Journal Pre-proof

Multi-modal Neuroimaging Feature Fusion for Diagnosis of Alzheimer's Disease

Tao Zhang, Mingyang Shi

PII: S0165-0270(20)30218-1

DOI: https://doi.org/10.1016/j.jneumeth.2020.108795

Reference: NSM 108795

To appear in: *Journal of Neuroscience Methods*

Received Date: 4 February 2020

Revised Date: 19 May 2020

Accepted Date: 19 May 2020

Please cite this article as: Zhang T, Shi M, Multi-modal Neuroimaging Feature Fusion for Diagnosis of Alzheimer's Disease, *Journal of Neuroscience Methods* (2020), doi: https://doi.org/10.1016/j.jneumeth.2020.108795

Multi-modal Neuroimaging Feature Fusion for Diagnosis of Alzheimer's Disease

Tao Zhang, Mingyang Shi*

School of Electronic and Information Engineering, Tianjin University,300387, Tianjin, China

*Correspondence: Mingyang Shi shimingyang@tju.edu.cn

Highlights

- Multi-modal classification can achieve better performance by fusing different information
- Expanding the early and the late fusion into a hierarchical fusion to effectively exploit low-level and high-level features
- The attention complementary strategy is introduced to extract the synergy between multi-modal images
- The attention strategy is introduced in the feature extraction task to suppresses the irrelevant information
- Experiments on ADNI show the effectiveness of proposed method and its superiority

Abstract

Background: Compared with single-modal neuroimages classification of AD, multi-modal classification can achieve better performance by fusing different information. Exploring synergy among various multi-modal neuroimages is contributed to identifying the pathological process of neurological disorders. However, it is still problematic to effectively exploit multi-modal information since the lack of an effective fusion method.

New method: In this paper, we propose a deep multi-modal fusion network based on the attention mechanism, which can selectively extract features from MRI and PET branches and suppress irrelevant information. In the attention model, the fusion ratio of each modality is assigned automatically according to the importance of the data. A hierarchical fusion method is adopted to ensure the effectiveness of Multi-modal Fusion.

Results: Evaluating the model on the ADNI dataset, the experimental results show that it outperforms the state-of-the-art methods. In particular, the final classification results of the NC/AD, SMCI/PMCI and Four-Class are 95.21%, 89.79%, and 86.15%, respectively.

Comparison with existing methods: Different from the early fusion and the late fusion, the hierarchical fusion method contributes to learning the synergy between the multi-modal data. Compared with some other prominent algorithms, the attention model enables our network to focus on the regions of interest and effectively fuse the multi-modal data.

Conclusion: Benefit from the hierarchical structure with attention model, the proposed network is capable of exploiting low-level and high-level features extracted from the multi-modal data and improving the accuracy of AD diagnosis .Results show its promising performance.

Keyword: Alzheimer's Disease; Deep Learning; Classification; Multi-modal Fusion; Attention Model;

Introduction

With the growth of population, the population aging has become a problem that cannot be ignored in social development. The aging population will bring a series of disease. Alzheimer's Disease (AD) is a potential onset neurodegenerative disease primarily characterized by progressive episodic memory loss and accompanied by several kinds of cognitive and functional impairments [23]. Mild Cognitive Impairment (MCI) is the transition period between Normal Control (NC) and possible AD. 44% of MCI patients may eventually convert to AD within a few years [14]. The diagnosis of AD mainly relies on clinical examinations and psychometric assessments, and there is no definitive and effective treatment for patients with advanced AD. Timely medical intervention could slow down the deterioration process, so it is of great significance to discover the irreversible change in the brain of patients before the onset of clinical symptoms.

With the rapid development of neuroimaging technology, neuroimaging has become the most intuitive and reliable method for the auxiliary diagnosis of AD. In neuroimaging methods, Magnetic Resonance Imaging (MRI) has high resolution for soft tissues of the brain, which can clearly distinguish the gray and white matter of the brain and reflect the degree of brain atrophy [9][10]. Positron Emission computed Tomography (PET) is also a common neuroimaging technique for diagnosis of AD, it can show the distribution of lesions and the rate of glucose metabolism by imaging agents [12]. Diffuse Tensor Imaging (DTI) can reflect the structural properties of white matter cellulose in the brain, so it is often applied to analyze water diffusion at the microstructural level of the brain for determining the abnormal diffusion pattern of AD [13]. Since the single-modal neuroimage only contains a part of information related to the brain atrophy, it may be insufficient

for the MCI conversion prediction. However, the multi-modal neuroimage can provide more complementary information, which could be fused to learn the synergy between different neuroimages.

Multi-modal fusion is one of the frontiers in multi-modal machine learning with the early, the late and the mixed fusion approaches. Multimodal learning is widely applied in image classification [5][18] and image registration [6]. Interests in multi-modal fusion arise from two main benefits. First, more robust predictions could be achieved from multiple modalities that observe the same phenomenon [8]. Second, the complementary information could be extracted from multiple modalities to improve the accuracy of classification results [11]. Liu et al. [15] used Sparse Auto Encoder (SAE) to obtain high-level features, and a zero-masking strategy is applied to extract the synergy between MRI and PET images. Suk et al. [16] used multi-modal Deep Boltzmann Machines (DBM) representation to perform AD classification from MRI and PET images. Zhang et al. [17] used a simple but effective Multiple Kernel Learning (MKL) method to combine three different biomarkers for classification. Zhou et al. [18] proposed a stage-wise deep feature learning and fusion framework. Each stage of the network learns feature representations for independent modality or different combinations of modalities. Due to registration deviation and noise interference, Regions of Interest (ROI) based feature vectors extracted by the above methods depend largely on image preprocessing, so ROI-based feature engineering requires knowledge from domain experts. Only a very small amount of training data can be used to learn discriminative patterns in high-dimensional feature spaces. Besides, the traditional multi-modal fusion algorithms ignore not only many available and extractable features in images, but also the differences in brain volume levels of different patients.

Deep learning methods, especially Convolutional Neural Network (CNN), outperform existing machine learning methods in image classification task [29]. In CNN, the original images are used directly as the input, and then learning is automatically conducted with the training data. In most recent studies, CNNs were used to extract the features of MRI and PET respectively. Cheng et al. [19] used image patches to transform the local images into high-level features from the 3D original MRI and PET images and combine their results to run 2D CNN. The early fusion focus on the combination of low-level features, but the original features will be destroyed after fusion. Liu et al.[20] used the correlation analysis to compute the consistency of two CNNs outputs. The late fusion focuses on the analysis of results but ignores the synergy between the low-level features. However, neither the early fusion nor the late fusion contributes much to the synergy between the multi-modal data. In other words, the current multi-modal fusion method cannot take full advantage of the complementarity between different neuroimages.

To solve the problems of traditional and current algorithms, this paper proposes a novel Deep Multi-modal Fusion Network (DMFNet), which has three branches corresponding to the data stream of MRI, PET and the merge information. More precisely, MRI and PET images are input into two independent branches for feature extraction. At each stage of DMFNet, an attention-based model is applied to fuse the information of the two modalities. The main contributions of this paper are as follows. 1) The attention complementary strategy is introduced in multi-modal fusion task to extract the synergy between multi-modal images. The fusion ratio of each modality is assigned automatically. 2) The attention strategy is introduced in the feature extraction task, which models the importance of each feature channel to enhance or suppress different channels for different classification tasks. 3) Expanding the early and the late fusion into a hierarchical fusion to effectively exploit low-level and high-level features and improve the accuracy of auxiliary diagnosis.

Material and Methods

Data acquire and Image preprocessing

Data used in the preparation of this paper are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological marker, clinical and psychometric assessment can be combined to measure the progressions of MCI and early AD.

ADNI assesses the process of NC to Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI) or AD through clinical, imaging, genetic and biospecimen biomarkers. Individuals with EMCI are still diagnosed with MCI, but have less memory impairment than MCI subjects recruited during ADNI1. For a comprehensive validation of the proposed method, we select 500 ADNI subjects including 163 NC, 113 EMCI, 105 LMCI and 119 AD subjects with both T1-weighted MRI scans and FDG-PET images. MCI may convert to AD in a few years. For further identifying the progress of MCI conversion, MCI subjects are divided into 153 Stable MCI (SMCI) and 65 Progressive MCI (PMCI) subjects according to the standard of ADNI. The definitions of SMCI and PMCI in both ADNI are based on whether MCI subjects would convert to AD within 36 months after the baseline time. Demographic and clinical information of the subjects are shown in Table 1.

Image pre-processing is performed for all MRI and PET images. First, Anterior Commissure (AC) – Posterior Commissure (PC) correction is implemented on all images, and the N3 algorithm [27] is applied to correct the intensity inhomogeneity. Next, skull stripping on structural MR images is conducted using CAT12 in the SPM package [25]. After the removal of the cerebellum, CAT12

in the SPM package is applied to segment structural MR images into three different tissues: grey matter, white matter, and cerebrospinal fluid. For PET image, head alignment (Realign) is applied to eliminate differences between the slices caused by slight brain shaking during scanning. Then PET image is aligned to its corresponding MR image of the same subject using a rigid transformation. The realigned image is normalized to MNI space. A spatial filter with a full width at half maxima of 6 is used to improve the signal-noise ratio of the image.

Network structure

The residual networks [4] have shown to be effective in training a deep network. The identity mapping and the bypass path play an important role in making the training of deep networks easy. Following the architecture of ResNet, DMFNet is proposed based on an attention mechanism to extract and fuse features of MRI and PET. There are three branches in DMFNet, two of which extract features of MRI and PET respectively, and the channel attention model [7] applied to extract the features of each branch and fuse the reweighted feature maps. The third branch is used to further extract fused features.

A new residual block is designed based on the basic block of ResNet. The proposed residual block is divided into four stages, and the corresponding numbers of channels are 64, 128, 256 and 512 respectively. Besides, an attention model is introduced to build our residual block. The structure of DMFNet is shown in Fig. 1. MRI and PET images are input to the network. During the process of network inference, a set of feature maps at each stage is output in each branch, and then weights are assigned to fuse the two sets of features maps through the attention model automatically. The reorganized feature maps and the fused feature map in the previous stage are fused together by concatenating and adding. In this way, low-level and high-level features could be utilized at the same time in DMFNet. Moreover, the combination of identity mapping and residual block ensures the effectiveness and the depth of the attention network.

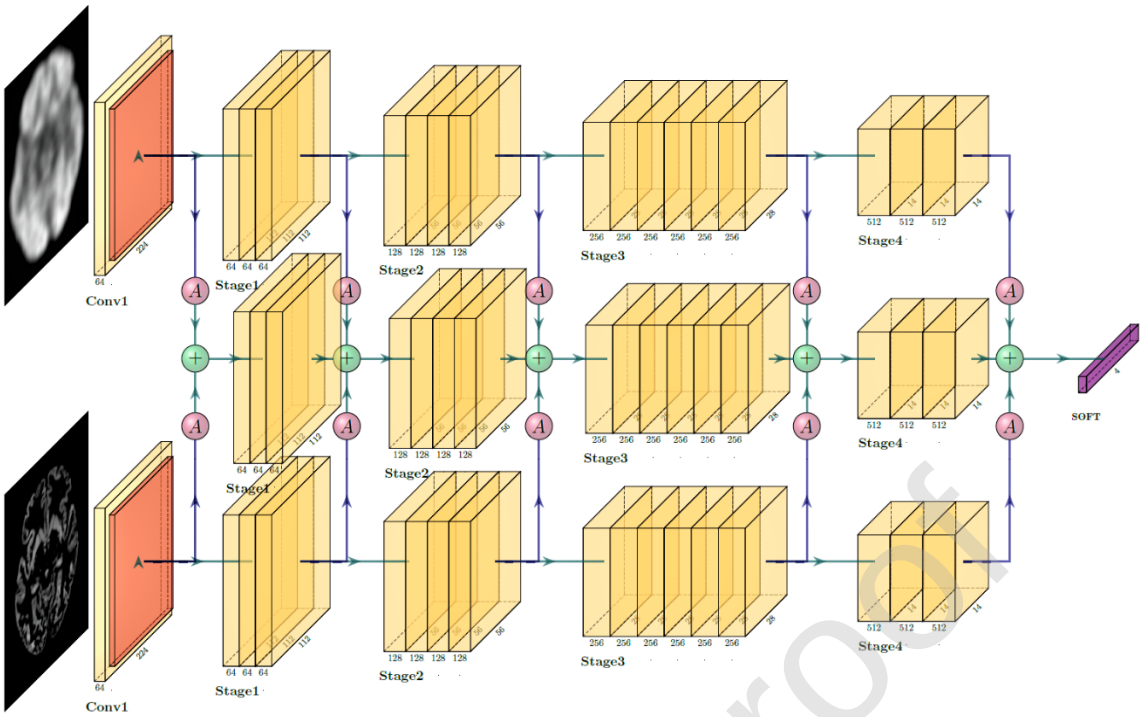


Figure 1 | The Architecture of the proposed DMFNet. MRI image and PET image are processed by two branches separately. 𝔸 represents the channel attention model. ⊕ represents the feature fusion operation, which contains addition and concatenation.

Attention Model

The essence of the attention mechanism is a series of attention distribution coefficients or weight parameters, which can be applied to enhance or select important information of the target, and suppress some irrelevant detailed information. For AD patients, the atrophy region is usually concentrated on the hippocampus and the entorhinal cortex, which means that most regions of the input image are unrelated to the disease. Indistinguishable information will increase the difficulty of AD classification, especially for images with an extremely high similarity of the human brain. In addition, the fusion ratios of each branch need to be assigned according to the importance of the features extracted from MRI and PET images. Therefore, the attention mechanism is introduced to explore potential characteristics of corresponding ROI, while different ratios are assigned automatically in each fusion branch. The architecture of the channel attention model is shown in Fig. 2.

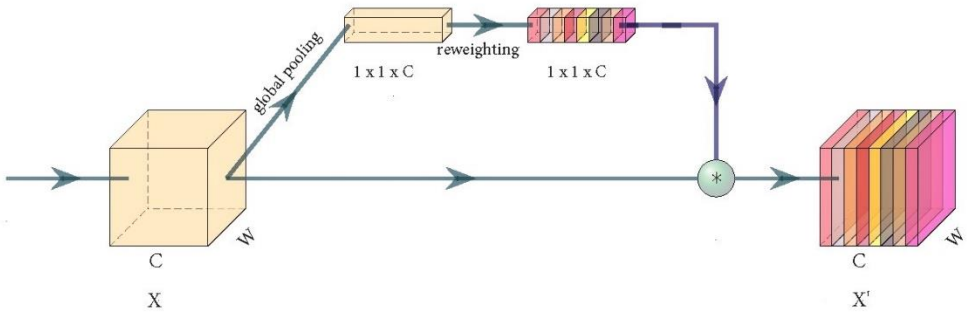


Figure 2 | The architecture of the channel attention model. X represents the input feature map. $\mathrm{X}^{'}$ represents the output feature map after channel reweighting.

Inspired by [7], an improved channel attention model is proposed. For an input feature map X with C channels, the deployment of the attention model is completed in the following three steps. 1)We perform feature compression along the spatial dimensions, transforming the two-dimensional feature of each channel into a real number. The real number has a global receptive field to some extent, and the dimensions of output are in accordance with the number of input feature channels. This process can be implemented by global pooling, and the corresponding equation of squeeze operation is shown as follows.

$$\mathrm{Z}_C = \mathrm{Fse} = \frac{1}{H * W} \sum_{\mathrm{i}}^{H} \sum_{\mathrm{j}}^{W} X_C(i, j) \tag{1}$$

Where H, W and C represent the height, the width and the number of channels of each feature map, respectively. 2)In order to map the importance of each channel with a compressed set of real numbers, new weights for each feature channel are generated to explicitly model the correlation between feature channels. A 1 * 1 convolution is able to explore the correlation among different channels, thereby obtaining the weight distribution of these channels. The corresponding calculation is formularized in Eqn. (2).

$$\mathrm{S}_C = \mathrm{Fex(Z)} = \delta(\mathrm{conv}(\mathrm{Z}_C)) \tag{2}$$

Where δ represents the Sigmod activation function. 3)Weights regenerated in the second step reflect the importance of each channel. Then the origin features are multiplied gradually to complete the redistribution of the original features in the channel dimension. The transition of the input feature map X_C to $\mathrm{X}^{'}_C$ can be expressed as Eqn. (3).

$$\mathrm{X}^{'}_C = \mathrm{Frw(x, s)} = \mathrm{S}_C \otimes \mathrm{X}_C \tag{3}$$

In this way, feature maps X_C are transformed into new feature maps $\mathrm{X}^{'}_C$ with reweighted channel information. To some extent, the attention model essentially introduces additional dynamic characteristics on the input, which can be considered as a self-attention function on the channel.

Attention for feature extraction

In ResNet, the network consists of a series of basic blocks and bottleneck blocks. Each residual block contains two subbranches: the identity mapping branch and the residual branch. The residual block is given as Eqn. (4).

$$X_{t+1} = H_t(X_t) + X_t \tag{4}$$

Where X_t represents the input of the t-th residual block, H_t represents the conversion equation corresponding to the t-th residual block. The attention model is embedded into each residual block. However, stacking attention model simply may not bring significant performance improvement [1]. There are two main reasons for this problem. First, repeating the dot product in the deep network will decrease the value of the feature maps. Second, the attention model potentially breaks the identity mapping structure of the residual branch, which increases the difficulty for the network expanding to a deeper level. Therefore, identity mapping and the attention model are combined into the new residual block, as defined in Eqn. (5).

$$X_{t+1} = A_t \otimes H_t(X_t) + H_t(X_t) + X_t \tag{5}$$

Where A_t represents the function of attention model. The network image can be expanded effectively to a deeper level by a similar identity mapping structure, obtaining high-level features. As shown in Fig. 3, in the basic block of ResNet, the feature map is convolved by two 3 * 3 convolutional layers, and the convolutional output is added to the input feature map to build the identity mapping. In the residual block of DMFNet, an attention-based model is connected for further feature extraction, and then the output of the attention model is added to the feature map and the convolutional output.

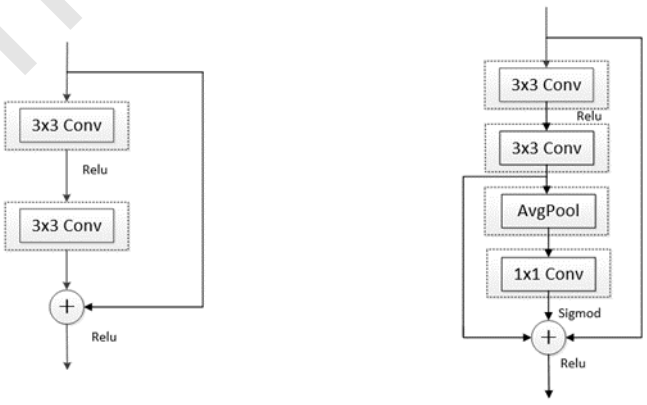


Figure 3 | The comparison of basic block and the proposed building block. Relu and Sigmod represent an activation function commonly used in artificial neural networks.

Attention for feature fusion

In most traditional multi-modal fusion algorithms, SAE or DBM is used to map features into a lower-dimensional space, and then the features for classification are fused. But the low-level feature fusion will destroy the original information of MRI and PET. Currently some studies vote on the classification results of two networks at the decision level or concatenate features extracted from two networks. This method ignores the synergy between different features, which leads to inefficient use of carried information.

The information characterized by MRI and PET images is not the same, so it is critical to use the complementary relationship between the two modalities. In past research, this point was usually ignored by researchers. It was used as hyperparameters to adjust the ratio of each modality, but the cost of such human intervention is usually time-consuming. This paper introduces an attention mechanism to automatically assign the proportion of each modality during fusion. It is defined in Eqn. (6).

$$\mathrm{X}_{F,t} = \mathrm{X}_{F,t-1} + \mathrm{A}_{M,t} \otimes \mathrm{X}_{M,t} + \mathrm{A}_{P,t} \otimes \mathrm{X}_{P,t} \tag{6}$$

Where $\mathrm{A}_{M,t}$ and $\mathrm{A}_{P,t}$ respectively represent attention models corresponding to the t-th building block of MRI and PET, $\mathrm{X}_{M,t}$, $\mathrm{X}_{P,t}$ and $\mathrm{X}_{F,t}$ represent the feature map corresponding to the t-th building block of MRI, PET, and the fusion branch respectively.

Experimental Results and Discussions

Slice Selection

The format of the original image archived from the ADNI database is NIfTI, so it is necessary to take out 2D slices from the 3D image for AD classification. In some studies based on 2D image classification [2], the selection of brain slices has not caught the attention of researchers. Introducing indiscriminate data into the dataset is usually unfriendly for image classification, so the slice with the subtle lesion region should be selected precisely. The numbers of slices of MRI and PET are different due to different scanning methods. To align the slices of the two modalities, the images of both modalities are resampled with the size of 180 * 180 * 180, and then 60 slices gathered in the image center are taken out from each axial. Although the spatial information of a sample in a 2D image taken from a 3D image will be greatly destroyed, slices from all axial are selected to minimize this loss. To screen out more descriptive slices, a slice screening network based on AlexNet [3] is designed. The slices located in the same position of all samples are taken out as a slice-dataset. Then the slices are filtered according to the classification accuracy corresponding to each slice-dataset.

Finally, 20 slices are selected out of 180 slices from each sample as the data set. The data set is divided into the training set, the validation set and the test set at a ratio of 6: 2: 2.

Training

For AD patients, there are usually large morphological differences between AD and NC, which makes the patients easier to be distinguished. But for MCI patients, the morphological differences between MCI and NC are insufficient to distinguish the patients with a simple method. In order to verify the performance of our method for predicting AD and MCI conversion, three sets of experiments: AD vs NC, SMCI vs PMCI, and four classifications are set up. In order to verify the superiority of our proposed multi-modal fusion method, several comparative experiments are set up based on the same dataset: 1) A single-modality MRI classification task based on ResNet (Baseline); 2) Multi-modal classification task based on ResNet (ResNetCom). Specifically, two branches are used to extract features of MRI and PET images, and the separate feature maps are set up and then SoftMax is used to complete the classification; 3) Multi-modal classification task based on DMFNet_v1. Compared with ResNetCom, the reorganized feature maps are fused at each stage of DMFNet; 4) Multi-modal classification task based on DMFNet_v2. Compared with ResNetCom, feature extraction and feature fusion based on attention mechanism are appended in DMFNet.

Implementation Details

Random scaling, cropping, and flipping are applied for data enhancement. The Adam optimizer with an initial learning rate of 0.002, a momentum of 0.9 and a weight decay of 1e-5 is applied in gradient descent. When performing batch training on an NVIDIA RTX2080Ti, the size is set to 32. The optimization of the learning rate is adjusted in two ways. One is to update by an exponential decay. The size of the learning rate is exponentially as the number of epochs increased. The other is to adjust the learning rate with the help of the train set. As the loss increases, the learning rate is further reduced with an exponential decay.

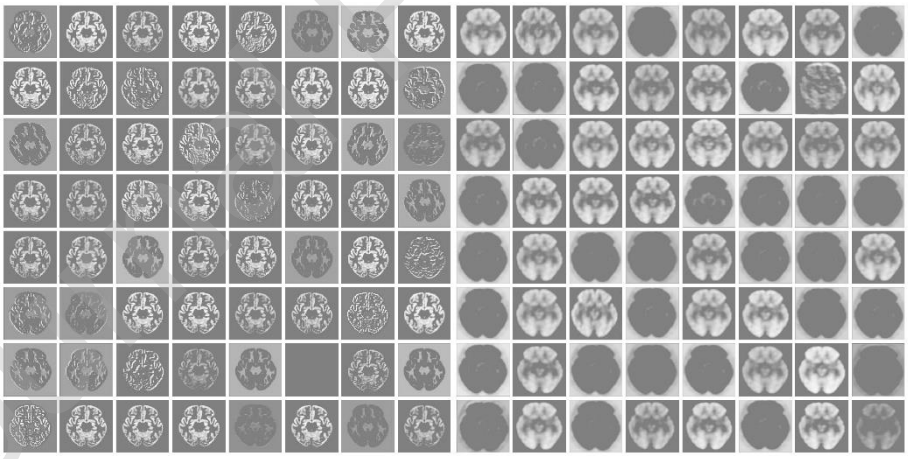
Results

As shown in Table 2, the final classification results of AD classification, MCI conversion prediction and Four-Class based on test set are 95.21 %, 89.79%, and 86.15%, respectively. The results of our proposed method are obviously better than other comparative methods in each group of experiments. In the experiments of AD vs NC, because of large differences between AD and NC images, which is easy to be distinguished, our algorithm does not have a significant performance improvement. Comparing single-modal and multi-modal based on ResNet, we can find that the simple decision-level fusion does not bring about significant performance improvement because the late fusion

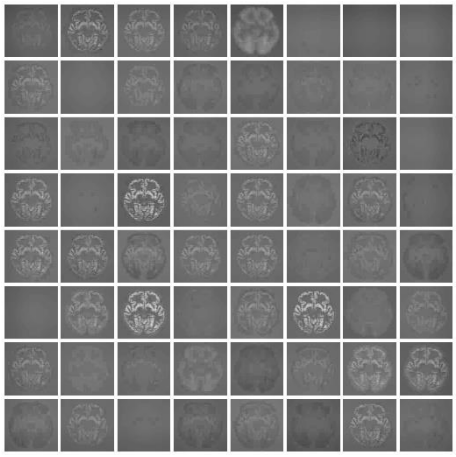
ignores the synergy between the low-level features. In the classification of SMCI vs PMCI, the classification accuracy of traditional algorithms has not been very high, because the samples in the stable and the progressive stages of MCI have no significant difference. According to the experimental results, our method is significantly better than single-modal and simple multi-modal fusion methods. The reason is that the attention model in the residual block can explore potential features between data. In the four-class classification, our method could provide great improvement. In general, the complexity of a network increases with its depth. When a complex model is applied to a simple classification problem, not much performance improvement could be achieved. In other words, a deep network tends to solve more complex classification task. As the category number of multi-category classification raises from 2 to 4, the classification accuracy improves significantly.

Data Visualization and Statistical Analysis

To understand how the attention model works better, we visualize the feature maps from Conv1 layer of DMFNet (shown in Fig. 4) since low-level features are more consistent with visual intuitions. Note that we visualize three branches feature maps for better illustration. The number of images corresponds to the number of channels. According to the Eqn. (6), we could find how do the features of MRI and PET impact on the fused features. The attention model tends to give a higher weight to the branch which contains more valid information, it means that the importance of feature channel plays a crucial role feature fusion.



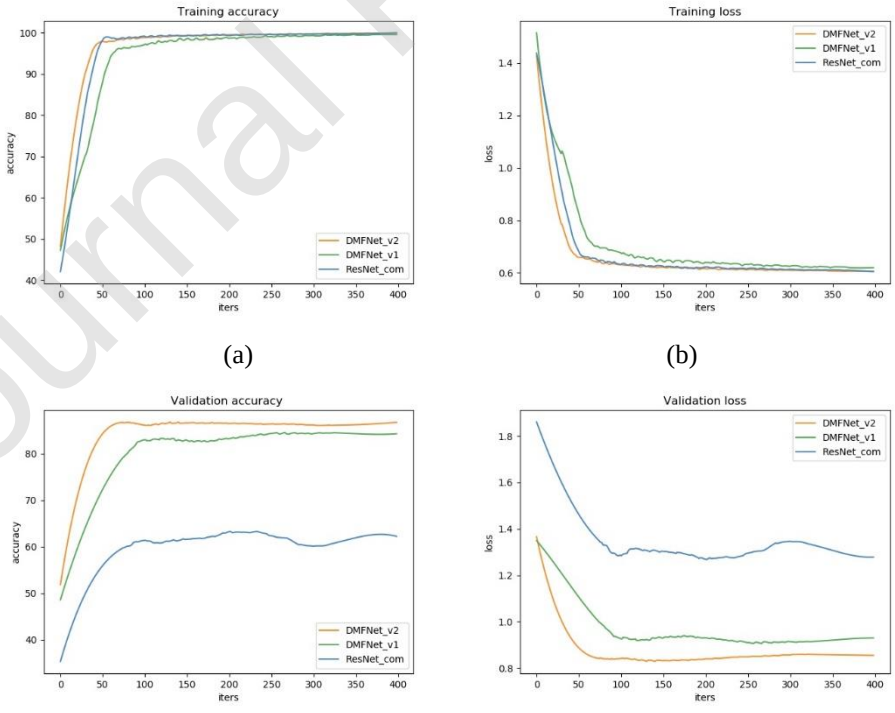
(a) (b)



(c)

Figure 4 | Visualization of local feature maps extracted by DMFNet. (a) feature maps extracted from MRI images, (b) feature maps extracted from PET images, (c) feature maps fused by (a) and (b).

The softmax layer outputs a probability distribution. Cross entropy indicates the distance between the output distribution and the original distribution. Cross entropy loss function is used to evaluate the training process of the neural network. Fig. 5 shows the trend of the observed value changing with the number of iterations. It can be seen that the DMFNet_v2 network converges after fewer iterations in training process. After each network model iterates for 100 cycles in the verification set, the loss function value tends to be stable. Due to the introduction of the attention model and the fused branch, the complexity of our model increases within an acceptable range. However, the convergence of the proposed model has been greatly improved.

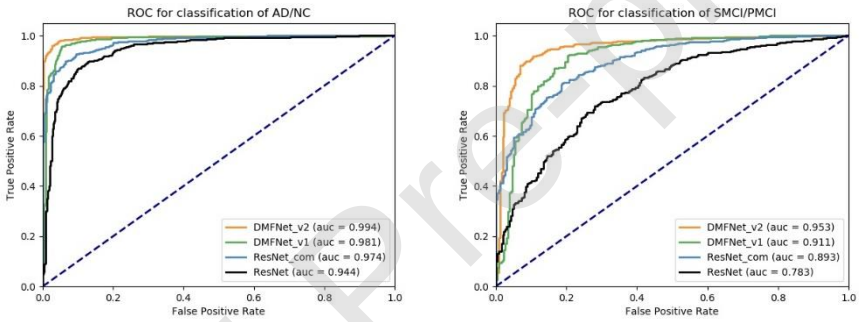


(a) (b)

(c) (d)

Figure 5 | Accuracy and loss curves achieved by three different methods in AD classification on ADNI. (a) and (b) indicate the trend of accuracy and loss in the training step, (c) and (d) indicate the trend of accuracy and loss in the verification step.

Receiver Operating Characteristic (ROC) curves and of different classification models tested on the 20% ADNI dataset are shown in Fig. 6. The Area Under the ROC Curve (AUC) for classification of AD/NC is 0.994, 0.981, 0.974 and 0.944 respectively. The AUC for classification of SMCI/PMCI is 0.953, 0.911, 0.893 and 0.783 respectively. To investigate the significance of classification performance between different methods, we have carried out a non-parametric statistical analysis, namely DeLong's test [33], for the comparison of each two ROC curves for classification of SMCI/PMCI on ADNI dataset, with a confidence interval of 95%. The results indicate that DMFNet_v2 performs significantly better than DMFNet_v1, ResNet_com and ResNet with p values = 0.003, 3.57×10^{-6} and $5{,}34\times10^{-9}$, respectively. The above AUCs and p values indicate that the proposed network has reasonable ability to distinguish the early AD patients. Especially in the classification of SMCI/PMCI, the DMFNet outperforms ResNet-based model.



(a) (b)

Figure 6 | ROC curves achieved by three different methods in AD classification on ADNI. (a) indicates ROC curves for classification of AD/NC, (b) indicates ROC curves for classification of SMCI/PMCI.

Classification Performance

Following a denoising fashion, Liu et al. [15] used SAE and a zero-mask strategy for feature fusion to extract complementary features from multi-modal data. SAE was applied to obtain high-level features in the unsupervised pretraining stage, which achieved an AD/NC classification accuracy of 91.4%. Liu et al. [31] proposed a classification framework based on combination of 2D CNN and Recurrent Neural Networks (RNNs), which learns the intra-slice and inter-slice features for classification after decomposition of the 3D PET image into a sequence of 2D slices, which achieved an AD/NC classification accuracy of 91.2%. Beheshti et al [32] used a voxel-based morphometry technique to investigate global and local gray matter atrophy. SVM is applied to learn the feature extracted from gray matter, which achieved an AD/NC classification accuracy of 93.01%. Liu et al.

[20] proposed a landmark-based deep multi-instance learning framework for brain disease diagnosis. A data-driven learning approach was used to select discriminative patches from MRI images based on AD-related anatomical landmarks identified, which achieved an AD/NC classification accuracy of 91.09% and a MCI conversion prediction accuracy of 76.9%. Lu et al. [21] proposed a deep multi-modal and multiscale neural network to discriminate individuals with AD and used SAE for pre-training, which achieved an AD/NC classification accuracy of 84.6% and a MCI conversion prediction accuracy of 82.93%. Shi et al. [28] used a multi-modal stacked deep polynomial networks algorithm to fuse and learn feature representation from multi-modal neuroimaging data, which achieved an AD/NC classification accuracy of 97.13% and a MCI conversion prediction accuracy of 78.88%. Zhang et al. [22] proposed a dual branch network based on VGG19, and the multi-modal medical images were trained by two independent branches. The results of multi-modal neuroimaging diagnosis were combined with the results of the clinical neuropsychological diagnosis, which achieved an AD/NC classification accuracy of 88.20%.

Deep learning has been applied to AD classification using original neuroimaging data without any feature selection procedures, which greatly preserves the information of the sample [25]. Table 3 shows the comparison of the state-of-the-art method and the proposed methods in this paper. Shi et al. [28] achieved an AD/NC classification accuracy of 97.13%, but their experimental results lacked the verification of proper untrained test data after cross-validation. Benefiting from the attention mechanism, our model has made great improvements in the MCI conversion prediction. We combine accuracy, specificity and sensitivity together to objectively evaluate the results, and the results of the proposed model show its superiority to the most state-of-the-art algorithms.

Conclusion

In this paper, a novel deep multi-modal fusion model is proposed for the early auxiliary diagnosis of AD and MCI conversion. MRI and PET images are trained to learn the synergy between the multi-modal data. The hierarchical architecture in the proposed model ensures that the features of MRI and PET are extracted independently without destroying the information of the original features. In the attention models, features are selectively extracted from MRI and PET branches and the irrelevant information is suppressed, and the weighted features are fused to build the fusion branch. In our model, potential features between different multi-modal data can be explored and combined. Low-level and high-level features are fused so that complementary data could be effectively exploited. The experiments show that the proposed model outperforms the state-of-the-art methods on the ADNI dataset.

In the future, more possibilities could be explored for AD classification and MCI conversion.

The following are our upcoming work. 1) Image segmentation based on Neural Network. The original 3D image of MRI is segmented into gray matter, white matter, etc. It not only means that the tedious work of segmentation using software such as SPM could be reduced, but also provides the possibility for the integration of scanning and predicting of neuroimaging of AD patients. 2) Multi-modal fusion based on 3D CNN. Relative to 2D slice images, 3D images completely preserve the spatial information of a sample, but this method has to face the dilemma of so few available data. 3) Further feature fusion with a new model. Complementary information between MRI and PET features could be learned so that when the inference exists, or even if a kind of modal data is missing, AD could still be predicted based on the single input and corresponding complementary information learned from the pre-trained model.

Credit Author Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. We certify that we have participated sufficiently in the work to take public responsibility for the appropriateness of the experimental design and method, and the collection, analysis, and interpretation of the data. All the authors listed have reviewed the final version of the manuscript and approved it for publication. To the best of our knowledge and belief, this manuscript has not been published in whole or in part nor is it being considered for publication elsewhere.

Declarations of interest

None.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company;

EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1]. Wang F, Jiang M, Qian C, et al. Residual Attention Network for Image Classification[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.]. Doi：10.1109/CVPR.2017.683

[2]. Hon M, Khan N. Towards Alzheimer's Disease Classification through Transfer Learning[J]. 2017.

[3]. Alex Krizhevsky, I Sutskever, G Hinton. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2). Doi：10.1145/3065386

[4]. He K, Zhang X, Ren S, and Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In ICCV, 2015. 2, 5 Doi: 10.1109/ICCV.2015.123

[5]. Wang J., Wang Q., Wang S., Shen D. (2017) Sparse Multi-view Task-Centralized Learning for ASD Diagnosis. In: Wang Q., Shi Y., Suk HI., Suzuki K. (eds) Machine Learning in Medical Imaging. MLMI 2017. Lecture Notes in Computer Science, vol 10541. Springer, Cham. Doi: 10.1007/978-3-319-67389-9_19

[6]. Jingfan Fan, Xiaohuan Cao, Qian Wang, Pew-Thian Yap, Dinggang Shen, Adversarial learning for mono- or multi-modal registration, Medical Image Analysis,Volume 58,2019,101545,ISSN 1361-8. Doi: 10.1016/j.media.2019.101545.

[7]. Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. Doi: 10.1109/TPAMI.2019.2913372

[8]. Baltrusaitis T, Ahuja C, Morency L P. Multimodal Machine Learning: A Survey and Taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,

2018:1-1. Doi：10.1109/TPAMI.2018.2798607

[9]. Hosseini-Asl E, Keynto R, El-Baz A. Alzheimer's Disease Diagnostics by Adaptation of 3D Convolutional Network[J]. In ICIP, 2016. Doi:10.1109/ICIP.2016.7532332

[10]. Yu Zhang, Han Zhang, Xiaobo Chen, Mingxia Liu, Xiaofeng Zhu, Seong-Whan Lee, Dinggang Shen, Strength and similarity guided group-level brain functional network construction for MCI diagnosis, Pattern Recognition,Volume 88,2019,Pages 421-430,ISSN 0031-3203. Doi: 10.1016/j.patcog.2018.12.001.

[11]. Bailey, D.L., Pichler, B.J., Gückel, B. et al. Combined PET/MRI: Multi-modality Multi-parametric Imaging Is Here. Mol Imaging Biol 17, 595–608 (2015). Doi: 10.1007/s11307-015-0886-9.

[12]. Singh, S., Srivastava, A., Mi, L., Caselli, R. J., Chen, K., Goradia, D., Reiman, E. M., & Wang, Y. (2017). Deep-learning-based classification of FDG-PET data for Alzheimer's disease categories. In 13th International Conference on Medical Information Processing and Analysis (Vol. 10572). [105720J] SPIE. doi:10.1117/12.2294537

[13]. Dyrba, M., Grothe, M., Kirste, T. and Teipel, S.J. (2015), Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. Hum. Brain Mapp., 36: 2118-2131. doi:10.1002/hbm.22759

[14]. Alzheimer’s Association, 2018 Alzheimer’s disease facts and figures, Alzheimers Dement. 14 (3) (2018) 367–429. Doi: 10.1016/j.jalz.2018.02.001

[15]. Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., et al. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer’s Disease. IEEE Trans. Biomed. Eng. 62, 1132–1140. Doi: 10.1109/TBME.2014.2372011

[16]. Suk H, Lee W, and Shen D, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage, 2014. Doi: 10.1016/j.neuroimage.2014.06.077

[17]. Zhang D, Wang Y, Zhou L, et al. Multimodal classification of Alzheimer\"s disease and mild cognitive impairment[J]. Neuroimage, 2011, 55(3):856-867.Doi : 10.1016/j.neuroimage.2011.01.008

[18]. Zhou, T, Thung, K-H, Zhu, X, Shen, D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum Brain Mapp. 2019; 40: 1001– 1016. Doi: 10.1002/hbm.24428

[19]. Cheng, D., and Liu, M. (2017). “CNNs based multi-modality classification for AD diagnosis,” in 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (Shanghai), 1–5.

[20]. Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018). Landmark-based deep multiinstance learning for brain disease diagnosis. Med. Image Anal. 43, 157–168. doi: 10.1016/j.media.2017.10.005

[21]. Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structuralMR and FDG-PET images. Sci. Rep. 8:5697. doi: 10.1038/s41598-018-22871-z

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

[23]. McDonald, C. R., Gharapetian, L., McEvoy, L. K., Fennema-Notestine, C., Hagler, D. J. Jr., Holland, D., et al. (2012). Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. Neurobiol. Aging 33, 242–253. Doi: 10.1016/j.neurobiolaging.2010.03.015

[24]. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[25]. Jo T, Nho K and Saykin AJ (2019) Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. Front. Aging Neurosci. 11:220. doi: 10.3389/fnagi.2019.00220

[26]. Friston, K. J. (2007). Statistical parametric mapping: the analysis of functional brain images. Neurosurgery 61, 216–216. Doi: 10.1016/B978-012372560-8/50002-4

[27]. Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97. Doi: 10.1109/42.668698

[28]. Shi J, Zheng X, Li Y, et al. Multimodal Neuroimaging Feature Learning with Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer\"s Disease[J]. IEEE Journal of Biomedical and Health Informatics, 2017:1-1. Doi: 10.1109/JBHI.2017.2655720

[29]. Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature 521:436. doi: 10.1038/nature14539

[30]. Folstein M. F, Folstein S. E., and McHugh P. R., "Mini-mental state".A practical method for grading the cognitive state of patients for the clinician," J. Psychiatric Res., vol. 12, no. 3, pp. 189–198, Nov. 1975.

[31]. Liu M, Cheng D and Yan W (2018) Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images. Front. Neuroinform. 12:35. doi: 10.3389/fninf.2018.00035

[32]. Beheshti I, Maikusa N, Daneshmand M, et al. Classification of Alzheimer's Disease and Prediction of Mild Cognitive Impairment Conversion Using Histogram-Based Analysis of Patient-Specific Anatomical Brain Connectivity Networks[J]. Journal of Alzheimers Disease, 2017, 60(1):295-304.

[33]. DeLong, E., DeLong, D., & Clarke-Pearson, D. (1988). Comparing the Areas

under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics, 44(3), 837-845. doi:10.2307/2531595

Table 1 | The statistical information for the subjects in ADNI. Values are reported as Mean ± Standard Deviation (Std); MMSE: mini-mental state examination.

Group	NC	EMCI	LMCI	AD
Sample size	163	113	105	119
Male/female	75/88	63/50	49/56	59/60
Age	76.74±6.80	71.45±7.09	71.66±7.67	74.67±+7.82
MMSE [30]	29.08±1.07	27.81±2.08	26.93±2.48	22.51±2.95

Table 2 | The Results of classification on ADNI among different methods.

Group	Baseline	ResNet_com	DMFNet_v1	DMFNet_v2
AD / NC	88.70%	85.37%	94.86%	95.21%
SMCI / MCI	77.87%	81.19%	87.27%	89.79%
Four-Class	62.10%	66.15%	84.30%	86.15%

Table 3 | The comparison results of AD classification, MCI conversion prediction with state-of-the-art methods.

References	Modality	NO. of subjects	method	ACC (AD vs NC)	SEN	SPE	ACC (SMCI vs PMCI)	SEN	SPE
Liu et al. (2015)	MRI, PET	800	SAE	*91.4*	*92.32*	*90.42*	*82.1 (MCI vs NC)*	*60.0*	*92.32*
Liu et al. (2018a)	PET	339	RNN	*91.2*	*91.4*	*91.0*	*78.9(MCI vs NC)*	*78.1*	*80.0*
Beheshti et al. (2017)	MRI	322	3D SVM	*93.01*	*89.13*	*96.80*	*75.00*	*76.92*	*73.23*
Liu et al. (2018b)	MRI	636	3D CNN	*91.09*	*88.05*	*93.50*	*76.9*	*42.11*	*82.43*
Lu et al. (2018)	MRI, PET	1242	DNN	*84.6*	*80.2*	*91.8*	*82.93*	*79.69*	*83.84*
Shi et al. (2018)	MRI, PET	202	DPN	***97.13***	*95.93*	*98.53*	*78.88*	*68.04*	*86.81*
Zhang et al. (2019)	MRI, PET	400	VGG19	*88.20*	*97.43*	*84.31*	*85.74 (MCI vs NC)*	*90.11*	*91.82*
proposed	MRI, PET	500	DMFNet	***95.21***	*93.56*	*97.48*	***89.79***	*81.15*	*93.46*

Note: SEN = TP/ (TP + FN), SPE = TN/ (TN + FP). TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative.