Multi-Modal Data Fusion with Missing Data Handling for Mild Cognitive Impairment Progression Prediction

Shuting Liu[1], Baochang Zhang[1], Veronika A. Zimmer[1,2], and Daniel Rueckert[1,2,3]

[1] School of Computation, Information and Technology, Technical University of Munich, Munich, Germany
shuting.liu@tum.de
[2] School of Medicine and Health, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
[3] Department of Computing, Imperial College London, London, UK

Abstract. Predicting Mild Cognitive Impairment (MCI) progression, an early stage of Alzheimer's Disease (AD), is crucial but challenging due to the disease's complexity. Integrating diverse data sources like clinical assessments and neuroimaging poses hurdles, particularly with data preprocessing and handling missing data. When data is missing, it can introduce uncertainty and reduce the effectiveness of statistical models. Moreover, ignoring missing data or handling it improperly can distort results and compromise the validity of research findings. In this paper, we introduce a novel fusion model considering missing data handling for early diagnosis of AD. This includes a novel image-to-graphical representation module that considers the heterogeneity of brain anatomy, and a missing data compensation module. In the image-to-graphical representation module, we construct a subject-specific graph representing the connectivity among 100 brain regions derived from structural MRI, incorporating the feature maps extracted by segmentation network into the node features. We also propose a novel multi-head dynamic graph convolution network to further extract graphical features. In the missing data compensation module, a self-supervised model is designed to compensate for partially missing information, alongside a latent-space transfer model tailored for cases where tabular data is completely missing. Experimental results on ADNI dataset with 696 subjects demonstrate the superiority of our proposed method over existing state-of-the-art methods. Our method achieves a balanced accuracy of 92.79% on clinical data with partially missing cases and an impressive 92.35% even without clinical data input.

1 Introduction

Alzheimer's Disease (AD), a progressive neurodegenerative disorder, presents a significant global public health challenge [3], with an expected substantial increase in affected individuals in the coming decades [9, 8]. Depending on the

criteria suggested by [1], current Mild Cognitive Impairment (MCI) patients can be categorized as either progressive MCI (pMCI) or stable MCI (sMCI) based on potential progression to AD within a typical 36-month follow-up period. However, accurately predicting MCI progression (i.e., pMCI or sMCI) to AD remains intricate and daunting due to the disease's heterogeneous nature and the multifactorial influences contributing to its onset and progression.

In recent years, Convolutional Neural Networks (CNN) have shown promising performance in brain disease diagnosis and prediction [7, 14]. CNN also progress in predicting AD conversion in advance from MCI [12, 15]. Compared with traditional machine learning methods [6], 3D CNNs excel at extracting high-level information about neuroanatomy from Magnetic Resonance Imaging (MRI). A multi-stream CNN fed with patch-based MRI imaging data has been proposed to classify stable MCI and progressive MCI [2]. Since the features of different parts make different contributions to the overall classification performance, a dual attention multi-instance deep learning model is proposed to automatically find and highlight the most informative points on feature maps for boosting the performance of pMCI/sMCI classification [17]. Also, several studies show that models pre-trained by normal controls (NC) and AD can further boost the prediction of the progression of MCI [13]. Instead of extracting features directly from 3D MRI data like the above methods, other studies focus on converting images into image-derived phenotypes in form of tabular features or conducting additional manual feature extraction. For instance, Zheng et al. [16] utilized Freesurfer to extract features and construct tabular information for further prediction. However, the aforementioned approaches all focus exclusively on neuroimaging data, ignoring the importance of clinical data. Some successful multi-modal models to fuse the different data sources have been proposed and achieved some improvements in predictive tasks, but image and clinical data are only integrated through simple concatenation [4]. Recently, a dynamic brain graph-based method for the fusion of imaging and tabular data [10] has been proposed for predicting cognitive outcomes in stroke cases, taking into account the specificity of a patient's brain anatomy. Although these fusion models achieve better performance for disease diagnosis, multi-modal data, i.e., medical examination data and neuroimaging data, are required as input, thus greatly limiting their applicability. Moreover, the acquisition of tabular patient data frequently lacks standardization, leading to incomplete or entirely absent information. Consequently, managing missing data in the context of multi-modal fusion models remains an open issue.

In this paper, we present a novel fusion model considering missing data handling for the early diagnosis of AD. Our contributions are as follows: (1) We introduce a novel missing data compensation module aimed at resolving issues of unavailability and incompleteness in clinical data, thereby enhancing the model's robustness to missing data and expanding its applicability. (2) In response to the heterogeneity of the brain anatomy, a novel image-to-graphical representation module is proposed to construct subject-specific dynamic brain region graphs, utilizing first-order statistical features calculated from raw T1-MRI data and segmentation related feature maps as node features. (3) We show comprehen-

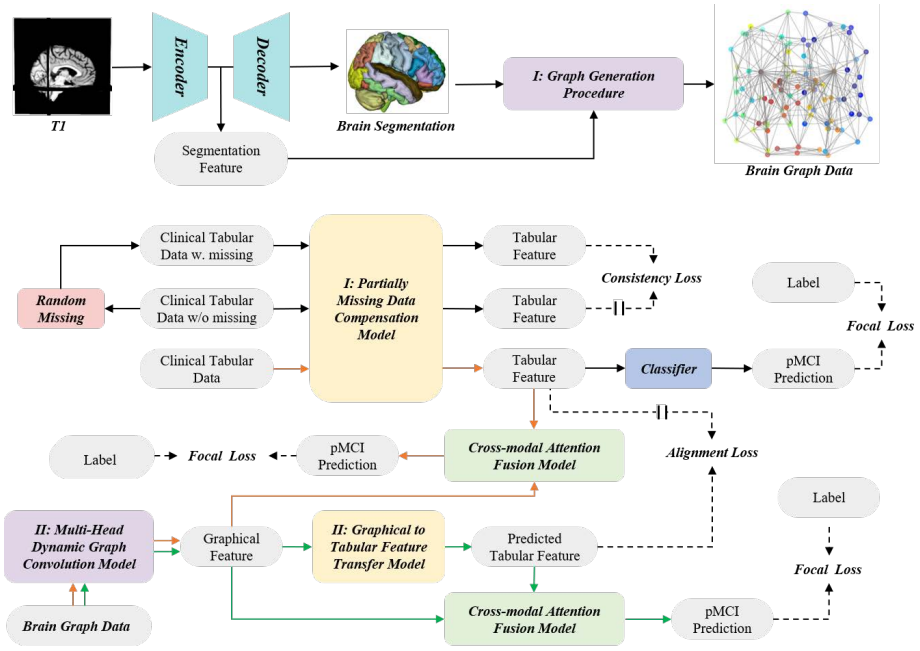

Fig. 1. The overview of our method for MCI progression prediction. Models in purple box belong to image-to-graphical representation module and models in yellow box belong to missing data compensation module. "Orange line" shows the flow with partially missing clinical data, and "Green line" means the flow without clinical data.

sive experimental results on the ADNI dataset and demonstrate the superior performance of our proposed method, significantly outpacing rival methods by a considerable margin. A detailed ablation study is conducted to verify the effectiveness of each proposed module.

2 Methodology

An overview of the proposed method is illustrated in Fig. 1. Considering the specificity of patient brain anatomy, brain graph features F_g are extracted via image-to-graphical representation module, as introduced in Section. 2.1. Meanwhile, accounting for the issues of unavailability and incompleteness in clinical data, the compensated tabular features F_t are extracted by a novel missing data compensation module, as outlined in Section. 2.2. In Section.2.3, a cross-modal attention based fusion model is employed to fuse F_t and F_g for pMCI prediction.

2.1 Image-to-Graphical Representation Module

Graph Generation. We represent the nodes of the graph with 100 structural brain regions segmented by FastsurferCNN [5]. The edges of the graph

are established based on the neighboring relationship of each brain structural area similar to [10]. FastsurferCNN, as a state-of-the-art brain region segmentation tool, enables to extract image-derived phenotypes across various anatomical regions, which facilitates the development of models that consider the heterogeneity of brain anatomy and the specificity of each subject. We extract the feature map output by the bottleneck of FastsurferCNN and interpolate it to the original image size. Based on segmentation result, the first-order statistical features are calculated from extracted feature map and T1 data at each brain structural region, which are used as graph features. Consequently, a total of 361 features are assigned to each node. The final brain graph data can be expressed as $G = \{F_g, E\}$, where $E \in \mathbb{R}^{N \times N}$ denotes the adjacency matrix, $F_g \in \mathbb{R}^{N \times 361}$ is graph feature, N is the number of nodes.

Multi-Head Dynamic Graph Convolution Network. In order to optimize the broadcasting of the graph and extract informative graphic feature, a multi-head dynamic graph convolution network is proposed, which has a linear layer and three graph convolution layers. For each graph convolution layer, it learns the mapping between the key $k_g \in \mathbb{R}^{N \times 256}$ and the query $q_g \in \mathbb{R}^{N \times 256}$ representation of updated graph feature $F_g \in \mathbb{R}^{N \times 256}$ by a linear layer. Then, the shapes of query q_g, key k_g and updated node feature D are reshaped as $[h, N, 256/h]$, where h is the number of heads. The dynamic edge weight $W \in \mathbb{R}^{h \times N \times N}$ is computed by measuring the similarity between q_g and k_g. Using the computed dynamic edge weight W and fixed adjacency matrix E, the graph will be updated and reshaped as $G = \{F_g, E | F_g \in \mathbb{R}^{N \times 256}\}$ for the preparation of starting the next graph convolution layer. Similar to [10], but we use sigmoid function instead of softmax function to scale the calculated edge weight. The whole process can be formulated as:

$$F_g = \left(\mathcal{R}(E) \bullet \sigma \left(q_g k_g^T / \sqrt{\frac{256}{h}} \right) \right) D \tag{1}$$

Here $\sigma()$ is the sigmoid function. $\mathcal{R}(E) \in \mathbb{R}^{h \times N \times N}$ is an operation to unsqueeze the adjacency matrix E in the first dimension and repeat it by h times. $\bullet$ denotes the element-wise multiplication operator.

2.2 Missing Data Compensation Module

Partially Missing Data. Rather than naively filling in missing elements with an average value or zero, a partially missing data compensation model is proposed, which takes the raw clinical data $X = \{x_i\}$ and a missing vector V as input. The missing vector V defines the missing values of the clinical data X. V has the same size as X and missing elements are represented as '1' in V, and the remaining elements are '0'. The tabular feature $F_t(X, V)$ extracted and compensated by the model from the given clinical data $\{X, V\}$ can be defined as:

$$F_t(X, V) = \mathcal{M}(BN(\Phi_1(\mathcal{A}(X) \bullet V) + \Phi_2(X))). \tag{2}$$

Here $\mathcal{A}$ is an auto-encoder built by two Multilayer Perceptrons (MLPs), ϕ_1 is a linear layer without considering the bias, ϕ_2 represents a normal linear layer, and $\mathcal{M}$ also is a MLP. Different with [10], the partially missing data compensation model is trained in self-supervised manner, where a consistency loss $\mathcal{L}_{con}$ based on L1-norm is employed to make the tabular feature $(F_t(X_c, V_c), F_t(X_m, V_m))$ extracted from the complete clinical data $\{X_c, V_c\}$ and the paired augmented random missing $\{X_m, V_m\}$ consistent. This can compensate for missing data and makes the model more robust towards missing data. The consistency loss $\mathcal{L}_{con}$ is formulated as:

$$\mathcal{L}_{con} = \|F_t(X_c, V_c) - F_t(X_m, V_m)\|_1 \tag{3}$$

Completely Missing Data. Clinical data plays a critical role in various medical analyses and decision-making processes. However, access to such data may sometimes be restricted or entirely absent. To address this challenge, a graphical-to-tabular feature transfer model is proposed. This innovative approach is designed to convert graphical features into a tabular format, thus facilitating the use of graphical representations in scenarios where traditional clinical data is inaccessible. The process begins with the utilization of a linear layer to aggregate the graph features with 100 nodes into a singular global node representation. Following this, an autoencoder comprised of two MLPs is employed to convert this global node representation into tabular features. During this training phase, an alignment loss function is implemented to minimize the discrepancy between the tabular features predicted by the model and those extracted from the paired clinical data through a partially missing data compensation model. This ensures that the transformed tabular features closely mirror the valuable information typically derived from comprehensive clinical data. The alignment loss $\mathcal{L}_{alg}$ is defined as:

$$\mathcal{L}_{alg} = \|G_T(F_g) - F_t(X, V)\|_1 \tag{4}$$

Here $G_T(F_g)$ is predicted tabular feature from graphical feature F_g by the the proposed graphical-to-tabular feature transfer model G_T.

2.3 Cross-Modal Attention based Fusion Model

Previous work simply multiplies or concatenates feature maps from the two modalities, but this fusion approach cannot model the long-range relations between clinical tabular data and T1 image data. Inspired by [10], a cross-modal attention based fusion model is proposed to fuse global tabular feature $F_t \in \mathcal{R}^{1\times 256}$ and local graphical feature $F_g \in \mathcal{R}^{N\times 256}$ to improve prediction performance. The query $q_t \in \mathcal{R}^{1\times 256}$ is the latent representation of the tabular feature F_t, and both the key $k_g \in \mathcal{R}^{N\times 256}$ and the value $v_g \in \mathcal{R}^{N\times 256}$ are the latent representations of the graphical feature F_g. Then the latent adaptation $F_{g\to t} \in \mathcal{R}^{1\times 256}$ from graphical feature to tabular feature is calculated via,

$$F_{g\to t} = softmax\left(\frac{q_t k_g^T}{\sqrt{256}}\right) v_g \tag{5}$$

Meanwhile, a linear layer is employed to aggregate the features of 100 nodes into a singular global node representation $F_g{}' \in \mathcal{R}^{1\times 256}$. Finally, we fuse the global graphical representation $F_g{}'$, the adaptation feature $F_{g\to t}$, and the tabular feature F_t via concatenation, which can be used as the input of an MLP for classification.

2.4 Implementation Details

For class imbalance issue, the focal loss is employed as shown in Fig. 1, which is formulated as:

$$\mathcal{L}_{focal} = -\alpha\,(1-p)^{\gamma}\,log\,(p) \tag{6}$$

Here, p is the predicted probability of the correct class, $\alpha = 0.33$ is the weighting factor for each sample, $\gamma = 0.05$ is a tunable focusing parameter. The proposed model is implemented on PyTorch library with one NVIDIA GPU (Quadro RTX A6000). The model is trained for 100 epochs with a batch size of 16, using the Adam optimizer (initial learning rate is 0.0005, decay of 0.85 per 5 epoch).

3 Experiments and Results

3.1 Dataset

The data utilized in our study comes from the public ADNI database [11]. 696 cases diagnosed with MCI who possessed baseline T1-weighted structural MRI scans are collected. All MCI cases were categorized into two groups based on clinical progression: (1) sMCI, encompassing individuals who did not progress to AD within a three-year follow-up period. Additionally, subjects diagnosed with MCI at least twice but eventually regressed to standard cognitive function were also classified as sMCI; (2) pMCI, comprising individuals diagnosed with MCI at the initial visit but exhibiting conversion to AD during longitudinal follow-up visits within three years. Among the 696 cases, 471 were classified as sMCI, while 225 were categorized as pMCI.

Simultaneously, we collected 12 demographic tabular records at basetime, including age, sex, and 10 variables of clinical indicators: functional activities questionnaire, AD assessment scale, mini-mental state examination, neuropsychiatric inventory questionnaire, geriatric depression scale, memory functioning, language, visuospatial functioning, and executive functioning I&II. Despite missing indicators in some cases, we included these records in our analysis. Additionally, we collected 2174 pieces of complete MCI clinical data for training the partially missing data compensation model.

3.2 Evaluation Metrics

Five-fold validation are used to evaluate the performance of our method, and six metrics are used, including Balanced Accuracy (BAcc), Classification Accuracy (Acc), Precision (Pre), Sensitivity (Sen), Specificity (Spe), and the area under

Table 1. Quantitative results when only using clinical data as input

Methods	***BAcc***	***Acc***	***AUC***	***Pre***	***Sen***	***Spe***
SVM	0.8106	0.8521	0.9108	0.8211	0.6933	0.9279
LR	0.8108	0.8492	0.9120	0.8088	0.7022	0.9194
MLP	0.8202	0.8319	0.9051	0.7240	0.7867	0.8536
MIC+MLP [10]	0.8332	0.8434	0.9113	0.7378	0.8044	0.8620
PMDC+MLP	**0.8869**	**0.9066**	**0.9519**	**0.8749**	**0.8311**	**0.9426**

Table 2. Quantitative results when only using Image data. $[I]$:input with T1 image, $[F]$:input with FastsurferCNN-extracted feature, $[f]$:input with cortex feature extracted by Freesurfer

Methods	***BAcc***	***Acc***	***AUC***	***Pre***	***Sen***	***Spe***
3D-ResNet$^{[I]}$	0.7358	0.8249	0.8823	**0.9689**	0.4845	**0.9872**
3D-ResNet$^{[F]}$	0.8981	0.9124	**0.9681**	0.8695	0.8578	0.9385
Transformer$^{[f]}$ [16]	0.8693	0.8923	0.9398	0.8550	0.8044	0.9342
DGNR$^{[F]}$ [10]	0.8875	0.8980	0.9429	0.8319	0.8578	0.9172
MHDGCN$^{[F]}$	**0.9168**	**0.9282**	0.9654	0.8952	0.8844	0.9492

the receiver operating characteristic curve (AUC). BAcc is a particularly important metric used to assess the effectiveness of a model on imbalanced datasets, ensuring that performance is not skewed by the disproportionate representation of classes.

3.3 Comparison Results and Ablation Study

First, we evaluate the effectiveness of our missing data compensation module when only clinical data is used as input. Table 1 presents the quantitative results. In this study, MLP is selected as baseline, and two traditional machine learning method are selected as comparison methods, including Support Vector Machine (SVM), Logistic Regression (LR). When without missing data compensation, baseline just shows slightly superior performance than SVM and LR with a BAcc of 0.8202. Considering missing data compensation, an MLP with our proposed Partially Missing Data Compensation model (PMDC+MLP) and a MLP with the Missing Information Compensation model (MIC+MLP) proposed by Liu et al. [10] are studied. Encouragingly, PMDC+MLP yielded significant improvements, achieving a BAcc of 0.8869. These results fully demonstrate the effectiveness of our proposed missing data compensation model.

Table 2 illustrates the performance of various methods applied to T1 imaging data. Initially, utilizing the TI image as input to the 3D-ResNet model yielded a BAcc of 0.7358. Then, taking FastsurferCNN-extracted feature as input, 3D-ResNet significantly improved BAcc, reaching 0.8981, as shown in the 2nd row. This improvement suggests that segmentation tasks can provide

Table 3. Quantitative results when using both Image data and clinical data. ∗: clinical data is completely missing .

Methods	*BAcc*	*Acc*	*AUC*	*Pre*	*Sen*	*Spe*
DGFusion [10]	0.8983	0.9095	0.9548	0.8544	0.8667	0.9299
Ours	0.9279	0.9354	0.9765	0.8967	0.9067	0.9491
DGFusion∗ [10]	0.6989	0.7888	0.7807	0.8721	0.4444	0.9534
Ours∗	0.9235	0.9325	0.9712	0.8959	0.8978	0.9491

Table 4. Ablation study results

Methods	*BAcc*	*Acc*	*AUC*	*Pre*	*Sen*	*Spe*
W/O PMDC	0.9114	0.9225	0.9671	0.8810	0.8800	0.9428
W/O MHDGCN	0.9131	0.9139	0.9703	0.8380	0.9111	0.9152
W/O CAF	0.9161	0.9211	0.9784	0.8631	0.9022	0.9301
Our	0.9279	0.9354	0.9765	0.8967	0.9067	0.9491

high-quality features for downstream classification task. In comparison, dynamic graph neural representation model (DGNR) proposed by [10] achieved an BAcc of 0.8875, while our proposed Multi-Head Dynamic Graph Convolution Network (MHDGCN) achieved an accuracy of 0.9168.

Table 3 illustrates the fusion model where taking both images and clinical data as input. When clinical data partially missing with a missing rate of 0.33, the DGFusion method [10] achieves a BAcc of 0.8983, and our methods achieves a higher BAcc score of 0.9279. When clinical data are not available, the performance of DGFusion drops dramatically, with BAcc dropping from 0.8983 to 0.6989. In contrast, our method can still provide a good performance, achieving a BAcc score of 0.9235.

Finally, we conduct an ablation study to evaluate the effectiveness of each component in our proposed method. Table 4 shows the ablation experiment results, where CAF means our proposed cross-modal attention based fusion model.

4 Conclusion

In summary, our fusion model, adept at handling missing data, offers a promising avenue for predicting MCI progression, crucial in AD diagnosis. Experimentation on the ADNI dataset showcased its superior accuracy, even in scenarios with incomplete clinical data. Outperforming existing methods, our model demonstrates robustness and potential for real-world application. Future research could explore additional data modalities and validate our approach on larger datasets, aiming to translate these advancements into practical tools for early AD detection and intervention.

Acknowledgments. The project was supported by the China Scholarship Council (File No.202106210062) and ERC Grant Deep4MI (884622).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. American Psychiatric Association, D., Association, A.P., et al.: Diagnostic and statistical manual of mental disorders: DSM-5, vol. 5. American psychiatric association Washington, DC (2013)
2. Ashtari-Majlan, M., Seifi, A., Dehshibi, M.M.: A multi-stream convolutional neural network for classification of progressive mci in alzheimer's disease using structural mri images. IEEE Journal of Biomedical and Health Informatics **26**(8), 3918–3926 (2022)
3. Atri, A.: The alzheimer's disease clinical spectrum: diagnosis and management. Medical Clinics **103**(2), 263–293 (2019)
4. Chen, H., Guo, H., Xing, L., Chen, D., Yuan, T., Zhang, Y., Zhang, X.: Multimodal predictive classification of alzheimer's disease based on attention-combined fusion network: Integrated neuroimaging modalities and medical examination data. IET Image Processing **17**(11), 3153–3164 (2023)
5. Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., Reuter, M.: Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. NeuroImage **219**, 117012 (2020)
6. Hojjati, S.H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., Initiative, A.D.N., et al.: Predicting conversion from mci to ad using resting-state fmri, graph theoretical approach and svm. Journal of neuroscience methods **282**, 69–80 (2017)
7. Huang, K., Lin, Y., Yang, L., Wang, Y., Cai, S., Pang, L., Wu, X., Huang, L., Initiative, A.D.N.: A multipredictor model to predict the conversion of mild cognitive impairment to alzheimer's disease by using a predictive nomogram. Neuropsychopharmacology **45**(2), 358–366 (2020)
8. Huang, S.Y., Hsu, J.L., Lin, K.J., Liu, H.L., Wey, S.P., Hsiao, I.T.: Characteristic patterns of inter-and intra-hemispheric metabolic connectivity in patients with stable and progressive mild cognitive impairment and alzheimer's disease. Scientific reports **8**(1), 13807 (2018)
9. Langa, K.M., Levine, D.A.: The diagnosis and management of mild cognitive impairment: a clinical review. Jama **312**(23), 2551–2561 (2014)
10. Liu, S., Zhang, B., Fang, R., Rueckert, D., Zimmer, V.A.: Dynamic graph neural representation based multi-modal fusion model for cognitive outcome prediction in stroke cases. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 338–347. Springer (2023)
11. Lowe, V.J., Peller, P.J., Weigand, S.D., Montoya Quintero, C., Tosakulwong, N., Vemuri, P., Senjem, M.L., Jordan, L., Jack Jr, C.R., Knopman, D., et al.: Application of the national institute on aging-alzheimer's association ad criteria to adni. Neurology **80**(23), 2130–2137 (2013)
12. Ocasio, E., Duong, T.Q.: Deep learning prediction of mild cognitive impairment conversion to alzheimer's disease at 3 years after diagnosis using longitudinal and whole-brain 3d mri. PeerJ Computer Science **7**, e560 (2021)

13. Wang, C., Lei, Y., Chen, T., Zhang, J., Li, Y., Shan, H.: Hope: Hybrid-granularity ordinal prototype learning for progression prediction of mild cognitive impairment. IEEE Journal of Biomedical and Health Informatics (2024)
14. Zhang, H., Song, R., Wang, L., Zhang, L., Wang, D., Wang, C., Zhang, W.: Classification of brain disorders in rs-fmri via local-to-global graph neural networks. IEEE Transactions on Medical Imaging **42**(2), 444–455 (2022)
15. Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D., Long, X.: A 3d densely connected convolution neural network with connection-wise attention mechanism for alzheimer's disease classification. Magnetic Resonance Imaging **78**, 119–126 (2021)
16. Zheng, G., Zhang, Y., Zhao, Z., Wang, Y., Liu, X., Shang, Y., Cong, Z., Dimitriadis, S.I., Yao, Z., Hu, B.: A transformer-based multi-features fusion model for prediction of conversion in mild cognitive impairment. Methods **204**, 241–248 (2022)
17. Zhu, W., Sun, L., Huang, J., Han, L., Zhang, D.: Dual attention multi-instance deep learning for alzheimer's disease diagnosis with structural mri. IEEE Transactions on Medical Imaging **40**(9), 2354–2366 (2021)