ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888742

Controllable Generative Model for Brain Evolution

Gengshuo Liu[§] Nikhil N. Chaudhari[§] Nikos Kanakaris Chenzhong Yin
Paul Bogdan Andrei Irimia

University of Southern California, Los Angeles, CA, USA

***Abstract*—Today's generative models can synthesize magnetic resonance images (MRIs) of the brain at specific ages. However, such models can neither map the aging process longitudinally within subjects, nor accommodate its variability across subjects. Such approaches also cannot predict anatomic features of aging in ways that can be validated retrospectively or trusted prospectively. We introduce a three-dimensional hybrid ControlNet + diffusion model that uses the baseline T_1-weighted MRIs of healthy adults to predict individual neuroanatomic aging trajectories, as reflected by follow-up MRIs. The approach captures individual anatomical changes with an average predicted voxelwise intensity error of 15% and structural similarity index of 93%. Unlike methods relying on qualitative validation, our approach quantifies the fidelity of prospective MRI synthesis using FreeSurfer volumetrics. Because brain atrophy reflects risk for Alzheimer's disease (AD), our model's ability to generate individual-specific prospective MRIs suggests its clinical potential to assist AD risk estimation.**

***Index Terms*—Brain evolution, diffusion, ControlNet, MRI.**

I. Introduction

Motivation. Aging is a primary risk factor for neurodegenerative diseases [1]–[4], whose likelihood increases with more brain tissue loss. The ability to anticipate future brain atrophy from baseline magnetic resonance imaging (MRI) can therefore help to assess the risk of such diseases [5]–[7] and to reduce it through lifestyle changes or prophylactic medical interventions. Advances in deep learning have enabled the synthesis of age-dependent MRIs, which could lead to better ability to map brain atrophy over time and to estimate the risk of aging-related disease [8], including Alzheimer's disease (AD). Ultimately, this could lead to better strategies for clinical disease prevention and earlier intervention.

Challenges in the state of the art & challenge. Generative models have synthesized T_1-weighted MRIs that are typical of participants at a chronological age (CA) specified by the user [9], while neglecting individual variability in aging [10], [11]. Enabling the prediction of individualized age-related anatomical changes can be beneficial for clinicians, since it would allow them to establish a baseline for typically aging brains and to track deviations from this baseline in patients with AD. Researchers have also explored ways to anticipate an individual's future MRIs. In [12], a variational autoencoder was used to synthesize future MRIs before AD. Producing clear, high-quality MRIs, and controlling the CA of the brain in the synthetic MRI was challenging. The authors of [13] showed that a variational encoder-based structural causal model can synthesize future MRIs. Nevertheless, a quantitative evaluation of generated MRIs was not provided. In [14], [15], generative adversarial networks (GANs) were used to synthesize 2D MRI slices of the brain across the lifespan using only cross-sectional MRIs. However, this ignores both 3D spatial information between slices and longitudinal information, which impacts quantitative validation. Additionally, the large computational requirements of GANs limit their ability to generate 3D MRIs. Several studies [16], [17] used patches/sub-volumes to overcome the burden of full volume synthesis at the cost of distant spatial information. Alternatively, they have sought to generate synthetic MRIs to increase training data sample sizes [9], rather than to predict brain aging for clinical applications.

Contributions. Our 3D diffusion-based ControlNet model predicts individual aging trajectories from their baseline brain MRIs to synthesize their future MRIs. Whereas predicted scans produced by other generative models have been appraised only *qualitatively*, we validate our predicted scans *quantitatively* against the ground truth provided by participants' longitudinal follow-up MRIs and by their segmentation-derived regional brain volumes. Predicting future brain scans may assist the prognosis of aging-related brain diseases, thus reducing the risk of neurodegenerative diseases including Alzheimer's disease.

II. Methods

Participants and neuroimaging. This study complies with the US Code of Federal Regulations (45 C.F.R. 46) and with the Declaration of Helsinki. Longitudinal T_1-weighted MRIs (baseline and two-year follow-up; interval μ = 2.25 years; σ = 0.11 years) were obtained from the UK Biobank (UKBB) [18]; ethical approval was obtained from the North West Multi-Centre Research Ethics Committee of the United Kingdom. All participants provided informed written consent (Table I); MRI acquisition parameters are available elsewhere [18].

Model overview. We design a subject-specific brain MRI generator using a deep generative diffusion model [19] that synthesizes participants' future MRIs from baseline MRIs by learning the intrinsic mapping between participants' CAs and their MRIs. The (input) baseline MRI is used as a prior within a ControlNet [20] to achieve subject-specific prediction of participants' follow-up MRIs. Starting from noise, the aggregate model (diffusion U-Net + ControlNet) uses baseline MRIs and follow-up CA as inputs to refine the synthetic MRI output iteratively. This produces MRIs that align with both individual structural features and age-related anatomy changes.

Conditional 3D diffusion. Inspired by [20], we design a 3D conditional diffusion model as a backbone for brain MRI generation. The diffusion process involves adding Gaussian noise progressively to a 3D MRI at each time step t, followed

[§]Equal contribution

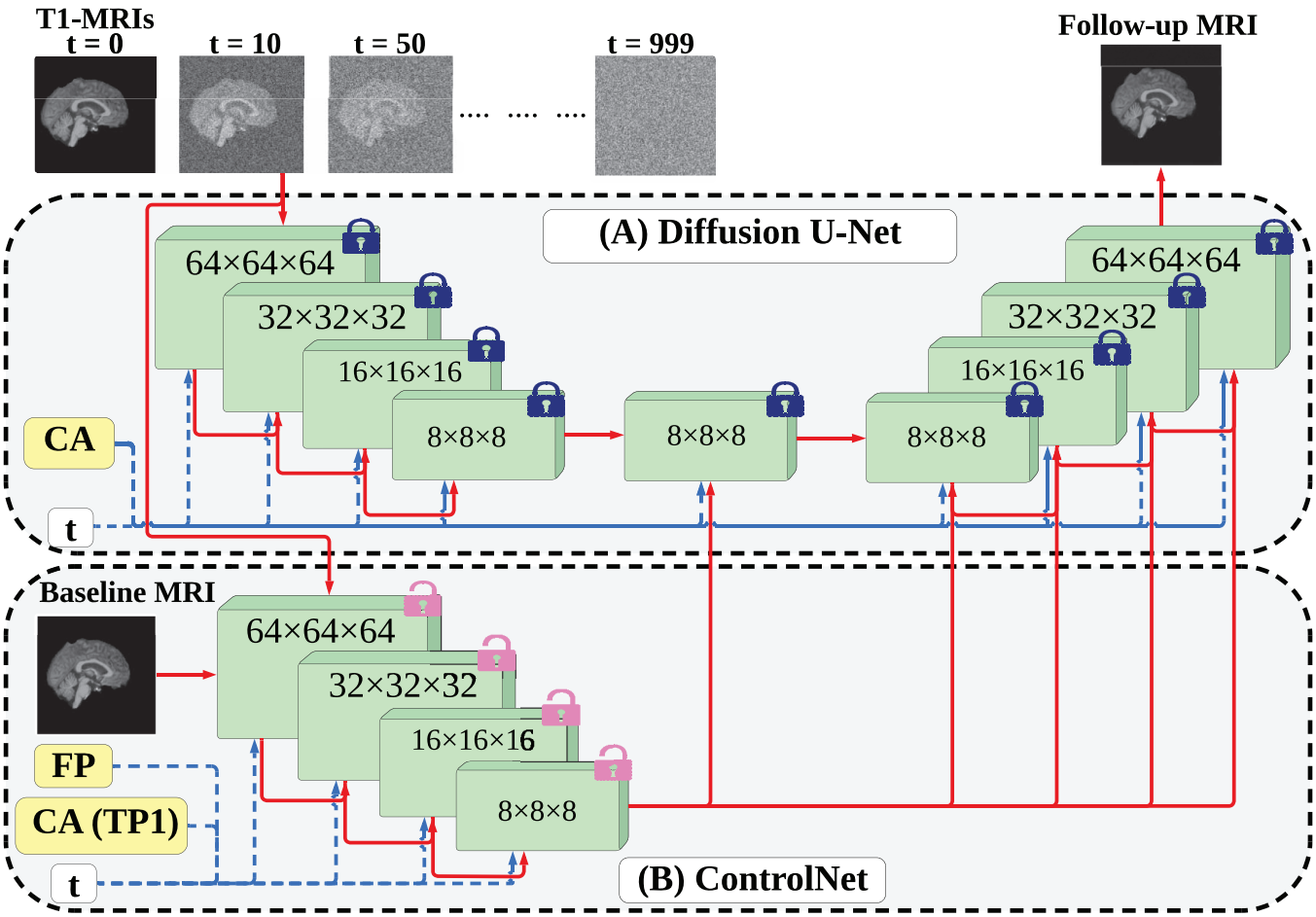

Fig. 1. **Model architecture**. Our model includes a 3D U-Net diffusion model (A) and a ControlNet model (B). Red solid lines represent forward propagation of MRI features, blue dashed lines indicate injection paths for time step t and conditional information from the MRI and chronological age (CA); the latter are conditional inputs for the diffusion backbone. The follow-up period is FP (the time in the future for which the synthetic MRI is generated) and the participant's age at baseline (time point 1) is $CA(TP1)$. These are conditional inputs for the ControlNet. A blue lock indicates that network parameters are frozen; a pink lock indicates that parameters are trainable.

by a reverse process to denoise and restore the original MRI. In the forward process, Gaussian noise ϵ is added to the input MRI $\mathbf{x}_0$, producing noisy MRIs $\mathbf{x}_t$ at each step according to $q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1-\alpha_t)\mathbf{I})$, where α_t is a diffusion parameter. The reverse process is performed by a 3D U-Net $\epsilon_\theta(\mathbf{x}_t, t)$, which predicts and removes noise iteratively, eventually synthesizing a noise-free 3D MRI (Figure 1). The U-Net has four 3D convolutional blocks, across which MRI size shrinks from 128^3 to 8^3 while the number of channels increases from 64 to 128. Each block includes residual blocks, starting with two in the first layer and decreasing to one in the final layer. Self-attention functionality in the last two layers enhance the ability to capture global information.

One key model feature is the integration of *conditional control* through age embedding. This propagates the effects of age, as a variable, across the high-dimensional representation space of the MRI, thereby facilitating the conditioning of the latter upon complex mapping relationships between aging and neuroanatomy. By concatenating age embeddings with time-step representations of the MRI, the model incorporates both MRI noise levels and age-specific features during denoising. This allows the generation of MRIs with age-related characteristics. During training, the model learns the anatomic features associated with various ages, including features that describe brain atrophy. This enables the age-conditioned 3D U-Net to guide synthetic MRI outputs such that the brains they display reflect CAs of the desired age (CA + FP).

ControlNet. The ControlNet has conditional inputs (baseline MRIs) to guide diffusion when synthesizing subject-specific follow-up MRIs. By using baseline MRI features as priors and incorporating follow-up age-specific characteristics, the ControlNet ensures that synthetic MRIs reflect individual brain evolution accurately. The ControlNet leverages our pretrained Conditional 3D diffusion model whose parameters are frozen to retain its strong backbone. A trainable copy is introduced, allowing adaptation to conditional inputs (such as baseline MRIs and follow-up ages). Downsampling blocks identical to those in the diffusion model process conditional inputs. This reduces MRI size gradually and extracts anatomic features into a middle block. The latter acts as a repository of structural features extracted from baseline MRIs, which provide prior knowledge for synthesizing realistic, participant-specific

TABLE I
PARTICIPANT DEMOGRAPHICS. DESCRIPTIVE STATISTICS OF TRAINING AND TESTING SAMPLES, INCLUDING SAMPLE SIZE N, AGE (MINIMUM, MAXIMUM, MEAN μ, AND STANDARD DEVIATION σ) AND THE MALE-TO-FEMALE (M:F) RATIO. ALL SUBJECTS WERE COGNITIVELY NORMAL.

		Age Statistics				ratio
Set	N	min	max	μ	σ	M:F
Diffusion training	8000	46	82.7	65.3	7.8	1:1.1
Diffusion testing	2000	46	82.7	65.3	7.5	1:1.1
ControlNet training	1000	48	80.3	63.1	7.3	1:1.1

follow-up MRIs. Once trained, the aggregate model (diffusion + ControlNet) generates MRIs of size 128^3 for any user-specified follow-up time point.

Training. The architecture was implemented in Python 3.7.16 and PyTorch 2.5.1 on dual 2.60 GHz Intel Xeon Platinum 8358 CPUs. Model training was accelerated by an NVIDIA A100 80GB GPU with CUDA 12.2. The diffusion model was trained on 80% of 10,000 cognitively normal (CN) participants with T_1-w MRIs (20% testing set). The ControlNet was trained on 1,000 pairs of longitudinal MRIs from CN participants.

Validation. The structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) allows assessment and comparison of the model to other approaches. SSIM, ranging from 0 (0%) to 1 (100%), quantifies similarity between synthetic and genuine MRIs while focusing on anatomic detail relevant to human visual perception rather than on brightness and contrast. Higher SSIM indicates that the synthetic MRI is more anatomically similar to the genuine MRI. PSNR is a voxelwise measure that quantifies MRI synthesis quality; the higher the PSNR, the smaller the error between synthetic and genuine MRIs. PSNR involves the ratio of maximum voxel intensity and mean squared error. It highlights overall MRI quality and fidelity, higher values indicating better MRI quality. Where SSIM emphasizes perceptual quality, PSNR focuses on quantitative errors. The recon-all function in FreeSurfer (FS) was used to segment and extract brain volumes from both genuine and synthetic follow-up MRIS for comparison of the two in 40 participants.

III. Results & discussion

Figure 2 facilitates inspection of model performance across 6 post-baseline timepoints by comparing genuine MRIs to synthetic MRIs and by quantifying the MRI intensity errors between them. Voxelwise errors are relatively lower in the deep white matter compared to the cortex. This is possibly due to the higher relative spatial homogeneity of the former compared to the latter, which has complex gyrification. Average errors are lowest at CA = 78.55 years because this corresponds to FP = 2 years and most longitudinal MRIs in the ControlNet training set have this follow-up interval. The mean follow-up interval ($\mu = 2.25$ years) of our training set limits our model's ability to synthesize MRIs at follow-ups that differ from it considerably. A longitudinal training set with a wider follow-up range can enable future MRI synthesis across other follow-up intervals.

TABLE II
MODEL COMPARISON.

Model	SSIM	PSNR [dB]
DaniNet [21]	0.782	-
CounterSynth [22]	0.871	-
HGAN [23]	0.668	26.09
FGAN [24]	0.640	25.10
RevGan [17]	0.675	24.97
This study	**0.932**	**29.42**

Models are compared in Table II on results from identical samples. Our model achieves leading performance according to both SSIM and PSNR. Mean total brain volume differs by 19.8% between synthetic and genuine follow-up MRIs. The means of cortical gray matter, cerebral white matter, cerebellum, and cerebellar cortex exhibit volume differences of 24%, 26%, 15% and 3%, respectively. Within subcortex, volumes of the right putamen (24%) and left pallidum (23%) differ most between genuine and synthetic MRIs. Volumes of the left hippocampus (2%) and right accumbens area (5%) differ least.

This study synergizes a 3D diffusion U-Net with a ControlNet to synthesize individuals' follow-up MRIs from their baseline MRIs. The original ControlNet [20] uses user input text to condition the synthetic output. Most approaches use conditional embedding through splicing [25], which does not extract baseline MRI features during inference. By contrast, our novel approach uses concatenation of inputs and synergizes the ControlNet (an explicit feature extraction network) with a 3D diffusion U-Net, where the baseline 3D MRI is used as a prior. Thereby, our method captures anatomic features and thereby extends the scope and flexibility of conditional generation.

IV. Conclusion

This work introduces a controllable brain evolution model by integrating a 3D diffusion model with ControlNet. From baseline MRIs, the model predicts follow-up MRIs, capturing age-related neuroanatomic changes while preserving individual-specific features. Although further research is needed to address model limitations, this research demonstrates the feasibility of using generative deep neural networks to predict individuals' future brain aging.

Acknowledgments

A.I. conducted part of this research while on sabbatical leave at King's College London. This research was funded by NIH grants R01 NS 100973, RF1 AG 082201, RF1 AG 054443, R01 AG 079512, R01 AG 079957, US DoD contract W81XWH-18-1-0413, by anonymous donors, by NSF CAREER award CPS/CNS-1453860 and grants CCF-1837131, MCB-1936775, CNS-1932620, by Okawa Foundation grant CMMI-1936624, by DARPA grant N66001-17-1-4044, by a 2021 USC Stevens Center TAG award, by an Intel Faculty Award and by a Northrop Grumman grant. This research has been conducted using the UK Biobank Resource, under application numbers 11559 and 47656.

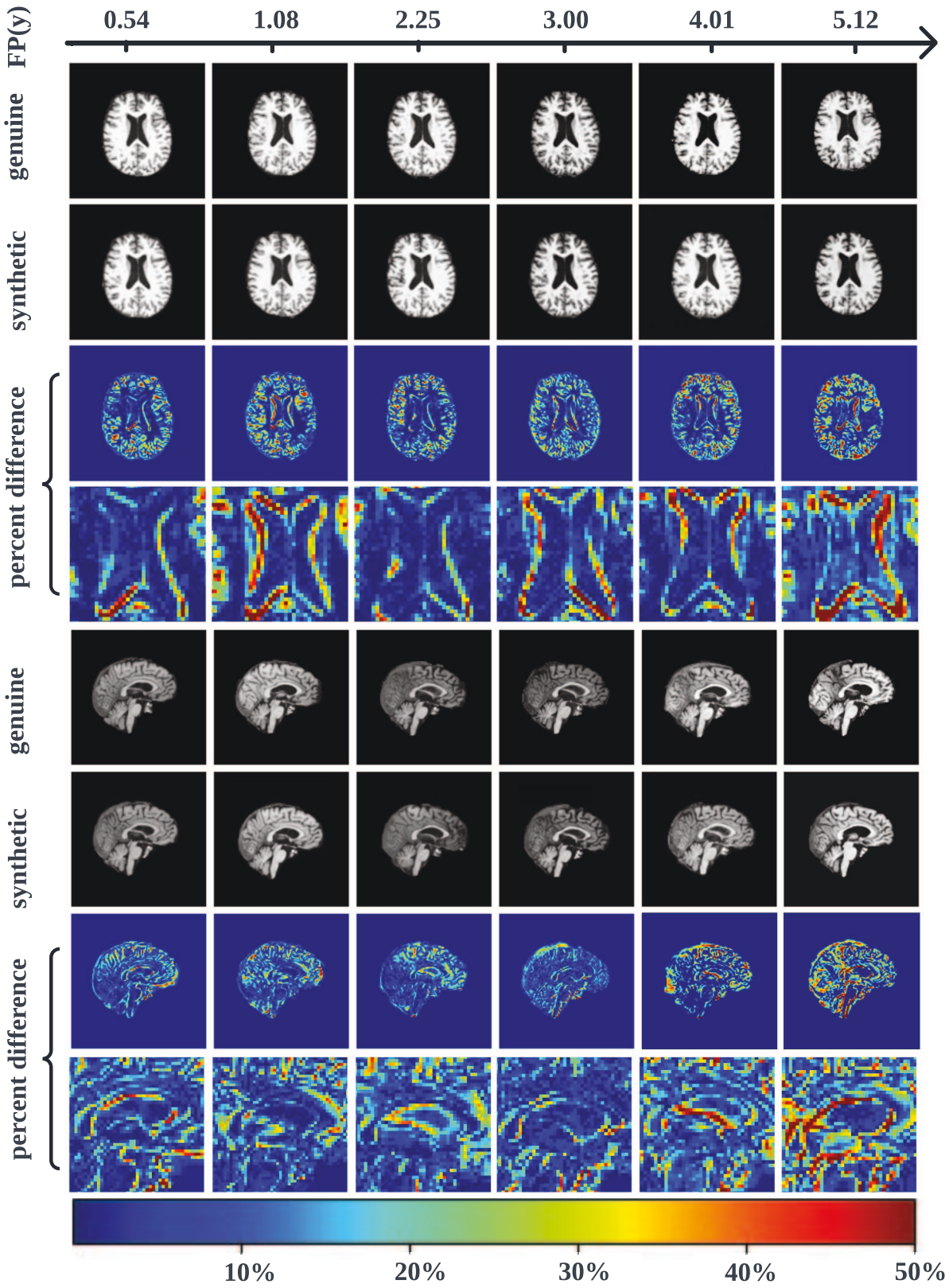

Fig. 2. **Brain evolution.** Comparison of genuine (row 1, row 5) and synthetic (row 2, row 6) for both axial and sagittal MRI slices for a participant aged 76.3 years at baseline. Future brain scans are predicted for six follow-up periods (FP) post-baseline (columns 1 through 6). Percentage differences between synthetic and genuine MRIs are displayed in row 3, 4, 7 and 8 with detailed views of the ventricles in rows 4 and 8.

References

[1] Shofiul Azam, Md Ezazul Haque, Rengasamy Balakrishnan, In-Su Kim, and Dong-Kug Choi, "The ageing brain: molecular and cellular basis of neurodegeneration," *Frontiers in cell and developmental biology*, vol. 9, pp. 683459, 2021.

[2] Dietmar Rudolf Thal, Kelly Del Tredici, and Heiko Braak, "Neurodegeneration in normal brain aging and disease," *Science of aging knowledge environment*, vol. 2004, no. 23, pp. pe26–pe26, 2004.

[3] Tahira Farooqui and Akhlaq A Farooqui, "Aging: an important factor for the pathogenesis of neurodegenerative diseases," *Mechanisms of ageing and development*, vol. 130, no. 4, pp. 203–215, 2009.

[4] Yujun Hou, Xiuli Dan, Mansi Babbar, Yong Wei, Steen G Hasselbalch, Deborah L Croteau, and Vilhelm A Bohr, "Ageing as a risk factor for neurodegenerative disease," *Nature Reviews Neurology*, vol. 15, no. 10, pp. 565–581, 2019.

[5] Susumu Mori, Kengo Onda, Shohei Fujita, Toshiaki Suzuki, Mikimasa Ikeda, Khin Zay Yar Myint, Jun Hikage, Osamu Abe, Hidekazu Tomimoto, Kenichi Oishi, et al., "Brain atrophy in middle age using magnetic resonance imaging scans from japan's health screening programme," *Brain Communications*, vol. 4, no. 4, pp. fcac211, 2022.

[6] CR Jack Jr, MM Shiung, SD Weigand, PC O'brien, JL Gunter, Bradley F Boeve, David S Knopman, GE Smith, RJ Ivnik, Eric George Tangalos, et al., "Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic mci," *Neurology*, vol. 65, no. 8, pp. 1227–1231, 2005.

[7] Lorenzo Pini, Michela Pievani, Martina Bocchetta, Daniele Altomare, Paolo Bosco, Enrica Cavedo, Samantha Galluzzi, Moira Marizzoni, and Giovanni B Frisoni, "Brain atrophy in alzheimer's disease and aging," *Ageing research reviews*, vol. 30, pp. 25–48, 2016.

[8] Allen Chang, Nikhil N. Chaudhari, and Andrei Irimia, "Neuroimaging data augmentation for robust brain age estimation using deep neural networks," in *IEEE Conÿerence on Computational Intelligence in Bioinformatics and Computational Biology*. IEEE, 2024, p. Oral online presentation.

[9] Walter HL Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M Jorge Cardoso, "Brain imaging generation with latent diffusion models," in *MICCAI Workshop on Deep Generative Models*. Springer, 2022, pp. 117–126.

[10] Andrei Irimia, "Cross-sectional volumes and trajectories of the human brain, gray matter, white matter and cerebrospinal fluid in 9473 typically aging adults," *Neuroinformatics*, vol. 19, no. 2, pp. 347–366, 2021.

[11] Nikhil N Chaudhari, Phoebe E Imms, Nahian F Chowdhury, Margaret Gatz, Benjamin C Trumble, Wendy J Mack, E Meng Law, M Linda Sutherland, James D Sutherland, Christopher J Rowan, et al., "Increases in regional brain volume across two native south american male populations," *GeroScience*, pp. 1–21, 2024.

[12] Diletta Milana, "Deep generative models for predicting alzheimer's disease progression from mr data," 2016.

[13] Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker, "Deep structural causal models for tractable counterfactual inference," *Advances in neural information processing systems*, vol. 33, pp. 857–869, 2020.

[14] Tian Xia, Agisilaos Chartsias, Chengjia Wang, Sotirios A Tsaftaris, Alzheimer's Disease Neuroimaging Initiative, et al., "Learning to synthesise the ageing brain without longitudinal data," *Medical Image Analysis*, vol. 73, pp. 102169, 2021.

[15] Shruti P Gadewar, Abhinaav Ramesh, Mengting Liu, Iyad Ba Gari, Talia M Nir, Paul Thompson, and Neda Jahanshad, "Predicting individual brain mris at any age using style encoding generative adversarial networks," in *18th International Symposium on Medical Information Processing and Analysis*. SPIE, 2023, vol. 12567, pp. 471–479.

[16] Jiancong Wang, Yuhua Chen, Yifan Wu, Jianbo Shi, and James Gee, "Enhanced generative adversarial network for 3d brain mri super-resolution," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3627–3636.

[17] Wanyun Lin, Weiming Lin, Gang Chen, Hejun Zhang, Qinquan Gao, Yechong Huang, Tong Tong, Min Du, and Alzheimer's Disease Neuroimaging Initiative, "Bidirectional mapping of brain mri and pet with 3d reversible gan for the diagnosis of alzheimer's disease," *Frontiers in Neuroscience*, vol. 15, pp. 646013, 2021.

[18] Fidel Alfaro-Almagro, Mark Jenkinson, Neal K Bangerter, Jesper LR Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Stamatios N Sotiropoulos, Saad Jbabdi, Moises Hernandez-Fernandez, Emmanuel Vallee, et al., "Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank," *NeuroImage*, vol. 166, pp. 400–424, 2018.

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[20] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3836–3847.

[21] Daniele Ravì, Stefano Blumberg, Silvia Ingala, Frederik Barkhof, Daniel Alexander, and Neil Oxtoby, "Degenerative adversarial neuroimage nets for brain scan simulations: Application in ageing and dementia," *Medical Image Analysis*, vol. 75, pp. 102257, 2021.

[22] Guilherme Pombo, Robert Gray, M Jorge Cardoso, Sebastien Ourselin, Geraint Rees, John Ashburner, and Parashkev Nachev, "Equitable modelling of brain imaging by counterfactual augmentation with morphologically constrained 3d deep generative models," *Medical Image Analysis*, vol. 84, pp. 102723, 2023.

[23] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Yong Xia, and Dinggang Shen, "Spatially-constrained fisher representation for brain disease identification with incomplete multi-modal neuroimages," *IEEE transactions on medical imaging*, vol. 39, no. 9, pp. 2965–2975, 2020.

[24] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, and Yong Xia, "Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019.

[25] Anna Zapaishchykova, Benjamin H Kann, Divyanshu Tak, Zezhong Ye, Daphne A Haas-Kogan, and Hugo JWL Aerts, "Synthbraingrow: Synthetic diffusion brain aging for longitudinal mri data generation in young people," in *MICCAI Workshop on Deep Generative Models*. Springer, 2024, pp. 75–86.