Dynamic Graph Neural Representation Based Multi-modal Fusion Model for Cognitive Outcome Prediction in Stroke Cases

Shuting Liu[1(✉)], Baochang Zhang[1], Rong Fang[2], Daniel Rueckert[1,3,4], and Veronika A. Zimmer[1,3]

[1] School of Computation, Information and Technology, Technical University of Munich, Munich, Germany
shuting.liu@tum.de

[2] Institute for Stroke and Dementia Research (ISD), LMU University Hospital, LMU Munich, Munich, Germany

[3] School of Medicine, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

[4] Department of Computing, Imperial College London, London, UK

Abstract. The number of stroke patients is growing worldwide and half of them will suffer from cognitive impairment. Therefore, the prediction of Post-Stroke Cognitive Impairment (PSCI) becomes more and more important. However, the determinants and mechanisms of PSCI are still insufficiently understood, making this task challenging. In this paper, we propose a multi-modal graph fusion model to solve this task. First, dynamic graph neural representation is proposed to integrate multi-modal information, such as clinical data and image data, which separates them into node-level and global-level properties rather than processing them uniformly. Second, considering the variability of brain anatomy, a subject-specific undirected graph is constructed based on the connections among 131 brain anatomical regions segmented from image data, while first-order statistical features are extracted from each brain region and internal stroke lesions as node features. Meanwhile, a novel missing information compensation module is proposed to reduce the impact of missing or incomplete clinical data. In the dynamic graph neural representation, two kinds of attention mechanisms are used to encourage the model to automatically localize brain anatomical regions that are highly relevant to PSCI prediction. One is node attention established between global tabular neural representation and nodes, the other is multi-head graph self-attention which changes the static undirected graph into multiple dynamic directed graphs and optimizes the broadcasting process of the graph. The proposed method studies 418 stroke patients and achieves the best overall performance with a balanced accuracy score of 79.6% on PSCI prediction, outperforming the competing models. The code is publicly available at github.com/fightingkitty/MHGSA.

Keywords: Post-stroke cognitive impairment · Graph neural representation · Fusion method

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14227, pp. 338–347, 2023.
https://doi.org/10.1007/978-3-031-43993-3_33

1 Introduction

Post-Stroke Cognitive Impairment (PSCI) is common following stroke, and more than half of people will suffer from PSCI in the first year after stroke. Therefore, predicting cognitive impairment after stroke is an important task [2,5]. However, the determinants and mechanisms of PSCI are insufficiently understood, and doctors cannot make a clear risk prediction based on single factors, e.g., the locations of infarct and White Matter Hyperintensities (WMH) lesions, resulting in frequent underdiagnosis of PSCI [16,17]. Therefore, accurate prediction of PSCI becomes challenging and critical for devising appropriate post-stroke prevention strategies.

In recent years, Convolutional Neural Networks (CNN) have shown promising performance in brain disease diagnosis and prediction [4]. Compared with traditional machine learning methods [14], CNNs excel at extracting high-level information about neuroanatomy from Magnetic Resonance Imaging (MRI) [11]. However, brain MRIs provide only a partial view of the underlying changes that lead to cognitive decline, and the extracted features using CNNs are difficult to interpret by clinicians. Hence, some studies use a pre-defined set of image features, like cortical features, extracted from MRIs via Freesurfer [18]. However, the aforementioned approaches all focus on neuroimaging data, ignoring the importance of clinical data such as patient demographics, family history, or laboratory measurements. Some successful multi-modal fusion models have been proposed and achieved some improvements in predictive tasks, but image and clinical data are only integrated through simple concatenation or affine transformations [3,12], without considering the specificity of patient brain anatomy [10].

To overcome the aforementioned issues of existing methods, we propose a novel multi-modal fusion model to solve this challenging task. Our main contributions are as follows: (1) Both MRI data and clinical data are studied in this task, where subject-specific anatomy, infarcts localization and WMH distribution are explored from MRIs, while clinical (tabular) data provides personal information and clinical indicators such as stroke history and stroke severity. (2) Dynamic graphs neural representation is first proposed for PSCI prediction, which can not only integrate multi-modal information but also take subject-specific brain anatomy into account. An effective missing information compensation module is proposed to reduce the impact of incomplete clinical data. (3) A detailed ablation study is conducted to verify the effectiveness of each proposed module. The proposed method outperforms competing models by a large margin. Additionally, the Top-15 brain structural regions strongly associated with PSCI are uncovered by the proposed method, which further emphasizes the relevance of the proposed method.

2 Methodology

Our approach consists of three steps (illustrated in Fig. 1): (i) graph construction and node feature extraction, (ii) missing information compensation module, and (iii) dynamic graph neural representation.

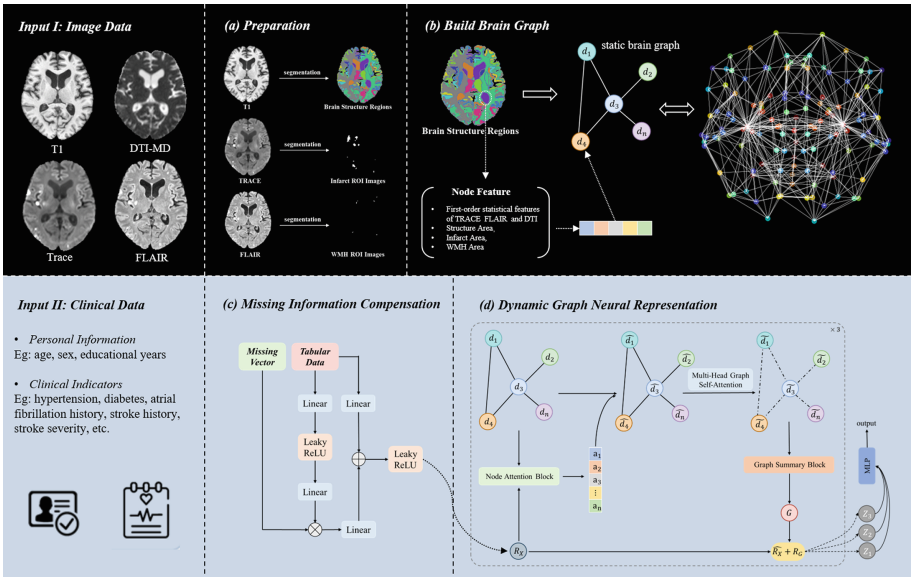

Fig. 1. The overview of our proposed method. (a) Data preparation, (b) graph building and node feature extraction, (c) missing information compensation module, (d) and dynamic graph neural representation.

2.1 Graph Construction and Node Feature Extraction

We represent the nodes of the graph with 131 structural brain regions obtained from the Hammers' brain atlas [7,9] (more details are given in Sect. 3.1). The edges of the graph are established based on the neighboring relationship of each brain structural area. Specifically, if two areas are adjacent, we create an edge between the corresponding nodes; otherwise, no edge is present. To each node, we assign a total of eighteen features, comprising fifteen first-order statistical features (maximum, minimum, median, mean, and variance) extracted from the corresponding brain structural regions in Trace Diffusion-Weighted MRI (DW-MRI) and Mean Diffusivity (MD) DW-MRI and FLAIR MRI, along with three stroke lesion features: structural volume, infarct volume, and WMH volume.

2.2 Missing Information Compensation Module

When dealing with clinical data, some patients may not have complete records, resulting in missing information. To mitigate the impact of missing data, we propose a module shown in Fig. 1(c). Rather than using a naive approach of filling in missing elements with an average value or zero, we suggest learning the offset value caused by missing elements based on the known information to compensate for the missing data. We denote a patient's 1D tabular record as $X = \{x_i\}$, and define a vector V with the same size as X. In V, bits for missing

elements are represented as '1', and the remaining bits are '0'. We denote the set of indices of missing elements as $M = \{m|v_m = 1\}$. As a complete X is processed by a linear layer, each element $x_i^{'}$ can be expressed as

$$x_i^{'} = \sum_j w_{ij} x_j + b_i \tag{1}$$

When the value of some input values $x_m, m \in M$ is unknown due to missing tabular data, we can only compute $x_i^{''} = \sum_{j \notin M} w_{ij} x_j + b_i$ and the difference $x_i^{'} - x_i^{''}$ is the impact the missing tabular data can have. We propose a module to learn this difference to mitigate the effect of missing values by

$$\phi_2 \left(\phi_1(X) \bullet V\right) = \sum_m w_{im} x_m \qquad \forall m \in \ M \tag{2}$$

Here, ϕ_1 represents two normal linear layers, ϕ_2 is a linear layer without considering the bias, ϕ_3 represents a normal linear layers, $LReLU$ is a leaky rectified linear activation, and $\bullet$ is the element-wise multiplication operator. When some elements are missing, this module will estimate the offset; and when processing a complete clinical record, the output of Eq. (2) should always be zero. Hence, the neural representation of each patient's clinical tabular record will be calculated using Eq. (3).

$$R_X = LReLU \left(\phi_2 \left(\phi_1(X) \bullet V\right) + \phi_3(X)\right) \tag{3}$$

2.3 Dynamic Graph Neural Representation

We regard the tabular's neural representation R_X as global information and the node feature d_i in each subject-specific graph $G = \{D, E|d_i \in D\}$ as local information (here, E denotes the adjacency matrix): First, we set up an attention mechanism between R_X and d_i, which will capture the contribution of each node and enhance node feature. Then, a multi-head graph self-attention is introduced for this task, which pays attention to the weight of edges in the graphs. It transforms the static graph to several dynamically subgraphs, influencing the broadcasting process of the graph. Next, we use graph summary block which consists of a normal linear layer and a leaky rectified linear unit activation function (LReLU), to summarize the graph G and output a graph-based global feature R_G, i.e., graph's neural representation. Finally, we fuse the graph's neural representation R_G and the updated tabular's neural representation $\hat{R_X}$ via concatenation, which can be used as the input of a multilayer perceptron (MLP) for classification as shown in Fig. 1. Here, focal loss function is employed, which is formulated as,

$$FL\left(p_t\right) = -\alpha \left(1 - p\right)^{\gamma} log\left(p\right) \tag{4}$$

Here p is the predicted probability of the correct class, $\alpha = 0.2$ is the weighting factor for each sample, $\gamma = 0.05$ is a tunable focusing parameter.

Node Attention Block. This is a parameterized function that learns the mapping between query ($q_t \in \mathbb{R}^{1\times C}$) coming from the tabular's neural representation R_X, and the corresponding key ($k_g \in \mathbb{R}^{N\times C}$) and value ($v_g \in \mathbb{R}^{N\times C}$) representations in graph G. Here, N is the number of nodes, C is the channel of node feature. Hence, the node attention block fuses information from tabular (global level) and image data (local/node level) by measuring the correlation between q_t and k_g, and the attention weight ($A_1 \in \mathbb{R}^{1\times N}$) is computed as follow,

$$A_1 = softmax\left(\frac{q_t k_g^T}{\sqrt{C}}\right) \tag{5}$$

Using the computed attention weight, the output of the node attention block is computed to update graph G' as,

$$G^{'} = (1 + A_1^T)v_g \tag{6}$$

Multi-head Graph Self-attention Block. Inspired by [15] and in order to optimize the broadcasting of the graph, the multi-head self-attention is proposed to learn the edge weights during training. Following with the node attention, it is also a parameterized function that learns the mapping between the reused key ($k_g \in \mathbb{R}^{N\times C}$) and the query ($q_g \in \mathbb{R}^{N\times C}$) representation in graph G ($G \in \mathbb{R}^{N\times C}$). Then, the shapes of query q_g, key k_g and updated graph feature G' are reshaped as $[h, N, C/h]$. Here, h is the number of heads. The attention weight ($A_2 \in \mathbb{R}^{h\times N\times N}$) is computed by measuring the similarity between q_g and k_g according to,

$$A_2 = softmax\left(\frac{q_g k_g^T}{\sqrt{\frac{C}{h}}}\right) \tag{7}$$

Using the computed attention weight and considering the fixed adjacency matrix E, the output of graph multi-head self-attention block $G'' \in \mathbb{R}^{h\times N\times C/h}$ is computed and reshaped as $[N, C]$ for the preparation of starting the next graph neural representation unit.

$$G^{''} = E \cdot A_2 G^{'} \tag{8}$$

3 Materials and Experiments

3.1 Data and Preparation

Our study utilized clinical data from 418 stroke patients, collected within five days of the onset of their acute stroke symptoms, and consisted of both clinical data and image data. Each clinical tabular records contains 20 variables, including 3 variables of basic personal information (age, sex, education) and 17 variables of clinical indicators (i.e., smoking, drinking, hypertension, diabetes, atrial fibrillation, previous stroke, BMI, circulating low-density lipoprotein cholesterol levels, stroke severity, pre-stroke function, cognitive impairment

in the acute post-stroke phase, stroke lesion volume, lacunes number, Fazekas score of WMH, Fazekas score of WMH in deep white matter, cerebral microbleeds number, perivascular spaces level). Patients were followed up, and clinical experts classified them as either PSCI negative or PSCI positive based on detailed neuropsychological tests conducted 12 months after stroke onset. Out of the 418 patients, there were 332 positive cases and 86 negative cases. All studies were approved by the local ethics committees. We split the patients into two groups with an equal proportion of PSCI candidates for five-fold cross validation, 80% for training and 20% for testing. In the preparation step, the MRIs are rigidly aligned using [1]. Based on each patient's T1 data, the whole brain was segmented into 131 structural regions using MALPEM [7,9], from which we constructed the subject-specific brain graphs. Meanwhile, based on each patient's Trace DW-MRI and FLAIR MRI, the infarct lesions and WMH lesions were semi-automatically segmented and manually corrected by a clinical expert.

3.2 Evaluation Measures

To validate the performance of our method, we employed five measures including Balanced Accuracy (BAcc), Classification Accuracy (Acc), Precision (Pre), Sensitivity (Sen), Specificity (Spe), and the area under the receiver operating characteristic curve (AUC). Among them, BAcc is a performance metric used to evaluate the effectiveness of a model on imbalanced datasets.

3.3 Experimental Design

To validate the performance of our method, we perform a series of experiments on our in-house dataset. Based on the multi-modal's information (i.e., clinical data and image data), we designed three experiments with different data inputs. (I) By only using the clinical data as input, the effectiveness of missing information compensation module is studied. (II) By only using the image-based features as input, the importance of graph-based analysis for PSCI prediction is investigated. (III) Using both as input, an ablation experiment is performed, where we learn the evolution process and the superiority of our proposed fusion model.

The proposed model is implemented on PyTorch library (version 1.13.0) with one NVIDIA GPU (Quadro RTX A6000), and trained for 400 epochs with a batch size of 32. Adam optimizer is employed with weight decay of 0.008, an initial learning rate of 0.0005; and CosineAnnealingLR scheduling technique is used to adjust the learning rate with the maximum number of iterations of 10.

4 Results

Table 1 summarizes the results of the five-fold cross-validation experiments. We used different models for the experiments, where $\boldsymbol{A}$ represents an MLP model, $\boldsymbol{B}$ represents an MLP model with the proposed missing information compensation module shown in Fig. 1(c), $\boldsymbol{C}$ represents another MLP model, $\boldsymbol{D}$ represents a

Graph Convolutional Network (GCN) [8], ***B***+***D*** represents a fusion model by concatenating the features from ***B*** and ***D***, ***B***+***E*** represents a fusion model by concatenating the features from ***B*** and ***E***, where ***E*** is a GCN with a multi-head graph self-attention block as explained in Sect. 2.3. Finally, ***F*** represents our proposed method as shown in Fig. 1.

Table 1. Experimental results of different methods. The input data is specified by *I* for clinical data only, *II* for imaging data only and *III* for both; **bold values** indicate best performance over all.

Exp.	Models	BAcc		Acc		AUC		Pre		Sen		Spe	
		mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
I	***A***	0.688	0.066	0.695	0.033	0.732	0.071	0.367	0.052	0.675	0.134	0.700	0.032
	B	0.709	0.090	0.689	0.060	0.741	0.072	0.372	0.076	**0.745**	0.162	0.674	0.056
II	***C***	0.671	0.035	0.785	0.027	0.648	0.064	0.493	0.089	0.477	0.093	**0.865**	0.046
	D	0.700	0.054	0.738	0.108	0.711	0.080	0.461	0.117	0.634	0.129	0.765	0.159
III	***B***+***D***	0.742	0.035	0.796	0.028	0.746	0.057	0.508	0.053	0.649	0.058	0.835	0.028
	B+***E***	0.761	0.055	0.785	0.034	0.770	0.067	0.490	0.055	0.721	0.133	0.802	0.053
	F	**0.796**	0.049	**0.832**	0.038	**0.800**	0.083	**0.580**	0.083	0.734	0.086	0.858	0.041

4.1 Effectiveness of Missing Information Compensation

To assess the efficacy of compensating for missing information, a standard MLP model is trained solely on clinical data with missing elements filled in with zeros. Subsequently, the same MLP model is trained with a missing information compensation module, utilizing the identical training set. Comparing the performances of model ***A*** and ***B***, it is obvious that the proposed missing information compensation module plays an important role, achieving a BAcc of 0.709 ± 0.054 and AUC score of 0.741 ± 0.072, outperforming Method ***A***.

4.2 Importance of Graph-Based Analysis

To investigate the importance of graph-based analysis for this task, an MLP model embedded with GCNs is trained only on imaging-based features, and the comparison method is a standard MLP model. Comparing the performances of model ***C*** and ***D***, it is clear that the suggested graph-based model makes an notable improvement on both BAcc and AUC, with the rise of BAcc from 0.671 to 0.700 and of Auc from 0.648 to 0.711, which demonstrates that the consideration of subject-specific brain anatomy is very important for PSCI prediction.

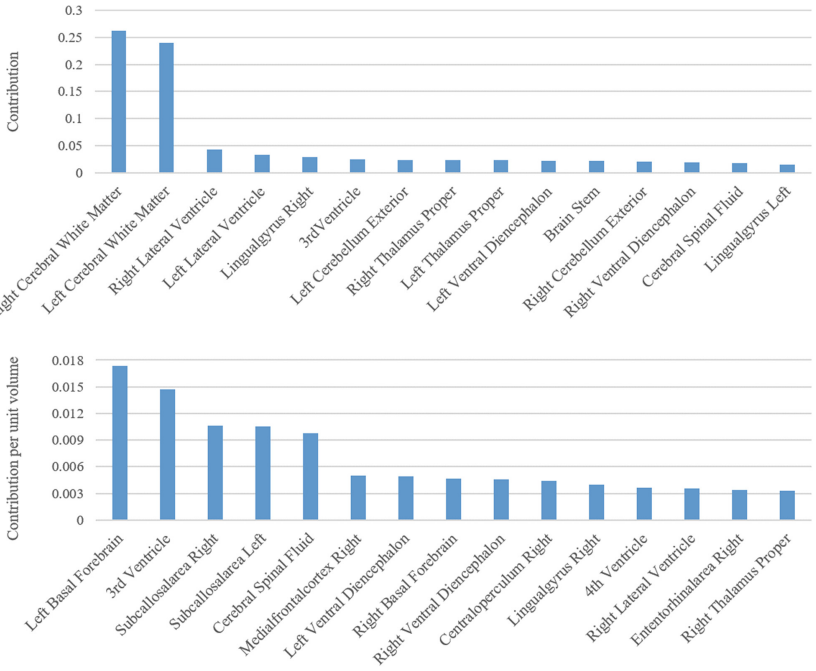

Fig. 2. The Top-15 brain structural regions ranking of contribution (top) and the Top-15 brain structural regions ranking of contribution per unit volume (bottom) that are strongly associated with PSCI discovered by our method.

4.3 Superiority of Proposed Fusion Model

Finally, we explore how to effectively fuse clinical data and image data for PSCI prediction. As a baseline approach, we directly concatenate the features learned from model $\boldsymbol{B}$ and $\boldsymbol{D}$. The first evolutionary process is to upgrade static brain map analysis to dynamic brain map analysis, thus the performance of model $\boldsymbol{B}$+$\boldsymbol{E}$ is explored, where the proposed multi-head graph self-attention block is first embedded into the GCN layers. Then, a better fusion pattern is discovered, i.e., dynamic graph neural representation in our final proposed method. The experimental results demonstrate that $\boldsymbol{B}$+$\boldsymbol{E}$ model results in a higher BAcc of 0.761 compared to baseline. Furthermore, our final proposed method outperforms all other methods, achieving a BAcc as high as 0.796 and an AUC of 0.800. To explore the contribution of each brain structural region in this task, we first average the attention map A_1 across all subjects and visualize it, as illustrated in Fig. 2(top). Additionally, to adjust for the influence of volume, we compute the contribution per unit volume for each region by dividing by its corresponding volume, as illustrated in Fig. 2(bottom). Our observations indicate that the network appears to concentrate on clinically significant regions that impact cognition, including the Cerebral White Matter, Basal Forebrain and 3rd Ventricle

[6,13]. Nevertheless, we have also identified some ambiguous relationships that necessitate further investigation.

5 Conclusion and Discussion

In this study, we tackle the challenge of predicting PSCI using both brain MRI and clinical non-imaging data. Our proposed method utilizes a dynamic graph neural representation that fully leverages the structural information of the brain. By incorporating node attention and self-attention, we effectively merge clinical tabular and image information. Our approach achieves a BAcc of 0.796 on the test dataset. Moreover, visual attention maps allow us to identify the contributions of different brain structures to this prediction task. The results of our study are consistent with prior medical research. However, there are certain brain regions that require further exploration by clinicians. In addition, there are several areas our proposed method is highly clinically relevant: (1) Risk prediction to identify (i) low-risk patients, and (ii) high-risk patients (for those early and targeted treatment and rehabilitation is recommended). (2) Detailed phenotyping and characterization of high-risk patients and follow-up. (3) Identification of novel biomarkers and risk factors.

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

References

1. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ANTS). Insight j **2**(365), 1–35 (2009)
2. Ball, E.L., et al.: Predicting post-stroke cognitive impairment using acute CT neuroimaging: a systematic review and meta-analysis. Int. J. Stroke **17**(6), 618–627 (2022)
3. Binzer, M., Hammernik, K., Rueckert, D., Zimmer, V.A.: Long-term cognitive outcome prediction in stroke patients using multi-task learning on imaging and tabular data. In: Rekik, I., Adeli, E., Park, S.H., Cintas, C. (eds.) PRIME 2022. LNCS, vol. 13564, pp. 137–148. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16919-9_13
4. Ebrahimighahnavieh, M.A., Luo, S., Chiong, R.: Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. Comput. Methods Programs Biomed. **187**, 105242 (2020)
5. Georgakis, M.K., et al.: Cerebral small vessel disease burden and cognitive and functional outcomes after stroke: a multicenter prospective cohort study. Alzheimer's Dementia **19**(4), 1152–1163 (2023)
6. Grothe, M., et al.: Reduction of basal forebrain cholinergic system parallels cognitive impairment in patients at high risk of developing Alzheimer's disease. Cereb. Cortex **20**(7), 1685–1695 (2010)
7. Heckemann, R.A., et al.: Brain extraction using label propagation and group agreement: pincram. PLoS ONE **10**(7), e0129211 (2015)

8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
9. Ledig, C., et al.: Robust whole-brain segmentation: application to traumatic brain injury. Med. Image Anal. **21**(1), 40–58 (2015)
10. Lim, J.S., Lee, J.J., Woo, C.W.: Post-stroke cognitive impairment: pathophysiological insights into brain disconnectome from advanced neuroimaging analysis techniques. J. Stroke **23**(3), 297–311 (2021)
11. Liu, M., et al.: A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. Neuroimage **208**, 116459 (2020)
12. Pölsterl, S., Wolf, T.N., Wachinger, C.: Combining 3D image and tabular data via the dynamic affine feature map transform. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 688–698. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_66
13. Ray, N.J., et al.: Cholinergic basal forebrain structure influences the reconfiguration of white matter connections to support residual memory in mild cognitive impairment. J. Neurosci. **35**(2), 739–747 (2015)
14. Uysal, G., Ozturk, M.: Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods. J. Neurosci. Methods **337**, 108669 (2020)
15. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
16. Verdelho, A., et al.: Cognitive impairment in patients with cerebrovascular disease: a white paper from the links between stroke ESO dementia committee. Eur. Stroke J. **6**(1), 5–17 (2021)
17. Weaver, N.A., et al.: Strategic infarct locations for post-stroke cognitive impairment: a pooled analysis of individual patient data from 12 acute ischaemic stroke cohorts. Lancet Neurol. **20**(6), 448–459 (2021)
18. Zheng, G., et al.: A transformer-based multi-features fusion model for prediction of conversion in mild cognitive impairment. Methods **204**, 241–248 (2022)