

# Prediction of Alzheimer's Disease Progression with Multi-Information Generative Adversarial Network

Yan Zhao, Baoqiang Ma, Pengbo Jiang, Debin Zeng, Xueting Wang, and Shuyu Li 

**Abstract**—Alzheimer's disease (AD) is a chronic neurodegenerative disease, and its long-term progression prediction is definitely important. The structural Magnetic Resonance Imaging (sMRI) can be used to characterize the cortical atrophy that is closely coupled with clinical symptoms in AD and its prodromal stages. Many existing methods have focused on predicting the cognitive scores at future time-points using a set of morphological features derived from sMRI. The 3D sMRI can provide more massive information than the cognitive scores. However, very few works consider to predict an individual brain MRI image at future time-points. In this article, we propose a disease progression prediction framework that comprises a 3D multi-information generative adversarial network (mi-GAN) to predict what one's whole brain will look like with an interval, and a 3D DenseNet based multi-class classification network optimized with a focal loss to determine the clinical stage of the estimated brain. The mi-GAN can generate high-quality individual 3D brain MRI image conditioning on the individual 3D brain sMRI and multi-information at the baseline time-point. Experiments are implemented on the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our mi-GAN shows the state-of-the-art performance with the structural similarity index (SSIM) of 0.943 between the real MRI images at the fourth year and the generated ones. With mi-GAN and focal loss, the pMCI vs. sMCI accuracy achieves 6.04% improvement in comparison with conditional GAN and cross entropy loss.

**Index Terms**—Alzheimer's disease, multi-information generative adversarial network, multi-class classification, disease progression.

## I. INTRODUCTION

**A**LZHEIMER'S disease (AD), one of the most common causes of dementia, is a neurodegenerative disease that

Manuscript received November 12, 2019; revised May 30, 2020; accepted June 29, 2020. Date of publication July 3, 2020; date of current version March 5, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 81622025 and 81972160, and in part by the State Key Laboratory of Software Development Environment (SKLSDE-2019ZX-21). (Corresponding author: Shuyu Li.)

The authors are with the School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China, and also with the Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China (e-mail: ymdxjan@163.com; 1137723824@qq.com; pengbojiang@buaa.edu.cn; debin.z@qq.com; xuetingwang@buaa.edu.cn; shuyuli@buaa.edu.cn).

Digital Object Identifier 10.1109/JBHI.2020.3006925

can progressively cause memory-related impairments [1]. Although pharmacologic treatments are still not available to reverse the progression of AD, early intervention and treatment at its prodromal stages can offer the best chance to slow or stop the progression of AD [2], [3]. Alzheimer's-related brain changes may begin 3 to 10 years, some even more than 30 years before symptoms appear [4]. Thus, it is definitely important to predict disease progression for early intervention. The structural magnetic resonance imaging (sMRI) is an excellent noninvasive examination method for the early detection of brain anatomical changes in the progression of AD. Tremendous efforts have been made to perform the precise prediction of AD progression based on sMRI [5]–[8]. Recently, with the impressive development of generative adversarial networks (GAN) [9], the estimation of the brain sMRI images is receiving a lot of attention.

GAN is composed of a generator and a discriminator to pit one against the other. It not only has an extensive influence on computer vision, but also attracts lots of interests in the medical imaging community such as medical image synthesis, segmentation, reconstruction, detection, registration, noise reduction, classification and other various applications [9]–[14]. Moreover, GAN-based methods have also been utilized to perform mild cognitive impairment (MCI) conversion prediction based on synthetic multi-modal images, for example, Pan [15] and Yan *et al.* [16] have proposed conditional GAN (cGAN) for generating the missing brain PET from its corresponding MRI. They have achieved the highest classification accuracy for stable MCI (sMCI) against progressive MCI (pMCI) when using the synthetic PET images.

More recently, a few studies have explored the application value of GANs in estimating brain structural images at certain future time-points to predict disease progression. Specifically, the degenerative adversarial neuroimage nets (DaniNet) model with biological constraints has been proposed to generate MRI images that mimic disease progression with aging [17]. The Wasserstein GAN (WGAN) model has been applied to generate the hippocampus and the ventricles MRI images at certain future time-points by subtracting the multiples of the latent encoding of the AD features [18]. Additionally, the Temporal-GAN model has been proposed to predict whether a MCI subject converted to AD within 36 months [19]. The generator of the Temporal-GAN model is trained to approximate the joint distribution of the T1-weighted MRI images and labels at adjacent time-points for producing more reliable samples [19]. These studies have proved

the application prospects of GAN-based methods in predicting disease progression.

However, most of the existing GAN-based models employ 2D-MRI slices as the input and output images when modeling disease progression, which can not reveal the progressive changes of the whole 3D brain. Despite these GAN-based models generate MRI images at future time-points, they do not further explore diagnostic application of the generated MRI images. Although DaniNet allows conditioning the disease progression on non-imaging characteristics (age and diagnosis) of individuals [17], it ignores the factors that can significantly impact the AD progression, such as gender [20], education level [21], and apolipoprotein E (APOE) gene [22]. Furthermore, most of the existing AD progression studies focus on distinguishing whether a MCI subject will convert to AD within three years [23], they do not consider other cases, such as whether a normal control (NC) will progress to MCI within a long-term interval [24], such as four years. Thus, how to generate high-quality realistic 3D whole MRI images with a long-term interval considering other non-imaging information and how to classify the generated images into different diagnostic stages still have huge challenges.

To overcome the above-mentioned limitations, in this paper, we propose a disease progression prediction framework that consists of two different models: the first one is a 3D GAN model to generate the 3D MRI images at future time-points; the second one is a 3D classification model to identify the stages of the generated MRI images. Specifically, a multi-information generative adversarial network, named mi-GAN, is proposed to produce estimated 3D sMRI at a long-term time-point, conditioning on corresponding 3T sMRI and multi-information at baseline time-point (BL). The multi-information includes the age, gender, education level, and APOE  $\epsilon 4$  allele status of an individual at BL. The 3D classification model is built based on Densely Connected Convolutional Networks (DenseNet) [25], to classify the generated images into different stages. In this way, we can predict the disease progression of an individual. The experimental results on Alzheimer's Disease Neuroimaging Initiative (ADNI) indicate that our estimated 3D sMRI images are considerably realistic and the proposed disease progression prediction framework is appealing. The main contributions of the paper are in threefold, listed as following:

- 1) We propose a novel disease progression framework to predict progression that covers five progression cases (i.e., MCI-AD, MCI-MCI, MCI-NC, NC-MCI and NC-NC);
- 2) We propose a novel mi-GAN model and employ multiple loss functions to generate individual high-quality 3D sMRI images at future time-point conditioning on multi-information at BL;
- 3) We develop a multi-class classification model with a focal loss to classify the clinical stage of the generated sMRI image and to further predict the disease progression.

The rest of the paper is arranged as follows. Section II introduces the dataset, image preprocessing, details of mi-GAN and multi-class classification models. Experimental results and analyses are presented in Section III. Finally, Section IV draws conclusions and future perspectives.

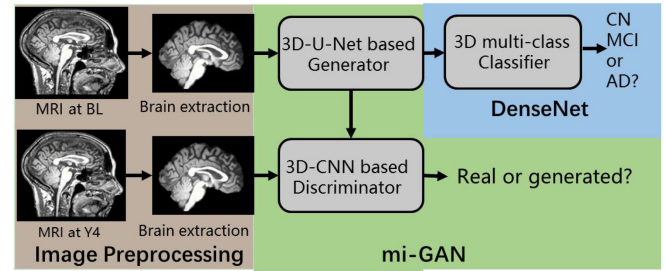


Fig. 1. The illustration of the progression prediction framework, comprising of image preprocessing (in brown), mi-GAN model (in green), and DenseNet based multi-class classification model (in blue).

A preliminary version of this work has been reported in a conference abstract [43].

## II. MATERIALS AND METHODS

The goal of this study is to construct a progression prediction framework that can generate the individual brain images at the future time-point as well as identify the brain stage from the predicted images. The proposed framework is illustrated in Fig. 1 including three main steps. First, image preprocessing is performed for the ADNI data. Then, a 3D patch-based mi-GAN model is constructed to generate estimated brain MRI image at a future time-point (i.e., Year 4, Y4), conditioning on the brain MRI image at baseline time-point (i.e., BL). Finally, a 3D DenseNet based multi-class classification model is established to identify the clinical stage of the estimated brain MRI. More details are provided in the following subsections.

### A. Dataset

We adopt the ADNI-GO and ADNI-2 (adni.loni.usc.edu) to train our models. The ADNI is an ongoing, multisite, longitudinal dataset. The time-point when the subject accomplishes the first screening is called baseline, denoted as BL. One year and four years later since BL are denoted as Y1 and Y4.

In this study, to build the mi-GAN model, we use 210 I subjects' T1-weighted MRI images at BL, Y1 and Y4 time-points and their corresponding multi-information (i.e., age, gender, education level, and APOE  $\epsilon 4$  allele status) at BL. The demographic characteristics of the above 210 subjects (named as dataset A) are displayed in Table I. Please note that only 205 subjects have been scanned at Y2 and 33 subjects at Y3, so we do not train mi-GAN models for generating estimated Y2 and Y3 images. A total of 2189 T1-weighted MRI images of other 603 ADNI subjects, scanned at multiple time-points are used to train the multi-class classification model. The demographics of the above 603 subjects (named as dataset B) at BL are shown in Table II. We use OASIS (<http://www.oasis-brains.org/>) to perform an independent cross-validation, by selecting subjects who are scanned with one or four years' interval. 48 subjects meet this criterion, and they are all in NC group at baseline and remain in NC group at Y1 or Y4, except 2 subjects convert to AD at Y1. The demographic characteristics of those 48 subjects (named as dataset C) at BL are shown in Table III.

**TABLE I**  
SUMMARY OF DATASET A SUBJECTS INFORMATION

Cases	MCI-MCI	NC-NC	MCI-AD	MCI-NC	NC-MCI
Number	78	67	37	19	9
Sex(F/M)	35/43	33/34	19/18	10/9	5/4
Age	70.0±6.8	73.9±6.5	70.4±7.4	67.0±7.3	76.0±6.5
Edu	16.4±2.5	16.9±2.4	15.8±2.5	17.3±2.4	15.1±2.1
APOE	44,25,9	50,16,1	8,19,10	10,9,0	6,2,1
MMSE	28.2±1.6	29.3±1.0	27.1±1.6	29.4±0.8	28.1±1.8
Interval	4.09±0.1	4.12±0.1	4.11±0.1	4.05±0.04	4.18±0.1

Key: The data is presented in a mean± standard deviation format. MCI-MCI and NC-NC indicate MCI and NC that not convert to other stages within 4 years; MCI-AD and MCI-NC denote that MCI progress to AD and NC within 4 years; NC-MCI denotes NC that progress to MCI within 4 years; F, female; M, male; Edu, education level (years); APOE, the distribution of the number of  $\epsilon 4$  alleles (0/1/2); MMSE, Mini-Mental State Examination score; Interval, duration (years) between BL and Y4.

**TABLE II**  
SUMMARY OF DATASET B SUBJECT INFORMATION

Category	NC	MCI	AD
Number	113	341	151
Sex	61/52	151/190	63/88
Age	72.8±6.0	72.3±7.6	74.7±8.2
Edu	16.5±2.6	16.1±2.7	15.8±2.7
APOE(0/1/2)	78,29,4,-	182,123,28,-	48,70,29,-
MMSE	28.9±1.3	28.0±1.7	23.1±2.1

Key: “-” denotes APOE  $\epsilon 4$  allele information is not clear.

**TABLE III**  
SUMMARY OF DATASET C SUBJECT INFORMATION AT BASELINE

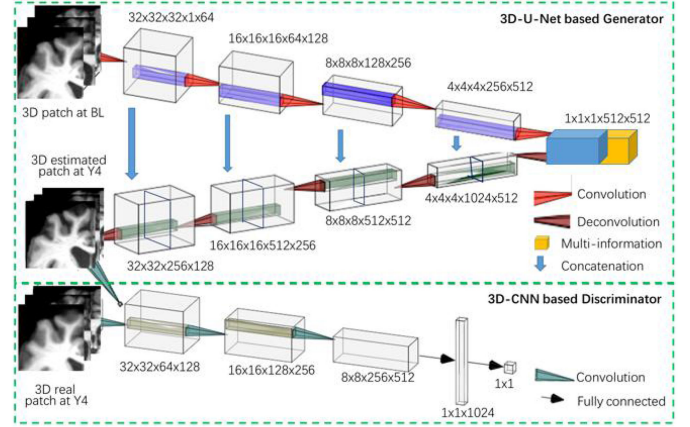
	F/M	Age	Edu	APOE	MMSE	Interval
BL-Y1	10/14	71.3±6.2	16.7±2.5	13,9,2	29.0±1.2	1.0±0.05
BL-Y4	11/13	70.4±7.4	16.6±2.7	20,4,0	29.4±0.7	4.0±0.05

Key: “BL-Y1/Y4” denotes the subjects scanned at BL and Y1/Y4.

## B. Image Preprocessing

Brain extraction is performed on sMRI image to strip the non-brain tissues, including: (1) segmenting the sMRI image into gray matter (GM), white matter (WM), cerebral spinal fluid (CSF) with VBM (<http://dbm.neuro.uni-jena.de/vbm>) toolbox [26]; (2) adding the GM, WM and CSF to obtain a brain mask; and (3) multiplying the brain mask with the MRI image to obtain an initial brain image. Then, the initial brain image is linearly registered to the MNI152\_T1\_1mm\_brain template with FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL>) to obtain a brain image. Finally, to remove uninformative background and reduce computation cost, all brain images are cropped from  $182 \times 218 \times 182$  to  $160 \times 192 \times 160$ . Moreover, to reduce calculation burden, we extract 80 patches (i.e.,  $4 \times 5 \times 4$ , with a stride of 32) of size  $64 \times 64 \times 64$  from  $160 \times 192 \times 160$  when training mi-GAN model, and downsample images from  $160 \times 192 \times 160$  to  $80 \times 96 \times 80$  before fed into the classification model.

Additionally, we further non-linearly register all the brain images at Y1 and Y4 in dataset A to their corresponding ones at BL by adopting the ANTs (<http://stnava.github.io/ANTs/>) to get more data (i.e.,  $Y1_{reg}$  and  $Y4_{reg}$ ) to verify the robustness of our proposed mi-GAN model.



**Fig. 2.** The structure of the mi-GAN model. The boxes represent feature maps. The yellow box represents individual multi-information at BL. The number above/below the feature maps denote the size of the 3D volumes and the input and output channels, i.e., Height  $\times$  Width  $\times$  Depth  $\times$  Number of input channel  $\times$  Number of output channel.

## C. 3D Patch-Based MI-GAN Model

Inspired by the pix2pix architecture [27], we propose a 3D patch-based mi-GAN model to generate high-quality images at future time-points with two improvements. They are (1) using a 3D U-Net [28] based network conditioning on the image patches and multi-information (i.e., age, gender, education level, and APOE) at BL; (2) adding a gradient difference loss ( $GDL$  loss) [29] and a mean square error loss [30] in both image space and frequency domain ( $L2$  loss) to the final objective function. The 3D U-Net based network can effectively utilize the multi-scale features of input image patches and multi-information. The  $GDL$  loss can enable generator to produce less blurry and more realistic images.

More specially, the mi-GAN is composed of 3D U-Net based generator  $G$  and a 3D CNN based discriminator  $D$ , which are trained simultaneously like playing a two-player min-max game. The structure of our proposed mi-GAN model is illustrated in Fig. 2. The image patches and corresponding multi-information at BL input to  $G$ . The output of  $G$  is the estimated image patches at a future time-point. Then, the estimated or the real image patches of the individual at a future time-point input to  $D$ . The output of  $D$  is a classification result indicating whether the input image patch of  $D$  is real or generated. The training of  $G$  and  $D$  is guided by an objective function and stops until  $G$  can generate enough realistic image patches that  $D$  cannot distinguish them from the real ones.

Given image patch  $x_{bl} \sim p_{bl}$  and its corresponding multi-information  $x_{mi} \in \mathbb{R}^{1 \times n}$  at BL, and its target real image patch  $y_i \sim p_{yi}$  at the  $i^{th}$  years of an individual, the final objective function of our proposed mi-GAN is formulated as:

$$\arg \min_{\theta_G} \max_{\theta_D} (\alpha L_{mi-GAN}(G, D) + \beta L_{L2}(G) + \gamma L_{GDL}(G)) \quad (1)$$

where  $\theta_G$  and  $\theta_D$  are the parameters to be optimized of  $G$  and  $D$ ,  $\alpha > 0$ ,  $\beta > 0$  and  $\gamma > 0$  are three tunable hyper-parameters to



balance the importance of mi-GAN loss,  $GDL$  loss and  $L2$  loss items. The mi-GAN loss is expressed as:

$$L_{mi-GAN}(G, D) = \mathbb{E}_{y_i \sim P_{y_i}} [\log D(y_i)] + \mathbb{E}_{x_{bl} \sim p_{bl}} [\log (1 - D(G(x_{bl}, x_{mi})))] \quad (2)$$

where  $\mathbb{E}$  is the expectation for a given probability distribution.  $p_{bl}$  and  $p_{y_i}$  denote distribution of inputting image patch  $x_{bl}$  and its targeting real image patch  $y_i$ .  $x_{mi}$  denotes multi-information.

The  $L2$  loss both in image space and in frequency domain is employed as:

$$L_{L2}(G) = \mathbb{E}_{y_i \sim p_{y_i}, x_{bl} \sim p_{bl}} [(fft(y_i) - fft(G(x_{bl}, x_{mi})))^2] + \mathbb{E}_{y_i \sim p_{y_i}} [(y_i - G(x_{bl}, x_{mi}))^2] \quad (3)$$

where  $fft(\cdot)$  means computing the discrete Fourier transform of  $y_i$  and  $G(x_{bl}, x_{mi})$  using a fast Fourier transform (FFT) algorithm to transform  $y_i$  and  $G(x_{bl}, x_{mi})$  into frequency domain.

Meanwhile, the  $GDL$  loss is also used to solve the over-smooth issue of the generated images, which is defined as:

$$L_{GDL}(G) = ||\nabla y_i - \nabla G(x_{bl}, x_{mi})||^2 \quad (4)$$

where  $\nabla$  represents the vector differential operator. This loss tries to minimize the differences of gradient magnitudes between the patches of ground-truth and estimated image patches. In this way, the estimated image patches will try to keep the regions with sharp edges.

### D. 3D Multi-Classification Model

Once 3D whole MRI brain images at Y4 obtained, we feed them into a trained 3D ternary classification model to classify their stages (i.e., AD, MCI or AD) and further identify their disease progression within four years.

It has been shown that the 2D DenseNet model outperforms state-of-the-art models [25], mainly because of three reasons: (1) it requires fewer parameters, as there is no need to re-learn redundant feature maps; (2) it is easier to train, due to improved information flow and gradients throughout the network; and (3) dense connections of DenseNet have a strong regularizing effect and can reduce overfitting on smaller training sets [31]. Thus, we extend the idea of 2D DenseNet to 3D volumetric image classification tasks and establish a 3D DenseNet based multi-classification model to classify the stage of the 3D generated brain image. The structure of 3D DenseNet based multi-classification model is shown in Fig. 3.

Specially, the 3D generated images are first fed into a  $3 \times 3 \times 3$  convolutional layer, and then input to a denseblock. In a denseblock, each dense layer receives feature maps of all previous dense layers by shortcut connection [32]. The dense layer is composed of two BN and ReLU, one  $1 \times 1 \times 1$  and one  $3 \times 3 \times 3$  convolutional layers and a dropout layer to avoid overfitting. After the denseblock, a transition layer is used to change the number of feature-volumes between blocks. The transition layer consists of one BN, one ReLU, one Conv and one average pooling ( $2 \times 2 \times 2$ ) operations. Then another two

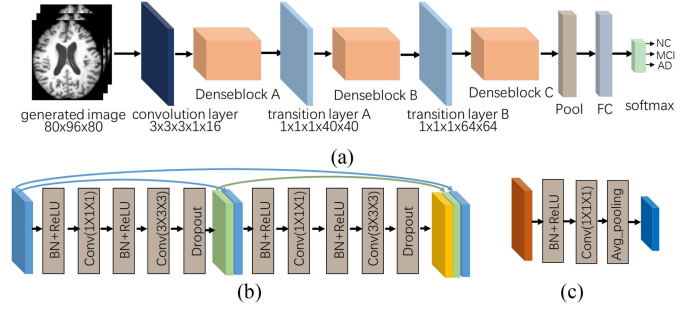


Fig. 3. The structure of 3D DenseNet based multi-classification model (a), which comprises three denseblocks (b) and two transition layers (c). The denseblock shown in (b) comprises 2-layer denseblock with bottleneck. The arrows indicate the information flow. The boxes represent feature maps. BN refers to a batch normalization layer. ReLU refers to a rectified linear unit activation function. Conv denotes a convolutional layer. Avg\_pooling denotes an average pooling layer. The number in subfigure (a) denote the size of filter and the number of input and output channels. The number in subfigures (b) and (b) denote the size of filter.

denseblocks and one transition layer follow as shown in Fig. 3. The growth rate and the dropout rate of each denseblock are set to 12 and 0.6, respectively. Finally, a  $5 \times 6 \times 5$  average pooling layer, a fully connected layer and a softmax layer are appended to make the final ternary classification.

Meanwhile, to achieve better classification results, we adopt the focal loss (FL) [33], [34], instead of the conventional softmax cross entropy loss (i.e., CEL), as the objective function of the 3D multi-classification model. The focal loss can address class imbalance problem and make hard images to be easily classified. The FL with an  $\alpha$ -balanced variant is expressed as follows:

$$FL(\mathbf{z}, y) = -\alpha_t \sum_{i=1}^3 (1 - p_i^t)^\gamma \log(p_i^t) \quad (5)$$

$$p_i^t = \text{sigmoid}(z_i^t) = \frac{1}{1 + e^{-z_i^t}} \quad (6)$$

$$z_i^t = \begin{cases} z_i & \text{if } i = y \\ -z_i & \text{otherwise} \end{cases} \quad (7)$$

where  $\mathbf{z}$  denotes a total number of classes, i.e.,  $\mathbf{z} = [z_1, z_2, z_3]$ .  $\gamma \geq 0$  refers to a tunable focusing parameter to modulate loss from hard and easy images.  $p_i^t$  is the probability of being classified to the true class.

## III. EXPERIMENTAL RESULTS

### A. Datasets and Implementation Details

Both datasets A and B are used to test the proposed disease progression prediction framework. The dataset A is used to verify the proposed mi-GAN model, which can generate high-quality estimated images at future time-points. The dataset B is used to train the 3D DenseNet based multi-classification model to accurately recognize the above generated image in which stage. We will describe the experiments and the results of these two models separately.

To train and test the mi-GAN model, we utilize a 7-fold cross-validation method. The dataset A is partitioned into 7 equal size folds. Each fold comprises 30 image pairs. Each image pair is composed of MRI image at BL and its corresponding MRI image at a future time-point. One fold is retained as the test data, and the other 6 folds are used as the training data, including one fold used as validation data. Considering the calculation cost and the limited training data, the mi-GAN model is trained and tested on image patches, as described in Section II. The generated image patches for each subject are then merged together by averaging the overlap to obtain the 3D brain MRI (with the size of  $182 \times 218 \times 182$ ) at a future time-point. Besides, we perform min-max normalization on individual image patches and multi-information at BL before inputting into the mi-GAN model. To train and test the 3D DenseNet based multi-classification model, we take the dataset B as training data, and the real and corresponding generated MRI images at a future time-point as validation data and test data, respectively.

All the experiments are implemented by TensorFlow deep learning framework, and carried out on a NVIDIA Titan Xp GPU with 12 GB memory. The mi-GAN model is trained using the Adam optimizer with a learning rate of 0.0002, exponential decay rate for the first moment of 0.5, and the mini-batch size of 20. The initial weights of the mi-GAN model are uniformed with  $\text{stddev} = 0.02$ . The scale parameter of each batch normalization is uniform with  $\text{stddev} = 0.02$  and  $\text{mean} = 1$ . The generator is trained using  $\alpha = 1$ ,  $\beta = 300$  and  $\gamma = 300$ . The number of training epochs is set to 200, but the training will stop if SSIM of the validation dataset is not improved in 6 epochs. The activation function for each neuron is leaky rectified linear unit with slide 0.2. Each epoch of training the GAN model costs about 50 minutes. The training of cGAN is much more time-consuming than mi-GAN.

Furthermore, the multi-classification model is trained with a Momentum optimizer with a momentum of 0.9, a learning rate of 0.01 at first 60 epochs and 0.001 at next 70 epochs, the mini-batch size of 10, and its initial weights are uniformed with  $\text{stddev} = 0.02$ . The  $\gamma$  and  $\alpha_t$  in the focal loss are set to 2 and 0.25, respectively. Early stopping is also used. The total training time of the multi-classification model costs about 2 hours.

### B. Performance of MI-GAN Model

The structural similarity index (SSIM) between the ground truth  $Y$  (i.e., real image) and its corresponding generated image  $\hat{Y}$  is adopted to evaluate the performance of the mi-GAN model. Theoretically, a higher SSIM means higher quality. The SSIM between  $Y$  and  $\hat{Y}$  can be defined as:

$$\text{SSIM}(Y, \hat{Y}) = \frac{(2\mu_Y\mu_{\hat{Y}} + C_1)(2\sigma_{Y\hat{Y}} + C_2)}{(\mu_Y^2 + \mu_{\hat{Y}}^2 + C_1)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + C_2)} \quad (8)$$

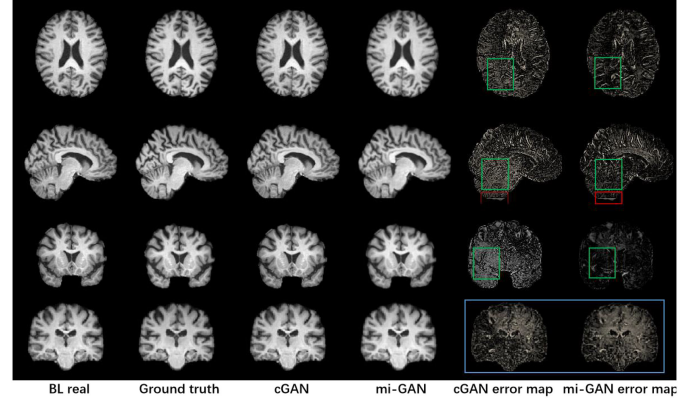
where  $\mu_Y$ ,  $\mu_{\hat{Y}}$ ,  $\sigma_Y$ ,  $\sigma_{\hat{Y}}$ , and  $\sigma_{Y\hat{Y}}$  are the means, standard deviations, and cross covariance of  $Y$  and  $\hat{Y}$ , and  $C_1$  and  $C_2$  are regularization constants, determined by the pixel value range.

Apart from 3D mi-GAN, we also constitute a 3D cGAN for comparison, which only conditions on the individual MRI images at BL. The experimental results of generated 3D MRI

**TABLE IV**  
SUMMARY OF SSIM RESULTS OF DIFFERENT DATA PAIR BY TWO 3D GAN MODELS

	$Y1 \leftrightarrow \hat{Y1}$	$Y1_{\text{reg}} \leftrightarrow \hat{Y1}_{\text{reg}}$	$Y4 \leftrightarrow \hat{Y4}$	$Y4_{\text{reg}} \leftrightarrow \hat{Y4}_{\text{reg}}$
cGAN	0.944±0.038	0.944±0.038	0.939±0.031	0.939±0.031
mi-GAN	<b>0.945±0.038</b>	<b>0.946±0.038</b>	<b>0.943±0.028</b>	<b>0.942±0.031</b>

Key: The data is presented in a mean  $\pm$  standard deviation format.  $\leftrightarrow$  denotes the SSIM between the ground truth  $Y$  and corresponding generated image  $\hat{Y}$ , where  $\cdot$  denotes different data pairs.  $Y_{\text{reg}}$  represents the nonlinearly registered images to BL ones.



**Fig. 4.** Qualitative comparison of MRI images of an EMCI-AD subject obtained by 3D cGAN and the proposed 3D mi-GAN. The 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>/4<sup>th</sup> rows show the axial, sagittal and coronal views, respectively. The first two columns represent the real images at BL and Y4. The middle two columns represent the estimated images generated by cGAN and mi-GAN. The last two columns represent the error map between the ground truth and the estimated images generated by cGAN and mi-GAN, respectively. The higher intensity (i.e., more white parts) means more error.

images at different future time-points (i.e.,  $Y1$  and  $Y4$ ) based on 3D cGAN and mi-GAN models are reported in Table IV. Compared with the best SSIM result (i.e., 0.62) achieved by the 2D based GAN model of Ravi [17], the results of both 3D cGAN and mi-GAN models are significantly improved. From Table IV, we can see that the SSIM between the ground truth and corresponding estimated images generated by 3D mi-GAN are all higher than the ones generated by 3D cGAN and there are significant differences between the SSIM obtained by mi-GAN and cGAN ( $p < 0.05$ ). The  $p$  values can be obtained by performing the two-tailed paired t test between the SSIM values of cGAN and mi-GAN. Furthermore, we adopt dataset C to validate our trained GAN models. The SSIM results of  $Y1 \leftrightarrow \hat{Y1}$  and  $Y4 \leftrightarrow \hat{Y4}$  using mi-GAN are  $0.9589 \pm 0.0140$  and  $0.9590 \pm 0.0138$ , while using cGAN are  $0.9576 \pm 0.0156$  and  $0.9569 \pm 0.0144$ , and their  $p$  values are 0.2260 and 0.0026, respectively.

To qualitatively compare the generated MRI images by 3D cGAN and mi-GAN models, we visualize the generated MRI images and their corresponding ground-truth of a subject (EMCI at BL and progress to AD at Y4) and the difference between those images in Fig. 4. In Fig. 4, we can observe that the error map shown in the first three rows indicate that the generated images based on mi-GAN are of higher quality than cGAN. For example, the intensity of the difference image based on cGAN is higher than that of mi-GAN, especially the regions limited by the green

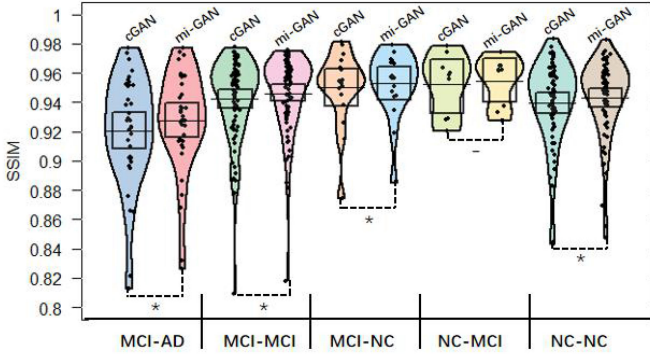


Fig. 5. SSIM distribution of the estimated images at Y4 generated by cGAN and mi-GAN for each case. Black points denote the SSIM between the generated image and real image at Y4 of all subjects in dataset A. Bar and band in the middle denote the mean and the Confidence Interval (CI) of SSIM for each case. Bean denotes the smoothed density curve showing the SSIM distribution for each case. The narrower bean means the less concentrative distribution. “\*” and “-” under the dash line denote the p value obtained by performing the paired t test on each case when  $p < 0.05$  and  $p > 0.05$  respectively.

boxes. Note that the error shown in the red boxes is caused by the operation of cropping the original whole brain images that the input brain images of cGAN or mi-GAN models are flawed. However, the error maps shown in the blue box indicate that this slice generated by both mi-GAN and cGAN are undesired, and cGAN seems better than mi-GAN for most regions. The reason might be that the training of the generator stops when the quality of most, but not all, slices reach an optimal level. The mi-GAN outperforms cGAN on most slices, but not on all slices.

To analyze the quality of the generated MRI images for each different disease progression case, we show the SSIM distribution of each case in Fig. 5. In Fig. 5, we can observe that the mean of SSIM obtained by mi-GAN is greater than cGAN and the deviation of SSIM obtained by mi-GAN is smaller than cGAN. And the SSIM distribution of mi-GAN is better than cGAN. Take the MCI-AD case for example, the mean of SSIM obtained by cGAN is 0.92, while mi-GAN is greater than 0.93; Even the worst SSIM obtained by mi-GAN is better than cGAN. There are significant differences ( $p < 0.05$ ) between the SSIM obtained by mi-GAN and cGAN in all cases except the NC-MCI. This may be due to a minimum of image pairs in the NC-MCI case, which limits the learning of GAN model in this case.

To evaluate the performance of the proposed mi-GAN in terms of volumetric features, we extract the gray matter, white matter, cerebral spinal fluid, hippocampus, parahippocampal gyrus and total intracranial volume ( $\text{cm}^3$ ) from the real MR images at Y4 and the generated counterparts using the SPM toolbox and perform statistical analysis on these features. Results are shown in Table V. We can observe that the performance of mi-GAN is dramatically better than cGAN in CSF and TIV, but not for all these structures. The reason may be that the target of GAN model is to generate the whole brain image, which is similar to the real whole brain image as much as possible. The performance of mi-GAN as a whole is better than cGAN, but not in all features extracted from the whole brain image. The performance on dataset C is much better than dataset A, which may be due

TABLE V  
PERFORMANCE OF MI-GAN AND CGAN IN TERMS OF VOLUMETRIC FEATURES ON DATASETS A AND C

Y1/Y4 (A&C)	Real	cGAN	mi-GAN	RMSE cGAN	RMSE mi-GAN	P
GM	764.6±63.1/	678.9±45.4/	670.1±40.1/	<b>104.1/</b>	111.0/	*/
	750.4±55.1&	677.1±45.5&	657.3±40.5&	<b>87.5&amp;</b>	102.8&	*&
	792.6±45.22/	753.3±58.6/	760.6±36.0/	60.5/	<b>39.2/</b>	-/
	782.6±45.9	743.0±42.2	751.3±40.3	44.7	<b>33.3</b>	-
WM	685.5±49.6/	673.2±36.1/	679.1±43.6/	42.7/	<b>41.7/</b>	*/
	678.8±37.9&	670.9±35.5&	685.9±34.2&	<b>29.3&amp;</b>	32.1&	*&
	689.5±30.8/	684.3±41.3/	695.1±25.4/	24.6/	<b>22.9/</b>	-/
	695.8±38.8	704.5±33.6/	703.8±27.8	<b>17.2</b>	19.5	-
CSF	395.6±50.3/	294.1±40.7/	312.9±37.2/	107.5/	<b>91.0/</b>	*/
	417.0±50.1&	293.0±39.4&	325.2±34.6&	128.7&	<b>99.0&amp;</b>	*&
	386.4±42.5/	346.8±48.3/	386.8±38.3/	55.2/	<b>11.4/</b>	*/
	390.5±48.2	335.2±49.7	382.3±43.2	57.6	<b>13.2</b>	*
HIPPO	9.35±1.53/	8.92±1.48/	8.93±1.50/	<b>0.81/</b>	0.83/	-/
	8.73±1.76&	8.83±1.37&	8.82±1.46&	0.76&	0.76&	-&
	9.63±1.28/	9.59±1.31/	9.83±1.22/	<b>0.36/</b>	1.97/	*/
	9.53±1.41	9.60±1.37	9.83±1.36	<b>0.40</b>	0.48	*
P_H	9.80±1.06/	8.90±1.08/	8.91±1.11/	1.28/	<b>1.25/</b>	-/
	9.67±0.94&	8.99±0.89&	8.98±0.90&	1.02/&	<b>0.96&amp;</b>	-&
	9.97±1.11/	9.83±1.02/	9.80±1.10/	0.44/	<b>0.39/</b>	-/
	9.96±0.66	9.86±0.58	9.89±0.60	0.24	<b>0.21</b>	-
TIV	1845.6±102.1/	1646.2±66.5/	1662.1±56.5/	230.1/	<b>215.3/</b>	*/
	1846.3±21.9&	1640.9±64.1&	1668.4±52.6&	214.4&	<b>185.6&amp;</b>	*&
	1868.5±22.8/	1784.4±95.2/	1842.5±19.9/	125.5/	<b>37.2/</b>	*/
	1868.9±16.5	1782.7±32.7	1837.4±12.6	93.1	<b>36.8</b>	*

Key: GM, WM, CSF, HIPPO, P\_H and TIV represent the gray matter, white matter, cerebral spinal fluid, hippocampus, parahippocampal gyrus and total intracranial volume respectively. RMSE represents root mean square error between the volume extracted from the real Y1/Y4 images (Real) and their counterparts generated by cGAN or mi-GAN on the datasets A and C (A&C). P is the p value obtained by performing the paired t test between the volumetric values of cGAN and mi-GAN, where “\*” and “-” denote  $p < 0.05$  and  $p > 0.05$ .

to only two subjects in dataset C convert to AD stage and other subjects maintain at NC stage.

### C. Performance of Multi-Classification Model

The performance of 3D DenseNet based multi-classification model not only reflects the quality of the generated MRI images, but also can represent the performance of the proposed disease progression prediction framework. We employ the overall accuracy, precision, recall, F1 score and pMCI vs. sMCI accuracy (MCI-AD vs. MCI-MCI) to measure the performance of the classification model, calculated by the following formulas [44]:

$$\text{ACC} = \frac{\sum_{i=1}^n TP_i}{N} \quad (9)$$

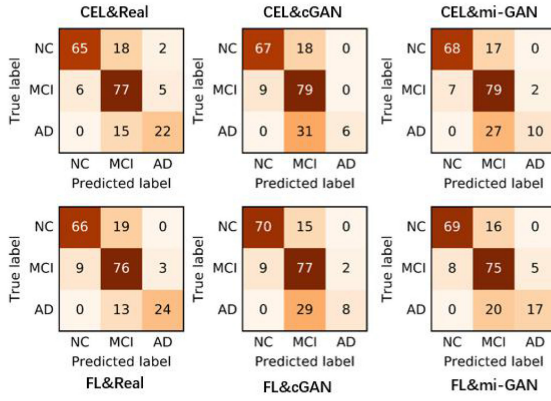
$$\text{Precision} = \frac{\sum_{i=1}^n \frac{TP_i}{TP_i + FP_i}}{n} \quad (10)$$

$$\text{Recall} = \frac{\sum_{i=1}^n \frac{TP_i}{TP_i + FN_i}}{n} \quad (11)$$

$$\text{F1score} = \frac{2\text{PrecisionRecall}}{\text{Precision} + \text{Recall}} \quad (12)$$

where N is the number of total samples in n classes.  $TP_i$ ,  $FP_i$ , and  $FN_i$  are true positive, false positive, and false negative for the  $i^{\text{th}}$  class. The pMCI vs. sMCI accuracy can be calculated when





**Fig. 6.** The confusion matrices of the ternary classification. The results acquired using the cross entropy loss (CEL) and the focal loss (FL) on real and generated image data are shown in 1<sup>st</sup> and 2<sup>nd</sup> rows, respectively. CEL& Real represents the multi-classification model that adopts CEL and performs classification on the real images at Y4 (Real). FL& mi-GAN represents the multi-classification model that adopts FL and performs classification on the estimated images at Y4 generated by mi-GAN.

**TABLE VI**

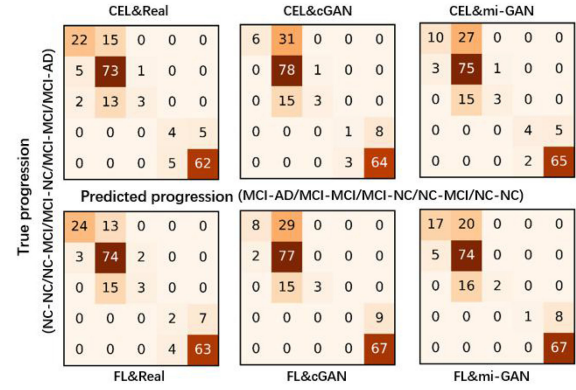
SUMMARY OF TERNARY AND PROGRESSION PREDICT CLASSIFICATION ON DIFFERENT IMAGES WITH DIFFERENT LOSS FUNCTIONS

	ACC (%)	Precision/Recall/F1 score (%)	pMCI vs. sMCI ACC (%)
CEL&Real	78.10	79.14/74.48/76.74 72.02/61.10/62.41	81.90
FL&Real	79.05	82.42/76.29/79.24 68.95/58.29/60.30	84.48
CEL&cGAN	72.38	<b>83.29</b> /61.60/70.83 70.36/47.65/47.90	72.41
CEL&mi-GAN	74.76	79.41/65.60/71.85 75.11/ <b>56.02</b> / <b>58.41</b>	73.28
FL&cGAN	73.81	77.41/63.82/69.97 66.36/47.15/46.66	73.28
FL&mi-GAN	<b>76.67</b>	78.15/ <b>70.78</b> / <b>74.28</b> <b>86.78</b> /52.37/54.06	<b>78.45</b>

Key: In 3<sup>rd</sup> column, 1<sup>st</sup> and 2<sup>nd</sup> rows show the precision, recall and F1 score of the ternary and progression predict classification, respectively. The best results obtained on the generated images are in bold.

$n$  equals 2, i.e., only considering the MCI-AD and MCI-MCI cases.

The confusion matrices of the ternary classification with the FL or CEL on the real or the generated images are shown in Fig. 6, and the classification results are shown in Table VI. Compared with the results achieved by some machine or deep learning based multi-class classifiers, such as 0.56, 0.51 and 0.64 F1 scores of NC, MCI and AD of [35], and 68.8%, 60.2% and 68.9% ternary classification accuracies of [36]–[38], the accuracy (76.67%) achieved by our multi-classification model on the real Y4 images is remarkable. From Table VI, we can observe that the accuracy obtained by using FL is higher than using CEL on the same image data, and the accuracy of the estimated images by mi-GAN is higher than c-GAN. For example, compared with the results using CEL on the estimated images generated by c-GAN, the accuracy, recall and F1 score using FL on the estimated images generated by mi-GAN improve 4.29%, 9.18% and 3.45%, respectively. The classification performance



**Fig. 7.** The confusion matrices of disease progression prediction.

indicates that our mi-GAN can generate more image details of future brain MRI, particularly in predicting subjects who will progress into AD stages in four years, from which the classification model can extract high-level features.

Once the stages of the generated images at Y4 are identified, we further constitute the confusion matrices of the disease progression prediction with FL or CEL on different images, which are shown in Fig. 7, and the progression classification results are shown in Table VI. The disease progression cases in our study include MCI-AD, MCI-MCI, MCI-NC, NC-MCI and NC-NC. No research considers five progression cases in one framework, so we cannot compare the performance of our disease progression prediction framework with others. Compared with the pMCI vs. sMCI accuracy achieved by other sMRI based studies, such as 75%, 75.44%, 72.4% and 74.82% accuracy of [39]–[42], the pMCI vs. sMCI accuracy (i.e., 78.45%) obtained by our proposed disease progression prediction framework is improved by 3.01%.

#### IV. CONCLUSION

This paper studies the disease progression by estimating brain appearance at future time-points and classifying the clinical stage of the estimated brain image. We propose a novel 3D patch based mi-GAN model, conditioned on the brain image and its corresponding multi-information at BL, to estimate the high-quality 3D brain sMRI images at future time-points. Additionally, we construct a 3D DenseNet based multi-classification model to evaluate the quality of the generated images and to further identify what the progression will be. The qualitative and quantitative experiments results demonstrate that the generated images by mi-GAN are acceptable to the multi-classification model with the focal loss, and our proposed disease progression prediction framework is effective. Even though the above results are obtained, the limitations such as the performance in the gray matter distribution prediction and short-term brain image prediction should be improved in the future work.

#### ACKNOWLEDGMENT

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National

Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F.Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California. Open Access Series of Imaging Studies (OASIS) is a longitudinal neuroimaging, clinical, cognitive, and biomarker dataset for normal aging and Alzheimer's Disease that aimed at making neuroimaging datasets freely available to the scientific community, which is funded by P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

## REFERENCES

- [1] "2018 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 14, no. 3, pp. 367–429, 2018.
- [2] C. Patterson, *World Alzheimer Report 2018—The State of the Art of Dementia Research: New Frontiers*. London, U.K.: Alzheimer's Disease International (ADI), 2018.
- [3] A. Bahar-Fuchs, L. Clare, and B. Woods, "Cognitive training and cognitive rehabilitation for persons with mild to moderate dementia of the Alzheimer's or vascular type: A review" (in English), *Alzheimer's Res. Therapy*, vol. 5, no. 4, pp. 35–35, 2013.
- [4] L. Younes, M. Albert, A. Moghekar, A. Soldan, C. Pettigrew, and M. I. Miller, "Identifying changepoints in biomarkers during the preclinical phase of Alzheimer's disease," *Front. Aging Neurosci.*, vol. 11, 2019, Art. no. 74.
- [5] P. Jiang, X. Wang, Q. Li, L. Jin, and S. Li, "Correlation-aware sparse and low-rank constrained multi-task learning for longitudinal analysis of Alzheimer's disease," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1450–1456, Jul. 2019.
- [6] D. P. Veitch *et al.*, "Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's disease neuroimaging initiative," *Alzheimer's Dementia*, vol. 15, pp. 106–152, 2019.
- [7] C. Platero, M. E. López, M. D. Carmen Tobar, M. Yus, and F. J. H. B. M. Maestu, "Discriminating Alzheimer's disease progression using a new hippocampal marker from T1-weighted MRI: The local surface roughness," *Hum. Brain Mapp.*, vol. 40, no. 5, pp. 1666–1676, 2019.
- [8] K. Li, R. O'Brien, M. Lutz, S. Luo, and I. Alzheimer's Disease Neuroimaging, "A prognostic model of Alzheimer's disease relying on multiple longitudinal measures and time-to-event data," *Alzheimer's Dement.*, vol. 14, no. 5, pp. 644–651, May 2018.
- [9] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [10] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552.
- [11] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, "Image synthesis in multi-contrast MRI with conditional generative adversarial networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2375–2388, Oct. 2019.
- [12] Y. Wang *et al.*, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose" (in English), *Neuroimage*, vol. 174, pp. 550–562, Jul. 1, 2018.
- [13] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification" (in English), *Neurocomputing*, vol. 321, pp. 321–331, Dec. 10, 2018.
- [14] D. Nie *et al.*, "Medical image synthesis with context-aware generative adversarial networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2017, pp. 417–425.
- [15] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, and D. Shen, "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 455–463.
- [16] Y. Yan, H. Lee, E. Somer, and V. Grau, "Generation of amyloid PET images via conditional adversarial training for predicting progression to Alzheimer's disease," in *Proc. Int. Workshop Predictive Intell. Med.*, 2018, pp. 26–33.
- [17] D. Ravi, D. C. Alexander, N. P. Oxtoby, and A. s. D. N. Initiative, "Degenerative adversarial neuroimage nets: Generating images that mimic disease progression," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2019, pp. 164–172.
- [18] C. Bowles, R. Gunn, A. Hammers, and D. Rueckert, "Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks," *Med. Imag.: Image Process. Int. Soc. Opt. Photon.*, vol. 10574, 2018, Art. no. 105741K.
- [19] X. Wang, W. Cai, D. Shen, and H. Huang, "Temporal correlation structure learning for MCI conversion prediction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 446–454.
- [20] L. A. Farrer *et al.*, "Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer Disease: A meta-analysis," *JAMA*, vol. 278, no. 16, pp. 1349–1356, 1997.
- [21] S. B. Sando *et al.*, "Risk-reducing effect of education in Alzheimer's disease," *Int. J. Geriatric Psychiatry*, vol. 23, no. 11, pp. 1156–1162, 2008.
- [22] S. J. van der Lee *et al.*, "The effect of APOE and other common genetic variants on the onset of Alzheimer's disease and dementia: A community-based cohort study," *Lancet Neurol.*, vol. 17, no. 5, pp. 434–444, May 2018.
- [23] S. Spasov, L. Passamonti, A. Duggento, P. Lio, N. Toschi, and I. Alzheimer's Disease Neuroimaging, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *Neuroimage*, vol. 189, pp. 276–287, Apr. 1, 2019.
- [24] M. Albert *et al.*, "Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years" (in English), *Brain*, vol. 141, no. 3, pp. 877–887, Mar. 2018.
- [25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2017, pp. 4700–4708.
- [26] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class Alzheimer's disease classification using image and clinical features" (in English), *Biomed. Signal Process. Control*, vol. 43, pp. 64–74, May 2018.
- [27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2017, pp. 1125–1134.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
- [29] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *Proc. Int. Conf. Learning Representations*, 2016.
- [30] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2016, pp. 2536–2544.
- [31] G. Huang, Z. Liu, G. Pleiss, L. V. Der Maaten, and K. Q. Weinberger, "Convolutional networks with dense connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019, p. 1.



- [32] R. Cui and M. Liu, "Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019.
- [33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 2980–2988.
- [34] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2019, pp. 9268–9277.
- [35] L. Brand, H. Wang, H. Huang, S. Risacher, A. Saykin, and L. Shen, "Joint high-order multi-task feature learning to predict the progression of Alzheimer's disease," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 555–562.
- [36] L. Sorensen, M. Nielsen, and A. D. N. Initia, "Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination" (in English), *J. Neurosci. Methods*, vol. 302, pp. 66–74, May 15, 2018.
- [37] V. Wegmayr, S. Aitharaju, and J. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," *SPIE Med. Imag.*, vol. 10575, Art. no. 105751S, 2018.
- [38] A. Khvostikov, K. Aderghal, A. S. Krylov, G. Catheline, and J. Benois-pineau, "3D inception-based CNN with sMRI and MD-DTI data fusion for Alzheimer's disease diagnostics," 2018, *arXiv:1809.03972*.
- [39] I. Beheshti, N. Maikusa, M. Daneshmand, H. Matsuda, H. Demirel, and G. Anbarjafari, "Classification of Alzheimer's disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks," *J. Alzheimer's Dis.*, vol. 60, no. 1, pp. 295–304, 2017.
- [40] D. Lu, K. Popuri, W. Ding, R. Balachandar, and M. F. Beg, "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 5697.
- [41] X. Zhu, H. I. Suk, L. Wang, S. W. Lee, D. Shen, and I. Alzheimer's Disease Neuroimaging, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Med. Image Anal.*, vol. 38, no. 6, pp. 205–214, May 2017.
- [42] H.-I. Suk, S.-W. Lee, D. Shen, and A. S. D. N. Initiative, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Med. Image Anal.*, vol. 37, pp. 101–113, 2017.
- [43] Y. Zhao, B. Ma, D. Zeng, and S. Li, "Long-term prediction of Alzheimer's disease progression with multi-information generative adversarial network" in *Proc. Int. Symp. Image Comput. Digit. Med.*, 2019, Paper P48.
- [44] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, 2009.