Machine Intelligence Research
www.mi-research.net

21(1), February 2024, 136-152
DOI: 10.1007/s11633-023-1442-8

Multimodal Fusion of Brain Imaging Data: Methods and Applications

Na Luo[1*] Weiyang Shi[1*] Zhengyi Yang[1] Ming Song[1] Tianzi Jiang[1,2,3,4]

[1] Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

[3] Center for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[4] Research Center for Augmented Intelligence, Zhejiang Laboratory, Hangzhou 311100, China

Abstract: Neuroimaging data typically include multiple modalities, such as structural or functional magnetic resonance imaging, diffusion tensor imaging, and positron emission tomography, which provide multiple views for observing and analyzing the brain. To leverage the complementary representations of different modalities, multimodal fusion is consequently needed to dig out both inter-modality and intra-modality information. With the exploited rich information, it is becoming popular to combine multiple modality data to explore the structural and functional characteristics of the brain in both health and disease status. In this paper, we first review a wide spectrum of advanced machine learning methodologies for fusing multimodal brain imaging data, broadly categorized into unsupervised and supervised learning strategies. Followed by this, some representative applications are discussed, including how they help to understand the brain arealization, how they improve the prediction of behavioral phenotypes and brain aging, and how they accelerate the biomarker exploration of brain diseases. Finally, we discuss some exciting emerging trends and important future directions. Collectively, we intend to offer a comprehensive overview of brain imaging fusion methods and their successful applications, along with the challenges imposed by multi-scale and big data, which arises an urgent demand on developing new models and platforms.

Keywords: Multimodal fusion, supervised learning, unsupervised learning, brain atlas, cognition, brain disorders.

Citation: N. Luo, W. Shi, Z. Yang, M. Song, T. Jiang. Multimodal fusion of brain imaging data: Methods and applications. *Machine Intelligence Research*, vol.21, no.1, pp.136–152, 2024. http://doi.org/10.1007/s11633-023-1442-8

1 Introduction

Neuroimaging provides a means for identifying and measuring the structure and function of the brain. Different non-invasive imaging measurements reveal different characteristics of the nervous system, e.g., architecture, activation, or structural and functional connectivity. Magnetic resonance imaging (MRI) is one of the most important neuroimaging technologies and widely used in neuroscience research and clinical settings. Structural MRI (sMRI) provides information about the tissue type of the brain. Functional MRI (fMRI) measures the hemodynamic response related to neural activity in the brain dynamically. Diffusion-weighted imaging (DWI) can additionally provide information on structural connectivity among brain regions. Typically, these data are analyzed separately in a single-modality fashion. While more recently, collecting multiple types of brain data from the same individual using various imaging techniques has become common practice[1]. Compared to single modality, the fusion of multiple modalities, which may capture cross-modal (both shared and complementary) information, is envisioned to provide more insights into the underlying problem.

The common goal of data fusion is to maximally dig out the joint information shared among modalities as well as the modality-specified complementary information. The past decades have witnessed significant improvements in learning-based fusion methods. Based on whether labels are used to guide the learning process, we subdivide the existing fusion technologies into unsupervised and supervised learning methods. The objective function for the supervised learning method is obvious and consistent to learn the mapping between input and labels, then the joint representations were optimized through reducing the difference between predicted labels and true la-

Review
Manuscript received on December 16, 2022; accepted on March 23, 2023
Recommended by Associate Editor Dao-Qiang Zhang
Colored figures are available in the online version at https://link.springer.com/journal/11633
* These authors contribute equally to this work
© The Author(s) 2024

bels. While for the unsupervised learning strategy, we further subdivide it into three categories according to different objective functions, correlation-based fusion, multi-view clustering and data reconstruction. Advanced methods in each category will be systematically reviewed later. The conventional fusion methods are commonly emphasized to maximally exploit the shared representation, whereas the modality-specified complementary information is often underutilized. Therefore, the variants of those methods towards exploring the complementary information will be discussed along the way.

Carrying out multimodal fusion analysis benefits a cumulative understanding of the complex brain networks on different temporal and spatial scales. First, the brain atlas is a prerequisite for studying brain networks, which plays a central role in neuroscience and clinical practice[2]. Though many extensively applied brain atlases segregate brain into distinct brain areas primarily by a single modality (cytoarchitecture, topography, function or connectivity), a series of recent studies shed light on more stable boundaries delineated by fusing various modalities[3, 4]. More importantly, constructing a reference brain atlas paves the way to fuse a large scale of information spanning from genes, proteins, synapses and neurons to areas, pathways and the whole brain, which provides the possibility to comprehensively explore neuroscience issues for both healthy development and clinical pathology via data fusion technology. Furthermore, the exploration of the mystery of cognition and development has always been a core topic in the field of neuroscience. Using multimodal data has achieved significantly higher accuracies compared to using unimodal data when predicting individuals′ behaviors and intelligence quotient scores in recent studies[5, 6]. Last but not least, encouraging efforts have been devoted to early diagnosis and prognosis of psychiatric disorders via multimodal fusion methods. During the period of growth, psychiatric symptoms frequently emerge with complex reasons. Typically, psychiatric disorders develop with a long process, imposing a great socioeconomic burden. Consequently, increasing attention is focused on detecting early abnormalities[7, 8], exploring potential subtypes[9, 10], as well as revealing possible neuroimaging biomarkers for predicting treatment outcomes.

In this review, four interrelated topics are covered as shown in Fig. 1: 1) Methodologies, which summarize the representative multimodal brain imaging fusion technologies in recent years; 2) Atlasing via multimodal brain imaging, which reviews brain parcellations at both macro-level and micro-level based on information of anatomical structure, function activation, connectivity or multiple modalities; 3) Multimodal fusion in studying cognition and development, which includes representative applications on how multimodal fusion methods help improve the prediction and understanding of behavioral phenotype and brain aging; 4) Multimodal fusion in brain disorders, which elaborates important applications on how multimodal fusion helps accelerate the exploration of underlying biological mechanisms of brain diseases.

2 Multimodal fusion methods

As brain imaging is often with three-dimension (3D) or higher dimension, it is difficult to determine the linkages via computing simple correlation. To effectively fuse multimodal data, various machine learning methodologies have been proposed. The common pipeline is to first

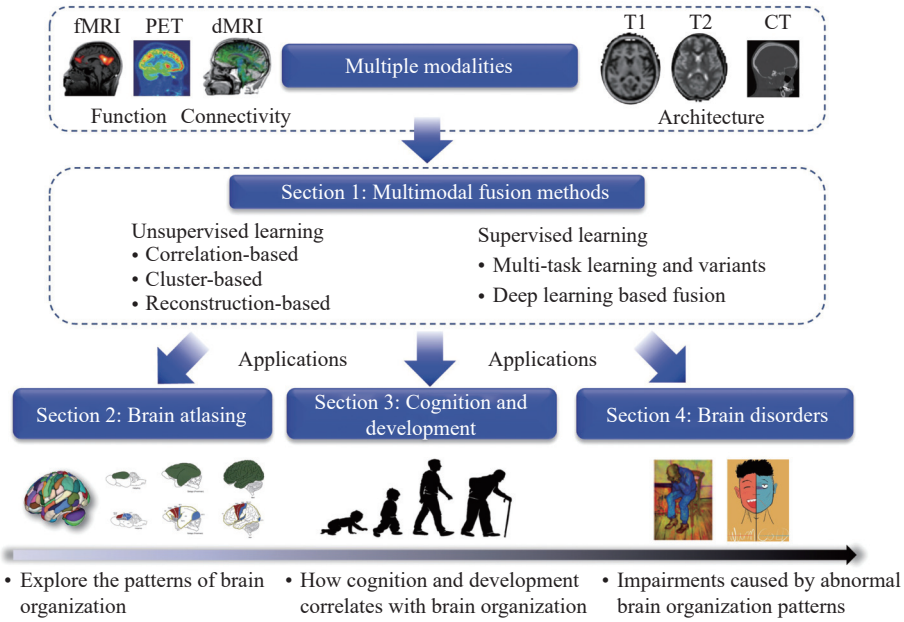

Fig. 1 Four interrelated topics covered in this review

transfer the high-dimension images to a 2D matrix. Supervised or unsupervised strategies are then adopted to reduce the dimension of the 2D matrix of different modalities to a common latent space. The inner associations between modalities are then explored in the latent space. In this section, we review some important advances of each category that have been successfully used for brain imaging data fusion (Fig. 2).

2.1 Unsupervised learning

Unsupervised learning is to discover latent representations and disentangle explanatory factors from rich and unlabelled data. It does not receive any kind of supervision from the target outputs (or labels) to guide the learning process. Based on the objective functions, unsupervised fusion approaches could be further differentiated as 1) correlation-based, 2) cluster-based, and 3) data reconstruction-based fusion. Correlation-based methods are particularly suitable for revealing the joint patterns shared among modalities. These methods can typically be combined with other algorithms for tasks such as prediction, clustering, or other applications[7]. Clustering-based methods are effective at exploring the clusters shared across different modalities, which have been successfully used for revealing subtypes of disease[11]. Data reconstruction-based methods perform well when the two modalities are across multiscales, i.e., when they have different resolutions or signal forms (such as imaging and data series)[12].

2.1.1 Correlation-based fusion

CCA and variants

Canonical correlation analysis (CCA) is a typical subspace learning approach that aims to find pairs of projections for different views with maximized linear correlations[13]. This method decomposes each dataset (X_1, X_2) into a set of components and their corresponding mixing profiles (W_1, W_2) that maximize inter-subject covariation across two datasets as (1).

$$\max_{W_1,W_2} W_1^{\mathrm{T}} X_1^{\mathrm{T}} X_2 W_2 \tag{1}$$

$$\text{s.t.}\ \ W_1^{\mathrm{T}} X_1^{\mathrm{T}} X_1 W_1 = W_2^{\mathrm{T}} X_2^{\mathrm{T}} X_2 W_2 = 1. \tag{2}$$

It can be extended to multi-set CCA (mCCA) to incorporate more than two modalities[14]. The largest shortcoming of CCA is that it only analyzes linear relationships between modalities. Therefore, sparse CCA and kernel CCA have been developed and applied in brain imaging analysis[15–17]. Sparse CCA are developed by introducing the sparse penalties into the traditional CCA model, mostly using the $l1$ norm (CCA-$l1$) or the combination of $l1$ and $l2$ norm (CCA-elastic net) penalties[15]. Group sparse CCA extends to explore the correlation of the group structure between the two modalities[16]. Kernel CCA employs feature mappings induced by positive-definite kernels[17]. Another drawback of CCA is that the intra-modality independence is overlooked. To extract complementary information between modalities, Sui et al.[18] then proposed to combine independent component analysis (ICA) with mCCA, where the output spatial components of mCCA were further concatenated and decomposed using ICA. This method allows both highly and weakly connected modulations as well as joint independent components. To uncover the neurocognitive mapping of specific clinical measurements, the prior information has further been employed as a reference to guide the multimodal data fusion process[19].

ICA and variants

Independent component analysis (ICA) discovers hidden features from observations which are assumed to be linear mixtures of independent sources. Joint ICA (jICA) is a second-level fMRI analysis method that assumes two or more features (modalities) share the same mixing matrix and maximizes the independence among joint components[20]. This method maximizes the independence among the concatenated multimodal features but generates the same mixing matrix for all modalities. A general frame-

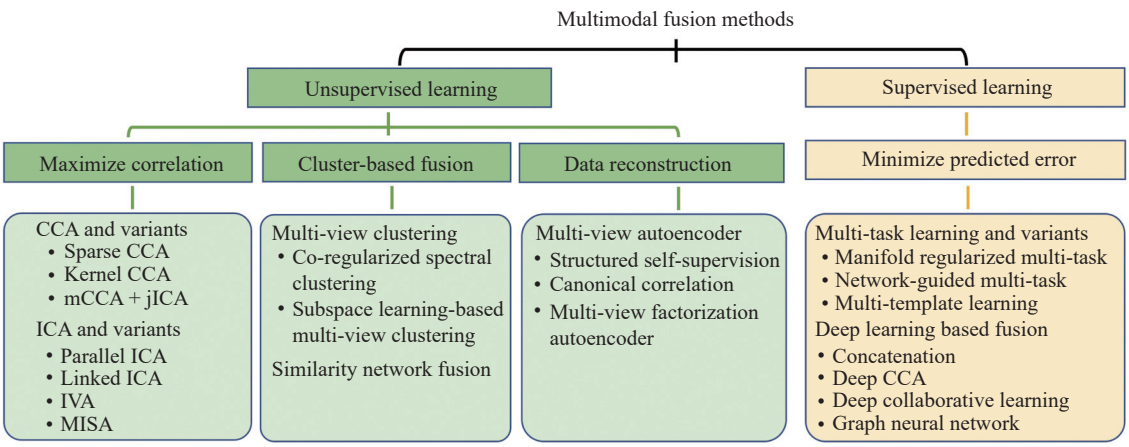

Fig. 2 Advanced multimodal fusion methods

work named disjoint subspace analysis using ICA (DS-ICA) was developed, which identifies and extracts not only the common but also the distinct components across multiple datasets[21]. The main idea for DS-ICA is to identify and split the common and distinct subspaces from the modalities and perform separate analyses.

Also, the strong regularization imposed by the jICA framework can be relaxed in a number of ways to allow for more flexibility in the estimation. One such approach is called parallel ICA. This method builds upon the blind matrix factorization techniques used in ICA to simultaneously extract latent statistically independent components and jointly identify mutual relationships between modalities[22]. It has also been generalized to include three modalities[23]. Group ICA has further been incorporated into parallel ICA to fuse the first-level 4D fMRI data[24]. Linked ICA is a probabilistic approach based on a modular Bayesian framework, which is designed for simultaneously modeling and discovering common characteristics across multiple modalities[25]. Linked ICA automatically determines the optimal weighting of each modality, and also can detect single-modality structured components when present. To face a computational challenge for thousands of subjects, more quick fusion strategies based on linked ICA have been proposed[26, 27].

Independent vector analysis (IVA), a multidataset extension of ICA, provides a natural and extendable way to directly link multivariate brain imaging data together. IVA extends the ICA model to multiple datasets, assuming a linear mixture of independent sources for each dataset. This collection of linked sources is defined as the source component vector (SCV)[28]. As a more general model, multidataset independent subspace analysis (ISA) solves multiple blind source separation problems (including ICA, IVA, ISA, and more) under the same framework with remarkable performance and improved robustness even at low signal to noise ratio (SNR)[29]. In addition, Avants et al.[30] also proposed a general fusion framework named similarity-driven multi-view linear reconstruction (SiMLR), which combines features from CCA, ICA and singular value decomposition (SVD) in an accessible and flexible joint dimensionality reduction algorithm.

2.1.2 Clustering-based fusion

Multi-view clustering and variants

For the multi-view clustering problem, the general assumption is that different views admit same underlying clustering of the data, so the goal of this kind of methods is to look for clusters that are consistent across the views. Among all the popular techniques, spectral clustering has gained considerable attention due to its good performance on arbitrary shaped clusters and well-defined mathematical framework. Existing multi-view clustering methods are often conducted based on the similarity matrix. One of the most representative multi-view spectral clustering methods is co-regularized spectral clustering[31]. It combines co-regularization with existing spectral clustering approaches to make the clustering hypotheses on different views agree with each other. For example, after calculating the similarity matrix, the cost function is measured as the disagreement between the clustering of two views:

$$D\left(U^{(v)}, U^{(w)}\right) = \left\| \frac{K_{U(v)}}{\|K_{U(v)}\|_F^2} - \frac{K_{U(w)}}{\|K_{U(w)}\|_F^2} \right\|_F^2 \tag{3}$$

where $K_{U(v)}$ and $K_{U(w)}$ are the similarity matrices for the eigenvector matrix $U^{(v)}$ and $U^{(w)}$, respectively. $\|\cdot\|_F$ denotes the Frobenius norm of the matrix. The similarity matrices are generally learned by most existing methods, which cannot well characterize the intrinsic geometric structure and the neighbor relationship. To address this issue, Xie et al.[32] proposed a novel subspace learning-based multi-view clustering method, which learns similarity matrix adaptively from the learned latent representation by manifold learning.

Similarity network fusion

To create a comprehensive view of a disease given a cohort of patients, Wang et al.[11] proposed similarity network fusion, which computes and fuses similarity networks of patients obtained from each of their data types separately, taking advantage of the complementary in the data. A sample-by-sample similarity matrix for each data type was first constructed, which is equivalent to a similarity network. The nodes are samples and the weighted edges represent pairwise sample similarities. The network-fusion step uses a nonlinear method based on message-passing theory to integrate a set of biological graphs into a single network in an iterative manner. The advantage of this procedure is that weak similarities disappear, helping reduce the noise, and strong similarities present in one or more networks are added to the others. This method makes full use of a network′s local structure, integrating common as well as complementary information across networks.

2.1.3 Data reconstruction

Multi-view autoencoder and variants

Autoencoder (AE) is an artificial neural network designed to learn latent data representations in an unsupervised manner, which can optimally reconstruct the original data[33]. AEs are composed by an encoder, which transforms the input into a latent representation, and a decoder, which reconstructs the input from this representation[34]. AEs are trained to minimize the reconstruction error. It has been demonstrated the capacity of reducing dimensionality and mining latent features. Multi-view autoencoder learns a representation with multi-encoders and then uses the shared representation for reconstruction. To fuse more complementary information from multi-views, usually constraints will be added in the latent space. As an example, Bao et al.[12] used structured self-supervision learning to encourage the structure of each modality to

be maintained in the joint latent space. Then, the joint representation is optimized by a common self-reconstruction loss and a structured self-supervision loss. Many other widely used multi-view learning methods focus on improving model performance by effectively utilizing feature correlation among different views in the latent space. Moreover, to deal with the issue that many proposed multi-view learning methods overlooked "biologically meaningful" features, Ma and Zhang[35] incorporated biological interaction networks as an "external" domain knowledge source into the models through network regularization. This method provides a framework for unifying data-driven and knowledge-driven approaches for mining multiple data with biological knowledge.

2.2 Supervised learning

Supervised learning denotes the class of problems where the original data and its corresponding target predictions (or labels) are provided for the learning system. The goal is to learn the mapping between input and label, so that the system is capable of performing predictions for previously unseen input data points. The best joint multimodal features are selected through maximizing the classification/prediction accuracy.

Multi-task learning and variants

The classification of multimodality data is formulated as a multitask learning problem. Classifications with different modalities are regarded as different tasks. Assume that there are M different modalities (i.e., tasks). The classical multi-task model is to solve the following optimization problem:

$$\min_{W} \frac{1}{2} \sum_{m=1}^{M} \|y - X^{(m)} w^{(m)}\|_2^2 + \beta \|W\|_{2,1} \tag{4}$$

where $X^{(m)}$ denotes the input feature vector of the m-th modality for all subjects. y is the response vector from these subjects. $w^{(m)}$ is the regression coefficient vector for the m-th modality. The $l_{2,1}$-norm encourages these predictors from different modalities to share similar parameter sparsity patterns. Zhang and Shen[36] first proposed a multimodal multi-task learning method to jointly predict multiple variables from multimodal data. The variables include both the clinical variables used for regression and the categorical variable for classification. It is a general learning framework which includes two major steps, i.e., the multi-task feature selection and multimodal support vector machine.

In the classical multi-task learning model, only the relation between data and the response values is considered, which ignores the structural information of the data, leading to large deviations. With the expectation that similar subjects should have similar response values, Jie et al.[37] proposed a manifold regularized multi-task learning model, which added manifold regularizer to consider the subject-subject relation within each modality. To further add the important mutual relation of subjects between modalities, Xiao et al.[38] proposed a new manifold regularized multi-task learning model which effectively considers both the relation of subjects within the same modality and that between modalities.

Another important extension of multi-task learning is multi-template learning. Since subjects are often acquired from a wide range of patients and controls with different ages, ethnicities, races, etc., feature representations generated from a single template may not be sufficient to reveal the underlying complex differences. Therefore, researchers have proposed several multi-template-based methods to compare group differences more efficiently. For example, Liu et al.[39] nonlinearly register each brain′s MR images separately onto multiple preselected templates and then fused the extracted multiple sets of features under a multi-task sparse feature selection framework with a support vector machine (SVM) classifier. The classification in each template space is treated as a specific task. The final result is achieved through using an ensemble classification to combine outputs of all SVM classifiers.

Deep learning-based fusion

Along with the accumulation of big data from various consortia, the number of neuroimaging studies using deep learning models has rapidly grown since 2014[40]. Compared to standard machine learning methods, deep learning approaches are highly flexible and can learn multi-level nonlinear abstract representations of the data. The simplest way of fusing multi-modality data in a supervised manner is to concatenate data before sending them to classification model (a prefusion strategy) or after learning feature representations of each modality (a postfusion strategy). A prefusion strategy is easy to implement but has limitations when the feature dimensionality of one modality is much higher than the others or when the concatenation is infeasible because of the heterogeneity in data format[41]. Moreover, such strategies cannot effectively explore the correlations and complementary characteristics across different views, and their explicit physical meanings of different types of features and biomarkers are not fully considered in the feature learning phase, resulting in great information loss[42]. Compared to prefusion, a postfusion framework is more flexible when dealing with diverse modalities but more laborious in finding the optimal architectures and hyperparameters.

Beyond the simple concatenation for postfusion, more advanced postfusion strategies, taking nonlinear cross-modality relationships into consideration, have been proposed. For example, deep CCA was proposed by Andrew et al.[43] to detect nonlinear correlations, which introduces a deep network representation for each modality and then fused later using a CCA framework. Collaborative regression incorporates a regression penalty into CCA

so that the model is capable of identifying correlated representations from different views, which are also equipped with discriminative power for phenotype[44]. Hu et al.[45] proposed a deep collaborative learning (DCL) framework, which seeks a deep network representation of two datasets, while incorporating their correlations into the model at the same time. Despite the deep learning model is capable of better performing data fusion by capturing the complex relationship between multimodal data, the interpretability of the model is poor. Focusing on this limitation, Hu et al.[46] further proposed Gradient-weighted Class Activation Mapping (Grad-CAM) guided convolutional collaborative learning on the basis of DCL, which enhanced the interpretability of the model to the results.

Besides, as an interesting branch of deep learning, graph neural network (GNN) is quite suitable for multimodal-fusion-based learning tasks, as the features of each node and edge can be extracted from different modalities. For instance, Chen et al.[47] utilized a GNN based method to integrate features extracted from sMRI and fMRI. Specifically, they constructed brain-level graph and took the functional connectivity calculated from fMRI as edge features, took T1-weighted intensity from sMRI and the amplitude of low-frequency fluctuation (ALFF) from fMRI as node features. As graph convolutional networks (GCN) have recently been demonstrated to be superior in capturing network representations tailored for identifying specific brain disorders, Yao et al.[48] proposed a mutual multi-scale triplet GCN method for brain disorder diagnosis, which outperforms several state-of-the-art methods in identifying three types of brain disorders through fusing multi-scale structural and functional connectivities.

3 Atlasing via multimodal brain imaging

The terms of brain template, brain parcellation and brain atlas are often confusing and occasionally used interchangeably. To be clear in this review, the brain template refers to a standardized 3D coordinate frame, which allows brain researchers to combine data from many subjects to detect group-averaged signals above the background noise[49]. Based on the brain template frame, researchers can compare findings from different imaging modalities and laboratories around the world, share large-scale neuroimaging databases, conduct parcellation, as well as constructing atlas maps across subjects. The term brain parcellation is a way of segmenting a particular region of the brain or the whole brain into sub-regions. They may not reflect a pre-defined ontology of brain structures, but better characteristics of the signal of interest[50]. Unlike brain parcellation, brain atlases refer to the segmentation or parcellation of the whole brain and the associated assignment of each subregion, accounting for a certain state of the knowledge of structures anatomically, functionally or based on connectivity.

Brodmann atlas is the most widely cited atlas of humans defined by the cytoarchitectural organization of neurons using the Nissl method of cell staining in 1909[51]. It split the cortex into 52 different areas. Although this atlas has provided invaluable information, their micro-scale cytoarchitectonics is insufficient to completely represent brain organization, especially under the recent progress of ultrathin section technology, staining technology and microscopic observation technology. Since most of the existing high-resolution histological imaging technologies need to slice the brain first, then the inevitable tissue defects and deformation in this process make it lack accurate 3D spatial information. In order to construct a brain atlas with both accurate 3-dimensional spatial information and high-resolution anatomical information, researchers usually fuse MRI with high-resolution histological images, which provides an important way to study the mesoscopic mechanism of macroscopic features and the macroscopic characterization of mesoscopic features. For example, Allen Institute recently constructed an average template brain named the Allen Mouse Brain Common Coordinate Framework at 10 μm voxel resolution by interpolating high resolution in-plane serial two-photon tomography images with 100 μm z-sampling from 1 675 young adult mice[52]. As a general reference framework, the atlas can integrate multimodal data such as histological staining, immunohistochemistry, transgene, in situ hybridization and tracer projection, and has a good visualization tool and interactive interface. Amunts et al.[53] published the latest human cytoarchitecture brain atlas (Julich-Brain atlas) via fusion MRI images and brain histological sections for microscopic imaging. This atlas is the first human three-dimensional atlas containing cytoarchitectonic maps of cortical areas and subcortical nuclei. It is open access to support neuroimaging studies as well as modeling and simulation. It is also interoperable, which enables it to connect to other atlases and resources.

Apart from building brain atlas based on micro-scale cytoarchitecture, another category is based on macro-scale connectional information, both structural and functional connectivity profiles derived from MRI. For example, the Yeo Atlases are designed to study intrinsic functional connectivity using resting-state fMRI, with each hemisphere being subdivided into either 7 or 17 functionally coupled regions across the cerebral cortex[54]. Structural connectivity derived from DTI is used to construct a fine-grained parcellation named the Human Brainnetome Atlas[55]. It provides a cross-validated atlas and contains information on both anatomical and functional connections. As child brain may have different topological patterns from adult, the same team recently released a child atlas named the Human Child Brainnetome Atlas[56], which provides a precise period-specific and cross-validated version of the brain atlas for the preadolescent individuals aging from 8 to 10 years old.

To fuse multiple modalities in the same parcellation task, Parisot et al.[3] proposed a graph-based multimodal parcellation framework to handle the large variety of available input modalities for parcellation task. This method designs an iterative framework to fuse different modalities, forcing the parcellations to converge towards a set of mutually informed modality specific parcellations. Quantitative and qualitative results show that integrating multi-modal information yields stronger and more robust connectivity networks that provide a better representation of the population. To further consider individual-specific topography of cerebral cartography for personalized medical practice, Ma et al.[4] developed a new tool called brain atlas individualization network (BAI-Net) to automatically parcellate individual cerebral cortex into segregated areas using structural and diffusion MRIs.

4 Multimodal fusion in studying cognition and development

4.1 Cognition

The mysteries of cognition have attracted wide attention. To clarify the potential source of cognition, some researchers have tried to mine the internal key factors by establishing multimodal fusion algorithms that can effectively predict cognitive ability. For example, through a standard post-fusion pipeline, Xiao et al.[57] established a multimodal model to predict the visual working memory (VWM) based on ALFF, gray matter volume (GM) and fraction anisotropy (FA) extracted from multimodal MRI data, and identified specific features that supported the human VWM. Besides, Xiao et al.[38] took the between-subject and between-modality relationships reflected by multimodalities into account, forming a manifold regularized multi-task learning model. Through this approach, they achieved better performance in the prediction of intelligence quotient (IQ) using both working memory task and emotion cognition task fMRI.

Furthermore, in order to effectively exploit the nonlinear relationship of multimodal data, deep learning technology has also been widely used in the prediction of cognitive ability. For example, Hu et al.[45] proposed the DCL framework to capture the nonlinear predictive relationship between multimodal data. They applied the method to several fMRI modalities and achieved higher accuracy over other baseline methods in tasks of classifying individuals into groups with different age and cognition level. To improve the interpretability of the model, they further proposed Grad-CAM guided convolutional collaborative learning on the basis of DCL[46]. They integrated fMRI and single nucleotide polymorphism (SNP) data to classify subjects into groups with low or high cognitive functions and identified associated salient features with each modality of input data. Besides, by adopting GNNs as a basic framework for multimodal fusion technology, Qu et al.[58] handled multi-diagram fMRI (emotion cognition task and working memory task fMRI), and conducted multimodal fusion, taking the relationships between subjects within and between modalities as manifold regularization term. The authors achieved satisfactory regression results in the task of predicting individual wide range achievement test (WRAT) score based on multimodal data, and also discovered relevant potential biomarkers.

4.2 Brain age

Brain age is defined as the estimated biological age based on the normal population[59]. Recently, using brain imaging to accurately predict brain age is another popular topic as it is closely associated with development, cognition and even disorders[60, 61]. Accurate and objective brain age prediction based on multimodal data can be used as a biomarker, which has important neurobiological and clinical significance[62, 63]. In order to improve the accuracy of brain age prediction, Liem et al.[64] adopted multimodal data (fMRI and sMRI) for predicting of brain age. They used support vector regression (SVR) to make predictions with unimodality and then integrated the predictions from each modality using random forest models. The experimental results showed that multimodal data improved brain age prediction and the discrepancy between brain age and the chronological age captured cognitive impairment. Based on these results, Engemann et al.[65] further integrated magnetoencephalography (MEG) with fMRI and sMRI to improve the performance of brain age prediction, which found that MEG carried non-redundant information with fMRI and the multimodal learning mitigated the impact of missing data on prediction results. More comprehensively, Niu et al.[66] conducted a study to systematically compare the effects of different combinations of machine learning methods and multimodal features on the performance of brain age prediction. They included features extracted from three modalities (T1, resting-state fMRI and DWI) and four machine learning algorithms (SVR, Gaussian processes regression, ridge regression and deep neural networks), forming a total of 36 combinations. Through a large number of experiments, they demonstrated that multimodal imaging features can predict brain age with better accuracy than unimodal imaging features.

In addition, aiming at the more challenging task of infant age prediction, Hu et al.[67] proposed a disentangled-multimodal adversarial autoencoder method (DMM-AAE) to disentangle complementary and shared information between modalities for effectively predicting infant age. The well-designed deep learning model considered both the complementary and shared information of multimodal data and brought new possibilities for improving the accuracy of brain age prediction.

5 Multimodal fusion in studying brain disorders

With the accumulation of clinical multimodal data, multimodal fusion technology has a wide range of applications in clinical scenarios (Fig. 3). On the one hand, through the multimodal fusion technology based on supervised learning, we can carry out efficient computer-aided diagnosis and identify key biomarkers. On the other, unsupervised fusion methods are expected to help us explore potential disease-related factors of complex brain diseases, of which pathogenesis of many diseases has not been fully understood, from multiple perspectives, thus enabling the development of clinical diagnosis and treatment.

5.1 Diagnosis

In terms of multimodal based diagnosis, supervised learning methods are widely used to guide the multimodal fusion process. Among them, the one simple and commonly used framework is to concatenate the features extracted from multimodalities and take the concatenated features as input of the classification algorithms, such as SVM[68] and decision tree[69]. Besides, increasing efforts have been made to take into account the correlational or complementary relationship between multimodal data. Focusing on the use of the information shared by multimodal data for Alzheimer′s disease (AD) classification, Ning et al.[70] mapped the features extracted from sMRI and positron emission tomography (PET) to a discriminative shared latent space. To improve the performance of representation learning and alleviate the risk of overfitting, serval relational constraints were introduced as regularization terms. Due to the end-to-end learning strategy, deep learning provides a scalable multimodal fusion framework for computer-aided diagnosis. Taking both edge and node features into account, Chen et al.[47] proposed an adversarial learning based node-edge graph attention network architecture and experimentally demonstrated the effectiveness of the model in handling multimodal fusion based autistic spectrum disorder (ASD) classification task. Simultaneously considering cross-modality information correlation and complementation, Zheng et al.[71] prosed a multimodal graph learning (MMGL) method constructed on a population-level graph to conduct ASD diagnosis. The authors fused demographic information, automated anatomical quality assessment metrics, automated functional quality assessment metric and functional connectivity with the attention mechanism, achieving modal-aware representation learning. Based on the modal-aware representation, then, the population-level adaptive graph structure learning was introduced to build the graph structure and perform the node classification task (diagnosis of ASD). Besides, the proposed method was also applied to the prediction task of AD, demonstrating its popularization.

Although the applications of multimodal fusion technology based on unsupervised learning in computer-aided diagnosis are relatively challenging, some studies have made attempts in this field and provided feasible research schemes. As the information contained in different modalities is not necessarily linearly related, inspired by similarity network fusion, Tong et al.[72] focused on nonlinear graph fusion which fused the inter-subject similarity matrix derived from different modalities (MRI, PET, cerebrospinal fluid (CSF) and genes) into a unified graph in an iteration manner so as to package the complementary information. The unified graph was then used to classify healthy controls, mild cognitive impairment (MCI) and AD subjects and the experimental results showed that this nonlinear graph fusion method is superior to

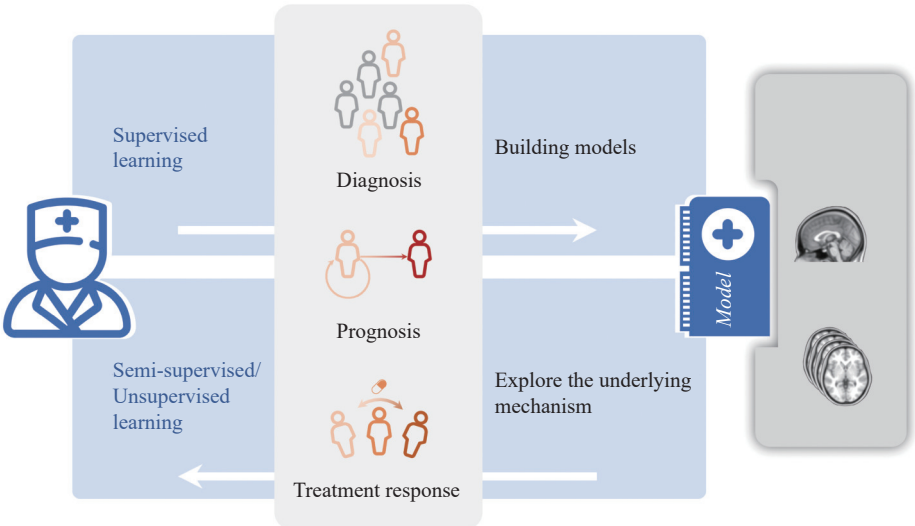

Fig. 3 Application scenarios of multimodal fusion in brain disorders

other methods based on single modality and the multimodal method based on linear fusion in the AD diagnosis task. Besides, Liu et al.[73] tried to provide a more comprehensive view of schizophrenia (SCZ) through unsupervised multimodal fusion technology. Aiming to uncover the co-vary abnormal patterns across different neuroimaging modalities of SCZ, they conducted a linked 4-way multimodal analysis in an unsupervised manner using MCCA + jICA. In this study, Zang et al.[74] identified the co-varying patterns cross four types of features, that is GM extracted from sMRI, FA extracted from dMRI and regional homogeneity (ReHo) and functional network connectivity extracted from rsfMRI. And multiple synchronous abnormal changes in functional and structural regions were suggested to be associated with SCZ, providing a unique perspective to bridge the functional and structural abnormalities in SCZ which cannot be achieved in unimodal studies. Moreover, through fusing sMRI, fMRI and genetics with a parallel ICA-based framework, Luo et al.[75] explored how the changes captured by different modalities are correlated along the progress of SCZ, and revealed that different modalities are differently sensitive to the duration of illness and disease stages.

5.2 Prognosis

Establishing accurate prognosis models (prediction of disease progression) can provide valuable advice for clinical treatment plans, and it is also an essential part of precision medicine. In this research field, due to high requirements for model interpretability and the scarcity of data resources, most studies have adopted relatively simple multimodal fusion strategies with strong interpretability. Kinreich et al.[76] used the least absolute shrinkage and selection operator (LASSO) to select discriminative ones from multimodal features extracted from electroencephalogram (EEG), polygenic risk scores, medications, and demographic information. And then linear SVM was used to predict the future status (continued or remitted) of participants with alcohol use disorder. And Luo et al.[77] explored the performance of ensemble learning based multimodal fusion methods in the prognosis of childhood onset ADHD, persisters or remitters. In this work, they enumerated the combinations of four ensemble learning strategies (voting, bagging, boosting and stacking) and seven commonly used classifiers. Finally, it was found that after training SVM classifiers respectively based on the features extracted from T1, DWI and the cued attention based fMRI, the best performance of outcome prediction with an area under the curve (AUC) score of 0.9 could be achieved by bagging based ensemble learning. Also, Song et al.[78] packaged resting-state fMRI and three clinical characteristics, and successfully identified whether the patients with disorders of consciousness could later recover consciousness with high accuracy of 88%, utilizing partial least squares regression.

Deep learning methods have also been proposed to fuse multimodal neuroimaging for better performance of prognosis prediction. For example, focusing on incomplete multimodal neuroimaging, Liu et al.[79] proposed a joint neuroimaging synthesis and representation learning framework to effectively integrate sMRI and PET in the scenarios with incomplete multimodal data, and predicted conversion from subjective cognitive decline (SCD) to MCI. The proposed framework was equipped with a generative adversarial network based image synthesis subnetwork and a fusion subnetwork, which took the concatenated multimodal features as inputs, finally achieved an AUC score of 0.747 in SCD conversion prediction.

5.3 Treatment response prediction

It is of great significance for clinical decision-making to establish a treatment response prediction model based on clinical data and machine learning, which still remains challenging. Due to the particularity of the application scenario, the models for predicting the treatment response need to be easy to understand. Accordingly, most of the relevant studies use algorithms with strong interpretability. For example, Luo et al.[80] integrated multimodal data in a simple way, logistic regression, and found that the multimodal model which combined the neuroimaging and behavioral information significantly improved the predictive performance of treatment response for cocaine dependence with 96% accuracy. Besides, Billot et al.[81] found that using multimodal features selected from demographic information, behavioral assessment, and features extracted from multimodal neuroimaging data, SVM or random forest can be used for better prediction of response to rehabilitation in chronic post-stroke than utilizing single modality data. Also, Schmitgen et al.[82] adopted a random forest model to predict the response of subjects with borderline personality disorder to dialectical behavior therapy based on previously identified multimodal features. Using the random forest algorithm, they identified multimodal features that have predictive power for the treatment from the feature set which contained features extracted from functional and structural MRI, as well as clinical and demographic information, providing the possibility for personalized psychiatric interventions.

6 Challenges and future directions

With the ability to revealing cross-modal information compared to single modality, multimodal fusion of brain imagings gained promising performances in studying brain parcellation, cognition and development, and brain disorders. However, with the random variation characteristics of some fusion strategies, small sample size may

lead to a false positive linkage. Then the first common challenge is to produce big data. Second, the development of imaging technology brings the brain imaging to a finer scale. How to bridge the gap of spatial alignment across-scales is challenging. Third, most current multimodal fusion methods are applied to solve macro-scale problems. With the release of micro-scale and meso-scale data, how to transfer the current fusion strategies for interaging multi-scale datasets is an important focus. Furthermore, the above discussed large cohorts, high-throughput and high-quality imaging, and new models are all creating substantial challenges for resources of storing, analyses and visualization of the data, leaving an increased demand for building new platforms.

6.1 Big data

Studies with small sample sizes are often vulnerable to sampling variability, the random variation of an association across population subsamples. Sampling variability decreases and associations stabilize with increasing sample sizes. A recent study which quantifies the effect sizes and reproducibility of brain-wide association studies (BWAS) as a function of sample size revealed that when sample sizes grew into the thousands, replication rates began to improve and effect size inflation decreased[83]. BWAS paves a way to reveal associations between inter-individual differences in brain structure or function and complex cognitive or mental health phenotypes. Consequently, achieving reliable association results from various modalities is also appealing for larger sample sizes. Encouragingly, it has gained increasing attention, for example, an increasing number of large consortia released multiple brain imaging modalities data from over thousands of participants involving lifespan development and typical mental illnesses, e.g., UK Biobank[84], Adolescent Brain Cognitive Development (ABCD)[85], Human Connectome Project (HCP)[86], etc. The Chinese government has also released ongoing projects to collect large brain imaging cohorts for children development and brain disorder exploration.

6.2 Multi-scale data

Multimodal data are always across scales, from genetic, molecular, cellular to macroscale level. With the development of imaging technology, observing the brain at much finer scales, such as visualizing individual brain cells and synapses, are made possible. For example, advanced optical imaging is capable of delineating cellular-level images of the mouse brain at a voxel size of $0.32 \times 0.32 \times 1.00\,\mu m^3$[87]. Large-scale neuronal tracing, fluorescent labelling techniques and CLARITY-tissue staining have proven useful in uncovering axonal projections between brain areas in mice[88] and macaques[89]. Nevertheless, histology requires sectioning tissue, hence distorting its 3D structure. To link the different spatial scales from the synaptic level to the whole brain on macroscopic scale is challenging, because data from different scales are always with various heterogeneities, uncomparable coordinates and different levels of resolution. In general, a volumetric MRI scan was often used as undistorted reference to guide the reconstructions of histology images[90]. For example, the Allen Institute for Brain Science defined the Allen Mouse Brain Common Coordinate Framework on the whole mouse brain[52], which provides a nice paradigm and resource for researchers. The open Big Brain allow for a unique comparison of cytoarchitectonic features to macroscale brain connectivity[91]. The latest human cytoarchitecture brain atlas (Julich-Brain atlas) fused MRI images with brain histological sections for microscopic imaging[53]. However, due to the relatively large volume and complex structure of primates, especially human brains, the slicing work is cumbersome and there will be more defects and deformation in the process, which makes it very difficult to obtain and reconstruct histological images.

6.3 New models

Whether now or in the future, how to mine complementary information between different modalities, and how to integrate different spatial resolution information at macro-, meso- and micro-scales are important questions in brain imaging fusion. With the advances of artificial intelligence (AI), it is an inevitable trend to fuse multimodal high-quality images of large samples using AI systems and algorithms. While brain-inspired and biologically constrained AI is promising in particular because of the biological meanings embedded. For example, all the anatomical nodes, function activations and connectivity derived from the multimodal multi-scale brain atlases, could be considered as constraints to the AI models for information fusion.

Meanwhile, brain modeling and simulation plays an increasingly important role in understanding the working mechanism of the brain, such as digital twin brain. It has been used to build computational models of complex systems with multiple modality information to predict with high precision how a certain therapeutic intervention would benefit or harm a particular patient[92]. However, the simulation of brain network function across scales is still in progress. Combining multiscale brain networks to a digital twin brain model and simulating brain functional activity across different resolutions, as well as developing targeted and integrated platforms to do so, has become an important direction of multimodal brain data fusion.

Moreover, although multimodal fusion has been widely used and is becoming a powerful tool for a wide range of

applications, it is not always the optimal choice. Not all problems require multiple sources of information, and in some cases, the fusion of all available modalities may even lead to a degradation of overall performance[81, 82]. Nevertheless, it is undeniable that multimodal fusion provides us with the possibility of a more comprehensive understanding of the research topic and effectively compensates for incomplete information in single modalities. It is worth noting that different modalities depict different information and have certain modality-specificity. Therefore, it is necessary to select appropriate multi-modal data according to different research tasks. This is also why we emphasize the need to further explore the sharing and complementary relationships of multimodal information while advocating multimodal fusion technology in our review. By exploring the best practices for multimodal fusion strategies, we can unlock new opportunities for scientific inquiry and gain a deeper understanding of the brain.

6.4 New platforms

Due to an increased demand for large cohorts, high-throughput and high-quality imaging, and models with a huge number of parameters, research on optimization for computational efficiency is certainly an important direction. For example, when reconstructing a fragment of the human cerebral cortex to three dimensions, a volume of about 1 mm^3 is correspond to a data volume of about 1.4 petabytes[93]. Requirements will further increase when adding more temporal changes to the current spatial scale data. Such large datasets also create substantial challenges for the analysis and visualization of the data[94]. The big data challenge in neuroscience demands modular and interactive concepts of future supercomputing for data processing, petabyte-scale data centers for data storage and sometimes even hard drives for long-distance data transmission. To deal with the challenges in storage and computational efficiency, new software and hardware platforms are in urgent demand. For example, the computing services of Human Brain Project[95] build on resources and services of the Fenix infrastructure. Within the Fenix infrastructure, six European supercomputing centres have agreed to align their services, supporting for interactive and scalable computing services, virtual machine services, active and archival data repositories, and other addition services. The raw data of the BRAIN initiative cell census network (BICCN) project[96] is stored in three data centers (University of Maryland, USA, Pittsburgh Supercomputing Center, USA, and Massachusetts Institute of Technology, USA). The code tools related to the project are updated on GitHub and customized based on Google Terra cloud native computing platform.

7 Conclusions

In this review, we attempted to conduct a comprehensive survey on the progress of the multimodal brain imaging fusion studies in recent years, aiming at tracking down the advanced fusion strategies and remarkable applications. Compared to unimodal, the multimodal strategies could better exploit both the shared representation between modalities and the modality-specific complementary information, hence improving the applications on building brain atlases, predicting phenotypic outcomes and classifying brain diseases. With the development of larger datasets, novel fusion models and supercomputing facilities, multimodal fusion could help better unveil the underlying mechanisms of human brain cognition and clinical disorders in the future.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 82001450 and 82151307); Science and Technology Innovation 2030 – Brain Science and Brain-Inspired Intelligence Project of China (No. 2021 ZD0200201), the National Key Research and Development Program of China (No. 2022YFC3601200), the China Postdoctoral Science Foundation (No. BX2020 0364), the Chinese Academy of Sciences, Science and Technology Service Network Initiative (No. KFJ-STS-ZDTP-078), the Strategic Priority Research Program of the Chinese Academy of Sciences, China (No. XDB3203 0200), the Scientific Project of Zhejiang Laboratory, China (No. 2022ND0AN01).

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article′s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article′s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

[1] J. Sui, T. Adali, Q. B. Yu, J. Y. Chen, V. D. Calhoun. A

review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*, vol. 204, no. 1, pp. 68–81, 2012. DOI: 10.1016/j.jneumeth.2011.10.031.

[2] T. Z. Jiang. Brainnetome: A new-ome to understand the brain and its disorders. *Neuroimage*, vol. 80, pp. 263–272, 2013. DOI: 10.1016/j.neuroimage.2013.04.002.

[3] S. Parisot, B. Glocker, S. I. Ktena, S. Arslan, M. D. Schirmer, D. Rueckert. A flexible graphical model for multi-modal parcellation of the cortex. *NeuroImage*, vol. 162, pp. 226–248, 2017. DOI: 10.1016/j.neuroimage.2017.09.005.

[4] L. Ma, Y. Zhang, H. T. Zhang, L. Q. Cheng, J. J. Zhuo, W. Y. Shi, Y. H. Lu, W. Li, Z. Y. Yang, J. J. Wang, L. Z. Fan, T. Z. Jiang. BAI-Net: Individualized anatomical cerebral cartography using graph neural network. *IEEE Transactions on Neural Networks and Learning Systems*, to be published. DOI: 10.1109/TNNLS.2022.3213581.

[5] J. Sui, R. T. Jiang, J. Bustillo, V. Calhoun. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises. *Biological Psychiatry*, vol. 88, no. 11, pp. 818–828, 2020. DOI: 10.1016/j.biopsych.2020.02.016.

[6] R. T. Jiang, V. D. Calhoun, Y. Cui, S. L. Qi, C. J. Zhuo, J. Li, R. Jung, J. Yang, Y. H. Du, T. Z. Jiang, J. Sui. Multimodal data revealed different neurobiological correlates of intelligence between males and females. *Brain Imaging and Behavior*, vol. 14, no. 5, pp. 1979–1993, 2020. DOI: 10.1007/s11682-019-00146-z.

[7] A. C. Feng, N. Luo, W. T. Zhao, V. D. Calhoun, R. T. Jiang, D. M. Zhi, W. Y. Shi, T. Z. Jiang, S. Yu, Y. Xu, S. Liu, J. Sui. Multimodal brain deficits shared in early-onset and adult-onset schizophrenia predict positive symptoms regardless of illness stage. *Human Brain Mapping*, vol. 43, no. 11, pp. 3486–3497, 2022. DOI: 10.1002/hbm.25862.

[8] N. Luo, J. Sui, J. Y. Chen, F. Q. Zhang, L. Tian, D. D. Lin, M. Song, V. D. Calhoun, Y. Cui, V. M. Vergara, F. F. Zheng, J. Y. Liu, Z. Y. Yang, N. M. Zuo, L. Z. Fan, K. B. Xu, S. F. Liu, J. Li, Y. Xu, S. Liu, L. X. Lv, J. Chen, Y. C. Chen, H. Guo, P. Li, L. Lu, P. Wan, H. N. Wang, H. L. Wang, H. Yan, J. Yan, Y. F. Yang, H. X. Zhang, D. Zhang, T. Z. Jiang. A schizophrenia-related genetic-brain-cognition pathway revealed in a large Chinese population. *Ebiomedicine*, vol. 37, pp. 471–482, 2018. DOI: 10.1016/j.ebiom.2018.10.009.

[9] W. Y. Shi, L. Z. Fan, H. Y. Wang, B. Liu, W. Li, J. Li, L. Q. Cheng, C. Y. Chu, M. Song, J. Sui, N. Luo, Y. Cui, Z. W. Dong, Y. H. Lu, Y. W. Ma, L. Ma, K. X. Li, J. Chen, Y. C. Chen, H. Guo, P. Li, L. Lu, L. X. Lv, P. Wan, H. N. Wang, H. L. Wang, H. Yan, J. Yan, Y. F. Yang, H. X. Zhang, D. Zhang, T. Z. Jiang. Two subtypes of schizophrenia identified by an individual-level atypical pattern of tensor-based morphometric measurement. *Cerebral Cortex*, vol. 33, no. 7, pp. 3683–3700, 2023. DOI: 10.1093/cercor/bhac301.

[10] N. Luo, X. S. Luo, S. L. Zheng, D. R. Yao, M. Zhao, Y. Cui, Y. Zhu, V. D. Calhoun, L. Sun, J. Sui. Aberrant brain dynamics and spectral power in children with ADHD and its subtypes. *European Child & Adolescent Psychiatry*, to be published. DOI: 10.1007/s00787-022-02068-6.

[11] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. W. Tu, M. Brudno, B. Haibe-Kains, A. Goldenberg. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*, vol. 11, no. 3, pp. 333–337, 2014. DOI: 10.1038/nmeth.2810.

[12] F. Bao, Y. Deng, S. Wan, S. Q. Shen, B. Wang, Q. H. Dai, S. J. Altschuler, L. F. Wu. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nature Biotechnology*, vol. 40, no. 8, pp. 1200–1209, 2022. DOI: 10.1038/s41587-022-01251-z.

[13] H. Hotelling. Relations between two sets of variates. *Biometrika*, vol. 28, no. 3–4, pp. 321–377, 1936. DOI: 10.1093/biomet/28.3-4.321.

[14] Y. O. Li, T. Adali, W. Wang, V. D. Calhoun. Joint blind source separation by multiset canonical correlation analysis. *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009. DOI: 10.1109/TSP.2009.2021636.

[15] D. D. Lin, V. D. Calhoun, Y. P. Wang. Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Medical Image Analysis*, vol. 18, no. 6, pp. 891–902, 2014. DOI: 10.1016/j.media.2013.10.010.

[16] X. W. Zhang, J. Pan, J. Shen, Z. U. Din, J. L. Li, D. W. Lu, M. X. Wu, B. Hu. Fusing of electroencephalogram and eye movement with group sparse canonical correlation analysis for anxiety detection. *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 958–971, 2022. DOI: 10.1109/TAFFC.2020.2981440.

[17] Z. M. Zhang, Z. D. Deng. A kernel canonical correlation analysis based idle-state detection method for SSVEP-based brain-computer interfaces. *Advanced Materials Research*, vol. 341–342, pp. 634–640, 2011. DOI: 10.4028/www.scientific.net/AMR.341-342.634.

[18] J. Sui, H. He, G. D. Pearlson, T. Adali, K. A. Kiehl, Q. B. Yu, V. P. Clark, E. Castro, T. White, B. A. Mueller, B. C. Ho, N. C. Andreasen, V. D. Calhoun. Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia. *NeuroImage*, vol. 66, pp. 119–132, 2013. DOI: 10.1016/j.neuroimage.2012.10.051.

[19] S. L. Qi, V. D. Calhoun, T. G. M. van Erp, J. Bustillo, E. Damaraju, J. A. Turner, Y. H. Du, J. Yang, J. Y. Chen, Q. B. Yu, D. H. Mathalon, J. M. Ford, J. Voyvodic, B. A. Mueller, A. Belger, S. McEwen, S. G. Potkin, A. Preda, T. Z. Jiang, J. Sui. Multimodal fusion with reference: Searching for joint neuromarkers of working memory deficits in schizophrenia. *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 93–105, 2018. DOI: 10.1109/TMI.2017.2725306.

[20] V. D. Calhoun, T. Adali, N. R. Giuliani, J. J. Pekar, K. A. Kiehl, G. D. Pearlson. Method for multimodal analysis of independent source differences in schizophrenia: Combining gray matter structural and auditory oddball functional data. *Human Brain Mapping*, vol. 27, no. 1, pp. 47–62, 2006. DOI: 10.1002/hbm.20166.

[21] M. A. B. S. Akhonda, B. Gabrielson, S. Bhinge, V. D. Calhoun, T. Adali. Disjoint subspaces for common and distinct component analysis: Application to the fusion of multi-task FMRI data. *Journal of Neuroscience Methods*, vol. 358, Article number 109214, 2021. DOI: 10.1016/j.jneumeth.2021.109214.

[22] J. Y. Liu, O. Demirci, V. D. Calhoun. A parallel independent component analysis approach to investigate genomic influence on brain function. *IEEE Signal Processing Letters*, vol. 15, pp. 413–416, 2008. DOI: 10.1109/LSP.2008.922513.

[23] V. M. Vergara, A. Ulloa, V. D. Calhoun, D. Boutte, J. Y. Chen, J. J. N. Liu. A three-way parallel ICA approach to analyze links among genetics, brain structure and brain function. *NeuroImage*, vol. 98, pp. 386–394, 2014. DOI: 10.1016/j.neuroimage.2014.04.060.

[24] S. L. Qi, R. F. Silva, D. Q. Zhang, S. M. Plis, R. Miller, V. M. Vergara, R. T. Jiang, D. M. Zhi, J. Sui, V. D. Calhoun. Three-way parallel group independent component analysis: Fusion of spatial and spatiotemporal magnetic resonance imaging data. *Human Brain Mapping*, vol. 43, no. 4, pp. 1280–1294, 2022. DOI: 10.1002/hbm.25720.

[25] A. R. Groves, C. F. Beckmann, S. M. Smith, M. W. Woolrich. Linked independent component analysis for multimodal data fusion. *NeuroImage*, vol. 54, no. 3, pp. 2198–2217, 2011. DOI: 10.1016/j.neuroimage.2010.09.073.

[26] W. K. Gong, C. F. Beckmann, S. M. Smith. Phenotype discovery from population brain imaging. *Medical Image Analysis*, vol. 71, Article number 102050, 2021. DOI: 10.1016/j.media.2021.102050.

[27] W. K. Gong, S. Bai, Y. Q. Zheng, S. M. Smith, C. F. Beckmann. Supervised phenotype discovery from multimodal brain imaging. *IEEE Transactions on Medical Imaging*, vol. 42, no. 3, pp. 834–849, 2023. DOI: 10.1109/TMI.2022.3218720.

[28] T. Adali, M. A. B. S. Akhonda, V. D. Calhoun. ICA and IVA for data fusion: An overview and a new approach based on disjoint subspaces. *IEEE Sensors Letters*, vol. 3, no. 1, Article number 7100404, 2019. DOI: 10.1109/LSENS.2018.2884775.

[29] R. F. Silva, S. M. Plis, T. Adalı, M. S. Pattichis, V. D. Calhoun. Multidataset independent subspace analysis with application to multimodal fusion. *IEEE Transactions on Image Processing*, vol. 30, pp. 588–602, 2021. DOI: 10.1109/TIP.2020.3028452.

[30] B. B. Avants, N. J. Tustison, J. R. Stone. Similarity-driven multi-view embeddings from high-dimensional biomedical data. *Nature Computational Science*, vol. 1, no. 2, pp. 143–152, 2021. DOI: 10.1038/s43588-021-00029-8.

[31] A. Kumar, P. Rai, H. Daumé. Co-regularized multi-view spectral clustering. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, Granada, Spain, pp. 1413–1421, 2011.

[32] D. Y. Xie, X. D. Zhang, Q. X. Gao, J. L. Han, S. Xiao, X. B. Gao. Multiview clustering by joint latent representation and similarity learning. *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4848–4854, 2020. DOI: 10.1109/tcyb.2019.2922042.

[33] G. E. Hinton, R. S. Zemel. Autoencoders, minimum description length and Helmholtz free energy. In *Proceedings of the 6th International Conference on Neural Information Processing Systems*, Denver, USA, pp. 3–10, 1993.

[34] R. Miotto, F. Wang, S. Wang, X. Q. Jiang, J. T. Dudley. Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018. DOI: 10.1093/bib/bbx044.

[35] T. L. Ma, A. D. Zhang. Integrate multi-omics data with biological interaction networks using multi-view factorization autoencoder (MAE). *BMC Genomics*, vol. 20, no. S11, Article number 944, 2019. DOI: 10.1186/s12864-019-6285-x.

[36] D. Q. Zhang, D. G. Shen. The Alzheimer′s Disease Neuroimaging Initiative. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer′s disease. *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012. DOI: 10.1016/j.neuroimage.2011.09.069.

[37] B. Jie, D. Q. Zhang, B. Cheng, D. G. Shen. The Alzheimer′s Disease Neuroimaging Initiative. Manifold regularized multitask feature learning for multimodality disease classification. *Human Brain Mapping*, vol. 36, no. 2, pp. 489–507, 2015. DOI: 10.1002/hbm.22642.

[38] L. Xiao, J. M. Stephen, T. W. Wilson, V. D. Calhoun, Y. P. Wang. A manifold regularized multi-task learning model for IQ prediction from two fMRI paradigms. *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 796–806, 2020. DOI: 10.1109/TBME.2019.2921207.

[39] M. X. Liu, D. Q. Zhang, D. G. Shen. Relationship induced multi-template learning for diagnosis of Alzheimer′s disease and mild cognitive impairment. *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1463–1474, 2016. DOI: 10.1109/TMI.2016.2515021.

[40] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, V. D. Calhoun. Deep learning for neuroimaging: A validation study. *Frontiers in Neuroscience*, vol. 8, Article number 229, 2014. DOI: 10.3389/fnins.2014.00229.

[41] W. Z. Yan, G. Qu, W. X. Hu, A. Abrol, B. Cai, C. Qiao, S. M. Plis, Y. P. Wang, J. Sui, V. D. Calhoun. Deep learning in neuroimaging: Promises and challenges. *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 87–98, 2022. DOI: 10.1109/MSP.2021.3128348.

[42] Z. Zhang, Q. Zhu, G. S. Xie, Y. Chen, Z. M. Li, S. H. Wang. Discriminative margin-sensitive autoencoder for collective multi-view disease analysis. *Neural Networks*, vol. 123, pp. 94–107, 2020. DOI: 10.1016/j.neunet.2019.11.013.

[43] G. Andrew, R. Arora, J. Bilmes, K. Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, USA, vol.3, pp. 1247–1255, 2013.

[44] S. M. Gross, R. Tibshirani. Collaborative regression. *Biostatistics*, vol. 16, no. 2, pp. 326–338, 2015. DOI: 10.1093/biostatistics/kxu047.

[45] W. X. Hu, B. Cai, A. Y. Zhang, V. D. Calhoun, Y. P. Wang. Deep collaborative learning with application to the study of multimodal brain development. *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 12, pp. 3346–3359, 2019. DOI: 10.1109/TBME.2019.2904301.

[46] W. X. Hu, X. H. Meng, Y. T. Bai, A. Y. Zhang, G. Qu, B. Cai, G. M. Zhang, T. W. Wilson, J. M. Stephen, V. D. Calhoun, Y. P. Wang. Interpretable multimodal fusion networks reveal mechanisms of brain cognition. *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1474–1483, 2021. DOI: 10.1109/TMI.2021.3057635.

[47] Y. Z. Chen, J. D. Yan, M. X. Jiang, T. Zhang, Z. B. Zhao, W. H. Zhao, J. Zheng, D. Z. Yao, R. Zhang, K. M. Kendrick, X. Jiang. Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. *IEEE Transactions on Neural Networks and Learning Systems*, to be published. DOI: 10.1109/TNNLS.2022.3154755.

[48] D. R. Yao, J. Sui, M. L. Wang, E. K. Yang, Y. Jiaerken, N. Luo, P. T. Yap, M. X. Liu, D. G. Shen. A mutual multiscale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity. *IEEE Transactions on Medical Imaging*, vol. 40, no. 4, pp. 1279–1289, 2021. DOI: 10.1109/tmi.2021.3051604.

[49] A. C. Evans, A. L. Janke, D. L. Collins, S. Baillet. Brain templates and atlases. *NeuroImage*, vol. 62, no. 2, pp. 911–922, 2012. DOI: 10.1016/j.neuroimage.2012.01.024.

[50] B. Thirion, G. Varoquaux, E. Dohmatob, J. B. Poline.

Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience*, vol. 8, Article number 167, 2014. DOI: 10.3389/fnins.2014.00167.

[51] K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde in Ihren Prinzipien Dargestellt auf Grund des Zellenbaues*, Leipzig, Germany: Barth, 1909.

[52] Q. X. Wang, S. L. Ding, Y. Li, J. Royall, D. Feng, P. Lesnar, N. Graddis, M. Naeemi, B. Facer, A. Ho, T. Dolbeare, B. Blanchard, N. Dee, W. Wakeman, K. E. Hirokawa, A. Szafer, S. M. Sunkin, S. W. Oh, A. Bernard, J. W. Phillips, M. Hawrylycz, C. Koch, H. K. Zeng, J. A. Harris, L. Ng. The Allen mouse brain common coordinate framework: A 3D reference atlas. *Cell*, vol. 181, no. 4, pp. 936–953.e20, 2020. DOI: 10.1016/j.cell.2020.04.007.

[53] K. Amunts, H. Mohlberg, S. Bludau, K. Zilles. Julich-Brain: A 3D probabilistic atlas of the human brain′s cytoarchitecture. *Science*, vol. 369, no. 6506, pp. 988–992, 2020. DOI: 10.1126/science.abb4588.

[54] B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. S. Liu, R. L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011. DOI: 10.1152/jn.00338.2011.

[55] L. Z. Fan, H. Li, J. J. Zhuo, Y. Zhang, J. J. Wang, L. F. Chen, Z. Y. Yang, C. Y. Chu, S. M. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. S. Yu, T. Z. Jiang. The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex*, vol. 26, no. 8, pp. 3508–3526, 2016. DOI: 10.1093/cercor/bhw157.

[56] W. Li, L. Z. Fan, W. Y. Shi, Y. H. Lu, J. Li, N. Luo, H. Y. Wang, C. Y. Chu, L. Ma, M. Song, K. X. Li, L. Q. Cheng, L. Cao, T. Z. Jiang. Brainnetome atlas of preadolescent children based on anatomical connectivity profiles. *Cerebral Cortex*, vol. 33, no. 9, pp. 5264–5275, 2023. DOI: 10.1093/cercor/bhac415.

[57] Y. Xiao, Y. Lin, J. J. Ma, J. H. Qian, Z. J. Ke, L. F. Li, Y. Y. Yi, J. B. Zhang, Z. J. Dai. Predicting visual working memory with multimodal magnetic resonance imaging. *Human Brain Mapping*, vol. 42, no. 5, pp. 1446–1462, 2021. DOI: 10.1002/hbm.25305.

[58] G. Qu, L. Xiao, W. X. Hu, J. Q. Wang, K. Zhang, V. D. Calhoun, Y. P. Wang. Ensemble manifold regularized multi-modal graph convolutional network for cognitive ability prediction. *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 12, pp. 3564–3573, 2021. DOI: 10.1109/TBME.2021.3077875.

[59] N. U. F. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar, K. A. Barnes, J. W. Dubis, E. Feczko, R. S. Coalson, J. R. Jr. Pruett, D. M. Barch, S. E. Petersen, B. L. Schlaggar. Prediction of individual brain maturity using fMRI. *Science*, vol. 329, no. 5997, pp. 1358–1361, 2010. DOI: 10.1126/science.1194144.

[60] J. H. Cole, K. Franke. Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends in Neurosciences*, vol. 40, no. 12, pp. 681–690, 2017. DOI: 10.1016/j.tins.2017.10.001.

[61] R. T. Jiang, D. Scheinost, N. M. Zuo, J. Wu, S. L. Qi, Q. H. Liang, D. M. Zhi, N. Luo, Y. C. Chung, S. Liu, Y. Xu, J. Sui, V. Calhoun. A neuroimaging signature of cognitive aging from whole-brain functional connectivity. *Advanced Science*, vol. 9, no. 24, Article number 2201621, 2022. DOI: 10.1002/advs.202201621.

[62] R. A. I. Bethlehem, J. Seidlitz, S. R. White, J. W. Vogel, K. M. Anderson, C. Adamson, S. Adler, G. S. Alexopoulos, E. Anagnostou, A. Areces-Gonzalez, D. E. Astle, B. Auyeung, M. Ayub, J. Bae, G. Ball, S. Baron-Cohen, R. Beare, S. A. Bedford, V. Benegal, F. Beyer, J. Blangero, M. Blesa Cábez, J. P. Boardman, M. Borzage, J. F. Bosch-Bayard, N. Bourke, V. D. Calhoun, M. M. Chakravarty, C. Chen, C. Chertavian, G. Chetelat, Y. S. Chong, J. H. Cole, A. Corvin, M. Costantino, E. Courchesne, F. Crivello, V. L. Cropley, J. Crosbie, N. Crossley, M. Delarue, R. Delorme, S. Desrivieres, G. A. Devenyi, M. A. Di Biase, R. Dolan, K. A. Donald, G. Donohoe, K. Dunlop, A. D. Edwards, J. T. Elison, C. T. Ellis, J. A. Elman, L. Eyler, D. A. Fair, E. Feczko, P. C. Fletcher, P. Fonagy, C. E. Franz, L. Galan-Garcia, A. Gholipour, J. Giedd, J. H. Gilmore, D. C. Glahn, I. M. Goodyer, P. E. Grant, N. A. Groenewold, F. M. Gunning, R. E. Gur, R. C. Gur, C. F. Hammill, O. Hansson, T. Hedden, A. Heinz, R. N. Henson, K. Heuer, J. Hoare, B. Holla, A. J. Holmes, R. Holt, H. Huang, K. Im, J. Ipser, C. R. Jack Jr, A. P. Jackowski, T. Jia, K. A. Johnson, P. B. Jones, D. T. Jones, R. S. Kahn, H. Karlsson, L. Karlsson, R. Kawashima, E. A. Kelley, S. Kern, K. W. Kim, M. G. Kitzbichler, W. S. Kremen, F. Lalonde, B. Landeau, S. Lee, J. Lerch, J. D. Lewis, J. Li, W. Liao, C. Liston, M. V. Lombardo, J. Lv, C. Lynch, T. T. Mallard, M. Marcelis, R. D. Markello, S. R. Mathias, B. Mazoyer, P. McGuire, M. J. Meaney, A. Mechelli, N. Medic, B. Misic, S. E. Morgan, D. Mothersill, J. Nigg, M. Q. W. Ong, C. Ortinau, R. Ossenkoppele, M. Ouyang, L. Palaniyappan, L. Paly, P. M. Pan, C. Pantelis, M. M. Park, T. Paus, Z. Pausova, D. Paz-Linares, A. Pichet Binette, K. Pierce, X. Qian, J. Qiu, A. Qiu, A. Raznahan, T. Rittman, A. Rodrigue, C. K. Rollins, R. Romero-Garcia, L. Ronan, M. D. Rosenberg, D. H. Rowitch, G. A. Salum, T. D. Satterthwaite, H. L. Schaare, R. J. Schachar, A. P. Schultz, G. Schumann, M. Schöll, D. Sharp, R. T. Shinohara, I. Skoog, C. D. Smyser, R. A. Sperling, D. J. Stein, A. Stolicyn, J. Suckling, G. Sullivan, Y. Taki, B. Thyreau, R. Toro, N. Traut, K. A. Tsvetanov, N. B. Turk-Browne, J. J. Tuulari, C. Tzourio, É. Vachon-Presseau, M. J. Valdes-Sosa, P. A. Valdes-Sosa, S. L. Valk, T. van Amelsvoort, S. N. Vandekar, L. Vasung, L. W. Victoria, S. Villeneuve, A. Villringer, P. E. Vértes, K. Wagstyl, Y. S. Wang, S. K. Warfield, V. Warrier, E. Westman, M. L. Westwater, H. C. Whalley, A. V. Witte, N. Yang, B. Yeo, H. Yun, A. Zalesky, H. J. Zar, A. Zettergren, J. H. Zhou, H. Ziauddeen, A. Zugman, X. N. Zuo, 3R-BRAIN, AIBL, Alzheimer′s Disease Neuroimaging Initiative, Alzheimer′s Disease Repository Without Borders Investigators, CALM Team, Cam-CAN, CCNP, COBRE, cVEDA, ENIGMA Developmental Brain Age Working Group, Developing Human Connectome Project, FinnBrain, Harvard Aging Brain Study, IMAGEN, KNE96, The Mayo Clinic Study of Aging, NSPN, POND, The PREVENT-AD Research Group, VETSA, E. T. Bullmore, A. F. Alexander-Bloch. Brain charts for the human lifespan. *Nature*, vol. 604, no. 7906, pp. 525–533, 2022. DOI: 10.1038/s41586-022-04554-y.

[63] N. Luo, J. Sui, A. Abrol, D. D. Lin, J. Y. Chen, V. M. Vergara, Z. N. Fu, Y. H. Du, E. Damaraju, Y. Xu, J. A. Turner, V. D. Calhoun. Age-related structural and functional variations in 5 967 individuals across the adult lifespan. *Human Brain Mapping*, vol. 41, no. 7, pp. 1725–1737, 2020. DOI: 10.1002/hbm.24905.

[64] F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. Kharabian Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock, S. Riedel-Heller, T. Luck, M. Loeffler, M. L. Schroeter, A. V. Witte, A. Villringer, D. S. Margulies. Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, vol. 148,

pp. 179–188, 2017. DOI: 10.1016/j.neuroimage.2016.11.005.

[65] D. A. Engemann, O. Kozynets, D. Sabbagh, G. Lemaître, G. Varoquaux, F. Liem, A. Gramfort. Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *eLife*, vol. 9, Article number e54055, 2020. DOI: 10.7554/eLife.54055.

[66] X. Niu, F. Q. Zhang, J. Kounios, H. L. Liang. Improved prediction of brain age using multimodal neuroimaging data. *Human Brain Mapping*, vol. 41, no. 6, pp. 1626–1643, 2020. DOI: 10.1002/hbm.24899.

[67] D. Hu, H. Zhang, Z. W. Wu, F. Wang, L. Wang, J. K. Smith, W. L. Lin, G. Li, D. G. Shen. Disentangled-multimodal adversarial autoencoder: Application to infant age prediction with incomplete multimodal neuroimages. *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4137–4149, 2020. DOI: 10.1109/TMI.2020.3013825.

[68] Y. C. Shi, Z. Wang, P. D. Chen, P. Y. Cheng, K. Zhao, H. X. Zhang, H. Shu, L. H. Gu, L. J. Gao, Q. Wang, H. S. Zhang, C. M. Xie, Y. Liu, Z. J. Zhang. Alzheimer's Disease Neuroimaging Initiative. Episodic memory-related imaging features as valuable biomarkers for the diagnosis of Alzheimer's Disease: A multicenter study based on machine learning. *Biological Psychiatry*: *Cognitive Neuroscience and Neuroimaging*, vol. 8, no. 2, pp. 171–180, 2023. DOI: 10.1016/j.bpsc.2020.12.007.

[69] S. G. Liang, Y. F. Li, Z. Zhang, X. Z. Kong, Q. Wang, W. Deng, X. J. Li, L. S. Zhao, M. L. Li, Y. J. Meng, F. Huang, X. H. Ma, X. M. Li, A. J. Greenshaw, J. M. Shao, T. Li. Classification of first-episode schizophrenia using multimodal brain features: A combined structural and diffusion imaging study. *Schizophrenia Bulletin*, vol. 45, no. 3, pp. 591–599, 2019. DOI: 10.1093/schbul/sby091.

[70] Z. Y. Ning, Q. Xiao, Q. J. Feng, W. F. Chen, Y. Zhang. Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis. *IEEE Transactions on Medical Imaging*, vol. 40, no. 6, pp. 1632–1645, 2021. DOI: 10.1109/TMI.2021.3063150.

[71] S. Zheng, Z. F. Zhu, Z. Z. Liu, Z. Y. Guo, Y. Liu, Y. C. Yang, Y. Zhao. Multi-modal graph learning for disease prediction. *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2207–2216, 2022. DOI: 10.1109/TMI.2022.3159264.

[72] T. Tong, K. Gray, Q. Q. Gao, L. Chen, D. Rueckert. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recognition*, vol. 63, pp. 171–181, 2017. DOI: 10.1016/j.patcog.2016.10.009.

[73] S. F. Liu, H. Y. Wang, M. Song, L. X. Lv, Y. Cui, Y. Liu, L. Z. Fan, N. M. Zuo, K. B. Xu, Y. H. Du, Q. B. Yu, N. Luo, S. L. Qi, J. Yang, S. M. Xie, J. Li, J. Chen, Y. C. Chen, H. N. Wang, H. Guo, P. Wan, Y. F. Yang, P. Li, L. Lu, H. Yan, J. Yan, H. L. Wang, H. X. Zhang, D. Zhang, V. D. Calhoun, T. Z. Jiang, J. Sui. Linked 4-way multimodal brain differences in schizophrenia in a large Chinese Han population. *Schizophrenia Bulletin*, vol. 45, no. 2, pp. 436–449, 2019. DOI: 10.1093/schbul/sby045.

[74] Y. F. Zang, T. Z. Jiang, Y. L. Lu, Y. He, L. X. Tian. Regional homogeneity approach to fMRI data analysis. *NeuroImage*, vol. 22, no. 1, pp. 394–400, 2004. DOI: 10.1016/j.neuroimage.2003.12.030.

[75] N. Luo, L. Tian, V. D. Calhoun, J. Y. Chen, D. D. Lin, V. M. Vergara, S. Q. Rao, J. Yang, C. J. Zhuo, Y. Xu, J. A. Turner, F. Q. Zhang, J. Sui. Brain function, structure and genomic data are linked but show different sensitivity to duration of illness and disease stage in schizophrenia. *NeuroImage*: *Clinical*, vol. 23, Article number 101887, 2019. DOI: 10.1016/j.nicl.2019.101887.

[76] S. Kinreich, V. V. McCutcheon, F. Aliev, J. L. Meyers, C. Kamarajan, A. K. Pandey, D. B. Chorlian, J. Zhang, W. P. Kuang, G. Pandey, S. S. S. D. Viteri, M. W. Francis, G. Chan, J. L. Bourdon, D. M. Dick, A. P. Anokhin, L. Bauer, V. Hesselbrock, M. A. Schuckit, J. I. Nurnberger, T. M. Foroud, J. E. Salvatore, K. K. Bucholz, B. Porjesz. Predicting alcohol use disorder remission: A longitudinal multimodal multi-featured machine learning approach. *Translational Psychiatry*, vol. 11, no. 1, Article number 166, 2021. DOI: 10.1038/s41398-021-01281-2.

[77] Y. Y. Luo, T. L. Alvarez, J. M. Halperin, X. B. Li. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. *NeuroImage*: *Clinical*, vol. 26, Article number 102238, 2020. DOI: 10.1016/j.nicl.2020.102238.

[78] M. Song, Y. Yang, J. H. He, Z. Y. Yang, S. Yu, Q. Y. Xie, X. Y. Xia, Y. Y. Dang, Q. Zhang, X. H. Wu, Y. Cui, B. Hou, R. H. Yu, R. X. Xu, T. Z. Jiang. Prognostication of chronic disorders of consciousness using brain functional networks and clinical characteristics. *eLife*, vol. 7, Article number e36173, 2018. DOI: 10.7554/eLife.36173.

[79] Y. B. Liu, L. Yue, S. F. Xiao, W. Yang, D. G. Shen, M. X. Liu. Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages. *Medical Image Analysis*, vol. 75, Article number 102266, 2022. DOI: 10.1016/j.media.2021.102266.

[80] S. X. Luo, D. Martinez, K. M. Carpenter, M. Slifstein, E. V. Nunes. Multimodal predictive modeling of individual treatment outcome in cocaine dependence with combined neuroimaging and behavioral predictors. *Drug and Alcohol Dependence*, vol. 143, pp. 29–35, 2014. DOI: 10.1016/j.drugalcdep.2014.04.030.

[81] A. Billot, S. Lai, M. Varkanitsa, E. J. Braun, B. Rapp, T. B. Parrish, J. Higgins, A. S. Kurani, D. Caplan, C. K. Thompson, P. Ishwar, M. Betke, S. Kiran. Multimodal neural and behavioral data predict response to rehabilitation in chronic poststroke aphasia. *Stroke*, vol. 53, no. 5, pp. 1606–1614, 2022. DOI: 10.1161/strokeaha.121.036749.

[82] M. M. Schmitgen, I. Niedtfeld, R. Schmitt, F. Mancke, D. Winter, C. Schmahl, S. C. Herpertz. Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging. *Brain and Behavior*, vol. 9, no. 9, Article number e01384, 2019. DOI: 10.1002/brb3.1384.

[83] S. Marek, B. Tervo-Clemmens, F. J. Calabro, D. F. Montez, B. P. Kay, A. S. Hatoum, M. R. Donohue, W. Foran, R. L. Miller, T. J. Hendrickson, S. M. Malone, S. Kandala, E. Feczko, O. Miranda-Dominguez, A. M. Graham, E. A. Earl, A. J. Perrone, M. Cordova, O. Doyle, L. A. Moore, G. M. Conan, J. Uriarte, K. Snider, B. J. Lynch, J. C. Wilgenbusch, T. Pengo, A. Tam, J. Z. Chen, D. J. Newbold, A. N. Zheng, N. A. Seider, A. N. Van, A. Metoki, R. J. Chauvin, T. O. Laumann, D. J. Greene, S. E. Petersen, H. Garavan, W. K. Thompson, T. E. Nichols, B. T. T. Yeo, D. M. Barch, B. Luna, D. A. Fair, N. U. F. Dosenbach. Reproducible brain-wide association studies require thousands of individuals. *Nature*, vol. 603, no. 7902, pp. 654–660, 2022. DOI: 10.1038/s41586-022-04492-9.

[84] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins. UK biobank: An open access resource for identifying the causes of a wide range of

complex diseases of middle and old age. *PLoS Medicine*, vol. 12, no. 3, Article number e1001779, 2015. DOI: 10.1371/journal.pmed.1001779.

[85] B. J. Casey, T. Cannonier, M. I. Conley, A. O. Cohen, D. M. Barch, M. M. Heitzeg, M. E. Soules, T. Teslovich, D. V. Dellarco, H. Garavan, C. A. Orr, T. D. Wager, M. T. Banich, N. K. Speer, M. T. Sutherland, M. C. Riedel, A. S. Dick, J. M. Bjork, K. M. Thomas, B. Chaarani, M. H. Mejia, D. J. Jr. Hagler, M. Daniela Cornejo, C. S. Sicat, M. P. Harms, N. U. F. Dosenbach, M. Rosenberg, E. Earl, H. Bartsch, R. Watts, J. R. Polimeni, J. M. Kuperman, D. A. Fair, A. M. Dale. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, vol. 32, pp. 43–54, 2018. DOI: 10.1016/j.dcn.2018.03.001.

[86] D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. E. J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss, S. Della Penna, D. Feinberg, M. F. Glasser, N. Harel, A. C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. E. Petersen, F. Prior, B. L. Schlaggar, S. M. Smith, A. Z. Snyder, J. Xu, E. Yacoub, WU-Minn HCP Consortium. The human connectome project: A data acquisition perspective. *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, 2012. DOI: 10.1016/j.neuroimage.2012.02.018.

[87] Q. Y. Zhong, A. N. Li, R. Jin, D. J. Zhang, X. N. Li, X. Y. Jia, Z. H. Ding, P. Luo, C. Zhou, C. Y. Jiang, Z. Feng, Z. H. Zhang, H. Gong, J. Yuan, Q. M. Luo. High-definition imaging using line-illumination modulation microscopy. *Nature Methods*, vol. 18, no. 3, pp. 309–315, 2021. DOI: 10.1038/s41592-021-01074-x.

[88] M. Goubran, C. Leuze, B. Hsueh, M. Aswendt, L. Ye, Q. Y. Tian, M. Y. Cheng, A. Crow, G. K. Steinberg, J. A. McNab, K. Deisseroth, M. Zeineh. Multimodal image registration and connectivity analysis for integration of connectomic data from microscopy to MRI. *Nature Communications*, vol. 10, no. 1, Article number 5504, 2019. DOI: 10.1038/s41467-019-13374-0.

[89] M. Zubair, S. R. Murris, K. Isa, H. Onoe, Y. Koshimizu, K. Kobayashi, W. Vanduffel, T. Isa. Divergent whole brain projections from the ventral midbrain in macaques. *Cerebral Cortex*, vol. 31, no. 6, pp. 2913–2931, 2021. DOI: 10.1093/cercor/bhaa399.

[90] M. Mancini, A. Casamitjana, L. Peter, E. Robinson, S. Crampsie, D. L. Thomas, J. L. Holton, Z. Jaunmuktane, J. E. Iglesias. A multimodal computational pipeline for 3D histology of the human brain. *Scientific Reports*, vol. 10, no. 1, Article number 13839, 2020. DOI: 10.1038/s41598-020-69163-z.

[91] K. Amunts, C. Lepage, L. Borgeat, H. Mohlberg, T. Dickscheid, M. É. Rousseau, S. Bludau, P. L. Bazin, L. B. Lewis, A. M. Oros-Peusquens, N. J. Shah, T. Lippert, K. Zilles, A. C. Evans. BigBrain: An ultrahigh-resolution 3D human brain model. *Science*, vol. 340, no. 6139, pp. 1472–1475, 2013. DOI: 10.1126/science.1235381.

[92] J. N. Acosta, G. J. Falcone, P. Rajpurkar, E. J. Topol. Multimodal biomedical AI. *Nature Medicine*, vol. 28, no. 9, pp. 1773–1784, 2022. DOI: 10.1038/s41591-022-01981-2.

[93] A. Shapson-Coe, M. Januszewski, D. R. Berger, A. Pope, Y. L. Wu, T. Blakely, R. L. Schalek, P. Li, S. H. Wang, J. Maitin-Shepard, N. Karlupia, S. Dorkenwald, E. Sjostedt, L. Leavitt, D. Lee, L. Bailey, A. Fitzmaurice, R. Kar, B. Field, H. Wu, J. Wagner-Carena, D. Aley, J. Lau, Z. D. Lin, D. Wei, H. Pfister, A. Peleg, V. Jain, J. W. Lichtman. A connectomic study of a petascale fragment of human cerebral cortex. *BioRxiv*, to be published. DOI: 10.1101/2021.05.29.446289.

[94] K. Amunts, T. Lippert. Brain research challenges supercomputing. *Science*, vol. 374, no. 6571, pp. 1054–1055, 2021. DOI: 10.1126/science.abl8519.

[95] K. Amunts, C. Ebell, J. Muller, M. Telefont, A. Knoll, T. Lippert. The human brain project: Creating a European research infrastructure to decode the human brain. *Neuron*, vol. 92, no. 3, pp. 574–581, 2016. DOI: 10.1016/j.neuron.2016.10.046.

[96] J. R. Ecker, D. H. Geschwind, A. R. Kriegstein, J. Ngai, P. Osten, D. Polioudakis, A. Regev, N. Sestan, I. R. Wickersham, H. K. Zeng. The BRAIN initiative cell census consortium: Lessons learned toward generating a comprehensive brain cell atlas. *Neuron*, vol. 96, no. 3, pp. 542–557, 2017. DOI: 10.1016/j.neuron.2017.10.007.

Na Luo received the Ph. D. degree in pattern recognition and intelligent system from University of Chinese Academy of Sciences, China in 2020. She is currently an associate professor with Institute of Automation, Chinese Academy of Sciences, China.

Her research interests include fusing multi-modal brain imaging and multi-omics data to help extend the understanding of brain in both health and disease status.

E-mail: luona2015@ia.ac.cn

ORCID ID: 0000-0002-2500-2083

Weiyang Shi received the Ph. D. degree in pattern recognition and intelligent system from Institute of Automation, Chinese Academy of Sciences, China in 2023. He is currently an assistant professor with Institute of Automation, Chinese Academy of Sciences, China.

His research interests include machine learning in medical image analysis, computational psychiatry and brain-inspired artificial intelligence.

E-mail: shiweiyang2017@ia.ac.cn

ORCID ID: 0000-0002-1805-6884

Zhengyi Yang received the Ph. D. degree in mechanical engineering from University of Hong Kong, China in 2003. He is currently an associate professor with Institute of Automation, Chinese Academy of Sciences, China.

His research interests include medical image analysis and applications on robot.

E-mail: zhengyi.yang@nlpr.ia.ac.cn

Ming Song received the Ph. D. degree in pattern recognition and intelligent system from University of Chinese Academy of Sciences, China in 2008. He is currently a professor with Institute of Automation, Chinese Academy of Sciences, China.

His research interests include brain imaging and analysis, brain network and brain-computer interface.

E-mail: msong@nlpr.ia.ac.cn

Tianzi Jiang (Fellow, IEEE) received the B. Sc. degree in computational mathematics from Lanzhou University, China in 1984, and the Ph. D. degree in computational mathematics from Zhejiang University, China in 1994. He is currently a member of the Academy of Europe, senior research professor, and director of Beijing Key Laboratory of Brainnetome and director of the Brainnetome Center at Institute of Automation, Chinese Academy of Sciences, China, and Adjunctive Senior Investigator of Zhejiang Laboratory in Hangzhou, China. He was the recipient of Hermann von Helmholtz Award (Lifetime Contribution Awards) of International Neural Network Society, Turan Itil Career Contribution Award of EEG & Clinical Neuroscience Society, Wu Wen-Jun AI Distinguishing Contribution Award of the Chinese Association of Artificial Intelligence, Natural Science Award of China, Beijing Natural Science Award (the first class), and Wu Wen-Jun AI Natural Science Award. He was elected as a fellow of IAPR and AIMBE. He is currently the Chair of Organization of Human Brain Mapping, an Associate Editor for *IEEE Transactions on Cognitive and Developmental Systems*, *Neuroscience Bulletin*, and Action Editor for *Neural Networks*.

His research interests include brainnetome atlas, neuroimaging, and their clinical applications in brain disorders.

E-mail: jiangtz@nlpr.ia.ac.cn (Corresponding author)
ORCID ID: 0000-0001-9531-291X

Machine Intelligence Research

Citation: N. Luo, W. Shi, Z. Yang, M. Song, T. Jiang. Multimodal fusion of brain imaging data: methods and applications. *Machine Inteligence Research*, vol.21, no.1, pp.136−152, 2024. https://doi.org/10.1007/s11633−023−1442−8

Articles may interest you

Machine learning for brain imaging genomics methods: a review. *Machine Intelligence Research*, vol.20, no.1, pp.57-78, 2023.
DOI: 10.1007/s11633-022-1361-0

Federated learning on multimodal data: a comprehensive survey. *Machine Intelligence Research*, vol.20, no.4, pp.539-553, 2023.
DOI: 10.1007/s11633-022-1398-0

Multimodal biometric fusion algorithm based on ranking partition collision theory. *Machine Intelligence Research*, vol.20, no.6, pp.884-896, 2023.
DOI: 10.1007/s11633-022-1403-7

Multitask learning with multiscale residual attention for brain tumor segmentation and classification. *Machine Intelligence Research*, vol.20, no.6, pp.897-908, 2023.
DOI: 10.1007/s11633-022-1392-6

Towards a new paradigm for brain-inspired computer vision. *Machine Intelligence Research*, vol.19, no.5, pp.412-424, 2022.
DOI: 10.1007/s11633-022-1370-z

Brain-inspired intelligent robotics: theoretical analysis and systematic application. *Machine Intelligence Research*, vol.20, no.1, pp.1-18, 2023.
DOI: 10.1007/s11633-022-1390-8

A dynamic resource allocation strategy with reinforcement learning for multimodal multi-objective optimization. *Machine Intelligence Research*, vol.19, no.2, pp.138-152, 2022.
DOI: 10.1007/s11633-022-1314-7

WeChat: MIR

Twitter: MIR_Journal