Individualized multi-horizon MRI trajectory prediction for Alzheimer's Disease

Rosemary He[1,2][0000−0001−8307−3958], Gabriella Ang[1][0009−0009−7901−6809], Daniel Tward[2,3,4][0000−0002−4607−6807], and for the Alzheimer's Disease Neuroimaging Initiative*

[1] Departments of Computer Science, University of California, Los Angeles
[2] Departments of Computational Medicine, University of California, Los Angeles
[3] Departments of Neurology, University of California, Los Angeles
[4] Corresponding author, DTward@mednet.ucla.edu

Abstract. Neurodegeneration as measured through magnetic resonance imaging (MRI) is recognized as a potential biomarker for diagnosing Alzheimer's disease (AD), but is generally considered less specific than amyloid or tau based biomarkers. Due to a large amount of variability in brain anatomy between different individuals, we hypothesize that leveraging MRI time series can help improve specificity, by treating each patient as their own baseline. Here we turn to conditional variational autoencoders to generate individualized MRI predictions given the subject's age, disease status and one previous scan. Using serial imaging data from the Alzheimer's Disease Neuroimaging Initiative*, we train a novel architecture to build a latent space distribution which can be sampled from to generate future predictions of changing anatomy. This enables us to extrapolate beyond the dataset and predict MRIs up to 10 years. We evaluated the model on a held-out set from ADNI and an independent dataset (from Open Access Series of Imaging Studies). By comparing to several alternatives, we show that our model produces more individualized images with higher resolution. Further, if an individual already has a follow-up MRI, we demonstrate a usage of our model to compute a likelihood ratio classifier for disease status. In practice, the model may be able to assist in early diagnosis of AD and provide a counterfactual baseline trajectory for treatment effect estimation. Furthermore, it generates a synthetic dataset that can potentially be used for downstream tasks such as anomaly detection and classification.

Keywords: Conditional VAE · Disease progression · Alzheimer's Disease.

---

1   Introduction

1.1   Background

Neurodegeneration as observed in magnetic resonance imaging (MRI) is recognized as a potential biomarker for diagnosing Alzheimer's disease (AD)[13]. While MRI has the advantage of being noninvasive, it is generally considered not specific enough, unlike biomarkers of amyloid or tau[14]. The potential for using machine learning approaches ("data-driven statistical approaches in which many different brain regions are evaluated simultaneously"[2]) has been recognized. However, an important challenge in making diagnoses from brain images is the large amount of interpersonal variability, compared to minor structural changes in the earliest stages of the disease such as the transentorhinal stage (Braak stage I and II)[5]. We hypothesize that this issue can be overcome by carefully modeling timeseries of imaging data, allowing each person to be compared to their own baseline, rather than to a population average.

1.2   Related work

One strategy for timeseries analysis in brain imaging is to quantify structural changes using the diffeomorphism group[4]. After defining an appropriate metric, geodesics can be computed to study changes over time[27,7,29]. More recently, deep learning has been applied to model disease progression via changing pixel values, with models including deep structural causal models[1], variational autoencoders (VAE)[33,24], generative adversarial networks (GAN)[31,21], and diffusion models[32,19]. Among these works, we highlight two that train on timeseries MRIs to predict the aging brain, conditioned on age and diagnosis. The first [24] combines a variational autoencoder with a mixed-effect model imposed on the latent space to simulate the aging process. By explicitly modeling age in the latent space, this model synthesizes images across 30+ years with atrophy patterns consistent with expectation. However, the model focuses on providing a population trajectory and is not specific enough for individual predictions, and images produced tend to be lower resolution as is typical of VAEs. In the second, 4D-DANI-Net, a GAN model with specified loss functions and a weight profile function, is proposed to simulate the aging process both globally and locally[22]. Additional techniques are used to synthesize images that are high resolution, more individualized, and consistent with atrophy patterns associated to AD. Unlike VAEs which typically impose a Gaussian distribution on the latent space through their encoder, GAN models typically offer a uniform distribution where no point is more likely than any other, and cannot be directly used for statistical modeling to classify trajectories as normal or abnormal.

1.3   Our Contribution

In this work, we aim to develop an approach that combines the complementary strengths of the two highlighted previous methods. We maintain an interpretable

latent space in the VAE framework, but introduce a new architecture to condition on past images leading to more realistic and higher resolution outputs. This is achieved by a novel double-encoder CVAE architecture to model changing pixel values. Our model takes in conditional inputs of two data types: a previous image and other demographic information (age, elapsed time and disease status), leading to a novel network architecture where the encoder and decoder don't exhibit the symmetry typical in VAE work. This approach allows prediction for an arbitrary elapsed time, i.e. multi-horizon prediction. Compared to previous VAE-based methods[24], we increase output resolution and subject specificity by incorporating a prior MRI. Compared to previous GAN-based approaches[22], our method produces a latent space, in which we can estimate a posterior probability of disease status for disease classification.

Second, we introduce a novel training strategy to reduce computational requirement per sample and augment the dataset, where the network is trained on all possible pairs of images in a timeseries. The model we develop can predict future brain images with good accuracy, and we validate it in terms of mean square error (MSE) on the held out test and independent external dataset, while comparing to other baseline methods. Lastly, if a pair of scans has already been obtained, we show how our encoder can be used to build a likelihood ratio classifier. By focusing on individual trajectories in timeseries of MRI rather than population averages, our work has the potential to improve neurodegeneration biomarkers for AD.

We list our contribution as follows: i) we design a novel architecture that directly takes in both conditional images and demographic variables, ii) we propose a novel training strategy for computational cost reduction and data augmentation, iii) we propose a novel utilization of the latent space produced by our encoder for disease classification.

2   Method

2.1   Data Preparation

We obtain our dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`), led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI is to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease. We include subjects from ADNI 1, 2 and 3 studies as of July 11, 2023, by searching for all MPRAGE scans. Here, we introduce a novel data preparation strategy to solve two problems we face in medical imaging datasets: data scarcity and computational limits. We structure our dataset so that each sample contains two images from the same patient, age and disease status when the first image was taken, and the time difference between the two images. For subjects with only one image, a pair of the same image is included, and the time difference is 0. For subjects with more than 2 images, we include all combinations of pairs both forward (positive time difference) and backward (negative time difference) in time. Including a negative time difference may or may not have direct clinical

utility, but doubles our sample size. For example, if a person has 3 images we take the pairs $\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$. This training strategy converts serial inputs into pairs of images, effectively augmenting our dataset and reducing computational costs per sample during training. We note this may introduce a bias where patients with more scans are overrepresented, and alternative sampling schemes will be the subject of future research. To simulate real data, often lower quality than carefully curated public datasets, we performed minimal quality check and did not discard any images.

For scalar conditional variables, we standardize age and time difference to mean 0 and variance 1. ADNI groups patients into 6 categories: cognitive normal, subjective memory complaint, early mild cognitive impairment, mild cognitive impairment, late mild cognitive impairment and AD. We follow the same grouping and assign a number from 0 to 5, noting this is a reasonable approximation of slightly different disease categories across ADNI studies. We assume that this progression is ordered, so we use a single number rather than a one hot encoded representation. We take only the diagnosis for the first scan of the pair, assuming future diagnostic labels are unknown. For the training set, we constructed 15,579 pairs from 907 subjects. For the test set, we held out 4 subjects at random from each category, for a total of 24 subjects, 180 images and 430 pairs. Characteristics of our ADNI cohort are shown in Table 1. In addition to ADNI, we obtain an independent set from OASIS3[18] with 1085 subjects for external validation. We follow the same preprocessing procedure as above. Since OASIS and ADNI have different conventions for disease status, we keep consistency by grouping OASIS subjects into cognitive normal, cognitive impairment and AD, and assign them a number of 0, 3 and 5.

Each image was rigidly aligned to the 2020 MNI extended nonlinear symmetric average template [10]. First, the template was aligned to each brain image using an affine transformation, accounting for contrast differences using code from [28] (`https://github.com/twardlab/emlddmm`). Second, a Procrustes method was used to project this transform onto the closest rigid transform[6] by minimizing sum of square error in voxel locations. This ensured that size and shape differences would be modeled by our network, and not "normalized away" by registration. We found through visual inspection that an *accurate* affine transformation followed by projection led to less variability, relative to an *inaccurate* rigid transformation directly. Finally, we cropped and downsampled all images to $80 \times 80 \times 80$, 2mm^3 voxels due to computational limitations.

2.2   Autoencoders, VAE and CVAE

First, we give an overview of autoencoder models and their extensions. Autoencoders are a class of neural network methods that learn a low-dimensional representation of high-dimensional structured data[8]. They consist of two parts: an encoder that projects high dimensional data into a latent space with lower dimensions, and a decoder that learns to map a point in the latent space back to its high dimensional representation. The latent distribution of an autoencoder is unknown, making inference difficult and prompting the need for VAEs[16].

Table 1: ADNI cohort characteristics stratified by disease status.

	Normal (n=219)	SMC (n=73)	EMCI (n=253)	MCI (n=86)	LMCI (n=140)	AD (n=136)
Male	104(47.5%)	34(46.6%)	133(52.6%)	53(61.6%)	71(50.7%)	71(52.2%)
Age*	75.51 [6.91]	71.87 [5.6]	70.84 [7.23]	76.27 [8.37]	72.02 [7.92]	74.92 [7.98]
50-70	49(22.4%)	35(47.9%)	123(48.6%)	21(24.4%)	54(38.6%)	29(21.3%)
70-80	112(51.1%)	33(45.2%)	99(39.1%)	33(38.4%)	63(45%)	75(55.1%)
80-90	56(25.6%)	4(5.5%)	31(12.3%)	32(37.2%)	22(15.7%)	31(22.8%)
90+	2(0.9%)	1(1.4%)	0	0	1(0.7%)	1(0.7%)
Scans per subject	4.74 [2.76]	2.44 [1.14]	4.7 [1.87]	6.53 [4.71]	4.49 [1.68]	3.53 [2.09]
Length (years)	2.48 [1.83]	1.61 [0.95]	2.48 [1.6]	1.98 [1.44]	2.24 [1.53]	0.88 [0.69]

*mean and [standard deviation] of age at first visit

VAE models assume the samples x are generated from a distribution conditioned on the latent space z, where $z \sim p_\theta(z)$ and $x \sim p_\theta(x|z)$[16]. In practice, posterior inference may be intractable and estimating the posterior distribution is challenging[16]. To address this problem, a more tractable distribution, $q_\phi(z|x)$, is used to approximate the true posterior[16]. Therefore, VAEs can be trained to efficiently sample from unknown distributions using the variational lower bound of the log-likelihood, which can be written as follows[16]:

$$\mathcal{L}(\theta, \phi, x) = -D_{KL}(q_\phi(z|x)||p_\theta(z)) + E_{q_\phi(z|x)}[\log p_\theta(x|z)], \qquad (1)$$

where D_{KL} is the Kullback–Leibler divergence[17] that measures how similar two distributions are, and $E_{q_\phi(z|x^i)}$ measures the sum of squared reconstruction error where we assume $p_\theta(x|z)$ is Gaussian with mean predicted by our decoder and fixed variance 1. We use the common choice that $p_\theta(z)$ is multivariate standard normal, with no learnable parameters.

As with many medical applications, conditional variables offer additional information and can improve parameter estimation. A natural extension to the VAE is CVAE[26], which includes an additional variable y as the conditional variable. The variational lower bound for CVAE to optimize is as follows[26]:

$$\mathcal{L}_{\text{CVAE}}(\theta, \phi, x, y) = -D_{KL}(q_\phi(z|x,y)||p_\theta(z)) + E_{q_\phi(z|x,y)}[\log p_\theta(y|x,z)]. \quad (2)$$

2.3   Our Double Encoder CVAE

We present our novel model architecture, inspired by CVAE[26], to generate 3D MR images. While previous methods have conditioned on a learned representation of an image, our architecture allows for a direct conditional image input. In our model, the encoder and decoder are not "symmetric", but rather the encoder takes the form of a standard CNN with two image inputs, and the decoder takes the form of a U-net[23]. The latter allows the prior MR image and conditional variables, as well as the latent space representation to be used for decoding. A point in the latent space does not represent an image, but a transformation between images, linking VAE modeling and classical work with diffeomorphisms.

Objective function We follow the same underlying assumptions of the distribution process as CVAE[26], where MR images are generated from the distribution $p_\theta(x|z)$ conditioned on 4 variables: the base image, age, disease status and elapsed time between two images. Though we have multi-modality conditional variables, the generation inference remains the same. Therefore, our empirical objective function follows closely to that of CVAE[26] in Equation (2).
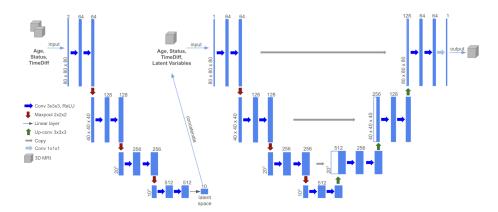

Fig. 1: Model architecture.

Model architecture Due to our conditional variables' multi-modality nature, we implement a novel architecture to incorporate scalar variables (age, elapsed time, disease status) and a conditional 3D image. The overall architecture consists of an encoder and a U-net[23] shaped decoder. The encoder consists of 4 blocks, where each block contains 2 convolutional layers and 1 max pooling layer. We choose a kernel size of 3 and stride of 1 for all convolutional layers except the last, and a kernel size of 2 and stride of 2 for max pooling layers. Each convolution is followed by a group normalization layer[30] with group size of 4 and a ReLU layer. During training, we take in an input of 2 stacked images and conditional variables, and project it down to a latent space of dimension 10 (i.e. we output 10 means and 10 log-variances). The decoder takes in the conditional image, latent space sample and scalar variables as inputs, and outputs the predicted image. Since the downsampling branch of the U-Net can be thought of as an encoder, we call this architecture a "double encoder CVAE". All non-image inputs are linearly transformed and added to each layer before the nonlinearity, following a strategy similar to "timestep embedding" common in modern diffusion models [11]. We show a visualization of our model in Fig. 1 and make our code available on GitHub at `https://github.com/rosie068/Double_Encoder_CVAE_AD`.

Training procedure We used the Adam optimizer[15] with learning rate 1e-5 and batch size of 4. We trained our model for 1000 epochs with early stopping on a single NVIDIA GeForce RTX 4090 GPU with a training time of 5 days.

2.4   Comparison to alternative methods

We attempted a comparison of our method to those described in section 1.2, but were unable to produce satisfactory results using publicly available code and documentation. We omit the comparisons, rather than casting these methods in a negative light. We suggest that a challenge style comparison, where authors can optimize parameter selection for their own methods and have them run automatically on a hidden test set, would be more appropriate for head to head comparisons. In this work we compare to several simpler models to put our method into context and understand typical values of our figures of merit. Our goal is to provide insight into the challenge of longitudinal image prediction by developing a new approach, not claim superiority or inferiority. For a simple baseline, we use the conditional image as a prediction of the future image, and the other methods we compare to are described below.

Low rank linear prediction For N individuals, we take the first and last images in their trajectories and form two matrices X and Y, both of size 80^3 (number of pixels) by N. We include demographic variables by appending additional rows to X for age, status, elapsed time and 1 (for mean). We perform singular value decomposition: $X = USV^T$, and take the first 10 (the same dimension as our CVAE latent space) or 100 (for better results) singular vectors as the latent space representation. To estimate the future MRI $\hat{Y}$ for individual i, we use the first few singular values (indicated by $\tilde{\cdot}$) and calculate as: $\hat{Y}_i = Y\tilde{V}\tilde{S}^{-1}\tilde{U}^T X_i$

VAE with linear mixed effect estimation We train an autoencoder using the architecture described in [24], and fit a linear mixed effects (LME) model in the latent space using the statsmodels package in python [25]. While other authors [33] have noted that this approach should be improved upon, it provides a simple model based on standard VAEs for comparison. Images are predicted using the following procedure. First a baseline image is passed through the encoder. Second, its latent space representation is shifted based on our LME model parameters, in a manner depending on time difference and disease status. Third, the resulting vector is passed through the decoder.

3   Results

3.1   Trajectory prediction

First, we compare results for trajectory reconstruction in the held-out test set. In Fig. 2, we visualize sample reconstructions for randomly selected healthy and

AD individuals in the test set. As a healthy brain does not change much over 1 year, the ground truth does not have much variation and both the 100 dim. SVD method and our method perform well. In the diseased trajectory, we observe a noticeable enlargement in the ventricles, which is predicted well by our method but not the linear methods. The 10 dimensional SVD and VAE+LME model did not produce satisfactory images, and we do not show their outputs.
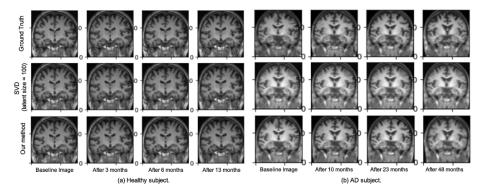

Fig. 2: Sample prediction in the held-out test set.

For a quantitative comparison, we calculate mean squared error (MSE) in test and external validation datasets for 5 methods: our double encoder CVAE, VAE+LME model, SVD with both latent space sizes, and baseline. We conduct comparisons in three regions of interest (ROIs) to reflect global and local performance: the entire image, a region surrounding the hippocampus and a region surrounding the ventricles (both associated with AD progression[9,3]). ROIs were identified during image registration, using an affine transform of the MNI atlas and dilating annotations by several pixels. Out of 230 pairs, our model achieves the best performance in 202 pairs (88%) in the whole brain and 218 pairs in both the hippocampus and ventricles (95%). In the OASIS validation set, our model achieves the lowest MSE across all three regions, but the distribution has longer tails. Out of 1085 pairs of images, our model achieves the best performance in 855 pairs (79%) in the whole brain, 889 pairs (82%) in the hippocampus, and 922 pairs (85%) in the ventricles. In Fig. 3, we visualize the distribution of MSE on a log scale and show our proposed model reconstructs trajectories with more accuracy globally and locally. While still performing best, our method seems to follow a bimodal distribution with long tails in OASIS data, an observation which warrants further investigation.

In addition, we demonstrate our model's ability to interpolate between images and extrapolate beyond the last image in the timeseries. We generate images by sampling a standard multivariate normal in the latent space, and pass it through the decoder with the original image and scalar variables. The random latent sample is fixed across all panels, and only the elapsed time is changed. In Fig. 4,
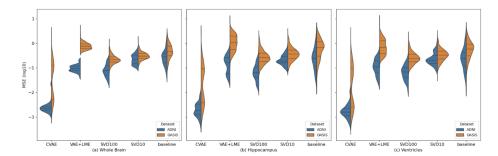
Fig. 3: MSE (log scale) in test (left) and external (right) validation set in 3 ROIs.

we predict the trajectory annually over 10 years for one individual in the test set with disease status 4 (LMCI) and only two real images available (extrapolation beyond the 4th year). We visualize structural changes over time by calculating a smooth optical flow estimate, and overlaying the divergence of the flow field. Values close to 0 are transparent, positive (expansion) is shown in blue, and negative (contraction) is shown in red. We observe expansion in the ventricular space over time including its inferior horn in the temporal lobe, and contraction of nearby brain structures, in accordance with the AD aging process.
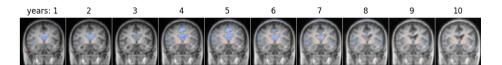

Fig. 4: Ten year prediction from one base image. Incremental changes (non-cumulative) shown via optical flow divergence, blue: expansion, red: contraction.

3.2   Latent space disease status probability estimation

In addition to trajectory estimation, we propose a practical use of the latent space to estimate a posterior probability of a subject's disease status given a pair of images. For subjects in the test set, we set the disease status to both normal and AD to obtain two sets of latent space representations. Let $\mu_{i,j}$ be the jth dimension of the encoder output when disease status is i (0 for healthy and 5 for AD), and define $f_i = \prod_{j=1}^{10} (2\pi)^{-1/2} \exp(-\mu_{ij}^2/2)$ (taking a product over the 10 dimensional outputs, reflecting our standard normal latent space distribution). Our posterior probability of disease status 0 is defined as $p_0 = f_0/(f_0 + f_5)$. We show two examples of disease classification analysis in Fig. 5, one healthy and one AD (same subjects as Fig. 2), where the posterior probability significantly favors the hypothesis corresponding to their true disease status.
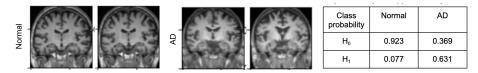
Class probability	Normal	AD
H_0	0.923	0.369
H_1	0.077	0.631

Fig. 5: Hypothesis testing for disease status, H_0: null, H_1: AD.

4   Discussion

In this work, we present our double encoder CVAE, a novel architecture that predicts 3D MRI at an arbitrary time given a conditional prior image, age, disease status, and the future time. We compared our model on two datasets against other simpler methods, and showed our model is better at tracking both global and local changes during the aging process associated with AD. Our goal is to design a predictive model that operates on images and demographics only. While ADNI provides other biomarker information such as amyloid and tau pathology, we do not include them in this work but will consider in the future. One strength of our work is that our model learns aging patterns from populations without losing specificity when making individualized predictions. Given a baseline image, it could be used to understand potential trajectories of healthy aging or disease. Another strength is that our model produces a latent space for downstream inference, as demonstrated by our likelihood ratio classifier example. Our model has characteristics similar to "image translation" like the popular pix2pix model[12]. However, pix2pix is deterministic[20], whereas our model allows users to explore uncertainty in future trajectories. Lastly, model predictions could be used for downstream analysis, including treatment effect comparison (where our model predicts the trajectory with no intervention).

One limitation of this work is that we were unable to perform head-to-head comparisons with state-of-the-art methods, and we suggest this could be addressed in the future in a challenge framework, where each author can run their own code. Another is that it is not straightforward to incorporate more than one previous scan when making predictions, unlike in [24]. Future work will incorporate serial conditional images to increase predictive performance. Lastly, our test set size was relatively small due to a smaller cohort size compared to some computer vision datasets. With potential treatments for AD now approved, early diagnosis has become critical. Our model can generate synthetic datasets that can potentially be used for downstream tasks such as anomaly detection and classification. This work, which seeks to improve measures of neurodegeneration as a biomarker of AD by leveraging timeseries analysis, has the potential to impact treatment decisions.

References

1. Abdulaal, A., Castro, D.C., Alexander, D.C.: Deep structural causal modelling of the clinical and radiological phenotype of alzheimer's disease. In: NeurIPS 2022 Workshop on Causality for Real-world Impact (2022)
2. Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., et al.: The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & dementia **7**(3), 270–279 (2011)
3. Apostolova, L.G., Green, A.E., Babakchanian, S., Hwang, K.S., Chou, Y.Y., Toga, A.W., Thompson, P.M.: Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (mci), and alzheimer disease. Alzheimer Disease & Associated Disorders **26**(1), 17–27 (2012)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision **61**, 139–157 (2005)
5. Braak, H., Braak, E.: Neuropathological stageing of alzheimer-related changes. Acta neuropathologica **82**(4), 239–259 (1991)
6. Challis, J.H.: A procedure for determining rigid body transformation parameters. Journal of biomechanics **28**(6), 733–737 (1995)
7. Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. International journal of computer vision **103**, 22–59 (2013)
8. Ehrhardt, J., Wilms, M.: Autoencoders and variational autoencoders in medical image analysis. In: Biomedical Image Synthesis and Simulation, pp. 129–162. Elsevier (2022)
9. Evans, M.C., Barnes, J., Nielsen, C., Kim, L.G., Clegg, S.L., Blair, M., Leung, K.K., Douiri, A., Boyes, R.G., Ourselin, S., et al.: Volume changes in alzheimer's disease and mild cognitive impairment: cognitive associations. European radiology **20**, 674–682 (2010)
10. Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., Group, B.D.C., et al.: Unbiased average age-appropriate atlases for pediatric studies. Neuroimage **54**(1), 313–327 (2011)
11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
12. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
13. Jack Jr, C.R., Albert, M.S., Knopman, D.S., McKhann, G.M., Sperling, R.A., Carrillo, M.C., Thies, B., Phelps, C.H.: Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & dementia **7**(3), 257–262 (2011)
14. Jack Jr, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., et al.: Nia-aa research framework: toward a biological definition of alzheimer's disease. Alzheimer's & Dementia **14**(4), 535–562 (2018)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

16. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)

17. Kullback, S.: Kullback-leibler divergence (1951)

18. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., et al.: Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv pp. 2019–12 (2019)

19. Puglisi, L., Alexander, D.C., Ravì, D.: Enhancing spatiotemporal disease progression models via latent diffusion and prior knowledge. arXiv preprint arXiv:2405.03328 (2024)

20. Raad, R., Ray, D., Varghese, B., Hwang, D., Gill, I., Duddalwar, V., Oberai, A.A.: Conditional generative learning for medical image imputation. Scientific Reports **14**(1), 171 (2024)

21. Ravi, D., Alexander, D.C., Oxtoby, N.P., Initiative, A.D.N.: Degenerative adversarial neuroimage nets: generating images that mimic disease progression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 164–172. Springer (2019)

22. Ravi, D., Blumberg, S.B., Ingala, S., Barkhof, F., Alexander, D.C., Oxtoby, N.P., Initiative, A.D.N., et al.: Degenerative adversarial neuroimage nets for brain scan simulations: Application in ageing and dementia. Medical image analysis **75**, 102257 (2022)

23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)

24. Sauty, B., Durrleman, S.: Progression models for imaging data with longitudinal variational auto encoders. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–13. Springer (2022)

25. Seabold, S., Perktold, J.: statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference (2010)

26. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. Advances in neural information processing systems **28** (2015)

27. Thomas Fletcher, P.: Geodesic regression and the theory of least squares on riemannian manifolds. International journal of computer vision **105**, 171–185 (2013)

28. Tward, D., Brown, T., Kageyama, Y., Patel, J., Hou, Z., Mori, S., Albert, M., Troncoso, J., Miller, M.: Diffeomorphic registration with intensity transformation and missing data: Application to 3d digital pathology of alzheimer's disease. Frontiers in neuroscience **14**, 52 (2020)

29. Tward, D.J., Sicat, C.S., Brown, T., Bakker, A., Gallagher, M., Albert, M., Miller, M., Initiative, A.D.N.: Entorhinal and transentorhinal atrophy in mild cognitive impairment using longitudinal diffeomorphometry. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring **9**(1), 41–50 (2017)

30. Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)

31. Xia, T., Chartsias, A., Tsaftaris, S.A., Initiative, A.D.N.: Consistent brain ageing synthesis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. pp. 750–758. Springer (2019)

32. Yoon, J.S., Zhang, C., Suk, H.I., Guo, J., Li, X.: Sadm: Sequence-aware diffusion model for longitudinal medical image generation. In: International Conference on Information Processing in Medical Imaging. pp. 388–400. Springer (2023)
33. Zhao, Q., Adeli, E., Honnorat, N., Leng, T., Pohl, K.M.: Variational autoencoder for regression: Application to brain aging analysis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 823–831. Springer (2019)