[AnoAlZ] Unsupervised 3D anomaly detection of Alzheimer's disease using structural MR Images

Junbeom Kwon
Seoul National University
kjb961013@snu.ac.kr

Jiho Kim
Seoul National University
gkmoney7@snu.ac.kr

Jun Young Park
Seoul National University
junyoung0131@snu.ac.kr

Inhoe Lee
Seoul National University
inhoelee@snu.ac.kr

Abstract

Alzheimer disease is neuro-degenerative disease that requires early detection and intervention. Unlike previous deep learning approaches that cannot discriminate the severity in the same classes determined by prescriptions, we propose 3D GAN-based anomaly detection model that can provide reliable anomaly scores for predicting Alzheimer's disease. We adapted two recent 3D GAN architecture, Cycle-Consistent GAN(CCEGAN) and α-WGAN-gp, to the f-AnoGAN framework to improve the quality of generated images and embedding vectors. Compared to baseline 3D CNN models, CCEGAN with f-Anogan show comparable performance in classifying the participants of Normal and Mild Cognitive Impairment groups.

1. Introduction

Alzheimer's disease (AD), the most common cause of dementia, is a neurodegenerative disease with multiple symptoms associated with cognitive impairment. Many patients with Alzheimer's disease have difficulty in remembering recent events and suffer from several behavioral and emotional problems as the symptoms become worse.

Alzheimer's disease is associated with changes in the brain structure resulting from the accumulation of amyloid-$\beta(A\beta)$ plaques and hyperphosphorylated tau. Imaging methodologies, such as Positron emission tomography (PET) or magnetic resonance imaging (MRI), are usually used to detect the biomarkers of the disease. Previous study suggests that Alzheimer's disease shows structural changes in medial temporal lobe involvement (hippocampus and entorhinal cortex) before cognitive impairment appear [7].

Recently, the FDA approved the first medicine for Alzheimer's disease, Aducanumab, which alleviates the symptoms and delays the progression by reducing the $A\beta$ plaques. However, the medicine focuses on delaying the progression, not a complete cure, so early detection and intervention is highly required. Image-derived features including cortical thickness, subcortical volumes, and cerebral $A\beta$ accumulation in the regions of interest (ROI) were widely used to detect AD [13].

Researchers have adopted deep learning technologies like Convolutional Neural Network (CNN) from the computer vision area to enhance the performance by automatically extracting features from the brain images [2]. However, clinical datasets usually confront the issue of imbalance among the normal and diseased groups, which makes it difficult to have desired performances [10]. In addition, previous researches have focused on classifying Alzheimer's Disease, Mild Cognitive Impairment, and Normal control, which could not represent the severity of subjects in the same group.

To tackle the issue, unsupervised anomaly detection can be a remarkable approach. Anomaly detection is defined as the identification of test data deviating from the normal data distribution learned during training. For our case, we can find the brain images with Alzheimer's diseases using anomaly scores, which represents the degree of deviance from the normal brain images [15].

Recent studies have tried to apply anomaly detection to enhance the classification of Alzheimer's disease using PET images [3, 4]. In particular, the research [3] used GAN-based anomaly detection with 2-dimension PET images beating the performances of Densenet169 in classification of Alzheimer's disease. However, the effect of anomaly detection for the structural MRI of Alzheimer's disease was not yet verified. It is expected that MRI containing anatomical information would provide great insight that was not included in PET images. Furthermore, previous 2D-based approaches overlook volumetric information in complex brain

anatomy which can be found in 3D MRI [18]. Moreover, they have the limited numbers of normal subjects because they usually utilize only one source of database such as ADNI(http://adni.loni.usc.edu/).

In this work, we propose a unsupervised 3D anomaly detection model that can detect the severity of Alzheimer's disease using 3D MR images. We adapt the structure of Cycle Consistent Embedding GAN(CCE-GAN) to the framework of f-AnoGAN. [16, 19] With cycle-consistency loss and Wasserstein loss, the network of CCE-GAN is trained to encode the informative representation required for generating the original images. After learning from training dataset, the encoder, generator, and discriminator from the GAN model are used to calculate anomaly score in accordance with f-AnoGAN. Contribution of our research is as follows:

- This is the first approach using a 3D anomaly detection model for Alzheimer prediction.
- We improved the quality of generated images by decreasing the Wasserstein loss and cycle-consistent loss.
- Practically, our model can provide reliable anomaly scores, which may represent the severity and progress of Alzheimer's disease.

2. Related works

3D classification model for Alzheimer's prediction. Image-derived features including cortical thickness, subcortical volumes, and cerebral $A\beta$ accumulation in the regions of interest (ROI) were widely used to detect AD [13]. Researchers have adopted deep learning technologies such as Convolutional Neural Network (CNN) from the computer vision area to enhance the performance by automatically extracting features from the brain images [2].

There are two distinct approaches for classifying Alzheimer's disease from brain MRI data using 3D convolutional neural networks. One method is to apply 3D-CNNs to extract generic features automatically from MRIs then add other classifiers. [8, 20] The other method trains 3D-CNNs end-to-end using only MRI scans and labels as input. [11] Hosseini-Asl et al. [8] evaluated a short stack of unsupervised 3D convolutional auto-encoders (3D-CAE) on MRI data using the entire brain as input. Notably, they obtained a classification accuracy of 61.1 percent but found overfitting in the training data. In more recent research, Abrol et al. [1] constructed 3D CNNs based on the ResNet architecture and experimented with many binary and multi-class tasks using ADNI dataset. Despite the fact that their findings were encouraging, they conducted no additional comparisons to other established grading systems.

Unsupervised anomaly detection models. Unsupervised learning for visual anomaly identification is used to various biomedical imaging datasets, most notably brain and retinal images. Schlegl et al. firstly used GANs to predict the distribution of healthy retinal patches and found anomalies by computing the difference between the retinal patch and a healthy equivalent restored by the GAN. Schlegl et al. [14] recently published a follow-up work of f-AnoGAN, began by modeling the distribution of healthy retinal anatomy dataset using the Wasserstein-GAN algorithm.

Unsupervised anomaly detection in Alzheimer prediction with brain imaging has seen a variety of approaches. In particular, Generative models, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), are widely used to learn the distribution of the normal samples. For example, Choi et al. [5] introduced anomaly detection model based on Variational Autoencoder to detect anomalies from Positron Emission Tomography(PET), providing abnormality scores for identifying Alzheimer's disease. Baydargil et al. [3] proposed GAN-based anomaly analysis beating the performances of Densenet169 in classification of Alzheimer's disease. Recently, Han et al. [9] introduced a new Wasserstein-based Medical Anomaly Detection Generative Adversarial Network(MADGAN) using self-attention to take multiple adjacent slices of T1-weighted MRI as the input for the reconstruction, showing the usefulness of MRI to the detection of Alzheimer's disease.

However, most anomaly detection work has focused on 2D brain slices overlooking volumetric information in complex brain anatomy. Most studies using 3D volumes are mostly limited to generating brain images through 3D GAN rather than anomaly detection [12, 19]. For the first 3D application of AnoGAN, Viana et al. [18] proposed a volumetric extension of the 2D f-AnoGAN model to detect traumatic brain injury(TBI) abnormalities, combining a state-of-the-art 3DGAN with refinement training steps. Still, to the knowledge of this researcher, studies on the application of 3D Anogan to Alzheimer's disease have not yet been published. Therefore, we aim to predict Alzheimer's disease with 3-dimensional t1-weighted MRI, applying the state-of-art 3D Generative Adversarial Networks to the anomaly detection framework proposed by Schlegl et al.

3. Method

3.1. 3D GAN

The core idea of our method is to train encoder and generator of 3D GAN with the healthy subjects necessary for calculating anomaly scores of the diseased subjects. To learn the distribution of healty subjects, we use two 3D GAN approaches, α-Wasserstein GAN with Gradient Penalty (α-WGAN-gp) [12] and Cycle Consistent Embedding GAN (CCE-GAN) [19], which were re-

ported previously. Both methods combined the advantages of Variational Auto-Encoder (VAE) and Wasserstein GAN (WGAN). In [12], they added an module "code discriminator", which played adversarial learning with the encoder module, and optimized encoder-generator and code discriminator, seperately.

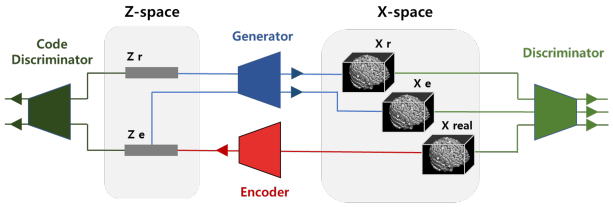

Figure 1. **3D GAN Model –** 3D α-WGAN-GP

We adopt 3D α-WGAN-GP architecture for training as Figure 1 with different parameter λ_2 in Equation 1.

$$\begin{aligned} L_D &= \mathbb{E}_{z_e}\left[D\left(G\left(z_e\right)\right)\right] + \mathbb{E}_{z_r}\left[D\left(G\left(z_r\right)\right)\right] - 2\mathbb{E}_{x_{\text{real}}}\left[D\left(x_{\text{real}}\right)\right] + \lambda_1 L_{GP-D} \\ L_G &= -\mathbb{E}_{z_e}\left[D\left(G\left(z_e\right)\right)\right] - \mathbb{E}_{z_r}\left[D\left(G\left(z_r\right)\right)\right] + \lambda_2 \left\|x_{\text{real}} - G\left(z_e\right)\right\|_{L1} \\ L_C &= \mathbb{E}_{z_e}\left[C\left(z_e\right)\right] - \mathbb{E}_{z_r}\left[C\left(z_r\right)\right] + \lambda_1 L_{GP-C} \end{aligned} \tag{1}$$

On the other hand, [19] trained the encoder and generator module at once with cycle consistency loss [21] to improve the embedding of the encoder module.

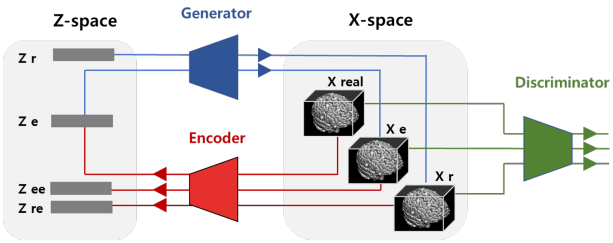

Figure 2. **3D GAN Model –** 3D CCE-GAN

We also adopt 3D CCE-GAN architecture for training as Figure 2 with different parameter λ_2 in Equation 2. With cycle-consistency loss and Wasserstein loss, the network of CCE-GAN is trained to encode the informative representation required for generating the original images.

$$\begin{aligned} &\underset{D}{\operatorname{argmin}} \mathbb{E}_{z_e}\left[D\left(X_e\right)\right] + \mathbb{E}_{z_r}\left[D\left(X_r\right)\right] - 2\mathbb{E}_{x_{\text{real}}}\left[D\left(x_{\text{real}}\right)\right] + \lambda_1 L_{gp}(D) \\ &\underset{G,E}{\operatorname{argmin}} - \mathbb{E}_{z_e}\left[D\left(X_e\right)\right] - \mathbb{E}_{z_r}\left[D\left(X_r\right)\right] + \lambda_2 \left\|X_r - X_e\right\|_1 + \\ &\lambda_3 \left\|z_r - z_{re}\right\|_2 + \lambda_3 \left\|z_e - z_{ee}\right\|_2 + \mathbb{W}_l\left(z_r, z_e\right) \end{aligned} \tag{2}$$

3.2. Anomaly detection

F-AnoGAN. F-AnoGAN is specialized in training a generator with an encoder that is used to transfer the testing images to the embedding space of training images. α-WGAN-gp also train encoder network which generates embedding from images. Therefore, we can use the generator, discriminator, and encoder of α-WGAN-gp trained with healthy subjects for calculating anomaly scores of the diseased subjects in the framework of f-AnoGAN.

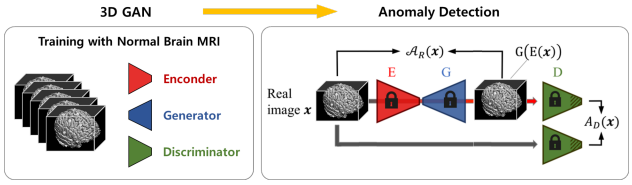

Figure 3. **3D f-ANoGAN Model**

Anomaly detection score. Anomalies are detected via a combined anomaly score based on the building blocks of the trained model – comprising a discriminator feature residual error and an image reconstruction error. The anomaly quantification formulation follows directly the specific definition of the loss used for encoder training (Equation 1, 2). For the proposed f-AnoGAN model, which implements the discriminator guided encoder training (Equation 1, 2), the final anomaly score $\mathcal{A}(\mathbf{x})$ for a new image $\mathbf{x}$ is defined by

$$\mathcal{A}(\mathbf{x}) = \mathcal{A}_R(\mathbf{x}) + \kappa \cdot \mathcal{A}_D(\mathbf{x}) \tag{3}$$

where

$$\mathcal{A}_R(\mathbf{x}) = \frac{1}{n} \cdot \|\mathbf{x} - G(E(\mathbf{x}))\|^2 \tag{4}$$

$$\mathcal{A}_D(\mathbf{x}) = \frac{1}{n_d} \cdot \|f(\mathbf{x}) - f(G(E(\mathbf{x})))\|^2 \tag{5}$$

and k is a weighting factor(Figure 3). In general, both formulations yield high anomaly scores on anomalous images and small anomaly scores on normal input images. Since the model is only trained on normal images, it only "reconstructs" an image visually similar to the input image and lying on the manifold of normal images X. The ability of reconstructing visually similar images is inversely proportional to the degree or distinction of anomaly. Normal query images result in small deviations whereas anomalous images are mapped to "reconstructions" yielding large deviations.

3.3. BaseLine Models

Machine learning(ML). T1-weighted images were preprocessed using recon-all script from FreeSurfer V7.0 [6]. We obtained cortical thickness values and subcortical volume features for grey matter(GM), white matter(WM) and cerebrospinal fluid (CSF) surface. Penalized logistic regressions (Ridge logistic regression and Lasso logistic regression) were built with above MRI features. The gridsearchCV on scikit-learn library from python was conducted to find the optimal hyper-parameter values.

3D CNN. For pre-processed MR images, we cropped each image to minimize background and normalized MR images to change the range of voxel intensity values from 0 to 1 per image using ScaleIntensity on Monai library. Finally all the images were resampled to the size of 96x96x96. We built two 3D CNN models. First, a simple CNN model was composed of 6 3D convolutional layers, 4 max pooling layers and 2 fully connected layers. Each 3D convolutional layer consisted of 3x3x3 kernel size, 1x1x1 stride and each max pooling layer with 2x2x2 size were constructed at the end of each two convolutional layers. After the final max pooling layer, two fully connected layers were adjusted with 4096 input nodes and 2 output nodes. Second, we conducted a 3D VGGNet11 [17], which added just 1 dimension compared to the original VGGNet11 to adjust 3D images. The 3D VGGNet11 consisted of 8 3D convolutional layers, 5 max pooling layers and 3 fully connected layers. The classifier of 3D VGGNet11 was replaced with three dense layers which composed of 4096 input nodes and 2 output nodes and finally a softmax layer was adjusted to predict the probability.

4. Experiments

4.1. Datasets

ADNI. From the Alzheimer's Disease Neuroimaging Initiative(ADNI), we use 3D T1-weighted MRI images of Normal Control(NC), Mild Cognitive Impairment(MCI), and Alzheimer's Disease(AD) participants. There are 1807 subjects for baseline year, and total 6414 volumes. We preprocess the datasets with recon-all function of FreeSurfer, which includes motion correction and intensity normalization. To decrease unnecessary memory use, we remove redundant background and resized the volumes into 64*64*64.

4.2. Image Generation

We adopt the structure of α-WGAN-gp and CCEGAN. We train α-WGAN-gp for 100,000 iterations and CCE-GAN for 50,000 iterations with batch size 8. To maximize the similarity between real images and fake images, we increase the relative importance of λ_2 in (1),(2). Plus, we leave out the intensity randomization proposed in the previous models. We sampled 148 participants as test dataset from 738 Control Normal(CN) participants. Including images collected several times from the same subjects, the test dataset has 453 CN images, and remaining 1887 images are used to train our GAN models. All the participants from MCI and AD is included into test dataset.(MCI:3003, AD:1071) Figure 4 are the comparison between original images and reconstructed images by Cycle Consistent Embedding GAN. Despite minor differences, it su. Figure 5 are slices from the sampled fake images.

Method	CN vs MCI		CN vs AD		MCI vs AD	
	Acc.	AUC	Acc.	AUC	Acc.	AUC
Baseline(ML)	0.57	0.61	0.76	0.84	0.68	0.76
Baseline(simple CNN)	0.52	0.52	0.80	0.75	0.70	0.67
Baseline(VGGNet11)	0.63	0.63	0.87	0.83	0.73	0.72
Ours(alpha+f)	0.56	0.45	0.69	0.42	0.34	0.48
Ours(CCE+f)	0.59	0.62	0.66	0.70	0.61	0.58

Table 1. **Classification results with anomaly score of the proposed model compared with baseline models**

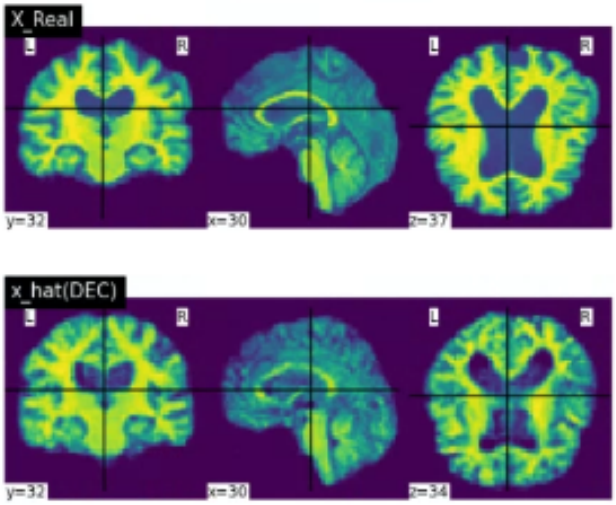

Figure 4. **Center Cut Slice Visualization –** Original and Reconstructed Images of Normal Control generated by CCE-GAN

4.3. Baseline Experiments

Unlike anomaly detection, we use the 1,807 volumes from the baseline year to avoid data leakage. Data for trainset and testset are randomly split up as 8:2 respectively. We measure accuracy, and area under the ROC curve (AUC) as the metrics for model performance. Experiments are repeated 10 times with several random seed and, finally, averaged each metric, respectively. For each model, the exactly same random seeds are adjusted.

4.4. Results

ML The classification results of our model are compared with baseline models (see Table 1). First the penalized logistic (machine learning method) with cortical thickness and subcortical volume features shows the accuracy of 56.8%, the AUC of 0.61 for CN vs MCI classification, the accuracy of 75.8%, the AUC of 0.75 for CN vs AD and the accuracy of 67.7%, the AUC of 0.76 for MCI vs AD.

3D CNN we compare our methods to two 3D CNN models with the pre-processed images. First, with the simple CNN model, we obtain the accuracy of 52.1%, the AUC of 0.52 for CN vs MCI classification, the accuracy of 80.2%,

Figure 5. **Slice Series Visualization –** Fake Images generated by 3D α-WGAN-GP

the AUC of 0.75 for CN vs AD and the accuracy of 72.9%, the AUC of 0.72 for MCI vs AD. The VGGnet11 model, which improve the accuracy and the AUC than the simple CNN model, show the accuracy of 63.2%, the AUC of 0.63 for classifying CN vs MCI, the accuracy of 86.5%, the AUC of 0.83 for CN vs AD and the accuracy of 72.9%, the AUC of 0.72 for CN vs MCI (see Table 1).

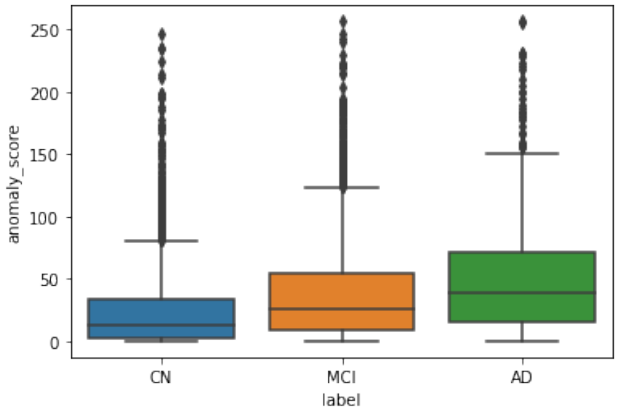

Figure 6. **CCE + f-ANoGAN Model box plot of anomaly score**

Anomaly detection Among the three intermediate scores, we decide to use anomaly score, which contains the l1 loss between original and reconstructed images, and difference between scores from discriminator for the two images. Figure 6 shows that the three groups are differed in anomaly scores to some extent. AD participants tend to have higher anomaly scores than MCI and CN participants. However, the boxes are overlapped among three groups, which means that the difference is not significant. To gauge more concrete measures, we choose a threshold that maximizes the difference between true positive rate and false positive rate as an optimal threshold. Table 1 shows that CCEGAN has better AUC over three groups than α-Wasserstein GAN. Furthermore, its performance is comparable to a general 3D classification model in classifying CN and MCI.

4.5. Conclusion and Discussions

Even though our model falls short of performance that can be utilized in the real world, it shows the possibility of unsupervised anomaly detection to Alzheimer's disease. Our model can encode and reconstruct high quality brain images from training dataset. In addition, there are room for improvement. Sufficient hyper-parameter tuning was not performed due to the extended learning time. We found that it is important to make the GAN focus on reconstruction by adjusting the loss differently. In addition, the embedding size seems to play a role in its capacity to learn representations of the training dataset. On top of that, we plan to use warped t1-weighted images in the future. It is usual that researchers warp 3D brain images into common space such as the MNI space when running a deep learning classifier. Considering the high variability in ADNI dataset, our GAN model might gets some benefits from spatial normalization. Furthermore, the spatial normalization can provide solution to supplement insufficient numbers of normal participants. We can utilize tremendous numbers of brain images from other sources such as UK Bio-bank. However, our model has some limitations in discriminating the abnormalities in unseen images. Learning the distribution of normal brain images does not prove that the model can disentangle the distribution in the unseen dataset. It seems that we have to develop a method that can control the output of generator.

Acknowledgments: The initial ADNI data was provided by Seonjoo Lee in Columbia University. Also, we thank Shiboxing, Gihyun Kwon for insightful discussions.

References

[1] Anees Abrol, Manish Bhattarai, Alex Fedorov, Yuhui Du, Sergey Plis, Vince Calhoun, Alzheimer's Disease Neuroimaging Initiative, et al. Deep residual learning for neuroimaging: An application to predict progression to alzheimer's disease. *Journal of neuroscience methods*, 339:108701, 2020. 2

[2] Jong Bin Bae, Subin Lee, Wonmo Jung, Sejin Park, Weonjin Kim, Hyunwoo Oh, Ji Won Han, Grace Eun Kim, Jun Sung Kim, Jae Hyoung Kim, and Ki Woong Kim. Identification of alzheimer's disease using a convolutional neural network model based on t1-weighted magnetic resonance imaging. *Scientific Reports*, 10(1):22252, Dec 2020. 1, 2

[3] Husnu Baris Baydargil, Jang-Sik Park, and Do-Young Kang. Anomaly analysis of alzheimer's disease in pet images using an unsupervised adversarial deep learning model. *Applied Sciences*, 11(5), 2021. 1, 2

[4] Ninon Burgos, M. Jorge Cardoso, Jorge Samper-González, Marie-Odile Habert, Stanley Durrleman, Sébastien Ourselin, Olivier Colliot, for the Alzheimer's Disease Neuroimaging Initiative, and the Frontotemporal Lobar Degeneration Neuroimaging Initiative. Anomaly detection for the individual analysis of brain PET images. 8(02), Apr. 2021. 1

[5] Hongyoon Choi, Seunggyun Ha, Hyejin Kang, Hyekyoung Lee, Dong Soo Lee, Alzheimer's Disease Neuroimaging Initiative, et al. Deep learning only by normal brain pet identify unheralded brain anomalies. *EBioMedicine*, 43:447–453, 2019. 2

[6] Bruce Fischl. FreeSurfer. 62(2):774–781, Aug. 2012. 3

[7] Giovanni B. Frisoni, Nick C. Fox, Clifford R. Jack, Philip Scheltens, and Paul M. Thompson. The clinical use of structural MRI in alzheimer disease. 6(2):67–77, Feb. 2010. 1

[8] Mohammed Ghazal. Alzheimer rsquo s disease diagnostics by a 3d deeply supervised adaptable convolutional network. 23(2):584–596, 2018. 2

[9] Changhee Han, Leonardo Rundo, Kohei Murao, Tomoyuki Noguchi, Yuki Shimahara, Zoltán Ádám Milacski, Saori Koshino, Evis Sala, Hideki Nakayama, and Shin'ichi Satoh. Madgan: unsupervised medical anomaly detection gan using multiple adjacent brain mri slice reconstruction. *BMC bioinformatics*, 22(2):1–20, 2021. 2

[10] Justin M. Johnson and Taghi M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27, Mar 2019. 1

[11] Alexander Khvostikov, Karim Aderghal, Jenny Benois-Pineau, Andrey Krylov, and Gwenaelle Catheline. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. *arXiv preprint arXiv:1801.05968*, 2018. 2

[12] Gihyun Kwon, Chihye Han, and Dae shik Kim. Generation of 3d brain MRI using auto-encoding generative adversarial networks. pages 118–126. Springer International Publishing, 2019. 2, 3

[13] Brandalyn C. Riedel, Madelaine Daianu, Greg Ver Steeg, Adam Mezher, Lauren E. Salminen, Aram Galstyan, and Paul M. Thompson and. Uncovering biologically coherent peripheral signatures of health and risk for alzheimer's disease in the aging brain. 10, Nov. 2018. 1, 2

[14] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 2

[15] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. pages 146–157. Springer International Publishing, 2017. 1

[16] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. pages 146–157. Springer International Publishing, 2017. 2

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. 4

[18] Jaime Simarro Viana, Ezequiel de la Rosa, Thijs Vande Vyvere, David Robben, Diana M. Sima, CENTER-TBI Participants, and Investigators. Unsupervised 3d brain anomaly detection. pages 133–142. Springer International Publishing, 2021. 2

[19] Shibo Xing, Harsh Sinha, and Seong Jae Hwang. Cycle consistent embedding of 3d brains with auto-encoding generative adversarial networks. In *Medical Imaging with Deep Learning*, 2021. 2, 3

[20] Chengliang Yang, Anand Rangarajan, and Sanjay Ranka. Visual explanations from deep 3d convolutional neural networks for alzheimer's disease classification. In *AMIA annual symposium proceedings*, volume 2018, page 1571. American Medical Informatics Association, 2018. 2

[21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2017. 3