Neural Networks 169 (2024) 442–452

Contents lists available at ScienceDirect

Neural Networks

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

A comparative study of GNN and MLP based machine learning for the diagnosis of Alzheimer's Disease involving data synthesis

Ke Chen [a,b,c], Ying Weng [a,*], Akram A. Hosseini [d], Tom Dening [c], Guokun Zuo [b], Yiming Zhang [a,c]

[a] School of Computer Science, University of Nottingham Ningbo China, Ningbo, 315100, China
[b] Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, 315201, China
[c] School of Medicine, University of Nottingham, Nottingham, NG7 2RD, UK
[d] Neurology Department, Nottingham University Hospitals NHS Trust, Nottingham, NG7 2UH, UK

ARTICLE INFO

Keywords:
Alzheimer's Disease (AD)
Synthesis model
Graph Neural Network (GNN)
Data fusion

ABSTRACT

Alzheimer's Disease (AD) is a neurodegenerative disease that commonly occurs in older people. It is characterized by both cognitive and functional impairment. However, as AD has an unclear pathological cause, it can be hard to diagnose with confidence. This is even more so in the early stage of Mild Cognitive Impairment (MCI). This paper proposes a U-Net based Generative Adversarial Network (GAN) to synthesize fluorodeoxyglucose - positron emission tomography (FDG-PET) from magnetic resonance imaging - T1 weighted imaging (MRI-T1WI) for further usage in AD diagnosis including its early-stage MCI. The experiments have displayed promising results with Structural Similarity Index Measure (SSIM) reaching 0.9714. Furthermore, three types of classifiers are developed, i.e., one Multi-Layer Perceptron (MLP) based classifier, two Graph Neural Network (GNN) based classifiers where one is for graph classification and the other is for node classification. 10-fold cross-validation has been conducted on all trials of experiments for classifier comparison. The performance of these three types of classifiers has been compared with the different input modalities setting and data fusion strategies. The results have shown that GNN based node classifier surpasses the other two types of classifiers, and has achieved the state-of-the-art (SOTA) performance with the best accuracy at 90.18% for 3-class classification, namely AD, MCI and normal control (NC) with the synthesized fluorodeoxyglucose - positron emission tomography (FDG-PET) features fused at the input level. Moreover, involving synthesized FDG-PET as part of the input with proper data fusion strategies has also proved to enhance all three types of classifiers' performance. This work provides support for the notion that machine learning-derived image analysis may be a useful approach to improving the diagnosis of AD.

1. Introduction

Dementia is a general term for a loss of memory, language, problem-solving and other thinking skills that are severe enough to affect daily life. The commonest causes of dementia include Alzheimer's disease (AD), vascular dementia (VaD) and dementia with Lewy bodies (DLB). Both the prevalence and incidence of dementia increase with advancing age, but about 5% of cases have onset at age <65 years. The World Alzheimer Report 2018 estimated that across the world a person will develop dementia every 3 s (Patterson, 2018). The number of people suffering from dementia is about 50 million in 2018 and will be more than treble to 152 million by 2050. AD is the commonest cause of dementia, accounting for 60%–80% of cases of dementia, and it is one of the main causes of physical and mental health disorders among older people worldwide.

People with AD may suffer from memory loss, progressive language impairment, a gradual decline in the ability to perform daily tasks and abnormal changes in personality and behavior. These symptoms can affect the person's life and severely reduce the life quality both for the person and their family. Moreover, AD is an irreversible disease and all current available treatments may only delay its progress. Nonetheless, the diagnosis of AD, especially in the early stages, is important, so that individuals and families can be aware and make adjustments to their lives as needed, but also since more precise diagnosis will be needed to develop new, potentially disease-modifying treatments.

In general, mild cognitive impairment (MCI) represents the early stage of AD. Patients diagnosed as MCI are more likely to progress to AD with a ratio of about 16% developing AD within about 4 years, compared with non-MCI subjects whose conversion rate is about 1%–2% in this period (Michaud et al., 2017). However, not all people

* Corresponding author.
E-mail addresses: ke.chen2@nottingham.edu.cn (K. Chen), ying.weng@nottingham.edu.cn (Y. Weng), akram.hosseini@nuh.nhs.uk (A.A. Hosseini), tom.dening@nottingham.ac.uk (T. Dening), moonstone@nimte.ac.cn (G. Zuo), yiming.zhang2@nottingham.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.neunet.2023.10.040
Received 12 May 2023; Received in revised form 23 September 2023; Accepted 25 October 2023
Available online 26 October 2023

diagnosed with MCI will progress to AD and some will remain stable or even return to normal cognition, which indicates that MCI can be further divided into two categories: progressive MCI (pMCI) and stable MCI (sMCI), according to the future risk of progressing to AD in future years. Therefore, identifying MCI and its sub-categories (pMCI and sMCI) should have a significant impact on the early stage of treatment to mitigate the progress of dementia.

AD, as a neurodegenerative disease whose pathologic cause is unclear, is hard to identify with certainty by non-invasive clinical investigations. Clinical diagnosis usually relies on a combination of history, mental state examination and cognitive testing. The most commonly used cognitive instruments for the diagnosis of AD and MCI are the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) (Pinto et al., 2019). These cognitive assessment tools are short questionnaires to test a range of cognitive domains. However, despite having relatively high sensitivity and specificity for dementia, they perform less well in the diagnosis of early-stage MCI. Further, their scores are affected by educational level and language, and therefore, they are insufficiently objective as a means of diagnosis.

Recently, machine learning technology has shown promising performance on AD incidence prediction with large-scale administrative health data (Park et al., 2020). A more objective approach to AD diagnosis is through biomarkers such as brain imaging (Khojaste-Sarakhsi et al., 2022; Wang et al., 2022), blood and cerebrospinal fluid examination. However, due to the unclear pathological cause, there is no uniform standard for AD diagnosis through biomarkers. Applying such tests for MCI is even less precise because of the heterogeneity of the syndrome of MCI. Consequently, with the development of artificial intelligence (AI) in the field of computer vision, computer-aided AD diagnosis, including prediction of AD and MCI, using medical images such as magnetic resonance imaging - T1 weighted imaging (MRI-T1WI), positron emission tomography (PET) (Sharma et al., 2023) has become a research hotspot in recent years(Tanveer et al., 2020). Some recent studies have shown its capability in AD diagnosis with various modalities of radiography as input including structural MRI (sMRI), functional MRI (fMRI) and positron emission tomography (PET).

sMRI is the most common modality for the diagnosis of AD due to its low cost for acquisition, non-invasiveness and accessibility. The feature extraction methods of sMRI can be roughly divided into three categories: density maps-based, cortical surface-based, and predefined regions-based. Multiple-set adaptive ROIs-based method was proposed with density maps to predict AD and MCI and achieved accuracy at 92.51% for AD/normal control(NC) classification and 78.99% for pMCI/sMCI classification (Liu et al., 2015). Not easily visible changes in temporal and parietal regions of the cortical surface at the early stage of AD can also be used in classification. In the study of Wee et al. (2013), the authors took cortical morphological patterns with the multi-kernel SVM-based method and achieved similar performance for AD/NC classification and pMCI/sMCI classification as well as two more classifications, AD/MCI and MCI/NC, with both accuracy above 75%. Besides, the predefined region-based approach exploits prior knowledge from previous studies in special patterns of AD. In other words, it focuses on some regions which are likely to be useful for classification. Many studies show that it is effective to use the predefined region. For instance, in the study of Chincarini et al. (2011), it achieved 97% accuracy for AD/NC classification. In addition to the prior feature-based methods, utilizing the whole 3D images as input also shows its capability in the differential diagnosis. A Content-Based Image Retrieval (CBIR) system was proposed using a 3D capsule network for AD diagnosis with sMRI and PET images as input and achieved an accuracy of 89.1% on AD/MCI/NC classification, 97.6% on AD/NC classification, 95% on AD/MCI classification and 90.8% on MCI/NC classification (Kruthika et al., 2019). In the study of Xia et al. (2020), the authors put forward a hybrid network involving both 3D Covolutional Neural Network (CNN) and 3D convolutional long short-term memory (3D CLSTM) to predict AD using sMRI and achieved an accuracy of 94.19% on AD/NC classification. A dual attention multi-instance deep learning method was presented for AD prediction and MCI conversion and achieved an accuracy of 92.4% for AD/NC classification, 80.2% for pMCI/sMCI classification (Zhu et al., 2021). A new method for AD diagnosis was described by estimating brain age from sMRI image instead of performing classification and achieved the mean absolute error at 2.428 years (Cheng et al., 2021). In the study of Park et al. (2023), they proposed a prospective classification of AD conversion from MCI and achieved an accuracy of 88.1% for AD conversion prediction. A Monte Carlo Ensemble Neural Network (MCENN) for AD diagnosis was proposed and achieved an accuracy of 88.7% for AD/MCI/NC classification with random feature sampling set to 32 (Liu et al., 2023). Combining sMRI features with phenotype features, Ho et al. (2022) proposed a forward-to-backward bi-directional network with integrative imputation to predict patients' next one-year biomarkers and clinical status, and achieved an accuracy of 58.47% in forecasting clinical status. An ensemble framework has also been proposed by Nguyen et al. (2022) using sMRI to diagnose AD and achieved an accuracy of 96% for AD/NC classification. Consequently, using sMRI in automated AD diagnosis shows promising results with various machine learning based approaches. However, the sMRI data can only reflect structural pathology while brain functional pathology cannot be reflected. Therefore, the result might be further improved when used together with other types of data that reflecting brain functional pathology, which is also a recent hot-spot in machine learning as multi-modal learning. In this work, we start with sMRI while involving the synthesized FDG-PET to improve the derived model's performance.

Apart from sMRI, AI application in other modalities such as fMRI and PET has also had some encouraging findings. Moreover, studies involving fMRI and PET even show higher accuracy compared with sMRI. Recent studies achieved accuracy at around 95% for AD/NC classification (Li et al., 2020; Parmar et al., 2020a, 2020b). In the study of Liu et al. (2015), the authors also achieved 91.2% accuracy on AD/NC classification and 78.9% accuracy on MCI/NC classification using pure fluorodeoxyglucose-positron emission tomography (FDG-PET) as input on their proposed model which was a combination of CNN and bidirectional gated recurrent unit (BGRU). In the study of Pan et al. (2021), the authors came up with a CNN-based multi-view pyramid network for AD prediction at the MCI stage using FDG-PET as input and achieved an accuracy of 93.13% for AD/NC classification and 83.05% for pMCI/sMCI classification. Tuan et al. (2022) proposed C-Atlas-based machine learning approaches to diagnose AD with FDG-PET and achieved an accuracy of 91.83%. Though using functional imaging such as fMRI and PET shows promising results in AD diagnosis, their high cost for acquisition as well as limitation of equipment still limits their usage. Therefore, using sMRI might maximize the generalizability of AI in AD diagnosis in real practice. In our work, we use the synthesized FDG-PET from MRI-T1WI as an additional reference for the model instead of the ground truth one due to its difficulties in acquisition. In other words, the only medical image used in our proposed method in the inference stage is sMRI.

Although sMRI, fMRI, PET have all shown promising results on AD diagnosis with machine learning, sMRI only has shown the anatomical structures in the brain while the functional or metabolic properties of brain tissues are not highlighted. In this work, we propose a U-Net based Generative Adversarial Network (GAN) to synthesize PET images from sMRI to obtain the expensive and higher AD-sensitive modality from the cheap and lower AD-sensitive modality as well as obtain functional or metabolic properties from anatomical structures. Compared to the other AI-based AD diagnosis applications, we further use the synthesized FDG-PET images amalgamated with its source sMRI images as the joint input for the classifier to improve the accuracy of diagnosis for AD as well as its early-stage MCI.

This paper describes the following work that has been undertaken:

Table 1
Demographic analysis of Paired Dataset with MRI-T1WI and FDG-PET as pair and Extended Dataset with MRI-T1WI only (Where 'Scans' for Paired Dataset denotes the number of image pairs and 'Scans' for Extended Dataset denotes the number of MRI-T1WI).

	Paired dataset	Extended dataset
Participants	332	760
Scans	1036	3948
Male (%)	62.65%	57.76%
Female (%)	37.35%	42.23%
Age in years, mean (SD)	76.59 ± 6.65	76.47 ± 6.85

- We propose a GAN model to generate synthesis images that can overcome the limitation of the missing PET images in AD diagnosis. The model is trained on the selected paired data of MRI-T1WI and FDG-PET from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and achieves an overall performance with MAE at 0.0141, MSE at 0.0014, ZNCC at 0.9871 and SSIM at 0.9714.
- We investigate the different machine learning models, including MLP and GNN, for AD classification. The models are trained on the MRI-T1WI data from the ADNI dataset and the various experiments are conducted. The best performance is achieved by GNN based model for node classification with 90.18% accuracy on NC/MCI/AD classification.
- We carry out additional experiments on identification of normal controls (NC), stable MCI (sMCI) and progressive MCI (pMCI), and achieve an accuracy of 82.77%
- Our work shows that the synthesized data can enhance the model performance for AD diagnosis and also provides support for the notion that machine learning-derived image analysis can be a useful approach to improve the diagnosis of AD.

2. Methods

2.1. Datasets

Data used in the preparation of this work are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

In this research, we have downloaded the paired MRI T1 weighted images (MRI-T1WI) and FDG-PET images for the task to synthesize FDG-PET from MRI-T1WI. The pairing procedure is based on the metadata of the downloaded images and the terms for pairing are the same subject ID and the interval of the age of obtaining image less than 30 days. The demographic analysis of the dataset with MRI-PET pairs is shown in Table 1.

Furthermore, we collect more data as an extended dataset containing MRI-T1WI images only because FDG-PET images can be synthesized from MRI-T1WI. The extended data are used for training the classifier since the classifier takes MRI-T1WI and synthesized FDG-PET images. In other words, the actual input is MRI-T1WI images for the whole diagnosis system. The demographic analysis of the extended dataset is shown in Table 1.

2.2. Data preprocessing

The data in the ADNI dataset is collected from different sources, which implies differences existing in voxel size, orientation, etc. Data preprocessing is carried out using FSL (Jenkinson et al., 2012). For both MRI-T1WI and FDG-PET images, we first remove non-brain tissues. The images are then registered to the standard MNI152 space. Moreover, brain region segmentation is applied to the registered images to obtain white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) as shown in Fig. 1.

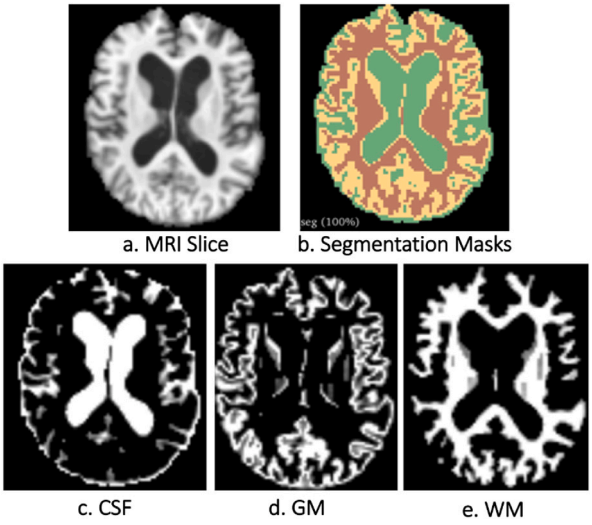

Fig. 1. MRI-T1WI segmentation. (a) is an MRI slice, (b) is the visualization of the corresponding segmentation masks where green refers to cerebrospinal fluid (CSF), yellow refers to grey matter (GM) and brown refers to white matter (WM). (c), (d), (e) are the visualization of CSF, GM, WM tissues respectively.

2.3. Synthesis model

For data synthesis from MRI-T1WI to FDG-PET, we propose a U-Net (Ronneberger et al., 2015) based Generative Adversarial Network (GAN) (Goodfellow et al., 2020). The overall architecture of GAN in this work is shown in Fig. 2 with a generator responsible for synthesizing data and a discriminator responsible for determining its input as the real data or the synthesized data. The advantage of using GAN for data synthesis is that the generator is not only optimized through predefined loss but also through the discriminator. As for the discriminator, the ability to identify whether an input is synthesized or not is also promoted following the growth of the generator.

Fig. 3 shows the detailed structure of the generator which is in the U-Net style with the left side as the contracting path and the right side as the expanding path. The contracting path consists of the repeated 3D CNN based down-sampling modules. In each down-sampling module, a 3D convolutional block, which consists of a 3D convolutional layer and a rectified linear unit (ReLU), is first performed with a stride of 2 to extract features and down-sample the input with a ratio of 2. The result is then added to the result from the following two 3D convolutional blocks with a stride of 1 to construct the residual connection which helps with training. The expanding path is the reverse of the contracting path which consists of repeated up-sampling modules with the same number as the contracting path. The up-sampling module is similar to the downsampling module except for replacing the 3D convolution block with a stride of 2 in an up-sampling layer and a 3D convolution block with a stride of 1. Moreover, the skipped connection exists between the contracting path and the expanding path by concatenating the corresponding feature representations from two paths as shown in Fig. 3. This skipped connection allows feature fusion between feature representations with the small receptive field and the large receptive field, which enhances the generator's ability to generate more precise and detailed synthesized FDG-PET images.

As for the discriminator, it aims at identifying whether the FDG-PET image is the corresponding ground truth image of the source MRI-T1WI

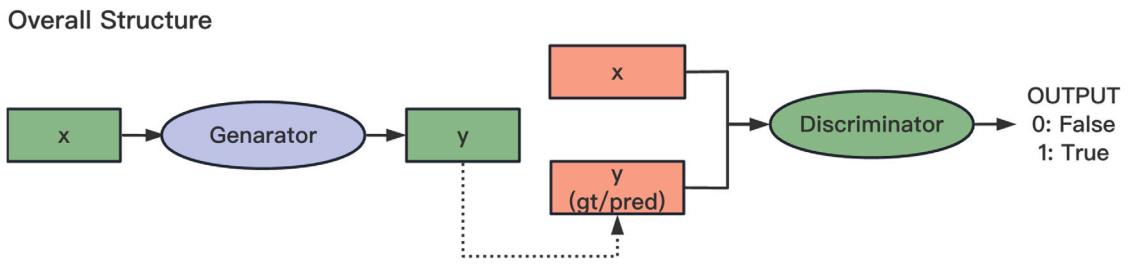

Fig. 2. General architecture of GAN implementation consists of a generator to generate the FDG-PET from MRI-T1WI and a discriminator to identify if the FDG-PET is synthesized or real.

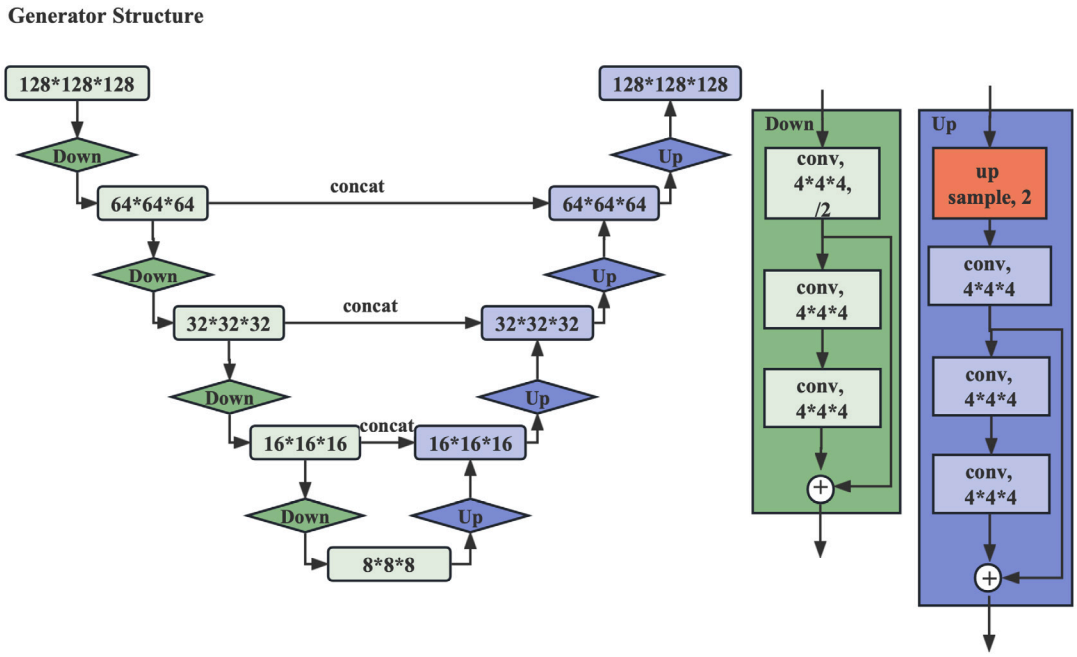

Fig. 3. Structure of the 3D U-Net generator consists of a contracting path and an expanding path with skipped connections to refine the prediction.

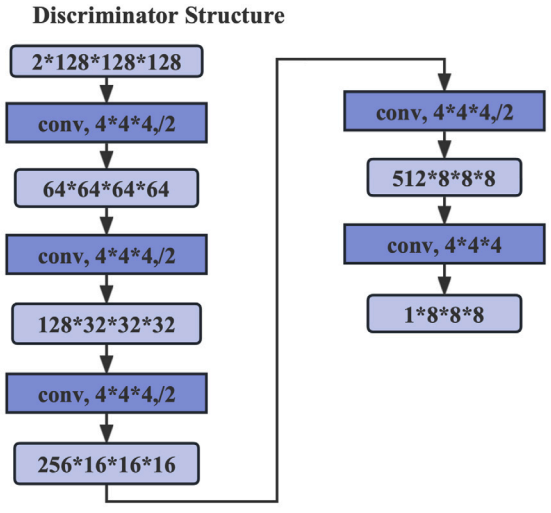

Fig. 4. Structure of the discriminator consists of several 3D convolutional modules.

image. Therefore, the discriminator takes two channels of input, which is the paired image set including the source MRI-T1WI image and its corresponding target FDG-PET image, either the ground truth FDG-PET image or the synthesized one. Fig. 4 shows the detailed structure of the discriminator. The first four 3D convolutional blocks perform feature extraction as well as downsampling to the size of $8 * 8 * 8$ while the last 3D convolutional block, which consists of a 3D convolutional layer with a stride of 1 and a sigmoid function, performs scoring for each sub-region.

As for updating, the loss of the discriminator is used to help the generator optimize not only from the predefined loss but also from the knowledge learned by the discriminator. Since the output of the discriminator is a $8 * 8 * 8$ matrix with each value ranging from 0 to 1 indicating the probability of the sub-region being part of the ground truth FDG-PET image. Therefore, a matrix V with a size of $8 * 8 * 8$ and all values at 1 is designed to represent the desired output of the image pair of the source MRI-T1WI image and the ground truth FDG-PET image. Similarly, the same-sized matrix F with all values at 0 represents the desired output of the image pair of the source MRI-T1WI image and the synthesized FDG-PET image. The loss of the discriminator is hence defined as below with mean squared error:

$$Loss(D) = \frac{MSE(D(gt), V) + MSE(D(pred), F)}{2} \tag{1}$$

Where D denotes the discriminator, gt denotes the ground truth FDG-PET image and $pred$ denotes the synthesized FDG-PET image. Moreover, the equation for Mean Squared Error (MSE) is as below:

$$MSE(X, Y) = \frac{\sum_{i=1}^{\xi}(y_i - x_i)^2}{\xi} \tag{2}$$

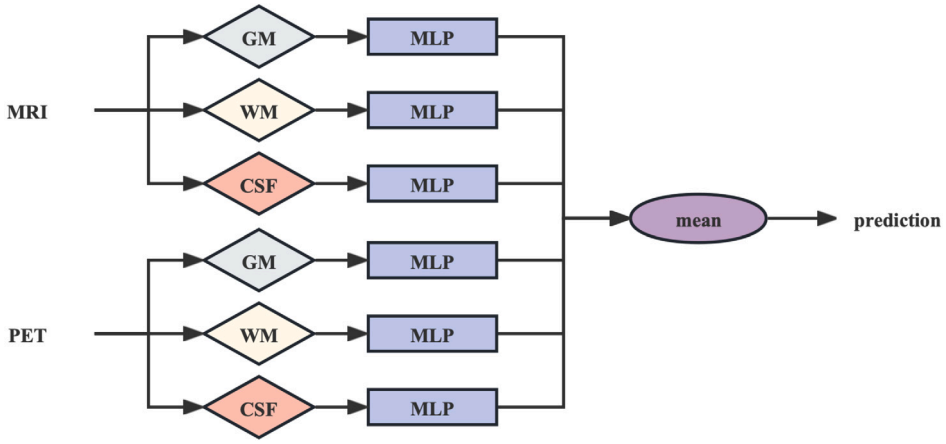

Fig. 5. Structure of MLP classifier with each MLP module consisting of 4 hidden layers, where the first hidden layer has 1024 units and halves for each of the following hidden layers.

Where ξ denotes the number of values x, y in their corresponding variables X, Y.

For the loss of the generator, it consists of two terms including the reversed loss of discriminator and the pixel-wise loss. The reversed loss of discriminator aims at enhancing the generator's ability to confuse the discriminator while the pixel-wise loss aims at accelerating the training. In this work, structural similarity index measure (SSIM) loss has been chosen to be the pixel-wise loss as it considers multiple aspects when measuring similarity between two images. Therefore, the loss for the generator is as below:

$$Loss(G) = \frac{SSIM(pred, gt) + MSE(D(pred), V)}{2} \tag{3}$$

Where G denotes the Generator and the equation of $SSIM$ is as follows:

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \tag{4}$$

Where $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ and $c_3 = c_2/2$ are used to stabilize the division with weak denominator and k_1 is set to 0.01, k_2 is set to 0.03. Meanwhile, L denotes the range of voxel values, μ denotes the mean value and σ denotes the standard deviation and σ_{XY} denotes the covariance of X, Y.

The MSE loss aims to reduce the per voxel error between predicted images and their corresponding ground truth images. However, the MSE score has a poor correlation with the perception of human visual system (HVS). According to HVS, some distortions are not clearly visible and other distortions are present but do not affect the image quality, which both cannot be reflected correctly by MSE. SSIM is a more HVS-correlated criterion that aims to extract structural information from an image by evaluating three independent highly structured parameters, namely, luminance, contrast and structure. In other words, the SSIM loss focuses on the overall image structural difference, or equivalently the inter-voxel errors while the MSE loss focuses on the per-voxel errors. Therefore, using the SSIM loss along with the MSE loss can help the model generate images more similar to the ground truth images both from the per-voxel perspective and the overall structural perspective.

Though the generator and the discriminator are both part of the U-Net-based GAN, the training is not a universe. Instead, the two parts take turns to train and optimize. In that way, the generator and the discriminator are both continuously improved in the adversarial process.

Furthermore, to evaluate the performance of the synthesized images, we also use mean average error (MAE) and Zero-Normalized Cross-Correlation (ZNCC) apart from MSE and SSIM mentioned above. The equations are as follows:

$$MAE(X,Y) = \frac{1}{\xi}\sum_{i=1}^{\xi}|y_i - x_i| \tag{5}$$

$$ZNCC(X,Y) = \frac{1}{\xi}\frac{1}{\sigma_X\sigma_Y}\sum_{i=1}^{\xi}(x_i - \mu_X)(y_i - \mu_Y) \tag{6}$$

2.4. Classifier

Three classifiers are designed in this study for the AD, MCI and NC classifications. One is a multi-layer perceptron (MLP) based classifier, and two are geometric neural network (GNN) based classifiers where one focuses on graph classification and the other focuses on node classification.

As for the MLP based classifier, it consists of several MLPs with each corresponding to one specific type of feature vector and returning the probability for each class. The results from each MLP are then averaged to produce the final prediction of class probabilities. Fig. 5 illustrates the structure of the MLP classifier with MRI-T1WI and FDG-PET images as the joint input. Furthermore, it can be adapted to single modality input. For instance, the classifier is suitable for the input of MRI-T1WI images only by eliminating the bottom three MLPs in Fig. 5 ; and similarly, when eliminating the top three MLPs in Fig. 5, it is suitable for the input of synthesized FDG-PET images only. The input of the MLP based classifier is designed for the brain atlases-based features extracted from different modalities with different brain segments, i.e., GM, WM and CSF, using AAL3 (Rolls et al., 2020) atlases. As for MRI-T1WI, the feature extraction is by counting non-zero voxels in each brain sub-region defined in the selected brain atlases as MRI-T1WI mostly reflects the degree of structural shrink of the brain in AD diagnosis. However, for the FDG-PET, feature extraction is by summing the values in each brain sub-region as the power reflects the degree of neuron activity which is an important factor in the diagnosis of neurodegenerative disease such as AD.

As for the two GNN based classifiers, both are inspired by GraphSAGE (Hamilton et al., 2017), which is a framework for inductive representation learning on large graphs extended from Graph Convolutional Network (GCN) (Kipf & Welling, 2016). The key idea of GraphSAGE is to aggregate feature representation from a node's local neighborhood. The new feature representation of the node is then derived from its previous feature representation and the aggregated feature representation of its neighbors. The equation to generate the new feature representation v_i' of a node i is given as below, where v denotes the original feature representation, $AGGR$ denotes the aggregation function, $CONCAT$ denotes the concatenation function, W

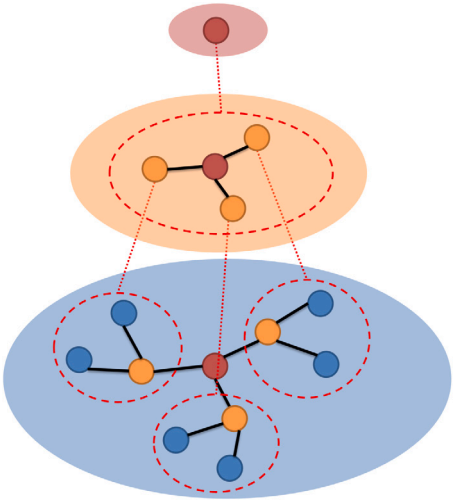

Fig. 6. Illustration for aggregation where the nodes information are aggregated in their neighborhood.

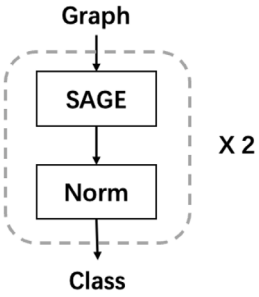

Fig. 7. Overall GraphSAGE based GNN architecture where *SAGE* represents the GraphSAGE module which is responsible for aggregation and *Norm* represents the normalization module.

denotes the weight matrices for linear transformation and σ denotes the non-linear activation function.

$$v_i' = \sigma(W * CONCAT(v_i, AGGR(v_j, \forall j \in N(i)))) \tag{7}$$

Fig. 6 shows the illustration of the aggregation procedure for the red node with twice aggregation from bottom to top. Each time a node's neighborhood information is aggregated to generate the new feature representation of that time. Therefore, with twice aggregation, the receptive field of the red node is expanded to its second-order neighborhood. The overall architecture of GNN used in this work is shown in Fig. 7 with two times of aggregation applied in the whole process.

However, for large and complex graphs, the nodes and edges can be at large size, high order aggregation with all neighbors aggregated could result in heavy memory load and long computing time, especially during the training process which requires all gradients of the weights retained for backpropagation. Using a randomly sampled fixed-size neighbor set during the training instead of the full neighbor set is one way to avoid out-of-memory and reduce computing time. Moreover, sampling can be seen as a data augmentation procedure for the graph by generating sub-graphs. With different sub-graphs fed into the model at each iteration of training, it also makes the trained model more robust. As for inferencing, the sampling strategies are removed to ensure the completeness of the data input and reduce bias.

As aforementioned, there are two GNN based classifiers designed in this research, one is responsible for graph classification, and the other is responsible for node classification. For GNN based graph classifier, a set of graphs are generated where each graph is responsible for one specific sample and reflects the brain features on the spatial structure. The nodes of the graph are the feature representations of brain regions defined in AAL3 atlases. The edges are undirected and established for the nodes which are adjacent to each other based on the AAL3 atlases. However, there are some brain regions independent and have no direct adjacent regions. Therefore, a loose adjacent definition that allows 5 voxels gap, equivalently 10 mm in the brain, is used to ensure no independent nodes. A visualization of the constructed graph within the brain is shown in Fig. 8. Since all graphs share the same structure and have only a small amount of nodes and limited edges, sampling strategies are not used during the training process, and the final classification is based on the pooling results of all nodes' feature representations in the last graph convolutional layer.

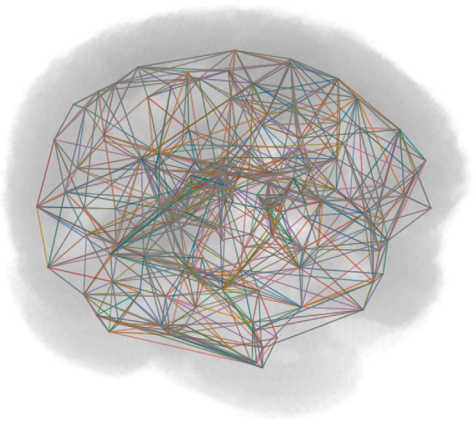

Fig. 8. A visualization of the constructed brain graph with each node containing the feature representation of the corresponding brain region and the line representing the edges in the graph.

For GNN based node classifier, only one graph is generated. The nodes of the graph are the feature representations of samples. And the edge is established with additional information including the ages at the image acquisition time and the genotypes of APOE A1 and A2 of the patients of the samples belonging to as well as sex. The edge establish algorithm is shown in Algorithm1 with an illustration of the graph construction workflow shown in Fig. 9.

The feature representations of inputs for both GNN based classifiers are extracted with the same strategy as the MLP based classifier, which counts the number of voxels for MRI-T1WI and sums the power of voxels for FDG-PET with different brain segments based on AAL3 atlases.

Furthermore, the loss function used for all classifiers is cross entropy as the below equation shows:

$$Loss(C) = -\sum_{i=1}^{c} p(x_i) log(q(x_i)) \tag{8}$$

Where c denotes the number of classes, p denotes the ground truth probability distribution and q denotes the predicted probability distribution.

3. Experimental results and analysis

3.1. Performance of synthesis model

We have trained the U-Net based GAN model on Nvidia V100 32 GB GPU for 400 epochs with the batch size set to 10. We have conducted 5-fold cross validation on the dataset of MRI-PET pairs with 80% for training and 20% for testing. During the training, we normalize the value of the input into the range 0-1 with the formula:

$$Normalize(X) = \frac{X - min(X)}{max(X) - min(X)} \tag{9}$$

Table 2
Comparison of works in Hu et al. (2022), Li et al. (2014), Sikka et al. (2018) and Zhang et al. (2022) with our proposed synthesis model for the prediction of FDG-PET from MRI-T1WI. For ZNCC and SSIM metrics, the higher value indicates the better performance of the prediction, and for MAE and MSE metrics, the lower value indicates the better performance of the prediction.

Authors	Methods	MAE	MSE	ZNCC	SSIM
Li et al. (2014)	3D CNN	0.0862	–	–	0.5419
Sikka et al. (2018)	3D U-Net	0.0422	–	–	0.8211
Hu et al. (2022)	Bidirectional Mapping GAN (BMGAN)	–	–	–	0.8900
Zhang et al. (2022)	Brain PET GAN (BPGAN)	0.0318	–	–	0.7294
Ours	GAN	**0.0141**	**0.0014**	**0.9871**	**0.9714**

Table 3
Experimental results for MLP based classifier with the input of MRI-T1WI only and the joint input of MRI-T1WI and synthesized FDG-PET.

	MRI-T1WI				Joint input			
	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
NC	0.9371	0.9400	0.9386		0.9393	0.9442	0.9417	
MCI	0.8420	0.8668	0.8542	87.21%	0.8622	0.8705	0.8663	**88.39%**
AD	0.8481	0.8081	0.8276		0.8568	0.8399	0.8483	

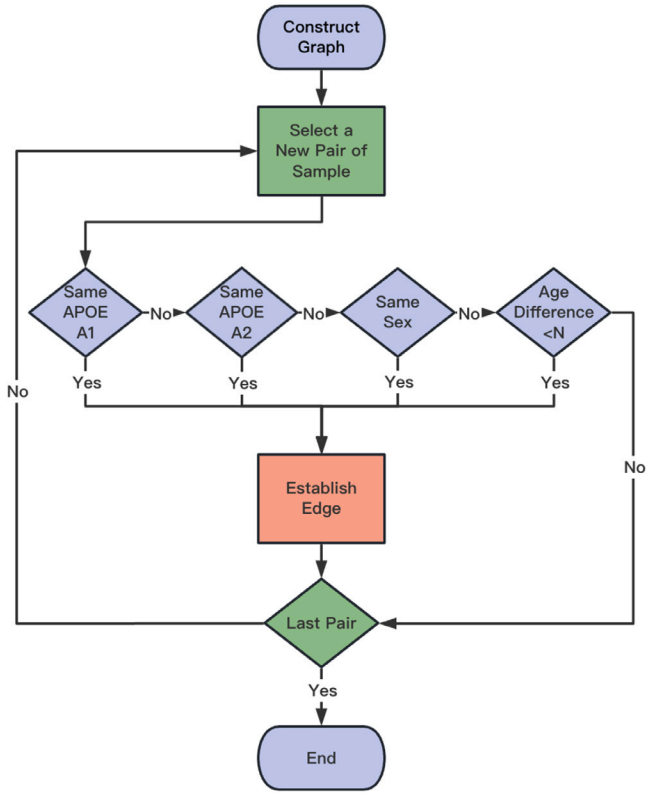

Fig. 9. Workflow of constructing a graph with each node as a feature representation of a specific sample with several criteria, namely APOE A1, APOE A2, sex and age difference.

The four criteria metrics including MAE, MSE, ZNCC and SSIM are used to evaluate the performance of the synthesized model. While computing the criteria metrics, the ground truth FDG-PET image and the synthesized one are also normalized to 0-1 to produce a standard comparison.

The qualitative results of our proposed synthesis model are shown in Fig. 10. Table 2 shows the performance of our proposed method with all four criteria metrics compared to others' works with criterion reported in their works. Our method achieves better performance compared with other works, which indicates that our method produces better synthesis results. Moreover, our methods achieve a ZNCC score of 0.9871.

Early works of synthesizing PET from MRI use plain 3D CNN (Li et al., 2014) and 3D U-NET based model (Sikka et al., 2018) with pixel-wise loss to extract features and reconstruct PET in 3D view. These 3D CNN based methods focus on pixel-wise loss while ignoring perceptual loss which infects their robustness. The latter two works (Hu et al., 2022; Zhang et al., 2022), which were based on GAN, also had 3D U-Net like backbones for generators. These two GAN based methods had better synthesizing performance compared to the previous works due to their involvements of adversarial loss and perceptual loss. Both of the works attempted to involve the latent vectors corresponding to the desired PET scans as part of the input, which increases the diversity. However, the latent vectors can be randomly generated in Gaussian space which produces results that are inconsistent over time, therefore insufficiently precise for use in the medical field. In our work, we only use MRI-T1WI as input to synthesize PET through the generator trained with pixel-wise loss, adversarial loss and perceptual loss.

3.2. Performance of classifier

We train the three types of classifiers on the extended dataset and the 10-fold cross validation experiments are conducted on all three types of experiments.

As for the MLP based classifier, we train 6 classifiers where each classifier is responsible for one type of feature extracted from the modalities of GM, WM and CSF of MRI-T1WI and synthesized FDG-PET. With an average of GM, WM and CSF prediction scores for MRI-T1WI, the final prediction score for MRI-T1WI is obtained. Furthermore, the final prediction score for the joint input of MRI-T1WI and synthesized FDG-PET is derived in the same way.

The experimental results for the MLP based classifier are shown in Table 3. With MRI-T1WI as input, the performance of the MLP based classifier has achieved an overall accuracy of 87.21% and all F1-scores are over 0.82. With MRI-T1WI and synthesized FDG-PET as the joint input, the performance of the MLP based classifier has achieved an overall accuracy of 88.39% and all F1-scores are over 0.84. Moreover, regarding all criteria metrics including precision, recall and F1, the MLP based classifier with the joint input setting performs better than the MRI-T1WI input setting. It proves the hypothesis that the synthesized FDG-PET can enhance the performance of the classifier for AD diagnosis including its early-stage MCI.

As for GNN based graph classifier whose inputs are brain graphs of samples, instead of 6 classifiers, we train the 2 classifiers, where each classifier is responsible for one modality in MRI-T1WI and synthesized FDG-PET. Due to the graph structure, if we train 6 classifiers with a similar strategy as that of MLP based classifier, the feature representation for each node should have only one value, which contains less information. Therefore, we decide to train only 2 classifiers as discussed

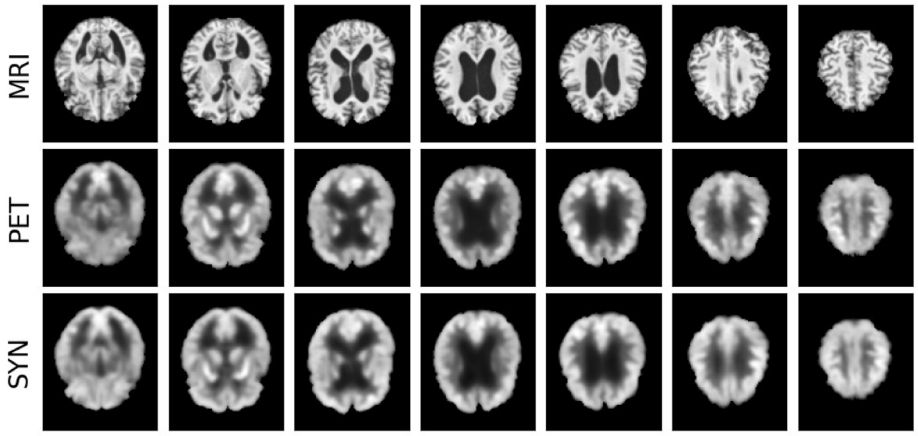

Fig. 10. Qualitative results of the synthesis model: images in the first row are slices of MRI, the second and third rows are PET and SYN (synthesized PET) slices corresponding to the first row.

Table 4
Experimental results for GNN based graph classifier with the input of MRI-T1WI only and the joint input of MRI-T1WI and synthesized FDG-PET.

	MRI-T1WI				Joint input			
	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
NC	0.8805	0.9297	0.9044		0.8840	0.9380	0.9102	
MCI	0.8403	0.7704	0.8038	**82.64%**	0.8405	0.7756	0.8067	82.23%
AD	0.7518	0.8004	0.7754		0.7662	0.8048	0.7850	

Table 5
Average edges per node for the different settings of age difference allowance.

Years	1	2	4	6
Edges	53.24	101.54	191.69	272.08

before. In that case, for each node, the feature representation has a length of 3 with feature values corresponding to WM, GM and CSF. The prediction for the joint input of MRI-T1WI and synthesized FDG-PET is obtained in the same way as that of MLP based classifier.

The experimental results for GNN based graph classifier have been shown in Table 4. The performance for input as MRI-T1WI has achieved an overall accuracy of 82.64% and all F1-score over 0.77. The performance for the joint input has achieved an overall accuracy of 83.23% and all F1-score over 0.78. With the synthesized FDG-PET as additional input, the performance is also enhanced, which is similar to that of MLP based classifier, and further proves that the synthesized FDG-PET helps AD diagnosis including its early-stage MCI.

As for GNN based node classifier whose inputs are large graphs with nodes containing feature representations of samples, we also train 6 classifiers with different modalities as input, i.e., GM, WM and CSF of MRI-T1WI and synthesized FDG-PET. The prediction score for the input of MRI-T1WI only and the joint input of MRI-T1WI and synthesized FDG-PET are derived in the same way as that for MLP based classifier by simply averaging the scores for their corresponding classifier set.

Furthermore, as shown in Fig. 9, the age difference allowance for constructing edge can be varied to adjust the degree of connection restriction between nodes. With the large age difference allowance, the connection restriction should be loose, which results in more edges for each node, and vice versa. In the experiments, we set the age difference allowance to 1, 2, 4 and 6 years. The average edges per node for 4 graphs generated under the age difference allowance settings (1, 2, 4 and 6 years) are shown in Table 5.

The joint input accuracy for age difference allowance set to 1 year is slightly lower than using MRI-T1WI as input. However, in other trials of experiments, using the joint input of MRI-T1WI and synthesized FDG-PET shows better or at least equal accuracy compared with using MRI-T1WI as input. Moreover, the average accuracy on 4 trials of experiments also shows that using the joint input of MRI-T1WI and synthesized FDG-PET performs better than using the input of MRI-T1WI only, which proves our hypothesis again.

The above experiments all involve multimodal data input including MRI-T1WI and synthesized FDG-PET and their subregions of GM, WM and CSF. And all experiments share the same data fusion strategies at the decision level by averaging the prediction score. We also experiment with a different data fusion strategy at the input level by concatenating different modal features into one feature vector.

The results for three types of classifiers with all modal features fused at the input level, including WM, GM and CSF of MRI-T1WI and synthesized FDG-PET, are shown in Table 7. The overall performance drops for both MLP based and GNN based graph classifiers compared with that in Table 3 and Table 4 while the performance enhanced for GNN based node classifier compared with that in Table 6.

According to our experimental results, we have found that GNN based node classifier performs better than MLP based classifier than GNN based graph classifier among all experiments with different settings. This might be due to the fact that each node in the brain graph contains less information which somehow affects its resistance to disturbance. Also, MLP based classifier can be seen as a very special node classifier with all nodes in the graph independent, equivalently no edges. With the additional information like APOE A1, A2 genotypes and sex introduced to generate edges, GNN based node classifier performs better than MLP based classifier as expected.

We also compare our work with other state-of-the-art methods and ours GNN based node classifier achieves the best performance for 3-class classification (AD/MCI/NC). Also, we have trained the binary classification models for comparison with other works. It also shows in Table 8 our work's robustness with AD/NC and MCI/NC achieves accuracy at 98.72% and 95.83% respectively. The accuracy for AD/MCI is slightly lower at 89.96% compared with (Kruthika et al., 2019). However, they used MRI and PET as input while only MRI is actually needed in our work for inference. Therefore, our work still has state-of-the-art performance.

To further validate the proposed methods, experiments for prediction AD conversion are performed by re-labeling some MCI subjects into sMCI and pMCI. The MCI subjects who progress to AD are labeled as

Table 6
Experimental results for GNN based node classifier with the input of MRI-T1WI only and the joint input of MRI-T1WI and synthesized FDG-PET.

		MRI-T1WI				Joint input			
		Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
1 year	NC	0.9536	0.9556	0.9551	**89.31%**	0.9466	0.9535	0.9500	89.10%
	MCI	0.8816	0.8712	0.8764		0.8822	0.8653	0.8737	
	AD	0.8645	0.8586	0.8525		0.8453	0.8629	0.8540	
2 year	NC	0.9490	0.9628	0.9559	88.82%	0.9477	0.9566	0.9521	88.85%
	MCI	0.8768	0.8639	0.8703		0.8786	0.8624	0.8704	
	AD	0.8399	0.8454	0.8462		0.8405	0.8553	0.8478	
4 year	NC	0.9419	0.9555	0.9487	88.23%	0.9527	0.9586	0.9557	88.73%
	MCI	0.8725	0.8558	0.8440		0.8729	0.8639	0.8683	
	AD	0.8333	0.8443	0.8388		0.8391	0.8465	0.8428	
6 year	NC	0.9479	0.9586	0.9532	88.63%	0.9468	0.9576	0.9511	88.63%
	MCI	0.8793	0.8580	0.8685		0.8770	0.8602	0.8685	
	AD	0.8319	0.8520	0.8418		0.8360	0.8498	0.8428	
Average					88.59%				**88.82%**

Table 7
Experimental results of data fusion at input level for three types of models.

			Joint input (input level data fusion)			
			Precision	Recall	F1	Accuracy
MLP based classifier		NC	0.9234	0.9349	0.8664	86.47%
		MCI	0.8626	0.8315	0.8468	
		AD	0.8072	0.8399	0.8323	
Graph classification classifier		NC	0.8815	0.9235	0.9020	83.08%
		MCI	0.8377	0.7785	0.8070	
		AD	0.7682	0.8103	0.7887	
Node classification classifier	1 year	NC	0.9477	0.9555	0.9516	**90.18%**
		MCI	0.8980	0.8742	0.8859	
		AD	0.8596	0.8860	0.8726	
	2 year	NC	0.9395	0.9628	0.9510	89.84%
		MCI	0.9035	0.8609	0.8817	
		AD	0.8487	0.8860	0.8670	
	4 year	NC	0.9437	0.9535	0.9486	89.13%
		MCI	0.8917	0.8543	0.9726	
		AD	0.8373	0.8805	0.8584	
	6 year	NC	0.9383	0.9586	0.9483	89.35%
		MCI	0.8930	0.8589	0.8759	
		AD	0.8471	0.8750	0.8608	

Table 8
Results for additional experiments on the identification of NC, sMCI and pMCI.

Methods	Modalities	AD/MCI/NC	AD/NC	AD/MCI	MCI/NC
Liu et al. (2018)	PET	n/a	91.2	n/a	78.9
Xia et al. (2020)	MRI	n/a	94.19	n/a	79.10
Hosseini-Asl et al. (2016)	MRI	n/a	91.4	70.1	77.4
Zhang et al. (2019)	MRI + PET	n/a	98.6	88.2	88.0
Kruthika et al. (2019)	MRI + PET	89.10	97.60	**95.00**	90.80
Ours	MRI + Synthesized PET	**90.18**	**98.72**	89.96	**95.83**

pMCI and those who do not progress to AD for more than 3 years are labeled as sMCI. Based on the re-labeled data, we trained a GNN based node classifier for identifying NC, sMCI and pMCI with joint input of MRI-T1WI features and synthesized FDG-PET features. The results are shown in Table 9 with an overall accuracy of 82.77%. The pMCI class performs lower among the three classes with an F1-score of 75.54 while the sMCI class has an F1-score of 91.08 and the NC class of 83.49. Though the model has shown weakness in identifying pMCI, it still has shown its feasibility in predicting AD conversion.

Table 9
Results for additional experiments on the identification of NC, sMCI and pMCI.

		Precision	Recall	F1	Accuracy
NC/sMCI/pMCI	NC	78.66	89.21	83.49	82.77%
	sMCI	85.43	97.53	91.08	
	pMCI	85.36	61.57	71.54	

4. Discussion

4.1. Advantages and limitation

We propose a U-Net based GAN model which takes advantage of both U-Net model and GAN model, where the former merges high-level and low-level features, the latter not only optimizes the generator from pixel-wise loss but also from the discriminator loss. Our model has been proven effective owing to the performance improved by machine learning algorithms with all three types of classifiers for AD diagnosis. Though the synthesized FDG-PET is with a very high similarity (SSIM 0.9714) to the ground truth (real image) and has proven the promising improvements on the performance of AD diagnosis, it can only be acted as an important intermediate step for machine learning models to aid the diagnosis of AD instead of its immediate clinical utility according to the rigor required in clinical practice. In other words, the synthesized

FDG-PET can only be treated as an additional reference in practice instead of a piece of evidence.

Moreover, the experimental results have demonstrated the advanced nature of our proposed GNN based node classifier with graphs constructed from the image and phenotypic data to achieve an accuracy of 90.18% for NC/MCI/AD classification. Our model has also shown a promising result in identifying NC/sMCI/pMCI with an accuracy of 82.77%. However, our proposed GNN based node classifier relies on the graph construction with patients' genotypes of APOE A1 and A2, which requires an additional step to acquire the genetic information.

In addition, the model is trained on data from the ADNI dataset, which has relatively higher-quality data. However, in actual practice, many medical centers have lower-quality data from the perspectives of either data completion or image resolution, which might make the model result less reliable, especially for those with much lower-quality data compared to the ADNI dataset.

4.2. Future directions

As for medical image synthesis, apart from the methods proposed in this work, diffusion model, a recent research hotspot in the field of natural image synthesis with representative works such as guided diffusion and stable diffusion, might also be a possible solution. Though diffusion model usually starts with noised images which might further introduce bias to the prediction, it is still worth trying due to its high performance in natural image synthesis. Meanwhile, synthesizing other modalities might also be a promising way to help improve AD diagnosis.

Regarding automated computer-aided diagnosis for AD, using more modalities of medical images, for instance, amyloid-PET and fMRI, might also have a positive effect on the results. Meanwhile, involving more phenotypic data might also improve the performance of automated computer-aided diagnosis for AD. Moreover, techniques to solve data incompletion, such as setting the default data and synthesizing the missing data, can also be applied by reason that some data has difficulty in acquisition in real practice, which might result in the derived model more robust and generalizable.

Apart from diagnosis, risk prediction for AD is also an important aspect in preventing progress to AD for patients at the pre-AD stage as well as for healthy people without any cognitive impairment. Herein, AI with medical imaging and other phenotypic data is a promising way to predict AD risk for the former; however, it might be more difficult to predict AD risk for the latter. Compared to medical imaging, genetic data might have advantages in AD risk prediction for the latter, which can be analyzed by algorithms such as fractional-order time-delayed genetic regulatory networks (Pratap et al., 2022). The derived result might be further used to help AI models for AD prediction at an earlier stage.

Moreover, arithmetic optimization algorithms such as prairie dog optimization algorithm (Ezugwu et al., 2022), dwarf mongoose optimization algorithm (Agushaka et al., 2022), aquila optimizer (Abualigah et al., 2021), Reptile Search Algorithm (Abualigah et al., 2022) and Ebola Optimization Search Algorithm (Oyelade et al., 2022), can be used to further optimize the hyper-parameters in the training process. These types of optimization algorithms can also be applied to the process of model space search to suggest a more robust model structure for the task.

5. Conclusion

This paper proposes a new method for synthesizing FDG-PET from MRI-T1WI in AD diagnosis including its early-stage MCI. With our proposed U-Net based GAN, we have delivered a state-of-the-art model with MAE at 0.0141, MSE at 0.0014, ZNCC at 0.9871 and SSIM at 0.9714. Furthermore, we have conducted comparative experiments on the synthesized FDG-PET images.

Moreover, we compare the performance of one MLP based classifier and two GNN based classifiers with different aims, where one is for graph classification and the other is for node classification. We set different inputs in the experiments, i.e., the input of MRI-T1WI only and the joint input of MRI-T1WI and synthesized FDG-PET, to validate if the synthesized FDG-PET aids the diagnosis of AD including its early-stage MCI. We also explore the effect of different data fusion strategies for multimodal input.

The experimental results show that the synthesized data enhances the models' performance for all three types of classifiers. The best-performing classifier is GNN based node classifier with feature-level data fusion and achieves the best accuracy at 90.18% for 3-class classification. This work provides support for the notion that machine learning-derived image analysis may be a useful approach to improving the diagnosis of AD.

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.

Data availability

Data will be made available on request.

Acknowledgments

This research is supported by Ningbo Major Science & Technology Project under Grant 2022Z126; Medical Research Council, UK under Grant MR/T005580/1; National Institute of Health/NIA, USA under Grant 1R56AG074467-01.

References

Abualigah, L., Abd Elaziz, M., Sumari, P., Geem, Z. W., & Gandomi, A. H. (2022). Reptile search algorithm (RSA): A nature-inspired meta-heuristic optimizer. *Expert Systems with Applications*, *191*, Article 116158. http://dx.doi.org/10.1016/j.eswa.2021.116158.

Abualigah, L., Yousri, D., Abd Elaziz, M., Ewees, A. A., Al-Qaness, M. A., & Gandomi, A. H. (2021). Aquila optimizer: a novel meta-heuristic optimization algorithm. *Computers & Industrial Engineering*, *157*, Article 107250. http://dx.doi.org/10.1016/j.cie.2021.107250.

Agushaka, J. O., Ezugwu, A. E., & Abualigah, L. (2022). Dwarf mongoose optimization algorithm. *Computer Methods in Applied Mechanics and Engineering*, *391*, Article 114570. http://dx.doi.org/10.1016/j.cma.2022.114570.

Cheng, J., Liu, Z., Guan, H., Wu, Z., Zhu, H., Jiang, J., Wen, W., Tao, D., & Liu, T. (2021). Brain age estimation from MRI using cascade networks with ranking loss. *IEEE Transactions on Medical Imaging*, *40*(12), 3400–3412. http://dx.doi.org/10.1109/TMI.2021.3085948.

Chincarini, A., Bosco, P., Calvini, P., Gemme, G., Esposito, M., Olivieri, C., Rei, L., Squarcia, S., Rodriguez, G., Bellotti, R., Cerello, P., De Mitri, I., Retico, A., & Nobili, F. (2011). Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *NeuroImage*, *58*(2), 469–480. http://dx.doi.org/10.1016/j.neuroimage.2011.05.083.

Ezugwu, A. E., Agushaka, J. O., Abualigah, L., Mirjalili, S., & Gandomi, A. H. (2022). Prairie dog optimization algorithm. *Neural Computing and Applications*, *34*(22), 20017–20065.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Vol. 30*, In *Advances in neural information processing systems*.

Ho, N.-H., Yang, H.-J., Kim, J., Dao, D.-P., Park, H.-R., & Pant, S. (2022). Predicting progression of Alzheimer's disease using forward-to-backward bi-directional network with integrative imputation. *Neural Networks*, *150*, 422–439. http://dx.doi.org/10.1016/j.neunet.2022.03.016.

Hosseini-Asl, E., Keynton, R., & El-Baz, A. (2016). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In *2016 IEEE international conference on image processing (ICIP)* (pp. 126–130). IEEE, http://dx.doi.org/10.1109/ICIP.2016.7532332.

Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., & Shen, Y. (2022). Bidirectional mapping generative adversarial networks for brain MR to PET synthesis. *IEEE Transactions on Medical Imaging*, *41*(1), 145–157. http://dx.doi.org/10.1109/TMI.2021.3107013.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, *62*(2), 782–790. http://dx.doi.org/10.1016/j.neuroimage.2011.09.015.

Khojaste-Sarakhsi, M., Haghighi, S. S., Ghomi, S. F., & Marchiori, E. (2022). Deep learning for Alzheimer's disease diagnosis: A survey. *Artificial Intelligence in Medicine*, *130*, Article 102332. http://dx.doi.org/10.1016/j.artmed.2022.102332.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Kruthika, K., Rajeswari, & Maheshappa, H. (2019). CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis. *Informatics in Medicine Unlocked*, *14*, 59–68. http://dx.doi.org/10.1016/j.imu.2018.12.001.

Li, W., Lin, X., & Chen, X. (2020). Detecting Alzheimer's disease based on 4D fMRI: An exploration under deep learning framework. *Neurocomputing*, *388*, 280–287. http://dx.doi.org/10.1016/j.neucom.2020.01.053.

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., & Ji, S. (2014). Deep learning based imaging data completion for improved brain disease diagnosis. In *International conference on medical image computing and computer-assisted intervention* (pp. 305–312). Springer, http://dx.doi.org/10.1007/978-3-319-10443-0_39.

Liu, M., Cheng, D., Yan, W., & Initiative, A. D. N. (2018). Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Frontiers in neuroinformatics*, *12*, 35. http://dx.doi.org/10.3389/fninf.2018.00035.

Liu, C., Huang, F., & Qiu, A. (2023). Monte Carlo ensemble neural network for the diagnosis of Alzheimer's disease. *Neural Networks*, *159*, 14–24. http://dx.doi.org/10.1016/j.neunet.2022.10.032.

Liu, M., Zhang, D., Shen, D., & Initiative, A. D. N. (2015). View-centralized multi-atlas classification for Alzheimer's disease diagnosis. *Human Brain Mapping*, *36*(5), 1847–1865.

Michaud, T. L., Su, D., Siahpush, M., & Murman, D. L. (2017). The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes. *Dementia and Geriatric Cognitive Disorders Extra*, *7*(1), 15–29. http://dx.doi.org/10.1159/000452486.

Nguyen, D., Nguyen, H., Ong, H., Le, H., Ha, H., Duc, N. T., & Ngo, H. T. (2022). Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. *IBRO Neuroscience Reports*, *13*, 255–263.

Oyelade, O. N., Ezugwu, A. E.-S., Mohamed, T. I. A., & Abualigah, L. (2022). Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm. *IEEE Access*, *10*, 16150–16177. http://dx.doi.org/10.1109/ACCESS.2022.3147821.

Pan, X., Phan, T.-L., Adel, M., Fossati, C., Gaidon, T., Wojak, J., & Guedj, E. (2021). Multi-view separable pyramid network for AD prediction at MCI stage by 18f-FDG brain PET imaging. *IEEE Transactions on Medical Imaging*, *40*(1), 81–92. http://dx.doi.org/10.1109/TMI.2020.3022591.

Park, J. H., Cho, H. E., Kim, J. H., Wall, M. M., Stern, Y., Lim, H., Yoo, S., Kim, H. S., & Cha, J. (2020). Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ Digital Medicine*, *3*(1), 46. http://dx.doi.org/10.1038/s41746-020-0256-0.

Park, S., Hong, C. H., gi Lee, D., Park, K., & Shin, H. (2023). Prospective classification of Alzheimer's disease conversion from mild cognitive impairment. *Neural Networks*, *164*, 335–344. http://dx.doi.org/10.1016/j.neunet.2023.04.018.

Parmar, H. S., Nutter, B., Long, R., Antani, S., & Mitra, S. (2020). Deep learning of volumetric 3D CNN for fMRI in alzheimer's disease classification. *Vol. 11317*, In *Medical imaging 2020: Biomedical applications in molecular, structural, and functional imaging* (pp. 66–71). SPIE, http://dx.doi.org/10.1117/12.2549038.

Parmar, H., Nutter, B., Long, R., Antani, S., & Mitra, S. (2020). Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data. *Journal of Medical Imaging*, *7*(5), Article 056001. http://dx.doi.org/10.1117/1.JMI.7.5.056001.

Patterson, C. (2018). *World alzheimer report 2018*, Alzheimer's Disease International.

Pinto, T. C. C., Machado, L., Bulgacov, T. M., Rodrigues-Júnior, A. L., Costa, M. L. G., Ximenes, R. C. C., & Sougey, E. B. (2019). Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? *International Psychogeriatrics*, *31*(4), 491–504. http://dx.doi.org/10.1017/S1041610218001370.

Pratap, A., Raja, R., Agarwal, R. P., Alzabut, J., Niezabitowski, M., & Hincal, E. (2022). Further results on asymptotic and finite-time stability analysis of fractional-order time-delayed genetic regulatory networks. *Neurocomputing*, *475*, 26–37. http://dx.doi.org/10.1016/j.neucom.2021.11.088.

Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J., & Joliot, M. (2020). Automated anatomical labelling atlas 3. *NeuroImage*, *206*, Article 116189. http://dx.doi.org/10.1016/j.neuroimage.2019.116189.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Sharma, R., Goel, T., Tanveer, M., Lin, C. T., & Murugan, R. (2023). Deep-learning-based diagnosis and prognosis of Alzheimer's disease: A comprehensive review. *IEEE Transactions on Cognitive and Developmental Systems*, *15*(3), 1123–1138. http://dx.doi.org/10.1109/TCDS.2023.3254209.

Sikka, A., Peri, S. V., & Bathula, D. R. (2018). MRI to FDG-PET: cross-modal synthesis using 3D U-Net for multi-modal Alzheimer's classification. In *International workshop on simulation and synthesis in medical imaging* (pp. 80–89). Springer, http://dx.doi.org/10.1007/978-3-030-00536-8_9.

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(1s), 1–35. http://dx.doi.org/10.1145/3344998.

Tuan, P. M., Phan, T.-L., Adel, M., Bourennane, S., Trung, N. L., & Guedj, E. (2022). C-Atlas: A brain mapping based on FDG-PET images for Alzheimer's disease diagnosis. In *2022 RIVF international conference on computing and communication technologies (RIVF)* (pp. 150–155). IEEE, http://dx.doi.org/10.1016/j.ibneur.2022.08.010.

Wang, C., Li, Y., Tsuboshita, Y., Sakurai, T., Goto, T., Yamaguchi, H., Yamashita, Y., Sekiguchi, A., Tachimori, H., & Initiative, A. D. N. (2022). A high-generalizability machine learning framework for predicting the progression of Alzheimer's disease using limited data. *NPJ Digital Medicine*, *5*(1), 43. http://dx.doi.org/10.1038/s41746-022-00577-x.

Wee, C.-Y., Yap, P.-T., Shen, D., & Initiative, A. D. N. (2013). Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. *Human Brain Mapping*, *34*(12), 3411–3425.

Xia, Z., Yue, G., Xu, Y., Feng, C., Yang, M., Wang, T., & Lei, B. (2020). A novel end-to-end hybrid network for Alzheimer's disease detection using 3D CNN and 3D CLSTM. In *2020 IEEE 17th international symposium on biomedical imaging (ISBI)* (pp. 1–4). http://dx.doi.org/10.1109/ISBI45749.2020.9098621.

Zhang, J., He, X., Qing, L., Gao, F., & Wang, B. (2022). BPGAN: brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis. *Computer Methods and Programs in Biomedicine*, *217*, Article 106676. http://dx.doi.org/10.1016/j.cmpb.2022.106676.

Zhang, F., Li, Z., Zhang, B., Du, H., Wang, B., & Zhang, X. (2019). Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing*, *361*, 185–195. http://dx.doi.org/10.1016/j.neucom.2019.04.093.

Zhu, W., Sun, L., Huang, J., Han, L., & Zhang, D. (2021). Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI. *IEEE Transactions on Medical Imaging*, *40*(9), 2354–2366. http://dx.doi.org/10.1109/TMI.2021.3077079.