Application of Deep Learning in Resting State Functional Magnetic Resonance Imaging Data Analysis

by

Donglin Wang

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy in Computational Sciences

Middle Tennessee State University

February, 2021

Dissertation Committee:

Dr. Don Hong, Chair

Dr. Qiang Wu, Co-chair

Dr. Zachariah Sinkala

Dr. Jing Kong

To the gift that God has granted me, my parents and my family.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my dissertation advisors, Dr. Don Hong and Dr. Qiang Wu, for their continuous encouragement, valuable guidance and discussions through the entire process with me. They have continuously provided me with timely guidance and suggestions to help me overcome the challenges one by one. Their broad knowledge and enthusiasm for research have deeply influenced me. I am inspired greatly by their devotion and passion in their research and feel so lucky to have them as advisors.

I would also like to give my sincere thanks to my dissertation committee members, Dr. Zachariah Sinkala and Dr. Jing Kong, for their insightful comments and suggestions. I give my special thanks to Dr. Sinkala for giving me encouragement and support from time to time. I also owe special thanks to Dr. Kong for providing the computer with a graphics processing unit (GPU) for my research. I really appreciate his support and friendship.

I am also grateful to the computational science Ph.D. program at Middle Tennessee State University, especially the program director, Dr. Wallin John, for introducing me to this program and his helpful guidance on courses. I would also like to give my thanks to my department for supporting me during my entire study period. In addition, I am grateful to my friends in the department for their kindness and support which makes my Ph.D study experience more enjoyable and memorable.

Last but not least, I would like to express my gratitude to my dear wife, Wandi, my

lovely kids, Forrest and Nathan, for their unconditional love, support and encouragement, without which it would have been impossible for me to finish this dissertation. I love you all!

ABSTRACT

Since the advent of deep learning, it has been used in a lot of fields like computer version, image recognition and speech recognition, etc. It has been made many achievements due to its outstanding ability on both classification and regression tasks. In this dissertation, we explore the use of both existing and newly designed deep learning techniques in resting state functional magnetic resonance imaging data analysis.

In resting state, the human brain is still active and different regions are functionally connected to each other through the intrinsic networks called Resting State Networks (RSNs). It is important to locate the RSNs and understand how these RSNs function between each other. Especially, in the clinical fields this will help provide better and more precise treatment plans for some mental disorder diseases. In this dissertation the Graph Sample and aggregate (graphSAGE), a typical graph neural network, is proposed for extracting RSNs. It treats each resting state functional magnetic imaging as a graph while analyzing the data. Compared with the classical methods such as seed-based method, independent component method, and dictionary learning method, the application of grahSAGE gives robust results. The classical methods need some user-defined prior values, such as thresholds p values for both single subject analysis and group level analysis and the number of components for independent component analysis (ICA) and dictionary learning analysis. The results depend on these prior assumptions and therefore subjective more or less. On contrast, graphSAGE does not need prior assumptions. The results are more objective and robust. Moreover, it can perform single subject analysis and group subject analysis simultaneously.

Attention Deficit Hyperactivity Disorder (ADHD) is a type of mental health disorder. It is a disease that can be seen from children to adults and affects the

patient's normal life. Therefore, the accurate diagnosis as early as possible is very important for the treatment of the patient in clinical applications. Some traditional classification methods, although having been shown powerful in many other classification tasks, are not as successful in the application of ADHD classification. In this dissertation, we designed two novel deep learning approaches, called ICA-CNN method and correlation-autoencoder method, respectively, for the ADHD classification task.

The ICA-CNN method first extracts independent components from each subject. These independent components are then fed into a convolutional neural network as input features to classify the ADHD patient from typical controls.

The correlation-autoencoder method calculates the correlation between regions of interest of the brain, which is then input of an autoencoder to learn the latent features. After the latent features are learned, they are used in classification task by a new neural network. Both methods significantly outperform the classical methods such as logistic regression, support vector machines, and other methods used in previous studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Research on human brain activity has always been one of the interests for scientists in neuroscience fields. The human brain is the most developed thinking organ of the human body. It controls all the activities of the human body and regulates the balance between the body and the surrounding environment. Therefore, the brain is the basis for all high-level neural activities in humans. Generally speaking, the structure of human brain can be split into three main parts: cerebrum, cerebellum and brainstem [2, 51]. The cerebrum is the largest part. It is composed of left hemisphere and right hemisphere divided by longitudinal fissure, and the cerebrum mainly includes the cerebral cortex, brain medulla and basal nucleus. The cerebral cortex is the gray matter covering the surface of the brain, and is mainly composed of neurons. The deep part of the cortex is composed of medulla or white matter formed by nerve fibers. The cortex is irregularly distributed on the surface of the brain, some are called gyri with unique folds or bumps and others are called sulci with grooves. Figure 1.1 shows the brief structure of human brain. All the gyri and sulci form the cortex into different brain functions. Although the brain only accounts for 2% of the body weight, the oxygen consumption on average is about 20% of the total body oxygen consumption [136].

The understanding of the anatomical structure and the working mechanism of the human brain will help us further understand various phenomena from sociology, psychology, pathology and neurology, especially for the study of diseases related to

neurological disorders and cognitive functions such as Parkinson's disease, depression, schizophrenia, multiple sclerosis. [39, 43, 65, 108, 118, 192]. In addition to the study of the anatomical structure of the brain, the functional study of the brain is currently one of the mainstreams of neurology.

There are several imaging methods that can reflect brain function. Usually there are two common types, one is invasive methods and the other is non-invasive methods. Non-invasive methods mainly include electroencephalography (EEG), magnetoencephalography (MEG) based on electromagnetic signal detection, functional magnetic resonance imaging (fMRI) based on measurement of blood oxygen level dependence, near-infrared spectroscopy (NIRS) based on metabolic level Measured near infrared spectroscopy and transcranial magnetic stimulation (TMS) based on magnetic fields to stimulate nerve cells in the brain [42]. Invasive methods mainly include positron emission tomography (PET) based on injecting a radionuclide labeled compound into the brain and measuring positron annihilation rays to obtain tomographic images [42], electrocorticography (ECoG) or intracranial electroencephalography (iEEG). As the name implies, the electrodes of the subcortical EEG are placed under the scalp. Most subcortical EEG studies are conducted in patients. When the patient must undergo craniotomy and implant electrodes, scientists can only collect relevant data after obtaining the patient's consent [13, 74, 125]. The invasive methods will cause more or less unknowable damage to the brain, although it may be very low. Most of the current researches on brain function use non-invasive methods. The results of all the above measurement methods can be displayed as images in two-dimensional or three-dimensional space.

Different neuroimaging methods have different spatial and temporal resolutions for measuring human brain functions, see Figure 1.2 for comparison among different

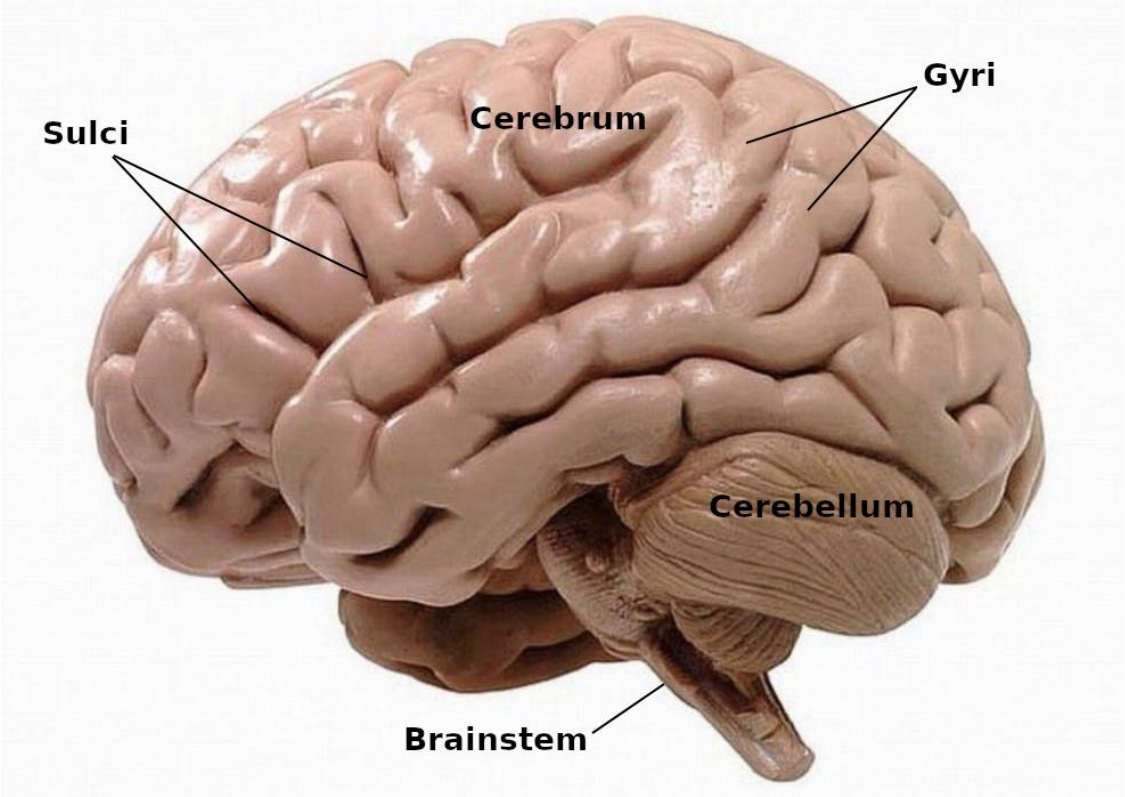


Figure 1.1. Structure of human brain: cerebrum, cerebellum and brainstem

techniques. Different research purposes or practical applications require different techniques to obtain corresponding brain images. While TMS, EGG, and MEG have a higher spatial resolution but a low temporal resolution, and PET has a high temporal resolution and spatial resolution, fMRI has a moderate temporal resolution and spatial resolution which makes it favored in many brain function studies [42, 66].

1.2 Functional magnetic resonance imaging (fMRI)

Since its development in 1990 functional magnetic resonance imaging (fMRI) [12, 23, 93, 121] is becoming a popular neuroimaging technique to measure human brain activities due to its safety, non-invasive nature and relatively high spatial resolution

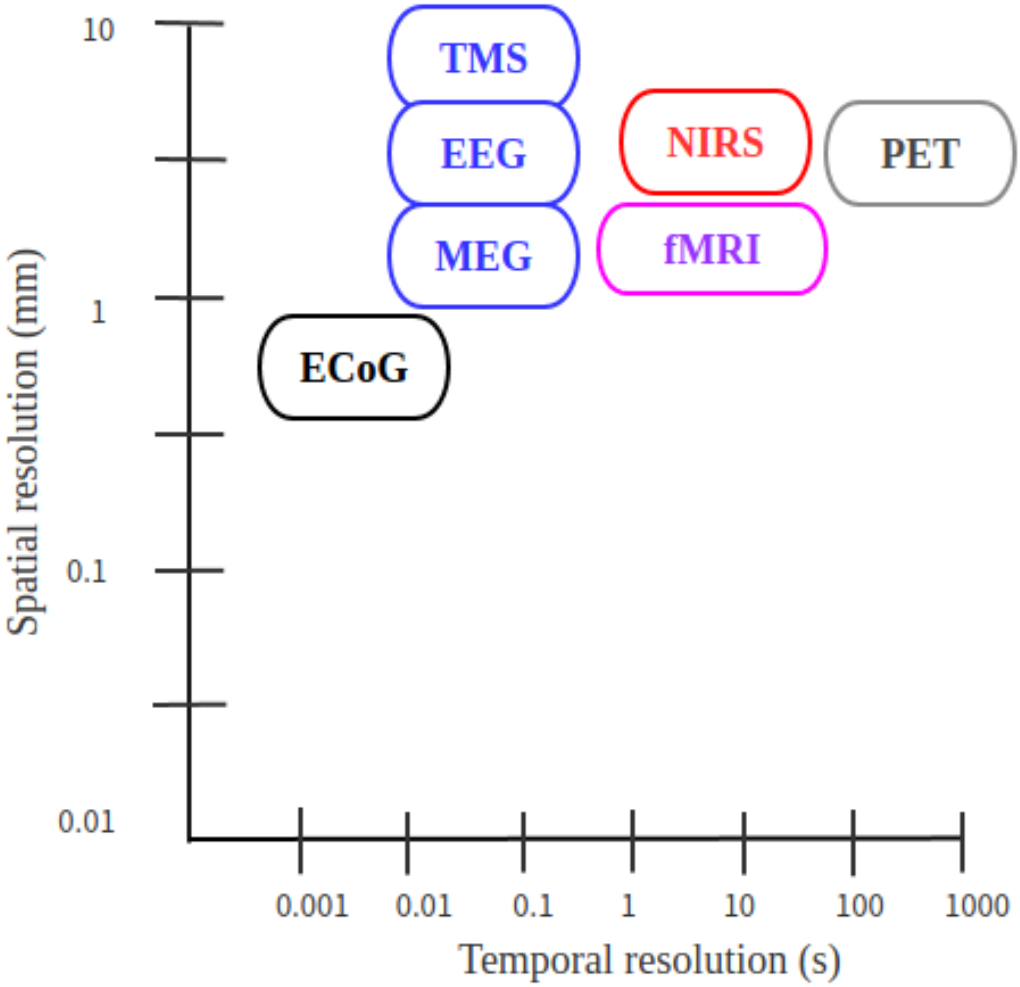


Figure 1.2. Spatial and temporal resolution among different neuroimaging techniques

which make it play an important role in both clinical and academic researchers [66, 115, 130]. It is an application based on magnetic resonance imaging (MRI) technique [179] and depends on Blood Oxygenation Level Dependent (BOLD) contrast [122]. There are two types of hemoglobins in blood: oxyhemoglobin and deoxyhemoglobin. Here is the simple idea how fMRI can be taken. When any part of human brain is activated, this kind of neural activity will consume oxygen from oxyhemoglobin of

blood. It causes increasing of deoxyhemoglobin and decreasing of oxyhemoglobin, and then oxygen needs to be transported through cerebral blood vessels to supply the consumption. Therefore the activity of brain will be accompanied by changes of concentration of oxygen in blood. Under the interference from external magnetic field this change can be created into images [19, 28, 29, 66, 95, 121]. See Figure 1.3 for an illustration. In general, the stronger the BOLD signal is in a certain area of the brain, the stronger the nerve activity is in that area.

Functional magnetic resonance imaging (fMRI) can incorporate many different types of contrast such as T1, T1 weighted(T1*), T2, T2 weighted(T2*) and fluid attenuated inversion recovery (FLAIR). All the types of images are based on two values; one is called repetition time(TR) which is the interval time between two continuous pulse sequence applied to the subject, the other is called time to echo(TE) which is the time between the radio frequency pulse emission and the acquisition of the echo signal. With the radio frequency pulses, the hydrogen proton absorption energy in human tissue is in an excited state. After the radio frequency pulse is terminated, the protons in the excited state return to their original state. This process is called relaxation. There are two types of relaxation of T1 and T2; T1 is longitudinal relaxation time of the excited protons returning to equilibrium and T2 is transverse relaxation time of the excited protons reaching equilibrium. By adjusting TR and TE an image highlighting a certain tissue can be obtained and this kind of images are called weighted images [22, 33, 45, 69, 76, 93, 100, 114, 142]. Table 1.1 gives a simple illustration for different types of images with different TR and TE. Each image has a different usage in different medical applications or research fields.

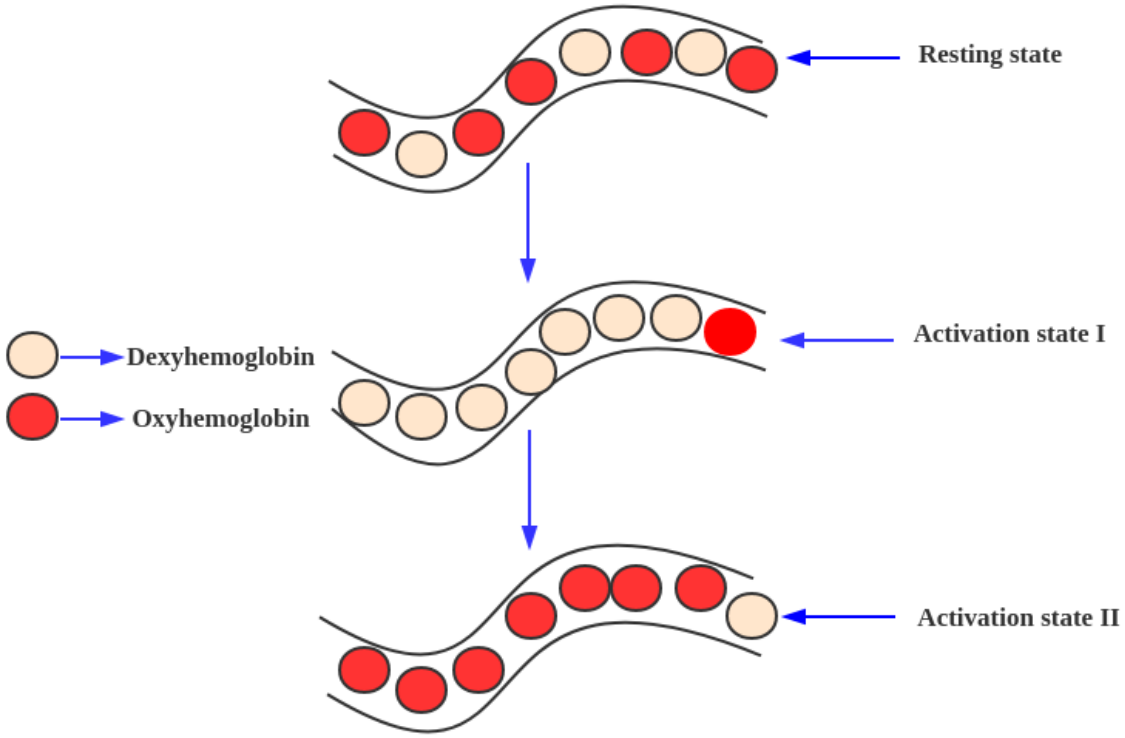


Figure 1.3. Simulation illustration of state of oxyhemoglobin and dexyhemoglobin in region of interest (ROI): *Resting state means no activation and oxyhemoglobin and dexyhemoglobin randomly flow; activation state I means the ROI is activated with oxyhemoglobin decreasing and deoxyhemoglobin increasing; activation II means a number of oxyhemoglobin is added to this ROI in a very short time.*

1.3 Data preprocessing and analysis of fMRI in literature

Since fMRI needS to scan the whole brain, it contains a three-dimensional spatial brain image. In addition it also needs to scan the brain for a while and this causes another dimension of time. Totally fMRI constitutes a four-dimensional data. To be specific, fMRI is a time series data. During the scanning process, different types of

Table 1.1. Approximation of TR and TE to be used for images [134]

	TR(msec)	TE(msec)	Description
T1-weighted image	500	14	short TR and short TE
T2-weighted image	4000	90	long TR and long TE
FLAIR image	9000	114	very long TR and very long TE

noise could be generated. For example, the noise could be from head movement of the subjects, the spontaneous fluctuation of neurons in the brain and some non-neuronal signals such as heartbeat and breathing, the differences in functional magnetic resonance imaging objects and the interference signals generated by the limitations of the magnetic resonance imaging instrument itself. The noise can affect or even destroy brain signals [104,106,133], so a preprocessing of fMRI data is necessary before doing any statistical analysis. With this process, to the greatest extent all kinds of noise from the fMRI signal will be removed. The preprocessing of fMRI usually follows but is not limited to the standard steps as follows [86,88,96,131,157]:

- **Distortion correction**. The raw image data sometimes shows some unusually steep peaks due to the inhomogeneity of the main magnetic field or body movement. For this kind of images, one will either use specific models to remove the artifacts or delete these images.

- **Slice timing correction**. With one TR interval, a complete scan for the whole brain is acquired by several slices in sequential or in interleaved order, and the acquisition time of each slice is inconsistent, so in order to eliminate the difference in the acquisition time of each slice, the time of acquiring each slice must be corrected to the same time point.

- **Motion correction**. Head movement is also a big problem for the subsequent

statistical analysis. During multiple scans, one assumes that the position of each voxel in the brain is unchanged. If the subject's head moves even slightly, the voxel may correspond to other positions in the brain for the next scan . Therefore, in order to avoid this problem, one need to convert and fix all the voxels in the brain in the same position all time.

- **Co-registration**. It refers to the registration of high-resolution structural brain-extracted image (T1) and functional images for the same subject. Usually, a structural image T1 is captured at the beginning of each round of the experiment. The co-registration makes each part of the brain more detailed. Meanwhile it is used for the further registration to the standard coordinate system.

- **Normalization**. The human brain has certain individual differences in anatomical structure such as shape and size. Before any statistical analysis, the same voxel position in all brain images from each subject must meet the same anatomical position, so all the brain images of the subjects should be transformed into a standardized template with the same size and direction, that is, on the same coordinate system. This process is called normalization. At present, there are two main standard brain templates. One is the Talairach space [160,161], which is based a single individual originally. The other is called Montreal Neurological Institute Brain Atlas (MNI) [58,94]. There are two types of templates: MNI305 and MNI152 templates. There are four methods to do normalization and they are landmark-based methods, volume-based methods, computational anatomy methods and surface-based methods.

- **Temporal filtering**. It help reduce the low-frequency noise from external and internal sources to improve the power of statistical analysis. Usually the noise

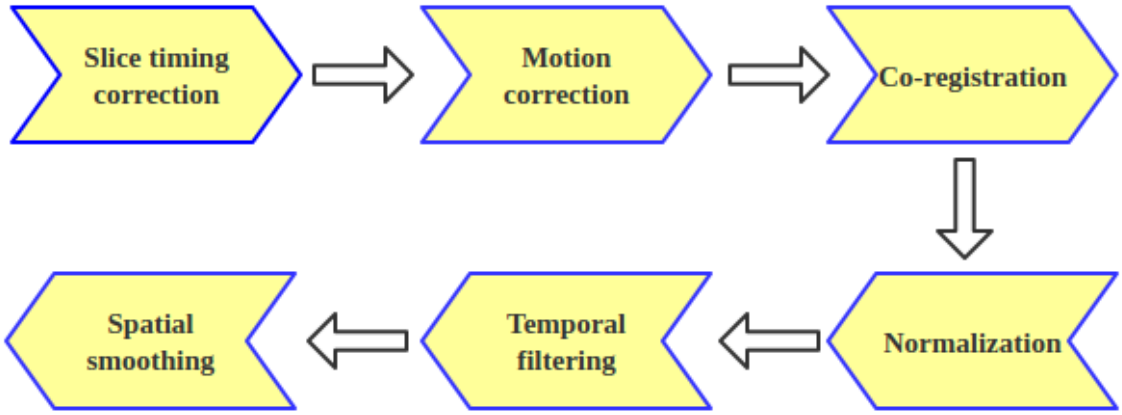


Figure 1.4. Illustration of a standard preproecessing of fMRI data

signal is from physiological noise from the subject and noise related to scanner.

- **Spatial smoothing**.

  This process of spatial smoothing mainly reduces the noise to improve signal to noise ratio (SNR) and increases the power of statistical analysis and the possibility of Gaussian distribution. The Gaussian kernel often is used when doing spatial smoothing. The size of the smoothing kernel or smooth range is determined by the full width at half maximum (FWHM). For the Gaussian distribution, the relationship between the FWHM and the standard deviation σ is as follows: [178]:

$$\sigma = \frac{FWHM}{2\sqrt{2\ln(2)}} \tag{1.1}$$

There are several softwares and packages which can deal with data preprocessing of fMRI data such as AFNI [41], SPM [128], FSL [87], FreeSurfer [59], ANTs [10], Nilearn [1], fMRIPrep [57], R [135], Python [169] and Matlab [111]. Figure 1.4 shows a standard workflow of preprocessing of fMRI.

After preprocessing of fMRI, one can start the statistical analysis. For fMRI analysis, there are mainly three fields [104, 129, 153], one is to detect activation of ROIs by a certain task, the second is to detect brain connectivity, especially functional connectivity, and the the third is to classify and predict psychological or disease state such as Alzheimer's disease [3,11,151], depression [171,181], anxiety [71,126], attention deficit and hyperactivity disorder (ADHD) [31, 113], and etc. There are two types of data of fMRI based on whether a task is involved. If a task is involved, the data is task-related fMRI. Otherwise the data is resting state fMRI (rs-fMRI). Recently rs-fMRI has been becoming more and more a popular research field [21] since the first study of rs-fMRI [20]. Resting state fMRI analysis mainly focuses on functional connectivity analysis between distinct brain regions and on classifying and predicting psychological or disease state [54, 73, 175, 176].

The article [168] gave a details review on functional connectivity of resting state fMRI and discussed the value in examining connectivity diseases such as Alzheimer's disease and schizophrenia with graph theory technique. In [92] the authors proposed an integrated strategy for improving functional connectivity mapping using multiecho fMRI with spatial independent component analysis (ME-ICA), which improved the functional connectivity on specificity compared with classical connectivity estimation methods. In [48] the study found 10 patterns with potential functional relevance including deafult mode network (DMN) and showed the Consistent resting state networks across healthy subjects. The article [52] used a robust method called probabilistic independent component analysis (PICA) to identify resting state networks (RSNs), and at least five different RSNs were identified across different subjects. In [144] the fully exploratory network independent component analysis (FENICA) was used on resting state fMRI data based on 28 healthy subjects to result in eight

distinct RSNs without manual selection of individual subject components. In [82] the authors combined the graph measures with a support vector machine model (SVM) to classify accurately the mild cognitive impairment (MCI) patients who develop to Alzheimer's disease (AD) from those who do not develop to Alzheimer's disease (AD). In the study [137] the authors used static and dynamic brain connectivity features to classify schizophrenia from bipolar patients and found that the accuracy of classification was higher based on the dynamic connectivity features than the static connectivity features. In [116] the authors proposed a connectome-convolutional neural network (CCNN) for functional connectome classification, and the model can efficiently classify mild cognitive impairment patients from control group. The article [8] used a convolutional neural network to identify Attention Deficit and Hyperactivity Disorder (ADHD) based on the correlation computed from different default mode network (DMN) regions, and the proposed method significantly improved the accuracy of identification of ADHD.

In this dissertation, we focus on two research objectives in the context of rs-fMRI data analysis. First, we propose the use of graph neural networks for resting state networks detection and compare it with classical methods. Second, we propose two deep learning approaches for the classification of Attention Deficit and Hyperactivity Disorder (ADHD).

Extracting or detecting resting state networks (RSNs) based on resting state functional magnetic resonance imaging (rs-fMRI) data is challenging. Classical methods such as seed-based correlation, independent component analysis, and dictionary learning, usually rely on user-defined prior assumptions and the results are subjective more or less. To overcome this problem and improve the robustness, we propose to use a graph neural network (GNN) called graphSAGE [72] to extract the default mode network

(DMN), one of the commonly recognized RSNs. Compared with classical methods, the GNN model extracts not only DMN but also several other RSNs. The results do not depend on prior assumptions and are therefore objective and robust.

Attention Deficit and Hyperactivity Disorder (ADHD) is a type of mental disorder. It happens both in children and adults. Accurate diagnosis and early treatment are important. The traditional methods, such as logistic regression model and support vector machine model, although have been shown powerful in many others classification tasks, their application in ADHD classification based on rs-fMRI data seems far from satisfactory. A plausible reason may be that the raw rs-fMRI data is very noisy while the traditional classification methods are not as powerful to denoise the data. The deep learning is an end-to-end learning method developed in the last decades. It can learn directly from raw data and automatically extract latent features based on a multi-layer nonlinear neural network and help denoise images. This motivates us to incorporate the deep neural network into the learning process and propose two novel deep learning methods for ADHD classification. Both methods significantly improve the ADHD classification accuracy and outperform the traditional methods.

1.4 Outline of this dissertation

In this chapter, the techniques of getting brain images are simply introduced, what functional maganetic resonance image (fMRI) is, how to do preprocessing of fMRI before statistical analysis, and we introduced the two fields of resting state fMRI analysis. The rest of this dissertation is organized as follows.

In Chapter 2 a deep learning algorithm called graph sample and aggregate (graphSAGE) [72] is proposed to extract the resting state networks (RSNs), mainly the default mode network (DMN) and compared with several common methods, namely seed-based

correlation method (SBC), the independent component analysis method (ICA), the dictionary learning method. The application of grahSAGE gives a robust result. The common methods need to make prior thresholds p values for single subject analysis and group level analysis and the number of components for ICA and dictionary learning. The result of analysis depends on these prior assumptions more or less. On contrast, graphSAGE does not need these assumptions and considers the single subject analysis and group subjects analysis simultaneously.

In Chapter 3 two methods called independent component analysis with convolutional neural network (ICA-CNN) and correlation with autoencoder model (Correlation-autoencoder) are proposed to classify the Attention Deficit Hyperactivity Disorder (ADHD) from typical controls, and then compared the performance with classical binary classification models such as logistic regression model and support vector machine model. The results of analysis show that the two proposed methods outperform the classical methods and other methods in previous studies.

In Chapter 4 The main contributions of this dissertation are summarised and the future work is discussed.

CHAPTER 2

FUNCTIONAL CONNECTIVITY TECHNIQUES

2.1 Overview

Functional connectivity is usually based on two approaches: task-based fMRI analysis and resting state fMRI analysis. In general, task-based fMRI not only depends on the ability of the subjects to follow the task procedure but also depends on the good design of the experiment, especially in clinical applications [6,46,190]. On the contrary, resting functional magnetic resonance imaging can measure spontaneous fluctuations of human brain and reflect the relationship among different networks [20, 61]. Thus, rs-fMRI is applicable to situations in which task-related fMRI may provide insufficient information or sometimes fail to perform [152]. In resting state the brain is still constantly active and the parts of the brain are connected by its intrinsic connections, and some connections even show stronger relationship when doing certain tasks, so it is more reasonable to use resting state to study basic functional connections, especially it is used to identify so called intrinsic connectivity networks or resting state networks (RSNs) [14, 52, 68, 147]. As [14, 38, 52, 61, 154] mentioned, RSNs are located on grey matter regions of human brain, and these RSNs reflect core perceptual and cognitive processes of brain functional systems.

There are several common RSNs [38, 101] which are often identified from resting state fMRI study. The most basic RSN is the default mode network (DMN) [7, 48, 68, 149, 164, 167] which is active when human brain is at rest without any external or attention-demanding tasks. It mainly includes the posterior cingulate cortex (PCC) and precuneus, the medial prefrontal cortex (MPFC), and the inferior parietal cortex.

Somatomotor network (SMN) is another network which includes somatosensory and motor regions related to motor tasks [20,34]. Visual network (VIN) includes main part of the occipital cortex, which deals with related visual activity [52, 80, 132, 154, 164]. The dorsal attention network (DAN), also known as visuospatial attention network, is to provide attention orientation and to participate in external tasks, it is mainly located in the intraparietal sulcus and the connection area between the central anterior sulcus and the forehead sulcus [17,84,102,148,149,172,173]. Salience netwrok (SAN) mainly evaluates external stimuli and internal events and helps directing attention by switching to related processing systems. This network includes the left and right anterior insula and the middle of anterior cingulate cortex (ACC) [24, 35, 77, 139, 146, 155, 165, 188]. Frontoparietal network (FPN) includes left and right parts, with the left part mainly dealing with voice and memory and right part mainly dealing with activity-related events. FPN mainly includes the dorsolateral prefrontal cortex (dlPFC), the dorsolateral prefrontal cortex and the inferior parietal lobule [97, 110, 145, 148, 172, 189]. Other networks from different studies are also identified such as auditory network [77, 188], language network [24, 83] and etc. The article [78] also gives a detailed description of RSNs and shows ten RSNs based on normal wakeful resting state which includes default mode network, left executive control network, right executive control network, salience network, sensorimotor network, auditory network, visual medial network, visual lateral network, visual occipital network and cerebellum. Figure 2.1 for the details. But it should be mentioned that even in the resting state it is impossible to detect what the human brain is really thinking or imagining about. This is one disadvantage of the resting state analysis [47].

2.2 Common methods review

There are several techniques to identify RSNs from rs-fMRI, which are reviewed below. The rest of the chapter is organized as follows. In Section 2.2.1, we review seed-based correlation method to extract RSNs; in Section 2.2.2 independent component analysis (ICA) is introduced; in Section 2.2.3 the dictionary learning is introduced; in Section 2.3 we proposed to use a graph neural network called graphSAGE [72] to help extract RSNs from resting state fMRI; in Section 2.4 we will compare all the above methods on two real data sets.

2.2.1 Seed-based correlation (SBC)

Seed-based correlation (SBC) method is the first method to be used for identifying resting state networks by [20]. It calculates the Pearson correlation coefficient between two voxels or region of interests (ROI) from the corresponding two time series, denoted as $\boldsymbol{x} = [x_1, x_2, x_3, ..., x_n]$ and $\boldsymbol{y} = [y_1, y_2, y_3, ..., y_n]$. The correlation coefficient is calculated as:

$$r_{\boldsymbol{xy}} = \frac{\sum_{i=1}^{n}(x_i - \overline{\boldsymbol{x}})(y_i - \overline{\boldsymbol{y}})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{\boldsymbol{x}})^2 \sum_{i=1}^{n}(y_i - \overline{\boldsymbol{y}})^2}} \tag{2.1}$$

where $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ are the average values of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

Normally there are about one million voxels or so in one entire brain scan, and the amount is based on the size of voxles and the scanned brain. It is possible to calculate all the correlations between any two voxels, but this will cause numerous pairs of comparisons. More often, this method uses a prior selected seed or ROI as reference and calculate the correlation with the rest voxels of the brain. A big correlation value means a closer connectivity between the two seeds or ROIs. This method has been applied in many research studies and has been identifying

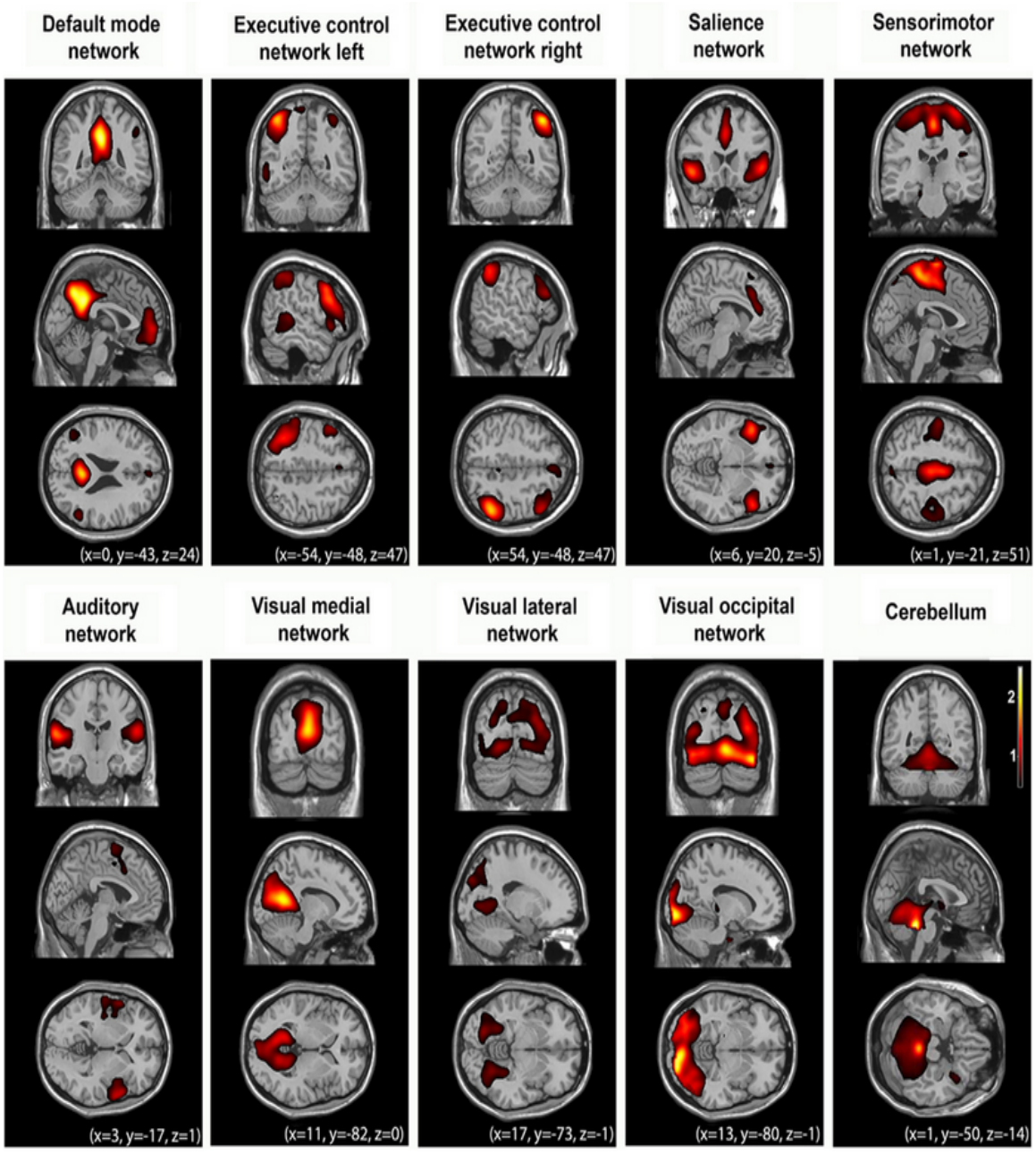


Figure 2.1. Ten RSNs based on fMRI from the article [78]

some functional connectivity based on rs-fMRI. In [60], the authors identified a bilateral dorsal attention system, a right-lateralized ventral attention system, and detected a potential mediating function in the prefrontal cortex. In [62] the authors identified two opposed brain networks based on correlations within each network and anticorrelations between networks, one network is related to task-related activations and the other is related to task-related deactivations. In [53] the authors used sensorimotor networks as the seed to extract similar RSNs from rs-fMRI with those from finger-tapping task fMRI data based on eight healthy volunteers. Besides a prior selected seed or ROI, this method also need a threshold to determine the significant voxels with the seed or ROI.

The main advantages of SCA are its algorithmic simplicity and straightforwd interpretation, which make it play an important role in functional connectivity field. The main disadvantage of this method is that all the relationship is only between the prior selected seed or ROI and the remaining other voxels of the brain, so the univariate analysis method completely ignores the possible relations between other seeds. In other words, some significant information between other voxels is ignored. This restricts more networks to be detected. Therefore, the choice of seed, size and position directly determine the interpretation of the results [26, 37].

2.2.2 Independent component analysis (ICA)

Independent component analysis (ICA) is one of the most popular decomposition method for analyzing fMRI data and it is a data-driven technique. Since this technique was first used in [112], it has been applying in many studies so far. ICA assumes the observed data is a linear combination of sources which are statistically independent [25]. The purpose of ICA is to extract independent components based on optimization

technique, such as maximum likelihood estimation (MLE) [156], minimum of the mutual information between sources (InfoMax) [16] or maximum of the non-gaussianity between sources (FastICA) [85].

Because ICA is looking for non-gaussian sources, most ICA algorithms often use principle component analysis (PCA) [163] to remove gaussian signals in the observed data. The difference between ICA and PCA is that ICA searches statistically independent components with maximum possibility and PCA searches uncorrelated components with orthogonal property and maximum of variance. Since fMRI is a four-dimension data, it has to be transferred into a two-dimension data when using ICA or PCA. ICA analysis can be one of two types of analysis which are spatial ICA and temporal ICA, and different types of analysis determine how to transfer the fMRI data. If the spatially independent components are detected, the columns of the transferred matrix are voxles and the rows are the time points; if the temporally independent components are detected, the columns of the transferred matrix are time points and the rows are the voxels. ICA can split into spatial ICA and temporal ICA, but usually spatial ICA is used often since it can produce as many components as time points [30, 112].

Here is the basic algorithm of spatial ICA [30, 112]. Let $\mathbf{X}$ be a matrix with dimension $t \times n$ ($\mathbf{X} \in \mathbb{R}^{t \times n}$), where t means the number of time points and n means the number of voxels. The goal is to solve the follow equation to find $\mathbf{W}, \mathbf{S}$:

$$\mathbf{X} = \mathbf{WS} \tag{2.2}$$

where $\mathbf{W}$ is called the matrix of mixing coefficient with dimension $t \times p$, $\mathbf{S}$ is called source matrix with dimension $p \times n$, and the rows of matrix $\mathbf{S}$ is representing spatially independent components. The basic algorithm of temporal ICA is similar with spatial ICA [30, 112], where $\mathbf{X}$ is the matrix with dimension $n \times t$ ($\mathbf{X} \in \mathbb{R}^{n \times t}$). The goal is

to solve the follow equation to find $\hat{\mathbf{W}}, \hat{\mathbf{S}}$:

$$\mathbf{X} = \hat{\mathbf{W}}\hat{\mathbf{S}} \tag{2.3}$$

where $\hat{\mathbf{W}}$ is still the matrix of mixing coefficient with dimension $n \times p$, $\hat{\mathbf{S}}$ is the source matrix with dimension $p \times t$, and the rows of matrix $\hat{\mathbf{S}}$ is representing the temporally independent time courses. Usually spatial ICA is used more often than temporal ICA.

There are many techniques applied in ICA. All algorithms fall into one of three groups. One is based on the projection pursuit [156] which basically extracts one component at one time. The second is based on infomax [16] which extracts multiple components at one time in a parallel way. The third is based on maximum likelihood estimation [156] which is a statistical tool to calculate the mixing coefficient matrix to best fit the observed data. There are many practical methods used in literature, for example, in [14] the authors proposed a probabilistic independent component analysis (PICA), compared with typical noise-free ICA, PICA tried to detect the independent components under additive noise interference. In [85] the authors introduced a widely used method called FastICA with kurtosis as the cost function. In [15] the authors used multiple subjects and dual regression to do a group comparison. To date, ICA still plays a important role in resting analysis.

ICA does not provide an order for independent components (ICs) unlike PCA which offers an order for principal components. This property makes it hard to tell which IC is more important. To solve the mixing coefficient $\mathbf{W}$ and sources $\mathbf{S}$, ICA need some prior assumptions. It is also a tricky part to determine the number of ICs, which will cause over-fitting or under-fitting problems.

2.2.3 Dictionary learning (DL)

Dictionary learning (DL) is another linear decomposition technique to extract the fundamental components. This technique emphasizes the sparsity between components unlike ICA which focuses on the independence between components or PCA which focuses on the orthogonal components. The idea of DL is to decompose the observed data into two matrices; a matrix is called dictionary $\mathbf{D}$ which is a collection of atoms and the other is the corresponding spare coefficients $\mathbf{A}$. Here is the basic algorithm of dictionary learning [89, 91, 109, 123, 124], let $\mathbf{Y} = [y_1, y_2, ..., y_n]$ be the observed data with $y_i \in \mathbb{R}^m$ and $\mathbf{Y} \in \mathbb{R}^{m \times n}$, and the dictionary matrix $\mathbf{D} \in \mathbb{R}^{m \times k}$ ($\mathbf{D} = [d_1, d_2, ..., d_k]$) with corresponding sparse coefficient matrix $\mathbf{A} \in \mathbb{R}^{k \times n}$ ($\mathbf{A} = [a_1, a_2, ..., a_n]$). The number of columns k in dictionary $\mathbf{D}$ are the number of atoms. The purpose is to find $\mathbf{D}$ and $\mathbf{A}$ from the following optimization problem:

$$\begin{aligned} \underset{\mathbf{D},\mathbf{A}}{\text{Minimize}} \quad & \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_2^2 \\ \text{Subject to} \quad & \|\mathbf{D}_{.,i}\|_2 \leq 1, \ \forall i = 1, \ldots, k, \\ & \|\mathbf{A}_{.,j}\|_0 \leq C, \ \forall j = 1, \ldots, n, \end{aligned} \tag{2.4}$$

where C is the parameter to control the sparsity for each column of $\mathbf{A}$.

There are several improved ways to solve the optimization problem. For example, in [4] the authors used stochastic gradient descent technique to solve this problem, in [99] the authors used the property of duality to solve the optimization problem. A widely used method is K-SVD method [5] which uses singular value decomposition to generate the K-meaning clustering. It is an iterative process updating sparse matrix and dictionary matrix alternatively.

2.3 Graph neural network

Graph neural network (GNN) is a type of deep neural network which is applied on graph-based data. To date, there are many variants of GNN which are successfully used in many fields. In [184, 191] almost all of GNN models are surveyed in details and compared. In graph theory, an undirected and unweighted graph is represented by $G(V, E)$, in which G means a graph, V means the nodes which often repents voxels or ROIs in fMRI analysis, and E means the edges connecting two nodes, which usually indicates correlation [27, 166] in fMRI analysis. An adjacent matrix $\mathbf{A}$ with dimension $N \times N$ is often used to describe this kind of graph. The entry of $\mathbf{A}$ is either one or zero based on whether there is an existing edge between two nodes. A matrix $X \in \mathbb{R}^{m \times |V|}$ often indicates the attribute or features of a graph, where m is the number of features of each node and $|V|$ indicates the number of nodes.

GraphSAGE (Sample and aggregate) [72] is an inductive learning technique to extract node embedding. GraphSAGE can be used in unsupervised and supervised learning. GraphSAGE for the unsupervised learning can be implemented as follows [72]:

1) Given a graph $G(V, E)$ and feature matrix $X \in \mathbb{R}^{m \times |V|}$; for $\forall v \in V$, each x_v has a vector of feature with m dimensions.

2) Set up a neural network with K layers and let $h^0 = x_v$ indicate original node embedding for $\forall v \in V$, and this is the node features or attribute with m dimensions originally.

3) For $k = 1$ layer, and for each $v \in \mathcal{V}$:

   (a) Let $\mathcal{N}(v)$ indicate local neighborhood of nodes of v, all information from the neighborhood is aggregated by a function indicated by f, then the

generalized aggregation is indicated as follows:

$$h^1_{\mathcal{N}(v)} = f^1(h^0_u), \quad \forall u \in \mathcal{N}(v) \tag{2.5}$$

the aggregate function f can be one of the three: i) mean value function $f = \sum_{u \in \mathcal{N}(v)} \frac{h^0_u}{|\mathcal{N}(v)|}$ which takes average among all the h^0_u; ii) pool value function $f = \max(W_{\text{pool}} h^0_u)$, $\forall u \in \mathcal{N}(v)$ which applies an element-wise max-pooling operation on neighbour set nodes; iii) a long short-term memory (LSTM) function $f = \text{LSTM}(h^0_u)$, $\forall u \in \pi(v)$ which applies LSTM to random permutation of neighbor set of nodes.

(b) After summarizing information from local neighbor sets by aggregate function, the embedding for each node is in the form below:

$$h^1_v = \sigma(W^1 \cdot \text{concatenate}(h^0_v, \quad h^1_{\mathcal{N}(v)})) \tag{2.6}$$

where σ is non-linearity function and W^1 is the weight matrix.

4) After all nodes though part a) and b) from step 3), the node embedding after the first layer is represented below:

$$h^1_v = \frac{h^1_v}{\|h^1_v\|_2}, \quad \forall v \in \mathcal{N}(v) \tag{2.7}$$

5) For $2 \leq k \leq K$ layers step 3) and 4) are repeated. The embedding for each node is in the form below:

$$\begin{aligned} h^k_{\mathcal{N}(v)} &= f_k(h^{k-1}_u), \ \forall u \in \mathcal{N}(v) \\ h^k_v &= \sigma(W^k \cdot \text{concatenate}(h^{k-1}_v, \ h^k_{\mathcal{N}(v)})) \\ h^k_v &= \frac{h^k_v}{\|h^k_v\|_2}, \ \forall v \in \mathcal{N}(v) \end{aligned} \tag{2.8}$$

and the common forms of aggregate functions are :

$$\begin{aligned} f_k &= \sum_{u \in \mathcal{N}(v)} \frac{h_u^{k-1}}{|\mathcal{N}(v)|} \\ f_k &= \max(W_{\text{pool}} h_u^{k-1}),\ \forall u \in \mathcal{N}(v) \\ f_k &= \text{LSTM}(h_u^{k-1}),\ \forall u \in \pi(v) \end{aligned} \tag{2.9}$$

6) The embedding of node after K layer is as follows:

$$z_v = h_v^K,\ \forall v \in \mathcal{V}. \tag{2.10}$$

7) The process is iterated to minimize the loss function below:

$$\mathcal{J}(z_u) = -\log(\sigma(z_u^\top z_v)) - Q \cdot \mathbb{E}_{v_n \sim P_n(v)} \log(\sigma(-z_u^\top z_{v_n})) \tag{2.11}$$

where σ is non-linearity function, z_u is the embedding of neighbor node of u, Q is the negative samples (non-neighbor node) of u and v_n is from negative sample distribution $P_n(v)$.

2.4 Case study

In this section, two real data sets are used to extract functional connectivity network, the default mode network, based on the methods mentioned above. A comparison will be made among the methods.

2.4.1 Case one

The first data set is from an open shared neuroimage data resource (`http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html`), namely, the 1000 Functional Connectomes Project.

(i) ***Data description***

The data has 84 subjects including 43 males and 41 females. Their ages range from 7 to 49. The repetition time (TR) is 2 seconds, the number of slice is 39 and the time points are 192. The structural brain of one random chosen subject with dimension $256 \times 176 \times 256$ is shown in Figure 2.2, and the functional brain of the same subject with dimension $64 \times 64 \times 39 \times 192$ is shown in Figure 2.3.

(ii) ***Data preprocessing***

Data preprocessing is an important step before doing any statistical analysis. As mentioned in Section 1.3 there are several softwares or packages which deal with data preprocessing. For this data set the DPABI [186], a toolbox of Matlab, is used. The first ten time points are removed. The pipeline for preprocessing is as follows; first, time correction is done; the following is realign, co-registration between T1 weighted and functional images; head motion model with Friston 24 parameters motion covariates is used to reduce head motion effect [63]; nuisance regression with linear trend, average cerebrospinal fluid (CSF) and white matters (WM) as nuisance regressors are to reduce espiratory and cardiac effects; finally normalization to MNI template is performed. The temporal filtering is to set in range of 0.01∼0.1Hz and a Gaussian kernel of full width with half maximum (FWHM) equal to 6mm is used for spatial smoothing.

(iii) ***Data analysis***

First, the SBC is performed in the toolbox CONN [180] in Matlab. The voxel threshold p-uncorrected values is set less than 0.0001 and Familywise Error Rate (FWER) is set less than 0.01 which attempts to control the probability of false positives. $|T|$ value is larger than 4.09 with degree of freedom 83 for two-sided

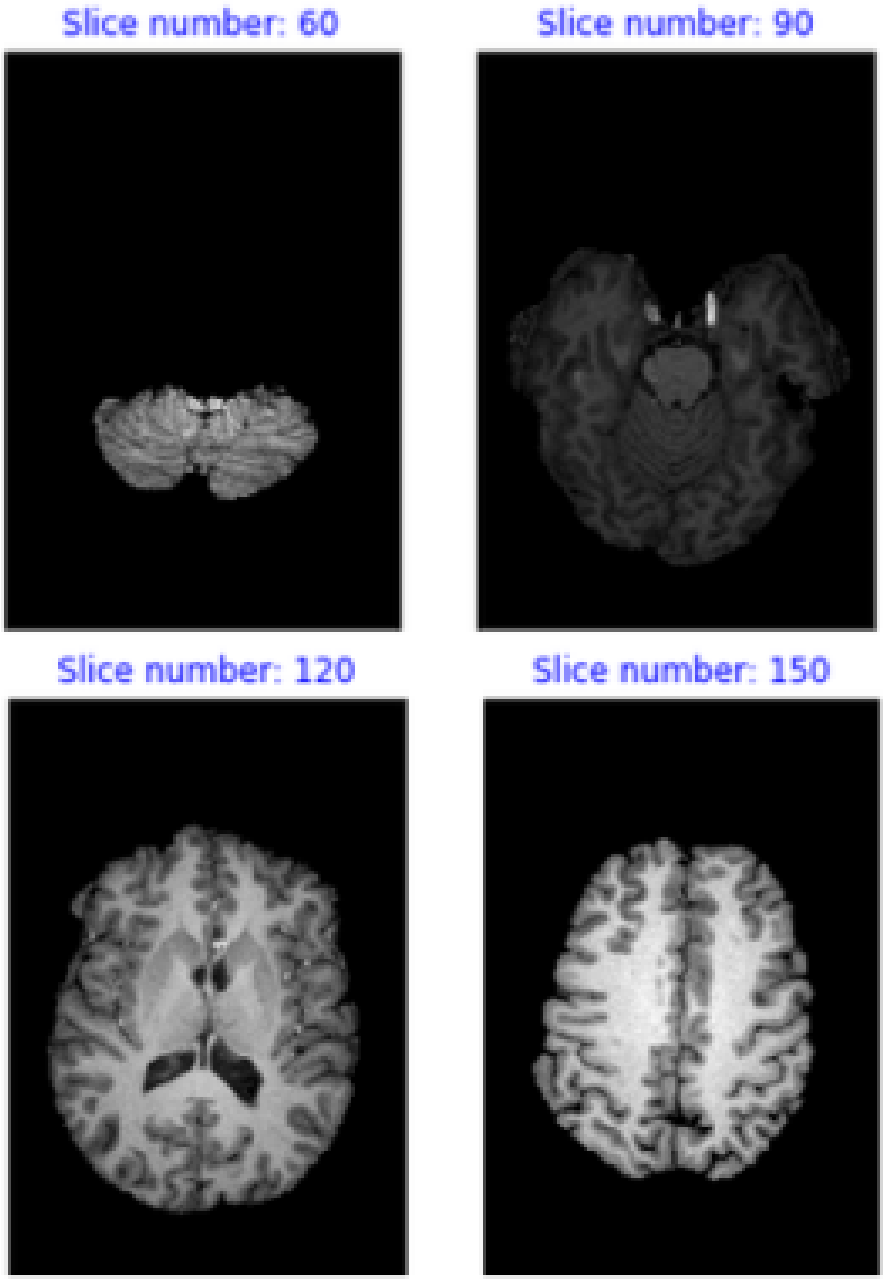


Figure 2.2. The structural brain of a randomly chosen subject

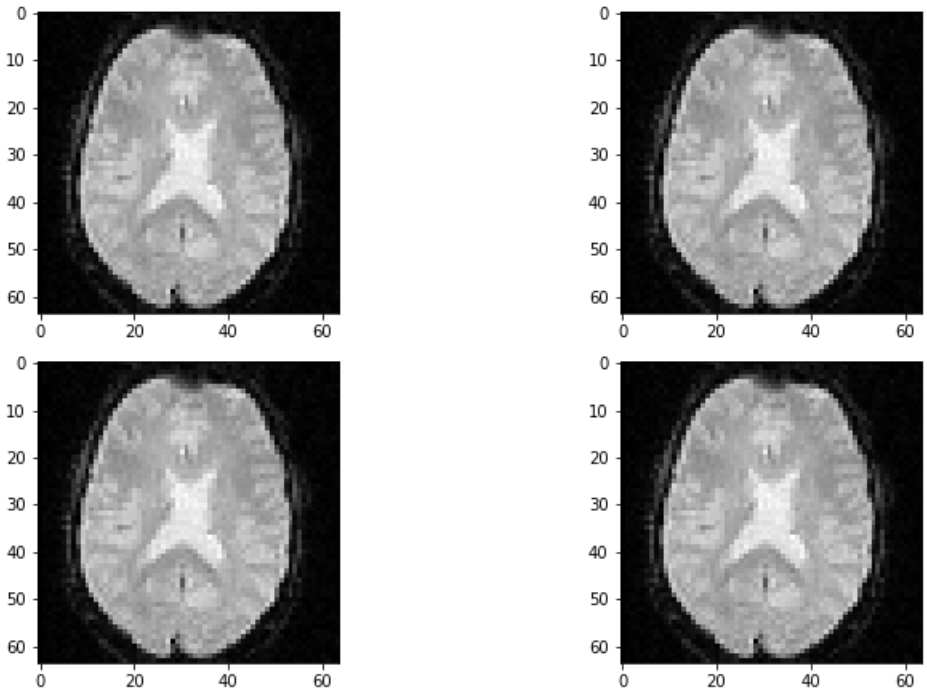


Figure 2.3. The functional brain of the randomly chosen subject in Figure 2.2

test. See Figures 2.4, 2.5, 2.6 for details based on the seed of posterior cingulate cortex (PCC) from the default mode network.

Secondly, the group fastICA is also performed in CONN [180] and the independent components are set to 10. Other parameters are the same as SBC. See the Figures 2.7, 2.8, 2.9 for details.

Thirdly, the dictionary learning is performed in Python with Nilearn [1] package. As mentioned in Section 2.2.3, dictionary learning focuses on sparsity between components. The component is set to 10 and the sparsity parameter is set to 15. See Figures 2.10, 2.11, 2.12 for details.

Lastly, the neural network method is performed through the package of StellarGraph [50] in Python. The 800 ROIs of brain are used, so there are 800 nodes in the

graph and each two makes an edge so that there are totally 319,600 edges in the graph. For each subject the correlation between any two ROIs is calculated based on 182 time series data. Each subject has 182 time points, meaning that each ROI of each subject has 182 values. Each subject has a correlation matrix with dimension of 800×800. There are 84 correlation matrices for 84 subjects. To reduce complexity and noise, all the correlation values less than 0.1 are replaced by zero and thus the associated edges are dropped. The correlation for each node is the node feature which is used in the subsequent analysis. A two-layer graph network is set for analysis. The number of walks for each node v is set to 10; the length of walk is 5; batch size is set to 10; 1-hop neighbors node for each node v is set to 50 and 2-hop neighbors node is set to 10 for each node v; the first layer is set to 200 hidden neurons; the second layer is set to 100 neurons; the learning rate is set to 0.01; epochs is set to 100 and dropout rate is set to 0.15. After analysis each node embedding is represented by a vector with 100 dimensions. To visualize the node embedding a principal component analysis (PCA) is used with 95% variation held and the first two components are used for plot. Figure 2.13 shows seven groups. Then K-means [107] cluster algorithm is performed with components extracted from PCA to group ROIs into seven parts. See Figures 2.14, 2.15, 2.16 for details. The comparison of DMN netwrok for all the algorithms are shown in Figure 2.17.

2.4.2 Case two

The second data set is also from the neuroimage data resource (`http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html`), namely, 1000 Functional Connectomes Project.

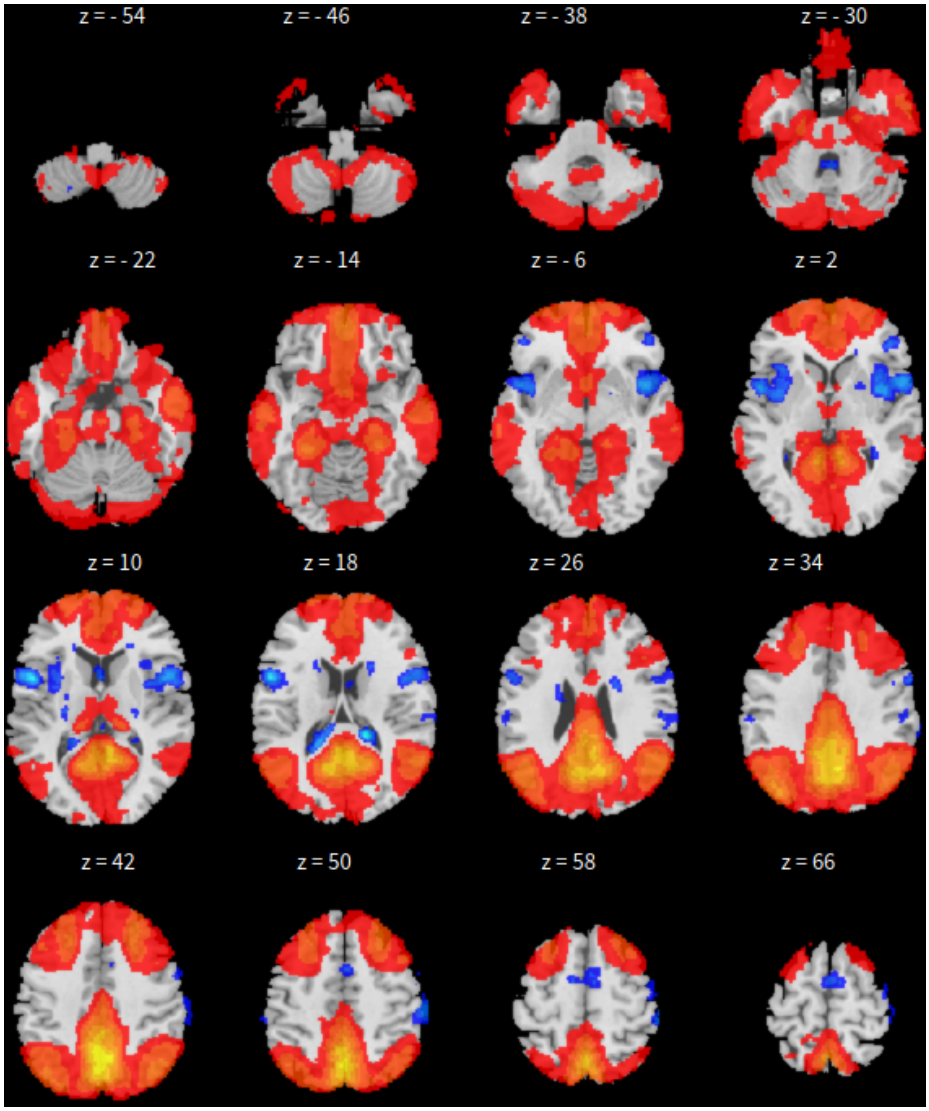


Figure 2.4. Axial display (xy-plane) view for SBC based on PCC seed

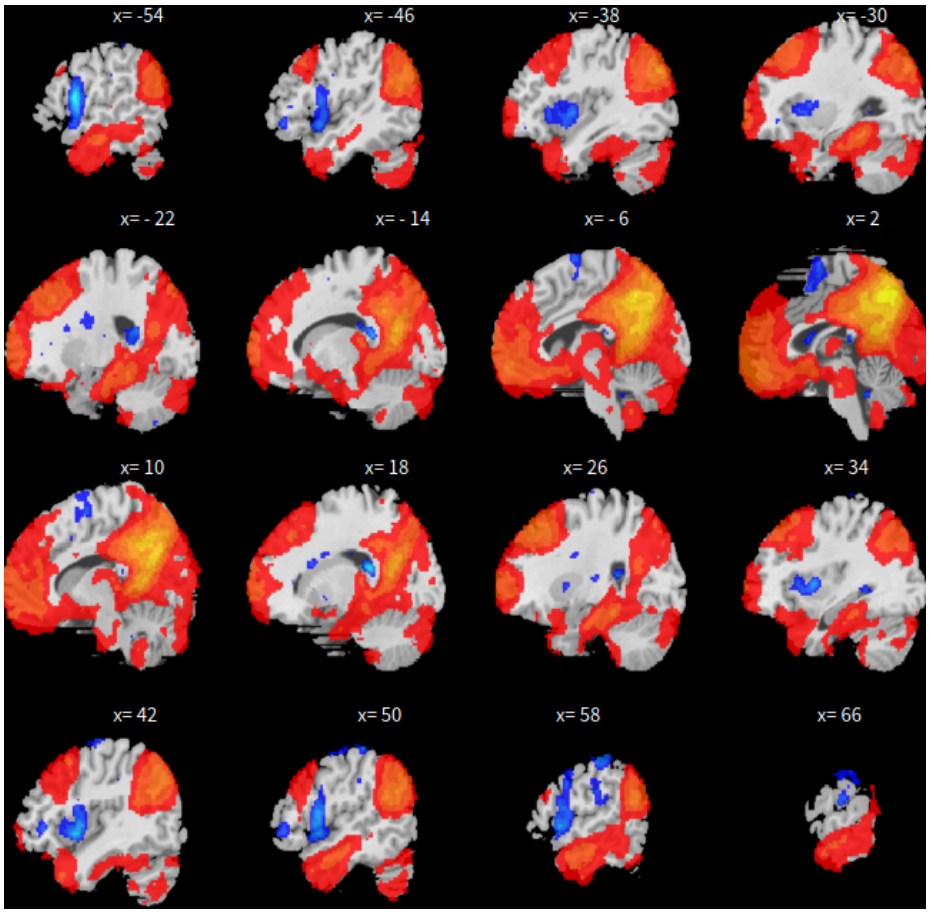


Figure 2.5. Sagittal display(yz-plane) view for SBC based on PCC seed

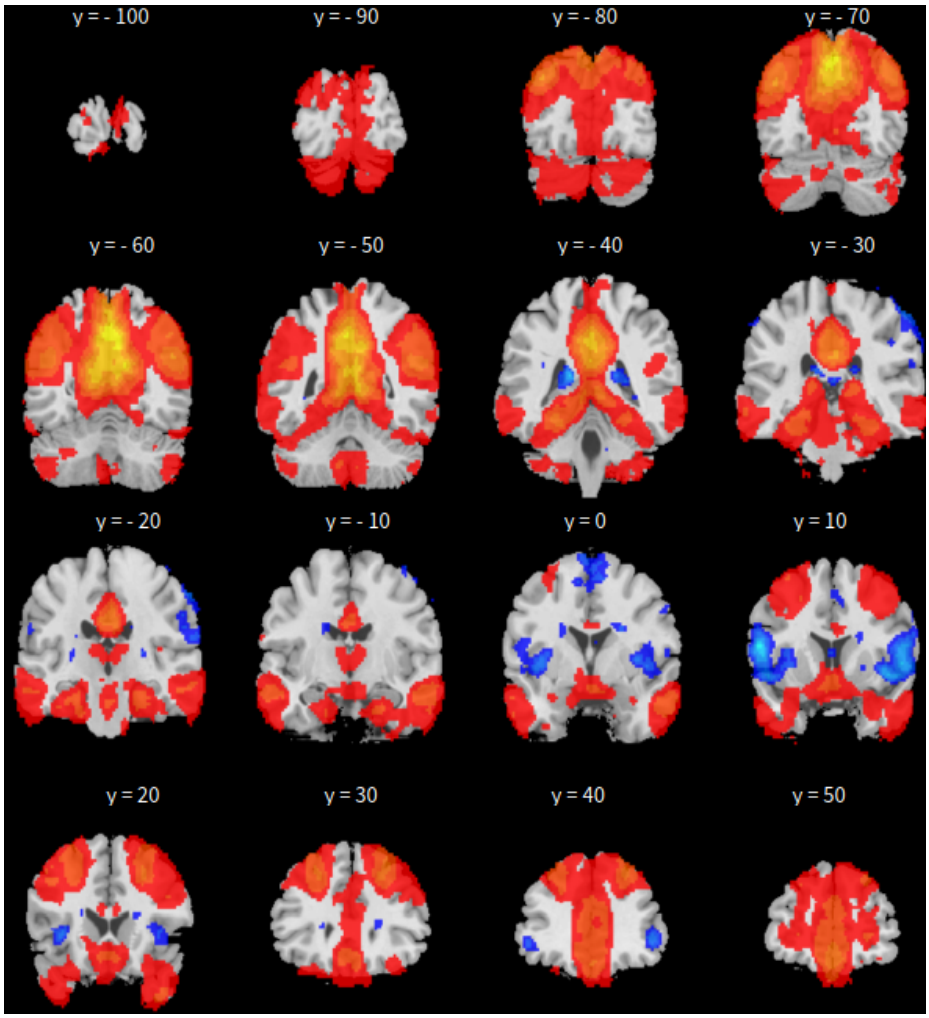


Figure 2.6. Coronal display (xz-plane) view for SBC based on PCC seed

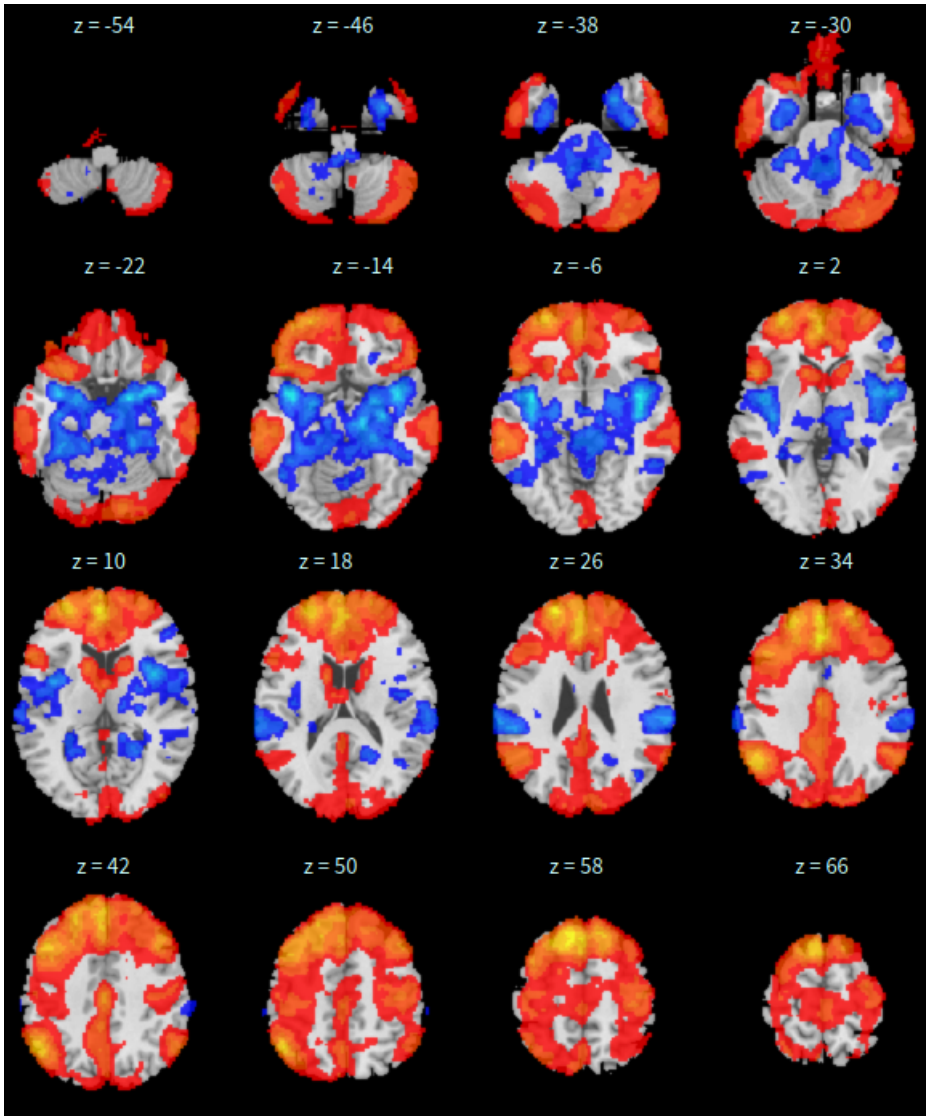


Figure 2.7. Axial display (xy-plane) view for ICA based on PCC

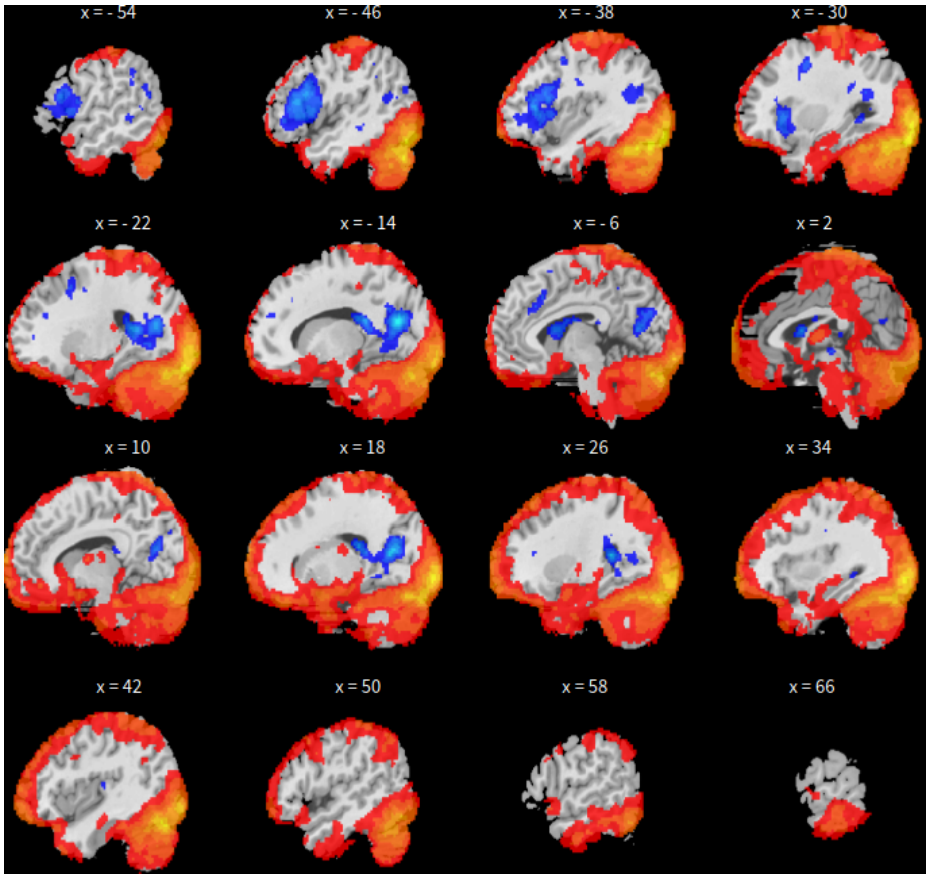


Figure 2.8. Sagitta display (yz-plane)view for ICA based on PCC

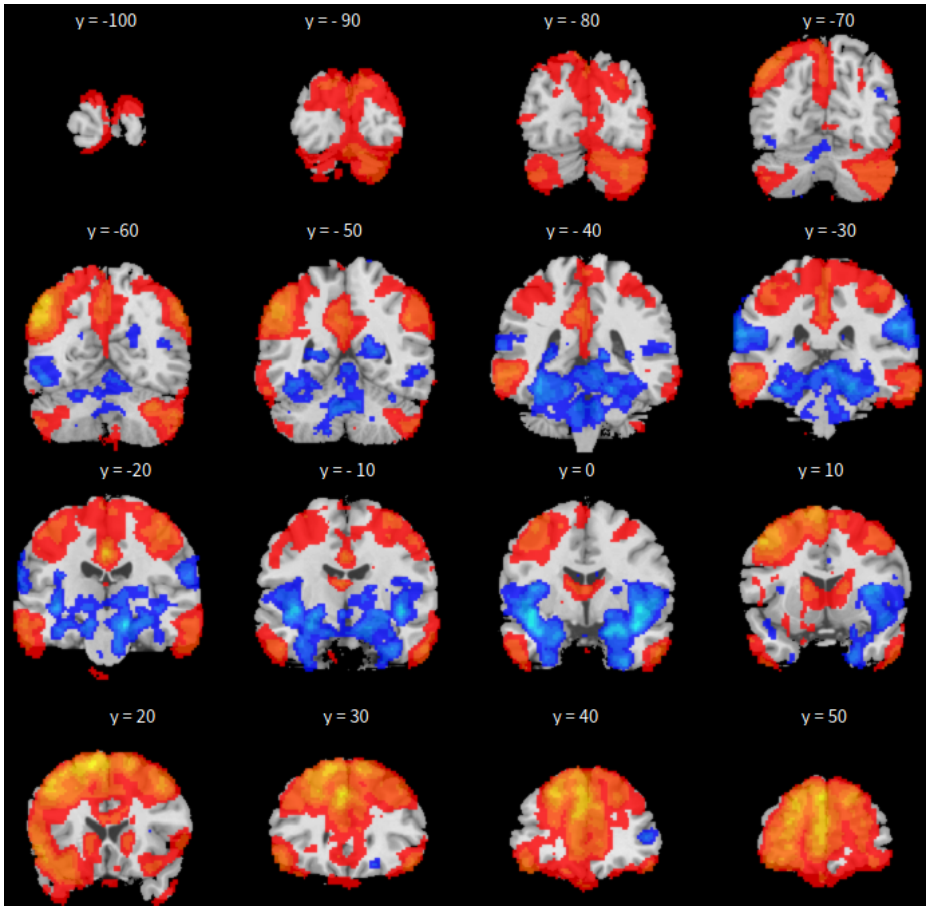


Figure 2.9. Coronal display (xz-plane) view for ICA based on PCC seed

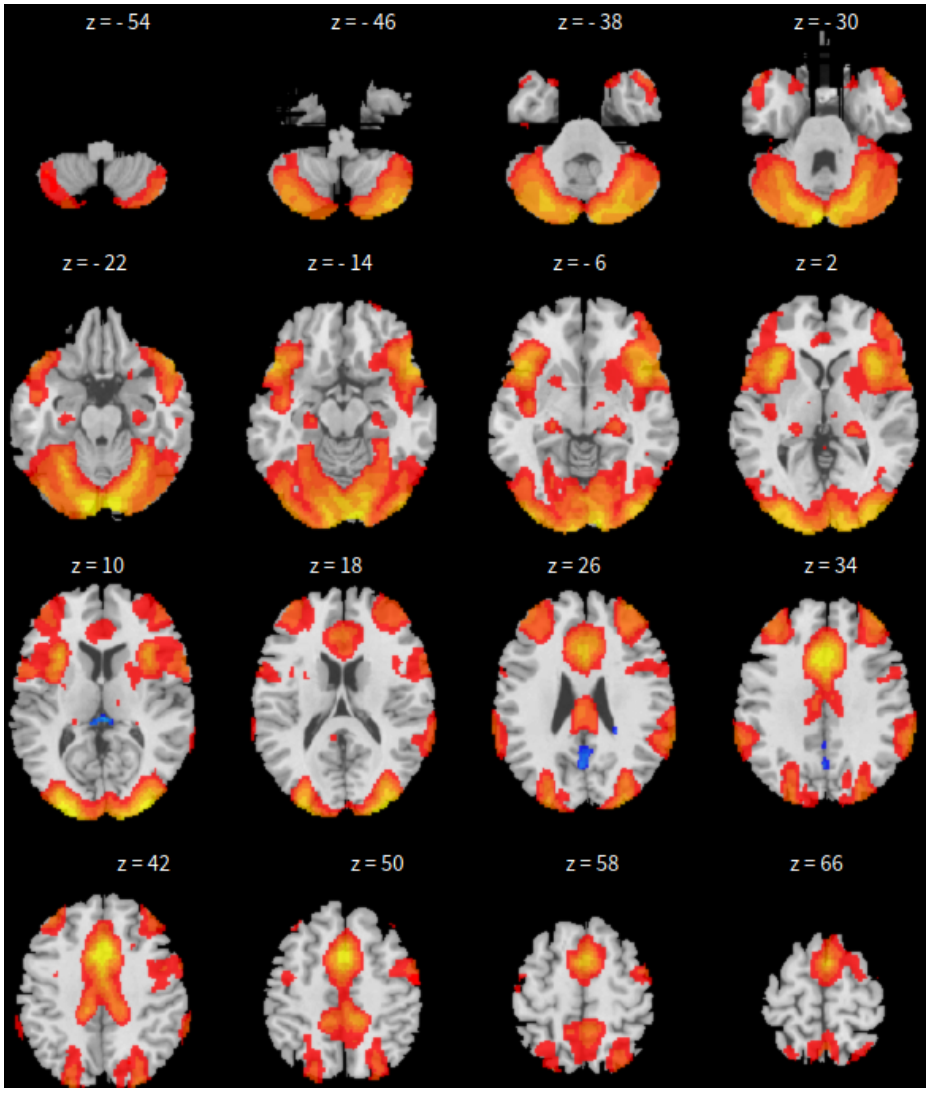


Figure 2.10. Axial display (xy-plane) view for dictionary learning based on PCC seed

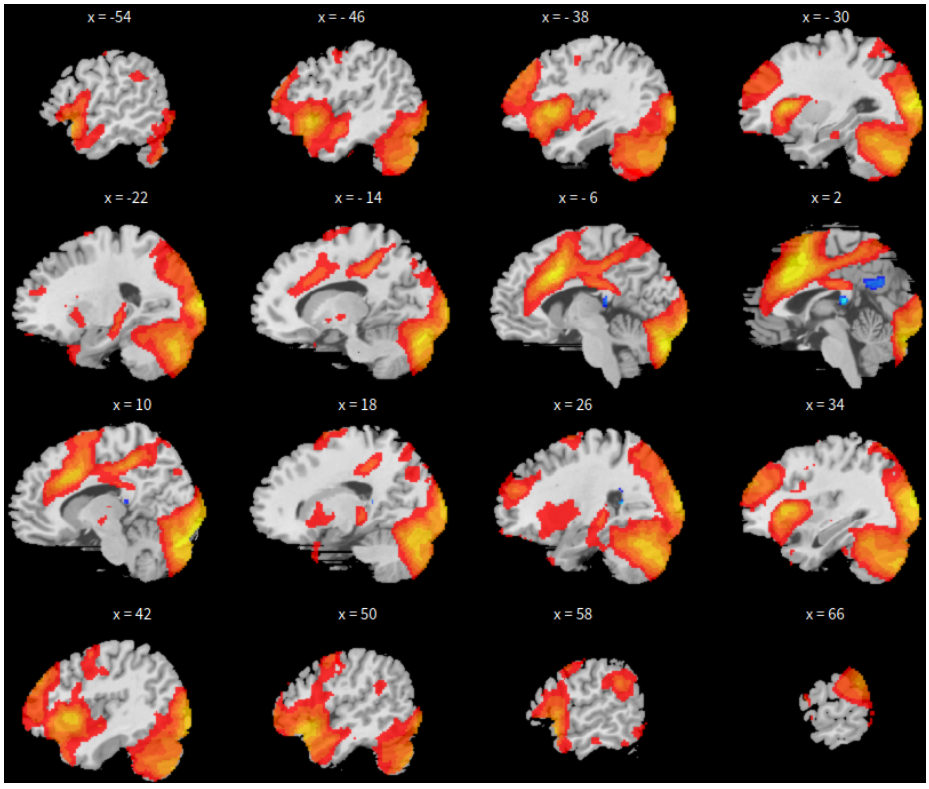


Figure 2.11. Sagittal display (yz-plane)view for dictionary learning based on PCC seed

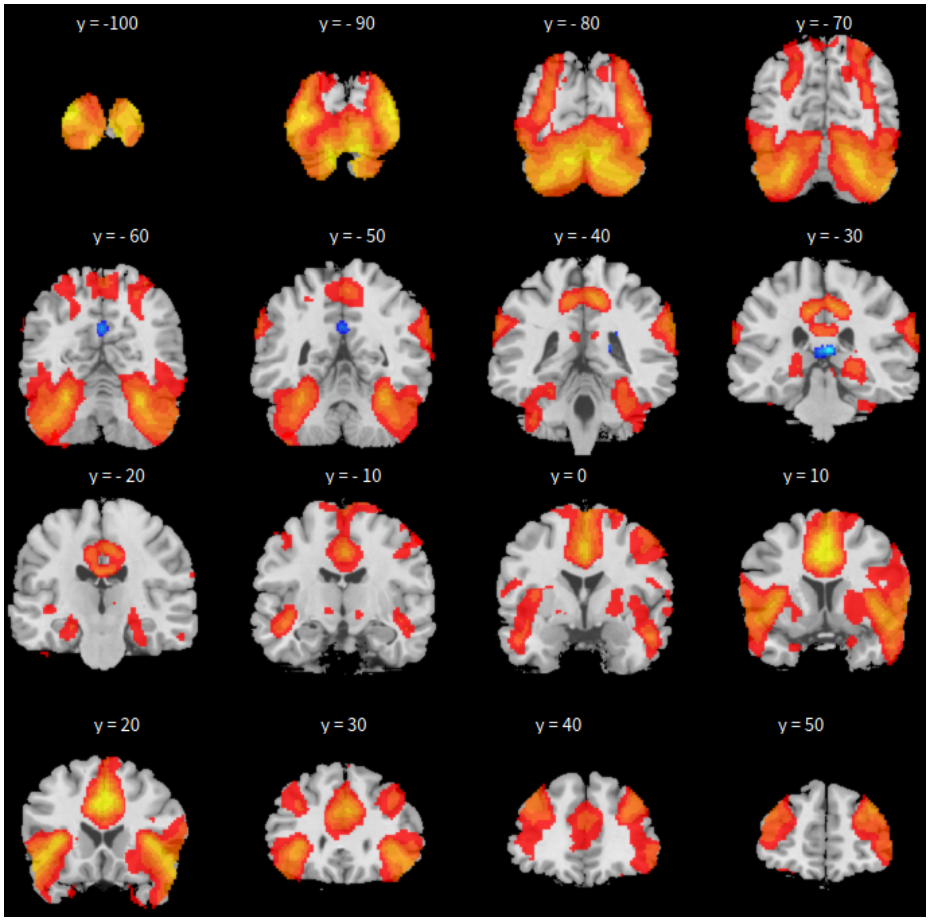


Figure 2.12. Coronal display (xz-plane) view for dictionary learning based on PCC seed

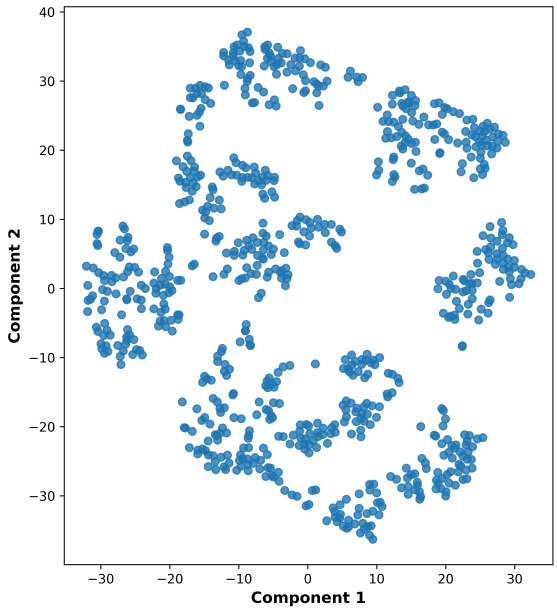


Figure 2.13. Plot of node embedding with two components

(i) ***Data description***

The second data has 46 subjects including 15 males and 31 females. Their ages range from 44 to 65. The repetition time (TR) is 2 seconds; the number of slice is 64 and the time points are 175. The structural brain of one randomly chosen subject with dimension $256 \times 256 \times 144$ is shown in Figure 2.18, and the functional brain image of the same subject with dimension $36 \times 64 \times 64 \times 175$ is shown in Figure 2.19.

(ii) ***Data preprocessing***

For this data set, the package of fMRIPrep [56] is used for data preprocessing. The first ten time points are removed. The pipeline for preprocessing is the same as case one.

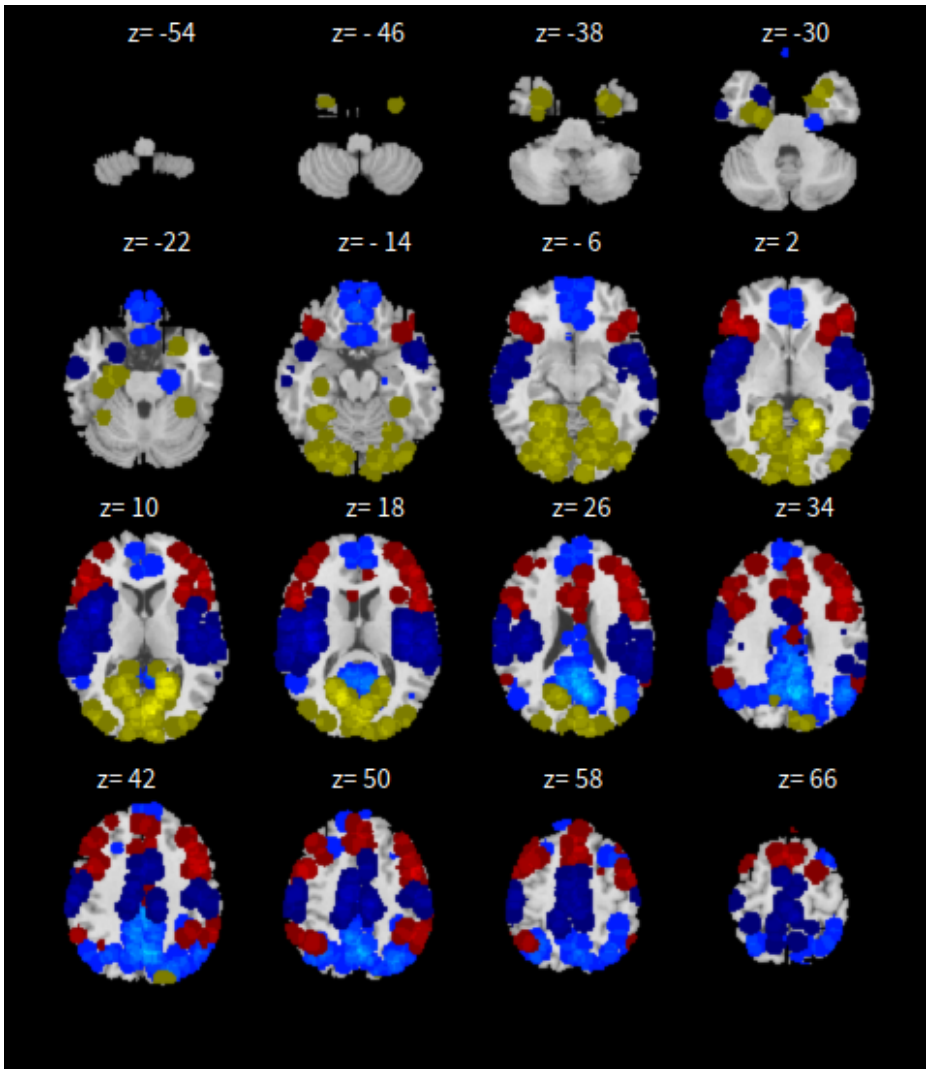


Figure 2.14. Axial display (xy-plane) display view for graphSAGE

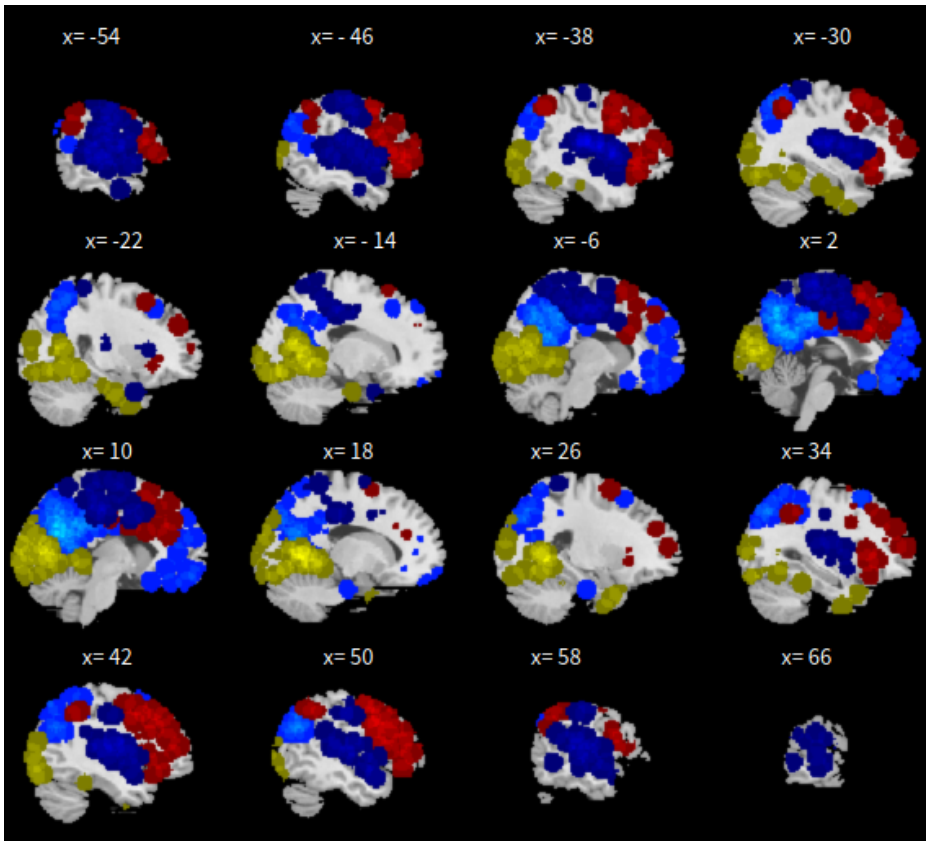


Figure 2.15. Sagittal display (yz-plane)view for graphSAGE

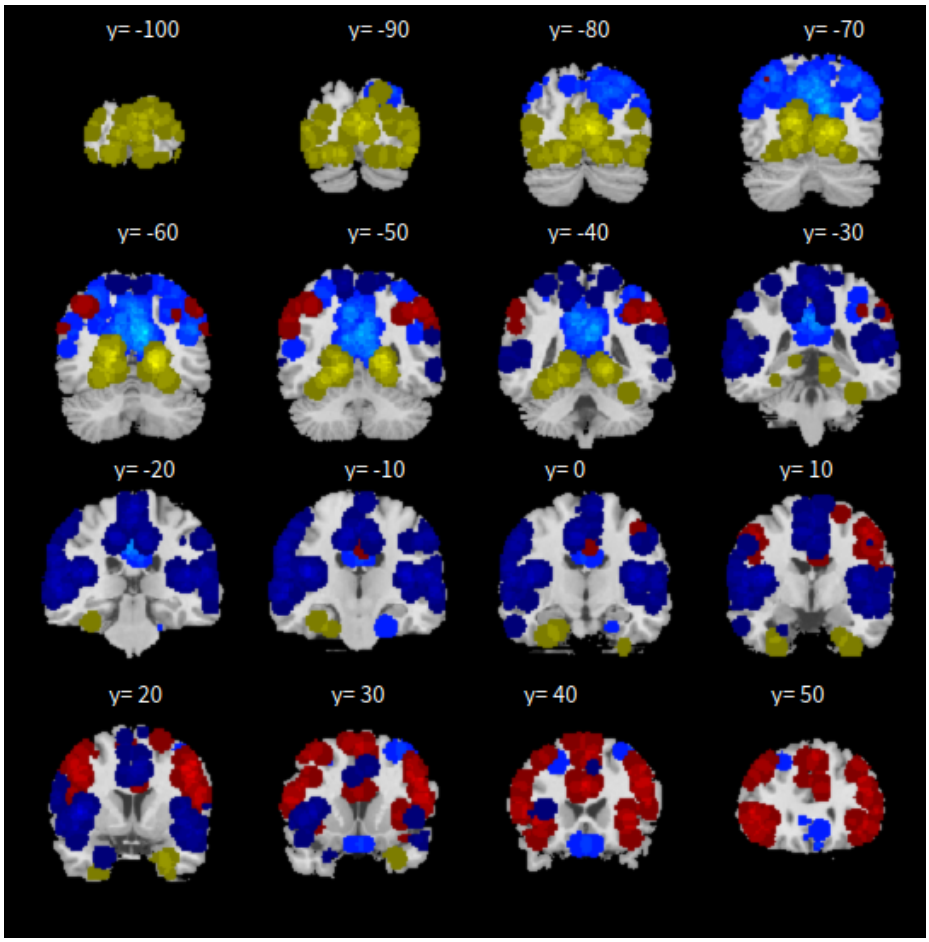
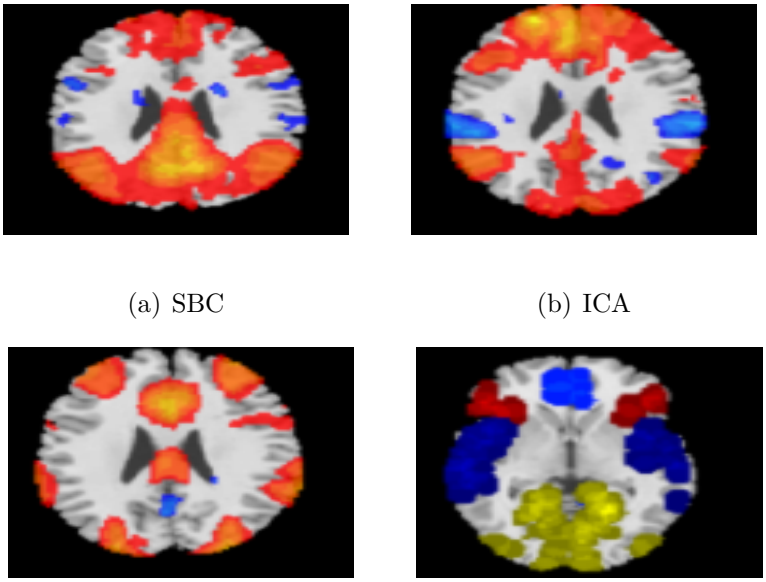


Figure 2.16. Coronal display (xz-plane) view for graphSAGE



(a) SBC (b) ICA

(c) DL (d) graphSAGE

Figure 2.17. Result comparison: DMN

(iii) ***Data analysis***

First, the SBC is performed in the toolbox CONN [180] in Matlab. The voxel threshold p-uncorrected values is set less than 0.0001 and Familywise Error Rate (FWER) sets at less than 0.01. $|T|$ value is larger than 4.29 with degree of freedom 45 for two-sided test. See Figures 2.21, 2.22, 2.23 for details based on the seed of posterior cingulate cortex (PCC) from the default mode network.

Secondly, the group fastICA is also performed in CONN [180] and the number of independent components are set to 10. Other parameters are the same as SBC. See Figures 2.24, 2.25, 2.26 for details.

Thirdly, the dictionary learning is performed in Python with Nilearn [1] package. As mentioned in Section 2.2.3, dictionary learning focuses on sparsity between

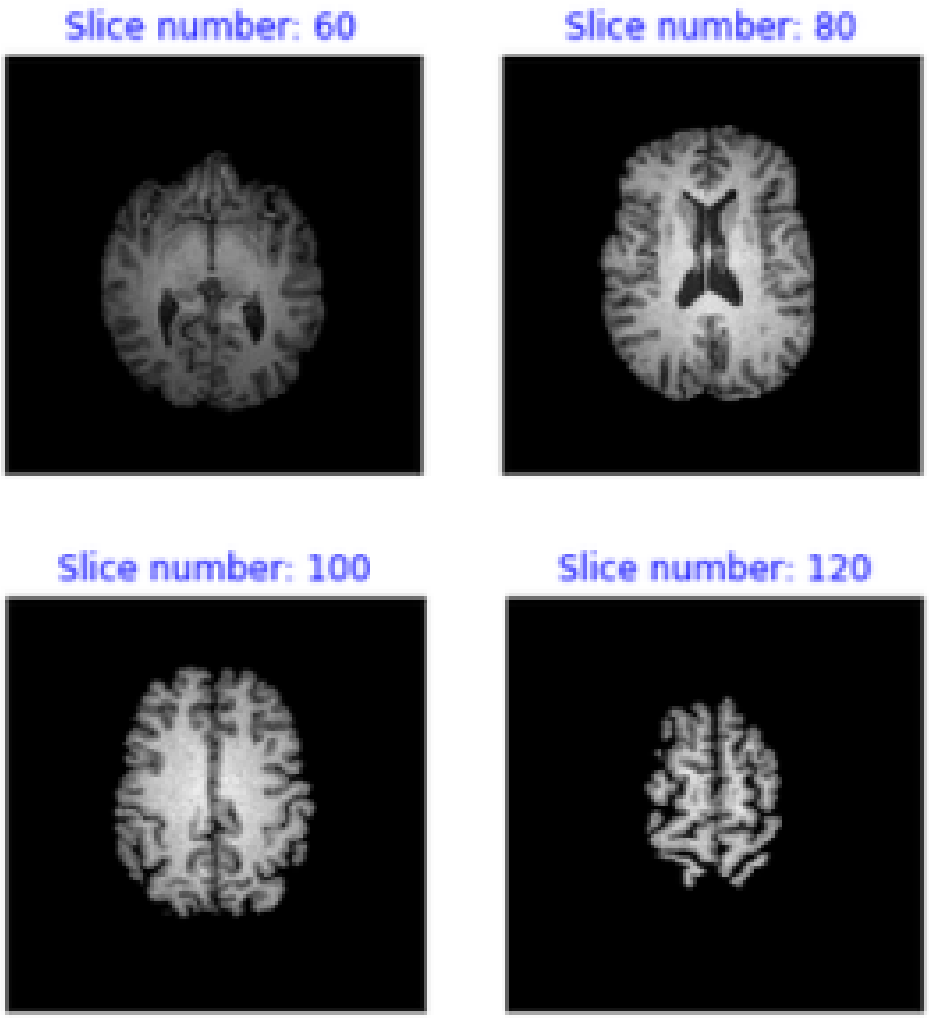


Figure 2.18. The structural brain of a randomly chosen subject

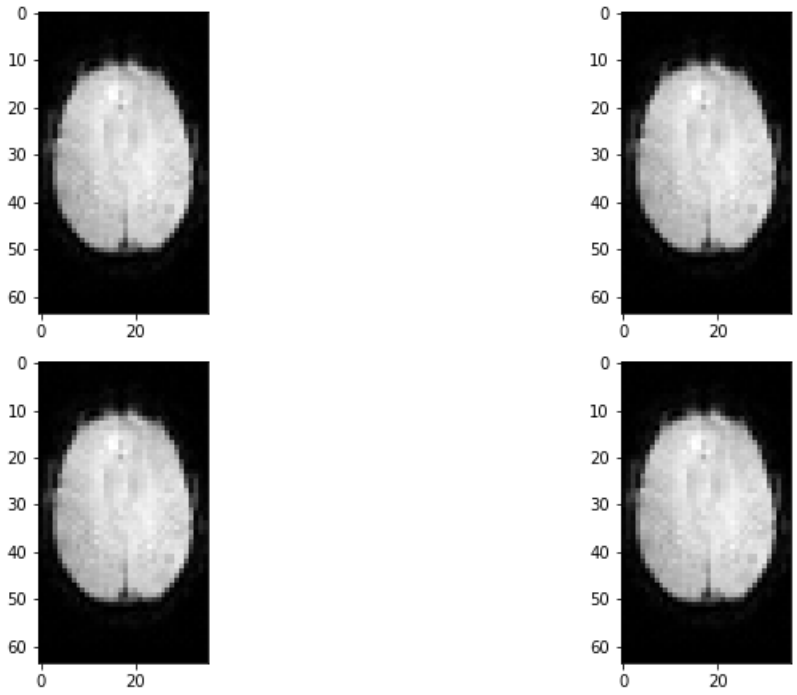


Figure 2.19. The functional brain of the randomly chosen subject in Figure 2.18

components. The component is set to 10, and the sparsity parameter is set to 15. See Figures 2.27, 2.28, 2.29 for details.

Lastly, the neural network method is performed through the package of StellarGraph [50] in Python. The 800 ROIs of brain are used, so there are 800 nodes in the graph and each two nodes make an edge so that there are 319,600 edges in the graph. For each subject the correlation between each pair of two ROIs is calculated based on 165 time series data. Each subject has a 165 time points, meaning that each ROI of each subject has 165 values. Each subject has a correlation matrix with dimension of 800×800. There are 46 correlation matrices for 46 subjects. To reduce complexity and noise, all the correlation values less than 0.10 are dropped and replaced by zero and thus the associated edges are dropped. The correlation for each node is the node feature which is

used in the subsequent analysis. A two-layer graph neural network is set for analysis. The number of walks for each node v are set to 10; the length of walk is 5; the batch size is set to 10; 1-hop neighbor node for each node v is set to 50 and 2-hop neighbors node is set to 10 for each node v; the first layer is set to 200 hidden neurons; the second layer is set to 200 neurons; learning rate is set to 0.01; epochs is set to 100 and dropout rate is set to 0.05. After analysis each node embedding is represented by a vector with 100 dimensions. To visualize the node embedding a principal component analysis (PCA) is used with 95% variation held and the first two components are used for the plot. Figure 2.20 shows seven groups. Then K-means [107] cluster algorithm is performed with components extracted from PCA to group ROIs into seven parts. See Figures 2.30, 2.31, 2.32 for details. The comparison of DMN network from different algorithms are shown in Figure 2.33.

2.5 Discussion and conclusion

We proposed to use graph neural network technique called graphSAGE to extract functional connectivity networks based on resting state functional magnetic resonance imaging data. The application of this technique in rs-fMRI can extract the default mode network (DMN). Seed-based method is a very classical way to detect RSNs. For all the two cases the DMN is found, which shows strong correlation with the prior selected seed of PCC. This technique can directly gives answer to the question but it is affected by the chosen seed and thus ignores other possible networks. Independent component method and dictionary learning method can also find default mode network in both cases, the two methods can give many networks based on prior assumption,

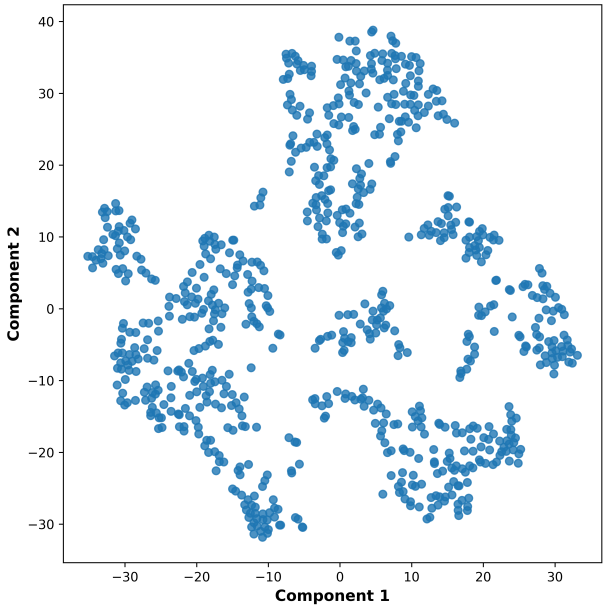


Figure 2.20. Plot of node embedding with two components

but some extracted networks are difficult to interpret and also the result relies on the number of components.

Compared with the seed-based method, independent component method and dictionary learning method, the application of grahSAGE gives a more robust result. The three existing methods need to make prior thresholds p values for single subject analysis and group level analysis and assume the number of components for ICA and dictionary learning. The result of analysis depends on these prior assumptions and is subjective more or less. On contrast, graphSAGE does not need these assumptions and considers the single subject analysis and group subjects analysis simultaneously. It gives more reliable results without subjective facts from researchers, but just like other deep learning methods the graphSAGE also need more data to train the model

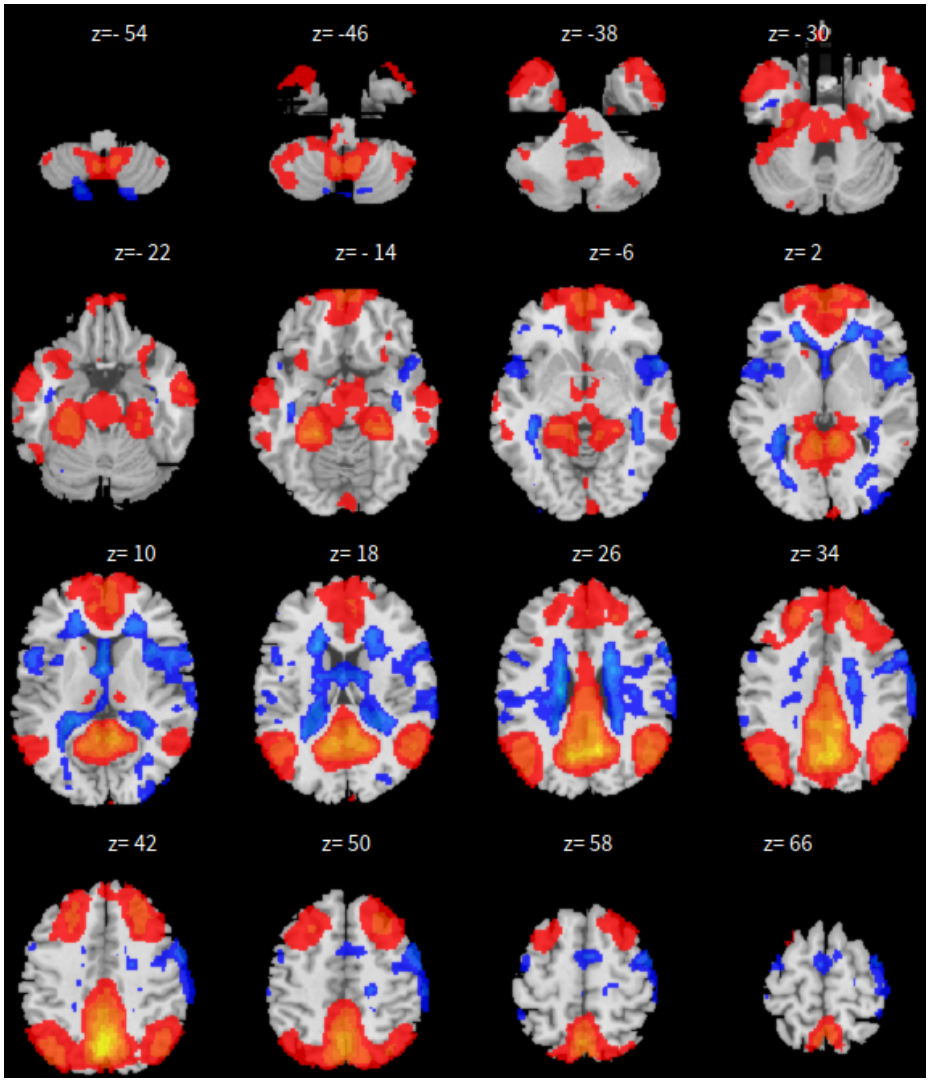


Figure 2.21. Axial display (xy-plane) view for SBC based on PCC seed

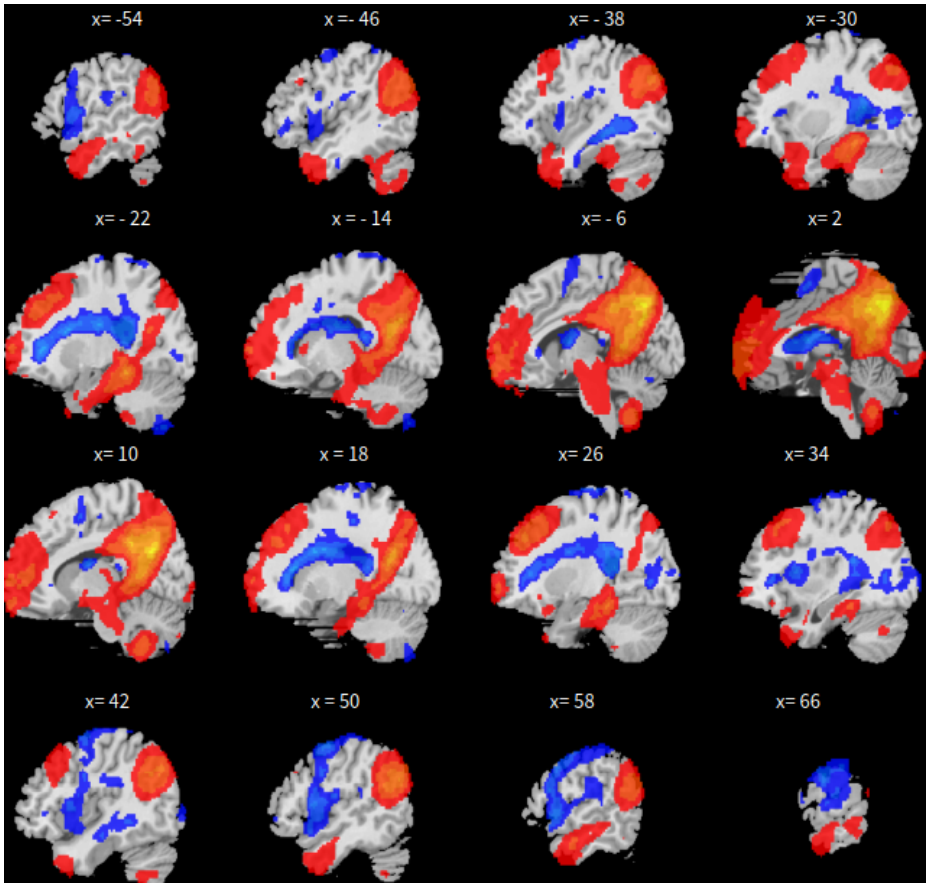


Figure 2.22. Sagittal display (yz-plane) view for SBC based on PCC seed

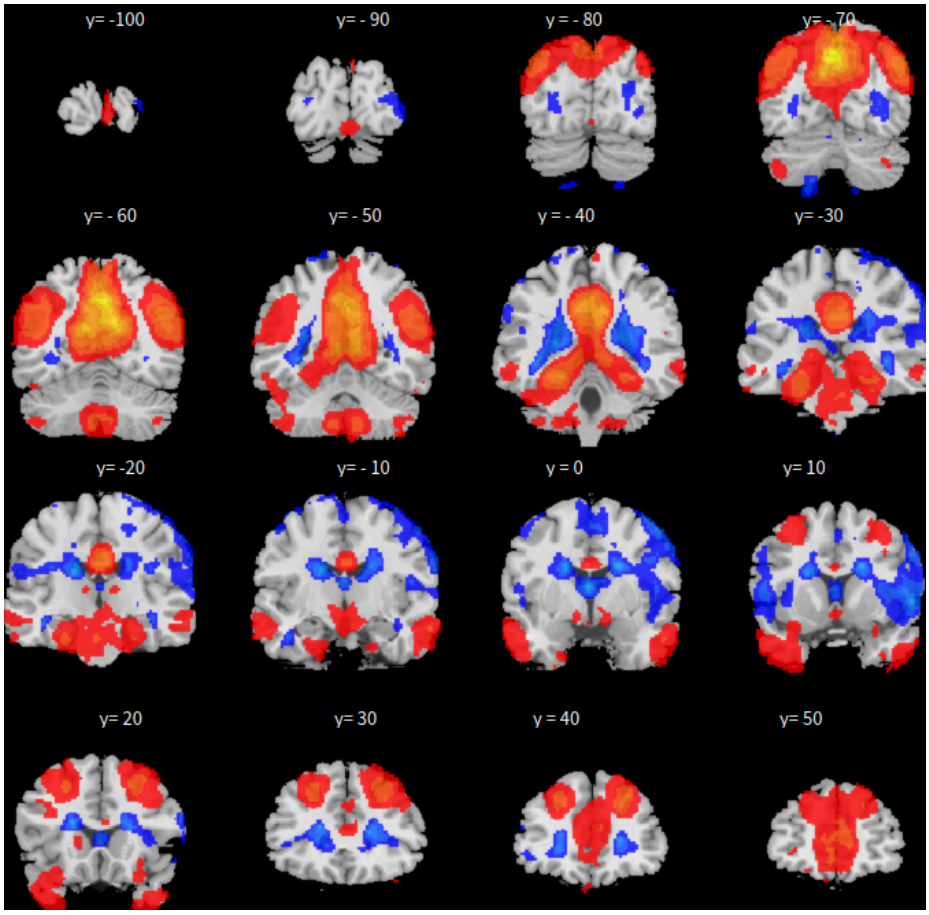


Figure 2.23. Coronal display (xz-plane) view for SBC based on PCC seed

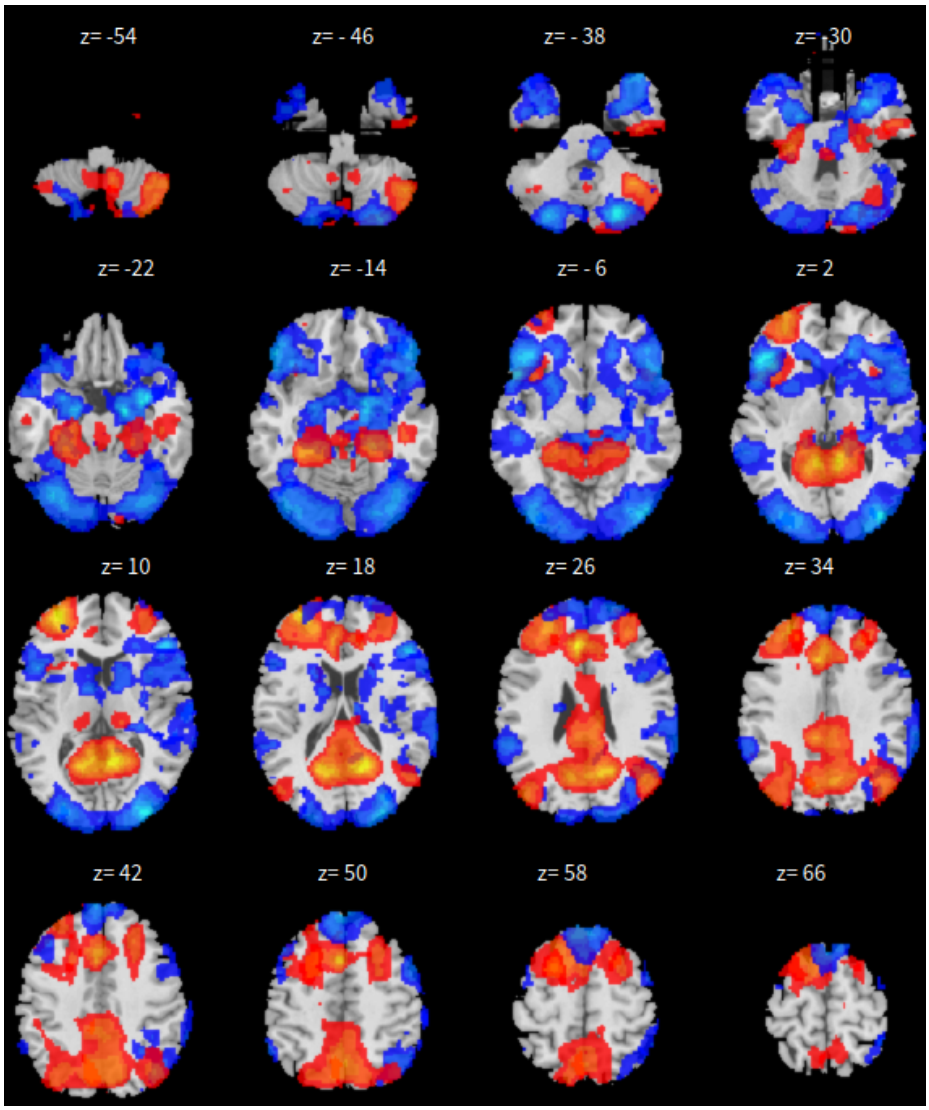


Figure 2.24. Axial display (xy-plane) view for ICA based on PCC seed

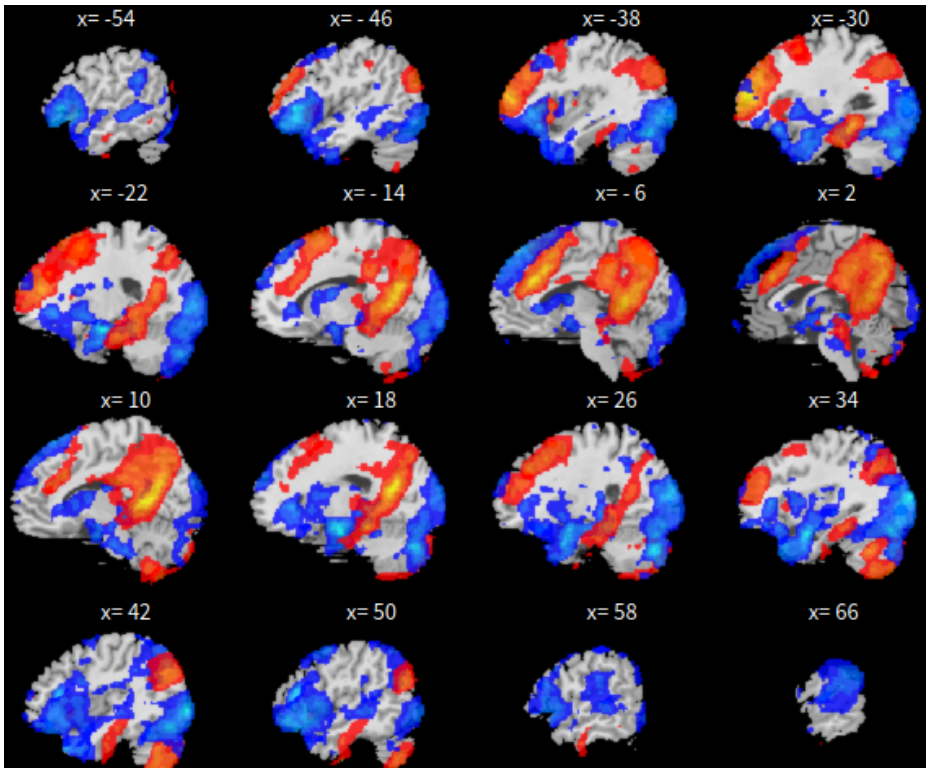


Figure 2.25. Sagittal display (yz-plane) view for ICA based on PCC seed

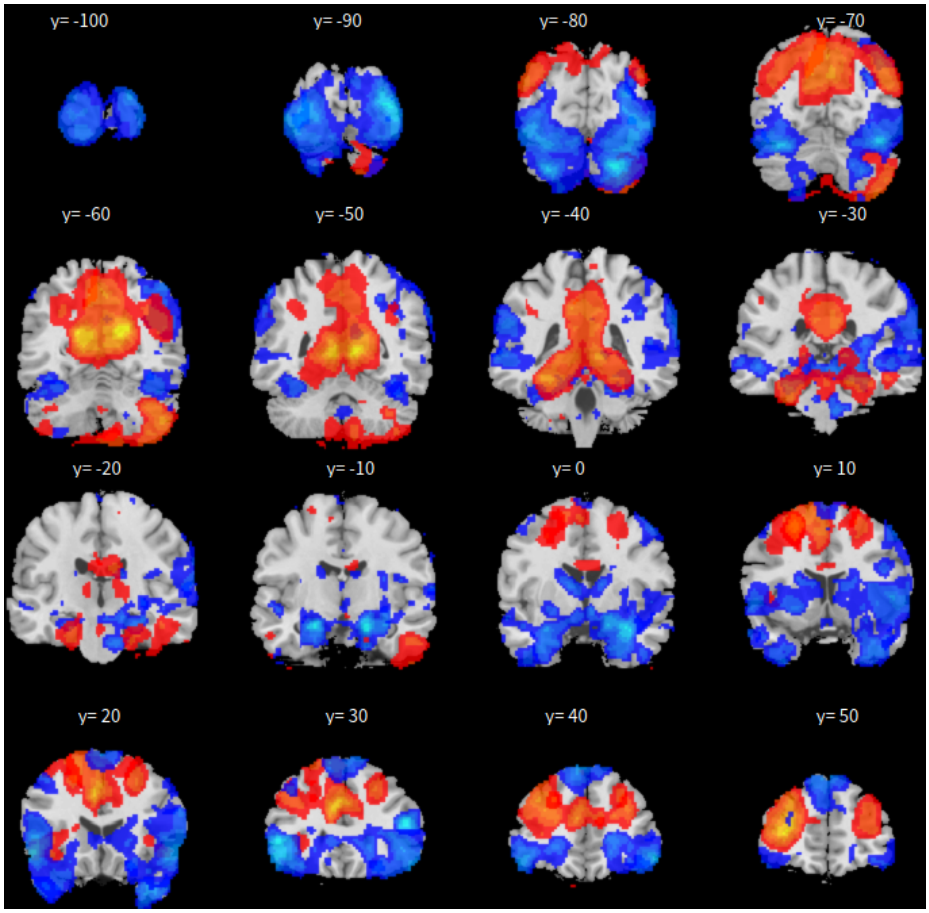


Figure 2.26. Coronal display (xz-plane) view for ICA based on PCC seed

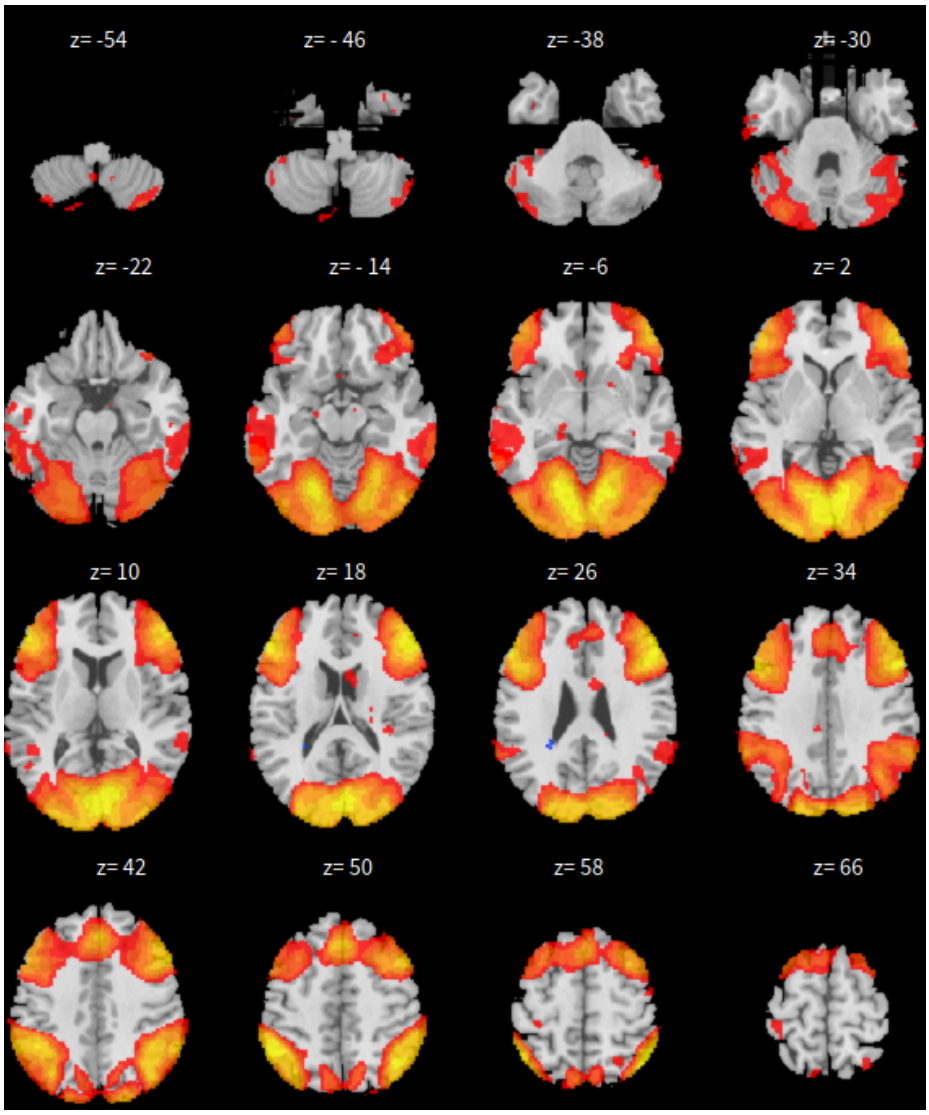


Figure 2.27. Axial display (xy-plane) view for dictionary learning based on PCC seed

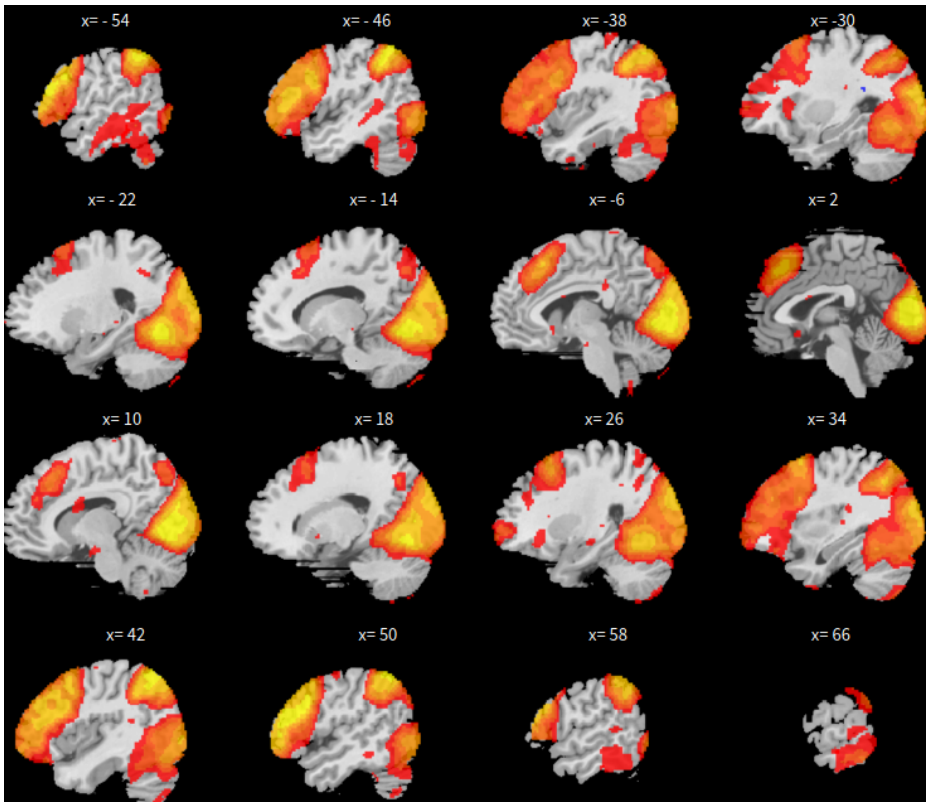


Figure 2.28. Sagittal display (yz-plane) view for dictionary learning based on PCC seed

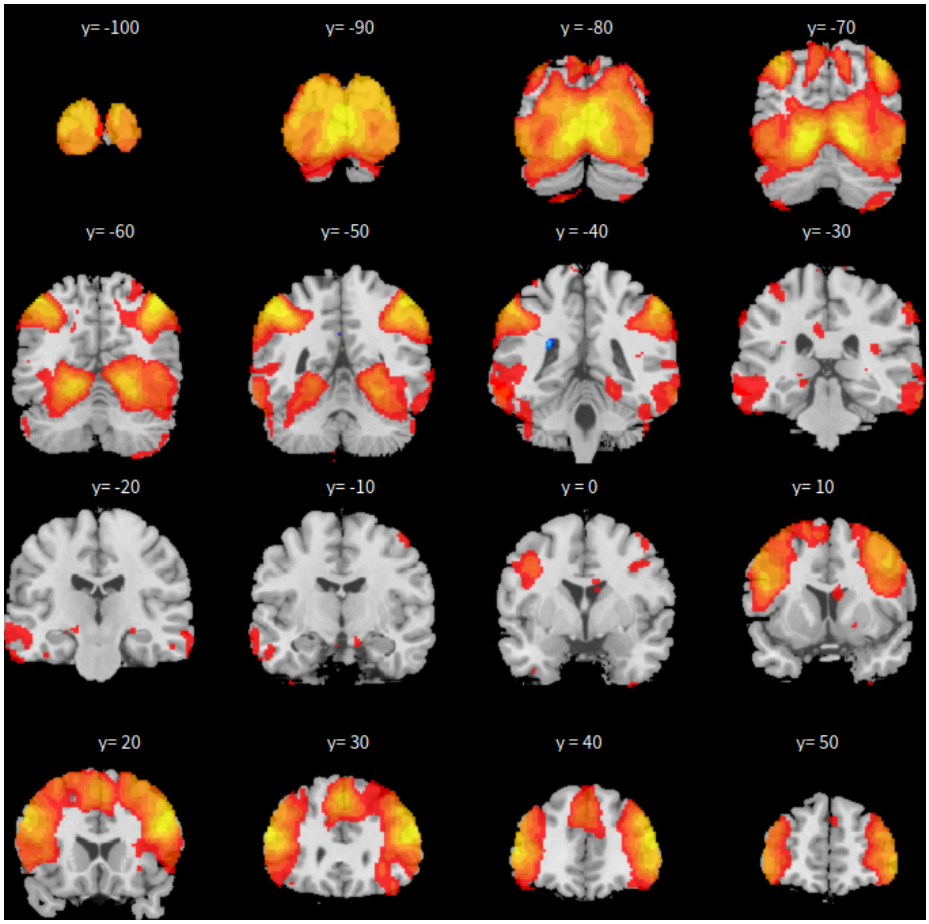


Figure 2.29. Coronal display (xz-plane) view for dictionary learning based on PCC seed

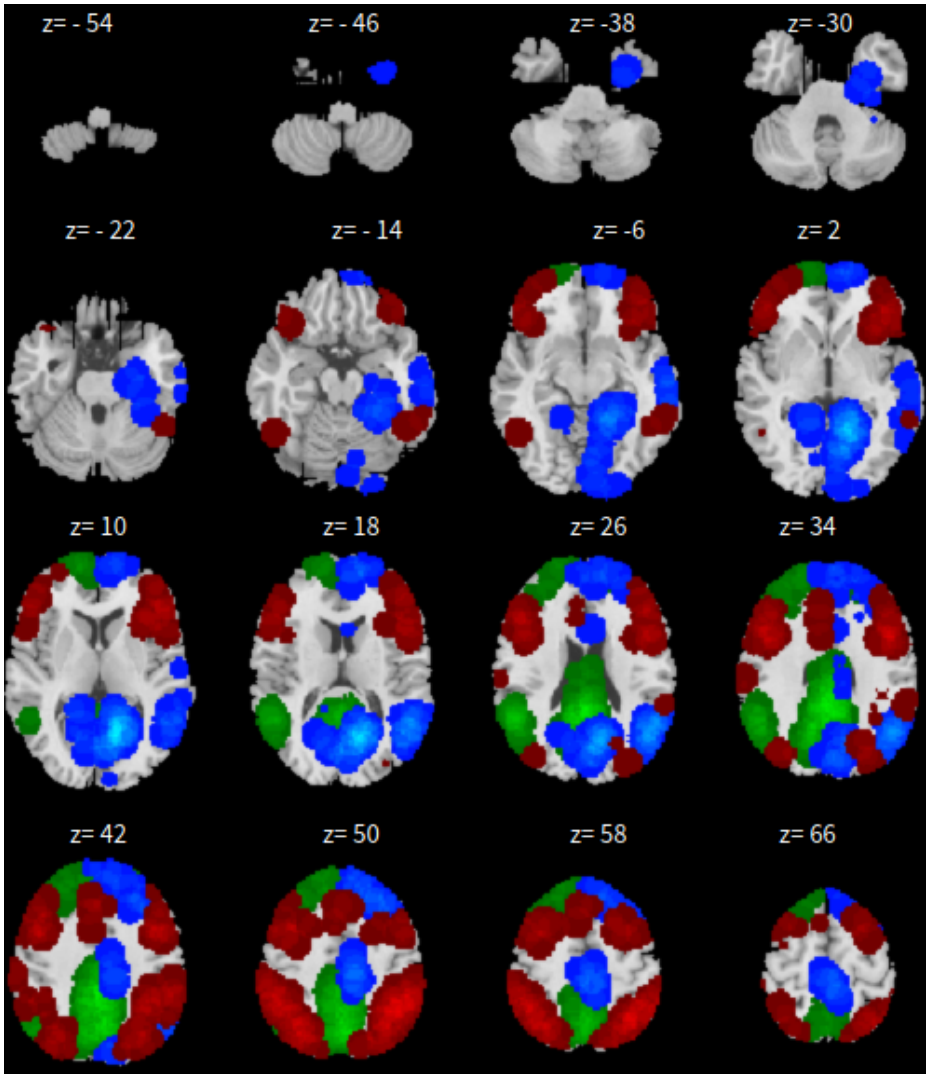


Figure 2.30. Axial display (xy-plane) view for graphSAGE

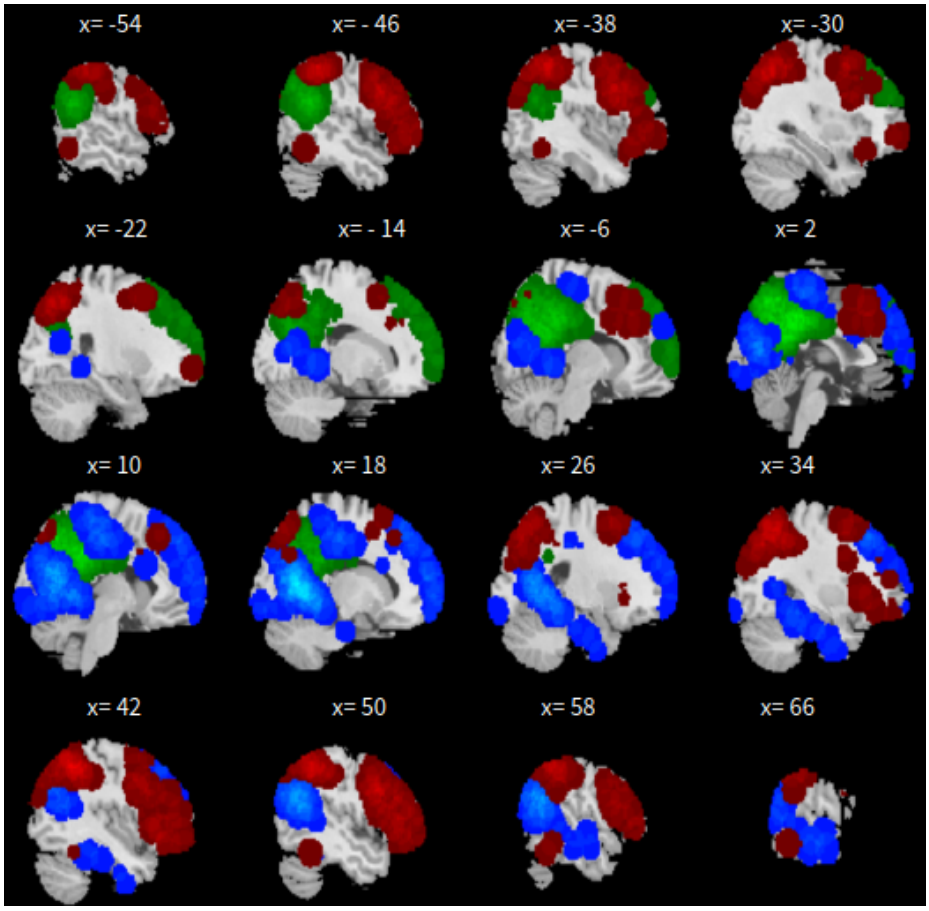


Figure 2.31. Sagittal display (yz-plane) view for graphSAGE

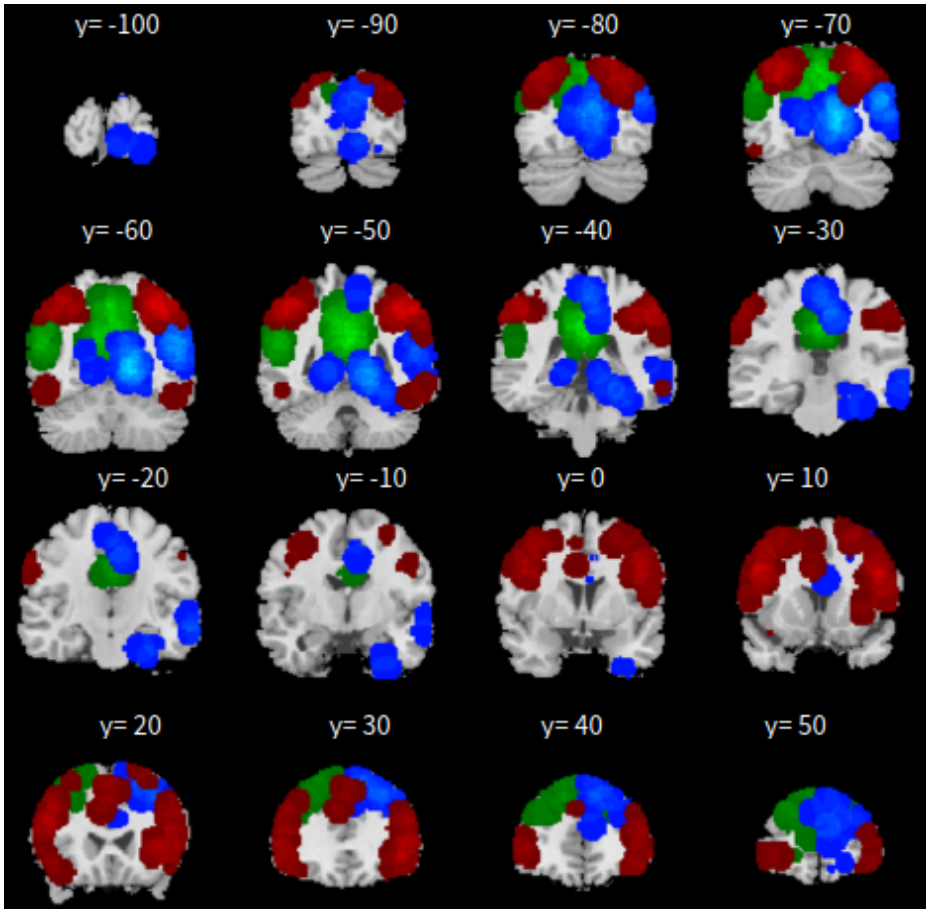
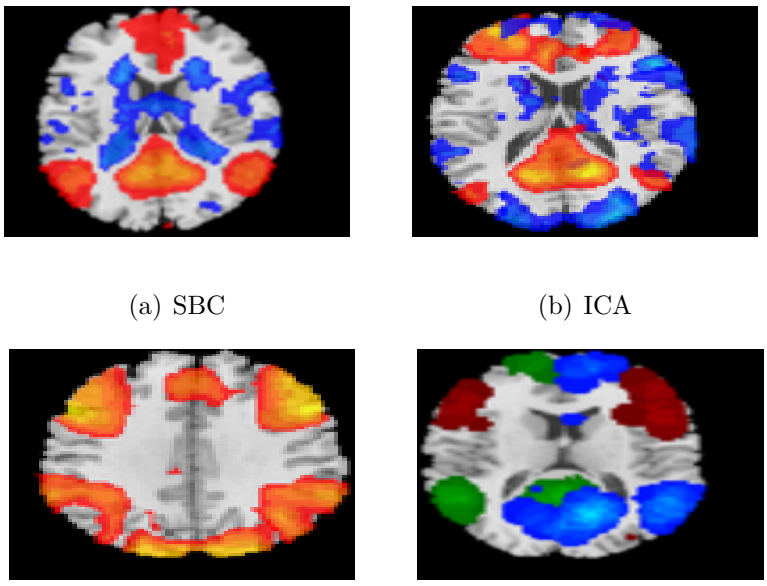


Figure 2.32. Coronal display (xz-plane) view for graphSAGE



(a) SBC (b) ICA

(c) DL (d) graphSAGE

Figure 2.33. Result comparison: DMN

to get an optimal solution to help extract the RSNs.

CHAPTER 3

ATTENTION DEFICIT AND HYPERACTIVITY DISORDER CLASSIFICATION

3.1 Overview

Attention Deficit and Hyperactivity Disorder (ADHD) is a mental disorder that commonly occurs in childhood. Based on American Psychiatric Association, the symptoms are characterized mainly in concentration with obvious difficulty, short duration of attention and hasty acting compared with children with the same age; 6.4 million children aging from 4 to 17 have been diagnosed ADHD and the average age of ADHD diagnosis is 7 years old. The prevalence of ADHD in the US is 5.5% to 9.3% in children. ADHD occurs not only in children but also in adults. An estimated 8.4% of children and 2.5% of adults have ADHD [49]. The study [55] mentioned that the estimated annual national extra cost for ADHD ranges from $143 billion to $266 billion. There are three main types of ADHD: inattentive type, Hyperactive/impulsive type and combined type.

ADHD negatively affects the patients both mentally and physically. It has become one of major problem in public health system. It is important to diagnose early and accurately whether or not one is in ADHD status. The common diagnosis of ADHD is often based on the Diagnostic and Statistical Manual of Mental Disorders (DSM), Published by the American Psychiatric Association (APA), and the symptoms are consistently exhibited for at least six months. The process of diagnosis of ADHD is a process of interview. It needs the clinician to make the decision based on the interview to the patient. This is somehow subjective [127] more or less since the right

decision needs a full evaluation for the patient. The functional magnetic resonance imaging (fMRI) teachnique offers a novel way to assess the patients with ADHD status. It has been used to study ADHD in different ways recently and obtained exciting results. The article [103] detected significant differences in cerebellum, motor cortex and temporal lobe between ADHD and normal humans between 73 children with ADHD and 76 normal children, and it used linear discriminant analysis classifier to distinguish health controls from ADHD children with an accuracy rate of 80.08% through 50 times of 2-fold validation. In the paper [117] the authors explored specific regions of interest (ROIs) to distinguish ADHD children from control children and reached an classification performance with a sensitivity rate of 90%. The paper [36] used both structural and functional magnetic resonance imaging data to predict diagnostic status of individuals with ADHD and got an accurate rate of 55%. There is no doubt that classification of ADHD is playing an essential role in the diagnosis applications.

3.2 Classical Methods

Several classical techniques can be used to classify ADHD patients from health controls. The techniques are reviewed in the following sections. In Section 3.2.1 the most common algorithm, the logistic model, is reviewed; in Section 3.2.2 the support vector machine (SVM) model including linear and non-linear format is reviewed; in Section 3.3 two deep learning algorithms are proposed; one is called independent component analysis with convolutional neural network (ICA-CNN), the other is called correlation-autoencoder model; in Section 3.4 a real data set is analyzed by all methods to compare their performance. Meanwhile the results are also compared with other previous studies for the same data set.

3.2.1 Logistic regression model

Logistic regression model is the most common technique to do classification task for binary outcomes. It plays an important role in machine learning. Given N samples (x_i, y_i), $i = 1, \cdots, N$; each x_i is a vector of p features and each y_i is the corresponding binary response with value 1 or 0. The purpose of logistic regression model is to model the quantity $\log(\frac{P(y=1|x)}{P(y=0|x)})$=$\log(\frac{P(y=1|x)}{1-P(y=1|x)})$ by a function $f(x, \theta)$, where θ is the estimated parameters for this function. This can be written as the follow form:

$$\log(\frac{P(y=1|x)}{1-P(y=1|x)}) = f(x,\theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p. \tag{3.1}$$

Equation (3.1) can be rewritten into another form as follows:

$$P(y=1|x) = \frac{1}{1+\exp(-\theta^\top x)}. \tag{3.2}$$

If the values is larger than 0.5, the model will classify the case into 1, otherwise the case belongs to 0. The optimal parameters θ is solved by the cost function:

$$\underset{\theta}{\text{Minimize}} \quad \frac{1}{N}\sum_{i=1}^{N}(y_i \log(\theta^\top x_i) + (1-y_i)\log(1-\theta^\top x_i)). \tag{3.3}$$

logistic regression often is penalized by ℓ_1 norm $\|\theta\|_1$ or ℓ_2 norm $\|\theta\|_2$.

3.2.2 Support vector machine model

Support vector machine (SVM) [40, 120] is another popular algorithm used in binary classification task and patter recognition. There are many studies related to support vector machine algorithm. It is a supervised machine learning technique and plays an important role in the family of machine learning. Due to its high

performance in most cases SVM model is used and developed in many fields in last two decades [32, 44, 64, 75, 159, 174, 182, 183].

Given N observations, each observation has p features with $x_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]$, y_i is the response value 1 or -1, $i = [1, 2, \ldots, N]$. The support vector machine finds the best hyperplane that maximizes the margins between two groups. If two groups are linearly separable, this leads to a hard margin optimization problem as follows:

$$\begin{aligned} &\underset{w}{\text{Minimize}} \quad \frac{1}{2}\|w\|_2^2 \\ &\text{Subject to} \quad y_i(w^\top x_i + b) \geq 1, \ \forall i = 1, 2, \ldots, N. \end{aligned} \tag{3.4}$$

If the two groups are not linearly separable, this leads to a soft margin optimization problem as follows:

$$\underset{w}{\text{Minimize}} \quad \frac{1}{N}\sum_{i=1}^{N}(1 - (w^\top x_i + b))_+ + \lambda\|w\|_2^2 \tag{3.5}$$

where λ is a regularization parameter and $(1 - (w^\top x_i + b))_+ = \max(0, 1 - (w^\top x_i + b))$ is the hinge loss function.

The Equation (3.4) and (3.5) are linear models but in some cases the linear model cannot reach an optimal solution and thus a nonlinear model of SVM via kernel tricks is applied. Let K indicate a kernel function [9] and two vectors a and b are given, the inner product for the kernel is in the form of $K(a, b) == \phi(a)^\top \phi(b)$, where ϕ is a map function; with the kernel function SVM can map features into high dimension space and makes classification task easier than in original space. The common used kernels in SVM are linear kernel $K(x, y) = x^\top y$, the Gaussian kernel $K(x, y) = \exp(-\sigma\|x - y\|_2^2)$ and the polynomial kernel $K(x, y) = (\sigma x^\top y + b)^d$. The

SVM based on kernel is given as follows:

$$\underset{f}{\text{Minimize}} \quad \frac{1}{N}\sum_{i=1}^{N}(1-(f(x_i)+b))_+ + \lambda\|f\|_K^2 \tag{3.6}$$

where $f = \beta^\top \phi(x)$ for β in the feature space.

3.3 Deep learning model

Deep learning, as a branch of machine learning, has been used in many fields and has made many excited achievements. In recent years deep learning was also used in classification of mental disorder based on fMRI data. Two major deep learning models, feedforward neural network (FNN) and convolutional neural network (CNN), were often used in previous studies. In the article [158] the authors used an autoencoder model to help examining the relations among the regions of brain to classify Mild Cognitive Impairment (MCI) patient. The authors of [79] used two encoders to classify autism spectrum disorder (ASD) patients from the database known as Autism Brain Imaging Data Exchange (ABIDE). In [141] the authors proposed to use CNN model and the architecture LeNet-5 to classify Alzheimer's disease (AD) patients from normal controls. The authors of [193] used a 3D-CNN to extract features from both functional and structural MRI and then used a CNN model to combine these features to classify ADHD patients from normal controls.

In this dissertation, two deep learning models are proposed to classify ADHD patients from normal controls. One is to incorporate independent component analysis with convolutional neural network (ICA-CNN) based on rs-fMRI data and the other is to incorporate correlation between nodes with one-dimension deep convolutional autoencoder model (Correlation-Autoencoder) based on rs-fMRI. The details of each framework is described as follows.

3.3.1 ICA-CNN model

Convolutional neural network (CNN), a popular deep learning technique, is used in many fields and has made many excited achievements especially in image recognition since it was proposed in [98]. One of the big advantages of CNN is the shared convolution kernel to extract features faster and effectively.

For a given ADHD data set, the data includes N subjects and each subject has n volumes with dimension of $l \times w \times h$, the following steps are performed.

1) First m independent components are extracted from each of N subjects based on the canonical ICA algorithm (canICA) [170].

2) After extracting m components from all subjects, a masker based on the m components is made to extract the corresponding signals for each subject.

3) After being extracted the signal based on the masker, each subject has a matrix with dimension $n \times m$, where n is the number of volumes. Totally, there are N matrices with dimension $n \times m$.

4) A convolutional neural network (CNN) is built with several convolutional layers including pooling layers, dropout layers, batch-normalization layer, flatten layer and dense layers. Each convolutional layer has activation of rectified linear activation unit (ReLU). The last layer is dense layer with the sigmoid activation function. The N matrices are as input for this CNN model and the output is the medical status of subjects. The value of 0 means normal control and 1 means ADHD status.

5) The model is trained by back-propagation algorithm with mini-batch method. The parameters w and bias b are updated according to the following rules until

the convergence,

$$w^l \rightarrow w^l - \frac{\lambda}{k}\sum_x \frac{\partial C(w,b)}{\partial w^l} \tag{3.7}$$

$$b^l \rightarrow b^l - \frac{\lambda}{k}\sum_x \frac{\partial C(w,b)}{\partial b^l} \tag{3.8}$$

where λ is the learning rate, k is the number of batch size, l is the lth layer, x is the data in a batch size, $C(w,b)$ is the binary cost function of the form:

$$C(w,b) = -ylog(\hat{y}) - (1-y)log(1-\hat{y}) \tag{3.9}$$

where $\hat{y}$=$f(w,b,x)$. The architecture of this proposed method is shown in Figure 3.1.

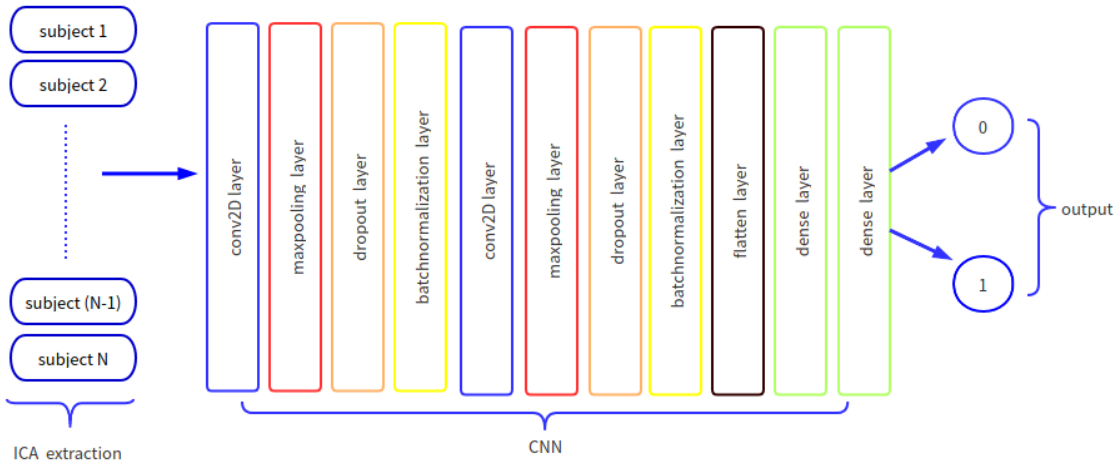


Figure 3.1. Framework of ICA-CNN method: *each subject is extracted a matrix with dimension of $n \times m$ as inputs for CNN model, n is the number of volumes and m is the number of components; the last dense layer has two nodes with sigmoid activation function.*

3.3.2 Correlation-autoencoder model

Autoencoder [67,105] is an unsupervised learning method. It consists of two neural networks: an encoder and a decoder. The main purpose of an autoencoder is to learn the latent representations denoted by μ in a lower dimension from the input denoted by x. Generally the encoder network compresses x into μ and the decoder network rebuilds μ into $\hat{x}$ so that $x \approx \hat{x}$. The autoencoder model is often used in dimension reduction [81,177], image reconstruction [162] and noise reduction [185,187].

For a given ADHD data set, the data includes N subjects and each has n volumes. A brain atlas is needed for partitioning the brain into the regions of interest (ROIs). In this application the Schaefer atlas [143] is used to partition the brain to extract the signal from the rs-fMRI data, and the algorithm follows the steps below.

1) First the Schaefer atlas [143] is used to partition the brain into r regions of interest (ROIs) and help extract the signal from rs-fMRI for each subject. Each ROI corresponds a time series data with length t, t is the number of time points of each subject.

2) For each subject the correlation is calculated between every two ROIs. This results in a square matrix with dimension $r \times r$ for each subject. Since this matrix is symmetric and only the upper triangular matrix is reserved and then it is flattened into a vector with length $(r^2+r)/2$. As a result the original rs-fMRI data of N subjects with n volume for each is transferred into N observations with each having $(r^2 + r)/2$ features.

3) The N observations, with $(r^2+r)/2$ features for each, are the input denoted by x for the encoder model and the input shape is $(r^2+r)/2$ by 1. The encoder model includes convolutional 1D layers, maxpooling 1D layers and batch-normalization

layers. The purpose of encoder model is to extract the representation of input in a lower dimension a, and $a \ll (r^2 + r)/2$.

4) Then the extracted representation $a \times 1$ is as input for the decoder model. The decoder model includes convolutional 1D layers, upsampling 1D layers and batch-normalization layers. The purpose of the decoder model is to rebuild the extracted representation a into $\hat{x}$ with dimension $(r^2 + r)/2$. The encoder and the decoder are trained together as an autoencoder model and the loss function is mean square error:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x_i})^2 \tag{3.10}$$

5) After the autoencoder model is trained, the encoder model is used to build the new network for binary classification. This new model first keeps all the layers from encoder model and the weights of the encoder model are fixed. In addition, the flatten layer and dense layers are added in the model. This new model is trained and used to do classification task with binary loss function:

$$C(w, b) = -ylog(\hat{y}) - (1 - y)log(1 - \hat{y}) \tag{3.11}$$

where $\hat{y}$=$f(w, b, x)$. The architecture of this proposed method is shown in Figure 3.2.

3.4 Case study

3.4.1 Data description

There is a publicly available ADHD-200 database, the 1000 Functional Connectomes Project, which is provided by `http://fcon_1000.projects.nitrc.org/indi/adhd200/`.

The dataset has 973 resting state fMRI acquisitions labeled with ADHD or typically developing controls (TDC) from subjects aging from 7 to 21. The dataset is collected from eight different international sites, which are NeuroImage group (NeuroImage), New York University Child Study Center (NYU), Peking University (Peking), Brown University (BU), Kennedy Krieger Institute (KKI), University of Pittsburgh (Pitts), Oregon Health and Science University (OHSU) and Washington University in St. Louis (WU). The ADHD-200 dataset was used in a competition to classify ADHD in 2011 and then the Preprocessed Connectomes Project (PCP) made the competition accessible to a broader range by preprocessing the data and sharing the results to public. Three different pipelines were used to preprocess the fMRI data. The preprocess data is used in the subsequent analysis based on Athena pipeline [18] which was performed by Cameron Craddock using AFNI and FSL running on the Athena computer cluster at Virginia Tech's ARC. There are more details on preprocessing for both structural data and functional data through the website of `https://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline`. The functional image is using a 6mm FWHM Gaussian filter and band pass filter is between 0.009 and 0.08 Hz.

These data was used for global competition of classification task in 2011, and each data was also provided a holdout test data. In the following analysis, the data from the site of Peking University is used. There are totally 198 subjects in the training data, including 123 typical controls and 75 ADHD (ADHD-Hyperactive/Impulsive, ADHD-inattentive and ADHD-combined) subjects, and the test data has 51 subjects including 24 ADHD subjects and 27 typical controls. In practice the training data only has 168 subjects due to preprocessing and the test data is still the same as the original test data. The final data to be used is summarized in Table 3.1. The

functional brain of one random chosen subject with dimension $49 \times 58 \times 47 \times 232$ is shown in Figure 3.3.

Table 3.1. Summary of original data and used data

	Training data			Test data		
	ADHD	Typical control	Total	ADHD	Typical control	Total
Original data	75	123	198	24	27	51
Used data	69	99	168	24	27	51

3.4.2 Data analysis

The data is analysed by the traditional logistic regression model and support vector machine (SVM) model as well as the two proposed methods. The extracted features are used in all the four methods. The two proposed methods are also compared with the traditional methods and other methods in the previous literature.

(i) **Analysis one**

In this dataset some subjects have 231 volumes and some subjects have 232 volumes, in order to extract the same number of independent components all subjects are set to 231 volumes by slicing. A number of 100 independent components are extracted from each volume for each subject, so each subject has 231 times series of length 100. Totally there are 168 subjects, so there are 168 matrices with dimension of 231×100 for training data and 51 matrices with dimension of 231×100 for testing data.

For logistic regression model, each volume of each subject is extracted 100 component, and each subject has 231 volumes so the average of the 100 components from all 231 volumes is as the input for logistic regression model. The model

gives an accuracy rate of 55%, sensitivity 33% and specificity 74%, the area under receiver operating characteristics (ROC) curve 0.54.

For support vector machine model, the features are also the average based on the 231 time series. This model with Gaussian kernel gives an accuracy rate of 55%, sensitivity 4% and specificity 100%, the area under receiver operating characteristics (ROC) curve 0.52.

For ICA-CNN model, the input is the matrix with dimension of 231×100 after extracting the independent components from each subject, and the convolutional neural network is used to classify the ADHD patient. The convolutional layers use 32 and 64 filters, kernel size is 3×3, maxplooing size is 2×2, stride step is 1, dropout rate is 0.2, the activation function is rectified linear unit (ReLU) [119], the learning rate is 0.001 and the optimization function is Adam [90], the last dense layer has two nodes with sigmoid function. The fine-tuned model gives an accuracy rate of 67%, sensitivity rate 42%, specificity rate 89% and the area under ROC curve 0.65. The ROC curves for logistic regression model, support vector machine model and ICA-CNN are shown in Figure 3.4. The results are also compared with other studies based on machine learning and deep learning algorithms in the literature. See Table 3.2.

(ii) **Analysis two**

In the second analysis each subject is partitioned into 300 regions of interest (ROIs) based on the Schaefer atlas [143]. Since each subject is set to 231 volumes there are 231 time series for each ROI. Then the correlation between each pair of ROIs is calculated. For each subject this gives us a square matrix of correlation with dimension 300×300. Since this matrix is symmetric and

Table 3.2. Comparison of results of different models

Model	Accuracy	Specificity	Sensitivity	ROC
Competition [140]	51.05%	-	-	-
Logistic	55%	74%	33%	0.54
SVM(Gaussian kernel)	55%	100%	4%	0.52
DeepFMRI [138]	62.7%	79.1%	48.1%	-
3D CNN [193]	62.95%	-	-	-
Deep forest [150]	64.87%	-	-	-
ICA-CNN	**67%**	**89%**	**42%**	**0.65**

only the upper triangular part is used and is flattened into a vector with the length of 45150. Totally there are 168 subjects with 45150 features for training data and 51 subjects with 45150 features for testing data.

For logistic regression model, a variable selection algorithm is used to choose the most important features from the 45150 features. The Recursive Feature Elimination algorithm (RFE) [70] is used to select the most important features. The basic idea of RFE is to build a full model at the beginning and drop the least important feature, then a new model is built again with the rest features, and the process is repeated until the necessary number of features are reached. The RFE algorithm is a backward selection feature method. Two options are used for logistic regression model, one used the most important 100 features and the other used the most important 200 features. The model with the most important 100 features gives an accurate rate of 61%, specificity rate 85%, sensitivity rate 33% and the area under ROC 0.59. The model with the most important 200 features gives an accurate rate of 65%, specificity rate 89%, sensitivity rate 38% and the area under ROC 0.63.

For SVM model with Gaussian kernel, the RFE algorithm again is used to choose the most important features from the 45150 features. The model with the most important 100 features gives an accuracy rate of 61%, specificity rate 85%, sensitivity rate 33% and the area under the ROC 0.59. The model with the most important 200 features gives an accurate rate of 65%, specificity rate 89%, sensitivity rate 38% and the area under ROC 0.63.

The correlation-autoencoder model includes three parts: encoder, decoder and prediction. The encoder and the decoder make an autoencoder model which takes all the 45150 features as input. The purpose of this autoencoder model is to extract the latent features which can represent the input as much as possible. The encoder includes convolutional layers with filters of 1024, 512 and 128, kernel size is 3, maxpooling size is 2 and all the activation functions are ReLU. The purpose of the decoder is to rebuilt the input features as much as possible based on the latent features. The decoder model includes convolutional layers with filters 256, 512 and 256, kernel size is 3, upsampling size is 2 and all the activation functions are ReLU. The learning rate is 0.001. After the autoencoder is trained and the encoder is to hold fixed, that is all the weights from the encoder part are fixed and are used for the prediction part. The prediction model includes the encoder part and an extra part with a flatten layer and two dense layers. For this prediction model the input is the same as autoencoder model and contains all the 45150 features. The response is the ADHD with value of 1 and typical controls with values 0. In this prediction model only the last two dense layers are trained, first dense layer has 128 nodes and the last dense layer has two nodes with the activation function of sigmoid. The fine-tuned model gives an accuracy rate of 69%, specificity rate 89%, sensitivity rate

46% and the area under the ROC curve 0.67. The results are also compared with other studies based on machine learning and deep learning algorithm in the literature. Details are shown in Table 3.3. The ROC curve is shown in Figure 3.5.

Table 3.3. Comparison of results of different models based on correlation

Model	Accuracy	Specificity	Sensitivity	ROC
Competition [140]	51.05%	-	-	-
Logistic with 100	61%	85%	33%	0.59
SVM with 100(Gaussian kernel)	61%	85%	33%	0.59
Logistic with 200	65%	89%	38%	0.63
SVM with 200(Gaussian kernel)	65%	89%	38%	0.63
DeepFMRI [138]	62.7%	79.1%	48.1%	-
3D CNN [193]	62.95%	-	-	-
Deep forest [150]	64.87%	-	-	-
Correlation-autoencoder	**69%**	**89%**	**46%**	**0.67**

3.5 Discussion and conclusion

Two proposed methods are used to classify the ADHD from typical controls in this chapter, one is ICA-CNN and the other is correlation-autoencoder. Both two methods outperform the existing models. Even the traditional binary classification model of logistic regression and SVM with the most important 200 features based on correlation also outperform the other models from previous studies. Based on 100 independent components extracted from each subject the ICA-CNN model gives an accuracy rate of 67%, and based on 300 regions of interest extracted from each subject the correlation-autoencoder model gives an accuracy rate of 69%. The correlation-autoencoder model gives a better prediction compared with the ICA-CNN model. For traditional binary classification algorithms the models based on correlation also give

better results compared with models based on independent components. SVM model and logistic regression model with the most important 200 features give an accuracy rate of 65% which is 10% higher compared with the logistic model and SVM with 100 independent components. The area under ROC is 0.67 for correlation autoencoder model and 0.65 for ICA-CNN model. Both models have the same specificity rate of 89%. Correlation autoencoder has a sensitivity rate of 46% and ICA-CNN has a sensitivity of 42%.

Both proposed methods give the accuracy rate less than 70%, to a certain extent it is some guidance for classification in clinical applications. Due to non-linear learning strategy with multiple layers, deep learning methods often give much better performance in many applications. But they usually need a big sample data to train thousands of parameters to learn the latent features. In this data analysis the training sample contains only 168 observations which are not enough to have the parameters trained very well. It is reasonable to believe that the performance based on the two proposed methods will get better if the training data is large and the computer resources are allowed.

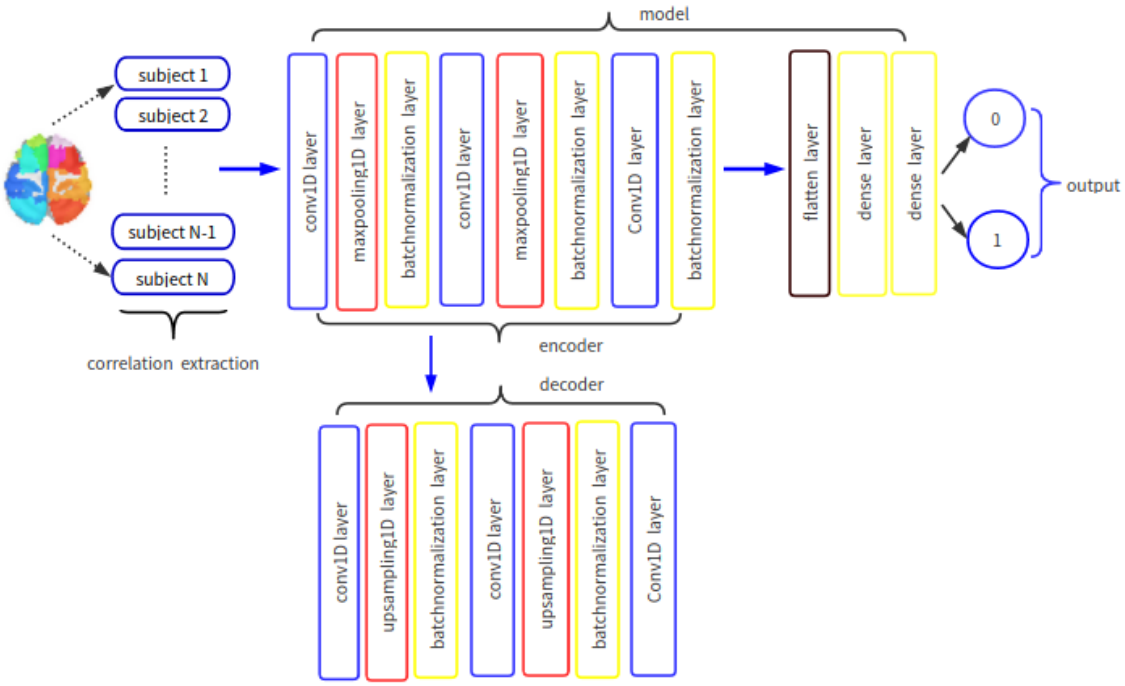


Figure 3.2. Framework of correlation autoencoder method: *each subject is partitioned into r ROIs, that is, $n \times r$ matrix for each subject, n is the number of volumes; then the correlation between the r ROIs is calculated, then the upper triangular matrix of $r \times r$ is flattened as input with length $(r^2 + r)/2$ for the autoencoder model; the last dense layer in the model has two nodes with sigmoid activation function.*

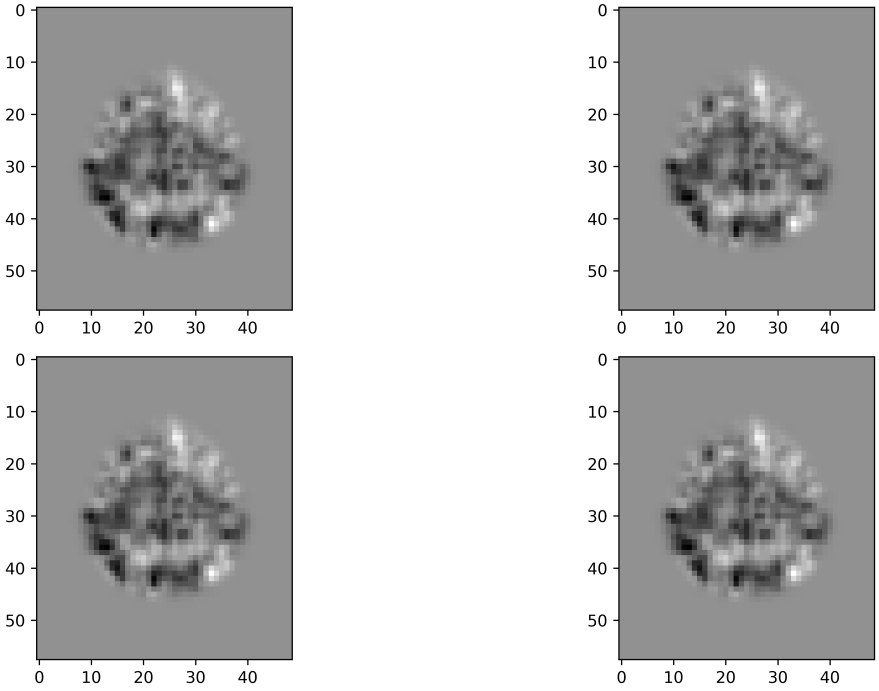


Figure 3.3. The functional brain of one random chosen subject

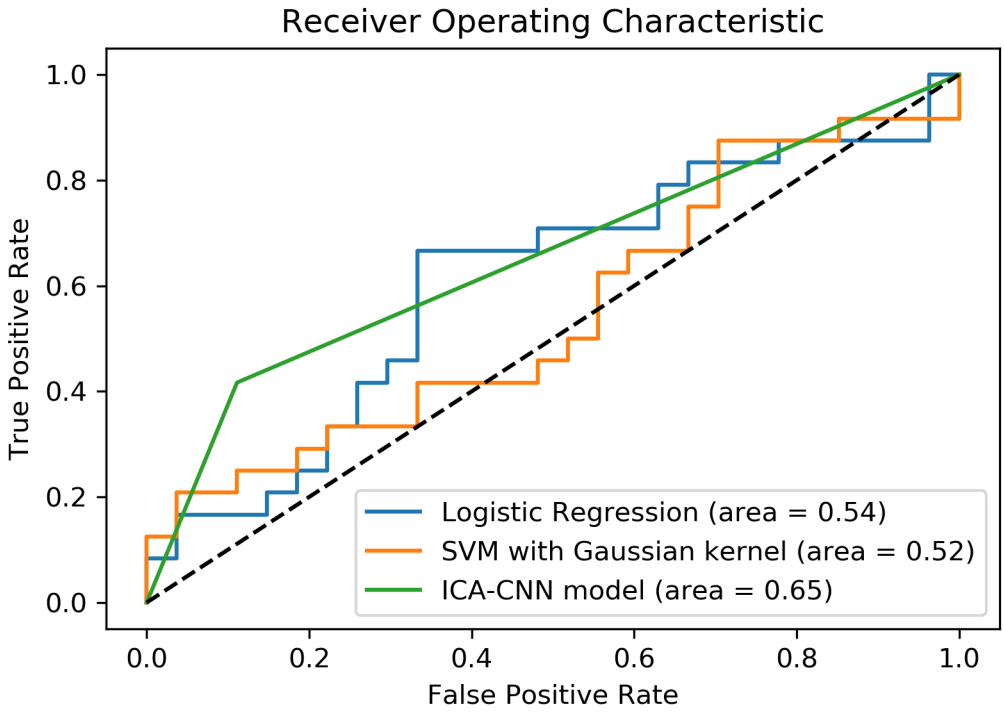


Figure 3.4. Receiver operating characteristics (ROC) curve for logistic regression, SVM and ICA-CNN models

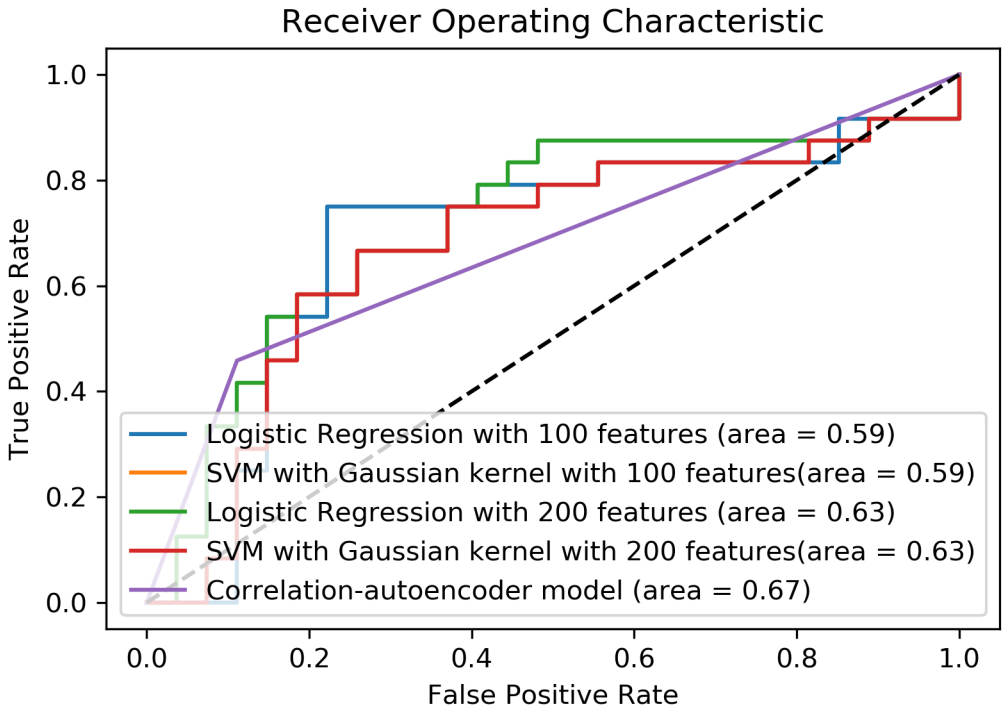


Figure 3.5. Receiver operating characteristics (ROC) curve for logistic regression, SVM and Correlation-autoencoder models

CHAPTER 4

SUMMARY AND FUTURE WORK

4.1 Summary

The resting state functional Magnetic Resonance Imaging data (rs-fMRI) has been used in exploring the functional connection information in many ways. In resting state the human brain is still active without any external interference and the parts of the brain are connected with its intrinsic connections. It is important to explore and identify these intrinsic connectivity networks or resting state networks (RSNs). This helps better understand the functional connections between RSNs, especially for discovering the difference of functional changes of RSNs between healthy and unhealthy individuals and thus helps provide a more accurate treatment plan in clinical applications. In this dissertation we proposed the application of a deep learning method called graphSAGE to extract ROIs based on two rs-fMRI data, both analyses showed the results as we expected, especially in extracting the default mode network (DMN). Compared with the traditional methods like seed-based analysis, ICA and dictionary learning method, the application of graphSAGE has less prior assumptions and can incorporate individual and group analysis. It simplifies the complexity of the analysis to a certain extent and makes the result more robust.

Attention deficit hyperactivity disorder (ADHD) is a type of mental disorder which often occurs in the young ages. One of the main symptoms of the patients with ADHD usually is difficult to focus on tasks and the patients can be distracted easily. Therefore, the accurate diagnosis as early as possible will be important for the treatment of the patient in clinical applications. In this dissertation we proposed two

methods to classify ADHD patients from typical controls. One method is called ICA-CNN model which uses convolutional neural network with extracted independent components as input from each subject to classify ADHD patients from typical controls. The other method is called correlation autoencoder model which uses an autoencoder model to extract the latent features based on the correlation between the regions of interest (ROIs), then the encoder part is used to build a new neural model to classify the ADHD patients from typical controls. Both methods give better classification performance than the traditional logistic regression model and SVM model. In addition, both methods give better performance than the previous studies. The two methods provide a way on classification task on ADHD.

4.2 Future work

The graphSAGE as a deep learning algorithm needs a large data to train the model to get the optimal solutions. As the large training sample becomes available it will further improve the application of graphSAGE. It will help not only to extract the default mode network (DMN) but also to extract other RSNs more accurately.

In the future work we can try different large ADHD data if data is available and computer resources are available. Another idea is to use different RSNs extracted to classify the ADHD patient from typical controls. We can aslo apply the proposed algorithms to the classification of mental diseases such as Alzheimer's disease and depression. The computation cost is expensive and incorporation of parallel computation framework may also be considered in the future work.

REFERENCES

[1] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8:14, 2014.

[2] Sandra Ackerman et al. *Discovering the brain*. National Academies Press, 1992.

[3] Federica Agosta, Michela Pievani, Cristina Geroldi, Massimiliano Copetti, Giovanni B Frisoni, and Massimo Filippi. Resting state fmri in alzheimer's disease: beyond the default mode network. *Neurobiology of aging*, 33(8):1564–1578, 2012.

[4] Michal Aharon and Michael Elad. Sparse and redundant modeling of image content using an image-signature-dictionary. *SIAM Journal on Imaging Sciences*, 1(3):228–247, 2008.

[5] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[6] Edson Amaro Jr and Gareth J Barker. Study design in fmri: basic principles. *Brain and cognition*, 60(3):220–232, 2006.

[7] Jessica R Andrews-Hanna, Jonathan Smallwood, and R Nathan Spreng. The default network and self-generated thought: component processes, dynamic

control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316(1):29, 2014.

[8] Gangani Ariyarathne, Senuri De Silva, Sanuwani Dayarathna, Dulani Meedeniya, and Sampath Jayarathne. Adhd identification using convolutional neural network with seed-based approach for fmri data. In *Proceedings of the 2020 9th International Conference on Software and Computer Applications*, pages 31–35, 2020.

[9] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[10] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.

[11] Feng Bai, Zhijun Zhang, Hui Yu, Yongmei Shi, Yonggui Yuan, Wanlin Zhu, Xiangrong Zhang, and Yun Qian. Default-mode network activity distinguishes amnestic type mild cognitive impairment from healthy aging: a combined structural and resting-state functional mri study. *Neuroscience letters*, 438(1):111–115, 2008.

[12] Peter A Bandettini, Eric C Wong, R Scott Hinks, Ronald S Tikofsky, and James S Hyde. Time course epi of human brain function during task activation. *Magnetic resonance in medicine*, 25(2):390–397, 1992.

[13] Alan Baumeister. The tulane electrical brain stimulation program a historical case study in medical ethics. *Journal of the History of the Neurosciences*, 9(3):262–278, 2000.

[14] Christian F Beckmann, Marilena DeLuca, Joseph T Devlin, and Stephen M Smith. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1001–1013, 2005.

[15] Christian F Beckmann, Clare E Mackay, Nicola Filippini, and Stephen M Smith. Group comparison of resting-state fmri data using multi-subject ica and dual regression. *Neuroimage*, 47(Suppl 1):S148, 2009.

[16] Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.

[17] P. T. Bell and J. M. Shine. Estimating Large-Scale Network Convergence in the Human Functional Connectome. *Brain Connect*, 5(9):565–574, Nov 2015.

[18] Pierre Bellec, Carlton Chu, Francois Chouinard-Decorte, Yassine Benhajali, Daniel S Margulies, and R Cameron Craddock. The neuro bureau adhd-200 preprocessed repository. *Neuroimage*, 144:275–286, 2017.

[19] Matt A Bernstein, Kevin F King, and Xiaohong Joe Zhou. *Handbook of MRI pulse sequences*. Elsevier, 2004.

[20] Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

[21] Bharat B Biswal, Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M Smith, Christian F Beckmann, Jonathan S Adelstein, Randy L

Buckner, Stan Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.

[22] Richard Bitar, General Leung, Richard Perng, Sameh Tadros, Alan R Moody, Josee Sarrazin, Caitlin McGregor, Monique Christakis, Sean Symons, Andrew Nelson, et al. Mr pulse sequences: what every radiologist wants to know but is afraid to ask. *Radiographics*, 26(2):513–537, 2006.

[23] Andrew M Blamire, Seiji Ogawa, Kamil Ugurbil, Douglas Rothman, Gregory McCarthy, Jutta M Ellermann, Fahmeed Hyder, Zachary Rattner, and Robert G Shulman. Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(22):11069–11073, 1992.

[24] S. L. Bressler and V. Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci. (Regul. Ed.)*, 14(6):277–290, Jun 2010.

[25] Glen D Brown, Satoshi Yamada, and Terrence J Sejnowski. Independent component analysis at the neural cocktail party. *Trends in neurosciences*, 24(1):54–63, 2001.

[26] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter. The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.*, 1124:1–38, Mar 2008.

[27] Carter T Butts. Social network analysis: A methodological introduction. *Asian Journal of Social Psychology*, 11(1):13–41, 2008.

[28] Richard B Buxton and Lawrence R Frank. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *Journal of cerebral blood flow & metabolism*, 17(1):64–72, 1997.

[29] Richard B Buxton, Eric C Wong, and Lawrence R Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic resonance in medicine*, 39(6):855–864, 1998.

[30] Vince Daniel Calhoun, T Adali, GD Pearlson, and James J Pekar. Spatial and temporal independent component analysis of functional mri data containing a pair of task-related waveforms. *Human brain mapping*, 13(1):43–53, 2001.

[31] F Xavier Castellanos and Yuta Aoki. Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: a science in development. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(3):253–261, 2016.

[32] Gert Cauwenberghs and Tomaso Poggio. Incremental and decremental support vector machine learning. In *Advances in neural information processing systems*, pages 409–415, 2001.

[33] Govind B Chavhan, Paul S Babyn, Bejoy Thomas, Manohar M Shroff, and E Mark Haacke. Principles, techniques, and applications of t2*-based mr imaging and its special applications. *Radiographics*, 29(5):1433–1449, 2009.

[34] Sneha Chenji, Shankar Jha, Dawon Lee, Matthew Brown, Peter Seres, Dennell Mah, and Sanjay Kalra. Investigating default mode and sensorimotor network connectivity in amyotrophic lateral sclerosis. *PLoS One*, 11(6), 2016.

[35] W. Chiong, S. M. Wilson, M. D'Esposito, A. S. Kayser, S. N. Grossman, P. Poorzand, W. W. Seeley, B. L. Miller, and K. P. Rankin. The salience network

causally influences default mode network activity during moral reasoning. *Brain*, 136(Pt 6):1929–1941, Jun 2013.

[36] John B Colby, Jeffrey D Rudie, Jesse A Brown, Pamela K Douglas, Mark S Cohen, and Zarrar Shehzad. Insights into multimodal imaging classification of adhd. *Frontiers in systems neuroscience*, 6:59, 2012.

[37] D. M. Cole, S. M. Smith, and C. F. Beckmann. Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Front Syst Neurosci*, 4:8, 2010.

[38] David M Cole, Stephen M Smith, and Christian F Beckmann. Advances and pitfalls in the analysis and interpretation of resting-state fmri data. *Frontiers in systems neuroscience*, 4:8, 2010.

[39] Roshan Cools, Roger A Barker, Barbara J Sahakian, and Trevor W Robbins. Mechanisms of cognitive set flexibility in parkinson's disease. *Brain*, 124(12):2503–2512, 2001.

[40] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[41] Robert W Cox and James S Hyde. Software tools for analysis and visualization of fmri data. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, 10(4-5):171–178, 1997.

[42] Bruce Crosson, Anastasia Ford, Keith M McGregor, Marcus Meinzer, Sergey Cheshkov, Xiufeng Li, Delaina Walker-Batson, and Richard W Briggs.

Functional imaging and related techniques: an introduction for rehabilitation researchers. *Journal of rehabilitation research and development*, 47(2):vii, 2010.

[43] Tim J Crow. Molecular pathology of schizophrenia: more than one disease process? *British medical journal*, 280(6207):66, 1980.

[44] Felipe Cucker and Ding Xuan Zhou. *Learning theory: an approximation theory viewpoint*, volume 24. Cambridge University Press, 2007.

[45] Thomas S Curry, James E Dowdey, and Robert C Murry. *Christensen's physics of diagnostic radiology*. Lippincott Williams & Wilkins, 1990.

[46] mohammad reza Daliri and Mehdi Behroozi. Advantages and disadvantages of resting state functional connectivity magnetic resonance imaging for clinical applications. *OMICS Journal of Radiology*, 3, 01 2014.

[47] MR Daliri and M Behroozi. Advantages and disadvantages of resting state functional connectivity magnetic resonance imaging for clinical applications. *OMICS Journal of Radiology*, 3(1):1–2, 2013.

[48] Jessica S Damoiseaux, SARB Rombouts, Frederik Barkhof, Philip Scheltens, Cornelis J Stam, Stephen M Smith, and Christian F Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*, 103(37):13848–13853, 2006.

[49] Melissa L Danielson, Rebecca H Bitsko, Reem M Ghandour, Joseph R Holbrook, Michael D Kogan, and Stephen J Blumberg. Prevalence of parent-reported adhd diagnosis and associated treatment among us children and adolescents, 2016. *Journal of Clinical Child & Adolescent Psychology*, 47(2):199–212, 2018.

[50] CSIRO's Data61. Stellargraph machine learning library. `https://github.com/stellargraph/stellargraph`, 2018.

[51] Graham C Davey. *Applied psychology*. John Wiley & Sons, 2011.

[52] MBCFDS De Luca, Christian F Beckmann, Nicola De Stefano, Paul M Matthews, and Stephen M Smith. fmri resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage*, 29(4):1359–1367, 2006.

[53] Bruno Goulart de Oliveira, José Osmar Alves Filho, Nathalia Bianchini Esper, Dario Francisco Guimaraes de Azevedo, and Alexandre R Franco. Automated mapping of sensorimotor network for resting state fmri data with seed-based correlation analysis. In *XXVI Brazilian Congress on Biomedical Engineering*, pages 537–544. Springer, 2019.

[54] Adriana Di Martino, Anouk Scheres, Daniel S Margulies, AMC Kelly, Lucina Q Uddin, Zarrar Shehzad, Bharat Biswal, Judith R Walters, F Xavier Castellanos, and Michael P Milham. Functional connectivity of human striatum: a resting state fmri study. *Cerebral cortex*, 18(12):2735–2747, 2008.

[55] Jalpa A Doshi, Paul Hodgkins, Jennifer Kahle, Vanja Sikirica, Michael J Cangelosi, Juliana Setyawan, M Haim Erder, and Peter J Neumann. Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the united states. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(10):990–1002, 2012.

[56] Oscar Esteban, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, Asier Erramuzpe,

Mathias Goncalves, Russell A. Poldrack, and Krzysztof J. Gorgolewski. poldracklab/fmriprep: 1.0.0-rc5, September 2017.

[57] Oscar Esteban, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, et al. fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111–116, 2019.

[58] Alan C Evans, D Louis Collins, SR Mills, ED Brown, RL Kelly, and Terry M Peters. 3d statistical neuroanatomical models from 305 mri volumes. In *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pages 1813–1817. IEEE, 1993.

[59] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[60] Michael D Fox, Maurizio Corbetta, Abraham Z Snyder, Justin L Vincent, and Marcus E Raichle. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, 103(26):10046–10051, 2006.

[61] Michael D Fox and Marcus E Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience*, 8(9):700–711, 2007.

[62] Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27):9673–9678, 2005.

[63] Karl J Friston, Steven Williams, Robert Howard, Richard SJ Frackowiak, and Robert Turner. Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, 35(3):346–355, 1996.

[64] Glenn M Fung and Olvi L Mangasarian. Multicategory proximal support vector machine classifiers. *Machine learning*, 59(1-2):77–97, 2005.

[65] Piotr Gałecki, Monika Talarowska, George Anderson, Michael Berk, and Michael Maes. Mechanisms underlying neurocognitive dysfunctions in recurrent major depression. *Medical science monitor: international medical journal of experimental and clinical research*, 21:1535, 2015.

[66] Gary H Glover. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, 22(2):133–139, 2011.

[67] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[68] Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258, 2003.

[69] Vijay PB Grover, Joshua M Tognarelli, Mary ME Crossey, I Jane Cox, Simon D Taylor-Robinson, and Mark JW McPhail. Magnetic resonance imaging: principles and techniques: lessons for clinicians. *Journal of clinical and experimental hepatology*, 5(3):246–255, 2015.

[70] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[71] Andreas Hahn, Patrycja Stein, Christian Windischberger, Andreas Weissenbacher, Christoph Spindelegger, Ewald Moser, Siegfried Kasper, and Rupert Lanzenberger. Reduced resting-state functional connectivity between amygdala and orbitofrontal cortex in social anxiety disorder. *Neuroimage*, 56(3):881–889, 2011.

[72] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034, 2017.

[73] Ying Han, Jinhui Wang, Zhilian Zhao, Baoquan Min, Jie Lu, Kuncheng Li, Yong He, and Jianping Jia. Frequency-dependent changes in the amplitude of low-frequency fluctuations in amnestic mild cognitive impairment: a resting-state fmri study. *Neuroimage*, 55(1):287–295, 2011.

[74] Marwan I HarIz, Patric Blomstedt, and Ludvic Zrinzo. Deep brain stimulation between 1947 and 1987: the untold story. *Neurosurgical focus*, 29(2):E1, 2010.

[75] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[76] David J Heeger and David Ress. What does fmri tell us about neuronal activity? *Nature Reviews Neuroscience*, 3(2):142–151, 2002.

[77] L. Heine, A. Soddu, F. G?mez, A. Vanhaudenhuyse, L. Tshibanda, M. Thonnard, V. Charland-Verville, M. Kirsch, S. Laureys, and A. Demertzi. Resting state networks and consciousness: alterations of multiple resting state network connectivity in physiological, pharmacological, and pathological consciousness States. *Front Psychol*, 3:295, 2012.

[78] Lizette Heine, Andrea Soddu, Francisco Gómez, Audrey Vanhaudenhuyse, Luaba Tshibanda, Marie Thonnard, Vanessa Charland-Verville, Murielle Kirsch, Steven Laureys, and Athena Demertzi. Resting state networks and consciousness. *Frontiers in psychology*, 3:295, 08 2012.

[79] Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, 17:16–23, 2018.

[80] Dries Hendrikx, Anne Smits, Mario Lavanga, Ofelie De Wel, Liesbeth Thewissen, Alexander Caicedo, Sabine Van Huffel, Gunnar Naulaers, et al. Measurement of neurovascular coupling in neonates. *Frontiers in physiology*, 10:65, 2019.

[81] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[82] Seyed Hani Hojjati, Ata Ebrahimzadeh, Ali Khazaee, Abbas Babajani-Feremi, Alzheimer's Disease Neuroimaging Initiative, et al. Predicting conversion from mci to ad using resting-state fmri, graph theoretical approach and svm. *Journal of neuroscience methods*, 282:69–80, 2017.

[83] J. S. Hutton, J. Dudley, T. Horowitz-Kraus, T. DeWitt, and S. K. Holland. Functional Connectivity of Attention, Visual, and Language Networks During Audio, Illustrated, and Animated Stories in Preschool-Age Children. *Brain Connect*, 9(7):580–592, 09 2019.

[84] John S Hutton, Jonathan Dudley, Tzipi Horowitz-Kraus, Tom DeWitt, and Scott K Holland. Functional connectivity of attention, visual, and language networks during audio, illustrated, and animated stories in preschool-age children. *Brain connectivity*, 9(7):580–592, 2019.

[85] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[86] H. A. Jaber, H. K. Aljobouri, İ. Çankaya, O. M. Koçak, and O. Algin. Preparing fmri data for postprocessing: Conversion modalities, preprocessing pipeline, and parametric and nonparametric approaches. *IEEE Access*, 7:122864–122877, 2019.

[87] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[88] Hang Joon Jo, Stephen J Gotts, Richard C Reynolds, Peter A Bandettini, Alex Martin, Robert W Cox, and Ziad S Saad. Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state fmri. *Journal of applied mathematics*, 2013, 2013.

[89] Mohsen Joneidi. Functional brain networks discovery using dictionary learning with correlated sparsity. *arXiv preprint arXiv:1907.03929*, 2019.

[90] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[91] Kenneth Kreutz-Delgado, Joseph F Murray, Bhaskar D Rao, Kjersti Engan, Te-Won Lee, and Terrence J Sejnowski. Dictionary learning algorithms for sparse representation. *Neural computation*, 15(2):349–396, 2003.

[92] Prantik Kundu, Noah D Brenowitz, Valerie Voon, Yulia Worbe, Petra E Vértes, Souheil J Inati, Ziad S Saad, Peter A Bandettini, and Edward T Bullmore. Integrated strategy for improving functional connectivity mapping using multiecho fmri. *Proceedings of the National Academy of Sciences*, 110(40):16187–16192, 2013.

[93] Kenneth K Kwong, John W Belliveau, David A Chesler, Inna E Goldberg, Robert M Weisskoff, Brigitte P Poncelet, David N Kennedy, Bernice E Hoppel, Mark S Cohen, and Robert Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679, 1992.

[94] Jack L Lancaster, Diana Tordesillas-Gutiérrez, Michael Martinez, Felipe Salinas, Alan Evans, Karl Zilles, John C Mazziotta, and Peter T Fox. Bias between mni and talairach coordinates analyzed using the icbm-152 brain template. *Human brain mapping*, 28(11):1194–1205, 2007.

[95] Paul C Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *nature*, 242(5394):190–191, 1973.

[96] Nicole Lazar. *The statistical analysis of functional MRI data*. Springer Science & Business Media, 2008.

[97] H. C. L?ckmann, H. I. Jacobs, and A. T. Sack. The cross-functional role of frontoparietal regions in cognition: internal attention as the overarching mechanism. *Prog. Neurobiol.*, 116:66–86, May 2014.

[98] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[99] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808, 2007.

[100] Joseph KT Lee. *Computed body tomography with MRI correlation*, volume 1. Lippincott Williams & Wilkins, 2006.

[101] Megan H Lee, Christopher D Smyser, and Joshua S Shimony. Resting-state fmri: a review of methods and clinical applications. *American Journal of neuroradiology*, 34(10):1866–1872, 2013.

[102] X. Lei, Y. Wang, H. Yuan, and D. Mantini. Neuronal oscillations and functional interactions between resting state networks. *Hum Brain Mapp*, 35(7):3517–3528, Jul 2014.

[103] Sheng-Fu Liang, Tsung-Hao Hsieh, Pin-Tzu Chen, Ming-Long Wu, Chun-Chia Kung, Chun-Yu Lin, and Fu-Zen Shaw. Differentiation between resting-state fmri data from adhd and normal subjects: based on functional connectivity and machine learning. In *2012 International conference on Fuzzy Theory and Its Applications (iFUZZY2012)*, pages 294–298. IEEE, 2012.

[104] Martin A Lindquist et al. The statistical analysis of fmri data. *Statistical science*, 23(4):439–464, 2008.

[105] Cheng-Yuan Liou, Wei-Chen Cheng, Jiun-Wei Liou, and Daw-Ran Liou. Autoencoder for words. *Neurocomputing*, 139:84–96, 2014.

[106] Thomas T Liu. Noise contributions to the fmri signal: an overview. *NeuroImage*, 143:141–151, 2016.

[107] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[108] Mark J Lowe, Micheal D Phillips, Joseph T Lurito, David Mattson, Mario Dzemidzic, and Vincent P Mathews. Multiple sclerosis: low-frequency temporal blood oxygen level–dependent fluctuations indicate reduced functional connectivity—initial results. *Radiology*, 224(1):184–192, 2002.

[109] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696, 2009.

[110] S. Marek and N. U. F. Dosenbach. The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin Neurosci*, 20(2):133–140, 06 2018.

[111] MATLAB. *9.7.0.1190202 (R2019b)*. The MathWorks Inc., Natick, Massachusetts, 2018.

[112] Martin J McKeown, Scott Makeig, Greg G Brown, Tzyy-Ping Jung, Sandra S Kindermann, Anthony J Bell, and Terrence J Sejnowski. Analysis of fmri data by blind separation into independent spatial components. *Human brain mapping*, 6(3):160–188, 1998.

[113] Kevin R McLeod, Lisa Marie Langevin, Bradley G Goodyear, and Deborah Dewey. Functional connectivity of neural motor networks is disrupted in children with developmental coordination disorder and attention-deficit/hyperactivity disorder. *NeuroImage: Clinical*, 4:566–575, 2014.

[114] Katie L McMahon, Gary Cowin, and Graham Galloway. Magnetic resonance imaging: the underlying principles. *journal of orthopaedic & sports physical therapy*, 41(11):806–819, 2011.

[115] Ravi S Menon. Postacquisition suppression of large-vessel bold signals in high-resolution fmri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(1):1–9, 2002.

[116] Regina J Meszlényi, Krisztian Buza, and Zoltán Vidnyánszky. Resting state fmri functional connectivity-based classification using a convolutional neural network architecture. *Frontiers in neuroinformatics*, 11:61, 2017.

[117] Yukifumi Monden, Ippeita Dan, Masako Nagashima, Haruka Dan, Minako Uga, Takahiro Ikeda, Daisuke Tsuzuki, Yasushi Kyutoku, Yuji Gunji, Daisuke Hirano, et al. Individual classification of adhd children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fnirs. *NeuroImage: Clinical*, 9:1–12, 2015.

[118] Gerwyn Morris and Michael Berk. The many roads to mitochondrial dysfunction in neuroimmune and neuropsychiatric disorders. *BMC medicine*, 13(1):68, 2015.

[119] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

[120] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[121] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

[122] Seiji Ogawa, Tso-Ming Lee, Asha S Nayak, and Paul Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine*, 14(1):68–78, 1990.

[123] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[124] Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325, 1997.

[125] Andre Palmini. The concept of the epileptogenic zone: a modern look at penfield and jasper's views on the role of interictal spikes. *Epileptic disorders*, 8(2):10–15, 2006.

[126] J Nienke Pannekoek, Ilya M Veer, Marie-José van Tol, Steven JA van der Werff, Liliana R Demenescu, André Aleman, Dick J Veltman, Frans G Zitman, Serge ARB Rombouts, and Nic JA van der Wee. Resting-state functional connectivity abnormalities in limbic and salience networks in social anxiety disorder without comorbidity. *European neuropsychopharmacology*, 23(3):186–195, 2013.

[127] Erik Parens and Josephine Johnston. Facts, values, and attention-deficit hyperactivity disorder (adhd): an update on the controversies. *Child and adolescent psychiatry and mental health*, 3(1):1, 2009.

[128] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[129] Russell A Poldrack. Region of interest analysis for fmri. *Social cognitive and affective neuroscience*, 2(1):67–70, 2007.

[130] Russell A Poldrack and Martha J Farah. Progress and challenges in probing the human brain. *Nature*, 526(7573):371–379, 2015.

[131] Russell A Poldrack, Jeanette A Mumford, and Thomas E Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.

[132] Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.

[133] Jonathan D Power, Mark Plitt, Timothy O Laumann, and Alex Martin. Sources and implications of whole-brain fmri signals in humans. *Neuroimage*, 146:609–625, 2017.

[134] David C Preston. Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics. `https://casemed.case.edu/clerkships/neurology/WebNeurorad/MRIBasics.htm`, 2006. [Online; accessed April-2020].

[135] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[136] Marcus E Raichle and Debra A Gusnard. Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences*, 99(16):10237–10239, 2002.

[137] Barnaly Rashid, Mohammad R Arbabshirani, Eswar Damaraju, Mustafa S Cetin, Robyn Miller, Godfrey D Pearlson, and Vince D Calhoun. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fmri brain connectivity. *Neuroimage*, 134:645–657, 2016.

[138] Atif Riaz, Muhammad Asad, Eduardo Alonso, and Greg Slabaugh. Deepfmri: End-to-end deep learning for functional connectivity and classification of adhd using fmri. *Journal of Neuroscience Methods*, 335:108506, 2020.

[139] Valentin Riedl, Lukas Utz, Gabriel Castrillón, Timo Grimmer, Josef P Rauschecker, Markus Ploner, Karl J Friston, Alexander Drzezga, and Christian Sorg. Metabolic connectivity mapping reveals effective connectivity in the resting human brain. *Proceedings of the National Academy of Sciences*, 113(2):428–433, 2016.

[140] ADHD200 sample. `http://fcon_1000.projects.nitrc.org/indi/adhd200/results.html`.

[141] Saman Sarraf and Ghassem Tofighi. Deep learning-based pipeline to recognize alzheimer's disease using fmri data. In *2016 Future Technologies Conference (FTC)*, pages 816–820. IEEE, 2016.

[142] Pascal Sati, Ilena C George, Colin D Shea, María I Gaitán, and Daniel S Reich. Flair*: a combined mr contrast technique for visualizing white matter lesions and parenchymal veins. *Radiology*, 265(3):926–932, 2012.

[143] Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

[144] V Schöpf, CH Kasess, R Lanzenberger, F Fischmeister, C Windischberger, and E Moser. Fully exploratory network ica (fenica) on resting-state fmri data. *Journal of neuroscience methods*, 192(2):207–213, 2010.

[145] M. Scolari, K. N. Seidl-Rathkopf, and S. Kastner. Functions of the human frontoparietal attention network: Evidence from neuroimaging. *Curr Opin Behav Sci*, 1:32–39, Feb 2015.

[146] W. W. Seeley, V. Menon, A. F. Schatzberg, J. Keller, G. H. Glover, H. Kenna, A. L. Reiss, and M. D. Greicius. Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.*, 27(9):2349–2356, Feb 2007.

[147] William W Seeley, Vinod Menon, Alan F Schatzberg, Jennifer Keller, Gary H Glover, Heather Kenna, Allan L Reiss, and Michael D Greicius. Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, 27(9):2349–2356, 2007.

[148] G. Shafiei, Y. Zeighami, C. A. Clark, J. T. Coull, A. Nagano-Saito, M. Leyton, A. Dagher, and B. Mi?ic. Dopamine Signaling Modulates the Stability and Integration of Intrinsic Brain Networks. *Cereb. Cortex*, 29(1):397–409, 01 2019.

[149] Golia Shafiei, Yashar Zeighami, Crystal A Clark, Jennifer T Coull, Atsuko Nagano-Saito, Marco Leyton, Alain Dagher, and Bratislav Mišić. Dopamine signaling modulates the stability and integration of intrinsic brain networks. *Cerebral Cortex*, 29(1):397–409, 2019.

[150] Lizhen Shao, Donghui Zhang, Haipeng Du, and Dongmei Fu. Deep forest in adhd data classification. *IEEE Access*, 7:137913–137919, 2019.

[151] Yvette I Sheline and Marcus E Raichle. Resting state functional connectivity in preclinical alzheimer's disease. *Biological psychiatry*, 74(5):340–347, 2013.

[152] Joshua S Shimony, Dongyang Zhang, James M Johnston, Michael D Fox, Abhik Roy, and Eric C Leuthardt. Resting-state spontaneous fluctuations in brain activity: a new paradigm for presurgical planning using fmri. *Academic radiology*, 16(5):578–583, 2009.

[153] Stephen M Smith. Overview of fmri analysis. *The British Journal of Radiology*, 77(suppl_2):S167–S175, 2004.

[154] Stephen M Smith, Peter T Fox, Karla L Miller, David C Glahn, P Mickle Fox, Clare E Mackay, Nicola Filippini, Kate E Watkins, Roberto Toro, Angela R Laird, et al. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31):13040–13045, 2009.

[155] Rosa Steimke, Jason S Nomi, Vince D Calhoun, Christine Stelzel, Lena M Paschke, Robert Gaschler, Thomas Goschke, Henrik Walter, and Lucina Q Uddin. Salience network dynamics underlying successful resistance of temptation. *Social cognitive and affective neuroscience*, 12(12):1928–1939, 2017.

[156] James V Stone. *Independent component analysis: a tutorial introduction*. MIT press, 2004.

[157] Stephen C Strother. Evaluating fmri preprocessing pipelines. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):27–41, 2006.

[158] Heung-Il Suk, Chong-Yaw Wee, Seong-Whan Lee, and Dinggang Shen. State-space model with deep learning for functional dynamics estimation in resting-state fmri. *NeuroImage*, 129:292–307, 2016.

[159] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[160] Jean Talairach. Co-planar stereotaxic atlas of the human brain-3-dimensional proportional system. *An approach to cerebral imaging*, 1988.

[161] Jean Talairach, P Tournoux, H Corredor, and T Kvasina. *Atlas d'anatomie stereotaxique du telencephale: etudes anatomo-radiologiques*. Masson, 1957.

[162] Chun Chet Tan and Chikkannan Eswaran. Reconstruction and recognition of face and digit images using autoencoders. *Neural Computing and Applications*, 19(7):1069–1079, 2010.

[163] Alaa Tharwat. Principal component analysis: An overview. *Pattern Recognition*, 3(3):197–240, 2016.

[164] BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125–1165, 2011.

[165] L. Q. Uddin. Salience processing and insular cortical function and dysfunction. *Nat. Rev. Neurosci.*, 16(1):55–61, 01 2015.

[166] M. P. van den Heuvel, C. J. Stam, M. Boersma, and H. E. Hulshoff Pol. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage*, 43(3):528–539, Nov 2008.

[167] Martijn Van Den Heuvel, Rene Mandl, and Hilleke Hulshoff Pol. Normalized cut group clustering of resting-state fmri data. *PloS one*, 3(4), 2008.

[168] Martijn P Van Den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: a review on resting-state fmri functional connectivity. *European neuropsychopharmacology*, 20(8):519–534, 2010.

[169] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[170] Gaël Varoquaux, Sepideh Sadaghiani, Jean Baptiste Poline, and Bertrand Thirion. Canica: Model-based extraction of reproducible group-level ica patterns from fmri time series. *arXiv preprint arXiv:0911.4650*, 2009.

[171] Ilya M Veer, Christian Beckmann, Marie-Jose Van Tol, Luca Ferrarini, Julien Milles, Dick Veltman, André Aleman, Mark A Van Buchem, Nic JA Van

Der Wee, and Serge AR Rombouts. Whole brain resting-state analysis reveals decreased functional connectivity in major depression. *Frontiers in systems neuroscience*, 4:41, 2010.

[172] J. L. Vincent, I. Kahn, A. Z. Snyder, M. E. Raichle, and R. L. Buckner. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.*, 100(6):3328–3342, Dec 2008.

[173] S. Vossel, J. J. Geng, and G. R. Fink. Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *Neuroscientist*, 20(2):150–159, Apr 2014.

[174] Donglin Wang, Honglan Xu, and Qiang Wu. Averaging versus voting: A comparative study of strategies for distributed classification. *Mathematical Foundations of Computing*, page 0, 2019.

[175] Jinhui Wang, Liang Wang, Yufeng Zang, Hong Yang, Hehan Tang, Qiyong Gong, Zhang Chen, Chaozhe Zhu, and Yong He. Parcellation-dependent small-world brain functional networks: A resting-state fmri study. *Human brain mapping*, 30(5):1511–1523, 2009.

[176] Kun Wang, Meng Liang, Liang Wang, Lixia Tian, Xinqing Zhang, Kuncheng Li, and Tianzi Jiang. Altered functional connectivity in early alzheimer's disease: A resting-state fmri study. *Human brain mapping*, 28(10):967–978, 2007.

[177] Wei Wang, Yan Huang, Yizhou Wang, and Liang Wang. Generalized autoencoder: A neural network framework for dimensionality reduction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 490–497, 2014.

[178] Eric W. Weisstein. "Gaussian Function" from *MathWorld*–a wolfram web resource. `https://mathworld.wolfram.com/GaussianFunction.html`. [Online; accessed April-2020].

[179] Catherine Westbrook. *Handbook of MRI technique*. John Wiley & Sons, 2014.

[180] Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2(3):125–141, 2012.

[181] Qi-Zhu Wu, Dong-Ming Li, Wei-Hong Kuang, Ti-Jiang Zhang, Su Lui, Xiao-Qi Huang, Raymond CK Chan, Graham J Kemp, and Qi-Yong Gong. Abnormal regional spontaneous neural activity in treatment-refractory depression revealed by resting-state fmri. *Human brain mapping*, 32(8):1290–1299, 2011.

[182] Qiang Wu, Yiming Ying, and Ding-Xuan Zhou. Multi-kernel regularized classifiers. *Journal of Complexity*, 23(1):108–134, 2007.

[183] Qiang Wu and Ding-Xuan Zhou. Analysis of support vector machine classification. *Journal of Computational Analysis & Applications*, 8(2), 2006.

[184] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[185] Chunbo Xiu and Xuemiao Su. Composite convolutional neural network for noise deduction. *IEEE Access*, 7:117814–117828, 2019.

[186] Chao-Gan Yan, Xin-Di Wang, Xi-Nian Zuo, and Yu-Feng Zang. Dpabi: data processing & analysis for (resting-state) brain imaging. *Neuroinformatics*, 14(3):339–351, 2016.

[187] Lev Yasenko, Yaroslav Klyatchenko, and Oksana Tarasenko-Klyatchenko. Image noise reduction by denoising autoencoder. In *2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, pages 351–355. IEEE, 2020.

[188] R. Yuan, X. Di, P. A. Taylor, S. Gohel, Y. H. Tsai, and B. B. Biswal. Functional topography of the thalamocortical system in human. *Brain Struct Funct*, 221(4):1971–1984, 05 2016.

[189] T. P. Zanto and A. Gazzaley. Fronto-parietal network: flexible hub of cognitive control. *Trends Cogn. Sci. (Regul. Ed.)*, 17(12):602–603, Dec 2013.

[190] Shu Zhang, Xiang Li, Jinglei Lv, Xi Jiang, Lei Guo, and Tianming Liu. Characterizing and differentiating task-based and resting state fmri signals via two-stage sparse representations. *Brain imaging and behavior*, 10(1):21–32, 2016.

[191] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434*, 2018.

[192] Xiqun Zhu, Yong Han, Jing Du, Renzhong Liu, Ketao Jin, and Wei Yi. Microbiota-gut-brain axis and the central nervous system. *Oncotarget*, 8(32):53829, 2017.

[193] Liang Zou, Jiannan Zheng, Chunyan Miao, Martin J Mckeown, and Z Jane Wang. 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE Access*, 5:23626–23636, 2017.