Journal of Mathematical Imaging and Vision (2024) 66:1033–1059
https://doi.org/10.1007/s10851-024-01211-z

Classification of Deformable Smooth Shapes Through Geodesic Flows of Diffeomorphisms

Hossein Dabirian[1] · **Radmir Sultamuratov**[2] · **James Herring**[4] · **Carlos El Tallawi**[3] · **William Zoghbi**[3] · **Andreas Mang**[2] · **Robert Azencott**[2]

Received: 21 August 2023 / Accepted: 2 August 2024 / Published online: 14 October 2024

Abstract

Let $\mathcal{D}$ be a dataset of smooth 3D surfaces, partitioned into disjoint classes CL_j, $j = 1, \dots, k$. We show how *optimized diffeomorphic registration* applied to large numbers of pairs (S, S'), $S, S' \in \mathcal{D}$ can provide descriptive feature vectors to implement automatic classification on $\mathcal{D}$ and generate classifiers invariant by rigid motions in $\mathbb{R}^3$. To enhance the accuracy of shape classification, we enrich the smallest classes CL_j by diffeomorphic interpolation of smooth surfaces between pairs $S, S' \in CL_j$. We also implement small random perturbations of surfaces $S \in CL_j$ by random flows of smooth diffeomorphisms $F_t : \mathbb{R}^3 \to \mathbb{R}^3$. Finally, we test our classification methods on a cardiology database of discretized mitral valve surfaces.

Keywords Diffeomorphic shape matching · Classification of deformable smooth shapes · Random forests · Data augmentation

1 Introduction

We present research on the classification of deformable mitral valve (MV) shapes based on features derived from diffeomorphic shape matching. In the past two decades, various research groups have developed mathematical algorithms for solving the problem of diffeomorphic registration of deformable 3D shapes (volumes or surfaces; examples are [1–19]). Our approach builds upon the rich mathematical framework referred to as *large deformation diffeomorphic metric mapping* (LDDMM) [4, 20–24]. These techniques have been applied to many biomedical datasets, mostly for a quantitative comparison of "soft" human organs across patients cohorts, or across time for specific, individual patients [25–31]. This field of study is commonly referred to as *computational anatomy* [16, 32–39]. Lucid monographs and reviews about recent developments in (diffeomorphic) registration and shape matching are [1–4]. Key areas of application are neuroimaging [40–48] or cardiac imaging [33, 49–58]. In the present work, we develop computational techniques for the automatic classification of smooth 3D anatomical surfaces extracted from cardiac imaging. The *main goal of this paper* is to explore how to efficiently combine diffeomorphic registration of smooth surfaces with supervised machine learning (ML) for automatic classification of smooth 3D surfaces. In particular, we classify surfaces obtained from time series of 3D echocardiography of MVs based on features derived from diffeomorphic shape matching. In addition, we present methods for data augmentation based on random diffeomorphic perturbation and diffeomorphic interpolation of surfaces. We demonstrate that the proposed methodology allows for a discrimination between *regurgitation* and *normal* (i.e., healthy) MVs by ML.

1.1 Outline of the Method

Let $\mathcal{S}$ denote the set of all compact smooth 3D surfaces properly embedded in $\mathbb{R}^3$, and having piecewise smooth boundaries, where the term "*smooth*" is short-hand for "*of class* C^∞." Let $\mathcal{D}$ be a dataset of discretized smooth 3D sur-

✉ Andreas Mang
andreas@math.uh.edu

Robert Azencott
razencot@math.uh.edu

[1] Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, 1301 Beal Avenue, Ann Arbor, MI 48109, USA

[2] Department of Mathematics, University of Houston, 3551 Cullen Boulevard, Houston, TX 77204, USA

[3] Houston Methodist DeBakey Heart and Vascular Center, Houston Methodist Hospital, 6565 Fannin Street, Houston, TX 77030, USA

[4] Slingshot Aerospace, 5475 Tech Center Dr Suite 225, Colorado Springs, CO (Colorado) 80919, US

faces. When $\mathcal{D}$ is already partitioned into a finite number of disjoint classes, automatic class prediction by supervised ML is a natural goal. However, classical ML classifiers such as *random forests* (RF) [59, 60], *multi-layer perceptrons* (MLPs) [61], *convolution neural networks* (CNNs) [62], or *support vector machines* (SVMs) [63] require the characterization of every surface $S \in \mathcal{D}$ by a computable feature vector $\mathrm{vec}(S) \in \mathbb{R}^N$, for some fixed $N \in \mathbb{N}$. We provide an innovative approach to construct feature vectors $\mathrm{vec}(S)$ boosting the accuracy of automatic classification for smooth surfaces.

We generate several families of intrinsic feature vectors $\mathrm{vec}(S)$ invariant by rigid motions of $\mathbb{R}^3$. To this end, we first construct a family Δ of dissimilarities $\delta(S, S')$ computable for all pairs $(S, S') \in \mathcal{S} \times \mathcal{S}$ of smooth 3D surfaces. These dissimilarities are easily computed after numerical diffeomorphic registration from S to S'. Here, we seek a diffeomorphic spatial transformation $\phi \in \mathrm{diff}(\mathbb{R}^3)$, $\mathrm{diff}(\mathbb{R}^3) \subset \{f : \mathbb{R}^3 \to \mathbb{R}^3\}$ that maps S to S', i.e., verifies $\phi(S) = S'$.

In practice, we cannot expect this equality to hold due to, e.g., errors in the data, numerical errors, and the ill-posedness of the problem. For a numerical treatment, this requirement is relaxed to $\phi(S) \approx S'$; in the variational problem formulation, we minimize a measure of the proximity of $\phi(S)$ and S'.

We define the map ϕ as the endpoint of a time-indexed flows $(F_t)_{t\geq 0}$, $F_t : \mathbb{R}^3 \to \mathbb{R}^3$, of smooth $\mathbb{R}^3$-diffeomorphism. More precisely, let $v_t(x) := v(t, x)$, $v \in L^2([0, 1], \mathcal{V})$, denote a smooth vector field, where $\mathcal{V}$ denotes some Hilbert space that is compactly embedded in the Cartesian product space of one time continuously differentiable functions from $\mathbb{R}^3$ to $\mathbb{R}^3$, who—along with their derivatives—vanish at infinity. We model F_t as the solution of the ordinary differential equation (ODE)

$$\partial_t F_t = v_t(F_t) \quad \text{for almost all} \quad t \in [0, 1], \tag{1}$$

with initial condition $F_0 = \mathrm{id}_{\mathbb{R}^3}$, where $\mathrm{id}_{\mathbb{R}^3}(x) = x$ is the identity transformation. The endpoints F_1 form a group $\mathcal{G} \subseteq \mathrm{diff}(\mathbb{R}^3)$ that, along with a geodesic distance $\rho_{\mathcal{G}} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$, defines a complete metric space $(\mathcal{G}, \rho_{\mathcal{G}})$. In this Riemannian framework, we can measure distances between shapes S, S' by computing the length of the geodesic that maps S to S'. The computation of this geodesic length involves the numerical evaluation of the *kinetic energy* given by

$$\mathrm{kin}(v_t) := \int_0^1 \|v_t\|_{\mathcal{V}}^2 \,\mathrm{d}t, \tag{2}$$

where $\|\cdot\|_{\mathcal{V}}^2$ denotes a fixed Hilbert-norm for smooth vector fields on $\mathbb{R}^3$. The evaluation of (2) requires the solution of a nonlinear optimal control problem. We refer to [4, 24, 64, 65] for a more rigorous discussion and provide additional details below.

For this paper, the family Δ includes the kinetic energy and all quantiles $QUANT_\alpha(S, S')$ of the isotropic strain values for the elastic deformation F_t. To overcome issues that are paramount to applying ML approaches to medical imaging datasets, we introduce two data augmentation techniques. (i) For most ML classifiers, unbalanced class sizes do degrade classification accuracy. To enrich any given class CL of smooth surfaces, we propose here a diffeomorphic interpolation algorithm: For any pair of surfaces $S, S' \in CL$ optimized diffeomorphic registration of S and S' generates a continuous time-indexed flow $(F_t)_{t\geq 0}$ of diffeomorphisms with $F_0(S) = S$ and $F_1(S) = S'$. When the kinetic energy is small enough, the smooth surfaces $S_t = F_t(S)$ can be added to class CL as virtual new cases. (ii) For ML classifiers, robustness is often enhanced when one enriches the training set by small random perturbations of existing training cases. To apply this approach to datasets of smooth surfaces, we generate continuous time flows F_t of *random smooth* $\mathbb{R}^3$ *diffeomorphisms*, by time integration of *smooth* Gaussian random vector fields $V_t(x)$ indexed by time $t \geq 0$ and $x \in \mathbb{R}^3$. For small $t > 0$, the smooth surfaces $F_t(S)$ are the small random deformations of S. Our numerical implementation involves a stochastic series expansion of smooth Gaussian vector fields on $\mathbb{R}^3$, which required the analysis of convergence speed.

1.2 Related Work

The present work is a continuation of our prior work on reconstructing and studying MV dynamics through the lens of geodesic flows of diffeomorphisms [18, 65–74]. The associated database of smooth 3D surfaces was cured and provided by our collaborators from the Houston Methodist DeBakey Heart and Vascular Center. In the present work, we use this dataset as a benchmark for the automatic classification of smooth 3D surfaces.

Pattern recognition, the classification of images, and, in particular, the classification of shapes, have a rich history in computer vision, with an extensive body of literature. Generic descriptors for shape recognition include Fourier descriptors [75], geodesic moments [76], curvature information [77], multiscale fractal dimensions [78, 79], or local contour signatures [80]. In the present work, we focus on the classification of smooth 3D surfaces representing anatomical shapes. In 3D medical imaging sequences, segmentation algorithms are often used to extract these smooth surfaces $S^1, \ldots, S^r \in \mathcal{S}$; they represent the outer surface of soft organs—such as human brains or hearts—as well as the boundaries of their internal chambers, cavities, or ventricles. Classification of these types of anatomical shapes poses significant challenges; they are typically smooth, with subtle differences in their appearance across individuals. These subtle differences necessitate the design of dedicated approaches for their classification. Based on our assessment, we anticipate that many automatic classifiers for 3D images of generic shapes, such as PointNet [81, 82], VoxNet [83, 84], DiffCONV [85], DiffusionNet [86], and similar machine learning (ML) classifiers trained on rigid body datasets like ShapeNet [87, 88] may, in general, not be well-suited for classification tasks involving deformable smooth shapes. These classifiers are primarily trained to quickly recognize rigid objects with bags of features, where many features typically contribute to identification by human vision, such as sharp edges, corners, colors, and textures. However, when dealing with 3D images of deformable soft organs, efficient shape features are then fundamentally different since they must remain reliable under possibly large elastic deformations of shapes. For comparison, we report results for PointNet for the considered benchmark dataset at the end of this manuscript.

There are two main approaches to studying shape variability in computational anatomy: (i) statistical shape models [89–93], and (ii) diffeomorphic shape matching [4, 20, 36–39, 65, 66, 94–98]

The former framework has been developed to aid image segmentation [99–103] or assess the health status of individual patients [104, 105]. The approach considered here falls into the latter category, where comparing two shapes involves the solution of a nonlinear, variational optimization problem. Adjoint-based numerical methods to solve this problem are, e.g., described in [8, 18, 20, 65, 66, 73, 74]. Methods that use automatic differentiation can be found in [106–108]. Instead of solving the optimization problem based on variational algorithms, some recent works try to estimate the diffeomorphic matching of two shapes using ML techniques [109–115]. Related work that quantitatively compares biomedical soft 3D shapes across patients and/or time based on diffeomorphic registration includes [25, 41–48, 110, 116, 117]. In the present work, we are computing families of dissimilarities Δ between all pairs of shapes (S, S'). An alternative approach is to construct a standard reference dataset–a so-called *atlas*–based on a healthy population, and subsequently compare new datasets to this reference [26, 31, 35].

1.3 Contributions

This paper builds upon our prior work on diffeomorphic shape matching and algorithmic reconstruction of unknown dynamics for MV leaflets [18, 65, 66, 66–74, 118]. Here, we are interested in developing a fully automatic computational framework for the classification of deformable shapes using tools derived from diffeomorphic registration of smooth 3D shapes [4, 20, 36–39, 65, 66]. Our *main contributions* are the following:

- We use diffeomorphic registration between pairs S, S' of smooth surfaces to compute several generic families of dissimilarities $\mathrm{dis}(S, S')$ invariant by the group $G3$ generated by rigid motions and homotheties in $\mathbb{R}^3$.
- For automatic classification of generic datasets of smooth 3D surfaces, we construct sets of $G3$-invariant feature vectors derived from the $G3$-invariant dissimilarities $\mathrm{dis}(S, S')$.
- To re-balance class sizes in generic databases of smooth 3D surfaces, we design two novel data enrichment approaches derived from diffeomorphic registrations. One approach is based on diffeomorphic shape interpolation. It extends the SMOTE enrichment technique [119] based on linear interpolation of feature vectors, which is restricted to Euclidean distances. The second approach uses random diffeomorphic shape perturbations by automated simulations of smooth Gaussian random vector fields in $\mathbb{R}^3$.
- For automatic classification in our benchmark dataset of MV surfaces, we develop a localization scheme for dissimilarity computations and tailor it to our discretized MV surfaces to improve the discriminating powers of our feature vectors.
- By implementing multiple RF classifiers based on our $G3$-invariant feature vectors, we perform a comparative importance analysis between nine groups of features to discover the most important of our dissimilarities, which turn out to be high quantiles of strain distributions.
- We successfully apply all our preceding approaches to perform automatic RF-classification of 800 MV surfaces into two classes (“regurgitation” vs “normal” cases), with high OOB accuracy.

1.4 Outline

In Sect. 2, we outline our methodology. We describe our benchmark dataset of smooth surfaces in Sect. 2.1. Our approach to diffeomorphic shape matching is described in Sect. 2.2; we formulate diffeomorphic registration as a classical variational problem and outline our algorithmic approach for diffeomorphic registration. Going beyond the minimal kinetic energy, we introduce families of $G3$-invariant surface dissimilarities, which we derive from diffeomorphic registration. In Sect. 2.3, we develop intrinsic families of $G3$-invariant feature vectors for the automatic classification of 3D surfaces. In Sect. 2.4, we outline two algorithms to enrich the classes of discretized 3D surfaces based on diffeomorphic shape interpolation and random diffeomorphic shape perturbations. In Sect. 3, we discuss our experimental setup and present results for our benchmark dataset of MV surfaces, which includes the overall classification approach (see Sect. 3.1), classification through RFs (see Sect. 3.2), as well as computing times (see Sect. 3.3). We conclude in Sect. 4.

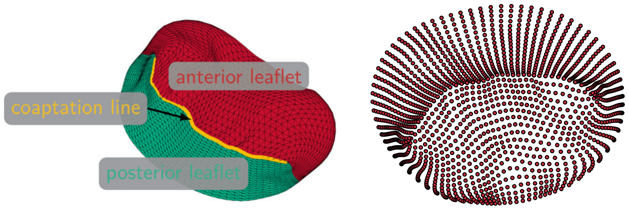


Fig. 1 Left: Anatomical regions for MV dataset. We highlight the posterior leaflet (teal color), the anterior leaflet (red color), and the coaptation line (gold). On the right, we show the discretized shape S represented by a grid $[x_1, \dots, x_{n(S)}]$ of $n(S) \in \mathbb{N}$ points in $\mathbb{R}^3$

2 Methods and Material

In the following sections, we discuss our benchmark data set of deformable shapes, our approach for computing various features in shape space, and our proposed methods for the classification of deformable soft shapes.

2.1 3D Echocardiographies of Human MVs

Our study is based on echocardiographic images acquired in vivo from 150 cardiology patients with potential MV complications. For these patients, transesophageal echocardiography provides dynamic 3D views of their MVs at rates of roughly 25 frames per heart cycle. A TOMTEC-Philips software is used to extract a 3D snapshot of the two MV leaflet surfaces per 3D frame. For frame time t, the extracted 3D snapshot displays the anterior (AL) and posterior leaflets (PL), denoted by $AL(t)$ and $PL(t)$, of the MV as smooth surfaces discretized by a dense grid of 800 points per leaflet. This database of 3 500 discretized 3D snapshots of MV leaflets was prepared and annotated by Dr. El-Tallawi.

During each heart cycle, the leaflets $AL(t)$ and $PL(t)$ close the MV at mid-systole ($t = t_{\text{midsys}}$) and open the MV at end-systole ($t = t_{\text{endsys}}$). These two leaflets define a deformable connected 3D surface $MV(t)$ bounded by a flexible ring (the "annulus"). At $t = t_{\text{midsys}}$, the leaflets $AL(t)$ and $PL(t)$ are in full contact along the *coaptation line* to tightly close the MV. The leaflets open the MV progressively until the end-systole, then remain fully open during diastole, and start closing again at the beginning of systole. When the MV is open, the coaptation line is split into two curved boundary segments $\partial AL(t)$, $\partial PL(t)$, sharing the same endpoints. We show representative patient data in Fig. 1. We note that some of the points in our mesh come with anatomical annotations (such as the coaptation line). However, there is no one-to-one correspondence between the mesh points. We need to establish pointwise correspondences through diffeomorphic shape matching.

In previous studies [18, 65, 66, 70–72], automatic *diffeomorphic registration* of the MV leaflets between $t = t_{\text{midsys}}$ and $t = t_{\text{endsys}}$ was developed and systematically implemented for this database of 3D image sequences, to compute the intensities and spatial distribution of the tissue strain induced by MV deformation at each heartbeat. This project aimed to provide cardiologists with patient-specific displays of MV leaflet strain intensities as a potential aid to evaluate and compare MV clinical cases [70–72]. Our current version of the solver performs these diffeomorphic registrations in less than 2 min per patient on a standard laptop, for pairs of surfaces discretized by 1 600 points each [66]. In the present paper, we develop diffeomorphic deformation techniques for the automatic classification of soft smooth shapes. The database of 3D MV snapshots serves as a benchmark to implement and test our approach.

2.2 Diffeomorphic Registration of 3D Surfaces

Below, we describe our approach for diffeomorphic shape matching. We refer to our past work for additional details [18, 65, 66, 118].

2.2.1 Diffeomorphic Deformations in $\mathbb{R}^3$

Recall the basic mathematical formalization of LDDMM [4, 20, 34, 36]. Fix a scale parameter $s > 0$ and let $K_s : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be the positive definite radial kernel

$$K_s(x, y) = \exp(-\|x - y\|_2^2 / s^2) \tag{3}$$

for all $x, y \in \mathbb{R}^3$. We call any smooth map $x \mapsto w_x$ from $\mathbb{R}^3$ to $\mathbb{R}^3$ such that w_x and all its derivatives tend to 0 as $\|x\| \to \infty$ a *smooth vector field* w on $\mathbb{R}^3$. For any such w, define the norm $\|w\|$ by

$$\|w\|^2 = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} K_s(x, y) \langle w_x, w_y \rangle_{\mathbb{R}^3} \, dx \, dy. \tag{4}$$

Endowed with this norm, the vector space V of smooth vector fields becomes a Hilbert space. Call *velocity flow* any set $v = (v_t)$ of time-indexed smooth vector fields $v_t \in V$ such that $t \mapsto v_t$ is a Lipschitz continuous map from $[0, 1]$ into V. Denote $\mathcal{V}$ the Hilbert space of all velocity flows $v = (v_t)$ having finite *kinetic energy* $\text{kin}(v)$ defined by

$$\text{kin}(v) = \int_0^1 \|v_t\|^2 \, dt.$$

Call *smooth deformation* of $\mathbb{R}^3$ any time-indexed flow (F_t), $t \in [0, 1]$, of smooth diffeomorphisms F_t from $\mathbb{R}^3$ to $\mathbb{R}^3$, such that F_0 is the identity map $\text{id}_{\mathbb{R}^3} : \mathbb{R}^3 \to \mathbb{R}^3$. As shown in [21, 64], for any velocity flow $v = (v_t)$ in $\mathcal{V}$, there is a unique smooth deformation (F_t) of $\mathbb{R}^3$ solving the ODE (1). Then, $\text{kin}(v)$ will also be called the *kinetic energy* of the smooth deformation (F_t).

2.2.2 Diffeomorphic Registration of Surfaces

To compare two surfaces S and Σ in $\mathcal{S}$, one seeks a smooth deformation (F_t) having *minimal kinetic energy* among all deformations verifying $F_1(S) = \Sigma$. This requires finding a velocity flow $v = (v_t)$ in $\mathcal{V}$ and an associated smooth deformation flow (F_t) solving the variational problem

$$\text{minimize}_{v \in \mathcal{V},\ (F_t) \in \mathcal{F}} \quad \text{kin}(v) \tag{5a}$$

under the nonlinear constraints

$$\partial_t F_t = v_t(F_t) \quad \text{for almost all} \quad t \in [0, 1], \tag{5b}$$

$$F_0 = \text{id}_{\mathbb{R}^3}, \tag{5c}$$

$$F_1(S) = \Sigma. \tag{5d}$$

To numerically solve the variational problem in (5), one has to relax the rigid matching constraint $F_1(S) = \Sigma$, replacing it with $F_1(S) \approx \Sigma$ [18, 20, 21, 23, 65]. This is typically accomplished by introducing a shape matching dissimilarity between $F_1(S)$ and Σ, as described in the following section.

2.2.3 Kernel-Based Dissimilarity between Smooth Surfaces

The set $\mathcal{S}$ of compact smooth 3D surfaces with boundaries can be endowed with many natural shape matching dissimilarities [23, 65, 120–122]. For a fast numerical solution of the variational problem (5), efficient differentiable shape dissimilarities are provided, as we now outline, via the self-reproducing Hilbert space (RKHS) associated with the radial Gaussian kernel $Q : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, $Q(x, y) := K_\tau(x, y)$, defined in (3) with any fixed scale parameter $\tau > 0$. The space of bounded Radon measures μ on $\mathbb{R}^3$ is a Hilbert space $\mathcal{H}$ for the norm $\|\mu\|$ defined by

$$\|\mu\|^2 = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} Q(x, y)\, \mathrm{d}\mu(x)\, \mathrm{d}\mu(y). \tag{6}$$

The Lebesgue measure of $\mathbb{R}^3$ induces on each surface $S \in \mathcal{S}$ a Riemannian surface element $\mathrm{d}\mu_S(z)$, which determines a bounded Radon measure $\mu_S \in \mathcal{H}$ with support equal to S. We rescale μ_S by imposing $\mu_S(S) = 1$. Define the Hilbertian *shape matching dissimilarity* $\text{HILB}(S, \Sigma)$ between any two surfaces $S, \Sigma \in \mathcal{S}$ by

$$\text{HILB}(S, \Sigma) = \|\mu_S - \mu_\Sigma\|^2.$$

When S, Σ are discretized by two finite grids of points $x_n \in S$, $1 \le n \le N$, and $y_m \in \Sigma$, $1 \le m \le M$, one approximates μ_S and μ_Σ by sums of Dirac masses

$$\nu_S = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_n} \qquad \text{and} \qquad \nu_\Sigma = \frac{1}{M} \sum_{m=1}^{M} \delta_{y_m},$$

respectively. This approximates $\text{HILB}(S, \Sigma)$ by $\|\nu_S - \nu_\Sigma\|^2$, which is a simple linear combination of all terms $Q(x_n, x_{n'})$, $Q(y_m, y_{m'})$, $Q(x_n, y_m)$ for $1 \le n, n' \le N$, $1 \le m, m' \le M$ [65].

We select this distance for several reasons: *(i)* most importantly, it is a natural distance measure for shapes parameterized by point clouds with a rich mathematical structure (as illustrated above) and *(ii)* it exhibits nice smoothness properties that make it suitable for gradient-based optimization algorithms. However, we also note that our framework may benefit from considering other measures as data attachment terms, such as currents [120] or varifolds [30, 121, 122]. Testing these types of data attachment models for the considered classification problem remains subject to future work.

2.2.4 Variational Problem Formulation

To soften the rigid matching constraint $F_1(S) = \Sigma$ in (5), fix a positive weight λ. Then, seek a velocity flow $v = (v_t)$, $v_t \in \mathcal{V}$, and an associated diffeomorphic flow (F_t), $F_t \in \mathcal{V}$, which solves the relaxed variational problem

$$\text{minimize}_{v \in \mathcal{V},\ (F_t) \in \mathcal{F}} \quad \text{kin}(v) + \lambda\, \text{HILB}(F_1(S), \Sigma) \tag{7a}$$

$$\text{subject to} \quad \frac{\mathrm{d}F_t}{\mathrm{d}t} = v_t(F_t) \tag{7b}$$

for almost all $t \in [0, 1]$ with initial condition $F_0 = \text{id}_{\mathbb{R}^3}$.

For fixed λ, after space-time discretization of S, Σ, and $[0, 1]$, the search for a vector field flow $v = (v_t)$ minimizing the cost functional is then implementable numerically by various gradient descent techniques [18, 20, 23, 65, 66, 73, 74]. Our numerical implementation of diffeomorphic registration for 3D surfaces is outlined further in Sect. 2.2.6, and provides a good approximation of the minimal kinetic energy $\text{KIN}(S, \Sigma)$ (which represents a short-hand notation for the evaluation of (2) for $v = (v_t)$ associated with the minimizing flow $F = (F_t)$ for (7) that maps S to Σ).

Remark 1 We note that the theoretically valid symmetry relation $\text{KIN}(S, \Sigma) = \text{KIN}(\Sigma, S)$ is only approximately true for numerical estimates. The average $(\text{KIN}(S, \Sigma) + \text{KIN}(\Sigma, S))/2$ would improve numeric accuracy at the cost of doubling computing times.

2.2.5 Strain Analysis

After computing an optimal diffeomorphic registration $F = (F_t)$ matching two surfaces $S, \Sigma \in \mathcal{S}$, the terminal $\mathbb{R}^3$-diffeomorphism $f = F_1$ is a smooth invertible map from S onto a smooth surface $S' = f(S)$, with very small Hausdorff distance $\text{HAUS}(S', \Sigma)$, where

$$\text{HAUS}(S, \Sigma) = \max_{x \in S} \min_{y \in \Sigma} \|x - y\|. \tag{8}$$

The minimal kinetic energy $\mathrm{KIN}(S, \Sigma)$ involves averages of squared velocities over the whole of S, and hence only provides a global dissimilarity between S and Σ. Strain analysis of f, which we now outline as in [18], generates the spatial distribution of local distortions between S and Σ: Fix any point $x \in S$ and let $y = f(x) \in f(S)$. Denote T_x, T_y, the tangent spaces to S, $S' = f(S)$, at x and y, respectively, endowed with local surface metrics. Since $f : S \to f(S)$ is a smooth bijection, the differential $f'(x)$ determines an invertible 2×2 linear map $J_x : T_x \to T_y$. For any tangent vector $u \in T_x$ with $\|u\| = 1$, the *directional strain* at x induced by deformation f in direction u is the length dilation (or contraction) factor $dirSTR(x, u) = |J_x u|$. Let $J_x^* : T_y \to T_x$ be the transpose of J_x, and denote $m_x \leq M_x$ the eigenvalues of the positive definite 2×2 matrix $J_x^* J_x$. The minimal and maximal directional strains around x are equal to $\sqrt{m_x}$ and $\sqrt{M_x}$, respectively. In general, one has $m_x < M_x$, and the directional strain at x depends on u. As in [18], to avoid this anisotropy, we focus on the *isotropic strain*

$$isoSTR_x = \sqrt{m_x M_x} = \sqrt{|\det(J_x)|}.$$

The measure $isoSTR_x$ has a simple geometric interpretation. For fixed $x \in S$, and any open patch $U_x \subset S$ around x, define the ratio of surface areas

$$\mathrm{rat}(U_x) = \mathrm{area}(f(U_x))/\,\mathrm{area}(U_x).$$

Then, $\mathrm{rat}(U_x)$ tends to $isoSTR_x^2$ when the diameter of U_x tends to 0. This provides the following fast numerical approximation of isotropic strain: After discretization of S by a finite grid $grid_S$ with small mesh size, and triangulation of $grid_S$, one simply sets U_x to be the union of all triangles with vertex x. Since $isoSTR_x$ is a dimensionless average length dilation (or contraction) factor around x, we convert it into an *isotropic strain intensity*

$$isi_x = |isoSTR_x - 1|.$$

Clearly, isi_x quantifies the intensity of local deformations around x by the diffeomorphic map $f = F_1$ from $f : S \to f(S) \approx \Sigma$.

As above, denote μ_S the probability distribution induced on S by the $\mathbb{R}^3$-Lebesgue measure. When $x \in S$ is selected at random—with probability distribution μ_S—the distribution of the random values isi_x is a probability $isi(S, \Sigma)$ on $\mathbb{R}^+$. For each percentile $0 < \alpha < 1$, the quantile q_α of $isi(S, \Sigma)$ can be viewed as a dissimilarity $QUANT_\alpha(S, \Sigma)$ between S and Σ. These dissimilarities are well approximated via quantiles of the finite samples isi_x, $x \in grid_S$ when $grid_S$ has a small mesh size. One can symmetrize $QUANT_\alpha$ by averaging its values for (S, Σ) and (Σ, S).

2.2.6 Diffeomorphic Registration Software

Prior versions of the diffeomorphic registration algorithms described in this section have been developed and tested in [18, 65]. These versions were implemented in MATLAB and applied to study 3D echocardiography data of MV patients [67–72]. As in our earlier papers, we implement a *discretize-then-optimize* approach for solving the optimization problem (7). The target shape Σ and the template shape S are represented by two grids of points $y_1 = \{y_1^i\}_{i=1}^n$ and $y_0 = \{y_0^i\}_{i=1}^m$ in $\mathbb{R}^3$, respectively. We model the velocity vector fields v_t as vectors belonging to an RKHS defined by a Gaussian kernel. Consequently, we write

$$v_t(z) = \sum_{i=1}^{m} K_\sigma(x_i(t), z) a_i(t) \tag{9}$$

for all $z \in \mathbb{R}^3$. Here, $K_\sigma : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ corresponds to the kernel in (3). The coefficients $\{a_i\}_{i=1}^m$, $a_i \in \mathbb{R}^3$, are then the new controls of the discrete optimization problem. We discretize the time interval [0, 1] by a nodal grid, resulting in q equispaced intervals. We collect the associated coefficients $a_i^j := a_i(t_j)$, $j = 0, \ldots, q$, in a vector (lexicographical ordering) a of size $c = 3m(q + 1)$. The ODE (i.e., the flow equation) in (7) is discretized using a first-order explicit Euler method. As for the control a, we collect the associated states (deformed shape) in a concatenated vector x of size $s = 3m(q + 1)$. With this, we can represent the forward Euler step as a linear system

$$\begin{bmatrix} G^x & G^a \end{bmatrix} \begin{bmatrix} x \\ a \end{bmatrix} = g$$

with state-control vector $(x, a) \in \mathbb{R}^{6m(q+1)}$. The moderate value $q = 4$ achieved a good compromise between numerical accuracy and computing time for all numerical diffeomorphic surface registrations considered here. Moreover, in our benchmark application, $m = n = 800$.

With slight abuse of notation, we arrive at the discrete version of (7) given by

$$\mathrm{minimize}_{a \in \mathbb{R}^c,\, x \in \mathbb{R}^s} \quad \mathrm{kin}(a) + \lambda\, \mathrm{HILB}(x^q, y_1) \tag{10a}$$

$$\text{subject to} \quad \begin{bmatrix} G^x & G^a \end{bmatrix} \begin{bmatrix} x \\ a \end{bmatrix} = g. \tag{10b}$$

We refer to [18, 66] for the precise form of the operators that appear in (10). To solve (10), we apply an operator splitting strategy typically referred to as the alternating direction method of multipliers [123–125] or Douglas–Rachford splitting [126]. In its modern form, this algorithm was introduced in [127, 128]. The proximal form of the consensus form of operator splitting for the discrete control problem in (10) at iteration k is given by [66]

$$\begin{bmatrix} x^{k+1} \\ a^{k+1} \end{bmatrix} = \operatorname{prox}_{\gamma(\operatorname{ind}_{\mathcal{C}}+\operatorname{kin})}\left(\tilde{x}^k + u^k, \tilde{a}^k + w^k\right) \tag{11a}$$

$$\begin{bmatrix} \tilde{x}^{k+1} \\ \tilde{a}^{k+1} \end{bmatrix} = \operatorname{prox}_{\gamma \operatorname{HILB}}\left(x^{k+1} - u^k, a^{k+1} - w^k\right) \tag{11b}$$

$$\begin{bmatrix} u^{k+1} \\ w^{k+1} \end{bmatrix} = \begin{bmatrix} u^k \\ w^k \end{bmatrix} + \begin{bmatrix} \tilde{x}^{k+1} \\ \tilde{a}^{k+1} \end{bmatrix} - \begin{bmatrix} x^{k+1} \\ a^{k+1} \end{bmatrix}. \tag{11c}$$

Here, $\gamma > 0$ is an algorithm parameter and $\operatorname{ind}_{\mathcal{C}}$ represents an indicator function for the set $\mathcal{C} \subseteq \mathbb{R}^{6m(q+1)}$ of state-control pairs (x, a) that satisfy the discretized dynamical system in (10), i.e.,

$$\mathcal{C} := \left\{ \begin{bmatrix} x \\ a \end{bmatrix} \mid \begin{bmatrix} G^x & G^a \end{bmatrix} \begin{bmatrix} x \\ a \end{bmatrix} = g \right\}.$$

We solve the first-order optimality conditions of subproblem (11a) for the state-control vector (x, a) using a matrix-free, preconditioned conjugate gradient method. Given the solution of (11a), we solve (11b) for the state-control vector $(\tilde{x}, \tilde{a})$ using a matrix-free Newton–Krylov method [129]. The last step in (11c) represents an update of the dual variables (u, w) associated with the consensus constraint $(x, a) = (\tilde{x}, \tilde{a})$. A more detailed discussion of this solver is beyond the scope of the present paper and is provided in [66]. For the sake of the present work, we consider it a black-box method that provides us with an efficient and fast diffeomorphic registration of 3D surfaces. We terminate this algorithm when the discretized, censored Hausdorff distance between the deformed shape x^q and the reference shape y_1 is of the order of the surface discretization mesh size (see [18, 65, 66] for details).

2.2.7 Choice of Meta-Parameters σ and τ

We explore the performance of our solver as a function of the scale parameters τ and σ in [66]. We found that our methodology is more sensitive to the scale parameter τ that enters the distance measure than the scale parameter σ for the parameterization of the diffeomorphism. Here, we propose heuristics for selecting these scale parameters based on the considered dataset. In typical applications comparing 3D images of soft organs across patients, surfaces S and Σ are discretized by grids of roughly equal sizes m. Preliminary homothetic transformations of S, Σ, to yield $\operatorname{area}(S) = \operatorname{area}(\Sigma)$ as done in our benchmark MV data set usually forces the mesh sizes h_S, h_Σ, of our two new grids to be of the same order. Let then $h = (h_S + h_\Sigma)/2$. We naturally want the kernel $K_\tau(x, z)$ defining the distance $\operatorname{HILB}(S, \Sigma)$ to remain sensitive to local matching discrepancies of the order of h, and to remain highly local. This is achieved by choosing $\tau \approx ch$ with $2.5 < c < 5$. Formula (9) for velocities $v_t(z)$ involves a priori m terms $K_\sigma(x_i(t), z)$. We want to ensure (uniformly in $z \in S$) moderate bounds of the form c'/σ for the norm of the differential $\partial_z v_t(z)$. To that end, we force (9) to only involve a small number of close neighbors $x_i(t)$ for each $z \in S$. This is achieved by choosing $\sigma = c'h$ with $2.5 < c' < 5$. Given a data set of surfaces to automatically classify, the practical choices of c, c' can be validated empirically by testing them on a small random sample of pairs S, Σ.

Remark 2 We note that a multiscale strategy (i.e., a continuation in the scale parameters) might improve the performance of our methodology, especially in cases where the initial misalignment between the datasets is significant or if the mesh points used for the parameterization of the surfaces are distributed more irregularly. The design of a policy for integrating scale continuation with parameter continuation for the distance measure (see, e.g., [13, 65]) remains subject to future work.

2.3 Shape Dissimilarities and Automatic Shape Classification

We discuss our proposed framework for the classification of deformable shapes next.

2.3.1 The Need for Intrinsic Feature Vectors

Consider any dataset $\mathcal{D} \subset \mathcal{S}$ of smooth 3D surfaces partitioned into a finite set of classes CL_j. We implement automatic classification in $\mathcal{D}$ by supervised ML. All well-known classifiers such as MLPs [61], RFs [59, 60] or SVMs [63] require the description of each surface $S \in \mathcal{D}$ by “natural” feature vectors $\operatorname{vec}(S)$ belonging to a Euclidean space of *fixed* dimension (or to some fixed RKHS). However, in practice each surface $S \in \mathcal{D}$ is available only through discretization by a finite grid $grid_S$ of points $x_n \in S$, $1 \leq n \leq N$, and the cardinality $card_S = N$ of $grid_S$ often varies with S. Naive direct indexation of S by the vector $X(S) = [x_1, x_2, \ldots, x_n, \ldots, x_N]$, is mathematically not sound since the dimension $3card_S$ of $X(S)$ may vary with S; any permutation of the points x_n would radically modify $X(S)$ while still defining the same discretized surface S. Intrinsic feature vectors $\operatorname{vec}(S)$ that describe the smooth shapes S discretized by finite grids $grid_S$ should at least remain stable under permutations of these grids, and exhibit some natural consistency when the mesh size of $grid_S$ tends to 0. Another natural requirement in many biomedical applications is the invariance of $\operatorname{vec}(S)$ when S is replaced by $\rho.S$, where ρ is any rigid motion in $\mathbb{R}^3$.

2.3.2 Families of Dissimilarities Invariant to Rigid Motions

To generate intrinsic feature vectors $\mathrm{vec}(S)$, we start by introducing multiple types of *dissimilarities* $\mathrm{dis}(S, \Sigma) = \mathrm{dis}(\Sigma, S) \geq 0$ between pairs of surfaces (S, Σ). These dissimilarities are not necessarily distances. While some of them are natural squared distances, in all our uses of dissimilarities one does not need the triangular inequalities to be verified. For the automatic classification of smooth surfaces, our generic dissimilarities are typically of the form

$$\mathrm{dis}(S, \Sigma) = u[d(S, \Sigma)], \tag{12}$$

where $u : \mathbb{R}^+ \to \mathbb{R}^+$ is any continuous increasing function verifying $u(0) = 0$, and $d(S, \Sigma) \geq 0$ is any bona fide distance between smooth surfaces S, Σ.

We have already defined four dissimilarities $\mathrm{dis}(S, \Sigma)$ between pairs of surfaces $S, \Sigma \in \mathcal{S}$, namely:

1. the Hausdorff distance $\mathrm{HAUS}(S, \Sigma)$,
2. the squared Hilbert distance $\mathrm{HILB}(S, \Sigma)$,
3. the kinetic energy $\mathrm{KIN}(S, \Sigma)$, and
4. the strain quantiles $\mathrm{QUANT}_\alpha(S, \Sigma)$.

Let SE(3) be the group of rigid motions in $\mathbb{R}^3$, generated by translations and rotations. Each one of these basic four dissimilarities $\mathrm{dis}(S, \Sigma)$ is invariant by rigid motion, i.e., verifies

$$\mathrm{dis}(S, \Sigma) = \mathrm{dis}(\rho.S, \rho.\Sigma) \tag{13}$$

for all $\rho \in \mathrm{SE}(3)$. This is an interesting property in the context of automatic classification for soft organ shapes observed across multiple patient groups since one expects the true class of a soft shape S to be invariant by all rigid motions of S. Within our benchmark dataset of discretized MV surfaces, we also wanted to mitigate the impact of patient height and weight by applying to each surface S an adequate homothetic transformation. This led us to construct strongly invariant dissimilarities as follows.

2.3.3 Strongly Invariant Dissimilarities

For any surface $S \in \mathcal{S}$, denote c_S its center of mass and ten_S its 3×3 tensor of inertia, which are given by

$$c_S = \int_{x \in S} x \, \mathrm{d}\mu_S(x)$$

and

$$ten_S = \int_{x \in S} \int_{y \in S} q(x, y) \, \mathrm{d}\mu_S(x) \, \mathrm{d}\mu_S(y),$$

where $q(x, y) = (x - c_S)(y - c_S)^*$, x, c_S are column vectors, and $*$ denotes transposition. When S is discretized by a finite grid $grid_S$, the intrinsic measure μ_S induced on S by $\mathbb{R}^3$-Lebesgue measure is simply the average of Dirac masses carried by the points of $grid_S$. The matrix ten_S has positive eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ and unit length eigenvectors η_1, η_2, η_3.

Fix an orthonormal basis e_1, e_2, e_3 in $\mathbb{R}^3$. For each $S \in \mathcal{D}$ define the $\mathbb{R}^3$-rotation rot_S and the rigid motion ρ_S by $rot_S \eta_j = e_j$ for $j = 1, 2, 3$, and $\rho_S y = rot_S(y - c_S)$ for $y \in \mathbb{R}^3$. To each surface S in our dataset $\mathcal{D}$, we will first associate the surface $S' = \rho_S.S$. Note that $c_{S'} = 0$, and that the inertia tensor $ten_{S'}$ has $\{e_1, e_2, e_3\}$ as ordered eigenvectors. Then, we transform S' into $S'' = h_S.S'$, where h_S is the homothety centered at 0 and such that $\mathrm{area}(S'') = 1$. The linear transformation $g_S = h_S \circ \rho_S$ belongs to the group $G3$ generated by the rotations, translations, and homotheties from $\mathbb{R}^3$ to $\mathbb{R}^3$, and we will systematically replace S by $S'' = g_S.S$. Let κ be either a translation, a rotation, or a homothety from $\mathbb{R}^3$ to $\mathbb{R}^3$. As is directly verified in each one of these three cases, the surface $(\kappa.S)'' = g_{\kappa.S}.(\kappa.S)$ is identical to S''. Hence, the property $(\kappa.S)'' = S''$ will also hold for all linear transformations $\kappa \in G3$.

Remark 3 The alignment based on moments described above does not work for all types of shapes. For the shapes considered in our work, we have developed a consistent strategy to order the eigenvectors $\{e_1, e_2, e_3\}$ based on anatomical features available to us. The true anatomic orientation of the eigenvectors is done by projecting on e_1 the centers of gravity of the anterior and posterior leaflets of S and projecting on e_2 the left and right commissures of S (endpoints of the coaptation line). We then impose $e_3 = e_1 \times e_2$. Let $\{u_1, u_2, u_3\}$ be a standard orthonormal basis of $\mathbb{R}^3$. There is then a unique rotation rot_S such that $\mathrm{rot}_S e_j = u_j$ for $j = 1, 2, 3$. In our MV benchmark data set, this approach was non-ambiguous because the eigenvalues of all inertia tensors ten_S are distinct, and the anatomic orientation of the first two eigenvectors of ten_S is straightforward. For generic smooth surfaces S, the eigenvalues of ten_S might be quite similar. Then, the reorientation strategy outlined above no longer directly applies. Consequently, other strategies such as the iterative closest point algorithm need to be considered for pre-alignment.

Any dissimilarity $\mathrm{dis}(S, \Sigma)$ defined for all pairs S, Σ in $\mathcal{S}$ and verifying the rigid motion invariance (13) will naturally define a standardized dissimilarity dis^+ by

$$\mathrm{dis}^+(S, \Sigma) = \mathrm{dis}(g_S.S, g_\Sigma.\Sigma) = \mathrm{dis}(S'', \Sigma''). \tag{14}$$

Thanks to (13), the dissimilarity $\mathrm{dis}^+(S, \Sigma)$ does not depend on the choice of orthonormal basis e_1, e_2, e_3. Moreover, for any pair of transformations $\kappa_1, \kappa_2 \in G3$, we have—as seen above—$(\kappa_1.S)'' = S''$ and $(\kappa_2.\Sigma)'' = \Sigma''$, so that (14) implies

$$\mathrm{dis}^{+}(S, \Sigma) = \mathrm{dis}^{+}(\kappa_1.S, \kappa_2.\Sigma)$$

for all $\kappa_1, \kappa_2 \in G3$.

To each one of the dissimilarities dis listed above as items 1 through 4, we apply the preceding construction to generate a corresponding strongly invariant dissimilarity dis^{+}. In what follows, we will adopt the following *simplified notations*:

$$\mathrm{haus}(S, \Sigma) = \mathrm{HAUS}^{+}(S, \Sigma), \tag{15a}$$

$$\mathrm{hilb}(S, \Sigma) = \mathrm{HILB}^{+}(S, \Sigma), \tag{15b}$$

$$\mathrm{kin}(S, \Sigma) = \mathrm{KIN}^{+}(S, \Sigma), \tag{15c}$$

$$\mathrm{quant}_{\alpha}(S, \Sigma) = \mathrm{QUANT}_{\alpha}^{+}(S, \Sigma). \tag{15d}$$

2.3.4 Construction of Intrinsic Feature Vectors for Shape Classification

Let $\mathcal{D}$ be a benchmark set of smooth surfaces partitioned into several disjoint classes CL_j of surfaces. To define intrinsic feature vectors describing these surfaces, we select (and fix) in $\mathcal{D}$ a finite *reference set* $REF = \{\Sigma_1, \ldots, \Sigma_r\}$ of $r \in \mathbb{N}$ surfaces. In our applications below, REF will simply be either the set of all training cases or a large subset of the set of "normal" cases.

Select any one of the strongly invariant dissimilarities $\mathrm{dis}(S, \Sigma)$ listed as items 1 through 4 above. Then, characterize any surface $S \in \mathcal{D}$ by the r-dimensional *feature vector* $\mathrm{vec}(S)$ having coordinates

$$\mathrm{vec}(S)_k = \mathrm{dis}(S, \Sigma_k), \quad k = 1, \ldots, r.$$

These feature vectors verify the $G3$ invariance $\mathrm{vec}(S) = \mathrm{vec}(\kappa.S)$ for all linear transformations $\kappa \in G3$. We extend this approach by selecting $p \in \mathbb{N}$ distinct dissimilarities $\mathrm{dis}_1, \ldots, \mathrm{dis}_p$ among all the dissimilarities listed above as items 1 through 4. Each associated dis_i defines—as above—feature vectors $\mathrm{vec}^i(S) \in \mathbb{R}^r$, which can naturally be concatenated into the pr-dimensional feature vector

$$\mathrm{VEC}(S) = \left[\mathrm{vec}^1(S), \ldots, \mathrm{vec}^p(S)\right] \in \mathbb{R}^{pr}.$$

This feature vector verifies the desirable properties outlined above for intrinsic feature vectors. To test and validate our approach, we have applied it to the automatic classification of MVs by RF classifiers (see Sect. 3).

2.4 Data Enrichment

Since our benchmark dataset included unbalanced classes, we implemented new diffeomorphic techniques for rebalancing small classes. This is outlined next.

Remark 4 When we enrich one class of cases, we control the individual surface perturbations to make sure that the perturbed surface remains sufficiently far from the surfaces belonging to the other class.

2.4.1 Enrichment by Diffeomorphic Interpolations

In automatic classification via ML by well-known classifiers such as MLPs [61], CNNs, RFs [59, 60], or SVMs [63], the strong imbalance between class sizes tends to degrade classification accuracy, specifically among the smallest classes. Since our benchmark dataset of MV surfaces was derived from unbalanced classes of patients, we have implemented several diffeomorphic deformation algorithms for the rebalancing of small classes. In most applications of automatic classifiers, all cases are described by feature vectors belonging to a fixed Euclidean space; enriching any small class CL of cases is often implemented via the SMOTE algorithm (see, e.g., [119]), which linearly interpolates between neighboring feature vectors of CL-cases. For datasets of discretized smooth 3D shapes, the SMOTE algorithm is not directly applicable because for intrinsic feature vectors (such as those constructed above) convex combinations of feature vectors produce vectors that are not necessarily associated with any smooth surface. So, to enrich any given finite class $CL \subset \mathcal{S}$ of smooth 3D surfaces, we proceed by *nonlinear diffeomorphic interpolation* between pairs of surfaces $S, \Sigma \in CL$ such that the associated Hausdorff distance is sufficiently small.

With the notations of Sect. 2.3.3, replace S and Σ by $S'' = g_S.S$ and $\Sigma'' = g_\Sigma.\Sigma$, where g_S and g_Σ are in the group $G3$. Compute an optimized diffeomorphic deformation flow $F = (F_t)$ such that $F_1(S'') \approx \Sigma''$. In practical applications below, time is discretized and we select an intermediary value $t \in [0, 1]$ between $1/2$ and $3/4$ before adding the new smooth surface $S_t = F_t(S'')$ to the class CL. See Fig. 2.

2.4.2 Enrichment by Random Diffeomorphic Deformations

Fix any given finite class CL of smooth 3D shapes. To enrich CL, we have also implemented small random diffeomorphic perturbations of the surfaces in CL. Our algorithms rely on simulating smooth random Gaussian vector fields indexed by time t and all points $x \in \mathbb{R}^3$, before integrating them in time.

Simulation of Smooth Gaussian Random Vector Fields Fix any integer $d > 0$, which in this paper will only take values $d \in \{1, 2, 3\}$. A random $\mathbb{R}^d$-valued vector field W, $x \mapsto W(x)$, indexed by all $x \in \mathbb{R}^3$ is for-

Fig. 2 Orange surface $S_{3/4}$ is created by diffeomorphic interpolation between the blue surface $S_0 = S$ and the yellow surface $S_1 = \Sigma$

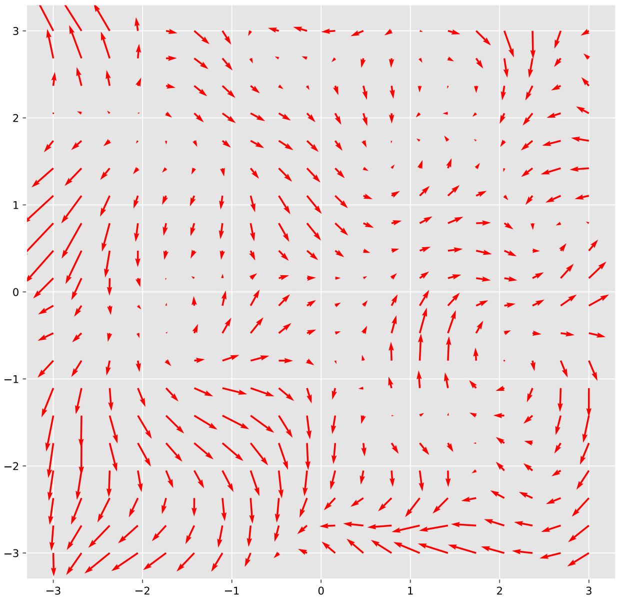


Fig. 3 An example of simulated two-dimensional smooth Gaussian random field $W(x) = [W^1(x), W^2(x)]$. For better visualization, all vectors $W(x)$ are normalized to have the same length

mally a set of random vectors $W(x) \in \mathbb{R}^d$ defined on the same probability space (Ω, P) such that $W(x, \omega)$ is a jointly measurable function of $(x, \omega) \in \mathbb{R}^3 \times \Omega$. Such a random vector field W is called *Gaussian* when for any finite set of points $\{x(1), \ldots, x(m)\}$ in $\mathbb{R}^{3m}$, the random vector $(W(x(1)), \ldots, W(x(m))$—which belongs to $\mathbb{R}^{dm}$—has a Gaussian distribution. Since any multi-dimensional Gaussian is characterized by its mean and its covariance matrix, the probability distribution of a Gaussian random vector field is fully determined by two deterministic functions, namely $x \mapsto u(x)$, $u(x) = \mathrm{E}[W(x)] \in \mathbb{R}^d$ and $(x, y) \mapsto \ker(x, y) = \mathrm{Cov}(W(x), W(y))$, where each $d \times d$ matrix $\ker(x, y)$ is positive semi-definite.

For easier simulations of Gaussian random vector fields $W : \mathbb{R}^3 \to \mathbb{R}^3$, we focus only on the case where $\mathrm{E}[W(x)] = 0$ for all x and the 3×3 covariance kernel $\ker(x, y)$ is a *diagonal matrix*:

$$\ker(x, y) = \mathrm{diag}[\kappa_1(x, y), \kappa_2(x, y), \kappa_3(x, y)] \tag{16}$$

for all $x, y \in \mathbb{R}^3$. For each $j = 1, 2, 3$, we fix a scale parameter $s_j > 0$ and we define the radial kernel $\kappa_j(x, y)$ for all $x, y \in \mathbb{R}^3$ by

$$\kappa_j(x, y) = K_\sigma(x/s_j, y/s_j), \tag{17}$$

with $K_\sigma : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ as in (3) and $\sigma = 1$. In "Appendix A.1," we outline our algorithm to numerically simulate a one-dimensional Gaussian random field $x \mapsto U(x) \in \mathbb{R}$ with mean 0 and covariance kernel $k(x, y) = \exp(-\|x - y\|^2)$, indexed by all $x \in \mathbb{R}^3$. Moreover, $x \mapsto U(x)$ is an almost surely smooth function of x. Our simulation algorithm involves the numerical summation over all triplets of non-negative integers (m, n, p) of the almost surely converging explicit series

$$U(x) = \sum_{m,n,p} Z_{m,n,p} u_{m,n,p}(x), \tag{18}$$

where each $u_{m,n,p}(x)$ is an explicit deterministic smooth function of $x \in \mathbb{R}^3$, and the $Z_{m,n,p}$ are independent Gaussian random variables having the same mean 0 and standard deviation 1. This numerical summation can be replicated 3 times (with new $Z_{m,n,p}$ each time), to simulate 3 *independent versions* $U^1(x)$, $U^2(x)$, $U^3(x)$ of the smooth Gaussian random field $U(x)$. We then rescale $U^j(x)$ by setting $W^j(x) = U^j(x/s_j)$ for all $x \in \mathbb{R}^3$ and each $j = 1, 2, 3$. Then, for $x \in \mathbb{R}^3$, we define the 3D smooth Gaussian random field $W(x) = [W^1(x), W^2(x), W^3(x)]$. Moreover, $W(x)$ has mean 0 and diagonal covariance kernel $\ker(x, y)$ given by equations (16) and (17). In Fig. 3, we display an example of smooth 2D Gaussian random field $[W^1(x), W^2(x)]$ simulated by numerical summation of the series (18).

Small Random Diffeomorphic Perturbations of Smooth Surfaces Let $W(x) \in \mathbb{R}^3$ be the just described smooth Gaussian random vector field indexed by $x \in \mathbb{R}^3$. Let L be any fixed 3×3 matrix. Then, for $x \in \mathbb{R}^3$, the affine function $x \mapsto Lx$ defines a deterministic vector field, and

$$V_t(x) = tLx + \sqrt{t}\,W(x) \quad \text{for all} \quad t \geq 0, x \in \mathbb{R}^3$$

defines a time-indexed flow $V_t(x)$ of smooth Gaussian random vector fields with deterministic mean vector field

$\mathrm{E}[V_t(x)] = tLx$ and diagonal covariance kernel $\ker(x, y)$ given by (16) and (17). For t small, the small perturbating random velocity fields V_t can be viewed as an infinite dimensional analog of a multivariate Brownian motion $w_t \in \mathbb{R}^k$ with an added drift $tBx \in \mathbb{R}^k$ linear in time and space. Since β_t and $\sqrt{t}\beta_1$ have the same Gaussian distribution, we naturally model the "Brownian" part of our small perturbating velocity fields by $\sqrt{t}W(x)$.

We have outlined above how to simulate $x \mapsto W(x)$, which then directly provides the values of $V_t(x)$ for all (t, x). We then numerically generate a stochastic flow F_t of random $\mathbb{R}^3$-diffeomorphisms by *pathwise* discrete integration in t of the stochastic ODE $\mathrm{d}F_t(x)/\mathrm{d}t = V_t(F_t(x))$, with $F_0 = \mathrm{id}_{\mathbb{R}^3}$. Time is discretized by fixing a moderate number of instants $t_j = j\delta$, $j = 0, 1, 2, \ldots$, with a small time step size $\delta > 0$. Since the random vector fields $V_t(x)$ are almost surely smooth in (t, x), discretized integration in time and space is mathematically stable when the mesh size in space and time tends to 0. To randomly perturb a discretized surface $S \in CL$, numerical ODE integration is done separately for each initial grid point $x \in S$. This will generate the points $y_j = F_{t_j}(x)$. One can then enrich the set CL by adding the (discretized) smooth surface $F_t(S)$ to CL as a virtual case, after checking that $F_t(S)$ is still close enough to S. Indeed, for bona fide enrichment of a given class of surfaces, one needs to stop the random diffeomorphic deformations $F_t(S)$ at moderate values of time t, as displayed for instance in Fig. 4.

In our application to the enrichment of small classes of discretized MV surfaces, we control random perturbations via simple geometric criteria such as triangulation homogeneity and the trimmed Hausdorff distance.

3 Setup, Experiments and Results

Next, we report the results of applying our shape classification methodology to our benchmark MV dataset.

3.1 Automatic Classification of MVs: Regurgitation vs. Normal

We have developed a methodology for the automatic classification of smooth 3D surfaces based on diffeomorphic registration of surfaces. We have implemented this automatic classification methodology for a subset of our MV surfaces, namely the union of two disjoint classes of MVs—normal cases vs. regurgitation cases.

3.1.1 Benchmark Classification Task

Our original dataset involves 3D views of human MVs and is acquired by echocardiography for 150 patients, with roughly 25 3D views per patient spanning one heart cycle. Each MV 3D view is discretized by a grid of 1 600 points in $\mathbb{R}^3$, with 800 points per MV leaflet. *MV regurgitation* occurs when at mid-systole the two MV leaflets AL and PL do not close properly around the coaptation line (instead of tightly closing as in normal patients). At mid-systole, this incomplete MV closure leaves a narrow gap between AL and PL, inducing a blood flow leak in the wrong direction during systole, which weakens the normal blood flow through arteries. MV regurgitation is not rare after age 60, and severe regurgitation requires surgical MV repair.

Our benchmark application is ML classification of regurgitation versus normal cases based on the diffeomorphic matching techniques introduced above. In 3D echocardiography, MV regurgitation is best visible at mid-systole, so we kept only the MV views acquired at times closest to mid-systole, namely four views per regurgitation case and two views per normal case. This defines an initial benchmark dataset $\mathcal{D}$ of 3D discretized MV surfaces, partitioned into 120 regurgitation cases and 200 normal cases.

We note that—strictly speaking—the datasets that display regurgitation have a different topology than those of healthy patients (due to the open gap at mid-systole). It is hard to guarantee anatomical integrity of the computed map $(F_t)_{t\geq 0}$ from a subject that has regurgitation to one that does not without including additional prior knowledge. That is, while our map is guaranteed to be a diffeomorphism (i.e., a smooth bijection that has a smooth inverse) from one point cloud to the other, we *cannot* guarantee that points located on one leaflet are mapped to points located on another leaflet if they are in close vicinity to one another (see Fig. 1 for an illustration). That is, points located on the AL could be mapped to points on the PL of another patient. One way of addressing this issue is to assign higher weights for points located on the boundary of the MV. The boundaries of the individual leaflets are well-identified even when the MV is closed. We have the anatomical annotations to do so, and our software supports this. Another approach (that our software supports) is to register the leaflets individually. This way, topology does not change. Lastly, we note that developing approaches that can handle topology changes is an active area of research; we refer to the works in [130–138] for examples.

3.1.2 Enrichment of Benchmark Data Set

Recall that the strongly invariant dissimilarity $\mathrm{haus}(S, \Sigma)$ is defined in 15a. As outlined in Sect. 2.4.1, we enrich the initial class of 120 regurgitation cases by diffeomorphic interpolation between pairs S, Σ of regurgitation 3D views having small dissimilarity $\mathrm{haus}(S, \Sigma)$. Our shape interpolation technique is thus applied to 80 pairs S, Σ of regurgitation MV surfaces, and generates 80 new virtual regurgitation cases. After this first enrichment of our benchmark dataset, the two

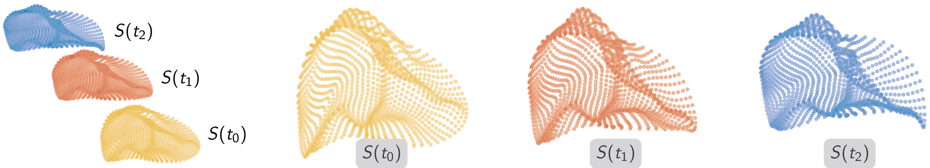


Fig. 4 Enrichment of shape classes through random diffeomorphic deformations $F_t(S)$. We display the initial surface $S(t_0)$ along with deformed surfaces $S(t_1)$ and $S(t_2)$ at times $0 \leq t_0 < t_1 < t_2$ in an overlay view on the left. The surfaces $S(t_1)$ and $S(t_2)$ are generated by applying a time-indexed flow of random diffeomorphic deformations to the initial surface $S(t_0)$. Enlarged views of the individual surfaces $S(t_0)$, $S(t_1)$, and $S(t_2)$ are shown to the right. The applied deformations remain moderate as long as $t_2 - t_0$ (and by that $t_1 - t_0$) is small

classes "regurgitation" and "normal" now have the same size of 200.

As outlined in Sect. 2.4.2, we have then implemented one small random diffeomorphic deformation for each one of these 400 smooth surfaces, to generate 200 new virtual "regurgitation" cases and 200 new virtual "normal" cases. After this second enrichment, our new benchmark dataset—still denoted $\mathcal{D}$ for simplicity—involves now a total of 800 smooth surfaces, namely 400 regurgitation cases and 400 normal cases.

3.1.3 3D Image Cropping of MV Surfaces

For expert cardiologists inspecting sequences of live echocardiography data in 3D, diagnosis of MV regurgitation includes visually checking if around mid-systole small gaps emerge along the coaptation line. We hence deliberately focus our MV surfaces analysis on an MV area close to the coaptation line, and we restrict our diffeomorphic shape registration techniques to *cropped 3D snapshots* of MV surfaces. Our image cropping keeps only the central half of each original discretized MV surface snapshot S, namely the 800 grid points of S which are roughly closest to its coaptation line. Cropping MV surfaces is primarily done to reduce the computational time. Indeed, our numerical implementation of diffeomorphic registration [65, 66] has a computational complexity dominated by the evaluation of the $n \times n$ kernel matrices K_σ, K_τ, so that keeping only 800 points reduces the computational time by a factor of four. We display a cropped MV surface in Fig. 5. After enrichment and image cropping, our final benchmark dataset, which we still denote $\mathcal{D}$, contains a total of 800 cropped smooth MV surfaces, namely 400 regurgitation cases and 400 normal cases.

3.1.4 Statistical Analysis of Kinetic Energy and Strain Quantiles

Let $G3$ be as above the group of linear transformations of $\mathbb{R}^3$ generated by rotations, translations, and homotheties. For our enriched benchmark dataset $\mathcal{D}$ of 800 cropped MV surfaces, we have studied the histograms of dissimilarity values taken by the following five strongly invariant dissimilarities:

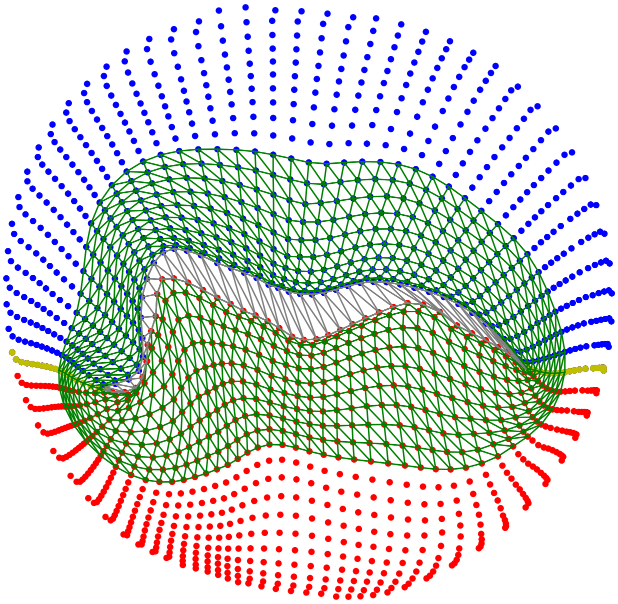


Fig. 5 Cropped MV surface snapshot for a severe MV regurgitation case: Only the 800 points closest to the coaptation line have been kept. We show in blue and red the points located on each leaflet. The subset of points kept that are located closest to the coaptation line are highlighted based on a Delaunay triangulation (in green). The opening of the MV due to regurgitation is highlighted in gray

1. the kinetic energy $\text{kin}(S, \Sigma)$.
2. the four strain intensity quantiles

$$\text{quant}_\alpha(S, \Sigma), \quad \alpha \in \{0.05, 0.50, 0.95, 0.99\}.$$

The construction of these dissimilarities was outlined in Sects. 2.3.2 and 2.3.3. In particular, for all $\kappa_1, \kappa_2 \in G3$, these dissimilarities remain unchanged when S, Σ are replaced by $\kappa_1.S, \kappa_2.\Sigma$. For fixed S, Σ, the actual computation of these five dissimilarities requires the numerical diffeomorphic registration of the two cropped MV surfaces $S'' = \rho_S.S$ and $\Sigma'' = \rho_\Sigma.\Sigma$ geometrically derived from S, Σ by specific $\rho_S, \rho_\Sigma \in G3$, as indicated in Sect. 2.3.3.

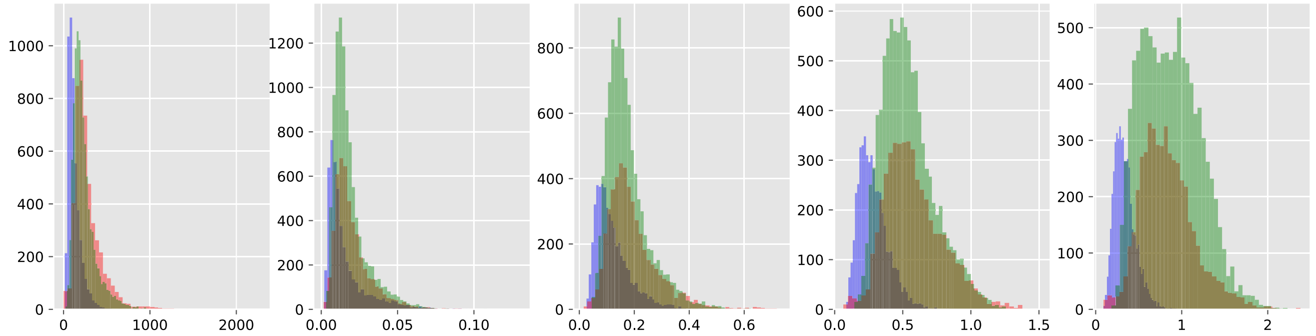


Fig. 6 Blue, red, and green bars correspond to the three sets Nor/Nor, Reg/Reg, Nor/Reg, respectively, of 10^4 pairs (S, Σ) in $\mathcal{D}$. The histograms of our five strongly invariant dissimilarities $\{\text{kin}, \text{quant}_{0.05}, \text{quant}_{0.50}, \text{quant}_{0.95}, \text{quant}_{0.99}\}$ are shown from left to right. The y-axis represents the number of pairs

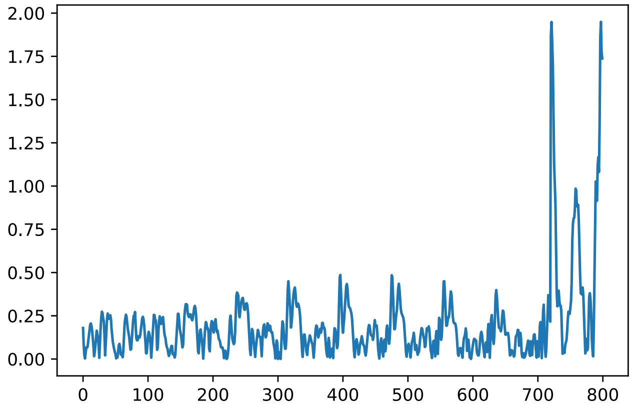


Fig. 7 An example of strain vector $isi(S, \Sigma)$ of dimension 800, where the cropped surfaces S and Σ are a regurgitation case and a normal case, respectively. The last 100 strain intensities correspond to the 100 points of S closest to the coaptation line and tend to have higher values

To study empirically the distributions of these five dissimilarities, we have picked in $\mathcal{D}$ a random subset of 100 normal cases and 100 regurgitation cases and implemented diffeomorphic registrations for three sets of 10^4 pairs (S, Σ), namely the three sets Nor/Reg, Reg/Reg, and Nor/Nor, corresponding to {Normal vs. Regurgitation}, {Regurgitation vs. Regurgitation}, and {Normal vs. Normal}, respectively. Within each one of these three sets, we have separately computed the histograms of our five dissimilarities

$$\{\text{kin}, \text{quant}_{0.05}, \text{quant}_{0.50}, \text{quant}_{0.95}, \text{quant}_{0.99}\}.$$

See Fig. 6, where we display five figures in each of the three histograms. Here, the blue, red, and green bins correspond to histograms computed within the sets Nor/Nor, Reg/Reg, and Nor/Reg, respectively.

For the pairs of Nor/Nor cases, the blue histograms show that our five dissimilarities are mostly concentrated around small values, indicating a rather tight grouping of normal cases in "dissimilarity space." For the pairs of Reg/Reg cases, the red histograms exhibit roughly larger values of our five dissimilarities, revealing a looser grouping of regurgitation cases in "dissimilarity space." But for the pairs of Nor/Reg cases, the green histograms show that all five dissimilarities exhibit fairly high values, and hence indicate a potentially good separability in "dissimilarity space." Moreover, the green histograms of quantiles dissimilarities quant_α observed for pairs of Nor/Reg cases exhibit a marked increase in values when the percentile α increases from 0.05 to 0.99. This points to a higher discriminating power of quant_α for higher percentiles α.

3.1.5 Selection of Dissimilarities to Improve Discrimination

When S is a regurgitation case and Σ is a normal case, as shown in Fig. 7, the presence of gaps along the coaptation line of Σ forces the strain intensities $isi_{x(j)}$ when the points $x(j)$ become closer to the coaptation line of S. This remark led us to focus on eight quantile dissimilarities defined by

$$medstrain_k(S, \Sigma) = q_{0.50}$$
$$highstrain_k(S, \Sigma) = q_{0.95},$$

where $q_{0.50}$ denotes the 50% quantile and $q_{0.95}$ the 95% quantile of $isi_{1:k}(S, \Sigma)$ for $k = 80, 160, 240, 800$, respectively.

Our choice of 50% and 95% quantiles was motivated by the histogram analysis outlined in Sect. 3.1.4. For better interpretability, we have also replaced the kinetic energy dissimilarity $\text{kin}(S, \Sigma)$ by its square root $\text{sqrtkin}(S, \Sigma) = \sqrt{\text{kin}(S, \Sigma)}$, which is a bona fide distance between smooth surfaces.

To decide which distances from (15) yield the most descriptive features, we have first compared RF classification accuracy when features are generated by three different groups of distances to the reference set of 100 surfaces: (i) Hausdorff & Hilbert distances yielding 200 features per case (ii) kinetic energy & strain quantiles yielding 900 features per case (iii) all eleven distances yielding 1 100 features per case. We provide more details about this performance analysis for a reduced set of features in Sect. 3.2. We report the associated OOB accuracy (see also Sect. 3.2.2) in Table 1.

Table 1 OOB accuracy for different choices of groups of dissimilarities to the reference set of 100 surfaces

Set of distances	OOB accuracy (%)
All distances	97.5
Kinetic energy and strain quantiles	97.5
Hausdorff and Hilbert distance	96.1

We can observe that the maximum OOB accuracy = 97.5% is already reached when using only kinetic energy and strain quantiles to generate features, and does not increase if we add the Hausdorff and Hilbert distances. We also note that jointly, Hausdorff and Hilbert distances already enable a high (but not maximal) classification accuracy, an interesting point since these two distances have low computing costs.

Classically, the importance of the group of 100 features derived from a single distance is defined as the OOB accuracy degradation generated when all values of these 100 features are randomly permuted. When comparing importances within a finite set of eleven distances, the eleven importances are normalized to add up to 100%. We have computed the relative importance of our eleven distances for RF classification; for Hilbert and Hausdorff distances, this yielded equal relative importance, which was inferior to the relative importance of each one of the five strain quantiles. We report additional results for a reduced set of features in Sect. 3.2.4.

Overall, from now on we consider only the nine strongly invariant dissimilarities

$$\begin{aligned} &D_1 = highstrain_{80}, \quad D_2 = highstrain_{160}, \\ &D_3 = highstrain_{240}, \quad D_4 = highstrain_{800}, \\ &D_5 = medstrain_{80}, \quad D_6 = medstrain_{160}, \\ &D_7 = medstrain_{240}, \quad D_8 = medstrain_{800}, \\ &D_9 = sqrtkin. \end{aligned} \tag{19}$$

3.1.6 Intrinsic Feature Vectors for Automatic Classification of MV Surfaces

To implement automatic classification of "regurgitation" versus "normal" MV surfaces within the enriched dataset $\mathcal{D}$ of 800 cropped MV surfaces, we first construct intrinsic feature vectors based on strongly invariant dissimilarities. To this end, we apply the generic approach outlined in section Sect. 2.3.4. Concretely, this involves two key steps: 1. Select and fix a *reference set* $REF = \{\Sigma_1, \ldots, \Sigma_r\}$ of r MV surfaces. 2. Fix the set of nine strongly invariant dissimilarities $DIS = [D_1, D_2, \ldots, D_9]$ listed in (19).

Each cropped MV surface S in $\mathcal{D}$ will be described by the group $\text{vec}(S)$ of $9 \times r$ *intrinsic features* defined by $\text{vec}(S)_{i,j} = D_i(S, \Sigma_j)$, where $i = 1, \ldots, 9$, and $j = 1, \ldots, r$. Our histogram analysis comparing several strongly invariant dissimilarities indicates that the dissimilarities $D_i(S, \Sigma)$ tend to be higher when the two MV surfaces (S, Σ) are in different classes as compared to when (S, Σ) are in the same class. This qualitative result indicates that for each S in $\mathcal{D}$, the dissimilarities between S and all the normal MV surfaces should play a key part in classifying S correctly by positively contributing to discrimination between "regurgitation" and "normal." This led us to select a reference set REF of $r = 100$ normal MV surfaces randomly extracted from our set of all 400 cropped normal MV surfaces. The group of intrinsic feature vectors $\text{vec}(S)$ then involves 900 features to describe each surface S. We have explored other choices for the reference set REF, as indicated further on. We now present our choice of ML classifiers.

3.2 RF Classification

Among ML classifiers, RFs have been applied widely with quite convincing performance. Generated by simultaneous training of large sets of decision trees, RFs were introduced by [139] and popularized by [60], for instance, as well as by the fast emergence of efficient RF software. In RF training, each decision tree is trained on a randomly selected training set, and each entropy-optimizing split of a tree node is based on a set of features randomly selected for each node. After training, the RF classifier combines the class predictions generated by each tree, usually by majority voting. For our dataset of 800 cropped MV surfaces, with each surface S described by a group $\text{vec}(S)$ of 900 features as just outlined, we have automatically trained distinct RF classifiers using the open source `rfpimp` software package [140, 141], which offers flexible tools dedicated to evaluating the importance of any given subgroup of features by randomly scrambling their values.

3.2.1 Meta-Parameters of RF Classifiers

Several well-known meta-parameters have to be specified for RF training. After empirical exploration of potential choices, we have selected and fixed the following RF meta-parameters:

- *Number of Trees* The number of trees is set to 300. For each tree training, the random training set has size $\frac{2}{3}800$. Class weights are used to compensate for the imbalance of the random training set.
- *Node Splitting* The node splitting is based on $30 = \sqrt{900}$ randomly selected features per node. Node impurity is quantified by its Gini index, which is given by $\alpha(1 - \alpha)$, where α denotes the frequency of regurgitation cases

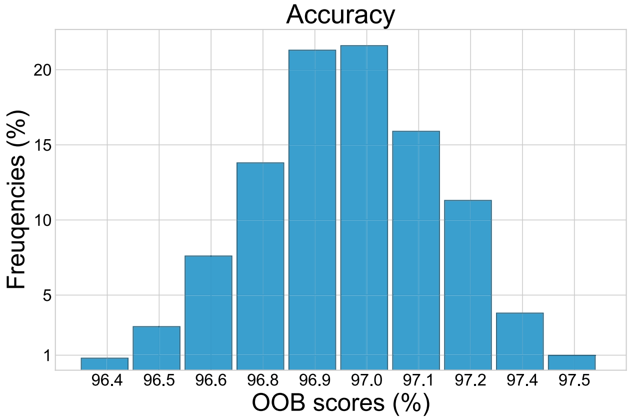


Fig. 8 For our benchmark dataset of 800 MV surfaces, we study the automatic classification of "regurgitation" MV surfaces versus "normal" MV surfaces. We have derived 900 $G3$-invariant feature vectors computed via diffeomorphic registration between pairs of surfaces. We have separately trained 1000 distinct RF classifiers. The histogram of their OOB accuracy is displayed here, with high OOB accuracy ranging from 96.1% to 97.6%

per node. Node splitting is accepted only if the splitting decreases node impurity by at least 0.002.

3.2.2 RF Performance Evaluation by OOB Accuracy

With the preceding specifications, for each fixed case S in $\mathcal{D}$, there is a random set $B(S)$ of trees whose training set does not include S. Here, the average size of $B(S)$ is roughly 37.4% of 300, i.e., 112 trees. After training, each tree $TR_i \in B(S)$ computes its own prediction $pred_i(S)$ for the true class $trueC(S)$ of S; the out-of-bag (OOB) prediction for $trueC(S)$ is then the class $\hat{C}(S)$, which occurs most often among all the $pred_i(S)$. The OOB accuracy of the RF classifier is the frequency of correct answers $\hat{C}(S) = trueC(S)$ over all cases $S \in \mathcal{S}$. OOB accuracy is known to be a fairly robust estimator for the generalization capacity of the trained RF classifier (see [142]).

3.2.3 OOB Accuracy Results for Benchmark Dataset

For our dataset of 800 cropped MV surfaces, equally split between regurgitation and normal cases, each case is described by the 900 intrinsic features described above. Training a single RF classifier having 300 trees and computing its OOB accuracy is quite fast. Since RF training is highly stochastic, we have repeated this operation 1000 times, which provided 1000 *distinct* RF classifiers. The OOB accuracies of these 1000 RF classifiers range between 96.1% and 97.6%. We display the histogram of OOB accuracies for these 1000 RF classifiers in Fig. 8.

The observed generality of RF classifiers with quite high OOB accuracy indicates that our intrinsic features based on strongly invariant dissimilarities between surfaces are quite efficient. The best of our 1000 RF classifiers, which we now denote RF^*, has a global OOB accuracy of 97.6%. The 2×2 confusion matrix $CONF$ of RF^* is given by

$$CONF = \begin{bmatrix} 97.25\% & 2.75\% \\ 2.25\% & 97.75\% \end{bmatrix},$$

where $CONF_{1,1} = 97.25\%$ and $CONF_{2,2} = 97.75\%$ corresponds to the OOB accuracy of correctly classified cases, among regurgitation and normal cases, respectively. As is well known, one can further improve classification accuracy by for instance "bagging" the five best RF classifiers via a standard majority vote. But we have preferred to study more precisely the geometric dissimilarities that have the strongest impact on RF^*.

Table 2 Influence of data enrichment (by diffeomorphic perturbations and/or interpolations) on classification accuracy

Training sets	OOB
200 original	96.5
200 original + 272 pert. cases	98.7
200 original + 328 interp. cases	97.3
200 original + 328 interp. + 272 pert. cases	97.6

We report the OOB accuracy obtained for four different training sets generated by enriching 200 original cases with 272 small random diffeomorphic perturbations (pert. cases) and/or 328 diffeomorphic interpolations (interp. cases). The table quantifies the beneficial effect of these types of diffeomorphic enrichment

To evaluate the influence of data enrichment (by diffeomorphic perturbations and/or interpolations) on the classification accuracy, we trained RF classifiers using only 200 original surfaces. After training on this reduced data set, the best RF classifier achieved only a 96.5% OOB score, which is lower than the 97.6% OOB reached by the best RF^* trained on 800 cases. See Table 2 for details of various intermediate levels of training set enrichment.

The results reported in Table 2 indicate that enrichment of a training set of smooth surfaces by diffeomorphic random perturbations or interpolations leads to improved classification accuracy.

So far, we have relied on the OOB score as a well-known robust estimate of any RF classifier's accuracy. To quantify the OOB robustness for our specific classification task, we have arbitrarily selected validation sets containing 15% of all cases, and trained our RF classifiers on the remaining 85% of all cases. The accuracy scores separately computed by OOB and via validation sets had identical medians 96.7% as well as very close means (96.5% and 96.4%). This confirms the robustness of OOB scores in our benchmark application.

Table 3 OOB accuracy obtained when the reference set of patients has sizes smaller than 100

n_r	Minimum	Median	Maximum
20	96.1	96.8	97.4
50	96.5	96.9	97.4

We report (from left to right) the number of reference patients, and the min, median, and max OOB score. Note that the OOB accuracy remains high even when the size of the reference set decreases

We performed additional experiments with a smaller reference set of patients. The best OOB scores as a function of the size of the reference set of patients are reported in Table 3.

In practical scenarios, where a reduction in computational cost is desirable during RF training, a smaller reference set can offer a sizable reduction of computational costs, at the price of a moderate degradation in classification accuracy. In our benchmark study, reducing the size of the random reference set from 100 to 20 cases reduces the computational cost by a factor of five but only degrades the best achievable OOB classification accuracy from 97.0% to 96.5%. This was verified by implementing 1 500 RF classifiers, involving 50 random reference sets of size 20 and 30 RF classifiers per reference set.In practical scenarios where computational cost reduction is desired during RF training, a smaller reference set can offer a sizable reduction of computational costs, at the price of a moderate degradation in classification accuracy.

Clearly, RF classifiers are not the only option. We have tested several other classifiers using the discriminating distance-based features as inputs. We have found that ridge classifiers and MLPs can reach an accuracy that is competitive with the RF classifier.

3.2.4 Importance Evaluation for Key Groups of Features

Recall that we had selected (see (19)) a set of nine strongly invariant dissimilarities $D_1, \dots, D_9$ computable by diffeomorphic matching between pairs of surfaces (S, S'), namely,[1]

$$\begin{aligned}&highstrain_{80}, \quad highstrain_{160}, \quad highstrain_{240},\\ &highstrain_{800}, \quad medstrain_{80}, \quad medstrain_{160},\\ &medstrain_{240}, \quad medstrain_{800}, \quad sqrtkin.\end{aligned}$$

For each $i = 1 \dots 9$, the dissimilarity D_i determines a group Γ_i of 100 features $\text{vec}_{i,j}(S) = D_i(S, \Sigma_j)$, where $j = 1 \dots 100$. A well-known technique to evaluate concretely the *importance* IMP_i of the group of features Γ_i is to compute the loss in OOB accuracy for each feature in Γ_i, one scrambles the 800 values $D_i(S, \Sigma_j)$ over all $S \in \mathcal{D}$. To implement this random scrambling, one selects 100 random permutations $PER_1, \dots, PER_{100}$ of our dataset of 800 cases, and then one replaces each $D_i(S, \Sigma_j)$ by $D_i(PER_j(S), \Sigma_j)$. This random scrambling, done for the fixed subgroup of features Γ_i naturally yields a decrease in OOB accuracy for the already trained classifier RF^*. For a more precise estimate of this accuracy decrease, the scrambling operation is repeated 100 times for each subgroup Γ_i and one computes the average OOB accuracy decrease IMP_i, which then quantifies the importance of the features subgroup Γ_i. The importance IMP_i of our nine subgroups of features Γ_i are reported in decreasing order in Table 4, which, hence, also ranks our nine dissimilarities D_i by decreasing importance for our discrimination task. These results identify the three most important dissimilarities D_1, D_2, D_9, with importances 28%, 13.5%, and 11%, respectively. The top two are the 95%-quantiles of strain intensities $D_1 = highstrain_{80}$ and $D_2 = highstrain_{160}$ focused n the 80 and 160 points closest to the coaptation line, respectively. The third top dissimilarity is the kinetic energy $D_9 = sqrtkin$.

Table 4 Importances IMP_i of our nine subgroups of features Γ_i

Dissimilarity	Importance (%)
$highstrain_{80}$	28.0
$highstrain_{160}$	13.5
$sqrtkin$	11.0
$highstrain_{240}$	9.0
$highstrain_{800}$	8.5
$medstrain_{160}$	8.5
$medstrain_{80}$	7.5
$medstrain_{800}$	7.0

The preceding importance analysis led us to also implement another RF classification based only on the three most important subgroups of intrinsic features, namely the three groups of 100 features each defined by the dissimilarities $D_1 = highstrain_{80}, D_2 = highstrain_{160}$ and the kinetic energy $D_9 = sqrtkin$. We trained 100 new RF classifiers based only on these three groups of 100 intrinsic features, and the best new RF^{**} classifier based on this reduced set of 300 intrinsic features reached a global OOB accuracy of 96.7%, which is very close to the 97.6% OOB accuracy of the best classifier RF^* based on 9 dissimilarities and the associated 900 intrinsic features. Thus, to discriminate efficiently between regurgitation and normal MVs, the strongest information derived from dissimilarities based on diffeomorphic registration is provided by the 95%-quantiles of the strain values observed very close to the coaptation line, and is effi-

---

[1] The first four D_i are the 95%-quantiles of strain intensities observed on four increasingly larger neighborhoods of the coaptation line on our MV surfaces; the next four D_i are the corresponding 50%-quantiles of strain intensities; the last D_9 is the square root of the kinetic energy required for optimal diffeomorphic matching of (S, S').

ciently completed by the kinetic energy of diffeomorphic registration between pairs of MV surfaces.

3.3 Computing Times

Most of the computing time involved in the preceding benchmark application was consumed by the implementation of diffeomorphic registration for about $800(100) = 80{,}000$ pairs of cropped MV surfaces, which took about 80 h of computing time on 80 nodes of the Opuntia cluster of the Computing Center at the University of Houston, as detailed below. After all diffeomorphic registrations were completed, automatic training of any RF classifiers was quite fast, with a runtime inferior to 30 s for each such training. This short RF training time was to be expected since we had only 800 cases and 300 trees.

Recall that our cropped MV surfaces were all discretized by 800-point triangulated grids. On a laptop with 1.4 GHz Quad-Core Intel Core i5, 16 GB memory, and 2133 MHz LPDDR3, the diffeomorphic registration of one pair of MV surfaces using our solver requires an average computing time of 20 s. The large number of such diffeomorphic registrations needed here were naturally distributed by deploying our solver on a cluster. For our benchmark dataset of 800 cropped MV surfaces, our solver had to be executed for 80,000 pairs of MV surfaces, requiring about 2 493 min for parallel computing on 80 nodes used at full availability. Since at any given time, we only had simultaneous use of about 40 nodes, completing these 80,000 diffeomorphic registrations took about 80 h of computing time.

4 Conclusions and Future Work

Given any dataset $\mathcal{D}$ of smooth 3D surfaces partitioned into a finite set of disjoint classes, implementing automatic class prediction by supervised ML classifiers such as RFs [59, 60], MLPs [61], or SVMs [63] usually requires the characterization of every surface $S \in \mathcal{D}$ by a computable feature vector $\text{vec}(S) \in \mathbb{R}^p$, for some fixed p. In this paper, we use the diffeomorphic registration of smooth surfaces to develop several algorithms dedicated to an efficient implementation of ML for the automatic classification of smooth 3D surfaces. In earlier research, we have developed a software solver to compute a diffeomorphic registration $f : S \rightarrow S'$ for any given pair (S, S') of smooth 3D surfaces, with f embedded in a smooth diffeomorphic flow with minimal kinetic energy $\text{kin}(S, S')$ [18, 65, 66]. We also compute the isotropic strain of f at each $x \in S$, and several quantiles $\text{quant}_\alpha(S, S')$ of all these strain values. From each $\text{quant}_\alpha(S, S')$, as well as from $\text{kin}(S, S')$, we derive a $G3$-invariant dissimilarity $\text{dis}(S, S')$ between surfaces S and S'. Here, $G3$ is the group of mappings from $\mathbb{R}^3$ to $\mathbb{R}^3$, generated by translations, rotations, and homotheties. We have then outlined how this family of dissimilarities $\text{dis}(S, S')$ can generate natural feature vectors $\text{vec}(S) \in \mathbb{R}^N$ defined for all surfaces $S \in \mathcal{D}$, and invariant by the group $G3$. This is an efficient first step to apply ML for supervised automatic classification in $\mathcal{D}$.

Moreover, since unbalanced class sizes tend to degrade classification accuracy by standard ML approaches, we develop two class enrichment algorithms derived from optimized diffeomorphic registration. Given two surfaces $S, S' \in \mathcal{D}$ belonging to the same given class CL we need to enrich, numerical diffeomorphic registration of S and S' automatically computes a set of surfaces $S_t \in \mathcal{D}$ depending smoothly on $t \in (0, 1)$, which are all diffeomorphic to S, and verify $S_0 = S$, $S_1 = S'$. Then, if S, S' are sufficiently close to each other, any interpolating surface S_t can be added to class CL as a virtual new case. Our diffeomorphic interpolation is a nonlinear extension of the well-known SMOTE enrichment technique [119], which is restricted to linear interpolation between Euclidean vectors.This interpolation technique enables a powerful extension of the well-known SMOTE enrichment technique [119] which is restricted to linear interpolation between Euclidean vectors. For class enrichment in $\mathcal{D}$, we have also developed and implemented small random perturbations of smooth 3D surfaces by applying flows of *random smooth diffeomorphisms* of $\mathbb{R}^3$, generated by time integration of smooth Gaussian random vector fields indexed by $\mathbb{R}^3$. Numerical implementation required a sophisticated convergence analysis for stochastic series expansions of smooth Gaussian vector fields. In our benchmark MV data set study, class enrichment by small random diffeomorphic perturbations was beneficial, while enrichment by diffeomorphic interpolation did not offer a clear improvement of the OOB accuracy. We anticipate that diffeomorphic interpolation would be more beneficial for larger data sets of smooth surfaces since interpolation could then be restricted to much closer pairs of surfaces.

The pre-treatment pipeline developed for our benchmark MV dataset uses several problem-specific steps that may not be generally applicable to arbitrary smooth surfaces S. For example, we use anatomical labels available for MV surfaces, such as the tagged coaptation line. This information can be used to improve the accuracy of the diffeomorphic matching and enables cropping to focus on a zone of interest, which speeds up computations. Such labels may not exist for other datasets. However, we expect that our general strategy applies to other datasets of smooth surfaces. In particular, the multiple distances derived from diffeomorphic registration such as strain quantiles, the characterization of each surface by its distances to a random finite set of prototype surfaces, and RF automatic classification based on these distances. Our class-enriching strategies based on diffeomorphic registration are also valid for quite generic sets of smooth surfaces. The discriminating power of our distance-based features will be application-dependent, and other datasets may require the design of different distances between surfaces.

We have successfully applied all the preceding tools and methodologies to the automatic classification of 800 MV surfaces into two classes: "regurgitation" versus "normal." The $G3$-invariant feature vectors we computed had dimension 900 and were derived from nine distinct dissimilarities (kinetic energy and eight strain quantiles). The well-known RF classifiers (here used with 300 trees) reached a high classification accuracy of 97.6%. A comparative importance analysis between nine groups of features (out of eleven candidate groups discussed in the present work) showed that the three most important dissimilarities were the 95%-quantiles of isotropic strain observed near the coaptation line of each MV surface and the kinetic energy. In fact, after the reduction of our feature vectors to the 300 features associated with the three most important dissimilarities, the RF classification accuracy was still quite high at 96.7%.

In our current work, we started to explore the use of optimal transport for computing distances between shape representations for classification. Our initial experiments based on off-the-shelf implementations for optimal transport did not yield satisfactory results.

On our benchmark data set, we have also tested the performance of neural networks classifiers, by training PointNet [82], which uses $\mathbb{R}^3$ coordinates of point clouds as features. To avoid over-fitting, we still had to enrich our data set of smooth MV surfaces by diffeomorphic perturbations and interpolations. The best PointNet architecture reached a high accuracy of 97.5%, and training time on our rather small dataset was fast, but not faster than for our RF classifiers. What was lost in the neural net approach was the evaluation of features relative importance, since each single point in a large point cloud has nearly zero importance.

In ongoing research, we also plan to apply the proposed techniques to automatic classification of much larger datasets of 3D smooth surfaces, into dozens of classes. This will require the development of ML techniques specifically dedicated to reducing the computing times required by large numbers of diffeomorphic registrations.

In our past work, we have focused on the design of effective numerical algorithms for the solution of the variational problem formulation [18, 65, 66]. Our methods have been developed in MATLAB. The computational kernels are consequently not necessarily optimized. In our future work, we intend to improve computational throughput by considering hardware-accelerated computational kernels. An example is the `KeOPs` library that implements efficient kernel matrix multiplication on GPUs [143].

Appendix A

A.1 Smooth One-Dimensional Gaussian Random Fields on $\mathbb{R}^3$

For $t \in \mathbb{R}$, the density function $\frac{1}{\sqrt{\pi}} \exp(-t^2)$ defines a measure γ of mass 2. Let $\Gamma = L_2(\mathbb{R}, \gamma)$. The Gaussian kernel $g(t, t') = \exp(-(t - t')^2)$ defines a linear operator $G : \Gamma \to \Gamma$ given by, for $\phi \in \Gamma$,

$$G\phi(t) = \int_{\mathbb{R}} g(t, t')\phi(t')\, \mathrm{d}\gamma(t')$$

for all $t \in \mathbb{R}$. Then, G is a Hilbert–Schmidt operator with known eigenvalues λ_n and eigenfunctions ϕ_n, which form an orthonormal basis of Γ. One has the converging series expansions

$$G\phi(t) = \sum_{n\geq 0} \lambda_n \langle \phi_n, \phi \rangle_\Gamma \phi_n(t) \tag{A1}$$

for all $\phi \in \Gamma$ and

$$g(t, t') = \sum_{n\geq 0} \lambda_n \phi_n(t)\phi_n(t') \tag{A2}$$

for all $t, t' \in \mathbb{R}$.

The ϕ_n are expressed below in terms of the Hermite polynomials $H_n(t)$, which are recursively given by $H_0(t) = 1$ and

$$H_{n+1}(t) = 2tH_n(t) - 2nH_{n-1}(t) \tag{A3}$$

with $n = 0, 1, \ldots$, for all $t \in \mathbb{R}$. Each $H_n(t)$ has degree n and leading term $(2t)^n$. The λ_n and ϕ_n are given by the known formulas (see [144])

$$\lambda_n = a^{n+1/2}, \tag{A4a}$$

$$\phi_n(t) = \frac{b}{\sqrt{2^n n!}} \exp(-ct^2) H_n(ht), \tag{A4b}$$

where $a = 1/(1 + \frac{\sqrt{5}}{2}) < 1/2$, $b = 5^{1/8}$, $c = (\sqrt{5} - 1)/2$, and $h = 5^{1/4}$. Rewrite the radial kernel $k(x, y) = \exp(-|x - y|^2)$, $x, y \in \mathbb{R}^3$, as follows:

$$k(x, y) = g(x_1, y_1)g(x_2, y_2)g(x_3, y_3), \tag{A5}$$

where x, y have coordinates x_i, y_i, $i = 1, 2, 3$. Endow $\mathbb{R}^3$ with the product measure $\theta = \gamma^3$. Define the Hilbert–Schmidt operator HS from $M = L_2(\mathbb{R}^3, \theta)$ into itself by

$$HS\psi(x) = \int_{\mathbb{R}^3} k(x, y)\psi(y)\, \mathrm{d}\theta(y)$$

for all functions $\psi \in M$. The eigenvalues and eigenfunctions of *HS* are given by

$$\tau_{m,n,p} = \lambda_m \lambda_n \lambda_p \tag{A6}$$

$$\psi_{m,n,p}(x) = \phi_m(x_1)\phi_n(x_2)\phi_p(x_3) \tag{A7}$$

for all integers m, n, p and all $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. The $\psi_{m,n,p}$ provide an orthonormal basis for M, and one has the converging series expansions

$$HS\psi = \sum_{m,n,p} \tau_{m,n,p} \langle \psi_{m,n,p}, \psi \rangle_M \psi_{m,n,p} \tag{A8}$$

for all $\psi \in M$ and

$$k(x, y) = \sum_{m,n,p} \tau_{m,n,p} \psi_{m,n,p}(x)\psi_{m,n,p}(y) \tag{A9}$$

for all $x, y \in \mathbb{R}^3$.

We now concretely construct an explicit stochastic series converging to a smooth $\mathbb{R}$ valued random Gaussian vector field $U : \mathbb{R}^3 \to \mathbb{R}$ with mean 0 and covariance kernel $k(x, y)$. The construction is summarized in Theorem 1. The speed of convergence for this series is studied in Theorem 2. The proofs of these two theorems are given in "Appendix A.3."

Theorem 1 *Let $Z_{m,n,p}$ be any sequence of standard independent Gaussian random variables, indexed by the integer triplets m, n, p, and defined on the same probability space (Ω, P). Then P-almost surely, the following stochastic series converges pointwise for all $x \in \mathbb{R}^3$ to a finite limit denoted*

$$U_x = \sum_{m,n,p} Z_{m,n,p} \sqrt{\tau_{m,n,p}}\, \psi_{m,n,p}(x). \tag{A10}$$

The one-dimensional random vector field $U_x \in \mathbb{R}$ defined by this series is Gaussian with mean 0 and covariance kernel $\mathrm{E}[U_x U_y] = k(x, y)$ for all $x, y \in \mathbb{R}^3$. The function $x \mapsto U_x$ is in $M = L_2(\mathbb{R}^3, \theta)$, and is P-almost surely smooth in $x \in \mathbb{R}^3$. Moreover, the series (A10) also converges to U in the $L_2(\mathbb{R}^3, \theta)$-norm.

Proof See "Appendix A.3." □

Theorem 2 *The partial sums $U(N)_x$ of the series U_x are denoted*

$$U(N)_x = \sum_{(m,n,p)\in B(N)} Z_{m,n,p} \sqrt{\tau_{m,n,p}} \psi_{m,n,p}(x), \tag{A11}$$

where $B(N) = \{m, n, p \mid \max(m, n, p) \le N\}$. With probability $q(N) > 1 - \frac{1}{5N^{10}}$, one has the uniform bound

$$|U_x - U(N)_x| \le c \exp(|x|^2/2) N^{7/10}/2^N, \tag{A12}$$

for all $x \in \mathbb{R}^3$ and all $N > 6$. Here, c is a universal constant that does not depend on N nor on x.

Proof See "Appendix A.3." □

A.2 Numerical Implementation of Random Diffeomorphic Deformations

Select independent standard Gaussian random variables $Z^j_{m,n,p}$ indexed by $j = 1, 2, 3$ and by all the triplets (m, n, p). Then as in (A10), define on $\mathbb{R}^3$ three independent $\mathbb{R}$-valued *smooth* Gaussian random fields U^j, $j = 1, 2, 3$, by the almost surely convergent series

$$U^j_x = \sum_{m,n,p} Z^j_{m,n,p} \sqrt{\tau_{m,n,p}} \psi_{m,n,p}(x) \tag{A13}$$

for all $x \in \mathbb{R}^3$. Fix any positive scale parameters s_1, s_2, s_3 and the three scaled Gaussian kernels

$$k_j(x, y) = k(x/s_j, y/s_j)$$

for $j = 1, 2, 3$, and for all $x, y \in \mathbb{R}^3$. Define the random vector field $W_x = [W^1_x, W^2_x, W^3_x]$ by

$$W^j_x = U^j_{x/s_j} \quad \text{for} \quad j = 1, 2, 3, \tag{A14}$$

and all $x \in \mathbb{R}^3$. Due to Theorem 1, we have (i) The random vector field $x \mapsto W_x$, $W_x \in \mathbb{R}^3$, is Gaussian with zero mean and independent coordinates. (ii) The covariance kernel of W is the 3×3 diagonal matrix $\mathrm{Cov}(W_x, W_y) = \mathrm{diag}(k_1(x, y), k_2(x, y), k_3(x, y))$ for all $x, y \in \mathbb{R}^3$. (iii) Almost surely, the function $x \mapsto W_x$ is infinitely differentiable in $x \in \mathbb{R}^3$.

For numerical computations of W_x when $x \in \mathbb{R}^3$ is given, the main point is to determine how many terms to keep in the basic stochastic series (A10) defining U_x. As seen in Theorem 1, for $N \ge 10$, with probability $q(N)$ practically equal to 1, the partial sums $U(N)_x$ approximate U_x at a speed faster than $c \exp |x|^2/2N^{7/10}/2^N$, where c is a numerical constant. Our explicit theoretical upper bound for c is too large for pragmatic estimates of the remainder $|U_x - U(N)_x|$. But our numerical experiments and more concrete estimates of Hermite polynomials (see [145]) indicate that for $|x| \le 4$, one can obtain a good approximation of the series U_x by keeping only the terms $T_{m,n,p}(x)$ such than $m, n, p \le 25$. The restriction $|x| \le 4$ is not a real constraint since after an adequate $\mathbb{R}^3$-homothety, any bounded surface S can become a surface Σ included in a small ball of radius 4 around its center of mass, and after implementing a numeric small random deformation of $\Sigma \to \Sigma(t)$, one can rescale $\Sigma(t)$ to give it a total area that is equal to 1.

A.3 Proofs of Theorems 1 and 2

Proof Clearly, $x \to U(N)_x$ is a random smooth function of $x \in \mathbb{R}^3$, and belongs to $L_2(\mathbb{R}^3, \theta)$. Since the $Z_{m,n,p}$ are decorrelated, one has for all $x, y \in \mathbb{R}^3$ that $\mathrm{E}[U(N)_x U(N)_y]$ is given by

$$\sum_{m,n,p \in B(N)} \tau_{m,n,p} \psi_{m,n,p}(x) \psi_{m,n,p}(y).$$

In view of the converging series expansion (A9) of $k(x, y)$, this implies

$$k(x, y) = \lim_{N \to \infty} \mathrm{E}[U(N)_x U(N)_y]. \tag{A15}$$

Define the random function $x \to T_{m,n,p}(x)$ for all $x \in \mathbb{R}^3$ by

$$T_{m,n,p}(x) = Z_{m,n,p} \sqrt{\tau_{m,n,p}} \psi_{m,n,p}(x). \tag{A16}$$

The function $T_{m,n,p}$ belongs to the space $L_2(\mathbb{R}^3, \theta)$, where its norm $A_{m,n,p}$ is given by $A_{m,n,p} = |Z_{m,n,p}| \sqrt{\tau_{m,n,p}}$. Thanks to equations (A6), (A4) one has

$$\sum_{m,n,p} E(A^2_{m,n,p}) = \sum_{m,n,p} \tau_{m,n,p} < \sum_{m,n,p} 1/2^{m+n+p} < \infty.$$

This implies since the random variables $A_{m,n,p}$ are independent, that P-almost surely, the following stochastic series converges to a (random) limit

$$\sum_{m,n,p} A_{m,n,p} = \sum_{m,n,p} \|T_{m,n,p}\|_{L_2(\mathbb{R}^3,\theta)}.$$

But whenever the series of $L_2(\mathbb{R}^3, \theta)$-norms $\|T_{m,n,p}\|$ converges, then the stochastic series $\sum_{m,n,p} T_{m,n,p}$ must also converge in $L_2(\mathbb{R}^3, \theta)$-norm to some random function $U \in L_2(\mathbb{R}^3, \theta)$. Hence, P-almost surely, we have

$$U = \lim_{N \to \infty} U(N) \tag{A17}$$

for convergence in $L_2(\mathbb{R}^3, \theta)$-norm. For short, we abbreviate P-almost surely as P-a.s. In what follows, we will also show that P-a.s. one also has the *pointwise* convergence

$$\lim_{N \to \infty} U(N)_x = U_x \quad \text{for all} \quad x \in \mathbb{R}^3. \tag{A18}$$

For faster exposition, we derive right away the main consequence of this P-a.s. pointwise convergence. In $\mathbb{R}^3$, fix any finite set of points $x(j)$, $j = 1, \dots, m$. Define the random vectors $X(N)$ and Y in $\mathbb{R}^m$ by their coordinates

$$X(N)_j = U(N)_{x(j)} \quad \text{and} \quad Y_j = U_{x(j)}. \tag{A19}$$

Denote $f_N(z) = E(\exp(i\langle z, X(N)\rangle))$ and $f(z) = E(\exp(i\langle z, Y\rangle))$ the characteristic functions of $X(N)$ and Y for all $z \in \mathbb{R}^m$. The $X(N)$ are Gaussian with mean 0, covariance matrix $Q(N)$, and tend P-a.s. to Y. Dominated convergence implies $f(z) = \lim_{N\to\infty} f_N(z)$ for all z. The formula $\log f_N(z) = z^* Q(N) z$, then forces $\log f(z) = z^* Q z$ with $Q = \lim_{N\to\infty} Q(N)$. Hence, Y is Gaussian with mean 0 and covariance matrix Q. This proves that the vector field $x \to U_x$ is Gaussian with mean zero and covariance kernel

$$E(U_x U_y) = \lim_{N \to \infty} E[U(N)_x U(N)_y] = k(x, y),$$

where the last equality is due to (A15). Since the covariance kernel $k(x, y)$ is infinitely differentiable in x and y, known results on random Gaussian fields (see [146]) show that one can find a version of the random field $x \mapsto U_x$, which will P-a.s. be smooth in $x \in \mathbb{R}^3$.

We still need to prove the P-a.s. pointwise convergence stated in (A18). We first derive bounds for the eigenfunctions ϕ_n given in (A4). The paper [147] provides sharp universal bounds for the Hermite polynomials $H_n(t)$. These bounds show that for all $n \geq 6$ and all $t \in \mathbb{R}$

$$|H_n(t)| < 9n^{-1/12} c_n \exp(t^2/2), \tag{A20}$$

where

$$c_n < \begin{cases} \sqrt{2} n^{1/4} \frac{n!}{(n/2)!} & \text{if } n \text{ is even,} \\ \sqrt{5}(n-1)^{3/4} \frac{(n-1)!}{((n-1)/2)!} & \text{if } n \text{ is odd.} \end{cases} \tag{A21}$$

Recall the Stirling formula, which states that for all $n > 1$

$$1 < \frac{n!}{\mathrm{stir}(n)} < \frac{3}{2} \tag{A22}$$

with $\mathrm{stir}(n) = (2\pi n)^{1/2} (n/e)^n$. This formula implies

$$\begin{cases} \frac{n!}{(n/2)!} < \frac{3}{2}(2n/e)^{n/2} & \text{for } n \text{ even,} \\ \frac{(n-1)!}{((n-1)/2)!} < \frac{3}{2}(\frac{2(n-1)}{e})^{(n-1)/2} & \text{for } n \text{ odd.} \end{cases} \tag{A23}$$

Combining the bounds (A21) and (A23), we get for all $t \in \mathbb{R}$ and all $n \geq 6$

$$|H_n(t)| < \begin{cases} 30\, n^{1/6} (2n/e)^{n/2} \exp(t^2/2) & \text{if } n \text{ is even,} \\ 30\, \xi(n) \exp(t^2/2) & \text{if } n \text{ is odd,} \end{cases}$$

with $\xi(n) = (n-1)^{2/3}(\frac{2(n-1)}{e})^{(n-1)/2}$ and hence a fortiori

$$|H_n(t)| < 30n^{2/3}(2n/e)^{n/2}\exp(t^2/2) \tag{A24}$$

for all $n \geq 6$, all $t \in \mathbb{R}$. (A4) yields two numerical constants c and h such that, for all $n \geq 1$, all $t \in \mathbb{R}$,

$$|\phi_n(t)| = \frac{b}{\sqrt{2^n n!}\exp(-ct^2)|H_n(ht)|}.$$

Moreover, one has $h^2/2 - c = 1/2$, which implies directly, in view of (A24),

$$|H_n(ht)| < 30n^{2/3}(2n/e)^{n/2}\frac{\exp(t^2/2)}{\exp(-ct^2)} \tag{A25}$$

for all $n \geq 6$, $t \in \mathbb{R}$.

From (A22) and the value of b, we get

$$\frac{b}{\sqrt{2^n n!}} < n^{-1/4}(2n/e)^{-n/2}$$

for all $n > 1$. So, we can finally bound the eigenfunctions ϕ_n by

$$|\phi_n(t)| < 30\exp(t^2/2)n^{1/2} \tag{A26}$$

for all $n \geq 6, t \in \mathbb{R}$. From (A7) and (A26), we derive the following bound, valid for all $m, n, p \geq 6$, all $x = [x_1, x_2, x_3] \in \mathbb{R}^3$,

$$\begin{aligned}|\psi_{m,n,p}(x)| &= |\phi_m(x_1)\phi_n(x_2)\phi_p(x_3)| \\ &< 30^3\exp|x|^2/2(mnp)^{1/2}.\end{aligned} \tag{A27}$$

From (A4), we get $\lambda_n < (1/2)^n$ and hence

$$\sqrt{\tau_{m,n,p}} < (1/2)^{m+n+p}. \tag{A28}$$

We now compute simultaneous probabilistic bounds for the $|Z_{m,n,p}|$. Define the following sequence of independent random events. $\square$

Definition 1 The event $E(m,n,p)$ is realized iff $\{|Z_{m,n,p}| \leq 5\sqrt{\log(mnp)}\}$.

Lemma 3 *Define the probabilities* $q(N)$ *by*

$$\begin{aligned}q(N) = P(&\textit{all events } E(m,n,p) \textit{ with} \\ &mnp > N \textit{ are realized} \\ &\textit{simultaneously}).\end{aligned} \tag{A29}$$

Then, for each $N \geq 4$, *one has*

$$q(N) > 1 - \frac{1}{5N^{10}}. \tag{A30}$$

Proof Any standard Gaussian random variable Z, verifies for all $t > 1$

$$\begin{aligned}P(|Z| > t) &< 2\int_{s>t}\frac{s}{\sqrt{2\pi}}\exp(-s^2/2)\,ds \\ &= \sqrt{2/\pi}\exp(-t^2/2)\end{aligned}$$

and hence for any integer $r \geq 2$

$$P(|Z| > 5\sqrt{\log(r)}) < \frac{2}{3r^{25/2}}. \tag{A31}$$

Since the $E(m,n,p)$ are independent, definition (A29) implies

$$q(N) = \prod_{m,n,p|mnp>N} P(E(m,n,p)).$$

Due to (A31), this yields for $N \geq 4$

$$q(N) > \prod_{m,n,p\,|\,mnp>N}\left[1 - \frac{1}{(mnp)^{25/2}}\right].$$

For $0 < u < 10^{-3}$, one has $1 - u > \exp(-2u)$ so that for $N \geq 4$

$$q(N) > \exp(-2s(N))$$

with

$$s(N) = \sum_{m,n,p|mnp>N}\frac{1}{(mnp)^{25/2}}.$$

For $N \geq 4$, one has

$$\begin{aligned}&\mathrm{card}\{m,n,p \mid mnp = N\} \\ &\quad < \sum_{m=1}^{N}\mathrm{card}\{n,p\,|\,np = N/m\} \\ &\quad < \sum_{m=1}^{N} N/m < N + N\log(N),\end{aligned} \tag{A32}$$

which implies for $N \geq 6$

$$\begin{aligned}s(N) &\leq \sum_{k>N}(k + k\log(k))/k^{25/2} \\ &\leq \sum_{k>N} 1/k^{11} \leq 1/10N^{10}.\end{aligned} \tag{A33}$$

Hence, for $N \geq 6$, we obtain

$$q(N) > \exp(-2/10N^{10}) > 1 - \frac{1}{5N^{10}}.$$

$\square$

We now study the series remainders

$$|U_x - U(N)_x|$$

for $N > 6$. By definition of $U(N)$, we have

$$|U_x - U(N)_x| < \sum_{(m,n,p) \in G(N)} |T_{m,n,p}(x)|,$$

where

$$G(N) = \{m, n, p \mid \max(m, n, p) > N\}.$$

Denote $\beta(x) = 30^3 \exp(|x|^2/2)$. With probability $q(N) > 1 - 1/5N^{10}$, we will have $|Z_{m,n,p}| \leq 5\sqrt{\log(mnp)}$ for all (m, n, p) such that $mnp > N$. In view of the two bounds, (A27) and (A28), we conclude that with probability $q(N)$ we will have, for all $x \in \mathbb{R}^3$ and all m, n, p verifying $(\min(m, n, p) > 6) \wedge (mnp > N)$,

$$\begin{aligned} |T_{m,n,p}(x)| &< 5\beta(x) \log(mnp) \sqrt{mnp} 2^{m+n+p} \\ &< \frac{10}{2^{m+n+p}} \beta(x) (mnp)^{7/10}, \end{aligned} \tag{A34}$$

where the second inequality follows from $\log(mnp) < 2(mnp)^{1/5}$. Since $mnp > N$, whenever $(m, n, p) \in G(N)$, we conclude that with probability $q(N)$ we will have for all $x \in \mathbb{R}^3$,

$$|U_x - U(N)_x| < 10\beta(x) r(N)$$

with

$$r(N) = \sum_{(m,n,p) \in G(N)} \frac{1}{2^{m+n+p}} (mnp)^{7/10}.$$

For $(m, n, p) \in G(N)$ one has $\max(m, n, p) > N$ and, hence,

$$r(N) < 3 \sum_{(m \leq p, n \leq p) \wedge (p > N)} \frac{1}{2^{m+n+p}} (mnp)^{7/10}.$$

Since

$$\sum_{m \leq p, n \leq p} \frac{1}{2^{m+n}} (mn)^{7/10} < \left[\sum_m \frac{1}{2^m} m^{7/10} \right]^2 < 4,$$

we obtain

$$r(N) < 12 \sum_{p > N} \frac{1}{2^p} p^{7/10} < \frac{24}{2^N} N^{7/10}.$$

Hence, with probability $q(N) > 1 - \frac{1}{5N^{10}}$, we will have for all $x \in \mathbb{R}^N$,

$$\begin{aligned} |U_x - U(N)_x| &< 10 r(N) \beta(x) \\ &< \frac{240}{2^N} N^{0.7} \beta(x) \\ &= 240(30^4 \exp(|x|^2/2)), \end{aligned}$$

which proves the announced bound (A12). Equation (A33) forces the series

$$\sum_{m,n,p} P(E(m, n, p))$$

to be finite. Hence, by Borel–Cantelli's lemma, there is a random integer *RAND* which is P-a.s. finite, and such that all the events $E(m, n, p)$ with $mnp > RAND$ are simultaneously realized. The arguments just used above show that the bound (A34) on $|T_{m,n,p}(x)|$ will hold for $mnp > RAND$. Whenever *RAND* is finite, this forces the pointwise convergence of the series U_x for all $x \in \mathbb{R}^3$. We have thus proved that P-a.s the series U_x will converge pointwise for all $x \in \mathbb{R}^3$. □

Acknowledgements This work was partly supported by the National Science Foundation under the Awards DMS-1854853, DMS-2012825, and DMS-2145845. Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of NSF. This work was completed in part with resources provided by the Research Computing Data Core at the University of Houston.

Author Contributions Hossein Dabirian involved in methodology, software, investigation, writing—original draft, writing—review and editing, and visualization. Radmir Sultamuratov involved in Methodology, software, investigation, writing—original draft, writing—review and editing, and visualization. James Herring involved in methodology, software, investigation, and writing—review and editing. Carlos El Tallawi involved in investigation, data curation, data acquisition, data annotation, and writing—review and editing. William Zoghbi involved in investigation, data curation, data acquisition, data annotation, and writing—review and editing. Andreas Mang involved in methodology, software, investigation, writing—original draft, writing—review and editing, visualization, supervision, and funding acquisition. Robert Azencott involved in methodology, investigation, writing—original draft, writing—review and editing, supervision, and funding acquisition.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, New York (2004)
2. Modersitzki, J.: FAIR: Flexible Algorithms for Image Registration. SIAM, Philadelphia (2009)

3. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. IEEE Trans. Med. Imaging **32**(7), 1153–1190 (2013)
4. Younes, L.: Shapes and Diffeomorphisms, 2nd edn. Springer, Berlin (2019)
5. Brunn, M., Himthani, N., Biros, G., Mehl, M., Mang, A.: Fast GPU 3D diffeomorphic image registration. J. Parallel Distrib. Comput. **149**, 149–162 (2021)
6. Brunn, M., Himthani, N., Biros, G., Mehl, M., Mang, A.: Multi-node multi-GPU diffeomorphic image registration for large-scale imaging problems. In: SC20: International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1–17. IEEE (2020)
7. Polzin, T., Niethammer, M., Heinrich, M.P., Handels, H., Modersitzki, J.: Memory efficient LDDMM for lung CT. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention, pp. 28–36 (2016)
8. Polzin, T., Niethammer, M., Vialard, F.-X., Modersitzki, J.: A discretize–optimize approach for LDDMM registration. In: Riemannian Geometric Statistics in Medical Image Analysis, pp. 479–532 (2020)
9. Sommer, S., Lauze, F., Nielsen, M., Pennec, X.: Kernel bundle EPDiff: evolution equations for multi-scale diffeomorphic image registration. In: Proceedings of International Conference on Scale Space and Variational Methods in Computer Vision, vol. 6667, pp. 677–688. LNCS (2011)
10. Mang, A., Ruthotto, L.: A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity-preserving diffeomorphic image registration. SIAM J. Sci. Comput. **39**(5), B860–B885 (2017)
11. Mang, A., Biros, G.: A semi-Lagrangian two-level preconditioned Newton–Krylov solver for constrained diffeomorphic image registration. SIAM J. Sci. Comput. **39**(6), B1064–B1101 (2017)
12. Mang, A., Gholami, A., Biros, G.: Distributed-memory large-deformation diffeomorphic 3D image registration. In: Proceedings of ACM/IEEE Conference on Supercomputing, pp. 842–853 (2016)
13. Mang, A., Gholami, A., Davatzikos, C., Biros, G.: CLAIRE: a distributed-memory solver for constrained large deformation diffeomorphic image registration. SIAM J. Sci. Comput. **41**(5), C548–C584 (2019)
14. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage **38**(1), 95–113 (2007)
15. Mang, A.: CLAIRE: scalable GPU-accelerated algorithms for diffeomorphic image registration in 3D (2024). arXiv:2401.17493 [math.OC]
16. Ashburner, J.: Computational anatomy with the SPM software. Magn. Reson. Imaging **27**(8), 1163–1174 (2009)
17. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging **23**(8), 995–1005 (2004)
18. Zhang, P., Mang, A., He, J., Azencott, R., El-Tallawi, K.C., Zoghbi, W.A.: Diffeomorphic shape matching by operator splitting in 3D cardiology imaging. J. Optim. Theory Appl. **188**(1), 143–168 (2021)
19. Burger, M., Modersitzki, J., Ruthotto, L.: A hyperelastic regularization energy for image registration. SIAM J. Sci. Comput. **35**(1), B132–B148 (2013)
20. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. **61**(2), 139–157 (2005)
21. Trouvé, A.: Diffeomorphism groups and pattern matching in image analysis. Int. J. Comput. Vis. **28**(3), 213–221 (1998)
22. Trouvé, A.: A Infinite Dimensional Group Approach for Physics Based Models in Pattern Recognition. Laboratoir d'Analyse Numerique CNRS URA, Universite Paris, Technical report (1995)
23. Glaunes, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. **2**, 712–718 (2004)
24. Glaunès, J., Qiu, A., Miller, M.I., Younes, L.: Large deformation diffeomorphic metric curve mapping. Int. J. Comput. Vis. **80**(3), 317–336 (2008)
25. Louis, M., Charlier, B., Durrleman, S.: Geodesic discriminant analysis for manifold-valued data. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 332–340 (2018)
26. Lee, S., Charon, N., Charlier, B., Popuri, K., Lebed, E., Sarunic, M.V., Trouvé, A., Beg, M.F.: Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (fshape) framework. Med. Image Anal. **35**, 570–581 (2017)
27. Campbell, K.M., Dai, H., Su, Z., Bauer, M., Fletcher, P.T., Joshi, S.C.: Structural connectome atlas construction in the space of Riemannian metrics. In: International Conference on Information Processing in Medical Imaging, pp. 291–303. Springer (2021)
28. Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. Neuroimage **59**(3), 2255–2265 (2012)
29. Hartman, E., Sukurdeep, Y., Klassen, E., Charon, N., Bauer, M.: Elastic shape analysis of surfaces with second-order Sobolev metrics: a comprehensive numerical framework. Int. J. Comput. Vis. **131**, 1183–1209 (2023)
30. Hsieh, D.-N., Arguillère, S., Charon, N., Younes, L.: Mechanistic modeling of longitudinal shape changes: equations of motion and inverse problems. SIAM J. Appl. Dyn. Syst. **21**(1), 80–101 (2022)
31. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: International Conference on Information Processing in Medical Imaging, pp. 37–48. Springer (2013)
32. Thompson, P.M., Toga, A.W.: A framework for computational anatomy. Comput. Vis. Sci. **5**(1), 13–34 (2002)
33. Guigui, N., Pennec, X.: Parallel transport, a central tool in geometric statistics for computational anatomy: application to cardiac motion modeling. Handb. Stat. **46**, 285–326 (2022)
34. Grenander, U., Miller, M.I.: Computational anatomy: an emerging discipline. Q. Appl. Math. **56**(4), 617–694 (1998)
35. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage **23**, S151–S160 (2004)
36. Miller, M.I.: Computational anatomy: shape, growth and atrophy comparison via diffeomorphisms. Neuroimage **23**(1), S19–S33 (2004)
37. Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler–Lagrange equations of computational anatomy. Annu. Rev. Biomed. Eng. **4**(1), 375–405 (2002)
38. Younes, L., Arrate, F., Miller, M.I.: Evolutions equations in computational anatomy. Neuroimage **45**, S40–S50 (2009)
39. Miller, M.I., Trouvé, A., Younes, L.: Hamiltonian systems and optimal control in computational anatomy: 100 years since D'Arcy Thompson. Annu. Rev. Biomed. Eng. **17**, 447–509 (2015)
40. Toga, A.W., Thompson, P.M.: The role of image registration in brain mapping. Image Vis. Comput. **19**(1–2), 3–24 (2001)
41. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)
42. Risser, L., Vialard, F.X., Wolz, R., Holm, D.D., Rueckert, D.: Simultaneous fine and coarse diffeomorphic registration: application to atrophy measurement in Alzheimer's disease. In: Pro-

ceedings of Medical Image Computing and Computer-Assisted Intervention, vol. 6362, pp. 610–617. LNCS (2010)

43. Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J.C., Csernansky, J.G., Miller, M.I.: Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. IEEE Trans. Med. Imaging **26**(4), 462–470 (2007)
44. Fox, N.C., Freeborough, P.A.: Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. J. Magn. Reson. Imaging **7**(6), 1069–1075 (1997)
45. Hua, X., Loew, A.D., Lee, S., Klunder, A.D., Toga, A.W., Lepore, N., Chou, Y.Y., Brun, C., Chiang, M.C., Barysheva, M., Jack, C.R., Bernstein, M.A., Britson, P.J., Ward, C.P., Whitwell, J.L., Borowski, B., Fleisher, A.S., Fox, N.C., Boyes, R.G., Barnes, J., Harvey, D., Kornak, J., Schuff, N., Boreta, L., Alexander, G.E., Weiner, M.W., Thompson, P.M.: 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage **41**(1), 19–34 (2008)
46. Joy, A., Nagarajan, R., Daar, E.S., Paul, J., Saucedo, A., Yadav, S.K., Guerrero, M., Haroon, E., Macey, P., Thomas, M.A.: Alterations of gray and white matter volumes and cortical thickness in treated HIV-positive patients. Magn. Reson. Imaging **95**, 27–38 (2023)
47. Misra, C., Fan, Y., Davatzikos, C.: Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage **44**(4), 1415–1422 (2009)
48. Davatzikos, C., Genc, A., Xu, D., Resnick, S.M.: Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. Neuroimage **14**(6), 1361–1369 (2001)
49. Bistoquet, A., Oshinski, J., Skrinjar, O.: Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model. Med. Image Anal. **12**(1), 69–85 (2008)
50. Gorce, J.-M., Friboulet, D., Magnin, I.E.: Estimation of three-dimensional cardiac velocity fields: assessment of a differential method and application to three-dimensional CT data. Med. Image Anal. **1**(3), 245–261 (1996)
51. Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N.: iLogDemons: a demons-based registration algorithm for tracking incompressible elastic biological tissues. Int. J. Comput. Vis. **92**(1), 92–111 (2011)
52. Sundar, H., Davatzikos, C., Biros, G.: Biomechanically constrained 4D estimation of mycardial motion. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention, vol. 5762, pp. 257–265. LNCS (2009)
53. Delingette, H., Billet, F., Wong, K.C.L., Sermesant, M., Rhode, K., Ginks, M., Rinaldi, C., Razavi, R., Ayache, N.: Personalization of cardiac motion and contractility from images using variational data assimilation. IEEE Trans. Biomed. Eng. **59**(1), 20–24 (2012)
54. Lombaert, H., Peyrat, J.-M., Croisille, P., Rapacchi, S., Fanton, L., Clarysse, P., Delingette, H., Ayache, N.: Statistical analysis of the human cardiac fiber architecture from DT-MRI. In: International Conference on Functional Imaging and Modeling of the Heart, pp. 171–179 (2011)
55. Vadakkumpadan, F., Arevalo, H., Ceritoglu, C., Miller, M., Trayanova, N.: Image-based estimation of ventricular fiber orientations for personalized modeling of cardiac electrophysiology. IEEE Trans. Med. Imaging **31**(5), 1051–1060 (2012)
56. Perperidis, D., Mohiaddin, R., Rueckert, D.: Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention. Lecture Notes In Computer Science, vol. 3750, pp. 402–410 (2005)
57. Bai, W., Shi, W., de Marvao, A., Dawes, T.J.W., O'Regan, D.P., Cook, S.A., Rueckert, D.: A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Med. Image Anal. **26**(1), 133–145 (2015)
58. Shen, D.G., Sundar, H., Xue, Z., Fan, Y., Litt, H.: Consistent estimation of cardiac motions by 4D image registration. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science, vol. 3750, pp. 902–910 (2005)
59. Cutler, A., Cutler, D.R., Stevens, J.R.: Random forests. In: Ensemble Machine Learning, pp. 157–175 (2012)
60. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
61. Rosenblatt, F.: The perceptron, a perceiving and recognizing automaton. Technical report, Cornell Aeronautical Laboratory (1957)
62. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al.: Recent advances in convolutional neural networks. Pattern Recogn. **77**, 354–377 (2018)
63. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
64. Dupuis, P., Gernander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. **56**(3), 587–600 (1998)
65. Azencott, R., Glowinski, R., He, J., Jajoo, A., Li, Y., Martynenko, A., Hoppe, R.H.W., Benzekry, S., Little, S.H.: Diffeomorphic matching and dynamic deformable surfaces in 3D medical imaging. Comput. Methods Appl. Math. **10**(3), 235–274 (2010)
66. Mang, A., He, J., Azencott, R.: An operator-splitting approach for variational optimal control formulations for diffeomorphic shape matching. J. Comput. Phys. **493**, 112463 (2023)
67. Zekry, S.B., Lawrie, G., Little, S., Zoghbi, W., Freeman, J., Jajoo, A., Jain, S., He, J., Martynenko, A., Azencott, R.: Comparative evaluation of mitral valve strain by deformation tracking in 3D-echocardiography. Cardiovasc. Eng. Technol. **3**(4), 402–412 (2012)
68. Zekry, S.B., Freeman, J., Jajoo, A., He, J., Little, S.H., Lawrie, G.M., Azencott, R., Zoghbi, W.A.: Patient-specific quantitation of mitral valve strain by computer analysis of three-dimensional echocardiography: a pilot study. Circ. Cardiovasc. Imaging **9**(1), e003254 (2016)
69. Zekry, S.B., Freeman, J., Jajoo, A., He, J., Little, S.H., Lawrie, G.M., Azencott, R., Zoghbi, W.A.: Effect of mitral valve repair on mitral valve leaflets strain: a pilot study. JACC Cardiovasc. Imaging **11**(5), 776–777 (2018)
70. El-Tallawi, K.C., Zhang, P., Azencott, R., He, J., Herrera, E.L., Xu, J., Chamsi-Pasha, M., Jacob, J., Lawrie, G.M., Zoghbi, W.A.: Valve strain quantitation in normal mitral valves and mitral prolapse with variable degrees of regurgitation. Cardiovasc. Imaging **14**(6), 1099–1109 (2021)
71. El-Tallawi, K.C., Zhang, P., Azencott, R., He, J., Xu, J., Herrera, E.L., Jacob, J., Chamsi-Pasha, M., Lawrie, G.M., Zoghbi, W.A.: Mitral valve remodeling and strain in secondary mitral regurgitation: comparison with primary regurgitation and normal valves. Cardiovasc. Imaging **14**(4), 782–793 (2021)
72. El-Tallawi, K.C., Zhang, P., Azencott, R., He, J., Herrera, E., Chamsi-Pasha, M., Jacob, J., Lawrie, G.M., Zoghbi, W.: Quantitation of mitral valve strain in normals and in patients with mitral valve prolapse. J. Am. Coll. Cardiol. **73**(9S1), 1953 (2019)
73. Freeman, J.: Combining diffeomorphic matching with image sequence intensity registration. PhD thesis, University of Houston (2014)
74. Jajoo, A.: Diffeomorphic matching and dynamic deformable shapes. PhD thesis, University of Houston (2011)

75. Osowski, S., Nghia, D.D.: Fourier and wavelet descriptors for shape recognition using neural networks—a comparative study. Pattern Recogn. **35**(9), 1949–1957 (2002)
76. Luciano, L., Ben Hamza, A.: Deep similarity network fusion for 3D shape classification. Vis. Comput. **35**(6), 1171–1180 (2019)
77. Wu, W.-Y., Wang, M.-J.J.: Detecting the dominant points by the curvature-based polygonal approximation. CVGIP Graph. Models Image Process. **55**(2), 79–88 (1993)
78. Torres, R.S., Falcao, A.X., Costa, L.F.: A graph-based approach for multiscale shape analysis. Pattern Recogn. **37**(6), 1163–1174 (2004)
79. Plotze, R.D.O., Falvo, M., Pádua, J.G., Bernacci, L.C., Vieira, M.L.C., Oliveira, G.C.X., Bruno, O.M.: Leaf shape analysis using the multiscale Minkowski fractal dimension, a new morphometric method: a study with Passiflora (Passifloraceae). Can. J. Bot. **83**(3), 287–301 (2005)
80. Junior, J.J., Backes, A.R., Bruno, O.M.: Randomized neural network based descriptors for shape classification. Neurocomputing **312**, 201–209 (2018)
81. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660 (2017)
82. Garcia-Garcia, A., Gomez-Donoso, F., Garcia-Rodriguez, J., Orts-Escolano, S., Cazorla, M., Azorin-Lopez, J.: PointNet: a 3D convolutional neural network for real-time object class recognition. In: 2016 International Joint Conference on Neural Networks, pp. 1578–1584. IEEE (2016)
83. Shen, X., Stamos, I.: Frustum VoxNet for 3D object detection from RGB-D or depth images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1698–1706 (2020)
84. Maturana, D., Scherer, S.: Voxnet: a 3D convolutional neural network for real-time object recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 922–928. IEEE (2015)
85. Lin, M., Feragen, A.: DiffConv: analyzing irregular point clouds with an irregular view. In: European Conference on Computer Vision, pp. 380–397. Springer (2022)
86. Sharp, N., Attaiki, S., Crane, K., Ovsjanikov, M.: DiffusionNet: discretization agnostic learning on surfaces. ACM Trans. Graph. **41**(3), 1–16 (2022)
87. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNET: an information-rich 3D model repository (2015). arXiv:1512.03012 [cs.GR]
88. Li, G., Choi, B., Xu, J., Bhowmick, S.S., Chun, K.-P., Wong, G.L.-H.: ShapeNET: a shapelet-neural network approach for multivariate time series classification. Proc. AAAI Conf. Artif. Intell. **35**, 8375–8383 (2021)
89. Lüdke, D., Amiranashvili, T., Ambellan, F., Ezhov, I., Menze, B.H., Zachow, S.: Landmark-free statistical shape modeling via neural flow deformations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 453–463. Springer (2022)
90. Heimann, T., Meinzer, H.-P.: Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. **13**(4), 543–563 (2009)
91. Ambellan, F., Lamecker, H., von Tycowicz, C., Zachow, S.: Statistical shape models: understanding and mastering variation in anatomy. In: Biomedical Visualisation, pp. 67–84 (2019)
92. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: 3D statistical shape models using direct optimisation of description length. In: European Conference on Computer Vision, pp. 3–20. Springer (2002)
93. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: A minimum description length approach to statistical shape modeling. IEEE Trans. Med. Imaging **21**(5), 525–537 (2002)
94. Bauer, M., Bruveris, M., Michor, P.W.: Overview of the geometries of shape spaces and diffeomorphism groups. J. Math. Imaging Vis. **50**, 60–97 (2014)
95. Bauer, M., Charon, N., Klassen, E., Kurtek, S., Needham, T., Pierron, T.: Elastic metrics on spaces of Euclidean curves: theory and algorithms. J. Nonlinear Sci. **34**(56), 1–37 (2024)
96. Marslanda, S., Sommer, S.: Riemannian geometry on shapes and diffeomorphisms. In: Riemannian Geometric Statistics in Medical Image Analysis, p. 135 (2019)
97. Bauer, M., Charon, N., Younes, L.: Metric registration of curves and surfaces using optimal control. In: Handbook of Numerical Analysis, vol. 20, pp. 613–646 (2019)
98. Miller, M.I., Younes, L.: Group actions, homeomorphism, and matching: a general framework. Int. J. Comput. Vis. **41**(1/2), 61–81 (2001)
99. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
100. Saito, A., Nawano, S., Shimizu, A.: Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs. Med. Image Anal. **28**, 46–65 (2016)
101. Kainmueller, D., Lamecker, H., Heller, M.O., Weber, B., Hege, H.-C., Zachow, S.: Omnidirectional displacements for deformable surfaces. Med. Image Anal. **17**(4), 429–441 (2013)
102. Kirschner, M.: The probabilistic active shape model: from model construction to flexible medical image segmentation. PhD thesis, Kirschner (2013)
103. Gollmer, S.T., Kirschner, M., Buzug, T.M., Wesarg, S.: Using image segmentation for evaluating 3D statistical shape models built with groupwise correspondence optimization. Comput. Vis. Image Underst. **125**, 283–303 (2014)
104. von Tycowicz, C., Ambellan, F., Mukhopadhyay, A., Zachow, S.: An efficient Riemannian statistical shape model using differential coordinates: with application to the classification of data from the Osteoarthritis Initiative. Med. Image Anal. **43**, 1–9 (2018)
105. Thomson, J., O'Neill, T., Felson, D., Cootes, T.: Automated shape and texture analysis for detection of osteoarthritis from radiographs of the knee. In: Medical Image Computing and Computer-Assisted Intervention, pp. 127–134. Springer (2015)
106. Hsieh, H.-W., Charon, N.: Diffeomorphic registration with density changes for the analysis of imbalanced shapes. In: International Conference on Information Processing in Medical Imaging, pp. 31–42. Springer (2021)
107. François, A., Gori, P., Glaunès, J.: Metamorphic image registration using a semi-Lagrangian scheme. In: International Conference on Geometric Science of Information, pp. 781–788. Springer (2021)
108. Bône, A., Louis, M., Martin, B., Durrleman, S.: Deformetrica 4: an open-source software for statistical shape analysis. In: International Workshop on Shape in Medical Imaging, pp. 3–13. Springer (2018)
109. Amor, B.B., Arguillère, S., Shao, L.: ResNet-LDDMM: advancing the LDDMM framework using deep residual networks. IEEE Trans. Pattern Anal. Mach. Intell. **45**(3), 3707–3720 (2022)
110. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE Trans. Med. Imaging **38**(9), 2165–2176 (2019)
111. Sun, S., Han, K., Kong, D., Tang, H., Yan, X., Xie, X.: Topology-preserving shape reconstruction and registration via neural diffeomorphic flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20845–20855 (2022)

112. Yang, X., Kwitt, R., Styner, M., Niethammer, M.: Quicksilver: fast predictive image registration—a deep learning approach. Neuroimage **158**, 378–396 (2017)
113. Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X.: Deep learning in medical image registration: a review. Phys. Med. Biol. **65**(20), 20TR01 (2020)
114. Mussabayeva, A., Kroshnin, A., Kurmukov, A., Denisova, Y., Shen, L., Cong, S., Wang, L., Gutman, B.A.: Image registration and predictive modeling: learning the metric on the space of diffeomorphisms. In: International Workshop on Shape in Medical Imaging, pp. 160–168. Springer (2018)
115. Joshi, A., Hong, Y.: R2Net: efficient and flexible diffeomorphic image registration using Lipschitz continuous residual networks. Med. Image Anal. **89**, 102917 (2023)
116. Bône, A., Colliot, O., Durrleman, S.: Learning the spatiotemporal variability in longitudinal shape data sets. Int. J. Comput. Vis. **128**(12), 2873–2896 (2020)
117. Charon, N., Younes, L.: Shape spaces: from geometry to biological plausibility. arXiv preprint arXiv:2205.01237, pp. 1–30 (2022)
118. Azencott, R., Glowinski, R., Ramos, A.M.: A controllability approach to shape identification. Appl. Math. Lett. **21**(8), 861–865 (2008)
119. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, P.: Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
120. Vaillant, M., Glaunes, J.: Surface matching via currents. In: International Conference on Information Processing in Medical Imaging, pp. 381–392. Springer (2005)
121. Kaltenmark, I., Charlier, B., Charon, N. A general framework for curve and surface comparison and registration with oriented varifolds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3346–3355 (2017)
122. Charon, N., Trouvé, A.: The varifold representation of nonoriented shapes for diffeomorphic registration. SIAM J. Imaging Sci. **6**(4), 2547–2580 (2013)
123. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **1**(3), 1–122 (2011)
124. O'Donoghue, B., Stathopoulos, G., Boyd, S.: A splitting method for optimal control. IEEE Trans. Control Syst. Technol. **21**(6), 2432–2442 (2013)
125. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**(3), 123–231 (2013)
126. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**(2), 421–439 (1956)
127. Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. ESAIM Math. Modell. Numer. Anal. **9**(R2), 41–76 (1975)
128. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
129. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)
130. Scheufele, K., Mang, A., Gholami, A., Davatzikos, C., Biros, G., Mehl, M.: Coupling brain-tumor biophysical models and diffeomorphic image registration. Comput. Methods Appl. Mech. Eng. **347**, 533–567 (2019)
131. Scheufele, K., Subramanian, S., Mang, A., Biros, G., Mehl, M.: Image-driven biophysical tumor growth model calibration. SIAM J. Sci. Comput. **42**(3), B549–B580 (2020)
132. Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: GLISTR: glioma image segmentation and registration. IEEE Trans. Med. Imaging **31**(10), 1941–1954 (2013)
133. Hogea, C., Davatzikos, C., Biros, G.: Brain–tumor interaction biophysical models for medical image registration. SIAM J. Sci. Comput. **30**(6), 3050–3072 (2008)
134. Li, X., Long, X., Laurienti, P., Wyatt, C.: Registration of images with varying topology using embedded maps. IEEE Trans. Med. Imaging **31**(3), 749–765 (2012)
135. Hsieh, H.-W., Charon, N.: Weight metamorphosis of varifolds and the LDDMM-Fisher-Rao metric. Calc. Var. Partial. Differ. Equ. **61**(5), 165 (2022)
136. François, A., Maillard, M., Oppenheim, C., Pallud, J., Bloch, I., Gori, P., Glaunès, J.: Weighted metamorphosis for registration of images with different topologies. In: International Workshop on Biomedical Image Registration, pp. 8–17. Springer (2022)
137. Antonsanti, P.-L., Glaunès, J., Benseghir, T., Jugnon, V., Kaltenmark, I.: Partial matching in the space of varifolds. In: International Conference on Information Processing in Medical Imaging, pp. 123–135. Springer (2021)
138. Sukurdeep, Y., Bauer, M., Charon, N.: A new variational model for shape graph registration with partial matching constraints. SIAM J. Imaging Sci. **15**(1), 261–292 (2022)
139. Ho, T.K.: Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282 (1995)
140. Parr, T., Turgutlu, K.: Feature importances for `scikit-learn` machine learning models. https://github.com/parrt/random-forest-importances. Accessed 01 Aug 2023
141. Parr, T., Turgutlu, K.: `rfpimp`. https://pypi.org/project/rfpimp. Accessed 01 Aug 2023
142. Breiman, L.: Out-of-bag estimation. Technical report, Statistics Department, University of California Berkeley (1996)
143. Charlier, B., Feydy, J., Glaunès, J.A., Collin, F.-D., Durif, G.: Kernel operations on the GPU, with autodiff, without memory overflows. J. Mach. Learn. Res. **22**(74), 1–6 (2021)
144. Fasshauer, G.E.: Positive definite kernels: past, present and future. Dolomit. Res. Notes Approx. **4**, 21–63 (2011)
145. Bonan, S.S., Clark, D.S.: Estimates of the Hermite and the Freud polynomials. J. Approx. Theory **63**(2), 210–224 (1990)
146. Stein, M.L.: Interpolation of Spatial Data. Springer, New York (1999)
147. Krasikov, I.: New bounds on the Hermite polynomials (2004). arXiv:math/0401310 [math.CA]

Hossein Dabirian is a Ph.D. student at the University of Michigan, Ann Arbor, where his research focuses on applied probability and game theory. Prior to joining the University of Michigan, he worked as a research assistant in the Department of Mathematics at the University of Houston, focusing on the application of machine learning to medical imaging. He also holds a master's degree in mathematics from Paris 7 University.

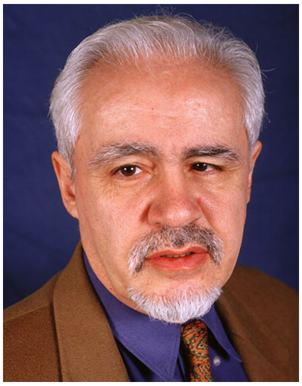
Radmir Sultamuratov is a Ph.D. candidate at the Department of Mathematics of the University of Houston. His research areas are medical image registration, machine learning, and computer vision.

James Herring is a Space Surveillance Engineer with Slingshot Aerospace specializing in image processing, telescope command & control, and telescope system design for novel SDA applications. He received his PhD in applied mathematics from Emory University in 2018 with a focus on optimization for large scale problems in imaging. From 2018 to 2019, he worked as a postdoctoral researcher at the University of Houston studying numerical optimization methods for PDE-constrained systems.

Carlos El Tallawi MD completed his medical school, internal medicine residency, and cardiology fellowship in Lebanon from 2006 to 2016. Following this, he undertook a three-year postdoctoral research fellowship at the Houston Methodist DeBakey Heart and Vascular Center, focusing on echocardiography in mitral valve prolapse. From 2019 to 2021, he pursued a two-year Advanced Cardiovascular Imaging fellowship at Houston Methodist. He is now an Assistant Professor of Medicine affiliated with Weill Cornell Medical College, engaged in clinical, research, and educational roles, with my research primarily centered on mitral valve prolapse and regurgitation.

William Zoghbi MD, holds the Elkins Family Distinguished Chair in Cardiac Health and is Professor of Medicine and chairman of the Department of Cardiology at Houston Methodist and the DeBakey Heart & Vascular Center. He attended the American University of Beirut and continued his medical studies at Meharry Medical College in 1979. He trained in Medicine and Cardiology at Baylor College of Medicine and joined its faculty from 1985 to 2005, where he was the John S. Dunn Professor of Medicine. In 2005, he joined Houston Methodist and was appointed Chairman of the Department of Cardiology in 2016. He is a renowned leader in cardiology, cardiac imaging, and global cardiovascular health. He has developed new techniques to evaluate heart function and valve disorders without resorting to invasive procedures. He is known for his research on ischemic heart disease, myocardial function and for his development of novel echocardiographic methods to evaluate the severity of valvular disease.

Andreas Mang is an Associate Professor at the Department of Mathematics at the University of Houston. He earned his Ph.D. from the University of Luebeck, Germany, in 2013. After that, he pursued a postdoctoral fellowship at the Oden Institute for Computational Engineering and Sciences at The University of Texas at Austin. In 2017, he joined the Department of Mathematics at the University of Houston, where he currently serves as an Associate Professor of Mathematics. His research interests include statistical and deterministic inverse problems, nonlinear optimal control, numerical optimization, data-enabled sciences, and parallel scientific computing.



Robert Azencott is currently Professor of Mathematics (University of Houston) since 2006 and Prof. Emeritus at Ecole Normale Superieure Paris-Saclay (France). He has held research and teaching positions at multiple institutions in the USA: University of California (Berkeley), Brandeis University (Boston) and in Paris, France: Universities Paris 7 and Paris 11, Ecole Polytechnique, Ecole Normale Superieure, (rue d'Ulm) and became Chair of the Maths Dept at Ecole Normale Superieure (formerly Cachan, now Paris-Saclay). His research interests are in machine learning, manifolds with negative curvature, probability and statistics, mathematical finance, shape analysis, computer vision. He has directed or co-directed so far more than 50 PhD theses in mathematics, as well as many applied mathematics projects focused on bio-medical imaging.