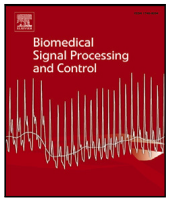
Biomedical Signal Processing and Control 95 (2024) 106404

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Multi-modal global- and local- feature interaction with attention-based mechanism for diagnosis of Alzheimer's disease

Nana Jia [a], Tong Jia [a,b,*], Li Zhao [c], Bowen Ma [a], Zheyi Zhu [a]

[a] Department of Artificial Intelligence, Northeastern University College of Information Science and Engineering, Shenyang, Liaoning, 504874, China
[b] Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Shenyang, Liaoning, 504874, China
[c] Department of Pulmonary and Critical Care Medicine, Shengjing Hospital of China Medical University, Shenyang, Liaoning, 504874, China

ARTICLE INFO

Keywords:
Alzheimer's disease diagnosis
Mild cognitive impairment
Multi-modal learning
Global–Local fusion

ABSTRACT

Alzheimer's disease is a complex neurodegenerative disease. Subjects with Mild Cognitive Impairment will progress to Alzheimer's disease, thus how to effectively diagnose Alzheimer's disease or Mild Cognitive Impairment using the clinical tabular data and Magnetic Resonance Images of the brain together has been a major concern of researches. Deep multi-modal learning-based methods can improve Alzheimer's disease diagnostic accuracy compared to the single modality-based methods. However, most existing multi-modal fusion methods only focus on learning global features fusion from image and clinical tabular data by concatenation, lacking the ability to jointly analyze and integrate global–local information of image with clinical tabular data. To address these limitations, this paper explored a novel Multi-Modal Global–Local Fusion method to perform multi-modal Alzheimer's disease classification through 3D Magnetic Resonance Images and clinical tabular data. Specifically, we adopt a global module that uses concatenation to fuse features to learn the global information. Moreover, we design an attention-based local module which encourages clinical tabular features to guide the learning of local 3D Magnetic Resonance Images information, thus, enhancing the power of features fusion from each modality. Our method considers both global and local information of the two modalities for multi-modal fusion. Experiment results show that our method in this paper is highly effective in combining 3D Magnetic Resonance Images and clinical tabular data for Alzheimer's disease classification with accuracy of 86.34% and 86.77% in ADNI and OASIS-1 datasets respectively, which outperforms the current state-of-the-art methods. Detailed ablation experiments are conducted to highlight the contribution of various components. code is available at: https://github.com/nananana0701/MMGLF.

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease with an insidious onset. Clinically, it is characterized by comprehensive dementia such as memory disorder, aphasia, agnosia, impairment of visuospatial skills, executive dysfunction, and personality and behavioral changes [1]. AD is one of the most growing health issue, which devastated many lives and will become a global burden in the coming decades. The number of people with Alzheimer's disease is approximately 50 million around the world in 2015 with more than half of them being early cases and is predicted to triple to 152 million by 2050 [1–3]. However, the basic understanding of the causes and mechanisms of AD are yet to be explored. Due to the rapid increase in the prevalence of AD, the accurate diagnosis of AD and its early stage, known as mild cognitive impairment (MCI), becomes very crucial for the timely treatment and possible delay of AD. In the past, the diagnosis of AD mainly relied on the evaluation of the individual's clinical tabular data, including patient medical history, clinical observation or cognitive evaluation [3]. Individual variables in tabular data capture rich clinical knowledge and thus high-quality clinical tabular data makes the diagnosis of AD more accurate, which can delay and control the further conversion of MCI into AD.

Recent studies have shown that Magnetic Resonance Imaging (MRI) is also used to detect the brain morphometric patterns for identifying disease-specific imaging biomarkers [4]. Numerous methods are introduced exploiting MRI data for distinguishing Alzheimer's Disease (AD), and its prodromal dementia stage, Mild Cognitive Impairment (MCI), from normal controls (CN) [5,6]. Leveraging the recent success of computer vision, deep convolutional neural networks (CNNs) have become the key technology for classification of Alzheimer's disease (AD) from

* Corresponding author at: Department of Artificial Intelligence, Northeastern University College of Information Science and Engineering, Shenyang, Liaoning, 504874, China.
E-mail address: jiatong@ise.neu.edu.cn (T. Jia).

https://doi.org/10.1016/j.bspc.2024.106404
Received 25 July 2023; Received in revised form 19 March 2024; Accepted 23 April 2024
Available online 7 May 2024

MRI of the brain. CNNs excel at extracting high-level information from MRI compared with manually extracted features [7–9]. However, brain MRI only offers a partial view on the underlying changes causing cognitive decline and studies that focus solely on the clinical tabular data neglect the brain MRI that also causes cognitive decline.

Therefore, clinicians and researches begin to use MRI and clinical tabular data together to classify Alzheimer's disease. The key idea of these methods is to learn the complementary fusion information from each modality to improve the classification performance. A typical way of learning fusion information is concatenating the feature vectors from different modalities. Each modality's feature vector provides information about different aspects of an object. However, clinicians always analyze global architecture changes and local distortions of a patient's 3D MRI with clinical tabular data together in an integrated manner for AD diagnosis. These existing fusion methods mainly focus on concatenated features by a global network using the whole image and clinical tabular data, which realize the global information fusion while ignoring the local information fusion between the two input modalities. Local information fusion can be leveraged to increase the confidence of the learned features for both modalities by encouraging the local information fusion of the feature vectors from two modalities. The local information is represented in multiple aspects, such as hippocampal morphosis and other potential shared characteristics between the two modalities; they are all critical for disease classification.

In this paper, we build a novel classification method, named global–local multi-modal fusion with attention mechanism, to learn the discriminative feature representations from 3D MRI and clinical tabular data. The flow chart of our method is shown in Fig. 1. We design a model in which a CNN's capacity can realize a global–local exchange of information from a patient's 3D MRI information and clinical tabular data instead of a typical concatenation of multiple data sources. This is achieved by a global module to concatenate the global features of each modality to obtain the global representation. Moreover, we propose a new local module to produce a representation by adopting attention-based fusion strategy to realize the clinical tabular features exchange with the local-information of 3D MRI feature maps. Lastly, we concatenate the global and local feature vectors of the two modalities to obtain more discriminative representations and feed them to a classifier for the final classification. In experiments on AD diagnosis, we show that our multi-modal global–local method leads to a superior predictive performance than using 3D MRI or clinical tabular data alone, and outperforms the state-of-the-art methods.

To summarize, the key contributions of the proposed multi-modal global–local framework for Alzheimer's disease classification are:

(1) a novel multi-modal fusion method is proposed to perform Alzheimer's disease Classification using 3D MRI and clinical tabular data. Its effectiveness is verified on a widely-used Alzheimer's disease Classification datasets, i.e., ADNI database;
(2) by adopting an attention-based mechanism strategy, our method can learn the local fusion information between the two modalities. More specifically, a modality discriminator is designed to guide the feature extractor to learn the local information explicitly;
(3) unlike most existing methods that only consider the global fusion information, our method considers the fusion between global–local information of 3D MRI and clinical tabular data.

The rest of this paper is organized as follows. First, a review of related work is provided in Section 2. In Section 3, we present the details of the material and our proposed method. In Section 4, we describe the experimental setups and report the experimental results. Section 5 offers Limitations and future work. Finally, Section 5 offers the conclusion.

2. Related work

This section reviews some related disease classification approaches, including single-modality and multi-modality Alzheimer's disease classification [10–13]. We will highlight how the proposed method differs from the existing methods.

2.1. Single-modality Alzheimer's disease classification

Identifying risk factors from patients' clinical tabular data is crucial as it helps AD management strategies, resulting in an improvement in the patients' life. Various risk factors have been previously identified in many studies, including patients' medical history, genetic data and cognitive evaluation and so on [14–16]. Researches attempted to pinpoint the risk factors with statistical tools in studies, such as logistic regression analyses [17,18]. In addition to that, the measurement of sensitive markers in the early stages of AD can help researches and clinicians develop new treatments and test their effectiveness. Various measurements such as structural atrophy, pathological amyloid deposition, and metabolic changes have already been shown to be sensitive to the diagnosis of AD and MCI [19]. Neuroimaging techniques [20–22] provide great help for the discovery of AD-related brain regions of interest (ROIs), which is a powerful instrument for classification of AD. For example, voxel-based measures extracted from structural MRI (VBM-MRI) and fluorodeoxyglucose positron emission tomography (FDG-PET), have been shown to be useful for investigating the neurophysiological features of AD and MCI [23–26].

In recent decades, machine learning and pattern recognition have been used in MRI for AD and MCI classification. For example, the researches extracted some features from certain ROI, such as the hippocampus on structure MRI [27] for the classification of AD. As the brain structure and clinical data related to AD are very complex, acquiring data from single modality (such as MRI or clinical tabular data alone) dose not provide enough sufficient feature information for AD diagnosis. Thus, researches begin to use MRI and clinical tabular data together for classification of AD. Numerous studies have shown that multi-modal data can provide complementary information, and the information fusion from different modalities can enhance classification performance. Thus, the accuracy of using multi-modal data for AD classification is better than that of single modality.

2.2. Multi-modality Alzheimer's disease classification

Existing models for image and clinical tabular data integration for disease classification can be divided into two categories [28]. The first category to combine the image and clinical tabular data is to directly concatenate tabular data and image features extracted using a CNN [29–32]. The authors of [33] first extracted regions of interest from brain MRI to obtain progression risk and then combined with baseline clinical data to predict progression to AD. However, image features are extracted independently of the clinical tabular data, which means the clinical tabular data may be used as redundant information, such as a patient's age, instead of complementing it. Thus, in [34], a single network is used to concatenate the clinical tabular data with the latent image representation prior to the last fully connected (FC) layer for survival prediction with histopathology images, genomic data, demographics and in [29,32], time-to-dementia was predicted with hippocampus shape and clinical markers. The disadvantage of this approach is that tabular data only contributes to the final prediction linearly. Based on this situation, a multi-layer perceptron (MLP) is used after concatenation to achieve non-linear relationships between image and tabular data. The authors of [31] concatenated digital pathology images and genomic data as inputs to MLP for cancer outcomes prediction, which is also used in [4,30] to learn from brain MRI and clinical markers for AD diagnosis. In addition to that, the authors of [35–37] firstly used a CNN and a MLP for the image data and

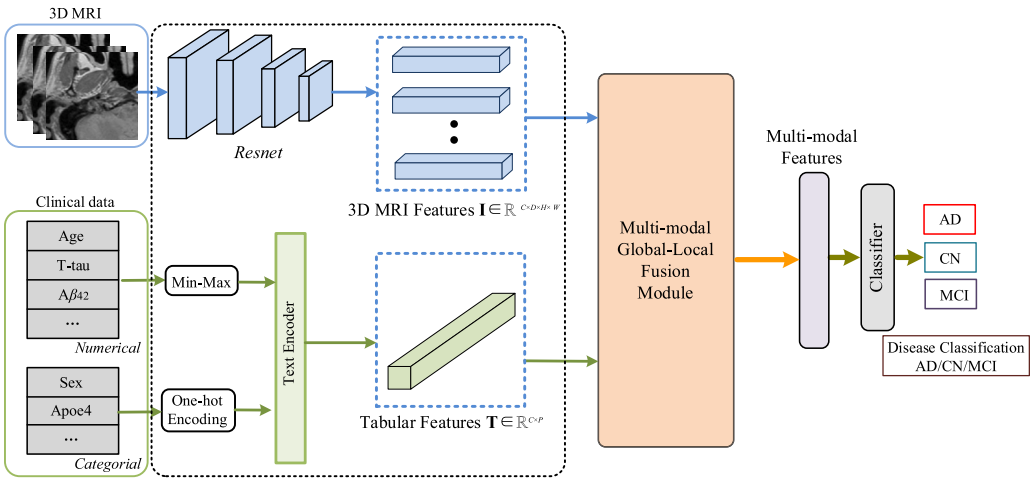


Fig. 1. The schematic illustration of our proposed multi-modal network for Alzheimer's classification based on 3D MRI and clinical tabular data.

tabular data respectively for feature extraction, and then concatenated extracted features to a MLP to achieve fusion for disease prediction. However, although using straightforward concatenation to exploit complementary modalities sometimes makes sense, it suffers from a major pitfall: straightforward concatenation means that multiple features are treated equally, which makes the complementary modalities have only minimal interaction and become incapable of being effectively utilized. In addition to that, concatenation based methods only integrate global information of each modality result in ignoring local information fusion of two modalities, so they cannot support local fine-grained interactions, which can lead to sub-optimal solutions.

The second type of methods are inspired by the mapping between language expressions and images [38–41]. For instance, the authors of [42] research the diagnosis of Alzheimer's disease. They construct an auxiliary neural network called DAFT, which dynamically incites or represses each feature map of a convolutional layer by utilizing tabular information that is complementary to the image information conditional on both image and tabular data. However, the DAFT network combined tabular data without encoding, which resulted the redundant information in tabular data cannot be removed and finally cause poor network performance. Meanwhile, this network uses the clinical tabular data to incite or repress each global image feature maps of a convolutional layer without exploring the local fine-grained fusion, which lead to performance degradation of fusion. Duanmu et al. [43] achieved the fusion model by channel-wise multiplication of the intermediate results of imaging and non-imaging clinical data branches for predicting response to chemotherapy. They use an auxiliary network that takes the tabular data and outputs a scalar weight for each feature map of every convolutional layer of their CNN. Thus, a patient's tabular data can amplify or repress the contribution of image-derived latent representations at multiple levels. The downside of this approach is that the number of weights in the auxiliary network scales quadratically with the depth of the CNN, which quickly becomes impracticable.

Above multi-modal fusion frameworks, most of them only focus on learning and using global information while the lack of global–local analysis may lead to sub-optimal performance. In our method, we consider both of global and local fusion strategy to learn high-level associations between 3D MRI and clinical tabular data and integrate them for the AD classification.

3. Material and method

3.1. Material

ADNI datasets. The data we used in this paper is collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). The ADNI project was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations with a 60 million, 5-year public–private partnership. The main purpose of this project is to verify whether brain MRI, other biomarkers, and neuropsychological assessments can be combined to measure the progression of AD and its early stage, MCI. In the current study, we included 1721 participants (Table 1) from the ADNI database. The datasets include all subjects from ADNI-1, ADNI-2, ADNI-3 and ADNI-GO, who had baseline MRI modality. We included three groups of participants: cognitively normal persons (CN), patients with Alzheimer's Disease (AD), and patients with mild cognitive impairment (MCI). Each participant contains MRI modality and metadata. Besides, AD is a form of dementia characterized by extracellular β-amyloid peptide plaque deposits and abnormal tau accumulation and phosphorylation which ultimately lead to neuronal and synaptic loss [44]. Thus, the clinical tabular data we used in this paper comprises 7 variables: age, sex, years in education, APoe4, cerebrospline fluid biomarkers Aβ, P-tau181, T-tau. There are 5 numerical and 2 categorical variables.

Table 1
The detailed attributes of ADNI and OASIS-1 datasets.

Datasets	Diagnosis	Subjects	Age	Male
ADNI(1/2/3GO)	AD	20.9%	74.7 ± 7.9	58.1%
	MCI	28.6%	74.9 ± 7.5	69.4%
	CN	50.5%	74.3 ± 7.9	49.6%
OASIS-1	Dementia	24.0%	78.9 ± 11.1	41.0%
	Non Dementia	76.0%	54.2 ± 26.2	37.7%

OASIS-1 datasets. The data can be obtained from the Open Access Series of Imaging Studies (www.oasis-brains.org), which is a project aimed at making magnetic resonance imaging (MRI) data sets of the brain freely available to the scientific community for Alzheimer's Disease. It is composed of 416 sample datasets (Table 1) divided into four subjects, which not only uses the brain MRI data (T1-weighted magnetic resonance imaging scans) from the original sagittal perspective as a modal data, but also preprocesses the text attribute information of each sample into another modal data. The text attribute information we used comprise 7 variables: sex, age, years in education, SES, MMSE, CDR, ASF, which are all related to Alzheimer's Disease. We included two groups of participants: Non Demented, and patients with Dementia.

3.2. Data preprocessing

Data preprocessing is the most essential step before applying feature extraction and fusion. It is not possible to utilize collected data directly

in the classification task, as it tends to be noisy, incomplete, and inconsistent. Therefore, a preprocessing step is applied to represent the data effectively for Alzheimer's disease classification. Data preprocessing includes missing data filtering, normalization and one-hot encoding of clinical tabular data and normalization of 3D MRI.

3.2.1. Missing data filtering of clinical tabular data

Since the datasets we used are an amalgamation of data from multiple related studies, most features are sparsely populated. Where measurements are missing, values are reconstructed using mean values from existing data which belongs to the same disease classification by using following equation:

$$\overline{X} = \frac{1}{n}\sum X_i. \tag{1}$$

where X_i represents the ith pattern of feature X within categories; $\overline{X}$ represents the mean of feature X under categories. In this work, $\overline{X}$ replaces the missing values of feature X within categories.

3.2.2. Normalization and one-hot encoding

We normalize images following the minimal pre-processing pipeline in [6]. All the input 3D MRI are resized to 64X64X64.

Clinical tabular data, such as ADNI datasets, contains a number of features, and we used in this paper includes 5 numerical and 2 categorical variables. Every numerical variable includes different numerical values, which increases the difficulties during the computation process. Therefore, a normalization technique is used to normalize numerical variable in the range between 0 and 1, as well as to decrease the numerical complexity during the computational process of Alzheimer's disease diagnosis. In this paper, the well-known min–max normalization method is used. The process of normalization is realized by using the following equation:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \times [new\_max - new\_min] + new\_min. \tag{2}$$

here, X_{norm}, X, X_{min}, and X_{max} are the normalized data value, the original data value, the minimum data value, and the maximum data value, respectively, in the entire datasets, while new_max and new_min indicate the range of the converted datasets. We use new_max = 1 and new_min = 0. Using this method, all the features' values lie within the interval [0, 1].

The categorical variables (eg, sex, APoe4) are presented by one-hot encoding, resulting in 6 binary features.

3.3. Framework of global–local multi-modal fusion with attention mechanism

The framework of our proposed is shown in Fig. 1, from which we can see that model contains two input branches: the 3D MRI branch and the clinical tabular data branch. When the 3D MRI and clinical tabular data are input into model, they will go through a ResNet [45] and a text encoder (a 1D convolution, batch normalization and sigmoid linear unit (SiLU)) for extracting 3D MRI features and clinical tabular features respectively. We obtain the feature maps within the last residual block as the extracted image representations and the last layer features of the text encoder as the extracted clinical tabular representations. Then our method will input the extracted representations into three modules: a global fusion module, a local attention-based fusion module, and a classification module.

These three modules and the feature extractor are trained jointly in an end-to-end manner to guide the feature extractor in learning both global and local fusion features. We will illustrate the details below.

Let $D_s = ((x^I, x^T, Y)_i)_{i=1}^N$ be the set of ADNI evaluation datasets, where $x^I \in \mathbb{R}^{w\times h\times d\times 1}$ and $x^I \in \mathbb{R}^{p\times 1}$ denote the ith 3D MRI and clinical tabular data respectively. w, h and d are the width, height and depth of the input 3D MRI with 1-channel. p denotes the number of tabular data features. $Y_i = \{0, 1, 2\}$ denotes the label for Alzheimer's disease classification task (0 = AD, 1 = CN, 2 = MCI). Lastly, N is the total number of subjects. The goal of our proposed method is to train the neural network as a function to map input 3D MRI and clinical tabular data from input space to its label space, where we need to obtain the trainable parameters of the neural network model.

3.3.1. Global multi-modal fusion

As mentioned above, clinicians always use global (e.g., architecture change of the whole brain area) and local information (e.g., hippocampal morphosis) of 3D MRI and clinical tabular data together to analyze Alzheimer's disease. To emulate the multi-modal global analysis of two modalities, our global fusion module is trained to achieve two objectives: (1) the global representations of two modalities should be similar to each other, and (2) the complementarity of both modalities has to be effectively characterized.

The global feature representations fusion of two modalities is implemented by the global fusion block, as shown in Fig. 2. We first transform the 3D MRI feature maps $I \in \mathbb{R}^{C\times D\times H\times W}$ into feature vectors $I^g \in \mathbb{R}^C$ by global max pooling and clinical tabular features $T \in \mathbb{R}^{C\times P}$ by text encoder. Where H, W and D are the spatial height, weight and depth, with C being channels and P being dimension of clinical tabular features.

To penalize differences between the two global representations, we define the mapping functions $I^I = f^{T-I}(T)$ that can learn a projection f from the clinical tabular representations to 3D MRI representations, and then we obtain the global feature representations by $U_{global-fusion} = [I^g, I^I]$, which characterize the complementary of both two inputs. Here, $[\cdot, \cdot]$ represents the concatenation operation. Next, the global fusion feature representations $U_{global-fusion}$combined two modalities features will join the next stage.

3.3.2. Attention mechanism-based local fusion

The local attention module aims to build feature vectors from the two inputs by analyzing interactions between samples from 3D MRI and clinical tabular data. We devise a local-aware fusion module to generate the 3D MRI embedding features with the guidance of the clinical tabular information, thus the output carries the information from each modality. Inspired by the recent application of attention mechanism in deep neural networks [9,46], we use the attention mechanism based fusion method to achieve fine-grained fusion of multi-modal features to learn the disease features. The benefit of our strategy is that we can use clinical tabular information to calculate weights represent the importance of each feature for 3D MRI. An overall representation of the input is then computed with the weights as a weighted combination of all the input 3D MRI features. Attention weights with greater values are higher priorities in determining the corresponding significant input 3D MRI features. Our proposed method comprises the attention module that generates local fusion features from clinical tabular features guided 3D MRI features for AD diagnosis classification task.

The schematic diagram of the local fusion operation is shown in Fig. 2. For a brief and detailed description, we introduce the feature map $I \in \mathbb{R}^{H\times W\times D\times C}$ along the channels, which we consider important for local fine-grained fusion. Starting from the 3D MRI backbone features I, we transform I into the main local feature matrix $I_{local} = [f^{1,1,1}, f^{1,2,1}, \ldots, f^{i,j,k}, \ldots, f^{H\times W\times D}]$, where $f^{i,j,k} \in \mathbb{R}^{1\times 1\times 1\times C}$ corresponding to the spatial location (i, j, k) with $i \in \{1, 2, \ldots, H\}$, $i \in \{1, 2, \ldots, W\}$ and $i \in \{1, 2, \ldots, D\}$. The clinical tabular features $T \in \mathbb{R}^{C\times P}$. To estimate fusion of the local 3D MRI information with clinical tabular features, let $f_m^{i,j,k}$ to be the m_{th} features of I_{local}, we apply the attention-based weighted fusion method to the representations:

$$Q_{local-fusion} = \sum_{k=1}^{n} a_k f_m^{i,j,k}. \tag{3}$$

$$a_k = \frac{exp((TW_q)\cdot(f_m^{i,j,k}))}{\sum_{k=1}^{K} exp(T\cdot f_m^{i,j,k})}. \tag{4}$$

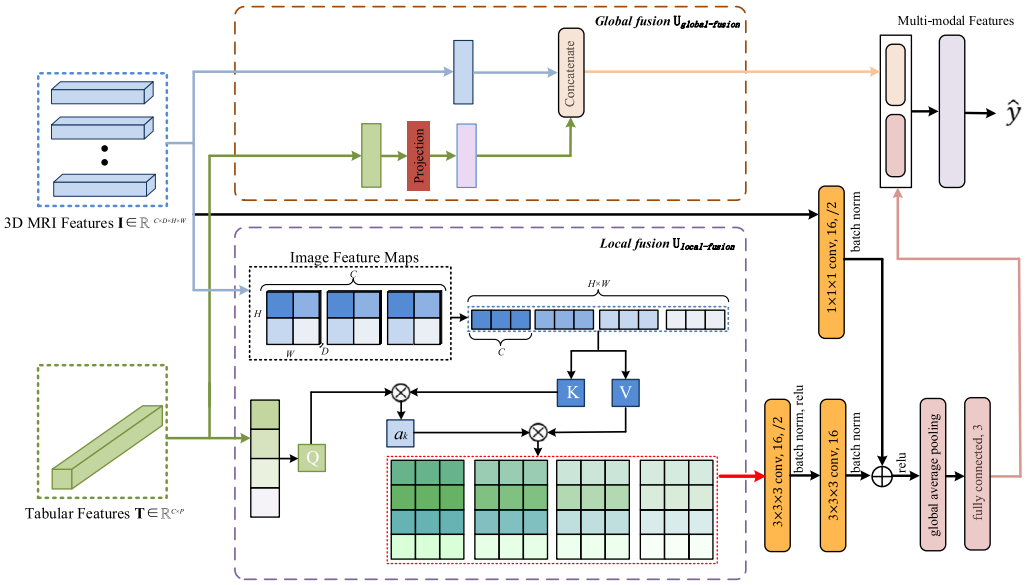


Fig. 2. The diagram of the fusion block.

where W_q, W_k^T are the linear layers. The local fusion features for two modalities are obtained as $Q_{local-fusion}$. Each $Q_{i,j,k}$ represents the depth fusion feature representation which contains feature information on the spatial position (i, j, k) of all channels C in 3D MRI feature maps and the clinical tabular features. After this operation, each feature maps carry clinical tabular information and 3D MRI information, and the two modalities realize complement each other. Then the fused local features $Q_{local-fusion}$ are passed through the fusion block's subsequent convolutional layer to obtain the final local fusion feature representations $Q^F_{local-fusion}$. the local features for each modality is obtained with $Q_{local-fusion} = GAP(Q^F_{local-fusion})$, GAP represents the global average pooling. Finally, the integration of the global and local module features is obtained via a concatenation fusion, where we concatenate $U_{global-fusion}$ and $U_{local-fusion}$ before applying the last MLP classification layer.

3.4. Classification module

The classification module is utilized to classify the multi-modal input feature maps into their corresponding categories. Firstly, this module concatenates the global features and local features from both 3D MRI and clinical tabular data to obtain global–local fusion feature. Then the concatenated feature vectors will go through a network of two layers. The first layer is a fully-connected layer, which is followed by the softmax layer that is used to classify the inputs into three disease categories.

In our method, we evaluate the multi-modal method on the task of diagnosing subjects as cognitively normal (CN), mild cognitive impaired (MCI), or demented (AD). We formulate the diagnosis task as a classification problem. The loss function of the classification module is the cross-entropy loss:

$$L_{log}(y, \hat{y}) = -(ylog(\hat{y}) + (1-y)log(1-\hat{y})). \quad (5)$$

where y is the ground truth label for task and $\hat{y}$ is the predict label for task.

3.5. Exploration of critical brain regions for classification with Class Activation Mapping (CAM)

CAM technique can provide which parts of the medical image affect the classification decisions made by methods for tasks. In [47], the authors have given the images obtained using the gradient-weighted class activation mapping (Grad-CAM) technique to inspect the network predictions for the COVID-19 detection task. Thus, we use class activation maps (CAM) to demonstrate our model's effectiveness visually to indicate the brain regions relevant for AD classification. In visualizing the activation for the network's decision, we explored class activation mapping (CAM), a technique that incorporates a global average pooling (GAP) layer succeeding the last convolutional layer in any image classification task. We visualize the activations to interpret our proposed model's robustness and performance without treating it as a black-box. This technique provides remarkable localization performance on discriminative features in the images by generating heat maps to buttress the network's performance. Specifically, we applied the CAM to highlight the parts of the brain that are discriminative for AD classification. The CAM technique is extended to a 3D architecture to produce the activations in the AD prediction task. The class activation map harnesses the activations produced by the last convolutional layers in visualizing the discriminative features. The method projects the class weights of the output layer onto the activation maps in the last convolutional layer. Furthermore, a weighted sum of the features in the last convolutional layer generates the activations. In implementing the 3D CAM for this work, we modified the last block by replacing the max-pooling layer with the global average layer to have the desired architecture in generating a class activation map [48].

4. Experiments study

4.1. Experimental settings

4.1.1. Training and test details

Our proposed network was implemented with PyTorch and trained on an Intel® CoreTM i5-4460 Processor paired with a NAVIDIA GEFORCE GTX 3060 GPU. During training process, our network is trained using the Adam optimizer, with an initial learning rate of 0.0023 to optimize the entire neural network, batch size of 8 and epochs of 30. In addition, We utilized the five-fold cross-validation, and divided all data into five folds at the subject level. In each training session, one fold was used for testing and the remaining four folds were used for training.

Table 2
Ablation results of global fusion module in terms of accuracy (%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

M	LFM only	Add	Multiply	GFM	AD	CN	MCI	Avg.	Ttrain	Ttest
I	✓				85.26 ± 0.55	85.77 ± 0.46	83.56 ± 0.49	84.86 ± 0.49	10.3 h	9.9 ms
II	✓	✓			85.56 ± 0.51	86.33 ± 0.36	83.73 ± 0.24	85.19 ± 0.37	10.4 h	10.3 ms
III	✓		✓		85.56 ± 0.48	86.88 ± 0.73	82.44 ± 0.64	84.96 ± 0.62	10.4 h	10.3 ms
Ours	✓			✓	86.84 ± 0.49	87.69 ± 0.57	84.22 ± 0.57	**86.34 ± 0.54**	10.4 h	11.3 ms

Note that 201800B1std2019 represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Table 3
The detailed ablation results of global fusion module in terms of Specificity (Spec.), precision (Prec.), AUC and F1-score(%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Method	Met.	AD	CN	MCI	Avg.
I	Spec.	95.68 ± 0.69	93.51 ± 0.50	88.86 ± 0.60	92.68 ± 0.60
	Prec.	85.79 ± 0.64	91.41 ± 0.35	76.56 ± 0.60	84.58 ± 0.53
	AUC	92.12 ± 0.29	96.24 ± 0.40	93.57 ± 0.36	**93.98 ± 0.35**
	F1-score	79.66 ± 0.55	90.38 ± 0.38	82.78 ± 0.24	84.27 ± 0.39
II	Spec.	95.91 ± 0.60	96.66 ± 0.42	84.54 ± 0.38	92.37 ± 0.47
	Prec.	81.79 ± 0.43	92.64 ± 0.38	74.62 ± 0.39	83.02 ± 0.40
	AUC	95.24 ± 0.42	96.71 ± 0.47	84.76 ± 0.57	92.23 ± 0.49
	F1-score	78.76 ± 0.61	89.82 ± 0.86	81.22 ± 0.38	83.27 ± 0.62
III	Spec.	97.08 ± 0.32	91.42 ± 0.42	89.34 ± 0.48	92.61 ± 0.41
	Prec.	89.95 ± 0.47	90.15 ± 0.42	76.02 ± 0.62	85.37 ± 0.50
	AUC	94.44 ± 0.36	96.82 ± 0.72	86.77 ± 0.68	92.68 ± 0.59
	F1-score	82.16 ± 0.60	89.76 ± 0.86	81.17 ± 0.57	84.36 ± 0.68
Ours	Spec.	97.56 ± 0.57	93.38 ± 0.38	88.86 ± 0.56	**93.27 ± 0.50**
	Prec.	91.59 ± 0.59	91.37 ± 0.55	78.74 ± 0.58	**87.22 ± 0.57**
	AUC	94.63 ± 0.58	96.87 ± 0.62	86.95 ± 0.47	92.82 ± 0.56
	F1-score	81.35 ± 0.42	91.36 ± 0.38	83.13 ± 0.58	**85.44 ± 0.46**

Note that 201800B1std2019 represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

4.1.2. Evaluation metrics

We use accuracy (Acc.), F1 score, specificity (Spec.), precision (Prec.) and the area under the receiver operator characteristic curve (AUC) as our evaluation criteria metrics. F1 score = 2 × precision × recall / (precision + recall). The definitions of accuracy, F1 score, specificity, and precision are as follows:

$$\begin{cases} Acc = \frac{TP+TN}{TP+FP+TN+FN}, \\ Spec = \frac{TN}{FP+TN}, \\ Prec = \frac{TP}{TP+FP} \\ F1score = 2 \times Prec \times \frac{Recall}{Prec + Recall} \end{cases} \tag{6}$$

where TP, FP, TN, and FN are the numbers of the true positive, false positive, true negative, and false negative samples, respectively. recall = TP/(FP+FN). For all these metrics, the larger values indicate the better performance.

4.2. Experimental results and analysis

4.2.1. Ablation study

To better understand the contributions of each module in our proposed method and which settings can best integrate clinical tabular data, we perform an ablation study to demonstrate the effectiveness of the global fusion module, the local fusion module, the text encoder module, and which the location of fusion block within the last residual block can get better performance.

1. Effectiveness of the global fusion module.

To evaluate the effectiveness of the global fusion module, we compared the performance of different global fusion methods. In model I, model II, and model III, we conducted experiments without global fusion module, add and multiply global feature representations of two modalities, respectively. Specifically, in Model I, only local fusion module is utilized to fuse multi-modal feature representations. In Model II, local fusion module and global fusion module that added the global representations of two modalities are utilized for AD classification. In Model III, local fusion module and global fusion module that multiplied the global representations of two modalities are utilized for AD classification.

The performance of these models is shown in Tables 2 and 3, "Avg". denotes the average score over the entire row. The performance of Model I is obviously worse than others, demonstrating that using only local fusion module to fuse multi-modal feature representations cannot obtain the best performance for AD classification. This is not surprising, since there is need in classifying AD using global fusion feature representations. Model II performs better than Model III, due to the fused feature representations are also zeros when the feature representations of a modality exists zero values using multiply for global feature fusion. Besides, the proposed method further improves Model I on most metrics (4 out of 5), verifying the effectiveness of using concatenation for global fusion and the effectiveness of using global and local fusion module together for AD classification.

2. Effectiveness of the local fusion module and text encoder module.

To validate the effectiveness of local fusion module, we discard local fusion module and perform global fusion module only for multi-modal fusion, as implemented by Model 1. Besides, we also compare our proposed local fusion module against a self-attention (SA) module, which integrates features of two modalities and performs the self-attention mechanism, as implemented by Model 2. Meanwhile, we conducted an ablation study to evaluate the effectiveness of text encoder module, as implemented in Model 3. In Model 3, the clinical tabular data is fused with features of 3D MRI without encoding. The performance of these models is shown in Tables 4 and 5. It reveals that the proposed method outperforms Model 1 with a significant AUC improvement of 4.54%, demonstrating that local fusion module for multi-modal fusion is helpful for AD classification. Moreover, comparing to Model 2, our proposed method achieved better performance on metrics (4 out of 5), which means the effectiveness of our proposed local fusion module based on attention. According to the results of Model 3, our proposed method obtains the better performance when encodering the clinical tabular data via text encoder module. All of those demonstrate our

Table 4
Ablation results of local fusion module and text encoder module in terms of accuracy (%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

M	GFM only	ASA	LFM	Text encoder	AD	CN	MCI	Avg.	Ttrain	Ttest
1	✓			✓	85.35 ± 0.31	84.71 ± 0.60	83.12 ± 0.49	83.79 ± 0.47	9.7 h	9.5 ms
2	✓	✓		✓	85.34 ± 0.45	86.76 ± 0.46	82.48 ± 0.47	84.86 ± 0.47	10.1 h	9.8 ms
3	✓		✓		83.44 ± 0.57	84.22 ± 0.36	82.87 ± 0.54	83.51 ± 0.49	10.3 h	10.1 ms
Ours	✓		✓	✓	86.84 ± 0.49	87.69 ± 0.57	84.22 ± 0.57	**86.34 ± 0.54**	10.4 h	11.3 ms

Table 5
The detailed ablation results of proposed global fusion module in terms of Specificity (Spec.), precision (Prec.), AUC and F1-score(%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Method	Met.	AD	CN	MCI	Avg.
1	Spec.	95.83 ± 0.47	91.94 ± 0.57	89.81 ± 0.53	92.53 ± 0.52
	Prec.	87.22 ± 0.53	89.56 ± 0.36	77.36 ± 0.48	84.71 ± 0.46
	AUC	94.36 ± 0.36	96.22 ± 0.44	74.34 ± 0.38	88.31 ± 0.39
	F1-score	81.71 ± 0.44	90.08 ± 0.33	80.52 ± 0.44	84.10 ± 0.40
2	Spec.	95.66 ± 0.38	96.63 ± 0.37	82.82 ± 0.28	91.70 ± 0.34
	Prec.	84.84 ± 0.55	94.31 ± 0.34	68.79 ± 0.48	82.64 ± 0.46
	AUC	95.12 ± 0.33	96.32 ± 0.36	90.72 ± 0.29	**94.05 ± 0.33**
	F1-score	76.69 ± 0.34	89.36 ± 0.42	77.91 ± 0.36	81.32 ± 0.37
3	Spec.	95.58 ± 0.38	93.84 ± 0.30	87.52 ± 0.38	92.31 ± 0.35
	Prec.	84.39 ± 0.45	91.46 ± 0.38	75.96 ± 0.24	83.93 ± 0.36
	AUC	93.38 ± 0.42	95.39 ± 0.24	93.32 ± 0.32	94.03 ± 0.33
	F1-score	79.72 ± 0.45	88.48 ± 0.55	81.76 ± 0.46	83.32 ± 0.49
Ours	Spec.	97.56 ± 0.57	93.38 ± 0.38	88.86 ± 0.56	**93.27 ± 0.50**
	Prec.	91.59 ± 0.59	91.37 ± 0.55	78.74 ± 0.58	**87.22 ± 0.57**
	AUC	94.63 ± 0.58	96.87 ± 0.62	86.95 ± 0.47	92.82 ± 0.56
	F1-score	81.35 ± 0.42	91.36 ± 0.38	83.13 ± 0.58	**85.44 ± 0.46**

Table 6
Ablation results of different location in terms of accuracy (%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Location	AD	CN	MCI	Avg.	Ttrain	Ttest
Location 1	83.88 ± 0.46	85.51 ± 0.32	83.12 ± 0.32	84.17 ± 0.37	10.7h	11.9 ms
Location 2	82.12 ± 0.29	83.63 ± 0.44	81.34 ± 0.38	82.36 ± 0.37	11.0h	12.9 ms
Location 3	83.13 ± 0.29	84.06 ± 0.21	82.34 ± 0.28	83.18 ± 0.28	10.5h	11.8 ms
Ours	86.84 ± 0.49	87.69 ± 0.57	84.22 ± 0.57	**86.34 ± 0.54**	10.4h	11.3 ms

proposed method achieves the best performance when utilizing both local fusion module and text encoder module.

3. Location of fusion block within the last residual block.

To better understand under which the location of fusion block within the last residual block can get better performance, we set the fusion module as positions before last residual block (Location 1), before the first convolutional layer (Location 2), before the second convilutional layer (Location 3), and before the first ReLU (MMGLF). The experimental results of different location in Tables 6 and 7 show that the proposed method is relatively robust to the choice of location. Moreover, when the fusion module block is located before the first ReLU within the last residual block, the results were further increased to the best performance.

4.2.2. Comparison with the state-of-the-art methods

To evaluate the performance of our proposed model in classifying Alzheimer's disease, we compared against existing methods on the ADNI datasets. All the methods are evaluated with the same datasets. The comparison methods include: (1) only using 3D MRI to classify Alzheimer's disease; (2) only using clinical tabular data to classify Alzheimer's disease; (3) the clinical tabular data is concatenated with the latent image feature vectors and then the concatenated vectors are fed directly to the final classification layer [29,32,34]; (4) the concatenated vector is fed to an FC bottleneck layer prior to the classification layer [4,30]; (3) and (4) fuse the representations of different modalities by concatenating the representations from different modalities at the end of feature extractor. (5) a general-purpose module for CNNs was proposed that dynamically rescales and shifts the feature maps of a convolutional layer, conditional on a patient's tabular clinical information [42]; (6) the channel-wise multiplication of the intermediate results of imaging and non-imaging data [43] was originally proposed for prediction of breast cancer, which trained MRI data and non-imaging clinical data with one informing the other at the intermediate stages of the CNN (Interactive-model) and then multi-modal features (original clinical tabular data and encoded image data) were multiplied at three levels for final prediction.

Firstly, the results of the three classification tasks in terms of accuracy are reported in Table 8. From Table 8, the average accuracy of different models indicate that our model is superior compared with the uni-modal and other multi-modal fusion approaches, followed by the second best [42] and the third best [43] multi-modal networks. Combining two modalities can effectively improve model performance according to the improved performance over uni-modal analysis, which indicates the significance of multi-modal fusion. According to Table 8, the baseline model is ResNet using image and text encoder using tabular data, which obtained the accuracy of 70.15% and 72.52% respectively. The reason of text encoder using clinical tabular data outperforms ResNet using 3D MRI is clinical tabular data comprises amyloid-specific measures derived from cerebrospinal fluid and PET images that are known to become abnormal before changes in MRI are visible. All Concatenation networks are succeed in extracting global complementary information, resulting in improved performance compared with the uni-modal. However, these results clearly demonstrate that the Concatenation networks fall behind compared with other fusion frameworks in this paper. The main reason why the concatenation-based fusion methods perform worse is that even

Table 7
The detailed ablation results of different location in terms of Specificity (Spec.), precision (Prec.), AUC and F1-score(%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Method	Met.	AD	CN	MCI	Avg.
Location 1	Spec.	97.42 ± 0.41	91.68 ± 0.55	87.46 ± 0.48	92.19 ± 0.48
	Prec.	89.52 ± 0.35	90.45 ± 0.68	74.84 ± 0.44	84.94 ± 0.49
	AUC	93.50 ± 0.40	96.12 ± 0.44	88.81 ± 0.58	92.81 ± 0.47
	F1-score	78.73 ± 0.33	90.47 ± 0.38	80.71 ± 0.62	83.30 ± 0.44
Location 2	Spec.	95.24 ± 0.48	87.48 ± 0.41	90.62 ± 0.47	91.11 ± 0.45
	Prec.	83.69 ± 0.43	85.76 ± 0.43	78.49 ± 0.50	82.62 ± 0.45
	AUC	92.21 ± 0.37	95.32 ± 0.34	91.43 ± 0.36	**92.99 ± 0.36**
	F1-score	79.52 ± 0.43	87.79 ± 0.50	78.94 ± 0.55	82.08 ± 0.49
Location 3	Spec.	95.52 ± 0.46	91.69 ± 0.36	88.42 ± 0.37	91.88 ± 0.40
	Prec.	84.64 ± 0.44	89.51 ± 0.52	75.74 ± 0.41	83.30 ± 0.46
	AUC	92.73 ± 0.48	95.76 ± 0.43	84.54 ± 0.57	91.01 ± 0.49
	F1-score	78.59 ± 0.41	88.64 ± 0.29	80.83 ± 0.50	82.69 ± 0.40
Ours	Spec.	97.56 ± 0.57	93.38 ± 0.38	88.86 ± 0.56	**93.27 ± 0.50**
	Prec.	91.59 ± 0.59	91.37 ± 0.55	78.74 ± 0.58	**87.22 ± 0.57**
	AUC	94.63 ± 0.58	96.87 ± 0.62	86.95 ± 0.47	92.82 ± 0.56
	F1-score	81.35 ± 0.42	91.36 ± 0.38	83.13 ± 0.58	**85.44 ± 0.46**

Table 8
Comparison with the SOTA methods by accuracy (%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Methods	AD	CN	MCI	Avg.	Ttrain	Ttest
Image only	70.64 ± 0.36	70.51 ± 0.42	69.31 ± 0.36	70.15 ± 0.38	8.2 h	7.3 ms
Tabular data only	72.34 ± 0.42	73.45 ± 0.44	71.76 ± 0.38	72.52 ± 0.41	0.08 h	0.5 ms
Concat-1 [29]	73.86 ± 0.36	74.62 ± 0.41	69.63 ± 0.48	72.70 ± 0.42	9.2 h	8.9 ms
Concat-2 [30]	81.52 ± 0.46	84.88 ± 0.44	79.25 ± 0.48	81.88 ± 0.46	9.5 h	9.1 ms
DAFT [42]	83.88 ± 0.50	87.49 ± 0.47	82.46 ± 0.40	84.61 ± 0.46	11.2 h	13.5 ms
Inter [43]	82.89 ± 0.53	85.52 ± 0.46	80.49 ± 0.29	82.97 ± 0.43	10.9 h	12.7 ms
Ours	86.84 ± 0.49	87.69 ± 0.57	84.22 ± 0.57	**86.34 ± 0.54**	10.4 h	11.3 ms

Table 9
The detailed ablation results of different location in terms of Specificity (Spec.), precision (Prec.), AUC and F1-score(%). Note that '±std' represents the empirical standard deviation across the 5 folds. The best-found scores are indicated in bold.

Method	Met.	AD	CN	MCI	Avg.
Image only	Spec.	70.88 ± 0.56	71.26 ± 0.51	70.67 ± 0.34	70.94 ± 0.47
	Prec.	70.22 ± 0.51	71.84 ± 0.51	63.46 ± 0.29	68.51 ± 0.44
	AUC	70.85 ± 0.49	71.73 ± 0.44	69.96 ± 0.24	70.85 ± 0.39
	F1-score	69.86 ± 0.41	70.25 ± 0.47	69.28 ± 0.29	69.80 ± 0.39
Tabular data only	Spec.	73.98 ± 0.31	73.67 ± 0.63	72.86 ± 0.48	73.50 ± 0.47
	Prec.	68.92 ± 0.29	72.16 ± 0.38	70.44 ± 0.48	70.51 ± 0.38
	AUC	74.42 ± 0.39	83.17 ± 0.61	68.80 ± 0.47	75.46 ± 0.49
	F1-score	71.16 ± 0.36	72.02 ± 0.39	70.82 ± 0.54	71.33 ± 0.43
Concat-1 [29]	Spec.	90.76 ± 0.48	91.78 ± 0.46	76.64 ± 0.54	86.39 ± 0.49
	Prec.	71.69 ± 0.42	74.52 ± 0.33	76.72 ± 0.45	74.31 ± 0.40
	AUC	79.34 ± 0.52	88.14 ± 0.33	71.15 ± 0.85	79.54 ± 0.57
	F1-score	67.66 ± 0.38	78.17 ± 0.66	73.78 ± 0.50	73.20 ± 0.51
Concat-2 [30]	Spec.	97.74 ± 0.52	84.32 ± 0.35	89.48 ± 0.39	90.51 ± 0.42
	Prec.	90.75 ± 0.56	82.47 ± 0.39	76.66 ± 0.43	83.29 ± 0.46
	AUC	91.34 ± 0.36	94.29 ± 0.23	89.27 ± 0.52	91.63 ± 0.37
	F1-score	76.67 ± 0.29	86.77 ± 0.46	78.58 ± 0.40	80.67 ± 0.38
DAFT [42]	Spec.	97.38 ± 0.38	94.30 ± 0.31	86.17 ± 0.40	92.62 ± 0.36
	Prec.	90.59 ± 0.40	92.47 ± 0.48	73.83 ± 0.46	85.63 ± 0.45
	AUC	92.64 ± 0.51	94.72 ± 0.45	83.01 ± 0.23	90.12 ± 0.40
	F1-score	79.18 ± 0.46	90.24 ± 0.43	82.95 ± 0.17	84.12 ± 0.35
Inter [43]	Spec.	98.62 ± 0.40	86.24 ± 0.30	88.42 ± 0.40	91.09 ± 0.37
	Prec.	94.54 ± 0.39	84.15 ± 0.40	75.85 ± 0.45	84.86 ± 0.41
	AUC	88.72 ± 0.33	93.75 ± 0.34	89.56 ± 0.37	90.68 ± 0.35
	F1-score	79.89 ± 0.54	87.37 ± 0.35	79.80 ± 0.41	82.35 ± 0.49
Ours	Spec.	97.56 ± 0.57	93.38 ± 0.38	88.86 ± 0.56	**93.27 ± 0.50**
	Prec.	91.59 ± 0.59	91.37 ± 0.55	78.74 ± 0.58	**87.22 ± 0.57**
	AUC	94.63 ± 0.58	96.87 ± 0.62	86.95 ± 0.47	**92.82 ± 0.56**
	F1-score	81.35 ± 0.42	91.36 ± 0.38	83.13 ± 0.58	**85.44 ± 0.46**

the concatenation-based fusion methods contain global complementary information of different modalities, but the feature from each modality are extracted separately, which cause the sub-optimal solution. Though our method also concatenates the feature maps, it employs attention mechanism-based local fusion to learn highly discriminative features which contains global and local information. Thus, our method achieves the best result among all the comparison methods. Integrating 3D MRI features with tabular data by DAFT, as done by [42], can improve performance compared with the Concatenation networks and network in [43], but appears worse than our approach. DAFT only used tabular data to fuse image global features without taking into account the local fine-grained fusion. The network in [43] only performs better than the Concatenation multi-modal fusion networks, which means integrating tabular data with both low and high-level descriptors of the image can severely deteriorate performance. Even though DAFT [42]

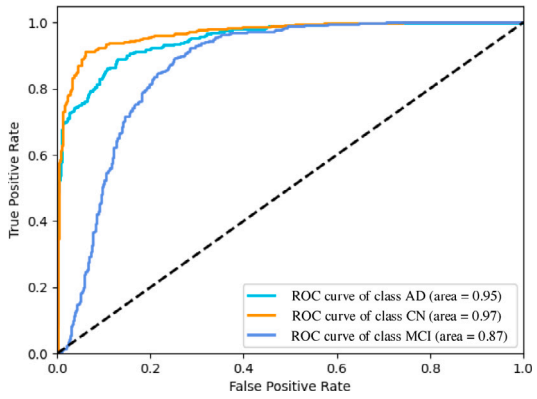


Fig. 3. ROC curves for each label in the AD classification task.

and Interactive [43] introduce a new method to fuse the features, it only learns the global complementary information.

Our proposed method is the only approach that excels at integrating 3D MRI and clinical tabular data for classification task by outperforming competing methods by a large margin. The main reason is that by employing an attention mechanism-based fusion method to guarantee representations from different modalities to gain local complementary information, which the other multi-modal methods lack. When comparing the performance of other classification (as shown in Table 9), our proposed method achieves the best performance metrics. Specifically, our proposed method improves more than 1% of F1-score on ADNI datasets and 2.76% AUC compared to the previous SOTA method (DAFT). We also plot the ROC curves in Fig. 3, from which one can see that the curves for all categories except the MCI category have similar area sizes (around 90.0%) under the ROC curve. For the MCI category, its ROC curve is obviously lower than other categories, its area under the ROC curve is less than 90.0%. Overall, these results show that our method can achieve a promising performance for AD classification. Finally, Tables 8 and 9 show the training and testing time of models, which shows that our model has advantages in model performance and real-time performance.

4.3. Validation on OASIS-1 datasets

After training and testing in ADNI datasets, we also utilized another OASIS-1 datasets which aimed at the research on Alzheimer's disease to validate our proposed method. The detailed description of OASIS-1 datasets has shown in Section 3.1. Table 10 shows the average metric values, our proposed method has better performance than other models with different settings, which are consistent with results of ablation experiments in Section 4.2.1 validated in ADNI datasets. It means the effectiveness of each part of our proposed method and the learning both the global and local fusion information can further improve classification performance.

Table 11 reports the detailed comparison results validated on OASIS-1 datasets. Our proposed method achieves better performance on most metrics. Specifically, our proposed method surpassed the AUC result by DAFT by 0.0%, and also outperformed INTER by 0.0%. All of those mean our proposed model is still effective on other datasets.

4.4. Statistical analysis

We adopted the Student t-test to determine whether the performance gain achieved by the proposed MMGLF framework over the competing methods is statistically significant. We assumed that the Acc/Spec/Prec/AUC/F1-score values of MMGLF and each competing method are random variables X_1 and X_2, respectively, each following a Gaussian distribution, i.e., $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$. The difference between X_1 and X_2 is defined as $\Delta = X_1 - X_2$. The hypotheses to be tested are H_0: $\mu_\Delta \leq 0$ versus H_1: $\mu_\Delta > 0$. To enhance the rigor of this statistical testing and control the overall false positive rate, we applied the Bonferroni correlation to adjust the significance level. To achieve this, we divided the original level of significance ($\alpha = 0.05$) by the total number of tests performed ($m = 6 \times 5$), which yielded a new significance threshold of $\alpha' = \alpha/m = 0.00167$. Our analysis, as presented in Table 12, indicates that for the vast majority of comparisons with competing methods, the calculated p=values were below the adjusted significance level of $\alpha' = 0.00167$. As a result, we were able to reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1), indicating that the MMGLF framework performed significantly better than the other competing methods in terms of five evaluation metrics.

Performing the same procedure on the ablation experiments, as shown in Table 13, the vast majority of comparisons with the ablation experiments, the calculated p-values were below the adjusted significance level of $\alpha' = \alpha/m = 0.00111$, where $\alpha = 0.05$, $m = 9 \times 5$. As a result, we were able to reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1), indicating that the MMGLF framework performed significantly better than the ablation experiments in terms of five evaluation metrics.

Table 10
The detailed classification results (%) of ablation modules validating on OASIS-1 datasets. The best-found scores are indicated in bold.

Model	Acc.	Spec.	Prec.	AUC	F1-score
I	84.27	82.07	89.87	91.05	83.85
II	84.34	82.54	90.09	92.43	84.02
III	84.29	81.33	90.70	91.63	83.33
1	83.22	80.33	88.66	90.66	80.34
2	85.43	83.32	90.89	**93.79**	81.99
3	83.91	81.84	89.68	91.25	82.06
Location 1	84.79	83.95	91.95	92.59	85.45
Location 2	84.26	82.33	91.26	91.93	82.94
Location 3	83.61	80.94	89.53	90.75	81.97
Ours	**86.77**	**88.67**	**92.42**	92.93	**87.33**

Table 11
The detailed classification results (%) of comparative methods validating on OASIS-1 datasets. The best-found scores are indicated in bold.

Model	Acc.	Spec.	Prec.	AUC	F1-score
Image only	69.70	70.09	72.25	73.05	71.52
Tabular data only	70.36	70.27	75.63	74.46	72.26
Concat-1 [29]	71.59	71.33	76.52	74.82	59.01
Concat-2 [30]	76.70	82.33	89.57	91.29	84.00
DAFT [42]	84.61	81.00	89.07	91.79	84.33
Inter [43]	81.25	75.67	82.31	92.03	**90.67**
Ours	**86.77**	**88.67**	**92.42**	**92.93**	87.33

4.5. Visualization

To evaluate the regions in the 3D MRI that the model considered essential for AD classification, we also generate heatmaps showing the location that the network paid attention to in the classification. It is well known that the Hippocampus and Amygdala to be strongly affected by AD [49–51].

Firstly, to evaluate the effectiveness of global and local information fusion of our proposed method, we visualize three cases for the classification task, as shown in Fig. ??. From Fig. ??, with the single-modal-based model, there are many regions outside the AD-related regions that appear to be mistakenly considered to be important. For other multi-modal-based model, there are fewer regions appearing to be mistakenly considered. With our MMGLF-model, attention regions inside and around the AD-related regions are most relevant in classifying AD. Besides, it can be observed that the discriminative regions within AD subject are more distinct than that of MCI subject. Considering the fact that structural changes caused by AD are relatively easier to be

Table 12
The p-values of the Student t-test performance compared with competing methods on AD classification task. The significance level is set to $\alpha' = 0.00167$ after Bonferroni correlation.

Methods	p-value (Acc.)	p-value (Spec.)	p-value (Prec.)	p-value (AUC)	p-value (F1-score)
Ours vs. Image only	6.01E−14	3.58E−15	4.76E−14	2.62E−14	1.93E−13
Ours vs. Tabular data only	5.63E−13	8.14E−15	1.63E−15	3.78E−15	1.11E−13
Ours vs. Concat-1[29]	8.40E−14	1.38E−9	1.89E−12	2.23E−11	1.39E−11
Ours vs. Concat-2[30]	2.49E−8	9.96E−9	2.44E−8	1.98E−6	3.40E−18
Ours vs. DAFT [42]	7.44E−5	3.39−3	2.27E−5	5.37E−8	2.64E−4
Ours vs. Inter [43]	2.51E−7	2.21E−6	1.612.27E−7	2.58E−7	2.89E−7

Table 13
The p-values of the Student t-test performance compared with ablation experiments on AD classification task. The significance level is set to $\alpha' = 0.00111$ after Bonferroni correlation.

Methods	p-value (Acc.)	p-value (Spec.)	p-value (Prec.)	p-value (AUC)	p-value (F1-score)
Ours vs. Model I	1.75E−4	1.62E−3	6.64E−9	2.91E−4	1.01E−3
Ours vs. Model II	9.60E−5	5.56E−5	5.71E−9	1.27E−2	4.30E−5
Ours vs. Model III	4.66E−4	6.38E−4	2.64E−6	3.27E−5	5.11E−4
Ours vs. Model 1	7.79E−7	2.61E−3	1.54E−6	6.07E−11	6.35E−5
Ours vs. Model 2	7.96E−5	1.25E−7	1.25E−6	4.65E−6	3.54E−9
Ours vs. Model 3	2.53E−7	2.29E−5	1.61E−9	4.90E−5	5.62E−4
Ours vs. Location 1	2.76E−7	5.20E−5	5.86E−6	8.57E−1	5.69E−7
Ours vs. Location 2	1.08E−9	6.01E−6	9.62E−11	6.34E−3	5.14E−7
Ours vs. Location 3	4.21E−9	1.12E−5	4.07E−8	2.23E−5	1.61E−6

detected than MCI, these results suggest that the learned attention maps of the proposed method are reasonable.

Secondly, we demonstrate several CAM-based images obtained from both 3D MRI and clinical tabular data to show the effectiveness of attention-based local fusion module. We compare the visualization of our proposed model with only the global fusion module model (as model 1 in the ablation study). The difference between our proposed model and only with global module model is that our proposed model used an attention mechanism-based local fusion module to realize fusion of local region features of 3D MRI and tabular features. We visualize situations of these two methods: (1) proposed method only with global fusion module(both methods classified correctly), (2) proposed method with both global and local fusion module(both methods misclassified). Fig. 5 shows the specific visualization results. For these two methods different situations, we can see that for both 3D MRI and clinical tabular data, the important areas focused by our proposed model with both global and local fusion module are more compact and centered on the areas strongly related to AD. Even when both methods misclassified the sample, our MMGLF can still focus on the partially correct regions.

The above visualization results verify the effectiveness of the attention mechanism-based local fusion module. As can be seen from Fig. 4 and Fig. 5, our proposed multi-modal learning method focuses on the areas strongly related to AD, which verifies that the model has been well trained, and the global–local fusion information has been well learned (see Fig. 4).

5. Limitations and future work

Although the proposed MMGLF model has obtained good performance in AD classification, there are still some limitation that need to be addressed in the future.

First, the feature extractor network is trained scratch in the current work. It is interesting to pretrain existing 3D CNNs on the large-scale 3D medical image datasets and fine-tune them on the ADNI datasets to further improve the classification performance. Second, only MRI and tabular data are considered in our current work, while PET may also play a role in AD prediction. It is interesting to incorporate PET to improve the classification results. Besides, our current model is mainly trained on one domain and transferred to other domain. As future work, one can study how to leverage multi-source domain learning to incorporate more diverse training sets into whole learning process to further enhance the robustness and transferability. Furthermore, missing modality data is common in clinical practice and it can result in the collapse of most previous methods relying on complete modality data. Thus, it is desired to design models that can learn multi-modal information for disease classification even if some modalities are missing.

6. Conclusion

In this paper, we proposed a novel multi-modal global–local neural network for Alzheimer's disease classification. Our proposed method can realize global–local information fusion of 3D MRI and clinical tabular data. Specifically, to learn the global fusion information, we adopted concatenation method to fuse global features of two modalities. Furthermore, to make the local information of 3D MRI and the clinical tabular features have a better fine-grained fusion, we designed an attention-based fusion network to force the network to learn the discriminate features. Then, we concatenated the extracted different modalities' features to obtain the final fusion features for classification. Through these proposed modules, the network effectively extracts and fuses features from 3D MRI and clinical tabular data. The effectiveness of our proposed multi-modal framework is validated on ADNI datasets and OASIS-1 datasets. Experiments on those two datasets demonstrate that our method could significantly improve performance compared with previous deep learning approaches that combine image and tabular data. Our framework can be extended to other multi-modal medical data.

CRediT authorship contribution statement

Nana Jia: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology. **Tong Jia:** Writing – review & editing, Methodology. **Li Zhao:** Writing – review & editing. **Bowen Ma:** Writing – review & editing. **Zheyi Zhu:** Writing – review & editing.

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

The data that has been used is confidential.

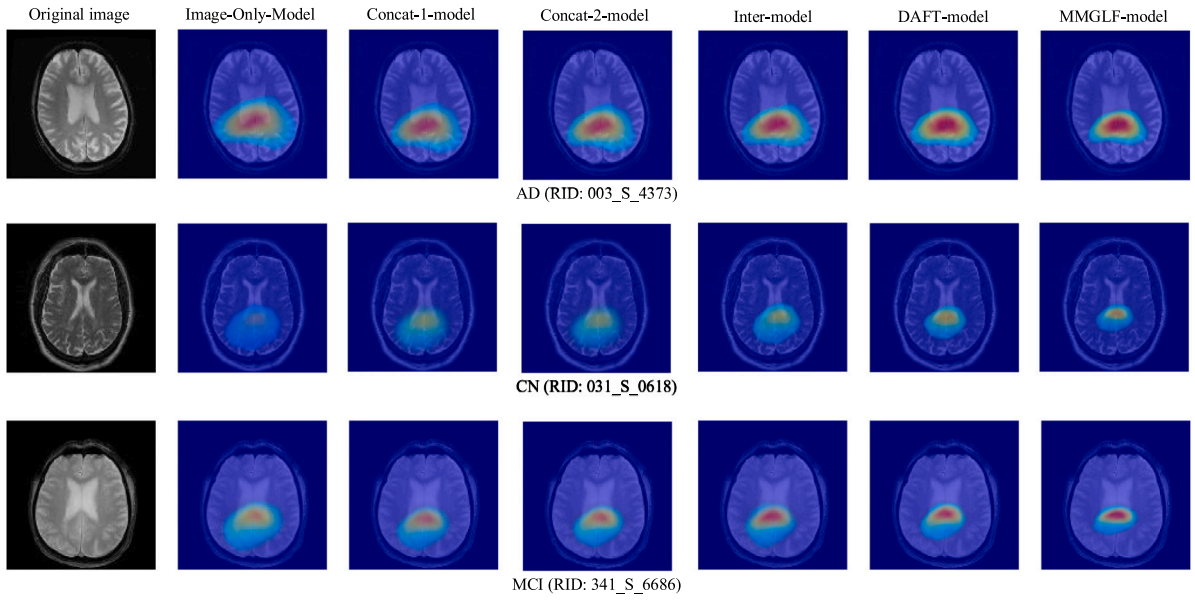


Fig. 4. Visualization results on evaluating the effectiveness of our proposed multi-modal network.

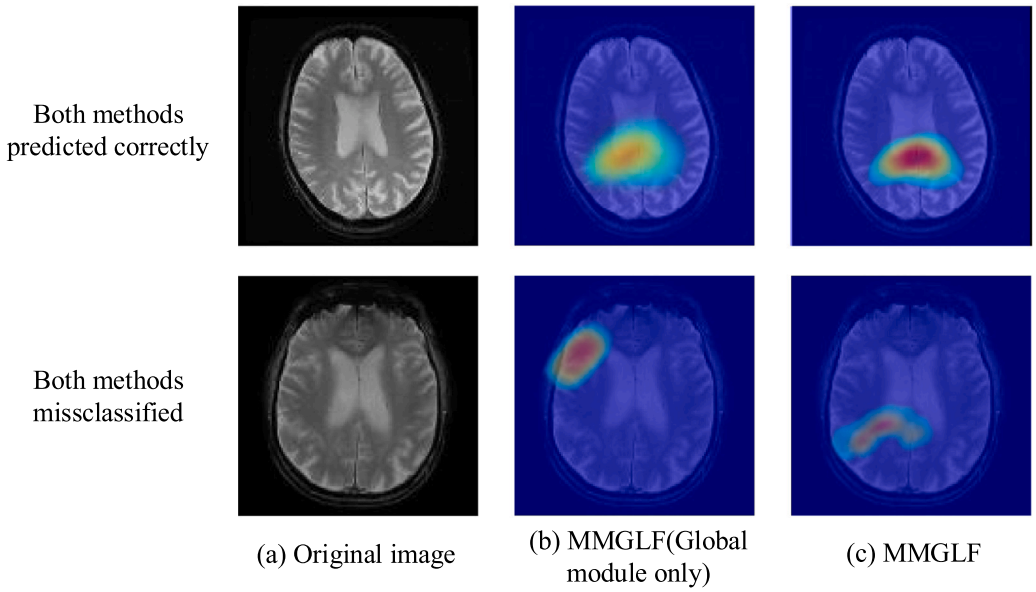


Fig. 5. Visualization results on evaluating the effectiveness of the local module based on attention mechanism under two different classification situations.

References

[1] Alzheimer's Association, et al., 2017 Alzheimer's disease facts and figures, Alzheimer's & Dementia 13 (4) (2017) 325–373.

[2] Clifford R Jack Jr., Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging: An Off. J. Int. Soc. Magn. Reson. Med. 27 (4) (2008) 685–691.

[3] Yuanpeng Zhang, Shuihua Wang, Kaijian Xia, Yizhang Jiang, Pengjiang Qian, Alzheimer's Disease Neuroimaging Initiative, et al., Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion, Inf. Fusion 66 (2021) 170–183.

[4] Soheil Esmaeilzadeh, Dimitrios Ioannis Belivanis, Kilian M Pohl, Ehsan Adeli, End-to-end Alzheimer's disease diagnosis and biomarker identification, in: Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9, Springer, 2018, pp. 337–345.

[5] Mr Amir Ebrahimighahnavieh, Suhuai Luo, Raymond Chiong, Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review, Comput. Methods Programs Biomed. 187 (2020) 105242.

[6] Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, Olivier Colliot, et al., Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation, Med. Image Anal. 63 (2020) 101694.

[7] Hui Cui, Yiyue Xu, Wanlong Li, Linlin Wang, Henry Duh, Collaborative learning of cross-channel clinical attention for radiotherapy-related esophageal fistula prediction from CT, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23, Springer, 2020, pp. 212–220.

[8] Phu-Hung Dinh, Combining Gabor energy with equilibrium optimizer algorithm for multi-modality medical image fusion, Biomed. Signal Process. Control 68 (2021) 102696.

[9] Phu-Hung Dinh, Combining spectral total variation with dynamic threshold neural P systems for medical image fusion, Biomed. Signal Process. Control 80 (2023) 104343.

[10] Fangyu Liu, Shizhong Yuan, Weimin Li, Qun Xu, Bin Sheng, Patch-based deep multi-modal learning framework for Alzheimer's disease diagnosis using multi-view neuroimaging, Biomed. Signal Process. Control 80 (2023) 104400.

[11] V.P. Subramanyam Rallabandi, Krishnamoorthy Seetharaman, Deep learning-based classification of healthy aging controls, mild cognitive impairment and

Alzheimer's disease using fusion of MRI-PET imaging, Biomed. Signal Process. Control 80 (2023) 104312.
[12] R. Divya, R. Shantha Selva Kumari, SUVR quantification using attention-based 3D CNN with longitudinal florbetapir PET images in Alzheimer's disease, Biomed. Signal Process. Control 86 (2023) 105254.
[13] Juan E. Arco, Javier Ramírez, Juan M. Górriz, María Ruz, Data fusion based on searchlight analysis for the prediction of Alzheimer's disease, Expert Syst. Appl. 185 (2021) 115549.
[14] Grazia Daniela Femminella, Denise Harold, James Scott, Julie Williams, Paul Edison, Alzheimer's Disease Neuroimaging Initiative, et al., The differential influence of immune, endocytotic, and lipid metabolism genes on amyloid deposition and neurodegeneration in subjects at risk of Alzheimer's disease, J. Alzheimer's Disease 79 (1) (2021) 127–139.
[15] Giuseppe Tosto, Christiane Reitz, Genome-wide association studies in Alzheimer's disease: a review, Current neurology and neuroscience reports 13 (2013) 1–7.
[16] Denise Harold, Richard Abraham, Paul Hollingworth, Rebecca Sims, Amy Gerrish, Marian L Hamshere, Jaspreet Singh Pahwa, Valentina Moskvina, Kimberley Dowzell, Amy Williams, et al., Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease, Nature Genet. 41 (10) (2009) 1088–1093.
[17] Xinchun Cui, Ruyi Xiao, Xiaoli Liu, Hong Qiao, Xiangwei Zheng, Yiquan Zhang, Jianzong Du, Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis, Chemometr. Intell. Lab. Syst. 215 (2021) 104316.
[18] Ruyi Xiao, Xinchun Cui, Hong Qiao, Xiangwei Zheng, Yiquan Zhang, Chenghui Zhang, Xiaoli Liu, Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net, Biomed. Signal Process. Control 66 (2021) 102362.
[19] Xiaoke Hao, Yongjin Bao, Yingchun Guo, Ming Yu, Daoqiang Zhang, Shannon L Risacher, Andrew J Saykin, Xiaohui Yao, Li Shen, Alzheimer's Disease Neuroimaging Initiative, et al., Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease, Med. Image Anal. 60 (2020) 101625.
[20] Saima Rathore, Mohamad Habes, Muhammad Aksam Iftikhar, Amanda Shacklett, Christos Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, NeuroImage 155 (2017) 530–548.
[21] Jieping Ye, Teresa Wu, Jing Li, Kewei Chen, Machine learning approaches for the neuroimaging study of Alzheimer's disease, Computer 44 (4) (2011) 99–101.
[22] Jing Sui, Tülay Adali, Qingbao Yu, Jiayu Chen, Vince D Calhoun, A review of multivariate methods for multimodal fusion of brain imaging data, J. Neurosci. Methods 204 (1) (2012) 68–81.
[23] G Chetelat, B Desgranges, V De La Sayette, F Viader, F Eustache, J-C Baron, Mild cognitive impairment: can FDG-PET predict who is to rapidly convert to Alzheimer's disease? Neurology 60 (8) (2003) 1374–1377.
[24] Ann D. Cohen, William E. Klunk, Early detection of alzheimer's disease using PiB and FDG PET, Neurobiol. Dis. 72 (2014) 117–122.
[25] Norman L Foster, Judith L Heidebrink, Christopher M Clark, William J Jagust, Steven E Arnold, Nancy R Barbas, Charles S DeCarli, R Scott Turner, Robert A Koeppe, Roger Higdon, et al., FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease, Brain 130 (10) (2007) 2616–2635.
[26] Biao Jie, Daoqiang Zhang, Bo Cheng, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, Manifold regularized multitask feature learning for multimodality disease classification, Hum. Brain Mapping 36 (2) (2015) 489–507.
[27] Giovanni B Frisoni, Nick C Fox, Clifford R Jack Jr., Philip Scheltens, Paul M Thompson, The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 67–77.
[28] Ting Xiao, Han Zheng, Xiaoning Wang, Xinghan Chen, Jianbo Chang, Jianhua Yao, Hong Shang, Peng Liu, Intracerebral haemorrhage growth prediction based on displacement vector field and clinical metadata, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 741–751.
[29] Philipp Kopper, Sebastian Pölsterl, Christian Wachinger, Bernd Bischl, Andreas Bender, David Rügamer, Semi-structured deep piecewise exponential models, in: Survival Prediction-Algorithms, Challenges and Applications, PMLR, 2021, pp. 40–53.
[30] Mingxia Liu, Jun Zhang, Ehsan Adeli, Dinggang Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.
[31] Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A Gutman, Jill S Barnholtz-Sloan, José E Velázquez Vega, Daniel J Brat, Lee AD Cooper, Predicting cancer outcomes from histology and genomics using convolutional networks, Proc. Natl. Acad. Sci. 115 (13) (2018) E2970–E2979.
[32] Sebastian Pölsterl, Ignacio Sarasua, Benjamín Gutiérrez-Becker, Christian Wachinger, A wide and deep neural network for survival analysis from anatomical shape and tabular clinical data, in: Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019, WÜRzburg, Germany, September 16–20, 2019, Proceedings, Part I, Springer, 2020, pp. 453–464.
[33] Hongming Li, Mohamad Habes, David A Wolk, Yong Fan, Alzheimer's Disease Neuroimaging Initiative, et al., A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data, Alzheimer's & Dementia 15 (8) (2019) 1059–1070.
[34] Jie Hao, Sai Chandra Kosaraju, Nelson Zange Tsaku, Dae Hyun Song, Mingon Kang, PAGE-Net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data, in: Pacific Symposium on Biocomputing 2020, World Scientific, 2019, pp. 355–366.
[35] Shaker El-Sappagh, Tamer Abuhmed, SM Riazul Islam, Kyung Sup Kwak, Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data, Neurocomputing 412 (2020) 197–215.
[36] Shuai Li, Haolei Shi, Dong Sui, Aimin Hao, Hong Qin, A novel pathological images and genomic data fusion framework for breast cancer survival prediction, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 1384–1387.
[37] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Lio, Nicola Toschi, Alzheimer's Disease Neuroimaging Initiative, et al., A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, Neuroimage 189 (2019) 276–287.
[38] Yulu Guan, Hui Cui, Yiyue Xu, Qiangguo Jin, Tian Feng, Huawei Tu, Ping Xuan, Wanlong Li, Linlin Wang, Been-Lirn Duh, Predicting esophageal fistula risks using a multimodal self-attention network, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 721–730.
[39] Hui Cui, Ping Xuan, Qiangguo Jin, Mingjun Ding, Butuo Li, Bing Zou, Yiyue Xu, Bingjie Fan, Wanlong Li, Jinming Yu, et al., Co-graph attention reasoning based imaging and clinical features integration for lymph node metastasis prediction, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 657–666.
[40] Hang Li, Fan Yang, Xiaohan Xing, Yu Zhao, Jun Zhang, Yueping Liu, Mengxue Han, Junzhou Huang, Liansheng Wang, Jianhua Yao, Multi-modal multi-instance learning using weakly correlated histopathological images and tabular clinical information, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24, Springer, 2021, pp. 529–539.
[41] Nathaniel Braman, Jacob WH Gordon, Emery T Goossens, Caleb Willis, Martin C Stumpe, Jagadish Venkataraman, Deep orthogonal fusion: multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 667–677.
[42] Sebastian Pölsterl, Tom Nuno Wolf, Christian Wachinger, Combining 3D image and tabular data via the dynamic affine feature map transform, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24, Springer, 2021, pp. 688–698.
[43] Hongyi Duanmu, Pauline Boning Huang, Srinidhi Brahmavar, Stephanie Lin, Thomas Ren, Jun Kong, Fusheng Wang, Tim Q Duong, Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using deep learning with integrative imaging, molecular and demographic data, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23, Springer, 2020, pp. 242–252.
[44] M. Paul Murphy, Harry LeVine III, Alzheimer's disease and the amyloid-β peptide, J. Alzheimer's disease 19 (1) (2010) 311–323.
[45] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS '17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 6000–6010.
[47] Coşku Öksüz, Oğuzhan Urhan, Mehmet Kemal Güllü, An integrated convolutional neural network with attention guidance for improved performance of medical image classification, Neural Comput. Appl. (2023) 1–33.
[48] Bernard M Cobbinah, Christian Sorg, Qinli Yang, Arvid Ternblom, Changgang Zheng, Wei Han, Liwei Che, Junming Shao, Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder, Med. Image Anal. 82 (2022) 102585.
[49] Endel Tulving, Hans J. Markowitsch, Episodic and declarative memory: role of the hippocampus, Hippocampus 8 (3) (1998) 198–204.
[50] Benno Roozendaal, Bruce S. McEwen, Sumantra Chattarji, Stress, memory and the amygdala, Nat. Rev. Neurosci. 10 (6) (2009) 423–433.
[51] M.W. Hopper, F.S. Vogel, The limbic system in Alzheimer's disease. a neuropathologic investigation., Amer. J. Pathol. 85 (1) (1976) 1.