LEARNING TRANSFERABILITY OF COGNITIVE TASKS BY GRAPH GENERATION FOR BRAIN DECODING

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

BY

BİLGİN COŞKUN

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

DECEMBER 2021

Approval of the thesis:

LEARNING TRANSFERABILITY OF COGNITIVE TASKS BY GRAPH GENERATION FOR BRAIN DECODING

submitted by **BİLGİN COŞKUN** 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** _____

Examining Committee Members:

Assoc. Prof. Dr. Sinan Kalkan
Computer Engineering, METU _____

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

Prof. Dr. Pınar Duygulu Şahin
Computer Engineering, Hacettepe University _____

Date: 10.12.2021

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, Surname: Bilgin Coşkun

Signature :

ABSTRACT

LEARNING TRANSFERABILITY OF COGNITIVE TASKS BY GRAPH GENERATION FOR BRAIN DECODING

Coşkun, Bilgin
M.S., Department of Computer Engineering
Supervisor: Prof. Dr. Fatoş T. Yarman Vural

December 2021, 59 pages

Brain decoding involves analyzing the cognitive states of human brain by using some statistical techniques in order to understand the relations among the cognitive states, based on neuroimaging data. A very powerful tool to acquire the brain data is functional magnetic resonance images (fMRI), which generates three-dimensional brain volume at each time instant, while a subject performs a cognitive task involving social activities, emotion processing, game playing, memory etc. However, it is very difficult and time-consuming to acquire data, which is statistically sufficient to train a deep neural network. An alternative to generate data is to reuse the available statistically similar data, so that the information of the available data can be transferred to relatively small datasets. In this thesis, we propose a pipeline method to create and verify a graph, which shows the relations among well-defined cognitive tasks, based on fMRI data. We measure the affinity between the cognitive tasks on an fMRI dataset gathered from multiple subjects using the adaptation of an existing end-to-end method which compares transferring performance between the cognitive tasks using learned latent representations. However, due to high variance between measurements on different subjects, the results vary greatly. In order to both verify and regulate the

differences, we use the performance of the binary classifiers trained on imbalanced data. In the last step, we generate a graph, where the edges are the possible relations between the cognitive tasks which have the potential to improve transfer learning in other datasets.

Keywords: fMRI, deep learning, transfer learning, cognitive state classification, Human Connectome Project

ÖZ

BEYİN ÇÖZÜMLEMESİ İÇİN BİLİŞSEL GÖREVLERİN AKTARILABİLİRLİĞİNİN GRAFİK ÜRETİMİ İLE ÖĞRENİMİ

Coşkun, Bilgin
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Fatoş T. Yarman Vural

Aralık 2021 , 59 sayfa

Beyin çözümlemesi beyin taraması verilerini temel alarak, bazı istatistiksel tekniklerin kullanımı ile bilişsel durumlar arasındaki ilişkiyi anlamak için insan beyninin bilişsel durumunu analiz etmeyi içermekte. Beyin verisini elde etmek için denek sosyal aktivite, duygu işleme, oyun oynama, hafıza gibi bilişsel görevleri yapmaktayken her zaman anında 3 boyutlu beyin hacmi üreten fonksiyonel manyetik rezonans görüntüleme (fMRI) güçlü bir araç olarak karşımıza çıkmakta. Ancak derin sinir ağlarını eğitmeye istatistiksel olarak yetecek veri elde etmek hem çok zor hem de zaman almakta. Veri üretmeye başka bir seçenek de halihazırda varolan istatistiksel olarak benzer veriyi tekrar kullanarak mevcut verideki bilgiyi görece küçük veri kümesine aktarmak. Bu tezde fMRI verilerine dayanarak, tanımlı bilişsel görevler arasındaki ilişkileri gösteren bir çizge oluşturmak ve doğrulamak için bir işlem dizisi yöntemi önermekteyiz. Bilişsel görevler arasındaki yakınlığı bir çok denekten toplanan fMRI veri kümesi üzerinde, aktarım performansını öğrenilmiş gizli sunumları kullanarak karşılaştıran mevcut bir uçtan uca yöntemi kullanarak ölçmekteyiz. Ama, farklı denekler üzerinde yapılan ölçümlerin değişkenliğinden dolayı sonuçlar da büyük oranda değişkenlik

göstermekte. Bu değişkenlikleri hem doğrulamak hem de düzenlemek için, dengesiz veri üzerinde eğitilmiş ikili sınıflandırıcılarının performansını kullanmaktayız. Son adımda, kenarları başka veri kümelerinde aktarım öğrenimini geliştirme potansiyeli olan bilişsel görevler arasındaki olası ilişkileri tanımlayan bir çizge oluşturmaktayız.

Anahtar Kelimeler: derin öğrenme,aktarım öğrenmesi, bilişsel durum sınıflandırması, Human Connectome Project

To my family

ACKNOWLEDGMENTS

Foremost, I would like to express my sincerest gratitude to my supervisor, Prof. Dr. Fatoş T. Yarman Vural. Without her wise guidance, endless patience and unlimited support, I have no doubt my career would take a very steep, and most probably to a very boring direction, turn a long time ago.

Also, without my parents, I might not have the strength to overcome some obstacles seemed insurmountable at first sight. I am glad they were here, by my side.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AAL	Automatic Anatomical Labeling
BOLD	Brain Oxygen Level Dependency
fMRI	functional Magnetic Resonance Imaging
HCP	Human Connectome Project
LRELU	Leaky Rectified Linear Unit
SR	Spatial Resolution
tfMRI	task-focused fMRI
TR	Temporal Resolution

CHAPTER 1

INTRODUCTION

We think, talk and dream using our brains. In a sense, our brain is the most important part of our body that makes us. As a burden of self-awareness, we also get curious about how does the brain work. Physiologically, we have some answers: we observe a complex network of cells with complex electrochemical activities. However, we are far from understanding the clockwork which makes us look over the horizon of the known.

One way to understand the mechanisms underlying the brain has been observing the instant changes in the anatomic regions of the brain during stimuli. Brain decoding aims to analyze the overall mental process better, by analyzing the activation patterns on the brain under well-defined neuroscientific experiments, each of which corresponds to an isolated cognitive task.

Although there are several, one of the brain imaging techniques, functional Magnetic Resonance Imaging (fMRI), gained popularity by enabling us to measure these patterns on very small brain volumes, called voxels, by inspecting the oxygen level changes within the volume and creating a 3 dimensional brain volume within an acceptable time interval in a non-invasive way.

One specific fMRI dataset group, task-focused fMRI (tfMRI) datasets, are especially useful to explore the functions of the anatomic regions of the brain by taking images when the subject doing a given cognitive task which can vary over a wide range such as language skills or emotional response.

Unfortunately, the downside of most of the fMRI datasets are statistical insufficient of the acquired datasets. The experimental setups are restricted due to the limitations of

the fMRI machines. Furthermore, it is very expensive and time-consuming to collect data from multiple subjects. As a result, most of the fMRI datasets are sparse both in terms of numbers and quality. Additionally, variation of the operating characteristics of fMRI machines makes different datasets incompatible with each other. Due to these severe limitations, fMRI experiments are usually terminated with small, hence insufficient data.

1.1 Motivation and Problem Definition

In order to overcome the insufficient data problem, one approach is to modify data using prior knowledge, such as generating synthetic data or modifying representation of data ([12]). However, this approach heavily leans on the assumptions based on neuroscientific experiments, which may be misleading or not be applicable to the current dataset at hand.

Another approach is to transfer knowledge between cognitive tasks. Transfer learning is a machine learning field that aims to improve performance by transferring knowledge with the assumption that there is an underlying common structure. By common sense, this common structure for information is surely utilized by the brain. Generally, Turkish speakers learn English harder than German speakers. Learning to play another musical instrument is easier than learning to play the first musical instrument. It is also shown empirically for specific cognitive task groups like visual tasks ([14]) or arithmetic processing ([10]) and by doing transfer learning successfully ([22], [35]).

However, in order to transfer learn between cognitive tasks, we need to define relations between cognitive tasks explicitly.

1.2 Proposed Methods and Models

In order to quantify the degree of relations between the cognitive states, we adapt the method from Zamir et al. ([38]) to Human Connectome Project (HCP) dataset to create an affinity matrix between its seven well-defined cognitive tasks. By measuring transfer learning performance between representations of different cognitive tasks, we

aim to find a set of explicit relations which can be utilized in transfer learning.

Since the suggested approach does not depend on prior knowledge, when discovering affinities between cognitive tasks, it has the potential to find obscure relations. If we can find relatively stable relations in between-subject experiments with abundant data, we can possibly utilize these relations in other, relatively small datasets also.

In order to measure the validity of found affinities of the cognitive tasks, we also create another affinity matrix by measuring classification performance on imbalanced data without transfer learning.

To the best of our knowledge, the only research similar to ours which applies to fMRI datasets is Wang et al. ([36]). When we compare our study to that of Wang et al., our major differences can be summarized as follows:

- Their research is focused on cognitive tasks only involving computer vision tasks, which is rather trivial to discriminate, while the HCP data set involves cognitive skills, captured by fMRI signals. It is well known that the cognitive tasks are very complicated and it is very difficult to disentangle them.

- Their experiments are done with the dataset of within-subject experiments. In other words, the vision tasks are classified only for a specific subject performing the experiments. On the other hand,in our suggested method we use the whole data collected fro multiple subjects. This approach enable us to estimate general affinities across the cognitive tasks. Thus it is possible to adopt the suggested method to a different dataset in the future.

- In their method, affinities between computer vision tasks are calculated based on the predicted brain activity and compared with the original method since tasks are same. In our proposed method, we calculate affinities between tasks based on the reconstruction loss of transfer learning between the cognitive tasks. In order to verify the results, we use the performance of binary classifiers trained with imbalanced data.

1.3 Contributions and Novelties

It is well-known that the success of the transfer learning methodologies heavily depend on the compatibility of the information content of the source dataset and the target dataset. Thus, it is very crucial to estimate the transferability of the information content of a source class to a target class.

In this study, we attempt to generalize the method suggested by Zamir et al. ([38]) to estimate the affinities between the cognitive tasks, across the multiple subjects. Considering the fact that there are large variations among the subjects, while performing a cognitive task, this is a very difficult problem. This difficulty is clearly observed, when we directly apply the method in [38], where there is practically no resemblance among the affinities between the cognitive tasks, across the folds. In order to estimate a generic set of affinities across the tasks, we suggest a voting schema, which brings all the estimated affinity matrices across the folds into the same ground. Although the suggested computational method does not require any prior assumptions, the resulting affinity matrices is consistent to the findings in experimental neuroscience to a certain degree. We hope that the suggested method is also applicable to relatively small fMRI datasets to estimate the affinities across the tasks and make transfer learning methods applicable for brain decoding .

In order to verify the affinity matrices created by the adapted method, we create another affinity matrix using the classifier performances trained with imbalanced data and examine the similarity between them.

1.4 The Outline of the Thesis

The outline of this thesis is given as follows:

In Chapter 2, we discuss the fMRI technology and the preprocessing steps for fMRI datasets, brain decoding, machine learning and especially deep learning methods on brain decoding, and transfer learning on brain decoding.

In Chapter 3, we introduce our proposed method of estimating the affinities among

the cognitive tasks and generate a relational graph among them. We also briefly discuss the structure of HCP dataset, together with the dimension reduction and sample selection techniques to reduce the curse of dimensionality problem.

In Chapter 4, we present results of our experiments on HCP dataset. We compare the relations estimated by our suggested method with neuroscientific evidence.

Finally, in Chapter 5 we conclude this thesis by discussing our findings, and we suggest possible future works.

CHAPTER 2

BACKGROUND FOR DECODING THE COGNITIVE TASKS OF BRAIN

In this chapter, we discuss the fMRI technology, the preprocessing steps for fMRI datasets and give the general overview of brain decoding literature. We also discuss machine learning on brain decoding with the emphasis on deep learning, since the ability to learn higher level representations of deep learning models in an unsupervised fashion enabled us to use the methods proposed in this thesis. Lastly, we summarize the methods utilizing transfer learning on brain decoding.

2.1 Nature of fMRI (functional Magnetic Resonance Imaging) Data

Basis of functional Magnetic Resonance Imaging (fMRI) stems from the fact that hemoglobin protein (Hb) in red blood cells which carries oxygen to the tissues through vessels have different magnetic characteristics when it is oxygenated (HbO_2) and deoxygenated ([16]). HbO_2 is diamagnetic like the brain tissue itself, thus indistinguishable from it. Fully deoxygenated Hb, on the other hand, is highly paramagnetic. Like all tissues, the brain requires energy, thus oxygen, to function whether it is in resting state or stimulated state. As expected, in the stimulated state, the brain requires more energy, hence more oxygen than in the resting state. What is unexpected, however, oxygen level actually increases in the stimulated areas on the brain compared to their resting state due to dilated vessels and increased blood flow and delivering more oxygen than consumed by the system. By measuring the Blood Oxygen Level Dependent (BOLD) changes between the resting and the stimulated state in unit volume (called voxel) we can construct a 3D image of the brain which shows which parts are activated with a given stimulus. Unfortunately, limitations of MR machines also create

limitations for the dataset created by using it. Main limitations are spatial resolution (SR) which defines the size of the voxel in the image and temporal resolution (TR) which determines the time interval between two consecutive brain images.

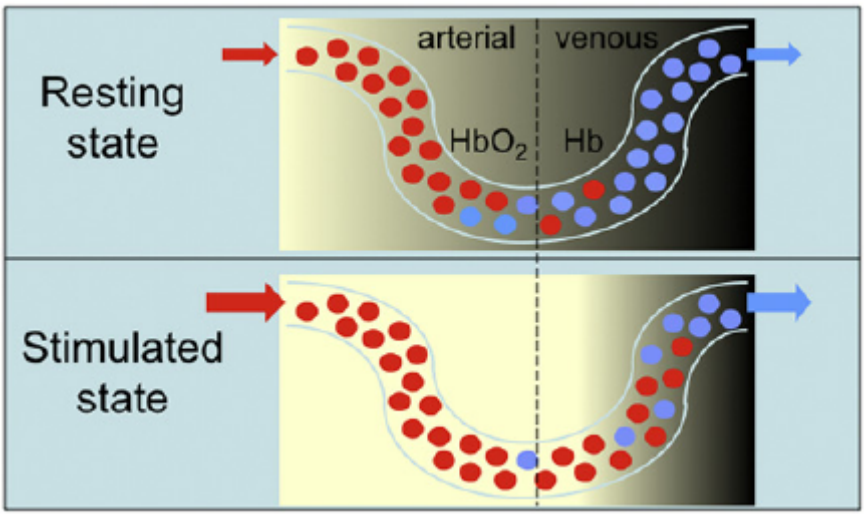

Figure 2.1: Magnetism Change in Brain Tissue Capillaries. When oxygen removed from hemoglobin, due to the changes in its magnetism. Hb acts as a contrast agent (shown darker). When the tissue is stimulated, due to the increased blood flow, HbO_2 increases and causes BOLD signal to increase. Taken from [16].

2.1.1 Acquisition and Preprocessing tfMRI Data

To acquire tfMRI datasets, there are four steps:

Stimulus: A subject is asked to do a cognitive task that creates stimuli on the subject's brain. The cognitive tasks in our dataset are explained in Section 3.1.

Acquisition: These stimuli create BOLD responses on the brain per voxel. The responses captured by the fMRI machine. Due to TR of the machine, these captures are not instant.

Image Construction: Due to the nature of fMRI machines, and involuntary movement of the subjects, these BOLD responses are needed to be preprocessed before creating a 3D brain image. The details of the preprocessing steps are explained in Glasse et al. ([15]).

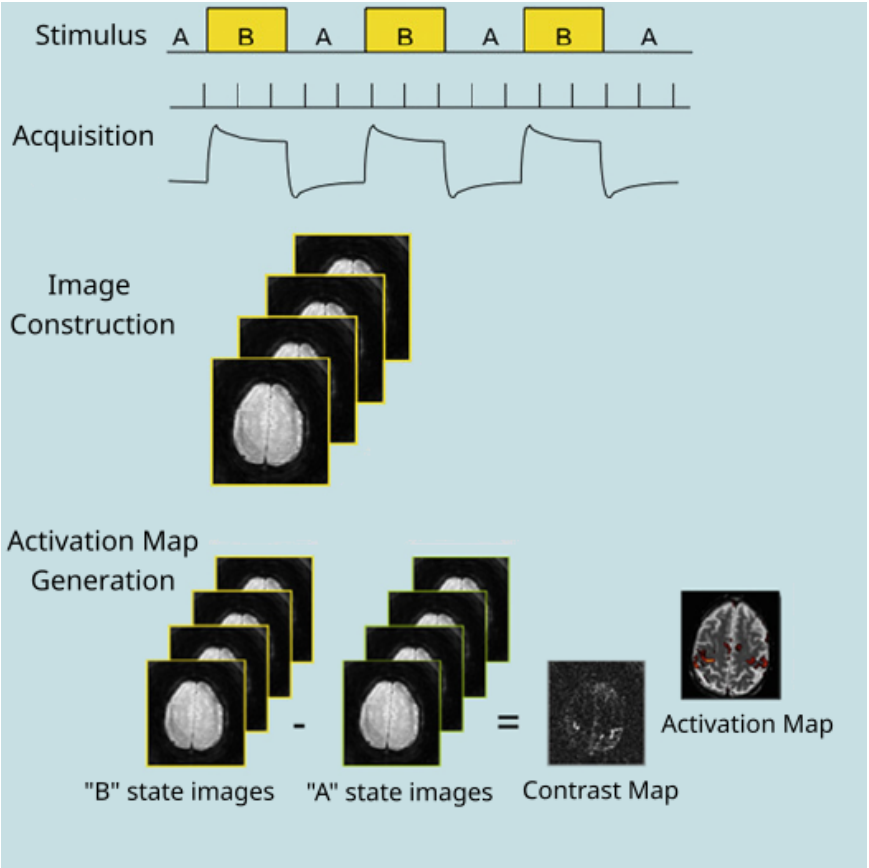

Figure 2.2: Steps to Acquire tfMRI Data. Based on the original image taken from [16].

In summary, these steps are:

- Realigning the voxels in different time instances
- Normalization of the images to a standard template
- Smoothing the images

Activation Map Generation: Created images does not explain which parts of the brain is active by itself. They need to be compared with an image captured during another state. For tfMRI datasets, this comparison image is captured during when the subject doing nothing, which is also called as resting state.

2.2 Brain Decoding Using Machine Learning

Decoding cognitive tasks of human brain involves finding the relevant cognitive states captured by the functional magnetic resonance imaging techniques, while the subject performs a predefined cognitive task.

Brain decoding techniques are very useful for analyzing the brain states by non-destructive methods. Thus, these techniques enable us to study the active brain regions, while the subject perform cognitive tasks. It is also used for diagnosing neurological diseases and mental disorders such as Alzheimer disease or Schizophrenia.

Classical brain decoding techniques involve some statistical analysis of the fMRI data. These techniques analyze the statistical properties of voxel intensity values, such as estimating the probability densities of the voxel intensity values and its moments. Then, the methods perform statistical significance analysis to relate the voxel intensity values to the associated cognitive task.

One example of the brain decoding technique is the study reported by Dehaene et al. ([9]) In this work, an fMRI dataset gathered by 7 subjects are classified to show the response on the motor cortex into simple left-hand vs right-hand movement by predicting the response with great accuracy.

The first pioneering work which applies the machine learning methods to the brain decoding problem is suggested by Haxby et al. ([18]). They try to analyze the patterns on the visual system by a classification algorithm. In the experiments, the subjects are asked to view faces, cats, five categories of man-made objects and nonsense pictures while the response in their brains is recorded using fMRI. Different patterns for every category can be recognized. Later, hundreds of studies employed similar techniques, called multi voxel pattern analysis.

Another pioneering work is suggested by Mitchell et al. ([24]). In this paper, they use a machine learning method to classify instantaneous cognitive states from consecutive brain volume instances. In their following study, Mitchell et al.([25]) try to classify several cognitive states from a single fMRI time instance for three different setups. In the first setup, the subjects are asked to look at a picture or a sentence. In the second

setup, the subjects are asked whether the sentence they are reading is ambiguous or non-ambiguous. In the last setup, the subjects are asked to determine which of 12 categories, the words they are presented belong to.

Kamitani and Tong ([19]) successfully classify 8 distinct orientations of the viewed object with support vector machines using the fMRI measurements from early visual areas of the brain. Some other studies using support vector machines include classifying true and false statements by Davatzikos et al. ([8]), distinguishing healthy subjects from the clinically depressed subjects using resting-state fMRI data by Craddock et al. ([7]) or dementia detection on resting-state fMRI data using the optimal features extracted from the graph measure by Khazaee et al. ([20]).

The above-mentioned methods, employ classical machine learning methods, such as Support Vector Machines, k-nearest neighbor methods, or few layered perceptron. They feed hand-crafted feature vectors to the input of the classifiers for training and testing the algorithms.

As the deep learning methods emerged in the literature of Machine Learning, the researchers adopted some of these methods for learning a representation for cognitive tasks or neurological diseases, as summarized below.

2.2.1 Brain Decoding Using Deep Learning

Ability to extract important features made deep learning methods dominant in many fields, such as classification problems in computer vision. However, it is well known that deep learning methods need abundant amount of statistically sufficient data. Unfortunately, fMRI data falls short to satisfy the needs of deep architectures. Due to this problem, the methods, which employ the deep neural networks, can be applied to a very limited extent on brain decoding problems.

Initial efforts to apply deep learning methods on fMRI data showed only a limited success. Plis et al. ([28]) showed that deep belief networks was able to compete with the state-of-art methods without a need to select features manually for fMRI data. Sarraf et al. ([32]) used convolutional neural network architecture to detect Alzheimer disease.

The above-mentioned methods suffer from the curse of dimensionality problem of the fMRI data. Although there are techniques, such as ANOVA, to reduce the size of the feature space, the data augmentation methods is not enough to generate informative samples to increase the size of the dataset. An alternative to attack the curse of dimensionality problem may be to encode the data in a source dataset and transfer this information to a target dataset. Thus, one needs to develop algorithms which transfers information across the cognitive tasks of small datasets. This requires adopting the transfer learning methods to the fMRI data sets.

2.3 Transfer Learning on Brain Decoding

Transfer learning aims to improve the performance of a target task using the knowledge from either a different domain or task. [26]. It is especially useful in fields where collecting data is expensive or time-consuming, like fMRI. For this section, we will solely focus on transfer learning applications on fMRI data using deep learning models.

Fırat et al. ([13]) use stacked autoencoder to reduce dimension number of the input with unsupervised learning and fine-tuned the architecture with few labeled samples to use in classification problem.

Koyadama et al. ([22]) is able to train a deep learning architecture that can classify the cognitive tasks on the samples from multiple subjects on the HCP dataset. This study shows that the information contained by the brain activation patterns can be transferred between subjects by extracting subject-independent features.

Velioğlu et al. ([35]) use 3-layered stacked denoising autoencoder to extract latent representations of the cognitive tasks on the HCP dataset and use these latent representations for training on a relatively small dataset.

The above-mentioned methods do not consider the affinities among the cognitive tasks. They blindly try to transfer the information from a source task to a target task. However, it is known that transfer learning algorithms are only successful, when the distribution of a task in a source dataset is compatible to the tasks of the target dataset.

Therefore, it is crucial to establish affinities among the tasks of the source and target datasets.

In this thesis, we try to find an affinity graph among the cognitive tasks, recorded by an fMRI experiments.

2.4 Chapter Summary

This chapter overviews the necessary background, which is used to develop the suggested global transferability graph of this thesis.

We start by discussing the fMRI technology, which measures the BOLD response via magnetic changes. We summarize the data generation steps of fMRI data, which consists of the experimental procedure to record a predefined stimulus, acquisition, image construction and activation map generation.

Then, we introduce the brain decoding problem with the popular studies in the literature. We discuss the classical machine learning techniques for brain decoding. Considering the recent trends in machine learning methods, we study the deep learning applications for brain decoding. Finally, we present some of the previous studies on transfer learning methods on brain decoding, which attacks the curse of dimensionality problem.

CHAPTER 3

TRANSFERABILITY AND CONFUSION GRAPH GENERATION ON TASK FMRI DATASET

In this chapter, first we describe the nature of the Human Connectome Project (HCP) tfMRI dataset. We use this dataset for studying the transferability properties of the underlying cognitive tasks performed by the subjects during the fMRI recordings. Then, transferability graph generation on a between-subject setup based on the approach by Zamir et al. ([38]) is explained. The transferability graph is generated by estimating an affinity matrix, which denotes expected transferring performance between cognitive tasks recorded by fMRI data. In order to verify the validity of the estimated transferability graph, we also create another affinity matrix using classification performances of binary classifiers trained with imbalanced training data without using transfer learning, which we dub as confusion graphs. We compare the expected transferring performance relatively from the same source cognitive tasks in the transferability graphs and the confusion graphs in order to create a global transferability graph which defines a transferability policy possibly applicable to other datasets.

3.1 Human Connectome Project (HCP) Dataset

In this thesis, we employ a well-known de facto standard dataset, called HCP (Human Connectome Project) dataset. In this dataset, 808 subjects are asked to perform seven cognitive tasks during fMRI scanning [3]. These tasks are called **Emotion Processing, Gambling, Language Processing, Motor, Relational Processing, Social Cognition and Working Memory**. Explanations of these cognitive tasks are given below:

3.1.1 Emotion Processing

Adapted from Hariri et al. [17], subjects are tasked to match one of two faces at the bottom of the screen with the face at the top of the screen where the facial expressions are either angry or fearful, or one of the shapes at the bottom of the screen with the shape at the top of the screen.

3.1.2 Gambling

Using the same experimental setup with Delgado et al. [11], subjects are asked to guess if the number on a card is higher or lower than 5 for monetary gain as a reward and monetary loss as a punishment.

3.1.3 Language Processing

Adapted from Binder et al. [4], subjects are presented with either a story or a math exercise verbally and are asked a question and are asked to select one of two possible answers.

3.1.4 Motor

Adapted from Buckner et al. [5], subjects are asked to move a body part (foot, hand or tongue) in a predefined way with a visual cue.

3.1.5 Relational Processing

Adapted from Smith et al. [33], subjects are presented with objects with different shapes or textures. In the relational condition, subjects are asked to show the difference between the pair at the top of the screen, which is either texture or shape is same with the difference between the pair at the bottom of the screen. In the control condition, subjects are asked if the object at the bottom of the screen matches either one of

the objects at the top of the screen for the specified feature, which is either texture or shape.

3.1.6 Social Cognition

Subjects are presented with short clips created by Castelli et al. ([6]) and Wheatley et al. ([37]), which objects (squares, circles, triangles) are interacting in some way, or moving randomly. Then, the subjects are asked to choose between 3 possibilities which objects had a social interaction, not sure, or no interaction.

3.1.7 Working Memory

Subjects are asked to complete N-back tasks (two-back and zero-back) where stimuli are pictures of faces, places, tools and body parts (Barch et al. ([3])).

Number of fMRI scans and durations per each task are given in Table 3.1.

Table 3.1: Scans and Durations per Each Cognitive Task in HCP Dataset.

	Scans	Duration
Emotion Processing	176	2:16
Gambling	253	3:12
Language	316	3:57
Motor	284	3:34
Relational Processing	232	2:56
Social Cognition	274	3:27
Working Memory	405	5:01

The major goal of this thesis is to estimate the transferability relations between the above mentioned cognitive tasks.

3.2 Presentation of fMRI Data as a Set of Samples in a Feature Space

In order to decode the cognitive tasks of HCP dataset with the deep learning models, we apply several preprocessing operations, as explained below.

3.2.1 Reducing Feature Space Dimensions

There are about 150,000 voxels in a brain volume, recorded for each time instance of each subject in HCP dataset. Due to the computational constraints, even if we use singular time instances, training a deep learning model on a dataset with 150,000 dimensional feature space is surely infeasible. In order to solve this problem, we select the most informative voxels and take their average intensity values at each anatomic region ([1]).

3.2.1.1 Selecting Informative Voxels

Suppose that, fMRI signals recorded during a cognitive task session, consists of time series $v_i(t)$ at each voxel coordinate i to represent the neural activity of the underlying cognitive process task. Most of the measured voxels are not useful for discriminating the cognitive tasks. Thus, in order to select most important voxels, we use ANOVA. We call the selected voxels as $v'_i(t)$.

3.2.1.2 Taking Average Intensity Values By Region

In order to reduce the feature space further, we use Automatic Anatomical Labeling (AAL) ([34]) to partition the selected voxels according to their respective anatomic regions. Then, we represent each anatomic region by a time series, by taking the average of the selected voxel time series as follows:

$$X_r(t) = \frac{1}{n_r} \sum_{\forall v'_i \in r} v'_i(t), \tag{3.1}$$

where n_r is the number of selected voxels and X_r is the resulting averaged time series for region r. There are total of 116 anatomic regions in AAL atlas. However,

Neuroscientific literature shows that the regions in Cerebellum and Vermis do not contribute to the cognitive processes of HCP dataset. Thus, we omit these regions and obtain 90 relevant anatomic regions. Therefore, we define our brain volume at time t as the concatenation of these 90 regions at time t:

$$V(t) = [X_1(t), X_2(t), ..., X_{90}(t)].$$

Note that, $V(t)$ represents a brain volume by 90 dimensional vector, where each entry corresponds to the average fMRI intensity value of a region, r, at time t.

3.2.2 Extracting Temporal Information

In order to define a labeled feature vector, we concatenate several time instances to capture the dynamic variations at each cognitive task. Therefore, we define the labeled dataset of feature vectors each of which consists of $Tx90$ dimensional labeled feature vectors,

$$D = \{f, l_i\},$$

where$f$ is defined as the feature vector, which consists of the volume time series, concatenated with T consecutive time instances in the same subject where the time instances do not overlap between the feature vectors, as follows;

$$f = [V(t), V(t+1), ...V(t+T)]$$

and $l_i, \forall i = 1, ...7$, are the cognitive task labels, described in the previous section.

For our experiments we empirically found that $T = 25$ yields the best performance, compared to $T = 10$, $T = 15$ and $T = 20$ as expected, since longer time series is more likely to capture more temporal information.

3.2.3 Balancing Dataset

Scan number per cognitive task is not equal in HCP dataset. The longest cognitive task (Working Memory) has more than double scans compared to the shortest cogni-

tive task (Emotion Processing). In order to prevent data imbalance problem, we select equal number of samples during all the experiments.

3.3 Transferability Graph Generation

In order to create a transferability graph for the seven cognitive tasks of the tfMRI dataset, we adopted the method suggested by Zamir et al. ([38]). By finding a latent representation for every source cognitive tasks and measuring the transferability between every source and target pair by using these latent representations, we create the transferability graph which is represented by an affinity matrix. Block diagram of the whole procedure given in Figure 3.1.

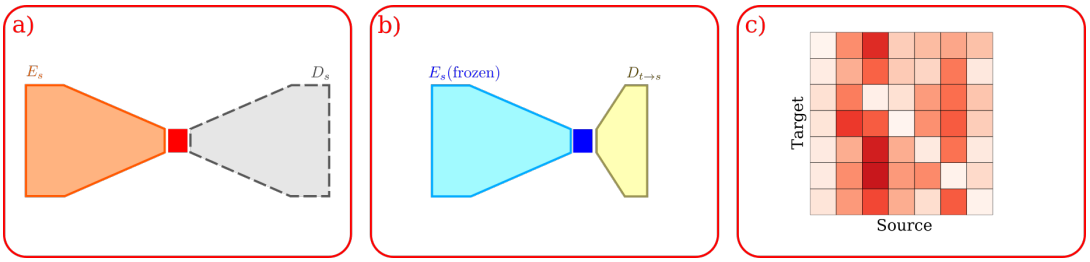

Figure 3.1: Steps to Generate Transferability Graph. There are three main steps to generate transferability graph: a)Learning a latent representation for every source cognitive task using deep denoising autoencoders. b) Measuring Transferability between every source and target by training a shallow decoder on the source encoder with frozen weights. c) Creating an affinity matrix using comparative performance of source tasks for the same target task.

In the following subsections, we explain each step of the transferability graph generation method introduced in this study.

3.4 Learning Latent Representation for Source Tasks

The method we use is based on the assumption that a deep learning model can learn higher level features of the input in an unsupervised fashion. In order to find a latent

representation for the source cognitive tasks we utilize deep denoising autoencoders. By shrinking sizes of fully connected layers gradually to create a bottleneck in the encoder part and by trying to reconstruct the input from the encoder output by expanding layer sizes gradually in the decoder part, deep autoencoders can learn a latent representation end-to-end. Due to the noisy nature of tfMRI data, by adding a static noise only to the input and using original input as the output, we train a denoising autoencoder for each of the seven cognitive tasks:

$$arg \min_{\theta_E, \theta_D} \mathbb{E}_{f \in \mathcal{S}}[L_2(D_s(E_s(\ltimes(f); \theta_E); \theta_D), f)] \tag{3.2}$$

where E_s , D_s and $\mathcal{S}$ are encoder, decoder and dataset for the source cognitive task s respectively, and $\ltimes$ is additive Gaussian noise $\mathcal{N}(0, 3)$. We have several reasons to use this architecture, instead of some architectures such as recurrent neural networks, or generational models like variational autoencoders, which are more convenient for modeling multivariate time series such as fMRI data:

- It is well known that the deep neural network architectures for modeling a string of symbols, are more resistant to problems, such as, vanishing gradient or posterior collapse.

- Since it involves unsupervised learning, autoencoders are relatively easy and stable to train with relatively smaller number of hyperparameters, compared to the deep neural network architectures based on supervised learning paradigm.

Given the number of networks needed to be trained, second part is especially important for experimentation.

Therefore, instead of using network models for time series representation, we prefer to use a set of independent fully connected autoencoders to learn a representation for each source cognitive task. General structure of source autoencoders given in Figure 3.2.

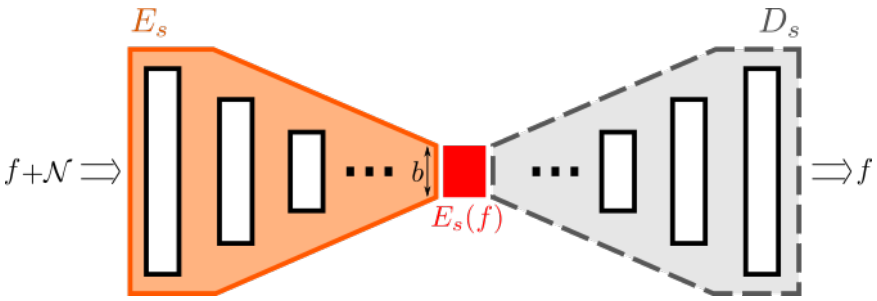

Figure 3.2: Source Autoencoder Architecture. This architecture finds a latent space representation ($E_s(f)$) of source task s. Every fully connected layer is followed by a batch normalization layer and a LRELU layer. These three layers are represented as a single white block for visual clarity. Block size is usually half of its predecessor to approximate wanted output size of encoder b. Added noise $\mathcal{N}$ and batch normalization layers help for more robust latent representation of f, $E_s(f)$.

In the encoder, size of a fully connected layer is usually selected as the half of the size of its predecessor until the selected encoder output size b is reached. Empirically we selected b as 180.

For activation functions leaky rectified unit (LRELU) is used with negative slope of 0.05 to prevent vanishing gradient and dead neuron problems.

Apart from noise, another issue for between-subject experiments are difference in distributions of samples from different subjects due to physiologic differences. To make the networks more robust against this problem, we used batch normalization after every fully connected layer before activation function.

To prevent overfitting, every autoencoder is trained for 300 epochs and weights with minimum test loss are selected to be used through rest of the experiments. We select epoch number large enough to observe overfitting with other hyperparameter values like smaller encoder output size. For training,batch size is about 10th of the training dataset in order to utilize batch normalization better.

Adamax ([21]) is selected as optimizer experimentally.

5-fold cross validation is used for splitting training and testing data.

Our final autoencoder architecture is with input size 2250 which is the concatenation of $T = 25$ time instances with 90 averaged regions, and fully connected layers with the output sizes 1125, 562, 281 and 180 in the encoder. The decoder is symmetric with the encoder. Output of the encoder part of the trained autoencoder is the latent representation of the cognitive task.

3.4.1 Measuring Transferability Across Cognitive Tasks

Our assumption is that, in order to transfer the learning capabilities from a source cognitive task to a target cognitive task, these cognitive tasks should share a common latent representation. In order to measure the transferring capabilities from a source cognitive task to a target cognitive task, we train a decoder which can reconstruct the target cognitive task from the output of the encoder part of the source autoencoder.

What we are trying to measure with the transfer autoencoder architecture is the reconstruction performance of the target task from the latent representation of the source task. Noise is not added to input in order to make autoencoder more robust to the inherent noise embedded in the dataset in the latent representation. Also, in the transfer autoencoders,we only train the decoder part. Furthermore, by using the decoder part as a simple read-out function by keeping it shallow, we ensure that output is constructed from the information extracted from the source task. So for transferring, we optimize:

$$arg\min_{\theta} \mathbb{E}_{f\in\mathcal{T}}[L_2(D_{s\to t}(E_s(f);\theta), f)], \tag{3.3}$$

where $\mathcal{T}$ is the dataset of cognitive target task t and $D_{s\to t}$ is the transferring decoder for every source s and target t pairs.

Like the source encoders, the transfer autoencoders are trained for 300 epochs with Adamax optimizer, and minimum test loss weights are used to prevent overfitting and for training,batch size is about 10th of the training dataset.

Testing dataset of the source autoencoder dataset of the same cognitive task is used as the target dataset where %50 is used for training and %50 for testing. For the scope of this thesis, we only limited transfers from a single source cognitive task.

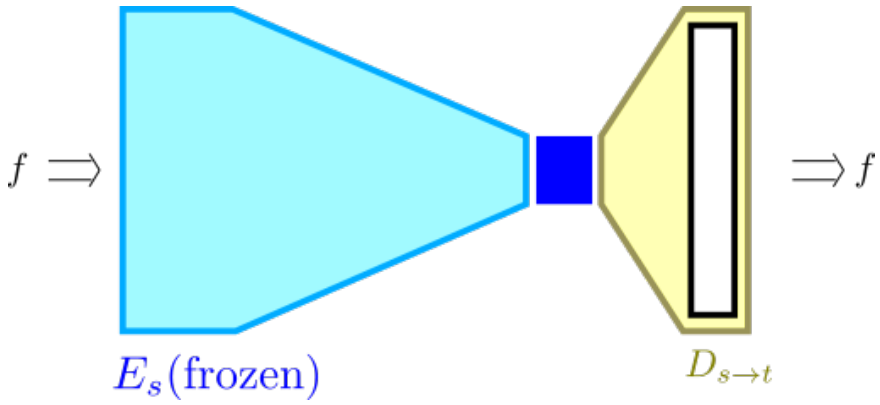

Figure 3.3: Target Autoencoder Architecture. Compared to the source autoencoder, in the transfer autoencoders, only the decoder part is trained. The encoder part is not trained and the weights from the encoder part of the source autoencoder is used. Also, the decoder consists of one fully connected layer followed by batch normalization and LRELU layers (the white block) because a shallow network is necessary to ensure learning is utilizing the source representation.

3.4.2 Creating Transferabilty Graphs

In order to create an affinity matrix, we use the performance of transfer models given in Eq.3.3 to estimate the degree of affinity between each pair of source and target cognitive tasks. For this purpose, we somehow need to normalize the performance of each model which is trained with a cognitive task and tested by some other cognitive task.

A simple normalization between cumulative errors would not suffice since different tasks have different metrics. Even for our problem, where samples come from the same type of dataset and the models use the same loss function, due to the difference of distributions of the cognitive tasks, this normalization cannot be accurate to measure the affinity between the tasks. The method suggested by Zamir et al. [38], tries to mitigate this problem by comparing network performances per sample basis.

For a rigorous normalization method, a score matrix W_t is defined for target t, where its $(i,j)^{th}$ element is:

$$w_{i,j} = \mathbb{E}_{f \in D_{test}}[T_{s_i \to t}(f) > T_{s_j \to t}(f)] \tag{3.4}$$

which defines the percentage of samples on the test dataset which have better reconstruction performance with the transfer autoencoder from source s_i to target t, ($T_{s_i \to t}$) than the transfer autoencoder from source s_j to target t ($T_{s_j \to t}$).

After clipping the entries of the matrix W_t into [0.001,0.999], we calculate W'_t which tells us s_i,s_j performance comparison which its element at (i,j) is:

$$w'_{i,j} = \frac{w_{i,j}}{w_{j,i}} = \frac{\mathbb{E}_{f \in D_{test}}[T_{s_i \to t}(f) > T_{s_j \to t}(f)]}{\mathbb{E}_{f \in D_{test}}[T_{s_j \to t}(f) > T_{s_i \to t}(f)]} \tag{3.5}$$

In order calculate our transferability graph P, we calculate the eigenvector of W'_t for every target cognitive task t and normalize the sum of the elements of the eigenvector to 1. We define $(i,j)^{th}$ element of P as:

$$p_{i,j} = e_{ij} \tag{3.6}$$

where e_i is the normalized eigenvector of W'_i and $p_{i,j}$ is the expected transfer performance from source cognitive task j to target cognitive task i . This process is derived from Analytic Hierarchy Process ([30]).

3.5 Confusion Graph Generation

How can we analyze the validity of the extracted transferability matrix? Our answer to this question is creating another affinity matrix without using transfer learning paradigm. For this purpose, we train binary classifiers with imbalanced data and test with balanced data. Our assumption is that in this configuration, similarity between two cognitive tasks directly impacts to the performance of the classifier.

To train a classifier without transfer learning, we simply optimize the following cost function:

$$arg \min_{\theta} \mathbb{E}_{f \in \mathcal{F}_{s \to t}}[\mathbb{C}(C_{s \to t}(f), l_f)], \tag{3.7}$$

where s and t represents a cognitive task pair, $\mathcal{F}_{s \to t}$ is the dataset consists of samples of s and t, l_f is the label of the sample, $\mathbb{C}$ is cross entropy loss and $C_{s \to t}$ is the binary classifier trained with imbalanced dataset.

5-fold cross validation is used for splitting training and testing data. For training, the ratio between the number of samples from s to the number of samples from t is 5. For testing, the number of samples from both cognitive tasks are equal.

Like the source and transfer autoencoders, the binary classifiers consist of fully connected layers where every connected layer followed by a batch normalization layer and a LRELU layer.

For the output layer of the binary classifiers, additional softmax layer is used. During our experiments, we observe that keeping LRELU before the softmax layer after the last fully connected layer, both improves the performance and the stability of the network for this configuration.

Every binary classifier is trained for 1000 epochs, and weights with minimum test loss are used to generate confusion graph. In the experiments we observe that most of the binary classifiers tend to overfit after 800 epochs.

Input sizes of the fully connected layers are: 2250, 200, 50 and 10.

Lastly, we define (i,j)th element of confusion matrix as the expected performance of $C_{j \to i}$ on the test dataset.

3.6 Global Transferability Graph Generation

Recall that our goal is to find the generic affinities between the cognitive tasks which is valid for all of the subjects in the HCP dataset. Unfortunately the method, suggested by Zamir et al. ([38]) fails to extract consistent set of affinity matrices, across the folds. This failure can be attributed to the nature of the HCP dataset, which consists of multiple subjects, who perform the same set of cognitive tasks. The great variations among the fMRI data which belong to different subjects prohibits us to estimate a robust set of affinities across the subjects.

In order to adapt the method of [38] to between-subject setup, we suggest a new approach which exposes relatively stable connections from the same source cognitive task between folds. This approach defines a generic graph, called global transferabil-

ity graph.

The global transferability graph is estimated by selecting a set of candidate connections from several transferability graphs, obtained from multiple folds and use them as a vote of confidence.

More specifically, we normalize all the weights to the interval [0,1] from the same source cognitive task excluding the self-transferring weights for every fold, prior to the selection of the candidate connections. This normalization process brings all the folds into the same scale. As the next step, we select the largest K-connections as the candidate connections. The threshold value, K is very crucial to eliminate the cognitive tasks of relatively small affinities. We find that the optimal value of $K = 3$ provides us the most relevant cognitive tasks, which are consistent with neuroscientific findings, available in the literature. However, this threshold value can be adjusted by using some other criteria, such as the consideration of the active brain regions relevant to a specific cognitive task, depending on the characteristics of the fMRI datasets, other than that of the HCP dataset.

We also observed that due to internal and external noise sources embedded in the fMRI data, the difference between two connections of the graph affinity matrix might be too small to be excluded when the threshold value K is fixed. This observation reveals that the $K = 3$ value can not be an absolute threshold to identify the high affinity cognitive task. Therefore, we need to also consider the difference in the affinity values between two consecutive connections in the sorted list of affinities.

Once we form the sorted list of the connections, in the affinity graph, we check for the distance between the connections. If the difference between a remaining connection and a selected connection is less than a threshold value, T, then, we also include these connections to the selected affinity list. This threshold value determines the number of selected affinities. Thus, resulting a varying number of affinities for each cognitive task. We observe that $T = 0.05$ provides us relatively stable value for selecting the cognitive tasks, which are " closest" to each cognitive task, with respect to the graph affinity connections.

The above approach, prunes the transferability graphs obtained from multiple folds

and generates a global transferability graph by including a candidate connection if and only if the same candidate connection is both suggested by a number of transferability graphs and the average confusion graph.

3.7 Summary

In this chapter, first we give a brief information about the nature of the HCP dataset, which we use to study the affinities among the cognitive tasks. Then, we discuss about the dimension reduction problem of this dataset and methods to mitigate this problem.

Finally, we introduce a new method to create a global transferability graph, which is adopted by post processing the method suggested by Zamir et al. ([38]). This computational method enables us to find a generic representation of the affinities among the cognitive tasks in a dataset, which consists of multiple subjects.

CHAPTER 4

EXPERIMENTAL RESULTS

In this chapter, we present the experimental results of the graph affinity matrices, estimated by the methods proposed in Chapter 3. We tested the validity of the suggested computational model on the HCP data set, which is recorded while the subjects perform seven cognitive tasks. Recall that, in HCP dataset, 808 subjects are asked to perform seven cognitive tasks during fMRI scanning [3]. These tasks are called **Emotion Processing, Gambling, Language Processing, Motor, Relational Processing, Social Cognition and Working Memory**. In this section, first, we estimate the affinities among these tasks, in which a source task can be transferred to another task in the target. Then, we analyse the validity of the proposed method by comparing this computational model with the evidences obtained from the experimental neuroscience.

In order to find the degree of transferability of information among these seven tasks, we define two separate affinity matrices, which represent transferability graph and confusion graph. We analyze the task affinities by comparing and analyzing the consistencies of these graphs.

The transferability graphs are generated using the method adapted from Zamir et al. ([38]) on a between-subject setup, which uses transfer learning performance between cognitive tasks in order to calculate the affinity between them. We generate the confusion graphs using the classification performances of binary classifiers trained with imbalanced data without transfer learning. Using these two types of affinity matrices, we generate a global transferability graph which defines the possible relations between the cognitive tasks that can be utilized by transfer learning in a different datasets than it generated. We also briefly discuss the effect of hyperparameter values on the affinity matrix generation algorithms.

4.1 Transferability Graph Generation

Recall that the first step to measure transferability performances between the cognitive tasks is to find latent representations of the cognitive tasks in HCP Dataset by training a denoising autoencoder per cognitive task, as suggested in Chapter 3.

When training the source autoencoders, about 4520 samples are used for training and about 1130 samples are used for testing. A static Gaussian noise, $\mathcal{N}(0, 3)$, is added to make the latent representation more robust against to the inherent noise in the dataset.As an example, train and test losses vs epoch graph for emotion processing cognitive task for the first fold can be seen in Figure 4.1.

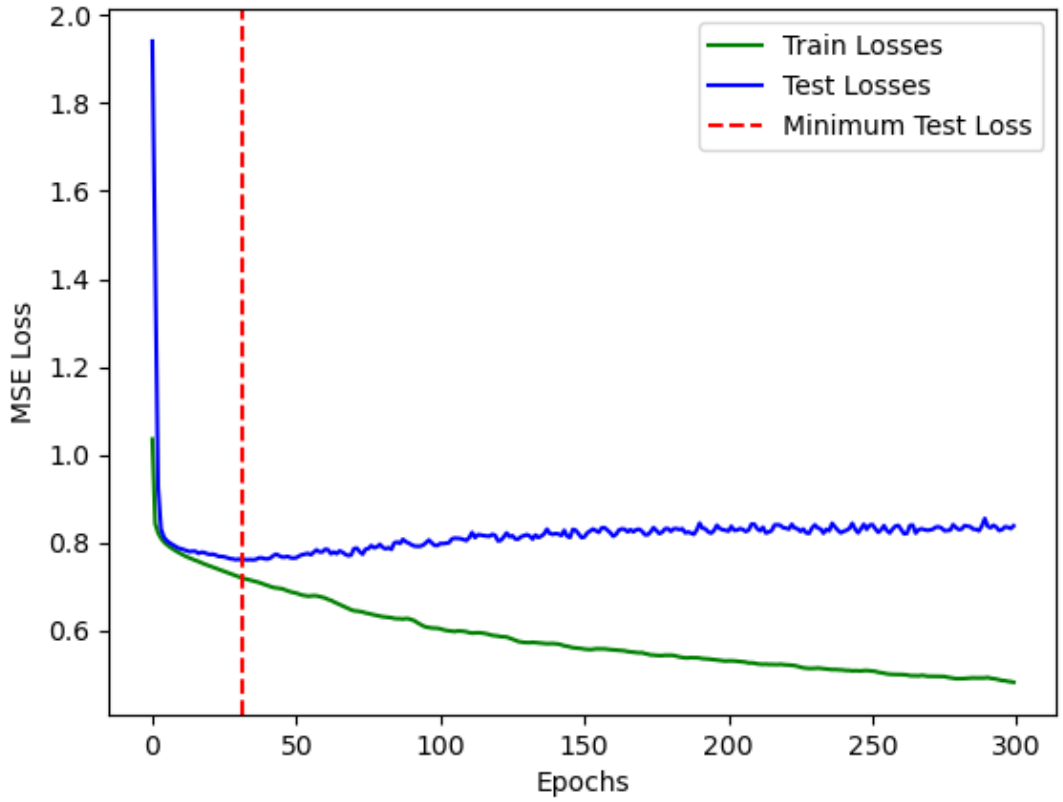

Figure 4.1: Train and Test Losses vs Epoch Graph for the Source Autoencoder for Emotion Processing Cognitive Task. Like other 6 source autoencoders, the minimum test loss is achieved before 60 epochs for all folds.

The loss plots of all 7 cognitive tasks have similar profile for all folds and starts to overfit less than 60 epochs without an exception. In order to prevent the overfitting

and to use optimal latent representation of the source cognitive tasks, instead of training every source autoencoder with equal number of epochs, we train the networks until the test loss is minimum for the source autoencoders. In order to measure the transfer learning performance between source and target cognitive task pairs, we train a transfer autoencoder for each pair of source cognitive task, s, and target cognitive task, t.

For a transfer autoencoder from source cognitive task s to target cognitive task t, the encoder part of the source autoencoder trained on samples from cognitive task s are reused as the encoder without retraining, and only the shallow decoder part which serves as a simple read-out function is trained. For training and testing the transfer autoencoder, same dataset for testing the source autoencoder for cognitive task t is used. Half of the data is used for training which is about 560 samples and the other half is used for testing. As an example, the loss plots for the transfer autoencoder from emotion processing as the source task and gambling as the target task for the first fold is shown in Figure 4.2.

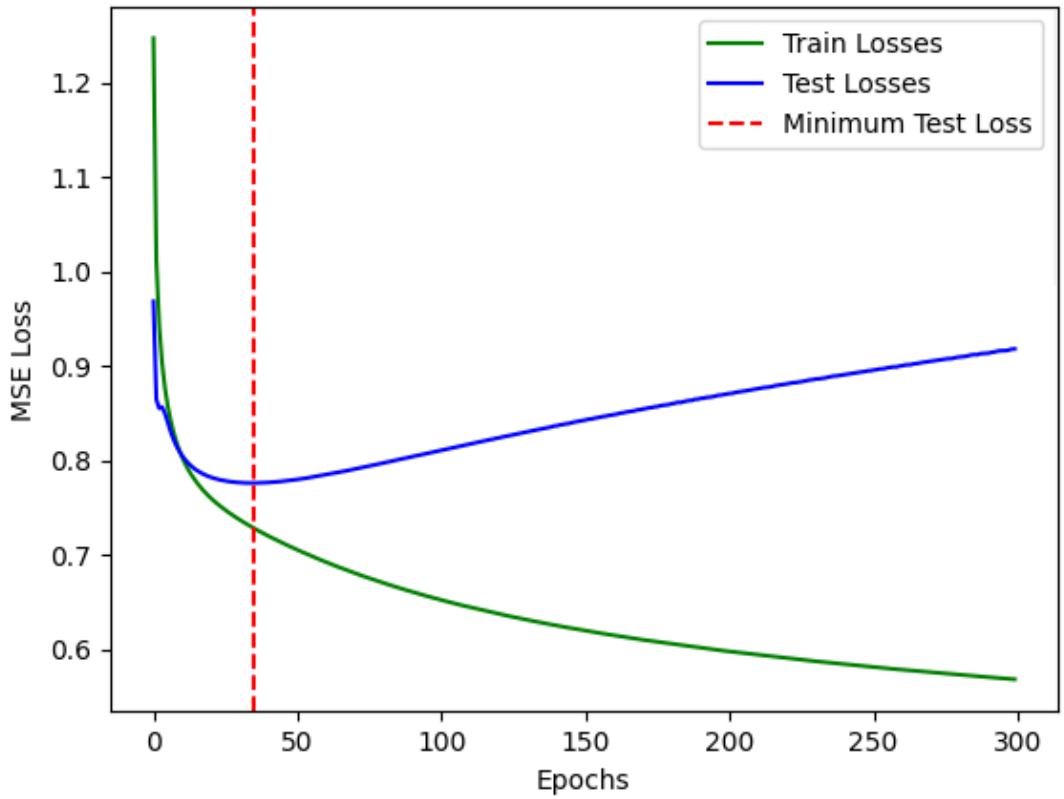

Figure 4.2: Train and Test Loss vs Epoch Graph for Transfer Autoencoder from Emotion Processing Cognitive Task to Gambling Cognitive Task. Like the source autoencoders, the transfer autoencoders begin to overfit even before 60 epochs.

As can be seen in Figure 4.2, the transfer autoencoders start to overfit in very small epochs similar to the source autoencoders. To prevent overfitting interfere with the transferability performances, minimum test loss weights are used when generating transferability graphs.

For every fold, a source autoencoder are trained for each cognitive task. Considering the fact that, we have 7 cognitive tasks, we train 7 autoencoders and 7x7= 49 transfer autoencoders.

Trained transfer autoencoders are used for generating a transferabilty graph using the algorithm explained in Section 3.4.2. For every target task, a score matrix, W_t for target cognitive task t, is defined which its elements are the expected transference performance of source cognitive task, compared to another source cognitive task (Eq. 3.4). W'_t defines the relative comparisons of two source cognitive tasks and calculated by dividing clipped W_t by the transpose of itself element-wise (Eq. 3.5). The resulting affinity matrix, which defines the transferability graph is generated by concatenating the normalized eigenvectors of all score matrices for every target cognitive task (Eq. 3.6).

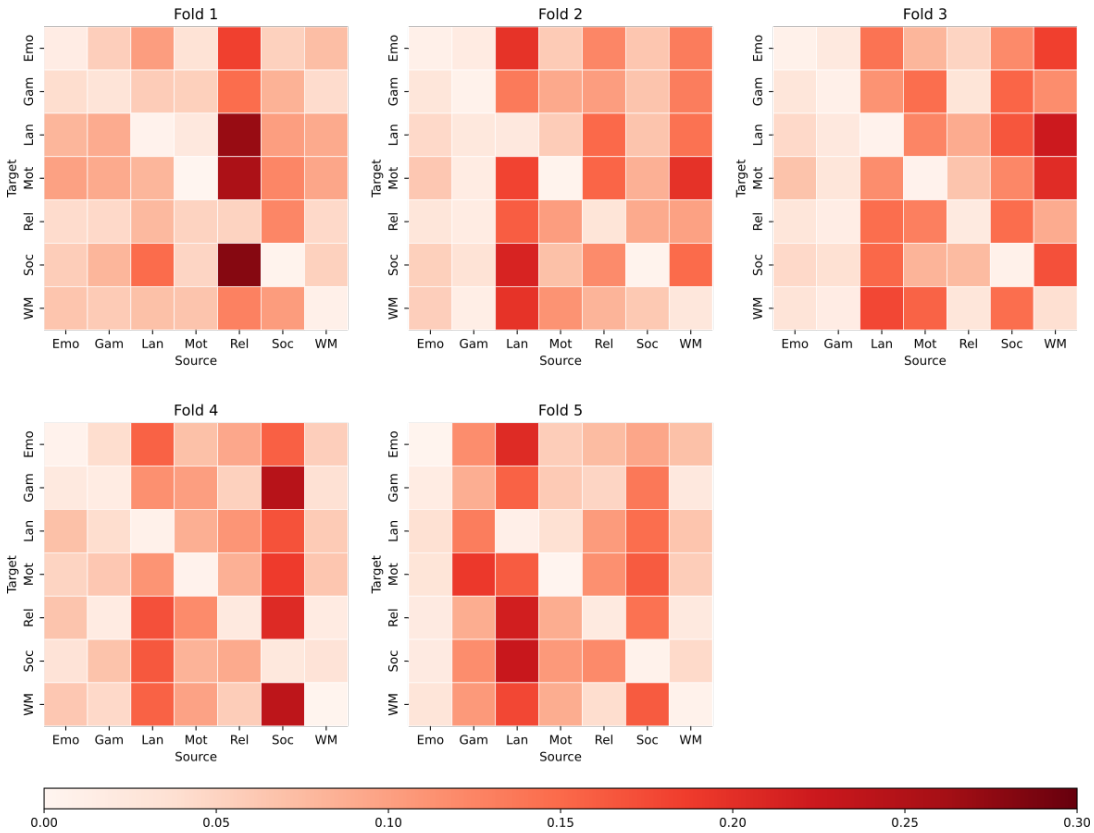

Figure 4.3: Generated Transferability Graphs per Fold. For visual clarity, the distance formula e^{-20x} is used for every cell x. Lower values mean better expected transferability. Note that apart from using the same cognitive task as both source and target having the best expected transferability, there is no common structure between folds. Although general overlap between resulting transferability graphs are sparse, it can be seen that relative expected transferring performance from the same source have similar characteristics throughout the folds.

We generate the transferability graphs for 5 different folds, each of which consists of multiple subjects performing the same set of cognitive tasks. In other words, we form the training set for each fold by randomly mixing the samples from each subject for all the tasks. This method enables us to see the variations of the affinity matrices among the folds.

The estimated affinity matrix for each fold is shown in Figure 4.3. Note that the transferability graphs for different folds does not share a common structure, apart from diagonals having the highest transference performance as expected, due to the problems arising from between-subject setup. In other words the high affinities in one fold may have very low affinities in another fold. Thus, it is not possible to find important affinities between the tasks which are valid for all folds.

The above observation shows that creating a global transferability graph using the method by Zamir et al. ([38]), is not possible. However, one interesting feature demonstrated by these graphs, even though transfer graphs are dissimilar, the relative performances from the same source cognitive task are relatively stable throughout the folds. In order to create global transferability graph, we expose this relative stability.

4.2 Confusion Graph Generation

Similar to the transferability graphs, the confusion graphs are generated using the performances of the neural network models in order to measure transferability performance. However, instead of autoencoders, binary classifiers are trained with imbalanced data. Same folds with the source autoencoders are used.

Our assumption is that in an imbalanced data situation if two cognitive tasks are similar, abundant number of samples from one cognitive task can compensate for scarce samples of other cognitive task. For this purpose, we create training datasets which uses all samples from the first task (about 4500 samples), and 1/5 of the samples from the second task (about 900 samples). Same number of samples are used (about 1130 for each) from both cognitive tasks for testing dataset. 42 models are trained for every fold. Cross entropy loss and performance vs. epoch plots for emotion processing as the first cognitive task and Language as the second cognitive task can be seen Figure

4.4.

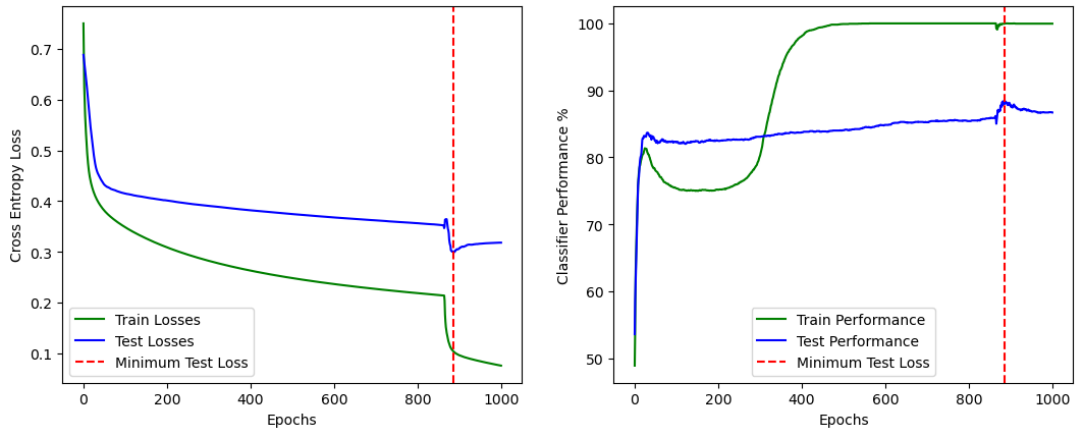

Figure 4.4: Train and Test Losses vs Epoch Graph (Top) and Train and Test Performances (Bottom) vs Epoch Graph for Binary Classifier as Emotion Processing as the First Cognitive Task and Language as the Second Cognitive Binary Classifier. Most of the trained classifiers have similar graph and tend to overfit above 800.

Classifiers are trained longer, since most of the classifiers tend to over fit about 800-900 epochs. As in the estimation of the transfer graphs, minimum test loss weights are used, when generating the confusion graphs.

For Confusion graph generation, we use the performance of the trained classifiers on the test datasets. Resulting confusion graphs can be seen in Figure 4.5.

For the same fold, the same samples are used when training and testing the source autoencoders, while training and testing the binary classifiers. However, unlike the transferability graphs, the confusion graphs are very stable across folds. This observation shows that for a multi-subject setup, the suggested method is very sensitive to distribution of data and other hyperparameters. Since the confusion graphs are very similar for different folds, when generating the global transferability graph, we use a single confusion graph which is the average of the confusion graphs from all folds. In other words, $(i,j)^{th}$ element of the confusion graph is defined by the following equation:

$$\bar{b}_{i,j} = \frac{1}{k}\sum_{n=1}^{k} b_{n(i,j)}. \tag{4.1}$$

where k is the fold number which is 5 for our experiments, $b_{n(i,j)}$ is the $(i,j)^{th}$ element

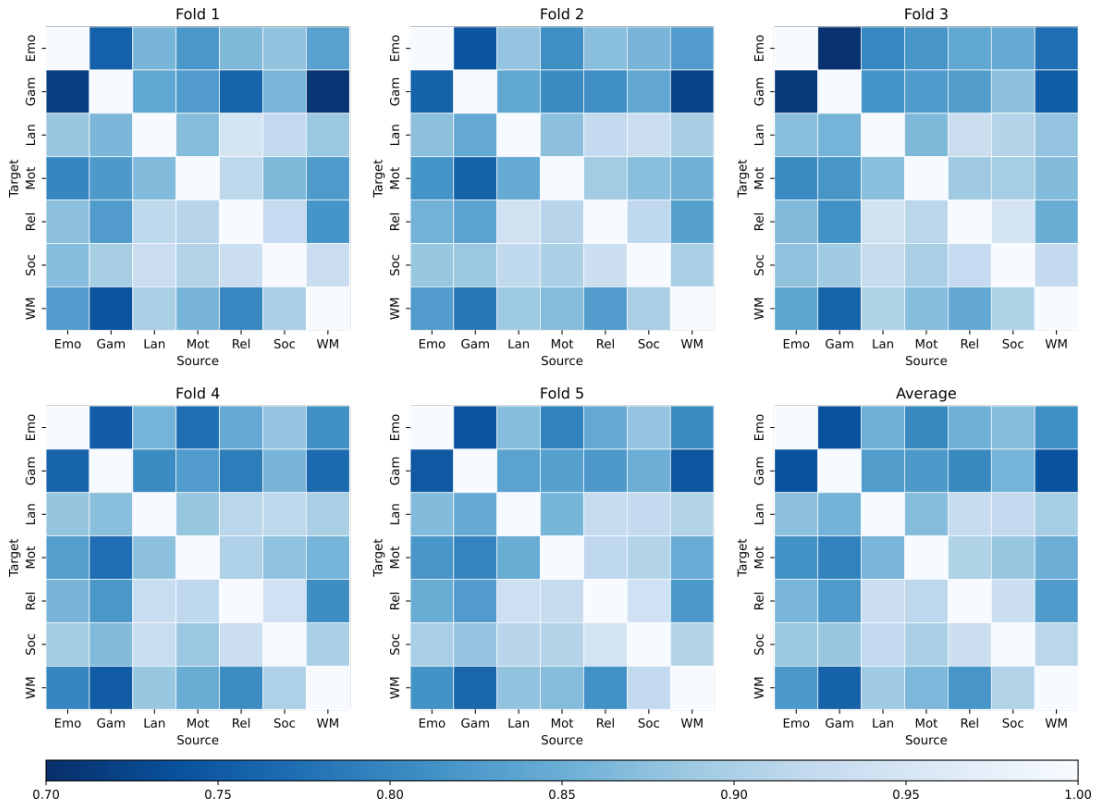

Figure 4.5: Generated Confusion Graphs per Fold and the Averaged Confusion Graph. Larger values means higher expected transferability between the cognitive tasks. Unlike the transferability graphs, the confusion graphs are very stable throughout the folds and tend to find same connections. Since first and second tasks are have similar relation as source and target as the transferability graphs, we label first tasks as Source and second tasks as Target respectively.

of the confusion matrix generated from n^{th} fold.

4.3 Global Transferability Graph Generation

As can be seen in Figure 4.3, the adapted method fails to find a common structure on a between-subject fMRI dataset due to the high variance. At the same time high variance is not a problem for the classifiers as can be seen in Figure 4.5. Although, the general structure is absent between the transferability graphs, expected transference performance from the same source cognitive task to different target cognitive tasks have similar orderings in multiple transferability graphs generated using differ-

ent folds.

In order to utilize this ordering and eliminate possible weak connections, we use the following algorithm:

- For the i^{th} cognitive task, select the i^{th} column from the affinity matrix which represents connections where their weights represent the expected transfer performance from i as the source cognitive task.
- Remove the self-transferring performance which is the i^{th} element from the selected column in order to make comparison of others easier.
- Normalize the remaining values into [0,1] interval.
- Select any connection as a candidate connection bigger than $a - t$, where a is the 3^{rd} biggest number in the normalized values and t is a threshold to prevent numerical errors to eliminate important candidate connections which we selected empirically as $t = 0.05$.

We apply this algorithm to every transferability graph estimated from different folds and the averaged confusion graph. Normalized relative transferability per source cognitive task charts are given in Figure 4.6, 4.7, 4.8, 4.9, 4.10, 4.11 and 4.12. Green bars show the candidate connections suggested by the transferability graph on that fold and yellow bars shows the candidate connections suggested by the average confusion graph. The candidate connections denotes possible transferring to that target cognitive task from the given source cognitive task.

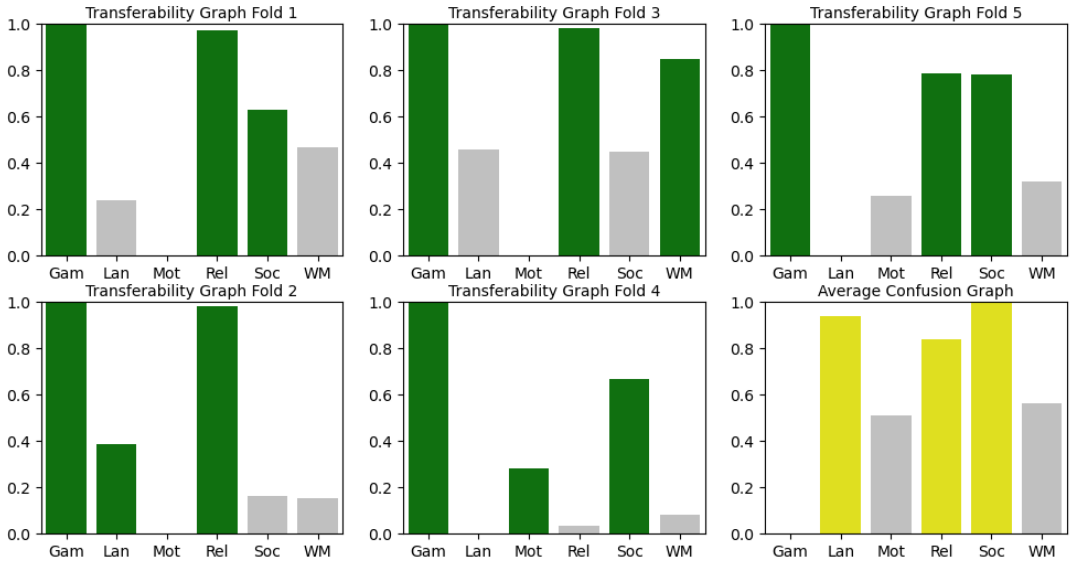

Figure 4.6: Normalized Relative Expected Transference Performance for Emotion Processing Source Cognitive Task. The strongest candidate connection suggested by all of the transferability graphs is gambling cognitive task. Interestingly gambling cognitive task has the lowest expected transference performance among the average confusion graph connections. Other candidate connections suggested by 4 and 3 transferability graphs respectively are relational processing and social cognition cognitive tasks which both are suggested by the average confusion graph. Language cognitive task on the other hand, which is the last connection suggested by average confusion graph, is only suggested by one transferability graph.

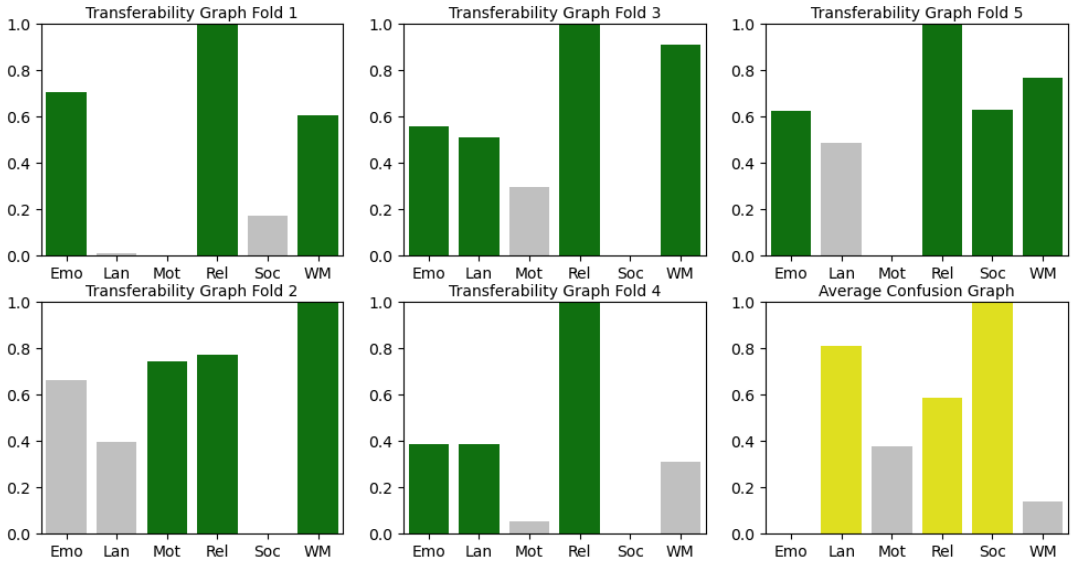

Figure 4.7: Normalized Relative Expected Transference Performance for Gambling Source Cognitive Task. Relational processing cognitive task is both suggested by all of the transfer graphs and the average confusion graph. Emotion processing and working memory cognitive tasks on the other hand, although suggested by 4 transferabilty graphs, have the least relative expected performance in the average confusion graph. Lastly language cognitive task is both suggested by the average confusion graph, and 2 transferability graphs.

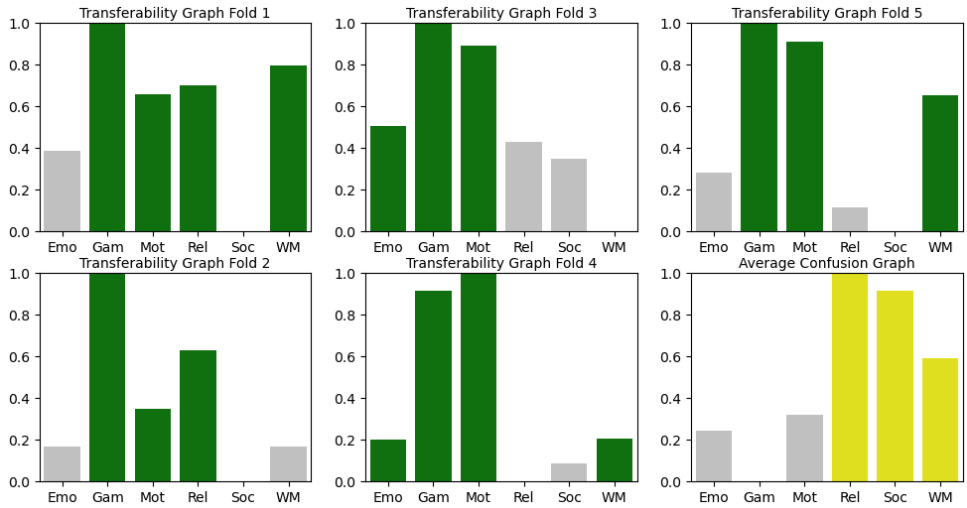

Figure 4.8: Normalized Relative Expected Transference Performance for Language Source Cognitive Task. Gambling and motor cognitive tasks are suggested by all 5 transferability graphs, but none of them are suggested by the average confusion graph. Working memory cognitive task on the other hand, is suggested by 3 transferability graphs and the average confusion graph. Emotion processing cognitive task is suggested with only two folds and is not considered by the average confusion graph. Relational processing cognitive task, on the other hand, although suggested by only 2 transferability graphs, is the strongest suggested connection in the average confusion graph.Social cognition cognitive task is the last candidate connection suggested by the average confusion graph and is not suggested by the transferability graphs.

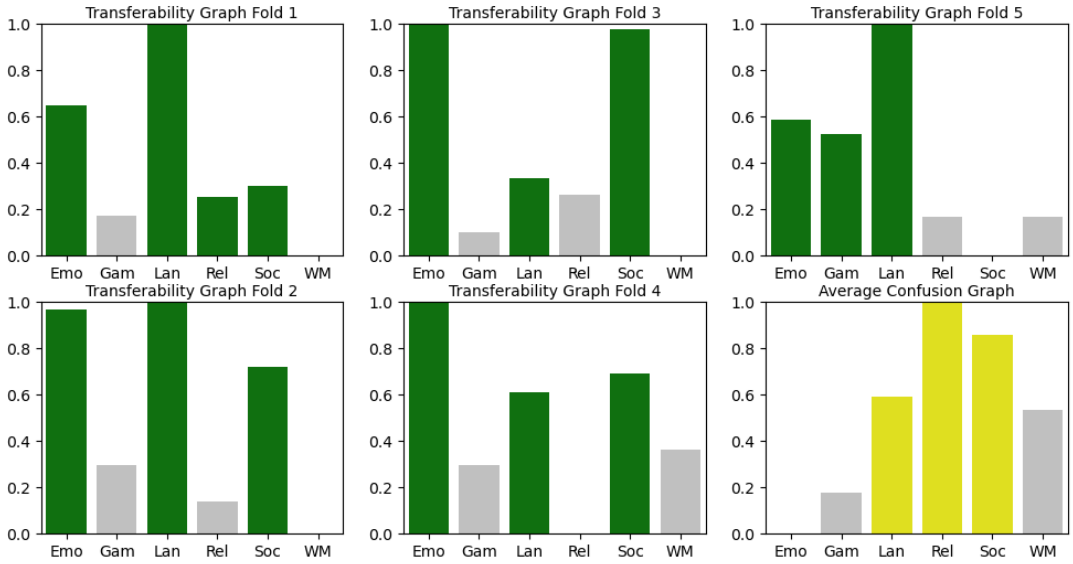

Figure 4.9: Normalized Relative Expected Transference Performance for Motor Source Cognitive Task. 2 out of 3 connections suggested by the average confusion graph, language and social cognition cognitive tasks, are also suggested by 5 and 4 transferability graphs respectively. Relational processing cognitive task on the other hand is only suggested by single transferability graph, hence is not a candidate connection. Another connection suggested by all of the transferability graphs is emotion processing cognitive task has the least expected transference performance in average confusion graph.

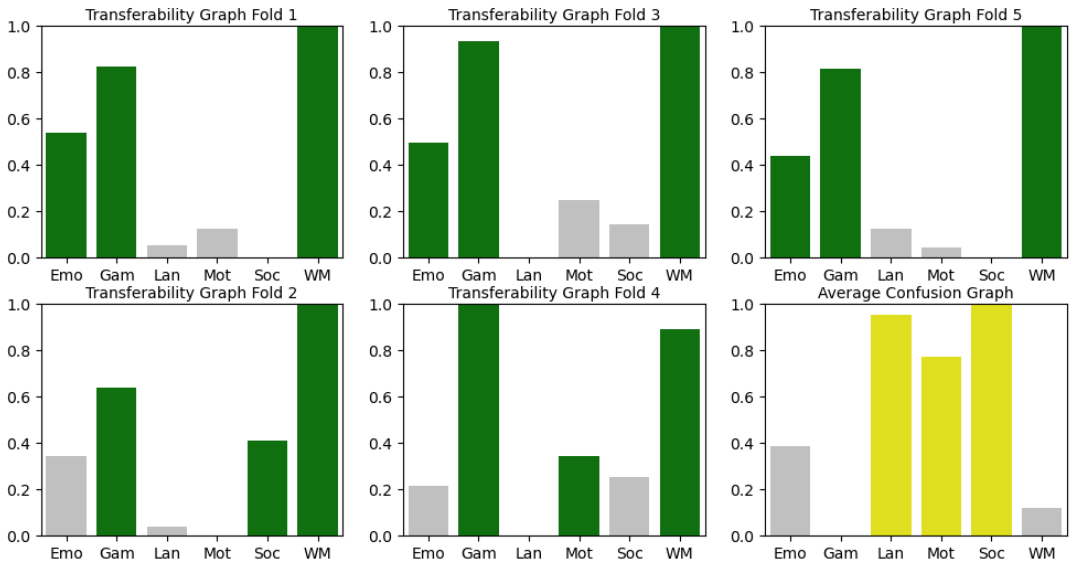

Figure 4.10: Normalized Relative Expected Transference Performance for Relational Processing Source Cognitive Task. The connections suggested by the average confusion matrix, is suggested by at most one transferability graph. One interesting observation is that, gambling and working memory cognitive tasks which are suggested by 5 transferability graphs have the smallest expected transferability performance in the average confusion graph followed by emotion processing cognitive task which is suggested by 3 transferability graphs.

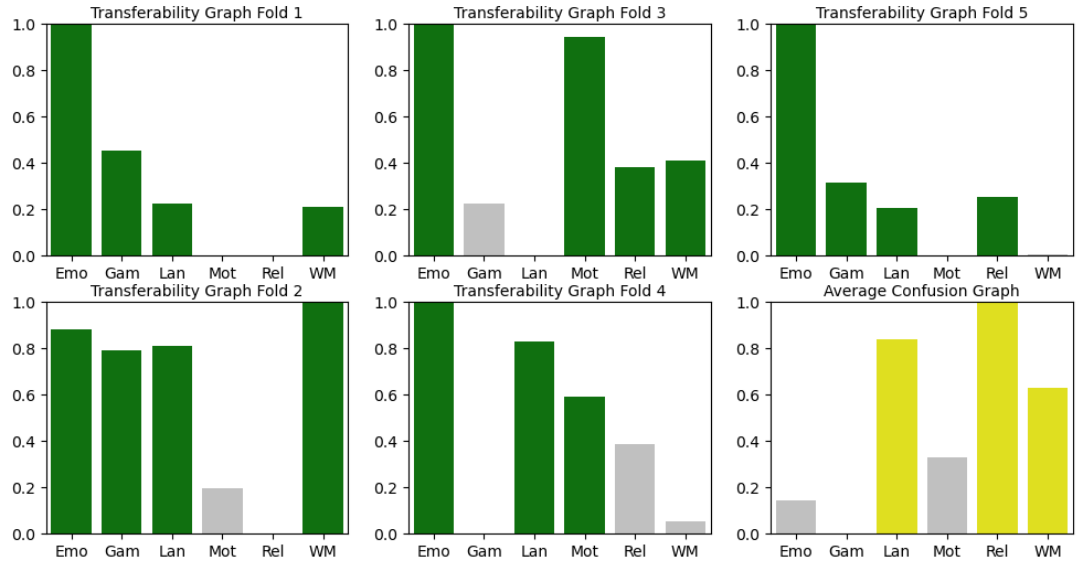

Figure 4.11: Normalized Relative Expected Transference Performance for Social Cognition Source Cognitive Task. Language and working memory cognitive tasks are suggested by the average confusion graph and 4 and 3 transferability graphs respectively. The other connection suggested by the average confusion graph is also suggested by the transferability graphs albeit only by 2. Emotion processing, gambling and motor connection cognitive tasks are also suggested by 5, 3 and 2 transferability graphs, are not suggested by the average confusion graph.

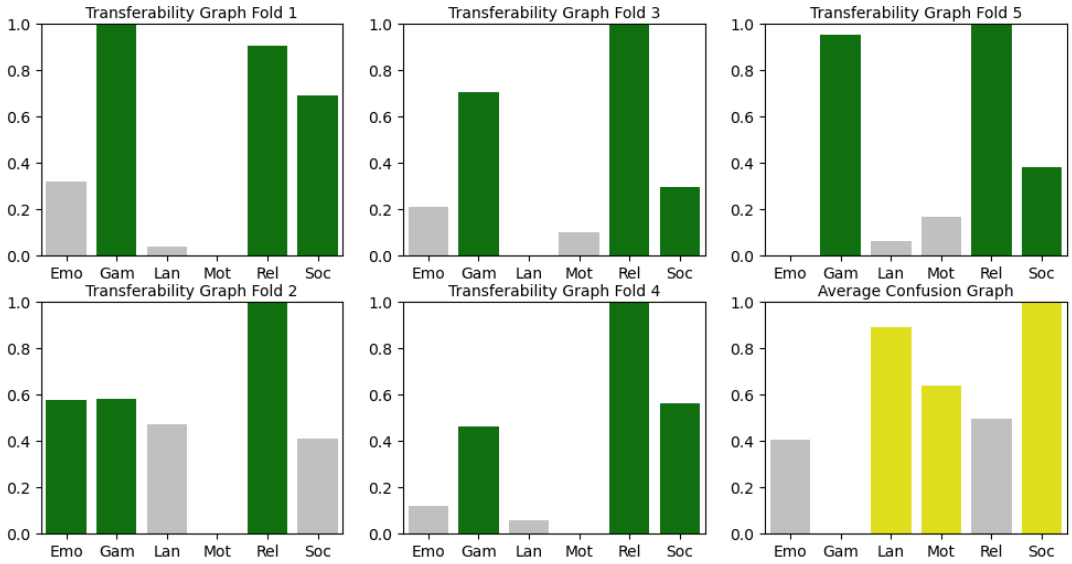

Figure 4.12: Normalized Relative Expected Transference Performance for Working Memory Source Cognitive Task. Although gambling and relational processing cognitive tasks are suggested by all of transferability graphs, none of them is suggested by the average confusion graph. Also language and motor cognitive tasks suggested by the average confusion graph is not suggested by any transferability graph. Social cognition cognitive task, on the other hand is the only connection suggested by both the average confusion graph and majority of transferability graphs.

In order to create the global transferability graph, which defines a possible global transfering policy instead of the current fold, we use the selected candidate connections as a vote of confidence. We select the candidate connections suggested by the affinity matrices with the following criterion:

- The candidate connection should be suggested by at least n transferability graphs. For this thesis we selected $n = 2$ and $n = 3$.

- The candidate connection should be suggested by the average confusion graph.

The resulting global transferability graphs with the connections suggested by at least 2 transferability graphs and 3 transferability graphs can be seen in Figure 4.13 and 4.14.

One interesting observation is that there is a strong disagreement between the average confusion graph and the transferability graphs, when we use the emotion processing and gambling as a target task. From almost every source cognitive task to emotion and gambling target cognitive tasks suggested by the transferability graphs. However, these transfers have usually the lowest expected transference performance in the average confusion graph. Hence, there are no connection to emotion processing and gambling in the global transferability graph.

The above observations reveal that the estimated global transferability graphs, seems to be generally consistent with the neuroscientific evidence. For example, the relationships between motor skills with language processing and social cognition are well studied and established in ([31],[29],[23]). Similarly, the relationship between the language with working memory or language with relational processing while relatively weaker, suggested in ([2]). Although some other relations like emotion processing and working memory could not be discovered by our method ([27]). Considering the fact that estimation of the global transferability graph is generated with no assumption on the cognitive tasks, the suggested computational model captures the affinities among the cognitive tasks of the HCP dataset.

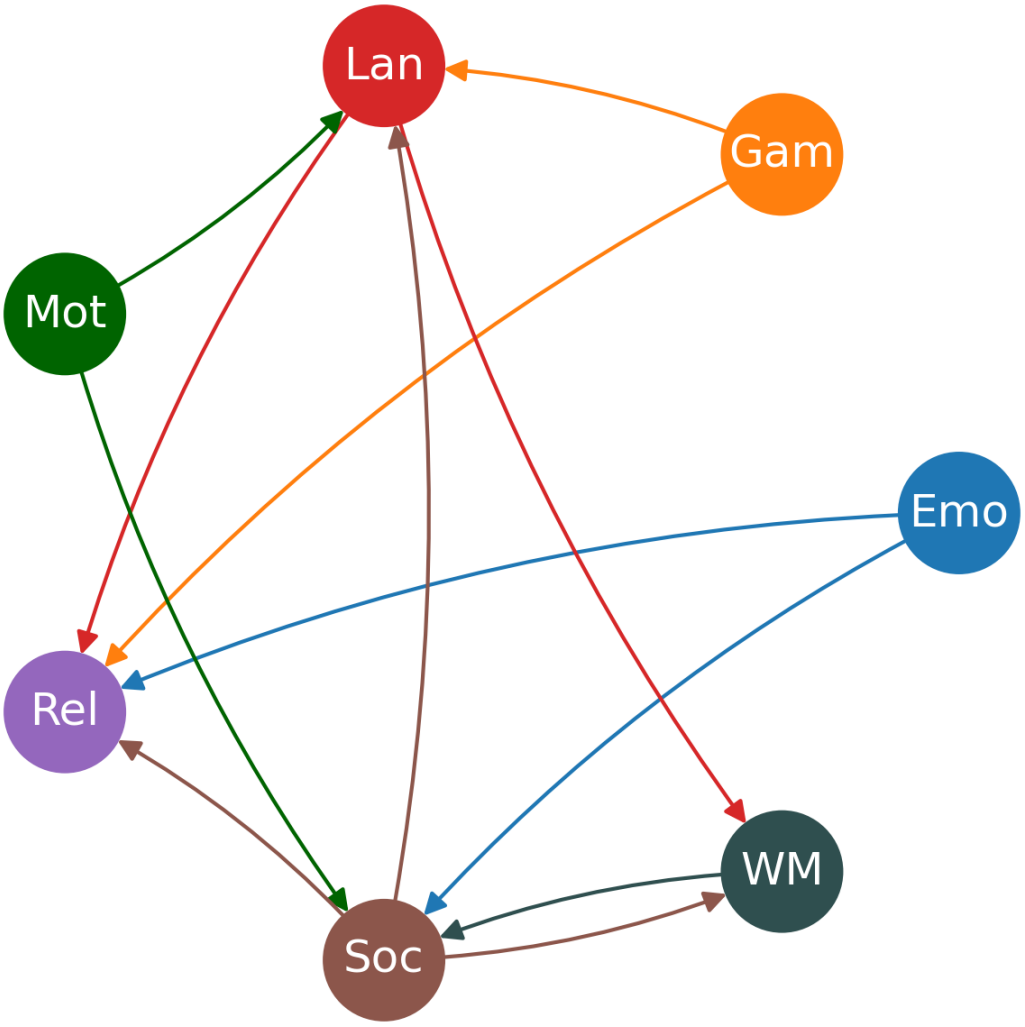

Figure 4.13: Resulting Global Transferability Graph with at least 2 Transferability Graph Suggestions. Majority of connections suggested by the average confusion graph also suggested by the transferability graphs except working memory as the source cognitive task which has only one transference in the global transference graph as the source cognitive task. All Connections which social cognition is the source cognitive task suggested by the average confusion graph on the other hand also suggested by the transferability graphs. For all other cognitive tasks as source cognitive task, 2 connection out of 3 suggested by the average confusion graph are selected as the connection in the global transferability graph.

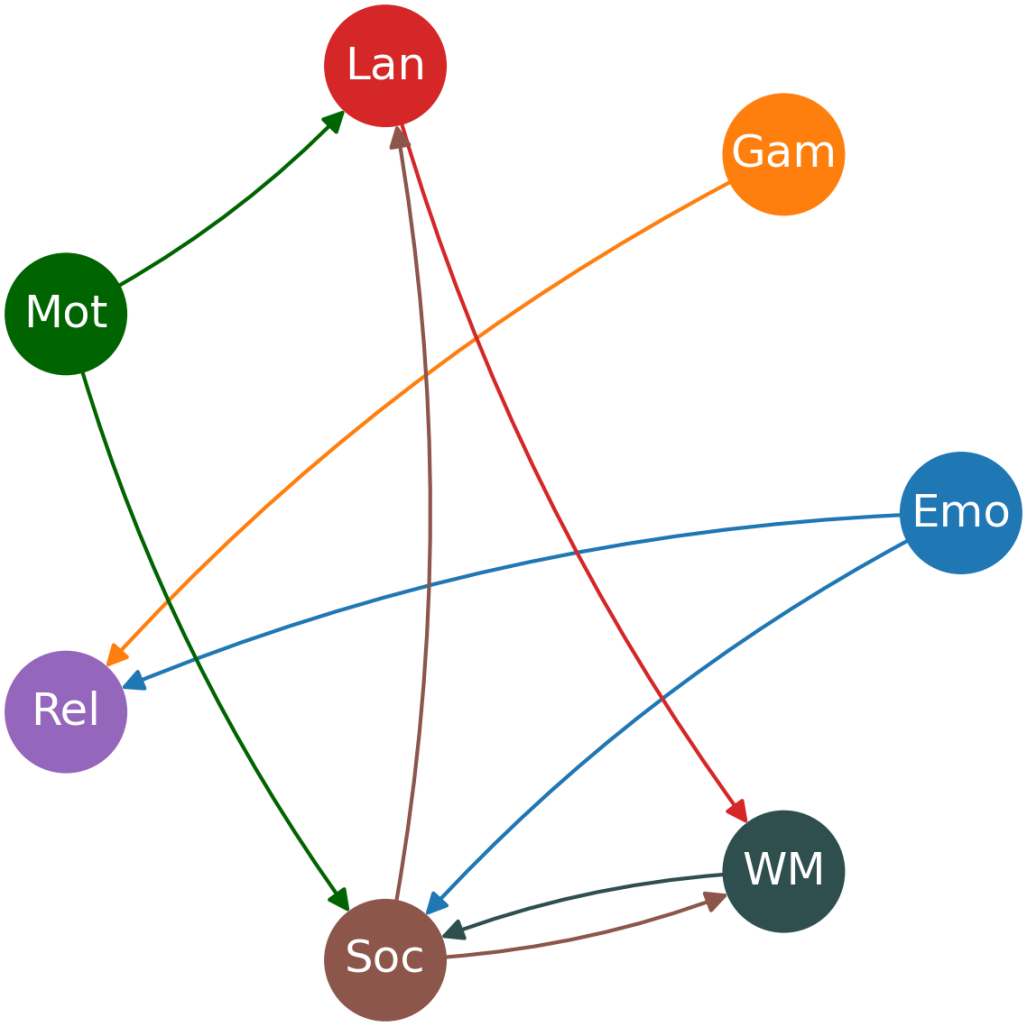

Figure 4.14: Resulting Global Transferability Graph with at least 3 Transferability Graph Suggestions. Most connections are preserved but weak connections like language cognitive task to relational processing cognitive task is pruned since they are not suggested by the majority of the transferability graphs.

4.4 Chapter Summary

In this chapter, we discuss the training process of deep learning models, the results of the generated affinity matrices and the global transferability graph and compare our results with the neuroscientific evidence in the literature. By not being based on any assumptions, our method can find some of important relations between cognitive tasks. However, conflicting values between different transferability graphs show the

difficulty of the method and high sensitivity to higher parameters due to the vast differences between different subjects. The same issue also causes some connections to be found in only small number of transferability graphs. It is clear that before doing transfer learning on the assumptions based on the global transferability graph, we need to make our algorithm more robust against between-subject variance.

CHAPTER 5

CONCLUSION AND SUGGESTIONS FOR FUTURE WORK

The major drawback of brain decoding methods using the fMRI data is the curse of dimensionality problem. Due to the limitations of the fMRI machines, only a limited amount of time, such as 10 minutes, can be used to record a cognitive process. On the other hand, the fMRI techniques generate almost 200,000 voxels at each time instant. This requires 200,000 dimensional feature space to represent the cognitive task recorded at each time instant.

One of the largest fMRI datasets available in fMRI literature is the HCP dataset. Considering the fact that the other datasets are relatively much more smaller than the HCP dataset, we explore the possibilities of transferring the information from one cognitive task to another, in this dataset. If we can develop an algorithm, which explores the affinities among the cognitive tasks, this technique can enable us to use transfer learning techniques for decoding the brain states from the HCP dataset to the other relatively smaller datasets. Therefore, the following question is very crucial to attack the curse of dimensionality problem:

Can we transfer information among cognitive tasks during brain decoding? For example, consider an important cognitive task of recording a particular information into memory. Also consider another cognitive task, such as emotion. These two tasks are widely studied in experimental neuroscience indepentently, yet, using small amount of data using some heuristic techniques. The reported literature consists of conflicting opinions about the nature of these cognitive tasks.

In this thesis, we try to develop a computational model, which estimates the affinities among the cognitive tasks. For this purpose, we employ the HCP dataset, in which

seven cognitive tasks are performed by multiple subjects, under an fMRI scanner.

We propose a pipeline method to discover and verify the transferability of a source task to a target task. The transferability among the cognitive tasks provides us a measure about the affinities among the tasks. Furthermore, assessing the transferability among the tasks enables us to design a robust training set for transfer learning algorithms, which improves the decoding performances.

Inspired from the pioneering work of Zamir et. al. ([38]), the proposed pipeline method extracts a global transferability graph among the cognitive tasks. In the first step, we train an autoencoder to learn a representation for each cognitive task. In the second step, we optimize a simple read-out function, which transfers the learned representation of a source task to a target task. The minimum loss values of this readout function indicates the degree of transferability among the cognitive tasks. In the third step, we normalize the minimum values of the readout function to obtain the graph affinity matrix.

Unfortunately, the above mentioned method is not robust to between subject variability of the recorded fMRI data for each cognitive task. When we perform k-fold cross validation to extract the affinity matrices , we could not obtain similar affinity values among the cognitive tasks. In other words, a pair of cognitive tasks, which transfers information with a relatively high affinity in a fold may not transfer any information at all in another fold. This fact can be attributed to the great variability among the subjects while they perform the same cognitive task. Additionally, the normalization method suggested by Zamir et. al.([38]) spoils the degree of transferability among the tasks in an uncontrolled fashion. This observation reveals that a further normalization step is required to obtain similar graph affinity matrices per task, across the subjects.

In order to attack the between-subject variability problem to asses the affinities among the cognitive tasks, we suggest a method, which generates a global transferability graph. This graph extracts the task affinities across the subjects. In this set-up, we rather estimate relatively consistent affinities from the same source cognitive task. For this purpose, we generate multiple transferability graphs, using their relative connection as a vote of confidence and verifying and pruning resulting connections using the generated average confusion graph.

The global transferability graph, suggested in this study is a simple normalization technique, which relies on empirically determined thresholds. It does not use any information based on experimental neuroscience. However, the highest affinities among the cognitive tasks are generally consistent with neuroscientific evidence.

The threshold values selected for pruning the global transferability graph defines the number of important connections between the cognitive tasks. The larger values of the threshold implies more affinities on the global transferability graph. As we decrease the threshold values, the suggested technique fails to find the crucial connections due to its conservative nature of pruning all connections. On the other hand, these connections may appear as high affinity values in the confusion graphs. Therefore, the inconsistency between the transferability graphs and the confusion graphs is still an unsolved problem.

In summary, the suggested global transferability graph is capable of extracting some important affinities among the cognitive tasks of HCP dataset. However, it loses transferability power of some tasks with respect to the confusion graph. This result reveals that a better thresholding method is needed to obtain a transferability graph.

5.1 Future Work

Due to insufficient data to extract temporal information, within-subject experiments are out of the scope of this thesis. On the other hand, the data per cognitive task, obtained from multiple subjects does not provide consistent affinities among the cognitive tasks, due to high between subject variance. Thus, the suggested method falls quite short to define generic measure for similarities among the cognitive tasks.

In order to solve the above mentioned problem, we need to further extend the current method, by training and testing source and transfer autoencoders within-subject datasets, while keeping a common score matrix for multiple subjects. We believe solving consistency issue will also help the conservative selection of edges. Additionally, in the future, we can verify the resulting graph on a smaller dataset to transfer learn instead of overcautiously pruning using the confusion graphs.

Also, we plan to extend our scope to higher order transferabilities by using multiple source latent representations, when training the transfer autoencoders, as proposed in the work of Zamir et. al ([38]).

REFERENCES

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

[2] G. Andrews, D. Birney, and G. S. Halford. Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & cognition*, 34(6):1325–1340, 2006.

[3] 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*, 80:169–189, 2013.

[4] J. R. Binder, W. L. Gross, J. B. Allendorfer, L. Bonilha, J. Chapin, J. C. Edwards, T. J. Grabowski, J. T. Langfitt, D. W. Loring, M. J. Lowe, et al. Mapping anterior temporal lobe language areas with fmri: a multicenter normative study. *Neuroimage*, 54(2):1465–1475, 2011.

[5] R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. Yeo. The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(5):2322–2345, 2011.

[6] F. Castelli, F. Happé, U. Frith, and C. Frith. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12(3):314–325, 2000.

[7] R. C. Craddock, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg. Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 62(6):1619–1628, 2009.

[8] C. Davatzikos, K. Ruparel, Y. Fan, D. Shen, M. Acharyya, J. W. Loughead,

R. C. Gur, and D. D. Langleben. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, 28(3):663–668, 2005.

[9] S. Dehaene, G. Le Clec'H, L. Cohen, J.-B. Poline, P.-F. van de Moortele, and D. Le Bihan. Inferring behavior from functional brain images. *Nature neuroscience*, 1(7):549–549, 1998.

[10] M. Delazer, A. Ischebeck, F. Domahs, L. Zamarian, F. Koppelstaetter, C. M. Siedentopf, L. Kaufmann, T. Benke, and S. Felber. Learning by strategies and learning by drill—evidence from an fmri study. *Neuroimage*, 25(3):838–849, 2005.

[11] M. R. Delgado, L. E. Nystrom, C. Fissell, D. Noll, and J. A. Fiez. Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of neurophysiology*, 84(6):3072–3077, 2000.

[12] I. O. Ertugrul, M. Ozay, and F. T. Y. Vural. Encoding the local connectivity patterns of fmri for cognitive state classification, 2016.

[13] O. Firat, L. Oztekin, and F. T. Y. Vural. Deep learning for brain decoding. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2784–2788. IEEE, 2014.

[14] G. Ganis, W. L. Thompson, and S. M. Kosslyn. Brain areas underlying visual mental imagery and visual perception: an fmri study. *Cognitive Brain Research*, 20(2):226–241, 2004.

[15] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

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

[17] A. R. Hariri, A. Tessitore, V. S. Mattay, F. Fera, and D. R. Weinberger. The amygdala response to emotional stimuli: a comparison of faces and scenes. *Neuroimage*, 17(1):317–323, 2002.

[18] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

[19] Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, 8(5):679–685, 2005.

[20] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi. Application of advanced machine learning methods on resting-state fmri network for identification of mild cognitive impairment and alzheimer's disease. *Brain imaging and behavior*, 10(3):799–817, 2016.

[21] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[22] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, and S. Ishii. Deep learning of fmri big data: a novel approach to subject-transfer decoding. *arXiv preprint arXiv:1502.00093*, 2015.

[23] H. C. Leonard and E. L. Hill. The impact of motor development on typical and atypical social cognition and language: A systematic review. *Child and Adolescent Mental Health*, 19(3):163–170, 2014.

[24] T. M. Mitchell, R. Hutchinson, M. A. Just, R. S. Niculescu, F. Pereira, and X. Wang. Classifying instantaneous cognitive states from fmri data. In *AMIA annual symposium proceedings*, volume 2003, page 465. American Medical Informatics Association, 2003.

[25] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman. Learning to decode cognitive states from brain images. *Machine learning*, 57(1):145–175, 2004.

[26] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[27] L. H. Phillips, S. Channon, M. Tunstall, A. Hedenstrom, and K. Lyons. The role of working memory in decoding emotions. *Emotion*, 8(2):184, 2008.

[28] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun. Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, 8:229, 2014.

[29] G. Rizzolatti and M. Fabbri-Destro. The mirror system and its role in social cognition. *Current opinion in neurobiology*, 18(2):179–184, 2008.

[30] R. W. Saaty. The analytic hierarchy process—what it is and how it is used. *Mathematical modelling*, 9(3-5):161–176, 1987.

[31] K. Sakreida, C. Scorolli, M. M. Menz, S. Heim, A. M. Borghi, and F. Binkofski. Are abstract action words embodied? an fmri investigation at the interface between language and motor cognition. *Frontiers in human neuroscience*, page 125, 2013.

[32] S. Sarraf and G. Tofighi. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*, 2016.

[33] R. Smith, K. Keramatian, and K. Christoff. Localizing the rostrolateral prefrontal cortex at the individual level. *Neuroimage*, 36(4):1387–1396, 2007.

[34] 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*, 15(1):273–289, 2002.

[35] B. Velioglu and F. T. Y. Vural. Transfer learning for brain decoding using deep architectures. In *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pages 65–70. IEEE, 2017.

[36] A. Wang, M. Tarr, and L. Wehbe. Neural taskonomy: Inferring the similarity of task-derived representations from brain activity. *Advances in Neural Information Processing Systems*, 32, 2019.

[37] T. Wheatley, S. C. Milleville, and A. Martin. Understanding animate agents: distinct roles for the social network and mirror system. *Psychological science*, 18(6):469–474, 2007.

[38] A. R. Zamir, A. Sax, W. Shen, L. J. Guibas, J. Malik, and S. Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3712–3722, 2018.