SUPERVISED AND UNSUPERVISED MODELS OF BRAIN NETWORKS FOR BRAIN DECODING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ABDULLAH ALCHIHABI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2018

Approval of the thesis:

SUPERVISED AND UNSUPERVISED MODELS OF BRAIN NETWORKS FOR BRAIN DECODING

submitted by **ABDULLAH ALCHIHABI** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Halit Oğuztüzün _____
Head of Department, **Computer Engineering**

Prof. Dr. Fatoş T. Yarman Vural _____
Supervisor, **Computer Engineering, METU**

Dr. Itır Önal Ertuğrul _____
Co-supervisor, **Robotics Institute, Carnegie Mellon University**

Examining Committee Members:

Assist. Prof. Dr. Hande Alemdar
Computer Engineering, METU _____

Prof. Dr. Fatoş T. Yarman Vural
Computer Engineering, METU _____

Assist. Prof. Dr. Gülşah Tümüklü Özyer
Computer Engineering, Atatürk University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ABDULLAH ALCHIHABI

Signature :

ABSTRACT

SUPERVISED AND UNSUPERVISED MODELS OF BRAIN NETWORKS FOR BRAIN DECODING

Alchihabi, Abdullah
M.S., Department of Computer Engineering
Supervisor : Prof. Dr. Fatoş T. Yarman Vural
Co-Supervisor : Dr. Itır Önal Ertuğrul

September 2018, 89 pages

In this thesis, we propose computational network models for human brain. The models are estimated from fMRI measurements, recorded while subjects perform a set of cognitive tasks. We employ supervised and unsupervised machine learning techniques to represent high level cognitive tasks of human brain by dynamic networks.

In the first part of this thesis, we propose an unsupervised multi-resolution brain network model. First, we decompose the signal into multiple sub-bands using Wavelet transform and estimate a set of local meshes at each sub-band. Then, we use stacked denoising auto-encoders to learn low-dimensional connectivity patterns from constructed mesh networks. Finally, learned connectivity patterns are concatenated across different frequency sub-bands and clustered using a hierarchical clustering method. Results show that our proposed model successfully decodes the cognitive states of Human Connectome Project, yielding high rand and adjusted rand indices.

In the second part of this thesis, we propose a supervised dynamic brain network model to decode the cognitive subtasks of complex problem solving. First, the raw

fMRI images are passed through a preprocessing pipeline that decreases their spatial resolution while increasing their temporal resolution. Then, dynamic functional brain networks are constructed using neural networks. Constructed networks successfully distinguish the phases of complex problem solving. Finally, we analyze the network properties of constructed brain networks to identify potential hubs and clusters of densely connected anatomic regions during planning and execution subtasks. Results show that there are more potential hubs during planning and that clusters are more strongly connected in planning compared to execution.

Keywords: Brain Networks, Brain Decoding, Complex Problem Solving, Human Connectome Project, fMRI

ÖZ

BEYİN DURUMU TANIMA İÇİN GÖZETİMLİ VE GÖZETİMSİZ OLARAK OLUŞTURULAN BEYİN AĞLARI

Alchihabi, Abdullah
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Fatoş T. Yarman Vural
Ortak Tez Yöneticisi : Dr. Itır Önal Ertuğrul

Eylül 2018 , 89 sayfa

Bu tezde insan beyni için işlemsel ağ modelleri önerildi. Modeller, denekler belli bilişsel görevleri gerçekleştirirken kaydedilen fMRI ölçümlerinden elde edildi. Dinamik ağlarla insan beynindeki üst seviye bilişsel görevler modellendi. Bu görevlerden temsil edebilmek için gözetimli ve gözetimsiz makine öğrenimi teknikleri kullanıldı.

Tezin ilk kısmında, gözetimsiz çok-çözünürlüklü beyin ağı modeli önerdik. İlk olarak, dalgacık (wavelet) dönüşümünü kullanarak sinyali çoklu alt-bantlara ayrıştırıp sonra her alt-bant için ayrı ayrı yerel beyin ağı oluşturduk. Daha sonra, oluşturulan beyin ağlarının alt boyutlardaki bağlanım örüntülerini öğrenmek için yığınlanmış gürültü giderici oto-kodlayıcı denilen derin öğrenme mimarisini kullandık. Son olarak, öğrenilen bağlanım örüntüleri farklı alt-bantlar boyunca uç uca eklenip hiyerarşik kümeleme ile gruplandı. Rand indeks ve ayarlanmış rand indeks değerlerini ölçerek, önerilen modelin "Human Connectome Project" verilerinin içerdiği bilişsel durumları başarılı şekilde ayrıştırdığını gösterdik.

Bu tezin ikinci kısmında, karmaşık problem çözmenin alt görevlerini modellemek

ve ayrıştırmak için gözetimli dinamik fonksiyonel beyin ağları oluşturduk. İlk olarak, fMRI görüntülerinin uzamsal çözünürlüğünü düşürüp zamansal çözünürlüğünü yükselten yeni bir ön işleme yöntemi önerdik. Sonra, yapay sinir ağları kullanarak dinamik fonksiyonel beyin ağları oluşturduk. Oluşturulan ağlar, karmaşık problem çözmenin fazlarını başarılı şekilde ayırt etti. Son olarak, oluşturulan beyin örüntülerinin ağ özelliklerini, planlama ve uygulama fazları sırasındaki merkezleri ve birbirine yoğun şekilde bağlı anatomik bölgeleri açığa çıkarmak için analiz ettik. Sonuçlar, planlama fazında, uygulama fazından daha fazla merkezin olduğunu ve kümelerin daha sık bağlanım yaptığını açıkça gösterdi.

Anahtar Kelimeler: Beyin Ağları, Beyin Durumu Tanıma, Karmaşık Problem Çözme, Human Connectome Project, fMRI

To my family

ACKNOWLEDGMENTS

I would like to express my deepest appreciation and gratitude to Prof. Dr. Fatoş T. Yarman Vural for all her help, guidance and supervision. This work would not have been possible without her continuous support and motivation.

I would like to thank my thesis committee members, Asst. Prof. Hande Alemdar, and Asst. Prof. Gülşah Tümüklü Özyer for their valuable feedback.

I would like to thank Prof. Sharlene D. Newman for her help and contribution to this thesis as well as the Imaging Research Facility at Indiana University for sharing the Tower of London (TOL) dataset.

I would like to thank my co-supervisor Dr. Itır Önal Ertuğrul for all her help and contribution to this work.

I would like to thank Ömer Ekmekci, Baran Barış Kıvılcım, Adnan Kılıç, Hazal Moğultay, Arman Afrasiyabi and Arash Rahnama for all their help and contribution to this work.

I acknowledge the support of TÜBİTAK (Scientific and Technological Research Council of Turkey) under grant No: 116E091.

I would like to express my gratitude to my family and friends for their continuous support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AAL	Automated Anatomical Labeling
ANOVA	Analysis of Variance
BOLD	Blood Oxygenation Level Dependent
fMRI	functional Magnetic Resonance Imaging
HCP	Human Connectome Project
MAD	Mesh Arc-weight Descriptors
MVPA	Multi-voxel Pattern Analysis
SDAE	Stacked Denoising Auto-encoders
SVM	Support Vector Machine
TOL	Tower of London

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

The cognitive task of human brain is the task that the brain performs while the subject is exposed to certain stimuli or while performing certain processes, such as problem solving, gambling, memory retrieval and so on. The cognitive state corresponding to a cognitive task is represented by the activation patterns of participating brain anatomic regions along with their interactions and dependencies.

The cognitive state can be measured to a certain extent, by functional magnetic resonance imaging (fMRI) technology. fMRI images partition the human brain into tiny cubes called voxels, where each voxel contains thousands of neurons. fMRI images measure the changes in the oxygenation level in the voxels of the brain as the oxygenation level and the activation of the neurons are tightly coupled.

Cognitive state decoding, also known as brain decoding, is the problem of identifying the cognitive state of human brain under a pre-defined set of stimuli. The aim is to develop computational models that make use of the fMRI recordings measured from human brain in order to distinguish and analyze the cognitive state of each fMRI recording.

The main reason that cognitive state decoding is a challenging problem is firstly due to the very high spatial resolution of the collected fMRI images, hundreds of thousands of voxels per image, which causes a curse of dimensionality as the number of features collected is significantly larger than the number of samples (number of subjects). Secondly, a relatively small number of the collected voxels actively partic-

ipates and is involved with the studied cognitive state. Thirdly, the remaining voxels that do not contribute to the studied task are not simply idle, they might be performing other involuntary tasks, such as regulating breathing and metabolism, collecting sensory information from the environment, which makes the collected data very noisy. Also, the very high inter-subject variability due to individual indifferences makes the problem of brain decoding significantly harder.

Various approaches have been proposed in the literature in order to solve the brain decoding problem. The most common approach is to use the raw voxel intensity values, also known as blood oxygenation level dependent (BOLD) response, as features in the computational model proposed. The proposed models can be categorized as supervised and unsupervised models, where the supervised models make use of the cognitive task labels during the training phase of the computational model. While the unsupervised models do not make use of the cognitive task labels during model development and training [1, 2, 3, 4, 5, 6, 7, 8].

Furthermore, numerous studies proposed models to construct brain networks in order to perform brain decoding rather than using the raw voxel intensity values, due to the findings that connectivity patterns in the brain provide more information about cognitive tasks compared to the isolated behavior of individual voxels groups or anatomic regions [3, 9, 6].

1.2 Proposed Computational Models

In this thesis, we suggest both supervised and unsupervised machine learning techniques to model the cognitive states of human brain as dynamical networks.

First, we propose an unsupervised multi-resolution model to construct brain functional networks that is built on the previous work in [10], in order to decode the cognitive state of human connectome project. Since the fMRI recordings are known to contain information at multiple frequency resolutions, we decompose the collected raw intensity values of anatomic regions into frequency sub-bands using discrete wavelet transform [10]. Then, we construct local mesh networks around each anatomic region at all frequency sub-bands, motivated by the finding that brain networks formed by

the correlation of voxel pairs' in fMRI signals provide more information for brain decoding than the temporal information of individual voxels [7, 8]. Since the data obtained from functional Magnetic Resonance Imaging (fMRI) is high-dimensional and sometimes not suitable for analyzing the cognitive states [11], learning efficient low-dimensional features from high-dimensional complex input spaces is crucial for decoding of cognitive processes. In order to achieve that, we train stacked denoising auto-encoders at all frequency sub-bands using the constructed ensembles of mesh networks in order to learn low-dimensional connectivity patterns (features) that capture the underlying activation patterns of the brain during the studied cognitive tasks. Finally, we concatenate the learned connectivity patterns from different frequency resolutions and cluster them using hierarchical clustering.

In the second part of this thesis, we suggest a supervised learning approach to analyze the complex problem solving task of human brain. For this purpose, we propose a dynamic mesh network representation of the brain in order to study the relative activation patterns of the brain during complex problem solving task. In order to achieve this goal, we first select the most informative voxels using mutual information aiming to reduce the spatial resolution of the fMRI images. Then, we find the closest neighbors of each one of the selected voxels within its brain anatomic region defined by the Automated Anatomical Labeling (AAL) atlas [12] in functional neighborhood using Pearson correlation. Then, we construct dynamic mesh brain networks around each selected voxel (seed voxel) using Levinson-Durbin autoregressive estimator. We concatenate the mesh networks of all the voxels located in the same anatomic region into vectors of ensembles of mesh networks. Finally, we calculate the accumulated cosine distances between the ensemble vectors across time during planning and during execution subtasks in order to measure the relative activation of each anatomic region during both subtasks.

In order to further analyze the activation patterns of anatomic regions during complex problem solving, we propose a supervised learning model to construct dynamic brain networks. At first, we pass the fMRI images through a preprocessing pipeline in order to reduce the spatial resolution of the images and increase their temporal resolution using Anova voxel selection method, cubic spline temporal interpolation and Gaussian colourful noise. Thus, making the fMRI images more suitable for con-

structing brain networks. Then, we construct dynamic mesh networks around each selected brain anatomic region using neural networks as suggested in [13]. The network properties of the constructed brain networks are studied in detail in order to identify potential hubs and densely connected anatomic regions during both planning and execution phases of complex problem solving task. The differences between planning and execution brain networks are examined as well.

1.3 Contributions and Novelties

The novelties of this thesis are as follows:

- A new unsupervised multi-resolution brain networks model to perform brain decoding is proposed. The proposed model makes use of discrete wavelet transform and ridge regression to construct brain networks at multiple frequency resolutions. Then, it employs a deep learning architecture namely stacked denoising auto-encoders in order to learn low-dimensional connectivity patterns (features) of human brain. The learned features obtained from stacked denoising auto-encoders, successfully cluster the cognitive tasks of human connectome project using a hierarchical clustering algorithm.
- A new dynamic mesh network representation of the brain that examines the relative activation patterns of brain anatomic regions during complex problem solving is presented. The model reduces the spatial resolution of the data using voxel selection then constructs local mesh networks around each selected voxel. Then, it organizes the constructed meshes into vectors of ensemble of mesh networks and measures the cosine distance between the vectors across time during both planning and execution subtasks.
- A new preprocessing pipeline to increase the temporal resolution and decrease the spatial resolution of the fMRI images is proposed. It applies Anova voxel selection method, temporal interpolation and colourful Gaussian noise to raw fMRI data. The preprocessing pipeline successfully increases the brain decoding power of raw fMRI images. It also allows us to construct dynamic functional brain networks using the tower of London dataset. The established net-

works successfully decode the cognitive phase of the complex problem solving task.

- The network properties of the constructed brain networks, during complex problem solving are studied in detail. Potential hubs and clusters of densely connected anatomic regions during both planning and execution phases are identified. Furthermore, the differences between the brain networks of planning and execution are examined.

The work presented in this thesis has appeared in the following publications:

- Alchihabi, A., Kivilicim, B. B., Ekmekci, O., Newman, S. D., & Vural, F. T. Y. (2018, May). Decoding cognitive subtasks of complex problem solving using fMRI signals. In 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE.

- Alchihabi, A., Kivilicim, B. B., Newman, S. D., & Vural, F. T. Y. (2018, April). A dynamic network representation of fMRI for modeling and analyzing the problem solving task. In Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on (pp. 114-117). IEEE.

- Rahnama, A., Alchihabi, A., Gupta, V., Antsaklis, P. J., & Vural, F. T. Y. (2017, October). Encoding Multi-Resolution Brain Networks Using Unsupervised Deep Learning. In Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference on (pp. 75-80). IEEE.

1.4 The Outline of the Thesis

Chapter 2 introduces the complex problem solving task and the various studies that used Tower of London (TOL) to study complex problem solving abilities. It, also, presents the cognitive tasks included in the Human Connectome Project (HCP), the technical details concerning both TOL and HCP datasets, their experimental setup and data collection methods. Then, chapter 2 provides an overview of the various

methods proposed in the literature in order to construct brain networks. It also discusses the examined network properties of brain networks in the literature and their neuro-biological interpretations.

Chapter 3 presents our unsupervised multi-resolution hierarchical model to decode the cognitive tasks of human connectome project. The details of the proposed model are provided, then the experimental results for brain decoding are introduced. Visualizations of the constructed brain networks across different subjects conclude this chapter.

Chapter 4 starts with our model for preliminary analysis of the Tower of London (TOL) problem. The results of the preliminary analysis concerning relative activation patterns of anatomic regions are presented. Then, our supervised brain decoding pipeline to construct dynamic functional brain networks is presented in detail. Next, the brain decoding powers of the proposed preprocessing pipeline and model are examined. Finally, the network properties of the constructed brain networks are studied in detail highlighting important anatomic regions to the complex problem solving task as well as the differences between the brain networks of the two phases of complex problem solving.

Chapter 5 provides a brief summary of the entire thesis, it also presents potential directions for future work that move forward with the proposed models in chapters 3 and 4.

CHAPTER 2

COGNITIVE TASKS AND MACHINE LEARNING TOOLS

In this chapter, we explore the basic literature of cognitive tasks including complex problem solving and its findings. Then, we focus on the Tower of London (TOL) game as a representative of complex problem solving tasks. We provide an overview of the various applications and experiments, where TOL game was performed by the subjects, in order to analyze complex problem solving abilities of human brain. Then, we discuss the cognitive tasks included in the Human Connectome Project (HCP) along with the related experimental setup. Next, we provide an overview of the brain networks' construction methods applied to fMRI data. Finally, we investigate the literature on network properties of the constructed brain networks and their neuroscientific meanings.

2.1 Cognitive Tasks of Human Brain

What is a cognitive task? Although there are many definitions, in this thesis when we mention a cognitive task, we mean a task of human brain which is accomplished while the subject is exposed to certain stimuli or performing certain task such as problem solving, gambling, memory retrieval etc. It can be roughly measured by the activation patterns of participating brain anatomic regions along with their interactions and dependencies.

2.1.1 Complex Problem Solving

Complex problem solving process of human brain has the been the focus of numerous studies in the field of neuroscience for over 30 years given the large number of high-level cognitive tasks that fall under its umbrella. Complex problem solving includes: strategy formation, coordination and sequencing of mental functions, and holding information on-line.

Complex problem solving arises from the cooperation among neurons within and across brain anatomic regions. This cooperation takes the form of functional connectivity between the participating anatomic regions that is dynamic and constantly changing across time adapting to both previous and current experiences.

This dynamically changing connectivity is an important trait of intelligence, namely fluid intelligence [14], where being able to switch between solution strategies for a given task contributes to the ability of solving problems. Artificial neural networks that are capable of adjusting their own connections as a response to environment stimuli score higher on fluid intelligence tests [15]. This highlights the importance of the functional connectivity as well as its dynamic nature in providing insights into complex problem solving and its biological underpinnings.

Nevertheless, current research in neuroimaging has not yet studied this dynamic connectivity during problem solving. The main reason behind that is that most studies aim to propose models that will support a given theoretical cognitive hypothesis which may produce misleading findings. On the other hand, using the nueroimaging data solely to investigate the underlying brain networks that support complex problem solving behavior can lead to more accurate results.

A problem-solving model was proposed in [16, 17] where complex problem solving is made up of three sequential cognitive subtasks. The first subtask is the construction of problem representation, where a memory representation of the problem is created that includes all the information available concerning the problem [18]. The second subtask is called elaboration, which includes searching for a suitable sequence of operations to solve the problem. The aforementioned two subtasks constitute what is referred to as planning subtask. The third cognitive subtask is execution where the

solution constructed in the elaboration phase is implemented [18, 19, 20].

However, planning sometimes occurs during the execution phase, which can be then referred to as online planning. In some cases, instead of planning the entire solution sequence, subjects plan only the initial sequence of moves then the planning and execution phases overlap [21, 22]. Several studies have supported this hypothesis with Tower of London in [23] and with Pittsburgh problems in [24]. This can explain the overlap of activation patterns of planning and execution subtasks in problems that requires the completion of multiple subgoals.

2.1.1.1 Tower of London Game

The Tower of London (TOL) game was developed by Tim Shallice in 1982 to examine patients with prefrontal damage [25], since then it has been one of the standard problems in the literature to study complex problem solving capabilities. The Tower of London problem consists of three bins having different capacities with coloured balls placed in the bins, where the aim is to rearrange the balls from their initial state to a predetermined goal state while moving one ball at a time and taking into consideration the limited capacity of each bin.

TOL game has been used to investigate the effect of various clinical disorders on problem solving abilities. It is utilized to identify executive dysfunction in children and adolescents suffering from epilepsy and seizures [26]. It is also used to compare the cognitive activation patterns of people suffering from depressions against those of healthy adults [27]. TOL is employed to examine the cognitive impairment in patient's diagnosed with Parkinson disease [28]. In another study, TOL along with functional Magnetic Resonance Imaging (fMRI) technology was employed to study the differences in the neural basis of planning and executive function between first-episode schizophrenia patients and healthy subjects [29]. Furthermore, it has been used to examine prefrontal processing in Parkinson's, schizophrenia, and autism patients [30, 31, 32].

Besides clinical disorders, TOL has also been employed to study the effect of various parameters on the performance of healthy subjects. The predictive power of working

memory, inhibition, and fluid intelligence on TOL performance is explored [33, 34]. Also, the effect of physical activity and exercise, age and impairment in the executive function due to aging process as well as gender on planning and problem-solving ability and its underlying neural basis have been studied [35, 36, 37, 38]. Furthermore, the performance of chess players against non-chess players is studied in an attempt to gain an insight into problem solving skills and whether practicing problem solving in a given setting would generalize to a better performance in other settings [39].

TOL problem itself has many variations due to its large number of parameters such as goal hierarchy, demand for subgoal generation, start position hierarchy, number of solution paths, and the existence of suboptimal alternatives. Several studies have examined the effect of the aforementioned structural parameters along with numerous other non-structural parameters including procedure instructions, experience, environment and problem-solving strategy on the performance of subjects when solving TOL puzzles, where performance is measured in preplanning time and accuracy [40, 41, 42, 23, 43, 44, 45].

Despite the popularity of TOL studies in the literature and the wide range of cases it is used to investigate, relatively few works have explored its underlying neural networks in detail. In [46], the involvement of the parietal cortex, prefrontal cortex, basal ganglia and anterior cingulate in the problem solving task has been reported. The activation patterns of the dorsolateral and rostrolateral subregions of the prefrontal cortex during planning has been examined [47]. The focus of another study has been the hemispheric deferences in the per-frontal cortex during planning and execution as well as the contribution of the superior parietal region to spatial working memory [48]. In addition, some work has been done to investigate the variance in the neural basis of planning between standard and expert subjects [49]. Expert subjects activated more spatially widespread regions of the left dorsolateral prefrontal cortes while standard subject had higher levels of activation in the anterior cingulate [49]. Given all of this literature, it is clear that a holistic understanding of the anatomical brain regions and their respective roles and interactions during complex problem-solving is both lacking and important.

2.1.1.2 Tower of London Dataset

In this section, we present the details of the TOL experiment and the corresponding dataset, which is provided by Indiana University [46]. At first, the participants in the experiments are introduced, then the experimental setup of the procedure is detailed. Finally, data collection and preprocessing methods are discussed.

Participants and Stimuli

18 college students aged between 19 and 38 participated in the experiment after signing informed, written consent documents approved by the Indiana university Institutional Review Board. The subjects solved a computerized version of TOL, two configurations are presented at the beginning of each puzzle: the initial state and the goal state. The subjects were asked to transform the initial state into the goal state using the minimum number of moves. However, the subjects were not informed of the minimum number of moves needed to solve a given puzzle nor of the existence of multiple solution paths.

Procedure

Each subject underwent a practice session before entering the scanning session aiming to get subjects acquainted with the TOL problem. The subjects were given the following instructions: "You will be asked to solve a series of puzzles. The goal of the puzzle is to make the 'start' or 'current' state match the 'goal' state (They were shown an example). Try to solve the problems in the minimum number of moves by planning ahead. Work as quickly and accurately as possible, but accuracy is more important than speed."

The practice session was made up of two parts, the first part consisted of 24 un-timed single-path 4-move puzzles. The second part was made up of 12 7-move puzzles with a time-limit of 15 seconds. Subjects were allowed to continue planning after the 5 seconds planning only time slot, if they chose to do so.

The scanning session consisted of 4 runs, each run included 18 timed puzzles, with a 5 seconds planning only time slot during which subjects are not allowed to move the balls. However, they were allowed to continue planning after the 5 seconds planning

only time slot, if they chose to do so. Following every puzzle, there was a 12-second resting period where subjects focused on a plus sign in the center of the screen. Each run was also followed by a 28-second fixation period.

Data acquisition & preliminary analysis

The fMRI images where collected using a 3T Siemens TRIO scanner with an 8-channel radio frequency coil located in the Imaging Research Facility at Indiana University. The images were acquired in 18 5 mm thick oblique axial slices using the following set of parameters: TR=1000 ms, TE=25 ms, flip angle=60°, voxel size = 3.125 mm × 3.125 mm × 5 mm with a 1 mm gap.

The statistical parametric mapping toolbox was used to perform the preliminary data analysis that included: image correction for slice acquisition timing, re-sampling to 2×2×2 mm voxels, spatial smoothing using a Gaussian filter of 8 mm at full-width at half maximum [50]. Then, the data was passed through a high-pass filter with 1/128 Hz cutoff frequency to discard low frequency signals. Next, motion correction and normalization to the Montreal Neurological Institute (MNI) EPI template were performed. Further details concerning the procedure and data acquisition can be found in [46].

2.1.2 Tasks of Human Connectome Project

We use the fMRI data from the Human Connectome Project (HCP) of 300 subjects performing different cognitive tasks. Each subject performs 7 distinct cognitive tasks during the experiment procedure namely Emotion Processing, Gambling, Language, Motor, Relational Processing, Social Cognition, and Working Memory (WM) [51].

In the emotion processing task, subjects are presented with two faces at the bottom of the screen and one face at the top of the screen, where the faces express fear or anger. Then, subjects are asked to determine which of the faces at the bottom match the face at the top of the screen. In the gambling task, subjects guess the number written on a card, where they win or lose money based on the correctness of their guess. The language task is made up of two parts. In the first part, subjects listen to a short story then answer questions related to it. As for the second part, subjects are

Table 2.1: Scans per task and duration for each task (min:sec).

	Emotion	Gambling	Language	Motor	Relational	Social	WM
Scans	176	253	316	284	232	274	405
Durations	2:16	3:12	3:57	3:34	2:56	3:27	5:01

presented with a math problem and they are asked to choose the correct answer from two possible choices. In the motor task, the subjects are asked to move one of their body parts from a list of five potential body parts. As for the relational processing task, the subjects are presented with objects and they are asked to determine whether the objects match or differ from each other given a predefined characteristic feature of the presented objects. In social cognition task, subjects determine whether a group of objects on the screen are interacting, moving randomly or subjects are not certain if the objects adhere to the first or second options. As for the working memory task, subjects complete two-back working and zero-back memory tasks using 4 different types of visual stimuli (places, tools, faces, body parts).

Different cognitive tasks have different duration length. However, the duration of each cognitive task is equal across all subjects. Table 2.1 shows the duration of each cognitive task along with the corresponding number of brain scans collected during the cognitive task [51].

Each task j consists of s_j scans of the brain volume representing the brain activation pattern during the task where the total number of scans across all tasks is $S = 1940$. As for the spatial resolution of the brain images collected, the BOLD responses of all the voxels located in each anatomic region defined by the Automated Anatomical Labeling (AAL) [12] are averaged in order to obtain the BOLD response of each anatomic region as shown in equation 2.1.

$$\boldsymbol{r_u}(t) = \frac{1}{tot\_vox} \sum_{\forall i \in u} \boldsymbol{v_i}(t), \tag{2.1}$$

where tot_vox represents the total number of voxels in region u, $\boldsymbol{v_i}(t)$ represents the BOLD response of voxel i at time t, and $\boldsymbol{r_u}(t)$ is the BOLD response of region u at time t. Then, the anatomic regions located in Cerebellum and Vermis regions are

discarded yielding a total of 90 anatomic brain regions. Further details concerning the experimental setup and data collection methods can be found in [51].

2.2 Extracting Brain Networks from fMRI Data

Numerous studies have proposed various computational models in order to build brain networks from fMRI images, both during cognitive tasks or during resting state. These works represent a shift in the literature towards brain decoding algorithms that are based on the connectivity patterns in the brain motivated by the finding that these patterns provide more information about cognitive tasks than the isolated behavior of individual voxel groups or anatomic regions [3, 9, 6, 7, 8].

Some of these studies focus on the pairwise relationships between voxels or brain anatomic regions. For example, Pearson correlation have been used in order to construct undirected functional connectivity graphs at different frequency resolutions in [5]. Also, pairwise correlations and mutual information have been used in order build functional brain networks in various studies aiming to investigate the network differences between patients with Schizophrenia or Alzheimer's disease and healthy subjects [52, 53, 54]. Partial correlation along with constrained linear regression was also used to generate brain networks in [55].

Other studies take advantage of the locality property of the brain by constructing local mesh networks around each brain anatomic region then representing the entire brain network as an ensemble of local meshes. In such studies, the blood oxygenation level dependent (BOLD) response of each brain region is estimated as a linear combination of the responses of its closest neighboring regions. Levinson-Durbin recursion has been applied in several studies in order to estimate the edge weights of each local star mesh, where the nodes are the neighboring regions of the seed anatomic brain region [56]. Ridge regression was also used to estimate edge weights while constructing local mesh networks across windows of time [57, 6].

Another study have established resting-state brain networks as sparse constrained networks using both L1 and L2 regularization to introduce sparsity and control for the cross-subject variability [58]. A two-step model was also used in order to build

functional brain networks, where at first sparse multivariate autoregressive model is employed with penalized regression to estimate the brain networks. Then, false discovery rate (FDR) is used to prune low probability connections to introduce sparsity in the brain network [59].

A recent study [13] has proposed a model to build brain networks using neural networks where the loss function is minimized jointly for all anatomic regions rather than estimating a mesh network for each anatomic region independently from other regions.

2.3 Connectivity Properties of Brain Networks

After constructing the brain networks, the statistical properties of the established networks are studied in order to obtain neuroscientific insights related to the experiment at hand. Several network properties have been investigated in detail in the context of brain networks given their neuroscientific interpretations[60].

The small-world property of the brain networks have been studied extensively in numerous studies, namely measured by the clustering coefficient, minimum path length, global and local efficiency [61, 62]. Also, the node degree distribution of the networks has also been investigated and compared to power law and truncated power law distributions[61]. The modularity of the brain networks, the existence of rich-club hubs as well as betweenness centrality of nodes have been explored [62, 63].

Other studies extended the literature to weighted functional brain networks and defined a null model for weighted undirected functional brain networks [64]. Further work focused on controlling for family-wise error (FWE) that complements false discovery rate (FDR) [65].

Several studies have compared the network properties of functional brain networks across different age groups [66, 63]. Other studies performed similar analysis to compare the properties between healthy individuals and those suffering from diseases related to cognitive impairment (Alzheimer, epilepsy, Schizophrenia) [67].

The following are the most commonly studied network properties of brain networks

categorized based on what they aim to quantify. Brain network $N = (V, W)$ is made up of the set of nodes of the network V where the nodes can be either voxels or brain anatomic regions, M is the total number of nodes in V. W is the set of directed weighted edges between the network nodes, where $w_{i,j}$ is the weight of the edge from node i to node j. Matrix A is defined as the binarized version of matrix W, where $a_{i,j}$ takes value 0 if $(w_{i,j} == 0)$ and takes value 1 otherwise.

2.3.1 Measures of Centrality

Measures of centrality aim to identify brain regions that play a central role in the flow of information in the brain network, or nodes that can be identified as hubs.

2.3.1.1 Node Degree

The degree of a node is the total number of its edges as shown in equation 2.2, where $degree_i$ is the degree of node i, V is the set of all nodes in the graph and $a_{i,j}$ is the binary edge between node i and node j.

$$degree_i = \sum_{j \in V} a_{i,j} \tag{2.2}$$

In the case of a directed graph, we distinguish two different metrics: node in-degree $degree_i^{in}$ and node out-degree $degree_i^{out}$ metrics which are shown in equations 2.3 and 2.4 respectively where $a_{i,j}$ is the directed edge from node i to node j.

$$degree_i^{in} = \sum_{j \in V} a_{j,i} \tag{2.3}$$

$$degree_i^{out} = \sum_{j \in V} a_{i,j} \tag{2.4}$$

Node degree is a measure of centrality of the given nodes, where it aims to quantify the hub brain regions interacting with a large number of anatomic brain regions. Thus, a node with high degree indicates its central role in the network.

2.3.1.2 Node Strength

Node strength is the sum of the weights of edges connected to a given node 2.5, where $w_{i,j}$ is the weight of the edge between node i and node j.

$$strength_i = \sum_{j \in V} w_{i,j} \tag{2.5}$$

Similar to node degree, node strength also distinguishes two metrics in the case of directed graphs namely: node in-strength $strength_i^{in}$ and out-strength $strength_i^{out}$ shown in equations 2.6 and 2.7 respectively, where $w_{i,j}$ is the weight of the edge from node i to node j.

$$strength_i^{in} = \sum_{j \in V} w_{j,i} \tag{2.6}$$

$$strength_i^{out} = \sum_{j \in V} w_{i,j} \tag{2.7}$$

Node strength is a node centrality measure that is similar to node degree, which is used in the case of weighted graphs. Nodes with large strength value are tightly connected to other nodes in the network forming hub nodes.

2.3.1.3 Node Betweenness Centrality

Betweenness centrality of node i is the fraction of the shortest paths in the network that pass through node i as shown in equation 2.8.

$$betweenness_i = \frac{1}{(M-1)(M-2)} \sum_{j,k \in V} \frac{\rho^i_{j,k}}{\rho_{j,k}} \tag{2.8}$$

where $\rho_{j,k}$ is the number of shortest paths between nodes j and k, $\rho^i_{j,k}$ is the number of shortest paths between nodes j and k that pass through node i, where nodes i, j and k are distinct nodes.

However, before measuring the betweenness centrality of a node, the connection weights matrix needs to be converted into connection lengths matrix since betweenness centrality is a distance-based metric. In the weights matrix, larger weights imply higher correlation and shorter distance while it is the opposite for the lengths matrix. A connection length matrix is obtained by inverting the weights of the connection weights matrix. The algorithm suggested in [68] can be applied in order to calculate the node betweenness centrality for each anatomic region.

Nodes with high centrality are expected to participate in many of the shortest paths of the networks. Thus, taking a crucial role in the information flow in the network.

2.3.2 Measures of Segregation

Measures of segregation aim to quantify the existence of subgroups within brain networks where the nodes are densely interconnected. These subgroups are commonly referred to as clusters or modules. The existence of such clusters in functional brain networks is a sign of interdependence among the nodes forming the cluster.

2.3.2.1 Clustering Coefficient

The clustering coefficient of a node i is the fraction of triangles around node i which is calculated as shown in equation 2.9 as proposed in [69]. It is equivalent to the fraction of the neighbors of node i that are also neighbors of each other.

$$C_i = \frac{\chi_i}{[(d^{out}_i + d^{in}_i)(d^{out}_i + d^{in}_i - 1) - 2\sum_{j \in V} a_{i,j}a_{j,i}]} \tag{2.9}$$

where d_i^{in} is the in-degree of node i and d_i^{out} is the out-degree of node i. χ_i is the weighted geometric mean of triangles around node i that is calculated as shown in equation 2.10.

$$\chi_i = \frac{1}{2} \sum_{j,h \in V} (w_{i,j} w_{i,h} w_{j,h})^{1/3} \tag{2.10}$$

2.3.2.2 Transitivity

Transitivity of a node is similar to its clustering coefficient. However, transitivity is normalized over all nodes while cluster coefficient for each node is normalized independently which makes clustering coefficient biased towards nodes with low degree. Transitivity can be expressed as the ratio of triangles to triplets in the network.

It is calculated as shown in equation 2.11 [69]:

$$T = \frac{\chi_i}{\sum_{j \in V} [(d_j^{out} + d_j^{in})(d_j^{out} + d_j^{in} - 1) - 2 \sum_{h \in V} a_{j,h} a_{h,j}]} \tag{2.11}$$

2.3.2.3 Global & Local Efficiency

The global efficiency of a brain network is a measure of its functional integration, where functional integration is a measure of how easy it is for brain anatomic regions to communicate with each other. The global efficiency is the average of the inverse shortest path length between all pairs of nodes in the network. It is closely related to small-world property of a network. Equation 2.12 shows how to calculate the global efficiency of a brain network, where $\varrho_{i,j}^w$ is the weighted shortest path length between the distinct nodes i and j [60].

$$E_{global} = \frac{1}{M} \sum_{i \in V} \frac{\sum_{j \in V} (\varrho_{i,j}^w)^{-1}}{M - 1} \tag{2.12}$$

As for the local efficiency of a network, it is the global efficiency calculated over the neighborhood of a single node. The local efficiency is thus a measure of segregation rather than functional integration as it is closely related to clustering coefficient. While the global efficiency is calculated for the entire network, the local efficiency is calculated for each node in the network [60].

CHAPTER 3

UNSUPERVISED BRAIN DECODING MODEL USING DEEP HIERARCHICAL MULTI-RESOLUTION MESH NETWORKS

In this chapter, we introduce our computational model, named Deep Hierarchical Multi-Resolution Mesh networks (DHMMNs) for unsupervised brain decoding. The proposed architecture is tested in task data-set of Human Connectome Project (HCP), where we extract multi-resolution low-dimensional connectivity patterns using deep learning architectures, each of which corresponds to a cognitive task.

At the first level of the architecture, we decompose the fMRI images into multiple sub-bands using Wavelet decomposition. At the second level, for each sub-band, we estimate a brain network, called mesh network, extracted from short time windows of the fMRI images. At the third level, we feed the adjacency matrices of each mesh network at each time-resolution to an unsupervised deep learning architecture, namely, Stacked Denoising Auto-Encoder (SDAE). The output of SDAE provides a compact low-dimensional connectivity pattern for each time window at each sub-band of the fMRI images. At the final level, we concatenate the learned representations of all sub-bands at each window, to cluster them using a hierarchical algorithm to find the natural groupings across the windows.

We visualize the median values and the precisions of the mesh networks at each component of the cluster mixture. The median mesh networks at cluster centers show between-cluster variances and the precision of each cluster shows the within class variability of mesh networks, across the subjects.

3.1 Deep Hierarchical Multi-Resolution Mesh networks (DHMMNs)

In this section, we introduce our proposed model to perform unsupervised cognitive state decoding. Our model firstly decomposes the original fMRI signal into multiple frequency sub-bands using discrete wavelet transform. Then, it constructs functional brain networks at every sub-band using Ridge regression and Pearson correlation. Next, it employs stacked denoising auto-encoders (SDAE) in order to learn an efficient low-dimensional representation of the established brain networks that captures that activation patters of each one of the cognitive tasks. Finally, the learned low-dimensional connectivity patterns at different frequency sub-bands are concatenated and clustered using hierarchical clustering in order to perform brain decoding.

Figure 3.1 shows the pipeline of our proposed model. The representative time series of each anatomic region is decomposed into a set of multi-resolution signals (orange box). Next, mesh networks at each sub-band are estimated from time windows of 30 samples at each subject (gray box). The stacked denoising auto-encoders are used to learn a set of compact connectivity patterns from the constructed mesh arch-weights (white box) at every frequency sub-band. The connectivity patterns are concatenated (blue box) and fed to a hierarchical clustering algorithm to obtain mixture density of connectivity patterns (right box).

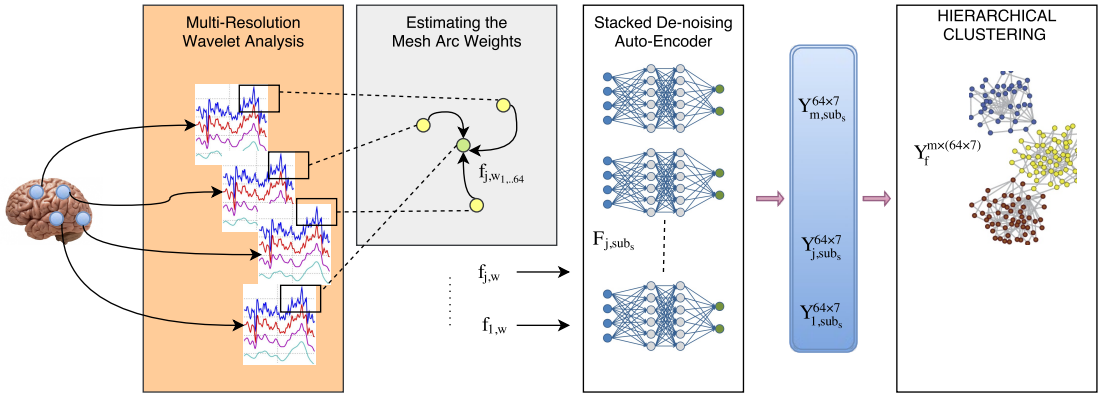


Figure 3.1: Overview of the proposed Deep Hierarchical Multi-Resolution Mesh networks (DHMMNs) model.

3.1.1 Multi-Resolution Signal Decomposition

At first, we decompose the representative time series of BOLD response of each anatomic region into a set of BOLD responses at different frequency resolutions. As it is expected, the BOLD response at each frequency sub-band provide complementary information about the underlying brain activities.

In order to represent the BOLD responses in multiple resolution, the representative time-series $\boldsymbol{r_u}(t)$, for each anatomic region u is decomposed into a set of time series at different resolution. This allows us to estimate and analyze how the anatomical regions process information in different frequency sub-bands [70]. We employ Discrete Wavelet Transform (DWT) as our main tool [71].

We apply the DWT to $\boldsymbol{r_u}(t)$ for all brain regions in order to decompose the signals into l sub-bands where $l = 1, 2, ..., L$, $(L = 11)$. At sub-band level l, we obtain two sets of orthonormal components named as sets of approximation coefficients $\mathcal{A} = \{a_{u,l,k}\}$ and detail coefficients $\mathcal{D} = \{d_{u,l,k}\}$ where k represents the location of the wavelet waveform in discrete-time [71]. These coefficients then may be utilized to reconstruct the fMRI signals at each frequency level, yielding the total of $(2 \times L) + 1$ fMRI time-series. Formally, the representative time-series at sub-band j ($j \in [0, 1, ..., 2L]$) may be defined as shown in equation 3.1.

$$\boldsymbol{r_{j,u}}(t) = \begin{cases} \boldsymbol{r_u}(t), & \text{if } j = 0 \\ \sum_k a_{u,l,k} \boldsymbol{\Phi_{l,k}}(t) \; and \; l = j & \text{if } 1 \leq j \leq L \\ \sum_k d_{u,l,k} \boldsymbol{\Psi_{l,k}}(t) \; and \; l = j - L + 1 & \text{if } j > L \end{cases} \tag{3.1}$$

where $\boldsymbol{\Phi_{l,k}}$ and $\boldsymbol{\Psi_{l,k}}$ are called the mother wavelet and the father wavelet respectively. Further details concerning applying discrete wavelet transform to decompose fMRI signals into multiple frequency sub-bands are available in [10].

3.1.2 Constructing Brain Mesh Networks

After decomposing the BOLD response of each anatomic region into multiple frequency sub-bands, we aim to build ensembles of local mesh networks at each of the frequency sub-bands. Building local mesh networks is motivated by the fact that the structure of the brain is highly interconnected and that neurons influence each other based on the strengths of their synaptic connections [72].

As the first step, we divide the BOLD response into non-overlapping fixed-sized windows. Then for each window, we build local mesh networks around each anatomic region by representing its BOLD response as a linear combination of its P closest neighbors. Next, we group the constructed local mesh networks of all anatomic regions to obtain an ensemble of local mesh networks in each window.

For this purpose, we divide the entire experiment session ($S = 1940$ number of scans) into unlabeled windows of length $Win\_Size = 30$ consisting of 30 discrete scans, thus obtaining a total of 64 windows, with indices $i = 1, ..., 64$ for each subject for the entire experiment. The length of the window, Win_Size, is determined empirically, as the shortest time-interval which provides the highest Rand index, at the output of clustering. It is important to note that the windows are unlabeled and may consist of overlapping data points from different cognitive tasks.

In order obtain a mesh network, we form a set of local meshes by connecting the node of an anatomic region to its p-nearest neighboring anatomic regions. The neighborhood system used is the functional neighborhood, where distance between each pair of anatomic regions is measured using Pearson correlation. As a result, the nearest p neighbors of anatomic region u are the ones with the largest Pearson correlation scores with region u.

For each mesh formed around an anatomic region u, the mesh arc weights for the window i are estimated at the sub-band j using the following regularized linear model 3.2

$$\boldsymbol{r_{j,i,u}} = \sum_{v \in \eta_p[u]} w_{j,i,u,v} \boldsymbol{r_{j,i,v}} + \lambda |w_{j,i,u,v}|^2 + \epsilon_{j,i,u} \tag{3.2}$$

where the regularization parameter is λ. The mesh arc weights $w_{j,i,u,v}$, defined in the $\eta_p[u]$ neighborhood of region u are estimated by minimizing the error $\epsilon_{j,i,u}$. $\boldsymbol{r_{j,i,u}}$ is a vector representing the average voxel time-series in region u at sub-band j for window i, such that:

$$\boldsymbol{r_{j,i,u}} = [r_{j,i,u}(1), r_{j,i,u}(2), ..., r_{j,i,u}(30)]. \tag{3.3}$$

The relation defined in equation 3.2 is solved for each region u with its neighbors, independently by minimizing the mean square error function. In other words, we obtain an independent local mesh around each region u.

After estimating all the mesh arc weights, we organize them under the vector $W_{j,i} = \{w_{j,i,u,v}\}_{u,v}^{M}$, called Mesh Arc Descriptors (MADs), where $M = 90$ is the total number of anatomic regions. Then, we represent the mesh network $N_{j,i}$ as an ensemble of all local meshes.

Note that, the local meshes are estimated for the original fMRI signal, and its approximation and detail parts of different resolutions. Consequently, we form $2L+1$ distinct mesh networks for the frequency sub-bands $\{A_0, A_1, A_2, ..., A_L, D_1, D_2, ..., D_L\}$. The multi-resolution mesh network for a subject is defined by a graph, $N_{j,i} = (V, W_{j,i})$, for each unlabeled window i and for each sub-band j.

The set of vertices V corresponds to the ordered set of anatomic regions and is of size M. Vertex attributes are the time-series $\boldsymbol{r_{j,i,u}}$ contained in the window i, at the sub-band j. The arc weights, $W_{j,i} = \{w_{j,i,u,v}\}_{u,v}^{M}$ between regions u and v, for each window i are obtained from the local meshes of the representative time-series data points at sub-band j.

This process results in $2L + 1$ distinct mesh networks represented by an adjacency matrix of size $M \times M$ made up of $(\forall_{u,v} w_{j,i,u,v})$ for each window i. We concatenate the arc weights under a vector $(f_{j,i})$ of size $1 \times M^2$ and embed the brain network for all windows, $i = 1, ..., 64$, at sub-band j. This means that for each level j and each subject, we represent the entire experiment by an unlabeled matrix of size $64 \times (M^2)$ i.e. $F_{j,sub_s} = [f_{j,1}, ..., f_{j,64}]^T$.

3.1.3 Representation Learning for the Connectivity Patterns by Stacked Denoising Auto-encoders

After constructing ensembles of local mesh networks at each frequency sub-band for all time windows, we aim to use deep learning architectures, namely, stacked denoising auto-encoders, in order to learn a set of compact connectivity patterns from the embedded brain networks. Consequently, clustering the windows w into similar connectivity patterns where each cluster of connectivity pattern corresponds to one of the cognitive tasks of HCP dataset.

The embedded mesh networks model the connectivity among the anatomic regions at different sub-bands of fMRI images under each window i, for each subject. We utilize a deep learning architecture to extract a set of compact connectivity patterns from the mesh networks where the learned connectivity patterns form natural clusters corresponding to the cognitive states of the underlying experiment.

In order to achieve that, we design a multi-layer stacked denoising sparse auto-encoder [73]. For each sub-band j, we train an auto-encoder that takes the windows in the embedded brain network associated with subject sub_s i.e. $f_{j,i} \in F_{j,sub_s}, i = 1, ..., 64$ as its input, and produces a vector y of size 1×7 where the size of output vector corresponds to the number of the cognitive tasks in HCP dataset.

The learned features represent the connectivity patterns at sub-band j for subject, sub_s, during time window i, as shown in equation 3.4:

$$Y_\theta(F_{j,sub_s}) = \Phi(\mathcal{W}F_{j,sub_s} + \mathcal{B}) \tag{3.4}$$

with the auto-encoder parameter set $\Theta = [\mathcal{W}, \mathcal{B}]$, where $\mathcal{W}$ is the collection of weights $\{\mathcal{W}_i\}_{1:4}$, $\mathcal{B}$ is the collection of biases $\{\mathcal{B}_i\}_{1:4}$ at each neuron and Φ represents the activation function $\arctan$. Our sparse auto-encoder design includes an input layer of size $f_{j,i}^T$ with three hidden layers $[500, 64, 21]$ and an output layer of size 7 and the sparsity parameter ρ. The output of each neuron y_i may be represented as $y_i = \sum_{j=1}^{n} w_j x_j + b_i$, where n and x_j's indicate the total number of neurons and the neurons' outputs from the previous layer respectively.

The objective function J is to minimize the mean-squared loss function $L(\mathcal{W}, \mathcal{B}|F_{j,sub_s})$ in the presence of an L_2-Ridge regularization with parameter λ_2 which adds stability and robustness to the learning process as shown in equation 3.5

$$J = \arg\min_{w_i, b_i}\{L(\mathcal{W}, \mathcal{B}|F_{j,sub_s}) + \lambda_2||W||_2^2\} \tag{3.5}$$

The training procedure of the stacked denoising auto-encoders is shown in algorithm 1 where $i = 4$ (3 hidden layer + 1 output layer). The algorithm 1 shows the steps of training the stacked auto-encoder layer by layer in order to minimize the mean-squared loss function. Then, after the training procedure is complete, the constructed mesh networks are fed to the stacked auto-encoder in order to obtain the learned connectivity patterns.

Algorithm 1 Training stacked denoising auto-encoders

Input: Mesh networks for subject sub_s at sub-band j (F_{j,sub_s})

Output: Learned connectivity patterns for subject sub_s at sub-band j (Y_{j,sub_s})

for $layer_i$ of stacked auto-encoder **do**

  Initialize $\mathcal{W}_i$ and $\mathcal{B}_i$ of $layer_i$

  if i == 1 **then**

    $input_i = F_{j,sub_s}$

  end if

  for each epoch **do**

    Feed $input_i$ to $layer_i$

    Obtain $output_i$

    Measure the mean-squared loss function $L(\mathcal{W}_i, \mathcal{B}_i|input_i)$

    Adjust weights $\mathcal{W}_i$ and $\mathcal{B}_i$ to minimize the loss function

  end for

  Feed $input_i$ to trained $layer_i$

  Obtained learned the features $output_i$ of $layer_i$

  $input_{i+1} = output_i$

end for

Feed F_{j,sub_s} to the trained stacked auto-encoder

Return the learned connectivity patterns Y_{j,sub_s}

Furthermore, we use dropout during training procedure where at each learning epoch, 20% of the data points are removed. Using dropout manages to regularize the data and adds a denoising effect that controls for over-fitting [74]. After training the aforementioned auto-encoder, we extract the feature matrices for subject sub_s at sub-band j to attain $(Y_{j,sub_s}^{(64\times7)})_f$.

The proposed deep learning architecture manages to remove the large intra-class variance among input mesh network representation as it gives an effective representation of the brain networks in a low-dimensional space, which can be expressed as a non-linear mapping model from a high-dimensional space to a low-dimensional space suited for clustering.

3.1.4 Multi-Resolution Hierarchical Clustering

Following the construction of low-dimensional representation of mesh networks using auto-encoders, we perform hierarchical clustering on a combination of learned connectivity patterns from different frequency levels in order to distinguish the different cognitive tasks.

The clustering algorithm clusters a subject's brain features matrix $Y_f^{64\times(m\times7)} = [(Y_{1,sub_s}^{(64\times7)})_f, ..., (Y_{m,sub_s}^{(64\times7)})_f]$ consisting of the concatenation of the feature matrices from m different frequency levels selected from the the frequency sub-bands $\{A_0, A_1, A_2, ..., A_{11}, D_1, D_2, ..., D_{11}\}$. The reason to concatenate the feature matrices from different frequency sub-bands is that each frequency level carries complementary information concerning the cognitive tasks performed during the experiment.

Given the 64 discrete-time windows, the clustering algorithm attempts to divide the data points into 7 clusters ($clust_h$, $h = 1, ..., 7$), by minimizing the following cost function 3.6 :

$$\Lambda = \sum_{h=1}^{7} \Lambda_h = \sum_{h=1}^{7} (\sum_{y_j \in clust_h \cap y_j \in Y_f} dis(y_j, clust_h)) \tag{3.6}$$

where the distance matrix $dis(y_j, clust_h)$ is based on the Pearson Correlation between

the learned representation vectors which closely models the functional connectivity pattern in the brain from one task to another. The distance metric used by the hierarchical clustering algorithm is show in equation 3.7

$$dis(y_j, clust_h) = 1 - Corr^2(y_j, clust_h) \tag{3.7}$$

The entries of the correlation matrix $Corr(y_j, clust_h)$ indicate the degree to which the learned connectivity pattern, y_j, for window j is correlated with cluster $clust_h$. The above equation captures the coupling between windows and consequently closely models the flow of change in brain features from one cognitive state to another [75].

We measure the performance of clustering using Rand index and adjusted Rand index. These indices measure the agreement between two partitions; one given by the clustering process and the other defined by the ground-truth labels of time windows. The values of Rand index (RI) lie between 0 and 1, while the Adjusted Rand index (ARI) may take negative values due to its sensitivity to randomness compared with RI [76].

3.2 Experiments and Results

In this section, we introduce the experimental results of our proposed model where we perform within-subject clustering and average the results over 100 subjects. We present the results of brain decoding before and after applying each step of our pipeline in order to show the incremental increase in brain decoding power as the computational model is applied. The Rand index (RI) and Adjusted Rand index (ARI) are used to measure the performance of the clustering process employed. Then, we visualize the constructed brain mesh networks across subjects and cognitive tasks to highlight the inter-task and inter-subject variabilities in the networks.

The design parameters of the proposed model are selected empirically through a cross validation process based on the obtained clustering performance. We search for the optimal design parameters based on the sets, $p \in \{10, 20, 30, 40, 50\}$, $\lambda \in \{16, 32, 128, 256\}$, $\rho \in \{0.01, 0.001, 0.0001\}$, and $\lambda_2 \in \{0.00001, 0.00055, 0.0001\}$.

Table 3.1: Clustering performance of raw fMRI images and mesh arc-weight descriptors.

	RI	ARI
Raw fMRI	0.68	-0.07
Mesh Arc-weight Descriptors	0.84	0.37

We select the following model parameters, $p = 40$ and $\lambda = 32$ for the mesh networks, and $\rho = 0.001$ and $\lambda_2 = 0.00055$ for the SDAE design as optimal values.

At first, we compare the clustering performance of the raw fMRI images against that of the constructed mesh networks at the original frequency sub-band A0. Table 3.1 shows the performance comparison between the clustering of the raw fMRI images (i.e. representative time series of each anatomic region) and the clustering of the mesh arc weights descriptors of constructed brain networks (MADs). It is clear from the table 3.1 that the mesh arc-weight networks increase the performance when compared with that of the raw fMRI images from 68% to 84% and from -0.07% to 0.37% of the Rand index and the Adjusted Rand index respectively.

Then, we compare the performance of mesh networks against the connectivity patterns obtained using the stacked denoising auto-encoders (SDAE) across all frequency sub-bands. Table 3.2 shows the clustering performance of both mesh networks and the connectivity patterns of SDAE measured in Rand index and Adjusted Rand index. The table clearly shows that, for both mesh networks and SDAE connectivity patterns, all frequency sub-bands obtain good clustering performance which highlights the existence of complementary information in different frequency sub-bands.

The table also shows that the clustering performance ranges between 68-86% measured in Rand index across all sub-bands for both the mesh networks and SDAE connectivity patterns. Note that, sub-bands **A5** to **A11**, **D5** to **D7** and **D9** to **D11** show relatively higher performance compared to the other sub-bands when SDAE connectivity patterns are used for clustering, indicating that these frequency bands contain more information than the rest of the sub-bands. Furthermore, the SDAE connectivity patterns obtain performance that is better than mesh networks in some frequency

sub-bands while obtaining slightly worse performance in other sub-bands this can be explained by using the same hyper-parameters for all stacked denoising auto-encoders across all sub-bands rather than fine-tuning the hyper-parameters of each frequency sub-band individually. However, given the significant reduction in dimensionality obtained by the SDAE connectivity patterns against the very high dimensionality of the mesh networks, the performance obtained is significantly superior to that of the mesh networks.

Then, we examine the clustering performance by ensembling multiple frequency sub-bands. Table 3.3 shows the clustering performance of concatenating the mesh networks and SDAE connectivity patterns at multiple frequency sub-bands. The obtained Rand index and adjusted Rand index values clearly show a significant boost in performance compared to using a single frequency sub-band. The Rand index values jumped from a score between 68-86% to 90-93%. This shows that not only the brain networks constructed at multiple time-resolutions provide complementary information for the clustering algorithm but that the proposed deep learning architecture is capable of detecting distinct connectivity patterns in the brain for a given cognitive task, independent of subjects.

The high ARI values in Table 3.3 confirm that utilizing the complementary information gained from different time-resolutions results in clusters with relatively low within-cluster variances and high between-cluster variances.

Furthermore, when we increased the number of subjects to 200, by fusing the brain networks obtained from the entire 23 sub-bands and clustering their connectivity pattern extracted by the SDAE architecture, we are able to achieve similar performance of 93% RI and 71% ARI. This clearly shows that increasing the number of subjects does not affect the clustering performance.

Finally, we aim to visualize the constructed brain mesh networks obtained in the original time resolution of fMRI images in order to observe the inter-task and inter-subject variability of the brain networks. Recall that we performed the clustering within the fMRI data of individual subjects. The reason behind performing within-subject clustering rather than across-subject clustering in this study is the high inter-subject variability, that may prevent the clustering algorithm from finding natural

Table 3.2: Clustering performance of mesh arc-weight descriptors (MAD) and learned connectivity patterns (SDAE). A_i indicates the approximation part and D_i indicates the detail part, while A_0 corresponds to the raw fMRI images.

MAD	RI	ARI	SDAE	RI	ARI
A0	0.84	0.37	A0	0.78	0.11
A1	0.83	0.34	A1	0.76	0.02
D1	0.81	0.28	D1	0.75	-0.04
A2	0.77	0.15	A2	0.74	-0.06
D2	0.86	0.47	D2	0.76	0.11
A3	0.75	0.12	A3	0.74	0.07
D3	0.72	0.15	D3	0.74	-0.34
A4	0.68	0.06	A4	0.77	0.06
D4	0.77	0.24	D4	0.78	0.15
A5	0.68	0.08	A5	0.80	0.17
D5	0.74	0.17	D5	0.80	0.16
A6	0.75	0.18	A6	0.81	0.20
D6	0.75	0.17	D6	0.80	0.20
A7	0.87	0.50	A7	0.80	0.21
D7	0.84	0.37	D7	0.82	0.26
A8	0.85	0.37	A8	0.80	0.16
D8	0.82	0.27	D8	0.79	0.14
A9	0.85	0.39	A9	0.83	0.30
D9	0.82	0.28	D9	0.80	0.12
A10	0.82	0.29	A10	0.86	0.41
D10	0.83	0.30	D10	0.84	0.20
A11	0.79	0.20	A11	0.82	0.25
D11	0.81	0.26	D11	0.83	0.29

Table 3.3: Clustering performance for combinations of sub-bands.

MAD	Rand	A. Rand	SDAE	Rand	A. Rand
All Sub-bands	0.91	0.64	All Sub-bands	0.93	0.71
Sub-bands 7-9	0.92	0.66	Sub-bands 7-9	0.90	0.59
Sub-bands 7-11	0.92	0.66	Sub-bands 7-11	0.91	0.60
Sub-bands 3-8	0.89	0.57	Sub-bands 3-8	0.91	0.64
Sub-bands 3-11	0.90	0.59	Sub-bands 3-11	0.91	0.63

groupings in the data.

In order to illustrate the inter-subject variability, we plot the mesh networks of three different subjects in Figure 3.2 for each cognitive task. Each row in Figure 3.2 corresponds to a distinct subject, and each column corresponds to one of the cognitive tasks of HCP dataset. The selected subjects have the RI of 99%, which indicates that the proposed model has successfully estimated the natural groupings for each one of the three subjects. The brain networks shown in the aforementioned figure represent the medoids of the clusters which corresponds to each one of the seven different tasks.

The mesh networks corresponding to each of the subjects have been pruned to reach 1% sparsity for simplification by eliminating the mesh arc weights with values less then a threshold thus deleting weak connection in the constructed brain networks. A detailed analysis of the mesh networks corresponding to each task for the subjects shows that the mesh networks corresponding to the same task have practically very small similarities across the three subjects. This validates our prior claim on the existence of very high inter-subject variability. In order to further investigate the inter-subject variability, we select a subgroup of subjects with Rand indices higher than 90% from the HCP task dataset of 300 individuals. Then, we define the precision of the mesh networks across the set of subjects as the inverse of variance across the selected subjects. Figure 3.3 shows the pruned precision of the mesh networks of the aforementioned set of subjects with 1% sparsity. The thickness and the colors of the edges are proportional to their corresponding precision values. It is clear from Figure 3.3 that the majority of the edges are thin-blue while very few of them are thick-red. This indicates that the majority of the mesh network connections have very

high variance across subjects.

3.3 Chapter Summary

In this chapter, we proposed a framework called Deep Hierarchical Multi-Resolution Mesh networks (DHMMNs) for constructing a set of brain networks in multiple time-resolutions in order to model and analyze the connectivity patterns between the anatomic regions for different cognitive states. We proposed an unsupervised deep learning architecture that utilized these brain networks in multiple frequency sub-bands in order to learn the natural groupings of connectivity patterns in the human brain for a given cognitive task. We showed that our proposed pipeline is capable of clustering the representative groupings into their corresponding cognitive states. We examined our suggested architecture on a task data-set from Human Connectome Project and achieved a clustering performance of 93% Rand index and 71% Adjusted Rand index for 100 subjects. In addition, we visualized the median networks and the precisions of the mesh networks at each component of the cluster mixture. We showed that the median mesh networks at each cluster have high inter-subject variabilities.

Finally, it is important to note that the nature of HCP dataset, large number of subjects and large number of fMRI recordings for each cognitive task, made it suitable for the type of analysis contained in our proposed model. As our proposed model requires abundant data to train the deep learning architectures, as well as large number of fMRI recordings per task in order to properly apply multi-resolution analysis. However, the proposed model in this chapter may not suit well datasets with small number of subjects or small number of fMRI recordings collected for the cognitive tasks studied. As applying multi-resolution analysis as well as employing deep architectures to such datasets is not appropriate. Therefore, in the following chapter we propose a different computational model more suitable for smaller datasets with smaller number of fMRI recordings per task.

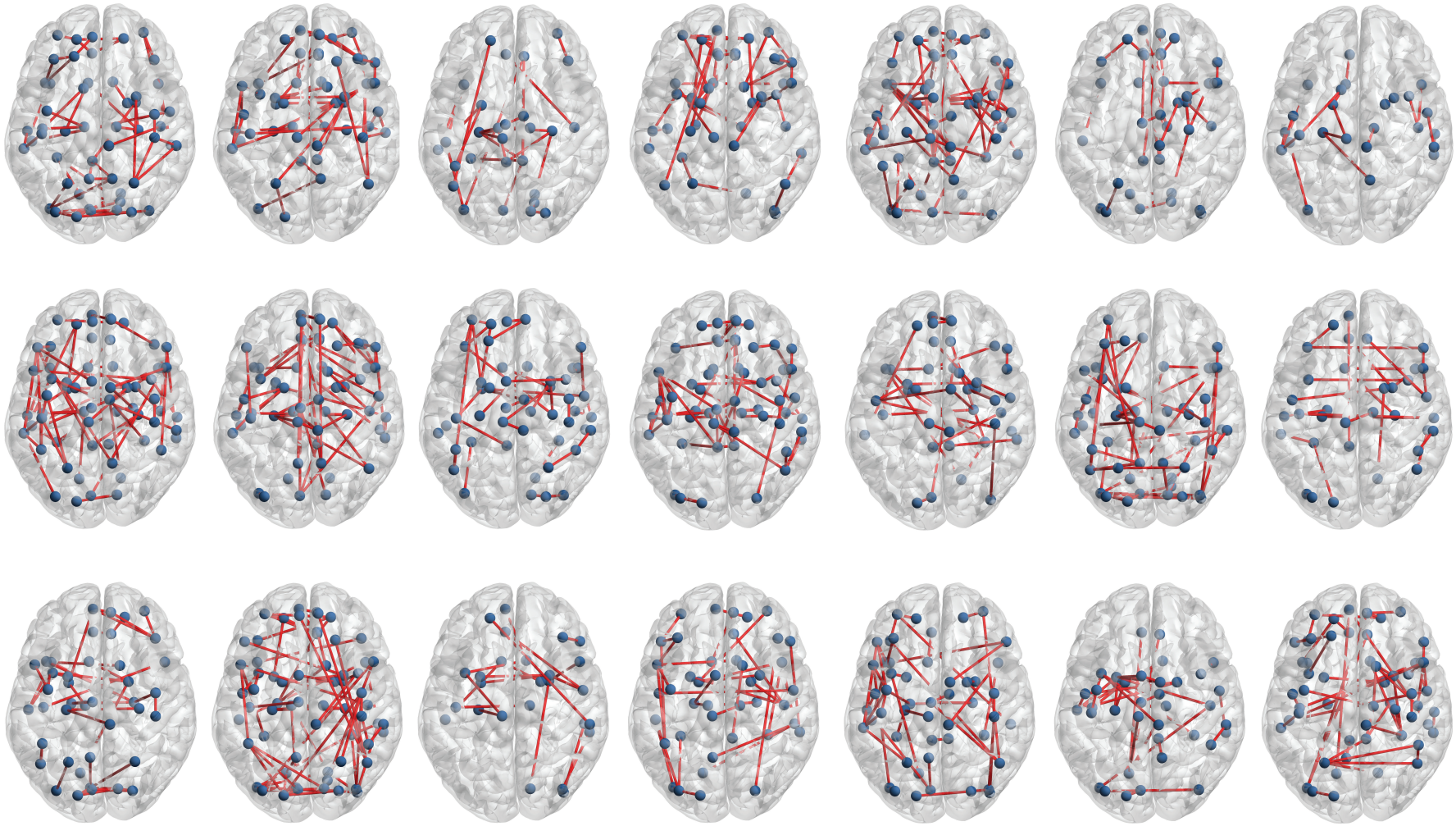


Figure 3.2: Brain mesh networks of 3 subjects.

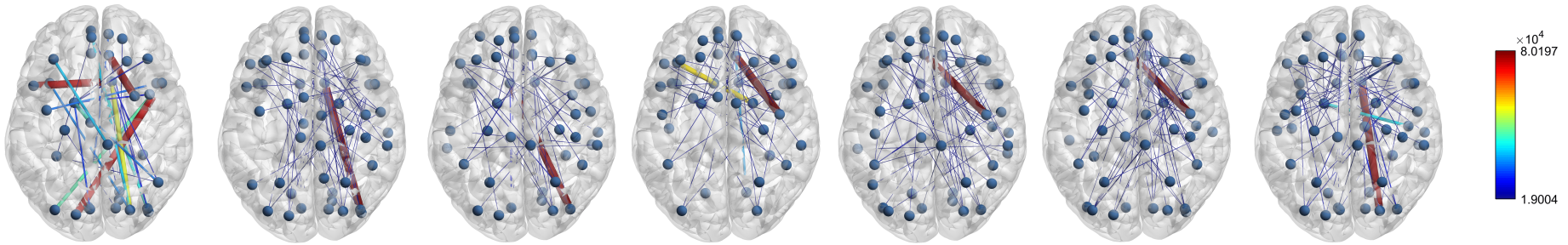


Figure 3.3: Precision of mesh networks across subjects.

CHAPTER 4

SUPERVISED BRAIN DECODING MODEL FOR COMPLEX PROBLEM SOLVING

In this chapter, we introduce our supervised network model for the complex problem solving task using the Tower of London (TOL) dataset. Firstly, we introduce our model for preliminary analysis of TOL that aims to identify the relative activation patterns during both phases of complex problems solving, namely planning and execution. Then, we present our proposed model to decode the cognitive subtasks of the complex problem solving task. The proposed model is made up of two parts, the first part is a preprocessing pipeline that reduces the spatial resolution of the data while increasing its temporal resolution. Then, we construct dynamic functional brain networks using neural networks that represent the underlying cognitive subtasks of problem solving. The results of the proposed preprocessing pipeline and brain decoding model are provided. Finally, the network properties of the constructed brain networks are examined in detail in order to identify potential hubs and clusters of densely connected brain regions during both planning and execution subtasks.

4.1 Preliminary Analysis of Tower of London Problem

In this section, we discuss the preliminary analysis performed on the Tower of London dataset. The aim of the preliminary analysis is to investigate the activation patterns and relations of brain anatomic regions during the planning and execution phases of the problem solving task.

In order to do that, we propose a dynamic sparse network representation estimated from the fMRI brain volumes at all time instances. This representation, called Dy-

namic Mesh Network, enables us to analyze the activation levels of the brain under planning and execution stages of a TOL problem. Before applying our model, we define the ground-truth labels for each puzzle by assuming that all the brain volumes prior to the first move of the participant are categorized as planning whereas the remaining volumes belong to the execution phase.

4.1.1 Dynamic Mesh Network Representation of Brain

The proposed network model is computed in three steps. First, we select the active voxels, from the entire brain volume, that contribute to the problem solving task. Second, we form a dynamic network and estimate its arc weights, at each time instant of fMRI brain volumes, for each of the 18 subjects across all the sessions. Third, we measure the cosine distance between the estimated arc weights of the selected voxels after grouping the voxels into their corresponding anatomic regions. The cosine distance is measured across time for both planning and execution tasks in order to evaluate the activity changes of each anatomic region.

4.1.1.1 Voxel Selection

Given the very large number of voxels (185,405 voxels per time instant), the computational and time complexity of any analysis to be conducted on the entire brain volume is infeasible. Thus, there is a crucial need to reduce the number of voxels by selecting a subset of the voxels containing the most informative ones.

A variety of methods have been applied to select the most informative discriminative voxels [77, 78, 79, 80, 4]. We apply mutual information method, which measures the reduction of the entropy of the class labels vector $\boldsymbol{y_{label}}$ given the BOLD response of vector i ($\boldsymbol{v_i}$) which indicates the amount of information that voxel i contains about class labels $\boldsymbol{y_{label}}$ [77]. Mutual information can be calculated as shown in equation 4.1.

$$I(\boldsymbol{y_{label}}; \boldsymbol{v_i}) = \sum_{\boldsymbol{y_{label}}=1}^{C} \int_{\boldsymbol{v_i}} p(\boldsymbol{v_i}, \boldsymbol{y_{label}}) \log \frac{p(\boldsymbol{v_i}, \boldsymbol{y_{label}})}{p(\boldsymbol{v_i})p(\boldsymbol{y_{label}})} \tag{4.1}$$

where $\boldsymbol{y_{label}}$ is the class label vector, $\boldsymbol{v_i}$ is the BOLD response of voxel i and C is the number of class labels which is equal to 2 in the case of problem solving (planning and execution). The joint distribution $p(\boldsymbol{v_i}, \boldsymbol{y_{label}})$ of $\boldsymbol{v_i}$ and $\boldsymbol{y_{label}}$ is calculated using the chain rule along with Parzen-Rosenblatt window kernel method [77]. At First, we get rid of all the voxels that have a constant value across time as they provide no information concerning whether the task is planning or execution. Then, we measure the mutual information between each one of the remaining voxels and the corresponding class labels. Finally, we retain only the top 20% of the voxels which amounts to over 36 thousand discriminating voxels.

4.1.1.2 Building Dynamic Mesh Networks

After having selected the most informative voxels, we build the mesh networks around each one of the selected voxels, which will be referred to as seed voxels. The motivation behind building mesh networks to represent a voxel rather than its corresponding blood oxygenation level dependent (BOLD) response is due to the finding that the connectivity between the voxels contains more information than the corresponding individual BOLD signals of the voxels [5, 3]. In order to build mesh networks, we first need to define the neighborhood for the voxels. For this purpose, we compute the functional connectivity matrix, where each row corresponds to the Pearson correlation value between the time series of a seed voxel and that of the rest of the voxels located in its anatomic region, which means that the distance between each pair of voxels is measured by the Pearson correlation score between their corresponding BOLD signals. Pearson correlation between two voxels BOLD signals can be calculated as depicted in equation 4.2.

$$cor(\boldsymbol{v_i}, \boldsymbol{v_j}) = \frac{cov(\boldsymbol{v_i}, \boldsymbol{v_j})}{\sigma(\boldsymbol{v_i})\sigma(\boldsymbol{v_j})} \tag{4.2}$$

where $\boldsymbol{v_i}$ represents the time series of BOLD response of voxel i, recorded during the cognitive states. $cov(\boldsymbol{v_i}, \boldsymbol{v_j})$, is the covariance between the corresponding BOLD responses of voxels i and j and σ is the standard deviation function. Thus, the higher the Pearson correlation between two voxels the closer they are to each other in functional neighborhood system. Then, we select the voxels with the highest Pearson correlation values around each seed voxel and connect the seed voxel to the neighboring voxels to form a local star mesh.

After having defined the functional neighborhood of the voxels, we now represent the BOLD response of each seed voxel as a linear combination of the BOLD responses of its (P) closest neighbors located within the same anatomical region as the seed voxel as shown in equation 4.3.

$$\hat{\boldsymbol{v}}_{\boldsymbol{i}}(t) = \sum_{j \in \eta_p[i]} w_{t,j,i} \boldsymbol{v_j}(t) + \epsilon_{t,i} \tag{4.3}$$

In equation 4.3, $\hat{\boldsymbol{v}}_{\boldsymbol{i}}(t)$ represents the estimated BOLD value of voxel i at time instance t. $j \in \eta_P[i]$ indicates voxel j in the P-closest functional neighbors of the seed voxel i. Number of neighbors around each seed voxel defines the size of the local mesh. As P gets higher the mesh size grows and a denser brain network is generated. In this study, P is empirically selected as 5. We refer to $w_{t,j,i}$ as the mesh arc-weight descriptors (MAD) between voxel pair (i, j), estimated at time t [10]. The reason we restrict the neighborhood of each voxel to its anatomical region is due to the high computational complexity of measuring the functional neighborhood at the entire brain level due to the very high number of voxels. As for the parcellation of the brain into anatomical regions, we use Automated Anatomical Labeling (AAL) to divide the brain into 116 different anatomical regions [12]. $\epsilon_{t,i}$ indicates the error of the representation.

In order to estimate the mesh arc weights $w_{t,j,i}$ of a seed voxel i as a linear combination of its P closest neighbors, for all neighboring voxels, $j \in \eta_P[i]$, at each time instant t, we use Levinson-Durbin autoregressive estimator to minimize the expected squared error, $E[(\epsilon_{t,i})^2]$,

$$E[(\epsilon_{t,i})^2] = E[(\hat{\boldsymbol{v}}_{\boldsymbol{i}}(t) - \sum_{j \in \eta_P[i]} w_{t,j,i} \boldsymbol{v}_{\boldsymbol{j}}(t))^2] \tag{4.4}$$

with respect to mesh arc weights $w_{t,j,i}$. Expectation is taken over all the voxels located in the same mesh. Finally, we aggregate all the local meshes obtained at time instant t to form the Dynamic Mesh Network $N(t) = (V, W(t))$, where the set of nodes, $v \in V$ represents the seed voxels together with their neighboring voxels. The set of edges, $w_{t,j,i} \in W(t)$ represents the estimated mesh arc weights. Note that the edges are directed from the neighbors to the seed voxels, and the arc weights dynamically change over time. Note also that, some of the nodes may correspond to unselected voxels if they are in the P-closest functional neighbors set of a seed voxel.

4.1.1.3 Cosine Distance Between Mesh Arc Weights

We measure the percentage change in local meshes during the planning and execution phases for each anatomic region. The aim is to estimate the relative activity of the anatomic regions during planning and execution subtasks.

First, we group the seed voxels into 116 anatomic regions using Automated Anatomical Labeling (AAL) and partition the Dynamic Mesh Network, $N(t) = (V, W(t))$ into 116 subnetworks $N^u(t) = (V^u, W^u(t))$, each of which represents an anatomic region. We group the selected voxels into their corresponding anatomic regions defined by the AAL template in order to investigate the high level functional activation patterns of the anatomic regions. Therefore, the suggested Dynamic Mesh Network can be considered as a collection of subnetworks, i.e. $N(t) = \{N^u(t)\}_{u=1}^{116}$. In subnetwork N^u, the set of nodes V^u consists only of the seed voxels located in region u and their P-closest functional neighbors. The arc weights W^u consists of the mesh arc weights which belong to the seed voxels located in region u.

In each anatomic region u, we concatenate the mesh arc weights of the local meshes obtained at each time point, as shown in equation 4.5, where the vector $\mathtt{w}_u(t)$ represents the vector of mesh arc weights of anatomic region u at time point t, obtained by concatenation of all (i, j) pairs of mesh arc weights, $w_{t,j,i}$ between seed voxels i and

its neighbors j at time point t. As a result of applying equation 4.5, each anatomic region is represented by a variable length vector of mesh arc weights as the length of the vector representation depends on the number of seed voxels located in that region.

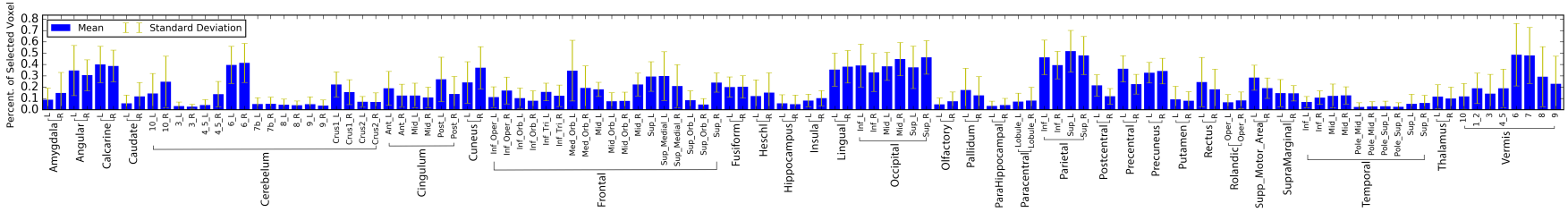
$$\mathtt{w}_u(t) = concat(w_{t,j,i}), \quad \forall i \in u, \forall j \in \eta_P[i] \tag{4.5}$$

Then, we measure the cosine similarity between each pair of consecutive time instances $\mathtt{w}_u(t-1)$ and $\mathtt{w}_u(t)$ to find the relative change of mesh arc weights in anatomic regions during the planning and execution subtasks. The angle θ is the acute angle between the two vectors defined as expressed in equation 4.6, where $y(t)$ is the task performed at time point t which can be either planning or execution. The accumulated cosine distance measured represents the activation level of an anatomic brain region during the TOL experiment. Regions which are active will have varying mesh vectors across time, thus the measured cosine distance will be large. While regions which are relatively inactive will have similar mesh vectors leading to small cosine distance.

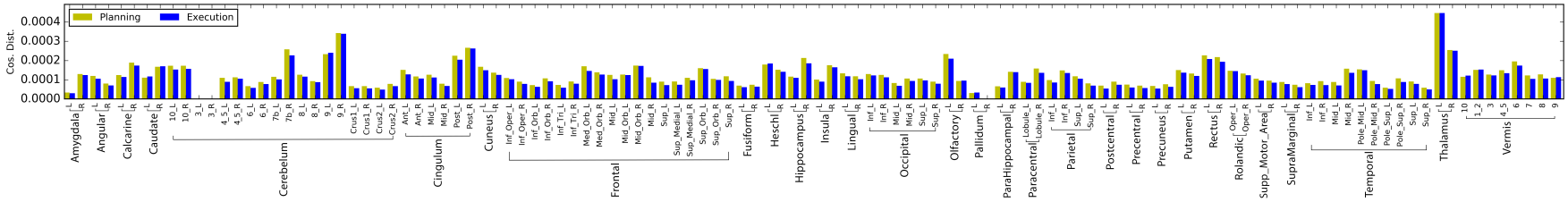
$$\begin{aligned} \theta_u(t) = accute - angle - between(\mathtt{w}_u(t-1), \mathtt{w}_u(t)) \\ where \quad y(t-1) = y(t) \end{aligned} \tag{4.6}$$

4.1.2 Experiments and Results

In this section, firstly we perform voxel selection and keep only the 20% most discriminating voxels, which amounts to more than 36 thousand voxels distributed all over the brain. Increasing the number of voxels more then 20% does not change the structure of the plots. Figure 4.1a shows the mean percentage of voxels selected from each anatomic region across all subjects and the corresponding standard deviation. It can be observed from the figure that superolateral and inferior parietal lobes and occipital lobe contribute between 30 to 50 percent of their voxels as well as Angular and Calcarine regions. While the temporal lobe contributes between 10 to 20%. The remaining regions have relatively low contributions, below 10%, and the cerebellum has varying contributions across different anatomic regions.



(a) Percentage of selected voxels from anatomic regions across subjects.



(b) Accumulated cosine distances of planning and execution tasks for anatomic region across subjects.

Figure 4.1: Distribution of selected voxels (Top) and accumulated cosine distances (Bottom) for anatomic regions.

Secondly, we build local meshes around each seed voxel by selecting its P=5 closest neighboring voxels. Taking higher values of P results in similar behavior of the plots, shown in Figure 4.1b. In this step, we reduce the number of selected voxels to the top 5% of the voxels for computational efficiency. Then, we measure the cosine distance between each consecutive pair of dynamic brain networks across time under both planning and execution tasks for all of the anatomic regions.

Figure 4.1b shows the mean accumulated cosine distances for each anatomic region across all subjects while Figure 4.2 shows the mean accumulated cosine distances for each brain lobe across all subjects. It is observed from both figures that all anatomic regions and brain lobes have higher accumulated cosine distance during the planning task compared to the execution task. This indicates that all regions are more active during the planning phase compared to execution phase.

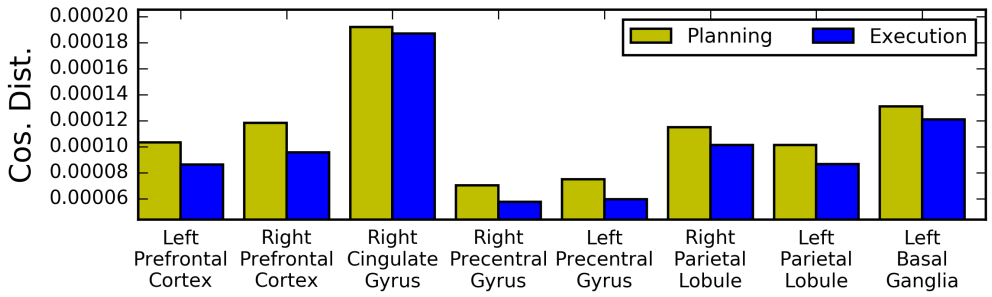


Figure 4.2: Accumulated cosine distances of planning and execution tasks for brain lobes of interest.

4.1.3 Brain Network Visualization of Planning and Execution Phases

In this section, we aim to explore the relationship among the anatomic regions, except for Vermis and Cerebellum regions, during the planning and execution tasks. For this purpose, we build a mesh network between the anatomic regions where the BOLD responses of the selected voxels are averaged into their corresponding anatomic regions. We construct two mesh networks from each puzzle, one that represents the planning phase and another for the execution phase. The constructed mesh networks are built in functional neighborhoods. However, rather than using a scalar represent-

ing the voxel intensity value at time-instance t, we use a vector representing the voxel time series during an entire phase of planning or execution. Therefore, we use ridge regression rather than Levinson-Durbin to approximate the activation of an anatomic region by its closest neighbors. Then, we prune the mesh arc weights by discarding the weakest connections within and across the subjects, where the strength of a connection is measured by its estimated mesh arc-weight value. We plot the resulting brain networks for planning and execution tasks where edge thickness and color indicate the strength of connection between two regions as shown in Figure 4.3 using [81]. The figure indicates that the planning network has several strong connections extending all across frontal lobe and cerebral cortex, while the execution network has a significantly smaller number of strong connections.

4.1.4 Section Summary

In this section, we proposed a **dynamic sparse** network model to represent the activity within and across the anatomic regions of the brain during problem solving task. Several steps introduce the sparsity into our model: selecting a subset of the voxels, building mesh networks of the P closest neighbors only, grouping the voxels into anatomic regions. Furthermore, using ridge regression to build the mesh networks leads to a sparser mesh network due to L2 regularization. We investigated the relative activation levels of the anatomic regions during both planning and execution phases as well as the contribution of each region. The proposed model highlights the brain anatomic regions that contribute to complex problem solving task. It also shows that most anatomic regions are more active during planning subtask than during execution subtask. We, finally, constructed the underlying mesh networks for both planning and execution tasks. The mesh networks indicate stronger connectivity during the planning phase compared to the execution phase.

4.2 Decoding the Phases of Complex Problem Solving

In this section, we introduce our proposed model for cognitive state decoding using brain networks. At first, we perform data preprocessing and data augmentation using

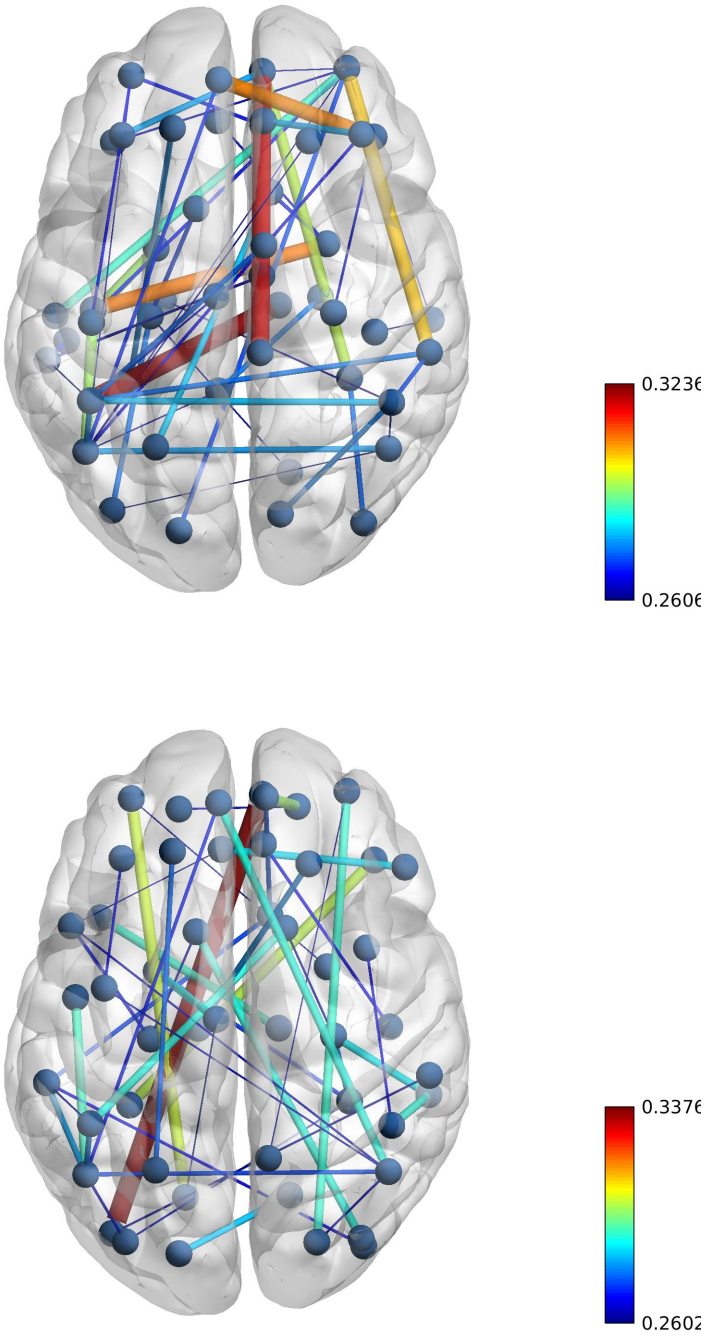


Figure 4.3: Underlying neural network of anatomic regions during planning (Top) and execution (Bottom) tasks.

voxel selection, temporal interpolation and injecting Gaussian colourful noise in order to adapt the data such that it is appropriate to build dynamic functional brain networks. Then, we construct directed weighted dynamic brain networks using neural networks in order to estimate the weights of the edges.

4.2.1 Preprocessing Pipeline

Given the small number of subjects in TOL dataset, and the large number of voxels in each brain image (185,405 voxels per time instant), voxel selection is used to reduce the spatial resolution of the collected brain images and dampen the noise that is inherent to voxel readings. Furthermore, due to the short duration of each puzzle (max 15 seconds) and the relatively low sampling rate (TR = 1 sec), temporal interpolation is needed in order to increase the number of brain volumes in each puzzle. Finally, Gaussian colourful noise is used in order to regularize the data and improve generalization.

4.2.1.1 Voxel Selection

In this first phase of data preprocessing, we use ANOVA feature selection method to choose the most discriminative subset of voxels and discard the remaining ones [77, 79, 80]. In order to do that, we calculate the f-value score of each voxel $\boldsymbol{v_i}$ as shown in equation 4.7:

$$F\_score_i = \frac{MSB(\boldsymbol{v_i}, \boldsymbol{y_{label}})}{MSW(\boldsymbol{v_i}, \boldsymbol{y_{label}})} \tag{4.7}$$

where $\boldsymbol{y_{label}}$ is the label indicating the subtask (Planning or Execution),
$MSB(\boldsymbol{v_i}, \boldsymbol{y_{label}})$ is the mean square value between voxel i and the label vector $\boldsymbol{y_{label}}$. While $MSW(\boldsymbol{v_i}, \boldsymbol{y_{label}})$ is the mean square value within voxel i and the label vector $\boldsymbol{y_{label}}$. Equations 4.8 and 4.9 show how to calculate the mean square value between

and mean square value within respectively.

$$MSB(\boldsymbol{v_i}, \boldsymbol{y_{label}}) = \frac{SSB(\boldsymbol{v_i}, \boldsymbol{y_{label}})}{df_{between}} \tag{4.8}$$

In equation 4.8, $SSB(\boldsymbol{v_i}, \boldsymbol{y_{label}})$ is the sum of squares between $\boldsymbol{y_{label}}$ and $\boldsymbol{v_i}$, $df_{between}$ is the number of groups minus one. As for equation 4.9, $SSW(\boldsymbol{v_i}, \boldsymbol{y_{label}})$ is the sum of squares within group and df_{within} is the degree of freedom within (total number of elements in $\boldsymbol{v_i}$ and $\boldsymbol{y_{label}}$ minus the number of groups).

$$MSW(\boldsymbol{v_i}, \boldsymbol{y_{label}}) = \frac{SSW(\boldsymbol{v_i}, \boldsymbol{y_{label}})}{df_{within}} \tag{4.9}$$

Then, we order the voxels according to their f-value scores. Next, the distribution of f-value scores of all voxels is plotted in order to determine the appropriate number of voxels to retain. Voxel selection is applied to the voxels of all brain regions except the ones located in Cerebellum and Vermis anatomic regions.

Voxel selection successfully manages to significantly reduce the number of voxels in each brain volume thus making the space and time complexity of the analysis on the dataset feasible, given the large total number of voxels in each brain volume, 185,405 voxels per time instant. Then, the BOLD response of the selected voxels is averaged into their corresponding anatomic brain regions defined by the automated anatomical labeling (AAL) atlas [12] as shown in equation 4.10:

$$\boldsymbol{r_u} = \frac{\sum_{i \in \zeta[u]} \boldsymbol{v_i}}{|\zeta[u]|} \tag{4.10}$$

where $\boldsymbol{r_u}$ is the BOLD response of region u, $\boldsymbol{v_i}$ is the BOLD response of voxel i and $\zeta[u]$ is the set of selected voxels located in region u. Averaging the selected voxels into their anatomic regions smooths the noise to a certain degree, and further reduces the dimensionality of each brain volume. As a result, each brain volume would be represented by one BOLD response for every anatomic region thus enabling us to investigate the role and contribution of each anatomic region to both phases of the problem solving task.

4.2.1.2 Temporal Interpolation

The TOL dataset has relatively low temporal resolution given the short duration of each puzzle (maximum of 15 seconds per puzzle) and the low sampling rate (TR = 1000 ms). Therefore, temporal interpolation is used to increase the temporal resolution by generating artificial brain volumes located between the measured brain volumes.

Using temporal interpolation, z artificial estimated brain volumes are inserted between each pair of measured brain volumes, where z depends on the interpolation rate. As a result, the total number of available brain volumes for each puzzle becomes $n + z * (n - 1)$ where n is the number of measured brain volumes of a given puzzle.

We use the cubic spline interpolation function rather than linear interpolation methods in order to prevent edge effects and smoothing out the spikes between the measured brain volumes [82]. Equation 4.11 shows the cubic spline interpolation function for a given piecewise cubic polynomial function between two measured brain volumes:

$$\boldsymbol{C_i}(t) = a_i t^3 + b_i t^2 + c_i t + d_i \tag{4.11}$$

where a_i, b_i, c_i and d_i are the estimated arguments of the spline function $\boldsymbol{C_i}$. Furthermore, we apply Fast Fourier Transform in order to analyze the amplitude spectrum of the signal before and after interpolation. The original single-sided amplitude of the signal and the one obtained after interpolation are compared in order to ensure that interpolation is preserving the smooth peaks of the data in the frequency domain [83, 84].

4.2.1.3 Colorful Gaussian Noise

The final phase of preprocessing is injecting noise to the BOLD response of each anatomic region. For this purpose, instead of just injecting white noise, a rather informed noise, *colorful noise*, is introduced. In order to reflect the corresponding brain region's properties, for each sample the additive noise is generated from a Gaussian

distribution having mean and variance of that anatomical region. This newly generated samples not only act like a natural regularizer to improve generalization but also help making mesh construction algorithms to be more stable while generating brain networks [85, 86].

Given a time series from a particular brain region, let it be anatomic region u, new samples are generated with vector addition of noise while preserving the signal-to-noise ratio (SNR) as in $\tilde{\boldsymbol{r}}_{\boldsymbol{u}} = \boldsymbol{r}_{\boldsymbol{u}} + \boldsymbol{\tau}_{\boldsymbol{u}}$, where $\boldsymbol{\tau}_{\boldsymbol{u}}$ is a noise vector sampled from $\mathcal{N}(\alpha_{noise}\ \mu(r_u),\ \beta_{noise}\ \sigma^2(r_u))$, α_{noise} and β_{noise} being scaling factors.

4.2.2 Building Brain Networks with Neural Networks

After applying the preprocessing pipeline, we aim to construct functional brain networks. In order to do that, we divide the BOLD response of the anatomic brain regions into fixed-size windows where each window $win(t)$ is centered around measured brain volume at time instance t. The size of each window equals $Win\_Size = z + 1$ brain volumes, z is the number of estimated brain volumes in each window. Equation 4.12 shows the time instances included in each window.

$$\boldsymbol{win(t)} = \left[t - \left\lfloor \frac{z}{2} \right\rfloor, .., t, .., t + \left\lceil \frac{z}{2} \right\rceil \right] \tag{4.12}$$

We construct brain networks $N(t) = (V, W(t))$ for each time window $win(t)$ where V is the set of nodes of the graph where each node corresponds to an anatomic brain region, while $W(t) = \{w_{t,v,u} | \forall u, v \in V\}$ represents the directed weighted edges between the nodes of the graph within time window $win(t)$. The nodes of the graph are the anatomic brain regions defined AAL [12], except for the regions located in Cerebellum and Vermis anatomic regions. The nodes are then pruned using voxel selection, as some anatomic regions contribute no voxels at all as a result get deleted from the set of nodes of the graph V.

In this section, we describe how we estimate the weights of the edges $W(t)$ of the brain graphs $N(t)$ where we employ the proposed method [13] in the literature in order to achieve that.

For each window $\boldsymbol{win(t)}$, we define the functional neighborhood matrix Ω_t which is an ensemble of the neighborhood matrices of each anatomic region u estimated using Pearson correlation, which means that the number of rows and number of columns of the matrix is equal to the number of anatomic regions M. The functional neighborhood matrix contains no self-connections $\Omega_t(u,u) = 0 \forall u \in [1,M]$. Also, the anatomic brain regions pruned by voxel selection contributing no voxels have no in-/out connections thus the corresponding entries in Ω_t are all zeros. Finally, for the remaining regions we calculate their corresponding entries in Ω_t using Pearson correlation as follows. At first, for every region u we measure the Pearson correlation between its BOLD response $\boldsymbol{r_{u,t}}$ and the BOLD responses of all the other remaining regions as shown in equation :

$$cor(\boldsymbol{r_{u,t}}, \boldsymbol{r_{v,t}}) = \frac{cov(\boldsymbol{r_{u,t}}, \boldsymbol{r_{v,t}})}{\sigma(\boldsymbol{r_{u,t}})\sigma(\boldsymbol{r_{v,t}})} \tag{4.13}$$

Where $\boldsymbol{r_{u,t}}$ is the BOLD response of anatomic region u across time window $win(t)$, $cov(\boldsymbol{r_{u,t}}, \boldsymbol{r_{v,t}})$ is the covariance between the corresponding BOLD responses of regions u and v, and σ is the standard deviation of the BOLD response of a given anatomic region. Thus, the higher the Pearson correlation between two regions the closer they are to each other in the functional neighborhood system.

Then, we select the P anatomic regions with the highest correlation scores with region u thus obtaining the set $\eta_p[u]$ which contains the P closest anatomic brain regions to region u. Finally, we fill the remaining entries in Ω_t using the constructed neighborhood sets as follows:

$$\Omega_t(u,v) = \begin{cases} 1, & \text{if } v \in \eta_p[u] \\ 0, & \text{otherwise} \end{cases} \tag{4.14}$$

After having determined the edges of the brain graph using the functional neighborhood matrix Ω_t, all that is left is to estimate the weights of these edges. In order to do that, we represent the response of each anatomic region u ($\boldsymbol{r_{u,t}}$) as a linear combination of its closest (P) functional neighbors as shown in equation 4.15,

$$\hat{\boldsymbol{r}}_{\boldsymbol{u,t}} = \sum_{v \in \eta_p[u]} w_{t,v,u} \boldsymbol{r}_{\boldsymbol{v,t}} + \epsilon_{u,t} \tag{4.15}$$

In equation 4.15, $\hat{\boldsymbol{r}}_{\boldsymbol{u,t}}$ represents the estimated BOLD response of anatomic region u within time window $win(t)$, $w_{t,v,u}$ are the estimated arc-weight descriptors. $\eta_p[u]$ is the (P) closest functional neighbors of region u.

Previous work [10] has estimated the mesh arc-weight descriptors for every anatomic region u for each time window $win(t)$ by minimizing the mean-squared error loss function using ridge regression. In there, the mean-squared error loss function is minimized for each region independently from other regions where the expectation is taken over all the time-instances in window $win(t)$ as shown in equation 4.16 .

$$E[(\epsilon_{u,t})^2] = E[(\hat{\boldsymbol{r}}_{\boldsymbol{u,t}} - \sum_{v \in \eta_p[u]} w_{t,v,u} \boldsymbol{r}_{\boldsymbol{v,t}})^2] + \lambda ||w_{t,v,u}||^2 \tag{4.16}$$

λ is the L2 regularization parameter whose value is optimized using cross-validation, L2 regularization is used in order to improve the generalization of constructed mesh networks.

However in this work, we use neural networks to estimate the value of mesh arc-weight descriptors for all anatomic regions jointly in each time window as proposed in [13]. In this method, we estimate the entire weight matrix $W(t) = \{w_{t,v,u} | v, u \in V\}$ using a feed forward neural network. The neural network used is made up of an input layer and an output layer both containing M nodes corresponding to each anatomic brain region. The edges of this network are constructed using the neighborhood matrix Ω_t, if and only if $\Omega_t(u, v) = 1$ then there is an edge between node u of the output layer and node v from the input layer.

Therefore, the loss function is as shown in equation 4.17, where W is the weight matrix of the entire neural network that corresponds to directed weighted edges of the graph and W_u is the entire row of matrix W corresponding to anatomic region u:

$$
\begin{aligned}
Loss(Output_u) &= E[(\epsilon_{u,t})^2] + \lambda W_u^T W_u \\
&= E[(\boldsymbol{r_{u,t}} - \sum_{v \in \eta_p[u]} w_{t,v,u} \boldsymbol{r_{v,t}})^2] + \lambda W_u^T W_u
\end{aligned}
\tag{4.17}
$$

We train the aforementioned neural network in order to obtain the weights of the network that minimize the mean squared error loss function by applying gradient descent optimization method as shown in equation 4.18 :

$$
w_{t,v,u}^{(\kappa)} = w_{t,v,u}^{(\kappa-1)} - \alpha_{learning} \frac{\partial E[(\epsilon_{u,t})^2]}{\partial w_{t,v,u}} \tag{4.18}
$$

where $w_{t,v,u}^{(\kappa)}$ is the weight of the edge from node v to node u at epoch (iteration) κ, $\alpha_{learning}$ is the learning rate. The number of epochs and learning rate used to train the network are optimized empirically using cross-validation.

Finally, when the network weights are obtained, our brain graphs are fully constructed. As a result, we have one brain graph $N(t) = (V, W(t))$ for each time window $win(t)$. Thus, we refer to the brain graphs using their window indices in order to obtain a set of dynamic brain networks $T = \{N(1), N(2), ...N(tot\_win)\}$, where $N(t)$ is the brain network for time window $win(t)$ and tot_win is the total number of time windows.

4.2.3 Section Summary

In this section, we introduced the three phases of our data preprocessing pipeline. At first, we performed feature selection using ANOVA to reduce the spatial dimensionality of each brain image. Then, we performed data augmentation using temporal interpolation in order to increase the sample size of brain images in each puzzle. Next, we regularized and smoothed our brain images by injecting them with Gaussian colorful noise.

Following the preprocessing pipeline, we constructed weighted directed mesh networks around each brain anatomic region centered at each measured brain image,

where we estimated the edge weights of those networks using neural networks, thus obtaining dynamic functional brain networks for each subject.

4.3 Experiments and Results

In this section, we introduce the experimental setup as well as the results obtained for each one of the preprocessing steps. Then, we quantify the improvement in brain decoding power of MVPA methods before and after our preprocessing pipeline. Next, we compare the brain decoding power of our dynamic functional brain networks which are constructed using neural networks against other methods proposed in the literature. Finally, we study the network properties of the established brain networks.

4.3.1 Voxel Selection

At first, we discard all of the voxels located in Vermis and Cerebellum regions. Then, we calculate the f-score for each one of the remaining voxels between its bold response and the y_{label} vector. Next, we order the obtained f-scores then we plot the ordered f-scores of the voxels in order to determine the appropriate number of voxels to retain. Figure 4.4 shows the ordered f-scores of the voxels averaged across all subjects. It is clear from this figure that a relatively small number of voxels is crucial for discriminating the subtasks of problem solving while the remaining voxels have very small contributions. Given the clear elbow distribution of the f-scores of voxels, we keep 10,000 voxels with the highest f-scores and discard the remaining voxels.

After selecting the 10,000 voxels with the highest f-scores in each session, we measure the contribution of each one of the 90 anatomic regions. In order to do that, we calculate the number of voxels selected from each anatomic region out of the selected 10,000 voxels as well as the percentage of voxels selected from each anatomic region relative to its total number of voxels. Figure 4.5a shows the average number of voxels contributed by each anatomic region across all subjects with its corresponding standard deviation, Figure 4.5b shows the average percentage of voxels contributed by each anatomic region across all subjects with its corresponding standard deviation.

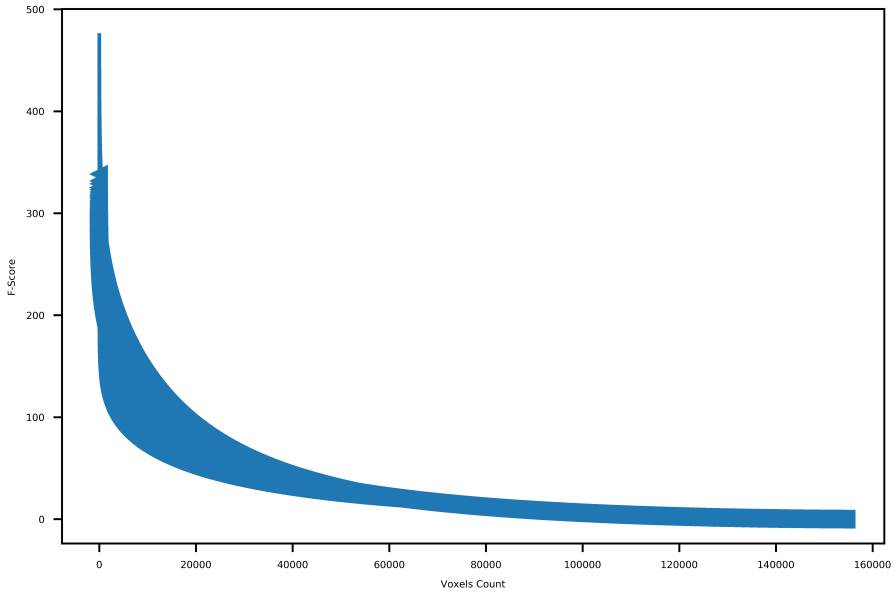


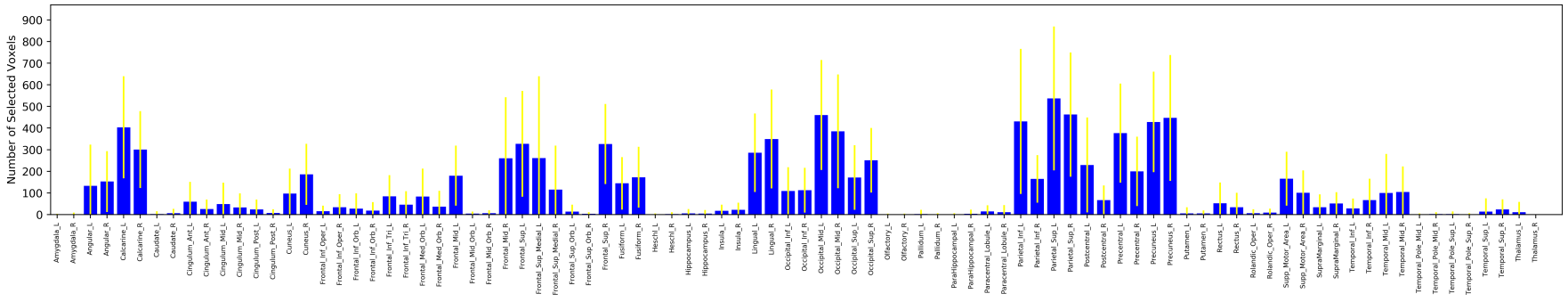
Figure 4.4: Ordered F-score of voxels across subjects.

It is clear from both figures that a large number of regions contribute little to no voxels at all, such as the Amygdala, caudate, Heschl, hippocampus, pallidum, putamen, temporal Pole, temporal Sup, olfactory, Thalamus and parahippocampal regions. While a small number of regions carry discriminative power concerning the phases complex problem solving as they contribute over 300 voxels each, such as the Occipital mid, Precentral, Precuneus and Parietal regions.

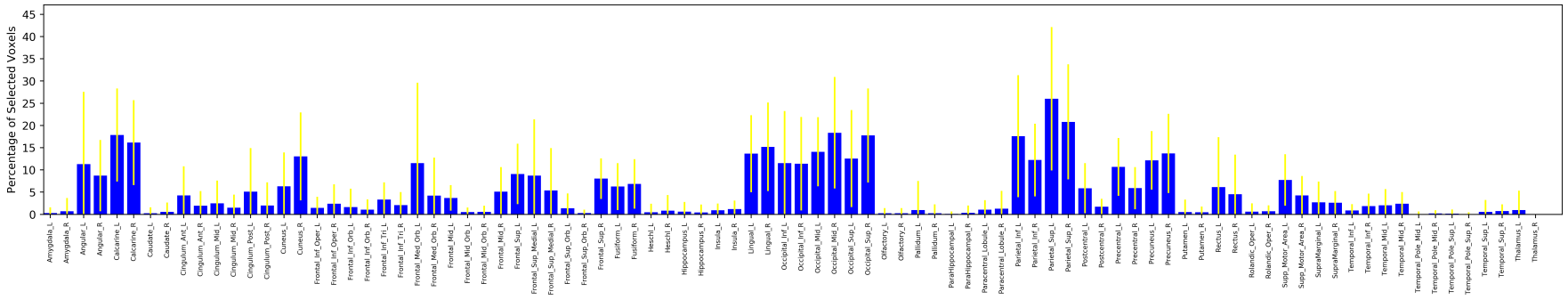
Furthermore, Figure 4.5b ensures that there is no bias against tiny anatomic regions with small number of voxels by normalizing the number of voxels selected from each anatomic region by its total number of voxels. Figure 4.5b clearly shows that in frontal med orb L and occipital inf regions a significant percentage of their voxels are active during complex problem solving task. Both figures also shows high standard deviation across subjects, which indicates high inter-subject variability.

4.3.2 Temporal Interpolation

After selecting the most discriminative voxels and averaging them into their corresponding anatomic regions, we use temporal interpolation to increase temporal resolution of the data.



(a) Average number of voxels selected from each anatomic region across all subjects.



(b) Average percentage of voxels selected from each anatomic region across all subjects.

Figure 4.5: Distribution of selected voxels across anatomic regions, measured by number of selected voxels (top) and percentage of selected voxels (bottom) from each anatomic region.

As a result, the total number of obtained brain volumes is equal to $n+z*(n-1)$ where n is the number of measured brain volumes of a given puzzle and z is the number of estimated brain volumes plugged between each pair of measured brain volumes. The optimal value of z is equal to 8 which is determined empirically using cross-validation. Figure 4.6 shows the interpolated BOLD response of a randomly selected anatomic region from a given subjects, where the blue dots represent the measured BOLD responses of the regions and the orange dashes are the interpolated values. It is clear from the figure that the interpolated points using cubic spline function do not introduce sharp edges nor do they smooth out the spikes between measured brain volumes.

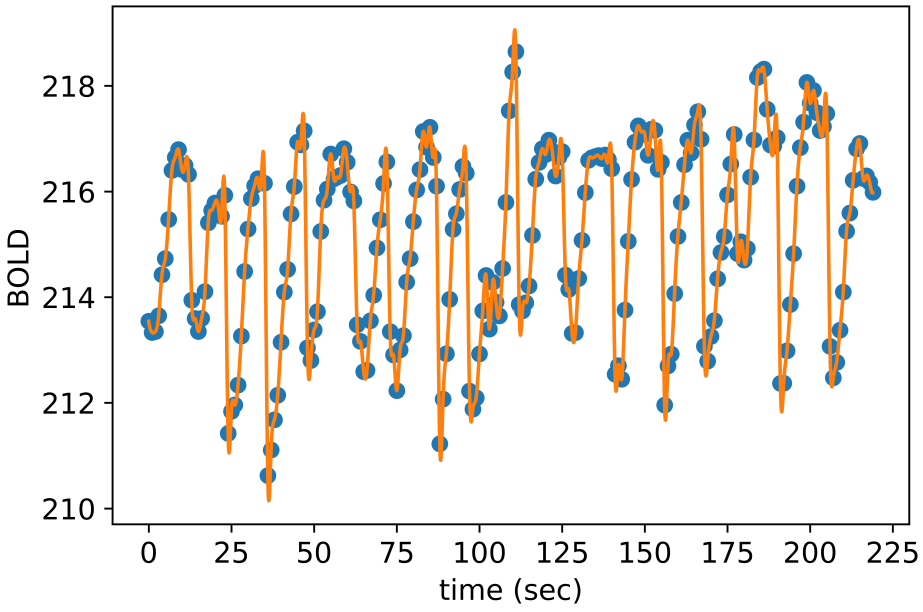


Figure 4.6: Interpolated BOLD response of anatomic region.

Furthermore, Figure 4.7 shows the single-sided amplitude spectrum of a randomly selected anatomic region from a given subject before interpolation, after interpolation and finally after adding Gaussian colourful noise. The figure clearly demonstrates that both interpolation and injecting Gaussian colourful noise preserve the smooth peaks of the signal in the frequency domain.

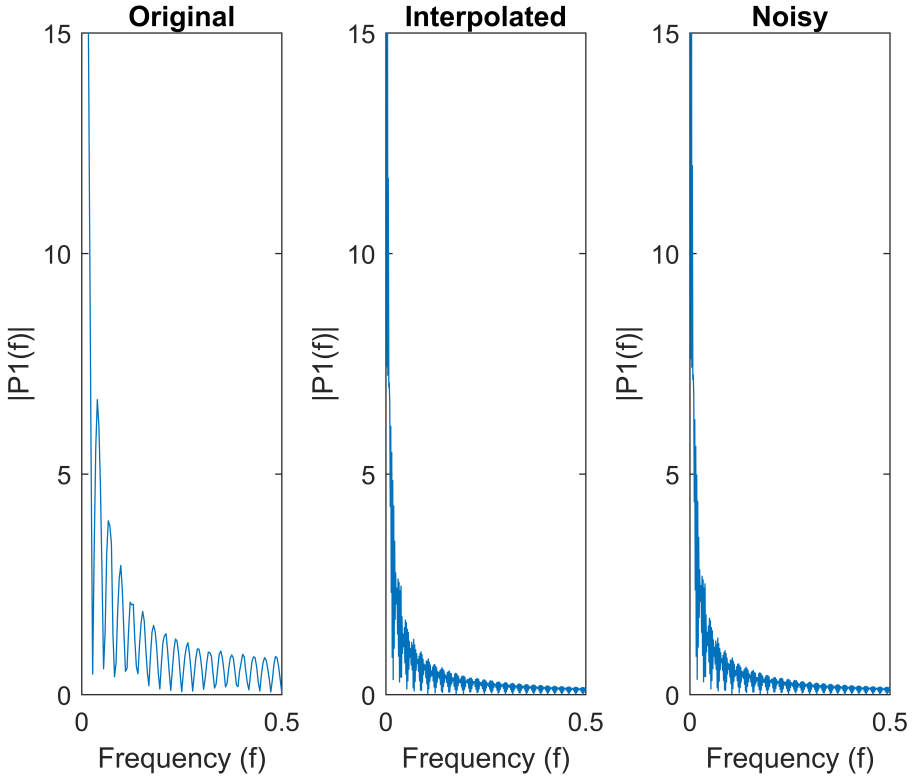


Figure 4.7: Single-Sided Amplitude Spectrum before interpolation (Left), after interpolation (Middle) and after injecting colourful Gaussian noise (Right).

4.3.3 Colorful Gaussian Noise

In order to control the signal-to-noise ration (SNR), we use cross-validation to choose the optimal pair of values for α_{noise} and β_{noise}, the ratios of mean and standard deviation of the added noise respectively. As a result, the optimal values obtained are $\alpha_{noise} = 0.025$ and $\beta_{noise} = 0.075$ from the following set of values $\alpha_{noise}, \beta_{noise} \in [0.025, 0.05, 0.075, 0.1]$.

4.3.4 Brain Decoding with Preprocessing Pipeline

In this section, we use brain decoding in order to quantify the effect of our proposed preprocessing steps on the TOL dataset. We aim to distinguish the two phases of complex problem solving namely: planning and execution.

At first, we use Anova to select the 10,000 voxels that discriminate the planning and execution phases with the highest f-scores then average the selected voxels into their corresponding anatomic regions defined by AAL [12]. Next, we use temporal interpolation to increase the temporal resolution of each puzzle by estimating $z = 8$ brain volumes between each pair of measured brain volumes. Finally, we use Gaussian colourful noise in order to regularize the BOLD responses of each anatomic regions to improve generalization.

K-fold Cross validation for each subject is used in all of the experiments introduced in this section, where $k = 8$. The results obtained are first averaged across different folds, then the average and standard deviation are calculated across all subjects. We use both supervised and unsupervised brain decoding methods, linear support-vector machines (SVM) [87] is used for supervised brain decoding, while K-means clustering is used for unsupervised brain decoding.

Table 4.1 shows the effect our preprocessing pipeline on brain decoding of complex problem solving subtasks. The first row shows the results of brain decoding of the raw dataset without any preprocessing, we simply average all of the voxels into their corresponding anatomic regions. While the second row shows the results of applying voxel selection then averaging the selected voxels into their corresponding anatomic regions. The third row shows the results of brain decoding after applying temporal interpolation, while the forth row shows the results after injecting Gaussian colourful noise.

It is clear from the table that voxel selection improves the brain decoding performance for both supervised and unsupervised methods from %60 to %74 and from %63 to %85 respectively. This can be attributed to voxel selection retaining only the most discriminative voxels and trashing the remaining uninformative ones. Also, voxel selection manages to sparsify the representation of the data since some brain anatomic regions contribute no voxels at all thus have a flat BOLD response.

The table also shows that temporal interpolation further improves the supervised brain decoding performance from %74 to %81, this increase is due to increasing the number of brain volumes thus increasing the number of training samples for the classifier. However, temporal interpolation slightly reduces the performance of unsupervised

methods from %85 to %84 which can be partially attributed to the estimated brain volumes during transitions between the two phases of problem solving, planning and execution reducing the separation between the two natural subgroups. This is due to the way estimated brain volumes are labeled, where an estimated brain volume takes the label of its closest neighboring measured brain volume.

Finally, injecting Gaussian colourful noise slightly boosts the performance of both supervised and unsupervised methods from %81 to %82 and from %84 to %85 respectively. The table also shows high standard deviation across subjects which ,similar to voxel selection plot, highlights the high inter-subject variability.

Table 4.1: Preprocessing pipeline brain decoding results.

Accuracy	SVM	k-Means
Raw data	0.60 ± 0.11	0.63 ± 0.09
Voxel Selection	0.74 ± 0.12	0.85 ± 0.06
Interpolation	0.81 ± 0.08	0.84 ± 0.06
Colorful Noise	0.82 ± 0.8	0.85 ± 0.06

4.3.5 Brain Decoding with Brain Networks

In this section, we compare our model for building dynamic functional brain networks with some of the other methods proposed in the literature in terms of their brain decoding power. Brain decoding can verify that the constructed brain networks are good representatives of the underlying cognitive subtasks. Otherwise, the constructed networks do not capture the connectivity and correlated activation patterns of the anatomic brain regions during the subtasks of complex problem solving.

In order to do that, we build brain networks as explained in the previous sections after having applied the preprocessing pipeline. The optimal values for learning rate $\alpha_{learning}$ and number of epochs are chosen empirically using cross-validation obtaining the following values respectively $1 * 10^{-8}$ and 10. As for p, the number of neighbors we use to represent each anatomic region, we chose p equal to the total number of anatomic regions which is 90, in this way we obtain a fully-connected brain network at each time window. However, the total number of nodes will be less than 90

given that some regions have flat BOLD responses, therefore they are pruned along with all their edges from the brain network.

We also construct brain networks using Pearson correlation and ridge regression as proposed in [5, 10] and [6, 57] respectively in order to compare the performance of our method with other works in the literature. First, we apply our proposed preprocessing pipeline to the raw fMRI recordings before constructing brain networks using Pearson correlation and ridge regression. In the case of Pearson correlation, the functional brain networks are constructed using Pearson correlation scores between each pair of anatomic brain regions [5, 10]. As for the case of ridge regression, the mesh arc-weight descriptors are estimated using ridge regression in order to estimate each anatomic region as a linear combination of its neighbors [6, 57].

Table 4.2 shows the brain decoding results of the aforementioned brain network construction methods compared against the results of multi-voxel pattern analysis (MVPA). The first row shows the brain decoding results of MVPA, while the second and third rows show the results of Pearson correlation and ridge regression methods respectively. The last row shows the brain decoding results of the neural networks model. The table clearly shows that both Pearson correlation and ridge regression fail to construct valid brain networks that are good representatives of the underlying cognitive tasks. However, our model manages to get brain decoding results similar or slightly better than those obtained from MVPA both in the cases of supervised and unsupervised methods. This can be attributed to the challenging nature of TOL dataset, Pearson correlation does not manage to capture the interdependencies between the anatomic regions over short time windows. While ridge regression fails to correctly estimate the mesh arc-weights as it estimates the arc-weights for each anatomic region independently of the other ones. Our proposed model, with a relatively small number of epochs manages to obtain mesh arc-weight values that capture the activation patterns of anatomic regions and their relationships.

4.3.6 Networks Properties of Brain Networks

In this section, we aim to analyze the network properties of the constructed functional brain networks. We investigate the network properties for each anatomic region dur-

Table 4.2: Brain networks brain decoding results.

Accuracy	SVM	k-Means
MVPA	0.82 ± 0.08	0.85 ± 0.06
Pearson	0.58 ± 0.05	0.57 ± 0.04
Ridge Regression	0.56 ± 0.05	0.55 ± 0.02
Neural Networks	0.82 ± 0.10	0.87 ± 0.06

ing both planning and execution subtasks in order to understand which regions are most active, which regions work together during each one of the two subtasks of complex problem solving.

Given that the constructed brain functional networks are both weighted, directed, fully-connected and contain both negative and positive weights, we preprocess the networks before we measure their network properties. At first, we get rid of all the negative weights by shifting all the mesh arc-weights values by a positive quantity equal to the absolute value of the largest negative arc-weight. Then, we normalize the weights of the mesh arc-weights to ensure that all of them are within the range of $[0, 1]$. Finally, we measure the network properties on the pruned brain network, where the the brain anatomic regions (nodes) which contribute no voxels (have a flat BOLD response) and all of their corresponding arc-weights (edges) are deleted from the brain graph. Thus, the brain graphs will contain less than 90 anatomic regions with their corresponding edges. We use brain connectivity toolbox to calculate the studied network properties [60].

As for measures of centrality, since we choose the number of neighbors for each anatomic region (P) to be equal to 89, which is equal to the total number of neighbors for any given node, since the total number of anatomic regions defined by the AAL atlas [12] after deleting the regions residing in Vermis and Cerebellum regions equals 90. Also, since we prune the nodes that correspond to anatomic regions from which no voxels are selected, as a result our constructed brain networks are weighted directed fully-connected networks. Therefore, the in-degree, out-degree and total degree of all nodes in the graph will be equal to the total number of anatomic regions remaining after voxel selection.

Therefore, we use node strength and node betweenness centrality to identify nodes with high centrality which are potential hubs in the brain networks controlling the flow of information in the network. In our proposed model, the node in-strength of node i is the sum of the mesh arc-weight values which are estimated using neural network method in order to minimize the reconstruction error of the BOLD response of anatomic region i using its neighbors. Thus, node in-strength is not used as part of our network properties analyses, we rather use node out-strength to measure the centrality of all anatomic regions.

As for measures of segregation, we aim to quantify the existence of subgroups within brain networks where the nodes are densely interconnected. These subgroups are commonly referred to as clusters or modules. The existence of such clusters in functional brain networks is a sign of interdependence among the nodes forming the cluster. Therefore, we measure clustering coefficient, transitivity and local efficiency to identify potential clusters with dense interconnections in the brain networks.

4.3.6.1 Planning Brain Networks & Execution Brain Networks

In this subsection, we aim to identify the brain anatomic regions with high centrality, as well as clusters of densely connected anatomic regions in both planning and execution networks. In order to do that, for each one of the aforementioned network metrics, we rank the brain anatomic regions in descending order according to their score on that network measure for all subjects across all sessions. Then, we retain the 10 anatomic regions with the highest scores. Next, we measure the frequency of occurrence of each brain anatomic region among the top 10 anatomic regions across all session aiming to identify the shared across subject patterns of activation for both planning and execution.

The following tables show the results of the aforementioned analysis aiming to identify important regions in both planning and execution networks. Table 4.3 shows the brain anatomic regions that have high scores for each one of the reported network properties during planning subtask. While table 4.4 shows the brain anatomic regions that have high scores for each of the reported network properties during execution subtask.

Table 4.3: Planning brain network, anatomic regions showing consistent network property.

Transitivity	**Local efficiency**	**Clustering coefficient**	**Betweenness**	**Out-strength**
Angular	Calcarine	Calcarine	Cueneus R	Cueneus R
Calcarine	Cuneus	Cuneus	Frontal Sup L	Frontal Sup L
Cingulum Ant	Frontal Mid R	Frontal Mid R	Fusiform R	Fusiform R
Cingulum Mid	Frontal Sup	Frontal Sup	Paracentral Lobule L	Paracentral Lobule L
Cuneus	Fusiform	Fusiform	Parietal Sup R	Supp Motor Area R
Frontal Inf Oper L	Occipital Inf R	Occipital Inf R	Precuneus L	Temporal Inf R
	Precentral	Parietal Sup R	Supp Motor Area R	Temporal Mid R
	Supp Motor Area R	Precentral	Temporal Inf R	
	Temporal Inf R	Supp Motor Area R	Temporal Mid R	
		Temporal Inf R		

The first three columns of table 4.3 show that the indicated anatomic regions, which are located in the parietal lobe, occipital lobe, frontal lobe and the temporal lobe, form a cluster of densely interconnected brain regions. The forth and fifth columns show the anatomic regions with high centrality that are potential hubs which are located in the occipital lobe, frontal lobe, parietal lobe and temporal lobe.

The same goes for the corresponding columns in table 4.4. The results presented in both tables are similar to an extent which can be caused by subjects performing online planning during execution phase. However, there are some differences to be noted between the two tables, Frontal Inf Oper R and Precentral L anatomic regions do not participate in the cluster of densely interconnected anatomic regions during execution phase. Also, the Parietal Sup R region does not have high centrality in execution subtask, but Fusiform L and Frontal Sup R are potential hubs in the case of execution subtask.

Figures 4.8a and 4.8b visualize the reported brain anatomic regions in tables 4.3 and 4.4 respectively using Brain Net Viewer [81]. In Figures 4.8a and 4.8b, the colour of the node (anatomic brain region) imply the following: the red colour indicates that the anatomic region has high transitivity, clustering coefficient or local efficiency. The green colour indicates that the node has high node centrality measured by node out-strength and node betweenness. As for the blue colour, it shows the nodes that have high node centrality and are part of subgroup of densely interconnected anatomic

Table 4.4: Execution brain network, anatomic regions showing consistent network property.

Transitivity	**Local efficiency**	**Clustering coefficient**	**Betweenness**	**Out-strength**
Angular	Calcarine L	Calcarine L	Cueneus R	Frontal Sup
Calcarine	Cuneus	Cuneus	Frontal Sup L	Fusiform
Cingulum Ant	Frontal Sup L	Frontal Mid R	Fusiform R	Paracentral Lobule L
Cingulum Mid	Fusiform	Frontal Sup	Paracentral Lobule L	Supp Motor Area R
Cuneus	Occipital Inf R	Fusiform	Precuneus L	Temporal Inf R
Frontal Inf Oper	Supp Motor Area R	Occipital Inf R	Supp Motor Area R	Temporal Mid R
	Temporal Inf R	Parietal Sup R	Temporal Inf R	
		Precentral R	Temporal Mid R	
		Supp Motor Area R		
		Temporal Inf R		

regions.

Results also show that all of the following brain anatomic regions do not play a significant role in discriminating the phases of complex problem solving task: Amygdala, Caudate, Heschl, Hippocampus, Olfactory, Pallidum, ParaHippocampal, Putamen Left, Temporal Pole Mid L, Tempole Pole Sup and Thalamus.

4.3.6.2 Differences Between Planning and Execution Brain Networks

In this subsection, we aim to highlight the differences between planning networks and execution networks. Therefore, for each session we take the difference between the network property scores for brain anatomic regions during planning and the network property scores for brain anatomic regions during execution. Then, we count the frequency of times a given anatomic region is more active during planning than execution and vice-versa in order to identify consistent patterns of the disagreements between planning networks and execution networks across all subjects.

Results obtained show that the following brain regions have higher node out-strength in planning networks than in execution networks: Calcarine, Cuneus, Frontal Sup R, Fusiform, Lingual, Parietal Sup, Precuneus. While the following regions have higher node out-strength in execution networks than in planning networks: Angular

L, Frontal Sup Medial.

As for node betweenness, the following brain regions have higher node betweenness in planning networks than in execution networks: Calcarine, Cuneus, Fusiform, Lingual, Occipital Mid R, Occipital Sup R, Parietal Sup, Postcentral L, Precuneus. While the following regions have higher node betweenness in execution networks than in planning networks: Frontal Sup Medial.

The results for node betweenness and node out-strength show that: Calcarine, Cuneus, Frontal Sup R, Fusiform, Lingual, Occipital Mid R, Occipital Sup R, Parietal Sup, Precuneus and Postcentral L brain anatomic regions play a more central role in the flow of information during planning phase compared to execution phase.

The reported functionality of the aforementioned anatomic regions in the literature are as follows: Calcarine, Cuneus, Occipital Mid and Occipital Sup are involved in visual processing. Frontal Sup region participates in coordination of action with sensory system [88]. Lingual participates in selective visual attention [89], logic-based conditions [90], and memorization [91]. Parietal Sup takes part in spatial orientation. Precuneus is involved with episodic memory retrieval [92], visuospatial processing, directing attention in space (both when an individual is making the movements and when imagining the movements) [93, 94] and shifting attention to different spatial locations [95]. As for the regions that have higher centrality during execution compared to planning. The Frontal Sup Medial is involved in executive functions and Angular controls attentions shift in space [96] and plays a role in memory retrieval [97].

Figures 4.9a and 4.9b visualize the brain anatomic regions with higher node out-strength during planning and during execution respectively.

Figures 4.10a and 4.10b visualize the brain anatomic regions with higher betweenness during planning and during execution respectively.

As for clustering coefficient, the following brain regions have higher clustering coefficient in planning networks than in execution networks: Cuneus, Occipital Mid L, Precuneus R. As for local efficiency, the following brain regions have higher local efficiency in planning than in execution networks: Cuneus, Occipital Mid L, Precuneus R. In the case of transitivity, the following brain regions have higher transitiv-

ity in planning than in execution networks: Angular, Calcarine, Cuneus, Frontal Mid, Frontal Sup, Fusiform, Lingual, Occipital Inf, Occipital Mid, Occipital Sup, Parietal Sup, Parietal Inf, Postcentral, Precentral, Precuneus, Supp Motor Area, Supramarginal R, Temporal Inf R, Temporal Mid R. While no brain anatomic regions have a clear pattern of having higher transitivity, higher local efficiency or higher clustering coefficient during execution phase compared to planning phase.

This demonstrates that the aforementioned brain regions are significantly more interconnected forming densely connected cluster during planning than during execution. Furthermore, even though there is a clear tendency of some brain anatomic regions to form connected clusters during execution phase as reported in the previous subsection, the clusters are significantly more interconnected during planning phase compared to the execution phase.

Figure 4.11a visualizes the brain anatomic regions with higher local efficiency and higher clustering coefficient during planning phase compared to execution phase. While Figure 4.11b visualizes the brain anatomic regions with higher transitivity during planning than during execution phase.

4.3.6.3 Global Efficiency

Since global efficiency is measured over the entire brain network, not for a given node in the network, we measure the global efficiency for all planning and execution brain networks within all sessions across subjects. Then, we compare the global efficiency of planning networks against that of execution networks.

Results show that the majority of the sessions have higher global efficiency scores during planning than during execution, 43 out of 72 sessions have higher global efficiency during planning than execution. Furthermore, table 4.5 shows the number of sessions where global efficiency was higher during planning and during execution subtasks across all subjects for all 4 sessions of each subject. The first column shows the number of subjects that have a higher global efficiency score during planning than during execution. The second column shows the number of subjects that have a higher global efficiency score during execution than during planning. From the table,

it is clear that the majority of the first sessions for all subjects have higher global efficiency during planning subtask compared to execution subtask. While in the case of the following sessions, the number of subjects that have an execution subtask with higher global efficiency than the planning subtask is almost equal to the number of subjects that have a planning subtask with higher global efficiency than the execution subtask.

The table 4.5 indicates that information flows more easily and brain anatomic regions are more interconnected during planning phase of the first sessions compared to execution phase across all subjects. This finding along with the results of the previous subsection which showed that brain anatomic regions are more interconnected and clustered together during planning than during execution, these two findings could explain the results of table 4.5 as follows. During the first session, subjects do not perform online planning during execution phase thus the brain anatomic regions during the planning phase is more interconnected compared to the execution phase. However, during the remaining sessions, subjects perform online planning relatively more during execution phase compared to the first session, thus global efficiency of those sessions during execution phase increases to be roughly similar to that during planning phase.

Table 4.5: Global efficiency of planning and execution brain networks.

Session Number	Planning	Execution
session 1	15	3
session 2	9	9
session 3	10	8
session 4	9	9

4.3.7 Section Summary

In this section, we discussed the experimental setup and the results obtained of the proposed supervised brain decoding model. The results of each step of the preprocessing pipeline are introduced along with their effect on the brain decoding power of the fMRI images. Then, the brain decoding power of the constructed functional

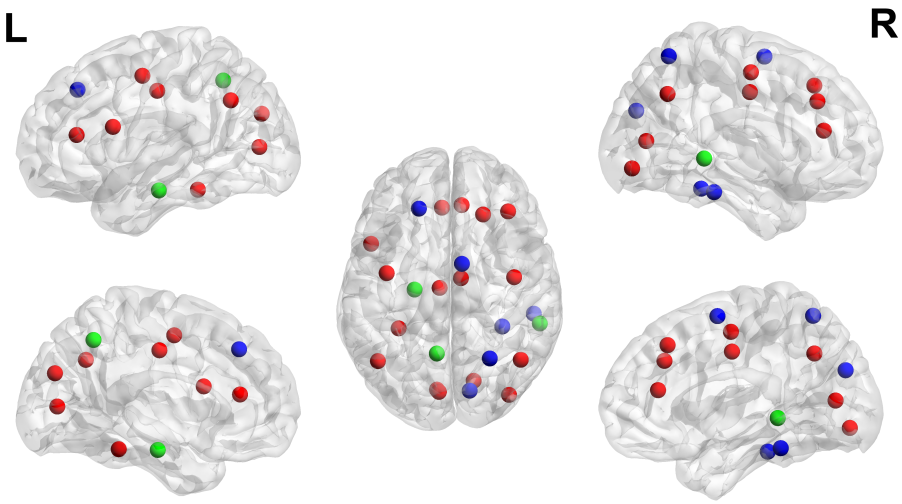
brain networks is compared against other methods in the literature. Next, the network properties of the constructed brain networks are studied identifying potential hubs and clusters of densely interconnected brain regions during both phases complex problem solving. Finally, the differences between the planning networks and execution networks are highlighted.

4.4 Chapter Summary

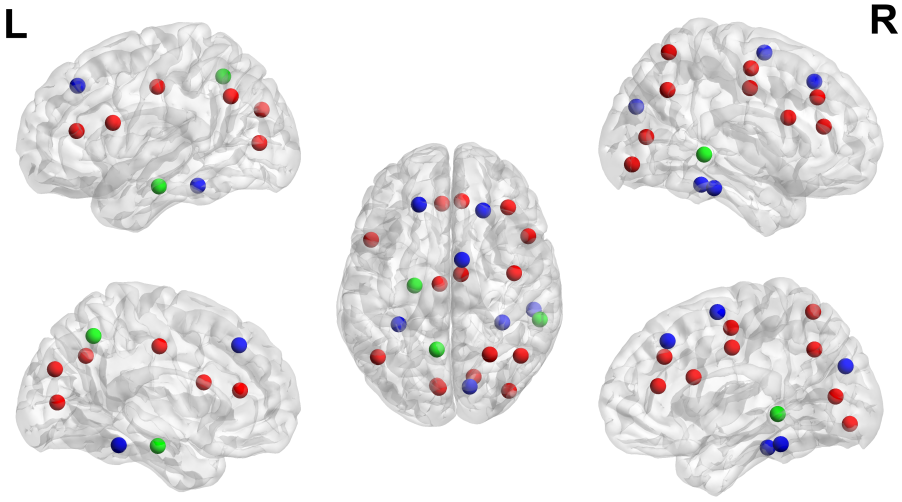
In this chapter, we have performed preliminary analysis on the TOL dataset aiming to identify the anatomic regions that contribute to the complex problem solving task. To do that, we have proposed a Dynamic Mesh Network representation of the brain that selects the most discriminative voxels. Then, it constructs dynamic mesh networks around each selected voxel. Finally, it measures the relative activation of each anatomic region during planning and execution phases using the accumulated cosine distance. Obtained results show that most anatomic regions are more active during planning phase compared to the execution phase.

Then, we proposed a supervised learning model to decode the cognitive subtasks of problem solving by constructing functional brain networks. At first, we passed the raw fMRI recordings through a preprocessing pipeline to reduce their spatial resolution and increase their temporal resolution. Next, we constructed dynamic functional brain networks using neural networks. The constructed brain networks successfully decode the phases complex problem solving task.

Finally, we studied the network properties of the established brain networks and identified potential hubs and clusters of connected anatomic regions for both planning and execution subtasks. The results of the conducted analysis showed an overlap between the planning brain networks and execution brain networks. In addition, it showed that there are more potential hubs during the planning phase as well as the clusters of densely interconnected anatomic regions are more strongly connected during the planning phase compared to the execution phase.

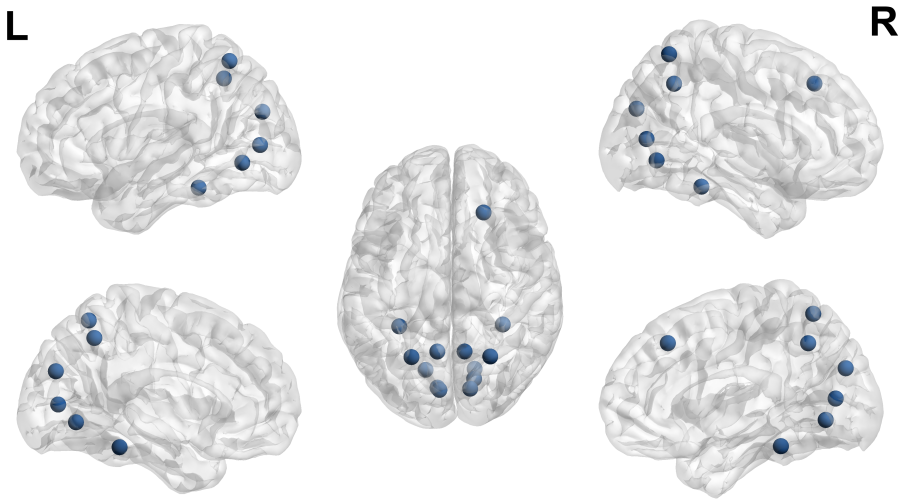


(a) Planning brain network, active anatomic regions.

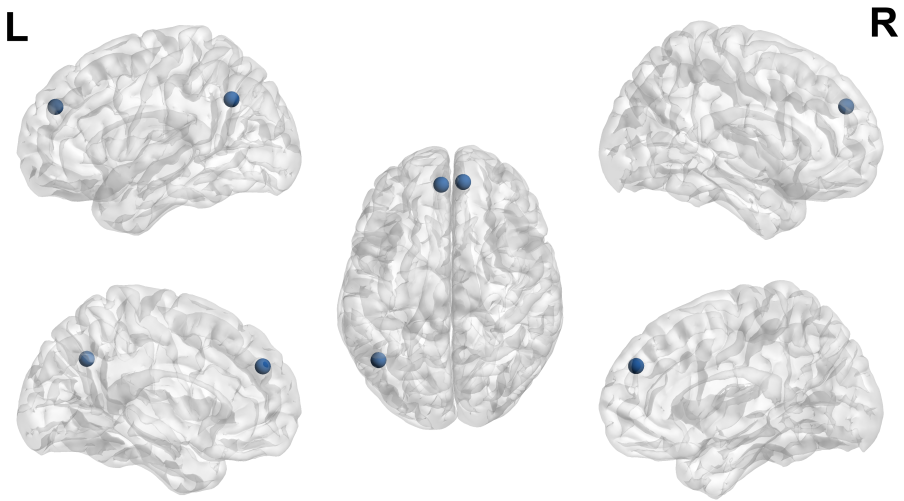


(b) Execution brain network, active anatomic regions.

Figure 4.8: Active anatomic regions of planning (Top) and execution (Bottom).

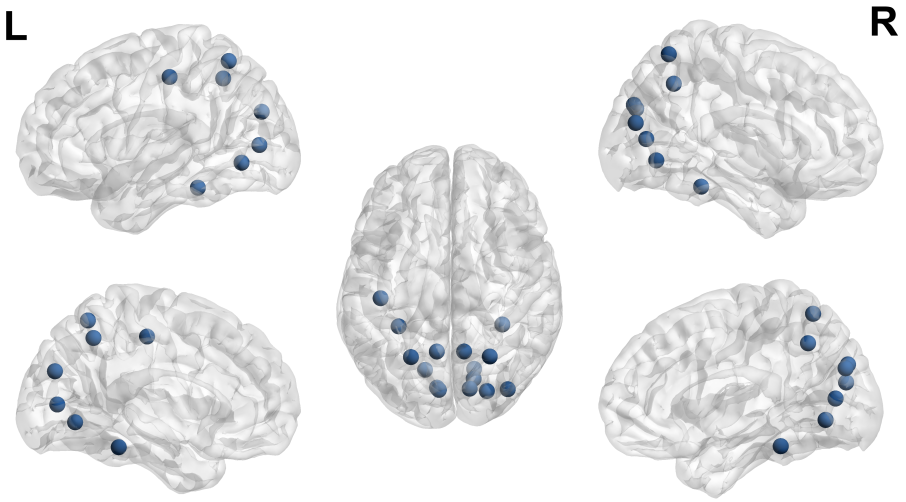


(a) Anatomic regions with higher node out-strength during planning.

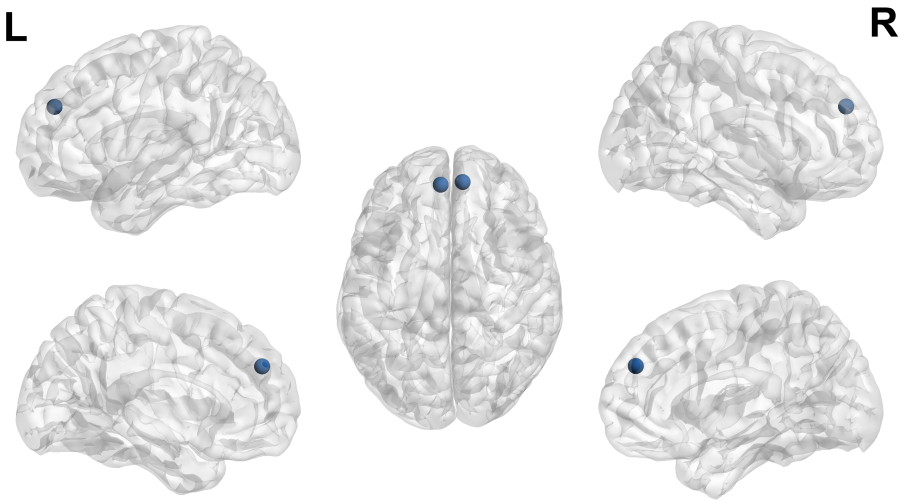


(b) Anatomic regions with higher node out-strength during execution.

Figure 4.9: Anatomic regions with higher node out-strength during planning (Top) and during execution (Bottom)

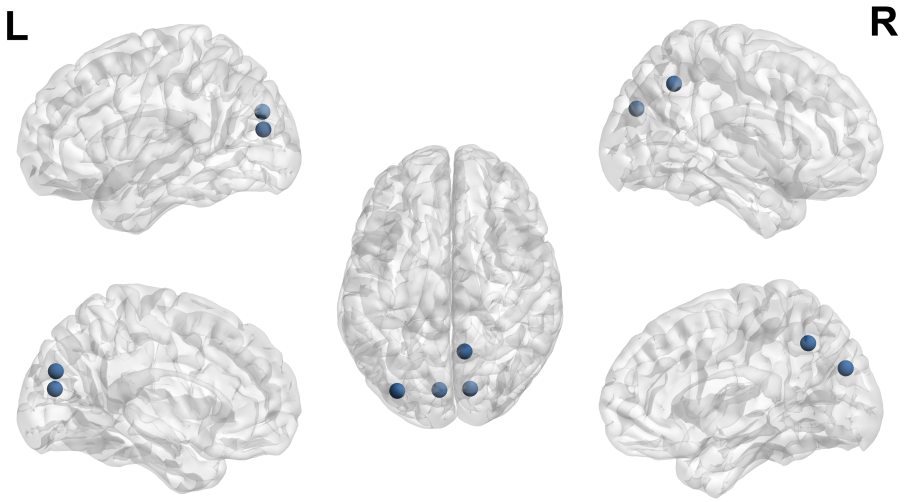


(a) Anatomic regions with higher node betweenness during planning.

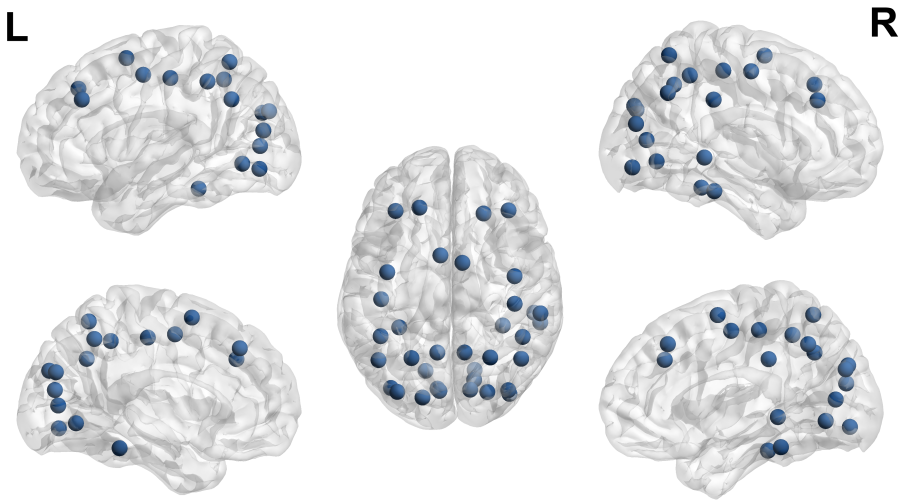


(b) Anatomic regions with higher node betweenness during execution.

Figure 4.10: Anatomic regions with higher node betweenness during planning (Top) and during execution (Bottom).



(a) Anatomic regions with higher local efficiency and clustering coefficient during planning.



(b) Anatomic regions with higher transitivity during planning.

Figure 4.11: Anatomic regions with higher local efficiency and clustering coefficient (Top) and higher transitivity (Bottom) during planning.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, we propose two different computational brain network models in order to represent, understand and analyze various high-level cognitive tasks represented by the Tower of London (TOL) game and the Human Connectome Project (HCP) datasets.

The first model is an unsupervised brain network model to study the cognitive tasks included in the HCP dataset. This model is built on top of the work proposed in [10], where in [10] the raw fMRI images are decomposed into different frequency sub-bands using discrete wavelet transform. Then, it builds a mesh network around each anatomic region using Pearson correlation and Ridge regression [10]. Our work employs a deep learning architecture, namely, stacked denoising auto-encoder, in order to learn lower dimensional connectivity patterns (features) from the constructed functional brain networks. Therefore, we train an auto-encoder for each subject at each frequency sub-band in order to learn a low-dimensional representation from the constructed brain mesh networks. Unfortunately, stacked denoising auto-encoders significantly increase the computational complexity of our proposed model, it also requires abundant data in order to successfully train the auto-encoders to capture the activation patterns of the anatomic regions. On the other hand, auto-encoders alleviate the curse of dimensionality problem and improve the brain decoding performance. After obtaining the connectivity patterns from the trained auto-encoders, we concatenate the learned features from all frequency sub-bands, then, employ a hierarchical clustering method in order to perform unsupervised brain decoding. Our proposed model successfully finds natural groupings in the learned connectivity patterns obtaining a Rand index and adjusted Rand index of 0.93 and 0.71 respectively.

When comparing our proposed model with the work presented in [10], our model is completely unsupervised, where the task labels are only used to evaluate the brain decoding performance of the proposed model. This allows our model to be applied to datasets where labeling the underlying cognitive tasks is challenging or the aim is to study the existence of natural subtasks (phases) within a high level cognitive task. On the other hand, our model is computationally more expensive compared to the model proposed in [10] as we train a stacked denoising auto-encoder for every subject at each frequency sub-band which is both computationally expensive and requires abundant data.

In order to create sufficient amount of training data, our model constructs the brain networks over significantly smaller time windows, compared with [10]. As a result, using smaller time windows increases the difficulty of the brain decoding problem as the activation patterns of each task are divided into shorter and harder to detect activation patterns. However, using shorter time windows provides our model with the flexibility and adaptiveness to be applied to a wider range of datasets where it is suitable to record the brain activation under the studied brain stimuli for relatively short durations of time.

Still, our proposed model can be improved by fine-tuning the hyper-parameters of the stacked denoising auto-encoders for each frequency sub-band individually in order to optimize the brain decoding performance. Furthermore, we aim to develop spatio-temporal models that take advantage of both the spatial and temporal information contained in the brain images while constructing brain networks. For this purpose, recurrent neural networks and recurrent auto-encoders are potential architectures to be incorporated with our current computational model.

The second proposed model is a supervised brain network model that aims to distinguish the two phases of complex problem solving, namely planning and execution. The model is made up of two parts, the first part is the preprocessing pipeline that processes the raw fMRI images in order to reduce their high spatial resolution and increase their low temporal resolution. The preprocessing pipeline achieves that using Anova voxel selection and cubic spline interpolation as well as Gaussian noise, where it increased the brain decoding power of the raw fMRI images from 0.6 to 0.82

with support vector machine and from 0.63 to 0.8 with k-means clustering.

The preprocessing pipeline provides a generic framework to process raw fMRI images by reducing the high spatial dimensionality while increasing their inherently low temporal resolution using cubic spline interpolation, which is one of the novelties of this work. It also regularizes the obtained images using Gaussian noise while preserving the signal-to-noise ratio. Still, the preprocessing pipeline can be further improved by using more adaptive methods to estimate brain volumes. For this purpose, learning methods can be employed in order to estimate brain volumes rather than deterministic methods such as cubic spline method which is used in this work.

The second part of this computational model is concerned with constructing weighted directed dynamic functional brain networks using neural networks. The constructed brain networks successfully classify the cognitive phases of complex problem solving task. The obtained brain decoding accuracy is 0.82 and 0.87 for SVM and k-means respectively. While other common solutions in the literature such as Pearson correlation [5, 10] and ridge regression [6, 57] fail to capture the interdependencies between the brain anatomic regions, the proposed model clearly succeeds in representing the underlying cognitive subtasks of complex problem solving.

In this thesis, we also studied the network properties of the constructed brain networks leading to successfully identifying potential hubs (using node out-strength and node betweenness as centrality measures) and clusters of densely interconnected brain anatomic regions (using transitivity, local efficiency and clustering coefficient as measures of segregation) during both planning and execution phases. Also, the differences and similarities between the planning and execution networks are highlighted, where obtained results clearly show that there are more potential hubs during planning compared to execution, and the clusters of densely connected anatomic regions are significantly more strongly interconnected during planning than during execution subtask.

Furthermore, future work on the tower of London (TOL) task will aim to identify online planning during the execution task. To achieve that, an unsupervised brain decoding model needs to be developed in order to successfully cluster the brain images of complex problem solving task into planning, execution and online planning sub-

tasks. After that, the network properties of each one of the subtasks is to be studied in detail and the differences as well as the similarities between the networks of the subtasks is to be examined. Furthermore, we will perform an in-depth analysis on the individual differences between the subjects aiming to identify the differences in the brain networks of planning and execution between good problem solver and bad problem solvers. In addition, we aim to study the differences between the network properties of each subject across different sessions, where the goal is to identify how experience and learning affect the brain networks of complex problem solving.

It is worth noting that the unsupervised and supervised network models that we propose in this thesis are not related, as we apply two different computational models to two different datasets. However, given the nature of each dataset (number of subjects and number of fMRI recordings per cognitive task) and the type of analysis included in each model, we were obligated to develop two different models. Therefore, we could not apply our proposed model in chapter 3 to the Tower of London dataset. We, also could not apply our proposed model in chapter 4 to the Human Connectome Project dataset.

REFERENCES

[1] S. M. LaConte, "Decoding fmri brain states in real-time," *Neuroimage*, vol. 56, no. 2, pp. 440–454, 2011.

[2] J.-D. Haynes and G. Rees, "Neuroimaging: decoding mental states from brain activity in humans," *Nature Reviews Neuroscience*, vol. 7, no. 7, p. 523, 2006.

[3] W. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. Greicius, "Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns," *Cerebral Cortex*, vol. 22, no. 1, pp. 158–165, 2012.

[4] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[5] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fmri connectivity graphs," *Neuroimage*, vol. 56, no. 2, pp. 616–626, 2011.

[6] I. Onal, M. Ozay, E. Mizrak, I. Oztekin, and F. T. Y. Vural, "A new representation of fmri signal by a set of local meshes for brain decoding," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 4, pp. 683–694, 2017.

[7] M. A. Lindquist, "The Statistical Analysis of fMRI Data," *Statistical Science*, pp. 439–464, 2008.

[8] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville, "Machine Learning with Brain Graphs: Predictive Modeling Approaches for Functional Imaging in Systems Neuroscience," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 58–70, 2013.

[9] M. Ekman, J. Derrfuss, M. Tittgemeyer, and C. J. Fiebach, "Predicting Errors

from Reconfiguration Patterns in Human Brain Networks," *Proceedings of the National Academy of Sciences*, vol. 109, no. 41, pp. 16714–16719, 2012.

[10] I. O. Ertugrul, M. Ozay, and F. T. Y. Vural, "Hierarchical multi-resolution mesh networks for brain decoding," *Brain imaging and behavior*, pp. 1–17, 2016.

[11] O. Firat, E. Aksan, I. Oztekin, and F. T. Y. Vural, "Learning Deep Temporal Representations for fMRI Brain Decoding," in *Medical Learning Meets Medical Imaging*, pp. 25–34, Springer, 2015.

[12] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[13] B. B. Kivilcim, I. O. Ertugrul, and F. T. Y. Vural, "Modeling brain networks with artificial neural networks," *arXiv preprint arXiv:1807.08368*, 2018.

[14] R. J. Sternberg and W. Salter, "Conceptions of intelligence. in r. j. sternberg, editor, handbook of human intelligence," *Cambridge University Press, New York*, pp. 3–28, 1982.

[15] D. Garlick, "Understanding the nature of the general factor of intelligence: the role of individual differences in neural plasticity as an explanatory mechanism.," *Psychological review*, vol. 109, no. 1, p. 116, 2002.

[16] H. A. Simon and A. Newell, "Human problem solving: The state of the theory in 1970.," *American Psychologist*, vol. 26, no. 2, p. 145, 1971.

[17] A. Newell, J. C. Shaw, and H. A. Simon, "Elements of a theory of human problem solving," tech. rep., RAND CORP SANTA MONICA CALIF, 1957.

[18] A. Newell, H. A. Simon, *et al.*, *Human problem solving*, vol. 104. Prentice-Hall Englewood Cliffs, NJ, 1972.

[19] G. Ward and R. Morris, "Introduction to the psychology of planning," *The cognitive psychology of planning*, pp. 1–34, 2005.

[20] J. R. Hayes, *The complete problem solver*. Routledge, 2013.

[21] F. Phillips, "A research note on accounting students' epistemological beliefs, study strategies, and unstructured problem-solving performance," *Issues in Accounting Education*, vol. 16, no. 1, pp. 21–39, 2001.

[22] L. H. Phillips, "The role of memory in the tower of london task," *Memory*, vol. 7, no. 2, pp. 209–231, 1999.

[23] S. D. Newman and G. Pittman, "The tower of london: A study of the effect of problem structure on planning," *Journal of Clinical and Experimental Neuropsychology*, vol. 29, no. 3, pp. 333–342, 2007.

[24] J. R. Anderson, M. V. Albert, and J. M. Fincham, "Tracing problem solving in real time: fmri analysis of the subject-paced tower of hanoi," *Journal of cognitive neuroscience*, vol. 17, no. 8, pp. 1261–1274, 2005.

[25] T. Shallice, "Specific impairments of planning," *Phil. Trans. R. Soc. Lond. B*, vol. 298, no. 1089, pp. 199–209, 1982.

[26] W. S. MacAllister, H. A. Bender, L. Whitman, A. Welsh, S. Keller, Y. Granader, and E. M. Sherman, "Assessment of executive functioning in childhood epilepsy: the tower of london and brief," *Child Neuropsychology*, vol. 18, no. 4, pp. 404–415, 2012.

[27] I. Goethals, K. Audenaert, F. Jacobs, C. Van de Wiele, H. Ham, H. Pyck, A. Vandierendonck, C. Van Heeringen, and R. Dierckx, "Blunted prefrontal perfusion in depressed patients performing the tower of london task," *Psychiatry Research: Neuroimaging*, vol. 139, no. 1, pp. 31–40, 2005.

[28] I. Rektorova, H. Srovnalova, R. Kubikova, and J. Prasek, "Striatal dopamine transporter imaging correlates with depressive symptoms and tower of london task performance in parkinson's disease," *Movement disorders*, vol. 23, no. 11, pp. 1580–1587, 2008.

[29] P. E. Rasser, P. Johnston, J. Lagopoulos, P. B. Ward, U. Schall, R. Thienel, S. Bender, A. W. Toga, and P. M. Thompson, "Functional mri bold response to tower of london performance of first-episode schizophrenia patients using cortical pattern matching," *Neuroimage*, vol. 26, no. 3, pp. 941–951, 2005.

[30] D. Carlin, J. Bonerba, M. Phipps, G. Alexander, M. Shapiro, and J. Grafman, "Planning impairments in frontal lobe dementia and frontal lobe lesion patients," *Neuropsychologia*, vol. 38, no. 5, pp. 655–665, 2000.

[31] M. A. Just, T. A. Keller, V. L. Malave, R. K. Kana, and S. Varma, "Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1292–1313, 2012.

[32] H. Srovnalova, R. Marecek, R. Kubikova, and I. Rektorova, "The role of the right dorsolateral prefrontal cortex in the tower of london task performance: repetitive transcranial magnetic stimulation study in patients with parkinson's disease," *Experimental brain research*, vol. 223, no. 2, pp. 251–257, 2012.

[33] J. M. Unterrainer, B. Rahm, C. P. Kaller, R. Leonhart, K. Quiske, K. Hoppe-Seyler, C. Meier, C. Müller, and U. Halsband, "Planning abilities and the tower of london: is this task measuring a discrete cognitive function?," *Journal of clinical and experimental neuropsychology*, vol. 26, no. 6, pp. 846–856, 2004.

[34] N. A. Zook, D. B. Davalos, E. L. DeLosh, and H. P. Davis, "Working memory, inhibition, and fluid intelligence as predictors of performance on tower of hanoi and london tasks," *Brain and cognition*, vol. 56, no. 3, pp. 286–292, 2004.

[35] Y.-K. Chang, C.-L. Tsai, T.-M. Hung, E. C. So, F.-T. Chen, and J. L. Etnier, "Effects of acute exercise on executive function: a study with a tower of london task," *Journal of Sport and Exercise Psychology*, vol. 33, no. 6, pp. 847–865, 2011.

[36] D. Albert and L. Steinberg, "Age differences in strategic planning as indexed by the tower of london," *Child development*, vol. 82, no. 5, pp. 1501–1517, 2011.

[37] N. Zook, M. C. Welsh, and V. Ewing, "Performance of healthy, older adults on the tower of london revised: Associations with verbal and nonverbal abilities," *Aging, Neuropsychology, and Cognition*, vol. 13, no. 1, pp. 1–19, 2006.

[38] A. Boghi, R. Rasetti, F. Avidano, C. Manzone, L. Orsi, F. D'agata, P. Caroppo, M. Bergui, P. Rocca, L. Pulvirenti, *et al.*, "The effect of gender on planning: An fmri study using the tower of london task," *Neuroimage*, vol. 33, no. 3, pp. 999–1010, 2006.

[39] J. M. Unterrainer, C. P. Kaller, U. Halsband, and B. Rahm, "Planning abilities and chess: A comparison of chess and non-chess players on the tower of london task," *British Journal of Psychology*, vol. 97, no. 3, pp. 299–311, 2006.

[40] C. P. Kaller, J. M. Unterrainer, B. Rahm, and U. Halsband, "The impact of problem structure on planning: Insights from the tower of london task," *Cognitive Brain Research*, vol. 20, no. 3, pp. 462–472, 2004.

[41] W. K. Berg, D. L. Byrd, J. P. McNamara, and K. Case, "Deconstructing the tower: Parameters and predictors of problem difficulty on the tower of london task," *Brain and Cognition*, vol. 72, no. 3, pp. 472–482, 2010.

[42] J. Unterrainer, B. Rahm, U. Halsband, and C. Kaller, "What is in a name: Comparing the tower of london with one of its variants," *Cognitive Brain Research*, vol. 23, no. 2, pp. 418–428, 2005.

[43] J. Unterrainer, B. Rahm, R. Leonhart, C. Ruff, and U. Halsband, "The tower of london: The impact of instructions, cueing, and learning on planning abilities," *Cognitive Brain Research*, vol. 17, no. 3, pp. 675–683, 2003.

[44] Z. Campbell, K. K. Zakzanis, D. Jovanovski, S. Joordens, R. Mraz, and S. J. Graham, "Utilizing virtual reality to improve the ecological validity of clinical neuropsychology: an fmri case study elucidating the neural basis of planning by comparing the tower of london with a three-dimensional navigation task," *Applied Neuropsychology*, vol. 16, no. 4, pp. 295–306, 2009.

[45] S. D. Newman, B. Pruce, A. Rusia, and T. Burns Jr, "The effect of strategy on problem solving: an fmri study," *The Journal of Problem Solving*, vol. 3, no. 1, p. 2, 2010.

[46] S. D. Newman, J. A. Greco, and D. Lee, "An fmri study of the tower of london: a look at problem structure differences," *Brain research*, vol. 1286, pp. 123–132, 2009.

[47] G. Wagner, K. Koch, J. R. Reichenbach, H. Sauer, and R. G. Schlösser, "The special involvement of the rostrolateral prefrontal cortex in planning abilities: an event-related fmri study with the tower of london paradigm," *Neuropsychologia*, vol. 44, no. 12, pp. 2337–2347, 2006.

[48] S. D. Newman, P. A. Carpenter, S. Varma, and M. A. Just, "Frontal and parietal participation in problem solving in the tower of london: fmri and computational modeling of planning and high-level perception," *Neuropsychologia*, vol. 41, no. 12, pp. 1668–1682, 2003.

[49] F. Cazalis, R. Valabregue, M. Pélégrini-Issac, S. Asloun, T. Robbins, and S. Granon, "Individual differences in prefrontal cortical activation on the tower of london planning task: implication for effortful processing," *European Journal of Neuroscience*, vol. 17, no. 10, pp. 2219–2225, 2003.

[50] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[51] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, *et al.*, "Function in the human connectome: task-fmri and individual differences in behavior," *Neuroimage*, vol. 80, pp. 169–189, 2013.

[52] M.-E. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. Bullmore, "Functional connectivity and brain networks in schizophrenia," *Journal of Neuroscience*, vol. 30, no. 28, pp. 9477–9487, 2010.

[53] V. Menon, "Large-scale brain networks and psychopathology: a unifying triple network model," *Trends in cognitive sciences*, vol. 15, no. 10, pp. 483–506, 2011.

[54] A. Kurmukov, M. Ananyeva, Y. Dodonova, B. Gutman, J. Faskowitz, N. Jahanshad, P. Thompson, and L. Zhukov, "Classifying phenotypes based on the community structure of human brain networks," in *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*, pp. 3–11, Springer, 2017.

[55] H. Lee, D. S. Lee, H. Kang, B.-N. Kim, and M. K. Chung, "Sparse brain network recovery under compressed sensing," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1154–1165, 2011.

[56] O. Fırat, M. Özay, I. Önal, İ. Öztekiny, and F. T. Y. Vural, "Functional mesh learning for pattern analysis of cognitive processes," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2013 12th IEEE International Conference on*, pp. 161–167, IEEE, 2013.

[57] I. Onal, M. Ozay, and F. T. Y. Vural, "Modeling voxel connectivity for brain decoding," in *Pattern Recognition in NeuroImaging (PRNI), 2015 International Workshop on*, pp. 5–8, IEEE, 2015.

[58] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Constrained sparse functional connectivity networks for mci classification," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 212–219, 2012.

[59] P. A. Valdés-Sosa, J. M. Sánchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez, "Estimating brain functional connectivity with sparse multivariate autoregression," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 360, no. 1457, pp. 969–981, 2005.

[60] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[61] D. S. Bassett and E. Bullmore, "Small-world brain networks," *The neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.

[62] H.-J. Park and K. Friston, "Structural and functional brain networks: from connections to cognition," *Science*, vol. 342, no. 6158, p. 1238411, 2013.

[63] J. D. Power, D. A. Fair, B. L. Schlaggar, and S. E. Petersen, "The development of human functional brain networks," *Neuron*, vol. 67, no. 5, pp. 735–748, 2010.

[64] M. Rubinov and O. Sporns, "Weight-conserving characterization of complex functional brain networks," *Neuroimage*, vol. 56, no. 4, pp. 2068–2079, 2011.

[65] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: identifying differences in brain networks," *Neuroimage*, vol. 53, no. 4, pp. 1197–1207, 2010.

[66] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS computational biology*, vol. 3, no. 2, p. e17, 2007.

[67] U. Braun, S. F. Muldoon, and D. S. Bassett, "On human brain networks in health and disease," *eLS*, 2015.

[68] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[69] G. Fagiolo, "Clustering in complex directed networks," *Physical Review E*, vol. 76, no. 2, p. 026107, 2007.

[70] W. H. Thompson and P. Fransson, "The Frequency Dimension of fMRI Dynamic Connectivity: Network Connectivity, Functional Hubs and Integration in the Resting Brain," *NeuroImage*, vol. 121, pp. 227–242, 2015.

[71] E. Bullmore, J. Fadili, V. Maxim, L. Şendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, "Wavelets and functional Magnetic Resonance Imaging of the Human Brain," *Neuroimage*, vol. 23, pp. S234–S249, 2004.

[72] S. P. Pantazatos, A. Talati, P. Pavlidis, and J. Hirsch, "Decoding Unattended Fearful Faces with Whole-Brain Correlations: An Approach to Identify Condition-Dependent Large-Scale Functional Connectivity," *PLoS Computational Biology*, vol. 8, no. 3, p. e1002441, 2012.

[73] C. Poultney, S. Chopra, Y. L. Cun, *et al.*, "Efficient Learning of Sparse Representations with an Energy-Based Model," in *Advances in neural information processing systems*, pp. 1137–1144, 2007.

[74] S. Wager, S. Wang, and P. S. Liang, "Dropout Training as Adaptive Regularization," in *Advances in neural information processing systems*, pp. 351–359, 2013.

[75] V. D. Calhoun, R. Miller, G. Pearlson, and T. Adalı, "The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery," *Neuron*, vol. 84, no. 2, pp. 262–274, 2014.

[76] G. W. Milligan and M. C. Cooper, "A study of the comparability of external criteria for hierarchical cluster analysis," *Multivariate Behavioral Research*, vol. 21, no. 4, pp. 441–458, 1986.

[77] A. Afrasiyabi, I. Onal, and F. T. Y. Vural, "A sparse temporal mesh model for brain decoding," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2016 IEEE 15th International Conference on*, pp. 198–206, IEEE, 2016.

[78] A. Afrasiyabi, I. Onal, and F. T. Y. Vural, "Effect of voxel selection on temporal mesh model for brain decoding," in *Signal Processing and Communication Application Conference (SIU), 2016 24th*, pp. 2249–2252, IEEE, 2016.

[79] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fmri: a tutorial overview," *Neuroimage*, vol. 45, no. 1, pp. S199–S209, 2009.

[80] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fmri)"brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex," *Neuroimage*, vol. 19, no. 2, pp. 261–270, 2003.

[81] M. Xia, J. Wang, and Y. He, "Brainnet viewer: a network visualization tool for human brain connectomics," *PloS one*, vol. 8, no. 7, p. e68910, 2013.

[82] S. McKinley and M. Levine, "Cubic spline interpolation," *College of the Redwoods*, vol. 45, no. 1, pp. 1049–1060, 1998.

[83] M. Frigo and S. G. Johnson, "Fftw: An adaptive software architecture for the fft," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 3, pp. 1381–1384, IEEE, 1998.

[84] W. T. Cochran, J. W. Cooley, D. L. Favin, H. D. Helms, R. A. Kaenel, W. W. Lang, G. C. Maling, D. E. Nelson, C. M. Rader, and P. D. Welch, "What is the fast fourier transform?," *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1664–1674, 1967.

[85] K. Matsuoka, "Noise injection into inputs in back-propagation learning," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 3, pp. 436–440, 1992.

[86] R. Reed, S. Oh, and R. Marks, "Regularization using jittered training data," in *Neural Networks, 1992. IJCNN., International Joint Conference on*, vol. 3, pp. 147–152, IEEE, 1992.

[87] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *Journal of machine learning research*, vol. 9, no. Aug, pp. 1871–1874, 2008.

[88] I. I. Goldberg, M. Harel, and R. Malach, "When the brain loses its self: prefrontal inactivation during sensorimotor processing," *Neuron*, vol. 50, no. 2, pp. 329–339, 2006.

[89] G. R. Mangun, M. H. Buonocore, M. Girelli, and A. P. Jha, "Erp and fmri measures of visual spatial selective attention," *Human brain mapping*, vol. 6, no. 5-6, pp. 383–389, 1998.

[90] E. Brunet, Y. Sarfati, M.-C. Hardy-Baylé, and J. Decety, "A pet investigation of the attribution of intentions with a nonverbal task," *Neuroimage*, vol. 11, no. 2, pp. 157–166, 2000.

[91] E. D. Leshikar, A. Duarte, and C. Hertzog, "Task-selective memory effects for successfully implemented encoding strategies," *PloS one*, vol. 7, no. 5, p. e38160, 2012.

[92] P. C. Fletcher, C. D. Frith, S. C. Baker, T. Shallice, R. S. Frackowiak, and R. J. Dolan, "The mind's eye—precuneus activation in memory-related imagery," *Neuroimage*, vol. 2, no. 3, pp. 195–200, 1995.

[93] R. Kawashima, P. E. Roland, and B. T. O'Sullivan, "Functional anatomy of reaching and visuomotor learning: a positron emission tomography study," *Cerebral Cortex*, vol. 5, no. 2, pp. 111–122, 1995.

[94] A. E. Cavanna and M. R. Trimble, "The precuneus: a review of its functional anatomy and behavioural correlates," *Brain*, vol. 129, no. 3, pp. 564–583, 2006.

[95] N. Wenderoth, F. Debaere, S. Sunaert, and S. P. Swinnen, "The role of anterior cingulate cortex and precuneus in the coordination of motor behaviour," *European Journal of Neuroscience*, vol. 22, no. 1, pp. 235–246, 2005.

[96] Q. Chen, R. Weidner, S. Vossel, P. H. Weiss, and G. R. Fink, "Neural mechanisms of attentional reorienting in three-dimensional space," *Journal of Neuroscience*, vol. 32, no. 39, pp. 13352–13362, 2012.

[97] M. L. Seghier, "The angular gyrus: multiple functions and multiple subdivisions," *The Neuroscientist*, vol. 19, no. 1, pp. 43–61, 2013.