Neurocomputing 492 (2022) 353–369

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A novel Alzheimer's disease detection approach using GAN-based brain slice image enhancement

Tian Bai [a], Mingyu Du [a], Lin Zhang [a,*], Lei Ren [a,*], Li Ruan [a], Yuan Yang [a], Guanghao Qian [a], Zihao Meng [a], Li Zhao [a], M. Jamal Deen [b]

[a] School of Automation Science and Electrical Engineering, Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Key Laboratory of Big Data-Based Precision Medicine, Ministry of Industry and Information Technology, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100191, China
[b] McMaster University, Hamilton, ON L8S 4K1, Canada

ARTICLE INFO

Article history:
Received 5 March 2021
Revised 17 November 2021
Accepted 3 April 2022
Available online 7 April 2022
Communicated by Zidong Wang

Keywords:
Alzheimer's disease diagnosis
Deep learning
Generative Adversarial Network
Convolutional Neural Network

ABSTRACT

With the prevalence and the enormous societal consequence on health of Alzheimer's disease (AD), diagnosis of AD and its prodromal form, mild cognitive impairment (MCI) is essential for patient care, and has been a research hotspot in recent years. Existing studies have applied machine learning methods to perform AD early diagnosis by analyzing various biomarkers. However, the difficulty in extracting the low-dimensional high-level brain features that accurately reflect main AD-related variations of anatomical brain structures becomes a bottleneck of the diagnosis performance in most of the existing researches. To overcome this bottleneck, this paper proposes a novel three-component adversarial network-based AD detection method (brain slice generative adversarial network for Alzheimer's disease detection, BSGAN-ADD) to predict the disease category. BSGAN-ADD combines generative adversarial network (GAN)-based brain slice image enhancement and deep convolutional neural network (CNN)-based AD detection. In BSGAN-ADD, under the restriction of the discriminator, the generator learns to integrate the disease category feedbacks from classifier into 2D-brain slice image reconstruction process for image enhancement in the training phase. In the prediction phase, the stacked CNN layers in the generator are used to extract high-level brain features from category-enhanced 2D-brain slice images. And the classifier receives the extracted brain features to output the posterior probabilities of diseased states (Normal, AD and MCI). Experimental results on two real-world datasets (Alzheimer's disease neuroimaging initiative, ANDI, Open Access Series of Imaging Studies OASIS) demonstrate that the new feature extraction process used in BSGAN-ADD can extract more representative high-level brain features to achieve a significant diagnosis performance gain compared with several typical methods.

1. Introduction

The Alzheimer's disease (AD), a neurological, progressive and irreversible brain disorder disease, which is the most common case of dementia mostly occurring in the late life and causes the death of nerve cells, thus affecting memory and thinking skills [1]. As a result of the increasing age of societies, the number of confirmed AD patients is expected to rise dramatically. And the worldwide prevalence of AD will double for the next 20 years, so that 1.2% of the global population will have the AD by 2046 [2].

Generally, the development of AD is divided into different stages according to deterioration degree of cognitive competence. In the initial stage, patients show memory impairment, socially awkward and apathy. Then as disease progresses, they gradually suffer incontinence and lose their autonomous behavior capacity. Ultimately, patients die of complications. It is easy to diagnose the patients at the end-stage of AD, while the stage of mild to moderate remains difficult to be distinguished. To prolong patients' lifespan and improve their lives' quality, the early diagnosis of AD identifies the high risk of progression, and allows patients to take preventive measures before irreversible brain damage occurs. As it has been confirmed that mild cognitive impairment (MCI) has a high risk of progression to AD [3], early diagnosis of Alzheimer's disease that is primarily associated to the detection of AD and MCI, and plays a significant role in patient care.

Recently, many studies have applied machine learning methods for computer-aided-diagnosis (CAD) of AD, and the accuracy can be increased than traditional doctor diagnosis. In these works, the diagnosis of AD can be modeled to be a multiclass classification problem [4–6]. But, a bottleneck of the diagnosis performance was shown in most of the existing researches, mainly due to the

* Corresponding authors.
E-mail addresses: zhanglin@buaa.edu.cn (L. Zhang), renlei@buaa.edu.cn (L. Ren).

https://doi.org/10.1016/j.neucom.2022.04.012
0925-2312/

expressiveness limitations of the chosen learning models. In this study, we design a novel deep learning-based model to overcome the bottleneck and aid the diagnosis of AD and its prodromal stage MCI by detecting the structural magnetic resonance imaging (sMRI) data of AD's brains, MCI's brains and healthy brains.

The two main contributions of this work can be summarized as follows:

(1) A novel three-component adversarial Alzheimer's disease detection (ADD) model (Called brain slice (BS) generative adversarial network for Alzheimer's disease detection, BSGAN-ADD) is proposed for early diagnosis of AD. The proposed ADD model can be regarded as a combination of the GAN-based brain slice image enhancement model and the deep convolutional neural network (DCNN)-based ADD model. In BSGAN-ADD, the pre-processed 2D-brain slice images (2D-brain feature images) are selected from 3D structural MRI (sMRI) as input, and the generator, discriminator and classifier compete in a three-player minimax game. And the posterior probability of disease category (Normal, AD, MCI) $P(c|Input)$ are outputs of the classifier.
(2) Extensive experiments on two real-world datasets (ANDI, OASIS) are carried out to verify the effectiveness of the proposed AD detection approach. And the proposed model was compared with other GAN-based models to indicate the classification performance. The experimental results demonstrate that BSGAN-ADD has better prediction performance than several state-of-the-art AD detection approaches.

2. Related Works

Existing methods can be divided into two categories. One is clinical methods and another is computer-aided methods.

2.1. Clinical methods

The diagnosis of prodromal Alzheimer's Disease (AD) remains difficult on purely clinical grounds [7]. Further, the accuracy of clinical diagnosis may be lower for patients with prodromal AD, besides it could be even lower in primary or secondary care settings than in the specialized AD centers [8]. The investigation by Beach et al. [9] showed that sensitivity of AD diagnostic ranged from 70.9% to 87.3% and specificity ranged from 44.3% to 70.8%. Sensitivity usually depends on the specific histopathologic diagnosis criteria used. Sacuiu et al. [10] made some research about chronic depressive symptomatology (chrDS). They found the relevance between chrDS and additional risk factor for conversion to dementia in MCI and chrDS wasn't a typical prodromal AD symptomatology. In [11], a timely diagnosis method was recommended to help early diagnosing prodromal AD. Different from early diagnosis, timely diagnosis appeals an earlier diagnosis and medical intervention. It is perceived that paying attention from the first time of symptom discovered can be helpful in early stage of dementia.

2.2. Computer-aided methods

In recent years, computer-aided methods have shown its promising results in AD diagnosis. Compared with traditional clinical diagnosis, computer-aided diagnosis needs less time and depends less on the experience of professional neuralimage doctors meanwhile provides better accuracy. The accuracy of this computer-based techniques improved with the development of machine learning.

In the traditional machine learning methods, unsupervised methods can be used to diagnose Alzheimer's Disease. Sun et al. [12] proposed a novel support vector machine (SVM)-based learning method integrating spatial-anatomical information for the classification of Alzheimer's disease (AD) and received a good classification performance between Normal and AD. A novel framework for estimating the hyper-connectivity network of brain functions was proposed by Jie et al. [13]. They extracted three sets of brain-region specific features from the connectivity hypernetworks, and used a multi-kernel SVM for classification. It helps discover disease-related biomarkers for disease diagnosis. To represent multivariate brain MRI features, Liu, X et al. [14] imported an unsupervised algorithm (locally linear embedding, LLE) into their classification. In their experience, classification with LLE performed better than ordinary classifications. In addition, supervised approaches are also important traditional machine learning methods to diagnose Alzheimer's Disease. Considering a high-level information inherent in the observations, Zhu et al. [15] designed a new loss function combined with a group lasso for joint sparse feature selection in the joint regression and classification (JRC) problem, and enhanced the performance of the regression and classification in AD/MCI diagnosis. Huang et al. [16] proposed a nonlinear supervised sparse regression-based random forest (RF) framework to avoid multiple limitations in the current models. A soft-split technique was utilized to assign probabilistic paths to a test sample in RF and the prediction accuracy was improved.

SVM, JRC and RF showed its ability in AD/Normal classification, however the difficulty in obtaining the brain representations that accurately reflect AD-related variations of anatomical brain structures becomes a main problem of the AD diagnosis.

Compared with traditional machine learning algorithms, neural networks can reach a higher accuracy in image recognition region, therefore many scholars fuse it with AD diagnosis and showed a promising effect.

Convolutional neural network (CNN) is most commonly used by researchers. Some studies [4,17–21], to obtain high accuracy under the circumstance of less labeled training samples, combined sparse auto-encoders with 3D convolutional neural networks (3D-CNNs), and performed better than 2D-CNNs on slices in their experiment, which shows a good potential of 3D approach to capture local 3D patterns. Cheng et al. [6] constructed a multiple 3D-CNNs and generic features from imaging data for classification can be automatically learned in their method. Korolev et al. [22] proposed two different 3D-CNNs architectures for brain MRI classification to overcome the problem about complex multistep pipelines for handcrafted feature generation and feature extraction from the data. Choi and Jin [23] applied a CNN based method to the prediction of cognitive decline, and they showed the strong correlation between CNN-extracted biomarker and future cognitive decline. Khvostikov et al. [24] designed a new 3D Inception-based convolutional neural network architecture and this CNN was demonstrated better compared with convolutional AlexNet-based network. Feng et al. [25] applied a spherical CNNs based framework on human cortex data from ADNI and they showed the potential of spherical CNNs in human cortex modeling and performing AD diagnosis. Aderghal et al. [26] fused 2D-CNNs from different direction in each brain projection, which not working with the whole brain volume, and the fusion methods showed better performance. Liu et al. [27] proposed a multi-channel learning framework to perform AD/MCI classification tasks. Specifically, a new data-driven method was applied to locate diseaseassociated image areas in the MRIs. And then, those sub-MRI scans were fed into a multi-channel CNN-based classifier for joint classification.

Deep convolution network (DNN) or recurrent neural network (RNN) can be used to extract image features. Cheng and Liu [28] proposed a combining convolutional and recurrent neural networks for PET images, and in their method, RNN was added to learn the features of sequential images and modeled the 3D structure of

medical images for segmentation and classification. Lu et al. [29] utilized a multi-modal and multiscale deep neural network. Features were extracted at coarse-to-fine structural scales in the multi-scale approach and the problem about loss of discriminative information like simple approach can be prevented.

In the field of transfer learning, Hon and Khan's study [30] showed that pre-trained weights with intelligently picked training data generalized very well for AD diagnosis even though the architectures were trained on a different domain.

And in the field of unsupervised learning, Li et al. [31] utilized a restricted Boltzmann machine (RBM) with dropout technique to prevent the overfitting by weight co-adaptation. They found a 5.9% improvement of accuracy compared with classical deep learning methods but the accuracy was still not satisfied.

Summarizing the existing AD detection methods, there are two main problems: 1) **Low accuracy and demanding expertise.** Traditional clinical diagnosis needs time and professional doctors with abundant neuroimaging experience. Systematic diagnosis process is not yet complete so the diagnosis accuracy is unsatisfied, which may leave the disease develop worse due to failure of recognizing the prodromal. 2) **Difficulty in representative brain feature extraction.** Computer-aided diagnosis becomes a prevail nowadays, owing to its fast and accurate diagnosis. The deep learning-based methods showed their great performance in this domain, however, the difficulty in extracting the high-level brain features that accurately reflect main AD-related variations of anatomical brain structures becomes a bottleneck of the diagnosis performance in these researches. Some researchers utilized Autoencoder (AE)-pre-trained CNNs to performs the transfer learning by automatically extracting discriminative AD features through an unsupervised learning way [18,19,6]. Unfortunately, the improvements of performance are limited mainly due to the back-propagation process of reconstruction loss cannot incorporate disease category information. More importantly, there still have a low accuracy of three-category classification in AD/MCI/Normal.

To overcome the above two limitations, the BSGAN-ADD is proposed for early diagnosis of AD, which learns to perform a novel brain feature extraction process based on an adversarial strategy. In BSGAN-ADD, different from the autoencoder-based image generation approach, the disease category feedbacks from classifier, influences the way of brain slice image reconstruction for generating brain slice images with more obvious category information. Then, it uses deep CNNs to extract high-level brain feature from the category-enhanced images. Such alteration of high-level brain feature extraction process allows generator can offer more representative features that are more helpful for classifier to make more accurate diagnosis of diseased state.

3. Methods

3.1. Dataset Introduction

The cases in this study were collected at baseline from the Alzheimer's disease neuroimaging initiative (ADNI) database with tests results and imaging outcomes comprising at least two years of follow-up. In this work, the efficiency of the proposed Alzheimer's disease detection model (BSGAN-ADD) was validated on 818 structural MRI (sMRI) samples in Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The detail demographic information is provided in Table 1.

In addition to experimented on ADNI dataset, we also tested this model on Open Access Series of Imaging Studies (OASIS) dataset to prove the universality. This dataset contains a cross-sectional collection of 416 subjects. These subjects are aged from 18 to 96 years. For every subject, 3 or 4 individual T1-weighted MRI scans are included that were acquired in single scan sessions. 100 out of the 416 subjects that are aged over 60 years have been diagnosed with Alzheimer's disease (AD), ranging from very mild to moderate level.

The proposed model was applied to four specific classification tasks: three binary ones (Normal vs AD, Normal vs MCI, AD vs MCI) and one ternary classification (Normal vs AD vs MCI). Classification performances were evaluated for each task by ten-fold cross-validation. Pytorch, one of the popular deep learning libraries in Python, was used to train the proposed Alzheimer's disease detection model.

3.2. Data pre-processing

To maximize the accuracy and generality of BSGAN-ADD, a general data pre-processing was used to obtain the brain gray matter (GM) image from sMRI data. The FSL-VBM [32] and ANTsR were used as pre-processing tools to analyze sMRI data. ANTsR, a high-dimensional brain mapping library in R language, was used to capture the GM from sMRI data. FMRIB's software library (FSL) provided a viable way to compare the GMs of different samples in the research group on a voxel-wise basis, and transform them into a standard space. It is used to eliminate the influence of brain size of different individuals on the classification model so as to obtain better classification accuracy and avoid over-fitting. Through pre-processing, 2D png images are obtained from the sMRI data of nii format. The size of each image is $3 \times 32 \times 32$. The images pre-processed from a sMRI data are shown in Fig. 1.

The 2D-brain feature images were divided into a training set and a test set. The random sampling method was applied to get 90% of the patient's brain feature image as training data, while the rest 10% was for testing. However, with the limited image data for training a neural network, it was easy to cause problems of underfitting or overfitting. To enlarge the amount of training data, data augmentation (DA) was performed in this work. Mirror flip was the first DA method to apply in processed images. The zero-mean Gaussian noise with variance of 0.005 was also used to generate new training images. The final DA method was brightness change. Image brightness was set from 90% to 110% in steps of 10%. The training dataset contained 58,800 2D-brain feature images after DA, as shown in Table 1.

3.3. Approach overview

In this section, the proposed brain slice (BS) GAN-based Alzheimer's disease detection (called BSGAN-ADD) model is introduced.

Table 1
Number of samples and images for each class before and after augmentation.

	AD	MCI	Normal
Samples	244	303	271
Samples for train	219	272	244
Samples for test	25	31	27
2D-brain slice images for train	3504	4352	3904
2D-brain slice images for test	400	496	432
2D-brain slice images for train after augmentation	17520	21760	19520

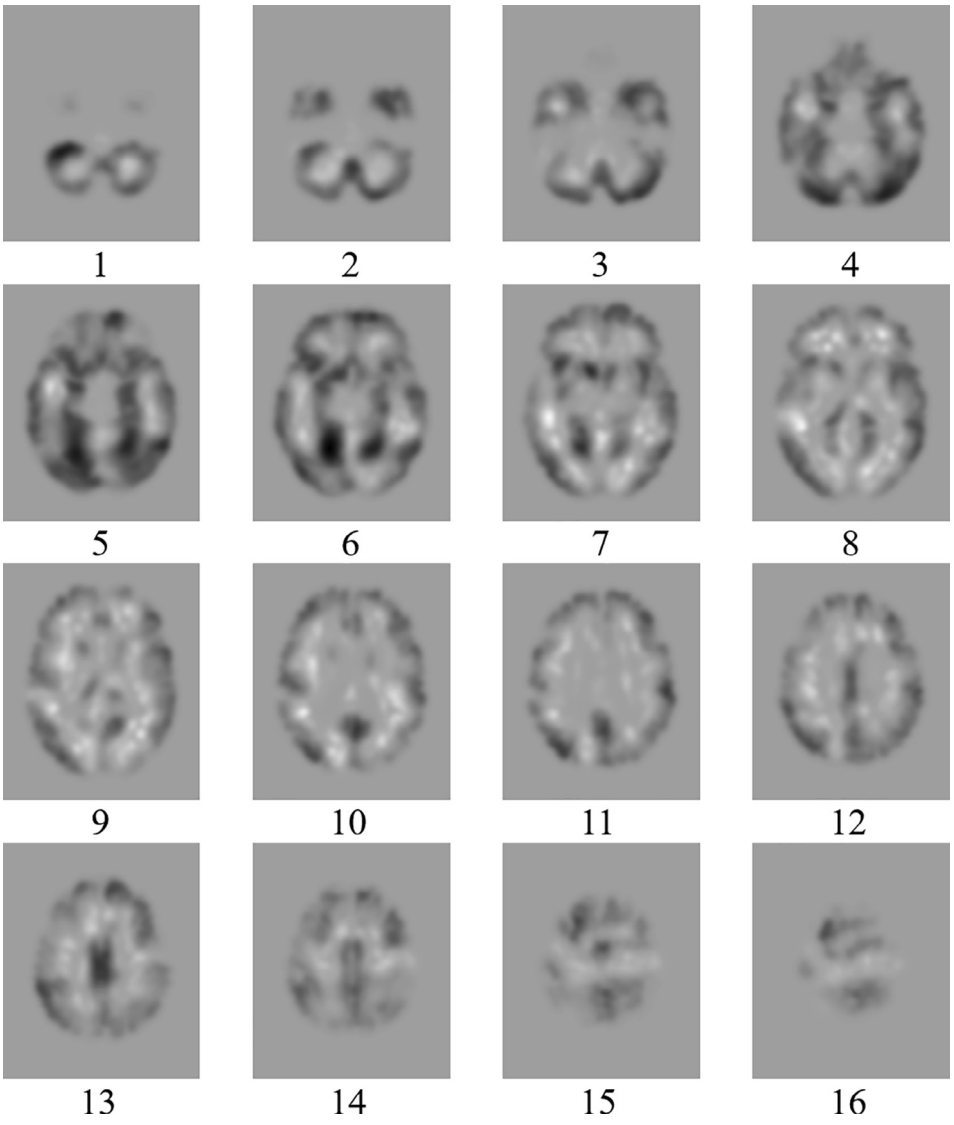

Fig. 1. The 2D-brain feature image.

BSGAN-ADD is a three-component adversarial detection model, which aims to perform the diagnosis of AD and its prodromal stage (Mild Cognitive Impairment, MCI). As shown in Fig. 2, BSGAN-ADD consists of three main components: 1) the encoder-decoder-encoder generator component; 2) the discriminator component; 3) and the classifier component.

The functions of these three components are now described. The generator component, under the constraint of specific brain slice locations, generates the pseudo 2D-brain feature images based on the pixel-wise distribution of the input images and the disease category Normal, AD and MCI. Moreover, the generator component learns the low-dimensional representations of the generated brain feature images by stacked CNN layers for obtaining the AD-related high-level brain features of the input brain feature images. The discriminator component is responsible for distinguishing the real and the pseudo 2D-brain feature images and output discriminant result y'_D. The classifier component constructs the mapping from the extracted high-level brain features output by the generator component to 3-dimensional posterior probability y'_C of one of the diseased states Normal, AD or MCI.

Next, we describe BSGAN-ADD components from the perspective of 2D-brain feature image enhancements for Alzheimer's disease (AD) detection. The discriminator component requires that the generator component generates the realistic pseudo images which have same distribution with the input images. At the same time, the classifier component influences the generation process of the pseudo 2D-brain feature image by integrating the disease category information into the generator component by means of supervised learning, thereby completing category enhancements of the input 2D-brain feature images. After that, the classifier component gives the disease category posterior probability according to the high-level brain features extracted from the category-enhanced brain feature images (the generated brain feature images).

As shown in Fig. 3, the generator component in BSGAN-ADD has three parts, an encoder network *Encoder1*, a decoder network *Decoder* and another encoder network *Encoder2*. In this work, because the pixel-wise distributions of 2D-brain slice images (2D-brain feature images) at different locations of Z axis are not the same, the brain slice location number of a 2D-brain feature image is coded

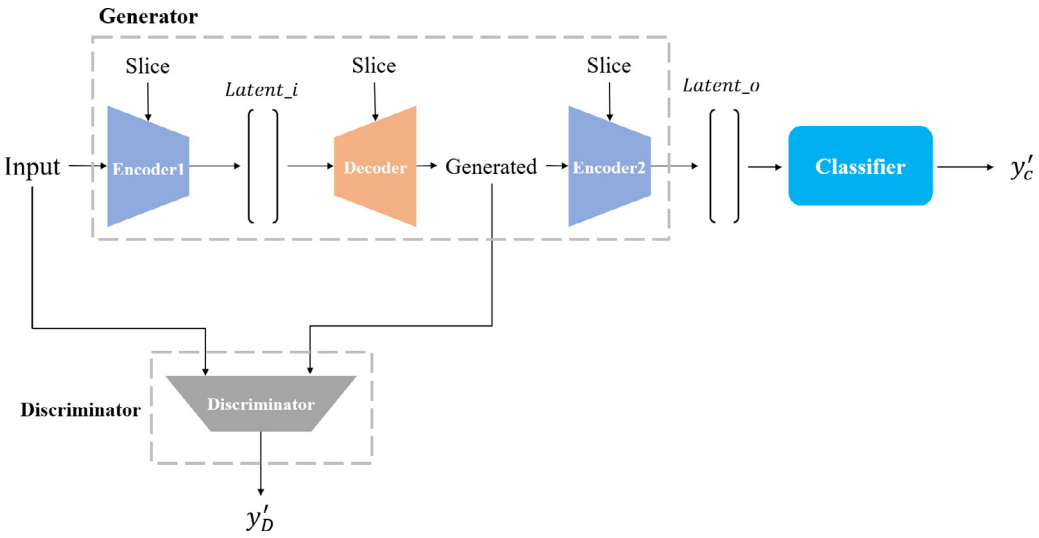

Fig. 2. The overall framework of the BSGAN-ADD.

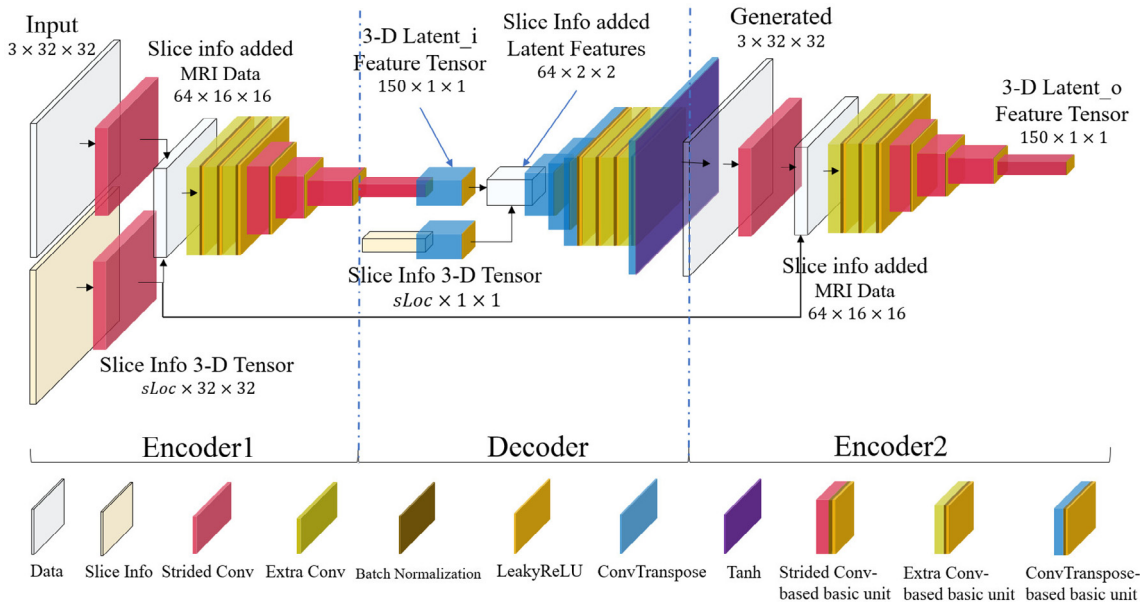

Fig. 3. The structure of the generator component in the BSGAN-ADD. A combination of convolutional layer, batch-norm and leakyReLU is called a basic unit.

into 3-dimensional slice matrix and fed into *Encoder1*, *Decoder* and *Encoder2*. This is to map the 2D-brain feature images to low-dimensional latent feature vectors according to their locations on the Z axis. For *Encoder1* network, the input 2D-brain feature images (*Input*) are transformed into the input latent feature vectors (*Latent_i*) through a learned distribution $P(Latent\_i|Input, Slice)$. And then, *Decoder* reconstructs the input images from the corresponding *Latent_i* to output the generated 2D-brain feature images (*Generated*) based on a learned distribution $P(Generated|Latent\_i, Slice)$. Finally, *Generated* are mapped to the output latent feature vectors (*Latent_o*) through *Encoder2*. Similar to the traditional GAN, the discriminator component takes real 2D-brain feature images and pseudo 2D-brain feature images generated by the generator component as input simultaneously, and learns for distinguishing between real and generated images. And the classifier component measures the posterior probability $P(c|Input)$ of diseased categories (Normal, AD, MCI) based on the output latent feature vectors (*Latent_o*) extracted by generator from the input 2D-brain feature images (*Input*). As shown by the dotted line in Fig. 5, for each of the three components in BSGAN-ADD, the loss values are used in back propagation process to update its parameters depending on the superposition of losses from itself and the other two networks, thus constituting a three-component adversarial model. And the parameter update strategy of each component will be discussed in detail in Section 3.5.

3.4. BSGAN-ADD pipeline

(1) Generator Component:

Fig. 3 illustrates the structure overview of the generator component, which contains two encoder networks and a decoder network.

Inspired by the deep convolutional GAN (DCGAN) [33], the convolutional layer followed by batch-norm and leakyReLU is regarded as a basic unit (BU) to extract brain image features. Three kinds of convolutional layers (Strided Conv, Extra Conv and

ConvTranspose) are used in our model. Strided convolution (SC) layer is non-pooling strided convolution layer whose shape of the output is half of the input. Extra convolution (EC) layer applies image padding to produce a set of feature images that have same shape as the input. And ConvTranspose (FC) is fractional-strided convolutional layer to reconstruct the input 2D-brain feature images (see Fig. 6).

The encoder network *Encoder1* takes 3-channel 32×32 2D-brain feature images (*Input*) and 3D-slice info matrix ($sLoc \times 32 \times 32$-dimensional matrix *Slice Info*, and 2D-matrix of the corresponding slice location in *Slice Info* is filled with one) as input, and feeds them into two separated SC layers. The outputs of two SC layers are combined to construct slice info-added MRI data, so that the generator component can effectively extract the high-level brain features from 2D-brain feature images based on different brain slice locations. And then, the encoder network integrates slice info-added MRI data to latent brain features vector *Latent_i* through several ECs and SC-based BUs.

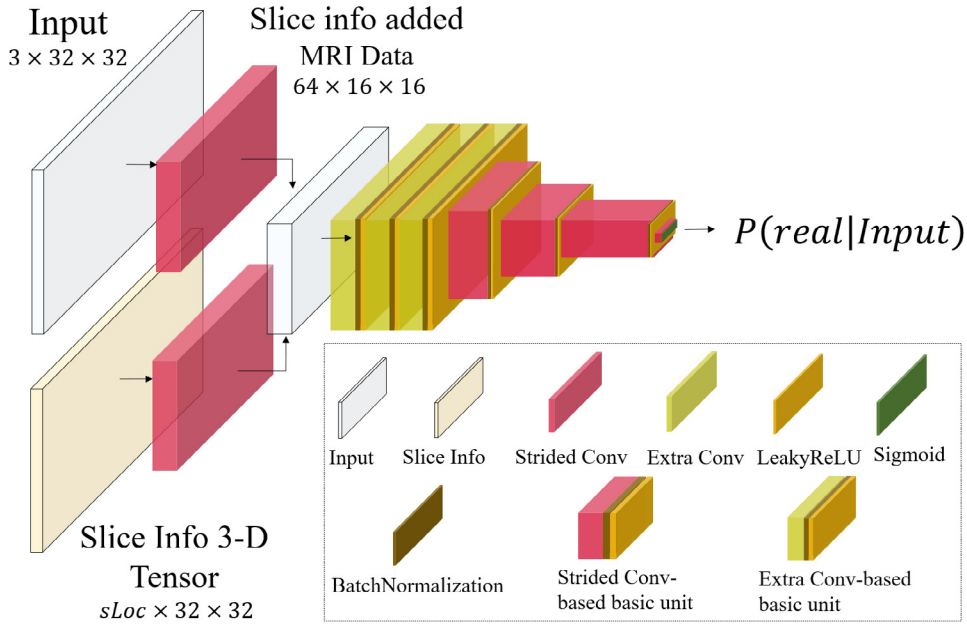

Fig. 4. The structure of the discriminator component in the BSGAN-ADD..

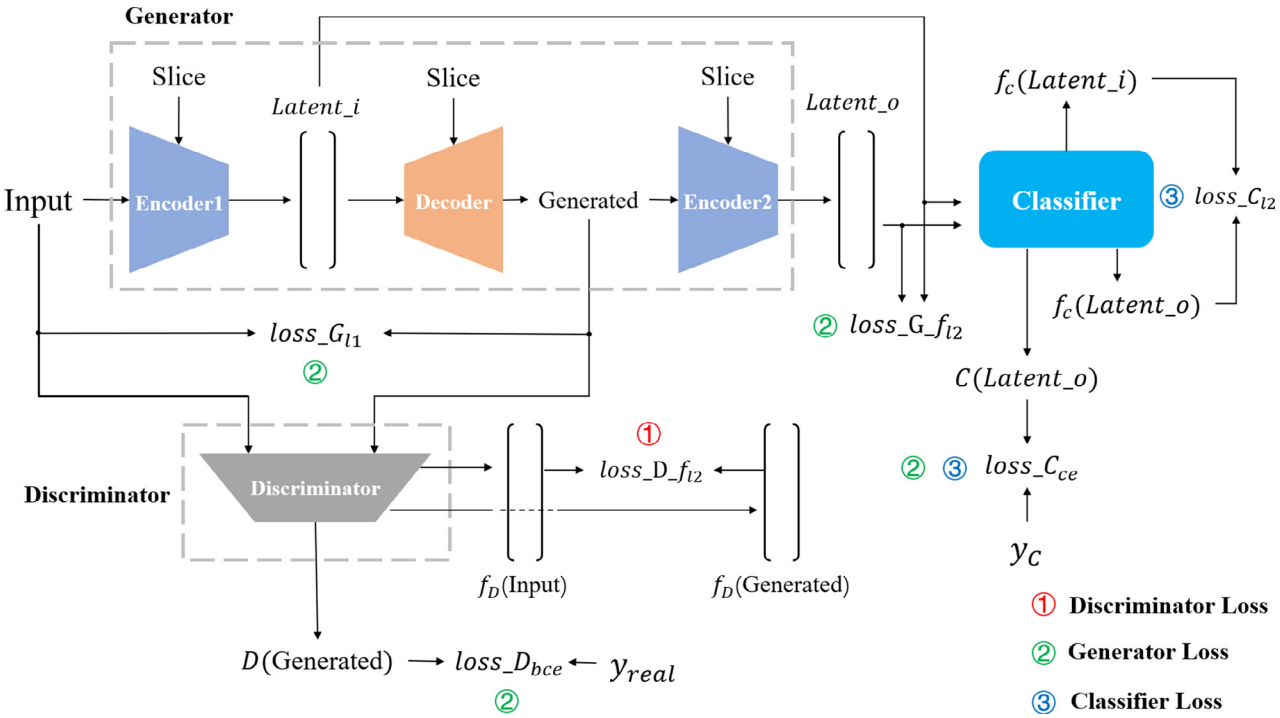

Fig. 5. The loss value structure of the BSGAN-ADD.

Decoder network *Decoder* adopts the architecture of DCGAN's generator, using fractional-strided convolutional layers (FC)-based BUs with a Tanh layer at the end. Similar to the encoder network, the decoder network also has a slice information integration module that effectively upscales the *Latent_i* to reconstruct the input 2D-brain feature image (*Input*) as the generated 2D-brain feature images (*Generated*). Another encoder network *Encoder2* extracts the output latent feature vectors (*Latent_o*) from the generated 2D-brain feature images to represent the input 2D-brain feature images.

(2) Discriminator Component:

Fig. 4 illustrates the structure overview of the discriminator component. The feature extraction layers (the intermediate layers) of the discriminator are similar to the encoder network in the generator component. The difference is that the discriminator component maps the input 3-channel 2D-brain feature image with the corresponding slice location matrix to a one-dimensional vector, and uses the Sigmoid function to output the posterior probability $P(real|Input)$to classify the real 2D-brain feature images and pseudo 2D-brain feature images.

(3) Classifier Component:

In this work, a DNN-based classifier component is proposed to generate the posterior probability $P(c|Input)$of diseased categories (Normal, AD, MCI). Specifically, the high-level brain feature vectors (*Latent_o*) output by the generator component taking 2D-brain feature images as input, are fed into a DNN to output the 3-dimensional vector. Each element of the output 3-dimensional vector represents the possibilities of belonging to Normal, AD and MCI, respectively.

3.5. BSGAN-ADD training strategy

As shown in Fig. 5, the loss values of each component in BSGAN-ADD depends on itself and the other two components. Each component updates the values of its network parameters by back propagation process of its losses. And all parameters and functions in loss calculations are shown in Table 2.

(1) Discriminator Component:

In traditional GANs [34], the discriminator network D tries to distinguish real data from pseudo data generated by the generator network G. Formally, the discriminator network minimizes the loss function $\mathscr{L}_D$ to formulate the objective mentioned above.

$$\mathscr{L}_D = -\mathbb{E}_{x\sim P_r}[\log D(x)] - \mathbb{E}_{z\sim P_z}[\log(1 - D(G(z)))] \tag{1}$$

In this work, inspired by the current trend within the new GAN training approach [35], the feature matching loss is used for learning discriminator parameters and to reduce the instability of training process. Formally, it assumes that $f_D$is a function that represents the intermediate layers of the discriminator component to output the feature representation for a given input image x drawn from the specific data distribution. The feature matching calculates the L_2distance between the feature representations of the real input and the generated images. Therefore, the loss function of the discriminator component in this work is defined as:

$$\mathscr{L}_D = \mathbb{E}_{x\sim P_r}||f_D(x, slice) - f_D(G(x), slice)||_2 \tag{2}$$

The parameter update strategy of the discriminator component is given as:

$$Param_D \leftarrow -\nabla \mathscr{L}_D, \tag{3}$$

where ∇ denotes the gradient values of the $\mathscr{L}_D$to the corresponding parameters (such as the weights and biases of the convolution layers in the discriminator network).

(2) Generator Component:

To optimize the generator component, the following four losses are combined.

Adversarial Loss. Similar to the traditional GAN model, the binary cross entropy (BCE) function is used for generator-

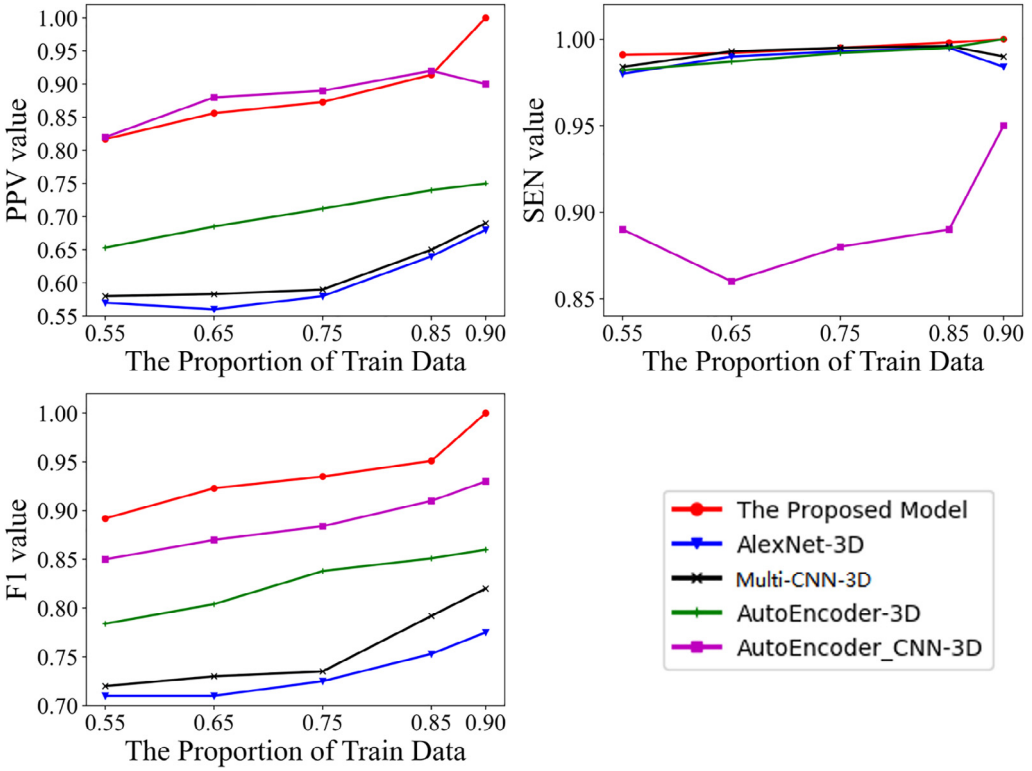

Fig. 6. The value changes of PPV, SEN and F1 on AD group of testing data when the proportion of training data varied from 55% to 90%.

Table 2
Description of parameters and functions.

Name	Description	Name	Description
$Param_D$	Parameters of discriminator network	y'_C	outputs of the classifier component
$Param_G$	Parameters of generator network	y_C	the true category labels of samples
$Param_C$	Parameters of classifier network	M	the number of classes of samples
Pr	the specific data distribution	y'_D	the discriminator output vector
$slice$	the slice locations of 2D-brain feature images	y_D	the label filled with 1
$\mathscr{L}_D$	total loss of *Discriminator*()	n	the number of dimensions of $y'_D$and y_D
$\mathscr{L}_G$	total loss of *Generator*()	$\mathscr{L}_{D\_bce}$	adversarial loss
$\mathscr{L}_C$	total loss of *Classifier*()	$\mathscr{L}_{G\_l1}$	reconstruction loss
$D()$	the discriminator function	$\mathscr{L}_{G\_l2}$	encoder loss
$G()$	the generator function	$\mathscr{L}_{C\_ce}$	classification loss
$C()$	the classifier function	$\mathscr{L}_{C\_l2}$	feature matching loss
$G_{E1}()$	the first encoder network *Encoder*1	ω_{bce}	the weight of $\mathscr{L}_{D\_bce}$
$G_{E2}()$	the second encoder network *Encoder*2	ω_{rec}	the weight of $\mathscr{L}_{G\_l1}$
$f_D()$	the intermediate layers of the discriminator component	ω_{en}	the weight of $\mathscr{L}_{G\_l2}$
$f_C()$	the intermediate layers of the classifier component	ω_{cla}	the weight of $\mathscr{L}_{C\_ce}$
$BCE()$	the binary cross entropy (BCE) function	ω_{fm}	the weight of $\mathscr{L}_{C\_l2}$

discriminator adversarial learning in the case of fixed network parameters of the discriminator component. Formally, the BCE function is used to calculate the distance $\mathscr{L}_{D\_bce}$ between the discriminator output vector (y'_D) determined by generated images and a real label vector (y_D, filling real label 1) that is consistent with the size of y'_D. The parameter values of the generator component are updated by minimizing the $\mathscr{L}_{D_bce}$to get the generated images that has same distribution of the real input images. The calculation process of $\mathscr{L}_{D_bce}$is defined as,

$$\begin{aligned} y'_D &= D(G(x), slice),\\ BCE(y'_D, y_D) &= -\frac{1}{n}\sum_i y_D^i * \log\left(y_D'^i\right)\\ &\quad + \left(1 - y_D^i\right) * \log\left(1 - y_D'^i\right),\\ \mathscr{L}_{D\_bce} &= \mathbb{E}_{x\sim P_r}\left[BCE(y'_D, y_D)\right] \end{aligned} \tag{4}$$

Reconstruction Loss. According to the work in [36], penalizing the generator of GAN by measuring the distribution differences between the input and the generated images can effectively improve the quality of generated images. Moreover, it was shown that the use of the L_1distance yields less blurry images than L_2. Hence, the L_1distance between the input 2D-brain feature images and the generated images, as a reconstruction loss $\mathscr{L}_{G\_l1}$, is used to penalize the generator component. $\mathscr{L}_{G_l1}$is defined as,

$$\mathscr{L}_{G\_l1} = \mathbb{E}_{x\sim P_r}||x, G(x)||_1, \tag{5}$$

Encoder Loss. As discussed above, this paper uses GAN-based 2D-brain feature image reconstruction for adding additional category features to the original 2D-brain feature images, and uses deep CNN networks to extract high-level brain features from the generated brain feature images to complete the classification task of the diseased state. To improve how *Encoder2* learns to extract brain features from the generated images for representing the input images, an additional encoder loss ($\mathscr{L}_{G\_l2}$) is employed to minimize the distance between the bottleneck features of the input (*Latent_i*) and the extracted features of the generated image (*Latent_o*). And $\mathscr{L}_{G_l2}$is defined as:

$$\begin{aligned} &Laten\_i = G_{E1}(x, slice),\\ &Latent\_o = G_{E2}(G(x)),\\ &\mathscr{L}_{G\_l2} = \mathbb{E}_{x\sim P_r}||Latent\_i, laten\_o||_2, \end{aligned} \tag{6}$$

Classification loss. This classification loss $\mathscr{L}_{C\_ce}$ is based on the current classifier parameters, which is obtained by comparing the outputs of the classifier component (y'_C) taking *Latent_o* as input with the true category labels (y_C) of the samples. The $\mathscr{L}_{C\_ce}$ is calculated by the multi-class cross entropy function:

$$\begin{aligned} &y'_C = C(Latent\_o)\\ &\mathscr{L}_{C\_ce} = -\sum_{i=0}^{M-1} y_C^i * \log\left(y_C'^i\right), \end{aligned} \tag{7}$$

where $i$is the i-th dimension of the $y_C$and y'_C.

Overall, the loss value $\mathscr{L}_G$of the generator component becomes the following:

$$\mathscr{L}_G = \omega_{bce} * \mathscr{L}_{D\_bce} + \omega_{rec} * \mathscr{L}_{G\_l1} + \omega_{en} * \mathscr{L}_{G\_l2} + \omega_{cla} * \mathscr{L}_{C\_ce}, \tag{8}$$

where $\omega_{bce}, \omega_{rec}, \omega_{en}$and $\omega_{cla}$are the weighting parameters adjusting the impact of individual losses to the overall loss value.

The parameter update strategy of the generator component is given as following:

$$Param_G \leftarrow -\nabla \mathscr{L}_G \tag{9}$$

where ∇denotes the gradient values of the $\mathscr{L}_G$to the corresponding parameters.

(3) Classifier Component:

In this work, the loss of the classifier component is mainly coming from two parts, the feature matching loss and the classification loss. Each of these losses is now described.

Feature matching loss. We employ a feature matching loss $\mathscr{L}_{C_l2}$to reduce the instability of updating the parameters of the classifier component.

$$\begin{aligned} &Latent\_i = G_{E1}(x, slice),\\ &Latent\_o = G_{E2}(G(x)),\\ &\mathscr{L}_{C\_l2} = \mathbb{E}_{x\sim P_r}||f_c(Latent\_i) - f_c(Latent\_o)||_2, \end{aligned} \tag{10}$$

Classification loss. The classification loss $\mathscr{L}_{C_ce}$of classifier component is based on the supervised learning, which is calculated by the multi-class cross entropy function and is:

$$\begin{aligned} &y'_C = C(Latent\_o)\\ &\mathscr{L}_{C\_ce} = -\sum_{i=0}^{M-1} y_C^i * \log\left(y_C'^i\right) \end{aligned} \tag{11}$$

where $i$is the i-th dimension of the $y_C$and y'_C.

Overall, the loss value $\mathscr{L}_C$of the generator component becomes the following:

$$\mathscr{L}_C = \omega_{cla} * \mathscr{L}_{C\_ce} + \omega_{fm} * \mathscr{L}_{C\_l2}, \tag{12}$$

where $\omega_{cla}, \omega_{fm}$are weighting parameters. The parameter update strategy of the classifier component is given by:

$$Param_C \leftarrow -\nabla \mathscr{L}_C, \tag{13}$$

where ∇ denotes the gradient values of the $\mathscr{L}_C$to the corresponding parameters.

The pseudo code of complete training process is shown blow.

Algorithm 1: Complete training process

Input: $X = \{x_1, x_2, \cdots, x_n\}$: source data; $Y_D = \{y_1, y_2, \cdots, y_n\}$: the labels of real source data; *G, D and C*: BSGAN-ADD model three components; G_{E1} and G_{E2}: *Encoder1* network and *Encoder2* network; θ_G , θ_D *and* θ_C: initial G network, D network and C network parameters; ϵ: Discriminator loss threshold; $f_D$and $f_C$are functions that represents the intermediate layers of the discriminator component and classifier component; ω_{bce}, ω_{rec}, ω_{en}, ω_{cla} and ω_{fm}: weighting parameters.

while $\theta_G$has not converged: **do**
  // Calculate Discriminator loss and update Discriminator
  $\mathscr{L}_D \leftarrow \|f_D(x, slice), f_D(G(x), slice)\|_2$
  $\theta_D \overset{+}{\leftarrow} -\nabla_{\theta_D}(\mathscr{L}_D)$
  // Calculate Generator loss and update Generator
  $y'_D \leftarrow D(G(x), slice)$
  $\mathscr{L}_{D\_bce} \leftarrow BCE(y_D, y'_D)$
  $\mathscr{L}_{G\_l1} \leftarrow \|x, G(x)\|_1$
  $Latent\_i \leftarrow G_{E1}(x, slice)$
  $Latent\_o \leftarrow G_{E2}(G(x))$
  $\mathscr{L}_{G\_l2} \leftarrow \|Latent\_i, Latent\_o\|_2$
  $\mathscr{L}_{C\_ce} \leftarrow -\log(C(Latent\_o))$
  $\theta_G \overset{+}{\leftarrow} -\nabla_{\theta_G}(\omega_{bce}\mathscr{L}_{D\_bce} + \omega_{rec}\mathscr{L}_{G\_l1} + \omega_{en}\mathscr{L}_{G\_l2} + \omega_{cla}\mathscr{L}_{C\_ce})$
  // Calculate Classifier loss and update Classifier
  $\mathscr{L}_{C\_l2} \leftarrow \|f_C(Latent\_i), f_C(Latent\_o)\|_2$
  $\theta_C \overset{+}{\leftarrow} -\nabla_{\theta_C}(\omega_{fm}\mathscr{L}_{C\_l2} + \omega_{cla}\mathscr{L}_{C\_ce})$
  if $\mathscr{L}_D < \epsilon$ **then** Initialize $D()$
  end if
end while

4. Experiments and Discussions

4.1. Evaluation metrics

The standard metrics were used to evaluate AD detection performance. These metrics are defined using accuracy (ACC), sensitivity (SEN) or recall, positive predictive value (PPV) or precision and F1-score.

$$\begin{aligned} ACC &= \tfrac{TP+TN}{TP+TN+FP+FN}, \\ SEN &= \tfrac{TP}{TP+FN}, \\ PPV &= \tfrac{TP}{TP+FP}, \\ F1 &= 2 * \tfrac{SEN*PPV}{SEN+PPV}, \end{aligned} \tag{14}$$

where TP is the number of true positive, and TN is the number of false positive. Similarly, let FP and FN denote true negative and false negative respectively. In addition, the classification performance was evaluated by the area under the receiver operating characteristic (ROC) curve (AUC). The abscissa of the ROC curve is false positive rate (FPR), and the ordinate is true positive rate (TPR).

$$\begin{aligned} TPR &= \tfrac{TP}{TP+FN}, \\ FPR &= \tfrac{FP}{FP+TN}, \end{aligned} \tag{15}$$

4.2. Performance comparison

(1) Non GAN-based Model Performance Comparison.

We evaluated the detection performance (classification accuracy) of the proposed Alzheimer's disease detection model (BSGAN-ADD) and several typical Alzheimer's disease detection models for each specific classification task. The experiment results are shown in Table 3. Through the experiments on two datasets, our BSGAN-ADD exhibits improvements over all competing methods on four different classification tasks (NL/AD/MCI classification task, NL/AD classification task, NL/MCI classification task and AD/MCI classification task). Moreover, the deep learning-based models (Li et al. [37], Hosseini-Asl et al. [19], Suk et al. [17], Wang et al. [38], Feng et al.[39], Emtiaz et al.[40], our BSGAN-ADD) have better performance than traditional classification models (Magnin et al. [41], Ben Ahmed et al. [42]) because of their good model expressiveness brought by deep structures. In deep learning methods, according to compare the parameters of model, the computational cost of proposed model is also at a low level.

Meanwhile, BSGAN-ADD has the highest accuracy in the AD/MCI and NL/MCI classification tasks, which verifies our AD detection model can identify MCI samples, thus to perform the early diagnosis of AD more effectively.

The proposed BSGAN-ADD model was further compared with four neural network-based methods (AlexNet3D [24], multi-CNN-3D [37], AutoEncoder-3D [17], AutoEncoder CNN-3D [19]) with respect to different proportion of training data. Specifically, we randomly chose 90% of samples as the training set (varied from 55% to 90% with a step size of 10%) and kept the remaining 10% for testing in each round. The changes of different evaluation metric values are shown in Fig. 7–9. Specifically, Fig. 7–8, show the value changes of PPV, SEN and F1 on Normal, AD, MCI group of testing data respectively. Fig. 9 shows the value changes of ACC and training loss of different AD detection methods.

It was shown (for example, Table 3) that the proposed AD detection model (BSGAN-ADD) outperformed the other detection methods on the third classification task (Normal vs AD vs MCI) with any percentage of the training data. It can also be found that as the proportion of training data increased, the proposed model improved the classification performance. Furthermore, the experimental results showed that the proposed model still had high classification accuracy with a small number of training samples.

(2) GAN-based Model Performance Comparison We discussed further the performance of AD detection models based on different GAN structures. Nine different GAN-based AD detection models were proposed to solve the binary AD classification problem (Normal vs AD). Specifically, four different state-of-the-art GAN structures DCGAN [43], WGAN [44], Conditional-GAN [45] and AEGAN [46], were used to construct AD detection models. For each GAN category, AD detection models were further divided into three sub-categories supervised learning-based AD detection model (SL); unsupervised learning-based AD detection model (USL); and three component adversarial AD detection model (TCA).

Compared the different kinds of GAN-based AD detection models mentioned above with the proposed Alzheimer's disease detection model (BSGAN-ADD) to analyze the effect of different GAN structures on AD detection performance. In Table 4, the class-wise classification performances of AD detection models based on different GAN structures are provided. From Table 4, five conclusions can be made.

(1) Compared with the direct use of GAN to measure the distribution gap between Normal and AD samples, it can be seen that GAN as a high-level brain feature extraction tool to generate latent features and feed them to the DNN-based classifier for supervised learning has better performance in this classification problem (Normal vs AD).

Table 3
Performance of the BSGAN-ADD on four classification tasks.

Approach	Modalities	NL/AD/MCI	NL/AD	NL/MCI	AD/MCI	Parameters
Magnin et al.[41]	MRI	-	0.902	-	-	
Ben Ahmed et al.[42]	MRI	-	0.854	0.722	0.663	
Khvostikov et al.[24]	PET + MRI	0.852	0.885	0.877	0.831	0.17 M
Li et al.[37]	MRI	0.867	0.907	0.893	0.848	4.51 M
Korolev et al.[22]	MRI	-	0.823	0.782	0.751	2.14 M
Hosseini-Asl et al.[19]	MRI	0.824	0.972	0.968	0.867	457.0 M
Suk et al.[17]	PET + MRI	0.915	0.942	0.936	0.912	0.7 M
Wang et al.[38]	MRI	0.975	0.988	0.984	0.936	5.3 M
Feng et al.[39]	MRI	0.957	0.991	0.989	0.894	221.5 M
Emtiaz et al.[40]	MRI	-	0.978	-	-	2.0 M
BSGAN-ADD model(ADNI)	MRI	**0.986**	**1.000**	**0.994**	**0.979**	2.7 M
BSGAN-ADD model(OASIS)	MRI	**0.983**	**0.998**	**0.991**	**0.981**	2.7 M

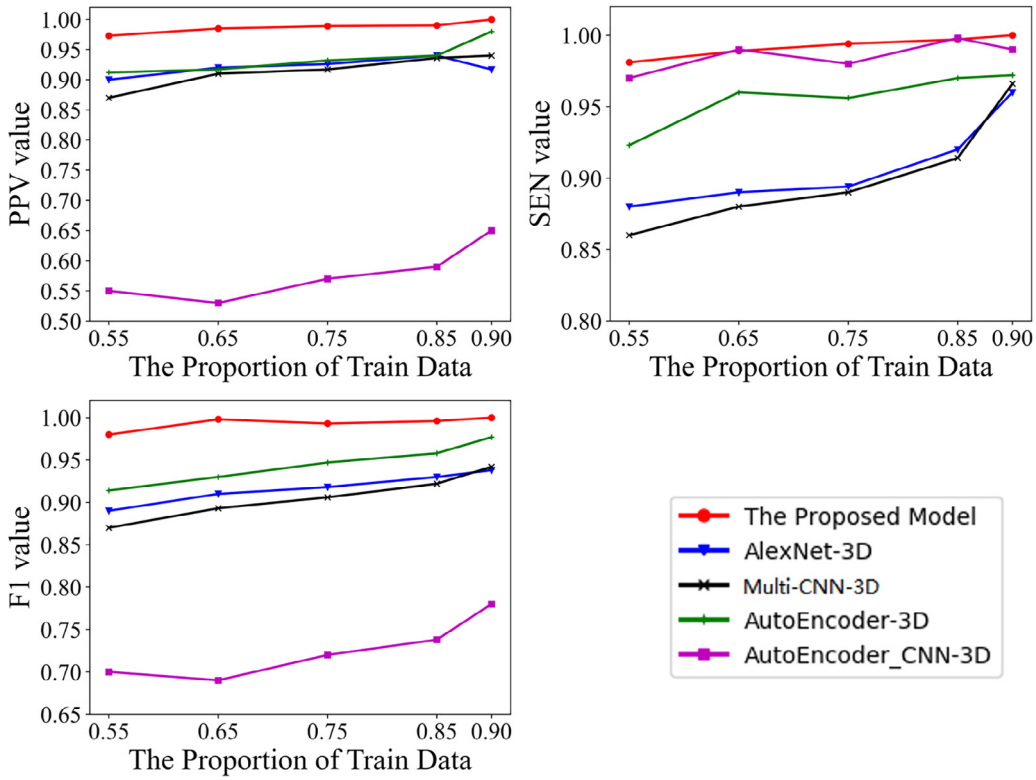

Fig. 7. The value changes of PPV, SEN and F1 on Normal group of testing data when the proportion of training data varied from 55% to 90%.

(2) By comparing the performance differences between the DCGAN-based and the Conditional-GAN-based AD detection model, it can be found that integrating the slice location information into the GAN-based AD detection model will improve the accuracy of classification model.
(3) By comparing the performance differences between the encoder-decoder GAN-based SL AD detection model and three-component adversarial AEGAN-based AD detection model, it was shown that adding classification loss into the loss calculation process of generator will help the model to extract more representative latent features.
(4) By comparing the performance differences between the DCGAN-based and the WGAN-based detection model, it shows that using different distance functions to measure the distribution gap between generated data and real data not significantly improves the classification performance.
(5) By comparing the performance differences between the proposed model which no *Encoder*2, no Encoder loss and no Classification loss respectively, it shows that *Encoder*2can significantly improve the performance of the model and Encoder loss and Classification loss can guide the model to better extract high-quality latent features. The proposed AD detection model (BSGAN-ADD) performs the category enhancement of brain slice images by integrating the classifier's feedbacks to the input 2D-brain feature image reconstruction process. The high-level brain features extracted from the category-enhanced brain slice images will be easier to correctly classify by the classifier, resulting in the BSGAN-ADD having better performance compared with competing GAN-based AD detection models.

4.3. Model performance

(1) Performances of the BSGAN-ADD:

The proposed Alzheimer's disease detection model (BSGAN-ADD) was applied to four specific classification tasks (three binary classifications and one ternary classification). The output of the BSGAN-ADD of the inference process is the NL/AD/MCI classifica-

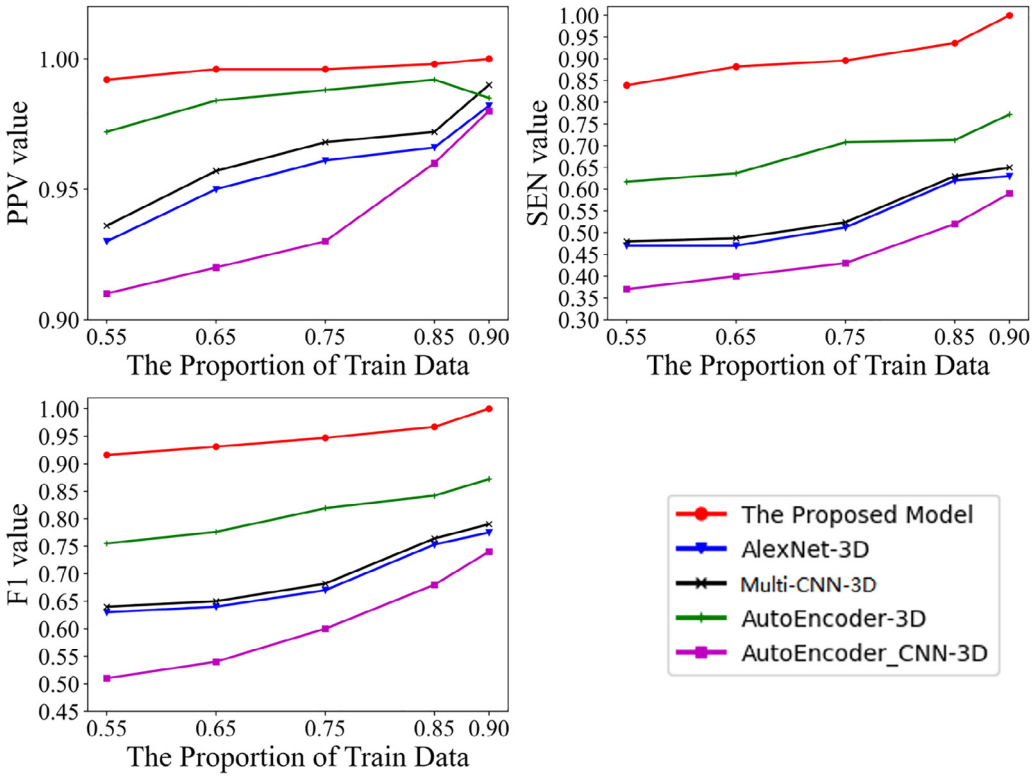

Fig. 8. The value changes of PPV, SEN and F1 on MCI group of testing data when the proportion of training data varied from 55% to 90%.

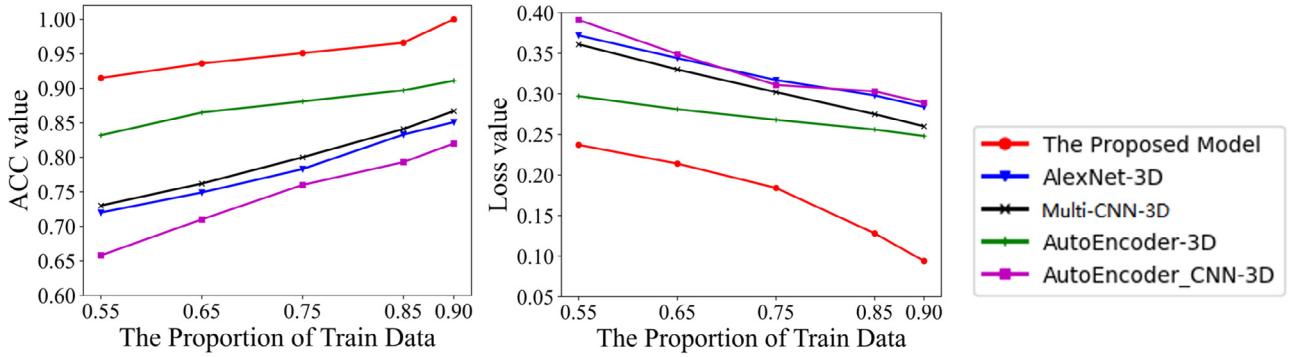

Fig. 9. The value changes of ACC and classification loss of different methods when the proportion of training data varied from 55% to 90%.

Table 4
The Class-wise classification performances of different GAN-based AD detection models [$Mean_{std}$].

Approach	PPV	NL/AD SEN	F1
DCGAN-SL	$0.615_{0.120}$	$0.578_{0.128}$	$0.596_{0.05}$
DCGAN-USL	$0.550_{0.02}$	$0.562_{0.413}$	$0.557_{0.238}$
WGAN-SL	$0.705_{0.124}$	$0.615_{0.120}$	$0.665_{0.08}$
WGAN-USL	$0.545_{0.04}$	$0.536_{0.369}$	$0.541_{0.192}$
Conditional-GAN-SL	$0.718_{0.07}$	$0.712_{0.057}$	$0.715_{0.03}$
Conditional-GAN-USL	$0.708_{0.099}$	$0.685_{0.289}$	$0.697_{0.113}$
AEGAN-SL	$0.852_{0.112}$	$0.908_{0.085}$	$0.875_{0.02}$
AEGAN-USL	$0.739_{0.079}$	$0.727_{0.151}$	$0.734_{0.092}$
AEGAN-TCA	$0.908_{0.061}$	$0.925_{0.042}$	$0.918_{0.014}$
Our Approach(No *Encoder*2)	$0.672_{0.109}$	$0.625_{0.211}$	$0.647_{0.150}$
Our Approach(No Encoder loss)	$0.932_{0.026}$	$0.951_{0.020}$	$0.943_{0.015}$
Our Approach(No Classification loss)	$0.927_{0.063}$	$0.925_{0.032}$	$0.926_{0.038}$
Our Approach(ADNI)	1.00_{0}	1.00_{0}	1.00_{0}
Our Approach(OASIS)	1.00_{0}	$0.997_{0.001}$	$0.998_{0.001}$

Table 5
Performance of the BSGAN-ADD on four classification tasks.

	NL/AD/MCI			NL/AD			NL/MCI			AD/MCI		
Class	PPV	SEN	F1	PPV	SEN	F1	PPV	SEN	F1	PPV	SEN	F1
Normal	0.982	0.997	0.991	1.000	1.000	1.000	0.989	0.996	0.993	-	-	-
AD	0.972	0.987	0.980	1.000	1.000	1.000	-	-	-	0.954	0.991	0.973
MCI	0.993	0.981	0.987	-	-	-	0.997	0.992	0.995	0.997	0.971	0.987
Mean	0.982	0.988	0.986	1.000	1.000	1.000	0.993	0.994	0.994	0.975	0.981	0.980

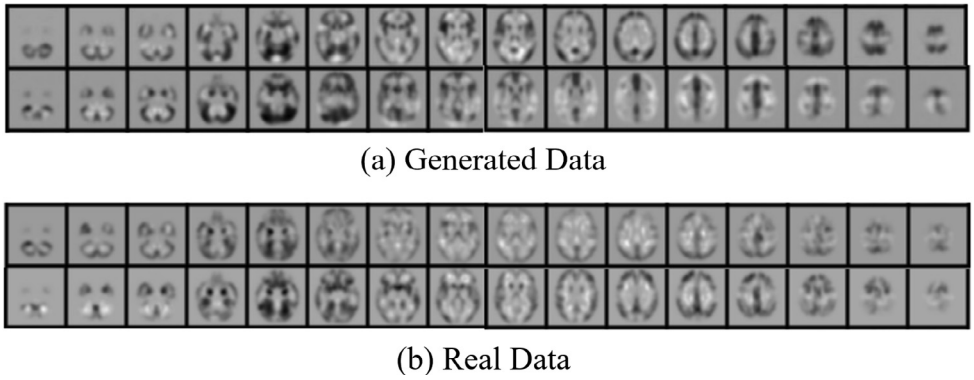

Fig. 10. The upper 16 images were the pixel-wise average images of 10 normal people on the corresponding slices, and the lower 16 images were the pixel-wise average images of 10 AD patients on the corresponding slices.

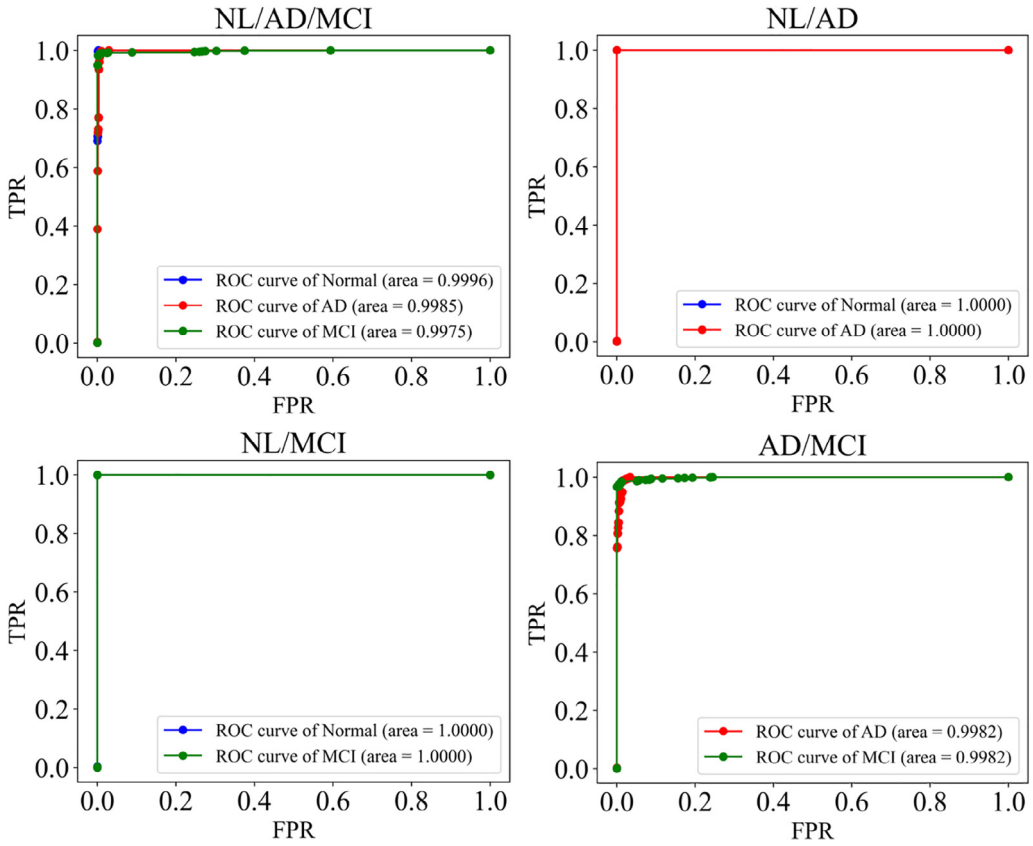

Fig. 11. The ROCs/ AUCs of four classification tasks from a single slice perspective.

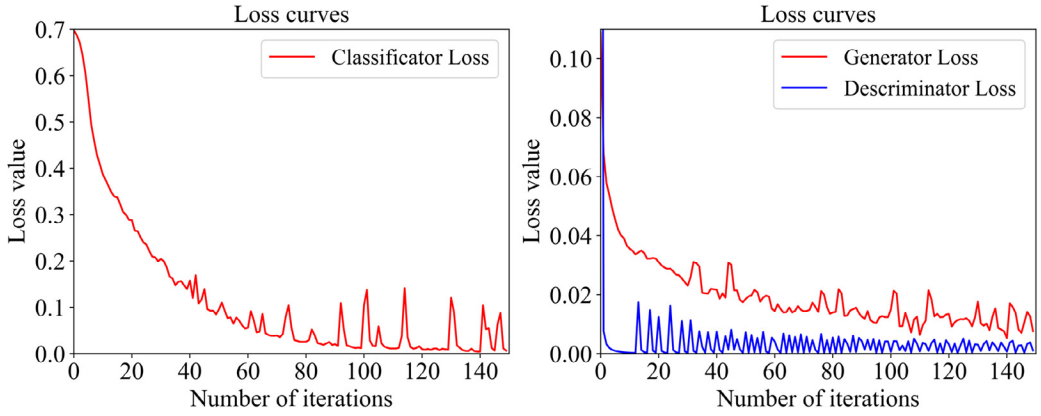

Fig. 12. The training loss curves of the three components (generator, discriminator and classifier) in BSGAN-ADD.

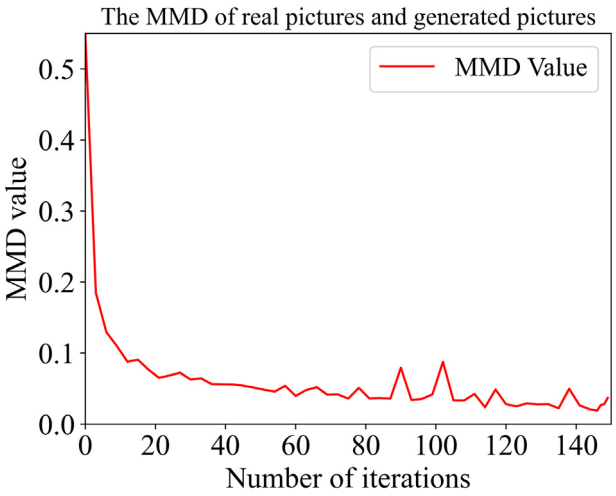

Fig. 13. The MMD values between the real images and the generated images across training iterations.

tion label of a PNG along the Z axis, and an MRI 3D data of a sample contains about 16 labels. The proposed model gives a diagnostic result for a specific sample base on the dominated AD predication labels of images along the Z axis. The detail class-wise experimental results were evaluated and the results shown in Table 5.

The ROCs/ AUCs of these classification tasks from a single slice (2D-brain slice image) perspective are shown in Fig. 11, and these results indicate that the proposed Alzheimer's disease detection model has great classification performance and make the AD detection tasks effectively.

From the results shown in Table 5 and Fig. 11, our BSGAN-ADD model obtains perfect results in the NL/AD two-category classification task. Excluding the factor of the limited test data, we believe that brain slice image enhancement achieved by the GAN-based image reconstruction helps the model gain the classification performance improvement. Using GAN-based approach to enhance the input 2D-brain slice images has three main advantages: **1) Stronger explicability of the high-level brain feature extraction process.** As shown in Fig. 10, the generator can reconstruct the input brain images with an encoder-decoder structure, which shows that the model can effectively capture the pixel-wise distribution of brain slice images, and brain slice images can be well represented by the high-level brain features extracted by the encoder. **2) Stronger model expressiveness.** Compared with the traditional CNN-based AD detection model, the proposed model learns how to perform image feature enhancement on the input brain feature images by integrating the disease category information to brain feature image reconstruction process. It can be observed in Fig. 10 that the pseudo 2D-brain feature images of normal people generated by the BSGAN-ADD were significantly different from those of AD patients, and the image differences were even greater than that between real images. Therefore, the controllable feature space of model can be extended from the extracted high-level

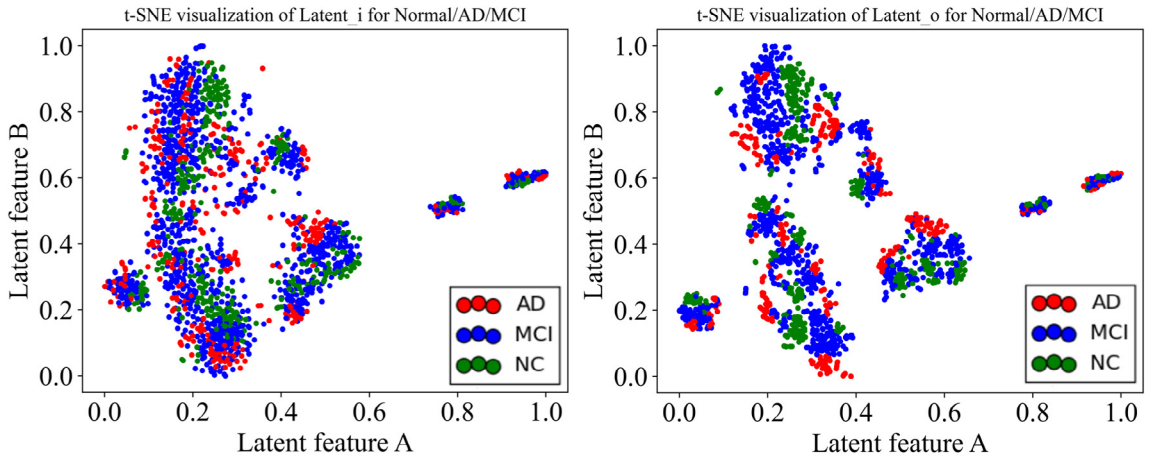

Fig. 14. The t-SNE projection-based feature visualizations of *Latent_i* and *Latent_o*.

brain feature space to contain the input brain slice image space, which greatly improves the model expressiveness. **3) Stronger model robustness.** According to [47], reconstructing the input data through adversarial training can effectively eliminate the influence of the trivial perturbations of the input data, and it allows us to navigate and manipulate the manifold of images.

Fig. 12 shows the training loss curves of the three components (generator, discriminator and classifier) in the BSGAN-ADD for the ternary classification problem (Normal vs AD vs MCI). As can be seen from Fig. 12, the proposed Alzheimer's disease detection model's components (generator, discriminator and classifier) can converge to low training loss values with the increase of training iteration number. In other words, Fig. 12 illustrates that the optimal model parameters can be obtained steadily using the model training strategy described in Section 3.5. To better illustrate that the generator of the BSGAN-ADD can output realistic 2D-brain feature images that have same data distribution with the input 2D-brain feature images, the maximum mean discrepancy (MMD) [48] was used to evaluate whether the proposed model had learned the data distributions of the input 2D-brain feature images. MMD is given by:

$$MMD(Input,Generated)=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K(Input_i,Generated_j)+\frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m}K(Generated_i,Generated_j)-\frac{2}{mn}\sum_{i=1}^{m}\sum_{j\neq i}^{m}K(Input_i,Generated_j), \tag{16}$$

The MMD values between the generated 2D-brain feature images and real images are plotted in Fig. 13 across training iterations.

According to Fig. 13, the MMD values tend to converge to a small value after 40 iterations. It can be concluded that the generator component in the proposed model can generate the pseudo-brain feature image similar to the input images, in other words, the proposed model can effectively learn the distributions of the original input 2D-brain feature images.

(2) Discussions of the BSGAN-ADD:

To verify that the high-level latent features (*Latent_o*) extracted from the pseudo 2D-brain feature images generated from the generator component are more advantageous for classifier to distinguish the diseased states than the highlevel latent features (*Latent_i*) extracted directly from the input 2D-brain feature images. For this, we plotted the t-SNE projection-based feature visualizations of the extracted high-level latent features and the visualizations of *Latent_i* and *Latent_o* are shown in Fig. 14. From Fig. 14, when we used *Latent_i* to represent the brain slices belonging to different diseased categories (Normal/AD/MCI) in low-dimensional latent space, there is almost no clear boundary between the data points of different diseased categories (the data points in red, blue and green colors belong to different categories). While using *Latent_o* to represent brain slice images is prone to map brain slice images of different disease categories to different regions in low-dimensional latent space, which is easier for classifier to determine the disease category.

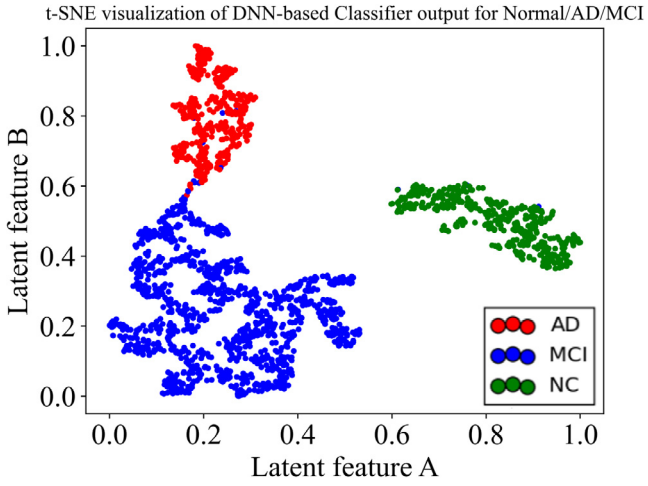

Fig. 16. The t-SNE projection-based visualizations of outputs of DNN-based classifier.

In order to show more clearly that it can make AD detection more effective by using the high-level latent features (*Latent_o*) extracted by deep CNNs from the generated 2D-brain feature images to represent the input 2D-brain feature images. The *Latent_o* and *Latent_i* were regarded as the brain slice image representations to train two independent SVM classifiers for the ternary classification task (Normal vs AD vs MCI) respectively. The visualization of the two linear classifiers' outputs is shown in Fig. 15. As shown in Fig. 15, the SVM classifier that takes *Latent_o* as input has better classification performance to accurately detect the diseased state. Moreover, it demonstrates to some extent that extracting high-level brain features from the category-enhanced brain images has substantial advantages over the traditional CNN-based extraction approaches that directly extract high-level brain features from

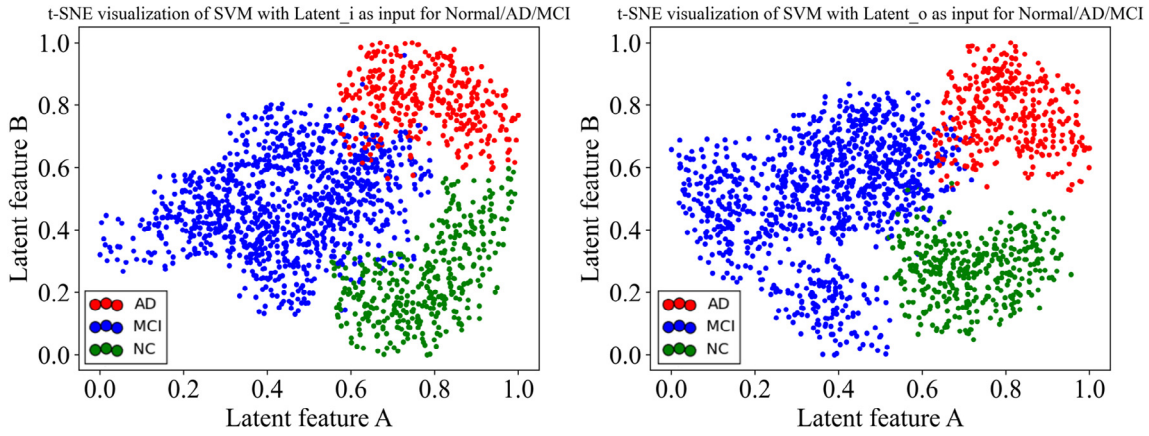

Fig. 15. The t-SNE projection-based visualizations of SVM with *Latent_i* and *Latent_o* as inputs.

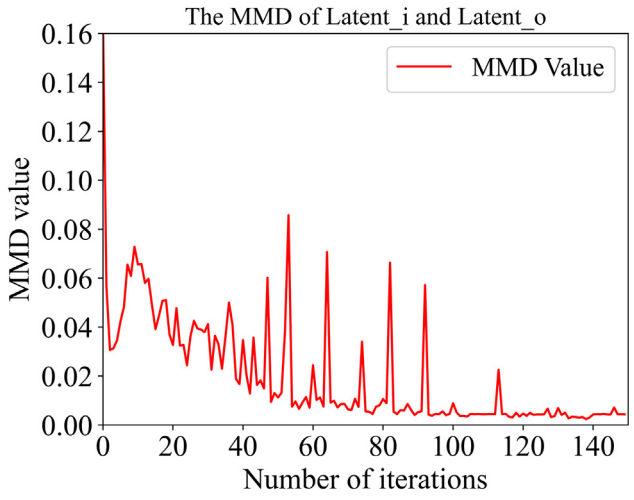

Fig. 17. The MMD values between Latent_i and Latent_o across training iterations.

the original brain feature images. We also plotted t-SNE projection-based visualization of the outputs of DNN-based classifier in the proposed Alzheimer's disease detection model. And it was shown in Fig. 16. These visualization results in Fig. 16 show that the DNN-based non-linear classifier has better classification performance in AD detection tasks than the linear classifier (SVM).

To further confirm the feasibility of our BSGAN-ADD model, we plotted that the MMD values between the high-level latent features (*Latent_i*) extracted from the input 2D-brain feature images and the high-level latent features (*Latent_o*) extracted from the generated 2D-brain feature images generated by the generator component. Fig. 17 shows that the MMD values between *Latent_i* and *Latent_o* can converge to a small value after about 100 training iterations. It proves that the extracted high-level features (Latent o) from the pseudo 2D-brain feature images have the same data distribution as the high-level latent features (*Latent_i*) directly extracted from the input 2D-brain feature images, so it is feasible represent the input 2D-brain feature images by *Latent_o*.

5. Conclusion

In this work, a novel feature-based and deep learning-based Alzheimer's disease detection (ADD) model (brain slice generative adversarial network for Alzheimer's disease detection, BSGAN-ADD) was proposed to perform computer-aided-diagnosis (CAD) of AD and its prodromal stage, Mild Cognitive Impairment, (MCI). In BSGAN-ADD, three components generator, discriminator and classifier compete in a three-player mini-max game. Performing the GAN-based brain slice image enhancement and the stacked CNN layers-based high-level brain feature extraction, to feed more representative brain features into DNN-based classifier to output the posterior probabilities of three diseased categories Normal, AD and MCI. Compared with several typical AD detection methods, our method showed a significant detection performance gain on two real-world datasets (ADNI, OASIS).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by Beijing Advanced Innovation Center for Big Date-based Precision Medicine, in part by the National Key RD Program of China [Grant No. 2018YFB0203900] and in part by the fund of the State Key Laboratory of Software Development Environment [Grant No. SKLSDE2017ZX10].

References

[1] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, Clinical diagnosis of alzheimer's disease: Report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease, Neurology 34 (7) (1984) 939.
[2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H.M. Arrighi, Forecasting the global burden of alzheimer's disease, Alzheimer's & dementia 3 (3) (2007) 186–191.
[3] S. Gauthier, B. Reisberg, M. Zaudig, R.C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, et al., Mild cognitive impairment, The lancet 367 (9518) (2006) 1262–1270.
[4] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, D. Feng, Early diagnosis of alzheimer's disease with deep learning, in: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), IEEE, 2014, pp. 1015–1018.
[5] M. Liu, D. Zhang, E. Adeli, D. Shen, Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis, IEEE Transactions on Biomedical Engineering 63 (7) (2015) 1473–1482.
[6] D. Cheng, M. Liu, J. Fu, Y. Wang, Classification of mr brain images by combination of multi-cnns for ad diagnosis, in: Ninth International Conference on Digital Image Processing (ICDIP 2017), Vol. 10420, International Society for Optics and Photonics, 2017, p. 1042042.
[7] B. Dubois, H.H. Feldman, C. Jacova, S.T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, et al., Research criteria for the diagnosis of alzheimer's disease: revising the nincds–adrda criteria, The Lancet Neurology 6 (8) (2007) 734–746.
[8] K. Blennow, B. Dubois, A.M. Fagan, P. Lewczuk, M.J. De Leon, H. Hampel, Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early alzheimer's disease, Alzheimer's & Dementia 11 (1) (2015) 58–69.
[9] T.G. Beach, S.E. Monsell, L.E. Phillips, W. Kukull, Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010, Journal of neuropathology and experimental neurology 71 (4) (2012) 266–273.
[10] S. Sacuiu, P.S. Insel, S. Mueller, D. Tosun, N. Mattsson, C.R. Jack Jr, C. DeCarli, R. Petersen, P.S. Aisen, M.W. Weiner, et al., Chronic depressive symptomatology in mild cognitive impairment is associated with frontal atrophy rate which hastens conversion to alzheimer dementia, The American Journal of Geriatric Psychiatry 24 (2) (2016) 126–135.
[11] B. Dubois, A. Padovani, P. Scheltens, A. Rossi, G. Dell'Agnello, Timely diagnosis for alzheimer's disease: a literature review on benefits and challenges, Journal of Alzheimer's disease 49 (3) (2016) 617–631.
[12] Z. Sun, Y. Qiao, B.P. Lelieveldt, M. Staring, A.D.N. Initiative, et al., Integrating spatial-anatomical regularization and structure sparsity into svm: Improving interpretation of alzheimer's disease classification, NeuroImage 178 (2018) 445–460.
[13] B. Jie, C.-Y. Wee, D. Shen, D. Zhang, Hyper-connectivity of functional networks for brain disease diagnosis, Medical Image Analysis 32 (2016) 84–100.
[14] X. Liu, D. Tosun, M.W. Weiner, N. Schuff, A.D.N. Initiative, et al., Locally linear embedding (lle) for mri based alzheimer's disease classification, Neuroimage 83 (2013) 148–157.
[15] X. Zhu, H.-I. Suk, D. Shen, A novel matrix-similarity based loss function for joint regression and classification in ad diagnosis, NeuroImage 100 (2014) 91–105.
[16] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, D. Shen, A.D.N. Initiative, et al., Longitudinal clinical score prediction in alzheimer's disease with soft-split sparse regression based random forest, Neurobiology of Aging 46 (2016) 180–191.
[17] H.-I. Suk, S.-W. Lee, D. Shen, Latent feature representation with stacked auto-encoder for ad/mci diagnosis, Brain Structure and Function 220 (2) (2015) 841–859.
[18] A. Payan, G. Montana, Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks, arXiv preprint arXiv:1502.02506 (2015).
[19] E. Hosseini-Asl, G. Gimel'farb, A. El-Baz, Alzheimer's disease diagnostics by a deeply supervised adapTable 3d convolutional network, arXiv preprint arXiv:1607.00556 (2016).
[20] T.D. Vu, H.-J. Yang, V.Q. Nguyen, A.-R. Oh, M.-S. Kim, Multimodal learning using convolution neural network and sparse autoencoder, in: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2017, pp. 309–312.
[21] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M.J. Fulham, et al., Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease, IEEE Transactions on Biomedical Engineering 62 (4) (2014) 1132–1140.

[22] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, Residual and plain convolutional neural networks for 3d brain mri classification, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017, pp. 835–838.
[23] H. Choi, K.H. Jin, A.D.N. Initiative, et al., Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging, Behavioural brain research 344 (2018) 103–109.
[24] A. Khvostikov, K. Aderghal, A. Krylov, G. Catheline, J. Benois-Pineau, 3d inception-based cnn with smri and md-dti data fusion for alzheimer's disease diagnostics, arXiv preprint arXiv:1809.03972 (2018).
[25] X. Feng, J. Yang, A.F. Laine, E.D. Angelini, Discriminative analysis of the human cortex using spherical cnns-a study on alzheimer's disease diagnosis, arXiv preprint arXiv:1812.07749 (2018).
[26] K. Aderghal, J. Benois-Pineau, K. Afdel, C. Gwenaëlle, Fuseme: Classification of smri images by fusion of deep cnns in 2d+ ε projections, in: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, 2017, pp. 1–7.
[27] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for alzheimer's disease diagnosis, IEEE Transactions on Biomedical Engineering 66 (5) (2018) 1195–1206.
[28] D. Cheng, M. Liu, Combining convolutional and recurrent neural networks for alzheimer's disease diagnosis using pet images, in: 2017 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2017, pp. 1–5.
[29] D. Lu, K. Popuri, G.W. Ding, R. Balachandar, M.F. Beg, Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural mr and fdg-pet images, Scientific reports 8 (1) (2018) 1–13.
[30] M. Hon, N.M. Khan, Towards alzheimer's disease classification through transfer learning, in: 2017 IEEE International conference on bioinformatics and biomedicine (BIBM), IEEE, 2017, pp. 1166–1169.
[31] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, J. Li, A robust deep model for improved classification of ad/mci patients, IEEE journal of biomedical and health informatics 19 (5) (2015) 1610–1616.
[32] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A. Smola, A kernel method for the two-sample-problem, Advances in neural information processing systems 19 (2006) 513–520.
[33] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434 (2015).
[34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems (nips), 2014.
[35] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, Cvae-gan: fine-grained image generation through asymmetric training, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2745–2754.
[36] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.
[37] F. Li, D. Cheng, M. Liu, Alzheimer's disease classification based on combination of multi-model convolutional networks, in: 2017 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2017, pp. 1–5.
[38] H. Wang, Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang, X. Zhao, Ensemble of 3d densely connected convolutional network for diagnosis of mild cognitive impairment and alzheimer's disease, Neurocomputing 333 (2019) 145–156.
[39] W. Feng, N.V. Halm-Lutterodt, H. Tang, A. Mecum, M.K. Mesregah, Y. Ma, H. Li, F. Zhang, Z. Wu, E. Yao, et al., Automated mri-based deep learning model for detection of alzheimer's disease process, International Journal of Neural Systems 30 (06) (2020) 2050032.
[40] E. Hussain, M. Hasan, S.Z. Hassan, T.H. Azmi, M.A. Rahman, M.Z. Parvez, Deep learning based binary classification for alzheimer's disease detection using brain mri images, in: 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2020, pp. 1115–1120.
[41] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, H. Benali, Support vector machine-based classification of alzheimer's disease from whole-brain anatomical mri, Neuroradiology 51 (2) (2009) 73–83.
[42] O.B. Ahmed, M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, C.B. Amar, A. D.N. Initiative, et al., Alzheimer's disease diagnosis on structural mr images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex, Computerized Medical Imaging and Graphics 44 (2015) 13–25.
[43] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, Computer ence (2015).
[44] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan (2017)..
[45] M. Mirza, S. Osindero, Conditional generative adversarial nets, Computer Science (2014) 2672–2680.
[46] S. Shen, G. Jin, K. Gao, Y. Zhang, Ae-gan: adversarial eliminating with gan (2017).
[47] B. Lindqvist, S. Sugrim, R. Izmailov, Autogan: Robust classifier against adversarial attacks, arXiv preprint arXiv:1812.03405 (2018).
[48] K.M. Borgwardt, A. Gretton, M.J. Rasch, H.-P. Kriegel, B. Schölkopf, A.J. Smola, Integrating structured biological data by kernel maximum mean discrepancy, Bioinformatics 22 (14) (2006) e49–e57.

Tian Bai is a Ph.D. candidate at School of Automation Science and Electrical Engineering in Beihang University. His research interests include deep learning, health data mining & management, service recommendation and service composition.

Mingyu Du was born in 1997. He received the M.S. degrees from School of Automation Science and Electrical Engineering in Beihang University, China. His research interests include deep learning, health data mining & management, and computer vision.

Lin Zhang is a professor of Beihang Unversity. He received the B.S. degree in 1986 from Nankai University, China, the M.S. and the Ph.D. degrees in 1989 and 1992 from Tsinghua University, China. He served as the President of the Society for Modeling and Simulation International (SCS) (2015–2016). He is a Fellow of SCS and ASIASIM, the excutive vice president of China Simulation Federation (CSF). He authored and co-authored 200 papers, 18 books and chapters. His research interests include service oriented modeling and simulation, model engineering, cloud manufacturing and simulation and their applications in health, etc.

Lei Ren received the Ph.D. degree in computer science from the Institute of Software, Chinese Academy of Sciences in 2009. He is currently an Associate Professor and the Deputy Head with the Cloud Manufacturing Research Center, School of Automation Science and Electrical Engineering, Beihang University, China. He has authored or coauthored over 60 papers and got over 2000 citations according to Google Scholar. His research interests include cloud manufacturing, industrial Internet of Things, and big data analytics.

Li Ruan is an assistant professor in state key laboratory of software development environment, school of computer science and engineering, Beihang university in China. Her current research focus on artificial intelligent, optimization theory, parallel and distributing computing, mobile computing, software system for Big Data applications, and performance optimization. Recent research topics include largescale cloud trace data mining, deep learning based task scheduling and load balance, runtime aware load balance for high performance computer. Her work has found broad applications including using AI and optimization in cloud platform, high performance computer etc. Her research work research record also includes two textbooks and more than twenty patents. Dr. Li received her PH.D degree in Computer Software from Institute of Software, Chinse Academy of Sciences in 2009. She received an master's degree in Computer Science from Chinse Academy of Sciences. Then she joined school of computer science and engineering, Beihang university, China.

Yuan Yang was born in 1993. He received the M.S. degrees from University of Science & Technology Beijing, China. He is a Ph.D. candidate in Beihang University. His research interests include health information management, wearable sensors, data mining and their applications to healthcare services.

Guanghao Qian is a Ph.D. candidate at School of Automation Science and Electrical Engineering in Beihang University. He received the B.S. degree in Beihang University. His research interests include data-driven methods, Internet of Things, modeling and simulation of complex systems.

Zihao Meng was born in 1996. He received the M.S. degrees from Beihang University, China. He is a Ph.D. candidate in Beihang University. His research interests include industry intelligence, naturel language processing, and soft sensors.

Li Zhao is a PhD candidate at School of Automation Science and Electrical Engineering in Beihang University. His research interests include deep learning, data-driven methods, and remaining useful life (RUL) prediction of lithium-ion battery.

M. Jamal Deen is a Distinguished University Professor and Senior Canada Research Chair in Information Technology, McMaster University, Canada. His current research interests are in nano-/optoelectronics, nanotechnology and data analytics and their applications in health and environment. He is an elected Fellow of ten national academies and professional societies including the Royal Society of Canada (the highest honor for academics, scholars and artists in Canada), IEEE, the American Physical Society and the Electrochemical Society. Over his career, he has won more than seventy-five awards and honors including the McNaughton Gold medal (highest award for Engineers), the Fessenden Medal and the Ham Education Medal, all from IEEE Canada. Most recently, he was elected to the Order of Canada, the highest civilian honor awarded by the Government of Canada.