Biomedical Signal Processing and Control 84 (2023) 104787

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

IDA-Net: Inheritable Deformable Attention Network of structural MRI for Alzheimer's Disease Diagnosis

Qin Zhao [a], Guoheng Huang [a,*], Pingping Xu [a], Ziyang Chen [b], Wenyuan Li [a], Xiaochen Yuan [c,*], Guo Zhong [d,*], Chi-Man Pun [e], Zhixin Huang [f,*]

[a] School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China
[b] Guangzhou KingMed Center for Clinical Laboratory Co., Ltd, Guangzhou 510317, China
[c] Faculty of Applied Sciences, Macao Polytechnic University, 999078, Macao Special Administrative Region of China
[d] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou 510006, China
[e] Department of Computer and Information Science, University of Macau, 999078, Macao Special Administrative Region of China
[f] Department of Neurology, Guangdong Second Provincial General Hospital, Guangzhou 510317, China

ARTICLE INFO

Keywords:
Alzheimer's Disease Diagnosis
Structural magnetic resonance image
Computer-aided diagnosis
Deep learning
Transformer
Self-attention
Deformation
Inheritance

ABSTRACT

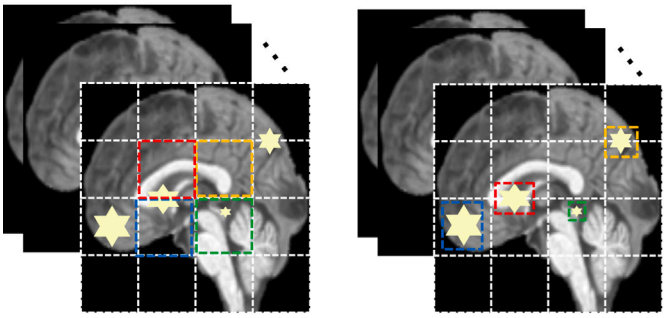
To precisely diagnose neurological diseases, such as Alzheimer's disease, clinicians need to observe the microstructural changes of local brain atrophy with the help of structural magnetic resonance image (sMRI). Some Convolutional Neural Networks (CNNs) have recently achieved excellent performance in auxiliary clinicians to provide the diagnosis suggestion. However, there still exist several challenges. Foremost, several researchers manually predefine some regions of interest (ROIs) as the input of the CNN-based networks, which impedes the model's robustness and interpretability of clinical applications. Second, since the position relevance of pathological features interferes with the surrounding tissue regions in ROIs, it is hard for the current CNN-based networks to extract the microstructural changes of these ROIs precisely. To address the above challenges, we optimize the Transformer structure for Alzheimer's Disease Diagnosis and propose an Inheritable Deformable Attention Network (IDA-Net). Specifically, the IDA-Net mainly comprises the 3D Deformable Self-Attention module and the Inheritable 3D Deformable Self-Attention module. The 3D Deformable Self-Attention module can automatically adjust the position and scale of the selected patches according to the structural changes in sMRI. Furthermore, the Inheritable 3D Deformable Self-Attention module can locate and output relatively important regions with discriminative features in sMRI, which can assist physicians in the clinical diagnosis. Our proposed IDA-Net method is evaluated on the sMRI of 2813 subjects from ADNI and AIBL datasets. The results show that our IDA-Net method behaves better than several state-of-the-art methods in classification performance and model generalization.

1. Introduction

Alzheimer's Disease (AD), one of the most severe neurological diseases among the elderly, is characterized by progressive memory-related impairments such as memory deterioration and cognitive deficits. The main reason AD dementia is an incurable and irreversible disease is that no pharmacologic treatments have proven effective enough in reversing AD progression. Therefore, early detection at AD dementia prodromal stages, i.e., Mild Cognitive Impairment (MCI), is of great clinical importance for applying some interventions and treatments to mitigate the progression of converting to AD dementia. Due to the related structural brain changes of AD emerge even earlier and are more obvious than the amnestic seeable symptoms. Then sMRI, which is focused on identifying subtle brain anatomical changes, has shown excellent clinical values in the progression of AD. Structural changes in the brain can be observed in sMRI scans of individuals with AD. As the disease progresses, there are characteristic changes in the brain structure that can be used as biomarkers to identify the disease and monitor its progression. One of the key structural changes observed in sMRI scans of individuals with AD is brain atrophy, which refers to a decrease in the size of the brain. This is particularly prominent in regions of the brain associated with memory and cognitive function, such as the hippocampus [1]. Another structural change that can be observed in sMRI scans of individuals with AD is a decrease in the cortical thickness of certain areas of the brain, such as the entorhinal cortex and the precuneus [2]. This decrease in cortical thickness is thought to reflect neuronal loss and is considered a hallmark of AD. In addition to

* Corresponding authors.
E-mail addresses: kevinwong@gdut.edu.cn (G. Huang), xcyuan@mpu.edu.mo (X. Yuan), yb77410@um.edu.mo (G. Zhong), hzxd6@163.com (Z. Huang).

https://doi.org/10.1016/j.bspc.2023.104787
Received 10 September 2022; Received in revised form 13 February 2023; Accepted 1 March 2023
Available online 11 March 2023



(a) Original patch-based methods (b) IDA-Net (ours)

Fig. 1. We give a comparison of IDA-Net with original patch-based methods. The stars represent the discriminative atrophic brain locations. (a) Original patch-based methods split the whole sMRI scan into patches of fixed location and scale, which may destroy the semantics of the discriminative pathological regions. (b) Our proposed IDA-Net locates and output discriminates atrophic brain locations by adjusting the selected patches' position and scale.

these structural changes, changes in the distribution of certain brain metabolites, such as N-acetylaspartate (NAA) and myoinositol (MI), can also be observed in sMRI scans of individuals with AD [3]. These changes in brain metabolites are thought to reflect changes in neuronal and glial cell metabolism and can be used as biomarkers to identify the disease and monitor its progression. Consequently, many studies have been made to investigate differences in typical structural brain changes between AD patients and normal controls (NC) and to predict the progression of MCI based on sMRI. The conventional AD diagnosis methods based on sMRI have made significant progress in the past few decades [4–6].

Most current sMRI-based AD diagnosis methods are based on handcrafted approaches designed by experts in related fields using domain knowledge [7,8]. According to the scales of handcrafted ROIs from sMRI for integrating feature extraction, sMRI-based AD diagnosis studies can be mainly subdivided into voxel-based, region-based, and patch-based methods. Selecting the right ROIs in sMRI scans is crucial in accurately diagnosing AD. sMRI scans provide detailed images of the brain and its structure, and the selection of ROIs allows for the focused analysis of specific regions that are associated with AD. Major ROIs selected in sMRI used for AD diagnosis include the hippocampus, entorhinal cortex, and amygdala [2]. These regions are responsible for memory and spatial navigation, and are often among the first to show signs of degeneration in individuals with AD. In regard to the voxel-based methods, the voxel features related to AD classification are extracted from sMRI and then combined into high-dimensional data [9–11]. For example, a linear support vector machine(SVM) was constructed to discriminate individuals with an MCI from controls by extracting the volume and geometry of each point on the surface of the cerebral cortex [12]. The measurement of gray matter (GM) density can provide important information about the structural organization and function of the brain, is a common approach used to assist AD diagnosis. The GM density map was used to perform AD classification by calibrating the spatial regularizer to linear programming boosting (LPboost-ing), such as Demiriz et al. [13] and Hinrichset al. [14]. Cho et al. [15] propose an incremental learning-based individual classification method for AD diagnosis and prediction using cortical thickness data. However, the voxel-level method only considers voxel features, ignoring their correlation. And since the voxel features are often very large, there are fewer images for training AD classification, which is easy to cause the overfitting phenomenon, resulting in a decrease in prediction accuracy. Thus, the region-based method is proposed to solve these problems

In region-based methods, features are extracted from multiple regions through pre-segmented ROIs, after which classifiers are constructed. For instance, Zhang et al. [16] and Magnin et al. [17] divided the whole brain into several non-overlapping regions, aligned the sMRI of each individual to an anatomically labeled map, and then extracted the brain region features to train the SVM classifier. After that, multicore-based approaches combining marginal Fisher analysis were proposed to achieve sparse dimensionality reduction of ROIs and capture the complex relationship between sMRI features and disease states [18]. Koikkalainen et al. [19] spatially normalized each individual sMRI space to multiple maps. They then extracted discriminative features of each map space to build an integrated classification model for AD diagnosis. Most other region-based methods, which extract features only from the ROIs, may not include other disease features. Moreover, the output of the classifier was combined to identify the normal control group (NC). These regions rely only on prior knowledge to instruct the selection of ROIs and features. However, since the pre-segmentation of ROIs needs to rely on the experience of experts, the definition, and segment of ROIs are resource-intensive. Furthermore, they only evaluate regions of interest, which can lead to ignoring other important pathological areas of AD. For example, regions like the angular gyrus and inferior parietal lobe are rarely assessed in ROIs, but these are also regions severely affected in AD [20]. Therefore, the patch-based methods between the voxel-based and region-based methods are proposed. It can more effectively solve the problem of ignoring structural changes.

In patch-based methods, the selection of the patches, i.e., Patch Location Proposals, can be determined by particular anatomical landmark detectors [21] or statistics methods [22]. Tong et al. [23] extracted local strength blocks as features and proposed to use a multiple instance learning (MIL) method for the detection of AD and its prodromal stage MCI. A dual attention multi-instance deep learning model (DA-MIDL) was proposed to confirm discriminative pathological locations for AD diagnosis [1]. A hierarchical full convolutional network (H-FCN) is proposed to automatically identify local patches and regions in whole brain sMRI and then co-learn multi-scale feature and fuse them into a hierarchical classification model for AD diagnosis [24]. However, determining the location and size of the patch is still a problem. As shown in Fig. 1(a), the size and position of the input patch are fixed. Moreover, changes in brain structure caused by brain atrophy may occur in areas of different scales and locations. Using fixed-size and fixed-position patches may disrupt the semantics of the discriminative pathological regions to represent various local features. Additionally, existing patch-based methods are highly flawed because the Patch Location Proposal is disconnected from subsequent network models, which means that they are not strictly end-to-end methods. At the same time, the Patch Location Proposal is a fixed parameter that cannot be learned and cannot be jointly optimized in real-time using the network model, which will limit the performance of the model. Then, the advent of Transformer brings the potential solution of a unified end-to-end approach to patch-based methods research.

In recent years, Transformer has become the leading model in natural language processing (NLP) [25]. Given the great success achieved in NLP tasks with the help of Transformers, researchers have begun exploring Transformers in computer vision. They found that the ability to model long-range dependencies and extract global information through the multi-head Self-Attention (MSA) in Transformers is also applicable for pixel-based image processing. Vision Transformer (ViT) was the first pioneer to apply a pure Transformer rightly to image patches for image recognition and achieve excellent performance in this task when pre-training on large datasets like imageNet-22K in advance [26]. Since then, a lot of Transformer variants have been derived and successfully applied in many vision tasks such as image classification [27–29], object detection [30–32] and image generation [33], etc. Inspired by these achievements, some attempts begin to be made on medical image analysis by combining the advantage of Transformer. For example, the Medical Transformer proposed a gated position-sensitive axial attention mechanism, a Local–Global (LoGo) training strategy to realize Transformer working flexibly well on relatively small medical image datasets [34]. The Swin-Unet [35] proposed

an Unet-like pure Transformer, which is a Transformer-based U-shaped Encoder–Decoder architecture motivated by U-net [36] and Swin Transformer [37] and outperforms many full-convolution or the combination of Transformer and convolution works in medical image segmentation task. Following Swin-Unet [35], DS-TransUNet is a further work that also combines the Swin Transformer [37] with the U-shaped architecture while proposing a dual-scale encoding mechanism to ensure the semantic consistency between the coarse and fine features and a TIF module to fuse multiscale contexts and yield high-quality semantic segmentation performance [38].

Reviewing the above literature, we can find that the secret to the success of Transformer is its structure, which also meets the requirements of the patch-based methods method in Alzheimer's Disease diagnosis well. Specifically, the patch-based methods method will firstly regard the entire sMRI as a series of patches and extract discriminative features from each patch, which can correspond to the patch embedding operation in Transformer. Besides, the Self-Attention module in Transformer can model remote dependencies in sequence patches. The multi-head mechanism in Transformer can enrich the diversity of modeling the patch relationship to extract more comprehensive information. The stacked attention module in Transformer can integrate the local feature representation of patches into the global feature representation of the whole sMRI. The Transformer method integrates the Patch Location Proposals, patch feature representation extraction, and the construction of the classifier in the patch-based methods method into an end-to-end strategy, which means that all the modules can be jointly optimized in real-time during the training.

Nevertheless, the Transformer still could not be directly applied to AD diagnosis for some reasons. Firstly, the critical challenge in visual recognition, especially in the computer-aided medical diagnosis domain, is adapting to learn geometric variations of ROIs' scale, shape, and part deformation. Specifically, the patch size and location in the vanilla Transformer are also fixed, which to some extent does not address the limitation of the patch-based methods approach. Moreover, in sMRI scans, only a few areas have noticeable structural changes which are highly related to the pathological characteristics, while the distinction information in other areas is very small. The original Transformer cannot locate relatively essential regions in global sMRI scans well, which cannot have specific outputs on the pathological part, ignoring the explanatory problem in medical practice.

In order to solve these problems, we do the following work. To begin with, to construct a unified end-to-end approach for sMRI-based AD diagnosis, we propose an Inheritable Deformable Attention Network (IDA-Net). With the Transformer as the backbone, the IDA-Net presents two new self-attention modules based on the above two problems. On the one hand, to solve the first problem mentioned above, the 3D Deformable Self-Attention module is proposed. As illustrated in Fig. 1(b), through the 3D Deformable Self-Attention module, IDA-Net can adaptively adjust the scale and position of the selected patch according to the input of sMRI scans to reduce the damage to atrophic brain structures caused by split patches. Structural MRI scans can reveal changes in brain structure that are characteristics of AD, including:brain atrophy which mainly refers a reduction of hippocampus and the temporal lobe in the size of brain regions; ventricular enlargement which indicates a loss of brain tissue, a significant sign of AD; white matter hyperintensities which is also related to AD patients [39]. In conclusion, sMRI scans can provide the mentioned valuable information on disease features that can be used to diagnose AD. So it is necessary to use sMRI scan for AD diagnosis.

On the other hand, to address the second problem, we introduce the Inheritance Patch module based on the 3D Deformable Self-Attention module and name it the Inheritable 3D Deformable Self-Attention module, which progressively localizes relatively important pathological regions in the global sMRI through Inheritable Patches operations across multiple stages. These possible pathological regions can be used as the discriminative basis for doctors to make a diagnosis in the clinical diagnosis stage. The key pathological change, brain atrophy would be observed in sMRI using the 3D Deformable Self-Attention module while comparing with subjective evaluation. By combining sMRI results with subjective evaluations, healthcare providers can get a more comprehensive understanding of the patient's condition and make more informed decisions about treatment. Our proposed IDA-Net method is evaluated on ADNI and AIBL datasets and multiple AD diagnosis tasks (e.g., AD classification, prediction of MCI conversion, and MCI discriminations). Experimental results show that our IDA-Net method behaves better than several state-of-the-art methods in classification performance and model generalization. The major contributions of this paper can be summarized as follows.

1. To construct a unified end-to-end approach to improve sMRI-based Alzheimer's Disease Diagnosis performance, an Inheritable Deformable Attention Network (IDA-Net) is proposed, which can automatically adjust the scale and position of each selected patch from sMRI scans and locate pathological areas with discriminative features to assist doctors in Alzheimer's Disease Diagnosis.
2. The 3D Deformable Self-Attention module is designed to automatically regulate the scale and position of each selected patch according to the different microstructure of brain atrophy to avoid the destruction of semantic features of brain atrophy regions.
3. The Inheritable 3D Deformable Self-Attention module is proposed to locate relatively important regions with discriminative features in the global sMRI scans and output the specific location of the pathological region, which explains the interpretability problem of Transformer structure application in clinical medical practice.
4. Lastly, the proposed method is evaluated on ADNI and AIBL datasets as well as on multiple AD-related diagnosis tasks, demonstrating its performance and generalization over several state-of-the-art methods, especially for predicting MCI conversion.

2. Method

2.1. Overall architecture of inheritable deformable attention network

As illustrated in Fig. 2, the overall architecture of the proposed Inheritable Deformable Attention Network consists of four stages with different attention blocks of Self-Attention modules, an Adaptive Average Pooling, and a Linear Classifier with Softmax. Specifically, an input sMRI scan with shape $112 \times 112 \times 112 \times 1$ is firstly embedded through non-overlapping 3D-convolution with kernel 2 and stride 2 to obtain $56 \times 56 \times 56 \times C$ patch embeddings, where C is the number of the channel dimension. Between two successive stages, a non-overlapping 3D-convolution with kernel 2 and stride 2 downsamples the patch embeddings to reduce the space size and increase the feature dimension. After each downsampled 3D convolution layer, a Layer Normalization layer and a GELU activation function layer are always present. Stage 1 is composed of only N1 stacked 3D Shift Window Attention module [40], while the other stages are composed of N* stacked with 3D Shift Window Attention module, 3D Deformable Self-Attention module, and Inheritable 3D Deformable Self-Attention module. The output of the last stage is first compressed from $7 \times 7 \times 7 \times 8C$ to $1 \times 8C$ by Adaptive Average Pooling and then fed into Linear Classifier layers with Softmax layers to complete the ADNI classification.

We add 3D Deformable Self-Attention and Inheritable 3D Deformable Self-Attention blocks in all stages of IDA-Net except the first stage, which only consists of 3D Shift Window Attention blocks. 3D Shift Window Attention allows for the extraction of features by considering multiple patches, which can be useful in capturing fine-grained details in sMRI. Since the first stage has not fully completed

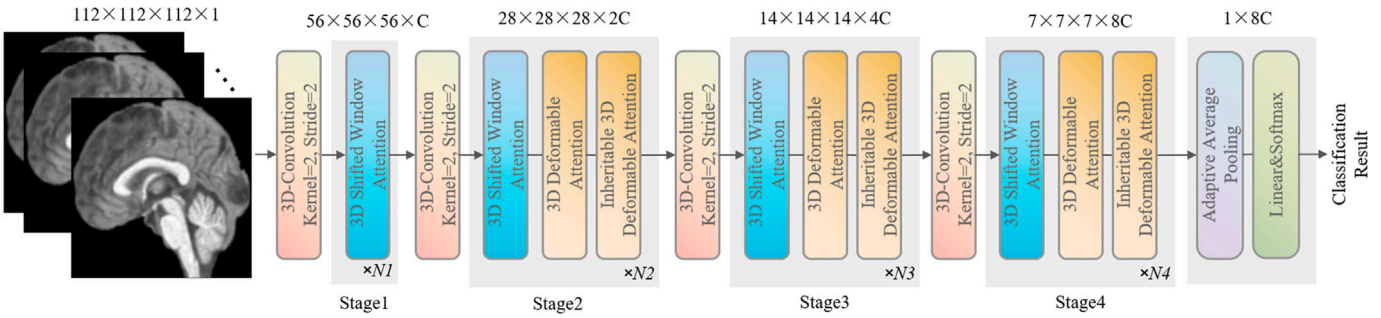


Fig. 2. The overall architecture of our Inheritable Deformable Attention Network (IDA-Net). C and N_i are the numbers of the channel dimension and stacked attention blocks in each stage. *Kernel* and *Stride* denote the kernel size and stride of the 3D-Convolution.

the feature extraction of local patch feature representation and the relational modeling of global feature representation, our deformable attention cannot obtain rich feature representation and thus ignore possible discriminative information. Further, the space size of the latter stages is smaller than that of the first stage, which can significantly reduce the computational cost of sampling and interpolation in our deformable attention modules. Thus, the first stage is only concatenated by multiple 3D Shift Window Attention modules for better feature extraction. In the remaining stages, we concatenate 3D Shift Window Attention, 3D Deformable Self-Attention, and Inheritable 3D Deformable Self-Attention modules into a multi-attention block in sequence.

2.2. 3D deformable self-attention module

In order to locate and extract more important features and utilize geometric variations adaptively, Deformable ConvNets [41,42], which enables free form deformation of the sampling grid in the traditional convolution, has been firstly proposed. Recently, many methods have emerged to fuse Deformable-Related Modules into Transformer-based Methods [43–45], notably DPT [46] and DAT [47]. Although the DePatch module proposed by DPT can automatically micro-adjust the scale and location of each selected patch according to the sMRI scans or input feature map, it does not integrate deformable operations into the Transformer backbone. In contrast, DAT manages to integrate deformable operations into the Transformer backbone by introducing a set of deformable keys and values. However, its deformable attention lacks coherence as deformable keys and values between different attention layers are learned separately without combining the previous deformable parameters. To address the issues above, by referring to DPT and DAT methods and proposing the Inheritance Patch module, the 3D Deformable Self-Attention module and Inheritable 3D Deformable Self-Attention module are proposed, respectively.

2.2.1. 3D deformable self-attention

To facilitate understanding of our method, we first review the Self-Attention mechanism in Vision Transformer. Specifically, let $X_{MRI} \in R^{H\times W\times D\times C}$ denote the input sMRI scans or feature map. X_{MRI} is first divided into a series of N patches with the same size and position by a patch embedding module with a linear layer. For convenience, we first assume $H = W = D$, and the embedding patch size is $s\times s\times s (s = \frac{H}{\sqrt[3]{N}})$. The entire sequence of patches can be viewed as a uniform grid of size $s\times s\times s$ in X_{MRI}. The series of patches is denoted as $\{X_i\}_{(0\le i\le N-1)}$. For each patch X_i, We denote its left-up-front coordinate and right-down-back coordinate as $\left(x_i^{left}, y_i^{up}, z_i^{front}\right)$ and $\left(x_i^{right}, y_i^{down}, z_i^{back}\right)$. These patches $\{X_i\}$ are then flattened and embedded into a flattened feature map $X \in R^{N\times d}$ by a linear layer with d output channels. A multi-head Self-Attention module with M heads takes $X \in R^{N\times d}$ as input and generates three learnable groups of representative features query $\left(q_m \in R^{N\times d_m}\right)$, key $\left(k_m \in R^{N\times d_m}\right)$, value $\left(v_m \in R^{N\times d}\right)$ by three linear embedding matrices $W_q \in R^{d\times d}$, $W_k \in R^{d\times d}$, $W_v \in R^{d\times d}$, where, $d_m = d/M$, respectively. The query features are used to determine the relevance of different regions of the sMRI. The key features are used to compare the query features to each region of sMRI and determine their similarity. The value features are used to aggregate information from different regions of sMRI and produce a weighted sum. The specific formula is as follows:

$$x_i^{right} = x_i^{left} + s \quad y_i^{down} = y_i^{up} + s \quad z_i^{back} = z_i^{front} + s \tag{1}$$

$$q = W_q X + P,\ k = W_k X + P,\ v = W_v X + P \tag{2}$$

$$Atten_m = Softmax\left(q_m k_m^T/\sqrt{d}\right)(m = 1, \ldots, M) \tag{3}$$

$$Z = Concat\left\{Atten_1 v_1, \ldots, Atten_M v_M\right\} W_o \tag{4}$$

where P is the Positional Encoding. $Atten_m$, which is used as a weighted weight for v_m, obtained by multiplying q_m and k_m, indicating the Similarity matrix between different patches. The new representative feature Z can be obtained by concatenating and transforming M heads $Atten_m v_m$.

While in the 3D Deformable Self-Attention module, the position offset and scale parameters will be generated for each selected patch in the input feature map. The position offset, and the scale parameter are represented by predictable parameters $\Delta position = (\Delta x, \Delta y, \Delta z)$, $\Delta scale = (\Delta h, \Delta w, \Delta d)$. For each deformed patch $\tilde{X}_i$, the updated left-up-front coordinate and right-down-back coordinate can be denoted as $\left(\tilde{x}_i^{left}, \tilde{y}_i^{up}, \tilde{z}_i^{front}\right)$ and $\left(\tilde{x}_i^{right}, \tilde{y}_i^{down}, \tilde{z}_i^{back}\right)$.

$$\tilde{x}_i^{left} = x_i^{left} + \Delta x,\ \tilde{x}_i^{right} = x_i^{left} + s + \Delta x + \Delta h \tag{5}$$

$$\tilde{y}_i^{up} = y_i^{up} + \Delta y,\ \tilde{y}_i^{down} = y_i^{up} + s + \Delta y + \Delta w \tag{6}$$

$$\tilde{z}_i^{front} = z_i^{front} + \Delta z,\ \tilde{z}_i^{back} = z_i^{front} + s + \Delta z + \Delta d \tag{7}$$

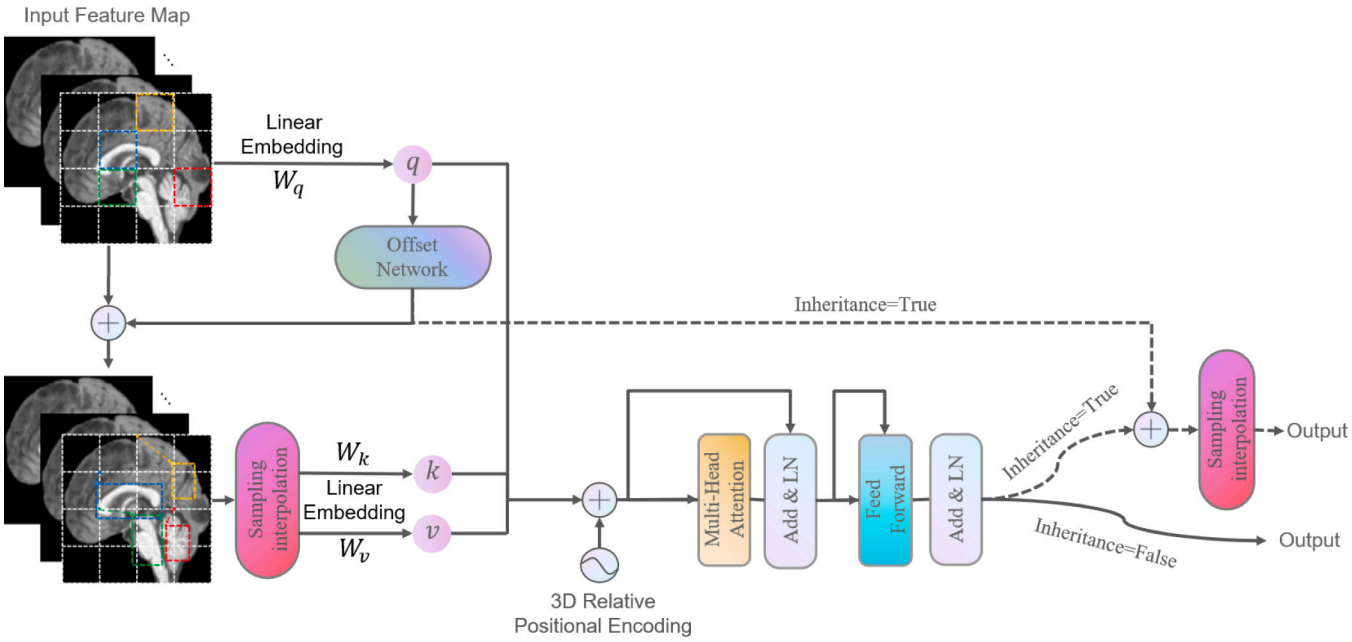
As illustrated in Fig. 3(a), to generate the position offset $\Delta position$ and the scale parameter $\Delta scale$, we add an Offset network ϕ_{Offset}. Firstly, the input feature map is embedded into features query $q = W_q X$, and then send into ϕ_{Offset} to predict offset parameters. After the deformed patches are determined, they will be projected linearly to deformed key $\tilde{k}$ and deformed value $\tilde{v}$ by sampling and interpolation $\Theta(\ldots)$:

$$q = W_q X + P,\ \tilde{k} = W_k \tilde{X} + \tilde{P},\ \tilde{v} = W_v \tilde{X} + \tilde{P} \tag{8}$$

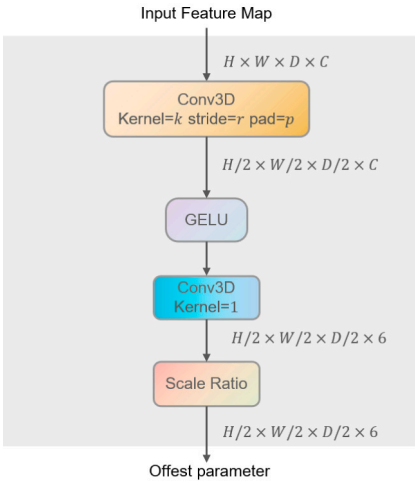
$$where\ \tilde{X} = \Theta(X, \Delta offset, \Delta scale) \tag{9}$$

where P and $\tilde{P}$ are the 3D Relative Positional Encoding and deformed 3D Relative Positional Encoding [47]. Then, we use a standard multi-head Self-Attention module with contrastive positional encoding, which can be rewritten as follows:

$$Atten_m = Softmax\left(q_m \tilde{k}_m^T/\sqrt{d}\right)(m = 1, ., M) \tag{10}$$



(a) Inheritable 3D Deformable Self-Attention module



(b) Offset Network

Fig. 3. An illustration of the proposed Inheritable 3D Deformable Self-Attention module. (a) present the Inheritable 3D Deformable Self-Attention module. Dotted lines are used to distinguish the Inheritable 3D Deformable Self-Attention module from the 3D Deformable Self-Attention module. After linear embedding, the offset parameter of each patch is learned from query q through the Offset Network. After that, the deformed patches are obtained by sampling and interpolating on the input feature map according to the offset parameter. After a standard self-attention module, if the predefined Inheritance parameter in the model is False, the result will be output directly, and the module at this time is the 3D Deformable Self-Attention module. While the Inheritance parameter is true, the module becomes an Inheritable 3D Deformable Self-Attention module and the output features will regenerate deformed patches. We only divide the input feature map into 16 patches and show 4 reference patches for clear display, and there will be more patches in the actual implementation. (b) show the structure of the Offset Network and annotate the size of the feature map. *kernel*, *stride* and *pad* denote the kernel size, stride, and padding of the 3D-Convolution.

$$Z = Concat\left\{Atten_1\tilde{v}_1, \ldots, Atten_M\tilde{v}_M\right\}W_o \tag{11}$$

The Deformable Transformer block adopts a standard Transformer block structure, which can be expressed as:

$$\dot{Z}_l = F_{MHSA}\left(F_{LN}\left(Z_{l-1}\right) + Z_{l-1}\right) \tag{12}$$

$$Z_l = F_{Feed-Forward}\left(F_{LN}\left(\dot{Z}_l\right)\right) + \dot{Z}_l \tag{13}$$

where Z_{l-1} and F_{LN} denote the output of $l-1$ th Deformable Transformer block and Layer Normalization. F_{MHSA} and $F_{Feed-Forward}$ denote the multi-head Self-Attention and the Feed-Forward network.

2.2.2. Offset network

Based on the output query of linear embedding, the Offset network will generate 3D offsets ($\Delta position, \Delta scale$) for each selected patch. Assuming the size of a feature map is H * W * D * C, using a patch of size 2, there are N (N = H/2 * W/2 * D/2) patches on a feature map. The offset network needs to predict N * 6 offset parameters in total. As illustrated in Fig. 3(b), the input feature map in which patch size is $2 \times 2 \times 2$, will be downsampled by a 3D convolution layer. Assuming the patch size of the input feature map is $2 \times 2 \times 2$, the kernel of the first 3D convolution layer is 6, its stride is 2 and its pad is 2. After being processed through the GELU activation function, it is then processed through another 3D convolution with a kernel size of 1, stride of 1,

padding of 0, and channel number of 6, resulting in a feature map of size H/2 * W/2 * D/2 * 6, which corresponds to the N * 6 offset parameters to be predicted. To avoid generating too large offsets, we use Scale Ratio (s_{po}, s_{sc}) to adjust the magnitude of $(\Delta position, \Delta scale)$:

$$(\Delta position, \Delta scale) = (s_{po}, s_{sc}) \tanh\left(\phi_{Offset}(q)\right) \tag{14}$$

At the beginning of the training phase, we randomly initialize the weight parameters of these layers. However, the bias value of a convolutional layer with a kernel size of 1 will be set to False to reduce the generated offset. The details of parameter settings for Scale Ratio will be explained in the following section.

2.2.3. Sampling and interpolation

Due to the problem that each deformed patch has a different size, we set Sampling and Interpolation method to extract the feature. Given the deformed coordinates $\left(\tilde{x}_i^{left}, \tilde{y}_i^{up}, \tilde{z}_i^{front}\right)$ and $\left(\tilde{x}_i^{right}, \tilde{y}_i^{down}, \tilde{z}_i^{back}\right)$, a uniform grid of $k \times k \times k$ are sampled within the deformed patch $\{\tilde{p}_j\}_{(0 \le j \le k^3-1)}$. The coordinates of the $k \times k \times k$ grid are often fractional. Therefore, we obtain the weighted average of the adjacent 8 points C^j by trilinear interpolation to represent $\tilde{p}_j$ as:

$$\tilde{p}^j = F_{Trilinear}\left(C^j, (x^j, y^j, z^j)\right)\left(j = 0, \ldots, k^3 - 1\right) \tag{15}$$

$$C^j = \left(C^j_{000}, C^j_{001}, \ldots, C^j_{110}, C^j_{111}\right) \tag{16}$$

2.3. Inheritable 3D deformable self-attention module

To promote the coherence of the Deformable Self-Attention module and solve the interpretability problem, we introduced an Inheritance patch module branch to implement information transfer between different Deformable Self-Attention layers. By choosing whether to enter this branch, our Deformable Self-Attention module can be subdivided into 3D Deformable Self-Attention and Inheritable 3D Deformable Self-Attention modules. We found that by concatenating multiple Inheritable 3D Deformable Self-Attention modules in the overall structure to inherit the past deformed patches continuously, the network can gradually locate relatively important regions in the global sMRI scan and finally output the location of the pathological region. The entire process of localization and output will be visualized in the next section.

As illustrated in Fig. 3(a), the dotted line represents the Inheritance patch module branch. When Inheritable is False, the feature map is output directly after the Deformable Self-Attention layer, without passing through the Inheritance patch module branch. When Inheritable is true, the output feature map of the Deformable Self-Attention layer will regenerate deformed patches based on the previous offset parameters. Compared with 3D Deformable Self-Attention, the parameter setting of Offset Network in Inheritable 3D Deformable Self-Attention modules is also different. Considering the Inheritable Patch Module may cause some regions to be unsampled in the feature map due to too large offsets, Scale Ratio (s_{po}, s_{sc}) in Inheritable 3D Deformable Self-Attention is set to be a smaller weight than 3D Deformable Self-Attention. Not only that, but for better learning Inheritable offset parameters by stacking multiple Self-Attention modules with 3D Shift Window Self-Attention, 3D Deformable Self-Attention, and Inheritable 3D Deformable Self-Attention.

3. Experiments and results

3.1. Materials and pre-processing

Two Alzheimer's Disease datasets, including Alzheimer's Disease Neuroimaging Initiative datasets (ADNI)[1] and Australian Imaging, Biomarker and Lifestyle Flagship Study of Aging datasets (AIBL)[2], are used to validate the classification performance and generalizability of the proposed Inheritable 3D Deformable Attention Network for Alzheimer's Disease Diagnosis. Shared through the LONI Image and Data Archive (IDA), ADNI is used to study the rate of progression of MCI and AD to improve diagnostics for early detection of Alzheimer's disease. The AIBL study, launched in 2006, will investigate biomarkers and psychometric tools to observe disease progression. It should be noted that ADNI contains 3 overlapping subsets, namely ADNI-1, ADNI-2, and ADNI-3. To avoid data leakage and ensure independent evaluation, if one subject's name is presented more than once in all subsets, only one subject examined at the earliest time will be kept. Since the primary goal of this paper is to predict MCI progression, only *baseline/screening* subjects (sMRI of the first inspection) will be considered. Therefore, only select subjects that meet the following requirements:

- AD (Alzheimer's Disease): subjects were diagnosed with AD at the first inspection and maintained AD at subsequent times.
- NC (Normal Controls): subjects were diagnosed with NC at the first inspection and maintained NC at subsequent times.
- pMCI (progressive MCI): subjects were diagnosed with MCI at the first inspection, but deteriorated to AD and maintained AD within 36 months.
- sMCI (stable MCI): subjects were diagnosed with MCI at the first inspection, but maintained MCI or eased to NC within 36 months.

After this processing, ADNI consists of 2248 subjects with 1.5T/3T T1-weighted sMRI scans, among which 419 AD, 832 NC, 297 pMCI and 700 sMCI. AIBL contains structural MRI (sMRI) from 565 subjects, including 79 AD, 450 NC, 12 pMCI and 24 sMCI. The demographic and clinical detail of ADNI and AIBL datasets can be found in Table 1.

In sMRI, there are often several types of noise present, including magnetic field inhomogeneity, acquisition noise, motion artifacts, etc. Therefore, all original sMRIs are then pre-processed following a standard pipeline, including: SANLM denoising after intensity normalization, Internal resampling (1.00*1.00*1.00 mm), Affine registration, non-uniform intensity normalization (N3), Gaussian smoothing, Global intensity bias correction [48,49], Skull-Stripping [22]. All preprocessing processes are implemented in MATLAB R2019b with Statistical Parametric Mapping Toolbox (SPM12) [50] and Computational Anatomy Toolbox (CAT12). Finally, we remove the uninformative background and resample all sMRI to the same resolution (resolution from [121,145,121] to [112,112,112]).

3.2. Experimental settings

Our proposed IDA-Net network is implemented with Pytorch 1.10.0. All experiments are performed on the Linux platform with a single NVIDIA RTX 3090 GPU by Python 3.9.0. The ADNI dataset is randomly split into the train, validate and test datasets with a percent of 60%, 20%, 20%. We also test our proposed IDA-Net model individually on the AIBL dataset to further evaluate the model generalization and robustness of our IDA-Net method. Overall, our IDA-Net method is validated on four Alzheimer's Disease Diagnosis tasks, including the AD classification task (AD versus NC), the prediction of MCI conversion task (pMCI versus sMCI), and the MCI discrimination task (sMCI versus NC, pMCI versus NC).

During the training phase, the Adam optimizer with a weight decay of $1e^{-5}$ for 100 epochs is employed. The Cosine Annealing Warm Restart scheduler with batch size 4. At the outset of training, the initial learning rate of the model is set to 1e−3. Subsequently, based on the Cosine Annealing Warm Restarts scheduler, the minimum learning rate is reduced to 1e−5. As for the IDA-Net parameter, the channel dimension C_i and the number of different blocks N_i are set to [96,192,384,768] and [2,1,6,1]. The weights of the Offset Network are initialized to zero.

[1] adni.loni.usc.edu

[2] aibl.csiro.au

Table 1
Demographic and clinical detail of ADNI and AIBL datasets, including category, clinical dementia rating (CDR), gender, mini-mental state examination (MMSE), age, education.

Dataset	Category	Gender (male/female)	Age (mean±std)	MMSE (mean±std)	CDR (mean±std)	Education (mean±std)
ADNI	AD	419(234/185)	74.97±8.12	23.09±2.06	0.80±0.26	15.72±2.72
	pMCI	297(173/124)	73.38+7.07	27.14±1.73	0.51±0.08	16.04±2.62
	sMCI	700(414/286)	71.82±7.57	28.26±1.66	0.47±0.15	16.23±2.67
	NC	832(352/480)	72.52±5.76	29.05±1.22	0.01±0.05	16.68±2.51
AIBL	AD	79(33/46)	73.34±7.82	20.42±5.50	0.93±0.54	–
	pMCI	12(8/4)	74.92±5.98	26.25±1.60	0.5±0.0	–
	sMCI	24(10/14)	74.96±6.92	28.04±1.65	0.44±0.17	–
	NC	450(184/266)	72.47±6.22	28.73±1.21	0.03±0.11	–

The Scale Ratios (s_{po}, s_{sc}) in Inheritable 3D Deformable Self-Attention and 3D Deformable Self-Attention are set to [2,2] and [4,4].

A vanilla Binary Cross Entropy Loss is used to train our model, which is described as:

$$Loss(P,T) = \frac{1}{n}\sum_{i=1}^{n}(P[i], T[i]) \tag{17}$$

$$l(p,t) = -\left[t * \log p + (1-t) * \log(1-p)\right] \tag{18}$$

where n is the number of sMRI scans, P ant T are the probability and the target to the sMRI.

Four metrics are employed to evaluate the classification performance of the IDA-Net, namely, Accuracy (ACC), Specificity (SPE), Sensitivity (SEN) and Area Under Curve (AUC). Each of them can be computed as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{19}$$

$$Specificity = \frac{TN}{TN+FN} \tag{20}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{21}$$

where TP, FP, FN, and TN are denoted as true positive, false positive, false negative, and true negative, respectively.

3.3. Competing methods

Our IDA-Net method is compared with multiple state-of-the-art patch-based methods. The selected comparison methods are briefly described below.

1. Multi-Instance Convolutional Neural Network (MICNN): A Multi-Instance CNN (MICNN) model [51] is proposed for AD-related diagnosis, which embeds L parallel CNN subnetworks in a series of 6 convolutional layers to learn specific patch representations. L parallel patch features are then concatenated into a bag-level feature, the input to an additional FC layer to capture complex relationships between image patches.;
2. Hierarchical Fully Convolutional Network (HFCN): In this work [24], the Hierarchical Fully Convolutional Network (HFCN) consists of a series of patch subnetworks and region subnetworks, which utilizes patch representations to provide more direct and higher semantic information under multi-scale supervision. Finally, the outputs are spatially aggregated and processed by a global subject network to predict the probability of the classification.;
3. Dual Attention Multi-Instance Deep Learning (DA-MIDL): Dual Attention Multi-Instance Deep Learning (DA-MIDL) [1], as the method we mainly use for comparison, is composed of the Patch-Net and the attention multi-instance learning pooling (MIL-pooling). The Patch-Net not only learns spatial attention local patch features, but also outputs a contribution score for each patch. The MIL-pooling is designed to compute the relative contributions of the selected patches and produces a globally distinct weighted feature representation according to the contribution scores generated by the Patch-Net.

Table 2
Performance comparison for AD classification task (AD versus NC) and the prediction of MCI conversion task (pMCI versus sMCI) on ADNI test dataset.

Methods	AD vs NC				pMCI vs sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
MICNN	0.902	0.881	0.915	0.943	0.772	0.685	0.827	0.798
H-FCN	0.898	0.841	0.927	0.925	0.797	0.741	0.804	0.805
DA-MIDL	0.918	**0.926**	0.924	0.947	0.813	0.786	0.824	0.853
IDA-Net	**0.927**	0.919	**0.946**	**0.972**	**0.835**	**0.802**	**0.855**	**0.877**

Table 3
Performance comparison for MCI discrimination task (sMCI versus NC, pMCI versus NC) on ADNI test dataset.

Methods	pMCI vs NC				sMCI vs NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
MICNN	0.877	0.758	0.915	0.900	0.806	0.758	0.837	0.817
H-FCN	0.892	0.801	0.923	0.912	0.818	0.799	0.815	0.823
DA-MIDL	0.894	0.813	0.927	0.904	0.823	0.786	0.833	0.858
IDA-Net	**0.913**	**0.846**	**0.949**	**0.931**	**0.852**	**0.838**	**0.851**	**0.884**

Note that the comparison methods selected above are reproduced from the detailed model structure diagram or official code provided by the original paper on the same train and test datasets used in our paper. The Patch Location Proposals module in the above methods uniformly uses the same Patch Location Proposals methods mentioned in Dual Attention Multi-Instance Deep Learning (DA-MIDL). Considering that the number of subjects and pre-processing methods of the datasets used in different comparison methods are different, these competing methods may not achieve the same level of performance as their paper.

3.4. Performance on ADNI

The performance of the proposed IDA-Net method and other comparison methods for AD classification (AD versus NC) and the prediction of MCI conversion (pMCI versus sMCI) on the ADNI test dataset is shown in Table 2.

As shown in Table 2, in the vast majority of cases, our IDA-Net method has stronger classification performance on both AD classification and the prediction of MCI conversion tasks. Specifically, ACC (0.927), SPE (0.946) and AUC (0.972) of our IDE-Net method have more robust classification performance than other methods in AD classification. Furthermore, in the prediction of the MCI conversion task, all metrics (ACC (0.835), SPE (0.855), SEN (0.802) and AUC (0.884)) of our IDA-Net method outperform the results achieved by the other four methods. Meanwhile, it should be noted that some of the competing methods we reproduced have higher scores than their original papers on the prediction of the MCI conversion task. The likely reason is that the number of subjects in pMCI and sMCI in our selected ADNI dataset is nearly double that of their original paper, alleviating their possible underfitting problem. Compared to the state-of-the-art methods, our IDA-Net method has stronger classification performance overall on both classification tasks, especially in the more difficult prediction of

Table 4
Generalization comparison for AD classification task (AD versus NC) and the prediction of MCI conversion task (pMCI versus sMCI) on AIBL dataset.

Methods	AD vs NC				pMCI vs sMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
MICNN	0.886	0.864	0.898	0.926	0.737	0.647	0.787	0.756
H-FCN	0.884	0.826	0.917	0.910	0.771	0.720	0.781	0.778
DA-MIDL	0.905	**0.918**	0.906	0.929	0.789	0.757	0.798	0.827
IDA-Net	**0.909**	0.903	**0.935**	**0.961**	**0.812**	**0.790**	**0.835**	**0.854**

Table 5
Generalization comparison for MCI discrimination task (sMCI versus NC, pMCI versus NC) on AIBL dataset.

Methods	pMCI vs NC				sMCI vs NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
MICNN	0.866	0.736	0.904	0.873	0.789	0.743	0.817	0.793
H-FCN	0.869	0.789	0.902	0.885	0.784	0.765	0.789	0.796
DA-MIDL	0.885	0.803	0.902	0.904	0.783	0.769	0.827	0.834
IDA-Net	**0.896**	**0.826**	**0.928**	**0.914**	**0.834**	**0.814**	**0.839**	**0.866**

MCI conversion task. The most likely reason may be that our IDA-Net model can more accurately locate the multi-scale microstructure of brain atrophy and extract discriminative features in the sMRI. This also explains why we have superior improvement on various metrics in the prediction of MCI conversion task because the microstructure of brain atrophy between pMCI and sMCI are very similar, making it challenging to discriminate effectively.

To further verify our above conjecture, we conducted additional experiments on the MCI discrimination tasks (i.e., sMCI versus NC and pMCI versus NC). The MCI discrimination task faces the same challenging problems as the prediction of MCI conversion; slight changes in brain structure in the early stage are similar and difficult to distinguish. As shown in Table 3, our IDA-Net method also has a clear lead in the performance of these additional MCI discrimination tasks. For example, our IDE-Net method achieves good results in all metrics (ACC (0.913), SPE (0.949), SEN (0.846) and AUC (0.931)) in the MCI discrimination task of classifying pMCI from NC. Additionally, in the MCI discrimination task of classifying sMCI from NC, the ACC (0.852), SPE (0.851), SEN (0.838) and AUC (0.851) of the IDA-Net method are also significantly better than the other competing methods.

3.5. Generalization on AIBL

To evaluate the robustness and generalization of our proposed IDA-Net method, our proposed IDA-Net method and other comparison methods trained on ADNI dataset are tested on the independent AIBL dataset. The experimental results of our IDA-Net method and other comparison methods for the AD classification task, prediction of MCI conversion task, and MCI discriminations tasks on the AIBL dataset are shown in Table 4 and Table 5.

In all AD-related diagnosis tasks, our proposed IDA-Net method behaves better than other comparison methods on most metrics. For example, as shown in Table 4, our IDA-Net method also achieved stronger classification results in the prediction of MCI conversion task (0.812, 0.835, 0.790 and 0.827 for ACC, SPE, SEN and AUC), outperforming DA-MIDL methods (ACC (0.789), SPE (0.798), SEN (0.757) and AUC (0.827)). On the AIBL dataset, our IDA-Net method has the best AUC (0.961) for the AD classification task, outperforming MICNN (0.926), H-FCN (0.910), and DA-MIDL (0.929). As shown in Table 5, ACC (0.896), SPE (0.928), SEN (0.826) and AUC (0.914) of our IDA-Net method achieves stronger classification results than other comparison methods in the pMCI versus NC classification task. For the MCI discrimination task of classifying sMCI from NC, our IDA-Net method also acquires better metrics, especially on ACC (0.834), SEN (0.814) and AUC (0.866). In addition, the performance metrics of our IDA-Net method for each AD-related diagnosis task are not significantly degraded compared to the results reported in Table 2 and Table 3. These experimental results generally indicate that our IDA-Net method has good robustness and generalization performance in AD-related diagnosis tasks.

3.6. Comparison with previous works

To compare our method more broadly with related research on sMRI-based AD diagnosis, in Table 6, we summarize several state-of-the-art methods for AD classification task and prediction of MCI conversion task based on structural MRI from ADNI database reported in relevant literature [57], including two 3D patches of Hippocampus methods [52,54], three 3D patches of whole brain method [1, 24,51], one 2D image method [56], one single cross-sectional brain method [55] and one 3D GM volume method [53]. It is worth noting that since these methods use different numbers of subjects, different pre-processing procedures, different training, valuing, and testing set partitioning strategies, and even different pMCI/sMCI definitions, a direct comparison between these methods is potentially unreasonable and misleading. However, we can still draw some rough conclusions by briefly comparing our study with these state-of-the-art methods.

As shown in Table 6, compared to other methods, the ADNI dataset we constructed has the largest number of subjects (2248 subjects), which means that the classification performance and generalization of our IDA-Net method will be relatively stable without overfitting. Moreover, compared with only a slight lead in the AD classification task, our IDA-Net method has about 3%–8% improvement in each metric in the MCI conversion prediction task. This means that our IDA-Net method is more sensitive to the microstructure of brain atrophy in sMRI due to the introduction of our proposed 3D Deformable Self-Attention module and the Inheritable 3D Deformable Self-Attention module.

4. Discussion

4.1. Ablation studies on different attention block structure

To verify the effectiveness of the 3D Deformable Self-Attention module and Inheritable 3D Deformable Self-Attention module in our method, we try different attention block structures of three Self-Attention modules (i.e., 3D Shift Window Self-Attention, 3D Deformable Self-Attention, and Inheritable 3D Deformable Self-Attention) used in IDA-Net, and the results are shown inTable 7.

As shown in Table 7, even the performance of the IDA-Net$_{(SSS)}$ composed of only 3D Shift Window Self-Attention modules is close to that of most comparison methods in the AD classification task and the prediction of MCI conversion task, which fully demonstrates the potential of the Transformer structure. By replacing the last two-layer 3D Shift Window Self-Attention of the attention block of the IDA-Net$_{(SSS)}$ with the 3D Deformable Self-Attention module, the indicators of the IDA-Net$_{(SNN)}$ are improved by nearly 2%. Comparing the results of IDA-Net, IDA-Net$_{(SNN)}$ and IDA-Net$_{(SSS)}$, we can conclude the effectiveness of our proposed 3D Deformable Self-Attention and Inheritable 3D Deformable Self-Attention modules. As shown in Table 7, by reasonably combining multiple Self-Attention modules, the classification performance of the model is often better than that of only using a single Self-Attention module. Suppose the 3D Shift Window Self-Attention module is used in the first layer of the attention block. In that case, it can bring about a 1% performance improvement, which indicates that the 3D Shift Window Self-Attention module can obtain rich feature representation, which can help 3D Deformable Self-Attention and Inheritable 3D Deformable Self-Attention modules learn more accurate offset information.

Table 6
A brief comparison of state-of-the-art methods on AD classification task and prediction of MCI conversion task based on sMRI-based ADNI datasets.

Reference	Method	Subjects	AD vs NC				pMCI vs sMCI			
			ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Li et al. [52]	3D Patches of Hippocampus, 3D DenseNet + BGRUs	194 AD + 164 pMCI + 233 sMCI + 216 NC	0.891	0.846	0.931	0.910	0.725	0.610	0.825	0.746
Cui et al. [53]	3D GM volume, CNN + RNN	198 AD + 167 pMCI + 236 sMCI + 229 NC	0.9133	0.8687	**0.9520**	0.9322	0.7171	0.6527	0.7627	0.7303
Lian et al. [24]	3D patches of whole brain, CNN	358 AD + 205 pMCI + 465 sMCI + 429 NC	0.903	0.824	0.965	0.951	0.809	0.526	0.854	0.781
Cui and Liu [54]	3D patches of Hippocampus, 3D DenseNet + Shape analysis	192 AD + 165 pMCI + 231 sMCI + 223 NC	0.9229	0.9063	0.9372	0.9695	0.7500	0.7333	0.7619	0.7970
Liu et al. [51]	3D patches of whole brain, landmark-based + multi-instance	358 AD + 205 pMCI + 465 sMCI + 429 NC	0.9109	0.8805	0.9350	0.9586	0.7690	0.4211	0.8243	0.7764
Zhu et al. [1]	3D patches of whole brain, multi-instance + Dual Attention	468 AD + 189 pMCI + 325 sMCI + 707 NC	0.924	0.910	0.938	0.965	0.802	0.771	0.826	0.851
Sylvia et al. [55]	single cross-sectional brain, 3D Convolutions Network	294 AD + 253 pMCI + 510 sMCI + 352 NC	0.992	0.989	0.995	–	0.751	0.748	0.753	–
Lim et al. [56]	2D image, VGG-16+ResNET-50	192 AD + 398 MCI + 229 NC	0.786	0.821	0.737	–	–	–	–	–
Proposed	3D patches of whole brain, Inheritable 3D Deformable Attention	498 AD + 309 pMCI + 724 sMCI + 1282 NC	**0.927**	**0.919**	0.946	**0.972**	**0.835**	**0.802**	**0.855**	**0.877**

Table 7
Results of ablation studies with different attention block structure of Self-Attention modules.

Ablation methods	Attention block structure	AD vs NC				pMCI vs sMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
IDA-Net	[3D Shifted Window, 3D Deformable, Inheritable 3D Deformable]	**0.927**	**0.919**	**0.946**	**0.972**	**0.835**	0.802	**0.855**	**0.877**
IDA-Net$_{(SSS)}$	[3D Shifted Window, 3D Shifted Window, 3D Shifted Window]	0.901	0.893	0.920	0.947	0.813	0.787	0.836	0.854
IDA-Net$_{(SNN)}$	[3D Shifted Window, 3D Deformable, 3D Deformable]	0.923	0.913	**0.946**	0.968	0.828	**0.806**	0.850	0.874
IDA-Net$_{(SII)}$	[3D Shifted Window, Inheritable 3D Deformable, Inheritable 3D Deformable]	0.903	0.894	0.923	0.949	0.813	0.782	0.828	0.853
IDA-Net$_{(NNN)}$	[3D Deformable, 3D Deformable, 3D Deformable]	0.912	0.898	0.930	0.950	0.818	0.786	0.836	0.858
IDA-Net$_{(III)}$	[Inheritable 3D Deformable, Inheritable 3D Deformable, Inheritable 3D Deformable]	0.883	0.883	0.908	0.932	0.800	0.763	0.825	0.843
IDA-Net$_{(NNI)}$	[3D Deformable, 3D Deformable, Inheritable 3D Deformable]	0.919	0.913	0.939	0.963	0.825	0.792	0.846	0.868

4.2. Ablation studies on different scale ratio

We further investigate the effect of different maximum offset range Scale Ratio (s_{po}, s_{sc}) for our proposed Self-Attention modules.

As shown in Table 8, we limited the range of (s_{po}, s_{sc}) from 0 to 8 and completed a total of 9 sets of ablation experiments. When the 3D Deformable Self-Attention module uses a larger (s_{po}, s_{sc}) range and a smaller (s_{po}, s_{sc}) in Inheritable 3D Deformable Self-Attention is set, the performance of the IDA-Net method will be improved to a certain extent. Our proposed IDA-Net method is relatively insensitive to the Scale Ratio range of values. In the end, we chose [4,4] in 3D Deformable Self-Attention and [2,2] in Inheritable 3D Deformable Self-Attention as the final model used in our paper, namely IDA-Net*42.

4.3. Visualization

To verify the effectiveness of Inheritable 3D Deformable Self-Attention and compare it with the results of the original patch-based method, we visualize the entire process of the model implementing adaptive patch position and scale adjustment by fusing the offsets of multiple attention block layers. Specifically, we choose stage 3 of the IDA-Net model, composed of 6* stacked attention block layers, to facilitate a more coherent presentation of the entire change process. The 14 × 14 × 14 feature map of stage 3 is divided into 343 patches of size 2 × 2 × 2. We record the offsets of each attention block's deformed patches and project it into 3D space by Plotly packages.[3] And

[3] https://github.com/plotly/plotly.py

Table 8
Results of ablation studies with different scale ratio in 3D Deformable Self-Attention module and inheritable 3D Deformable Self-Attention module.

Ablation methods	Scale Ratio in	AD vs NC				pMCI vs sMCI			
	3D Deform&Inher 3D Defrom	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
IDA-Net^{*00}	[0,0] [0,0]	0.914	0.900	0.927	0.959	0.826	0.783	0.847	0.862
IDA-Net^{*11}	[1,1] [1,1]	0.921	0.912	0.941	0.965	0.831	0.799	0.851	0.873
IDA-Net^{*21}	[2,2] [1,1]	0.924	0.918	**0.946**	0.969	0.836	0.799	0.855	**0.878**
IDA-Net^{*22}	[2,2] [2,2]	**0.928**	0.914	0.945	0.970	**0.837**	0.799	0.854	0.874
IDA-Net^{*42}	[4,4] [2,2]	0.927	**0.919**	**0.946**	**0.972**	0.835	0.802	**0.855**	0.877
IDA-Net^{*33}	[3,3] [3,3]	0.926	0.908	0.930	0.971	0.836	**0.803**	0.851	0.875
IDA-Net^{*63}	[6,6] [3,3]	0.918	0.912	0.941	0.968	0.827	0.789	0.843	0.873
IDA-Net^{*44}	[4,4] [4,4]	0.920	0.911	0.942	0.967	0.829	0.793	0.850	0.874
IDA-Net^{*84}	[8,8] [4,4]	0.906	0.911	0.920	0.954	0.824	0.790	0.834	0.853

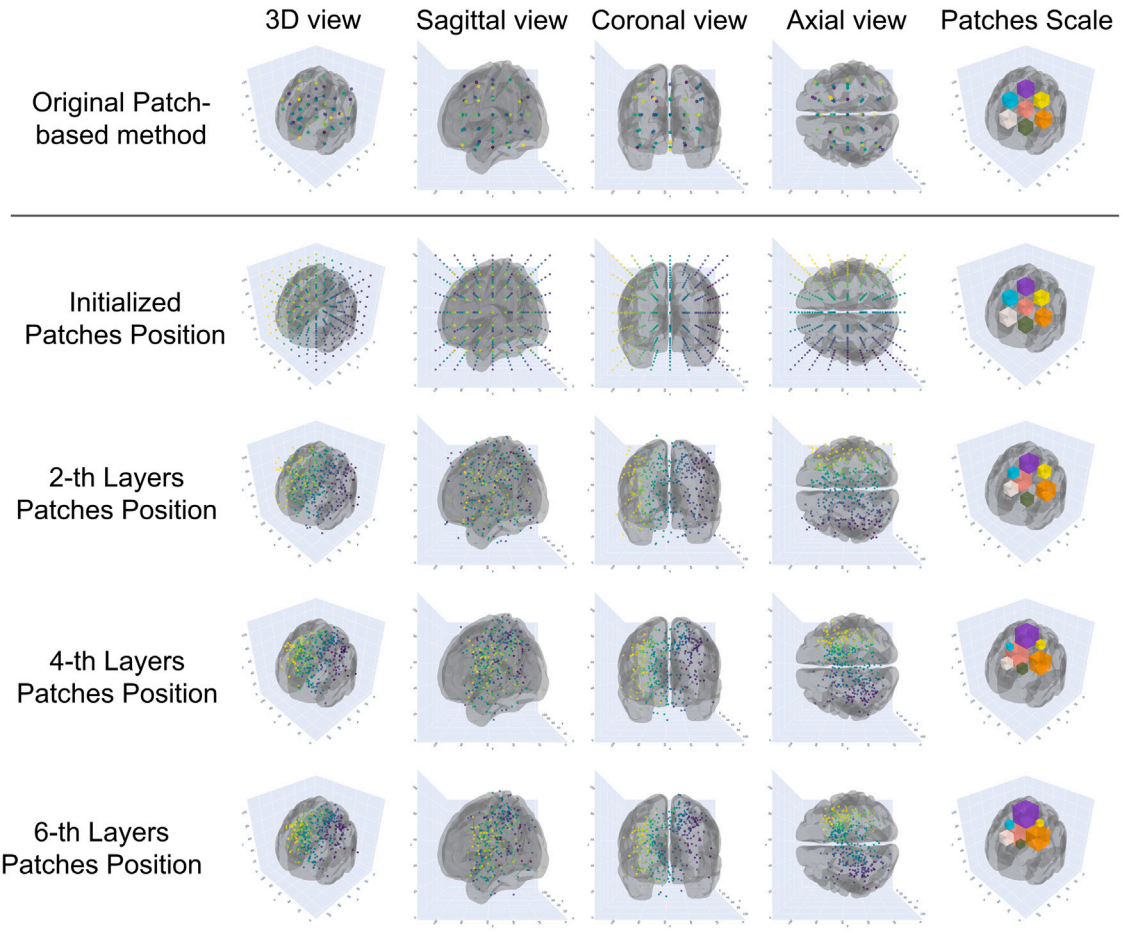


Fig. 4. Comparison of the position and scale of each patch between our proposed IDA-Net method and Patches Location Proposals in the original patch-based method (DA-MIDL) on one AD subject. The first four columns are visualizations of the coordinates of the center point position of each selected patch in different 3D perspectives. The last column visualizes the position and scale of some patches. The first row above the line represents the 60 patches selected by Patches Location Proposals based on Student's t-test and *p*-value ranking. The four rows below the line are visualizations of deformed patches after passing through which layer of attention blocks in stage 3 of IDA-Net.

for the visualization of the original patch-based method, we selected 60 patches by Student's t-test and p-value ranking based on the Patches Location Proposals proposed by DA-MIDL and visualized these patches using the same method.

As shown in Fig. 4, the first four columns are the coordinate visualization of the center point position of each selected patch in different 3D views. For intuitive observation, we only show the position and scale changes of a few patches in the last column of Fig. 4. The first row at the top of the line represents the 60 patches selected by Patches Location Proposals based on Student's t-test and p-value ranking. The four rows below visualize the deformed patches after which layer of attention blocks passed in the stage 3 of IDA-Net. As shown in the first row of Fig. 4, the position and scale of the patch selected by the original patch-based method are fixed, which leads to a lack of flexibility in patch selection and limits the classification performance of the model. In contrast, observing the movement trend of the center point of each patch in each attention block layer, the deformed patches are mainly concentrated in the upper middle region (i.e., frontal lobe, parietal cortex) and lower middle area (i.e., hippocampus, amygdala, and thalamus) of the whole brain, where neuritic plaque [58] and neurofibrillary tangles [59] which are highly related to Alzheimer's Disease usually appear. As shown in the last column of Fig. 4, different deformed patches can be viewed as different local feature extractors, which adaptively adjust their position and scale to capture essential features. These mean that our proposed IDA-Net method can locate and output possible discriminative pathological locations from the whole sMRI scans, which can be used as an optional auxiliary diagnosis method for doctors in the clinical diagnosis stage. Thus, our proposed IDA-Net method also solves the interpretability problem of Transformer structure application in clinical medical practice to a certain extent.

4.4. Limitations and future work

Although our IDA-Net method has excellent classification performance in various Alzheimer's Disease Diagnosis tasks, there still exist some limitations that may limit the performance of IDA-Net method. The following main limitations and potentially effective solutions are listed below. (1) As shown in Fig. 4, there are some deformed patches with considerable overlap, which brings too much repetitive information to our method and weakens the method's ability to extract the global feature of the entire sMRI. In future works, we can predict a contribution value for each patch and reduce those patches that overlap with large areas by merging or removing multiple overlap patches. (2) Although our proposed IDA-Net method has a better classification improvement over state-of-the-art methods in the MCI conversion prediction task. However, all metrics of the MCI conversion prediction task are still much lower than that of the AD classification task because the discriminative pathological features of pMCI and sMCI are very similar and difficult to distinguish. Therefore, Supervised Contrastive Learning methods [60] may be adopted to make the model more sensitive to discriminative features between pMCI and sMCI.

5. Conclusion

In this paper, inspired by Transformer, we proposed a unified end-to-end Inheritable Deformable Attention Network (IDA-Net) for sMRI-based Alzheimer's Disease diagnosis, which includes two major components: (1) 3D Deformable Self-Attention module for automatically adjusting the position and scale of the selected patch. (2) Inheritable 3D Deformable Self-Attention module for locating possible important regions with discriminative features in the global sMRI scans and outputting the specific location of the pathological regions. Our proposed IDA-Net method was evaluated on ADNI and AIBL datasets in multiple Alzheimer's Disease Diagnosis tasks. The experimental results of AD-related diagnosis demonstrated that our IDA-Net method acquires better classification performance and generalization than several state-of-the-art methods, especially for the prediction of MCI conversion.

CRediT authorship contribution statement

Qin Zhao: Conceptualization, Methodology, Software, Writing – original draft, Data curation. **Guoheng Huang:** Writing – review & editing, Supervision, Project administration, Validation, Resources, Funding acquisition. **Pingping Xu:** Writing – review & editing. **Ziyang Chen:** Writing – review & editing. **Wenyuan Li:** Writing – review & editing. **Xiaochen Yuan:** Writing – review & editing, Visualization. **Guo Zhong:** Writing – review & editing, Resources, Investigation. **Chi-Man Pun:** Formal analysis, Writing – review & editing. **Zhixin Huang:** Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Guoheng Huang reports financial support was provided by Guangdong University of Technology. Guoheng Huang reports a relationship with Guangdong University of Technology that includes: employment.

Data availability

The dataset is a public dataset from Internet. If necessary, we can share the link.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China-Guangdong Joint Fund under Grant U2001201, the Key-Area Research and Development Program of Guangdong Province under Grant 2018B010109007 and 2019B010153002, the Guangdong Provincial Key Laboratory of Cyber-Physical System under Grant 2020B1212060069, the Science and Technology Projects of Guangzhou under Grant 202007040006, Science and Technology Research in Key Areas in Foshan under Grant 2020001006832, the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515012534, and the National Statistical Science Research Project of China (No. 2022LY096).

References

[1] Wenyong Zhu, Liang Sun, Jiashuang Huang, Liangxiu Han, Daoqiang Zhang, Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI, IEEE Trans. Med. Imaging 40 (9) (2021) 2354–2366.

[2] Stephanos Leandrou, Styliani Petroudi, Panayiotis A. Kyriacou, Constantino Carlos Reyes-Aldasoro, Constantinos S. Pattichis, Quantitative MRI brain studies in mild cognitive impairment and Alzheimer's disease: a methodological review, IEEE Rev. Biomed. Eng. 11 (2018) 97–111.

[3] Qianyun Chen, Jill Abrigo, Wanting Liu, Elyia Yixun Han, David Ka Wai Yeung, Lin Shi, Lisa Wing Chi Au, Min Deng, Sirong Chen, Eric Yim Lung Leung, et al., Lower posterior cingulate N-acetylaspartate to creatine level in early detection of biologically defined Alzheimer's disease, Brain Sci. 12 (6) (2022) 722.

[4] Zhiwei Qin, Zhao Liu, Qihao Guo, Ping Zhu, 3D convolutional neural networks with hybrid attention mechanism for early diagnosis of Alzheimer's disease, Biomed. Signal Process. Control 77 (2022) 103828.

[5] Jayanthi Venkatraman Shanmugam, Baskar Duraisamy, Blessy Chittattukarakkaran Simon, Preethi Bhaskaran, Alzheimer's disease classification using pre-trained deep networks, Biomed. Signal Process. Control 71 (2022) 103217.

[6] Stephanos Leandrou, Demetris Lamnisos, Panicos A. Kyriacou, Stephanie Constanti, Constantinos S. Pattichis, et al., Alzheimer's Disease Neuroimaging Initiative, Comparison of 1.5 T and 3 T MRI hippocampus texture features in the assessment of Alzheimer's disease, Biomed. Signal Process. Control 62 (2020) 102098.

[7] Nitika Goenka, Shamik Tiwari, AlzVNet: A volumetric convolutional neural network for multiclass classification of Alzheimer's disease through multiple neuroimaging computational approaches, Biomed. Signal Process. Control 74 (2022) 103500.

[8] V.P. Subramanyam Rallabandi, K. Seetharaman, et al., Alzheimer's Disease Neuroimaging Initiative (ADNI), Classification of cognitively normal controls, mild cognitive impairment and Alzheimer's disease using transfer learning approach, Biomed. Signal Process. Control 79 (2023) 104092.

[9] Mingxia Liu, Daoqiang Zhang, Dinggang Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, IEEE Trans. Med. Imaging 35 (6) (2016) 1463–1474.

[10] Maria Vounou, Eva Janousova, Robin Wolz, Jason L. Stein, Paul M. Thompson, Daniel Rueckert, Giovanni Montana, et al., Alzheimer's Disease Neuroimaging Initiative, Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease, Neuroimage 60 (1) (2012) 700–716.

[11] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, et al., Alzheimer's Disease Neuroimaging Initiative, Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database, Neuroimage 56 (2) (2011) 766–781.

[12] Shuyu Li, Xiankun Yuan, Fang Pu, Deyu Li, Yubo Fan, Liyong Wu, Wang Chao, Nan Chen, Yong He, Ying Han, Abnormal changes of multidimensional surface features using multivariate pattern classification in amnestic mild cognitive impairment patients, J. Neurosci. 34 (32) (2014) 10541–10553.

[13] A. Demiriz, K.P. Bennett, J. Shawe-Taylor, Linear programming boosting via column generation, Mach. Learn. 46 (2002) 225–254.

[14] Chris Hinrichs, Vikas Singh, Lopamudra Mukherjee, Guofan Xu, Moo K. Chung, Sterling C. Johnson, Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset, Neuroimage 48 (1) (2009) 138–149.

[15] Youngsang Cho, Joon-Kyung Seong, Yong Jeong, Sung Yong Shina, Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data, Neuroimage 59 (3) (2012) 2217–2230.

[16] Daoqiang Zhang, Yaping Wang, Luping Zhou, Hong Yuan, Dinggang Shen, Multimodal classification of Alzheimer's disease and mild cognitive impairment, Neuroimage 55 (3) (2011) 856–867.

[17] Benoît Magnin, Lilia Mesrob, Serge Kinkingnéhun, Mélanie Pélégrini-Issac, Olivier Colliot, Marie Sarazin, Bruno Dubois, Stéphane Lehéricy, Habib Benali, Support vector machine-based classification of alzheimer's disease from whole-brain anatomical MRI, Neuroradiology 51 (2) (2009) 73–83.

[18] Peng Cao, Xiaoli Liu, Jinzhu Yang, Dazhe Zhao, Min Huang, Jian Zhang, Osmar Zaiane, Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures, Comput. Biol. Med. 91 (2017) 21–37.

[19] Juha Koikkalainen, Jyrki Lötjönen, Lennart Thurfjell, Daniel Rueckert, Gunhild Waldemar, Hilkka Soininen, Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease, Neuroimage 56 (3) (2011) 1134–1144.

[20] Jean-Claude Baron, Gaël Chételat, Béatrice Desgranges, G. Perchey, Brigitte Landeau, Vincent de La Sayette, Francis Eustache, In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease, Neuroimage 14 (2) (2001) 298–309.

[21] Jun Zhang, Yue Gao, Yaozhong Gao, Brent C. Munsell, Dinggang Shen, Detecting anatomical landmarks for fast Alzheimer's disease diagnosis, IEEE Trans. Med. Imaging 35 (12) (2016) 2524–2533.

[22] Manhua Liu, Daoqiang Zhang, Pew-Thian Yap, Dinggang Shen, Hierarchical ensemble of multi-level classifiers for diagnosis of Alzheimer's disease, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2012, pp. 27–35.

[23] Tong Tong, Robin Wolz, Qinquan Gao, Ricardo Guerrero, Joseph V. Hajnal, Daniel Rueckert, Multiple instance learning for classification of dementia in brain MRI, Med. Image Anal. 18 (5) (2014) 808–818.

[24] Chunfeng Lian, Mingxia Liu, Jun Zhang, Dinggang Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2018) 880–893.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[26] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.

[27] Chun-Fu (Richard) Chen, Quanfu Fan, Rameswar Panda, CrossViT: Cross-attention multi-scale vision transformer for image classification, in: ICCV, IEEE, 2021, pp. 347–356.

[28] Xiaoyue Chen, Sei-Ichiro Kamata, Weilian Zhou, Hyperspectral image classification based on multi-stage vision transformer with stacked samples, in: TENCON, IEEE, 2021, pp. 441–446.

[29] Hoang D. Nguyen, Xuan-Son Vu, Duc-Trong Le, Modular graph transformer networks for multi-label image classification, in: AAAI, AAAI Press, 2021, pp. 9092–9100.

[30] Lu He, Qianyu Zhou, Xiangtai Li, Li Niu, Guangliang Cheng, Xiao Li, Wenxuan Liu, Yunhai Tong, Lizhuang Ma, Liqing Zhang, End-to-end video object detection with spatial-temporal transformers, in: ACM Multimedia, ACM, 2021, pp. 1507–1516.

[31] Zhengyi Liu, Yuan Wang, Zhengzheng Tu, Yun Xiao, Bin Tang, TriTransNet: RGB-D salient object detection with a triplet transformer embedding network, in: ACM Multimedia, ACM, 2021, pp. 4481–4490.

[32] Zhigang Dai, Bolun Cai, Yugeng Lin, Junying Chen, UP-DETR: unsupervised pre-training for object detection with transformers, in: CVPR, Computer Vision Foundation / IEEE, 2021, pp. 1601–1610.

[33] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, Dustin Tran, Image transformer, in: International Conference on Machine Learning, PMLR, 2018, pp. 4055–4064.

[34] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, Vishal M. Patel, Medical transformer: Gated axial-attention for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 36–46.

[35] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, Manning Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, 2021, arXiv preprint arXiv:2105.05537.

[36] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[37] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.

[38] Ailiang Lin, Bingzhi Chen, Jiayu Xu, Zheng Zhang, Guangming Lu, David Zhang, DS-TransUNet: Dual swin transformer U-net for medical image segmentation, IEEE Trans. Instrum. Meas. 71 (2022) 1–15, http://dx.doi.org/10.1109/TIM.2022.3178991.

[39] Baiwen Zhang, Lan Lin, Shuicai Wu, A review of brain atrophy subtypes definition and analysis for Alzheimer's disease heterogeneity studies, J. Alzheimer's Dis. 80 (4) (2021) 1339–1352.

[40] Himashi Peiris, Munawar Hayat, Zhaolin Chen, Gary Egan, Mehrtash Harandi, A volumetric transformer for accurate 3d tumor segmentation, 2021, arXiv preprint arXiv:2111.13300.

[41] Xizhou Zhu, Han Hu, Stephen Lin, Jifeng Dai, Deformable ConvNets V2: More deformable, better results, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019.

[42] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, Yichen Wei, Deformable convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV, 2017.

[43] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, Jifeng Dai, Deformable DETR: Deformable transformers for end-to-end object detection, in: International Conference on Learning Representations, 2020.

[44] Jiahao Huang, Xiaodan Xing, Zhifan Gao, Guang Yang, Swin deformable attention U-net transformer (SDAUT) for explainable fast MRI, 2022, arXiv preprint arXiv:2207.02390.

[45] Jiashun Chen, Donghuan Lu, Yu Zhang, Dong Wei, Munan Ning, Xinyu Shi, Zhe Xu, Yefeng Zheng, Deformer: Towards displacement field learning for unsupervised medical image registration, 2022, arXiv preprint arXiv:2207.03180.

[46] Zhiyang Chen, Yousong Zhu, Chaoyang Zhao, Guosheng Hu, Wei Zeng, Jinqiao Wang, Ming Tang, Dpt: Deformable patch-based transformer for visual recognition, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 2899–2907.

[47] Zhuofan Xia, Xuran Pan, Shiji Song, Li Erran Li, Gao Huang, Vision transformer with deformable attention, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 4794–4803.

[48] John G. Sled, Alex P. Zijdenbos, Alan C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Trans. Med. Imaging 17 (1) (1998) 87–97.

[49] Jorge Jovicich, Silvester Czanner, Douglas Greve, Elizabeth Haley, Andre van Der Kouwe, Randy Gollub, David Kennedy, Franz Schmitt, Gregory Brown, James MacFall, et al., Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data, Neuroimage 30 (2) (2006) 436–443.

[50] William D. Penny, Karl J. Friston, John T. Ashburner, Stefan J. Kiebel, Thomas E. Nichols, Statistical Parametric Mapping: the Analysis of Functional Brain Images, Elsevier, 2011.

[51] Mingxia Liu, Jun Zhang, Ehsan Adeli, Dinggang Shen, Landmark-based deep multi-instance learning for brain disease diagnosis, Med. Image Anal. 43 (2018) 157–168.

[52] Fan Li, Manhua Liu, et al., Alzheimer's Disease Neuroimaging Initiative, A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease, J. Neurosci. Methods 323 (2019) 108–118.

[53] Ruoxuan Cui, Manhua Liu, et al., Alzheimer's Disease Neuroimaging Initiative, RNN-based longitudinal analysis for diagnosis of Alzheimer's disease, Comput. Med. Imaging Graph. 73 (2019) 1–10.

[54] Ruoxuan Cui, Manhua Liu, Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis, IEEE J. Biomed. Health Inf. 23 (5) (2018) 2099–2107.

[55] Silvia Basaia, Federica Agosta, Luca Wagner, Elisa Canu, Giuseppe Magnani, Roberto Santangelo, Massimo Filippi, et al., Alzheimer's Disease Neuroimaging Initiative, Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, NeuroImage: Clin. 21 (2019) 101645.

[56] Bing Yan Lim, Khin Wee Lai, Khairunnisa Haiskin, KA Saneera Hemantha Kulathilake, Zhi Chao Ong, Yan Chai Hum, Samiappan Dhanalakshmi, Xiang Wu, Xiaowei Zuo, Deep learning model for prediction of progressive mild cognitive impairment to Alzheimer's disease using structural MRI, Front. Aging Neurosci. 14 (2022).

[57] M. Khojaste-Sarakhsi, Seyedhamidreza Shahabi Haghighi, S.M.T. Fatemi Ghomi, Elena Marchiori, Deep learning for Alzheimer's disease diagnosis: A survey, Artif. Intell. Med. (2022) 102332.

[58] Pietro Tiraboschi, L.A. Hansen, L.J. Thal, Jody Corey-Bloom, The importance of neuritic plaques and tangles to the development and evolution of AD, Neurology 62 (11) (2004) 1984–1989.

[59] J.-P. Brion, Neurofibrillary tangles and Alzheimer's disease, Eur. Neurol. 40 (3) (1998) 130–140.

[60] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, Dilip Krishnan, Supervised contrastive learning, Adv. Neural Inf. Process. Syst. 33 (2020) 18661–18673.