

DGM: Disentangled Generative Model for Detecting AD Individualized Pathological Changes via Pseudo-Healthy Synthesis

Zhuangzhuang Li¹, Kun Zhao^{1,*}, Dong Wang¹, and Yong Liu¹

¹School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China
kunzhao@bupt.edu.cn

Abstract. Alzheimer’s disease (AD) is a complicated, heterogeneous neurodegenerative disease associated with cognitive decline, behavioral impairment, and brain atrophy. Detecting individualized pathological changes from cognitive normal (CN) to AD is critical for targeted treatment. Current existing methods face challenges, including biases toward specific pathology profiles. To this end, we proposed a disentangled generative model (DGM) to generate pseudo-healthy images and disease-related residual maps that accurately detect universal pathological changes. The framework of DGM consists of three modules: pseudo-healthy MRI synthesis, residual map synthesis, and input reconstruction modules. We take into account both the healthiness and subject identity to validate the biological validity of synthetic pseudo-healthy images. Our experiments demonstrated the effectiveness of the DGM in reconstructing healthy brain anatomy, preserving subject identity, and highlighting its direct application in anomaly pathological detection across the transitions from CN to MCI and from CN to AD. Code is available at https://github.com/zhonghuaajiuzhou12138/DDGM_disease_stage_model.

Keywords: Alzheimer’s disease · Pathological changes · Disentangled representation learning · Healthiness · Subject identity.

1 Introduction

Alzheimer’s disease (AD) is a progressive neurodegenerative disease, characterized by memory impairment, executive dysfunction, and behavior changes [1, 2]. Magnetic resonance imaging (MRI)-based biomarkers converging on structural changes stand out as predominant hallmarks for AD [3–5]. Modeling AD-related structural changes enables quantification of disease severity, thereby guiding clinical management strategies such as treatment planning. However, high heterogeneity in the underlying neuropathological changes poses challenges to understanding the underlying mechanisms. [6, 7]. Therefore, it is essential to detect the individual pathological changes for advancing precision medicine in AD.

Generative artificial intelligence (AI) has brought a novel perspective to anomaly pathological detection by adeptly capturing the subtle changes of what

is considered ‘normal’ in medical images [8–10]. The essence of generative AI lies in normative representation learning, which uncovers characteristics of a healthy population through training on large-scale neuroimaging datasets. Anomaly pathological detection assesses how much patients deviate from the healthy distribution [11–13]. The pseudo-healthy synthesis method synthesizes a pathology-free image from a pathological image by image-to-image translation technique [14–17]. It has been widely used in medical image segmentation [18], detection [19], and computer-aided diagnosis [20]. Previous studies have verified the healthiness of synthetic images in the pseudo-healthy image synthesis task [14]. Nonetheless, pseudo-healthy synthesis is an ill-posed inverse problem as a pathological image can synthesize many healthy-looking images. The medical images were very similar between the different subjects, so the minor generation deviation can seriously affect the histopathological features of the images and undermine their biological authenticity [19]. The healthiness and subject identity are essential and indispensable for the pseudo-healthy synthesis method. However, the above methods face challenges in ensuring both of these aspects. Disentangled representation learning methods can preserve the characteristic of subject identity while generalizing to data from different acquisitions and cohorts [21,22].

To address issues, we proposed a disentangled generative model (DGM) to capture the AD individualized pathological changes. The proposed model can generate pseudo-healthy images and disease residual/saliency maps. The disease residual/saliency maps can also assist radiologists in interpreting the MRI images and predicting the severity of the disease. We validated the biological validity of synthetic pseudo-healthy images generated by DGM. Finally, we applied the proposed DGM approach to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset to capture the pathological changes from cognitive normal (CN) to AD and from CN to MCI.

2 Methodology

2.1 Problem formulation

Let $X \in P$ and $Y \in H$ denote the pathological domain and healthy domain, respectively. In the traditional image-to-image translation task, the aim is to learn a mapping $P \rightarrow H$. Here, we are interested in generating pseudo-healthy images from diseased images as inputs. In this paper, we assume that the MRI Y of the disease state plus the residual map X_a represents the pseudo-healthy MRI X of the health state. Based on the above assumptions, this study aims to generate pseudo-healthy images and residual maps from disease images to obtain disease-related abnormalities and quantify the severity of AD. The underlying generative factors include disease-related factors and those shared by both the disease and CN states.

2.2 Disentangled Generative Model(DGM)

As shown in Fig. 1, the proposed DGM consists of 1) pseudo-healthy MRI synthesis module, 2) residual map synthesis module for detecting pathological changes, and 3) input reconstruction module.

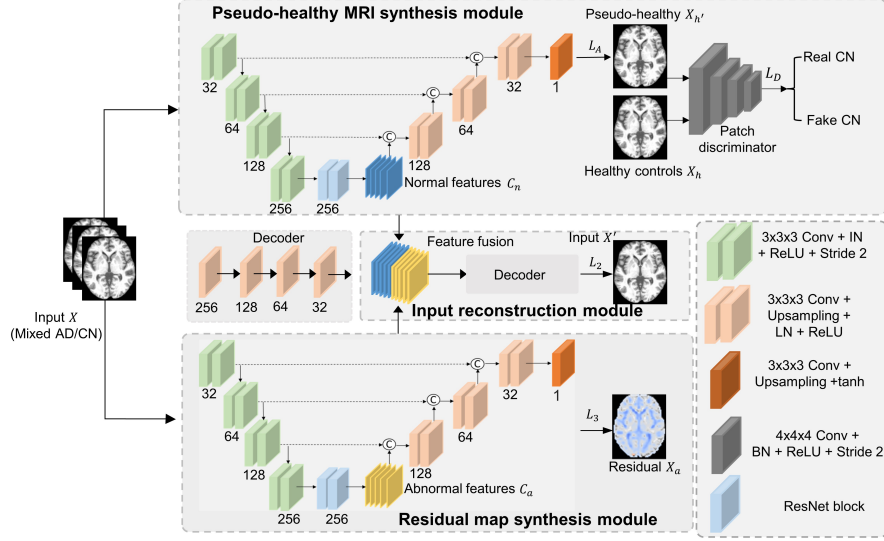


Fig. 1. Schematic of the disentangled generative model (DGM). The pseudo-healthy MRI synthesis module consists of a generator and a patch discriminator. This module can generate pseudo-healthy images by an adversarial training strategy. The input reconstruction module is a decoder to reconstruct the input images and facilitate the training process. The residual map synthesis module is a generator for synthetic residual maps.

Pseudo-healthy MRI Synthesis. The pseudo-healthy MRI synthesis consists of a generator N and a patch discriminator D . The generator can generate pseudo-healthy images by an adversarial training strategy. The generator employs an encoder-decoder mechanism. The encoder compresses an input image into a normal latent representation C_n . Then, a decoder restores a pseudo-healthy image. The network architecture of the generator comprises four down-sampling convolution blocks (DCBs), two residual network blocks, and four up-sampling convolution blocks (UCBs). The DCBs have two convolution layers and each convolution layer is followed by instance normalization. The UCBs contain two transposed convolution layers and each transposed convolution layer is followed by Lay Normalization (LN). To preserve the details of synthetic images, the encoder-decoder network has long skip connections between the DCBs and UCBs. The patch discriminator D was used to distinguish the real normal MRI

images from the pseudo-healthy ones. It can guarantee the orthogonality between normal features C_n and abnormal features C_a . The discriminator network follows the PatchGAN structure to preserve the high-frequency information of input images. The architecture of the discriminator has three DCBs and two convolution layers.

Residual Map Synthesis. Residual map synthesis module R captures a disease-related latent representation C_a , and then produces voxel-level residual maps. The network architecture of residual map synthesis follows the U-net structure, which is similar to the generator of the pseudo-healthy MRI synthesis module.

Input Reconstruction. The input reconstruction module E is used to reconstruct the input images and facilitate the training process. We integrate disease-related features and features shared by AD and CN to reconstruct the input MRI.

2.3 Loss Function

The DGM was trained using a multi-component loss function, including adversarial loss, reconstruction loss, penalty loss, residual loss and discriminator loss.

Adversarial Loss. Adversarial loss guarantees the healthiness of pseudo-healthy images. The adversarial loss function is defined as follows:

$$L_A = E_{X \sim P_X}(-\log(D_{\text{real / fake}}(N(X)))) \quad (1)$$

Penalty Loss. Since the input of the DGM is a mixture of AD and CN, we propose a penalty loss that forces the residual map X_a to be zero when the input is from CN. Penalty loss L_1 is defined as:

$$L_1 = \begin{cases} 0 & X \in AD \\ \|R(X)\|_1 & X \in CN \end{cases} \quad (2)$$

Reconstruction Loss. The reconstruction loss is defined as L1 distance to capture the high-frequency information from input images. The reconstruction loss is defined as follows:

$$L_2 = \|E(C_n, C_a) - X\|_1 \quad (3)$$

Residual loss. Building upon the assumption presented in this study, we define a loss function to supervise the residual map, which can accurately capture brain structural changes in patients with AD.

$$L_3 = \|N(X) - X - R(X)\|_1 \quad (4)$$

where $N(X)$ is the pseudo-healthy image generated by the generator N within the pseudo-healthy MRI synthesis module. $R(X)$ is the residual map by residual map synthesis module R .

Hybrid Loss Function. The final objective function of the DGM is defined by

$$L(N, R, E, D_J) = \lambda_1 (L_1 + L_2 + L_3) + \lambda_2 L_A \quad (5)$$

Table 1. The quantitative results of the healthiness.

Method	Convert	SSIM(CN)%	PSNR(CN)	NCC
DCAE	CN to MCI	47.97 ± 1.10	23.95 ± 0.25	0.02 ± 0.13
	CN to AD	48.50 ± 1.06	18.91 ± 0.23	0.01 ± 0.12
CycleGAN	CN to MCI	85.17 ± 3.81	23.95 ± 0.25	0.19 ± 0.07
	CN to AD	89.07 ± 2.13	18.91 ± 0.23	0.27 ± 0.07
VA-GAN	CN to MCI	39.41 ± 0.88	24.25 ± 0.15	0.10 ± 0.05
	CN to AD	38.23 ± 1.35	19.17 ± 0.11	0.16 ± 0.06
HealthyGAN	CN to MCI	86.74 ± 1.46	24.55 ± 0.74	0.18 ± 0.05
	CN to AD	87.93 ± 1.88	28.80 ± 0.67	0.25 ± 0.07
cDPM	CN to MCI	88.45 ± 1.69	24.68 ± 0.63	0.21 ± 0.11
	CN to AD	84.58 ± 0.15	24.48 ± 0.75	0.16 ± 0.09
DGM without DRL	CN to MCI	87.45 ± 1.53	24.08 ± 0.46	0.17 ± 0.06
	CN to AD	87.54 ± 1.06	27.82 ± 0.55	0.23 ± 0.04
DGM with DRL (ours)	CN to MCI	88.85 ± 0.93	33.76 ± 0.44	0.21 ± 0.07
	CN to AD	88.42 ± 1.05	28.21 ± 0.52	0.28 ± 0.08

Discriminator Loss. For patch discriminator D optimization, the cross-entropy loss is adopted as the discriminator loss:

$$\begin{aligned} \mathcal{L}_D = & \mathbb{E}_{X \in P_X} [\log (1 - D_{\text{real/fake}}(N(X)))] \\ & + \mathbb{E}_{X_h \in H} [\log (D_{\text{real/fake}}(X_h))] \end{aligned} \quad (6)$$

3 Experiments and Results

3.1 Datasets and Data Preprocessing

Datasets. We used 1436 T1-weighted structural MRI data of 1265 subjects from three distinct datasets: European DTI Study on Dementia (EDSD), in-house dataset, and ADNI dataset. The EDSD dataset and in-housed dataset are used to train the DGM, including 603 AD patients and 573 CN. Longitudinal data (CN converted to MCI or AD) from the ADNI dataset was employed to evaluate the biological validity of synthetic pseudo-healthy images.

Data preprocessing. The T1-weighted structural MRI data was using the CAT12 toolbox with the following pipelines: (1) denoising, (2) Interpolation, (3) affine preprocessing, (4) skull-stripping, (5) spatial normalization to Montreal Neurological Institute. The normalization volumetric images were resliced into $96 \times 96 \times 96$. We clipped the intensities of resliced images to $[0, V_{0.995}]$ and then rescaled the intensities to the range $[-1, 1]$.

Implementation Details. The proposed method is implemented in Tensorflow. The generator and discriminator were alternately trained until the loss

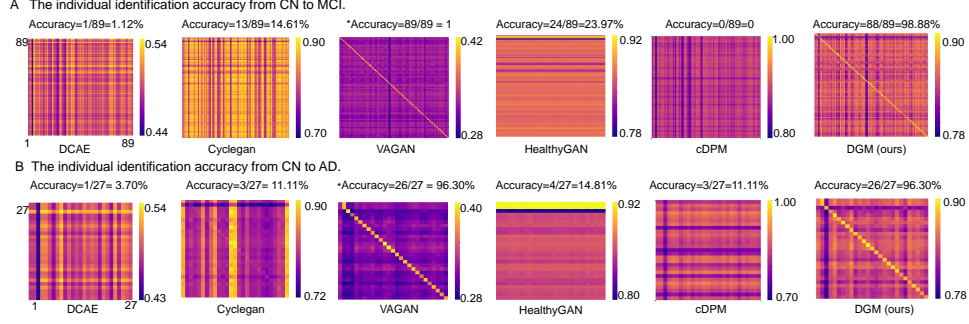


Fig. 2. Quantitative evaluation of individual identification accuracy compared with existing methods. *Accuracy denotes unreliable results when SSIM values fall below 0.5. The larger the diagonal value in the SSIM matrices, the more accurate the individual identification.

functions converged. During the training phase, we update the generator 100 iterations when the discriminator is updated once. The DGM is trained using 50 epochs with regularization weights $\lambda_1 = 5$, $\lambda_2 = 1$. We used the Adam optimizer with the learning rates of the generator 1×10^{-4} and discriminator 3×10^{-5} , respectively. We decay the learning rate by 0.5 for every 25 epochs.

Evaluation Metrics. In this paper, we adopted a comprehensive measurement of the biological validity of the generated pseudo-healthy images from four different perspectives of indicators. We adopted the Peak Signal-to-Noise Ratio (PSNR) to measure the quality of the generated images. Structure Similarity Index Measure (SSIM) is used to measure the image similarity between the generated pseudo-healthy images and corresponding longitudinal images. We used the NCC score to measure the image similarity between residual maps and longitudinal changes from CN to MCI or AD. The individual identification accuracy was employed to evaluate the consistency of subject identity. We calculated the inter-subject and intra-subject SSIM values between pseudo-healthy and longitudinal healthy images across the testing data. The SSIM value between the subjects and themselves is greater than the SSIM value between the subjects, the individual identification can be considered accurate.

3.2 Quantitative Results of Synthetic Images

Little work has been done to decompose the disease image into a pseudo-healthy image and a residual map using a disentangled representation learning algorithm. Some normative modeling and image-to-image translation methods can generate pseudo-healthy images. We compared the proposed DGM method with deep convolutional autoencoder (DCAE) [23], Cycle-GAN [24], visual attribution GAN (VA-GAN) [25], healthy-GAN [16], and conditional diffusion probabilistic model (cDPM) [26]. The residual map can be obtained by the difference between the input image and the pseudo-healthy image. VA-GAN can directly generate

the abnormal residual maps and implicitly get the pseudo-healthy images by summing the input abnormal MRI and the generated residual maps.

The Validation of the Healthiness. We calculated the PSNR and SSIM between pseudo-healthy images and longitudinal healthy images when patients' images are fed into DGM. As shown in **Table 1**, the proposed DGM outperformed the other comparative methods in assessing the quality and health characteristics of synthetic pseudo-healthy images. The PSNRs of reconstruction in both the conversion from CN to MCI and from CN to AD are $33.76 \pm 0.044\%$ and $28.21 \pm 0.52\%$, respectively. The mean SSIM values of DGM between pseudo-healthy images and longitudinal healthy images were $88.85 \pm 0.93\%$ and $88.42 \pm 1.05\%$ from CN to MCI and CN to AD. The CycleGAN and HealthyGAN methods achieved good performances in measuring the healthiness of synthetic pseudo-healthy images. The VA-GAN and DCAE suffer from a degree of spatial warping and blurring.

The Validation of Subject Identity. Medical images from different subjects often exhibit a high degree of similarity, and even minor generative deviations can have a substantial impact on the biological validity of the generated images. Consequently, it is important to preserve the subject identity for the pseudo-healthy synthesis. We used the individual identification accuracy to evaluate the subject identity quantitatively. As shown in **Fig. 2**, the SSIM matrices reflect the similarity between each subject's pseudo-healthy image and all subjects' actual healthy images. It is observed that VA-GAN and DGM achieved good individual identification accuracy compared with other comparison methods. However, the result of VA-GAN is unreliable between the pseudo-healthy images and the actual healthy images when SSIM values fall below 0.5. DDPMs suffer from identity inconsistency due to stochasticity, making them less suitable for capturing subject-specific AD pathology. Other contrast methods did not effectively synthesize realistic pseudo-healthy images with subject identity due to poor individual identification accuracy.

The similarity between longitudinal changes and residual maps. We used the NCC scores to quantify the similarity of generated residual maps and longitudinal change maps. The mean NCC scores from CN to MCI and CN to AD are 0.21 ± 0.07 and 0.28 ± 0.08 , respectively (**Table 1**). It is well established that diffuse atrophy occurs throughout the brain, accompanied by ventricular enlargement with the progression of AD. Normal brain tissue typically exhibits relatively high signal intensity in T1-weighted images. For AD patients, the signal intensity decreases due to cortical atrophy and white matter degeneration. As brain tissue atrophies, the contrast between the ventricles and surrounding brain tissue becomes more pronounced in T1-weighted images. As shown in **Fig. 3**, our proposed method has good agreement with the actual longitudinal progression change. The result demonstrated that the DGM method can accurately capture individual pathological changes with the disease progression compared with other competing methods.

Feature visualization in the latent space. The DGM can disentangle the patient images into normal and abnormal features. For feature visualization, we

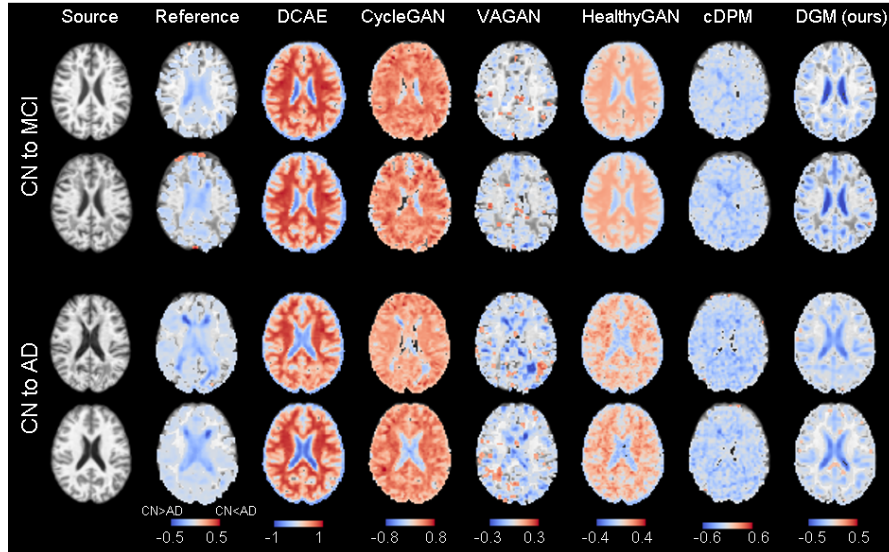


Fig. 3. The examples of individual changes from CN to MCI and CN to AD in different methods.

utilize the t-SNE algorithm to reduce the dimensionality of these features and map them into 2D space. As shown in **Fig. 4**, the proposed DGM can clearly distinguish normal features from abnormal features in both the conversion from CN to MCI and from CN to AD.

Ablation study. To investigate the effect of disentangled representation learning, we compared the quality of synthetic images and longitudinal data with and without DRL methods. As the model without DRL contains only the pseudo-healthy synthesis module in **Fig. 1**, we produce the residual maps through the difference between the input AD images and the pseudo-healthy images. **Table 1** shows that the performance of the disentangled representation learning model outperformed that of the model without DRL. This highlights that the DRL model is robust for different populations, acquisitions, and cohorts.

4 Conclusion

In this work, we proposed a disentangled generative framework designed to disentangle the AD-specific features from the shared features by both AD and CN. The proposed model can generate pseudo-healthy images and disease residual/saliency maps. The residual maps generated by AD-specific features can help neurologists understand pathological changes with disease progression. This, in turn, could be instrumental in generating novel hypotheses and advancing medical research, ultimately contributing to the development of improved patient care strategies.

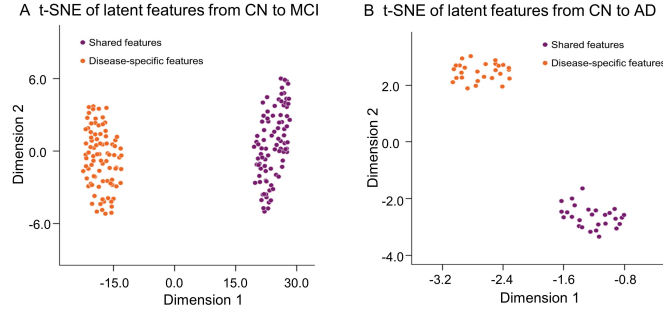


Fig. 4. T-SNE results of latent features. The normal and abnormal features from the pseudo-healthy synthesis and residual map modules were visualized using the t-SNE algorithm.

Acknowledgments

This work was supported by the Beijing Municipal Natural Science Foundation (No. 7244519), the National Natural Science Foundation of China (No. T2425027, No.62333002, No. 82401858) and BUPT Excellent Ph.D. Students Foundation (No. CX2023117).

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. R. U. Haque and A. I. Levey, "Alzheimer's disease: A clinical perspective and future nonhuman primate research opportunities," *Proceedings of the National Academy of Sciences* **116**(52), 26224-26229 (2019).
2. Zhao K., Wang D.W., Wang D. et al: Macroscale connectome topographical structure reveals the biomechanisms of brain dysfunction in Alzheimer's disease. *Science Advances* **10**(41), (2024).
3. O. L. Lopez et al., "Amyloid deposition and brain structure as long-term predictors of MCI, dementia, and mortality," *Neurology* **90**(21), 1920-1928 (2018).
4. K. Zhao et al., "Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer's disease: diagnosis, longitudinal progress and biological basis," *Sci Bull* **65**(13), 1103-1113 (2020).
5. K. Zhao et al., "A neuroimaging biomarker for Individual Brain-Related Abnormalities In Neurodegeneration (IBRAIN): a cross-sectional study," (in English), *Eclinicalmedicine* **67**, 102276 (2023).
6. M. Habes, M. J. Grothe, B. Tunc, C. McMillan, D. A. Wolk, and C. Davatzikos, "Disentangling Heterogeneity in Alzheimer's Disease and Related Dementias Using Data-Driven Methods," *Biol Psychiatry*, **88**(1), 70-82 (2020).

7. Verdi, S., Marquand, A.F., Schott, J.M., Cole, J.H.: Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. *Brain* **144**(10), 2946–2953 (2021)
8. Yi, F., Zhang, Y., Yuan, J., Liu, Z., Zhai, F., Hao, A., Wu, F., Somekh, J., Peleg, M., Zhu, Y.C.: Identifying underlying patterns in alzheimer’s disease trajectory: a deep learning approach and mendelian randomization analysis. *Eclinicalmedicine* **64** (2023)
9. Maheux, E., Koval, I., Ortholand, J., Birkenbihl, C., Archetti, D., Bouteloup, V., Epelbaum, S., Dufouil, C., Hofmann-Apitius, M., Durrleman, S.: Forecasting individual progression trajectories in alzheimer’s disease. *Nature Communications* **14**(1), 761 (2023)
10. Verdi, S., Kia, S.M., Schott, J.M., Marquand, A.F., Cole, J.H., Initiative, A.D.N.: Mapping individualised patterns of atrophy in alzheimer’s disease using neuroanatomical normative models. *Alzheimer’s Dementia* **18**, e060306 (2022)
11. Rutherford, S., Kia, S.M., Wolfers, T., Frazza, C., Zabihi, M., Dinga, R., Berthet, P., Worker, A., Verdi, S., Ruhe, H.G.: The normative modeling framework for computational psychiatry. *Nature protocols* **17**(7), 1711–1734 (2022)
12. Wolfers, T., Doan, N.T., Kaufmann, T., Alnæs, D., Moberget, T., Agartz, I., Buitelaar, J.K., Ueland, T., Melle, I., Franke, B.: Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA psychiatry* **75**(11), 1146–1155 (2018)
13. Di Biase, M.A., Tian, Y.E., Bethlehem, R.A., Seidlitz, J., Alexander-Bloch, A.F., Yeo, B.T., Zalesky, A.: Mapping human brain charts cross-sectionally and longitudinally. *Proceedings of the National Academy of Sciences* **120**(20), e2216798120 (2023)
14. Zhang, Y., Lin, X., Zhuang, Y., Sun, L., Huang, Y., Ding, X., Wang, G., Yang, L., Yu, Y.: Harmonizing pathological and normal pixels for pseudo-healthy synthesis. *IEEE Transactions on Medical Imaging* **41**(9), 2457–2468 (2022)
15. Siddiquee, M.M.R., Shah, J., Wu, T., Chong, C., Schwedt, T.J., Dumkrieger, G., Nikolova, S., Li, B.: Brainomaly: Unsupervised neurologic disease detection utilizing unannotated t1-weighted brain mr images. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* 2024. pp. 7573–7582. <https://doi.org/10.48550/arXiv.2302.09200>
16. Rahman Siddiquee M. M., Shah J., Wu T., Chong C., Schwedt T., Li B.: Healthy-GAN: Learning from Unannotated Medical Images to Detect Anomalies Associated with Human Disease. In: *Simulation and Synthesis in Medical Imaging 2022*. springer, Singapore (2022). https://doi.org/10.1007/978-3-031-16980-9_5
17. Tang Y., Zhu Y., Xiao J., Summers R.M.: A disentangled generative model for disease decomposition in chest X-rays via normal image synthesis. *Med Image Anal* **67**, 101839 (2021)
18. Xia, T., Chatsias, A., Tsiftaris, S.A.: Pseudo-healthy synthesis with pathology disentanglement and adversarial learning. *Medical Image Analysis* **64**, 101719 (2020)
19. Tsunoda, Y., Moribe, M., Orii, H., Kawano, H., Maeda, H.: Pseudo-normal image synthesis from chest radiograph database for lung nodule detection. *Advanced Intelligent Systems*, 147–155.
20. Oh, K., Yoon, J.S., Suk, H.I.: Learn-explain-reinforce: counterfactual reasoning and its guidance to reinforce an alzheimer’s disease diagnosis model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(4), 4843–4857 (2022)
21. Kobayashi, K., Hataya, R., Kurose, Y., Miyake, M., Takahashi, M., Nakagawa, A., Harada, T., Hamamoto, R.: Decomposing normal and abnormal features of medical

- images for content-based image retrieval of glioma imaging. *Medical image analysis* **74**, 102227 (2021)
22. Yang, M., Liu, F., Chen, Z., Shen, X., Hao, J., Wang, J.: Causalvae: Disentangled representation learning via neural structural causal models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021*. pp. 9593–9602. IEEE, Nashville (2021). <https://doi.org/10.1109/CVPR46437.2021.00947>
 23. Qian, F., Guo, W., Liu, Z., Yu, H., Zhang, G., Hu, G.: Unsupervised erratic seismic noise attenuation with robust deep convolutional autoencoders. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–16 (2022)
 24. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision 2017*. pp. 2223–2232. IEEE, Montreal (2017). <https://doi.org/10.1109/ICCV.2017.244>
 25. Baumgartner, C.F., Koch, L.M., Tezcan, K.C., Ang, J.X., Konukoglu, E.: Visual feature attribution using wasserstein gans. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 8309–8319. IEEE, Salt Lake (2018).
 26. Peng, W., Adeli, E., Bosschieter, T., Park, S.H., Zhao, Q., Pohl, K.M.: Generating realistic brain mris via a conditional diffusion probabilistic model. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention 2023*. pp.14–24. Springer, Vancouver (2023). https://doi.org/10.1007/978-3-031-43993-3_2