

Longitudinal self-supervision to disentangle inter-patient variability from disease progression

Raphaël Couronné^{1*}, Paul Vernhet¹, and Stanley Durrleman¹

Inria - Aramis project-team, Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France

Abstract. The problem of building disease progression models with longitudinal data has long been addressed with parametric mixed-effect models. They provide interpretable models at the cost of modeling assumptions on the progression profiles and their variability across subjects. Their deep learning counterparts, on the other hand, strive on flexible data-driven modeling, and additional interpretability - or, as far as generative models are involved, *disentanglement* of latent variables with respect to generative factors - comes from additional constraints. In this work, we propose a deep longitudinal model designed to disentangle inter-patient variability from an estimated disease progression timeline. We do not seek for an explicit mapping between age and disease stage, but to learn the latter solely from the ordering between visits using a differentiable ranking loss. Furthermore, we encourage inter-patient variability to be encoded in a separate latent space, where for each patient a single representation is learned from its set of visits, with a constraint of invariance under permutation of the visits. The modularity of the network architecture allows us to apply our model on various data types: a synthetic image dataset with known generative factors, cognitive assessments and neuroimaging data. We show that, combined with our patient encoder, the ranking loss for visits helps to exceed models with supervision, in particular in terms of disease staging.

Keywords: Longitudinal model · Disentanglement · Medical Imaging

1 Introduction

Understanding the progression of diseases is essential for accurate early diagnosis, prognosis, and patient monitoring. Often, there is a strong interplay between the pathological progression and the inter-subject variability, which makes it all the more necessary to characterize the contribution of each factor. Typically, in the context of neurodegenerative diseases, we may ask whether the atrophy of a particular brain region is predictive of a specific patient advancement in the disease, or rather can be dismissed as a specific characteristic of the individual.

Longitudinal data analysis has been usually addressed in the framework of parametric mixed-effect models. For instance, geometric approaches have been

* The first two authors have equal contribution.

proposed either for the progression of biomarkers [21] or shape changes [4]. This family of models assumes that each subject follows a curve on a Riemannian manifold which translates from a common geodesic. They also assume that the direction of translation is orthogonal to the direction of the progression curve, which ensures that the changes due to the progression of the disease are disentangled from the effects of different physiological or anatomical characteristics of the patient. The family of progression profiles is constrained, e.g. sigmoid curves for biomarkers changes, and an affine function maps the age of the subject to a disease stage.

Generative models such as variational auto-encoders (VAE) [13] have been consistently used in deep learning as they offer a flexible learning framework, in which disentanglement may be enforced through soft constraints and optimization schemes, as in β -VAE and their extensions [10,12,18]. With time series, however, separating static and dynamic representations without inductive bias still remains a challenge. In most research works, authors disentangle time-varying from time-invariant information by leveraging time labels explicitly: in literature focused on style and content of videos [9,15], in face ageing progression [11] or medical data, where age is used for supervision [19,22]. These previous methods are not directly transferable to longitudinal data where duration between visits differs. In [1], authors seek an age direction in the latent *a posteriori*, while in [24] they estimate the latent age regression jointly with the reconstruction task in a supervised fashion. A Riemannian manifold learning point of view, in the spirit of parametric models is proposed in [17] as it estimates both a static representation and an affine time reparametrization per patient. All these methods assume that age at observation is a direct marker for the progression timeline, which is not the case for most neurodegenerative disorders. In the recent work closest to ours [25], the authors propose to learn the disease stage without relying on the patient age, in a self-supervised fashion. They use a cosine loss to enforce progression in a specific direction of the latent space, learned during optimization. They do not study the disentanglement of their model but rather focus on the correlation with a disease progression timeline.

In this paper, we propose a generic deep longitudinal model, designed to disentangle inter-patient variability from an estimated disease progression timeline. We learn a disease stage as a flexible function that does not rely on age, but solely on the individual order between visits using a differentiable ranking loss, leveraging a much weaker prior. The remaining latent space is further favored to produce representations independent of the progression thanks to a DeepSet network which acts as a permutation invariance function on visits. The main contributions of this paper are therefore (i) an architecture that is tailored to disease progression modeling and disentangles the changes due to progression from the changes due to phenotypic differences across subjects; (ii) a modular method with decoders adapted to data types; (iii) an application on synthetic and real datasets - including imaging and clinical data - showing that one direction of the latent space alone describes temporal progression.

2 Methodology

The proposed generic deep longitudinal model is summarized in Fig 1.

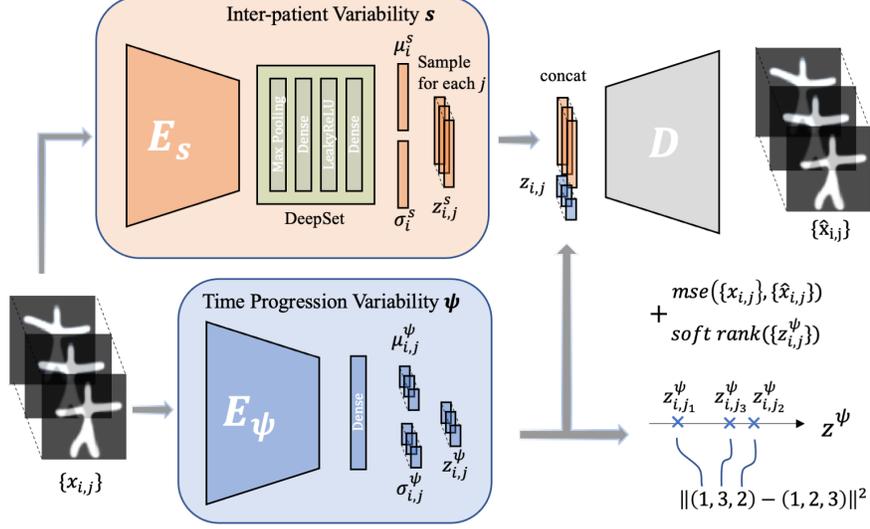


Fig. 1: Input data $\mathbf{x}_i = \{x_{i,j}, \forall j \in [1, n_i]\}$ is encoded simultaneously in a space encoder (Deepset) and a point-wise time encoder to get respectively (μ_i^s, σ_i^s) and $(\mu_{i,j}^\psi, \sigma_{i,j}^\psi)$. (μ_i^s, σ_i^s) can be computed from *any* subset of visits, and in practice randomized fixed-size subsets of visits are drawn in the spirit of stochastic optimization. Note that we sample a $z_{i,j}^s$ per visit, but z_i^s could be sampled once for a patient. Decoder can be either agnostic, or specific (e.g., velocity fields for deformations).

2.1 Longitudinal progression model

In this section, we propose a temporal latent variable model that encodes the disease progression in a low-dimensional probabilistic space. It assumes that a sequence of observations is generated as the combination of an intrinsic code z^s (as in *space shift*) and a disease progression factor z^ψ (where we use ψ instead of t to clearly distinguish the stage from the temporality of visits).

Generative disease progression model Let $\{(t_{i,j}, x_{i,j})\}_{1 \leq i \leq N}$ be a set of N subjects, each observed at the age of $t_{i,j}$ for $1 \leq j \leq n_i$ visits. We assume that the patient i 's observations $\mathbf{x}_i = \{x_{i,j}, \forall j \in [1, n_i]\}$ are generated from a Bayesian generative model as follows:

$$x_{i,j} \stackrel{\text{iid}}{\sim} \mathcal{N}\left\{\Phi(z_i^s, z_{i,j}^\psi); \epsilon^2 \mathbb{I}\right\} \text{ with } z_i^s \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \lambda_s^2 \mathbb{I}) \text{ and } z_{i,j}^\psi \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \lambda_\psi^2) \quad (1)$$

Here Φ denotes an unknown non-linear transform from the strongly decoupled generative factor space, also called latent space, $\mathcal{Z} = \mathcal{Z}^s \times \mathcal{Z}^\psi$, towards our observation space of scores or images \mathcal{X} . The generative factor z^ψ is assumed to be independent from the individual variability z^s factor. Notice that we do not assume any relationship between the ages t_{ij} and the associated observations x_{ij} . Parameters ϵ^2 , λ_ψ^2 and λ_s^2 are the diagonal Gaussian variance priors.

Variational Inference with VAE The inference is conducted within the VAE paradigm. The function Φ is approximated by a parametric class of neural network Φ_θ , the decoder. Two neural network encoders are used to approximate the intractable posterior distribution $p(z_i^s, z_i^\psi | \mathbf{x}_i; \theta)$ with $\mathbf{z}_i^\psi = \{z_{i,j}^\psi, \forall j \in [1, n_i]\}$. They respectively model the latent distributions of space shifts and disease progression, such that the approximated parametric distribution can be factorized as a product of independent Gaussians $\mathcal{N}(\mu^s, \Sigma^s) \otimes \mathcal{N}(\mu^\psi, \Sigma^\psi)$.

Set-invariant representation for \mathcal{Z}^s The strong condition on \mathcal{Z}^s is that it should extract from any time-series a time-invariant representation. To do so, we choose to learn the posterior $q_{\eta^s} \equiv \mathcal{N}(\mu^s, \Sigma^s)$ as a DeepSet encoder network [23] acting on any unordered subset of visits. This rewriting of our disentanglement hypothesis generalizes the use of simple operators such as averaging or maxing out of visits in an intermediate latent representation, or even more elaborate inverse Gaussian product of group-wise non-*iid* distributions [5]; all these can indeed be cast as specific choices of permutation-invariance operators, which were shown in [23] to be universally approximated by DeepSets.

Ranking visits in \mathcal{Z}^ψ as a regularization constraint The remaining generative factor of our one-dimensional \mathcal{Z}^ψ space must encode the dynamic of the progression preferentially. A disease progression constraint, $\mathcal{C}^{\text{ranking}}$, aims at favoring a natural ordering of visits along the temporal latent dimension in a self-supervised way. Unlike in [25], it builds upon the soft-ranking differentiable loss of [3] to enforce monotonic individual progression. The individual ranking errors of visits are penalized according to $\sum_j \|r(z_j^\psi) - j\|_2^2$. Note that this loss depends on the number of visits of patient i , and may be rescaled accordingly (with e.g. the expectation assuming random ordering). We found it more practical to choose a fixed number of 3 visits per patient at each iteration. The stochastic gradient descent randomly selects these visits at each iteration. From a theoretical perspective, the minimization of this loss is similar to maximizing the Spearman correlation between different visits of a given subject i : it is therefore to be understood as a soft supervision which only relies on the ordering of visits and not on the times of observations, which are never seen by the model.

Final objective The final loss can be written as the sum of two terms: the evidence lower bound, written as the sum of the KL-divergence $\text{KL}[q_\eta(\mathbf{z}_i^\psi, z_i^s | x_i) || p(\mathbf{z}_i)]$ and the data attachment term $-\mathbb{E}[\log p_\theta(\mathbf{x}_i | \mathbf{z}_i)]$ which is proportional to the ℓ^2 reconstruction loss, and the self-supervised ranking of visits $\mathcal{C}^{\text{ranking}}$ discussed previously with a weight γ to cross-validate (in practice we choose $\gamma = 0.1$):

$$L = \sum_{i=1}^N \text{KL}[q_{\eta}(z_i^{\psi}, z_i^s | \mathbf{x}_i) || p(z_i)] - \sum_{j=1}^{n_i} \mathbb{E}[\log p_{\theta}(x_{i,j} | z_{i,j})] + \gamma \mathcal{C}_i^{\text{ranking}} \quad (2)$$

2.2 Modularity

The model we propose in this paper can be seen as a “*meta*” architecture that can be instantiated according to the datatype: clinical data (1D), images (2D and 3D). We may even use decoders that are specifically designed for data, such as diffeomorphometry for brain grey matter. We detail the latter case, directly in the spirit of classical models with stationary velocity fields [7, 14]. An additional parameter, a template \mathcal{T} , is learned at the centered reference disease stage $z^{\psi} = 0$. From this, any observation x_{ij} can be reconstructed from the latent code $(z_{i,j}^{\psi}, z_i^s)$ by a deformation field Φ_v , parametrized with a velocity decoder v , acting on \mathcal{T} .

3 Experimental results

The network E_{ψ} is a classical CNN encoder with LeakyReLU non-linearities and a final dense layer toward \mathcal{Z}^{ψ} . The spatial encoder E_s is composed as a DeepSet, whose output does not depend upon the ordering of its inputs: it can be written as $\rho \circ \max_{j \sim \text{visits}}(f)$. We chose f to be a convolutional encoder network which outputs an intermediate representation (per observation), from which the max operator (over visits) retrieves a permutation invariant code. The latter is eventually mapped, *via* a MLP ρ , into the space shifts domain \mathcal{Z}^s . We choose a latent space dimension of $p = 5$ for Starmen and ADNI cognitive scores, and of $p = 8$ for ADNI MRIs. Inference was performed using Adam optimizer from PyTorch library version 1.7 with a learning rate of 0.01 and a batch size of 32. To promote fair comparison between the different models of the benchmark, the same architecture were used for encoders, as well as decoders.

3.1 Validation on synthetic data

To validate the disentangling ability of our model, we first generated a synthetic longitudinal dataset of starmen images, based on the longitudinal diffeomorphic model of [6]. From a given reference template y_0 , the cross-sectional variability of our population is prescribed by a diffeomorphism localized at four control points: the head, right arm and legs. The common progression timeline, on the other hand, is generated through a displacement of the left arm only.

The dynamics of progression is given by an affine reparametrization of the age t_{ij} at visit j , characterized by individual onset τ_i and acceleration α_i factors, such that the true disease progression is given by $\psi_{ij}^* = t_0 + \alpha_i(t_{ij} - \tau_i - t_0)$. We sample variables in a similar fashion as in [6] to obtain a dataset of $N = 1000$ subjects, each with $n = 10$ visits (Fig 2).

Metric	β -VAE	ML-VAE	LR-AE	AR-VAE	LSSL	Ours	Ours (wD)	Ours (woR)
MSE (10^{-3})	7.90 ± 0.57	22.7 ± 1.51	10.9 ± 1.53	8.26 ± 0.62	7.32 ± 0.379	8.83 ± 0.88	6.22 ± 1.23	14.2 ± 5.46
PLS z^ψ/z^s	- -	0.660 ± 0.343	0.137 ± 0.209	0.125 ± 0.117	0.098 ± 0.047	0.083 ± 0.026	0.083 ± 0.025	0.149 ± 0.131
Staging ψ^*	0.263 ± 0.348	0.030 ± 0.028	0.971 ± 0.024	0.984 ± 0.008	0.994 ± 0.003	0.997 ± 0.001	0.996 ± 0.002	0.524 ± 0.464

Table 1: Benchmark of proposed methods on Starmen dataset

We benchmark contender approaches mentioned in the introduction on the Starmen dataset (see Table 1): β -VAE [10], ML-VAE [5], LSSL [25] and both supervised “Longitudinal Riemannian VAE” (LR-AE) [16] and “Age Regression VAE” (AR-VAE) [8]. We compare with our model in its generic iconic form on pixels and its diffeomorphic version (wD). The version without ranking loss (woR) is added for ablation purposes. Beyond the reconstruction quality (Mean Square Error), which reveals only LSSL and our model are below the baseline of β -VAE, we are interested in disentanglement capacity. It can be measured by correlations between the estimated staging z^ψ and the latent space code z^s with a partial least square regression analysis (PLS), so as to ensure the independence of $\mathcal{Z}^s \times \mathcal{Z}^\psi$ (2nd row). Among all methods, ours performs best: even though LSSL uses a similar loss, it does not constrain its orthogonal directions to be independent. Other methods, especially supervised, naturally learn correlated representations.

Finally, the proper staging of ψ^* is evaluated by computing the Spearman ranking correlation between ψ^* and z^ψ : it evaluates the monotonicity of individual trajectories. Only methods with time supervision or ranking strategy (*ie* all but first two columns) manage to grasp a staging close to one (row 3).

In complement to the previous metrics, we visualized in Fig 3 the effect of specific directions in the latent space *via* gradient maps. The β -VAE is a low baseline as it does not model the progression and only views the data as static representations: we plot the PCA in the whole space, and observe that no principal direction correlates with the left-arm progression. It is interesting to note that the benefit of the ranking loss (when no supervision is available) is made clear by the study of our model without it (woR), and ML-VAE. They both focus on group-structure only, and fail to grasp progression. LR-AE and AR-AE, because they use supervision of time, are displaying time-related correlations in a space shift direction (last row).

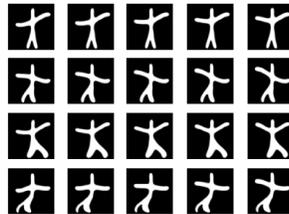


Fig. 2: Each row represents a synthetic subject across time.

3.2 Application to Alzheimer’s disease

Cognitive scores We apply our model on four subtest scores of the ADAS-Cog scale obtained from the ADNI dataset, namely concentration, praxis, memory and language; normalized between 0 and 1, with higher values indicating lower performance. 248 patients with mild cognitive impairment (MCI) converting to Alzheimer’s disease (AD) during the study are followed for an average of 6 visits over 3.5 years.

MSE (10^{-3}) on 5-fold cross validation yields 7.47 ± 0.778 for our model: slightly less than β -VAE (3.78 ± 0.562) and LSSL (3.64 ± 0.429), but on par with a parametric model of reference, Leaspy [21] (8.21 ± 0.155). Additionally, we can predict future visits from previous ones: in this scenario, our model reaches a lower MSE (10^{-3}) 29.1 ± 5.53 than LSSL 32.4 ± 5.93 .

Figure 4 illustrates the estimated average time progression for each model, as well as the effects of orthogonal directions in the latent space. An agnostic β -VAE fails at extracting a consistent dimension for the time progression while still pro-

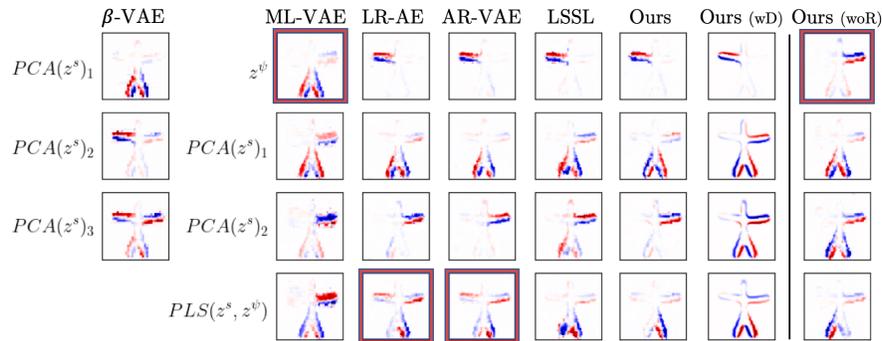


Fig. 3: Gradient directions in the latent space (extracted from a forward pass in the decoder). Row 1: gradient wrt to the latent space associated with disease progression. Rows 2 and 3: first two principal directions of the PCA in the orthogonal of the latent time (Z^s for us). 4th row: the direction in the orthogonal of the latent time that correlates most with it (PLS), as a way to challenge the model disentanglement.

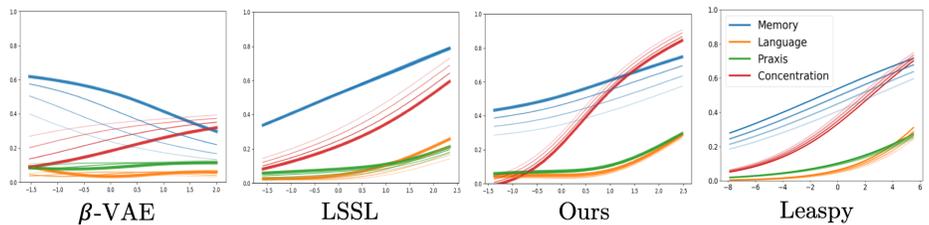


Fig. 4: Estimated average trajectory of scores. The effects of latent dimensions z^s (resp. z with β -VAE) are shown with degraded colors.

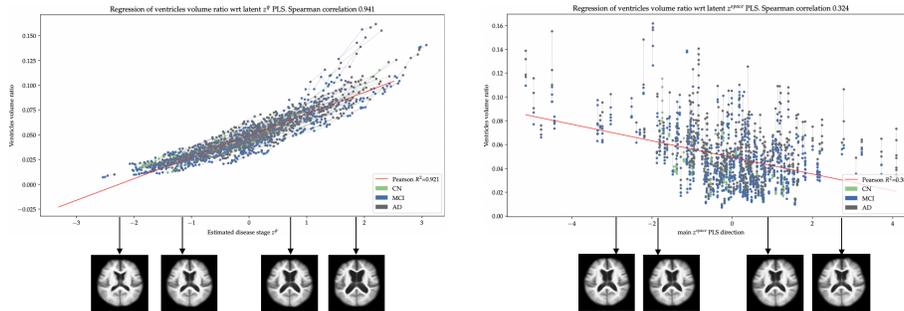


Fig. 5: PLS analysis with respect to ventricle volume ratio \mathcal{V} : z^ψ (left), z^s (right).

viding a good reconstruction of data. Our proposed model learns monotonicity in an unsupervised fashion, and is able to estimate a consistent time progression of the ADAS-Cog scores from longitudinal measurement in small number. Interestingly, despite the noisiness of cognitive scores, our model generates a progression very similar to that of Leaspy’s sigmoid geodesics: in particular, the first PCA directions have the same effects on scores (cf degraded color effects).

Neuroimaging data We selected 356 subjects MCI converters from ADNI, with a total of 1898 visits. The 1898 T1-weighted MRI were preprocessed using Clinica [20] (non-rigid alignment, skull stripping, intensity rescaling), and converted into mid axial slices of dimension 128×128 .

We do not have access to the true disease progression, however the disease severity at the MCI stage can be monitored through cognitive scores (ADAS-Cog) or markers of the morphological evolution such as atrophy of the hippocampi and increase of ventricle volumes. We computed the ratio of ventricle volumes by brain volumes, as a covariate factor, noted \mathcal{V} , which we assimilate with a good proxy of the disease progression.

Figure 5 shows interactions between our two latent spaces and the disease stage proxy \mathcal{V} via a correlation (PLS) analysis. First, z^ψ exhibits a quasi-linear regression fit associated with a high Spearman ranking (0.934 ± 0.025), suggesting that z^ψ has indeed captured the main disease progression trend \mathcal{V} and the individual ordering of visits. On the other hand, the best correlated direction (measured as the PLS main direction) between z^s and \mathcal{V} is not localized on the ventricle. This result further implies that \mathcal{Z}^s has indeed captured the variability necessary to perform a good fit of the data without correlating significantly with the disease stage marker.

Furthermore, z_ψ behaves as a clinical score informative of the onset: its distribution is significantly earlier for women ($p < 3.83e^{-2} \pm 7.05e^{-2}$ for Mann-Whitney U test), as observed from clinical data. Its derivative $\frac{\partial z_\psi}{\partial t}$ correlates with the pace of the disease: it is significantly ($p < 4.82e^{-2} \pm 1.06e^{-2}$) faster for APOE4 carriers (1 or 2 alleles), a result in accordance with well-documented disease progression patterns in AD [2].

4 Conclusion

In this paper, we proposed a generative variational autoencoder architecture that leverages the repetition of measurements per individual to disentangle between the global disease timeline and inter-patient variability. The one-dimensional time variability is captured by a differentiable ranking loss, while a permutation invariant function reduces the remaining information in a representative space. As we further demonstrate, inductive biases on the data itself (such as using diffeomorphometry for structural medical imaging) are completely synergetic and improve the quality of the representations learned. A very interesting avenue would be to extend our method to account for non-monotonic progression priors. This work has been partly funded by the European Research Council (ERC) under Grant Agreement No. 678304, European Union’s Horizon 2020 research and innovation program under Grant Agreement No 826421 (TVB-Cloud), and the program “Investissements d’avenir” ANR-10-IAIHU-06 (IHU-A-ICM) and ANR-19-P3IA-0001 (PRAIRIE 3IA Institute).

References

1. Berchuck, S.I., Mukherjee, S., Medeiros, F.A.: Estimating Rates of Progression and Predicting Future Visual Fields in Glaucoma Using a Deep Variational Autoencoder. *Scientific Reports* **9**(1), 18113 (Dec 2019). <https://doi.org/10.1038/s41598-019-54653-6>, <https://www.nature.com/articles/s41598-019-54653-6>, number: 1 Publisher: Nature Publishing Group
2. Bigio, E., Hynan, L., Sontag, E., Satumtira, S., White, C.: Synapse loss is greater in presenile than senile onset alzheimer disease: implications for the cognitive reserve hypothesis. *Neuropathology and applied neurobiology* **28**(3), 218–227 (2002)
3. Blondel, M., Teboul, O., Berthet, Q., Djolonga, J.: Fast Differentiable Sorting and Ranking. In: International Conference on Machine Learning. pp. 950–959. PMLR (Nov 2020), iSSN: 2640-3498
4. Bône, A., Louis, M., Martin, B., Durrleman, S.: Deformetrica 4: an open-source software for statistical shape analysis. In: ShapeMI @ MICCAI 2018. Granada, Spain (Sep 2018), <https://hal.inria.fr/hal-01874752>
5. Bouchacourt, D., Tomioka, R., Nowozin, S.: Multi-level variational autoencoder: Learning disentangled representations from grouped observations. *CoRR* **abs/1705.08841** (2017)
6. Bône, A., Colliot, O., Durrleman, S.: Learning Distributions of Shape Trajectories From Longitudinal Datasets: A Hierarchical Model on a Manifold of Diffeomorphisms. pp. 9271–9280 (2018)
7. Dalca, A.V., Rakic, M., Guttag, J., Sabuncu, M.R.: Learning Conditional Deformable Templates with Convolutional Networks. arXiv:1908.02738 [cs, eess] (Oct 2019), arXiv: 1908.02738
8. Gao, L., Pan, H., Liu, F., Xie, X., Zhang, Z., Han, J., Initiative, A.D.N., et al.: Brain disease diagnosis using deep learning features from longitudinal mr images. In: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data. pp. 327–339. Springer (2018)
9. Grathwohl, W., Wilson, A.: Disentangling space and time in video with hierarchical variational auto-encoders. arXiv preprint arXiv:1612.04440 (2016)

10. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: beta-vae: Learning basic visual concepts with a constrained variational framework (2016)
11. Hsu, W.N., Zhang, Y., Glass, J.: Unsupervised learning of disentangled and interpretable representations from sequential data. *Advances in neural information processing systems* **30**, 1878–1889 (2017)
12. Kim, H., Mnih, A.: Disentangling by Factorising. arXiv:1802.05983 [cs, stat] (Jul 2019), arXiv: 1802.05983
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
14. Krebs, J., Delingette, H., Ayache, N., Mansi, T.: Learning a Generative Motion Model from Image Sequences based on a Latent Motion Matrix. arXiv:2011.01741 [cs] (Nov 2020), arXiv: 2011.01741
15. Li, Y., Mandt, S.: Disentangled sequential autoencoder. arXiv preprint arXiv:1803.02991 (2018)
16. Louis, M., Charlier, B., Durrleman, S.: Geodesic discriminant analysis for manifold-valued data. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. pp. 332–340 (2018)
17. Louis, M., Couronné, R., Koval, I., Charlier, B., Durrleman, S.: Riemannian Geometry Learning for Disease Progression Modelling. In: *Information Processing in Medical Imaging*, vol. 11492, pp. 542–553. Springer International Publishing (2019), series Title: Lecture Notes in Computer Science
18. Mathieu, E., Rainforth, T., Siddharth, N., Teh, Y.W.: Disentangling disentanglement in variational autoencoders. In: *International Conference on Machine Learning*. pp. 4402–4412. PMLR (2019)
19. Ravi, D., Alexander, D.C., Oxtoby, N.P.: Degenerative adversarial neuroimage nets: Generating images that mimic disease progression (2019)
20. Routier, A., Burgos, N., Guillon, J., Samper-González, J., Wen, J., Bottani, S., Marcoux, A., Bacci, M., Fontanella, S., Jacquemont, T., Wild, A., Gori, P., Guyot, A., Lu, P., Díaz, M., Thibeau-Sutre, E., Moreau, T., Teichmann, M., Habert, M.O., Durrleman, S., Colliot, O.: Clinica: an open source software platform for reproducible clinical neuroscience studies (Oct 2019), <https://hal.inria.fr/hal-02308126>
21. Schiratti, J.B., Allasonniere, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajectories from manifold-valued longitudinal data. In: *Advances in Neural Information Processing Systems*. pp. 2404–2412 (2015)
22. Xia, T., Chatsias, A., Tsiftaris, S.A.: Consistent brain ageing synthesis. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. pp. 750–758. Springer International Publishing, Cham (2019)
23. Zaheer, M., Kottur, S., Ravanbakhsh, S., Póczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*. vol. 30, pp. 3391–3401. Curran Associates, Inc. (2017)
24. Zhang, Z., Song, Y., Qi, H.: Age Progression/Regression by Conditional Adversarial Autoencoder. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4352–4360. IEEE, Honolulu, HI (Jul 2017). <https://doi.org/10.1109/CVPR.2017.463>
25. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: LSSL: Longitudinal Self-Supervised Learning. arXiv:2006.06930 [cs, stat] (Jun 2020), <http://arxiv.org/abs/2006.06930>, arXiv: 2006.06930