

---

# 3D Causal-StyleGAN3 for Synthesizing MR Images of Alzheimer’s Disease

---

**Marcel Fernandez Rosas**

Hasso Plattner Institute

14482 Potsdam, Germany

[marcel.fernandezrosas@student.hpi.de](mailto:marcel.fernandezrosas@student.hpi.de)

## Abstract

This report presents a novel approach for synthesizing medical images of Alzheimer’s disease using a 3D causal generative model. The model is designed to generate images from underrepresented populations, including younger patients and those with less advanced disease stages, providing valuable insights into the early stages of the disease. The model uses a causal graph to represent the relationships between age, sex, clinical dementia rating, and brain volumes, and can intervene on these variables to generate images with specific features. The model is trained on the ADNI dataset and evaluated using the Fréchet Inception Distance. Results show that the model can generate images with varying interventions.

## 1 Introduction

Medical imaging plays a crucial role in the diagnosis and monitoring of Alzheimer’s disease (AD), a neurodegenerative disorder that affects cognitive functions such as memory and behavior.

**Diagnosing Alzheimer’s disease** Medical imaging techniques are used to detect and monitor the pathological changes in the brain of Alzheimer’s patients. These techniques provide insights into the structural and functional changes that occur in the brain during the progression of Alzheimer’s disease. MRI images are used to detect changes in the size and shape of the brain regions affected by Alzheimer’s disease. Among other things, the hippocampus and ventricles give good indications of disease progression (Blinkouskaya et al., 2021). Clinical Dementia Rating (CDR) is an instrument used to measure the severity of dementia and monitor the progression of AD.

**Motivation** Age is the most significant risk factor for developing AD, and the prevalence of the disease increases with age (Fiest et al., 2016). Also, older patients are more likely to have advanced disease stages, making it easier to detect and measure the changes in the brain associated with AD. However, it is important to note that younger people can also suffer from dementia and AD. About 5% to 6% of Alzheimer’s cases involve early-onset Alzheimer’s disease with an onset age of less than 65 years (Mendez, 2019).

**Contribution** Our contribution is the development of a causal generative model that can synthesize medical images from underrepresented populations, including younger patients of different sex and those with less advanced disease stages. This may provide valuable insights into the early stages of Alzheimer’s disease and improve early diagnosis and personalized treatment approaches for patients at different stages of the disease.

## 2 Related work

Generative algorithms for image generation are already being researched with vigor. We will look at generative adversarial networks (GAN's), peculiarities in the generation of medical images and the causality in GAN's.

**Generative adversarial architectures** Goodfellow et al. (2014) first proposed the idea of a generative and a discriminative model playing an adversarial game, where the generator tries to fool the discriminator, and the discriminator tries to detect the generator's fakes. Karras et al. (2018) introduced the StyleGAN and re-designed the generator architecture so that the image synthesis is more controllable. The problem with this architecture is that the generated images have an artifact that resembles a water bubble. Karras et al. (2019) solved this by changing the normalization in StyleGAN2. Training GAN's with too little data leads to overfitting of the discriminator. Karras et al. (2020) proposed a discriminator augmentation mechanism (ADA) with StyleGAN2-ADA to apply a set of augmentations to all images shown to the discriminator. With StyleGAN3, Karras et al. (2021) presented a solution to the aliasing effect of StyleGAN2 using pointwise nonlinearities. They changed the generator architecture by replacing the learned input constant with Fourier features and achieve translation and rotation equivariance.

**Generative networks for medical imaging** Special research on the generation of medical image data, was conducted by Hong et al. (2021). They extended the StyleGAN2 architecture to allow synthesis of 3D images and evaluated the results using several public datasets of MR T1 images of the brain. Woodland et al. (2022) evaluated the performance of StyleGAN2-ADA on multiple datasets by processing the three-dimensional images per slice in two dimensions.

**Causality** In conventional GAN architectures, there is no way to control the modes of the generated data. Mirza and Osindero (2014) proposed to condition the generator and discriminator on additional information to make it possible to direct the data generation. With conditioning, however, it is not possible to generate data with labels that were not present during training. To draw samples from distributions other than that of the dataset, causality comes into play. The approach of causalGAN was proposed by Kocaoglu et al. (2017). In this approach, a causal controller models the causal graph and generates the conditioning labels for the generator. These causal models allow for sampling from various distributions.

## 3 Methodology

### 3.1 Conditioning and causality

Conditioning involves providing additional information, such as class labels, to the generator network, which enables the generation of images with specific features, such as images of Alzheimer's disease from different age groups or disease stages. However, with conditioning, we can only generate images from the distribution of the original dataset. Causality refers to the relationship between cause and effect. Unlike conditioning, which focuses on correlations between input and output of the generative model, causality considers how changes in one variable cause changes in another. This requires more complex models that can be used to determine causal relationships from observational data.

### 3.2 Causal controller

The causal controller serves as a model to accurately represent the underlying causal mechanisms of AD. To achieve this, a causal graph is constructed that represents the causal relationships between age, sex, CDR, and brain volumes of left lateral ventricle, right lateral ventricle, left cerebral cortex, and right cerebral cortex (Figure 1). The causal graph is modeled using maximum likelihood estimation, which estimates the parameters of the graph based on the observed data. To avoid class imbalance, CDR values are clipped at 1.0, resulting in classes present: 0, 0.5, and 1.0. The probability of sex is measured as the mean, while age is modeled using a beta distribution with  $\epsilon = 0.0001$  and the method of moments. To build the relationship between age, sex, and CDR, a multinomial logistic regression is chosen, with sex and age as input variables, and CDR as the classes. Finally, the volumes

are modeled with Gaussian mixture models, where age, sex, and CDR are the input variables, and the brain volumes are the output variables.

**Domain reasoning** With increasing age, the brain undergoes various structural and functional changes. These changes may be morphological (e.g., reduction of brain volume or enlargement of ventricles) or pathophysiological (e.g., shrinkage of neurons) (Blinkouskaya et al., 2021). Research suggests that sex and gender have an impact on the development of AD. This may be due to gender-specific risk factors such as menopause or testosterone loss, but also to risk factors that occur in both sexes, such as diabetes or depression (Nebel et al., 2018). Age and sex can impact brain volumes through various mechanisms. Among others, changes in hormone levels, menopause, and cardio-metabolic health profile are associated with changes in brain volume (Than et al., 2021, 2022). The CDR value has also been associated with changes in brain volumes. It is assumed that a higher CDR value can be associated with enlarged ventricles and reduced size of the hippocampus (Chou et al., 2009; Wang et al., 2023; Jaroudi et al., 2017).

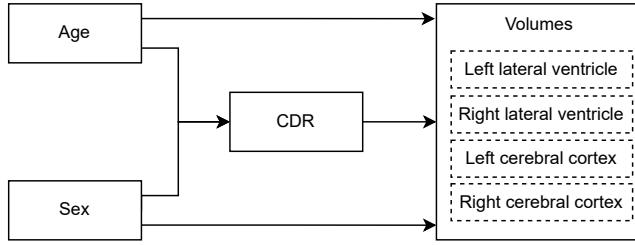


Figure 1: Causal graph whose relationships are learned from observational training data.

### 3.3 ADNI dataset

The Alzheimer’s Disease Neuroimaging Initiative (ADNI)<sup>1</sup> is a large-scale, longitudinal dataset that includes clinical, cognitive, genetic, and neuroimaging data from individuals with normal cognition, mild cognitive impairment, and Alzheimer’s disease. In our study, we use the MPRAGE images because they are considered the best in the quality ratings. We performed further preprocessing steps using the FMRIB Software Library (FSL) (Jenkinson et al., 2012). This included nonlinear image registration of the scans to MNI152 and debiasing to correct for intensity variations across different MRI scanners. We have split the data from 9185 subjects with an 80/20 split, obtaining 7348 subjects for the training set and 1837 subjects for the test set.

**Dataset statistics** Among the 7348 subjects used for training, ~52% are male and ~48% female. For CDR, ~35% are rated 0, ~54% are rated 0.5 and ~11% with 1.0. In the test set, there are ~51% male and ~49% female subjects. Also, ~36% with CDR 0, ~54% with CDR 0.5, and ~10 % with CDR 1.0. The distribution of the age for training and test set can be seen in Figure 2. To reduce the complexity, we limited the number of slices to the maximum range where ventricles are visible, resulting in a total of 123 slices per scan. We used the SynthSeg algorithm (Billot et al., 2021) to segment the ventricles.

## 4 Results

In this section, we present the results of our experiments in which we trained four different 3D models with resolutions of 32x32, 64x64, 128x128, and 256x256. We calculated the FID value from the three slices in the middle of the scan. We observed that with lower resolutions, the learning progress was better measured by the Fréchet Inception Distance (FID) (Figure 3). The 256x256 resolution model failed to make significant learning progress. In order to evaluate the causal mechanism, we intervened on sex, age, and CDR variables. In the following sections, we compare the FID values for different numbers of generated images and present histograms of age values. Additionally, we present samples of generated images for different resolutions and interventions.

<sup>1</sup><https://adni.loni.usc.edu>

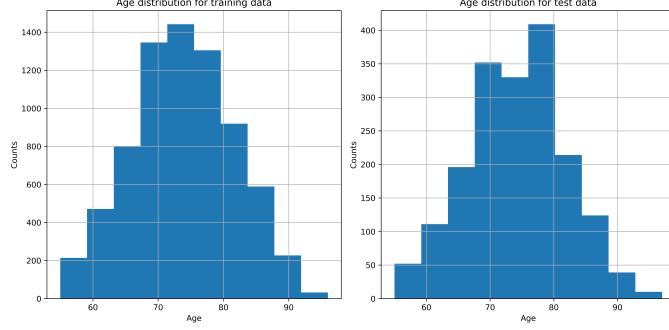


Figure 2: Distribution of age in the training dataset (left) and in the test dataset (right).

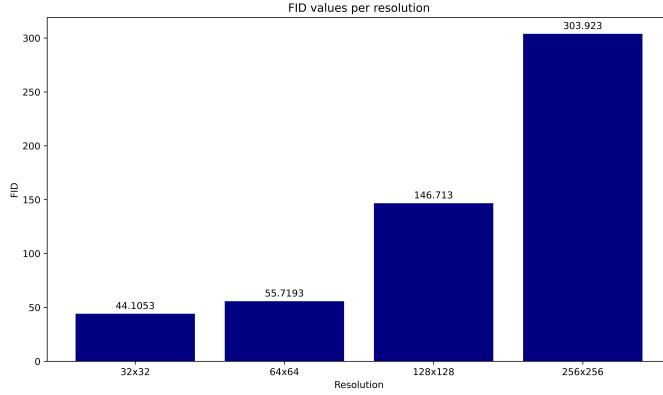


Figure 3: FID values for 10000 generated images per image resolution.

**Intervention at sex node** To begin, we first report the FID scores for male and female after intervening at the sex node of the causal graph (Figure 4). For resolutions 32x32, 64x64 and 256x256, we see a slightly better FID for male than for female. This could be due to the fact that men are slightly overrepresented in the dataset (see 3.3). From Table 1 we can observe, that for resolutions 32x32 and 64x64, increasing the number of generated images improves the FID. This indicates, that the causal mechanism work and more images are generated that compare to the quality of the test set. However, we can also see, that with increasing resolution (and decreasing model performance), this effect is not linear. It is also important to mention, that the high resolution models generate images of very poor quality.

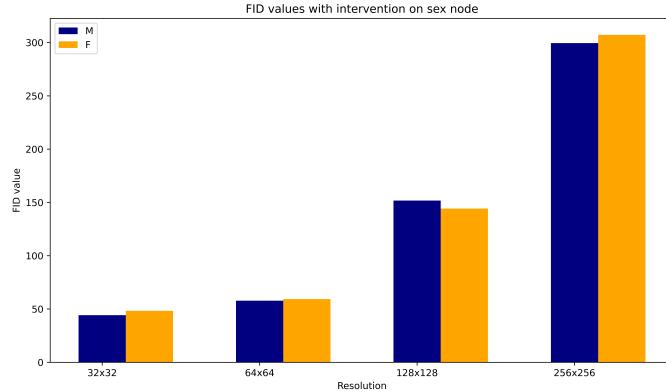


Figure 4: FID values for 10000 generated images per image resolution when intervening at the sex node of the causal graph.

Resolution	Sex	2000	4000	6000	10000
32x32	M	45.642998	44.524759	44.244783	44.155795
32x32	F	49.951874	48.264799	47.964436	48.254386
64x64	M	58.549210	57.826321	57.789532	57.765610
64x64	F	60.626389	60.460030	59.707472	59.292958
128x128	M	151.818835	150.985041	152.760994	151.712638
128x128	F	144.570702	144.461856	145.832073	144.207695
256x256	M	301.419985	300.193172	300.223747	299.505281
256x256	F	308.001684	306.636934	306.711354	307.238373

Table 1: FID values for 2000, 4000, 6000, and 10000 generated images for different image resolution models with intervention on the sex node of the causal graph.

**Intervention at age node** Figure 5 shows that the FID scores for the 32x32 images reveal a value of less than 50 for age groups up to 90 years. Notably, the FID score is even better for the 55-65 age group, with a value below 45. However, the FID for age groups beyond 90 years is considerably large, a likely consequence of the scarcity of training examples in this category. Regarding the 64x64 images, our analysis shows that the optimal FID scores are evident for the age groups between 69 and 86. Additionally, this age range comprises the highest number of subjects in the dataset, as illustrated in Figure 2. With respect to the 128x128 and 256x256 images, a clear pattern emerges, indicating that the FID values improve as age increases.

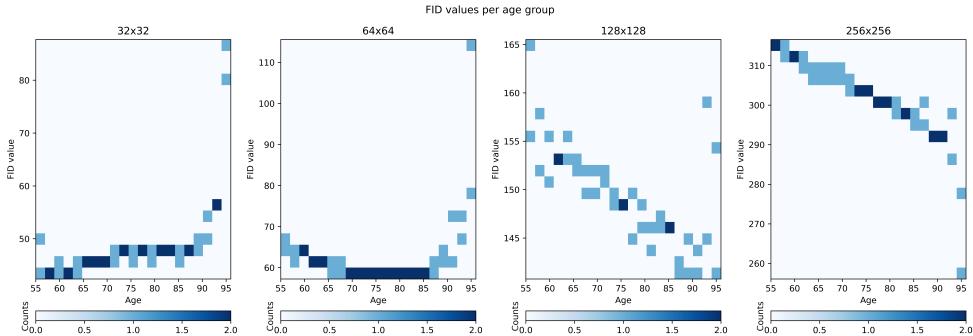


Figure 5: 2D histogram of FID values for 10000 generated images per image resolution when intervening at the age node of the causal graph.

**Intervention at CDR node** We perform an intervention at the inner node while leaving the child nodes unaffected. The relationship between CDR value and FID score is illustrated in Figure 6. Specifically, lower CDR values correspond to lower FID scores for 32x32 resolution images, while higher CDR values correspond to lower FID scores for the highest resolution (256x256) images. From Table 2 we can see no discernible pattern in terms of the FID score for different amounts of generated images at different interventions of CDR values. Nevertheless, we may infer that a greater number of generated images contributes to a more resilient FID value.

**Generated images** We present generated images with conditional and causal labels compared to the original image of one subject in Figure 7. We further generated images of label combinations not present in the data distribution in Figure 8.

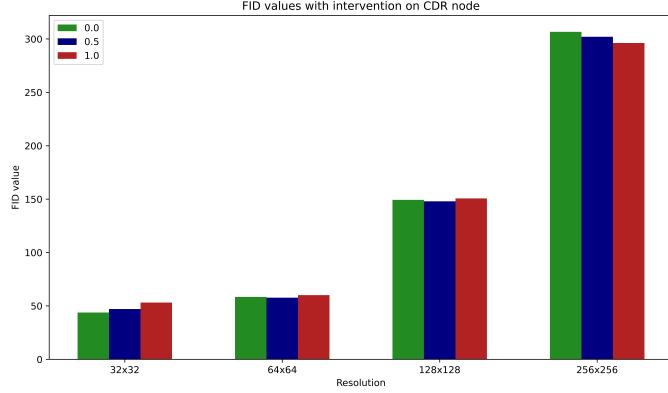


Figure 6: FID values for 10000 generated images per image resolution when intervening at the CDR node of the causal graph.

Resolution	CDR	2000	4000	6000	10000
32x32	0.0	43.877868	44.361938	44.005843	43.800415
32x32	0.5	48.061737	47.013607	46.770705	47.063430
32x32	1.0	53.597903	52.916813	53.318257	53.156807
64x64	0.0	59.584875	58.803185	58.864842	58.373499
64x64	0.5	59.236368	58.274918	58.021347	57.668443
64x64	1.0	61.642244	60.648928	60.202661	60.045319
128x128	0.0	151.781627	148.441448	147.859354	149.307447
128x128	0.5	149.284727	149.271740	148.871717	147.914494
128x128	1.0	151.171371	152.746966	150.049447	150.635541
256x256	0.0	307.269410	307.057655	306.836652	306.676725
256x256	0.5	303.022868	301.886058	302.133295	302.063998
256x256	1.0	295.776630	296.692138	296.652086	296.216598

Table 2: FID values for 2000, 4000, 6000, and 10000 generated images for different image resolution models with intervention on the CDR node of the causal graph.

## 5 Conclusion

In conclusion, we have presented a novel approach for synthesizing medical images of Alzheimer’s disease using a causal generative model. Our model allows for the generation of images from underrepresented populations, including younger patients and those with less advanced disease stages, providing valuable insights into the early stages of the disease. By incorporating causality into the generative model, we can control the modes of the generated data and sample from various distributions. Our approach has the potential to improve early diagnosis and personalized treatment approaches for patients at different stages of the disease.

## 6 Future work

Future work includes enhancing the quality of the high resolution model, evaluating the performance of our model on larger datasets, including the hippocampus in the causal controller and exploring the use of our approach for other medical imaging applications.

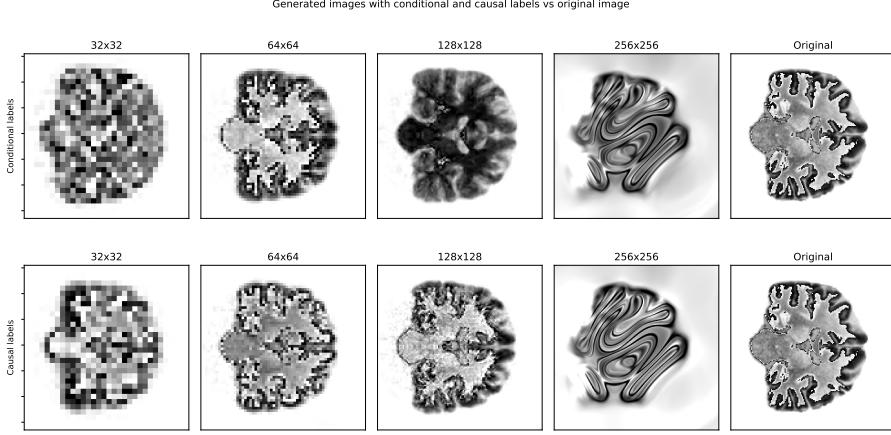


Figure 7: Visualization of the middle slice of generated images in different resolution. The conditional labels are chosen from one subject of the test set with age 75, sex female, CDR 0.5 and normalized ventricles -0.2504, -0.7107, 0.2859, 0.8365. The causal labels were generated by intervening on the age and sex node, with age 75, sex female. The CDR was estimated 0.5 and the normalized ventricles 0.5246, 0.2703, -0.7454, -0.9141

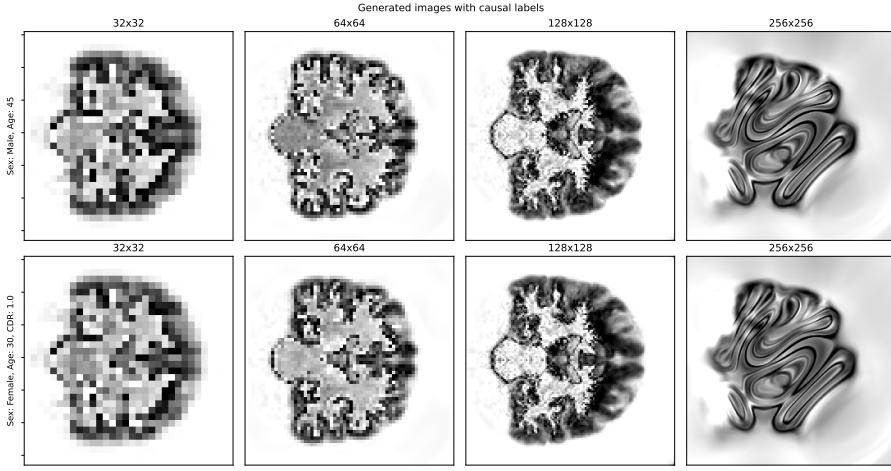


Figure 8: Visualization of the middle slice of generated images in different resolution. Column 1: Intervention at age and sex node, with age 45 and sex male. Generated causal labels are 0.5 CDR and normalized ventricles -0.7692, -1.2912, -0.0200, -0.1131. Column 2: Intervention at age, sex and CDR node with age 30, sex female and CDR 1.0. Generated causal labels for ventricles are -1.2397, -1.5880, -0.3577, -0.1819

## References

- Benjamin Billot, Douglas N. Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V. Dalca, and Juan Eugenio Iglesias. Synthseg: Domain Randomisation for Segmentation of Brain MRI Scans of any Contrast and Resolution. *arXiv:2107.09559 [cs]*, 2021.
- Yana Blinkouskaya, Andreia Caçoilo, Trisha Gollamudi, Shima Jalalian, and Johannes Weickenmeier. Brain aging mechanisms with mechanical manifestations. *Mechanisms of Ageing and Development*, 200:111575, 2021. ISSN 0047-6374. doi: <https://doi.org/10.1016/j.mad.2021.111575>. URL <https://www.sciencedirect.com/science/article/pii/S0047637421001470>.
- Yi-Yu Chou, Natasha Leporé, Christina Avedissian, Sarah K Madsen, Neelroop Parikshak, Xue Hua, Leslie M Shaw, John Q Trojanowski, Michael W Weiner, Arthur W Toga, Paul M Thompson, and

Alzheimer's Disease Neuroimaging Initiative. Mapping correlations between ventricular expansion and CSF amyloid and tau biomarkers in 240 subjects with alzheimer's disease, mild cognitive impairment and elderly controls. *Neuroimage*, 46(2):394–410, June 2009.

Kirsten M. Fiest, Nathalie Jetté, Jodie I. Roberts, Colleen J. Maxwell, Eric E. Smith, Sandra E. Black, Laura Blaikie, Adrienne Cohen, Lundy Day, Jayna Holroyd-Leduc, and et al. The prevalence and incidence of dementia: a systematic review and meta-analysis. *Canadian Journal of Neurological Sciences*, 43(S1):S3–S50, 2016. doi: 10.1017/cjn.2016.18.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. URL <https://arxiv.org/abs/1406.2661>.

Sungmin Hong, Razvan Marinescu, Adrian V. Dalca, Anna K. Bonkhoff, Martin Bretzner, Natalia S. Rost, and Polina Golland. 3d-stylegan: A style-based generative adversarial network for generative modeling of three-dimensional medical images, 2021. URL <https://arxiv.org/abs/2107.09700>.

Wafa Jaroudi, Julia Garami, Sandra Garrido, Michael Hornberger, Szabolcs Keri, and Ahmed A Moustafa. Factors underlying cognitive decline in old age and alzheimer's disease: the role of the hippocampus. *Rev. Neurosci.*, 28(7):705–714, October 2017.

Mark Jenkinson, Christian F Beckmann, Timothy E J Behrens, Mark W Woolrich, and Stephen M Smith. FSL. *Neuroimage*, 62(2):782–790, August 2012.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2018. URL <https://arxiv.org/abs/1812.04948>.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan, 2019. URL <https://arxiv.org/abs/1912.04958>.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data, 2020. URL <https://arxiv.org/abs/2006.06676>.

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks, 2021. URL <https://arxiv.org/abs/2106.12423>.

Murat Kocaoglu, Christopher Snyder, Alexandros G. Dimakis, and Sriram Vishwanath. Causalgan: Learning causal implicit generative models with adversarial training, 2017. URL <https://arxiv.org/abs/1709.02023>.

Mario F Mendez. Early-onset alzheimer disease and its variants. *Continuum (Minneapolis, Minn.)*, 25(1):34–51, February 2019.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014. URL <https://arxiv.org/abs/1411.1784>.

Rebecca A. Nebel, Neelum T. Aggarwal, Lisa L. Barnes, Aimee Gallagher, Jill M. Goldstein, Kejal Kantarci, Monica P. Mallampalli, Elizabeth C. Mormino, Laura Scott, Wai Haung Yu, Pauline M. Maki, and Michelle M. Mielke. Understanding the impact of sex and gender in alzheimer's disease: A call to action. *Alzheimer's Dementia*, 14(9):1171–1183, 2018. ISSN 1552-5260. doi: <https://doi.org/10.1016/j.jalz.2018.04.008>. URL <https://www.sciencedirect.com/science/article/pii/S1552526018301304>.

Stephanie Than, Chris Moran, Richard Beare, Amanda J Vincent, Taya A Collyer, Wei Wang, Michele L Callisaya, Russell Thomson, Thanh G Phan, Alex Fornito, and Velandai K Srikanth. Interactions between age, sex, menopause, and brain structure at midlife: A UK biobank study. *J. Clin. Endocrinol. Metab.*, 106(2):410–420, January 2021.

Stephanie Than, Chris Moran, Taya A Collyer, Richard J Beare, Emma M Lane, Amanda J Vincent, Wei Wang, Michele L Callisaya, Russell Thomson, Thanh G Phan, Alex Fornito, and Velandai K Srikanth. Associations of sex, age, and cardiometabolic risk profiles with brain structure and cognition: A UK biobank latent class analysis. *Neurology*, 99(17):e1853–e1865, October 2022.

Catherine Wang, Sasha Kravets, Abhishek Sethi, Mark A Espeland, Louis R Pasquale, Stephen R Rapp, Barbara E Klein, Stacy M Meuer, Mary N Haan, Pauline M Maki, Joelle A Hallak, and Thasarat Sutabutr Vajaranant. An association between large optic cupping and total and regional brain volume: The women's health initiative. *Am. J. Ophthalmol.*, 249:21–28, January 2023.

McKell Woodland, John Wood, Brian M. Anderson, Suprateek Kundu, Ethan Lin, Eugene Koay, Bruno Odisio, Caroline Chung, Hyunseon Christine Kang, Aradhana M. Venkatesan, Sireesha Yedururi, Brian De, Yuan-Mao Lin, Ankit B. Patel, and Kristy K. Brock. Evaluating the performance of StyleGAN2-ADA on medical images. In *Simulation and Synthesis in Medical Imaging*, pages 142–153. Springer International Publishing, 2022. doi: 10.1007/978-3-031-16980-9\_14. URL [https://doi.org/10.1007%2F978-3-031-16980-9\\_14](https://doi.org/10.1007%2F978-3-031-16980-9_14).