

**Shape modeling of longitudinal medical images
from diffeomorphic metric mapping to deep learning**

Tay, Edwin; Tümer, Nazli; Zadpoor, Amir A.

DOI

[10.3389/frai.2025.1671099](https://doi.org/10.3389/frai.2025.1671099)

Publication date

2025

Document Version

Final published version

Published in

Frontiers in Artificial Intelligence

Citation (APA)

Tay, E., Tümer, N., & Zadpoor, A. A. (2025). Shape modeling of longitudinal medical images: from diffeomorphic metric mapping to deep learning. *Frontiers in Artificial Intelligence*, 8, Article 1671099. <https://doi.org/10.3389/frai.2025.1671099>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



OPEN ACCESS

EDITED BY

Yassine Himeur,
University of Dubai, United Arab Emirates

REVIEWED BY

Adel Oulefki,
University of Sharjah, United Arab Emirates
Yassine Habchi,
University Center Salhi Ahmed Naama, Algeria

*CORRESPONDENCE

Edwin Tay
✉ e.w.s.tay@tudelft.nl

†These authors have contributed equally to this work and share last authorship

RECEIVED 22 July 2025

ACCEPTED 10 October 2025

PUBLISHED 30 October 2025

CITATION

Tay E, Tümer N and Zadpoor AA (2025) Shape modeling of longitudinal medical images: from diffeomorphic metric mapping to deep learning. *Front. Artif. Intell.* 8:1671099. doi: 10.3389/frai.2025.1671099

COPYRIGHT

© 2025 Tay, Tümer and Zadpoor. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Shape modeling of longitudinal medical images: from diffeomorphic metric mapping to deep learning

Edwin Tay*, Nazli Tümer† and Amir A. Zadpoor†

Department of Biomechanical Engineering, Faculty of Mechanical Engineering, Delft University of Technology (TU Delft), Delft, Netherlands

Living biological tissue is a complex system, constantly growing and changing in response to external and internal stimuli. These processes lead to remarkable and intricate changes in shape. Modeling and understanding both natural and pathological (or abnormal) changes in the shape of anatomical structures is highly relevant, with applications in diagnostic, prognostic, and therapeutic healthcare. Nevertheless, modeling the longitudinal shape change of biological tissue is a non-trivial task due to its inherent nonlinear nature. In this review, we highlight several existing methodologies and tools for modeling longitudinal shape change (i.e., spatiotemporal shape modeling). These methods range from diffeomorphic metric mapping to deep-learning based approaches (e.g., autoencoders, generative networks, recurrent neural networks, etc.). We discuss the synergistic combinations of existing technologies and potential directions for future research, underscoring key deficiencies in the current research landscape.

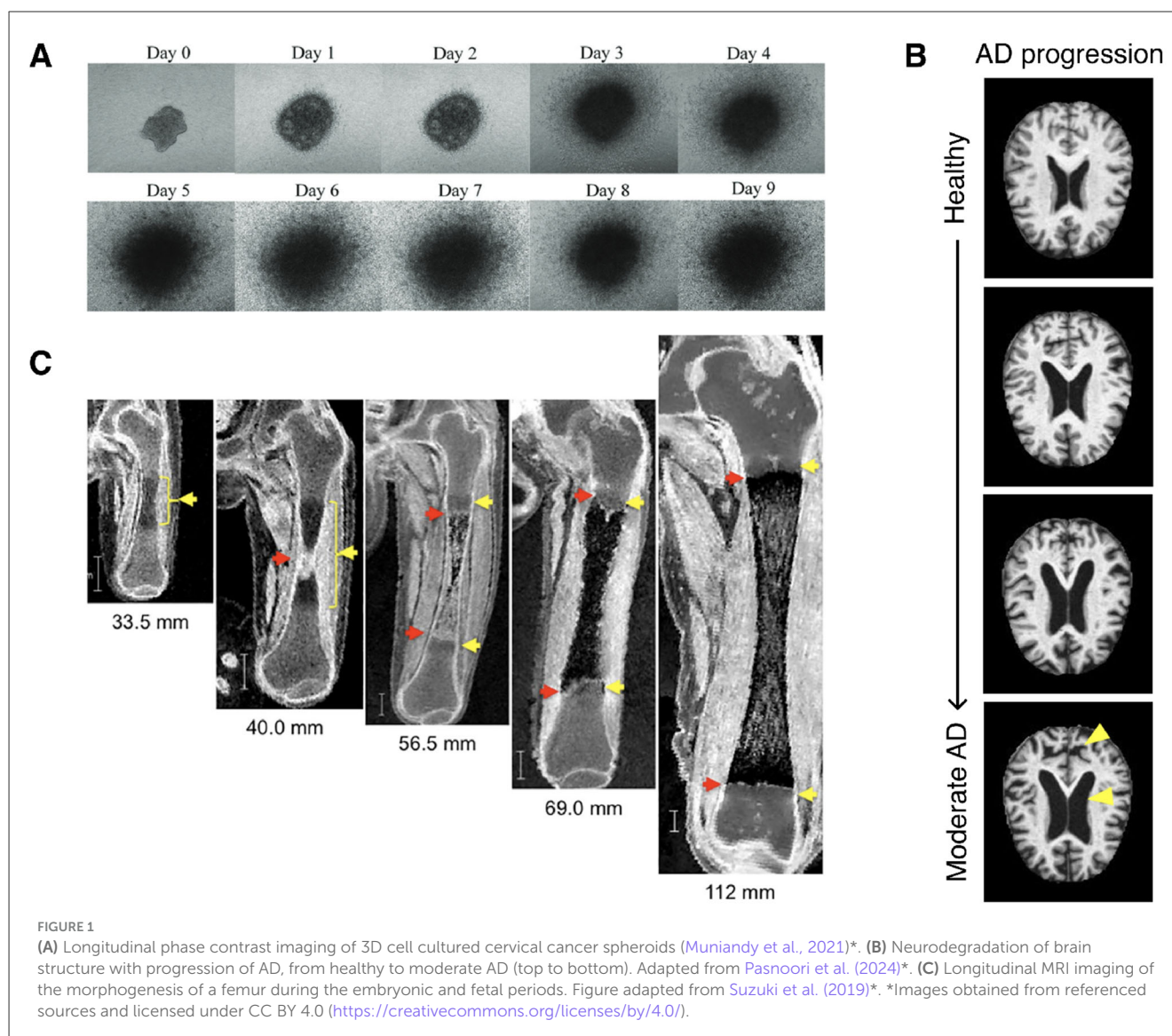
KEYWORDS

deep learning, shape modeling, spatiotemporal, medical imaging, diffeomorphisms, longitudinal data

1 Introduction

“*Form follows function*,” although originally a perennial maxim coined by architect Louis Sullivan in reference to pragmatic architectural design, it has been adopted by the biomedical engineering community in reference to nature and its adaptability (Sullivan, 1896; Russell et al., 2000). This phrase is often used in reference to natural materials, which have optimized their shape and structures over millennia of evolution and adapted to their specialized tasks (Wegst et al., 2014). While studies have investigated both *form* and its effect on *function* (Libonati and Buehler, 2017; Wang Y. et al., 2020), how it *follows* remains nebulous. In particular, the way in which the shapes of anatomical structures change over time has long interested the biomedical engineering community, dating back to and even predating the seminal works of Darwin (2009) and Thompson (1992). Modeling and predicting the evolving characteristics of anatomical geometry is relevant, with applications for clinical diagnoses, prognoses, and interventional treatments. Therefore, uncovering the underlying processes governing shape change of anatomical structures over time remains a highly relevant and developing domain of research.

Longitudinal changes in the shapes of anatomical structures are relevant in a myriad of clinical applications, especially for early diagnosis and disease prognosis (Figure 1). Developmental bone growth, for example, is a highly complex process wherein deficiencies or deviations from nominal standards could result in long-term health ramifications



(Parfitt et al., 2000; Weaver and Fuchs, 2014). Some examples of such disorders include but are not limited to developmental hip dysplasia, osteogenesis imperfecta, scoliosis, and clubfoot (Morcuende and Weinstein, 2003). Early diagnosis could enable non-surgical treatments. Therefore, accurate ways of quantifying normal development and identifying abnormal variations is paramount (Semler et al., 2019; Marzin and Cormier-Daire, 2020; Newsome et al., 2016). Another example is Alzheimer's disease (AD), one of the most common age-related neurodegenerative diseases (Scheltens et al., 2021). Commonly used techniques for early diagnosis of AD, such as neuropsychological tests, are unreliable and cerebrospinal-fluid biomarker measurements are intrusive and costly (Alberdi et al., 2016). In contrast, novel techniques examining structural brain changes from MRI can diagnose AD early and pre-symptomatically, while also informing future prognoses (Mueller et al., 2005; Pegueroles et al., 2016; Blinkouskaya and Weickenmeier, 2021). Yet another example is tumor growth, wherein growth rates and tumor sizes inform cancer severity and prognoses (Clark, 1991; Morikawa et al., 2011; Kuroishi et al., 1990). Thus, developing spatiotemporal growth

models of tumors has been a long-standing field of research, ranging from early simplified deterministic 1-D models to more complex probabilistic simulations (Adam, 1986; Jiang et al., 2005; Rejniak and Anderson, 2010; Gerlee, 2013; Benzekry et al., 2014). While not an exhaustive list, these examples demonstrate the wide-ranging applications and clinical relevance of developing robust spatiotemporal shape modeling tools and methodologies.

Early spatiotemporal shape modeling can be linked to morphometrics, wherein researchers attempted to analyze biological shape variation using statistical methods. Generally speaking, researchers analyzed variations of common anatomical landmarks across a population (Slice, 2007). These analyses examined variations in the coordinates of landmarks themselves, distances or relative angles between them, or metrics calculated from a combination thereof (Rohlf, 1990; Slice, 2005; Bookstein, 1982). In the femur, for example, measurements such as the whole femur length, diaphyseal length, subtrochanteric anteroposterior and mediolateral diameters, anteroposterior physeal angles, alpha angle, and vertical diameter of the femoral head are some of the measurements used to characterize femoral anatomy (Toogood

et al., 2008; Wescott, 2005; Rissech et al., 2008). These methods, however, are time-consuming and landmark placement can be unreliable, inconsistent, and fail to capture holistic spatial arrangement. Developments in computer vision and mathematical modeling tools have led to the development of computational anatomy (CA) (Miller et al., 1997; Grenander and Miller, 1998; Miller, 2004). Therein, the concept of shape manifolds and diffeomorphic transformations became central in describing anatomical shape variability over time. These tools have developed greatly in recent years. Furthermore, with the advent of deep learning (DL), novel methodologies have surfaced. Outlining available methodologies along with their strengths, weaknesses, and potential synergies is, thus, required.

In this review, we seek to highlight and discuss alternative techniques, methodologies, and tools used to model the changing shape of anatomical structures over time. For simplicity and due to the varied terminologies found in the literature, we use the terms spatiotemporal shape modeling and longitudinal shape models interchangeably. We also focus mainly on the techniques and tools themselves as opposed to their clinical applications. However, we highlight applications as necessary to enhance the descriptive value of the presented concepts. We also neglect exhaustive discussions on these methods' mathematical background and derivations, and instead refer the readers where necessary. We do, however, provide some further mathematical detail surrounding the discussed techniques in the accompanying supplementary document which is organized in parallel to the main text. Here, we refer to *shape* in both a geometric sense (i.e., a set of points in n -dimensional Euclidean space ($\mathbb{R}^n$) with defined connections) and also in the intuitive sense of a visual boundary defining an object of interest within an image. This is important as both definitions play a role in the differing techniques we explore. A relatively similar review was carried out by Harie et al. (2023), however they explicitly focused on growth modeling and mainly discussed DL-based generative networks. In contrast, this review focuses on shape change over time in general, thus encompassing both growth and alternative biological processes (e.g., degeneration). Furthermore, this review does not focus exclusively on DL-based methods and also covers alternatives. This review begins with a discussion on diffeomorphisms and large deformation diffeomorphic metric mapping (LDDMM) framework, the most common early framework for spatiotemporal shape modeling. Then, we discuss deep learning-based tools, focusing on autoencoders, generative adversarial networks, recurrent neural networks, transformers, and diffusion models. Finally, we discuss the strengths and drawbacks of each tool generally, highlighting similarities and potential synergies. We also speculate on potential future outlooks and directions for research into spatiotemporal shape modeling.

2 Large deformation diffeomorphic metric mapping

Of the many ways to describe variations in shapes in biology, a longstanding idea was first proposed by Thompson in his influential work “*On Growth and Form*” in 1917 (Thompson, 1992). Therein, he argued that variations in the shapes of biological organisms can be best described by geometrical transformations.

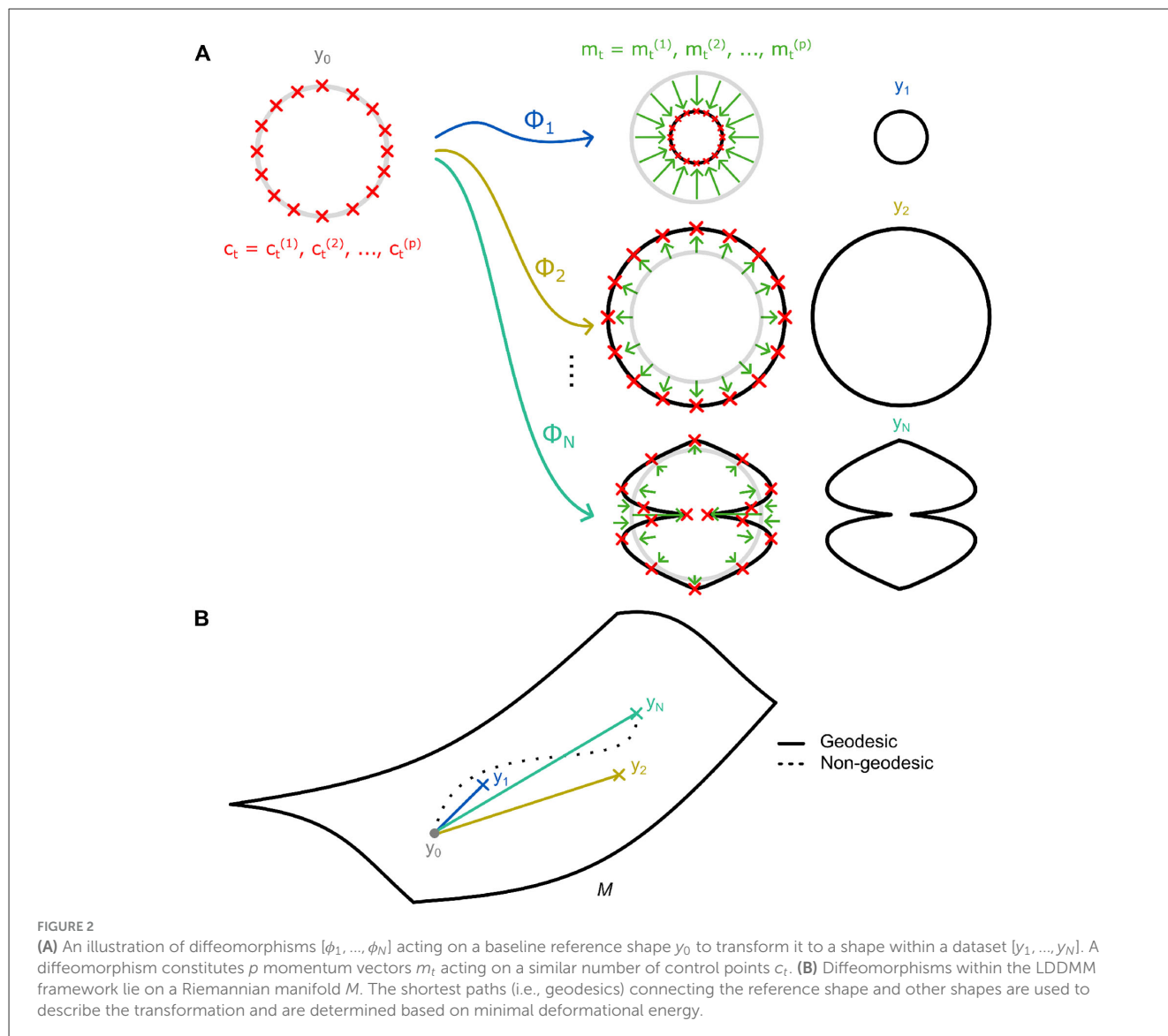
This pioneering theory formed the basis for CA decades later with the development of computer vision and mathematical tools. In essence, CA assumes that individual shapes are described as diffeomorphic transformations of an underlying reference shape. As, in principle, an infinite number of diffeomorphisms can act on a reference shape, sets of diffeomorphisms can then be considered as an infinite dimensional manifold (Marsland and Sommer, 2020). Accordingly, all the possible variations of a given shape can be represented within these manifolds, which are termed as “shape spaces” (Kendall, 1984; Monteiro et al., 2000; Rohlf, 2000). These manifolds can then be enriched with Riemannian metrics which enable quantitative comparison of these shapes and further mathematical operations (Younes, 2010; Miller et al., 2002; Miller, 2004). This constitutes the basis for large deformation diffeomorphic metric mapping (LDDMM) framework (Glaunès et al., 2008; Durrleman et al., 2014). Wherein, variations of anatomical shape in a population are described *via* diffeomorphisms acting on an underlying reference template. These diffeomorphisms then make up the shape space, an infinite dimensional Riemannian manifold describing all possible variations of a shape in a population. The LDDMM framework can then be extended further for longitudinal shape modeling as we will discuss.

2.1 Geodesics

For an initial reference shape y_0 and target shape y_1 , a diffeomorphism ϕ_1 exists which can be applied to transform the former to the latter (Figure 2A). Following the convention of Bône et al. (2020a), we denote this as $y_1 = \phi_1 \star y_0$. These diffeomorphisms ϕ_t can be difficult to obtain and describe, especially for complex shapes and deformations. Nevertheless, Miller et al. (2006) demonstrated that these complex deformations could be succinctly described by utilizing the principle of conservation of momentum (Vaillant et al., 2004). Specifically, by discretizing them as Gaussian convolutions g of p momentum vectors $m_t = m_t^{(1)}, \dots, m_t^{(p)} \in \mathbb{R}^d$ acting over a set of corresponding control points $c_t = c_t^{(1)}, \dots, c_t^{(p)} \in \mathbb{R}^d$ (Figure 2A). Nevertheless, solutions for ϕ_t are non-unique due to the infinite-dimensional nature of the underlying shape space manifold. Thus, the geodesic, that is the diffeomorphism requiring the least amount of deformational energy (Figure 2B), is utilized (Miller et al., 2002; Durrleman et al., 2014). Of note is that the geodesics' control points and momenta are also fully determined by their initial values (see Supplementary material S1.1 for further detail). This is particularly notable as, then, the system of initial momenta and control point locations $S_0 = \{c_0, m_0\}$, fully parametrize the entire flow of diffeomorphisms.

2.2 Geodesic regression

In representing each shape in a longitudinal dataset as a diffeomorphism of a template shape, the challenge remains in establishing the relationships between each shape. This is particularly important as deriving any underlying relationships between different shapes and independent variables (i.e., time)



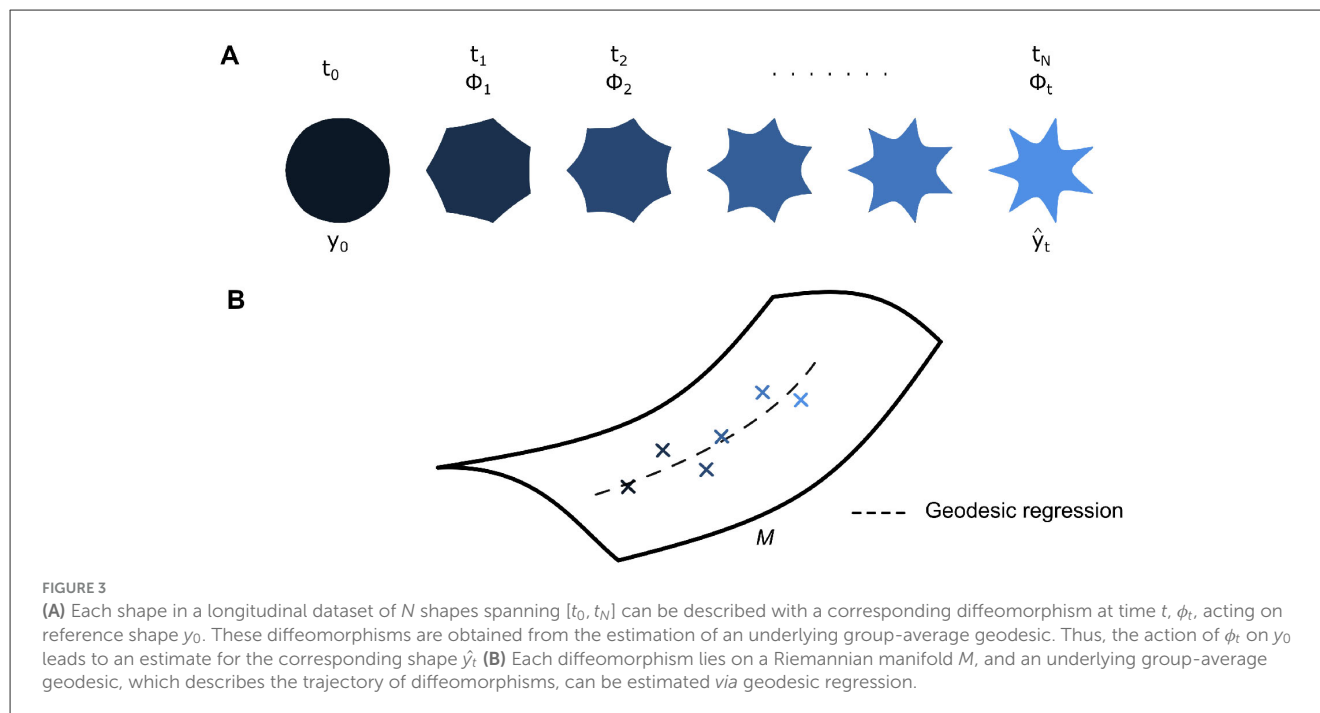
is essential for spatiotemporal shape modeling. Acquiring these relationships *via* standard regression techniques, for instance, is non-trivial due to the non-Euclidean structure of the Riemannian manifolds of diffeomorphisms. Nonetheless, Fletcher proposed an extension of standard linear regression to be applicable in a manifold-based setting, termed geodesic regression (Fletcher, 2011; Thomas Fletcher, 2012). Their technique was then developed further for a variety of applications, but the developments of Fishbaugh et al. (2013) and Fishbaugh et al. (2017) for use in longitudinal shape modeling are of particular interest for our purposes.

In detail, for a longitudinal dataset of shapes with N number of observations in the time range $[t_0, t_N]$, shape change over time is taken as a baseline shape y_0 being continuously deformed at each time point t by a corresponding diffeomorphism ϕ_t (Figure 3A). In principle, ϕ_t should lead to the baseline shape morphing to completely match the observed shape $y_t = \phi_t \star y_0$. However, in this context of estimating a holistic group-average geodesic (Figure 3B), we note that a diffeomorphism instead leads to an estimation of

the observed shape at time t instead ($\hat{y}_t = \phi_t \star y_0$). A regression criterion can then be expressed as follows (Equation 1) (Fishbaugh et al., 2013, 2017).

$$E(y_0, S_0) = \sum_{i=1}^N \frac{1}{2\gamma^2} \|(\phi_{t_i} \star y_0) - y_{t_i}\|^2 + L(S_0), \quad (1)$$

where, $L(S_0)$ represents a regularity term for the time-varying deformation, determined by the kinetic energy of the control points at S_0 (Supplementary Equation S3). γ^2 represents a term used to balance the importance between the data and regularity terms. Thus, given a dataset of longitudinal shapes, during the minimization of Equation 1 the baseline shape y_0 , initial control point locations c_0 , and initial momenta m_0 are the parameters which are estimated. This general form of geodesic shape regression was then developed further to incorporate both shape and image data based on a weighted joint optimization routine (Fishbaugh et al., 2014). Their multimodal approach demonstrated improved



performances as opposed to exclusive shape or image approaches. Nevertheless, optimization schemes to solve for the underlying geodesic regressions are computationally expensive, especially for large-scale image datasets. Recently, [Ding et al. \(2019\)](#) have also proposed methods to enhance their speed and effectiveness using DL. They demonstrated that the use of encoder-decoder networks with GPU acceleration could increase computation speeds, enabling the scaling up of studies toward larger datasets encompassing more subjects or longer timescales. Developments notwithstanding, these methods were limited to single subject regressions; geodesic regression up to this point has mainly captured spatiotemporal variability of only single subjects, thus these methods were extended further to capture populations and their intervariabilities.

2.3 Hierarchical models

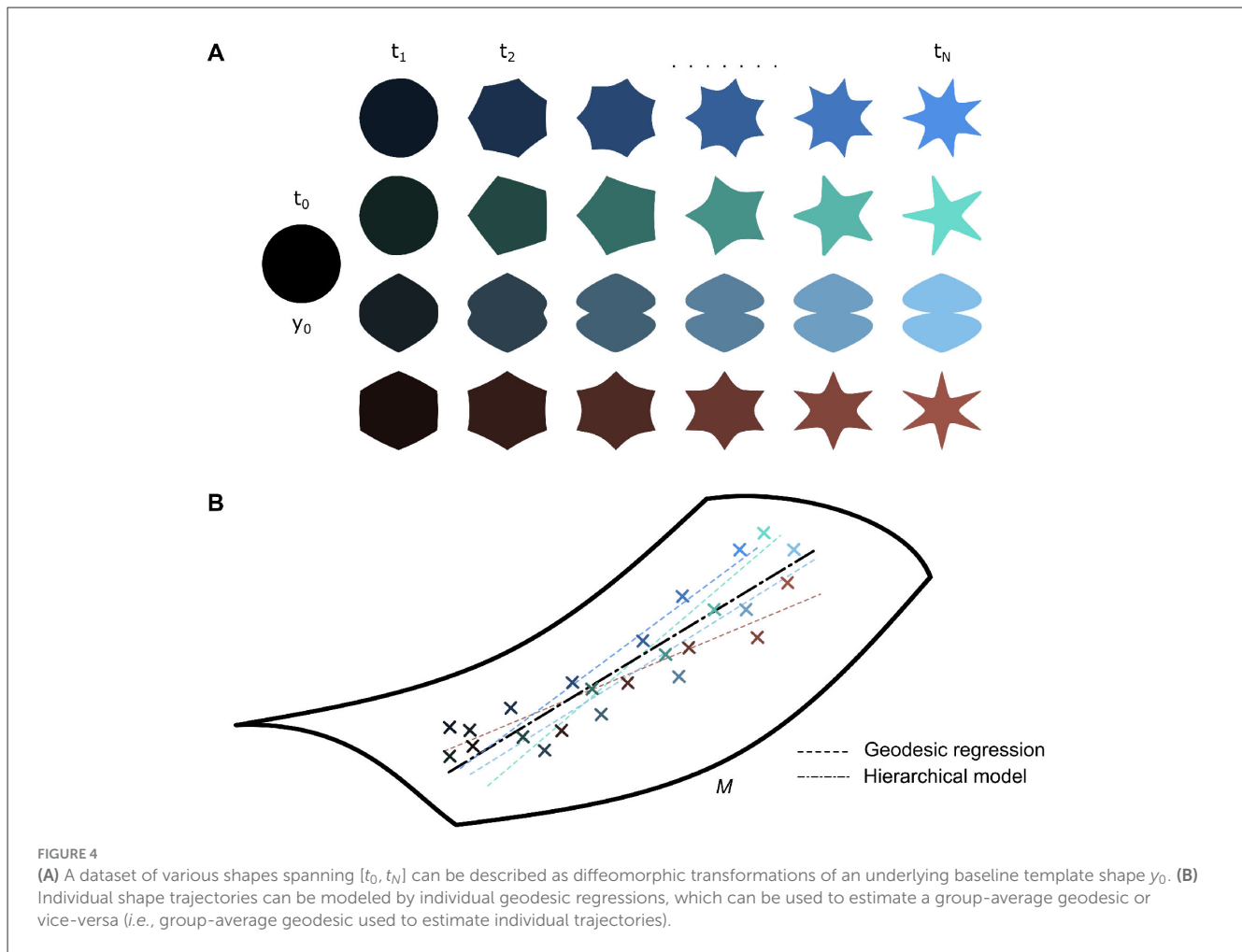
While geodesic regression can describe an object's longitudinal trajectory over time, it is insufficient to describe the longitudinal characteristics of multiple objects in a large dataset ([Figure 4A](#)). Thus, the LDDMM framework was further extended toward hierarchical models. Early work done by [Muralidharan and Fletcher \(2012\)](#) could estimate an underlying groupwise mean geodesic based on individual geodesics ([Figure 4B](#)). They did this with a least squares estimation of the underlying mean geodesic, using Sasaki metrics to compare individual trends. This was developed further by [Singh et al. \(2015\)](#) as a generalization of hierarchical linear models to a manifold-based setting. [Schiratti et al. \(2017\)](#) took a slightly different modeling approach, wherein they first found the underlying group average spatiotemporal trajectory and represented individual trajectories within the dataset as space and time transformations of this group-average. This approach

offers more flexibility as, unlike the former approach, it is not heavily dependent on initial time point choice, easing time reparametrization. [Bone et al. \(2018\)](#) and [Bône et al. \(2020a\)](#) developed this approach further for shape data within the LDDMM framework specifically.

Briefly, the hierarchical generative longitudinal models of [Bone et al. \(2018\)](#) and [Bône et al. \(2020a\)](#) rely on exp-parallelization ($\text{ExpP}_{\gamma}^{v_i}$) and a time warp function (ψ_i) to account for individual spatial and temporal differences respectively ([Equation 2](#)). Exp-parallelization essentially offers a tool to define parallel curves on a manifold whilst retaining the underlying structure ([Schiratti et al., 2015, 2017](#)). This enables us to define individual trajectories traversing the manifold as a variation of a group average. A time warp, on the other hand, accounts for the temporal characteristics of each individual's trajectory (*i.e.*, onset time, and rate of progression).

$$\begin{aligned} \text{ExpP}_{\gamma}^{v_i} [\psi_i(t_{i,j})] \star y_0 &\sim \mathcal{N}_{\epsilon}(y_{i,j}, \sigma_{\epsilon}^2) \\ \text{where } \psi_i: t &\rightarrow \alpha_i \cdot (t - \tau_i) + t_0 \\ |v_i = \text{Conv}(c_0, m_i), \quad m_i &= A_{0,m_0^{\perp}} \cdot s_i \end{aligned} \quad (2)$$

In turn, each component of the model ([Equation 2](#)) is as follows. A prediction for shape observation j of subject i , $y_{i,j}$ is modeled as a noisy estimate with variance σ_{ϵ}^2 . $y_{i,j}$ itself is predicted as a diffeomorphic transformation of a baseline reference shape y_0 transformed by an underlying group average geodesic γ space-shifted by exp-parallelization to match an individual's trajectory v_i . The time warp function ψ_i accounts for temporal characteristics, where α_i denotes progression rate, τ_i is onset time, and t_0 is the reference time. v_i accounts for the individuals' spatial variability and, in essence, is [Supplementary Equation S2](#) with some additional constraints. Namely, the momenta m_i are obtained from



a mixing matrix $A_{0,m_0^\perp}$ and q source parameters $s_i = s_i^{(1)}, \dots, s_i^{(q)}$. Note that the mixing matrix serves to project the source parameters into the higher dimensional momentum space (see Bone et al., 2018; Bône et al., 2020a for further detail). The parameters to be estimated which define individual trajectories are modeled as independent samples from normal distributions:

$$\begin{aligned} \alpha_i &\stackrel{\text{iid}}{\sim} \mathcal{N}_{[0,+\infty]}(1, \sigma_\alpha^2) \\ \tau_i &\stackrel{\text{iid}}{\sim} \mathcal{N}(t_0, \sigma_\tau^2) \\ s_i &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1) \end{aligned} \quad (3)$$

Taken together, a mixed effects model can be defined for the gathered parameters. Fixed effects, which account for parameters affecting the trajectories of all the subjects, can be denoted as $\theta = (\theta_1, \theta_2)$. Where, $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ and $\theta_2 = (y_0, c_0, m_0, A_0)$. The random effects z_i account for variations for each subject, where $z_i = (\alpha_i, \tau_i, s_i)$. This nonlinear multi-parameter optimization task is computationally complex and expensive and relies on a multi-step *calibration*, *personalization*, and *simulation* scheme detailed further in Bône et al. (2020a). In brief, it utilizes a novel Monte Carlo Markov Chains-Stochastic Approximation

Expectation Maximization-Gradient Descent (MCMC-SAEM-GD) algorithm detailed further in the reference.

2.4 Applications and further works

Overall, the use of hierarchical models provides us with a structured framework to characterize longitudinal data, both on an individual and group level. The use of a group-average trajectory enables us to quantify the variation of an individual's progression from a normative scenario (Kim et al., 2017). This also has the potential for prognostic benefits. For example, Cury et al. (2016) could detect shape changes in the thalamus of patients suffering from dementia 10 years prior to clinical symptoms by comparing healthy and diseased spatiotemporal trajectories. Bône et al. (2017) have demonstrated the use of exp-parallelization and time reparametrization to transport a population average trajectory onto new subjects. Thus, they demonstrated that population-average normative trajectories can be leveraged to predict trends in shape change or disease progression for new, unseen subjects. Similarly, Koval et al. (2018) implemented a manifold-based hierarchical model but in the context of graph networks. Specifically, they derived a population-based estimate for cortical atrophy dynamics

and demonstrated the capability to characterize patient-specific atrophy dynamics. They further extended this work to account for multimodal data such as biomarker levels and cognitive impairment scores to develop a comprehensive spatiotemporal atlas of Alzheimer's disease (Koval et al., 2021). This method of integrating the use of biomarkers (i.e., genetic and clinical factors) alongside imaging has gained traction and not only demonstrates soundness in and of itself (Dalca et al., 2015) but also has the potential to enhance the predictive capacity of existing frameworks with multimodality. Couronne et al. (2019) further demonstrated the efficacy of multimodal models in the context of Parkinson's disease prognosis. Utilizing both imaging and neurophysiological test score data, they demonstrated the robustness and efficacy of multimodality to improve predictive performance.

Nevertheless, the hierarchical model framework is still being developed further to refine its modeling efficacy and integrate newer technological developments. As opposed to modeling correlations along a manifold as quasi-linear in the manner of geodesics, Hanik et al. (2022) proposed to utilize generalized Bézier curves to model nonlinear relationships with the rationale that many biological processes are nonlinear (e.g., cardiac motion). Their initial work demonstrated the potential for extending this principle further and potentially decomposing longitudinal trends (i.e., disease progression) into different components of a nonlinear curve, enabling more granular analyses. Hong et al. (2019) also investigated the effects of subject-specific characteristics by including multivariate intercept models in their formulation of a hierarchical geodesic model. Debavelaere et al. (2020) developed a methodology to investigate datasets with heterogeneous populations (i.e., a dataset with diverging longitudinal dynamics). They developed an unsupervised algorithm that is able to detect clusters of subgroups within a dataset and differentiate their trajectories, accounting for diverging or converging trajectories from a population normal. Gaudferneau et al. (2023) also extended the LDDMM framework via multiscale representations of images and demonstrated improved results on fetal brain growth estimation, a comparatively more difficult task. Furthermore, the advent of DL has led to augmentations of the LDDMM framework due to its increased computational efficiency of processing large datasets (Yang et al., 2023; Ben Amor et al., 2023). Bône et al. (2019) demonstrated the use of autoencoders to learn an atlas and class of diffeomorphisms that describe a dataset of shapes and meshes. They further extended their work to also account for the texture (i.e., appearance) of images (Bône et al., 2020b). Pathan and Hong (2018) also demonstrated the potential of using DL to learn vector momenta utilized in the LDDMM framework. Other novel developments include the utilization of implicit neural representations (INRs) (Sitzmann et al., 2020). Dummer et al. (2023) demonstrated the potential of using INRs to extend the LDDMM framework toward increased robustness and resolution independence.

To surmise, the LDDMM framework is a powerful tool for representing and modeling a dataset of shapes. Assuming an underlying template shape, the LDDMM framework represents individual shapes as diffeomorphic transformations of this template. These diffeomorphisms lie on an infinite dimensional Riemannian manifold, thus relying on geodesic regression

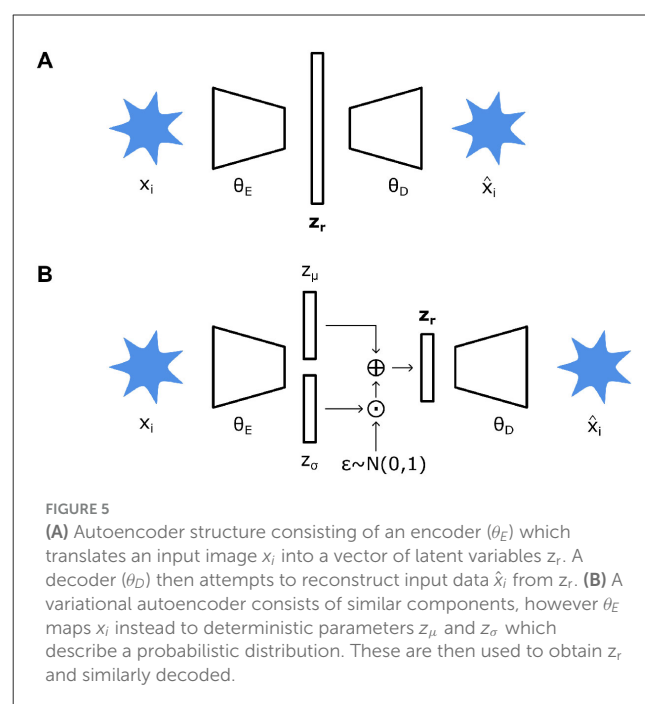
and parallel transport tools to estimate the longitudinal trajectories traversing the underlying data manifold. Hierarchical models can then be utilized to model differing spatiotemporal trajectories of a population, capable of estimating population average spatiotemporal trajectories and also quantifying intra and inter-individual differences. Whilst, in recent years, the proliferation of DL-based techniques has seemingly eclipsed LDDMM-based techniques, the framework is continuously developing. In fact, many of the developments seek to utilize DL tools to accelerate the framework and increase its efficacy. LDDMM methods are readily available in several software packages and applications such as Deformetrica (Bône et al., 2018), Leaspy (<https://leaspy.readthedocs.io/en/stable/>), and Morphomatics (Ambellan et al., 2021).

3 Deep learning

In medical imaging, DL-based solutions have pushed the state-of-the-art further for a variety of tasks. From image segmentation, disease diagnosis, and prognosis to synthetic image synthesis, DL represents a powerful paradigm for the future of medical image analysis (Shen et al., 2017; Wang T. et al., 2020). In this section, we highlight alternative network architectures that have been utilized for spatiotemporal shape modeling.

3.1 Autoencoders

Autoencoders (AEs) are a neural network (NN) architecture consisting of an encoder and decoder module (Figure 5A). This architecture, in principle, seeks to compress data to a low-dimensional latent space, reducing them to r number of latent



variables, $\mathbf{z}_r$. These latent variables themselves can then be utilized for other tasks as they represent, in essence, a compressed low-dimensional representation of higher-dimensional data. Thus, the weights of the encoder θ_E and decoder θ_D modules are learned to accurately de-construct input data down into a latent representation and re-construct them into the original input data, respectively (Lopez Pinaya et al., 2020). The objective when training an AE is then to minimize the loss function $\mathcal{L}^{rec}$, which takes the form of a dissimilarity function or reconstruction loss, to find θ_E and θ_D (Equation 4). Details on the loss function and structuring of regularization can be found in the [Supplementary material S2.1](#).

$$\min_{\theta_E, \theta_D} \mathcal{L}(\theta_E, \theta_D) = \min_{\theta_E, \theta_D} \sum_{i=1}^N \mathcal{L}^{rec}(x_i, \hat{x}_i) \quad (4)$$

where $\hat{x}_i = \theta_D(\theta_E(x_i))$

A variation of AEs is variational autoencoders (VAEs) which are similar but treat encoding and decoding in a probabilistic manner (Figure 5B) (Rezende et al., 2014). Instead of directly mapping input data to latent variables, VAEs map input data to probabilistic distributions of their corresponding latent variables. Briefly, θ_E maps input data to deterministic parameters, mean $z_\mu(x)$ and standard deviation $z_\sigma(x)$, which describe an underlying probabilistic distribution (usually Gaussian) of the latent space. These deterministic parameters are then injected with stochasticity sampled from a fixed normal distribution, where $\odot$ denotes a Hadamard product (Equation 5). This configuration is necessary to preserve the stochasticity within the latent space while enabling gradient-based backpropagation during training (Ehrhardt and Wilms, 2022). In turn, the loss function (Equation 4) is now modified to consider both reconstruction quality and regularity of the latent space (Equation 6). The latter is usually represented by a Kullback-Leibler divergence, detailed elsewhere (Ehrhardt and Wilms, 2022). Overall, a probabilistic treatment of latent variables and the spaces they inhabit leads to more structured, compact, and continuous latent spaces. This, in turn, leads to a smoother sampling of latent variables for generative processes and representation learning in general.

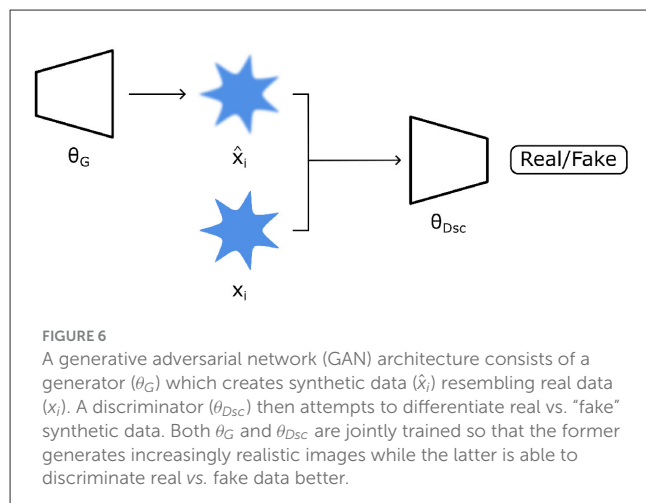
$$z = z_\mu(x) + z_\sigma(x) \odot \epsilon \quad \text{with } \epsilon \sim \mathcal{N}(0, 1) \quad (5)$$

$$\mathcal{L}^{VAE} = \mathcal{L}^{rec} + \mathcal{L}^{KL} \quad (6)$$

AEs, VAEs, and variations thereof have many applications in generative frameworks and tasks involving reduced dimension representations of high dimensional data such as images. The strengths of these architectures in modeling complex data within low-dimensional representations could lend themselves well to capturing the complex nonlinearities inherent in longitudinal datasets. Latent variables and the spaces they inhabit have been utilized as parameters to be fit to existing models. Sauty and Durrleman (2022) utilized a VAE to learn latent variables representing images within a longitudinal dataset. These latent variables are then fitted to a linear longitudinal mixed-effects

progression model similar to those of the LDDMM framework. Chadebec and Allasonnière (2023) utilized normalizing flows to model latent variables representing spatiotemporal data, thus imposing temporal structure onto the latent space. Kapoor et al. (2025) presented MRExtrap, a framework wherein they utilize linear models to model latent variables extracted from a regularized AE. Their framework successfully estimated longitudinal trajectories via a progression rate variable, from a single scan, based on population and subject-specific priors which can be updated dynamically with new data. Nevertheless, as latent variables lack any specific underlying physical meaning, developments in techniques to identify what these variables represent have also been made. Mouches et al. (2021) utilizes an invertible latent space disentanglement module within an autoencoder framework to determine latent variables that affect age-related changes. Isolated age-related latent variables can then be varied, with age-unrelated components kept constant, to simulate the aging of a particular individual. Following a similar vein, Zhao et al. (2021) utilized a cosine-based loss function to disentangle brain age from image representation. They did so with a self-supervised learning methodology, optimizing the correspondence between the “directionality” of latent variables in the latent space and physical developmental trajectories. As opposed to fitting models to latent variables themselves, structuring the latent space during training via conditional priors or regularization is a common and effective technique. He et al. (2025) developed a conditional VAE architecture capable of predicting follow-up MRI scans of the human brain. Ong et al. (2024) incorporated the use of linear mixed models as conditional priors on the latent space of VAEs. Chen et al. demonstrated the use of orthogonality mixed-effects constraints to structure the latent space of an autoencoder. Their method could robustly identify both global and local longitudinal trajectories, with enhanced classification outcomes (Chen et al., 2025).

Overall, AEs and VAEs represent powerful tools for reducing high-dimensional data into a low-dimensional latent space, efficiently encapsulating longitudinal data into compressed latent variables. Nevertheless, latent variables and their spaces are solely reduced dimension representations of the original input data (Ehrhardt and Wilms, 2022); latent variables have no underlying physical meaning. For example, the distribution of latent variables has been shown to be affected by training parameters, demonstrating their capricious nature (Lapenda et al., 2020). Thus, latent variables cannot be considered spatiotemporal variables. However, rational structuring and regularization to ensure that latent spaces are enriched with physical meaning can lead to better outcomes. Another point of concern is that both AEs and VAEs generally treat latent spaces and variables in a Euclidean manner, when in fact research has shown that a manifold-based approaches may be more prudent (Connor et al., 2021). These problems remain active fields of research, with solutions such as regularization and explicitly structuring latent spaces deterministically being continually developed (Tschannen et al., 2018; Ghosh et al., 2020). Nonetheless, existing works for spatiotemporal shape modeling demonstrated the potential applicability for autoencoders and learned latent variables to model longitudinal trajectories.



3.2 Generative adversarial networks

Generative Adversarial Networks (GANs) are neural network architectures first proposed by Goodfellow et al. (2014). In principle, they consist of generator θ_G and discriminator θ_{Dsc} networks being trained simultaneously (Figure 6). Therein, the former is trained to create new synthetic images, whilst the latter is trained to detect if an image is real or fake. In detail, θ_G maps random input variables v (sampled from a prior distribution $p(v)$) to the data space in an attempt to generate data $\hat{x}_G$ resembling data from a real dataset x_r . In turn, both types of data are fed into θ_{Dsc} , whereby θ_{Dsc} is trained to determine if data is real or fake. The objective function used to train both networks simultaneously is then a minimax problem (Equation 7).

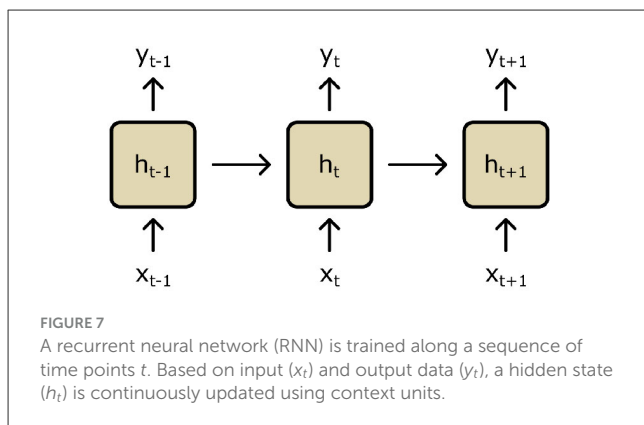
$$\min_{\theta_G} \max_{\theta_{Dsc}} V(\theta_G, \theta_{Dsc}) = E_{x \sim p_{data}(x_r)} [\log D(x)] + E_{z \sim p(v)} [\log(1 - D(G(v)))] \quad (7)$$

This architecture is very powerful, as the adversarial training configuration leads to the generator module being capable of generating realistic synthetic images that are indistinguishable from real data (Wang K. et al., 2017; Gui et al., 2023). Trained generators are then useful for many applications. In the context of medical imaging, examples include image synthesis, segmentation, and classification, among others (Yi et al., 2019). Many variants exist, and more are continually being developed, for which the reader is referred to other papers for further details (Jabbar et al., 2021). In the context of spatiotemporal shape modeling, GANs represent a powerful tool. Similar to previously discussed AEs and VAEs, their generative capacity can potentially be utilized to capture the underlying spatiotemporal trajectories.

Elazab et al. (2020) used a stack of 3D GANs to predict brain tumor growth. Specifically, with an input image and physiological feature maps, a generator predicted a brain scan at the proceeding time point whose accuracy was evaluated by a discriminator. Their results outperformed contemporary methods but relied on stacking and training consecutive GANs, which is computationally inefficient. Alternatively, Zhang et al. (2017) utilized GANs to uncover the underlying data manifold of

longitudinal progression for face aging. They first encode images to latent vectors, which are concatenated with age-related feature vectors and then mapped onto a manifold. Discriminators ensure regularized latent vector generation and image realism of the generators. Based on this, Ravi et al. (2019) developed a 2D framework to model age-related brain degeneration in the context of Alzheimer's diagnosis. They incorporated further voxel-based and region-level constraints which acted as biological constraints to model Alzheimer's progression, leading to improved prognoses. They developed this work further to examine 3D MRIs for a more holistic view of the brain (Ravi et al., 2022). Utilizing a 3D training consistency mechanism and a super-resolution module led to a full 4D model of brain aging without a loss in anatomical detail. Following the same principle of temporal embedding within a latent space, Schön et al. (2023) similarly implemented a GAN-based network for embedding temporal directionality in generators. Alternative GAN architectures have also been investigated. Wasserstein GANs (WGANs) utilize Wasserstein distances as a loss function as opposed to regularly used Jensen-Shannon divergence (Arjovsky et al., 2017). This architecture leads to more stable training outcomes and was utilized by Wegmayr et al. (2019) as a recursive generator model to predict time steps in brain aging. Combined with a classifier network, they present a framework for both predicting aged brain images and Alzheimer's prognosis, outperforming standard methods. In StyleGAN and derivatives thereof, the principle of style transfer and additional, intermediate latent spaces is utilized to improve generator architectures and disentangle latent space components and their effects on synthesized images (Karras et al., 2019, 2020; Fetti et al., 2020). Han et al. (2022) developed a framework for image-based osteoarthritis prognosis using StyleGAN as the generative architecture. This enabled them to construct the underlying manifold of longitudinal knee aging, and furthermore, they demonstrated that their model outperforms human radiologists in early diagnosis of osteoarthritis. Similarly, Gadewar et al. (2023b) and Gadewar et al. (2023a) utilized StarGAN-v2, a similar style-based generator architecture, to predict aging in structural MRIs of the brain.

In short, GAN-based architectures and adversarial training represent powerful tools for spatiotemporal shape modeling. In particular, discriminators support the structuring and regularization processes so that the latent space of generator modules is physically meaningful, similar to previously discussed regularized AEs. While GANs have their own challenges in terms of training stability, mode collapse, convergence, and image fidelity, continual developments in training schemes, architectures, and loss functions have led to continuous improvements (Yi et al., 2019; Gui et al., 2023; Saxena and Cao, 2021). Generators with well-defined and structured latent spaces, and rational generative processes enable us to predict growth trajectories. Said structuring of latent spaces is facilitated by discriminators and loss functions, which allow us to ensure smooth latent spaces that are temporally consistent and valid. In essence, in helping structure latent spaces, discriminators implicitly define the underlying manifold of spatiotemporal shape progression. Similar to previously discussed AEs, this structuring process ensures that latent spaces and variables therein can be endowed with meaningful physical characteristics.



3.3 Recurrent neural networks

Recurrent neural networks (RNNs) are a type of NN that are used to model sequential data such as a time series (Lipton et al., 2015; Salehinejad et al., 2018; Staudemeyer and Morris, 2019). They do so by considering data along a whole sequence's trajectory during training and inference; RNNs are designed explicitly with features that connect and consider data inputs across longitudinal sequences by maintaining memory (i.e., a hidden internal state h_t which is continually updated at each time point t). Early RNNs utilized simple "context units," which are units independently connected to nodes in the hidden layer of an NN (Figure 7) (Elman, 1990). These context units are then updated along steps in a data sequence *via* activation functions as the RNN is trained along a sequence. These simple context units were then developed to more complex long short-term memory (LSTM) cells to address practicalities surrounding network training (further details in Supplementary material S2.2).

Thus far, LSTMs have been used for natural language processing or other tasks examining relatively low-dimensional data. In the context of images and CNNs, LSTMs have been adapted for image inputs in the form of the convolutional LSTM (ConvLSTM) (Shi et al., 2015). ConvLSTM is able to capture temporal information and dependencies in a sequence of images while ensuring that spatial information is preserved during encoding. This has led to its use and marked effectiveness in video prediction tasks (Lotter et al., 2016; Lu et al., 2017).

In the context of longitudinal medical imaging, RNNs have improved the outcomes of segmentation (Gao Y. et al., 2018) and disease stage classification tasks (Santeramo et al., 2018; Gao L. et al., 2018; Cui and Liu, 2019; Ouyang et al., 2021; Ding et al., 2023). In explicitly modeling shape change using ConvLSTMs and its derivatives, however, RNNs have seen comparatively less uptake potentially due to the significantly high GPU memory requirements (Ma et al., 2022). Some studies nevertheless utilize RNNs as components within larger frameworks to avoid this obstacle. For example, Pathan and Hong used LSTMs to predict the vector momentum sequences to deform a longitudinal baseline image in an LDDMM framework (Pathan and Hong, 2018). This approach leverages the effectiveness of the LDDMM framework to predict changes over time without loss of detail and the computational efficiency of DL. Louis et al. (2019) utilized RNNs

to encode longitudinal trajectories into a latent space. These encoded trajectories are then decoded to construct the manifold and the Riemannian metrics lying on this manifold. Ma M. et al. (2023) utilized ConvLSTMs alongside a transformer in a "growth prediction module" to predict tumor growth. They demonstrated that utilizing both components in a unified module leads to better-predicted growth morphologies. Zhang et al. (2020) extended the ConvLSTM framework with the goal of modeling spatiotemporal sequences (ST-ConvLSTM). Their ST-ConvLSTM units learn both temporal and spatial dependencies in a sequence; for a 3D image slice, ST-ConvLSTM learns both the changes over time for that slice and accounts for the adjacent slices.

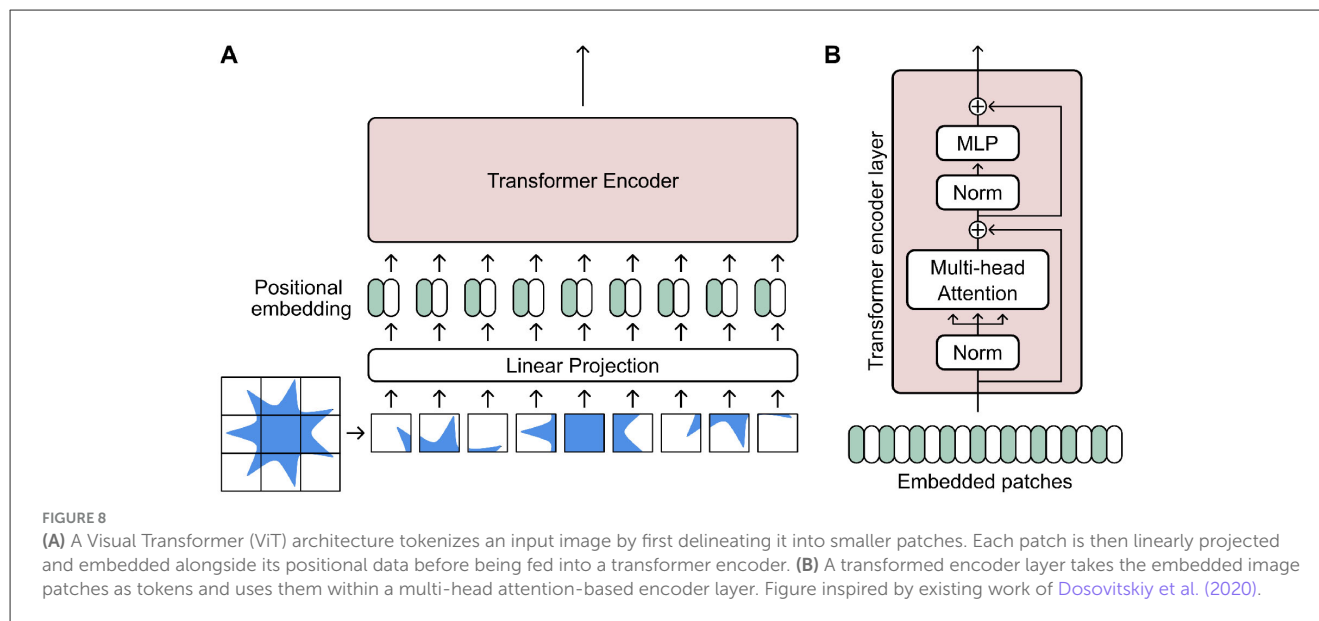
To surmise, RNNs represent a powerful network architecture for capturing temporal dependencies within a longitudinal dataset. Nevertheless, the issue of high GPU memory requirements for imaging data persists. This particular requirement precludes the use of RNNs for longitudinal shape modeling. Nevertheless, Ma et al. (2022) and Ma Z. et al. (2023) sought to address this by developing multi-scale RNN frameworks, which demonstrably improve performance with much lower GPU memory costs. Chen et al. (2024) demonstrated the use of signed distance function-based representations with ConvLSTMs to predict longitudinal changes in the shape of vestibular schwannoma. They demonstrated a proof of concept for using signed distance functions, which could address issues of large memory requirements of conventional ConvLSTMs operating directly on images. All in all, developments in using LSTMs for medical imaging datasets are relatively recent and have yet to be fully investigated in the context of longitudinal medical image shape modeling.

3.4 Transformers

Transformers are a relatively recent development in DL. Originally designed for natural language processing (NLP) tasks (Vaswani et al., 2017), they utilize a novel attention mechanism based on saliency, which can capture long-range dependencies in data sequences. The architecture was later developed further specifically for image data with the Vision Transformer (ViT) framework (Dosovitskiy et al., 2020). In any case, transformer networks rely firstly on tokenization of input data (Figure 8) (Shamshad et al., 2023; Islam et al., 2024; Torralba and Isola, 2024). This process essentially entails subdividing input data into "tokens," wherein each token is passed to an attention module where they can be used to calculate an attention score (further details in Supplementary material S2.3)

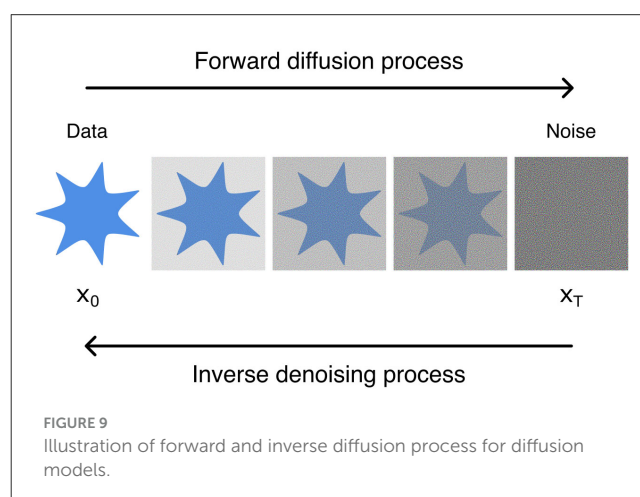
These tokenized representations and attention modules are then integrated into various NN architectures and can be configured for many applications, especially in medical image analysis (Azad et al., 2024). In particular, the capability to capture long-range dependencies and focus on salient features across long input sequences could potentially be applicable for predicting shape changes over time sequences.

In the context of longitudinal shape modeling, the use of transformers are still relatively unexplored. Sarasua et al. (2021) was one of the first to apply transformers to model longitudinal shape trajectories. They forecasted the change in the shape of



meshes of the left hippocampus in an encoder-decoder-style architecture utilizing a bidirectional transformer encoder. They extended this work further by explicitly embedding Alzheimer's cognitive impairment scores and utilizing pre-trained transformers ([Sarasua et al., 2022](#)). The latter method revolved around freezing most layers of a pre-trained transformer and fine-tuning it on a selected task to decrease the number of trainable parameters ([Lu et al., 2021](#)). The former method of embedding cognitive scores was also similarly utilized by [Xia et al. \(2021\)](#) to synthesize longitudinal brain images. With an input baseline brain image, their transformer architecture embeds a health state and age progression to synthesize changes over time. To improve the quality of their predicted progressions, they trained their networks in an adversarial manner with additional loss functions to preserve subject identity. [Wang et al. \(2022\)](#) developed a comprehensive transformer-based framework to predict tumor growth. Their so-called static-dynamic framework utilizes a transformer-based module to first encode and enhance high-level features of detected tumors. Then, a transformer-based growth estimation module is employed to predict growth based on the aforementioned extracted features.

Nevertheless, applications of transformers for longitudinal shape modeling is still in its relative infancy. Advances in transformer architectures, such as incorporating multi-scale convolutions for enhanced time-series prediction, could potentially be applied to imaging data as well ([Wang and Guan, 2023](#)). However, there are a number of caveats to the enhanced performance of transformer-based networks ([Li et al., 2023](#)). Firstly, the nature of the transformer architecture leads to lower degrees of inductive bias, necessitating larger amounts of training data for better performance. This could potentially be addressed with pre-training as demonstrated by [Lu et al. \(2021\)](#), but nevertheless remains a consideration. Furthermore, training transformer architectures is computationally expensive, requiring significant computing resources, especially if applied to 3D volumetric medical imaging. In fact, a relatively high number



of studies in the field are focused on reducing this computational burden ([Xia and Wang, 2023](#)). This heightened computational resources required thus present a barrier, prohibiting widespread development and applications to new data. Early studies have already demonstrated promising results, and transformer architectures could present a future avenue for spatiotemporal shape modeling.

3.5 Diffusion models

Diffusion models (DMs) are a type of generative DL architecture similar to aforementioned GANs and AEs. In contrast, however, DMs function on the principle of noise addition and removal ([Fuest et al., 2024](#); [Croitoru et al., 2023](#); [Kazerouni et al., 2023](#)); DMs consist of forward and inverse processes, wherein noise is added onto input data in successive steps, and the resulting noise is reversed to reform the input data ([Ho et al., 2020](#)) ([Figure 9](#)).

These processes are Markovian in nature, and the forward process is generally handcrafted (i.e., manually chosen or optimized for). The inverse process, however, is what is learned by the network. In detail, the process is a Markov chain which starts from a data distribution $q(x_0)$ and a sequence of T steps corrupting it to $\mathcal{N}(0, I)$, a Gaussian distribution, with Markov diffusion kernels $q(x|x_{t-1})$ (Equation 8).

$$q(x_1, x_2, \dots, x_T | x_0) = \prod_{t=1}^T q(x_t | x_{t-1})$$

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (8)$$

Where β_t is the variance of noise and I is the identity matrix. The reverse denoising process is what is learned by the model, that is inverting the diffusion process and turning the latent noise variable $p_\theta(x_t)$ back into the data distribution $p_\theta(x_0)$ parameterized by θ (Equation 9).

$$p(x_0, x_1, \dots, x_{T-1} | x_T) = \prod_{t=1}^T p_\theta(x_{t-1} | x_t)$$

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_0(x_t, t), \sigma_0^2(x_t, t) I)$$

$$p(x_t) = \mathcal{N}(x_t; 0, I) \quad (9)$$

Where, $\sigma_0^2(x_t, t)$ is variance at step t and $\mu_0(x_t, t)$ is the mean of the Gaussian distribution. Thus, from a randomly sampled noise vector, novel samples can be generated (for further mathematical detail, readers are referred to Fuest et al., 2024; Croitoru et al., 2023; Kazerouni et al., 2023).

DMs have led to state-of-the-art high resolution visual generative networks (Rombach et al., 2021; Dhariwal and Nichol, 2021), and existing works have demonstrated the potential for use in a variety of spatiotemporal modeling tasks (Yang et al., 2024; Rühling Cachay et al., 2023). Yoon et al. (2023) demonstrated the use of a sequence-aware diffusion model (SADM) to generate longitudinal medical images. Their framework utilized a sequence-aware transformer as the conditional module for a diffusion model, demonstrating effective data generation capabilities for longitudinal 3D medical imaging sequences, even with missing data. Litrico et al. (2025) utilized patient metadata and age gaps to condition their diffusion model, demonstrating effective results. Lozupone et al. (2025) took an alternative approach, wherein they applied the diffusion process to a compressed latent representation of their images as opposed to the images themselves. This approach enables computationally efficient processing of 3D medical imaging and demonstrated impressive results on disease classification and temporal trajectory prediction. Puglisi et al. (2025) followed a similar approach in their Brain Latent Progression (BrLP) framework. Alongside training a DM to operate on latent variables, they further condition it with additional patient metadata and anatomical measures. Operating within their wider framework, they demonstrate state-of-the-art results in trajectory prediction.

While a promising avenue of development, DMs are still a developing field with many deficiencies to be addressed. Namely, a main issue is with computational efficiency (Croitoru et al., 2023; Guo et al., 2023). The multi-step noising and denoising

processes takes more computational time and resources compared to other generative networks, precluding potential real-time diagnostic applications. Training these networks also requires considerable computational resources which potentially surpass the requirements of alternative networks. Whilst developments in improving efficiency of DMs have been made, it is still a growing field of interest (Shen et al., 2025). Furthermore, to ensure the validity of generated data, auxiliary networks such as ControlNet and variants thereof have been developed to enhance the tractability of DMs' generative processes (Zhang et al., 2023; Yang et al., 2025). Integration of DMs into wider frameworks such as those incorporating attention-based mechanisms has also demonstrated to be effective, and further developments could also be promising (Wu and Gong, 2024).

3.6 Limitations of deep learning approaches

Although the presented DL techniques have led to great strides forward in the state-of-the-art of spatiotemporal shape modeling, several challenges persist. These challenges hinder the widespread adoption of DL both in general and medical image analysis specifically. Thus, practitioners should be aware that DL cannot be simply considered a panacea for their tasks.

Firstly, DL models are inherently “black boxes” (Castelvecchi, 2016): while their outputs might be accurate, valid, and valuable according to many objective metrics, the inscrutability of how these outputs arise is a predominant concern. The opaque nature of how trained models arrive at their solutions engender doubts regarding trustworthiness due to the lack of explainability (Li et al., 2022; Xu and Yang, 2025). Especially for medical applications, interpretability and understanding is paramount. In modeling the progression of diseases for example, both for observational and prognostic applications, developing an understanding of the phenomena being modeled can be more important than the final model output itself (Young et al., 2024). Thus, efforts have been made to increase the explainability of these models and increase levels of trust with clinical end users (Singh et al., 2020). For example, a popular technique for image classification is via class activation maps (Teng et al., 2022). This visualization technique essentially highlights the specific discriminative image regions which influence final classification output. Many alternative methods exist such as utilizing accompanying language models to provide elaborative textual explanations (i.e., captions) describing visual results (Patrício et al., 2023). Schutte et al. (2021) also demonstrated the use of StyleGAN to generate counterfactual images as an alternative illustrative means of increasing interpretability.

Another concern is the implicit biases embedded into trained DL models. As DL models generally have no underlying physical grounding, their overall performance is entirely dependent on the quality of input training data and susceptible to biases or errors in the data itself (the garbage in, garbage out principle) (Geiger et al., 2021). This is related to the disproportionate prevalence of White, Educated, Industrialized, Rich, and Democratic (WEIRD) datasets in the field of behavioral sciences (Henrich et al.,

2010). This is an issue given that while WEIRD populations do not represent the global norm, they are overrepresented in academic research. This is also an underlying problem with DL and available training data. Septiandri *et al.* found that a majority of datasets utilized by researchers at two AI-focused conferences were WEIRD. This could risk under-representing less privileged populations, impeding equal availability of state-of-the-art models or even leading to harmful outcomes (Mihalcea *et al.*, 2025). For example, Puyol-Antón *et al.* (2021) and Puyol-Antón *et al.* (2022) demonstrated clear racial biases in segmentation models, attributed to training data composition. Nevertheless, they demonstrated the use of several alternative strategies to address this bias such as training separate population-specific models and ensuring population-balanced training data. Regardless of specific strategies, practitioners should be cognizant of this issue for all DL applications, including spatiotemporal shape modeling, and actively work to address and mitigate these biases.

Domain shift is yet another significant barrier to widespread adoption of DL. Simply put, domain shift refers to dissimilar training and target datasets of a DL model, leading to a lack of generalizability (Guan and Liu, 2022). Especially for medical image analysis, domain shift is an underlying issue and arises from several issues. For example, MRI and X-ray data gathered from different centers examining similar structures can exhibit differences due to differing scanners or acquisition protocols. These differences, while seemingly negligible, greatly degrade downstream performance on trained models (Guo *et al.*, 2024; Pooch *et al.*, 2020). In histopathological data as well, data acquired from different scanners or subject to different pre-processing steps also exhibit degraded downstream performance (Stacke *et al.*, 2019). Nevertheless, addressing this issue (i.e., domain adaptation) remains an active field of research and an underlying consideration to develop robust DL models (Guan and Liu, 2022; Singhal *et al.*, 2023).

Lastly, an unavoidable consideration for many practitioners is the resource requirements for DL. State-of-the-art networks are continually growing in size and complexity, and require an unsustainably increasing amount of compute resources to train (Thompson *et al.*, 2020). Whilst algorithmic improvements to decrease these compute costs for training are being developed, this remains a problem for widespread adoption (Bartoldson *et al.*, 2023). End users of trained models could also encounter high computational costs for model inference, depending on their size and complexity. Higher compute requirements also generally translate to higher monetary costs to access said resources. Similarly, methods are being developed to address efficiency from a monetary as opposed to a compute standpoint (Klemetti *et al.*, 2023). Nevertheless, these requirements could present a simple but significant barrier to resource-limited practitioners seeking to develop or train models further.

4 Discussion and conclusions

Several approaches for spatiotemporal shape modeling of anatomical structures were discussed in this review. Rapid developments in the field, especially in recent years, have been fueled by advancements in DL and are set to only continually

progress further. Nevertheless, the works found in the existing literature have been mainly focused on incremental developments in methodology or applications of novel new tools. This is in contrast with applying already developed tools to existing or novel clinical challenges. This seeming reluctance of the medical imaging community toward application-based research could stem from a multitude of reasons, but a simple lack of data could be the main factor, as we will discuss shortly. Deficiencies notwithstanding, in this section, we will discuss key concepts of spatiotemporal shape modeling uncovered from our review. We will then outline several key barriers to further research in the field before speculating on future research directions.

4.1 Nonlinear shape manifolds

From our review, it is clear that anatomical shape variation is highly nonlinear. This nonlinearity is further compounded by the additional nonlinear dynamics of growth and changing biological structures over time, leading to an intricate and complex outlook. Thus, the best-suited models for spatiotemporal progression are those that lie on non-Euclidean manifolds as they best capture this inherently high dimensional problem. A potential reason for this could be the manifold hypothesis, wherein it is postulated that all high-dimensional data lie on an embedded low-dimensional manifold (Fefferman *et al.*, 2016; Narayanan and Mitter, 2010). The task of spatiotemporal shape modeling can then be reduced to identifying and characterizing these manifolds, either implicitly or explicitly. The LDDMM framework discussed in Section 2, for example, explicitly seeks to uncover spatiotemporal trajectories of diffeomorphisms traversing across a manifold. DL techniques discussed in Section 3 also implicitly benefit from manifolds, as the efficacy of DL techniques has been attributed to their capability to uncover and disentangle underlying the manifolds of complex data (Brahma *et al.*, 2016).

In contrast to manifold-based techniques, several works do exist that have attempted to extend linear (PCA-based) statistical shape models toward spatiotemporal shape models. These, however, fall short when compared to LDDMM and DL-based solutions as they effectively only serve to compare differences across and interpolate between time points as opposed to true longitudinal forecasting (Hamarneh and Gustavsson, 2004; Kasahara *et al.*, 2018; Binte Alam *et al.*, 2020; Saito *et al.*, 2019). Due to their reliance on landmarks, these methods do not effectively work if anatomies significantly change over time, as is the case, especially in early development. Furthermore, they are incapable of separating groupwise vs. individual developmental trends, nor are they capable of effective data imputation (Adams *et al.*, 2023). Therefore, while these methods might be effective for comparing shape variation across time points, they are not as effective for shape trajectory forecasting as manifold-based methods.

Comparatively, manifold-based techniques are more effective as the longitudinal trajectories traversing the shape space yield an effective description of shape variation over time. LDDMM techniques offer a structured framework to describe shape variation, and furthermore, the geodesic trajectories themselves are clinically relevant as they offer prognostic and diagnostic

utility. When utilized within a hierarchical model that incorporates many trajectories for a population average, new trajectories can be estimated for unseen data, which could offer prognostic significance. Furthermore, trajectories can be compared using relational transport operators to diagnose if a trajectory is irregular compared to population averages. Similarly, DL methods mostly operate directly on medical imaging data with convolutional networks. This allows us a way to extract hidden features from images which could also influence spatiotemporal trajectories, otherwise lost during parameterization processes necessary for LDDMM or PCA-based models. In encoding networks especially, the latent space encompassed by these extracted latent variables can be structured to construct a physically meaningful underlying spatiotemporal manifold. The inductive capacity of DL methods with such structured latent spaces is then superior to linear methods, capable of imputing missing data and predicting spatiotemporal trajectories.

4.2 Paucity of longitudinal datasets

Another clear deficiency is the lack of large, open-source, and high-quality longitudinal imaging datasets. Existing datasets used in studies are generally small, in-house, cover a short time span, and are limited to very specific clinical conditions (Table 1). This is, of course, understandable as it is extremely difficult to gather longitudinal data. Issues such as participant attrition (Young et al., 2006) and ethical concerns (Tinker et al., 2009) are just two examples of difficulties that hamper the execution of effective studies. An exception to this is the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which is a large multimodal database of longitudinal biomarker and neuroimaging data tracking the progression of AD (Jack et al., 2008). This dataset is particularly outstanding due to its size and comprehensiveness, leading to many studies covered in this review validating their methods on the ADNI dataset. Nevertheless, this dataset remains unique and standout compared to others. This paucity of longitudinal datasets, especially for medical imaging, impairs the efficacy of both LDDMM and DL techniques covered in this study.

DL techniques are notoriously data hungry, with larger dataset sizes contributing significantly toward improved efficacy of networks (Sun et al., 2017; Cho et al., 2015). While techniques such as transfer learning (Alzubaidi et al., 2020) and data augmentation (Mumuni and Mumuni, 2022) seek to ameliorate this issue, it remains pervasive. Conversely, whilst the LDDMM framework is comparatively not as data-hungry, sufficiently sized datasets are also essential. Adequately sized and diverse datasets are vital to ensure that the estimated population average trajectories are reflective of the entire population. Solutions such as GAN-based frameworks discussed in Section 3.2 are shown to be helpful in addressing the issue of data paucity. Therein, generative processes and adversarial training frameworks can increase the generalizability of networks. The latter is particularly useful as the adversarial process assists in regularizing and structuring the latent space, implicitly learning the underlying spatiotemporal manifold. Nonetheless, the lack of datasets presents another issue of validity.

In essence, the impressive performance on specific datasets could be a function of the dataset and not the frameworks themselves. Thus, exploring their efficacies on additional anatomical structures and imaging modalities is also prudent. Initiatives to compile multimodal datasets to train and test frameworks in a challenge-like style such as the Medical Segmentation Decathlon (MSD) could be warranted to ensure that future developments in methodology are sufficiently valid (Antonelli et al., 2022).

Nevertheless, longitudinal datasets, be it open-source or in-house, remain scarce. Gathering additional longitudinal data remains the most ideal option, however the aforementioned practical difficulties in data gathering present a significant barrier. In the medium to long term, additional initiatives resembling ADNI could be warranted to gather high-quality, longitudinal, multi-center data for other diseases and disorders benefiting from spatiotemporal shape analyses. Furthermore, these data should be multi-modal, encompassing both imaging and also biomarker data as these have been shown to work synergistically when incorporated into joint frameworks, improving their efficacy. In the meantime, efforts to compile existing data into a large open-source database could be more warranted. This could resemble, for example, the aforementioned MSD. Nonetheless, the impetus to gather and unite such datasets is lacking, especially in the face of general (un)willingness to openly share rare datasets (Tedersoo et al., 2021). Furthermore, medical imaging datasets face strict international data privacy regulations (Lo, 2015; Phillips, 2018). Regulations such as General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) are compounded with ethical concerns and additional practicalities such as ensuring patient anonymity (Lo, 2015; Larson et al., 2020; Banja et al., 2021). Overall, these considerations are non-trivial and present significant barriers to nurturing a culture of open science for spatiotemporal shape modeling.

4.3 Comparison of methods

This review focused on two main avenues for spatiotemporal shape modeling, including the LDDMM framework and varying DL approaches. Each strategy presents distinct advantages and disadvantages as we will discuss (Table 2).

From the computational complexity viewpoint, the state-of-the-art LDDMM approaches we describe can be computationally expensive. Even with parallelized processing utilizing both CPU and GPU, moderately sized data require up to a day of compute time (Bône et al., 2020a). As discussed in Section 3.6, state-of-the-art DL approaches (e.g., transformers, diffusion models, etc.) also require extensive computational resources for training and potentially inference. Nevertheless, as briefly discussed in the same section, decreasing the cost of training and inference is an active field of research and heavily architecture-dependent.

On interpretability, the LDDMM framework is based on a clear underlying mathematical framework. The uncovered underlying longitudinal trajectories generally present directly interpretable and understandable outputs. Especially with regards to the discussed hierarchical models, which explicitly separate population-level and individual-level variations, the overarching findings taken

TABLE 1 A summary of several longitudinal medical imaging datasets.

Name	Anatomy	Modality	Age range		Number of subjects	Repeated measurements
			Youngest	Oldest		
OASIS-2	Brain	MRI	60	96	150	≤5
OASIS-3	Brain	MRI	42	95	1,378	≤7
HABS-HD	Brain	MRI	40	92	3,838	≤3
Harvard aging brain study	Brain	MRI	62	90	>290	≤3
ADNI-4	Brain	MRI	55	90	>2,400	≤6
UK Biobank	Brain	MRI	44	87	>5,000	≤2
	Whole body	DXA			≥5,156	
	Abdomen	MRI			≥113,65	
	Heart	MRI			≥5,100	
MCSA	Brain	MRI	30	89	1,802	Unknown

OASIS, Open Access Series of Imaging Studies; HABS-HD, Health and Aging Brain Study-Health Disparities; ADNI, Alzheimer’s disease neuroimaging initiative; MCSA, Mayo clinic study of aging.

TABLE 2 Comparison of approaches for spatiotemporal shape modeling.

Aspect	LDDMM approach	Deep learning approaches
Temporal modeling	Explicit via diffeomorphic trajectories and time warping	Implicit through network architecture (RNNs, transformers) or latent space structuring
Data requirements	Moderate—Can work with smaller datasets due to strong priors	High—Generally require large datasets for good performance
Computational complexity	High—Requires iterative optimization for hierarchical models, computationally expensive geodesic calculations	High for training, moderate for inferences; varies by architecture (transformers >RNNs >CNNs)
Scalability	Limited—Struggles with large datasets, optimization becomes prohibitive	Good—Naturally handles large datasets
Interpretability	High—Explicit mathematical framework, geodesics have geometric meaning, trajectories directly interpretable	Low—Black box nature, latent variables lack physical meaning, requires <i>post-hoc</i> interpretability methods
Clinical applicability	Strong for trajectory analysis, population vs individual variation assessment	Strong for end-to-end tasks (e.g., prediction) and multimodal data integration

as immediately interpretable by clinical end users. In contrast, DL models are “black boxes” with opaque decision processes as previously discussed in Section 3.6.

For clinical translation, both approaches offer benefits for different applications. The LDDMM framework allows us to elucidate upon the underlying mechanisms of shape change over time by presenting us with a foundation to separate and study both population and individual dynamics mathematically. In contrast, DL demonstrates superior performance for specific tasks such as prediction or imputation via well-regulated generative architectures. Furthermore, DL models can extract

hidden features from imaging data which are not specifically accounted for, compared to an explicit mathematical model for example. Nevertheless, as the decision-making process is opaque without additional *post-hoc* techniques to increase interpretability, it remains difficult to directly apply to a clinical setting as previously discussed (Section 3.6). Thus, the clinical translatability of both approaches is heavily dependent on the end-user and their requirements, be it end-to-end tasks (e.g., *classification, segmentation, etc.*) or understanding and characterization.

4.4 Future outlook and directions

In this review, we focused mainly on the development of LDDMM and DL-based techniques. We did so because these were considered the most versatile for generalizable spatiotemporal shape modeling. This is opposed to alternative methods seeking to model shape changes of specific anatomical structures over time from a mechanistic standpoint. For example, many early works on spatiotemporal shape modeling of tumors attempted to develop models uncovering the underlying mechanistic cause and effects governing their growth (Jarrett et al., 2018). Similar works also exist focusing on cardiac tissue remodeling (Wang V. Y. et al., 2017) and bone remodeling (Kameo et al., 2020). Whilst these varied mechanistic models are inherently different, they generally revolve around shape change as a consequence of mechanical and biochemical stimuli or a combination thereof. Thus, these models seek to uncover the underlying formulae governing these interactions and their relationships. This is in contrast with the LDDMM framework, which operates solely from a geometric perspective in uncovering the trajectories of diffeomorphic transformations. In other words, the LDDMM framework does not explicitly consider the underlying physical laws governing the biological processes that lead to the resulting shape changes. Therefore, this approach potentially neglects key information that may affect how reflective the LDDMM approach is in said processes and, therefore, its accuracy. Similarly,

DL techniques are opaque, often referred to as black boxes (Castelvecchi, 2016). Therein, the model layers, in effect, operate on hidden features uncovered during training processes. These have, in principle, no physical meaning and are not always explainable, engendering issues of trust and validity.

A compromise and potential future direction of research is via physics-informed neural networks (PINNs) (Cuomo et al., 2022). Therein, the strengths of DL to process large datasets are utilized to solve underlying physical equations that describe the physics of a system. PINNs are particularly useful even, for example, to uncover underlying dynamics of systems that were previously obscured under high dimensional nonlinear data (Lagergren et al., 2020). Tajdari et al. (2021) and Tajdari et al. (2022) demonstrated the applicability of PINN principles within frameworks to model the longitudinal progression of adolescent idiopathic scoliosis, outperforming traditional methods. While their works were mainly concerned with the mechanistic effects of loading on spinal outcomes, their efficacy also lends itself to potential benefits for spatiotemporal shape modeling. Nevertheless, PINNs remain an unexplored avenue and warrant further study.

Another developing field is utilizing and exploiting causality in the form of causal deep learning. In essence, causality and structural causal models (SCMs) seek to capture and model the chain of causality and inter-variability of multivariate systems (Peters et al., 2017; Pearl, 2013; Pearl et al., 2016). This is useful as it enables us to interrogate models to obtain counterfactuals (i.e., if X was different, what is the effect on Y ? Or more relevantly, “How would injury A affect bone development of pediatric subject B ?”). For spatiotemporal shape modeling specifically, this could be used to obtain predictions of shape change over time as a counterfactual from existing data. Traditional methods relied on a system of structural equations with computation graphs, but this precluded the use of higher dimensional data such as images. In recent years, several studies have explored extending SCMs toward being supported by DL [i.e., Deep Structural Causal Models (DSCMs)], enabling the use of hidden features identified via DL (Lore et al., 2018; Pawlowski et al., 2020; Berrevoets et al., 2023). Zhou et al. (2023) reviewed the synergistic capabilities of generative models and causality, specifically highlighting the applicability of the latter in enhancing the interpretability of generative processes. Further works such as by Reinhold et al. (2021) demonstrated the capability of DSCMs to generate counterfactual brain MRIs of patients with multiple sclerosis. They were able to manipulate demographic and disease covariates and observed their effects on MRI imaging in a novel proof-of-concept. Rasal et al. (2022) further extended DSCMs toward shape modeling, specifically 3D meshes, demonstrating the extendability and scalability of the principle toward more complex data types. Nevertheless, the field is still in its relative infancy, with further developments and refinements in the DSCM framework potentially leading to enhanced efficacy for longitudinal shape modeling.

In terms of direct clinical translation, spatiotemporal shape modeling is yet to be fully explored. Existing works covered in this review highlighted, for example, capabilities to stratify patient cohorts both temporally and by subtype (Young et al., 2024; Puglisi et al., 2025). This has the potential capability to enhance the efficacy of clinical trials, for example, by enabling more precise targeting of

treatments. To our knowledge, use of longitudinal data modeling is still relatively theoretical and whilst research has projected its utility, direct translations to clinical practice remain difficult as is the case for most biomedical research (Finney Rutten et al., 2024). For our specific context of spatiotemporal shape modeling, our proposed application scenarios (Section 1) are also simply hypothetical at this juncture. Nevertheless, the clinical applicability of spatiotemporal shape modeling remains an unexplored yet promising research domain.

In summary, this paper mainly reviewed the LDDMM framework and DL-based techniques for longitudinal shape modeling. Both achieve their remarkable state-of-the-art performance as they function on similar principles of uncovering the underlying nonlinear spatiotemporal data manifold. Whilst promising, the LDDMM framework is computationally expensive and inefficient due to the exhaustive optimization procedure necessary to calculate smooth and invertible diffeomorphisms. It, nevertheless, demonstrates strong capabilities to establish hierarchical models that differentiate individual and population-level temporal trajectories. Conversely, DL-based techniques are powerful but data-hungry and lack underlying physical meaning. Network architectures have been developed to predict shape changes in anatomical structures. Nevertheless, the underlying data manifolds and spatiotemporal trajectories governing these predictions are obscured by the “black box” nature of DL architectures. This affects the interpretability of these predictions, especially if the longitudinal trajectories themselves are important. Nevertheless, the capability of DL architectures to identify hidden features from input images and implicitly map the underlying data manifolds denote their importance for spatiotemporal shape modeling. State-of-the-art developments in DL such as via foundation models, highly sophisticated pre-trained feature extractors which can extract rich representations from data, are also a promising direction of exploration (Ma et al., 2024; van Veldhuizen et al., 2025; Homayounfar et al., 2026). Our review highlights that hybrid techniques that amalgamate both approaches’ strengths are more desirable. Furthermore, frameworks incorporating multi-modal data improved generalizability. Thus, further works should not neglect the utility of auxiliary data (e.g., biomarker levels, demographic information, etc.). Many studies discussed in our review utilized multimodal data, in LDDMM, DL, and hybrid frameworks. Thus, multimodality represents a clear path forward for state-of-the-art development (Nakach et al., 2024). Finally, we theorize that utilizing mechanistic models in a manner similar to PINNs or structured causal frameworks could also further improve the predictive capacities of future spatiotemporal shape models.

Author contributions

ET: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. NT: Conceptualization, Supervision, Writing – review & editing. AZ: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the

support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2025.1671099/full#supplementary-material>

References

- Adam, J. A. (1986). A simplified mathematical model of tumor growth. *Math. Biosci.* 81, 229–244. doi: 10.1016/0025-5564(86)90119-7
- Adams, J., Khan, N., Morris, A., and Elhabian, S. (2023). Learning spatiotemporal statistical shape models for non-linear dynamic anatomies. *Front. Bioeng. Biotechnol.* 11:1086234. doi: 10.3389/fbioe.2023.1086234
- Alberdi, A., Aztiria, A., and Basarab, A. (2016). On the early diagnosis of Alzheimer's disease from multimodal signals: a survey. *Artif. Intell. Med.* 71, 1–29. doi: 10.1016/j.artmed.2016.06.003
- Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., Santamaria, J., Duan, Y., et al. S. (2020). Towards a better understanding of transfer learning for medical imaging: a case study. *Appl. Sci.* 10:4523. doi: 10.3390/app10134523
- Ambellan, F., Hanik, M., and von Tycowicz, C. (2021). *Morphomatics: Geometric Morphometrics in Non-Euclidean Shape Spaces*. Available online at: <https://morphomatics.github.io/> (Accessed January 5, 2025).
- Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., et al. (2022). The medical segmentation decathlon. *Nat. Commun.* 13:4128. doi: 10.1038/s41467-022-30695-9
- Arjovsky, M., Chintala, S., and Bottou, L. (2017). "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning, Volume 70 of Proceedings of Machine Learning Research*, eds. D. Precup, and Y. W. Teh (Sydney, NSW: PMLR), 214–223.
- Azad, R., Kazerouni, E., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., et al. (2024). Advances in medical image analysis with vision transformers: a comprehensive review. *Med. Image Anal.* 91:103000. doi: 10.1016/j.media.2023.103000
- Banja, J., Rousselle, R., Duszak, R., Safdar, N., and Alessio, A. M. (2021). Sharing and selling images: ethical and regulatory considerations for radiologists. *J. Am. Coll. Radiol.* 18, 298–304. doi: 10.1016/j.jacr.2020.08.003
- Bartoldson, B. R., Kailkhura, B., and Blalock, D. (2023). Compute-efficient deep learning: algorithmic trends and opportunities. *J. Mach. Learn. Res.* 24, 1–77. Available online at: <https://dl.acm.org/doi/abs/10.5555/3648699.3648821>
- Ben Amor, B., Arguillere, S., and Shao, L. (2023). Resnet-ldmm: advancing the ldmm framework using deep residual networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 3707–3720. doi: 10.1109/TPAMI.2022.3174908
- Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J. M. L., Hlatky, L., et al. (2014). Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput. Biol.* 10:e1003800. doi: 10.1371/journal.pcbi.1003800
- Berrevoets, J., Kacprzyk, K., Qian, Z., and van der Schaar, M. (2023). *Causal Deep Learning*. arXiv preprint. doi: 10.48550/arXiv.2303.02186
- Binte Alam, S., Nii, M., Shimizu, A., and Kobashi, S. (2020). Spatiotemporal statistical shape model for temporal shape change analysis of adult brain. *Curr. Med. Imaging* 16, 499–506. doi: 10.2174/1573405615666181120141147
- Blinkouskaya, Y., and Weickenmeier, J. (2021). Brain shape changes associated with cerebral atrophy in healthy aging and Alzheimer's disease. *Front. Mech. Eng.* 7:705653. doi: 10.3389/fmech.2021.705653
- Bone, A., Colliot, O., and Durrleman, S. (2018). "Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE). doi: 10.1109/CVPR.2018.00966
- Bône, A., Colliot, O., and Durrleman, S. (2020a). Learning the spatiotemporal variability in longitudinal shape data sets. *Int. J. Comput. Vis.* 128, 2873–2896. doi: 10.1007/s11263-020-01343-w
- Bône, A., Louis, M., Colliot, O., and Durrleman, S. (2019). *Learning Low-Dimensional Representations of Shape Data Sets with Diffeomorphic Autoencoders*. Cham: Springer International Publishing, 195–207. doi: 10.1007/978-3-030-20351-1_15
- Bône, A., Louis, M., Martin, B., and Durrleman, S. (2018). *Deformetrica 4: An Open-Source Software for Statistical Shape Analysis*. Cham: Springer International Publishing, 3–13. doi: 10.1007/978-3-030-04747-4_1
- Bône, A., Louis, M., Routier, A., Samper, J., Bacci, M., Charlier, B., et al. (2017). *Prediction of the Progression of Subcortical Brain Structures in Alzheimer's Disease from Baseline*. Cham: Springer International Publishing, 101–113. doi: 10.1007/978-3-319-67675-3_10
- Bône, A., Vernhet, P., Colliot, O., and Durrleman, S. (2020b). *Learning Joint Shape and Appearance Representations with Metamorphic Auto-Encoders*. Cham: Springer International Publishing, 202–211. doi: 10.1007/978-3-030-59710-8_20
- Bookstein, F. L. (1982). Foundations of morphometrics. *Annu. Rev. Ecol. Syst.* 13, 451–470. doi: 10.1146/annurev.es.13.110182.002315
- Brahma, P. P., Wu, D., and She, Y. (2016). Why deep learning works: a manifold disentanglement perspective. *IEEE Trans. Neural Netw. Learn. Syst.* 27, 1997–2008. doi: 10.1109/TNNLS.2015.2496947
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature* 538, 20–23. doi: 10.1038/538020a
- Chadebec, C., and Allassonnière, S. (2023). *Variational Inference for Longitudinal Data Using Normalizing Flows*. arXiv preprint. doi: 10.48550/arXiv.2303.14220

- Chen, M., Bian, Y., Chen, N., and Qiu, A. (2025). Orthogonal mixed-effects modeling for high-dimensional longitudinal data: an unsupervised learning approach. *IEEE Trans. Med. Imaging* 44, 207–220. doi: 10.1109/TMI.2024.3435855
- Chen, Y., Wolterink, J. M., Neve, O. M., Romeijn, S. R., Verbist, B. M., Hensen, E. F., et al. (2024). “Vestibular schwannoma growth prediction from longitudinal mri by time conditioned neural fields,” in *Medical Image Computing and Computer Assisted Intervention* (Cham: Springer). doi: 10.1007/978-3-031-72384-1_48
- Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. (2015). *How Much Data is Needed to Train a Medical Image Deep Learning System to Achieve Necessary High accuracy?* *arXiv preprint*. doi: 10.48550/arXiv.1511.06348
- Clark, W. (1991). Tumour progression and the nature of cancer. *Br. J. Cancer* 64, 631–644. doi: 10.1038/bjc.1991.375
- Connor, M., Canal, G., and Rozell, C. (2021). “Variational autoencoder with learned latent structure,” in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, Volume 130 of Proceedings of Machine Learning Research*, eds. A. Banerjee, and K. Banerjee (San Diego, CA: PMLR), 2359–2367.
- Couronne, R., Vidailhet, M., Corvol, J. C., Lehericy, S., and Durrleman, S. (2019). “Learning disease progression models with longitudinal data and missing values,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (Venice: IEEE). doi: 10.1109/ISBI.2019.8759198
- Croitoru, F.-A., Hondru, V., Ionescu, R. T., and Shah, M. (2023). Diffusion models in vision: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 10850–10869. doi: 10.1109/TPAMI.2023.3261988
- Cui, R., and Liu, M. (2019). Rnn-based longitudinal analysis for diagnosis of Alzheimer’s disease. *Comput. Med. Imaging Graph.* 73, 1–10. doi: 10.1016/j.compmedimag.2019.01.005
- Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: where we are and what’s next. *J. Sci. Comput.* 92:88. doi: 10.1007/s10915-022-01939-z
- Cury, C., Lorenzi, M., Cash, D., Nicholas, J. M., Routier, A., Rohrer, J., et al. (2016). *Spatio-Temporal Shape Analysis of Cross-Sectional Data for Detection of Early Changes in Neurodegenerative Disease*. Cham: Springer International Publishing, 63–75. doi: 10.1007/978-3-319-51237-2_6
- Dalca, A. V., Sridharan, R., Sabuncu, M. R., and Golland, P. (2015). *Predictive Modeling of Anatomy with Genetic and Clinical Data*. Cham: Springer International Publishing, 519–526. doi: 10.1007/978-3-319-24574-4_62
- Darwin, C. (2009). *The Origin of Species: By Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511694295
- Debavelaere, V., Durrleman, S., and Allassonnière, S. (2020). Learning the clustering of longitudinal shape data sets into a mixture of independent or branching trajectories. *Int. J. Comput. Vis.* 128, 2794–2809. doi: 10.1007/s11263-020-01337-8
- Dhariwal, P., and Nichol, A. (2021). *Diffusion Models Beat GANs on Image Synthesis*. *arXiv preprint*. doi: 10.48550/arXiv.2105.05233
- Ding, H., Wang, B., Hamel, A. P., Melkonyan, M., Ang, T. F. A., Au, R., et al. (2023). Prediction of progression from mild cognitive impairment to Alzheimer’s disease with longitudinal and multimodal data. *Front. Dement.* 2:1271680. doi: 10.3389/frdem.2023.1271680
- Ding, Z., Fleishman, G., Yang, X., Thompson, P., Kwitt, R., Niethammer, M., et al. (2019). Fast predictive simple geodesic regression. *Med. Image Anal.* 56, 193–209. doi: 10.1016/j.media.2019.06.003
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. *arXiv preprint*. doi: 10.48550/arXiv.2010.11929
- Dummer, S., Strisciuglio, N., and Brune, C. (2023). RDA-INR: riemannian diffeomorphic autoencoding via implicit neural representations. *SIAM* 17. doi: 10.1137/24M1644730
- Durrleman, S., Prastawa, M., Charon, N., Korenberg, J. R., Joshi, S., Gerig, G., et al. (2014). Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage* 101, 35–49. doi: 10.1016/j.neuroimage.2014.06.043
- Ehrhardt, J., and Wilms, M. (2022). *Autoencoders and Variational Autoencoders in Medical Image Analysis*. Amsterdam: Elsevier, 129–162. doi: 10.1016/B978-0-12-824349-7.00015-3
- Elazab, A., Wang, C., Gardezi, S. J. S., Bai, H., Hu, Q., Wang, T., et al. (2020). GP-GAN: brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images. *Neural Netw.* 132, 321–332. doi: 10.1016/j.neunet.2020.09.004
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1
- Fefferman, C., Mitter, S., and Narayanan, H. (2016). Testing the manifold hypothesis. *J. Am. Math. Soc.* 29, 983–1049. doi: 10.1090/jams/852
- Fetty, L., Bylund, M., Kuess, P., Heilemann, G., Nyholm, T., Georg, D., et al. (2020). Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Z. Med. Phys.* 30, 305–314. doi: 10.1016/j.zemedi.2020.05.001
- Finney Rutten, L. J., Ridgeway, J. L., and Griffin, J. M. (2024). Advancing translation of clinical research into practice and population health impact through implementation science. *Mayo Clin. Proc.* 99, 665–676. doi: 10.1016/j.mayocp.2023.02.005
- Fishbaugh, J., Durrleman, S., Prastawa, M., and Gerig, G. (2017). Geodesic shape regression with multiple geometries and sparse parameters. *Med. Image Anal.* 39, 1–17. doi: 10.1016/j.media.2017.03.008
- Fishbaugh, J., Prastawa, M., Gerig, G., and Durrleman, S. (2013). *Geodesic Shape Regression in the Framework of Currents*. Cham: Springer Berlin Heidelberg, 718–729. doi: 10.1007/978-3-642-38868-2_60
- Fishbaugh, J., Prastawa, M., Gerig, G., and Durrleman, S. (2014). “Geodesic regression of image and shape data for improved modeling of 4D trajectories,” in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (Beijing: IEEE). doi: 10.1109/ISBI.2014.6867889
- Fletcher, T. (2011). “Geodesic regression on Riemannian manifolds,” in *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy - Geometrical and Statistical Methods for Modelling Biological Shape Variability* (Toronto, ON), 75–86.
- Fuest, M., Ma, P., Gui, M., Schusterbauer, J., Hu, V. T., Ommer, B., et al. (2024). *Diffusion Models and Representation Learning: A Survey*. *arXiv preprint*. doi: 10.48550/arXiv.2407.00783
- Gadewar, S. P., Ramesh, A., Liu, M., Ba Gari, I., Nir, T. M., Thompson, P., et al. (2023a). “Predicting individual brain mris at any age using style encoding generative adversarial networks,” in *18th International Symposium on Medical Information Processing and Analysis*, eds. M. G. Linguraru, L. Linguraru, N. Lepore, E. Romero Castro, J. Brieva, and P. Guevara (Valparaiso: SPIE). doi: 10.1117/12.2669741
- Gadewar, S. P., Zhu, A. H., Somu, S., Ramesh, A., Gari, I. B., Thomopoulos, S. I., et al. (2023b). *Normative Aging for an Individual’s Full Brain MRI Using Style GANs to Detect Localized Neurodegeneration*. Cham: Springer Nature Switzerland, 387–395. doi: 10.1007/978-3-031-45676-3_39
- Gao, L., Pan, H., Liu, F., Xie, X., Zhang, Z., Han, J., et al. (2018). “Brain disease diagnosis using deep learning features from longitudinal mr images,” in *Lecture Notes in Computer Science* (Cham: Springer International Publishing), 327–339. doi: 10.1007/978-3-319-96890-2_27
- Gao, Y., Phillips, J. M., Zheng, Y., Min, R., Fletcher, P. T., Gerig, G., et al. (2018). “Fully convolutional structured LSTM networks for joint 4D medical image segmentation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC: IEEE). doi: 10.1109/ISBI.2018.8363764
- Gaudferrau, F., Allassonnière, S., and Le Pennec, E. (2023). “A multiscale algorithm for computing realistic image transformation: application to the modelling of fetal brain growth,” in *Medical Imaging 2023: Image Processing*, eds. I. Išgum, and O. Colliot (San Diego, CA: SPIE), 3. doi: 10.1117/12.2654259
- Geiger, R. S., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J., et al. (2021). “Garbage in, garbage out” revisited: what do machine learning application papers report about human-labeled training data? *Quant. Sci. Stud.* 2, 795–827. doi: 10.1162/qss_a_00144
- Gerlee, P. (2013). The model muddle: in search of tumor growth laws. *Cancer Res.* 73, 2407–2411. doi: 10.1158/0008-5472.CAN-12-4355
- Ghosh, P., Sajjadi, M. S. M., Vergari, A., Black, M., and Scholkopf, B. (2020). “From variational to deterministic autoencoders,” in *International Conference on Learning Representations*. doi: 10.48550/arXiv.1903.12436
- Glaunès, J., Qiu, A., Miller, M. I., and Younes, L. (2008). Large deformation diffeomorphic metric curve mapping. *Int. J. Comput. Vis.* 80, 317–336. doi: 10.1007/s11263-008-0141-9
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, Vol. 27, eds. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger (Red Hook, NY: Curran Associates, Inc).
- Grenander, U., and Miller, M. I. (1998). Computational anatomy: an emerging discipline. *Q. Appl. Math.* 56, 617–694. doi: 10.1090/qam/1668732
- Guan, H., and Liu, M. (2022). Domain adaptation for medical image analysis: a survey. *IEEE Trans. Biomed. Eng.* 69, 1173–1185. doi: 10.1109/TBME.2021.3117407
- Gui, J., Sun, Z., Wen, Y., Tao, D., and Ye, J. (2023). A review on generative adversarial networks: algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* 35, 3313–3332. doi: 10.1109/TKDE.2021.3130191
- Guo, B., Lu, D., Szumel, G., Gui, R., Wang, T., Konz, N., et al. (2024). *The Impact of Scanner Domain Shift on Deep Learning Performance in Medical Imaging: an Experimental Study*. *arXiv preprint*. doi: 10.48550/arXiv.2409.04368
- Guo, Z., Liu, J., Wang, Y., Chen, M., Wang, D., Xu, D., et al. (2023). Diffusion models in bioinformatics and computational biology. *Nat. Rev. Bioeng.* 2, 136–154. doi: 10.1038/s44222-023-00114-9
- Hamareh, G., and Gustavsson, T. (2004). Deformable spatio-temporal shape models: extending active shape models to 2d+time. *Image Vis. Comput.*, 22, 461–470. doi: 10.1016/j.imavis.2003.11.009
- Han, T., Kather, J. N., Pedersoli, F., Zimmermann, M., Keil, S., Schulze-Hagen, M., et al. (2022). Image prediction of disease progression for osteoarthritis

- by style-based manifold extrapolation. *Nat. Mach. Intell.* 4, 1029–1039. doi: 10.1038/s42256-022-00560-x
- Hanik, M., Hege, H.-C., and Tycowicz, C. v. (2022). “A nonlinear hierarchical model for longitudinal data on manifolds,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (Kolkata: IEEE). doi: 10.1109/ISBI52829.2022.9761465
- Harie, Y., Gautam, B. P., and Wasaki, K. (2023). Computer vision techniques for growth prediction: a prisma-based systematic literature review. *Appl. Sci.* 13:5335. doi: 10.3390/app13095335
- He, R., Ang, G., and Tward, D. (2025). *Individualized Multi-horizon MRI Trajectory Prediction for Alzheimer's Disease*. Cham: Springer Nature Switzerland, 26–37. doi: 10.1007/978-3-031-84525-3_3
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X
- Ho, J., Jain, A., and Abbeel, P. (2020). *Denosing Diffusion Probabilistic Models*. *arXiv preprint*. doi: 10.48550/arXiv.2006.11239
- Homayounfar, M., Bierma-Zeinsträ, S., Zadpoor, A. A., and Tümer, N. (2026). PedVision: a manual-annotation-free and age scalable segmentation pipeline for bone analysis in hand X-ray images. *Biomed. Signal Process. Control* 112:108569. doi: 10.1016/j.bspc.2025.108569
- Hong, S., Fishbaugh, J., Wolff, J. J., Styner, M. A., and Gerig, G. (2019). “Hierarchical multi-geodesic model for longitudinal analysis of temporal trajectories of anatomical shape and covariates,” in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2019* (Cham: Springer), 57–65. doi: 10.1007/978-3-030-32251-9_7
- Islam, S., Elmekki, H., Elsebaï, A., Bentahar, J., Drawel, N., Rjoub, G., et al. (2024). A comprehensive survey on applications of transformers for deep learning tasks. *Expert Syst. Appl.* 241:122666. doi: 10.1016/j.eswa.2023.122666
- Jabbar, A., Li, X., and Omar, B. (2021). A survey on generative adversarial networks: variants, applications, and training. *ACM Comput. Surv.* 54, 1–49. doi: 10.1145/3463475
- Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691. doi: 10.1002/jmri.21049
- Jarrett, A. M., Lima, E. A., Hormuth, D. A., McKenna, M. T., Feng, X., Ekrut, D. A., et al. (2018). Mathematical models of tumor cell proliferation: a review of the literature. *Expert Rev. Anticancer Ther.* 18, 1271–1286. doi: 10.1080/14737140.2018.1527689
- Jiang, Y., Pjesivac-Grbovic, J., Cantrell, C., and Freyer, J. P. (2005). A multiscale model for avascular tumor growth. *Biophys. J.* 89, 3884–3894. doi: 10.1529/biophysj.105.060640
- Kameo, Y., Miya, Y., Hayashi, M., Nakashima, T., and Adachi, T. (2020). *In silico* experiments of bone remodeling explore metabolic diseases and their drug treatment. *Sci. Adv.* 6:eaa0938. doi: 10.1126/sciadv.aax0938
- Kapoor, J., Macke, J. H., and Baumgartner, C. F. (2025). *MRExtrap: Longitudinal Aging of Brain MRIs using Linear Modeling in Latent Space*. *arXiv preprint*. doi: 10.48550/arXiv.2508.19482
- Karras, T., Laine, S., and Aila, T. (2019). “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA: IEEE). doi: 10.1109/CVPR.2019.00453
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., et al. (2020). “Analyzing and improving the image quality of StyleGAN,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA: IEEE). doi: 10.1109/CVPR42600.2020.00813
- Kasahara, K., Saito, A., Takakuwa, T., Yamada, S., Matsuzoe, H., Hontani, H., et al. (2018). A spatiotemporal statistical shape model of the brain surface during human embryonic development. *Adv. Biomed. Eng.* 7, 146–155. doi: 10.14326/abe.7.146
- Kazerouni, A., Aghdam, E. K., Heidari, M., Azad, R., Fayyaz, M., Hachililoglu, I., et al. (2023). Diffusion models in medical imaging: a comprehensive survey. *Med. Image Anal.* 88:102846. doi: 10.1016/j.media.2023.102846
- Kendall, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* 16, 81–121. doi: 10.1112/blms/16.2.81
- Kim, H. J., Adluru, N., Suri, H., Vemuri, B. C., Johnson, S. C., Singh, V., et al. (2017). “Riemannian nonlinear mixed effects models: analyzing longitudinal deformations in neuroimaging,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: IEEE). doi: 10.1109/CVPR.2017.612
- Klemetti, A., Raatikainen, M., Myllyaho, L., Mikkonen, T., and Nurminen, J. K. (2023). Systematic literature review on cost-efficient deep learning. *IEEE Access* 11, 90158–90180. doi: 10.1109/ACCESS.2023.3275431
- Koval, I., Bône, A., Louis, M., Lartigue, T., Bottani, S., Marcoux, A., et al. (2021). AD Course map charts Alzheimer's disease progression. *Sci. Rep.* 11:8020. doi: 10.1038/s41598-021-87434-1
- Koval, I., Schiratti, J.-B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., et al. (2018). Spatiotemporal propagation of the cortical atrophy: population and individual patterns. *Front. Neurol.* 9:235. doi: 10.3389/fneur.2018.00235
- Kuroishi, T., Tominaga, S., Morimoto, T., Tashiro, H., Itoh, S., Watanabe, H., et al. (1990). Tumor growth rate and prognosis of breast cancer mainly detected by mass screening. *Jpn. J. Cancer Res.* 81, 454–462. doi: 10.1111/j.1349-7006.1990.tb02591.x
- Lagergren, J. H., Nardini, J. T., Baker, R. E., Simpson, M. J., and Flores, K. B. (2020). Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLoS Comput. Biol.* 16:e1008462. doi: 10.1371/journal.pcbi.1008462
- Lapenda, L. V. N., Monteiro, R. P., and Bastos-Filho, C. J. A. (2020). “Autoencoder latent space: an empirical study,” in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (Canberra, ACT: IEEE). doi: 10.1109/SSCI47803.2020.9308551
- Larson, D. B., Magnus, D. C., Lungren, M. P., Shah, N. H., and Langlotz, C. P. (2020). Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology* 295, 675–682. doi: 10.1148/radiol.2020192536
- Li, J., Chen, J., Tang, Y., Wang, C., Landman, B. A., Zhou, S. K., et al. (2023). Transforming medical imaging with transformers? A comparative review of key properties, current progresses, and future perspectives. *Med. Image Anal.* 85:102762. doi: 10.1016/j.media.2023.102762
- Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., et al. (2022). Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowl. Inf. Syst.* 64, 3197–3234. doi: 10.1007/s10115-022-01756-8
- Libonati, F., and Buehler, M. J. (2017). Advanced structural materials by bioinspiration. *Adv. Eng. Mater.* 19:1600787. doi: 10.1002/adem.201600787
- Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). *A Critical Review of Recurrent Neural Networks for Sequence Learning*. *arXiv preprint*. doi: 10.48550/arXiv.1506.00019
- Litrico, M., Guarnera, F., Giuffrida, M. V., Ravi, D., and Battiatto, S. (2025). *Temporally-Aware Diffusion Model for Brain Progression Modelling with Bidirectional Temporal Regularisation*. Cham: Springer. doi: 10.1007/978-3-031-72069-7_42
- Lo, B. (2015). Sharing clinical trial data: maximizing benefits, minimizing risk. *JAMA* 313:793. doi: 10.1001/jama.2015.292
- Lopez Pinaya, W. H., Vieira, S., Garcia-Dias, R., and Mechelli, A. (2020). *Autoencoders*. Amsterdam: Elsevier, 193–208. doi: 10.1016/B978-0-12-815739-8.00011-0
- Lore, K. G., Stoecklein, D., Davies, M., Ganapathysubramanian, B., and Sarkar, S. (2018). A deep learning framework for causal shape transformation. *Neural Netw.* 98, 305–317. doi: 10.1016/j.neunet.2017.12.003
- Lotter, W., Kreiman, G., and Cox, D. (2016). *Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning*. *arXiv preprint*. doi: 10.48550/arXiv.1605.08104
- Louis, M., Couronné, R., Koval, I., Charlier, B., and Durrleman, S. (2019). “Riemannian geometry learning for disease progression modelling,” in *Information Processing in Medical Imaging* (Cham: Springer International Publishing), 542–553. doi: 10.1007/978-3-030-20351-1_42
- Lozupone, G., Bria, A., Fontanella, F., Meijer, F. J. A., De Stefano, C., and Huisman, H. (2025). *Latent Diffusion Autoencoders: Toward Efficient and Meaningful Unsupervised Representation Learning in Medical Imaging*. *arXiv preprint*. doi: 10.48550/arXiv.2504.08635
- Lu, C., Hirsch, M., and Scholkopf, B. (2017). “Flexible spatio-temporal networks for video prediction,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: IEEE). doi: 10.1109/CVPR.2017.230
- Lu, K., Grover, A., Abbeel, P., and Mordatch, I. (2021). *Pretrained Transformers as Universal Computation Engines*. *arXiv preprint*. doi: 10.48550/arXiv.2103.05247
- Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B., et al. (2024). Segment anything in medical images. *Nat. Commun.* 15:654. doi: 10.1038/s41467-024-44824-z
- Ma, M., Zhang, X., Li, Y., Wang, X., Zhang, R., Wang, Y., et al. (2023). Convlstm coordinated longitudinal transformer under spatio-temporal features for tumor growth prediction. *Comput. Biol. Med.* 164:107313. doi: 10.1016/j.compbiomed.2023.107313
- Ma, Z., Zhang, H., and Liu, J. (2022). *MS-RNN: A Flexible Multi-Scale Framework for Spatiotemporal Predictive Learning*. *arXiv preprint*. doi: 10.48550/arXiv.2206.03010
- Ma, Z., Zhang, H., and Liu, J. (2023). MS-LSTM: exploring spatiotemporal multiscale representations in video prediction domain. *Appl. Soft Comput.* 147:110731. doi: 10.1016/j.asoc.2023.110731
- Marsland, S., and Sommer, S. (2020). *Riemannian Geometry on Shapes and Diffeomorphisms*. Amsterdam: Elsevier, 135–167. doi: 10.1016/B978-0-12-814725-2.00011-X
- Marzin, P., and Cormier-Daire, V. (2020). New perspectives on the treatment of skeletal dysplasia. *Ther. Adv. Endocrinol. Metab.* 11:204201882090401. doi: 10.1177/2042018820904016
- Mihalcea, R., Ignat, O., Bai, L., Borah, A., Chiruzzo, L., Jin, Z., et al. (2025). Why AI Is WEIRD and shouldn't be this way: towards AI for everyone, with everyone, by everyone. *Proc. AAAI Conf. Artif. Intell.* 39, 28657–28670. doi: 10.1609/aaai.v39i27.35092
- Miller, M., Banerjee, A., Christensen, G., Joshi, S., Khaneja, N., Grenander, U., et al. (1997). Statistical methods in computational anatomy. *Stat. Methods Med. Res.* 6, 267–299. doi: 10.1177/096228029700600305

- Miller, M. I. (2004). Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage* 23, S19–S33. doi: 10.1016/j.neuroimage.2004.07.021
- Miller, M. I., Trounev, A., and Younes, L. (2002). On the metrics and euler-lagrange equations of computational anatomy. *Annu. Rev. Biomed. Eng.* 4, 375–405. doi: 10.1146/annurev.bioeng.4.092101.125733
- Miller, M. I., Trounev, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.* 24, 209–228. doi: 10.1007/s10851-005-3624-0
- Monteiro, L. R., and Bordin, B. Furtado dos Reis, S. (2000). Shape distances, shape spaces and the comparison of morphometric methods. *Trends Ecol. Evol.* 15, 217–220. doi: 10.1016/S0169-5347(99)01775-9
- Morcuende, J. A., and Weinstein, S. L. (2003). Developmental skeletal anomalies. *Birth Defects Res. Part C: Embryo Today: Rev.* 69, 197–207. doi: 10.1002/bdrc.10011
- Morikawa, T., Kuchiba, A., Qian, Z. R., Mino-Kenudson, M., Hornick, J. L., Yamauchi, M., et al. (2011). Prognostic significance and molecular associations of tumor growth pattern in colorectal cancer. *Ann. Surg. Oncol.* 19, 1944–1953. doi: 10.1245/s10434-011-2174-5
- Mouches, P., Wilms, M., Rajashekar, D., Langner, S., and Forkert, N. (2021). “Unifying brain age prediction and age-conditioned template generation with a deterministic autoencoder,” in *Proceedings of the Fourth Conference on Medical Imaging with Deep Learning, Volume 143 of Proceedings of Machine Learning Research* (PMLR), 497–506.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). *Alzheimers Dement.* 1, 55–66. doi: 10.1016/j.jalz.2005.06.003
- Mumuni, A., and Mumuni, F. (2022). Data augmentation: a comprehensive survey of modern approaches. *Array* 16:100258. doi: 10.1016/j.array.2022.100258
- Muniandy, K., Asra Ahmad, Z., Annabel Dass, S., Shamsuddin, S., Mohana Kumaran, N., Balakrishnan, V., et al. (2021). Growth and invasion of 3D spheroid tumor of HeLa and CasKi cervical cancer cells. *Oncologie* 23, 279–291. doi: 10.32604/Oncologie.2021.015969
- Muralidharan, P., and Fletcher, P. T. (2012). “Sasaki metrics for analysis of longitudinal data on manifolds,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (Providence, RI: IEEE). doi: 10.1109/CVPR.2012.6247780
- Nakach, F.-Z., Idri, A., and Goceri, E. (2024). A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. *Artif. Intell. Rev.* 57:327. doi: 10.1007/s10462-024-10984-z
- Narayanan, H., and Mitter, S. (2010). “Sample complexity of testing the manifold hypothesis,” in *Advances in Neural Information Processing Systems*, Vol. 23, eds. J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta (Red Hook, NY: Curran Associates, Inc).
- Newsome, R., Green, M., Bell, N., Chagunda, M., Mason, C., Rutland, C., et al. (2016). Linking bone development on the caudal aspect of the distal phalanx with lameness during life. *J. Dairy Sci.* 99, 4512–4525. doi: 10.3168/jds.2015-10202
- Ong, P., Haufmann, M., Lönnroth, O., and Lähdesmäki, H. (2024). *Latent Mixed-effect Models for High-dimensional Longitudinal Data*.
- Ouyang, J., Zhao, Q., Sullivan, E. V., Pfefferbaum, A., Tapert, S. F., Adeli, E., et al. (2021). Longitudinal pooling & consistency regularization to model disease progression from mris. *IEEE J. Biomed. Health Inform.* 25, 2082–2092. doi: 10.1109/JBHI.2020.3042447
- Parfitt, A., Travers, R., Rauch, F., and Glorieux, F. (2000). Structural and cellular changes during bone growth in healthy children. *Bone* 27, 487–494. doi: 10.1016/S8756-3282(00)00353-7
- Pasnoori, N., Flores-Garcia, T., and Barkana, B. D. (2024). Histogram-based features track Alzheimer's progression in brain MRI. *Sci. Rep.* 14:257. doi: 10.1038/s41598-023-50631-1
- Pathan, S., and Hong, Y. (2018). *Predictive Image Regression for Longitudinal Studies with Missing Data*. *arXiv preprint*. doi: 10.48550/arXiv.1808.07553
- Patrício, C., Neves, J. C., and Teixeira, L. F. (2023). Explainable deep learning methods in medical image classification: a survey. *ACM Computing Surveys*, 56, 1–41. doi: 10.1145/3625287
- Pawlowski, N., Castro, D. C., and Glocker, B. (2020). *Deep Structural Causal Models for Tractable Counterfactual Inference*. *arXiv preprint*. doi: 10.48550/arXiv.2006.06485
- Pearl, J. (2013). Structural counterfactuals: a brief introduction. *Cogn. Sci.* 37, 977–985. doi: 10.1111/cogs.12065
- Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Hoboken, NJ: John Wiley & Sons.
- Pegueroles, J., Vilaplana, E., Montal, V., Sampedro, F., Alcolea, D., Carmona-Iragui, M., et al. (2016). Longitudinal brain structural changes in preclinical Alzheimer's disease. *Alzheimers Dement.* 13, 499–509. doi: 10.1016/j.jalz.2016.08.010
- Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA: The MIT Press.
- Phillips, M. (2018). International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR). *Hum. Genet.* 137, 575–582. doi: 10.1007/s00439-018-1919-7
- Pooch, E. H. P., Ballester, P., and Barros, R. C. (2020). *Can We Trust Deep Learning Based Diagnosis? The Impact of Domain Shift in Chest Radiograph Classification*. Cham: Springer International Publishing, 4–83. doi: 10.1007/978-3-030-62469-9_7
- Puglisi, L., Alexander, D. C., and Ravi, D. (2025). Brain latent progression: individual-based spatiotemporal disease progression on 3D Brain MRIs via latent diffusion. *Med. Image Anal.* 106:103734. doi: 10.1016/j.media.2025.103734
- Puyol-Antón, E., Ruijsink, B., Mariscal Harana, J., Piechnik, S. K., Neubauer, S., Petersen, S. E., et al. (2022). Fairness in cardiac magnetic resonance imaging: assessing sex and racial bias in deep learning-based segmentation. *Front. Cardiovasc. Med.* 9:859310. doi: 10.3389/fcvm.2022.859310
- Puyol-Antón, E., Ruijsink, B., Piechnik, S. K., Neubauer, S., Petersen, S. E., Razavi, R., et al. (2021). *Fairness in Cardiac MR Image Analysis: An Investigation of Bias Due to Data Imbalance in Deep Learning Based Segmentation*. Cham: Springer International Publishing, 413–423. doi: 10.1007/978-3-030-87199-4_39
- Rasal, R., Castro, D. C., Pawlowski, N., and Glocker, B. (2022). *Deep Structural Causal Shape Models*. Tel Aviv. doi: 10.1007/978-3-031-25075-0_28
- Ravi, D., Alexander, D. C., and Oxtoby, N. P. (2019). *Degenerative Adversarial NeuroImage Nets: Generating Images that Mimic Disease Progression*. Cham: Springer International Publishing, 164–172. doi: 10.1007/978-3-030-32248-9_19
- Ravi, D., Blumberg, S. B., Ingala, S., Barkhof, F., Alexander, D. C., Oxtoby, N. P., et al. (2022). Degenerative adversarial neuroimage nets for brain scan simulations: Application in ageing and dementia. *Med. Image Anal.* 75:102257. doi: 10.1016/j.media.2021.102257
- Reinhold, J. C., Carass, A., and Prince, J. L. (2021). *A Structural Causal Model for MR Images of Multiple Sclerosis*. Cham: Springer International Publishing, 782–792. doi: 10.1007/978-3-030-87240-3_75
- Rejniak, K. A., and Anderson, A. R. A. (2010). Hybrid models of tumor growth. *WIREs Syst. Biol. Med.* 3, 115–125. doi: 10.1002/wsbm.102
- Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). *Stochastic Backpropagation and Approximate Inference in Deep Generative models*. *arXiv preprint*. doi: 10.48550/arXiv.1401.4082
- Rissech, C., Schaefer, M., and Malgosa, A. (2008). Development of the femur – Implications for age and sex determination. *Forensic Sci. Int.* 180, 1–9. doi: 10.1016/j.forsci.2008.06.006
- Rohlf, F. J. (1990). Morphometrics. *Annu. Rev. Ecol. Syst.* 21, 299–316. doi: 10.1146/annurev.es.21.110190.001503
- Rohlf, F. J. (2000). On the use of shape spaces to compare morphometric methods. *Hystrix Ital. J. Mammal.* 11, 9–25. doi: 10.4404/hystrix-11.1-4134
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2021). *High-Resolution Image Synthesis with Latent Diffusion Models*. New Orleans, LA. doi: 10.1109/CVPR52688.2022.01042
- Rüthling Cachay, S., Zhao, B., Joren, H., and Yu, R. (2023). “DYfusion: a dynamics-informed diffusion model for spatiotemporal forecasting,” in *Advances in Neural Information Processing Systems*, Vol. 36, eds. A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Red Hook, NY: Curran Associates, Inc), 45259–45287.
- Russell, B., Motlagh, D., and Ashley, W. W. (2000). Form follows function: how muscle shape is regulated by work. *J. Appl. Physiol.* 88, 1127–1132. doi: 10.1152/jappl.2000.88.3.1127
- Saito, A., Kishimoto, M., Kasahara, K., Tsujikawa, M., Takakuwa, T., Yamada, S., et al. (2019). “Spatiotemporal statistical models of a human embryo,” in *International Forum on Medical Imaging in Asia 2019*, eds. H. Fujita, F. Lin, and J. H. Kim (Singapore: SPIE). doi: 10.1117/12.2522101
- Salehinejad, H., Sankar, S., Barfett, J., Colak, E., and Valaee, S. (2018). *Recent Advances in Recurrent Neural Networks*. *arXiv preprint*. doi: 10.48550/arXiv.1801.01078
- Santeramo, R., Withey, S., and Montana, G. (2018). “Longitudinal detection of radiological abnormalities with time-modulated LSTM,” in *Lecture Notes in Computer Science* (Cham: Springer International Publishing), 326–333. doi: 10.1007/978-3-030-00889-5_37
- Sarasua, I., Pölsterl, S., and Wachinger, C. (2021). “TransforMesh: a transformer network for longitudinal modeling of anatomical meshes,” in *Machine Learning in Medical Imaging* (Cham: Springer), 209–218. doi: 10.1007/978-3-030-87589-3_22
- Sarasua, I., Pölsterl, S., and Wachinger, C. (2022). *CASHformer: Cognition Aware SHape Transformer for Longitudinal Analysis*. Cham: Springer Nature Switzerland, 44–54. doi: 10.1007/978-3-031-16431-6_5
- Sauty, B., and Durrleman, S. (2022). *Progression Models for Imaging Data with Longitudinal Variational Auto Encoders*. Cham: Springer Nature Switzerland, 3–13. doi: 10.1007/978-3-031-16431-6_1
- Saxena, D., and Cao, J. (2021). Generative adversarial networks (GANs): challenges, solutions, and future directions. *ACM Comput. Surv.* 54, 1–42. doi: 10.1145/3446374

- Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., et al. (2021). Alzheimer's disease. *Lancet* 397, 1577–1590. doi: 10.1016/S0140-6736(20)32205-4
- Schiratti, J.-B., Allasonnière, S., Colliot, O., and Durrleman, S. (2017). A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *J. Mach. Learn. Res.* 18, 1–33. Available online at: <https://jmlr.org/papers/volume18/17-197/17-197.pdf>
- Schiratti, J.-B., Allasonnière, S., Routier, A., Colliot, O., and Durrleman, S. (2015). *A Mixed-Effects Model with Time Reparametrization for Longitudinal Univariate Manifold-Valued Data*. Cham: Springer International Publishing, 564–575. doi: 10.1007/978-3-319-19992-4_44
- Schön, J., Selvan, R., Nygård, L., Vogelius, I. R., and Petersen, J. (2023). *Explicit Temporal Embedding in Deep Generative Latent Models for Longitudinal Medical Image Synthesis*. *arXiv preprint*. doi: 10.48550/arXiv.2301.05465
- Schutte, K., Moindrot, O., Hérent, P., Schiratti, J.-B., and Jégou, S. (2021). *Using StyleGAN for Visual Interpretability of Deep Learning Models on Medical Images*. *arXiv preprint*. doi: 10.48550/arXiv.2101.07563
- Semler, O., Rehberg, M., Mehdiani, N., Jackels, M., and Hoyer-Kuhn, H. (2019). Current and emerging therapeutic options for the management of rare skeletal diseases. *Pediatr. Drugs* 21, 95–106. doi: 10.1007/s40272-019-00330-0
- Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., et al. (2023). Transformers in medical imaging: a survey. *Med. Image Anal.* 88:102802. doi: 10.1016/j.media.2023.102802
- Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248. doi: 10.1146/annurev-bioeng-071516-044442
- Shen, H., Zhang, J., Xiong, B., Hu, R., Chen, S., Wan, Z., et al. (2025). *Efficient Diffusion Models: A Survey*. *arXiv preprint*. doi: 10.48550/arXiv.2502.06805
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C. (2015). “Convolutional LSTM network: a machine learning approach for precipitation nowcasting,” in *Advances in Neural Information Processing Systems*, Vol. 28 (Montreal, QC).
- Singh, A., Sengupta, S., and Lakshminarayanan, V. (2020). Explainable deep learning models in medical image analysis. *J. Imaging* 6:52. doi: 10.3390/jimaging6060052
- Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2015). Hierarchical geodesic models in diffeomorphisms. *Int. J. Comput. Vis.* 117, 70–92. doi: 10.1007/s11263-015-0849-2
- Singhal, P., Walambe, R., Ramanna, S., and Kotecha, K. (2023). Domain adaptation: challenges, methods, datasets, and applications. *IEEE Access* 11, 6973–7020. doi: 10.1109/ACCESS.2023.3237025
- Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G. (2020). “Implicit neural representations with periodic activation functions,” in *Advances in Neural Information Processing Systems*, Vol. 33 (Curran Associates, Inc.), 7462–7473.
- Slice, D. E. (2005). *Modern Morphometrics*. New York, NY: Kluwer Academic Publishers-Plenum Publishers, 1–45.
- Slice, D. E. (2007). Geometric morphometrics. *Annu. Rev. Anthropol.* 36, 261–281. doi: 10.1146/annurev.anthro.34.081804.120613
- Stacke, K., Eilertsen, G., Unger, J., and Lundström, C. (2019). *A Closer Look at Domain Shift for Deep Learning in Histopathology*. doi: 10.48550/arXiv.1909.11575
- Staudemeyer, R. C., and Morris, E. R. (2019). *Understanding LSTM-A Tutorial into Long Short-Term Memory Recurrent Neural Networks*. doi: 10.48550/arXiv.1909.09586
- Sullivan, L. H. (1896). The tall office building artistically considered. *Lippincott's Monthly Magazine* 339, 403–409.
- Sun, C., Shrivastava, A., Singh, S., and Gupta, A. (2017). “Revisiting unreasonable effectiveness of data in deep learning era,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Venice: IEEE). doi: 10.1109/ICCV.2017.97
- Suzuki, Y., Matsubayashi, J., Ji, X., Yamada, S., Yoneyama, A., Imai, H., et al. (2019). Morphogenesis of the femur at different stages of normal human development. *PLOS ONE* 14:e0221569. doi: 10.1371/journal.pone.0221569
- Tajdari, M., Pawar, A., Li, H., Tajdari, F., Maqsood, A., Cleary, E., et al. (2021). Image-based modelling for adolescent idiopathic scoliosis: mechanistic machine learning analysis and prediction. *Comput. Methods Appl. Mech. Eng.* 374:113590. doi: 10.1016/j.cma.2020.113590
- Tajdari, M., Tajdari, F., Shirzadian, P., Pawar, A., Wardak, M., Saha, S., et al. (2022). Next-generation prognosis framework for pediatric spinal deformities using bio-informed deep learning networks. *Eng. Comput.* 38, 4061–4084. doi: 10.1007/s00366-022-01742-2
- Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, A., et al. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Sci. Data* 8:192. doi: 10.1038/s41597-021-00981-0
- Teng, Q., Liu, Z., Song, Y., Han, K., and Lu, Y. (2022). A survey on the interpretability of deep learning in medical diagnosis. *Multimed. Syst.* 28, 2335–2355. doi: 10.1007/s00530-022-00960-4
- Thomas Fletcher, P. (2012). Geodesic regression and the theory of least squares on riemannian manifolds. *Int. J. Comput. Vis.* 105, 171–185. doi: 10.1007/s11263-012-0591-y
- Thompson, D. W. (1992). *On Growth and Form*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107325852
- Thompson, N. C., Greenewald, K., Lee, K., and Manso, G. F. (2020). *The Computational Limits of Deep Learning*. *arXiv preprint*. doi: 10.48550/arXiv.2007.05558
- Tinker, A., Mein, G., Bhamra, S., Ashcroft, R., and Seale, C. (2009). Retaining older people in longitudinal research studies: some ethical issues. *Res. Ethics* 5, 71–74. doi: 10.1177/174701610900500206
- Toogood, P. A., Skalak, A., and Cooperman, D. R. (2008). Proximal femoral anatomy in the normal human population. *Clin. Orthop. Relat. Res.* 467, 876–885. doi: 10.1007/s11999-008-0473-3
- Torralla, A., and Isola, P. (2024). *Foundations of Computer Vision*. London: .
- Tschannen, M., Bachem, O., and Lucic, M. (2018). *Recent Advances in Autoencoder-Based Representation Learning*. *arXiv preprint*. doi: 10.48550/arXiv.1812.05069
- Vaillant, M., Miller, M., Younes, L., and Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *NeuroImage* 23, S161–S169. doi: 10.1016/j.neuroimage.2004.07.023
- van Veldhuizen, V., Botha, V., Lu, C., Cesur, M. E., Lipman, K. G., de Jong, E. D., et al. (2025). *Foundation Models in Medical Imaging-A Review and Outlook*. *arXiv preprint*. doi: 10.48550/arXiv.2506.09095
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need in *Advances in Neural Information Processing Systems*, Vol. 30 (Red Hook, NY: Curran Associates, Inc).
- Wang, H., Xiao, N., Zhang, J., Yang, W., Ma, Y., Suo, Y., et al. (2022). Static-dynamic coordinated transformer for tumor longitudinal growth prediction. *Comput. Biol. Med.* 148:105922. doi: 10.1016/j.combiomed.2022.105922
- Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., Wang, F.-Y., et al. (2017). Generative adversarial networks: introduction and outlook. *IEEE/CAA J. Autom. Sin.* 4, 588–598. doi: 10.1109/JAS.2017.7510583
- Wang, T., Lei, Y., Fu, Y., Wynne, J. F., Curran, W. J., Liu, T., et al. (2020). A review on medical imaging synthesis using deep learning and its clinical applications. *J. Appl. Clin. Med. Phys.* 22, 11–36. doi: 10.1002/acm2.13121
- Wang, V. Y., Hussain, J. R., Yousefi, H., Bradley, C. P., Hunter, P. J., Nash, M. P., et al. (2017). Modelling cardiac tissue growth and remodelling. *J. Elasticity* 129, 283–305. doi: 10.1007/s10659-017-9640-7
- Wang, Y., Naleway, S. E., and Wang, B. (2020). Biological and bioinspired materials: structure leading to functional and mechanical performance. *Bioact. Mater.* 5, 745–757. doi: 10.1016/j.bioactmat.2020.06.003
- Wang, Z., and Guan, Y. (2023). Multiscale convolutional neural-based transformer network for time series prediction. *Signal Image Video Process.* 18, 1015–1025. doi: 10.1007/s11760-023-02823-5
- Weaver, C. M., and Fuchs, R. K. (2014). *Skeletal Growth and Development*. Amsterdam: Elsevier, 245–260. doi: 10.1016/B978-0-12-416015-6.00012-5
- Wegmayr, V., Hörold, M., and Buhmann, J. M. (2019). *Generative Aging of Brain MR-Images and Prediction of Alzheimer Progression*. Cham: Springer International Publishing, 247–260. doi: 10.1007/978-3-030-33676-9_17
- Wegst, U. G. K., Bai, H., Saiz, E., Tomsia, A. P., and Ritchie, R. O. (2014). Bioinspired structural materials. *Nat. Mater.* 14, 23–36. doi: 10.1038/nmat4089
- Wescott, D. (2005). Population variation in femur subtrochanteric shape. *J. Forensic Sci.* 50:JFS2004281-8. doi: 10.1520/JFS2004281
- Wu, J., and Gong, K. (2024). *LDM-Morph: Latent Diffusion Model Guided Deformable Image Registration*. *arXiv preprint*. doi: 10.48550/arXiv.2411.15426
- Xia, K., and Wang, J. (2023). Recent advances of transformers in medical image analysis: a comprehensive review. *MedComm - Future Med.* 2:e38. doi: 10.1002/mef2.38
- Xia, T., Chartsias, A., Wang, C., and Tsiftaris, S. A. (2021). Learning to synthesise the ageing brain without longitudinal data. *Med. Image Anal.* 73:102169. doi: 10.1016/j.media.2021.102169
- Xu, B., and Yang, G. (2025). Interpretability research of deep learning: a literature survey. *Inf. Fusion* 115:102721. doi: 10.1016/j.inffus.2024.102721
- Yang, H., Lyu, J., Tam, R., and Tang, X. (2023). *A Survey on Deep Learning-Based Diffeomorphic Mapping*. Cham: Springer International Publishing, 1289–1321. doi: 10.1007/978-3-030-98661-2_108
- Yang, H., Tan, T., Tan, S., Yang, W., Cai, K., Chen, C., et al. (2025). *MambaControl: Anatomy Graph-Enhanced Mamba ControlNet with Fourier Refinement for Diffusion-Based Disease Trajectory Prediction*. doi: 10.48550/arXiv.2505.09965
- Yang, Y., Jin, M., Wen, H., Zhang, C., Liang, Y., Ma, L., et al. (2024). *A Survey on Diffusion Models for Time Series and Spatio-Temporal Data*. *arXiv preprint*. doi: 10.48550/arXiv.2404.18886
- Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: a review. *Med. Image Anal.* 58:101552. doi: 10.1016/j.media.2019.101552

- Yoon, J. S., Zhang, C., Suk, H.-I., Guo, J., and Li, X. (2023). "SADM: sequence-aware diffusion model for longitudinal medical image generation," in *Information Processing in Medical Imaging* (Cham: Springer Nature Switzerland), 388–400. doi: 10.1007/978-3-031-34048-2_30
- Younes, L. (2010). *Shapes and Diffeomorphisms*. Cham: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-12055-8
- Young, A. F., Powers, J. R., and Bell, S. L. (2006). Attrition in longitudinal studies: who do you lose? *Aus. N. Z. J. Public Health* 30, 353–361. doi: 10.1111/j.1467-842X.2006.tb00849.x
- Young, A. L., Oxtoby, N. P., Garbarino, S., Fox, N. C., Barkhof, F., Schott, J. M., et al. (2024). Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. *Nat. Rev. Neurosci.* 25, 111–130. doi: 10.1038/s41583-023-00779-6
- Zhang, L., Lu, L., Wang, X., Zhu, R. M., Bagheri, M., Summers, R. M., et al. (2020). Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. *IEEE Trans. Med. Imaging* 39, 1114–1126. doi: 10.1109/TMI.2019.2943841
- Zhang, L., Rao, A., and Agrawala, M. (2023). *Adding Conditional Control to Text-to-Image Diffusion Models*. Paris: IEEE. doi: 10.1109/ICCV51070.2023.00355
- Zhang, Z., Song, Y., and Qi, H. (2017). "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI: IEEE). doi: 10.1109/CVPR.2017.463
- Zhao, Q., Liu, Z., Adeli, E., and Pohl, K. M. (2021). Longitudinal self-supervised learning. *Med. Image Anal.* 71:102051. doi: 10.1016/j.media.2021.102051
- Zhou, G., Xie, S., Hao, G., Chen, S., Huang, B., Xu, X., et al. (2023). *Emerging Synergies in Causality and Deep Generative Models: A Survey*. doi: 10.48550/arXiv.2301.12351