Information Fusion 109 (2024) 102415

Contents lists available at ScienceDirect

Information Fusion

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

A coupled-GAN architecture to fuse MRI and PET image features for multi-stage classification of Alzheimer's disease

Chandrajit Choudhury [a], Tripti Goel [a,*], M. Tanveer [b]

[a] Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Assam, 788010, India
[b] Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore, 453552, India

ARTICLE INFO

Keywords:
Adversarial learning
Alzheimer's disease
Magnetic resonance imaging
Positron emission tomography

ABSTRACT

Alzheimer's disease (AD) is a degenerative neurological ailment that begins with memory loss and ultimately leads to a total loss of mental capacity. Researchers are interested in using magnetic resonance imaging (MRI) and positron emission tomography (PET) to find people with mild cognitive impairment (MCI), which is a stage before Alzheimer's disease (AD). Significant hippocampal loss and temporal lobe atrophy characterize the transition from MCI to AD, which can be visualized using T1-W structural MRI. PET visualizes brain glucose metabolism, which indicates neuronal activity, making it a viable neuroimaging method for AD diagnosis. The extraction and fusion of structural and metabolite information about brain alterations contained in multimodal data is crucial for achieving an appropriate classification result. Therefore, in this work a new end-to-end coupled-GAN (CGAN) architecture is introduced. The proposed CGANC network consists of two sub-models: a CGAN for extraction of fused features from multimodal data, and a CNN classifier to classify these features. The proposed CGAN model is trained to encode MRI and PET images into a shared latent space. The fused features are extracted from this shared latent space and then are classified according to particular stage of AD. In order to test the effectiveness of the suggested approach, experiments are done on the publicly available ADNI dataset and compared with state-of-the-art methods. The proposed method's source code will be made freely available at https://github.com/ChandrajitChoudhury/CGAN-AD.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder that primarily affects the elderly and causes a gradual decline in cognitive ability over time [1]. Cortical anatomical anomalies are permanent and part of the cognitive decline condition. AD accounts for 60–80 percent of all dementia cases. The estimated global cost of dementia care in 2023 is one trillion dollars [2]. Forecasts indicate that by 2030, this economy will have more than doubled in value from where it is now. It is crucial to identify AD in its prodrome, specifically in its transition condition known as mild cognitive impairment (MCI), in order to begin treatment and stall the progression of the disease. As there is no agreed-upon set of AD symptoms, the diagnosis was initially difficult [3]. Dementia with AD is caused by the well-known degeneration of cell loss in the brain, most notably in the cortical region, due to malfunctioning brain proteins [4]. Plaques and tangles have been cited as two key figures in AD development. Plaques and tangles are accumulated in the brain's hippocampus, basal ganglia and cortex regions of AD patients, resulting in structural atrophies of these regions. Brain plaque and tangle accumulation can be quantified by molecular studies, which are often conducted through autopsy or biopsy. However, these techniques are very painful, and AD patients are not comfortable with them.

Magnetic resonance imaging (MRI) provides detailed regional tissue characterization and increased soft-tissue contrast to understand brain structure and function. MRI imaging [5], [6] shows brain tissue in 3D and distinguishes gray and white matter in the cerebral cortex. MRI [7] allows for easy identification and assessment of brain areas, including the amygdala and thalamus, that contribute to AD. Neurodegeneration is indicated by increased tau or p-tau levels in cerebrospinal fluid, metabolism of the cortex, forebrain, and temporal cortex in positron emission tomography (PET) [8]. Cortical shrinkage, loss of gray matter volume in the medial parietal lobes, posterior cingulate and lateral temporal cortex atrophy, and hippocampus atrophy also are main biomarkers of AD disease. This paper suggests fusing MRI and PET images to efficiently diagnose AD by extracting the brain's structural and metabolic characteristics.

Many multimodal data fusion classification algorithms have been suggested for integrating MRI and PET data to take advantage of the rich discriminative information contained in both sets of data.

* Corresponding author.
E-mail addresses: chandrajit@ece.nits.ac.in (C. Choudhury), triptigoel@ece.nits.ac.in (T. Goel), mtanveer@iiti.ac.in (M. Tanveer).

https://doi.org/10.1016/j.inffus.2024.102415
Received 4 December 2023; Received in revised form 13 March 2024; Accepted 7 April 2024
Available online 10 April 2024

Typically, these techniques fall into one of three types: transform-based fusion, spatial domain-based fusion, or deep learning-based fusion. Spatial domain-based fusion involves using a specific fusion approach to combine the pixel values of each point in the multimodal images, to provide a fusion image that is more informative than each data source alone. The transform domain-based image fusion method involves converting the source image from one domain to another, such as the time domain to the frequency domain, to acquire the high-frequency coefficient and low-frequency coefficient. Both images will be fused in the frequency domain and then reconstructed. Both the transform-based and the spatial domain-based approaches of image fusion depend heavily on features that have been handcrafted, rendering them unable to adaptively characterize the inherent characteristics of various images. Further, from the fused image, original MRI and PET images cannot be reconstructed using transform-based and spatial domain-based image fusion approaches.

Research into multimodal image fusion and analysis using deep learning (DL) [9] techniques has increased in popularity [10] in the last several years. While traditional feature extraction approaches rely on hand-crafted features, DL methods show promise in multimodal image classification due to their superior capacity to discover deep features from images. DL networks use convolutional neural networks (CNN) that have the remarkable capacity to learn intrinsic information extraction and classification tasks from images in an automated and adaptive manner. A CNN model [11] with appropriate architecture can be trained to extract and down-sample local as well as global characteristics from images of various kinds, which makes the network robust. Further in-depth research is needed to determine the best way to extract modal-specific complementary information and related semantic information from multi-modal data to improve fusion classification accuracy.

Recently, adversarial learning [12] has been gaining popularity in multi-modal image fusion and classification problems to enhance learning model performance using adversarial training. The generative adversarial network (GAN), a classical adversarial learning paradigm, enhances model resilience and generalization through adversarial training between the generator and the discriminator. The adversarial training and deep neural network methods make the fusion and classification of multi-modal images more reliable and useful. GAN is employed to accomplish MRI and PET image fusion and also reconstruction of both images, with the two tasks being trained simultaneously to acquire more discriminative features by combining labeled and unlabeled examples. Li et al. [13] proposed a fusion model for PET and MRI data that uses a dense CNN network with dual attention. To generate the fused image from the information extracted from the MRI and PET images, the authors use an encoder network and a decoder made of densely connected neural networks. To further integrate local characteristics with their global dependencies adaptively, a dual-attention module is simultaneously implemented in the encoder and decoder. Nandhini et al. [14] fused the input MRI and PET images using GAN. The generator receives the concatenated input MRI and PET images. The generated image is then compared to the input image to determine the output image using adversarial learning between the discriminator and generator. Both the MRI and fused images are fed into one discriminator, while the other receives the PET and fused images as input. The latent representation from auto-encoders, especially GAN's, has been used as a feature vector for image classification in many reported works [15]. In the works of [14,16], GAN has been used to create fusion of PET and MRI images. The input here is concatenation of the MRI and PET image pair and the output is the fused image. In such approaches the latent vector's information content depends on the kind and extent of fusion produced at the output. Kang et al. [16] proposed tissue-aware conditional generative adversarial networks (TA-cGANs) for merging MRI and PET scans of the brain. The discriminator seeks to maximize the objective function by encouraging the fused image to incorporate more structural information from MRI, while the generator aims to decrease the objective function by generating a fused image mostly including PET metabolic information. Tissue label maps are produced from MRI images, and TA-cGAN's discriminator and generator are trained in a back-and-forth fashion using joint loss. Consequently, investigating adversarial learning models for processing and analyzing multi-modal data is a potential avenue for investigation.

Motivated by the above analysis, we present a new end-to-end coupled generative adversarial network-based classification (CGANC) architecture that aims to better and more reliable fusion of MRI and PET data and AD classification. The proposed network consists of two sub-networks: a multi-input, multi-output coupled GAN (CGAN) subnetwork for the fusion of features extracted from multimodal data and a DL-based classifier subnetwork to classify the fused features into particular categories. CGAN subnetwork consists of dual convolutional autoencoders and discriminators to fuse MRI and PET images. Each set of convolutional autoencoders and discriminators undergoes adversarial training to ensure that only the most relevant details from each set of data are retained in the latent space. Among these retained features, the mutually exclusive information of the two modalities are also included. The adversarial training of the CGAN, to reconstruct the input images at the output, will tend to optimize the shared-latent-space's features to contain the most comprehensive information from the images. These features will then ensure appropriate classification of the images according to the AD stages. In our work, we indirectly force the latent space to contain comprehensive and complimentary information from both MRI and PET images. The fused latent representation generated from this configuration can then be used for effective detection of the stages of AD. The adversarial part of the learning is added to get better convergence in learning the latent space. The main contributions of the paper are as follows:

- In this paper, an adversarial learning-based fusion of MRI and PET scans has been proposed for AD diagnosis. MRI scans contain structural atrophies, and PET scans consist of metabolic information. Therefore, we aim to extract structural as well as metabolic information for effective diagnosis of AD at an early stage.
- As both images are captured from different scanners with different acquisition parameters, preprocessing is done to make both images suitable for the fusion. Both scans are normalized, realigned, and registered on the standard MNI template and then co-registered with each other to make the scans suitable for image fusion.
- After co-registration, MRI and PET images are fused using adversarial learning-based CGANC which uses dual convolutional autoencoders and disciminators to fuse both MRI and PET images. Latent space features of dual encoders will be fused and fed to classifier for AD diagnosis. Images from the fused features are separated using dual decoders. Adversarial learning is added using dual discriminators to the proposed network to extract the most informative features.
- To validate the efficacy of the proposed work, significant experiments are conducted using MRI and PET images extracted from the ADNI dataset. Extensive experiments and comparing our model's performance with previous works revealed that our model outperformed all other models.

The rest of the organization for this research paper is as follows: In Section 2, we discussed the previous related works. Section 3 discusses the methodology of the proposed network, CGANC. Experiments and results are given in detail in Section 4. In Section 5 we draw the conclusion.

2. Literature review

As MRI advances, more imaging options become available to help the diagnosis of cognitive impairment. Imaging techniques such as diffusion tensor imaging (DTI), magnetic resonance spectroscopy (MRS),

single-photon emission computed tomography (SPECT), etc., help in analyzing the alterations in the brain of AD subjects. Alterations in the anterior cingulum, frontal white matter, and corpus callosum can be tracked by DTI. The frontal lobe, cingulate gyrus, parahippocampal gyrus, temporal lobe, and other brain regions show abnormalities in metabolite levels when analyzed with MRS and SPECT. Iron overload in the brain is a known contributor to mental deterioration. Many types of neurodegenerative disorders are associated with variable amounts of iron in the brain. Several studies have shown that changes in the brain's metabolism happen before the first signs of AD show up. PET imaging shows the brain's resting metabolic rates of glucose, which is a sign of neuronal activity and makes it a very promising neuroimaging tool for diagnosing AD. Thus, it is crucial to look into the possibility of a fusion of multimodality data for early-stage AD diagnosis.

Methods of medical image fusion are mainly categorized into spatial domain-based, transform-based, and deep learning-based. To generate fused images, spatial domain fusion approaches simply apply the fusion rules to the pixels of the input image. The spatial domain fusion methods include average, maximum, and minimum selection methods, intensity hue saturation (IHS) model, principal component analysis (PCA), and high-pass filtering. The IHS model mostly used for spatial domain fusion, is based on the human visual system. It has two traits: (1) intensity is unrelated to image color; and (2) hue and saturation have a profound connection to color perception. Researchers employ this model to address image fusion with color information. Haddadpour et al. [17] carried out multimodal image fusion using integrating the IHS model with the two-dimensional Hilbert transform (HT). As the discrepancy value decreases, the spectral resolution increases. This approach retains spectral properties while obtaining a low disparity. Along with successfully keeping spatial information, the approach also achieved a satisfactory average gradient. The drawback of the suggested fusion method is its low information entropy. Further, using the IHS model and Log-Gabor transform, Chen [18] proposed a new method for MRI-PET fusion, decomposing the PET picture with IHS. To determine high-frequency and low-frequency subbands in MRI and PET pictures, the Log-Gabor transform is used to decompose the intensity components, which indicate image brightness. High-frequency sub-band fusion uses maximum selection, while low-frequency sub-band fusion uses a new method called two-level fusion, combining visibility measurement and the weighted average rule. The inverse Log-Gabor transformed component and original hue and saturation components are converted to create a fused image. It effectively preserves source image structures and features while minimizing color distortion. The spatial domain fused research suffers from issues with spectral and spatial distortion. Therefore, researchers shift their attention to the transform domain to improve fusion effects.

In recent years, image fusion algorithms in the transform domain have focused on multiscale transform. The transform-based fusion method involves three steps: decomposition, fusion, and reconstruction. The transform domain image fusion approach involves transforming the image from time to frequency or other domains to produce low-frequency (LF) and high-frequency (HF) coefficients. Frequency-transform-based fusion methods have been widely used which transform the different modalities from the spatial domain into the frequency domain. Then both images will be fused in the frequency domain and transformed back to the spatial domain using the inverse transform. Shahdoosti and Mehrabi [19] introduced the modified dual ripplet-II transform using dual-tree complex wavelet, to solve the ripplet-II transform shift variance problem. The MRI and PET images are fused using the structural tensor and dual ripplet-II transform. Analyses reveal that the suggested strategy enhances visual quality and quantitative criteria based on mutual information, edge information, spatial frequency, and structural similarity. Ouerghi et al. [20] proposed Non-subsampled shearlet transform (NSST) and simplified pulse-coupled neural network model (S-PCNN) to fuse MRI–PET images. The PET image is converted to YIQ-independent components first. NSST decomposes the source registered MRI image and PET image Y-component into LF and HF subbands. LF coefficients are fused utilizing weight region standard deviation (SD) and local energy, whereas HF coefficients are mixed using S-PCCN, inspired by an adaptive-linking strength coefficient. Final inverse NSST and inverse YIQ are utilized to merge the image. Aymaz and Kose [21] presented hybrid super-resolution approach for MRI-PET fusion. First, all source images are super-resolved to improve the contrast. After that, stationary wavelet transform (SWT) is used which divides source images into four subbands after decomposing them. These subbands are LL, LH, HL, and HH. LL is the source image approximation coefficient, and others are its detail coefficients. PCA is used to pick the maximum eigenvector of each sub-band of source images to fuse images. Finally, inverse-SWT (ISWT) reconstructs fused sub-bands in the spatial domain. Dwivedi et al. [22] suggested discrete wavelet transform (DWT) to convert MRI and PET into frequency domain, then fuse both images in the frequency domain and transformed back to spatial domain using inverse wavelet transform (IWT). Fused image features have been extracted using DL network and classified using robust energy least square twin support vector machine classifier. Sharma et al. [23] proposed wavelet packet transform (WPT) domain-based fusion techniques. WPT, based on wavelet packet decomposition, addresses the drawbacks of wavelet transform (WT). Transform-based image fusion can enhance multimodal image classification performance by avoiding spatial distortions. However, transform-based fusion methods primarily use manually designed characteristics, which cannot adjust to the varied inherent properties of the image. Another issue with transform-based fusion methods is the assumption that distortions follow a Gaussian distribution, which might cause a model mismatch.

Recent research has extensively studied DL methods for multimodal image processing. Using DL networks to combine PET and MRI scans in a spatial domain may solve the issues that arise due to transform-based fusion approaches. DL-based multi-focus multimodal fusion was suggested by Liu et al. [24] as a way to fix the problem of spatial distortion for classification. With their superior ability to discover deep features from images, DL approaches show promise in improving multimodal image classification tasks, as compared to traditional manually constructed feature extraction methods. Ma et al. [25] proposed FusionGAN, a generative adversarial network to fuse infrared and visible light images. The proposed model sets up an adversarial game between a generator and a discriminator, with the former trying to produce combined images with strong infrared intensities and more visible gradients, and the latter trying to make sure that the fused image has more visible details. Because of this, the combined image can retain both the texture information of visible image and the thermal radiation information of an infrared image. Adversarial learning is gaining popularity in multimodal picture fusion and classification problems to enhance learning model performance through training. The GAN, a classical adversarial learning paradigm, enhances model resilience and generalization through adversarial training between the generator and the discriminator.

Ma et al. [26] proposed detail-preserving adversarial learning for fusing infrared and visible images. To enhance the quality of detail information and sharpen the edge of infrared targets within the framework of a GAN, the authors designed two loss functions—the detail loss and the target edge enhancement loss. The fused image keeps both the visible image's rich textural details and the infrared image's thermal radiation while sharpening the infrared target boundaries. Hang et al. [27] used GAN, to train both labeled and unlabeled samples to learn more discriminative features. Zhang et al. [28] improve the classification accuracy by fusing multimodal images with GAN-based data augmentation methods. In order to enhance fused image classification performance in situations when there are limited labeled examples, Wang et al. [29] utilized GAN to create artificial samples for data augmentation.

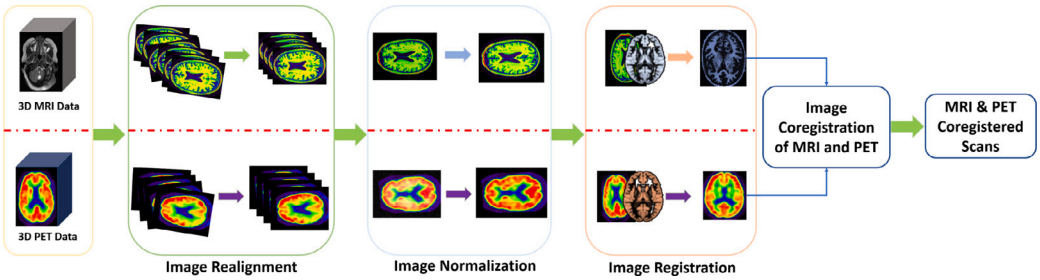


Fig. 1. Preprocessing pipeline of MRI and PET scans.

The deep neural network and adversarial training method enhance the robustness and generalization of the learned model. Exploring adversarial training models for multimodal data processing and analysis is an interesting field. There is very limited research on using adversarial learning for the fusion of medical images. Motivated by the above literature analysis, the GAN network utilizing adversarial learning produces more significant results for fusing two different modality images. Therefore, in the present paper, we utilize the advantage of adversarial learning to fuse MRI and PET images using two encoders and decoders, and two discriminators. Further, the fused data is classified using a CNN-based classifier.

In this section, we review the literature on multimodality image fusion, next section discusses the proposed methodology for MRI-PET image fusion and classification.

3. Methodology

MRI and PET images carry mutually exclusive information that are useful in detection of AD and its various stages. Therefore for a comprehensive automated diagnosis, it is necessary to consider both the structural and metabolic information from the MRI and PET images. As both image sources are different (magnet detectors and radiotracers), their acquisition parameters, acquisition machine, matrix size, voxel size, echo time, repetitive time, echo sequence, and slice number are different. Therefore, certain pre-processing has been done to make both images suitable for the fusion. After pre-processing, the adversarial network is used for image fusion and classification. The details of preprocessing and the proposed network for fusion and classification have been given in the following subsections.

3.1. Pre-processing

Pre-processing is necessary before fusing MRI and PET images due to differences in acquisition parameters, slice count, and matrix size. All 3D MRI and PET scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [30] data-set are preprocessed with the Statistical Parametric Mapping (SPM12) toolbox [31]. The preprocessing steps are shown in Fig. 1

All 3D MRI and PET scans undergo image realignment in order to eliminate motion artifacts. Normalization aims to align all input images to the standard MNI-152 template. Image normalization standardizes the intensity values in both scans. After that, both scans are registered on the standard MNI-152 template. After the image registration, the size of both scans becomes $212 \times 256 \times 256$.

To ensure that the two modalities fit together accurately, the co-registration of both scans has been done by taking the MRI image as the reference image. The co-registration entails matching each scan to the reference image slice by slice so that all scans have the same dimension.

Processing entire 3D brain scans can be a resource intensive and time-consuming. Key slice selection is employed to mitigate these challenges and achieve greater precision. The Grey Level Co-Occurrence Matrix (GLCM) is utilized to derive statistical texture characteristics that best represent the image's information. The proposed model extracts significant slices based on entropy and energy-based features. These features tend to have low values in areas of high diversity and high values in regions of low diversity. K-means clustering is applied to select meaningful slices based on differences in texture feature information, particularly benefiting slices affected by atrophy. Ten significant axial slices are extracted from each image and further processed for subsequent fusion.

3.2. Proposed architecture for fusion and classification

To fuse the mutually exclusive features from PET and MRI images, we propose to couple two Convolution Auto-Encoders (CAE) at their latent space. Auto-encoders are known to encode the input data, non-linearly, into a compressed latent vector while retaining most of the information content. In our problem, to extract a fusion of feature sets from MRI and PET images, we propose to encode the MRI and PET image separately into two separate latent vectors. These vectors are then added to form a fused latent representation. This fused vector is decoded separately by two decoder networks. One of the decoders should reproduce the input MRI image, while the other decoder should reproduce the input PET image. If this network configuration with the encoder pair and the decoder pair is trained to achieve the desired output, the fused latent space should logically contain the information required to reconstruct the MRI as well as the PET image separately. More importantly, the latent representations learned in this manner will contain sufficient information about one image type while retaining the information exclusively contained by the other image type.

For a better convergence of the training process we propose to further add adversarial learning to the above CAE model. The details of architecture of the proposed coupled generative-adversarial-network (CGAN) model and the training details are described in the following sub-sections.

3.2.1. Coupled-GAN (CGAN)

The coupled generator (CAE) part of the proposed architecture consists of two encoders and two decoders. Both the encoders have same architecture. Each of the encoder networks can be perceived as cascade of four blocks, marked as 'A', 'B', 'C' and 'D' in Fig. 2. Each of these blocks consists of a 2D convolution layer, one Batch-normalization layer, one activation layer with ReLU function and one 2D max-pooling layer. The convolution layers of the blocks A,B,C and D have $16, 32, 64$ and 128 kernels respectively. Size of each of these kernels is (3×3). All the max-pooling layers use a window size of (2×2). The output of each of the encoder is of size $(13 \times 16 \times 128)$. These two outputs are added. The final output of the encoder block i.e. the latent representation from our coupled generator network is of size $(13 \times 16 \times 128)$. This latent representation is decoded separately by two decoders to retrieve the input MRI and PET images.

The two decoder networks have the same architecture. Each of the decoder consists of four blocks, labeled 'E', 'F', 'G' and 'H' in Fig. 2. The blocks E and F consist of one up-sampling layer, one de-convolution layer, one batch normalization and one activation layer with ReLU

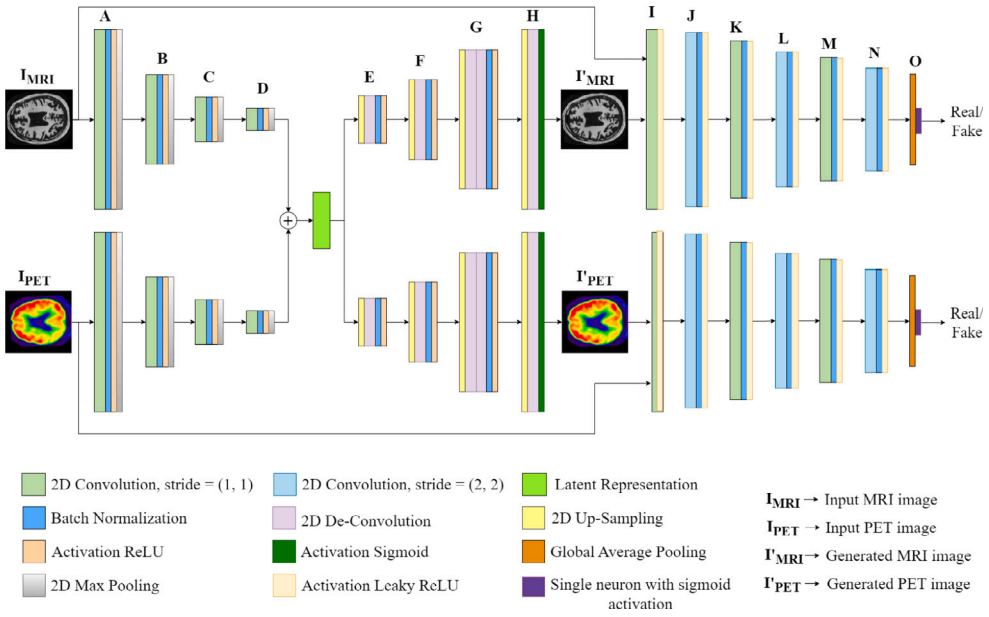


Fig. 2. Proposed GAN architecture.

function. The de-convolution layers of E and F have 64&32 kernels respectively, each with size (3 × 3). The block G has an up-sampling layer followed by two de-convolution layers, one batch normalization layer and one ReLU activation layer. The first de-convolution layer has 16 kernels of size (3×3). And the second one has 8 kernels of size (3×1). The last block 'H' has one up-sampling layer, one de-convolution layer with one kernel of size (3 × 3), followed by an activation layer with sigmoid function. Moreover in the decoder all the up-sampling layers have kernel size of (2 × 2). This generator model will yield two outputs of same shape as the inputs. The values in the two output matrices will lie in [0, 1].

The proposed CGAN architecture is trained such that the output pair, (I'_{MRI}, I'_{PET}), of the generator network is similar to the input pair, (I_{MRI}, I_{PET}). The output pair (I'_{MRI}, I'_{PET}) is criticized by two separate critic networks, one for each of the generate output images. Both the critics have the same architecture. Each of the critic has seven blocks, labeled as 'I', 'J', 'K', 'L', 'M', 'N' & 'O', as shown in Fig. 2. The block 'I' has one convolution layer and one Leaky-ReLU activation layer. Each of the blocks J, K, L, M & N has one convolution layer, one batch normalization layer and one Leaky-ReLU activation layer. All the convolution layers in all the blocks have 16 kernels of size 3 × 3. However, the convolution layers of the blocks J, L & N have stride size of (2, 2), while rest of the convolution layers have a stride size of (1, 1). The final block 'O' consists of a Global Average Pooling layer that averages and vectorizes the output feature map from block N, followed by the output layer consisting of one neuron, with sigmoid activation function. The output of the discriminator models will be a scalar value lying in [0, 1].

3.2.2. Training

A two-fold cost function is used for training this architecture. The output, (I'_{MRI}, I'_{PET}), of the generator is compared with the input in terms of mean square error (mse). And, binary cross entropy is used as the cost function for the critic networks. The generator is trained with the following cost function:

$$Cost_G = \frac{\alpha}{N}\sum_{i=1}^{N}(\{I_{MRI}\}_i - \{I'_{MRI}\}_i)^2 + (\{I_{PET}\}_i - \{I'_{PET}\}_i)^2 + \frac{(1-\alpha)}{N}\sum_{i=1}^{N} -\{log(p_i^1) + log(p_i^2)\} \quad (1)$$

where N is the number of training samples, $(\{I_{MRI}\}_i \& \{I_{PET}\}_i)$ are the MRI and PET image pair of the ith training sample. $(\{I'_{MRI}\}_i \& \{I'_{PET}\}_i)$ is the corresponding output pair of the generator model. Here, $p_i^1 \& p_i^2$ are the outputs of the two critics when the outputs of the generator are fed to the critics.

$$p_i^1 = C_1(\{I'_{MRI}\}_i)$$
$$p_i^2 = C_2(\{I'_{PET}\}_i) \quad (2)$$

where, C_1 and C_2 represent the two critic networks. For the critic the output value of '1' signifies 'real' data and the value of '0' signifies 'fake' data. The cost function for generator (1) consists of two parts: one part is computed w.r.t. the two outputs of the generator and the other part is w.r.t. the two discriminators. The parameter α represents the weightage of these two parts in the final cost function. In our experiments we have taken α to be 0.5.

For training the critics we have taken the binary-cross-entropy as the cost function. The total cost function of both the critics can be represented as:

$$Cost_{C_1,C_2} = \frac{1}{N}\sum_{i=1}^{N} -\{y_i^1 \times log(C_1(\{I_{MRI}\}_i)) + (1-y_i^1) \times log(1 - C_1(\{I'_{MRI}\}_i))\} + \frac{1}{N}\sum_{i=1}^{N} -\{y_i^2 \times log(C_2(\{I_{PET}\}_i)) + (1-y_i^2) \times log(1 - C_2(\{I'_{PET}\}_i))\} \quad (3)$$

where C_1 and C_2 represent the two critic networks and, $C_1(x)$ and $C_2(x)$ represent the output of the critics for a given input x. The variable y_i^1 and y_i^2 represent the label for the input to the two critics. If for the ith data input, the image input to the kth critic is a real image, then $y_i^k = 1$, else if the input image is fake, then $y_i^k = 0$.

The generator and the critic are trained alternately. In our experiment for every iteration over the entire training data sample, first the critic is trained for three iterations and then the generator is trained for five iterations.

3.2.3. Classifier

The latent representation from the above CGAN network is of size (13 × 16 × 128). Based on this representation the classification is done. For classification a separate convolution neural network (CNN) is built.

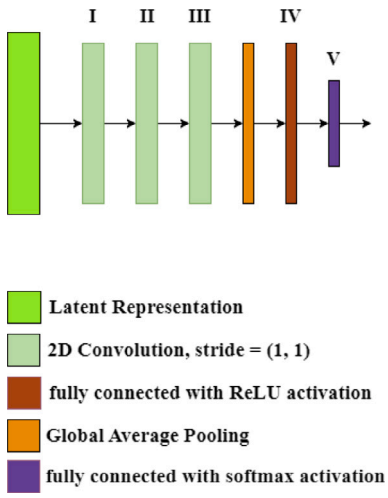


Fig. 3. Proposed CNN architecture for classifier.

The architecture of this CNN is shown in Fig. 3. The classifier consists of three 2D convolution layers, labeled as 'I', 'II', 'III' in Fig. 3. The layers have 128, 256 & 512 kernels of size (3×3). The layer III is followed by a Global Average Pooling layer and then a fully connected network. This fully connected sub-network consists of two layers, labeled as 'IV' and 'V' in Fig. 3. The layer 'IV' has 512 nodes each with ReLU activation function, and the layer 'V' has 3 nodes with softmax activation. The output of this network will be the probability distribution of the classes for a given input data.

The cost function used for training this network is categorical-cross-entropy:

$$Cost_{classifier} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{3} -\{y_i^j \times log(Class_j(L_i))\} \tag{4}$$

where y_i^j is the one-hot encoded class label for the ith input.

$$y_i^j = \begin{cases} 1 & \text{if } i\text{th input sample belongs to } j\text{th-class} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Here, L_i is the latent representation corresponding to the ith input pair $(\{I_{MRI}\}_i \& \{I_{PET}\}_i)$, extracted from the above CGAN network. $Class_j(L_i)$ is an indicator function is L_i belongs to jth class or not.

$$Class_j(L_i) = \begin{cases} 1 & \text{if } L_i \text{ belongs to } j\text{th-class} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The indicator function $Class_j(L_i)$ represents the class assigned by the classifier. For training the classifier the same training data-set is used as for training the CGAN model. Once the CGAN is trained the latent representations for the training data-set are generated from the encoder part of the CGAN. These latent representations are then used to train the classifier CNN, independent of the CGAN.

The pseudocode for the training process of CGAN model is given in Algorithm 1 and for the classifier model, is given in Algorithm 2.

The proposed network architecture is evaluated using a publicly available ADNI dataset. The next section discusses the experimental results and the comparison with state-of-the-art techniques.

Algorithm 1 Training proposed CGAN network for MRI-PET image fusion and AD diagnosis.

Network Model: CGAN network contains *a coupled generator* **G** *and two discriminators* $\mathbf{C_1}$ *and* $\mathbf{C_2}$
G contains: *two encoders* $\mathbf{E_1}$ *and* $\mathbf{E_2}$*, one fusion (addition) layer, and two decoders* $\mathbf{D_1}$ *and* $\mathbf{D_2}$.
Input: MRI Image: $\mathbf{I_{MRI}}$; PET image: $\mathbf{I_{PET}}$; Training Labels: **Y**; Number of iterations: $\mathbf{N_{iter}}$.
Output: Reconstructed MRI and PET images: $\mathbf{I'_{MRI}}$ and $\mathbf{I'_{PET}}$
procedure PREPROCESS MRI AND PET IMAGES; AND SPLIT THEM INTO TRAINING AND TEST DATA.
    Initialize the weights and Biases of CGANC network.
    while *iteration* $< N_{iter}$ **do**;
        for 3 epochs **do**
            $[I'_{MRI}, I'_{PET}] \leftarrow G[I_{MRI}, I_{PET}]$
            Using $[I'_{MRI}, I_{MRI}]$ and $[I'_{PET}, I_{PET}]$, train C_1 and C_2 to minimize eqn. (3)
            Freeze the weights of C_1 and C_2
            for 5 epochs **do**
                Train G: to minimize eqn. (2).
            end for
        end for
    end while
end procedure

Algorithm 2 Training proposed CNN Classifier

Network Model: CGAN network with trained weights from Algorithm 1
Input: MRI Image: $\mathbf{I_{MRI}}$; PET image: $\mathbf{I_{PET}}$; Training Labels: **Y**; Number of iterations: $\mathbf{M_{iter}}$.
Output: Classification into four pre-defined classes.
procedure PREPROCESS MRI AND PET IMAGES AND SPLIT THEM INTO TRAINING AND TEST DATA.
    Initialize CGANC network with parameters from k^{th} training iteration.
    $\mathbf{f_{MRI}} = \mathbf{E_1}(\mathbf{I_{MRI}})$
    $\mathbf{f_{PET}} = \mathbf{E_1}(\mathbf{I_{PET}})$
    $\mathbf{f} = \mathbf{A}(\mathbf{f_{MRI}}, \mathbf{f_{PET}}) = \mathbf{f_{MRI}} + \mathbf{f_{PET}}$
    while *iteration* $< M_{iter}$ **do**
        Train the classifier network (fig.3) with **f** and **Y**
    end while
end procedure

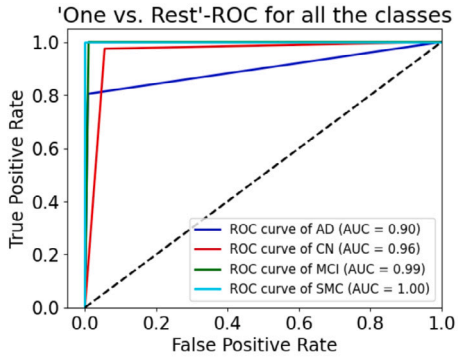
4. Experiment results

4.1. Dataset

For our experimentation we have considered the ADNI dataset [30]. The study includes both males and females with a follow-up during the last 18 months with an age range of 55 to 90 years. We only select bias-corrected MRI scans. The field strength utilized for MRI acquisition was 3.0 T, with 1 mm pixel spacing. The sagittal plane has been selected as the acquisition plane for both the MRI and PET scans. One hundred subjects have been chosen from each group—AD, MCI, subjective memory concern (SMC), and cognitive normal (CN) to conduct the experiments. Ten significant slices have been extracted from each MRI and PET after the preprocessing. Finally, a subset of 4000 MRI images and 4000 PET images are taken from the dataset. These 4000 images of each type comprise 1000 images of four classes: CN, MCI, SMC and AD. The images are of resolution: (212×256), and the pixel values were normalized to the range $[0, 1]$.

4.2. Implementation details

From each class of both MRI and PET images, we have taken 800 image pairs for training purpose and the rest 200 image pairs for testing purpose. We have trained the proposed CGAN for 70 iterations. For the training data set, the latent representations of Generator network, of the CGAN, are computed and then used for training the CNN classifier. The classifier has been trained for 50 iterations. However, the best



(a) ROC Curve

Multimodality				
AD	161	33	6	0
CN	5	195	0	0
MCI	0	0	200	0
SMC	0	0	0	200

(b) Confusion Matrix

Fig. 4. Performance of the proposed method.

Table 1
Performance parameters achieved by the proposed method.

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
AD	80.5%	81%	99.19%	97%	88%
CN	97.5%	97%	93.5%	86%	91%
MCI	100%	100%	92.67%	97%	99%
SMC	100%	100%	92.67%	100%	100%
Overall	**94.5%**	**95%**	**94.49%**	**95%**	**94%**

Table 2
Comparison of single and multi-modality.

Modality	Accuracy	Sensitivity	Specificity	Precision	F1-score
MRI	81.88%	82%	81.87%	82.5%	82%
PET	74.13%	74%	74.5%	74.5%	74%
Fused	**94.5%**	**95%**	**94.49%**	**95%**	**94%**

results for classification are achieved for the 59th iteration of the CGAN network and 17th iteration for the classifier network.

4.3. Performance results

The results achieved by the proposed method have been shown in Table 1. Table 1 lists the accuracy, sensitivity, specificity, precision, and F1 score for the proposed method for the classification between AD, CN, SMC and MCI. Accuracy signifies the total correct decision taken by the classifier. Sensitivity signifies the classifier's capacity to recognize the positive class accurately. Specificity accounts for the classifier's capability to correctly recognize the negative class. Precision indicates the accuracy of the positive prediction. As depicted in Table 1, the overall accuracy of the proposed model is 94.5%, which indicates a good performance. The receiver operating characteristics (ROC) curve and confusion matrix for the proposed fusion method are shown in Fig. 4.

4.4. Comparison of single and multi modality

We utilized the same CGAN and classifier architecture to compare single- and multi-modality AD diagnoses. A single encoder has been utilized to represent single-modality, MRI or PET, in latent space. A decoder is used to reproduce the image by decoding the latent vector. The most significant latent representations are generated through adversarial learning with the help of a discriminator network. Table 2, shows the performance results of both the single- and multi-modalities. When compared to MRI and PET imaging, the combined image's accuracy is substantially greater and more favorable. Fig. 5 shows the ROC comparison and Fig. 6 shows the confusion matrix comparison between single and multi-modality.

Table 3
Comparison of different CGAN architecture.

Architecture	Accuracy	Sensitivity	Specificity	Precision	F1-score
VGG16	64.75%	65%	64.75%	67%	66%
Resnet50	69.75%	70%	62%	73%	70%
Proposed	**94.5%**	**95%**	**94.49%**	**95%**	**94%**

4.5. Comparison with different CGAN architecture

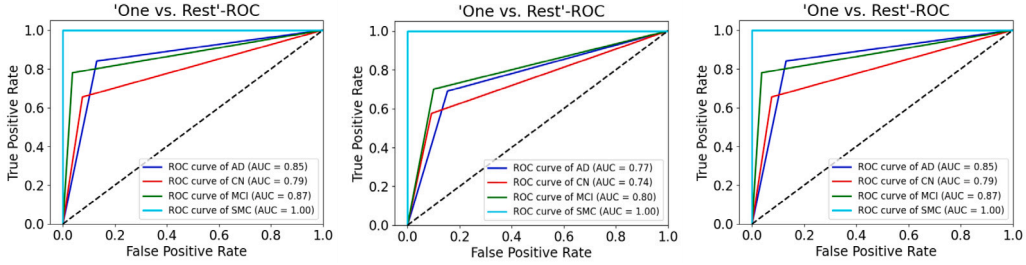
Also, the suggested CGAN architecture's performance is evaluated against various deep learning architectures. We have used same architecture as our proposed method, but encoder networks have been replaced with different standard architectures. Here, the findings for VGG16 and ResNet50 are presented in Table 3. For extraction of feature vectors from MRI and PET images, pre-trained VGG16 network is employed at the two encoder branches separately. The feature vector outputs from the VGG16 encoders, for both the PET and MRI images, are added together to form a shared latent representation. The corresponding input MRI and PET images are then reproduced from this common latent representation using the decoder networks of the proposed CGAN architecture. This network yielded an accuracy of 64.5%. A similar architecture was formed using pretrained Resnet-50 model. With Resnet-50 as encoders, 70% accuracy was attained in classification. The ROC curve and confusion matrix are shown in Figs. 7 and 8 respectively.

4.6. Computational complexity

The computational complexity for evaluating test case samples are presented in Table 4. Here, the time complexity for various fusion based methods, that we have experimented with, are presented. The computation cost has been computed as average over processing of the test dataset. This experimentation has been carried out on Nvidia RTX A6000 workstation with Intel Xeon-silver-4214 CPU and 64 GB RAM. The first column of Table 4 states the name of the approach/algorithm. The second column represents frames-per-second (fps) or, the number of test image samples the algorithm is able to process per second. The third column represents the time taken (in seconds) by the algorithm to process one test sample. Also, for ease of reference, the accuracy of the models are also listed in the final column. From this comparison it is understood that the proposed method performs better in terms of time complexity while achieving better accuracy.

4.7. Model parameter uncertainty

To determine the extent of uncertainty in the model parameters, we repeated the experimentation with the proposed CGAN and CNN



(a) MRI (b) PET (c) Mutli-modality

Fig. 5. ROC curve for Single and Multi Modality.

MRI

AD	168	25	7	0
CN	54	131	15	0
MCI	24	20	156	0
SMC	0	0	0	200

PET

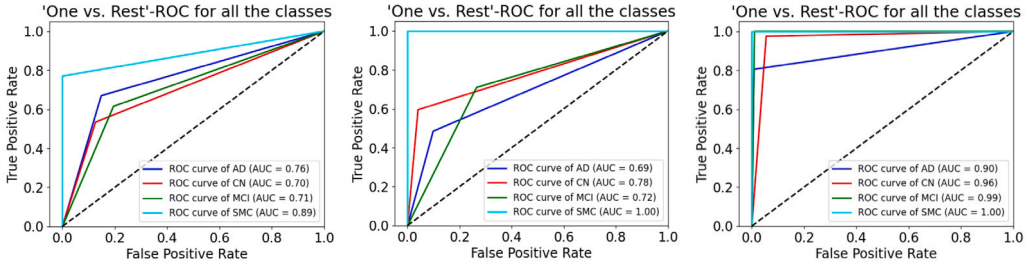
AD	138	35	27	0
CN	52	115	33	0
MCI	40	20	140	0
SMC	0	0	0	200

Multimodality

AD	161	33	6	0
CN	5	195	0	0
MCI	0	0	200	0
SMC	0	0	0	200

(a) MRI (b) PET (c) Mutli-modality

Fig. 6. Confusion matrix for Single and Multi Modality.



(a) VGG16 (b) Resnet50 (c) Proposed

Fig. 7. ROC curve for different architecture.

VGG16

AD	134	14	52	0
CN	32	107	61	0
MCI	43	34	123	0
SMC	14	28	4	154

Resnet50

AD	97	11	92	0
CN	14	119	67	0
MCI	45	13	142	0
SMC	0	0	0	200

Multimodality

AD	161	33	6	0
CN	5	195	0	0
MCI	0	0	200	0
SMC	0	0	0	200

(a) VGG16 (b) Resnet50 (c) Proposed

Fig. 8. Confusion matrix for different architectures.

Table 4
Comparison of time complexity of different fusion techniques.

Architecture	fps	Test time/ image (s)	Accuracy
VGG16	84.33	0.0199	64.75%
Resnet50	97.23	0.0103	69.75%
Proposed	**137.79**	**0.0073**	**94.5%**

classifier models multiple times. For each instance of experimentation, the models were trained on the same set of training data, from the scratch. Their performance was also tested on the same test data set. During the training process the weights of the proposed networks are randomly initiated according to zero-mean Normal distribution with standard deviation value 0.1. The biases are initiated as zeros. Though, the performance of the CGAN in terms of mse of the reconstructed output varied minimally, the classifier's performance over the latent space representation remained unchanged.

4.8. Discussion

AD includes structural, functional, metabolite, and chemical changes in the brain of the affected person. Diagnosis accuracy will increase if both the structural as well as metabolic alterations in the brain are considered. Therefore, in this study, we fused the MRI as well as PET images to get information on structural atrophies as well as metabolic changes.

The idea of this work is to fuse the MRI and PET images such that their feature set complement each other for classification of different stages of AD. In the proposed method, the adversarial learning of the GAN framework forces the output set to be same as the input set. However, a shared latent space, as in the proposed CGAN model, makes

sure that the complementary features of the MRI and PET images are included in the latent representation along with common features. Only then the decoders are able to reconstruct back the input MRI and PET images separately. The unsupervised learning paradigm of the coupled Generator model helps in fusing the complementary as well as common features of both the modalities. The supervised learning paradigm of two separate discriminators enhances the learning of the coupled Generator.

Table 1 shows the overall performance of the proposed architecture using data from ADNI dataset. Table 2 shows the comparison of the multimodal data with the single modality. Multimodal data produces much better performance than the single modality (MRI or PET). Table 3 shows the comparison with different CGAN architectures using VGG16 net and Resnet50, which evidences the efficiency of the proposed architecture.

Further, Table 4 demonstrates the computational complexity of the proposed network and compares it with different deep learning-based networks. In addition, we also tested the model's independence of parameter initialization by repeated-random-initialization of the weights and biases of the network, during training.

4.9. Limitation

In the proposed method the latent space is shaped according to the input, output and overall architecture of the CGAN model. However, for medical diagnosis problems with larger number of possible classes the proposed method may not be suitable. With increase in number of classes the ambiguity between the latent space representation of the classes will also increase. In such a case, it may be more suitable to add further constraints on the latent space, by including it in the cost function for training the GAN model.

5. Conclusion

In this work we have presented a novel semi-supervised learning-based method for comprehensive fusion of features from PET and MRI images. We demonstrated that the extracted fused feature set enables more accurate classification of AD stages than using features from only MRI or PET images. The proposed method uses adversarial learning to extract more of a complete feature set. This feature set includes the details that are common as well as exclusive in PET and MRI images. This is established from the fact that the feature set is used to reconstruct the MRI and PET images separately. The repeated experimentation with the architecture, with random initialization within a permissible limit, shows unchanged classification results. Which implies that the proposed architecture behaves in a robust manner with permissible deviations in initialization. Finally, analysis of computational complexity shows that the proposed architecture has better performance than standard deep learning architectures.

In the future the cost function of the CGAN can be extended to accommodate constraints on the latent space representation so that the proposed approach can be used for classification of data with larger number of classes. Also, this architecture can then be used for tasks like segmentation. Moreover, additional data modalities, such as functional MRI, diffusion tensor imaging, FLAIR, etc., may be explored for the AD diagnosis.

CRediT authorship contribution statement

Chandrajit Choudhury: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Tripti Goel:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **M. Tanveer:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization.

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

No data was used for the research described in the article.

Acknowledgments

This work is supported by Core Research Grant to the Science and Engineering Research Board (SERB) for funding under Grant No. CRG/2022/006866.

References

[1] M. Orouskhani, C. Zhu, S. Rostamian, F.S. Zadeh, M. Shafiei, Y. Orouskhani, Alzheimer's disease detection from structural MRI using conditional deep triplet network, Neurosci. Inform. 2 (4) (2022) 100066.
[2] M.D. Mulligan, R. Murphy, C. Reddin, C. Judge, J. Ferguson, A. Alvarez-Iglesias, E.R. McGrath, M.J. O'Donnell, Population attributable fraction of hypertension for dementia: global, regional, and national estimates for 186 countries, EClinicalMedicine 60 (2023).
[3] S. Asher, R. Priefer, Alzheimer's disease failed clinical trials, Life Sci. (2022) 120861.
[4] Y.L. Lo, S.-H. Cheng, Iron and Alzheimer's disease, in: Brain-Iron Cross Talk, Springer, 2022, pp. 139–170.
[5] G.B. Frisoni, N.C. Fox, C.R. Jack Jr., P. Scheltens, P.M. Thompson, The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 67–77.
[6] L. Fang, C. Yin, J. Zhu, C. Ge, M. Tanveer, A. Jolfaei, Z. Cao, Privacy protection for medical data sharing in smart healthcare, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16 (3) (2020) 1–18.
[7] M. Ganaie, A. Kumari, A. Girard, J. Kasa-Vubu, M. Tanveer, Alzheimer's Disease Neuroimaging Initiative, et al., Diagnosis of Alzheimer's disease via Intuitionistic fuzzy least squares twin SVM, Appl. Soft Comput. 149 (2023) 110899.
[8] A. Nordberg, J.O. Rinne, A. Kadir, B. Långström, The use of PET in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 78–87.
[9] R. Sharma, T. Goel, M. Tanveer, C. Lin, R. Murugan, Deep learning based diagnosis and prognosis of Alzheimer's disease: A comprehensive review, IEEE Trans. Cogn. Dev. Syst. (2023).
[10] J. Ma, H. Xu, J. Jiang, X. Mei, X.P. Zhang, DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion, IEEE Trans. Image Process. 29 (2020) 4980–4995.
[11] Z. Yue, S. Ding, L. Zhao, Y. Zhang, Z. Cao, M. Tanveer, A. Jolfaei, X. Zheng, Privacy-preserving time-series medical images analysis using a hybrid deep learning framework, ACM Transactions on Internet Technology (TOIT) 21 (3) (2021) 1–21.
[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[13] B. Li, J.-N. Hwang, Z. Liu, C. Li, Z. Wang, PET and MRI image fusion based on a dense convolutional network with dual attention, Comput. Biol. Med. 151 (2022) 106339.
[14] R. Nandhini Abirami, P. Durai Raj Vincent, K. Srinivasan, K.S. Manic, C.Y. Chang, et al., Multimodal medical image fusion of positron emission tomography and magnetic resonance imaging using generative adversarial networks, Behav. Neurol. 2022 (2022).
[15] L. Tran, X. Yin, X. Liu, Disentangled representation learning GAN for pose-invariant face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017.
[16] J. Kang, W. Lu, W. Zhang, Fusion of brain PET and MRI images using tissue-aware conditional generative adversarial network with joint loss, IEEE Access 8 (2020) 6368–6378.
[17] M. Haddadpour, S. Daneshvar, H. Seyedarabi, PET and MRI image fusion based on combination of 2-D Hilbert transform and IHS method, Biomed. J. 40 (4) (2017) 219–225.
[18] C.-I. Chen, Fusion of PET and MR brain images based on IHS and log-Gabor transforms, IEEE Sens. J. 17 (21) (2017) 6995–7010.
[19] H.R. Shahdoosti, A. Mehrabi, MRI and PET image fusion using structure tensor and dual ripplet-II transform, Multimedia Tools Appl. 77 (2018) 22649–22670.
[20] H. Ouerghi, O. Mourali, E. Zagrouba, Non-subsampled shearlet transform based MRI and PET brain image fusion using simplified pulse coupled neural network and weight local features in YIQ colour space, IET Image Process. 12 (10) (2018) 1873–1880.

[21] S. Aymaz, C. Köse, A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion, Inf. Fusion 45 (2019) 113–127.
[22] S. Dwivedi, T. Goel, M. Tanveer, R. Murugan, R. Sharma, Multimodal fusion-based deep learning network for effective diagnosis of Alzheimer's disease, IEEE MultiMedia 29 (2) (2022) 45–55.
[23] R. Sharma, T. Goel, M. Tanveer, P. Suganthan, I. Razzak, R. Murugan, Conv-ervfl: Convolutional neural network based ensemble RVFL classifier for Alzheimer's disease diagnosis, IEEE J. Biomed. Health Inf. (2022).
[24] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (2017) 191–207.
[25] J. Ma, W. Yu, P. Liang, C. Li, J. Jiang, FusionGAN: A generative adversarial network for infrared and visible image fusion, Inform. Fusion 48 (2019) 11–26.
[26] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, J. Wu, J. Jiang, Infrared and visible image fusion via detail preserving adversarial learning, Inf. Fusion 54 (2020) 85–98.
[27] R. Hang, F. Zhou, Q. Liu, P. Ghamisi, Classification of hyperspectral images via multitask generative adversarial networks, IEEE Trans. Geosci. Remote Sens. 59 (2) (2020) 1424–1436.
[28] L. Zhang, Q. Nie, H. Ji, Y. Wang, Y. Wei, D. An, Hyperspectral imaging combined with generative adversarial network (GAN)-based data augmentation to identify haploid maize kernels, J. Food Comp. Anal. 106 (2022) 104346.
[29] W.Y. Wang, H.C. Li, Y.J. Deng, L.Y. Shao, X.Q. Lu, Q. Du, Generative adversarial capsule network with ConvLSTM for hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 18 (3) (2020) 523–527.
[30] C.R. Jack Jr., M.A. Bernstein, N.C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P.J. Britson, J. L. Whitwell, C. Ward, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging 27 (4) (2008) 685–691.
[31] W.D. Penny, K.J. Friston, J.T. Ashburner, S.J. Kiebel, T.E. Nichols, Statistical Parametric Mapping: The Analysis of Functional Brain Images, Elsevier, 2011.