and knee osteoarthritis (OA) from the perspective of individuals living with this condition. *Rheumatology* 44 (12), 1531–1537.
- Hootman, J.M., Helmick, C.G., Barbour, K.E., Theis, K.A., Boring, M.A., 2016. Updated projected prevalence of self-reported doctor-diagnosed arthritis and arthritis-attributable activity limitation among US adults, 2015–2040. *Arthritis Rheumatol.* 68 (7), 1582–1587.
- Huang, C., Shan, L., Charles, H.C., Wirth, W., Niethammer, M., Zhu, H., 2015. Diseased region detection of longitudinal knee magnetic resonance imaging data. *IEEE Trans. Med. Imag.* 34 (9), 1914–1927.
- Hubert, L., Arabie, P., 1985. Comparing partitions. *J. Classif.* 2 (1), 193–218.
- Hyun, J.W., Li, Y., Huang, C., Styner, M., Lin, W., Zhu, H., Initiative, A.D.N., et al., 2016. Stgp: spatio-temporal gaussian process models for longitudinal neuroimaging data. *Neuroimage* 134, 550–562.
- Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning. In: *International Conference on Machine Learning*. PMLR, pp. 2127–2136.
- Jafarzadeh, S.R., Felson, D.T., 2017. Corrected estimates for the prevalence of self-reported doctor-diagnosed arthritis among US adults. *Arthritis Rheumatol.* 69 (8), 1701.
- Jin, I.H., Yuan, Y., Bandyopadhyay, D., 2016. A bayesian hierarchical spatial model for dental caries assessment using non-gaussian markov random fields. *Ann. Appl. Stat.* 10 (2), 884–905.
- Karim, M.R., Jiao, J., Döhmen, T., Cochez, M., Beyan, O., Rebholz-Schuhmann, D., Decker, S., 2021. Deepkneeexplainer: explainable knee osteoarthritis diagnosis from radiographs and magnetic resonance imaging. *IEEE Access* 9, 39757–39780.
- Kellgren, J.H., Lawrence, J.S., 1957. Radiological assessment of osteo-arthritis. *Ann. Rheum. Dis.* 16 (4), 494.
- Kijowski, R., Demehri, S., Roemer, F., Guermazi, A., 2020. Osteoarthritis year in review 2019: imaging. *Osteoarthritis Cartil.* 28 (3), 285–295.
- Kim, H., Park, H., 2007. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23 (12), 1495–1502.
- Liu, R., Zhu, H., for the Alzheimer's Disease Neuroimaging Initiative, 2021. Statistical disease mapping for heterogeneous neuroimaging studies. *Can. J. Stat.* 49 (1), 10–34.
- Lorensen, W.E., Cline, H.E., 1987. Marching cubes: a high resolution 3d surface construction algorithm. *Comput. Graph. (ACM)* 21 (4), 163–169.
- Metcalfe, A.J., Andersson, M.L.E., Goodfellow, R., Thorstensson, C.A., 2012. Is knee osteoarthritis a symmetrical disease? analysis of a 12 year prospective cohort study. *BMC Musculoskelet. Disord.* 13 (1), 153.
- Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S., 2014. Global image registration using a symmetric block-matching approach. *J. Med. Imaging* 1 (2), 024003.
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. *Comput. Methods Programs Biomed.* 98 (3), 278–284.
- Nevitt, M., Felson, D., Lester, G., 2006. The osteoarthritis initiative. Protocol for the Cohort Study 1.
- Nie, J., Xue, Z., Liu, T., Young, G.S., Setayesh, K., Guo, L., Wong, S.T.C., 2009. Automated brain tumor segmentation using spatial accuracy-weighted hidden markov random field. *Comput. Med. Imaging Graph.* 33 (6), 431–441.
- Niethammer, M., Kwitt, R., Vialard, F.-X., 2019. Metric learning for image registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8463–8472.
- Norman, B., Pedoia, V., Majumdar, S., 2018. Use of 2d u-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology* 288 (1), 177–185.
- Pedoia, V., Gallo, M.C., Souza, R.B., Majumdar, S., 2017. Longitudinal study using voxel-based relaxometry: association between cartilage t1ρ and t2 and patient reported outcome changes in hip osteoarthritis. *J. Magn. Reson. Imaging* 45 (5), 1523–1533.
- Pedoia, V., Lee, J., Norman, B., Link, T.M., Majumdar, S., 2019. Diagnosing osteoarthritis from t2 maps using deep learning: an analysis of the entire osteoarthritis initiative baseline cohort. *Osteoarthritis Cartil.* 27 (7), 1002–1010.
- Peterfy, C.G., Schneider, E., Nevitt, M., 2008. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis Cartil.* 16 (12), 1433–1441.
- Raj, A., Vishwanathan, S., Ajani, B., Krishnan, K., Agarwal, H., 2018. Automatic knee cartilage segmentation using fully volumetric convolutional neural networks for evaluation of osteoarthritis. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, pp. 851–854.
- Ramsay, J., Silverman, B.W., 2006. Functional data analysis. Springer New York.

- Raynauld, J.-P., 2003. Quantitative magnetic resonance imaging of articular cartilage in knee osteoarthritis. *Curr. Opin. Rheumatol.* 15 (5), 647–650.
- Roach, K.E., Pedoia, V., Lee, J.J., Popovic, T., Link, T.M., Majumdar, S., Souza, R.B., 2021. Multivariate functional principal component analysis identifies waveform features of gait biomechanics related to early-to-moderate hip osteoarthritis. *J. Orthop. Res.* 39 (8), 1722–1731.
- Roemer, F.W., Demeiri, S., Omoumi, P., Link, T.M., Kijowski, R., Saarakkala, S., Crema, M.D., Guermazi, A., 2020. State of the art: imaging of osteoarthritis-revisited 2020. *Radiology* 296 (1), 5–21.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Schaefer, L.F., Sury, M., Yin, M., Jamieson, S., Donnell, I., Smith, S.E., Lynch, J.A., Nevitt, M.C., Duryea, J., 2017. Quantitative measurement of medial femoral knee cartilage volume—analysis of the OA biomarkers consortium FNIH study cohort. *Osteoarthritis Cartil.* 25 (7), 1107–1113.
- Schiphof, D., de Klerk, B.M., Kerkhof, H.J.M., Hofman, A., Koes, B.W., Boers, M., Bierma-Zeinstra, S.M.A., 2011. Impact of different descriptions of the kellgren and lawrence classification criteria on the diagnosis of knee osteoarthritis. *Ann. Rheum. Dis.* 70 (8), 1422–1427.
- Segal, N.A., Nevitt, M.C., Gross, K.D., Hietpas, J., Glass, N.A., Lewis, C.E., Torner, J.C., 2013. The multicenter osteoarthritis study: opportunities for rehabilitation research. *PM&R* 5 (8), 647–654.
- Shen, Z., Han, X., Xu, Z., Niethammer, M., 2019. Networks for joint affine and non-parametric image registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4224–4233.
- Sutton, C., McCallum, A., Rohanimanesh, K., 2007. Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data. *J. Mach. Learn. Res.* 8 (Mar), 693–723.
- Tan, C., Yan, Z., Zhang, S., Li, K., Metaxas, D.N., 2019. Collaborative multi-agent learning for MR knee articular cartilage segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 282–290.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2008. Symmetric log-domain diffeomorphic registration: A demons-based approach. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 754–761.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage* 45 (1), S61–S72.
- Wang, Y., Ji, Q., 2005. A dynamic conditional random field model for object segmentation in image sequences. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 264–270.
- Wang, Y., Loe, K.-F., Wu, J.-K., 2006. A dynamic conditional random field model for foreground and shadow segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (2), 279–289.
- Wirth, W., Duryea, J., Le Graverand, M.-P.H., John, M.R., Nevitt, M., Buck, R.J., Eckstein, F., Group, O.A.I.I., et al., 2013. Direct comparison of fixed flexion, radiography and MRI in knee osteoarthritis: responsiveness data from the osteoarthritis initiative. *Osteoarthritis Cartil.* 21 (1), 117–125.
- Wirth, W., Eckstein, F., 2008. A technique for regional analysis of femorotibial cartilage thickness based on quantitative magnetic resonance imaging. *IEEE Trans. Med. Imag.* 27 (6), 737–744.
- Xu, Z., Shen, Z., Niethammer, M., 2018. Contextual Additive Networks to Efficiently Boost 3D Image Segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, pp. 92–100.
- Yekutieli, D., Benjamini, Y., 1999. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan. Inference* 82 (1–2), 171–196.
- Yin, J., Hu, D.H., Yang, Q., 2009. Spatio-temporal event detection using dynamic conditional random fields. In: *Proceedings of the 21st International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., pp. 1321–1326.
- Yin, Y., Zhang, X., Williams, R., Wu, X., Anderson, D.D., Sonka, M., 2010. Logismos-layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. *IEEE Trans. Med. Imag.* 29 (12), 2023–2037.
- Yuan, Y., Gilmore, J.H., Geng, X., Martin, S., Chen, K., Wang, J.-L., Zhu, H., 2014. Fmem: functional mixed effects modeling for the analysis of longitudinal white matter tract data. *Neuroimage* 84, 753–764.
- Zhang, J., Chen, J., 2007. Statistical inference for functional data. *Ann. Stat.* 35, 1052–1079.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.* 20 (1), 45–57.
- Zhu, H., Chen, K., Luo, X., Yuan, Y., Wang, J.-L., 2019. Fmem: functional mixed effects models for longitudinal functional responses. *Stat. Sin.* 29 (4), 2007–2033.
- Zhu, H., Gu, M., Peterson, B., 2007. Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm. *Stat. Comput.* 17 (2), 163–177.
- Zhu, H., Kong, L., Li, R., Styner, M., Gerig, G., Lin, W., Gilmore, J.H., 2011. Fadts: functional analysis of diffusion tensor tract statistics. *Neuroimage* 56 (3), 1412–1425.
- Zhu, H., Li, R., Kong, L., 2012. Multivariate varying coefficient model for functional responses. *Ann. Stat.* 40, 2634–2666.
- Zhu, H., Styner, M., Li, Y., Kong, L., Shi, Y., Lin, W., Coe, C., Gilmore, J.H., 2010. Multivariate varying coefficient models for DTI tract statistics. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 690–697.