

On-Manifold Variational Learning with Heat-Kernel Priors

Jiarui Xing¹, Tal Zeevi¹, Nian Wu², and Jian Wang³

¹ Yale School of Medicine, New Haven CT, 06510, USA

² University of Virginia, Charlottesville, VA, 22903, USA

³ Harvard Medical School, Boston MA, 02115, USA

jiarui.xing@yale.edu

Abstract. Learning unsupervised representations of medical imaging cohorts can reveal anatomically meaningful prototypes without expert labels, which are often noisy and fail to capture true pathological heterogeneity. However, existing deep latent-variable models estimate Gaussian mixture priors via Euclidean averaging, producing prototypes that drift off the curved data manifold and degenerate as the number of sub-populations grows. We propose a manifold-anchored variational framework built on a geometry-aware Expectation-Maximization (EM) algorithm, whose M-step selects each sub-population prototype as the graph medoid with the highest diffusion centrality on a heat-kernel-weighted latent graph, ensuring that every prototype remains on-manifold. A Dirichlet energy regularizer enforces geometric smoothness of the latent space, and a per-sub-population uncertainty score enables label-free quality assessment. The manifold-anchored EM is a general-purpose geometric tool that extends standard EM and applies readily to other latent-variable models beyond this setting. On cardiac scar and brain MRI benchmarks, our framework attains the highest accuracy among all compared methods, produces the sharpest prototypes reported to date, and remains stable at large sub-population counts where all baselines degenerate.⁴

Keywords: Unsupervised Representation Learning · Variational Inference · Manifold Geometry

1 Introduction

Learning unsupervised representations that organize medical images into clinically meaningful sub-populations is central to disease staging [9], atlas construction [7,13], and cohort selection for clinical trials [26]. Supervised methods assume that expert-provided labels faithfully delineate these groups. In practice, however, diagnostic labels across a wide range of clinical domains are unreliable, and the learning algorithms that consume them inherit failure modes that go beyond label noise alone.

⁴ Code and implementation details are available at <https://github.com/jr-xing/On-Manifold-Variational-Learning-with-Heat-Kernel-Priors>.

Consider two representative scenarios. In cardiac imaging, myocardial scar sub-typing from late gadolinium enhancement MRI is inherently ambiguous: scar morphologies form a continuum from dense core fibrosis to diffuse border zone with poorly reproducible delineation [18]. In neuroimaging, the diagnostic labels accompanying large-scale brain MRI cohorts (e.g., OASIS [20], ADNI [14]) reflect cognitive and behavioral assessments rather than direct image morphology, and are further shaped by inter-rater subjectivity and evolving diagnostic criteria [5]. Analogous ambiguity pervades tumor grading [19], retinal staging [10], and psychiatric neuroimaging [21]. Label noise is well-recognized; less appreciated is that current unsupervised approaches suffer from a deeper structural failure. Euclidean prototype estimation in latent space produces redundant, overlapping mixture components [29,22] whose arithmetic means may fall in low-density regions corresponding to no observed subject [2,3]. These failures compound: off-manifold averaging induces component overlap, forcing models to allocate additional components that further erode interpretability [15,8].

Recent advances address only part of this problem. Contrastive methods [6,11,4] learn transferable representations by enforcing invariance across augmented views, but do not explicitly model population-level structure. Variational autoencoders [17] provide a generative framework, yet the standard isotropic Gaussian prior fails to capture multi-modal distributions, producing entangled representations where sub-populations overlap. Hybrid variational Gaussian mixture models [15,8,24] introduce discrete latent structure, but their prototypes remain Euclidean averages, inheriting the off-manifold problem. Deformable atlas building models [7,1] construct population templates via spatial transformations, but are fundamentally limited to diffeomorphic deformations that preserve topology; when sub-populations exhibit distinct topological characteristics, such as the presence or absence of myocardial scar, these methods cannot represent the variation within a single atlas framework. Diffusion models [12,27] and their medical imaging adaptations [16,25] achieve state-of-the-art sample fidelity but lack an explicit mechanism for discovering sub-populations and do not naturally produce discrete prototypical representations for cohort stratification. Spectral and manifold-based methods [23] respect intrinsic data geometry but operate on fixed feature spaces, lacking end-to-end generative capacity for atlas synthesis. None of these paradigms simultaneously addresses label-free representation learning, manifold-aware prototype estimation, and generative modeling across topologically heterogeneous populations.

We propose a unified variational framework that bridges probabilistic latent modeling with spectral manifold geometry for label-free representation learning of medical images. Our contributions are threefold. First, we develop a manifold-anchored variational framework that constrains prototypes to the data manifold, resolving prototype degeneracy and off-manifold drift while accommodating sub-populations with distinct topological characteristics. At its core is a manifold-anchored EM algorithm that extends standard EM with a geometry-aware M-step; the underlying geometric tools, including heat-kernel-based priors and manifold-constrained inference, are general-purpose and readily applicable

to other latent-variable models, establishing a reusable foundation for robust representation learning. Second, we introduce a spectral regularization strategy that preserves the intrinsic geometric structure of the latent space, enabling anatomically coherent transitions between learned sub-populations. Third, we derive per-sub-population uncertainty quantification that surfaces heterogeneous or ambiguous sub-populations, providing clinicians with an automatic quality flag for cohorts requiring re-examination. On cardiac scar and brain MRI benchmarks, our method achieves the highest accuracy among all compared methods, produces anatomically faithful prototypical atlases, and delivers clinically meaningful stratification, all without diagnostic labels.

2 Background: Probabilistic Latent-Variable Modeling via GMM

We model the distribution of N medical images in a d -dimensional latent space $\mathcal{Z} \subseteq \mathbb{R}^d$, positing that latent representations $\mathbf{z} \in \mathcal{Z}$ arise from a *Gaussian Mixture Model* (GMM) with K components corresponding to anatomical or pathological sub-populations,

$$p(\mathbf{z}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{z} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad \text{s.t.} \sum_{k=1}^K \pi_k = 1, \pi_k \geq 0, \quad (1)$$

where π_k denotes the mixing coefficient, and $\boldsymbol{\mu}_k \in \mathbb{R}^d$, $\boldsymbol{\Sigma}_k \in \mathbb{S}_{++}^d$ are the mean and covariance of the k -th component. Maximum-likelihood estimation via Expectation-Maximization (EM) yields the standard M-step updates,

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^N \gamma_{ik} \mathbf{z}_i}{\sum_{i=1}^N \gamma_{ik}}, \quad \boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^N \gamma_{ik} (\mathbf{z}_i - \boldsymbol{\mu}_k)(\mathbf{z}_i - \boldsymbol{\mu}_k)^\top}{\sum_{i=1}^N \gamma_{ik}}, \quad (2)$$

where $\gamma_{ik} \triangleq p(k \mid \mathbf{z}_i; \Theta)$ is the posterior responsibility. While this formulation ensures statistically coherent grouping, the Euclidean averaging of $\boldsymbol{\mu}_k$ disregards the intrinsic geometry of $\mathcal{Z}$, often producing prototypes that lie *off-manifold*—i.e., in regions of negligible data density.

3 Methodology

Our core insight is that latent prototypes should be *manifold-resident representations* rather than Euclidean averages that drift into low-density regions. We construct a heat-kernel graph over latent embeddings (Sec. 3.1) and integrate it into an end-to-end variational framework (Sec. 3.2).

3.1 Manifold Regularization via Spectral Geometry

Heat Kernel Affinity and Dirichlet Energy. Let $\mathcal{M}$ denote the manifold underlying the latent space $\mathcal{Z}$. We approximate $\mathcal{M}$ as a weighted graph $\mathcal{G} =$

$(\mathcal{V}, \mathcal{E}, W)$ over the embeddings $\mathcal{V} = \{\mathbf{z}_i\}_{i=1}^N$, with edge weights given by the heat kernel,

$$W_{ij} = \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{4t}\right), \quad (3)$$

where $t > 0$ denotes a bandwidth parameter, which is adaptively chosen as the median of the pairwise distances within each mini-batch, thereby ensuring that the affinity scale dynamically reflects the evolving latent geometry. From W , we form the graph Laplacian $\mathbf{L} = \mathbf{D} - W$ with $\mathbf{D} = \text{diag}(\sum_j W_{ij})$, and define the Dirichlet energy,

$$\mathcal{E}(\mathbf{Z}) = \frac{1}{2} \sum_{i,j} W_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|^2 = \text{Tr}(\mathbf{Z}^\top \mathbf{L} \mathbf{Z}), \quad (4)$$

where $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]^\top \in \mathbb{R}^{N \times d}$. Minimizing $\mathcal{E}(\mathbf{Z})$ penalizes abrupt transitions across neighbors and preserves the diffusion geometry of $\mathcal{M}$. In practice, W and $\mathcal{E}$ are computed over mini-batches, treated as local, bounded approximations rather than global manifold recovery; batch size B incurs only $O(B^2 d)$ time and $O(B^2)$ memory, avoiding the $O(N^2)$ cost of a global graph.

Manifold-Anchored EM. We replace the standard EM updates (Eq. 2) with geometry-aware alternatives that anchor prototypes to $\mathcal{M}$. We denote the resulting parameters $\Theta^* = \{\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*, \pi_k^*\}_{k=1}^K$. The GMM parameters are periodically refit at fixed training intervals via the following procedure.

E-step. Posterior responsibilities under the current parameters Θ ,

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{c=1}^K \pi_c \mathcal{N}(\mathbf{z}_i \mid \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)}. \quad (5)$$

M-step. Let $\mathcal{C}_k = \{\mathbf{z}_i \mid k = \arg \max_c \gamma_{ic}\}$ denote the hard assignment induced by maximum responsibility. Instead of the arithmetic mean, we select the prototype as the *graph medoid*—the point in $\mathcal{C}_k$ with maximal aggregate heat-kernel affinity,

$$\boldsymbol{\mu}_k^* = \arg \max_{\mathbf{z}_i \in \mathcal{C}_k} \sum_{\ell \in \mathcal{C}_k} W_{i\ell}. \quad (6)$$

The medoid is a discrete analogue of the Fréchet mean on the graph: maximizing aggregate affinity is equivalent to minimizing a sum of diffusion-kernel distances to all members, since $W_{i\ell}$ is a monotonically decreasing function of $\|\mathbf{z}_i - \mathbf{z}_\ell\|^2$. This ensures $\boldsymbol{\mu}_k^* \in \mathcal{C}_k \subset \mathcal{M}$ by construction, yielding prototypes that are manifold-resident rather than artifacts of Euclidean averaging.⁵

The covariance and mixing coefficients retain soft responsibilities γ_{ik} ,

$$\boldsymbol{\Sigma}_k^* = \frac{\sum_{i=1}^N \gamma_{ik} (\mathbf{z}_i - \boldsymbol{\mu}_k^*)(\mathbf{z}_i - \boldsymbol{\mu}_k^*)^\top}{\sum_{i=1}^N \gamma_{ik}} + \epsilon \mathbf{I}, \quad \pi_k^* = \frac{1}{N} \sum_{i=1}^N \gamma_{ik}, \quad (7)$$

⁵ The maximisation in Eq. 6 is over a finite candidate set and a bounded objective, so the M-step is guaranteed to attain its optimum.

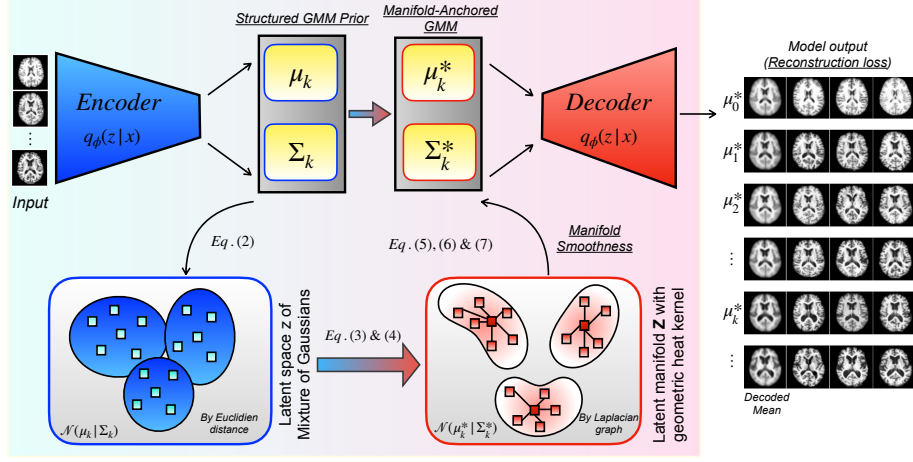


Fig. 1. Overview of the proposed variational framework. The encoder maps input images to a structured latent space governed by manifold-aware Gaussian mixtures, enabling end-to-end modeling of sub-populations with distinct topological characteristics; the decoder reconstructs from the latent embeddings.

where $\epsilon > 0$ ensures $\Sigma_k^* \succ 0$. The prototype uses hard assignments to remain anchored on the manifold, while the covariance retains soft responsibilities to capture the spread of boundary samples across sub-populations. After each MA-EM cycle, the parameters are updated as $\Theta \leftarrow \Theta^*$. In early training, when hard assignments may be noisy due to uncertain responsibilities ($\gamma_{ik} \approx \gamma_{ic}$ for multiple c), the soft covariance estimates mitigate the effect of occasional misassignments.

3.2 Variational Framework via Deep Neural Networks

We now specify the variational model that produces the latent embeddings on which the spectral regularization and MA-EM operate. We define a structured prior,

$$p(\mathbf{z}) = \sum_{k=1}^K \pi_k^* \mathcal{N}(\mathbf{z} \mid \mu_k^*, \Sigma_k^*) \quad (8)$$

directly informed by the MA-EM updates. An encoder parameterizes the approximate posterior $q_\phi(\mathbf{z} \mid \mathbf{x}) = \mathcal{N}(\mathbf{z}; \mu_\phi, \text{diag}(\sigma_\phi^2))$, and a decoder D_θ parameterizes the likelihood $p_\theta(\mathbf{x} \mid \mathbf{z})$ via $\hat{\mathbf{x}} = D_\theta(\mathbf{z})$.

Training Objective. A standard VAE regularizes toward an isotropic prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$, which is incompatible with the multimodal structure of Eq. 8. We instead regularize each sample toward its assigned manifold-anchored component. For each training sample, we identify its most probable component $\hat{k} = \arg \max_k \gamma_{ik}$ and regularize toward $\mathcal{N}(\mu_{\hat{k}}^*, \Sigma_{\hat{k}}^*)$. With $\mathbf{z} \sim q_\phi(\mathbf{z} \mid \mathbf{x})$, the training

objective is,

$$\mathcal{J} = \underbrace{-\log p_\theta(\mathbf{x} | \mathbf{z})}_{\text{reconstruction}} + \underbrace{\beta D_{\text{KL}}\left(q_\phi(\mathbf{z} | \mathbf{x}) \parallel \mathcal{N}(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*)\right)}_{\text{manifold-anchored regularization}} + \underbrace{\lambda \mathcal{E}(\mathbf{Z})}_{\text{manifold smoothness}}. \quad (9)$$

Under a Gaussian likelihood, $-\log p_\theta(\mathbf{x} | \mathbf{z})$ reduces to the mean squared error $\|\mathbf{x} - D_\theta(\mathbf{z})\|_2^2$ up to a constant. The Dirichlet energy $\mathcal{E}(\mathbf{Z})$ (Eq. 4) complements the KL term by promoting topological smoothness across sub-population boundaries. Hyperparameters β and λ balance posterior concentration against geometric regularity.⁶

3.3 Predictive Uncertainty via Monte Carlo Sampling

For each input $\mathbf{x}_i$, we draw M samples $\{\mathbf{z}_i^{(m)}\}_{m=1}^M$ from the approximate posterior $q_\phi(\mathbf{z} | \mathbf{x}_i)$ and decode each to obtain $\hat{\mathbf{x}}_i^{(m)} = D_\theta(\mathbf{z}_i^{(m)})$. We assess predictive uncertainty at two complementary scales. The pixel-wise uncertainty $\mathcal{U}_i(\mathbf{p}) = \text{Std}_{m=1}^M[\hat{\mathbf{x}}_i^{(m)}(\mathbf{p})]$ captures spatial variability under posterior perturbation. The structural uncertainty $\mathcal{S}_i = M^{-1} \sum_{m=1}^M \text{DSSIM}(\hat{\mathbf{x}}_i^{(m)}, \bar{\mathbf{x}}_i)$, where $\bar{\mathbf{x}}_i = M^{-1} \sum_m \hat{\mathbf{x}}_i^{(m)}$ and $\text{DSSIM} = (1 - \text{SSIM})/2$ in which $\text{SSIM} \in [-1, 1]$, captures spatially coherent deviations in local contrast and structure that pixel-wise measures may miss. Aggregating over each sub-population $\mathcal{C}_k$ yields per-sub-population profiles $\bar{\mathcal{U}}_k(\mathbf{p}) = |\mathcal{C}_k|^{-1} \sum_{i \in \mathcal{C}_k} \mathcal{U}_i(\mathbf{p})$ and $\bar{\mathcal{S}}_k = |\mathcal{C}_k|^{-1} \sum_{i \in \mathcal{C}_k} \mathcal{S}_i$ that summarize reconstruction consistency within each sub-population.

4 Experimental Evaluation

Datasets. We evaluated on three datasets of increasing clinical complexity, all split 70%/15%/15% for training/validation/testing. MNIST: 60,000 images across 10 digit classes, serving as a controlled benchmark. Synthetic Cardiac: 5,000 cardiac masks across five morphological classes (healthy, subendocardial, transmural, mid-wall, and epicardial). OASIS Brain MRI [20]: T1-weighted axial slices from 185 subjects, skull-stripped, affinely registered, and resampled to 128×128 , split at the subject level to prevent data leakage.

Experimental Setup. We evaluated representation quality, the fidelity of estimated prototypes, uncertainty quantification, and clinical relevance. All methods were trained under identical conditions with the same K . We compared against VAE-GMM [15] and Diffusion-GMM [28], two generative baselines that share our encoder-decoder backbone and match our multi-task objective of joint sub-population modeling, prototype generation, and uncertainty quantification; this design enables a direct evaluation of the manifold-anchored prior. Non-generative

⁶ Regularising toward a single mixture component is a standard variational approximation that yields a valid evidence lower bound [15]. The Dirichlet energy acts as a pairwise MRF log-prior over $\mathbf{Z}$.

methods such as DEC [29] or spectral clustering provide useful partition metrics but lack the native capacity to synthesize anatomical atlases or spatial uncertainty maps, and therefore fall outside our evaluative scope. Among generative approaches, these are the strongest available baselines: VAE-GMM surpasses discriminative methods such as DEC, and Diffusion-GMM further improves upon it. A standard GMM in image space averaged below 50% accuracy and is omitted from visual comparisons. On labeled datasets (MNIST, Cardiac), we assessed agreement with ground-truth classes quantitatively. On the unlabeled OASIS dataset, we examined estimated prototypes, uncertainty maps, and the association between discovered sub-populations and clinical diagnosis scores.

Implementation Details. All models used Adam with learning rates of 10^{-3} (MNIST, Cardiac) and 10^{-4} (OASIS); our model also used gradient clipping and L_2 weight decay for stable manifold traversal. The variational objective balances reconstruction, posterior concentration (β), and manifold smoothness (λ), tuned per dataset (e.g., $\beta=0.5, \lambda=1.0$ for OASIS). We selected K by the elbow method, where intra-component distance plateaued, yielding $K=10$ for MNIST, $K=5$ for Cardiac, and $K=8$ for OASIS, matching ground-truth class counts for labeled datasets. For unlabeled OASIS, $K=8$ was determined purely by this criterion, independent of clinical metadata, with prototypes stable across neighboring K . We used $M = 30$ Monte Carlo posterior samples for stable uncertainty estimates. Multiple random seeds confirmed consistent gains. At inference, fixed prior parameters require no graph reconstruction, making our framework more efficient than iterative-denoising diffusion baselines.

Evaluation Metrics. For labeled datasets (MNIST, Cardiac), we report Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI), measuring the agreement between discovered sub-populations and ground-truth classes. For the unlabeled OASIS dataset, we use the silhouette score to assess latent sub-population cohesion and separation.

Results. Fig. 2 summarizes results: (a) our method consistently achieves the best performance across datasets; (b) healthy-like archetypes have lower clinical scores and disease-like archetypes higher scores, indicating clearer clinical stratification than baselines; (c) our model produces the sharpest prototypes with preserved scar morphologies, whereas baselines degenerate; and (d) uncertainty maps are well-localized to pathologically salient regions, unlike diffuse, uninformative baseline maps.

Fig. 3 presents the results on the OASIS dataset. The top panel shows that our method produces the sharpest prototypes with well-preserved anatomical details, reflecting better preservation of cortical structure, whereas baselines yield blurred, degenerate prototypes with limited inter-prototype differentiation. The bottom panel compares uncertainty maps: baselines show diffuse, spatially uniform uncertainty across sub-populations, whereas our method concentrates uncertainty around high-variability regions such as ventricular boundaries and cortex. Its higher pixel-wise uncertainty reflects a well-calibrated posterior rather than inferior reconstruction: unlike baselines that collapse the posterior into over-

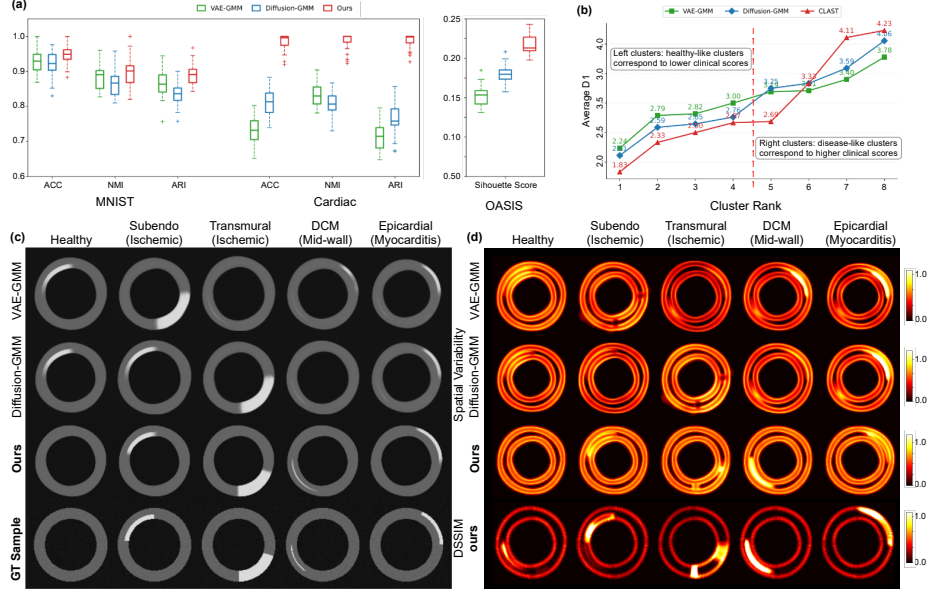


Fig. 2. (a) Quantitative performance across models, (b) clinical relevance by diagnosis scores and (c) estimated prototypes with (d) uncertainty maps.

confident, smoothed outputs, our manifold-anchored prior preserves meaningful variance in biologically complex regions. Together with superior quantitative performance (Fig. 2), this confirms disentanglement of individual anatomical variation from population-level structure, supporting robust clinical stratification.

5 Conclusion

We presented a manifold-anchored variational framework that eliminates off-manifold drift by selecting prototypes as heat-kernel medoids on the latent manifold. The manifold-anchored EM at its core is a general-purpose geometric extension of standard EM that is readily applicable to other latent-variable models beyond this setting. Experiments on cardiac scar and brain MRI confirm consistent gains in accuracy, sharper prototypical atlases, and well-calibrated uncertainty estimates, with no reliance on diagnostic labels. The per-sub-population uncertainty scores offer a practical tool for flagging ambiguous sub-populations that may warrant clinical re-examination, and the label-free nature of the framework makes it directly applicable to cohort stratification in settings where expert annotations are scarce, costly, or poorly reproducible. Looking ahead, quantitative validation of uncertainty quality, such as calibration or correlation with reconstruction error, would further strengthen its clinical utility; coupling the learned sub-populations with deformable registration would yield voxel-wise templates

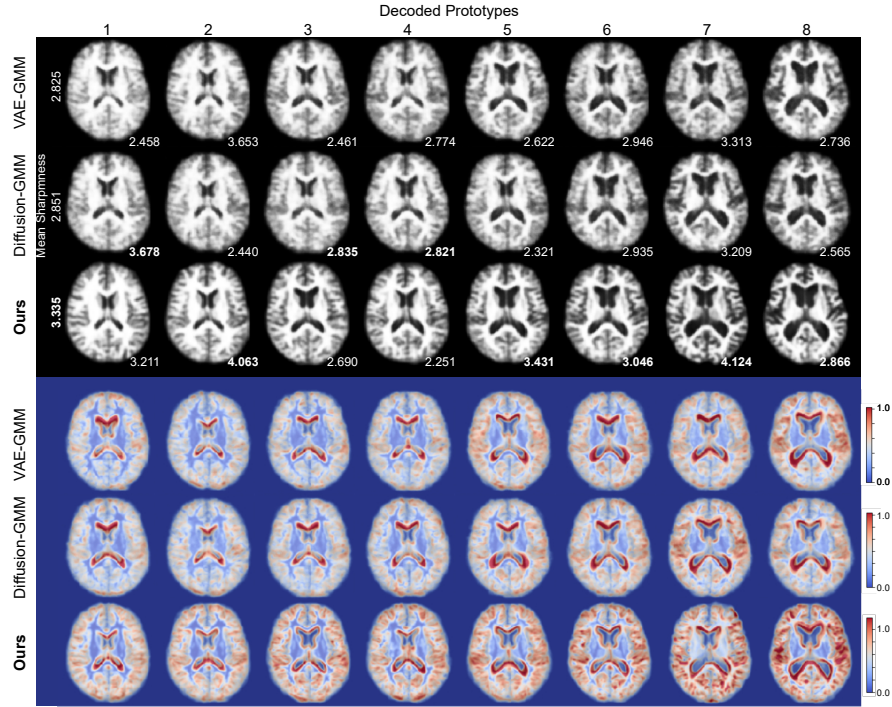


Fig. 3. Top: estimated brain prototypes with corresponding sharpness estimates across all models. Bottom: uncertainty maps estimated from all methods.

per sub-population, while applying the framework to longitudinal data could enable data-driven monitoring of disease progression.

Disclosure of Interests The authors declare that they have no competing interests.

References

1. Abulnaga, S.M., Hoopes, A., Dey, N., Hoffmann, M., Fischl, B., Gutttag, J., Dalca, A.: Multimorph: On-demand atlas construction. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30906–30917 (2025)
2. Arvanitidis, G., Hansen, L.K., Hauberg, S.: Latent space oddity: on the curvature of deep generative models. In: International Conference on Learning Representations (ICLR) (2018)
3. Arvanitidis, G., Hauberg, S., Schölkopf, B.: Geometrically enriched latent spaces. In: International Conference on Artificial Intelligence and Statistics (AISTATS). pp. 631–639. PMLR (2021)
4. Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., Natarajan, V., Norouzi, M.: Robust

- and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nature Biomedical Engineering* **7**(6), 756–779 (2023)
5. Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W.: Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *Journal of Neuropathology & Experimental Neurology* **71**(4), 266–273 (2012)
 6. Chen, T., Kornblith, S., Norbert, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: *International Conference on Machine Learning (ICML)*. pp. 1597–1607. PMLR (2020)
 7. Dalca, A.V., Rakic, M., Gutttag, J., Sabuncu, M.R.: Learning conditional deformable templates with convolutional networks. *Advances in Neural Information Processing Systems* **32** (2019)
 8. Dilokthanakul, N., Mediano, P.A.M., Garnelo, M., Lee, M.C.H., Salimbeni, H., Arulkumaran, K., Shanahan, M.: Deep unsupervised clustering with Gaussian mixture variational autoencoders. *arXiv preprint arXiv:1611.02648* (2016)
 9. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115–118 (2017)
 10. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**(22), 2402–2410 (2016)
 11. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 9729–9738 (2020)
 12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 33, pp. 6840–6851 (2020)
 13. Iglesias, J.E., et al.: A probabilistic histological atlas of the human brain for MRI segmentation. *Nature* (2025)
 14. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., et al.: The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* **27**(4), 685–691 (2008)
 15. Jiang, Z., Zheng, Y., Tan, H., Tang, B., Zhou, H.: Variational deep embedding: an unsupervised and generative approach to clustering. In: *International Joint Conference on Artificial Intelligence (IJCAI)*. pp. 1965–1972 (2017)
 16. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D.: Diffusion models in medical imaging: a comprehensive survey. *Medical Image Analysis* **88**, 102846 (2023)
 17. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. In: *International Conference on Learning Representations (ICLR)* (2014)
 18. Li, L., Wu, F., Yang, G., Xu, L., Wong, T., Mohiaddin, R., Firmin, D., Keegan, J., Zhuang, X.: Recent advances in fibrosis and scar segmentation from cardiac MRI: a state-of-the-art review and future perspectives. *Frontiers in Physiology* **12**, 709230 (2022)
 19. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al.: The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncology* **23**(8), 1231–1251 (2021)

20. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* **19**(9), 1498–1507 (2007)
21. Marquand, A.F., Rezek, I., Buitelaar, J., Beckmann, C.F.: Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biological Psychiatry* **80**(7), 552–561 (2016)
22. Min, E., Guo, X., Liu, Q., Zhang, G., Cui, J., Long, J.: A survey of clustering with deep learning: from the perspective of network architecture. *IEEE Access* **6**, 39501–39514 (2018)
23. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 14 (2001)
24. Ou, Z., Jiang, C., Liu, Y., Zhang, Y., Cui, Z., Shen, D.: A graph-embedded latent space learning and clustering framework for incomplete multimodal multi-class Alzheimer’s disease diagnosis. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. pp. 45–55. Springer (2024)
25. Peng, W., Adeli, E., Zhao, Q., Pohl, K.M.: Generating realistic 3D brain MRIs using a conditional diffusion probabilistic model. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI)*. pp. 14–24. Springer (2023)
26. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Weiner, M.W.: Alzheimer’s Disease Neuroimaging Initiative (ADNI): clinical characterization and 12-month follow-up. *Neurology* **74**(3), 201–209 (2010)
27. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 10684–10695 (2022)
28. Wang, P., Zhang, H., Zhang, Z., Chen, S., Ma, Y., Qu, Q.: Diffusion models learn low-dimensional distributions via subspace clustering. *arXiv preprint arXiv:2409.02426* (2024)
29. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: *International Conference on Machine Learning (ICML)*. pp. 478–487. PMLR (2016)