Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Brain Morphogenesis via Physics-Transfer Learning

Zhiping Xu

xuzp@tsinghua.edu.cn

Tsinghua University

Yingjie Zhao

Tsinghua University https://orcid.org/0009-0000-8336-275X

Yicheng Song

Fudan University

Fan Xu

Fudan University https://orcid.org/0000-0003-3910-5398

Article

Keywords:

Posted Date: February 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8599229/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Computational Science on September 9th, 2026. See the published version at https://doi.org/10.1038/s43588-026-01040-7.

Predicting Brain Morphogenesis via Physics-Transfer Learning

Yingjie Zhao[1], Yicheng Song[2], Fan Xu*[2], and Zhiping Xu*[1]

[1]Applied Mechanics Laboratory, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China
[2]Institute of Mechanics and Computational Engineering, Department of Aeronautics and Astronautics & College of Intelligent Robotics and Advanced Manufacturing, Fudan University, Shanghai 200433, China
*Corresponding authors: Fan Xu (fanxu@fudan.edu.cn), Zhiping Xu (xuzp@tsinghua.edu.cn).

Abstract

Brain morphology emerges from the interplay of genetic programming and mechanical forces. However, its fractal-like folding patterns, regional anisotropy, and uneven curvature distributions pose longstanding challenges for quantitative analysis in medical and developmental studies. These difficulties are exacerbated by the limited availability of labeled data for existing statistical learning approaches. Here we introduce a theory-grounded physics-transfer (PT) learning framework that enables generalization analysis and prediction in complex physical systems by leveraging the continuity of underlying physical laws across different levels of complexity. Specifically, mechanistic insights into nonlinear elasticity of simple, analytically tractable geometries are embedded directly into neural networks and then transferred to brain models through a sequence of physics-anchored domains, providing, for the first time, a formulation that integrates explicit physics constraints into learning theory. In this sense, PT advances classical statistical learning and domain adaptation theories by enabling generalization analysis in regimes where purely statistical assumptions are insufficient. Building on PT, we derive a generalization bound that explains the model's strong performance in feature characterization and morphogenesis prediction for brain. The results reveal that localized deformation modes, rather than global geometry, dominantly shape cortical folding dynamics. Beyond predictive accuracy, the framework also yields reduced-dimensional evolutionary representations that distill the essential physics of the highly folded cerebral cortex. Validation against medical imaging data and expert clinical assessment demonstrates the promise of physics-aware digital-twin technologies for understanding, diagnosing, and ultimately intervening in the morphological complexity of the developing and diseased brain.

In the scientific development and engineering applications, theoretical frameworks have traditionally followed an observation-assumption-modeling paradigm. From Galileo's early investigations into beam bending to Cauchy's formulation of continuum mechanics, this approach has proven effective for problems in mechanics with low-dimensional parameter spaces. In such cases, analytical models have often sufficed to capture the essential physical behaviors (Fig. 1). However, as scientific inquiry advances into the multiscale analysis of matter, the well-known 'curse of complexity' emerges. The material behavior exhibits pronounced multiscale coupling and non-equilibrium characteristics, which pose significant challenges to the conventional observation-assumption-modeling paradigm.

A striking example lies in brain development, a multi-scale physical process that spans from molecular gene expression and protein folding to cellular division, differentiation, and migration, culminating in macroscopic cortical folding (Fig. S1) [1, 2]. Genetic programs that direct anisotropic growth interact with mechanical forces to sculpt the brain's structure. These processes unfold over time and space, giving rise to the evolving morphology observed throughout life [1, 2]. The brain's geometry is highly intricate. Neurons are embedded in a dense, interconnected network, and its surface is patterned with gyri and sulci [3], which are key features intimately linked to cognition and function [4, 5]. Morphological patterns also serve as sensitive indicators of neurodevelopment, disease progression, and aging [6, 7]. Understanding how these features emerge and evolve could illuminate mechanisms of morphogenesis, functional specialization, and pathology. Progress hinges on integrating high-resolution observations with models capable of predicting individualized developmental trajectories.

Magnetic resonance imaging (MRI) is the primary tool for capturing in vivo brain morphology [8]. Yet, MRI studies are predominantly cross-sectional, with longitudinal datasets remaining scarce (Fig. 2) [9]. Resolution limits hinder the capture of fine cortical folding patterns, which are non-uniform, hierarchical, and individual-specific. Moreover, MRI provides geometry but not underlying physiological drivers, such as growth stresses responsible for folding (Fig. 2) [10, 11], making predictive diagnosis and prognosis difficult [12, 13]. Population-averaged models offer valuable baselines [14] but obscure the marked variability in neuroanatomical maturation between individuals. This variability limits the diagnostic value of normative curves and hampers early detection of atypical development [15, 16]. Computational approaches capable of forecasting individual-specific morphological trajectories could enable earlier intervention, better disease subtype stratification, and personalized therapeutic design. Continuum mechanics offers a natural multiscale framework for brain morphogenesis [17], incorporating genetic and biochemical cues via growth tensors, and mechanical properties, such as stiffness and viscoelasticity, via constitutive laws. These elements together drive local expansion, deformation, and mechanical instabilities. High-fidelity simulations can recover characteristic patterns such as sulci, gyri, and bifurcations [18–22]. However, the brain's geometric complexity makes such simulations computationally expensive, and the paucity of high-quality MRI data further impedes comprehensive digital reconstructions.

Recent advances in artificial intelligence (AI) and machine learning (ML) offer a promising data-driven alternative to traditional approaches. While often constrained by data density and coverage, these emerging methods are becoming increasingly powerful end-to-end solutions as data quality and quantity improve beyond a critical level. Enhanced interpolation and extrapolation capabilities allow AI and ML to augment conventional theoretical techniques, particularly in scenarios where direct solutions are impractical or intractable [23–26]. Despite the expressive power of well-trained neural networks (NNs), their internal representations often lack interpretability, making it challenging to uncover underlying physical principles and ensure generalizability [27]. Generative AI offers an appealing route to augment datasets, but purely data-driven approaches trained only on morphological data struggle with generalization and physical plausibility [28].

Embedding physical priors into AI models helps mitigate challenges related to generalization and interpretability [29]. There are two main approaches to integrating physics, depending on the strength of the imposed physics constraints. Physics-informed neural networks (PINNs) impose strong constraints on NN predictions in a mathematical form by embedding the governing partial differential equations (PDEs) or conservation laws, and bound-

ary conditions, e.g., directly into the loss function [30]. Strong physics constraints can also be incorporated by embedding symmetry into NN architectures [31]. However, many complex physical systems are governed by nonlinear, often ill-posed or partially known PDEs, where symmetries may be broken and high-quality data are scarce (Fig. 1a). This makes it impractical to impose strong physical constraints on NNs in tasks such as predicting brain morphogenesis [32], where explicit conservation laws are unavailable and limited high-fidelity data further hinder direct ML training and deployment. To address these challenges, particularly poor generalization and data scarcity, we propose the physics-transfer (PT) concept (Fig. 1b) [33]. In PT learning, a comprehensive digital library (DL) containing high-density and large-coverage data is first constructed to enable ML models to learn the fundamental physics in simplified or well-characterized settings (Fig. 1c). These learned physics are then transferred and adapted to more complex, data-scarce scenarios through a chain of anchored domains (CADs). A recent study demonstrated this concept by transferring crystal plasticity physics from low-fidelity force-fields to a model fed with high-fidelity features, achieving strength predictions with quantum-chemistry-level accuracy while maintaining a low computational cost for dataset construction [34]. This success highlights the feasibility and necessity of a unified theoretical framework for physics-transfer learning.

In this work, we present the PT theory, and apply the framework to resolve the physical complexity underlying brain morphological development. The PT approach provides provably bounded generalization error and uncertainty quantification through the analysis of neural activation patterns. The key idea is to first learn the nonlinear deformation and bifurcation physics from simplified geometries using high-fidelity continuum mechanics, then transfer this knowledge, encoded in NN weights, to predict individual brain evolution. We construct a dense, high-coverage DL of spherical and ellipsoidal growth models as the anchor domains to capture universal bifurcation behaviors, and then apply the learned physics to human brain geometries. Analysis of this data-driven model uncovers the physical principles encoded within the NN, emphasizing the predominant influence of cortical folding over background configurational curvature. These results underscore the pivotal role of localized deformation in driving morphogenesis. Predictions are validated against MRI-derived data, achieving high accuracy while offering interpretable insights. Our results demonstrate that PT learning bridges data scarcity and physical complexity, enabling realistic, temporally continuous reconstructions of brain development. This approach paves the way for digital twin models of the brain, linking morphology, mechanics, and function across scales.

Results

Physics-transfer learning theory

Statistical learning (SL) theory provides the theoretical foundation for modern ML by guaranteeing bounded generalization error when the training and test data are independently and identically distributed (i.i.d.), thereby explaining much of ML's empirical success over the past decades [35]. Domain adaptation (DA) theory extends this framework to settings where the training and test distributions differ, enabling analysis and learning under distributional shifts [36]. Specifically, the generalization error of DA can be bounded by

$$\varepsilon_{\mathcal{T}}(h) \leq \widehat{\varepsilon}_{\mathcal{S}}(h) + d(\widehat{\mathcal{D}}_{\mathcal{S}}^{X}, \widehat{\mathcal{D}}_{\mathcal{T}}^{X}) + \lambda^{*} + O\left(\sqrt{\frac{d\log n + \log(1/\delta)}{n}}\right), \tag{1}$$

where $\varepsilon_{\mathcal{T}}(h)$ denotes the expected error of hypothesis h on the target domain ($\mathcal{T}$), $\widehat{\varepsilon}_{\mathcal{S}}(h)$ is the empirical error on the source domain ($\mathcal{S}$), $d(\widehat{\mathcal{D}}_{\mathcal{S}}^{X}, \widehat{\mathcal{D}}_{\mathcal{T}}^{X})$ measures the divergence between the source and target distributions on inputs $\mathbf{x} \in \mathcal{X}$, $\lambda^{*} = \min_{h\in\mathcal{H}}\left(\varepsilon_{\mathcal{S}}(h) + \varepsilon_{\mathcal{T}}(h)\right)$ is the ideal joint error of the optimal hypothesis across both domains, and the last term is a statistical learning term that captures the generalization uncertainty induced by the hypothesis space complexity (measured by the Vapnik–Chervonenkis (VC) dimension d), sample size n, and

confidence parameter δ. Existing SL theories offer a limited quantitative understanding of λ^*, lack mechanisms to anticipate or diagnose failures when their assumptions no longer hold, and provide few systematic strategies to address such issues [36–38]. These limitations largely stem from the inherent constraints of purely statistical formulations, especially in the absence of labeled data from the target domain with the same or different physics.

This limitation becomes particularly pronounced in complex scientific problems, such as biological evolutionary dynamics and brain development, where direct solutions are intractable and labeled target data are scarce or unobtainable. To address these challenges, we propose the PT learning theory, which extends DA learning theory by incorporating a physical interpretation of the conditional probability and leveraging the hierarchical structure of physical principles (see Methods for details). Specifically, we formalize the conditional probability ($P(Y|X)$) as an instance of the underlying physical mapping ($\mathcal{P}$), enabling us to more naturally identify and define anchor domains ($\mathcal{A}$) to evaluate λ^* between the source and target domains beyond statistical approaches. The CADs ($\mathcal{A}_k, k = 1, ..., K$) are defined as

$$\widehat{\mathcal{A}}_k = \left\{ \left(\mathbf{x}_k^i, \mathbf{y}_k^i\right) \right\}_{i=1}^{l_k}, \ \mathbf{y}_k^i = \mathcal{P}\left(\mathbf{x}_k^i; \xi_{\mathcal{A}_k}\right), \ \xi_{\mathcal{A}_k} \in [\xi_{\mathcal{S}}, \xi_{\mathcal{T}}], \tag{2}$$

where $\mathcal{P}$ denotes the shared physical mapping underlying both the source and target domains, l_k is the number of samples in the anchor domain $\mathcal{A}_k$, k indexes multiple anchor domains that can be found between the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$, and ξ is the fidelity or complexity. Importantly, ξ implicitly restricts the hypothesis class to physically realizable mappings, thereby reducing the effective model complexity compared to unconstrained statistical learning. The CADs can be constructed by monotonically increasing the fidelity parameter ξ, ensuring a smooth transition from the source domain $\mathcal{S}$ to the target domain $\mathcal{T}$ with progressively rising complexity (Fig. 1d, and see Method for details). The validity of the CADs can be assessed through zero-shot prediction performance between neighboring domains.

Within this framework, we can define indicators ($\lambda^{\mathcal{A}_k}$ and $\delta^{\mathcal{A}_k}$) to quantitatively estimate the generalization error and assess the feasibility of physics transfer, that is

$$\lambda^{\mathcal{A}_k} = \frac{1}{2}\left[d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Y}, \widehat{\mathcal{D}}_{\mathcal{A}_k}^{Y}) - d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{A}_k}^{Z})\right]^2, \tag{3}$$

$$\delta^{\mathcal{A}_k} = \frac{d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{T}}^{Z})}{d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{A}_k}^{Z})}, \tag{4}$$

where Z is the learned latent physical representation (see Methods for details). Rather than directly bounding the ideal joint error λ^*, which is generally unobservable, PT learning introduces physically motivated indicators $\lambda^{\mathcal{A}_k}$ and $\delta^{\mathcal{A}_k}$ that serve as operational surrogates for assessing the generalization. PT learning employs the physical mapping $\mathcal{P}$ as a weak physics constraint to learn these physical representations Z in the latent space, where the overall process can be expressed as

$$X \xrightarrow{\mathcal{P}_{\text{enc}}} Z \xrightarrow{\mathcal{P}_{\text{dec}}} Y, \text{ with } \quad Z = \mathcal{P}_{\text{enc}}(X; \xi), \ Y = \mathcal{P}_{\text{dec}}(Z; \xi), \tag{5}$$

and $\mathcal{P} = \mathcal{P}_{\text{dec}} \circ \mathcal{P}_{\text{enc}}$ represents the complete physical mapping from input X to output Y. Leveraging this weak physics constraint as well as the high data density and coverage of low-fidelity models (the source domain), PT learning naturally achieves alignment in the latent space without introducing additional discrepancy or adversarial losses, avoiding distortion of the underlying physical patterns in the target domain (Fig. 1d). While classical generalization bounds explicitly depend on sample size n, PT learning extends explicit sample complexity with anchor domain coverage (l_k) and physical continuity in CADs, providing a pathway to control generalization uncertainty in low-label or zero-label regimes.

%What PT theory enables? Overall, PT learning theory provides a principled framework to analyze and control generalization in complex physical systems where geometry, scale, and model fidelity evolve, thereby enabling learning and prediction in regimes with little or no labeled target data. Compared with classical SL and DA theories, PT theory targets physical problems in which SL and DA become theoretically ineffective by reinterpreting source-target discrepancy not as an abstract distribution shift with an unobservable ideal joint error, but as a complexity-driven transformation parameterized by ξ and systematically controlled through CADs. Through these anchor domains, generalization can be operationally assessed and diagnosed, transforming an otherwise intractable theoretical quantity into physically interpretable indicators. A small value of $\lambda^{\mathcal{A}_k}$ and $\delta^{\mathcal{A}_k}$ indicates a higher likelihood of successful physics transfer and a lower generalization error (see Methods for details). Here, $\lambda^{\mathcal{A}_k}$ measures the feasibility of transferring physics from the source domain to the anchor domain, while $\delta^{\mathcal{A}_k}$ quantifies the physical dissimilarity between the anchor and target domains. More fundamentally, PT learning elevates physics from a role traditionally limited to regularization terms, loss functions, or priors to a structured domain chain that explicitly restricts the hypothesis space and provides estimators of generalization, enabling physically governed but statistically intractable shifts to be understood and controlled. In this sense, PT theory extends classical SL and DA theories by broadening generalization analysis to regimes where i.i.d. assumptions and purely statistical notions of distribution shift are insufficient.

The effectiveness of this framework in tackling complex problems is demonstrated through its application to resolve the complexity of brain morphological development. The morphogenesis problems can be characterized by varying geometric complexity ξ, and the underlying physics $\mathcal{P}$ involves nonlinear elasticity and bifurcation behaviors. Multiple anchor domains $\mathcal{A}_k$ could in principle be introduced, such as ellipsoids with different axis ratios, or saddle surfaces with negative Gaussian curvature, or even an animal's brain providing intermediate geometric features (see Methods for details). For simplicity, in this study, we train the models only on the source domain $\mathcal{S}$ (sphere), select only a single, simple ellipsoid as the anchor domain $\mathcal{A}_1$ to evaluate the feasibility of PT, and directly apply the trained model to predict brain morphology in the target domain $\mathcal{T}$.

Brain models

Brain morphological development arises from the interplay of physical complexity, encompassing genetic regulation and mechanical forces, and geometric complexity. The multiscale physics governing this process can be described within the framework of continuum mechanics, in which the growth tensor serves as a key physiological parameter for quantifying tissue growth kinetics (see Methods for details) [15, 17, 19–21, 39]. These parameters bridge microscale cellular behaviors and microstructural organization, enabling multiscale models that capture morphological instabilities [40]. Within this framework, the nonlinear deformation physics underlying cortical pattern formation, manifested in sulci, gyri, and bifurcation structures, can be faithfully reproduced (Fig. 2). When coupled with ML, these mechanical models offer a foundation for digital brain twins to inform patient-specific clinical decisions and precision medicine, though further refinement is required [15, 41].

Yet, modeling brain morphological development remains challenging, owing to the intrinsic complexity of neurodevelopment and the scarcity of longitudinal MRI datasets. Finite element analysis (FEA) is widely used to simulate brain growth and generate synthetic data, but geometrical nonlinearity leads to low computational efficiency and poor convergence (Fig. 3a) [19]. Consequently, most studies avoid direct simulation of realistic brain geometries, focusing instead on idealized two-dimensional (2D) shells or three-dimensional (3D) spheres and ellipsoids [17, 20–22]. This geometric complexity hampers the construction of a comprehensive digital brain library from limited MRI datasets, resulting in severe data sparsity (Fig. 3b).

Interestingly, even such simplified geometries exhibit spatiotemporal features reminiscent of brain morphology, including vertex-ridge networks and bifurcation patterns. Motivated by these parallels, we construct a high-density, high-coverage DL of morphological evolution using simplified geometries, and transfer the learned physics

to complex brain structures. This strategy overcomes data sparsity and enables physics-informed modeling of realistic brain development (Fig. 3c).

Digital libraries

Brain development data are both distinct and scarce. Even the most comprehensive MRI repository, comprising $123,984$ scans, contains only $\sim 10\%$ longitudinal records, often acquired at irregular intervals [9]. For brain morphology specifically, a single atlas is currently available [42]. This scarcity of high-quality, temporally consistent datasets poses a critical barrier to modeling and understanding brain morphogenesis.

To address this gap, we construct digital libraries of morphological patterns based on spheres and ellipsoids (Fig. S2, and see Methods for details). These are modeled using a representative core-shell framework [18, 22, 43–46], widely employed to investigate mechanical instabilities in cortical folding [15, 17, 19–21, 39]. In this framework, the outer shell represents the cerebral cortex (gray matter) and the inner core the white matter. Both are modeled as modestly compressible, hyperelastic Neo-Hookean materials with distinct growth rates [19]. Following experimental observations [11, 47–50], cortical thickness varies from $0.03 - 1.63\,\mathrm{mm}$, encompassing both normal and abnormal human measurements, while the relative shear modulus ($G_{\mathrm{grey}}/G_{\mathrm{white}}$) spans $0.65 - 1$.

In neurodevelopment, biomarkers are quantifiable anatomical or physiological features indicative of underlying processes or deviations. Cortical thickness is a key biomarker, influencing brain morphology and serving in the diagnosis of cortical malformations, including lissencephaly, polymicrogyria, and focal cortical dysplasia [22]. To simulate the growth stresses driving pattern evolution, we adopt the tangential growth (TG) model, which captures the cellular mechanisms that generate morphological instabilities [2, 18, 19].

For predictive modeling, ML models are trained on datasets derived from these simple geometries, then applied zero-shot to brain morphology, avoiding the high computational costs of direct modeling, which must resolve both global geometry (structure, shape, size) and local features (cortex, subcortical structures, ventricles) (see Methods for details). Sampling from the morphological development space of simple geometries is markedly more efficient than from brain tissues (Fig. 3c), while preserving spatiotemporal similarities to brain development. This enables robust generalization to out-of-distribution (OOD) brain data.

The PT approach proceeds by training on in-distribution (ID) simple geometries with an enriched sampling space, allowing the model to learn essential physical principles underlying morphological change. These learned representations are then transferred to brain models via NNs with fixed weights that encode the core physics. This strategy ensures that the robustly generalizes from the simple geometries (ID) to that of the more complex and data-scarce brain morphologies (OOD). Verified predictive accuracy and validation with experimental data demonstrate the success in resolving the accuracy-performance dilemma imposed by geometric complexity.

Predictions of curvature maps and 3D morphology

We begin with a simple scenario where the goal is to predict biomarkers such as the curvature map from 3D morphological data, formulated as a regression task (Figs. 4a, S3, and see Methods for details). Using an encoder-decoder graph neural network (GNN) architecture (Fig. S4), we represent morphology as a graph where each node encodes features including spatial coordinates and normal vectors. The model predicts local curvature values, which have been demonstrated as key descriptors of morphological complexity [51, 52], closely associated with neurodevelopmental trajectories, and serve as indicators of neurological function and disease.

As a control, we trained a SL model that excluded normal vectors from the inputs, retaining only morphological data and thereby limiting its ability to capture the physics underlying curvature, which depends on nonlinear elasticity (see Methods for details). Both SL and PT methods accurately predict curvature, computed as the sum

of absolute mean and Gaussian curvatures, $|H| + |K|$ (Fig. 4b). However, when applied to ellipsoids and experimental human brain data obtained from experimental MRI images, PT learning significantly outperforms SL, achieving mean absolute errors (MAEs) between the predicted and ground-truth curvature maps (reconstructed from MRI images) below 0.5 mm^{-1}, compared to 7.01 mm^{-1} for the latter (Figs. 4c and 4d). This validation using model ellipsoidal data and experimental brain data underscores the outstanding generalizability provided by the physical principles embedded within the NNs (Figs. 4e-4g).

We then extend to a spatiotemporal prediction problem to forecast future morphological development from existing data. This capability is vital for early diagnosis, disease monitoring, and personalized treatment planning. Due to the scarcity of longitudinal brain data, models are pretrained on simple geometry libraries before transferring learned physics to brain morphologies.

An encoder-decoder GNN [53] processes morphology graphs whose nodes carry spatial and normal features, while edges represent mesh connectivity or contact events, characterized by nodal distances in reference and current configurations. The encoder module embeds input features into a latent space through message-passing in graph convolution. The decoder module decodes latent features into nodal acceleration at each node. An integration module based on Newton's second law is embedded into the network to compute the position increments of the nodes and update their positions at the next time step [53]. During inference, an autoencoder or iterative rollout is employed, enabling a digital twin approach to predict brain morphological development for medical applications (Fig. 4h).

For one-step prediction (i.e., predicting brain morphology one week ahead), the PT models achieve high accuracy, with MAE between prediction and the ground truth reconstructed from MRI images below 0.01 (Figs. 4i and S5). Although errors increase over longer horizons (e.g., five weeks), visual inspection confirms preservation of key structural features. Clinical relevance dictates error thresholds. Global metrics like head circumference allow higher tolerance, whereas diagnoses sensitive to local detail, such as schizophrenia, require stricter limits [54]. Beyond these thresholds, manual corrections guided by the digital twin framework become necessary (Figs. 4h and 4i).

Finally, we quantify the relationship between latent space features and model generalizability to enable a priori estimation of model performance (Fig. S6a, and see Methods for details). The latent-space distance between training spheres and testing brain data positively correlates with prediction error (Figs. S5 and S6b). Short-term inputs (1-step) exhibit minimal latent distance and lowest errors, whereas longer-term inputs (7-step) show increased distance and degraded generalization (Figs. S6c and S6d). This latent distance metric provides an interpretable estimate of prediction uncertainty and serves as an early warning for potential model failure.

Discussion

Alignment with medical images

MRI brain imaging provides crucial, non-invasive insights into neurological anatomy and function, aiding diagnosis [14, 42]. However, predictive modeling of neurological diseases is limited by the scarcity and low fidelity of MRI data [12, 13]. In particular, longitudinal MRI scans for individuals are sparse, hindering accurate tracking of brain morphological development. Moreover, MRI primarily captures geometric features and lacks detailed information on biomechanical forces driving cortical folding, constraining morphology prediction from images alone (Fig. 2).

By integrating these limited clinical images with a richly sampled DL and PT, we overcome these challenges. Our DL provides dense, comprehensive data that encode nonlinear elasticity and folding mechanics within ML models. This fusion enables PT models to accurately predict brain morphology using sparse imaging data. Further-

more, augmenting these data with multimodal inputs, such as genetic, cellular, and biomechanical information, promises to enhance predictive diagnostics and data integration in neurodevelopment.

Reduced-dimensional representations

Reduced-dimensional models are essential for advancing scientific understanding and engineering of complex systems. They distill underlying physical principles from observational data, deepening insight and enabling tasks such as control, design, and decision-making. Compared to full-order models, reduced models offer dramatic gains in computational efficiency, often reducing dimensionality by several orders of magnitude, thereby supporting scalability and real-time simulation. Yet, constructing such models remains challenging, especially for systems governed by nonlinear, multiscale, and non-equilibrium physics.

Brain morphological development exhibits evolutionary patterns of vertex–ridge networks resembling those seen in spherical growth and shrinkage (Fig. 5a), suggesting that a reduced-dimensional description of these dynamics is feasible. Quantitative metrics, including vertex and ridge counts, track morphological evolution over time (Figs. 5b, 5c, and Supplementary Note S1) [55]. However, deriving governing equations is complicated by nucleation, annihilation, mutual interactions, and coupling with the surrounding tissue environment. Our PT framework produces rich, high-fidelity reduced-dimensional representations of brain morphology, laying the groundwork for future data-driven discovery of dynamical equations governing vertex–ridge network evolution [27].

Model interpretability

On the other hand, AI and ML provide powerful tools for reduced-dimensional modeling by enabling automatic extraction of governing dynamics from observational data. From a connectionist viewpoint, NNs approximate complex mappings between system variables, capturing intricate, high-dimensional behaviors [56]. Despite the expressive power of well-trained NNs, their internal representations often lack interpretability, making it challenging to extract physically meaningful laws and underscoring the persistent gap between connectionist learning and the understanding required for scientific discovery.

To close this gap, we apply dimensionality reduction to the internal representations of well-trained NNs to uncover interpretable physical insights emerging organically within the network (Figs. 5d, 5e, and Supplementary Note S1) [57]. For brain morphology, the PT-trained model exhibits similar distributions of weights Θ, i.e., $p(\Theta|\mathcal{D}_{\mathcal{S}}^{'}) \approx p(\Theta|\mathcal{D}_{\mathcal{A}}^{'})$), across datasets of varying geometrical complexity (spheres and ellipsoids), evaluated by mean layer-wise weights evaluated through the mean value μ (Fig. 5d). In contrast, SL models show marked differences in parameter distributions between low- and high-complexity datasets ($p(\Theta|\mathcal{D}_{\mathcal{S}}) \neq p(\Theta|\mathcal{D}_{\mathcal{A}})$) as compared to PT learning (Fig. 5e). This consistency in PT weight distributions reflects improved generalizability by aligning physics across domains [58].

Neuron activation patterns further illuminate PT's superior OOD performance (Figs. 5f and 5g), consistent with curvature mapping results (Fig. 4d). PT models trained on spherical data produce activation states resembling those elicited by brain morphology data (Fig. 5f), whereas SL models exhibit divergent activations across these domains (Fig. 5g).

In the latent space, the distribution distances measured by the Jensen-Shannon divergence (JS) distance (see Methods for details) between the anchor domain (ellipsoid) and the source domain (sphere), $d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{A}}^{Z})$, are 0.288 and 0.251 for PT and SL, respectively (Figs. 4e and 4f). The distribution distance of data in the output space is $d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Y}, \widehat{\mathcal{D}}_{\mathcal{A}}^{Y}) = 0.106$ (Fig. S7). Because the target-domain output space ($\widehat{\mathcal{D}}_{\mathcal{T}}^{Y}$) lacks ground-truth annotations, the discrepancy $d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Y}, \widehat{\mathcal{D}}_{\mathcal{T}}^{Y})$ cannot be directly evaluated. This limitation prevents a complete assessment of the generalization bound in the DA framework. The comparable and relatively small $\lambda^{\mathcal{A}} < 0.02$ values (see Meth-

ods for details) indicate that transferring physics from the sphere to the ellipsoid is feasible for both PT and SL, as the geometric deviation between the two domains is modest (Fig. 4c). In contrast, the distribution distances between the source domain (sphere) and the target domain (brain) in the latent space are $d(\widehat{\mathcal{D}}^Z_{\mathcal{S}}, \widehat{\mathcal{D}}^Z_{\mathcal{T}}) = 0.378$ for PT ($\delta^{\mathcal{A}} = 1.31$) and 0.581 for SL ($\delta^{\mathcal{A}} = 2.32$), respectively, representing increases of 37.7% and 131.47% compared to the anchor domain (Figs. 4e and 4f). The substantial increase observed for SL suggests a pronounced distribution mismatch in the latent space, implying poor generalization of SL in the target domain (Figs. 4d and 4g). Conversely, the relatively small increase for PT indicates a minor distribution shift, supporting its effective prediction performance in the target domain (Figs. 4d and 4g).

Brain morphogenesis features structural hierarchies spanning multiple spatial scales, including large-scale configuration curvature and smaller-scale deformation curvature such as wrinkling wavelength and radius (Fig. S8a). To quantify the relative importance of these scales, we leverage ML models coupled with information bottleneck (IB) theory [59, 60]. The latent features in ML models encode curvatures at different scales in distinct ways, enabling the model to disentangle and assess their respective contributions to morphogenesis. Within this framework, the mutual information between input and latent representations, denoted as $I(X;H)$, quantifies information compression during training (Fig. S8b). As training progresses, $I(X;H)$ converges, indicating that the model has distilled essential latent features that capture the fundamental physical principles of morphogenesis (Figs. S8c, S8d, and Supplementary Note S2). Monitoring information compression shows that background configurational curvature undergoes greater compression than deformation curvature, indicating that deformation curvature plays a more critical role in morphogenesis (Fig. S8d, and Supplementary Note S2).

Perspectives and limitations

In medical applications, while conventional diagnostic approaches rely primarily on structural MRI-based anatomical features, our PT framework integrates the underlying physical principles of brain morphogenesis, providing predictive modeling of developmental trajectories. This physics-informed approach complements existing clinical diagnostic pipelines (Fig. S9, and Supplementary Note S3).

Our model exhibits limitations in predicting brain morphological changes arising from rare events such as severe traumatic brain injuries (e.g., motor vehicle accidents, gunshot wounds, high-impact falls, explosions), uncommon infectious diseases (e.g., Creutzfeldt-Jakob disease, cerebral malaria), and exposure to extreme environments such as spaceflight (Fig. S9) [61–63]. Accurate diagnosis of brain conditions linked to these infrequent occurrences requires further evaluation that integrates patient history and clinical presentation under expert medical guidance.

The deformation physics in our DL is based on established constitutive models of brain development, which involve complex formulations and parameter uncertainties [17]. Improved understanding through theoretical, experimental, and ML-driven models will enhance accuracy [64]. By integrating diverse state-of-the-art models, the PT framework surpasses reliance on any single FEA approach. This also applies to growth factors linked to biomolecular kinetics [2, 17], where multi-scale data integration can deepen insight into brain morphogenesis.

The predictive accuracy of PT depends on the quality of its digital libraries, constrained by existing theoretical and experimental knowledge. Expanding MRI datasets, such as via the Developing Human Connectome Project [65], could further improve predictions, beyond this study's scope.

Conclusion

We present a PT learning framework that integrates high-fidelity continuum mechanics modeling with NNs to advance the understanding and prediction of brain morphogenesis. Exploiting shared biophysical principles across

simple and complex geometries, the approach overcomes challenges of data scarcity and geometric complexity. Unlike generative models that map distributions statistically, PT learning synthesizes developmental trajectories by embedding the physics of spatiotemporal evolution derived from spherical and ellipsoidal geometries. The resulting low-dimensional representations isolate the essential mechanisms of cortical folding in brain development, with large-scale configurational curvature shown to be irrelevant. Validation against medical imaging demonstrates the framework's predictive fidelity. At the theoretical level, PT learning establishes a physics-grounded generalization principle that extends classical statistical learning and domain adaptation theories by rendering generalization analyzable and controllable through chains of physics-anchored domains, even in regimes where purely statistical assumptions break down. These results underscore the potential of physics-informed digital twins to bridge mechanistic modeling and data-driven inference, enabling deeper insight and improved forecasting of brain developmental dynamics.

Methods

Physics-transfer learning framework

Problem settings

We begin by formulating the problem within the context of physical sciences based on the framework of statistical learning (SL) theory [35]. The input of models $\mathbf{x} \in \mathcal{X}$ can represent a scalar (e.g., a one-dimensional spatial or temporal coordinate), a vector (e.g., high-dimensional coordinates, a set of features or descriptors), a matrix (e.g., microstructures in pixel images), or even a higher-order tensor (e.g., microstructure in voxels or its evolution in time series). There is an unknown target function $f\colon \mathcal{X} \to \mathcal{Y}$, representing a physical mapping or operator, where $\mathcal{X}$ denotes the input space and $\mathcal{Y}$ the output space (e.g., physical quantities, fields, or evolving microstructures). A dataset $\mathcal{D}$, consisting of input-output pairs $(\mathbf{x}_1, \mathbf{y}_1), \cdots, (\mathbf{x}_N, \mathbf{y}_N)$, is typically obtained from experiments or simulations. For each $n = 1, \ldots, N$, the output satisfies $\mathbf{y}_n = f(\mathbf{x}_n)$. A learning algorithm $\mathcal{G}$ uses the dataset $\mathcal{D}$ to select a function $h\colon \mathcal{X} \to \mathcal{Y}$ that approximates the unknown mapping f. The function h is selected from a predefined set of candidate functions, referred to as the hypothesis set $\mathcal{H}$. The goal of SL theory is to ensure that the learned function $g \in \mathcal{G}$, trained on the dataset $\mathcal{D}$, can generalize reliably to unseen samples drawn independently from the same distribution.

SL theory typically assumes that training and test data are independently and identically distributed (i.i.d.) to bound generalization errors [35]. However, in complex problems in physical sciences that are characterized by multiscale behaviors, multiphysics coupling, and non-equilibrium dynamics, the target domain often lacks direct solutions as labeled data for training. This limitation reflects the inherent trade-off between accuracy and efficiency in simulations and the limited coverage and quality of experimental measurements. For example, tasks such as modeling non-equilibrium phenomena with quantum-chemistry accuracy, characterizing atomic stresses in materials, or analyzing biomechanical stresses in the brain often lack direct solutions. Inspired by the progressive problem-solving paradigm in the physical sciences, where difficult problems are tackled by building on earlier results and established physical laws, we use NNs to digitalize this process. Specifically, we first train models in domains where direct solutions are available, then transfer the learned physics to domains where such solutions are difficult or impossible to obtain.

In this setting, the learning task can be formulated as follows. The source domain dataset $\widehat{\mathcal{S}}$ is generated using low-fidelity or low-accuracy methods that compute both inputs $\mathcal{X}$ and corresponding solutions $\mathcal{Y}$. These methods enable efficient labeling of non-equilibrium responses or evolving microstructures due to their high computational efficiency,

$$\widehat{\mathcal{S}} = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^n , \tag{6}$$

where n denotes the number of labeled samples. In contrast, the target domain dataset $\widehat{\mathcal{T}}$ is is obtained from high-fidelity or high-accuracy methods,

$$\widehat{\mathcal{T}} = \left\{\mathbf{x}_i^t\right\}_{i=1}^m , \tag{7}$$

which contains m unlabeled samples, constrained by the accuracy-efficiency trade-off or experimental limitations. The objective is to ensure that the physics learned from the source domain $\mathcal{S}$ can generalize effectively to the target domain $\mathcal{T}$, even under distributional shifts and physical bias relative to the training data.

In the machine learning (ML) community, domain adaptation (DA) or domain generalization address the challenge of distributional shift [36–38, 66–68]. However, when labeled data from the target domain are unavailable, purely statistical approaches face inherent limitations. Theoretical guarantees remain loose, and failures often occur under conceptual shifts in the conditional distribution $P(Y|X)$) [66, 67]. These challenges highlight the need to move beyond purely SL frameworks and to incorporate physical principles as complementary theoretical

foundations.

Embedding physics in statistical learning

Incorporating physical priors into ML models helps mitigate challenges of generalization and data scarcity inherent in SL [29]. Physics priors can be introduced through three main pathways: observational bias, inductive bias, and learning bias [29]. Observational bias is introduced via data that encode the underlying physics. Empirical physical knowledge can be used to transform the dataset (from $\mathcal{D}$ to $\mathcal{D}'$) to more effectively reflect the underlying physical patterns. Inductive bias arises from embedding physical assumptions directly into the model architecture through tailored design choices that enforce invariance, covariance, or conservation laws [31]. This narrows the hypothesis space (from $\mathcal{H}$ to $\mathcal{H}'$), and enhances generalizability. In contrast to these weakly constrained approaches, physics-informed neural networks (PINNs) and symmetry-enforcing architectures incorporate physical constraints directly into both the loss function and the network architecture. [30]. The optimization algorithms (from $\mathcal{G}$ to $\mathcal{G}'$) can steer the model solutions toward those that satisfy the governing physical laws expressed in explicit mathematical formulations. However, such methods are not always feasible, particularly for highly complex systems, where deriving or implementing these formulations may be intractable.

However, for complex systems in physical sciences, the governing PDEs are often ill-defined or unknown, limiting the applicability of strong physics-based constraints (Fig. 1a). In such cases, PT learning offers a promising alternative to address data sparsity and limited generalization by leveraging approximate or related physical knowledge to guide learning across complex domains (Fig. 1b).

Generalization error bound

The DA theory provides a generalization error bound for the target domain (Eq. 1) [36]. Conventional analyses typically assume that the λ^* term is small, implying that the conditional distribution $P(Y \mid X)$ remains approximately invariant between the source and target domains [36–38]. This assumption, however, imposes significant limitations. Existing theories lack quantitative characterization of λ^*, offer no means to anticipate or diagnose failures when the assumption breaks down, and provide few principled strategies to mitigate them. These challenges arise from the fundamental constraints of purely statistical approaches, particularly when labeled data from the target domain are unavailable.

Beyond the purely statistical view of the conditional distribution $P(Y \mid X)$, we introduce a physical interpretation. Prior physical knowledge allows qualitative assessment of the λ^* term and estimation of potential failure risk. Leveraging principles from the natural sciences and physics-based transfer, one can identify an anchor domain (Eq. 2) to quantitatively evaluate a lower bound for λ^*. A large λ^* signals failure of physics transfer, while this indicator can also guide iterative refinement or supplementation of physical knowledge. Once the process converges, the stabilized value of λ^* provides a principled criterion for assessing the feasibility of physics transfer.

Next, we present the key components of the PT framework for reducing generalization error in the target domain.

(1) *Source-domain training*. We first collect high-density, wide-coverage source data using low-fidelity or low-complexity approaches, exploiting the high data acquisition efficiency in the source domain (computational or experimental). These data train ML models to reduce the source error $\widehat{\varepsilon}_{\mathcal{S}}(h)$ and the statistical term $O\left(\sqrt{\frac{d\log n+\log(1/\delta)}{n}}\right)$.

(2) *Latent representation learning*. Essential latent representations Z are extracted by incorporating physics priors, primarily as data priors (Eq. 3). Conventional methods reduce domain discrepancy through explicit alignment of latent features, typically using distance-based losses such as maximum mean discrepancy [36, 37]. In contrast, PT learning aligns data of varying fidelity through weak physics constraints, without control through

explicit mathematical forms. This implicit alignment enhances generalization while preserving the underlying physical patterns. The framework can further incorporate model architectures or optimization strategies when stronger physics priors are available.

We also leverage the information bottleneck (IB) principle to guide and monitor learning. Mutual information between Z and the input X, denoted as $I(Z;X)$, as well as between Z and the output Y, denoted as $I(Z;Y)$, is tracked during training to ensure that latent representations capture the most predictive, compressed features of the input, thereby reducing $d(\widehat{\mathcal{D}}_{\mathcal{S}}^{X}, \widehat{\mathcal{D}}_{\mathcal{T}}^{X})$ (Fig. 1c).

(3) *Anchor domains.* If the underlying physics are well preserved in both source and target domains ($\lambda^* \to 0$), generalization is guaranteed. A large λ^* signals failure of physics transfer, indicating missing physical mechanisms or incomplete understanding. Theoretical lower bounds exist [67] as

$$\lambda^* \geq \frac{1}{2}\left[d(\mathcal{D}_{\mathcal{S}}^{Y}, \mathcal{D}_{\mathcal{T}}^{Y}) - d(\mathcal{D}_{\mathcal{S}}^{Z}, \mathcal{D}_{\mathcal{T}}^{Z})\right]^2, \tag{8}$$

but they cannot be directly evaluated due to the lack of labeled target data. To address this issue, we introduce the chain of anchor domains (CADs, $\mathcal{A}_k$) as the intermediate bridge between source and target (Eq. 2) by leveraging the inherent hierarchy of the physical problem across different levels of fidelity or complexity, represented by the parameter ξ (Fig. 1d). Specifically, the anchor domains can be defined as

$$\widehat{\mathcal{A}}_k = \left\{\left(\mathbf{x}_k^i, \mathbf{y}_k^i\right)\right\}_{i=1}^{l_k}, \quad \mathbf{y}_k^i = \mathcal{P}\left(\mathbf{x}_k^i; \xi_{\mathcal{A}_k}\right), \quad \xi_{\mathcal{A}_k} \in [\xi_{\mathcal{S}}, \xi_{\mathcal{T}}], \tag{9}$$

where $\mathcal{P}$ denotes the shared physical mapping underlying both the source and target domains, l_k is the number of samples in the anchor domain $\mathcal{A}_k$, and k indexes multiple anchor domains that can be found between the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$.

We then can introduce indicators $\lambda^{\mathcal{A}_k} = \frac{1}{2}\left[d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Y}, \widehat{\mathcal{D}}_{\mathcal{A}_k}^{Y}) - d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{A}_k}^{Z})\right]^2$ to measure the feasibility of physics transfer between the source and anchor domains. Specifically, we compute the distributional distance d between the source and anchor domains in both the latent feature space Z and the label space Y. A small indicator $\lambda^{\mathcal{A}_k}$ is therefore a necessary, but not sufficient, condition for feasible physics transfer (i.e., for generalization or extrapolation). In particular, even when $\lambda^{\mathcal{A}_k}$ is small, target-domain predictions may still fail if no labeled data are available in the target domain, or if the distance between the anchor and target domains is not explicitly accounted for in $\lambda^{\mathcal{A}_k}$. To further assess feasibility, one can compute relative indicators $\delta^{\mathcal{A}_k} = d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{T}}^{Z}) / d(\widehat{\mathcal{D}}_{\mathcal{S}}^{Z}, \widehat{\mathcal{D}}_{\mathcal{A}_k}^{Z})$ to compare latent-space distances between anchor and target domain. In summary, a small $\lambda^{\mathcal{A}_k}$ and $\delta^{\mathcal{A}_k}$ indicate a higher likelihood of successful physics transfer and reduced generalization error (Fig. 1c).

The PT theoretical theory guides the selection of a more appropriate source domain. Specifically, one can start by identifying an arbitrary anchor domain between the source and target domains to evaluate the feasibility of PT. If the transfer is found infeasible (e.g., a large $\lambda^{\mathcal{A}_k}$), the current anchor domain is replaced by the new source domain, thereby narrowing the geometric complexity interval $[\xi_{\mathcal{S}}, \xi_{\mathcal{T}}]$. A new anchor domain is then selected within the reduced interval and re-evaluated. This process is repeated until PT is assessed to be feasible, or no further anchor domain can be found, at which point PT is considered infeasible. In such cases, the infeasibility of PT implies an insufficient understanding of the underlying physics, indicating the need to refine or supplement the physical knowledge before further transfer can be achieved.

The underlying physics can be transferred from the source domain through multiple anchor domains via stepwise fine-tuning, or an ensemble modeling approach can be used to combine predictions from models trained on the source and anchor domains. Specifically, to estimate the target-domain conditional distribution $P_{\mathcal{T}}(Y|X)$, we can calculate the distributions $P_{\mathcal{A}_k}(Y|X)$ from multiple anchor domains. Then, the target-domain distribution can be expressed as a weighted combination of anchor-domain distributions

$$P_{\mathcal{T}}(Y|X) = \frac{\sum_{k=1}^{K} w_k(X,Y)\, P_{\mathcal{A}_k}(Y|X)}{\sum_{k=1}^{K} w_k(X,Y)}. \tag{10}$$

The weights are determined by the domain discrepancy ($\delta^{\mathcal{A}_k}$) and PT feasibility ($\lambda^{\mathcal{A}_k}$)

$$w_k(X,Y) = \exp\Big[-\alpha\, \delta^{\mathcal{A}_k} - \gamma\, \lambda^{\mathcal{A}_k} \Big], \tag{11}$$

where α and γ are hyperparameters controlling the relative importance of the two indicators. Anchor domains with high PT feasibility (small $\lambda^{\mathcal{A}_k}$) and strong physical similarity to the target domain (small $\delta^{\mathcal{A}_k}$) receive higher weights in the combination.

Verification, validation, and uncertainty quantification

Beyond comparing distributional discrepancies in X, Y, and Z between the source and anchor domains, analyzing the parameter distributions of trained neural networks (NNs) provides an additional means to validate the effectiveness of physics transfer. Typically, ML models trained on different domains exhibit distinct parameter distributions

$$p(\Theta \mid \mathcal{D}_{\mathcal{S}}) \neq p(\Theta \mid \mathcal{D}_{\mathcal{T}}), \tag{12}$$

where Θ denotes model parameters and $p(\cdot \mid \cdot)$ is a conditional probability distribution, highlighting the challenges of transferring models across domains.

However, when underlying physical laws $\mathcal{P}$, govern the dataset $\mathcal{D}$, physics transfer can enable reliable extrapolation. A physical relationship between input features $\mathbf{x}$ and target predictions $\mathcal{O}$ can be formally represented as

$$\mathbf{x} \xrightarrow{\mathcal{P}} \mathcal{O}, \tag{13}$$

and the essential latent subspace can be defined as

$$\mathbf{x} \cap \mathcal{O} = \mathcal{D}' \subset \mathcal{D}, \tag{14}$$

where $\mathcal{D}'$ captures key physical correlations, representing the learned latent representations Z. Models ($h \in \mathcal{H}$) designed to encode these physical dynamics tend to converge to similar parameter distributions across domains,

$$p(\Theta \mid \mathcal{D}'_{\mathcal{S}}) \approx p(\Theta \mid \mathcal{D}'_{\mathcal{T}}). \tag{15}$$

Models trained on the source-domain subspace $\mathcal{D}'_{\mathcal{S}}$, characterized by parameter distributions $p(\Theta \mid \mathcal{D}'_{\mathcal{S}})$, can be extrapolated to predict high-fidelity, high-complexity scenarios.

Analyzing parameter distributions and connectivity patterns within trained NNs provides insight into model generalization and uncertainty. Using complex network theory, a deep NN can be represented as a directed graph $G = \langle N, E \rangle$, where each node $n_i^{[\ell]} \in N$ corresponds to a neuron in the ℓ-th hidden layer, and each directed edge ($e_{n_i^{[\ell]}, n_j^{[\ell+1]}} \in E$) encodes a weighted connection with real-valued weight $\omega_{i,j}^{[\ell]}$. Statistical metrics of the complex network can be used to characterize the structures of NNs, such as the mean and variance of weights. For

the ℓ-th layer, these quantities are defined as

$$\mu^{[\ell]} = \frac{1}{N^{[\ell]}N^{[\ell+1]}} \sum_{i=1}^{N^{[\ell]}} \sum_{j=1}^{N^{[\ell+1]}} \left(\omega_{i,j}^{[\ell]} + \beta_i^{[\ell]}\right), \tag{16}$$

$$\delta^{[\ell]} = \frac{1}{N^{[\ell]}N^{[\ell+1]}} \sum_{i=1}^{N^{[\ell]}} \sum_{j=1}^{N^{[\ell+1]}} \left[\left(\omega_{i,j}^{[\ell]} + \beta_i^{[\ell]}\right) - \mu^{[\ell]}\right]^2, \tag{17}$$

where $N^{[\ell]}$ is number of neurons in layer ℓ, $N^{[\ell+1]}$ is number of neurons in layer $\ell+1$, $\omega_{i,j}^{[\ell]}$ is weight of the connection from neuron i in layer ℓ to neuron j in layer $\ell+1$, $\beta_i^{[\ell]}$ is bias term associated with neuron i in layer ℓ, $\mu^{[\ell]}$ is average of all weights and biases feeding from layer ℓ to $\ell+1$, and $\delta^{[\ell]}$ is variance of those values relative to the mean. By evaluating these metrics between trained models on the source ($\mathcal{S}$) and anchor ($\mathcal{A}$) domains, using multiple random initializations to estimate the parameter distributions, one can assess both the feasibility of physics transfer and the model's ability to generalize beyond the training data.

Model uncertainty in extrapolative predictions is further evaluated via latent-space distributions. The activation of neuron k in the ℓ-th layer is

$$a_k^{[\ell]} = f^{[\ell]}\left(\sum_{i=1}^{N^{[\ell]}} z_i^{[\ell-1]}\omega_{i,k}^{[\ell]} + \beta_k^{[\ell]}\right). \tag{18}$$

These formulations exhibit that the activation $a_k^{[\ell]}$ is influenced by a weighted combination of preceding activations $z_i^{[\ell-1]}$, adjusted by the weight matrix $\omega_{i,k}^{[\ell]}$ and bias term $\beta_k^{[\ell]}$, and passed through the nonlinear activation function $f^{[\ell]}$. In cases where the input x is unfamiliar to the network (i.e., OOD), neurons in the latent space may demonstrate unusually high node strengths, reflecting significant deviations from activation patterns observed during training, an indicator of elevated epistemic uncertainty. However, even well-aligned latent spaces do not guarantee accurate predictions. Physical discrepancies between domains, quantified by $\lambda^{\mathcal{A}}$, must also be considered.

Overall, we term this approach PT learning, a theoretical framework that integrates prior physical knowledge with SL to assess and bound generalization. It provides principled criteria for failure detection and feasibility evaluation, extending beyond purely statistical methods. The framework is particularly suited to complex scientific problems where direct labels are unavailable or costly due to simulation limits, experimental constraints, or measurement challenges. By leveraging incremental prior knowledge, PT learning digitalizes the scientific process, representing a step toward building an AI scientist [69].

Continuum modeling

Brain development is regulated by genetic, molecular, cellular, and mechanical factors across multiple spatiotemporal scales [2, 70], and the differential tangential growth hypothesis is commonly used [2, 19, 70]. Finite element analysis (FEA) can model morphological evolution during brain growth at the continuum level [17–20, 22]. Tangential growth (TG) of the outer gray matter is faster than the inner white matter, and its implementation in FEA is known as the TG model [19]. Compression resulting from the mismatch in deformation may then lead to mechanical instabilities of the brain surface, forming characteristic sulci and gyri structures [17–22, 39].

In continuum modeling, the reference configuration can be mapped to the current one through the deformation

gradient tensor as

$$\mathbf{F} = \mathbf{F}^{\mathrm{e}} \cdot \mathbf{G}, \tag{19}$$

where $\mathbf{F}^{\mathrm{e}}$ is the elastic deformation gradient and $\mathbf{G}$ is the growth term. In the TG model, the growth tensor $\mathbf{G}$ is

$$\mathbf{G} = g\mathbf{I} + (1-g)\hat{\mathbf{n}} \otimes \hat{\mathbf{n}}, \tag{20}$$

where $\hat{\mathbf{n}}$ is the surface normal of the reference configuration, $\mathbf{I}$ is the unit tensor, and

$$g = 1 + \frac{\alpha_t}{1 + e^{10(\frac{y}{T}-1)}} \tag{21}$$

is the growth coefficient, where α_t controls the magnitude of local cortical expansion. The constant factor of 10 in the exponential term governs the steepness of the sigmoidal transition in g, producing a sharp yet smooth change near $y = T$ that delineates the gray matter layer from the underlying white matter [18, 19, 22]. There is a smooth transition from the surface of the gray matter layer to the white matter layer with a gradually decreasing growth coefficient. y is the distance to the surface, and T is the thickness of the cortex. The brain is modeled as a nonlinear neo-Hookean hyperelastic material, where the strain energy density is

$$W = \frac{G}{2}[\mathrm{Tr}(\mathbf{F}^{\mathrm{e}}\mathbf{F}^{\mathrm{eT}})J^{-2/3} - 3] + \frac{K}{2}(J-1)^2, \tag{22}$$

where G is the shear modulus, J is the determinant of Jacobian matrix, K is the bulk modulus.

For brain growth, a core-shell structure with a spherical geometry is used for its simplicity. The outer radius is 10 mm and the shell thickness ranges from 0.03 to 1.63 mm, which are determined from the measurements of abnormal and normal human cerebral cortices [22, 47]. 4-node tetrahedral elements with a density of 10^6 tetrahedra/cm^3 for discretization with the convergence confirmed [19, 22]. The morphogenesis of brains is triggered by internal elastic stresses generated from differential core-shell growth. The interaction between surfaces is modeled with an energy penalty via vertex-triangle contact, which prevents the nodes from penetrating the faces of elements [71]. An explicit solver is used to minimize the total (elastic and contact) energy of the quasi-static system. The time step $\Delta t = 0.05a\sqrt{\rho/K}$ is set to ensure the convergence, where a is the mesh size and ρ is the mass density [72].

Assigning material models and parameters to brain tissue regions is challenging due to intra-regional variability and differences across individual brains. Additionally, properties change with development or aging. The alternative approach taken here, which is the current state of the art, involves assigning typical properties for a tissue type and age, using experimental data that closely approximate the specific loading conditions. The bulk modulus of the core and shell is 5 times the shear modulus [19]. Following the experimental evidence, the relative shear modulus ($G_{\mathrm{shell}}/G_{\mathrm{core}}$) ranges from 0.65 to 1 [50].

Machine learning models

By designing NN architectures ($h \in \mathcal{H}$), one can capture the physics of bifurcation and morphological features from simple geometries with low complexity. Models trained on these data, characterized by parameter distributions $p(\Theta|\mathcal{D}'_S)$, can then be extrapolated to predict brain morphological development with a much more complex geometry.

In FEA, morphological data are meshed into discretized tetrahedral elements. The representation can be directly translated into graphs, where the nodes correspond to the vertices of the elements, and the edges correspond to the edges of the elements. Graph neural networks (GNNs) are then constructed to extract key features from the graphs [73]. We utilize an encoder-decoder architecture to learn the complexity of morphological development

(Fig. S4). The input to the model is a graph representation of the morphology, with node features such as the coordinates and normal directions. The output is the local curvatures. The encoder employs three message-passing blocks to transform the input features, comprising 3D positions and normals (6-dimensional), into a latent representation with 256 dimensions. In the first message-passing block, a multilayer perceptron (MLP) processes the 6D input using 64 neurons in both the hidden and output layers. This is followed by two additional message-passing blocks with 128 and 256 neurons, respectively, progressively increasing the expressive capacity of the node features. After message passing, several fully connected layers further refine the learned representations. A global feature is produced via a fully connected layer (256 to 1 neuron) followed by global pooling. Local curvature is predicted using a separate 256-to-1 neuron layer, and a 256-to-512 layer encodes the final latent geometry representation. The decoder maps this 512-dimensional latent code to reconstruct the 3D coordinates using an MLP with hidden layers of sizes 512, 256, 128, and 32, culminating in a 3D output. Training is guided by a composite loss function that incorporates errors from the reconstructed 3D coordinates, global gyrification index, and local curvatures, with all NN parameters updated using standard backpropagation. The adaptive moment estimation (ADAM) optimizer is used to train the NN parameters, with hyperparameter settings: a learning rate of 0.001, and first- and second-moment decay rates $\beta_0 = 0.9, \beta_1 = 0.999$ [74]. The algorithm is implemented in PyTorch, and the hyperparameters are set using widely adopted configurations from prior works rather than through extensive tuning [53].

Distance metrics

To assess the generalization ability of a model prior to its deployment, we compute the distance between training samples (sphere data) and testing samples (brain data) in the latent space. The underlying hypothesis is that a greater distance indicates a larger domain shift, which typically leads to poorer model generalization. Therefore, this distance can be used to estimate model performance *a priori*.

To quantify the distance between the training and testing distributions, we first project both datasets into a low-dimensional latent space using principal component analysis (PCA) [75], which retains the most significant features while reducing noise and redundancy. Next, we apply kernel density estimation (KDE) to the training data to estimate its probability density function in the latent space [76, 77]. Given a test sample $\tilde{q}$ and a set of N projected training samples $\{\tilde{p}_i\}_{i=1}^{N}$, the estimated density at $\tilde{q}$ is calculated as

$$\hat{f}(\tilde{q}) = \frac{1}{N} \sum_{i=1}^{N} K\left(\frac{\|\tilde{q} - \tilde{p}_i\|}{h} \right), \tag{23}$$

where $K(\cdot)$ is the Gaussian kernel function defined as

$$K(u) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} u^2 \right). \tag{24}$$

Here, h is the bandwidth parameter controlling the smoothness of the density estimate. A lower estimated density $\hat{f}(\tilde{q})$ implies a larger distance from the training distribution and suggests a higher likelihood of prediction error.

This KDE-based distance measure provides a quantitative approach to evaluate the proximity of testing data to the training distribution. It enables the estimation of prediction uncertainty and can serve as an early warning indicator for potential model failure due to domain shift.

In addition to point-wise distance evaluation between testing and train samples, the Jensen-Shannon (JS) diver-

gence is employed to quantify the distributional discrepancies between the source and target domains across the input, latent, and output space [67, 78]. The JS divergence is a symmetrized and smoothed version of the Kullback-Leibler divergence (KL), which is defined as:

$$\mathrm{JS}(\mathcal{S}\|\mathcal{T}) = \frac{1}{2}\mathrm{KL}(\mathcal{S}\|\mathcal{M}) + \frac{1}{2}\mathrm{KL}(\mathcal{T}\|\mathcal{M}), \tag{25}$$

where

$$\mathcal{M} = \frac{1}{2}(\mathcal{S} + \mathcal{T}), \tag{26}$$

and $\mathrm{KL}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence. A smaller JS value indicates higher similarity between the two distributions. The $d = \sqrt{\mathrm{JS}(\mathcal{S}\|\mathcal{T})}$ is used as the distance measure to quantify the distributional discrepancies between the source and target domains.

Collection of medical data

We collect experimental data on the human brain to validate our PT approach. The data of human brain morphologies are rare, especially for the sequences of individual brain morphologies (Fig. S2) [9, 42]. We use high-resolution MRI data of brain anatomy from open-source brain structural atlases [42], which are then translated to 3D model geometries using a pipeline involving cortical and sub-cortical volume segmentation and cortical surface extraction [65].

Specifically, we used a triangular surface mesh representation of the human cortical surface, consisting of $32,492$ vertices per hemisphere, derived from a population-averaged atlas covering 21 to 36 weeks of gestation (`https://doi.gin.g-node.org/10.12751/g-node.qj5hs7/`) [79].

Declaration of Competing Interest

The authors declare that they have no competing financial interests.

Credit authorship contribution statement

Z.X. conceived the concept. Y.Z. developed the methods and performed the research under the supervision of Z.X. and F.X. All authors analysed the results and wrote the manuscript.

Data Availability

The digital libraries of brain models and simulation data are provided, and all scripts that can be used to reproduce the results in this work are available on GitHub at `https://github.com/zhaoyj21/PT-brain`.

Acknowledgements

This work was supported by the National Natural Science Foundation of China through grants 12425201, 52090032, 12125206, 124B2035, 12425204, 12122204 and 12372096, Shanghai Pilot Program for Basic Research-Fudan University (grant no. 21TQ140010021TQ010), and Shanghai Municipal Education Commission (grant no. 24KXZNA14). The computation was performed on the Explorer 1000 cluster system of the Tsinghua National Laboratory for Information Science and Technology.

References

[1] John C Silbereis, Sirisha Pochareddy, Ying Zhu, Mingfeng Li, and Nenad Sestan. The cellular and molecular landscapes of the developing human central nervous system. *Neuron*, 89(2):248–268, 2016.

[2] Cristina Llinares-Benadero and Víctor Borrell. Deconstructing cortical folding: Genetic, cellular and mechanical determinants. *Nat. Rev. Neurosci.*, 20(3):161–176, 2019.

[3] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342(6158):1238411, 2013.

[4] Nikolaus Kriegeskorte and Xue-Xin Wei. Neural tuning and representational geometry. *Nat. Rev. Neurosci.*, 22(11):703–718, 2021.

[5] James C Pang, Kevin M Aquino, Marianne Oldehinkel, Peter A Robinson, Ben D Fulcher, Michael Breakspear, and Alex Fornito. Geometric constraints on human brain function. *Nature*, 618(7965):566–574, 2023.

[6] Renske Oegema, Tahsin Stefan Barakat, Martina Wilke, Katrien Stouffs, Dina Amrom, Eleonora Aronica, Nadia Bahi-Buisson, Valerio Conti, Andrew E Fry, Tobias Geis, et al. International consensus recommendations on the diagnostic work-up for malformations of cortical development. *Nat. Rev. Neurosci.*, 16(11):618–635, 2020.

[7] Chloe Hutton, Bogdan Draganski, John Ashburner, and Nikolaus Weiskopf. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage*, 48(2):371–380, 2009.

[8] Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.

[9] Richard AI Bethlehem, Jakob Seidlitz, Simon R White, Jacob W Vogel, Kevin M Anderson, Chris Adamson, Sophie Adler, George S Alexopoulos, Evdokia Anagnostou, Ariosky Areces-Gonzalez, et al. Brain charts for the human lifespan. *Nature*, 604(7906):525–533, 2022.

[10] Gang Xu, Philip V Bayly, and Larry A Taber. Residual stress in the adult mouse brain. *Biomech. Model. Mechanobiol.*, 8:253–262, 2009.

[11] Gang Xu, Andrew K Knutsen, Krikor Dikranian, Christopher D Kroenke, Philip V Bayly, and Larry A Taber. Axons pull on the brain, but tension does not drive cortical folding. *J. Biomech. Eng.*, 132(7):071013, 2010.

[12] Mohammad R Arbabshirani, Sergey Plis, Jing Sui, and Vince D Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145:137–165, 2017.

[13] Gaël Varoquaux and Veronika Cheplygina. Machine learning for medical imaging: Methodological failures and recommendations for the future. *NPJ Digit. Med.*, 5:48, 2022.

[14] Mariasavina Severino, Ana Filipa Geraldo, Norbert Utz, Domenico Tortora, Ivana Pogledic, Wlodzimierz Klonowski, Fabio Triulzi, Filippo Arrigoni, Kshitij Mankad, Richard J Leventer, et al. Definitions and classification of malformations of cortical development: Practical guidelines. *Brain*, 143(10):2874–2894, 2020.

[15] Mireia Alenyà, Xiaoyu Wang, Julien Lefèvre, Guillaume Auzias, Benjamin Fouquet, Elisenda Eixarch, François Rousseau, and Oscar Camara. Computational pipeline for the generation and validation of patient-specific mechanical models of brain development. *Brain Multiphys.*, 3:100045, 2022.

[16] Guillermo Lorenzo, David A Hormuth, Chengyue Wu, Graham Pash, Anirban Chaudhuri, Ernesto ABF Lima, Lois C Okereke, Reshmi Patel, Karen Willcox, Thomas E Yankeelov, et al. Validating the predictions of mathematical models describing tumor growth and treatment response. *arXiv preprint arXiv:2502.19333*, 2025.

[17] Mohsen Darayi, Mia E Hoffman, John Sayut, Shuolun Wang, Nagehan Demirci, Jack Consolini, and Maria A Holland. Computational models of cortical folding: A review of common approaches. *J. Biomech.*, 139:110851, 2022.

[18] Tuomas Tallinen, Jun Young Chung, John S Biggins, and Lakshminarayanan Mahadevan. Gyrification from constrained cortical expansion. *Proc. Natl. Acad. Sci. USA*, 111(35):12667–12672, 2014.

[19] Tuomas Tallinen, Jun Young Chung, François Rousseau, Nadine Girard, Julien Lefèvre, and Lakshminarayanan Mahadevan. On the growth and form of cortical convolutions. *Nat. Phys.*, 12(6):588–593, 2016.

[20] Silvia Budday and Paul Steinmann. On the influence of inhomogeneous stiffness and growth on mechanical instabilities in the developing brain. *Int. J. Solids Struct.*, 132:31–41, 2018.

[21] Lucas da Costa Campos, Raphael Hornung, Gerhard Gompper, Jens Elgeti, and Svenja Caspers. The role of thickness inhomogeneities in hierarchical cortical folding. *NeuroImage*, 231:117779, 2021.

[22] Xiaoyu Wang, Julien Lefèvre, Amine Bohi, Mariam Al Harrach, Mickaël Dinomais, and François Rousseau. The influence of biophysical parameters in a biomechanical model of cortical folding patterns. *Sci. Rep.*, 11:7686, 2021.

[23] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and EJPRL Weinan. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.*, 120(14):143001, 2018.

[24] He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and Yong Xu. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.*, 2(6):367–377, 2022.

[25] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[26] Alireza Ghafarollahi and Markus J Buehler. Automating alloy design and discovery with physics-aware multimodal multiagent AI. *Proc. Natl. Acad. Sci.*, 122(4):e2414074122, 2025.

[27] Zhao Chen, Yang Liu, and Hao Sun. Physics-informed learning of governing equations from scarce data. *Nat. Commun.*, 12:6136, 2021.

[28] Lemuel Puglisi, Daniel C Alexander, and Daniele Ravì. Brain latent progression: Individual-based spatiotemporal disease progression on 3D brain MRIs via latent diffusion. *arXiv preprint arXiv:2502.08560*, 2025.

[29] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nat. Rev. Phys.*, 3(6):422–440, 2021.

[30] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.

[31] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.*, 13(1):2453, 2022.

[32] Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nat. Mach. Intell.*, 6(10):1256–1269, 2024.

[33] Yingjie Zhao and Zhiping Xu. Physics-transfer learning: A framework to address the accuracy-performance dilemma in modeling morphological complexities in brain development. In *ICLR 2025 Workshop on Machine Learning Multiscale Processes*, 2025.

[34] Yingjie Zhao, Hongbo Zhou, Zian Zhang, Zhenxing Bo, Baoan Sun, Minqiang Jiang, and Zhiping Xu. Discovering high-strength alloys via physics-transfer learning. *Matter*, 8(9):102377, 2025.

[35] Yaser S Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning from Data*, volume 4. AMLBook New York, 2012.

[36] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Mach. Learn.*, 79:151–175, 2010.

[37] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *Int. Conf. Mach. Learn. (ICML)*, pages 97–105. PMLR, 2015.

[38] Keru Wu, Yuansi Chen, Wooseok Ha, and Bin Yu. Prominent roles of conditionally invariant components in domain adaptation: Theory and algorithms. *J. Mach. Learn. Res.*, 26(110):1–92, 2025.

[39] Georg F Striedter, Shyam Srinivasan, and Edwin S Monuki. Cortical folding: When, where, how, and why? *Annu. Rev. Neurosci.*, 38:291–307, 2015.

[40] Kara E Garcia, Xiaojie Wang, and Christopher D Kroenke. A model of tension-induced fiber growth predicts white matter organization during brain folding. *Nat. Commun.*, 12:6681, 2021.

[41] Jorge Corral-Acero, Francesca Margara, Maciej Marciniak, Cristobal Rodero, Filip Loncaric, Yingjing Feng, Andrew Gilbert, Joao F Fernandes, Hassaan A Bukhari, Ali Wajdan, et al. The 'Digital Twin' to enable the vision of precision cardiology. *Eur. Heart J.*, 41(48):4556–4564, 2020.

[42] Tommaso Ciceri, Luca Casartelli, Florian Montano, Stefania Conte, Letizia Squarcina, Alessandra Bertoldo, Nivedita Agarwal, Paolo Brambilla, and Denis Peruzzo. Fetal brain MRI atlases and datasets: A review. *NeuroImage*, 292:120603, 2024.

[43] Fan Xu, Yangchao Huang, Shichen Zhao, and Xi-Qiao Feng. Chiral topographic instability in shrinking spheres. *Nat. Comput. Sci.*, 2(10):632–640, 2022.

[44] Jie Yin, Zexian Cao, Chaorong Li, Izhak Sheinman, and Xi Chen. Stress-driven buckling patterns in spheroidal core/shell structures. *Proc. Natl. Acad. Sci. USA*, 105(49):19132–19135, 2008.

[45] Junjie Lin, Qiaohang Guo, Shiwen Dou, Nengbin Hua, Chan Zheng, Yian Pan, Youting Huang, Zi Chen, and Wenzhe Chen. Bistable structures with controllable wrinkled surface. *Extreme Mech. Lett.*, 36:100653, 2020.

[46] Lifeng Wang, Carlos E Castro, and Mary C Boyce. Growth strain-induced wrinkled membrane morphology of white blood cells. *Soft Matter*, 7(24):11319–11324, 2011.

[47] Bruce Fischl and Anders M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. USA*, 97(20):11050–11055, 2000.

[48] Bernard S Chang, Fusun Duzcan, Seonhee Kim, Mine Cinbis, Abha Aggarwal, Kira A Apse, Osman Ozdel, Munevver Atmaca, Sevil Zencir, Huseyin Bagci, et al. The role of reln in lissencephaly and neuropsychiatric disease. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, 144:58–63, 2007.

[49] Julien Dervaux and Martine Ben Amar. Morphogenesis of growing soft tissues. *Phys. Rev. Lett.*, 101(6):068101, 2008.

[50] Silvia Budday, Richard Nay, Rijk De Rooij, Paul Steinmann, Thomas Wyrobek, Timothy C Ovaert, and Ellen Kuhl. Mechanical properties of gray and white matter brain tissue by indentation. *J. Mech. Behav. Biomed. Mater.*, 46:318–330, 2015.

[51] Yingjie Zhao, Jianshu Qin, Shijun Wang, and Zhiping Xu. Unraveling the morphological complexity of two-dimensional macromolecules. *Patterns*, 3(6):100497, 2022.

[52] Yingjie Zhao and Zhiping Xu. A data-driven approach to morphogenesis under structural instability. *Cell Rep. Phys. Sci.*, 5(3):101872, 2024.

[53] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *Int. Conf. Learn. Represent. (ICLR)*, 2020.

[54] Silvia Budday, Charles Raybaud, and Ellen Kuhl. A mechanical model predicts morphological abnormalities in the developing human brain. *Sci. Rep.*, 4:5644, 2014.

[55] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowl. Discov. Data Min. (KDD)*, volume 96, pages 226–231, 1996.

[56] Geoffrey E Hinton. Connectionist learning procedures. In *Machine Learning*, pages 555–610. Elsevier, 1990.

[57] Emanuele La Malfa, Gabriele La Malfa, Giuseppe Nicosia, and Vito Latora. Deep neural networks via complex network theory: A perspective. *arXiv preprint arXiv:2404.11172*, 2024.

[58] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Mach. Learn.*, 79:151–175, 2010.

[59] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

[60] Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *IEEE Inf. Theory Workshop (ITW)*, pages 1–5. IEEE, 2015.

[61] Torgeir Hellstrøm, Lars Tjelta Westlye, Solrun Sigurdardottir, Cathrine Brunborg, HL Soberg, Øyvor Holthe, ANDRES Server, Martina Jonette Lund, Ole Andreas Andreassen, and Nada Andelic. Longitudinal changes in brain morphology from 4 weeks to 12 months after mild traumatic brain injury: Associations with cognitive functions and clinical variables. *Brain Inj.*, 31(5):674–685, 2017.

[62] Donald A Collie, Robin J Sellar, Martin Zeidler, Alan CF Colchester, R Knight, and RG Will. MRI of Creutzfeldt–Jakob disease: Imaging features and recommended MRI protocol. *Clin. Radiol.*, 56(9):726–739, 2001.

[63] Sahar Rezaei, Homa Seyedmirzaei, Esmaeil Gharepapagh, Fateme Mohagheghfard, Zahra Hasankhani, Mahsa Karbasi, Sahar Delavari, and Mohammad Hadi Aarabi. Effect of spaceflight experience on human brain structure, microstructure, and function: Systematic review of neuroimaging studies. *Brain Imaging Behav.*, 18:1256–1279, 2024.

[64] Kevin Linka, Sarah R St Pierre, and Ellen Kuhl. Automated model discovery for human brain using constitutive artificial neural networks. *Acta Biomater.*, 160:134–151, 2023.

[65] Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.

[66] Junguang Jiang, Yang Shu, Jianmin Wang, and Mingsheng Long. Transferability in deep learning: A survey. *arXiv preprint arXiv:2201.05867*, 2022.

[67] Han Zhao, Remi Tachet Des Combes, Kun Zhang, and Geoffrey Gordon. On learning invariant representations for domain adaptation. In *Int. Conf. Learn. Represent. (ICLR)*, pages 7523–7532. PMLR, 2019.

[68] Yifei He, Haoxiang Wang, Bo Li, and Han Zhao. Gradual domain adaptation: Theory and algorithms. *J. Mach. Learn. Res.*, 25(361):1–40, 2024.

[69] Zhiping Xu. Resolving physical complexities with machine intelligence. *Adv. Mech.*, 55:1–50, 2025.

[70] Esther Klingler, Fiona Francis, Denis Jabaudon, and Silvia Cappello. Mapping the molecular and cellular complexity of cortical malformations. *Science*, 371(6527):eaba4517, 2021.

[71] Christer Ericson. *Real-Time Collision Detection*. CRC Press, Florida, 2004.

[72] Ted Belytschko, Wing Kam Liu, Brian Moran, and Khalil Elkhodary. *Nonlinear Finite Elements for Continua and Structures*. John Wiley & Sons, New Jersey, 2014.

[73] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Trans. Neural. Netw.*, 20(1):61–80, 2008.

[74] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[75] Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 2(11):559–572, 1901.

[76] Emanuel Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33(3):1065–1076, 1962.

[77] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. In *Int. Conf. Learn. Represent. (ICLR)*, 2018.

[78] Corinna Cortes and Mehryar Mohri. Domain adaptation and sample bias correction theory and algorithm for regression. *Theor. Comput. Sci.*, 519:103–126, 2014.

[79] Vycheslav Karolis, Logan Williams, Vanessa Kyriakopoulou, Jelena Bozek, Alena Uus, Antonios Makropoulos, Andreas Schuh, Lucilio Cordero Grande, Emer Hughes, Anthony Price, Maria Deprez, Mary Rutherford, et al. Developing Human Connectome Project spatio-temporal surface atlas of the fetal brain. Dataset, G-Node, 2023. https://doi.org/10.12751/g-node.qj5hs7.

Figures and Figure Captions

Figure 1. Model complexity and physics-transfer (PT) learning.

Figure 2. Brain development complexity and challenges arising from experimental low fidelity and data scarcity.

Figure 3. PT approach for brain morphological development.

Figure 4. Descriptive and predictive performance of PT models.

Figure 5. Reduced-dimensional models and neural network analysis.

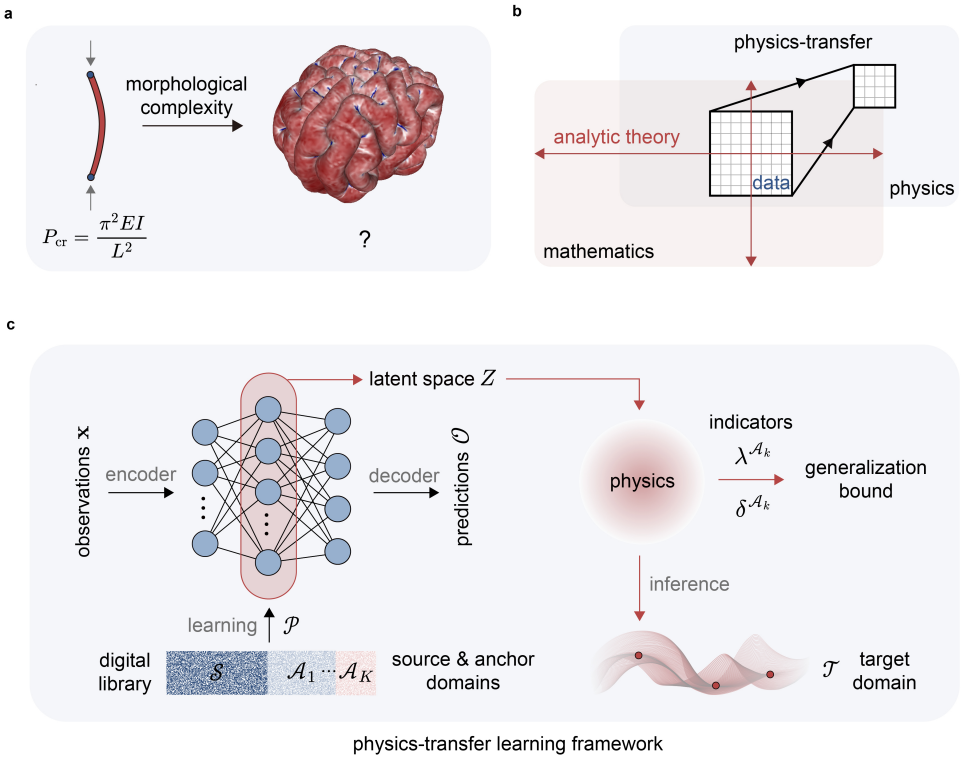

Figure 1. Model complexity and physics-transfer (PT) learning. **(a)** Analytical theories fail to address problems with multiscale complexity. **(b)** PT maps physics from simple to complex problems. **(c)** Schematic of the PT learning framework. A comprehensive digital library (DL) with high-density, large-coverage data is first constructed to enable the models to learn fundamental physics in simplified or well-characterized settings, which are then transferred and adapted to more complex, data-scarce scenarios. **(d)** Implementation of PT through chains of anchor domains (CADs) and weak physics constraints.

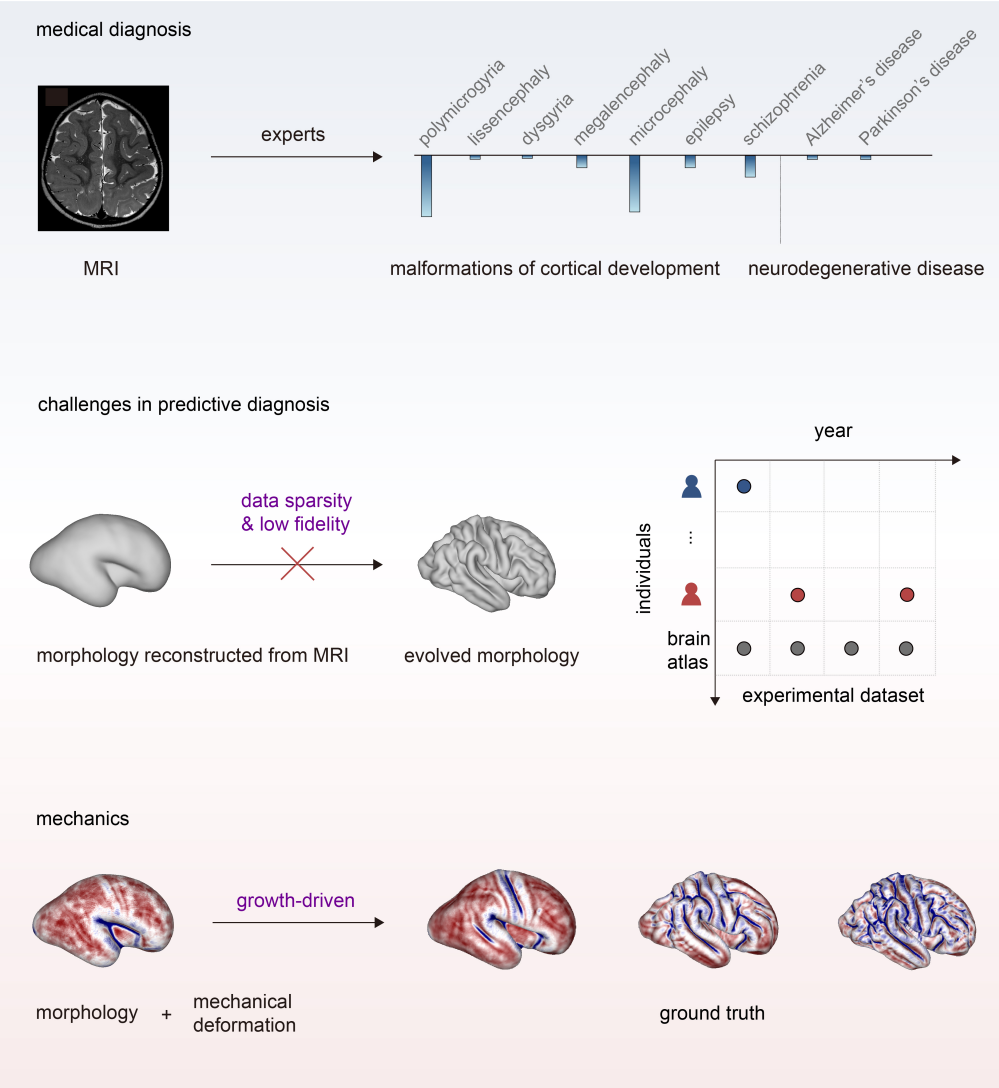

Figure 2. Complexity in brain development and challenges from experimental low fidelity and data scarcity. Current diagnostic practices primarily rely on expert interpretation of magnetic resonance imaging (MRI) scans, which provide geometric information based on voxel intensity but lack biomechanical fidelity. Moreover, the majority of MRI data in both clinical and experimental contexts are cross-sectional, limiting the ability to investigate temporal morphological changes. This scarcity of longitudinal data poses a significant barrier to the deployment of machine learning (ML) methods for modeling brain development or disease progression. By leveraging continuum mechanics, it is possible to reconstruct high-fidelity stress fields from observed geometries, offering a pathway to more comprehensive insights into brain morphogenesis.

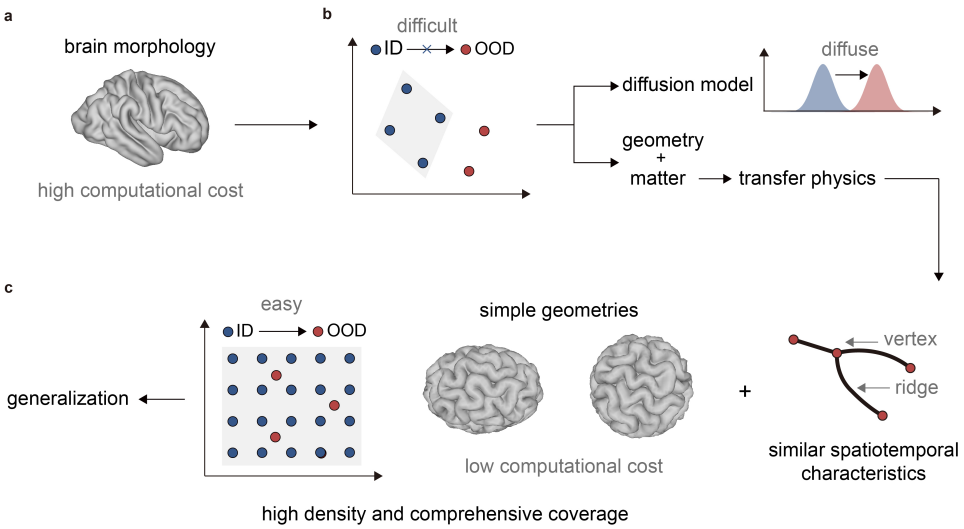

Figure 3. PT approach to predicting brain morphological development. (a) The inherent geometric complexity of brain anatomy poses a significant challenge to constructing a comprehensive DL from limited MRI datasets, resulting in data sparsity. ID: in distribution. OOD: out of distribution. **(b)** While generative AI holds promise, purely data-driven approaches based on statistical learning (SL) from morphological data remain constrained by the low fidelity and scarcity. **(c)** To address this, we decouple physical and geometric complexity by first modeling morphological development using simplified geometries. This enables the creation of a dense, well-sampled DL, from which the learned physics of nonlinear elasticity can be transferred to anatomically complex brain data. Sampling from the morphological space of simplified geometries proves significantly more efficient than direct sampling from brain tissue models. These simplified domains provide high coverage and density, and their spatiotemporal similarity to brain morphological evolution supports robust generalization of the underlying physics.

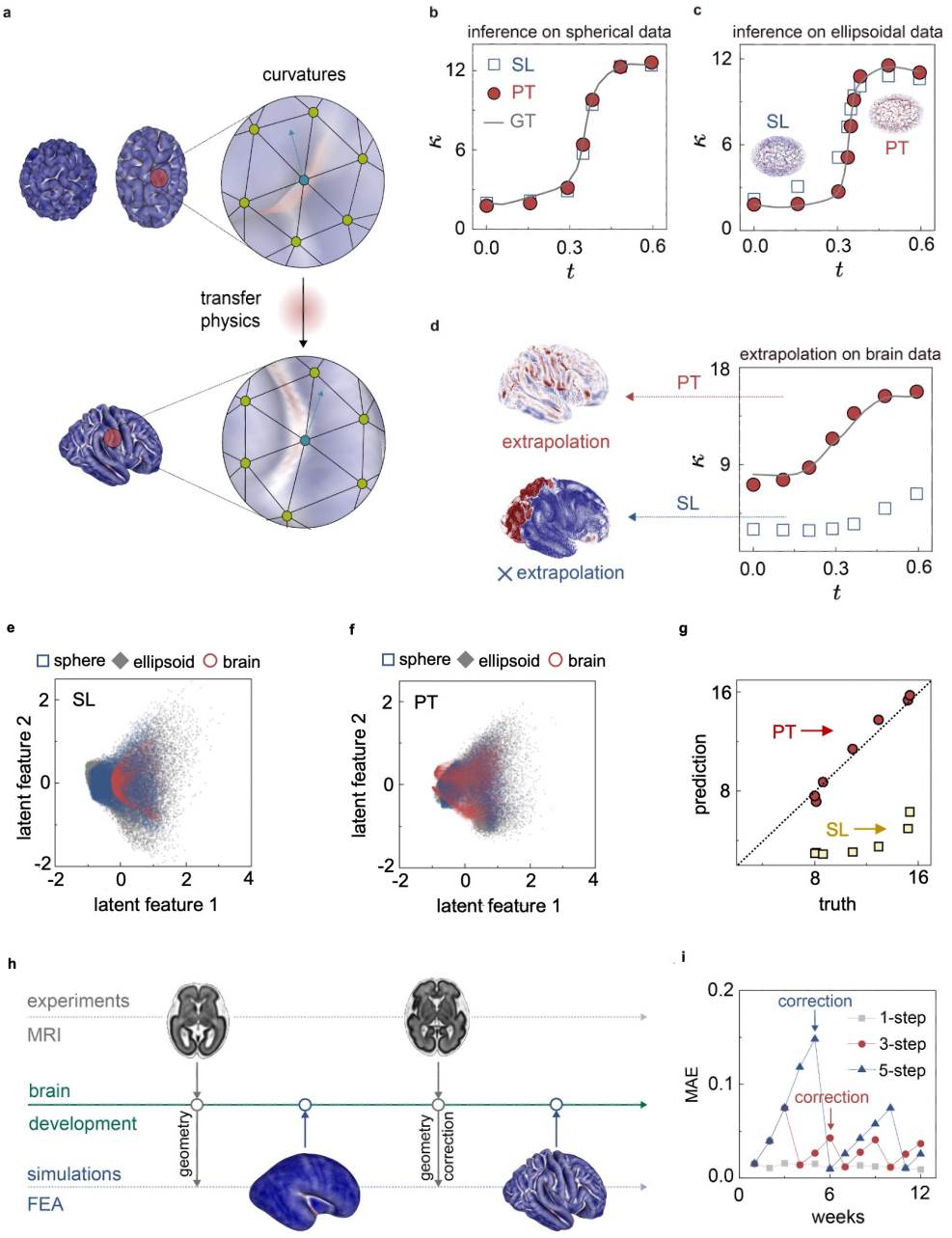

Figure 4. Descriptive and predictive performance of PT models. **(a)** PT models applied to descriptive curvature maps. **(b)** Interpolative predictions for spherical data. **(c, d)** Extrapolative predictions for ellipsoidal data **(c)** and the development of brain morphologies **(d)**. The insets are colored by the local curvature metrics, $|H|+|K|$, where H and K are the mean and Gaussian curvatures, respectively. **(e, f)** Latent feature distribution in SL **(e)** and PT **(f)** learning across domains with varying complexities. **(g)** Performance of curvature prediction in PT and SL based on brain development data. **(h)** PT models applied to predictive morphological development, enabling a digital twin approach for medical applications. **(i)** Predictive performance of PT models in brain morphological development.

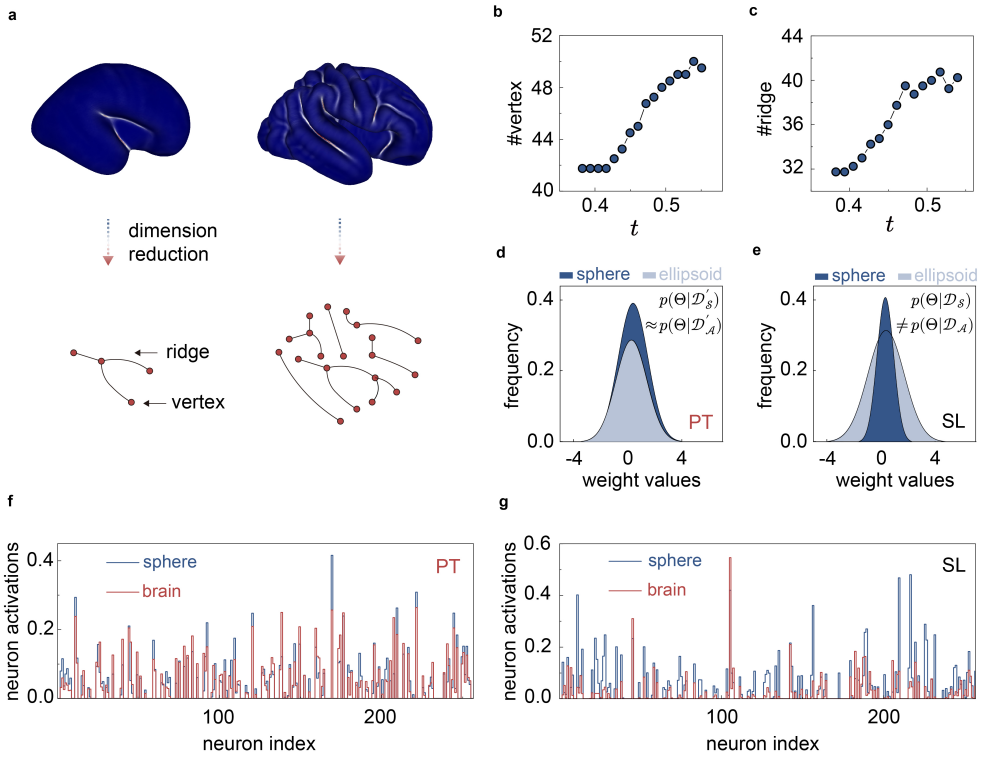

Figure 5. Reduced-dimensional models and neural network analysis. **(a)** The vertex-ridge network emerging from the machine learning (ML) model. **(b, c)** Dimension-reduction metrics used to track the evolutionary dynamics in brain morphological development, which include the number of the vertices **(b)** and the number of ridges **(c)**. **(d, e)** Weights parameters distribution of ML models trained on the spherical and ellipsoidal data for PT learning **(d)** and SL with only the morphological data used in the learning process **(e)**. **(f, g)** Neuron activations of the ML models trained on spherical data, when inference on both spherical and brain data for PT learning **(f)** and SL **(g)**.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- suppinfo.pdf