

REPRESENTATION OF HUMAN BRAIN BY MESH NETWORKS

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

BY

ITIR ÖNAL ERTUĞRUL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

JULY 2017

Approval of the thesis:

REPRESENTATION OF HUMAN BRAIN BY MESH NETWORKS

submitted by **ITIR ÖNAL ERTUĞRUL** in partial fulfillment of the requirements
for the degree of **Doctor of Philosophy in Computer Engineering Department,**
Middle East Technical University by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering**

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

Examining Committee Members:

Prof. Dr. Tolga Can
Computer Engineering Department, METU

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

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

Assist. Prof. Dr. Mete Özay
Graduate School of Information Sciences, Tohoku University

Assist. Prof. Dr. Aykut Erdem
Computer Engineering Department, Hacettepe 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: ITIR ÖNAL ERTUĞRUL

Signature :

ABSTRACT

REPRESENTATION OF HUMAN BRAIN BY MESH NETWORKS

Önal Ertuğrul, İtir

Ph.D., Department of Computer Engineering

Supervisor : Prof. Dr. Fatoş T. Yarman Vural

July 2017, 157 pages

In this thesis, we propose novel representations to extract discriminative information in functional Magnetic Resonance Imaging (fMRI) data for cognitive state and gender classification. First, we model the local relationship among a set of fMRI time series within a neighborhood by considering temporal information obtained from all measurements in time series. The estimated local relationships, called Mesh Arc Descriptors (MADs), are employed to represent information in fMRI data. Second, we adapt encoding methods frequently used in Computer Vision, namely Fisher Vectors (FV), Vector of Locally Aggregated Descriptors (VLAD) and Bag-of-Words (BoW) to encode local MADs. We show that employing MADs outperform state-of-the-art fMRI representations and encoding them further with FV gives superior performance over MADs. Then, we propose a hierarchical framework, called Hierarchical Multi-resolution Mesh Networks (HMMNs), in which the fMRI signal is decomposed into multiple subbands and mesh networks are constructed for each subband separately. We fuse the decisions of classifiers trained with multi-resolution mesh-networks in the final step of the framework. We show that Hierarchical Multi-resolution Mesh Net-

works outperform mesh-networks constructed from original fMRI signal. Finally, we adapt multi-resolution approach for gender classification using fMRI data. We fuse the decisions of classifiers trained with multi-resolution multi-task mesh networks in a 2-level hierarchical architecture to discriminate gender. The proposed gender classification framework performs better compared to single layer architectures fusing only multi-task or only multi-resolution mesh networks.

Keywords: fMRI, Mesh Networks, Multi-resolution Analysis, Fisher Vector, Cognitive State Classification, Gender Classification

ÖZ

İNSAN BEYNİNİN ÖRGÜ AĞLARI İLE GÖSTERİMİ

Önal Ertuğrul, İtir

Doktora, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Fatoş T. Yarman Vural

Temmuz 2017, 157 sayfa

Bu tezde, bilişsel durum ve cinsiyet sınıflandırmada kullanılmak üzere fonksiyonel Manyetik Rezonans Görüntüleme (fMRG) verisindeki ayımsayıcı bilgiyi çıkarmak için yeni gösterimler önermekteyiz. İlk olarak, zaman serilerindeki tüm ölçümleri içeren zamansal bilgiyi göz önünde bulundurarak fMRG zaman serileri kümesindeki yerel ilişkileri bir komşulukta modellemekteyiz. Örgü Yay Betimleyicileri (ÖYB) adı verilen yerel ilişkiler, fMRG verisindeki bilginin gösterimi için kestirilmiştir. İkinci olarak, bilgisayarla görme alanında yaygın olarak kullanılan Fisher Vektörü (FV), Yerel Kümelenmiş Betimleyiciler Vektörü ve Kelime Çantası kodlama yöntemlerini yerel ÖYB'leri kodlamak için uyarlamaktayız. ÖYB'lerin kullanımının güncel fMRI gösterimlerinden daha iyi performans verdiğini ve bunların daha sonra FV ile kodlanmalarının ÖYB'ye göre daha üstün performans verdiğini göstermekteyiz. Ardından, fMRG sinyalinin çoklu alt bantlara ayrıldığı ve her bir altband için ayrı ağlar oluşturulduğu Hiyerarşik Çok-çözünürlüklü Örgü Ağları (HÇÖA) adı verilen bir yapı önermekteyiz. Yapının son aşamasında, çok çözünürlüklü örgü ağları ile eğitilmiş sınıflandırıcıların kararlarını birleştirmekteyiz. Hiyerarşik Çok-çözünürlüklü Örgü Ağ-

larının orjinal fMRG verisi için kurulan örgü ağlarından daha iyi performans verdiğini göstermektedir. Son olarak, fMRG verisi kullanılarak cinsiyet sınıflandırma için çok-çözünürlü yaklaşımı uyarlamaktayız. Çok çözünürlüklü ve çok görev altında kurulan örgü ağları ile eğitilen sınıflandırıcıları, cinsiyet sınıflandırma için 2 katmanlı hiyerarşik bir mimari ile birleştirmektedir. Önerilen cinsiyet sınıflandırma yapısı, yalnızca çok görev altında kurulan ya da yalnızca çok çözünürlüklü örgü ağlarını tek katmanda birleştiren mimarilere göre daha başarılıdır.

Anahtar Kelimeler: fMRG, Örgü Ağları, Çok Çözünürlüklü Analiz, Fisher Vektörü, Bilişsel Durum Sınıflandırma, Cinsiyet Sınıflandırma

To my family

ACKNOWLEDGMENTS

With my deepest gratitude, I would like to thank my supervisor Prof. Dr. Fatoş Yarman Vural for her endless support and wise guidance throughout the course of this thesis. She has not only been a supervisor, but also been a role model to me with her stance in life, immense wisdom, enthusiasm to learn and teach and kind heart. It was a great pleasure and honor to have her guide me in my research and in my life.

I would like to express my gratitude to my three colleagues that affected my research career. First, I would like to thank Assist. Prof. Dr. Mete Özey for being such an encouraging and farsighted mentor. His inspiring knowledge and vision have enlightened my way through my research and have shaped the borders of this thesis. Second, I thank Dr. Orhan Fırat for setting the example of how a researcher should be and for his kind support from the very beginning of my research journey. Third, I wish to thank Dr. Hamdi Dibeklioglu not only for his guidance and hospitality during my visiting research in TU Delft, but also for his immense support in shaping my postdoctoral journey.

I would like to thank Assoc. Prof. Dr. Sinan Kalkan for his valuable comments and precious suggestions during the thesis monitoring committee meetings. I also appreciate the feedback offered by my thesis defense committee members Prof. Dr. Tolga Can and Assist. Prof. Dr. Aykut Erdem.

I owe special thanks to my friends in Image Processing and Pattern Recognition Laboratory, Burak Velioğlu, Emre Aksan, Dr. Serdar Çiftçi, Hazal Moğultay, Ozan Yıldız, Barış Nasır, Güneş Sucu, Arman Afrasiyabi and Sarper Alkan for providing a nice working environment. Thank you for all the joy and all the conversations in coffee breaks and barbecues in our terrace.

I would like to express my deepest gratitude to my family, for making me who I am today. I am deeply grateful to my dearest mom Deniz and my dearest dad Mustafa for

giving me all the love, peace, confidence and encouragement that a human being can give. I am thankful to my brother Mert, to my sister-in-law Figen and to my nephew Orhun for their motivation, support and joy. I also thank my parents-in-law Dursun and Fulya, and my sister-in-law Naz for their sincere support and making their home my home.

Last but surely not the least, my heartfelt appreciation goes to my beloved husband Mert for encouraging me to dream, dreaming with me and helping me to make my dreams come true. It was his never-ending support, limitless love and joyful companionship that turned all the disappointments into hope, all the failures into success and all the tough days into good ones, not only during my Ph.D. journey, but also throughout my life. While closing this chapter of my life, I am lucky and excited to open new ones with him. Cheers to new journeys!

I acknowledge the support of The Scientific and Technological Research Council of Turkey (TÜBİTAK) BİDEB for 2211E graduate student fellowship during my Ph.D. education. I thank TÜBİTAK also for supporting this work by the project (no 114E045) titled "Multi-layered cognitive learning model".

TABLE OF CONTENTS

ABSTRACT	v
ÖZ	vii
ACKNOWLEDGMENTS	x
TABLE OF CONTENTS	xii
LIST OF TABLES	xix
LIST OF FIGURES	xxi
LIST OF ABBREVIATIONS	xxviii
NOMENCLATURE	xxix

CHAPTERS

1	INTRODUCTION	1
1.1	Motivation and Problem Definition	1
1.2	Contributions of the thesis	3
1.3	Organization of the thesis	4

2	AN OVERVIEW OF REPRESENTATION AND ENCODING OF FMRI DATA FOR COGNITIVE STATE AND GENDER CLASSIFICATION	7
2.1	Multi-Voxel Pattern Analysis	7
2.2	Representation of fMRI Signal	10
2.3	Encoding Methods	12
2.4	Multi-resolution Representation of fMRI Data	14
2.5	Fusion of Classifier Decisions	16
2.6	Gender Classification	16
2.7	Chapter Summary	18
3	EXPERIMENTAL SETUP AND FMRI DATA COLLECTION	19
3.1	Visual Object Recognition Experiment	19
3.2	Emotional Memory Retrieval (EMR) Experiment	20
3.3	Human Connectome Project (HCP) task fMRI Data	21
3.4	Preprocessing and Analysis of Neuroimaging Data	23
4	A NEW REPRESENTATION OF FMRI DATA AND ITS ENCODING	25
4.1	Mesh Representation of fMRI Signal with Respect to Functional and Spatial Neighborhoods	25
4.1.1	Node Time Series	27
4.1.2	Neighborhood Systems	27

4.1.3	Construction of Local Meshes	29
4.1.4	Estimation of Mesh Arc Descriptors (MADs)	32
4.1.5	Classification of Cognitive States using Mesh Net- works	33
4.2	Generation of Brain Connectivity Dictionary and Encoding of Mesh Arc Descriptors	34
4.2.1	Fisher Vector (FV) Encoding	36
4.2.2	Vector of Locally Aggregated Descriptors (VLAD)	37
4.2.3	Bag-of-Words (BoW)	38
4.3	Chapter Summary	38
5	HIERARCHICAL MULTI-RESOLUTION MESH NETWORKS (HMMNS)	39
5.1	Multi-resolution Representation of fMRI Signals	40
5.2	Estimation of Multi-resolution Mesh Arc Descriptors	43
5.3	Network Topology and Connectivity Analysis of HMMNs . . .	45
5.3.1	Connectivity Analysis by Degree of Nodes of Mesh Network	45
5.3.2	Connectivity Analysis by Strength of Nodes of Mesh Network	46
5.3.3	Connectivity Analysis by Betweenness Centrality of Nodes of Mesh Network	46
5.3.4	Connectivity Analysis by Global Efficiency	47

5.4	Multi-resolution Mesh Networks for Cognitive State Classification	47
5.4.1	Graph Embedding of Multi-resolution Mesh Networks	48
5.4.2	Fusion of Multi-resolution Brain Networks under a Hierarchical Ensemble Learning Architecture . .	49
5.4.3	Significance of Meta-layer Features Representing Class Membership Values Obtained from Classifiers Trained with Multi-resolution Mesh Networks	50
5.4.4	Significance of Base-layer Features Representing Connectivity Among Anatomical Regions	51
5.4.5	Diversity of Base-layer Classifiers	52
5.5	Multi-resolution Mesh Networks for Gender Classification .	55
5.5.1	Fusion of Multi-resolution Multi-task Mesh Networks under Fuzzy Stacked Generalization (FSG) .	56
5.5.1.1	Fusion of Decisions of All Base-Layer Task-Classifiers	58
5.5.1.2	Fusion of Decisions of All Base-Layer Subband-Classifiers in a Task	58
5.5.1.3	Fusion of Base-Layer Task-Classifiers in a Subband Followed by Fusion of Meta-Layer Subband-Classifiers . . .	59

5.5.1.4	Fusion of Base-Layer Subband-Classifiers in a Task Followed by Fusion of Meta- Layer Task-Classifiers	61
5.5.2	Gender Discriminative Power of MADs	62
5.6	Chapter Summary	63
6	EXPERIMENTS AND RESULTS	65
6.1	Experiments and Results of Employing Mesh Arc Descriptors	66
6.1.1	Comparison of Mesh Arc Descriptors with State- of-the-art fMRI Features	67
6.1.2	Classification Results for Brain Decoding	68
6.1.3	Analysis of Similarity Measures for Neighborhood Selection to Form Local Meshes	76
6.1.4	Analysis of Local Meshes	77
6.1.4.1	Statistical Analysis of Correlations of Voxels	77
6.1.4.2	Statistical Analysis of Models	80
6.2	Experiments and Results of Encoding the Mesh Arc Descrip- tors	81
6.2.1	Classification Results	82
6.2.2	Analysis of Codewords	86
6.3	Experiments on Performances of Multi-resolution Mesh Net- works for Cognitive State Classification	90

6.3.1	Performances of Cognitive State Classification . . .	90
6.3.2	Analysis and Interpretation of Topology of Mesh Network	94
6.3.3	Significance Analysis of Meta-layer Features . . .	106
6.3.4	Significance Analysis of Base-layer Features . . .	107
6.3.5	Analysis of Diversity of Base-layer Classifiers . . .	113
6.4	Experiments and Results of Multi-resolution Mesh Networks for Gender Classification	114
6.4.1	Gender Classification Experiments and Results . .	114
6.4.1.1	Performances Obtained for Each Cog- nitive Task	115
6.4.1.2	Performances Obtained for Each Sub- band of Each Cognitive Task	115
6.4.1.3	Performances Obtained by 1-level Fu- sion of Task-Classifiers	118
6.4.1.4	Performances Obtained by 1-level Fu- sion of Subband-Classifiers	118
6.4.1.5	Performances Obtained by 2-level Ar- chitecture Containing Fusion of Task- Classifiers Followed by Fusion of Subband- Classifiers	121

6.4.1.6	Performances Obtained by 2-level Architecture Containing Fusion of Subband-Classifiers Followed by Fusion of Task-Classifiers	123
6.4.2	Analysis of Gender Discriminative Power of Mesh Arc Descriptors	124
6.5	Chapter Summary	130
7	CONCLUSION AND FUTURE WORK	131
7.1	Future Work	133
	REFERENCES	135
	APPENDICES	
	CURRICULUM VITAE	153

LIST OF TABLES

TABLES

Table 3.1	Number of scans per session and its duration for each task (min:sec) for HCP data.	23
Table 6.1	Classification performance (%) of SVM classifier computed in visual object recognition experiment.	70
Table 6.2	Classification performance (%) of SVM classifier computed in emotional memory retrieval experiment (2-class).	72
Table 6.3	Classification performance (%) of SVM classifier computed in emotional memory retrieval experiment (4-class).	73
Table 6.4	Classification performance (%) of SVM classifier for HCP task fMRI data.	74
Table 6.5	Classification performance (%) of SVM classifier computed in visual object recognition experiment.	77
Table 6.6	Classification performance (%) obtained using MADs	82
Table 6.7	Classification performance (%) computed for different encoding methods obtained using HCP and EMR datasets.	83
Table 6.8	Running times of each step of the framework measured in seconds. .	84
Table 6.9	Classification accuracy (%) measured with state-of-the-art fMRI representations.	85

Table 6.10 Results of McNemar’s test with Yates’ correction to compare significance of using Fisher vector encoding of MADs over using MADs . . .	86
Table 6.11 Classification performances (%) of approximation and detail parts of decomposition levels $l \in [0, 6]$ at the output of base-layer classifiers of FSG, trained and tested by three sets of feature vectors, namely, mesh arc descriptors (MAD), pairwise correlation (PC) and representative time series (RTS). The performances are obtained on HCP dataset.	92
Table 6.12 Classification performances (%) of fuzzy stacked generalization (FSG), majority voting (MV) and weighted majority voting (WMV) methods, obtained using mesh arc descriptors, pairwise correlation (PC) and representative time series (RTS). Numbers with + or - sign show the standard deviation of the performances obtained during the cross validation. .	93
Table 6.13 Diversity measures obtained for ensemble of classifiers trained with mesh arc descriptors. The arrow specifies whether diversity is larger if the measure is smaller ($\downarrow$) or larger ($\uparrow$).	114
Table 6.14 Classification performances (%) obtained with classifiers trained with mesh networks formed for different cognitive tasks. We compare the classification performances obtained using mesh arc descriptors (MAD), pairwise correlations (PC) and representative time series (RTS)	115
Table 6.15 Gender classification performances obtained by the 2-level fusion of <i>task-classifiers</i> followed by fusion of <i>subband-classifiers</i>	122
Table 6.16 Gender classification results obtained by the 2-level fusion of <i>subband-classifiers</i> followed by fusion of <i>task-classifiers</i>	123

LIST OF FIGURES

FIGURES

Figure 2.1	An example MVPA pipeline [33]. Classes of the presented images are used as labels and the measurements recorded from a number of voxels during the presentation of images are used as inputs to train a classifier. During test time, the classifier directly assigns a label to a new set of measurements recorded from voxels.	8
Figure 2.2	Steps of encoding local features using BoW and VLAD.	12
Figure 3.1	Visual Object Recognition Experiment. Participant is presented with bird or flower pictures. The participant is expected to indicate by pressing the button when the current stimulus matches the one from previous step. In this example, second and fifth pictures pointed with red arrows have the same categories (bird and flower, respectively) with their predecessors.	20
Figure 4.1	The flowchart of the proposed approach for obtaining local meshes and classification of cognitive states.	26
Figure 4.2	A local mesh formed by (a) spatial neighbors and (b) functional neighbors of a seed node. In both figures, the node denoted with yellow color at the center is the seed node. In (a), red nodes denote the spatial neighbors of the yellow node within a radius of 1 unit and blue nodes have $\sqrt{2}$ unit distance to yellow node. In (b), a mesh is formed with functional neighbors around the seed node denoted with yellow color.	29

Figure 4.3 Design of a spatially local mesh for modeling the relationship among the BOLD responses of nodes. (a) For simplicity, a seed node is depicted in an axial slice at the center of a mesh (red) with its four spatially nearest neighbors (yellow). Neighboring nodes are selected within a 3D brain volume. (b) fMRI time series are recorded at each node, for each stimulus for bird and flower classes. (c) A local mesh is computed using the responses obtained from a seed node $x_{w,i}$ and its four nearest neighbors for window w . We also visualize node intensity values of a sample response $x_{w,j}$ obtained from node j 30

Figure 4.4 Design of a functionally local mesh using the voxel time series. (a) A sample functional connectivity matrix formed using Pearson correlation. p -functionally nearest neighbors of a voxel are denoted as the ones with the dark-to-light red colors. (b) fMRI measurements are recorded for each stimulus for bird and flower classes. (c) We depict a functional mesh formed around a seed voxel, in star topology. The edge weights are computed by using the response obtained from a seed node $x_{w,i}$ and the responses of its four functionally nearest neighbors, $x_{w,j}$, $x_{w,u}$, $x_{w,y}$, $x_{w,z}$, for window w 31

Figure 4.5 An overview of our proposed framework which is employed to encode MADs. (a) For each cognitive task, we first form local meshes around all anatomical regions using their functionally nearest neighbors (In the figure, we represent only five representative meshes). (b) Next, we estimate the mesh arc descriptors (MADs) for all tasks of all subjects. (c) We apply PCA or ICA on mesh arc descriptors obtained from training dataset to decorrelate them. (d) Then, we employ GMMs and k-means clustering methods on the descriptors using training dataset and find codewords of our dictionary. (e) Finally, we obtain FV, VLAD and BoW encoding for each task of training and test datasets. 35

Figure 5.1 Overview of the proposed hierarchical framework (HMMNs). A random seed region and its three functionally nearest neighbors (with smaller sizes) are denoted on the template brain. In the multi-resolution analysis step, the average fMRI signal of each region is decomposed into a set of coarse-to-fine signals. Then, meshes are formed around each seed region for all subbands, and mesh arc descriptors are estimated. Finally, the estimated arc descriptors of the meshes are ensembled under a network structure and fed as an input to the hierarchical ensemble learning architecture, called, FSG, where the final decision is computed by a meta-layer classifier.	41
Figure 5.2 Cubic Battle-Lemarie filters (Father wavelet (Φ) on the left and mother wavelet (Ψ) on the right)	43
Figure 5.3 A sample mesh formed around an anatomical region i with its three neighboring regions j , y and z for subband s and window w . We represent $x_{s,w,i}$ as a weighted linear combination of $x_{s,w,j}$, $x_{s,w,y}$ and $x_{s,w,z}$. We estimate mesh arc descriptors $a_{s,w,i,j}$, $a_{s,w,i,y}$ and $a_{s,w,i,z}$ using ridge regression and represent the corresponding mesh in terms of these descriptors. We form meshes around all anatomical regions and the set of meshes form a mesh network for subband s and window w	44
Figure 5.4 1-level fusion of decisions of 7 base layer <i>task-classifiers</i> in a subband	57
Figure 5.5 1-level fusion of decisions of base layer <i>subband-classifiers</i> in a task	59
Figure 5.6 2-level fusion of all classifiers. In the first level, decisions of all base layer <i>task-classifiers</i> are fused within each subband. In the second level, decisions of meta layer <i>subband-classifiers</i> are fused to give the final decision.	60
Figure 5.7 2-level fusion of all classifiers. In the first level, decisions of all base layer <i>subband-classifiers</i> are fused within each task. In the second level, decisions of meta layer <i>task-classifiers</i> are fused to give the final decision.	61

Figure 6.1	Mean and standard deviation of classification performances of SVM computed in visual object recognition experiment.	74
Figure 6.2	Mean and standard deviation of classification performances of SVM computed in emotional memory retrieval experiment (2-class).	75
Figure 6.3	Mean and standard deviation of classification performances of SVM computed in emotional memory retrieval experiment (4-class).	75
Figure 6.4	BOLD responses of a voxel set consisting of 10 voxels, each of which is plotted with a different color.	78
Figure 6.5	Histograms of correlations between seed voxels and surrounding voxels selected using (a) Spatial closeness, (b) Granger causality, (c) Partial correlation and (d) Pearson correlation.	79
Figure 6.6	Histograms of R^2 values computed when the meshes are formed using (a) Spatial neighbors and (b) Functional neighbors.	81
Figure 6.7	Classification accuracy (%) measured for each task of HCP dataset.	87
Figure 6.8	Codewords of brain connectivity dictionary obtained by FVs for encoding of MADs with $p = 40$. Codewords are visualized using Brain-Net Viewer [155], and depicted in an order such that Codeword1 and Codeword20 represent the codeword of GHE, and GLE, respectively.	88
Figure 6.9	Effect of removing Gaussian mixture components from representation on classification accuracy. Results of removal of Gaussians, whose FV entries have high energy (GHE) are plotted with solid lines while results of removal of Gaussians, whose FV entries have low energy (GLE) are plotted with dashed lines. The numbers next to GHE and GLE correspond to the number of neighboring nodes in mesh (p). For example, GHE10 represents removal of Gaussians, whose FV entries have high energy such that GMMs are estimated for MADs formed with $p = 10$. It can be observed that removal of GHE decreases the accuracy much more compared to removal of GLE.	89

Figure 6.10 Overlay of the average out-degrees of nodes of mesh networks formed for HCP dataset on human brain templates for approximation ($[A_1 - A_6]$) and detail ($[D_1 - D_6]$) subbands. Darker blue colors denote smaller out-degrees while darker red colors denote larger out-degrees.	95
Figure 6.11 Average out-degrees of nodes (anatomical regions) of mesh networks formed at subbands $[A_1, A_6]$. Maximum values show the massively connected anatomic regions whereas the minimum values show the rarely connected regions.	96
Figure 6.12 Average out-degrees of nodes (anatomical regions) of mesh networks formed at subbands $[D_1, D_6]$. Maximum values show the massively connected anatomic regions whereas the minimum values show the rarely connected regions.	97
Figure 6.13 Overlay of the average out-strengths of nodes of mesh networks formed for HCP dataset on human brain templates for approximation ($[A_1 - A_6]$) and detail ($[D_1 - D_6]$) subbands. Darker blue colors denote smaller out-strengths while darker red colors denote larger out-strengths.	98
Figure 6.14 Average strength of nodes (anatomical regions) of mesh networks formed at subbands $[A_1, A_6]$. Note that the strength and variations among the anatomical regions decrease as the level increases.	99
Figure 6.15 Average strength of nodes (anatomical regions) of mesh networks formed at subbands $[D_1, D_6]$	100
Figure 6.16 Average total strength of all mesh networks formed for each task category, over the subbands, for approximation and detail parts.	102
Figure 6.17 Overlay of the average betweenness centrality of nodes of mesh networks formed for HCP dataset on human brain templates for approximation ($[A_1 - A_6]$) and detail ($[D_1 - D_6]$) subbands. Darker blue colors denote smaller betweenness centrality while darker red colors denote larger betweenness centrality.	103

Figure 6.18 Average betweenness centrality of nodes (anatomical regions) of mesh networks formed at subbands $[A_1, A_6]$	104
Figure 6.19 Average betweenness centrality of nodes (anatomical regions) of mesh networks formed at subbands $[D_1, D_6]$	105
Figure 6.20 Average global efficiency of all mesh networks formed for each task category, over the subbands, for approximation and detail parts. . . .	106
Figure 6.21 Significance of each task at each subband. Maximum z values of class memberships obtained using absolute value two-sample t-test with pooled variance estimate for subbands $[A_1, A_6]$ and $[D_1, D_6]$. z values are computed under one versus all condition. The bars show the maximum of z -values obtained from the membership values during one-versus-all task classification for each task category.	107
Figure 6.22 The most significant mesh arc descriptors that discriminate a specific task from others for different approximation subbands. Color of the mesh arc indicates the task that is discriminated from the others by the corresponding mesh arc descriptor.	108
Figure 6.23 The most significant mesh arc descriptors that discriminate a specific task from others for different detail subbands. Color of the mesh arc indicates the task that is discriminated from the others by the corresponding mesh arc descriptor.	109
Figure 6.24 Average z -values of top-20 mesh arc descriptors that discriminate a specific task from others for different subbands.	112
Figure 6.25 Gender classification performances obtained for each subband and each task separately, using three different feature sets, namely, mesh arc descriptors (MAD), pairwise correlation (PC) and representative time series (RTS).	116
Figure 6.26 Gender classification results performances by the 1-level fusion of decisions of <i>task-classifiers</i>	117

Figure 6.27 Gender classification performances obtained by the 1-level fusion of decisions of <i>subband-classifiers</i> using fuzzy stacked generalization (FSG).	119
Figure 6.28 Gender classification performances obtained by the 1-level fusion of decisions of <i>subband-classifiers</i> using majority voting (MV).	120
Figure 6.29 Top 50 gender discriminative MADs obtained for Emotion	125
Figure 6.30 Top 50 gender discriminative MADs obtained for Gambling	126
Figure 6.31 Top 50 gender discriminative MADs obtained for Language Pro- cessing	126
Figure 6.32 Top 50 gender discriminative MADs obtained for Motor	127
Figure 6.33 Top 50 gender discriminative MADs obtained for Relational Pro- cessing	127
Figure 6.34 Top 50 gender discriminative MADs obtained for Social Processing	128
Figure 6.35 Top 50 gender discriminative MADs obtained for Working Memory	128

LIST OF ABBREVIATIONS

fMRI	Functional Magnetic Resonance Imaging
MAD	Mesh Arc Descriptor
FV	Fisher Vector
VLAD	Vector of Locally Aggregated Descriptors
BoW	Bag of Words
GMM	Gaussian Mixture Model
PCA	Principal Component Analysis
ICA	Independent Component Analysis
DWT	Discrete Wavelet Transform
MVPA	Multi-voxel Pattern Analysis
SVM	Support Vector Machines
FLM	Functionally Local Meshes
SLM	Spatially Local Meshes
PC	Pairwise Correlation
RTS	Representative Time Series

NOMENCLATURE

$\alpha_{r,l,q}$	Approximation coefficient for region r level l and q^{th} location to which the wavelet is translated in time
β	Parameter set of all u_k 's
$\mathcal{G}_{\mu_k}^A$	Gradient of GMM with respect to μ_k
$\mathcal{G}_{\sigma_k}^A$	Gradient of GMM with respect to σ_k
$\mathcal{V}^A$	VLAD encoding
$\mathbf{v}_k^A$	VLAD encoding columns obtained for k^{th} Gaussian
μ_k	Mean vector of k^{th} Gaussian
Σ_k	Covariance matrix of k^{th} Gaussian
$\mathbf{a}_{w,i}$	Mesh Arc Descriptor vector of node i for window w
$\mathbf{a}_i$	MAD vector formed around node i (simple representation)
B^A	BoW encoding
G_{β}^A	Derivatives of log-likelihood of GMM with respect to the parameters β
G_{β}^A	Fisher vector containing derivatives of log-likelihood of GMM with respect to the parameters β
$\mathbf{x}_{w,i}$	Vector representing intensity values of the BOLD response measured at a node i within window w
$\delta_{r,l,q}$	Detail coefficient for region r level l and q^{th} location to which the wavelet is translated in time
$\eta(i, \pi)$	Spatial neighborhood centered in node i with a radius π

η_p^{func}	p functional neighborhood
η_p^{spat}	p spatial neighborhood
$\eta_p[i]$	p -neighborhood of node i
$\gamma_i(k)$	soft assignment of the MAD vector $\mathbf{a}_i$ to the k^{th} Gaussian
κ	Number of cross validation folds
λ	Regularization parameter
$\mathcal{A}$	A set of approximation coefficients
$\mathcal{D}$	A set of detail coefficients
$\mathcal{E}_i$	Set of arcs of mesh formed around node i
$\mathcal{M}_i$	Local mesh formed around node i
$\mathcal{N}$	Set of nodes
$\mathcal{N}_i$	Set of nodes of mesh formed around node i
ω_k	Mixture weight of k^{th} Gaussian
$\Phi_{l,q}(t)$	Father wavelet
π	Radius
$\Psi_{l,q}(t)$	Mother wavelet
θ	Measure of difficulty
A	Set of MADs obtained from a single task of a single subject
A_l	Approximation subband of level l
$a_{s,w,i,j}$	Mesh Arc Descriptor estimated between seed node i and neighboring node j for window w at subband s
$a_{w,i,j}$	Weight of an arc between seed node of a mesh i and a neighboring voxel j for window w
b_k^A	BoW encoding columns obtained for k^{th} Gaussian

$bc(i)$	Betweenness centrality of node i
C	Number of classes
c	Class label
D	Dimension of MADs
$d(i, j)$	Distance between the spatial coordinates of nodes i and j in 3-dimensional space
d_{ij}^{sh}	Length of the shortest path between nodes i and j
D_l	Detail subband of level l
D_w	Number of measurements recorded in window w
$deg_{in}(i)$	In-degree of node i
$deg_{out}(i)$	Out-degree of node i
E	Number of classifiers
e_{ij}	Arc between nodes i and j
F_w	Feature vector for window w obtained by concatenating all MADs
$F_{s,w}$	Feature vector of MADs extracted for subband s and window w
$G_{s,w}$	Mesh network formed for subband s , window w
$globe f$	Global efficiency of mesh network
i	Node identifier
j	Node identifier
K	Number of clusters or Gaussians
k	Cluster or Gaussian identifier of k-means and GMM algorithms
L	Number of decomposition levels
l	Decomposition level
M	Number of samples

$NN(\mathbf{a}_i)$	Nearest codeword of MAD $\mathbf{a}_i$
p	Mesh size (number of neighboring nodes)
q	Location to which the wavelet is translated in time
$Q_{w,i}$	Matrix consisting of BOLD responses obtained from p -nearest neighbors of a seed node i for window w
r	Anatomical region
s	Subband
$str_{out}(i)$	Out-strength of node i
T	Duration of a signal
t	Time
u_β	Gaussian mixture
u_k	k^{th} component of GMM
v	Voxel
w	Window
$X_i(t)$	Time series of node i
$X_r(t)$	Time series of region r
$X_v(t)$	Time series of voxel v
$x_{s,i}(t)$	Reconstructed signal of node i at subband s
$x_{s,w,i}(t)$	Time series of node i obtained for window w at subband s
$x_{w,i}(t)$	Intensity value of the BOLD response measured at time t from node i within window w

CHAPTER 1

INTRODUCTION

I am a brain, Watson. The rest of me is a mere appendix. Therefore, it is the brain I must consider.

Arthur Conan Doyle

1.1 Motivation and Problem Definition

How human brain works has been an enigma, a center of interest, a research object for many centuries. In the the early nineteenth century, the bumps and flattening of skull was associated to the amount of feelings such as love or honesty. Later on, scientists have explored physiology of brain and conducted invasive experiments on animal brains to relate brain regions with brain functions. Yet, since invasive experiments could not be conducted on humans, a large number of cognitive aspects of human brain has remained unacknowledged.

With the advent of technology, researchers are now able to take measurements from human brain via Electroencephalography (EEG), Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), etc. while the participants are at rest or performing cognitive tasks. The non-invasive nature of these modalities have led to a boost in the studies trying to understand the functions of human brain. Although thousands of studies have approached how human brain works from different perspectives, mankind is still very far from fully understanding it. Barack Obama summarizes the current situation of brain research by: "As humans, we can iden-

tify galaxies light-years away, we can study particles smaller than an atom. But we still haven't unlocked the mystery of the three pounds of matter that sits between our ears". Currently, majority of the studies benefit from machine learning, statistical data analysis, signal processing and data mining tools to explore the functions of human brain.

Studies focusing on how human brain works mainly rely on comparison. In order to understand active regions under a specific cognitive task, researchers mainly compare brain activation maps while the participants perform a cognitive task with the ones while the participants are at rest. Moreover, to identify causes of diseases, researchers compare the brain activations from people having diseases versus healthy people. Even in smaller granularity, within a visual object recognition experiment, stimuli belonging to different categories such as face or house are presented to the participants. To understand the effect of emotion, participants are exposed to emotional versus neutral stimuli. In addition, studies aim to discover gender, age or ethnicity differences in brain activation under a number of cognitive tasks. As a result, the current findings are based on the differences between man vs. women [40, 75], young vs. old [30], emotional vs. neutral [91, 97], healthy vs. diseased [5, 28, 73, 128, 147], and so on. The goals of the current studies are to find better representations, to design better classifiers, to generate more discriminative frameworks and at the end, to discriminate these conditions better.

In this thesis, we swim in the shore of the ocean of how human brain works. We employ fMRI data as our modality and focus on finding better representations and better frameworks to discriminate cognitive states and gender by benefiting from machine learning and signal processing. We construct this thesis on two contrasting concepts namely *similarity* and *diversity*.

In a recent study, it is shown that although weak connections far outnumber strong connections, each neuron receives the majority of its local excitation from a small number of strong inputs provided by the few neurons with similar responses to visual features [31]. Following that finding, at coarser scales we employ the *similarity* between neighboring voxels or anatomical regions within a locality for representation. We focus on modeling their relationships in time to obtain more discriminative

representations than the state-of-the-art.

On the other hand, during the construction of our frameworks, we follow the approach that *diversity* brings more information. Wisdom of the crowd principle states that the collective opinion of a group of individuals are as good as, and often better than, that of a single expert. By training classifiers with multi-resolution mesh networks, we obtain *diversity* and we fuse the diverse information obtained from a group of classifiers for better discrimination of cognitive tasks and gender.

1.2 Contributions of the thesis

We can list the contributions of this thesis as follows:

1. We propose to model relationship among a set of fMRI time series within a neighborhood by considering temporal information obtained from all of the measurements in time series. The estimated relationships, called Mesh Arc Descriptors (MADs) outperform state-of-the-art fMRI representations including pairwise correlations or raw fMRI time series.
2. We adapt encoding approaches, which are mainly employed to encode local features in computer vision domain, to encode our local MADs. We obtain better representation power by encoding MADs with Fisher Vectors compared to raw MADs and other state-of-the-art feature representations to classify cognitive states.
3. We propose a *brain connectivity dictionary* by using the codewords, which are the means of Gaussians in Gaussian Mixture Models fit to MADs. We visualize the codewords, revealing connectivity patterns encoded in MADs on human brain template.
4. We identify a relationship between the energy of Fisher Vector columns obtained from the Gaussians in the mixture and classification accuracy. Therefore, we can analyze and visualize the codewords representing connectivity patterns in MADs based on their discriminative power (identified by the energy of FV columns of corresponding Gaussian).

5. We propose a framework called Hierarchical Multi-resolution Mesh Networks (HMMNs) in which we fuse the diverse information obtained from multi-resolution mesh networks under a Fuzzy Stacked Generalization (FSG) architecture. We show that different subbands represent complementing information and fusion of mesh arc descriptors (MADs) extracted from multiple subbands boosts the classification accuracy.
6. We propose a two-level hierarchical FSG architecture to fuse the decisions of multi-task multi-resolution mesh networks for gender classification. We show that, fusing the representations computed at multiple resolutions for multiple tasks provides the best gender classification accuracy compared to employment of single-resolution single-task mesh networks and fusion of representations obtained using only multi-task or only multi-resolution data.

1.3 Organization of the thesis

The outline of this thesis is as follows:

In Chapter 2, we present the literature and state-of-the-art knowledge about Multi-voxel Pattern analysis (MVPA), fMRI representations employed for MVPA, encoding methods applied on local descriptors in various domains, multi-resolution fMRI analysis, methods for fusion of classifier decisions and gender classification.

In Chapter 3, we introduce the experimental setup and collection of fMRI datasets employed in this thesis.

In Chapter 4, we propose Mesh Arc Descriptors (MADs) that models relationships within spatial and functional neighborhood as our fMRI representation. We then propose a method to encode our local MADs using Fisher Vectors, Vector of Locally Aggregated Descriptors and Bag-of-Words methods.

In Chapter 5, we propose a novel hierarchical framework to fuse the information obtained from multi-resolution mesh networks for cognitive state classification. We further propose a two-level hierarchical framework to fuse the information obtained from multi-task multi-resolution mesh networks for gender classification.

In Chapter 6, we perform classification experiments and significance analyses to show the success of proposed MADs, encoding of MADs and the proposed hierarchical frameworks to discriminate cognitive state and gender.

In Chapter 7, we summarize our methods and findings and conclude the thesis by introducing future directions.

CHAPTER 2

AN OVERVIEW OF REPRESENTATION AND ENCODING OF FMRI DATA FOR COGNITIVE STATE AND GENDER CLASSIFICATION

A dwarf on a giant's shoulders sees
farther of the two.

George Herbert

We begin this chapter by surveying the literature on Multi Voxel Pattern Analysis (MVPA), a supervised classification problem where a classifier attempts to capture the relationships between spatial pattern of fMRI activity and cognitive experiments. We review the state-of-the-art representations of fMRI signal used for MVPA. Then, we overview the applications of popular encoding approaches, namely Fisher Vector, Vector of Locally Aggregated Descriptors and Bag-of-Words methods, in various domains. We follow this overview with the related work on multi-resolution analysis of fMRI data. We also present the studies in which multiple classifiers are trained with fMRI data and their decisions are fused. Finally, we end this chapter with the related work on gender differences captured from fMRI data and gender classification using fMRI.

2.1 Multi-Voxel Pattern Analysis

The techniques employed to study the brain activity patterns from the BOLD signals using machine learning methods, called Multi-Voxel Pattern Analysis (MVPA), be-

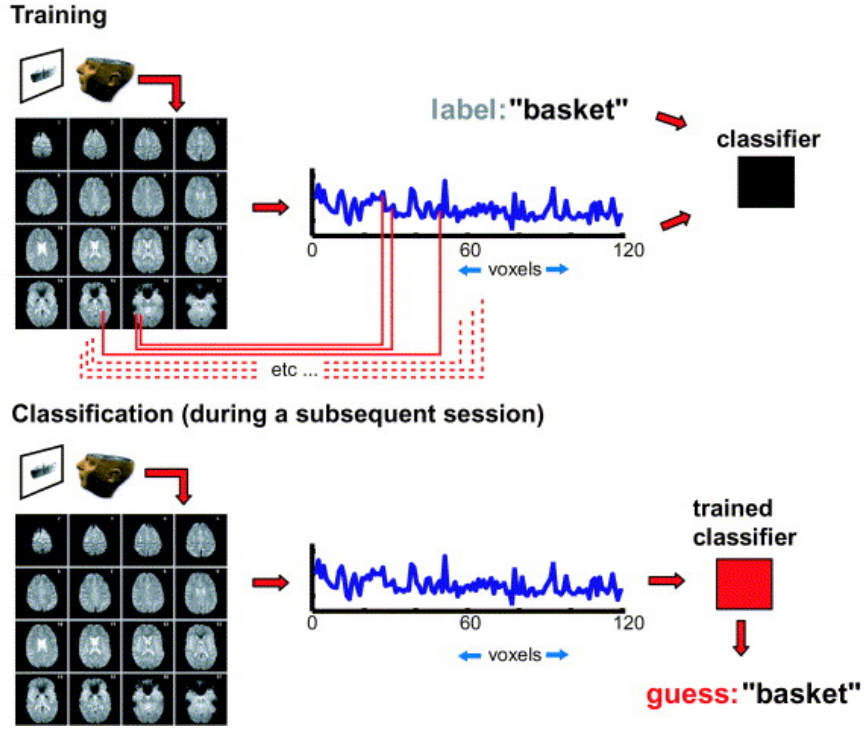


Figure 2.1: An example MVPA pipeline [33]. Classes of the presented images are used as labels and the measurements recorded from a number of voxels during the presentation of images are used as inputs to train a classifier. During test time, the classifier directly assigns a label to a new set of measurements recorded from voxels.

came popular in the last few decades (see Figure 2.1). Haxby et al. [59] suggested the pioneering method showing how multi-voxel patterns of brain activity in ventral temporal cortex can be used to discriminate between different cognitive tasks. In their study, subjects viewed objects from different categories which are faces, cats, non-sense objects and five categories of man-made objects. Following this pioneering study, machine learning techniques have been used for diagnosing disorders [74, 106], hypothesis validation [110] and classifying cognitive states which is a method for predicting cognitive states, called brain decoding.

There is a vast body of literature on MVPA that focus on discriminating various cognitive states. Among them, Kamitani et al. [69] proved the ability of MVPA to decode the eight orientations the subject was viewing. Moreover, in their following study, Haynes et al. [60] showed that, multi-voxel patterns of brain activity in visual cortex can be used to discriminate between unconscious representation of orientations.

Davatzikos et al. [35] used MVPA approach to discriminate patterns of brain activities measured by fMRI during truth-telling or lying experiment. Mitchell et al. [89] proved that MVPA methods can be used to decode whether subject is presented a sentence or a picture, whether the subject is presented an ambiguous or non-ambiguous sentence and which of the twelve categories (fruits, tools, etc.) does the presented picture belong. Polyn et al. [118] showed that MVPA methods can be used in memory retrieval tasks. Subjects were presented pictures from three different categories namely faces, houses and objects in the encoding phase. Then, they were expected to recall them in the retrieval phase. In this work, MVPA was used to detect how similar the patterns of brain activity belonging to a category are in the encoding and retrieval phases. In one of the recent studies, Hu et al. [64] examined whether power spectral density (PSD) profiles can be decoded from brain activities during naturalistic auditory experience. Moreover, Yoshida et al. [157] trained SVM classifiers to decode human emotion induced by visual stimuli using fMRI brain signal. Kubilius et al. [76] showed how MVPA allows to compare various possibilities of part combinations into wholes, such as taking the mean, weighted mean, or the maximum of responses to the parts. Additionally, Floren et al. [48] decoded visual information from fMRI data obtained in a realistic virtual environment. Toivainen et al. [143] investigated neural correlates of musical feature processing with a decoding approach. They decoded musical features from fMRI data collected using a naturalistic paradigm, in which brain responses of participants were recorded while they were continuously listening to pieces of real music.

In several studies, MVPA methods were used as powerful diagnosis tools. Craddock et al. [34] trained Support Vector Machine (SVM) classifiers with resting state functional connectivity patterns from healthy participants and patients having major depression. They proved that, activations of patients and healthy participants can be discriminated using MVPA methods. Moreover, Shen et al. [134] proved that, using patterns of resting state functional connectivity in machine learning tools can discriminate schizophrenic patients from healthy participants. Furthermore, brain activations measured by fMRI of patients having Autism Spectrum Disorder (ASD) and that of healthy people were discriminated using MVPA methods [32].

2.2 Representation of fMRI Signal

In traditional MVPA approaches, cognitive states are usually represented by concatenating the selected voxel intensity values to construct a vector in a feature space. The time series recorded at each voxel can be represented by various methods, such as computing maximal or mean values. Also, dimensions of the feature space can be reduced by Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Then, a classifier, such as, Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) or Naive Bayes, is trained on the feature vectors.

State-of-the-art methods use machine learning algorithms to design models for brain decoding [114]. For example, Graphnet [56] and TV-11 [7, 54] methods employ a spatial regularization technique for classification and regression problems of brain imaging.

In addition to the studies that design models, there exists a number of techniques to extract meaningful information from fMRI signals. A popular method for extracting spatial features from fMRI data is to average the BOLD responses measured at each voxel to represent a stimulus [33, 69, 70]. In their studies, Ng et al. [93, 95] consider each brain volume as a sample. On the other hand, Mitchell et al. [89] and Ng et al. [94] concatenate the BOLD responses measured in the same trial to represent spatio-temporal information.

Various studies [6, 45, 46, 111, 120] compute the pairwise correlations between the responses of voxels or anatomical regions to extract the temporal information from the fMRI data. Among these studies, Pantazatos et al. [111] represent temporal relationships by using pairwise correlations between anatomical regions as features for brain decoding. Moreover, Richiardi et al. [120] construct a functional connectivity graph using the pairwise correlations between anatomical regions, and they decode the cognitive states using graph matching algorithms. Vertices or edge weights of the constructed brain graphs are used as features for cognitive state classification [119]. At a finer level, Firat et al. [45] suggested to employ pairwise correlations of voxels as features whereas Baldassano et al. [6] extract the functional connectivity structures by modeling pairwise connectivity among voxels. Also, a minimum spanning

tree of a brain graph constructed using a functional connectivity matrix between all pairs of voxels is used to represent the brain states for cognitive state classification [46]. Notice that, the above mentioned studies propose temporal methods, which model only the pairwise relationships between voxels or anatomical regions. Although these methods represent the temporal information to a certain degree, none of them represents the temporal relationships within a neighborhood.

Functional and spatial similarities of voxel BOLD responses are observed in a pre-defined locality. Bazargani et al. [11] indicate that neighboring voxels belonging to a homogeneous region-of-interest have Hemodynamic Response Function (HRF) with the same shape and with slightly varying amplitude. In other words, spatially close voxels give similar BOLD responses to the same stimuli. Moreover, Kriegeskorte et al. [96] propose Searchlight analysis, in which maps are produced by measuring the information in small spherical subsets (searchlight) centered on every voxel. The map value for each voxel is obtained using the information present in its searchlight, not from the voxel individually. They report that univariate model of activations, which do not benefit from the local combination of signals, performs worse than Searchlight methods to detect informative regions.

Ozay et al. [108] remark that the voxel time series do not differ significantly to discriminate cognitive states, yet slight variations exist among the intensity values of voxels located in a spatial neighborhood. They propose Local Mesh Model (LMM) to model the relationship among the spatially neighboring voxels. They show that features of LMM perform better than the raw voxel intensity values for the classification of cognitive states. However, they employ only the brain volume, which is obtained 6 seconds after the stimulus presentation, for the construction of meshes while the remaining brain volumes are discarded under the canonical HRF assumption.

Finally, Firat et al. [47] propose a Functional Mesh Model (FMM), in which local meshes are constructed around voxels by selecting the neighbors within a functional neighborhood. Their experimental results show that FMM features are more discriminative than LMM features. However, they also estimate the mesh weights using only a single intensity value for each voxel.

In none of the aforementioned models, the time series recorded at all the voxels are

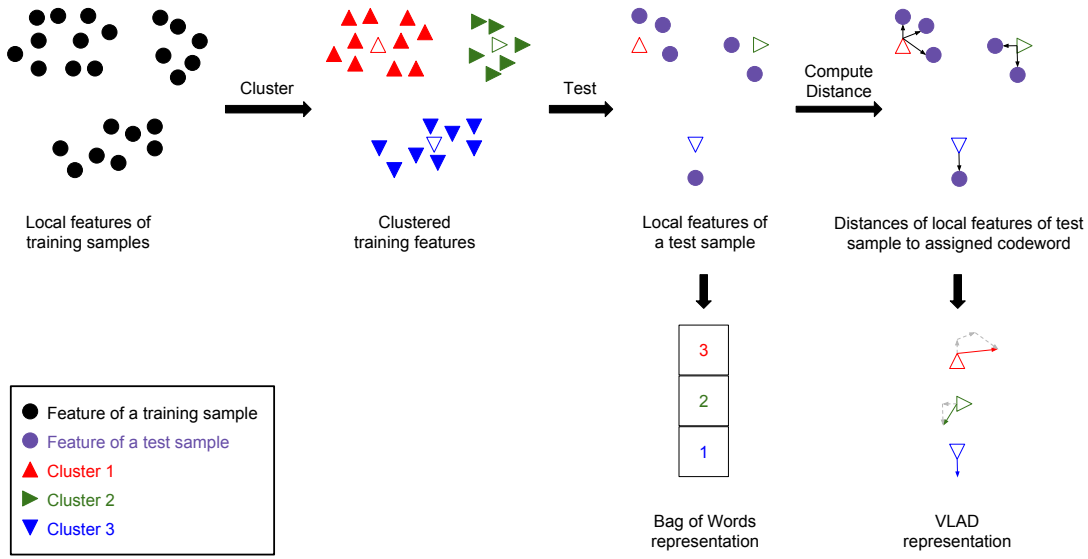


Figure 2.2: Steps of encoding local features using BoW and VLAD.

fully employed to represent the spatial relationships among the voxels of the fMRI data.

2.3 Encoding Methods

In this thesis, we use the term *encoding* to mean "learning a dictionary from the data and representing samples based on the codewords of the dictionary". Encoding approaches are widely used in pattern recognition literature to improve the representation power of local descriptors. In a popular encoding approach, called Bag-of-Words (BoW), first local descriptors are clustered using the k-means clustering method. Then, cluster centroids are defined as codewords to form a dictionary. Codewords are utilized as *textual words* in natural language processing, and *visual words* in image processing. Each descriptor is assigned to the closest codeword, and a sample is represented by a histogram of codewords, to which its local descriptors belong (see Figure 2.2). BoW approach has been used to detect diseases [138] and cognitive states [139] using fMRI data, and to classify EEG time series data [149, 88].

In another encoding approach, called Vector of Locally Aggregated Descriptors (VLAD),

codewords are computed similar to BoW. However, VLAD aims to accumulate the difference between the codewords and local descriptors assigned to codewords [68]. It has been used in various applications including image [37] and video [1] processing, yet it has not been used to encode local descriptors obtained from fMRI data.

Fisher Vector (FV) encoding methods are employed for statistical data analysis by making use of the benefits of generative and discriminative models [26, 126]. FV encoding is considered as an extension of BoW such that, rather than encoding the relative frequency of the descriptors, it encodes the information on distribution of the descriptors [116]. In other words, FV represents deviation of a sample from a generative model, where the deviation is computed as the gradient of the sample log-likelihood with respect to the model parameters [126].

Fisher kernels have been used to compute FVs utilizing a mechanism that incorporates generative probability models into discriminative classifiers in a number of applications. Jaakkola et al. [66, 67] applied Fisher kernel to detect remote protein homologies that perform well to classify protein domains by SCOP (Structural Classification of Proteins) superfamily and showed that using Fisher kernel boosts the classification accuracy. Moreno et al. [92] applied Fisher kernel for large scale Web audio classification. Moreover, Fine et al. [44] built a hybrid Support Vector Machine classifier that combines Gaussian mixture model (GMM) and Fisher kernel and applied it to speaker identification. In one of the recent studies, Oneata et al. [105] proposed an efficient action recognition system that combines three state-of-the-art low-level descriptors namely, dense motion boundary histogram (MBH), scale-invariant feature transform (SIFT) and mel-frequency cepstral coefficients (MFCC) with Fisher vector representation. Sekma et al. [133] suggested a multi-layer Fisher vector encoding method based on trajectory descriptors for human action recognition. FV encoding has proven to be among the most effective robust feature pooling techniques for object recognition [29]. FVs have also been used for image classification [84, 127, 135] and 3D object retrieval [130].

Broadersen et al. [18] applied generative embedding methods to neuroimaging data by using dynamic causal models of local field potential recordings for single-trial decoding of cognitive states. Moreover, generative embedding methods were used

to train discriminative classifiers for clinical applications, e.g. for classification of disease mechanisms in individual patients [19]. Yet, these studies did not form FV to represent the data. FVs have been applied to model effective connectivity of the networks using MRI and PET data [158, 159]. However, FVs have not been used to encode the connectivity patterns of fMRI data, which is very crucial to analyze the behavior of brain during cognitive tasks.

2.4 Multi-resolution Representation of fMRI Data

It is known that brain processes information in multiple frequency bands, and different frequencies of neuronal activity have been linked to the BOLD signal [72, 141]. Features of spoken sentences, visual stimuli or development of social interaction may unfold over distinct time scales [72]. Kauppi et al. [71] report that distinct regions exhibit inter-subject correlations at low, medium or high frequencies. These studies imply that different regions of the brain discriminate the conditions in different sub-bands. Therefore, multi-resolution analysis of fMRI signal is crucial for analyzing and decoding the cognitive states.

There has been a growing interest in multi-resolution analysis on fMRI data, especially resting state fMRI data. Low-frequency oscillations (LFO) have frequency specific property that contributes differently to functional connectivity [124]. In their study, Wee et al. [152] employed a frequency-specific approach to better characterize the subtle BOLD signal variations related to Mild Cognitive Impairment (MCI) and employed a graph measure to extract topological information of brain functional connectivity networks. Motivated by this study, Li et al. [83] decomposed the frequency interval into five equally non-overlapping frequency sub-bands to extract complex and subtle pathology information associated with MCI, enabling a more frequency specific analysis for the regional mean time series. Recent fMRI studies have extended the concept that independent frequency bands are generated by distinct oscillators, and each of them has specific properties and physiological functions [23], by decomposing the fMRI signals into multiple bands [9, 61, 160].

Wavelet transforms are widely used to represent the fMRI signals in multiple res-

olutions with approximately decorrelated coefficients [145]. Bullmore et al. [21] reported that the brain has fractal property (also called $1/f$ -like property), where the statistical properties that describe the structure of a system, in time or space, do not change over a range of different scales [86]. Wavelet analysis is similar to Fourier analysis in the sense that it decomposes a signal into its basic constituent components [156]. However, compared to smooth and infinitely periodic sine wave decomposition of Fourier transform, wavelet transform is irregular in shape, which is suitable for the analysis of non-stationary signals [39]. Therefore, adaptivity of wavelets to local or non-stationary features of an fMRI signal makes them suitable choices for the analysis of the fMRI signal, which is expected to include non-stationary features of interest at several scales [21]. Furthermore, Discrete Wavelet Transform (DWT) has decorrelating capability for a wide class of signals having $1/f$ -like property. In other words, even if the data is highly correlated, the correlations computed between wavelet coefficients are generally small. There have been previous fMRI studies of frequency dependency of functional coherence between regions in the Fourier domain [125, 140]. On the other hand, Achard et al. [3] have applied DWT to fMRI time series, acquired from healthy volunteers in the resting state, to estimate frequency-dependent correlation matrices characterizing functional connectivity between anatomical regions. They have found a small-world topology of sparse connections most salient in the low-frequency interval [0.03–0.06 Hz].

In this study, we assume that fMRI signals, which are reconstructed by the decorrelated wavelet coefficients for different time resolutions, carry complementary information in the corresponding feature spaces. This assumption is supported by the study of Richiardi et al. [120] which shows that the multi-resolution signals obtained with an orthogonal DWT are quasi class-conditionally independent. Therefore, we expect that the classifiers trained using multi-resolution signals are diverse, and fusion of their decisions would yield high performance for decoding the brain signals. This approach requires a fast and efficient decision fusion method to ensemble the classifiers trained by the fMRI signals at different resolutions.

2.5 Fusion of Classifier Decisions

Classifiers that are not very accurate individually but tend to make mistakes on different objects may form a very accurate ensemble [79]. Ensemble learning classifiers are used in many fMRI studies, including brain decoding. In a pioneering study by Kuncheva et al. [81], multiple classifiers are trained and their decisions are ensembled using bagging [15], Adaboost [49], Random Subspace [62], Random Forest [16], Rotation Forest [122] and Random Oracle [80, 121] methods. They observed that random subspace ensembles of SVM classifiers and Rotation forest give the best accuracy. Recently, Alkan et al. [4] partition the brain into homogeneous regions with respect to functional similarity of voxel time series and train a different classifier at each region. Multiple classifiers are also trained by a set of complementing stimuli [24] or subbands [120]. Then, the results of the classifiers are ensembled by majority voting techniques. These ensemble learning methods have been shown to outperform the methods based on a single classifier.

Fuzzy classifiers are used for classification of fMRI patterns. Wang et al. [150] employed distance-based fuzzy classifier for group separation of pediatric epilepsy patients into left dominant, right dominant, bilateral and others using fMRI patterns. Moreover, Liu et al. [85] used fuzzy SVM for multi-class brain state decoding. Kuncheva [78] state that fuzzy combination methods for ensemble of classifiers fare better than non-fuzzy ones. Fuzzy combination of classifiers is also used to classify fMRI data. Parida et al. [112] used Fuzzy Integral method to fuse decision of ensembles for cognitive state classification. Ozay et al.[107] proposed a hierarchical ensemble learning method, called, fuzzy stacked generalization (FSG), which fuses the decisions of multiple classifiers by a meta classifier. This method is shown to outperform a number of ensemble learning methods including Adaboost, Random Subspace, Rotation Forest methods in Multi-Voxel Pattern Analysis (MVPA) [109].

2.6 Gender Classification

The advent of brain-imaging techniques such as positron emission tomography (PET) and functional Magnetic Resonance Imaging (fMRI) has increased the awareness of

sex differences by demonstrating sex influences on brain functions for which the sex of participants was previously assumed to matter little [25]. Early studies have focused on identifying gender differences based on the activations in anatomical brain regions. They revealed sex differences under a range of conditions including creative thinking [2], lying [87], language processing [10], autobiographical memory [117], emotion processing [63], planning [14].

Gender differences have also been analyzed in functional brain networks. These studies mainly identify gender differences using resting state functional MRI (R-fMRI) data. Biswal et al. [13] demonstrated evidence of systematic sex differences in R-fMRI functional connectivity. Moreover, Tial et al. [142] investigated gender related differences in the topological organization of human brain functional networks formed using R-fMRI. Wu et al. [154] analyzed the effect of sex on the developmental changes of functional brain networks derived from functional connectivity. Wang et al. [151] employed a support vector machine-based MVPA approach to identify the complex patterns of sex differences in brain structure and resting-state function. Satterthwaite et al. [129] showed that sex differences in patterns of brain connectivity can be identified at late childhood, and these differences are related to sex-specific profiles of cognitive performance.

Contrary to R-fMRI studies, several studies have examined gender differences in patterns of functional connectivity in task conditions [53]. Among them, Schmithorst et al. [131] investigated gender differences of functional connectivity during silent verb generation task. They identified a significant interaction among sex, age and IQ in the functional connectivity between several brain regions in the left hemisphere. The same group also observed gender differences in the interaction between intelligence and functional connectivity for narrative comprehension task [132]. Moreover, Butler et al. [22] showed that only women had anticorrelated functional connectivity between the ventral anterior cingulate cortex (vACC) and the dorsal ACC (dACC), during a visuospatial task of mental rotation. They denoted that gender difference in the vACC-dACC connectivity may represent gender specificity in the interaction between cognition and emotion. These studies indicate that functional connectivity differences are observed for different genders under different cognitive tasks.

Recent studies have approached gender analysis as a classification problem. Duarte-Carvajalino et al. [40] classified individuals in terms of their sex using the brain connectivity networks. Vogelstein et al. [148] performed graph classification using signal subgraphs and classify brain graphs according to sex. Kulkarni et al. [77] showed that the connectivity structure of the brain admits classification of connectomes into sexes based on their connectivity characteristics. Ktena et al. [75] proposed a topological graph kernel for gender classification of functional brain networks. Results of these studies imply that gender can be identified by the classification of brain networks.

2.7 Chapter Summary

In this chapter, we provided an overview of representations and encoding methods for cognitive state and gender classification. We first introduced Multi-Voxel Pattern Analysis (MVPA) and studies employing patterns of fMRI data for brain decoding or identification of diseases. Next, we briefly explained different representations of fMRI data, containing or lacking spatial or temporal information, for classification. We figured out that, representation of fMRI signal affects the success of the classifiers and considering temporal information and relationship within a neighborhood is crucial when obtaining representations. Then, we presented the studies employing popular encoding methods namely FV, VLAD and BoW in various domains. The power of Fisher Vectors in various domains led us to explore the potential of the FV encoding on local features obtained from fMRI data.

We reviewed studies performing multi-resolution analysis using fMRI data and the suitability of wavelet transform for multi-resolution analysis of fMRI signal. We deduced that since the brain processes information in distinct frequencies, multi-resolution analysis of fMRI data is important. Then, we overviewed the crisp and fuzzy ensembling methods applied for fMRI classification. We figured out that, fusing the decisions of classifiers leads to better classification accuracy compared to the decision of a single expert classifier. Finally, we surveyed the studies attacking gender differences from fMRI data and machine learning studies performing gender classification using fMRI data. Our findings from literature encouraged us to discover the potential of gender classification under multi-resolution and multi-task setting.

CHAPTER 3

EXPERIMENTAL SETUP AND FMRI DATA COLLECTION

Hiding within those mounds of data is
knowledge that could change the life of
a patient, or change the world.

Atul Butte

In this chapter, we introduce the cognitive experiments and the preprocessing steps of fMRI datasets employed in this thesis. Three sets of fMRI data are tested to evaluate the proposed approaches. The first two datasets, namely, visual object recognition and emotional memory retrieval experiments, are designed and recorded by our team, in Middle East Technical University and Koc University. The third dataset is taken from Human Connectome Project (HCP). In the following sections, we will explain fMRI data collection and pre-processing steps in detail.

3.1 Visual Object Recognition Experiment

In this experiment, fMRI measurements were recorded while participants performed a one-back repetition detection task. The stimuli consist of gray-scale images belonging to two categories, namely, birds and flowers. Each trial lasts for 12, 14 or 16 seconds containing stimulus presentation for 4 seconds, and a following rest period of 8, 10 or 12 seconds. Scanning was performed on a Siemens 3T Magnetom TRIO MRI system. Functional images were acquired using a gradient EPI sequence (TR = 2000 msec, TE = 30 msec, flip angle = 90° , 34 interleaved axial slices, voxel size = $3\text{ mm} \times 3\text{ mm} \times 3\text{ mm}$ with 0.3 mm interslice gap). Since our TR = 2 seconds, we

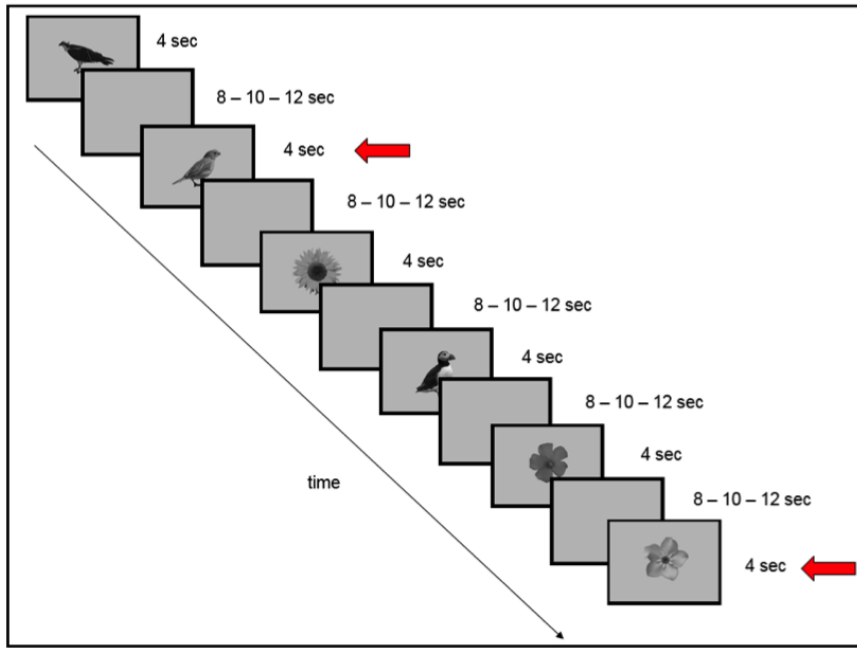


Figure 3.1: Visual Object Recognition Experiment. Participant is presented with bird or flower pictures. The participant is expected to indicate by pressing the button when the current stimulus matches the one from previous step. In this example, second and fifth pictures pointed with red arrows have the same categories (bird and flower, respectively) with their predecessors.

obtain 6 measurements (scan volumes) for each trial.

This experiment consisted of 6 runs (sessions), and each run contains 36 trials. Therefore, we obtained a total of 216 trials (each having 6 measurements corresponding each time instant) for a single participant. A total of 1512 images were collected for a single participant. We collected data from 5 participants.

3.2 Emotional Memory Retrieval (EMR) Experiment

In this experiment, the stimuli consist of two neutral (Kitchen utensils and furniture) and two emotional (Fear and Disgust) categories of images. Each trial started with a 12 seconds fixation period followed by a 6 seconds encoding period. In the encoding period, participants were presented with 5 images from the same category, each image lasting 1200ms on the screen. Following the fifth image, a 12 seconds delay period

was presented in which participants solved three math problems consisting of addition or subtraction of two randomly selected two-digit numbers. Following the third math problem, a 2 seconds retrieval period started in which participants were presented with a test image from the same category, and indicated whether the image was a member of the current study list or not. For a similar experimental setting, please refer to [90].

This experiment consisted of 6 19-min runs and each run contains 35 trials. A total of 210 trials and 3456 images were obtained for a single participant. Scanning was performed on a Siemens 3T Magnetom TRIO MRI system. Functional images were acquired using a gradient EPI sequence (TR = 2000 msec, TE = 30 msec, flip angle = 90° , 34 interleaved axial slices, voxel size = $3\text{ mm} \times 3\text{ mm} \times 3\text{ mm}$ with 0.3 mm interslice gap). We collected data from 13 participants. We employed 6 measurements (6 seconds of encoding and 6 seconds of following rest) of each trial for classification. Notice that, we only employed the data obtained during the encoding phase.

3.3 Human Connectome Project (HCP) task fMRI Data

We use task fMRI data of Human Connectome Project (HCP) collected from 808 participants. During the experiments the participants performed seven tasks namely Emotion Processing, Gambling, Language, Motor, Relational Processing, Social Cognition, Working Memory (WM) [8].

We summarize the experimental design for each task below:

- **Emotion:** This task is a modified version of the emotion task employed in Hariri et al. [58]. The participants are presented with blocks of trials that ask them to decide either which of two faces presented at the bottom of the screen match the face at the top of the screen (faces have fearful or angry expressions), or which of two shapes presented at the bottom of the screen match the shape at the top of the screen.
- **Gambling:** In this task, the participants play a card guessing game where they are asked to guess the number on a mystery card (represented by a “?”) to win

or lose money. Details of this experiment can be found in the study of Delgado et al. [36].

- **Language Processing:** This task consists of two subtasks namely a story task and math task. After the participants listen a brief story, they answer the question by selecting either alternative answers in the story task. In math task, they are presented a mathematical question and similarly they select the result among two choices. For more details about the experiment, please refer to Binder et al. [12].
- **Motor:** In this task, participants are instructed by a visual cue to move one of five body parts (left or right finger, left or right toe, or tongue). For details, please see Buckner et al. [20].
- **Relational Processing:** This task consists of relational processing and control matching conditions. In the former one, participants are shown two pairs of objects. First they decide whether the first pair differs in either the shape or texture. Then they decide whether the second pair also differs in the same dimension. In the latter one, participants are presented a pair of objects at top of the screen, an object at the bottom of screen and a word (shape or texture) in the middle. They first decide whether the pair differs in the dimension written in the middle. Then they decide whether the object at the bottom has the same dimension (shape or texture written in the middle) with either objects residing at the top. More details can be found in Smith et al. [136].
- **Social Processing:** The participants are presented with video clips of objects (squares, circles or triangles) either interacting in some way, or moving randomly. They can choose among the following three alternative options: i) the objects have a social interaction, ii) not sure and iii) the objects are moving randomly. The movie clips were originally prepared by Castelli et al. [27] and Wheatley et al. [153].
- **Working Memory:** In this task, participants are asked to complete two-back working memory tasks and zero-back tasks with four different types of image stimuli namely places, tools, faces or body parts (see Barch et al. [8]).

Note that the number of scans and their duration vary for each task yet we obtain equal duration for all participants (see Table 3.1). Although, there are total of 900 subjects in HCP task fMRI data release, 808 of them performed all of the 7 tasks. Therefore, we use the data obtained from 808 participants, and we discard the data obtained from the remaining participants.

In this experiment, we define *supervoxels* such that we average BOLD responses of voxels residing in the same anatomical region, and denote each *supervoxel* with that average time series. We use $R = 90$ anatomical regions of 116 AAL [144], after removing anatomical regions of Cerebellum and Vermis. Therefore, in our experiments, we have 90 *supervoxels*. Since our aim is to classify cognitive experiments rather than cognitive states within an experiment, we work in coarser granularity. Moreover, using this dataset, we perform inter-subject classification to see how well our framework performs for multi-subject classification.

Table 3.1: Number of scans per session and its duration for each task (min:sec) for HCP data.

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

3.4 Preprocessing and Analysis of Neuroimaging Data

SPM8 toolbox is used for preprocessing and analysis of fMRI datasets¹. Preprocessing of images consists of (a) correction of slice acquisition timing across slices, (b) realigning the images to the first volume in each fMRI run to correct for head movement, (c) normalization of functional and anatomical images to a standard template EPI provided by SPM2, and (d) smoothing images with a 6-mm full-width half-maximum isotropic Gaussian kernel. Finally, we extract 116 Automated Anatomical Labeling (AAL) regions [144] using Marsbar [17]. For the visual object recognition and emotional memory retrieval datasets, General Linear Model is implemented to find the distribution of active voxels across the anatomic regions. The active voxels in

¹ <http://www.fil.ion.ucl.ac.uk/spm/>

object recognition dataset are mostly located in occipital lobe, which is selected as the region of interest (ROI) for this experiment. On the other hand, the active voxels in the emotional memory retrieval dataset are scattered all over the brain regions. The most discriminative voxels are, then, selected by t-test among all the voxels of this dataset. We first split samples into two groups according to their class labels. Then, we compute p-values for each voxel that reflects how the two groups are well-separated by each voxel. For MAD analysis we select 800 voxels since almost 800 of them have $p < 0.005$. For encoding experiments, we select 100 voxels with $p < 0.001$ to work on a similar scale with HCP dataset. In both cases, we perform voxel selection repetitively for each training split of nested cross validation. Preprocessing details of HCP dataset are explained in [52]. For HCP dataset, rather than selecting the active voxels, we employ all of the 90 anatomical regions in the form of *supervoxels*, each of which is represented by the average region time series. This is a widely-accepted approach for the datasets, such as HCP, which consist of task experiments of long duration.

CHAPTER 4

A NEW REPRESENTATION OF FMRI DATA AND ITS ENCODING

Weak contacts in the brain have little
impact, despite being in the majority.

Thomas Mrsic-Flogel

In this chapter, we first focus on representing fMRI signals by a set of Mesh Arc Descriptors (MADs). That procedure is based on constructing local meshes around voxels or anatomical regions within a neighborhood and estimating the MADs representing their relationship. Following MAD representation, we will explain how we further encode MADs using state-of-the-art encoding methods including Fisher Vectors (FV), Vector of Locally Aggregated Descriptors (VLAD) and Bag of Words (BoW).

4.1 Mesh Representation of fMRI Signal with Respect to Functional and Spatial Neighborhoods

In this section, we propose a new approach to represent fMRI signals as a set of Mesh Arc Descriptors [101]. The major motivation of this mesh representation follows the observation that the strength of synaptic connections fundamentally determines how neurons influence each other's firing [31]. Although a neuron makes weak connections with a large number of other neurons, it obtains its local excitation from strong inputs provided by few neurons that give similar responses to the same stim-

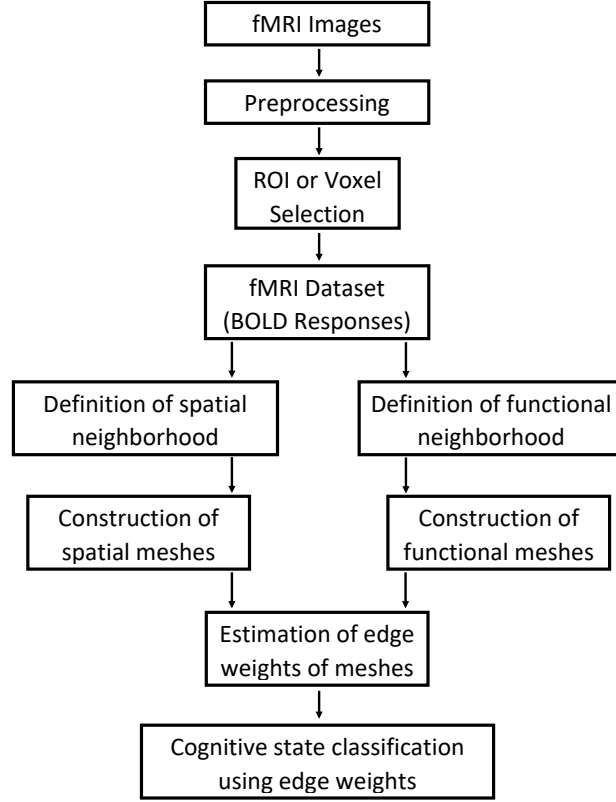


Figure 4.1: The flowchart of the proposed approach for obtaining local meshes and classification of cognitive states.

uli. Therefore, there exist spatial and temporal relationships among neurons residing within a locality, and the strength of these relationships determines the strength of neural response. It is observed that, at a coarser level, voxel BOLD responses are functionally and spatially correlated in pre-defined local structures, i.e. functional and spatial neighborhoods. Therefore, we propose representations which reflect the relationships in a locality.

The basic steps of the proposed method are summarized in Figure 4.1. In the first step, we preprocess the fMRI measurements and obtain voxel intensity values. In the second step, we select voxels or anatomic regions to reduce the redundancies and noise in our dataset. Then, we define spatial and functional neighborhood systems, and construct meshes, accordingly. Finally, we estimate arc weights of meshes, and input them to train and test a classifier to label the cognitive states.

The local mesh representation of fMRI signal, proposed in this study, is constructed

using a collection of nodes located within a neighborhood of each node, where nodes can be voxels or anatomical regions. For this purpose, first we make a formal definition of node time series. Next, we define locality by defining spatial and functional neighborhood systems. Then, we define local meshes and based on the neighborhood we obtain two types of meshes namely, spatially and functionally local meshes. In the following sub-sections, we provide the definitions of node time series, neighborhood systems, local meshes and mesh network.

4.1.1 Node Time Series

In this study, we form meshes around nodes, where a node i can be either a voxel or an anatomical region at different granularity levels.

Definition 1: Node Time Series: Let the time series, called BOLD signal, obtained at each voxel v be represented by a function of time t with $X_v(t)$. When we work at voxel granularity level, nodes correspond to voxels and node time series is denoted with $X_i(t) = X_v(t)$. On the other hand, we obtain a representative region time series $X_r(t)$ for each anatomical region r by spatially averaging the time series of voxels residing in that region using

$$X_r(t) = \frac{1}{V_r} \sum_{v \in r} X_v(t), \quad (4.1)$$

where V_r denotes the number of voxels residing in region r . When we work in anatomical region granularity, then nodes are regions and node time series $X_i(t) = X_r(t)$.

4.1.2 Neighborhood Systems

One of the most important contributions of this study is to construct a linear relationship among the BOLD responses of voxels by employing "locality" properties of the voxels. In order to achieve this goal, we define two types of neighborhood systems, namely, spatial and functional neighborhoods. In the case of spatial neighborhood, we compare 3-dimensional coordinates of voxels and in the functional case, we define similarity measures to compare voxel time series.

Definition 2: Spatial Neighborhood: For each node i , we define the spatial neighborhood centered at i with a radius π by $\eta(i, \pi) = \{j \in \mathcal{N} | d(i, j) < \pi\}$ where $\mathcal{N}$ denotes the set of nodes and $d(i, j)$ denotes the Euclidean distance between the spatial coordinates of nodes i and j in 3-dimensional space. If the number of nodes falling in the radius is p , then we use $\eta_p^{spat}[i]$ to denote p spatial neighborhood of node i .

Definition 3: Functional Neighborhood: Recall that, $X_i(t)$ denote the time series of BOLD responses obtained from node i , which is recorded as the response of a series of stimulus during the entire sequence of a cognitive experiment. In order to find the functionally nearest neighbor of a node i , we compute similarity (sim) between node time series $X_i(t)$ and $X_j(t)$ for all the node pairs $i, j \in \mathcal{N}$. Then, we select p -functionally nearest neighbors of a node as the nodes having p of the highest similarity with i .

A functionally nearest neighbor of a node i is defined as:

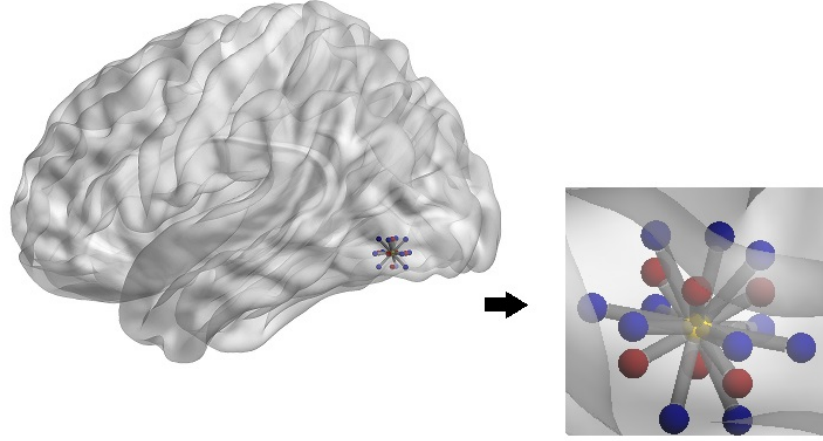
$$\eta_1^{func}[i] = \{j : sim(X_i(t), X_j(t)) \geq sim(X_i(t), X_o(t)), \forall o \in \mathcal{N}\}. \quad (4.2)$$

Finally, p -functionally nearest neighbors of the node i is recursively defined as follows;

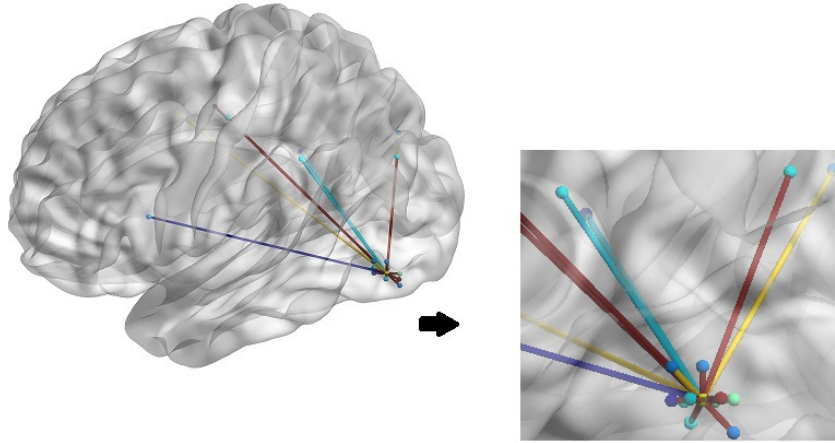
$$\eta_p^{func}[i] = \{j \cup \eta_{p-1}^{func}[i] : sim(X_i(t), X_j(t)) \geq sim(X_i(t), X_o(t)), \forall o \in \eta_{p-1}^{func}[i]'\}, \quad (4.3)$$

where $\eta_{p-1}^{func}[i]'$ is the set complement of $\eta_{p-1}^{func}[i]$. In our experiments, we employ Pearson correlation, partial correlation and Granger causality as functional similarity measures, since these measures are frequently used to compute pairwise similarities between nodes in the literature.

Figure 4.2 shows examples of meshes formed within spatial and functional neighborhood. Note that functionally p -nearest neighbors of a node may or may not be the same as the spatially p -nearest neighbors. Practical evidence indicates that most of the spatially close nodes are also functionally similar. However, some of the node pairs are spatially far apart, yet they are functionally close to each other (see Figure 4.2b).



(a) A spatially local mesh formed around a node.



(b) A functionally local mesh formed around a node.

Figure 4.2: A local mesh formed by (a) spatial neighbors and (b) functional neighbors of a seed node. In both figures, the node denoted with yellow color at the center is the seed node. In (a), red nodes denote the spatial neighbors of the yellow node within a radius of 1 unit and blue nodes have $\sqrt{2}$ unit distance to yellow node. In (b), a mesh is formed with functional neighbors around the seed node denoted with yellow color.

4.1.3 Construction of Local Meshes

Based upon the neighborhood systems introduced in the previous section, the concept of locality is represented by a set of meshes defined over a sequence of brain volumes. For this purpose, two types of meshes, namely, spatially and functionally local meshes, are established.

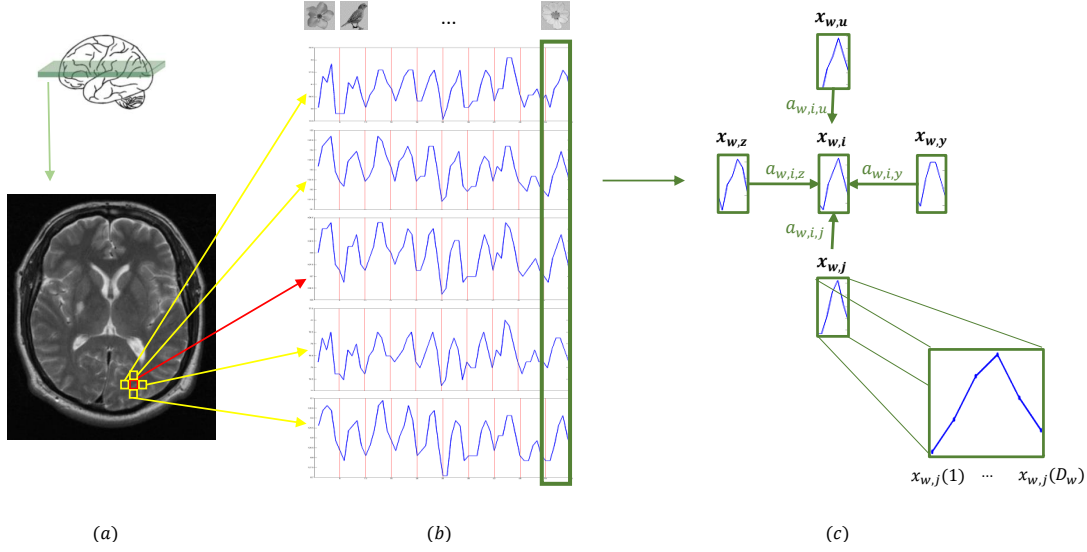


Figure 4.3: Design of a spatially local mesh for modeling the relationship among the BOLD responses of nodes. (a) For simplicity, a seed node is depicted in an axial slice at the center of a mesh (red) with its four spatially nearest neighbors (yellow). Neighboring nodes are selected within a 3D brain volume. (b) fMRI time series are recorded at each node, for each stimulus for bird and flower classes. (c) A local mesh is computed using the responses obtained from a seed node $x_{w,i}$ and its four nearest neighbors for window w . We also visualize node intensity values of a sample response $x_{w,j}$ obtained from node j .

Both types of meshes are constructed around each node of the brain volume using a predefined neighborhood system. Once we select the neighborhood system, we form a local mesh around each node by connecting it to its neighboring nodes in a star topology. The node located at the center of a mesh is called the *seed node*. Since we form the meshes around all of the nodes in the brain volume, the meshes defined over a neighborhood system may overlap. A seed node in a mesh may become a neighboring node of a seed voxel in a different mesh.

The formal definition of local meshes are given below:

Definition 4: Local Meshes: For each node i in the brain volume, a local mesh is defined as $\mathcal{M}_i = (\mathcal{N}_i, \mathcal{E}_i)$ where $\mathcal{N}_i = \{i \cup \eta_p[i]\}$ represents a set of the nodes of the mesh, and $\mathcal{E}_i = \{e_{ij} : \forall j \in \eta_p[i]\}$ is a set of arcs, called Mesh Arc Descriptors (MADs) formed between the seed node i of the mesh and its p -nearest neighbors.

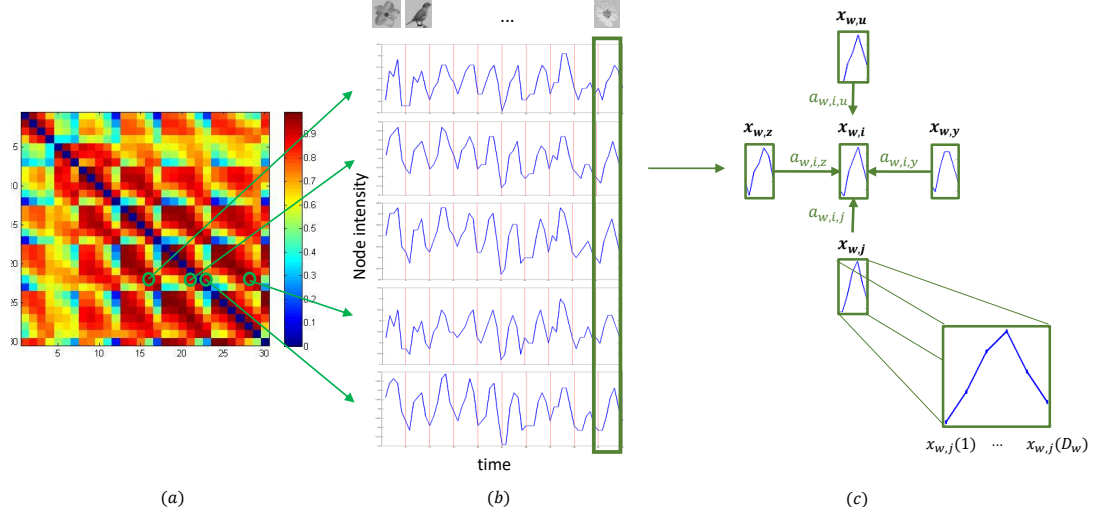


Figure 4.4: Design of a functionally local mesh using the voxel time series. (a) A sample functional connectivity matrix formed using Pearson correlation. p -functionally nearest neighbors of a voxel are denoted as the ones with the dark-to-light red colors. (b) fMRI measurements are recorded for each stimulus for bird and flower classes. (c) We depict a functional mesh formed around a seed voxel, in star topology. The edge weights are computed by using the response obtained from a seed node $x_{w,i}$ and the responses of its four functionally nearest neighbors, $x_{w,j}$, $x_{w,u}$, $x_{w,y}$, $x_{w,z}$, for window w .

Notice that η_p corresponds to η_p^{spat} if meshes are formed considering spatial neighborhood, and corresponds to η_p^{func} if the meshes are formed considering functional neighborhood.

When the spatial neighbors are used to form meshes, we call them Spatially Local Meshes (SLM) [104] (see Figure 4.3). On the other hand, when the nearest neighbors are selected based on the similarity of voxel time series, we call the meshes Functionally Local Meshes (FLM) [102] (see Figure 4.4).

For FLM, the same star topology is defined for all meshes with a fixed mesh size p . On the other hand, for SLM, different nodes may have different number of neighbors due to the definition of spatial neighborhood and topology of 3-dimensional locations of nodes. In both cases, the nodes and the arcs formed between them in a mesh do not change in a predefined time interval. However, the intensity values measured for each entry of the time-series representing the BOLD response, change in time. Since

we estimate a set of arc weights for each cognitive stimulus, they remain the same for the duration of a stimulus and only change across the stimuli.

Once we define a mesh around each node, we address the problem of estimating the arc weights. The proposed arc estimation method is explained next.

4.1.4 Estimation of Mesh Arc Descriptors (MADs)

Suppose that we record the BOLD response at a node i given to a stimulus within a window w in order to measure the brain activation for a predefined cognitive state with label c . We denote the intensity values of the BOLD response measured at a node i within window w by the vector $\mathbf{x}_{w,i} = [x_{w,i}(t)]_{t=1}^{D_w}$. For each window, we record D_w measurements.

We form meshes around nodes for each window. For a window, we denote the estimated weight of an arc between seed node of a mesh i and a neighboring voxel j by $a_{w,i,j}$. We estimate the arc weights, called Mesh Arc Descriptors (MADs) for both spatially and functionally local meshes in the same way. When we concatenate the BOLD responses of all stimuli, we obtain a vector $X_i(t) = [\mathbf{x}_{w,i}]_{\forall w}$ of BOLD responses measured from a node i for all windows of training samples. Recall that, we employ $X_i(t)$ to select functional neighbors using similarity measures.

We represent a mesh formed among the BOLD response of a seed node i of the mesh, and the BOLD responses of its p -nearest neighbors for a window w as:

$$\hat{\mathbf{x}}_{w,i} = \sum_{j \in \eta_p[i]} a_{w,i,j} \mathbf{x}_{w,j}, \quad (4.4)$$

where $\hat{\mathbf{x}}_{w,i}$ denotes the estimated BOLD response vector of node i for window w . We estimate the arc weights of a mesh formed around a node i for a window w by minimizing the regularized linear model error as follows:

$$\min_{a_{w,i,j}} (||\hat{\mathbf{x}}_{w,i} - \mathbf{x}_{w,i}||^2 + \lambda ||a_{w,i,j}||^2), \quad (4.5)$$

where $\lambda \in \mathbb{R}$ denotes regularization parameter.

When we perform within-subject classification with event related datasets, number of measurements recorded for each stimulus presentation is very small. Therefore,

selecting functional neighbors separately for each sample by computing correlations with small windows would not yield meaningful results. In such cases, we select the functional neighbors for each node using the measurements recorded from the participant during the entire course of the experiment. Since the neighboring nodes are the same for all windows, we compute a MAD vector $\mathbf{a}_{w,i} = [a_{w,i,1}, a_{w,i,2}, \dots, a_{w,i,p}]$ using ridge regression as

$$\mathbf{a}_{w,i} = (Q_{w,i}^T Q_{w,i} + \lambda I)^{-1} Q_{w,i}^T \mathbf{x}_{w,i}, \quad (4.6)$$

where $Q_{w,i}$ is a $D_w \times p$ matrix consisting of BOLD responses obtained from p -nearest neighbors of a seed node i for window w such that

$$Q_{w,i} = [\mathbf{x}_{w,j}], \quad \forall j \in \eta_p[i]. \quad (4.7)$$

When the number of measurements per each stimulus or task are long enough to compute correlation, we select the functional neighbors separately for each task or stimulus. In such cases, for each $\mathbf{a}_{w,i}$, we allocate a MAD vector of size $1 \times N$, where N is the number of nodes, and fill the corresponding entries with the mesh arc descriptors $a_{w,i,j}$. In other words, $\mathbf{a}_{w,i}(j) = a_{w,i,j}$ if $j \in \eta_p[i]$ and $\mathbf{a}_{w,i}(j) = 0$ if $j \notin \eta_p[i]$.

4.1.5 Classification of Cognitive States using Mesh Networks

Definition 5: Mesh Network: The mesh defined in (4.4) is solved for each node i with its neighbors, separately. In other words, we obtain an independent local mesh around each node, i . After we estimate all of the mesh arc descriptors, $\mathcal{E}_w = \{a_{w,i,j}\}_{i,j=1}^N$, we represent the connectivity graph, $G_w = \{V, \mathcal{E}_w\}$, as an ensemble of all local meshes. Thus, we call $G_w = \{V, \mathcal{E}_w\}$ a mesh network. Note that, mesh network is directed, since the relationship between two nodes i and j differs based on the seed node. In other words, $a_{w,i,j} \neq a_{w,j,i}$ since i is the seed node used in the former representation, and j is the seed node used in the latter one.

The representation power of the proposed mesh network is analyzed in the fMRI recordings during a number of brain decoding tasks. The major question is then how well a cognitive state can be represented by MADs. The answer to this question can

be partly observed from the performance of the classifier trained with MADs of mesh network.

Recall that for each cognitive stimulus or task, we record a sequence of brain volumes obtained for a time window. We estimate MADs for each window w , where w covers the BOLD responses obtained for each stimulus for visual recognition and emotional memory retrieval experiments. For HCP dataset, window w covers the number of scans of a cognitive task, where number of scans per session can be found in Table 3.1. Each stimulus or task is associated with a label c . Finally, each sample in the dataset is represented by a vector with the entries of the estimated MADs in the input space of a classifier.

Formally speaking, for each window w with label c , the arc descriptors $\mathbf{a}_{w,i}$ are used to construct a feature vector $F_w = [\mathbf{a}_{w,1}, \mathbf{a}_{w,2}, \dots, \mathbf{a}_{w,N}]$ by concatenating all MAD vectors, where N denotes the total number of nodes in a brain volume. We employ F_w features to train and test classifiers.

4.2 Generation of Brain Connectivity Dictionary and Encoding of Mesh Arc Descriptors

Encoding methods aim to summarize a number of local feature descriptors in vectorial statistics. They are based on generating a dictionary containing codewords, each of which represent the frequently observed descriptors in the dataset. These methods allow us to obtain a representation in terms of the distances of the local descriptors to the codewords or the number of occurrences of local descriptors assigned to these codewords. Following the success of summarization of local information using encoding methods in various domains, we propose to encode our local mesh arc descriptors using a number of encoding methods namely, Fisher Vectors (FV), Vector of Locally Aggregated Descriptors (VLAD) and Bag-of-Words (BoW).

In this section, we introduce our framework to encode MADs [41]. After we estimate MADs to represent the local connectivity patterns of cognitive tasks, we change the MAD space using Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Then, we fit GMMs to the set of MADs and cluster MADs

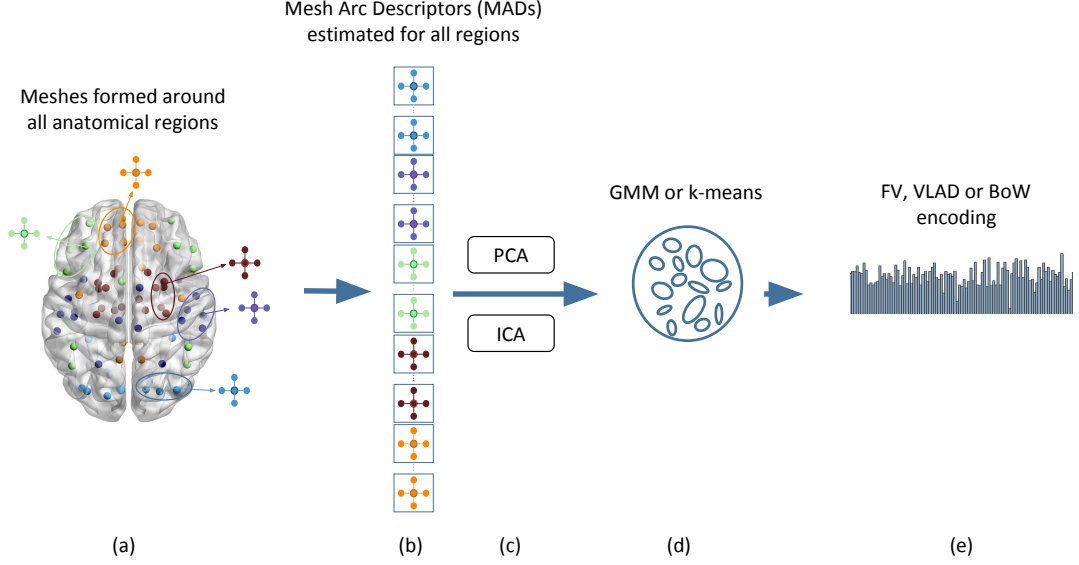


Figure 4.5: An overview of our proposed framework which is employed to encode MADs. (a) For each cognitive task, we first form local meshes around all anatomical regions using their functionally nearest neighbors (In the figure, we represent only five representative meshes). (b) Next, we estimate the mesh arc descriptors (MADs) for all tasks of all subjects. (c) We apply PCA or ICA on mesh arc descriptors obtained from training dataset to decorrelate them. (d) Then, we employ GMMs and k-means clustering methods on the descriptors using training dataset and find codewords of our dictionary. (e) Finally, we obtain FV, VLAD and BoW encoding for each task of training and test datasets.

using k-means clustering method. Centers of Gaussians in the mixture model or cluster centroids correspond to the codewords of our *brain connectivity dictionary*. Finally, we encode MADs using Fisher Vectors (FV) by employing their distance to the centers of GMMs, and using Vector of Locally Aggregated Descriptors (VLAD) and Bag of Words (BoW) by employing their distance to cluster centroids. The encoded MADs are used to train an SVM classifier for cognitive task classification. Figure 4.5 represents the steps of our encoding framework.

For simplicity, while encoding MADs, we remove window w subscript of $\mathbf{a}_{w,i}$ in (4.6) and represent MAD vector as $\mathbf{a}_i = [a_{i,1}, a_{i,2}, \dots, a_{i,N}]$ of size $1 \times N$, for each cognitive task.

4.2.1 Fisher Vector (FV) Encoding

Given a set of vectors, FV method encodes deviation of distribution of the vectors from a dictionary, which is typically described by a diagonal Gaussian mixture model (GMM) [26]. In the proposed framework, our vectors are MADs obtained from fMRI data. We obtain a *brain connectivity dictionary* of MADs by modeling the features using a GMM. In order to satisfy the assumption of diagonal covariance matrix of GMM and obtain linearly uncorrelated features, we apply PCA to MADs. For simplicity, we use the same notation for MADs at the output of PCA, such that $\mathbf{a}_i$ represents the projection of a MADs in a D -dimensional space.

Formally speaking, let $A = \{\mathbf{a}_i \in \mathbb{R}^D\}_{i=1}^N$ denote a set of D -dimensional MADs which are obtained from a single task of a single subject, and are sampled from the set of all MADs, $\mathbb{A}$. Also, let u_k be the k^{th} component of GMM, which models the generative process for elements belonging to $\mathbb{A}$ ($\forall a \in \mathbb{A}$). We denote the parameter set of all u_k 's by $\beta = \{\omega_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$, where ω_k , $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mixture weight, mean vector and covariance matrix of the k^{th} Gaussian, respectively. We compute the mixture u_β by $u_\beta(a) = \sum_{k=1}^K \omega_k u_k(a)$, where $u_k(a)$ denotes the k^{th} Gaussian in the mixture, and represented by

$$u_k(a) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(a - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (a - \boldsymbol{\mu}_k)\right). \quad (4.8)$$

We estimate the parameters of GMM on a training set of MADs using Expectation Maximization (EM) algorithm [50]. We assume that diagonal covariance matrices are identified by $\boldsymbol{\sigma}_k^2$ which denotes a vector of the diagonal entries of $\boldsymbol{\Sigma}_k^2$. Then, we compute the derivatives of log-likelihood of GMM with respect to the parameters using $\mathbf{G}_\beta^A = N^{-1} \sum_{i=1}^N \nabla_\beta \log u_\beta(\mathbf{a}_i)$.

The derivatives are only taken with respect to Gaussian means and variances, since the gradients computed with respect to the weight parameters, ω_i , provide *less information* [116]. Therefore, we obtain a representation that captures the average first and second order differences between MADs and each of the GMM centers. D -dimensional gradients obtained in [115] are as follows:

$$\mathbf{g}_{\boldsymbol{\mu}_k}^A = \frac{1}{N\sqrt{\omega_k}} \sum_{i=1}^N \gamma_i(k) \left(\frac{\mathbf{a}_i - \boldsymbol{\mu}_k}{\boldsymbol{\sigma}_k} \right) \quad (4.9)$$

and

$$\mathcal{G}_{\sigma_k}^A = \frac{1}{N\sqrt{2\omega_k}} \sum_{i=1}^N \gamma_i(k) \left[\frac{(\mathbf{a}_i - \boldsymbol{\mu}_k)^2}{\sigma_k^2} - 1 \right], \quad (4.10)$$

where $\gamma_i(k)$ represents the soft assignment of the MAD vector $\mathbf{a}_i$ to the k^{th} Gaussian, and is defined by

$$\gamma_i(k) = \frac{\omega_k u_k(\mathbf{a}_i)}{\sum_{\ell=1}^K \omega_\ell u_\ell(\mathbf{a}_i)}. \quad (4.11)$$

A Fisher vector $\mathbf{G}_\beta^A$ is obtained by concatenating the gradients under a vector of size $2KD$ $\mathbf{G}_\beta^A = [\mathcal{G}_{\boldsymbol{\mu}_1}^A, \mathcal{G}_{\sigma_1}^A, \mathcal{G}_{\boldsymbol{\mu}_2}^A, \mathcal{G}_{\sigma_2}^A, \dots, \mathcal{G}_{\boldsymbol{\mu}_K}^A, \mathcal{G}_{\sigma_K}^A]$. We obtain Fisher Vectors for both training and test data. In order to obtain better accuracy, we further apply l_2 normalization and square-root transformation on FVs as suggested in [126].

4.2.2 Vector of Locally Aggregated Descriptors (VLAD)

VLAD encodes a set of descriptors into a dictionary, which is computed by k-means clustering method. In the proposed framework, we first perform k-means clustering of MADs on the training data. Hence, the cluster centroids $\{\boldsymbol{\mu}_k\}_{k=1}^K$ correspond to our codewords of *brain connectivity dictionary*. Then, we associate each MAD $\mathbf{a}_i$ to its nearest codeword $NN(\mathbf{a}_i)$ in the dictionary, such that $NN(\mathbf{a}_i) = \underset{\boldsymbol{\mu}_k}{\operatorname{argmin}} \|\mathbf{a}_i - \boldsymbol{\mu}_k\|_2$, where $\|\cdot\|_2$ is the Euclidean norm.

Recall from the previous subsection that, $A = \{\mathbf{a}_i\}_{i=1}^N$ represents the set of D -dimensional MADs belonging to a single task of a single subject, each of which is sampled from $\mathbb{A}$ that contains all MADs belonging to all tasks of all subjects. For each codeword $\boldsymbol{\mu}_k$, we compute the sum of the differences, $(\mathbf{a}_i - \boldsymbol{\mu}_k)$, of the descriptors $\mathbf{a}_i$, which are assigned to the k^{th} cluster by

$$\mathbf{v}_k^A = \sum_{\mathbf{a}_i: NN(\mathbf{a}_i)=\boldsymbol{\mu}_k} (\mathbf{a}_i - \boldsymbol{\mu}_k). \quad (4.12)$$

We concatenate the D -dimensional vectors $\mathbf{v}_k^A$ for all clusters, and obtain a KD dimensional VLAD encoding $\mathcal{V}^A = [\mathbf{v}_1^A, \mathbf{v}_2^A, \dots, \mathbf{v}_K^A]$. Note that, we obtain VLAD encoding for each A , computed using training and test data.

4.2.3 Bag-of-Words (BoW)

We also encode MADs using the BoW approach, where the words correspond to a set of selected MADs. We first cluster MADs obtained from training data using a k-means clustering method. Then, for each A , we compute the number of MADs, b_k^A , belonging to the k^{th} cluster. Finally, we represent each sample by a K -dimensional vector such that $\mathbf{B}^A = [b_1^A, b_2^A, \dots, b_K^A]$.

4.3 Chapter Summary

In this chapter, we first formally defined node time series, spatial and functional neighborhood of a node and local meshes. In this study, nodes can be either voxels or anatomical regions. Spatial neighbors of a node are selected based on Euclidean distance while functional neighbors of a node are selected based on correlation. After we identified spatial and functional neighbors of a seed node, we formed spatially and functionally local meshes around it. We represented the seed node as a weighted combination of its neighboring nodes and estimate the arc weights of the mesh, called Mesh Arc Descriptors, using ridge regression. Finally, we concatenated MADs obtained for all nodes to represent a cognitive state and use these representations to train and test classifiers.

After obtaining MADs, we further encoded them using Fisher Vectors, Vector of Locally Aggregated Descriptors and Bag-of-Words methods. First, we decorrelated columns of MADs using PCA or ICA. After that, we fitted GMM to MADs and cluster MADs using k-means clustering method. We encoded MADs using FV by computing their distances to centers of GMMs. Moreover, we encoded MADs using VLAD and BoW by computing their distances to the cluster centroids.

CHAPTER 5

HIERARCHICAL MULTI-RESOLUTION MESH NETWORKS (HMMNS)

The many (hoi polloi), of whom none is individually an excellent (spoudaios) man, nevertheless can, when joined together, be better than those [the excellent few].

Aristotle, Politics

The major goal of the suggested HMMNs model is to represent a cognitive state of human brain by a set of coarse-to-fine brain networks, each of which is expected to signify complementing information about the underlying brain activities [103]. This model enables us to study the activities and interactions of brain regions in multiple resolutions. HMMNs also enable us to boost the brain decoding performance due to the fusion of diverse connectivity structures represented in mesh networks.

In the following sections, first we explain the multi-resolution representation of fMRI data. Next, we represent each cognitive task by a set of multi-resolution brain networks. Then, we analyze the network topology of multi-resolution brain networks using connectivity measures including node degree, node strength, node betweenness centrality and global efficiency. After that, we explain a hierarchical learning architecture, called fuzzy stacked generalization (FSG), which fuses the brain networks, obtained in different resolution for brain decoding. Finally, we form a 2-level architecture, in which we employ FSG to fuse information obtained from multi-resolution and multi-task networks at each level for gender classification.

5.1 Multi-resolution Representation of fMRI Signals

The first step for construction of the proposed HMMNs is to decompose the fMRI recordings into a set of signals in different time resolutions. The multi-resolution representation of fMRI data enables us to estimate and analyze the time scale(s) in which the anatomical regions process a particular information [120, 141, 72, 71]. Discrete Wavelet Transform (DWT) is preferred for multi-resolution analysis of fMRI data due to its decorrelating property [43] and adaptivity to non-stationary features of fMRI signals [21].

In this chapter, node defined in Section 4.1 correspond to anatomical region. The average signal $X_i(t)$ obtained for each anatomical region in (4.1), represents the information processed in node i . As the next step, we decompose all the representative signals, $X_i(t)$, for $i = 1, 2, \dots, N$, into a set of signals using subband coding along the temporal dimension to analyze the regional brain activities at different resolutions. In multi-resolution analysis step of HMMNs, shown in Figure 5.1, we define the mother wavelet by,

$$\Psi_{l,q}(t) = \frac{1}{\sqrt{2^l}} \Psi \left(\frac{t - 2^l q}{2^l} \right), \forall t = 1, 2, \dots, T, \quad (5.1)$$

where $\int \psi(t)dt = 0$, $\forall l = 1, 2, \dots, L$ and $\forall q = 1, 2, \dots, Q$. Similarly, we define the father wavelet that satisfies $\int \phi(t)dt = 1$ by

$$\Phi_{l,k}(t) = \frac{1}{\sqrt{2^l}} \Phi \left(\frac{t - 2^l q}{2^l} \right), \forall t, l, q. \quad (5.2)$$

There are a total of L decomposition levels, where $l = 1, 2, \dots, L$. Also, q represents the location to which the wavelet is translated in time, where $q = 1, 2, \dots, Q$. The value of Q changes for each l such that $Q = T/2^l$, and T denotes the duration of the signal, recorded during a cognitive task. Note that each high or low frequency sub-band is spanned by translated versions of a single mother wavelet or father wavelet, respectively.

We apply discrete wavelet transform (DWT) to the average fMRI signal $X_i(t)$ obtained from node i to decompose the signal into two sets of orthogonal components, namely, a set of approximation coefficients $\mathcal{A} = \{\alpha_{r,l,q}\}$, and a set of detail coefficients $\mathcal{D} = \{\delta_{r,l,q}\}$. While the detail coefficients represent the information about

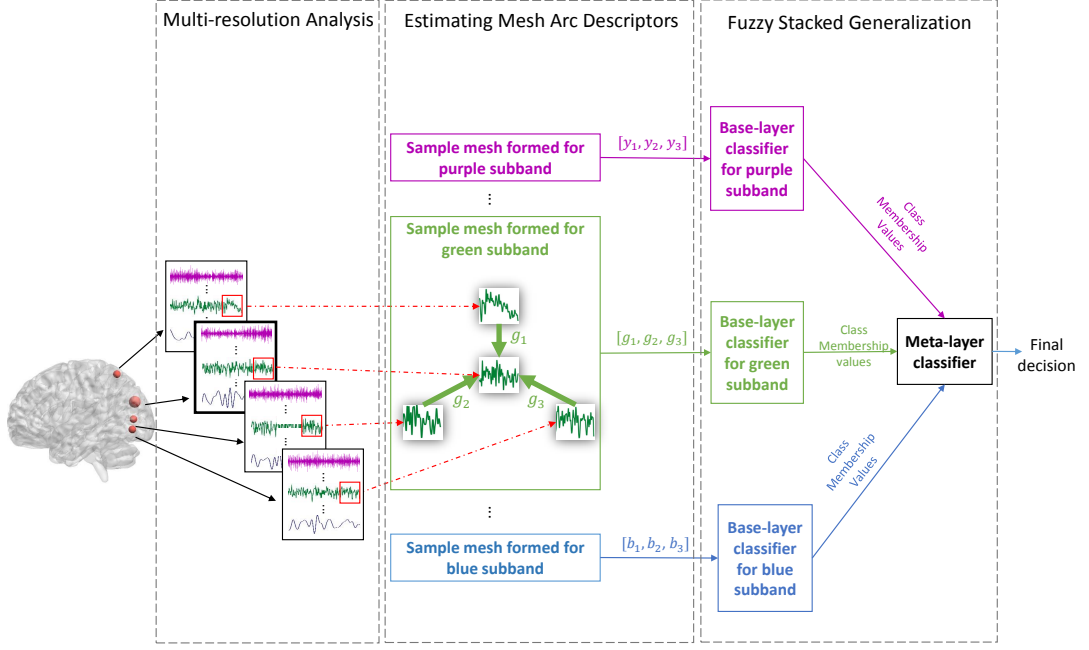


Figure 5.1: Overview of the proposed hierarchical framework (HMMNs). A random seed region and its three functionally nearest neighbors (with smaller sizes) are denoted on the template brain. In the multi-resolution analysis step, the average fMRI signal of each region is decomposed into a set of coarse-to-fine signals. Then, meshes are formed around each seed region for all subbands, and mesh arc descriptors are estimated. Finally, the estimated arc descriptors of the meshes are ensembled under a network structure and fed as an input to the hierarchical ensemble learning architecture, called, FSG, where the final decision is computed by a meta-layer classifier.

variation in the data at a given scale, the approximation coefficients contain the residual of the signal after the information in the provided scale and all finer scales are removed [21]. Then, we can recover the original signal $X_i(t)$ obtained from region r by

$$X_i(t) = \sum_q \alpha_{r,l,q} \Phi_{l,q}(t) + \sum_{l' \leq l} \sum_q \delta_{r,l',q} \Psi_{l',q}(t). \quad (5.3)$$

As it can be seen from (5.3), we can recover the original signal by first multiplying the father wavelet with the approximation coefficient at a particular decomposition level l , and then adding them to the multiplication of mother wavelet and detail coefficients of all levels greater than or equal to this particular level, $\forall l' \leq l$.

We can also reconstruct only the approximation or detail parts of the signal at a decomposition level. For example, if we decompose the signal in the first level, then we can reconstruct the approximation, A_1 , and detail, D_1 , parts using inverse discrete wavelet transform. If we continue decomposing the approximation part A_1 one more level, then, we obtain the second level approximation A_2 and detail D_2 parts. Therefore, for a given signal, we can obtain various signal components, each of which contains the information about components of the signal in different time resolutions.

Formally speaking, let $s \in \mathcal{S} = \{0, 1, \dots, 2L\}$ be the index of the entries of the ordered set of subbands $\{A_0, A_1, \dots, A_L, D_1, \dots, D_L\}$. Then, we can reconstruct a particular approximation or detail part of a signal using approximation and detail coefficients, respectively, as follows:

$$x_{s,i}(t) = \begin{cases} X_i(t), & \text{if } s = 0 \\ \sum_q \alpha_{i,s,q} \Phi_{s,q}(t), & \text{if } 1 \leq s \leq L \\ \sum_q \delta_{i,s-L+1,q} \Psi_{s-L+1,q}(t), & \text{if } s > L \end{cases} \quad (5.4)$$

Notice from (5.4) that $s = 0$ already corresponds to the original signal. Subbands that correspond to $s \in [1, L]$ are approximation parts of the signal, and we reconstruct them using only the approximation coefficients. Finally, subbands that correspond to $s \in [L + 1, 2L]$ are detail parts of signals, which are reconstructed using detail coefficients.

In the fMRI data acquisition experiments, we decompose the signals obtained during each task separately for each participant. Since the minimum number of measurements is recorded during Emotion experiment and equals to 176, decomposition of signals into more than 6 levels is not possible. Therefore, we use $L = 6$ in this study. For each level, we decompose a representative signal into $l + 1$ dyadic frequency subbands, and reconstruct $l + 1$ signals that contain a single approximation part A_l and detail parts of that level and finer levels $\{D_1, \dots, D_l\}$. We use cubic Battle-Lemarie wavelets shown in Figure 5.2, which are frequently used in wavelet analysis of fMRI signals [119, 120].

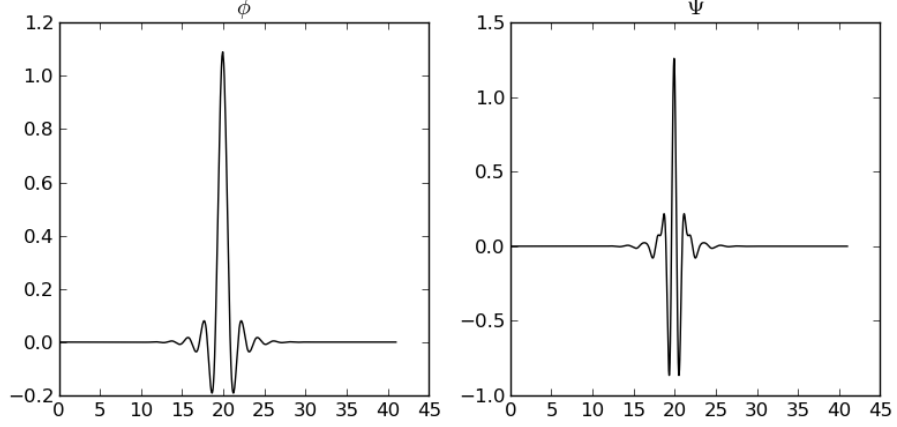


Figure 5.2: Cubic Battle-Lemarie filters (Father wavelet (Φ) on the left and mother wavelet (Ψ) on the right)

5.2 Estimation of Multi-resolution Mesh Arc Descriptors

In this section, since anatomical regions are not in a regular grid, we do not form meshes in spatial neighborhood. Instead, we form meshes only in a functional neighborhood of each anatomic region. Each region is connected to its p -nearest neighboring nodes to form a star mesh in its functional neighborhood as shown in Figure 5.3. The neighboring regions of each anatomic region are the ones having p -largest Pearson correlation with that region. We represent the BOLD response of a seed region $\mathbf{x}_{s,w,i}$ of the mesh, as weighted linear combination of the BOLD responses of its p -nearest neighbors for the w^{th} window at the s^{th} subband as:

$$\hat{\mathbf{x}}_{s,w,i} = \sum_{j \in \eta_p[i]} a_{s,w,i,j} \mathbf{x}_{s,w,j}, \quad (5.5)$$

where $\hat{\mathbf{x}}_{s,w,i}$ is the estimated response for node i , for the s^{th} subband and for the w^{th} window. For each mesh formed around a node i , we estimate the arc descriptors minimizing the following cost function for each local mesh formed around node i with respect to $\{a_{s,w,i,j}\} \forall s, w, i, j$;

$$\min_{a_{s,w,i,j}} \left(\|\hat{\mathbf{x}}_{s,w,i} - \mathbf{x}_{s,w,i}\|^2 + \lambda \sum_j \|a_{s,w,i,j}\|^2 \right), \quad (5.6)$$

where λ is the regularization parameter. In (5.5), $\mathbf{x}_{s,w,i}$ is a vector that represents the time series of node i for the s^{th} subband and for the w^{th} window.

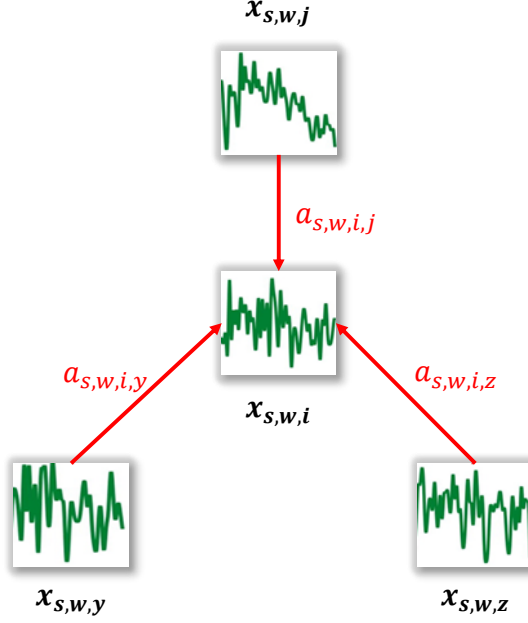


Figure 5.3: A sample mesh formed around an anatomical region i with its three neighboring regions j , y and z for subband s and window w . We represent $x_{s,w,i}$ as a weighted linear combination of $x_{s,w,j}$, $x_{s,w,y}$ and $x_{s,w,z}$. We estimate mesh arc descriptors $a_{s,w,i,j}$, $a_{s,w,i,y}$ and $a_{s,w,i,z}$ using ridge regression and represent the corresponding mesh in terms of these descriptors. We form meshes around all anatomical regions and the set of meshes form a mesh network for subband s and window w .

We form meshes around each node i with its neighbors separately. Following the estimation of all of the mesh arc descriptors, $\mathcal{E}_{s,w} = \{a_{s,w,i,j}\}_{i,j=1}^N$, we represent the mesh network, formed for each window w and each subband s ($G_{s,w} = \{V, \mathcal{E}_{s,w} : \forall s \in \mathcal{S}\}$), as an ensemble of all local meshes. Note that, mesh network is directed, since the relationship between two nodes i and j differ based on the seed node. In other words, $a_{s,w,i,j} \neq a_{s,w,j,i}$ since i is the seed node used in the former representation, and j is the seed node used in the latter one.

Mesh networks are estimated for both original fMRI signal, and its approximation and detail parts of different resolutions. In other words, we form $2L + 1$ distinct mesh networks for the frequency subbands $\{A_0, A_1, A_2, \dots, A_L, D_1, D_2, \dots, D_L\}$. Recall that, in our dataset, we use $L = 6$, and $l = 0$ corresponds to the raw fMRI signal.

5.3 Network Topology and Connectivity Analysis of HMMNs

In this sub-section, we analyze the network properties of the proposed HMMNs, in different resolutions. Recall that the mesh networks defined by an ensemble of local meshes are weighted and directed. We compute graph connectivity measures, namely, node degree, node strength and node betweenness centrality and global efficiency using Brain Connectivity Toolbox [123]. Note that, for simplicity, we do not include subband s and window w subscripts in the definitions given in this section.

5.3.1 Connectivity Analysis by Degree of Nodes of Mesh Network

Degree of a node of a mesh network represents the number of links connected to a node. While in-degree is the number of inward links, out-degree is the number of outward links. In this study, we select the number of nearest neighbors as $p = 40$, i.e. in-degree equals to 40 for each node. Yet, we wish to explore the nodes (anatomic regions) in a mesh network, which are connected to many nodes through the outgoing mesh arc descriptors. In order to analyze this property, we compute the out-degree distribution of nodes.

In order to formally define degree of a node, first we define the function $adj(i, j)$ as follows:

$$adj(i, j) = \begin{cases} 0, & \text{if } j \notin \eta_p[i] \\ 1, & \text{otherwise} \end{cases} \quad (5.7)$$

After that, we define in-degrees (deg_{in}) and out-degrees (deg_{out}) of node i in mesh network G by

$$deg_{in}(i) = \sum_{\forall j} adj(i, j) \quad (5.8)$$

and,

$$deg_{out}(i) = \sum_{\forall j} adj(j, i). \quad (5.9)$$

Note that, while computing in-degrees of node i in mesh network G , we sum the number of edges connected to node i where node i is the seed node of the mesh. On the other hand, while computing out-degrees of node i in mesh network G , we

sum the number of edges connected to node i where node i is a neighboring node of meshes.

5.3.2 Connectivity Analysis by Strength of Nodes of Mesh Network

Strength of a node in mesh network represents the sum of the arc descriptors connected to a node where node out-strength is the sum of arc descriptors of outward links. In this subsection, we compute the node out-strengths for each node and for each subband. Then, we overlay them on a brain template.

We formally define out-strength (str_{out}) of a node i in mesh network G as follows:

$$str_{out}(i, j) = \sum_{\forall j} a_{j,i}, \quad (5.10)$$

where, $a_{j,i}$ represents the arc descriptor between seed node j and its neighboring node i . Recall that, if nodes i and j are not neighbors, then $a_{j,i} = 0$.

5.3.3 Connectivity Analysis by Betweenness Centrality of Nodes of Mesh Network

Betweenness centrality of a node of mesh network represents how central a node is by considering the fraction of all shortest paths containing that node. In other words, nodes with large betweenness centrality take part in large number of shortest paths. Nodes with large betweenness centrality are candidates to be hubs, and expected to be vital for the flow of information through network [141].

Formally speaking, let $SP_{yz} = (v_1, v_2, \dots, v_n)$ denote the shortest path defined between node y and node z such that v_i and v_{i+1} are adjacent. We can find $SP_{yz} = (v_1, v_2, \dots, v_n)$, where $v_1 = y$ and $v_n = z$ by minimizing $\sum_{i=1}^{n-1} a_{v_i, v_{i+1}}$. Recall that, $a_{v_i, v_{i+1}}$ represents MAD formed between node v_i and node v_{i+1} .

We can formally define betweenness centrality of a node i as follows:

$$bc(i) = \sum_{y \neq i \neq z} \frac{N_{yz}^{sh}(i)}{N_{yz}^{sh}}, \quad (5.11)$$

where N_{yz}^{sh} represents the number of shortest paths formed between nodes y and z (SP_{yz}) and $N_{yz}^{sh}(i)$ represents the number of shortest paths formed between nodes y and z that pass through node i .

5.3.4 Connectivity Analysis by Global Efficiency

Global efficiency measures how efficiently a network exchanges information [82]. It is a measure that quantifies small-world property of a network, and it is calculated as the average inverse shortest path length in a network. We define the global efficiency (*globe f*) of our mesh network as follows:

$$globe f = \frac{1}{N} \sum_{\forall i \in \mathcal{N}} \frac{\sum_{j \in \mathcal{N}, j \neq i} d_{ij}^{sh}}{n - 1}, \quad (5.12)$$

where d_{ij}^{sh} represents the length of the shortest path between nodes i and j (SP_{ij}).

If nodes are disconnected, then the shortest path between them has an infinite length, leading to zero efficiency [123]. In this study, we analyze how global efficiency differs for each task and for each subband.

5.4 Multi-resolution Mesh Networks for Cognitive State Classification

In this section, first, we obtain graph embedding of multi-resolution mesh networks by concatenating the entries of adjacency matrix of MADs under a vector. Then, we employ the embeddings obtained for a number of subbands to train and test different classifiers. We fuse the decisions of classifiers, trained with MAD vectors of multi-resolution mesh networks for cognitive state classification. We also analyze the significance of meta-layer features (class membership values obtained at the output of base-layer classifiers) and base-layer features (MADs of multi-resolution mesh networks) obtained from two layers of FSG architecture. Finally, we explore whether the classifiers trained with multi-resolution mesh networks are diverse by applying a number of classifier diversity measures.

5.4.1 Graph Embedding of Multi-resolution Mesh Networks

To this end, we form a mesh network for each task experiment, as ensemble of local meshes and estimate mesh arc descriptors at each subband. Then, we concatenate the estimated MADs under a structured feature matrix to form the adjacency matrix of the mesh network at each subband.

We cannot perform multi-resolution analysis on visual object recognition and emotional memory retrieval datasets since their experiments are event-related and only a few measurements are collected during each stimulus presentation. In that case, window length is very short and we cannot decompose the fMRI signals into multiple subbands. Instead, we construct multi-resolution mesh networks for HCP dataset. Therefore, each window covers a task experiment, whose number of measurements are given in Table 3.1.

For each window, we form a set of multi-resolution mesh networks whose arc descriptors remain the same during the experiment. Recall that, arc descriptors represent the relationship between seed region and its functionally nearest neighbors. Given a window w corresponding to a cognitive task, if two nodes i and j of the mesh network G are not neighbors, then $a_{s,w,i,j} = 0, \forall s$, for that window.

We form an adjacency matrix of size $N \times N$ for mesh network G using the estimated arc descriptors $(\forall_{i,j} a_{s,w,i,j})$, where N denotes the total number of nodes in mesh network G . By concatenating the arc descriptors under a vector of size $1 \times N^2$, we embed the mesh network for a window w at subband s , by the feature vector $F_{s,w} = [a_{s,w,i,j}]_{i,j=1}^N$.

Note that, we compute a graph embedding vector $F_{s,w}$ for each task experiment recorded in time window w obtained from a subject in the dataset. When we concatenate all feature vectors, we obtain a feature matrix $\mathcal{F}_s = [F_{s,w}]_{\forall w}$, for the s th subband. We assign a class label $c \in [1, 7]$ to each mesh network G extracted from the window w covering the task experiment. Therefore, each mesh network is associated with a class label.

5.4.2 Fusion of Multi-resolution Brain Networks under a Hierarchical Ensemble Learning Architecture

One of the major problems of representing the cognitive tasks by multi-resolution mesh networks is the increased complexity of the analysis and decoding of the underlying task. Furthermore, the estimated arc descriptors in different resolutions should complement each other in some way, so that fusion process boosts the performance of the decoding task. In this study, we propose to use a fast and efficient ensemble learning method for fusion of the mesh networks obtained in multiple resolution by adapting a state-of-the-art ensemble learning method, called fuzzy stacked generalization (FSG), which was shown to perform better than various classifiers for MVPA [109]. We propose a two-phase approach to employ the basic structure of FSG in our proposed framework as follows (see Figure 5.1):

First, we train an individual classifier, such as logistic regression using the mesh arc descriptors obtained for each subband s at the base-layer of an FSG. The dataset is partitioned into κ -folds, where $\kappa - 1$ of them is used for training and the κ^{th} fold is used to test the classifiers. Then, we validate classification (decision) hypothesis of each classifier using the samples belonging to the κ^{th} chunk of the data. Each classifier outputs an estimate of the posterior probabilities of a mesh network belonging to a particular cognitive task. We put the posterior probabilities in a membership vector of size $1 \times C$ for all tasks, where C is the number of classes. Thereby, each base layer classifier maps N^2 dimensional MAD feature vectors obtained from each subband to a C dimensional decision space. Recall from Table 3.1 that, the number of cognitive tasks in HCP dataset is 7 ($C = 7$). This mapping trains the base layer classifiers to become experts for capturing diverse connectivity properties of fMRI data signified in different subbands. Note that, we perform κ -fold cross validation on the training dataset to obtain class membership vectors for all chunks of training dataset.

Next, we compute class membership vectors of test samples using decisions of the trained classifiers for test data. At the meta-layer, we obtain a membership vector of size $1 \times CE$ for each sample, by concatenating membership vectors obtained from all base-layer classifiers, where E is the number of classifiers (i.e., number of the subbands used to train base-layer classifiers).

Then, we train and test meta-layer classifier using the concatenated membership vectors of the training and test data, respectively. In the experiments, we provide classification results using Support Vector Machines (SVM) classifiers with linear kernels at the meta-layer. E is varied by selecting various combinations of subbands to see the effect of approximation and detail parts on the classification performance.

Note that, fusion of the collaboratively trained expert classifiers are shown to provide higher performance than the majority voting methods in [109]. In the experimental analyses (Section 6.3), we will examine this observation using the proposed framework, and compare the results with state-of-the-art majority voting methods for fMRI data analysis.

5.4.3 Significance of Meta-layer Features Representing Class Membership Values Obtained from Classifiers Trained with Multi-resolution Mesh Networks

Which meta-layer features can discriminate specific cognitive tasks from others? In order to answer this question, we investigate the discriminative properties of meta-layer features (class membership values) in different subbands.

Recall that, each base-layer classifier of FSG assigns a class membership vector to a mesh network extracted from a different subband of fMRI signal, recorded under a cognitive stimulus. Therefore, our meta-layer features are class membership values for each mesh network, representing a cognitive task at a specific resolution. We estimate the significance of class membership values using absolute value two-sample t-test with pooled variance estimate [137]. First, we concatenate the class membership vectors obtained at the output of base-layer classifiers under a feature matrix. Then, we divide the feature matrix into two groups, where the first group contains the tasks belonging to a specific task category c and the second group contains the remaining tasks. Then, we compute t-test and obtain z value for each class membership entry as explained in Algorithm 1. Notice that, class membership with larger value of z is the one that rejects the null hypothesis that two groups come from the same distribution with more confidence.

Algorithm 1 Finding significance of class membership values using absolute value two-sample t-test with pooled variance estimate

Require: Feature matrix of class membership values $\mathcal{M}$ for all tasks ($\forall q$)

Require: Class label vector $\mathcal{C}$ of all tasks ($\forall q$)

Ensure: Cell array $\mathcal{T}$ containing z_c values, for all classes ($\forall c$) such that $\mathcal{T}\{c\} = z_c$

```

1: for  $c = 1 \rightarrow 7$  do
2:    $groupinds \leftarrow (\mathcal{C} == c)$ 
3:    $nongroupinds \leftarrow (\mathcal{C} \sim c)$ 
4:    $n_1 \leftarrow$  number of tasks with class label  $c$ 
5:    $n_0 \leftarrow$  number of tasks with class label other than  $c$ 
6:    $m_1 \leftarrow \mathcal{M}_{groupinds}$ 
7:    $m_0 \leftarrow \mathcal{M}_{nongroupinds}$ 
8:    $z_c \leftarrow \frac{\mu(m_1) - \mu(m_0)}{\sqrt{\frac{\sigma(m_1)^2}{n_1} + \frac{\sigma(m_0)^2}{n_0}}}$  where  $\mu(\cdot)$  denotes mean  $\sigma(\cdot)$  denotes standard deviation.
9:    $\mathcal{T}\{c\} \leftarrow z_c$ 
10: end for

```

5.4.4 Significance of Base-layer Features Representing Connectivity Among Anatomical Regions

The proposed multi-resolution mesh network models the multiple connectivity among the anatomic regions at each subband. The mesh arc descriptors estimated during a cognitive task indicate the degree of connectivity among the anatomic regions. In order to observe the strongly connected anatomic regions for a particular cognitive task, we measure the significance of connectivity with respect to tasks. For this purpose, we identify the mesh arc descriptors that discriminate a specific task from others and analyze the relationship among the anatomical regions connected by these mesh arcs. We measure the significance of mesh arc descriptors using absolute value two-sample t-test with pooled variance estimate. We divide the feature matrix formed by mesh arc descriptors into two groups, where the first group contains the tasks belonging to a specific task category c and the second group contains the remaining tasks. Then, we compute t-test and obtain a z -value for each mesh arc descriptor. We take the best 20 mesh arc descriptors with maximum z -values and we identify the ones that are

significant to discriminate only a specific class from others in this set. Note that, we identify these mesh arc descriptors separately for each task and each subband. We select the number as 20 to make sure that for all of the 7 tasks we obtain MADs, which are significant to discriminate only a specific class from others.

5.4.5 Diversity of Base-layer Classifiers

The major assumption of this study is that each subband encodes different information which completes each other. If this assumption is valid, we expect a boost in the brain decoding performances when all the subbands are fused in an ensemble learning architecture. The validity of this assumption is demonstrated in the next section. Then, the next question is that what is the degree of information diversity embedded at each subband? In order to analyze diversity, we measure the degree of collaboration and cooperation among the base-layer classifiers trained with the mesh networks of different subbands. We compute popular diversity measures, suggested in Diversity Toolbox [79]. We analyze the relationship between the state-of-the-art diversity measures and classification accuracy of the proposed HMMN framework.

Recall that, each base-layer classifier outputs a membership value for a mesh network to belong to one of the c task categories. While measuring the diversity of classifiers, we use the oracle outputs which assign correct/incorrect decision of classifiers. If the membership value is greater than 0.5, then we assume that the input mesh network is generated by this class and assign 1 to the output. Otherwise, the value of the output is 0. The collaboration among the base-layer classifiers are quantified by both pairwise and non-pairwise diversity measures. Pairwise measures include the double-fault measure (**DF**), the correlation coefficient (**Mean ρ**); and non-pairwise measures include coincident failure diversity (**CFD**), generalized diversity (**GD**) and the measure of difficulty (**θ**) [79], which will be defined next.

Let C_i and C_j denote a pair of classifiers. Then, let N^{00} , N^{01} , N^{10} and N^{11} denote the number of samples misclassified by both C_i and C_j ; misclassified by C_i and correctly classified by C_j ; correctly classified by C_i and misclassified by C_j and correctly classified by both C_i and C_j , respectively. We average the values of pairwise measures across all pairs of base-layer classifiers.

- **Double-fault measure (DF):** This measure arose from the intuition that two classifiers should perform differently to be diverse. Giacinto and Roli [51] claimed that the more different two classifiers are, the fewer the coincident errors between them. Double-fault measure is used to form a pairwise diversity matrix for a classifier pool and to select classifiers that are least related [51]. It is computed as the proportion of the samples misclassified by both classifiers over all samples as follows:

$$DF_{i,j} = \frac{N^{00}}{N^{00} + N^{01} + N^{10} + N^{11}}. \quad (5.13)$$

If classifiers, trained with multi-resolution mesh networks make less coincident errors, then their diversity increases. Therefore, we obtain more diversity with higher DF values.

- **Correlation coefficient ρ :** Correlation coefficient is a number that quantifies a type of correlation and dependence, meaning statistical relationships between two or more values in fundamental statistics. The correlation coefficient ρ is computed between a pair of classifiers C_i and C_j as follows:

$$\rho_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{(N^{11} + N^{01})(N^{00} + N^{01})(N^{11} + N^{10})(N^{00} + N^{10})}}. \quad (5.14)$$

If classifiers, trained with multi-resolution mesh networks assign same class labels to the same samples, then correlation between them increases, leading to a decrease in the diversity. Therefore, we obtain more diversity with lower ρ values.

- **Generalized Diversity (GD) [113]:** Let p_i denote the probability that i classifiers among all of the classifiers fail on a randomly chosen sample. Also let $p(i)$ denote the probability that i *randomly chosen* classifiers will fail on a randomly chosen sample. We compute $p(1)$ and $p(2)$ by,

$$p(1) = \sum_{i=1}^E \frac{i}{E} p_i \quad (5.15)$$

and

$$p(2) = \sum_{i=1}^E \frac{i(i-1)}{E(E-1)} p_i, \quad (5.16)$$

where E indicates the number of classifiers at the base layer which corresponds to total number of subbands. Then, the generalized diversity is computed as follows:

$$GD = 1 - \frac{p(2)}{p(1)}. \quad (5.17)$$

Assume that we randomly select two subbands and corresponding classifiers with the mesh networks formed at these two subbands. When failure of one of these classifiers is always accompanied by correct labeling by the other classifier, then maximum diversity occurs and the probability of failing of both classifiers is $p(2) = 0$. On the other hand, when failure of one of these classifiers is always accompanied by failure of the other, then minimum diversity occurs. In that case the probability of both classifiers failing is the same as the probability of one randomly picked classifier $p(1) = p(2)$.

Generalized diversity measures the collaboration among classifiers trained with multi-resolution mesh networks. The cognitive tasks represented in different subbands complete each other for higher values of GD, whereas lower values of GD is an indicator of similar behavior of classifiers.

- **Coincident Failure Diversity (CFD) [113]:** The minimum value of CFD=0 is achieved when all classifiers are always correct or when all classifiers are simultaneously correct or simultaneously wrong. The maximum value of CFD=1 is achieved when at most one classifier fails on any randomly chosen task. CFD is computed as follows:

$$CFD = \begin{cases} 0 & p_0 = 1 \\ \frac{1}{1 - p_0} \sum_{i=1}^E \frac{E - i}{E - 1} p_i & p_0 < 1 \end{cases} \quad (5.18)$$

Recall that in the above equation $p(i)$ denote the probability that i *randomly chosen* classifiers fail on a randomly chosen sample and E is the total number of classifiers (i.e., subbands). In our case, if at most one classifier trained with mesh network of a specific subband fails, we obtain the maximum CFD, which results in maximum diversity.

- **Measure of difficulty θ [57]:** Let X be a discrete random variable taking values in $\{0, \frac{1}{E}, \frac{2}{E}, \dots, 1\}$ and denoting the proportion of classifiers that correctly

classify a randomly chosen sample. Then, difficulty θ is defined as the variance of X . If the same samples are difficult for all classifiers, and the others are easy for all classifiers, variance of the distribution will be large, leading to large value of θ . Hence, small values of θ indicate large diversity.

In our case, if there exists some cognitive tasks, which cannot be identified by any of the classifiers trained for all subbands and there exists some cognitive tasks which are easily identified by all of these classifiers, the difficulty measure will be high, leading to a low diversity.

5.5 Multi-resolution Mesh Networks for Gender Classification

One of the most crucial questions in brain sciences is the similarities and distinctions between males and females while they are performing a series of cognitive tasks. This has been studied through experimental psychology for a while using some surveys and specifically designed experiments. In recent years, researchers investigate the fMRI data for understanding gender differences under different cognitive tasks. Yet, only a few studies approach the exploration of gender differences as a gender classification problem from machine learning perspective.

In this study, we propose a hierarchical framework that classifies individuals based on their gender using the brain networks formed at different resolutions, while performing a variety of cognitive tasks. This approach enables us to study the gender discrimination while they perform different tasks at different resolution.

We first decompose the fMRI signal into multiple subbands using DWT. Then, we represent the fMRI signal as a set of local meshes for multiple tasks and resolutions, separately. For each resolution and each task, we form Functionally Local Meshes (FLM) around anatomical regions and represent each region as a regularized linear combination of their functionally neighboring regions. We estimate the MADs, which denote the relationship among the seed node of the mesh and its neighbors, by solving ridge regression. Then, using MADs of all meshes formed around all nodes, we train a logistic regression classifier at each resolution for different tasks. Finally, we fuse the complementing information obtained from mesh networks formed for differ-

ent tasks and different resolutions under a 2-level fuzzy stacked generalization (FSG) architecture. In this architecture, one layer is responsible for the fusion of complementary information obtained from different cognitive tasks for gender classification, while the other layer fuses diverse information coming from distinct resolutions.

We test our approach on Human Connectome Project (HCP) task-fMRI dataset, in which subjects perform 7 cognitive tasks. The results of our gender classification experiments are provided in Section 6.4. Details of the proposed 2-level FSG architecture are explained next.

5.5.1 Fusion of Multi-resolution Multi-task Mesh Networks under Fuzzy Stacked Generalization (FSG)

Gender differences have been generally explored while participants are performing a specific cognitive task or they are at rest. Since different anatomical regions and their relationships discriminate gender under different cognitive tasks, we hypothesize that fusion of classifiers trained with multi-task mesh networks would better discriminate gender of participants. Additionally, each network in a set of coarse-to-fine mesh networks is expected to signify complementing information about the underlying brain activities [103]. Following the observation that FSG was shown to perform better than various classifiers for multi-voxel pattern analysis (MVPA) [109], in this study, we adapt fuzzy stacked generalization (FSG) for fusion of the mesh networks obtained for multiple tasks and multiple resolutions.

We have multiple mesh networks obtained for different tasks or different subbands. In order to ensemble them using FSG, first, we train an individual logistic regression classifier for each dataset in the ensemble at the base-layer of FSG. We partition each dataset into κ -folds, where $\kappa - 1$ of them is used for training and the κ^{th} fold is used to test the classifiers. Then, we validate classification (decision) hypothesis of each classifier using the samples belonging to the κ^{th} chunk of the data. Each classifier outputs the estimate of the posterior probabilities of mesh networks in the κ^{th} chunk belonging to particular genders. For each mesh network, we put the posterior probabilities of being female or male in a 2-dimensional membership vector. Thereby, each base layer classifier maps N^2 dimensional MAD feature vectors obtained from each

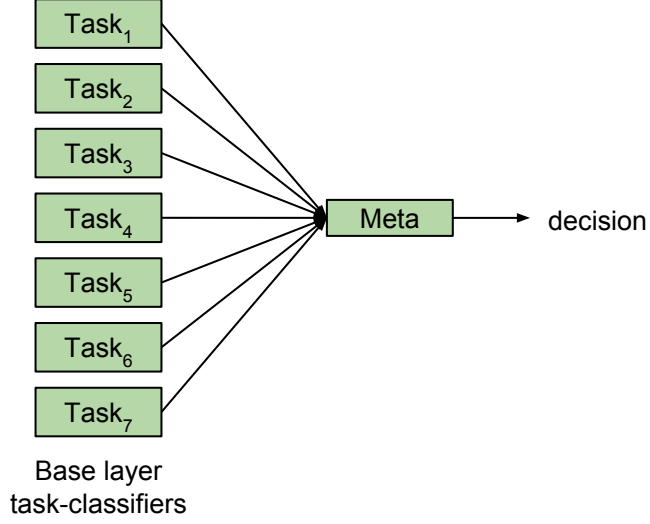


Figure 5.4: 1-level fusion of decisions of 7 base layer *task-classifiers* in a subband

mesh network to a 2-dimensional decision space. This mapping trains the base layer classifiers to become experts for capturing diverse connectivity properties of fMRI data signified in different subbands or different tasks. Note that, we perform $\kappa - 1$ -fold cross validation on the training dataset to obtain class membership vectors for all chunks of training dataset.

Next, we compute class membership vectors of test samples using decisions of the trained classifiers for test data. At the meta-layer, we obtain a membership vector of size $1 \times 2E$ for each mesh network, by concatenating membership vectors obtained from all base-layer classifiers, where E is the number of classifiers. Then, we train and test meta-layer classifier using the concatenated membership vectors of the training and test data, respectively. In the experiments, E is varied by selecting various combinations of subbands or tasks.

From now on, we represent a classifier trained with mesh networks formed for a task and a subband as *task-classifier* and *subband-classifier*, respectively. In this study, we adapt FSG to fuse the complementing information of different tasks and subbands in four different ways explained in the following subsections. In the following subsections, we will explain 1-level and 2-level architectures to fuse decisions of *task-classifiers* and *subband-classifiers*.

5.5.1.1 Fusion of Decisions of All Base-Layer Task-Classifiers

In this subsection, we aim to explore whether fusion of only *task-classifiers* is enough to discriminate gender. By comparing the results of the architecture given in this subsection with the proposed 2-level architectures in Section 5.5.1.3 and Section 5.5.1.4, we can explore if the multi-resolution analysis brings extra information for gender classification. For that purpose, in this subsection we use only original fMRI signals and do not decompose them into multiple subbands.

In this architecture, we only fuse the information coming from *task-classifiers* trained for a single subband as presented in Figure 5.4. Since subjects perform 7 distinct cognitive tasks, we train 7 classifiers with mesh networks formed at each cognitive task in the base layer. Then, gender posterior probabilities obtained from all base-layer classifiers are concatenated to form a 14-dimensional feature vector since we have two probability entries per gender for each of the 7 tasks. Using these representations, we train and test a single meta-layer classifier that assigns a gender label to individuals.

The performances provided in Section 6.4.1.3 will reveal that fusion of decisions of *task-classifiers* trained with mesh networks formed for multiple cognitive tasks is crucial for gender classification.

5.5.1.2 Fusion of Decisions of All Base-Layer Subband-Classifiers in a Task

In this subsection, our purpose is to understand if fusion of only *subband-classifiers* trained with mesh networks formed while individuals perform a single cognitive task is enough to discriminate gender. By comparing the results of the proposed architecture with the 2-level architectures in Section 5.5.1.3 and Section 5.5.1.4, we can understand if mesh-networks formed while individuals perform multiple-tasks contain complementing information for gender classification.

In this architecture, we fuse the information obtained from only *subband-classifiers* trained for a single task as shown in Figure 5.5. Number of classifiers trained in the base-layer varies with the decomposition level. We train $2L$ classifiers using approximation and detail parts of signals for each layer. After that, we concatenate

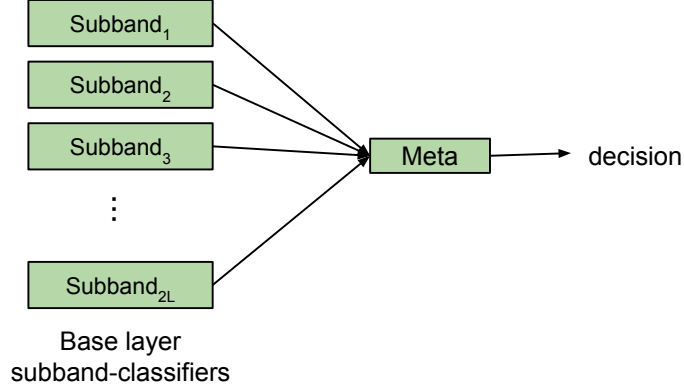


Figure 5.5: 1-level fusion of decisions of base layer *subband-classifiers* in a task

2-dimensional gender posterior probabilities obtained from all base-layer classifiers and form a $4L$ -dimensional feature vector. Using these representations, we train and test a single meta-layer classifier that assigns a gender label to individuals.

As it will be shown in Section 6.4.1.4, fusing the decisions of *subband-classifiers* trained with multi-resolution mesh networks boosts gender classification accuracy.

5.5.1.3 Fusion of Base-Layer Task-Classifiers in a Subband Followed by Fusion of Meta-Layer Subband-Classifiers

In this subsection, we propose a two-level fusion architecture, which aims to combine information obtained from multi-resolution and multi-task mesh networks in a hierarchical manner. By comparing the results of the proposed architecture shown in Figure 5.6 with the 1-level architectures shown in Figure 5.4 and Figure 5.5, we can identify the importance of multi-resolution or multi-task analysis. Moreover, by comparing the results of the proposed architecture with the one in Figure 5.7, we can explore whether the order of fusion of multi-resolution or multi-task mesh networks affect the gender classification accuracy.

In the first level of this architecture, we fuse the information obtained from base layer *task-classifiers*, each of which is trained for a single subband. In the second level, we fuse the decisions of meta-layer *subband-classifiers* as shown in Figure 5.6. In the base layer, we train $7 \times 2L = 14L$ classifiers. In the first meta layer, we train

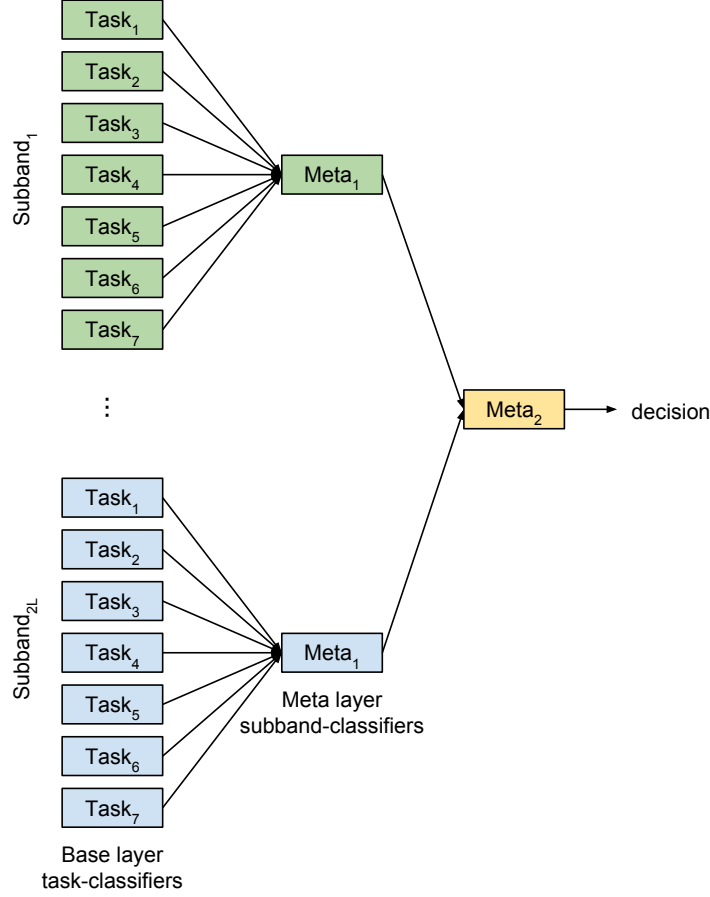


Figure 5.6: 2-level fusion of all classifiers. In the first level, decisions of all base layer *task-classifiers* are fused within each subband. In the second level, decisions of meta layer *subband-classifiers* are fused to give the final decision.

$2L$ *subband-classifiers*, each of which fuses the information obtained from all tasks of the corresponding subband. Note that, inputs of each meta-layer classifier are 14-dimensional vectors whose entries are gender posterior probabilities obtained from 7 base layer *task-classifiers*. Finally, second meta layer classifier fuses the decisions of $2L$ first meta layer classifiers. Therefore, inputs of second meta layer classifier are $4L$ -dimensional vectors. Final gender decision is made by the second meta layer classifier.

The classification performances given in Section 6.4.1.5 will indicate that fusion of decisions of classifiers trained with multi-resolution, multi-task mesh networks outperform 1-level fusion architectures.

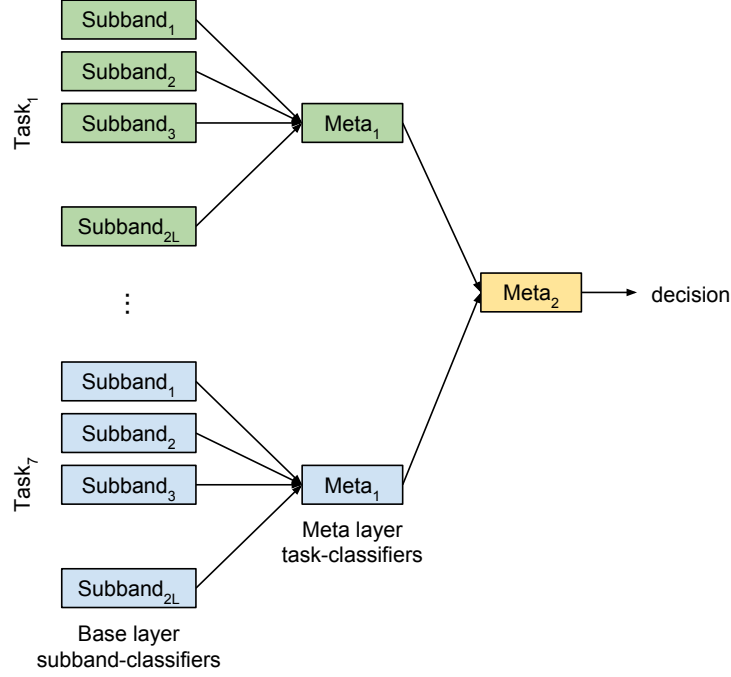


Figure 5.7: 2-level fusion of all classifiers. In the first level, decisions of all base layer *subband-classifiers* are fused within each task. In the second level, decisions of meta layer *task-classifiers* are fused to give the final decision.

5.5.1.4 Fusion of Base-Layer Subband-Classifiers in a Task Followed by Fusion of Meta-Layer Task-Classifiers

Similar to previous subsection, we propose a two-level fusion architecture that combines the information obtained from multi-task and multi-resolution mesh networks in a hierarchical manner. The only difference is that, in this architecture (see Figure 5.7) we change the order of fusion compared to the architecture given in Figure 5.6.

In this architecture, we first fuse the information obtained from base layer *subband-classifiers*, each of which is trained for a single task. Then, we fuse the decisions of meta-layer *task-classifiers* in the second level as presented in Figure 5.7. We train $14L$ base layer classifiers where we train $2L$ classifiers for each task. In the first meta layer, we train 7 *task-classifiers*, each of which fuses the information obtained from all subbands. Note that, inputs of each first meta layer classifier are $4L$ -dimensional vectors of gender posterior probabilities obtained from $2L$ base layer *subband-classifiers*.

The outputs of first meta layer classifiers are concatenated under 14-dimensional vectors and these vectors are used to train and test a second meta layer classifier. By fusing the decisions of first meta layer classifiers, second meta layer classifier gives the final decision.

The performances given in Section 6.4.1.6 will validate that fusion of decisions of classifiers trained with multi-resolution, multi-task mesh networks outperform 1-level fusion architectures or mesh networks formed using original fMRI signals.

5.5.2 Gender Discriminative Power of MADs

In this subsection we explore the relationships among the anatomical regions by investigating the MADs that carry gender discriminative information. We analyze the MADs that discriminate gender under only specific tasks. We also investigate the ones that carry gender discriminative power under all of the tasks.

We estimate the gender discriminative power of MADs using absolute value two-sample t-test with pooled variance estimate. Note that, for each of the 7 tasks, we have $2L$ feature matrices to train base-layer classifiers obtained from distinct subbands. Therefore, for each cognitive task, we identify the most gender discriminative MADs using all of the $2L$ feature matrices. We first divide our feature matrix into two groups, such that each group contains feature vectors of a specific gender. Then, we compute z -values for each MAD separately as follows:

$$z \leftarrow \frac{\mu(g_1) - \mu(g_0)}{\sqrt{\frac{\sigma(g_1)}{n_1} + \frac{\sigma(g_0)}{n_0}}}, \quad (5.19)$$

where g_0 and g_1 denote samples in zeroth and first group, respectively. n_0 and n_1 represent the number of samples belonging to zeroth and first group. Mesh arc descriptors with larger z -values are the ones rejecting the null hypothesis that two groups come from the same distribution with more confidence. In other words, MADs with high z -values discriminate gender compared to the ones having low z -values. For each task we sort the mesh arc descriptors obtained for all subbands, based on their z -values and keep the top-50 mesh weights with maximum z -values.

5.6 Chapter Summary

In the first part of this chapter, we introduced a new representation, called hierarchical multi-resolution mesh networks. First, we decomposed the fMRI signal into multiple subbands using discrete wavelet transform (DWT). Next, by forming local meshes around all nodes, we constructed a mesh network for each subband. We, then, fused the decisions of classifiers trained with multi-resolution mesh networks for cognitive state classification. We proposed to analyze the significance of base-layer and meta-layer features for discriminating cognitive states using t-test. We also computed a number of measures to explore the diversity of classifiers.

In the second part, we fused the decisions of classifiers trained with multi-resolution, multi-task mesh networks under a 2-level FSG architecture for gender classification. For comparison, we proposed 1-level fusion architectures that fuses only multi-resolution or only multi-task mesh networks. We also proposed to analyze the subbands and corresponding anatomical regions of the most gender discriminative of MADs.

CHAPTER 6

EXPERIMENTS AND RESULTS

We have to remain humble about our understanding of the brain, because even our most powerful tools remain pretty blunt instruments for decoding the brain. In fact, we still do not know how to decipher the basic language of how the brain works.

Thomas Insel

In this chapter, we first analyze the cognitive state classification performance of MADs formed within functional (FLM) and spatial (SLM) neighborhoods. We compare the results of MADs with state-of-the-art fMRI representations. We, also, statistically analyze local meshes about how well the seed nodes are represented in terms of their nearest neighbors. Next, we provide encoding results of MADs using Fisher Vectors (FV), Vector of Locally Aggregated Descriptors (VLAD) and Bag-of-Words (BoW) methods. We compare the discriminative power of MAD encodings by MADs and state-of-the-art fMRI representations. We also use the codewords obtained by GMM and k-means clustering to create a *brain connectivity dictionary*. Finally, we put forth a relationship between the energy of FV columns and classification accuracy.

After the analysis of MADs extracted from original fMRI signals, we perform multi-resolution analysis. We first share the cognitive state classification results obtained by the fusion of multi-resolution mesh-networks. We compare the results of multi-resolution experiments with single-resolution experiment. In order to investigate

the power of MADs compared to state-of-the-art representations, we train and test classifiers with pairwise correlations (PC) and representative time series (RTS) to classify cognitive states. We further compare the fusion of classifiers by FSG with majority voting approaches. After sharing the classification performance comparisons, we explain how mesh networks of multiple resolutions change by analyzing degree, strength, betweenness centrality of nodes of mesh networks and global efficiency of mesh networks. Additionally, we test the significance of meta-layer and base-layer features to discriminate tasks. In order to validate our assumption that multi-resolution signals contain complementing information about cognitive tasks, we measure the diversity of classifiers trained with multi-resolution mesh networks.

The suggested HMMN is also tested in another problem namely, gender classification. We provide gender classification results obtained by fusing multi-resolution multi-task mesh networks under a two-layer FSG architecture. We compare our results with single-resolution experiments, 1-level fusion of only multi-resolution or only multi-task experiments. Similar to cognitive state classification experiments, we compare MADs with PC and RTS and we compare fusion of classifiers by FSG with majority voting. We also analyze the significance of MADs that carry the most gender discriminative information. We identify significant anatomical region pairs such that MADs formed between them carry gender discriminative power under specific tasks or under all tasks.

6.1 Experiments and Results of Employing Mesh Arc Descriptors

In order to observe the validity of the proposed mesh representation, we perform two groups of experiments. In the first group, we train and test a Support Vector Machine (SVM) classifier implemented in Liblinear [42] by using the labeled fMRI data recorded using the experimental setups explained in Section 4.1. In this group of experiments, we examine the power of the mesh arc descriptors for representing the brain states. In the second group of experiments, we analyze the validity of the local meshes by exploring the similarities among the node time series in each mesh, and the distribution of the error of the proposed regularized linear regression model, used for estimating the mesh arc descriptors.

6.1.1 Comparison of Mesh Arc Descriptors with State-of-the-art fMRI Features

We compare the mesh arc descriptors of the proposed spatial and functional local meshes with the state-of-the-art features, by measuring the performance of the classifiers trained by these features. We train an SVM classifier with linear kernel, and classify an unknown stimulus. We perform seven sets of classification experiments on three different groups of fMRI datasets:

- In the first and second set of experiments, we represent a cognitive stimulus using the arc descriptors of meshes. The classification performances are measured when the arc descriptors obtained from the spatially (**SLM**) and functionally local meshes (**FLM**) are fed as the input of an SVM classifier.
- In the third and fourth set of experiments, we tested the classification performances of SVM by employing features obtained using Local Mesh Model (**LMM**) and Functional Mesh Model (**FMM**). Recall that both **LMM** and **FMM** employ only a single measurement from the BOLD response for each voxel omitting the rest of the signals measured during a stimulus.
- In the fifth group of experiments, we test the performance of SVM classifiers which employ features that are computed using pairwise Pearson correlation (**PC**). First, we compute the pairwise correlations between the BOLD responses given by seed nodes and each of their p -nearest neighbors for each stimulus. In other words, first meshes are constructed around seed nodes, and then pairwise correlations are computed between node pairs within meshes instead of estimating the arc descriptors within a neighborhood. Then, we form our feature vector by concatenating the correlation values in all meshes. Note that, we obtain feature vectors whose sizes are equal to that of **SLM** and **FLM** by concatenating only the distances within meshes to make a fair comparison.
- In the sixth set of experiments, we analyze the performance of SVM classifiers where representative time series (RTS) containing raw node intensity values are used as features to train and test classifiers. For **RTS-mid**, we only used the third instance $x_{w,i}(3)$ in the middle of a BOLD response $x_{w,i}$ following

our assumption that if a voxel becomes active, then its HRF reaches to its peak value after 5-6 seconds (around $t = 3$). On the other hand, **RTS-mean** depicts the case where we employed the average of D_w measurements (number of measurements in window w) of $x_{w,i}$. We also concatenate each of the D_w measurements of BOLD response $x_{w,i}(t)_{t=1}^{D_w}$ to obtain results for **RTS-all**. The dimension of a feature vector constructed for **RTS-mid** and **RTS-mean** equals to 1254 for visual object recognition experiment, 800 for emotional memory retrieval (EMR) experiment and 90 for HCP task fMRI data. Moreover, since we concatenate all measurements for **RTS-all**, the size of the feature vectors used for **RTS-all** equals to D_w times the size of features vectors for **RTS-mean** and **RTS-mid**. Since each task of HCP data has different durations and we do not perform event-related task classification, we cannot employ **RTS-mid** and **RTS-all** experiments on HCP data.

- In the seventh set of experiments, we compare mesh arc descriptors to two popular fMRI data representation methods namely, Principal Component Analysis (**PCA**) and spatial Independent Component Analysis (**ICA**). We map the fMRI data onto the first 100 components that retain 99% of the variance. Following **PCA**, we applied spatial **ICA** to our raw data using FastIca [65].

6.1.2 Classification Results for Brain Decoding

Since visual object recognition and emotional memory retrieval experiments are event-related and classes are discriminated at finer granularity, we use voxels as nodes in this subsection for these two datasets. On the other hand, for HCP dataset, we aim to classify cognitive tasks, which can be discriminated at coarser granularity. Therefore, we use anatomical regions as nodes. Also, in this subsection we perform intra-subject classification using SVM classifiers with visual object recognition and emotional memory retrieval experiments since the number of subjects is small in these datasets. On the other hand, we perform inter-subject classification using HCP task fMRI dataset since the number of subjects is large.

In visual object recognition and emotional memory retrieval experiments, we perform nested cross validation across runs [146]. The outer loop runs for 6 folds,

and within each fold of the outer loop, an inner loop runs for 5 folds for parameter tuning. For each fold of the outer loop, we employ measurements obtained from 5 runs for decoding, and 1 run for test. For each fold of the inner loop, we divide the decoding part into 4 runs of training set and 1 run of validation set. We use the validation set to optimize the cost parameter, C of SVM and the regularization parameter λ . Similarly, we also optimize the number of neighbors p used for computation of functional neighborhoods and radius of neighborhoods π used for computation of spatial neighborhoods. Note that, since optimization of p using information criteria does not yield a significant change in the accuracy [98, 99, 100], we optimize it using cross validation. After we estimate the parameters in the inner loop, we train classifiers using decoding part with the estimated *optimal* parameters, and measure the test performance using the unseen test set. We searched for the optimal values of λ in $\{0, 0.125, 0.25, 0.5, 1, 2, 4, 8\}$, SVM cost parameter C in $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ and π in $\{1, \sqrt{2}, \sqrt{3}\}$. We optimize the number of neighbors p in $\mathcal{P}^1 = \{2, 3, \dots, 30\}$ for visual object recognition experiment, and p in $\mathcal{P}^2 = \{5, 6, \dots, 15\}$ for emotional memory retrieval experiment.

For HCP task fMRI experiments, we split our data into 8 chunks, where each chunk contains fMRI data collected from 101 participants. We perform nested cross validation in which the outer loop runs for 8 folds, and within each fold of the outer loop, an inner loop runs for 7 folds for parameter tuning. We search for the optimal values of λ in $\{0, 32, 64, 128, 256, 512\}$, C in $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ and p in $\mathcal{P}^3 = \{5, 10, 15, 20, 25, 30\}$ on the validation set. Since the data scale in these datasets is at region level, we searched for the optimal λ and p values in a different range during these experiments. Note that, unlike voxels, *supervoxels* are not located on a full grid. Therefore, we select the number of spatial neighbors from the set $\mathcal{P}^3 = \{5, 10, 15, 20, 25, 30\}$ for HCP dataset.

We analyze the effect of similarity measures on defining the neighborhood system of local meshes in subsection 6.1.3. As will be explained in this subsection, when we use functional similarity between the nodes, Pearson correlation provides a relatively better decoding performance compared to partial correlation and Granger causality [55]. For this reason, we conduct the experiments on **FLM** by just measuring Pearson correlation to select nearest neighbors.

Table 6.1: Classification performance (%) of SVM classifier computed in visual object recognition experiment.

Participant	FLM	SLM	FMM	FMM	LMM	LMM
			mean	mid	mean	mid
1	91.42	93.14	73.14	80.00	70.29	74.28
2	83.81	81.43	67.14	68.09	68.57	73.33
3	92.86	83.33	67.62	75.71	68.57	78.09
4	89.05	76.67	71.90	75.23	73.33	74.76
5	78.09	81.43	68.57	75.71	63.33	71.43
Avg.	87.04	83.20	69.67	74.95	68.82	74.38
Participant	PC	RTS	RTS	RTS	PCA	ICA
		mean	mid	all		
1	72.00	74.85	78.29	75.43	73.71	86.29
2	71.90	71.90	74.76	76.19	73.33	79.52
3	53.33	70.00	75.71	76.67	70.48	79.52
4	62.38	74.29	78.57	70.48	72.86	87.62
5	57.14	68.09	70.48	67.62	68.57	75.71
Avg.	63.35	75.56	74.68	73.27	71.79	81.73

Visual Object Recognition Experiment: We give the average classification performances of 6-fold obtained for visual object recognition experiment in Table 6.1. Results reflect that **FLM** and **SLM** features discriminate brain states much better than other features. FLM gives slightly better accuracy (87.04%) compared to SLM (83.20%) and these two representations outperform variants of FMM and LMM representations significantly (12% - 19% increase). These results suggest that, employing all of the temporal measurements during the estimation of relationship among nodes brings significant information. While employing ICA performs similar to SLM, variants of RTS and PCA perform much worse. Moreover, the worst results (63.35%) are obtained when pairwise correlations (PC) are employed as features since the number of measurements in predefined windows is not enough to capture pairwise rela-

tionships.

Emotional Memory Retrieval Experiment: We provide performances for 2-class and 4-class classification experiments. In 2-class experiments the classes correspond to neutral and emotional tasks (Table 6.2). In 4-class experiments, we classify kitchen utensil, furniture, fear and disgust, where kitchen utensil and furniture belong to the neutral category, and fear and disgust belong to the emotional category (Table 6.3).

The results of 2-class and 4-class classification experiments show that we obtain substantially better performances using features extracted for **FLM** (90.70% for 2-class and 66.56% for 4-class) and **SLM** (89.71% for 2-class and 62.86% for 4-class) compared to the performances of other features given in Table 6.2 and 6.3. We validate the importance of employing all of the temporal measurements during the estimation of relationship among nodes on this dataset by showing in both Table 6.2 and 6.3 that FLM and SLM outperform variants of FMM and LMM. ICA perform similar to SLM in 2-class classification whereas, the feature spaces of **RTS-mid**, **RTS-mean**, **RTS-all** and PCA give poor performance.

The emotional memory retrieval experiment data has a drawback for 4-class brain decoding. The number of samples at each run is nearly the same in emotional and neutral categories, which provides a balanced dataset. However, runs are mostly unbalanced in four subcategories. As a consequence, we observe nearly chance results with raw data when we perform 4-class classification. In 4-class classification, ICA also do not perform well.

Human Connectome Project (HCP) task fMRI Data: Results in Table 6.4 reflect that extracting **FLM** and **SLM** features from HCP data represents the cognitive tasks in inter-subject experiments much better compared to the methods given in Table 6.4. **FMM-mean**, **LMM-mean** and **RTS-mean** give nearly chance results. On the other hand, employing **PC** features lead to a good classification accuracy since number of measurements in a window is large enough to represent relationships between anatomical regions. This result indicates that a functional connectivity feature, such as Pearson correlation, can capture the information content of the data. Considering the long time series of responses in HCP dataset, this result is expected.

Table 6.2: Classification performance (%) of SVM classifier computed in emotional memory retrieval experiment (2-class).

Participant	FLM	SLM	FMM mean	FMM mid	LMM mean	LMM mid
1	97.62	97.14	80.48	83.33	78.10	80.48
2	92.86	90.95	79.52	80.48	77.62	79.52
3	73.33	69.52	67.14	55.71	60.48	50.48
4	87.62	87.62	73.33	65.71	77.62	67.14
5	92.38	91.90	77.62	61.43	72.86	56.19
6	98.10	97.62	71.43	74.76	81.90	81.90
7	92.86	92.86	65.24	67.62	64.29	65.24
8	90.00	89.05	66.19	63.33	68.10	56.67
9	90.00	90.00	79.52	72.86	80.48	69.52
10	93.81	89.52	71.43	65.24	72.38	54.76
11	95.24	95.24	69.05	70.00	67.62	67.62
12	78.57	77.62	71.90	61.90	76.67	62.38
13	96.67	97.14	53.33	59.05	67.62	65.71
Avg.	90.70	89.71	71.25	67.80	72.75	65.97

Participant	PC	RTS mean	RTS mid	RTS all	PCA	ICA
1	84.29	81.43	76.19	92.86	74.76	94.29
2	76.19	83.81	75.24	90.48	73.81	82.85
3	61.90	57.14	50.48	67.62	58.57	64.76
4	76.67	62.38	56.67	73.33	65.24	78.57
5	84.29	66.19	54.76	71.43	64.76	83.81
6	89.05	66.19	68.57	79.05	72.86	88.57
7	80.00	63.81	56.67	81.90	72.86	85.24
8	79.05	67.14	60.00	79.05	63.33	78.57
9	78.57	79.05	72.38	94.76	71.90	90.48
10	74.76	61.43	59.52	74.29	63.33	79.52
11	89.05	75.71	78.10	87.14	79.05	93.81
12	78.10	63.81	66.67	75.71	80.95	98.10
13	88.10	58.57	55.24	69.52	58.57	90.00
Avg.	80.00	68.21	63.88	79.78	69.23	85.27

Table 6.3: Classification performance (%) of SVM classifier computed in emotional memory retrieval experiment (4-class).

Participant	FLM	SLM	FMM mean	FMM mid	LMM mean	LMM mid
1	81.43	76.67	60.48	57.62	41.43	53.81
2	75.24	68.57	51.91	41.43	48.09	38.57
3	40.95	41.43	34.76	32.38	32.86	32.38
4	61.90	57.62	43.33	32.86	40.48	31.43
5	73.33	67.62	47.62	31.43	35.24	32.86
6	85.71	77.14	49.05	49.52	47.62	47.62
7	59.05	57.62	40.95	34.29	32.86	36.67
8	64.29	64.76	46.19	28.10	43.33	30.00
9	51.43	50.95	35.24	34.29	29.05	24.29
10	65.71	54.29	37.62	27.62	36.19	32.86
11	69.05	70.95	45.71	44.29	42.38	38.57
12	54.76	51.43	40.00	38.10	42.86	45.71
13	82.38	78.10	43.33	37.14	50.95	33.33
Avg.	66.56	62.86	44.32	37.62	40.26	36.78
Participant	PC	RTS mean	RTS mid	RTS all	PCA	ICA
1	33.81	22.38	22.38	22.38	21.43	20.95
2	48.57	20.95	19.52	31.91	21.90	58.57
3	26.19	22.86	22.86	22.86	22.86	22.86
4	32.86	21.90	21.90	38.10	21.90	22.38
5	22.86	21.90	21.90	21.90	22.38	21.43
6	30.95	21.90	23.33	31.43	21.90	22.38
7	31.43	30.95	30.00	30.00	30.48	21.90
8	27.14	22.38	21.90	22.38	21.90	22.38
9	39.05	23.81	24.29	30.95	21.90	22.38
10	47.14	26.19	24.76	35.71	22.38	49.52
11	61.90	22.38	22.38	24.76	22.38	22.38
12	47.14	32.38	37.14	36.67	21.90	20.95
13	63.33	21.90	21.43	21.90	20.95	21.43
Avg.	39.41	23.99	24.14	28.54	22.64	26.89

Table 6.4: Classification performance (%) of SVM classifier for HCP task fMRI data.

FLM	SLM	FMM	LMM	PC	RTS
		mean	mean		mean
96.32	96.83	27.28	14.63	90.09	14.28

In HCP dataset, each window covers a large duration of the entire duration of cognitive task instead of short BOLD response given to a stimulus. Therefore, for this data set, we cannot perform experiments using the middle values of signals (**FMM-mid**, **LMM-mid**, **RTS-mid**). Moreover, since the number of measurements collected for each cognitive task is different, we cannot concatenate them under a feature vector as in **RTS-all**.

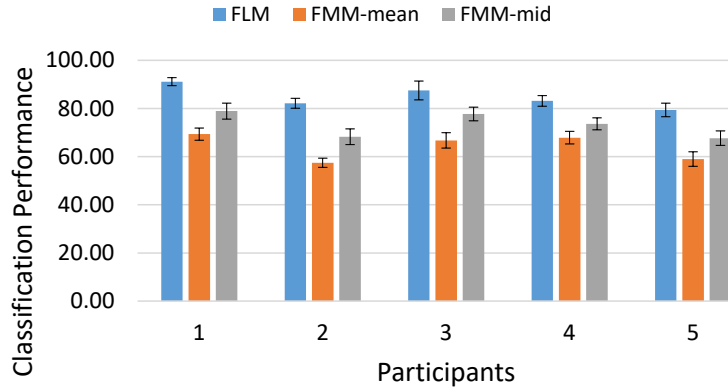


Figure 6.1: Mean and standard deviation of classification performances of SVM computed in visual object recognition experiment.

We provide the mean and standard deviations of classification performances obtained by different meshes that are constructed with $p \in \mathcal{P}^1 = \{2, 3, \dots, 30\}$ for visual object recognition experiment in Figure 6.1, and with $p \in \mathcal{P}^2 = \{5, 6, \dots, 15\}$ for emotional memory retrieval experiment in Figure 6.2 and Figure 6.3. In Figure 6.1, bars represent the mean classification performance averaged over all $p \in \mathcal{P}^1$, whereas in Figure 6.2 and Figure 6.3, bars represent the mean classification performance averaged over all $p \in \mathcal{P}^2$. In these figures, each error bar represents standard deviation of the corresponding performance value. For each participant, the first bar depicts the performance of MADs obtained by functional neighborhood in **FLM**, which are

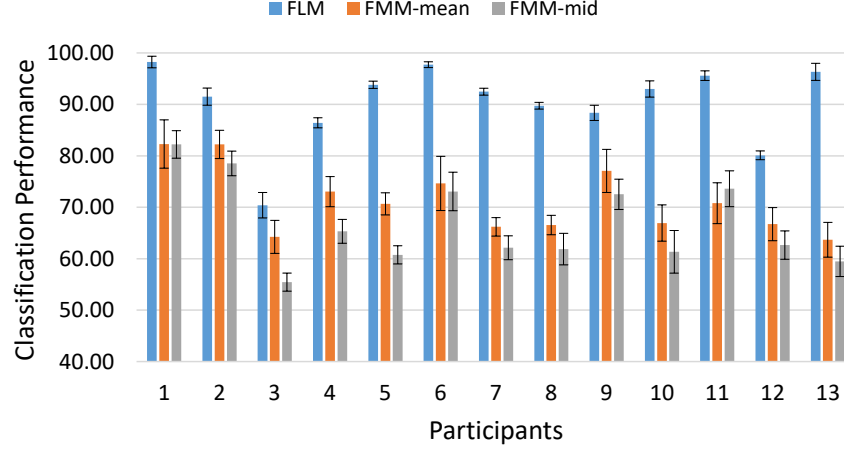


Figure 6.2: Mean and standard deviation of classification performances of SVM computed in emotional memory retrieval experiment (2-class).

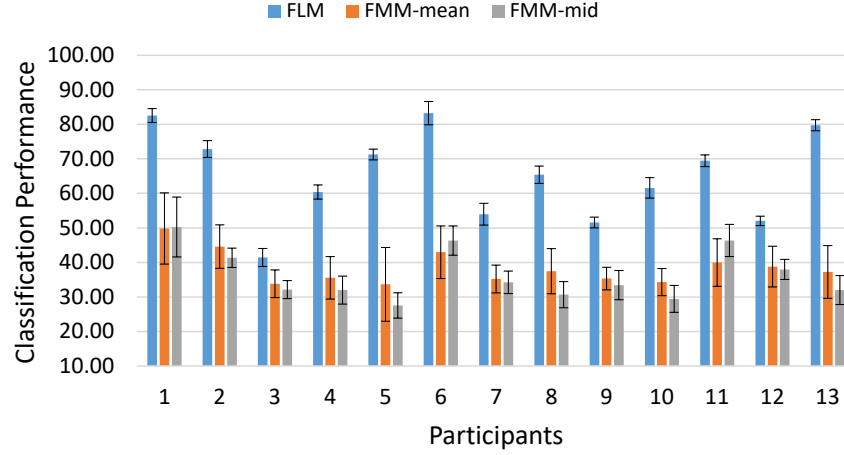


Figure 6.3: Mean and standard deviation of classification performances of SVM computed in emotional memory retrieval experiment (4-class).

extracted using the entire temporal measurements of the BOLD responses, while the last two bars depict performances obtained using the mean or middle values of BOLD responses. We observe that standard deviation of performance values obtained using various mesh sizes decreases, when temporal measurements are considered. In other words, the methods employing temporal measurements are more robust to changes of mesh size p than the ones which do not employ temporal measurements.

Notice that, the entries of BOLD response vector, $\mathbf{x}_{w,i} = [x_{w,i}(t)]_{t=1}^{D_w}$, at each neighboring node, contribute to the estimation of mesh arc descriptors in the regularized linear model of (4.4). If a node has a noisy entry in the BOLD response vector,

the effect of this entry on the estimation of mesh arc descriptors will be relatively small considering the entire time sequence of BOLD response. However, when we take a single value, such as the mean or middle measurement, not only we lose important information about the BOLD response, but also the mesh representation becomes vulnerable to artifacts. Averaging the BOLD response may avoid the effect of noise with the price of excessive smoothing, which results in losing the discriminating information of fMRI data among the cognitive states. This fact can be roughly observed when we compare the subject dependent performances of **FLM** and **SLM** to that of **FLM-mean**, **FLM-mid**, **SLM-mean** and **FLM-mid** methods. For example, **FLM** performance of first participant in Figure 6.2 varies between 96% - 99% whereas **FMM-mean** and **FMM-mid** performances vary between 75% - 91% and 79% - 87%, respectively. The high variance in performances of the mesh representations estimated by single BOLD response value may be attributed to the effect of artifacts hidden in fMRI data.

6.1.3 Analysis of Similarity Measures for Neighborhood Selection to Form Local Meshes

Recall that, we form a local mesh around each node to form a mesh network. The concept of locality is based on a neighborhood system, which is defined by the similarity between node time series. In this subsection we analyze the effect of various similarity measures on the classification performance. We compare the performance of **FLM** formed using Pearson correlation, partial correlation and Granger causality as the similarity measure to select functionally nearest neighbors in the visual recognition experiment.

Our results on visual object recognition experiment reflects that selecting functional neighbors based on Pearson correlation leads to better accuracy compared to partial correlation or Granger causality to select the neighbors of seed voxels. We also observe that meshes formed by Granger causality has the lowest classification performance on this dataset (see Table 6.5). Therefore, for the rest of the experiments, we suffice to employ Pearson correlation to select functionally nearest neighbors while forming a local mesh.

Table 6.5: Classification performance (%) of SVM classifier computed in visual object recognition experiment.

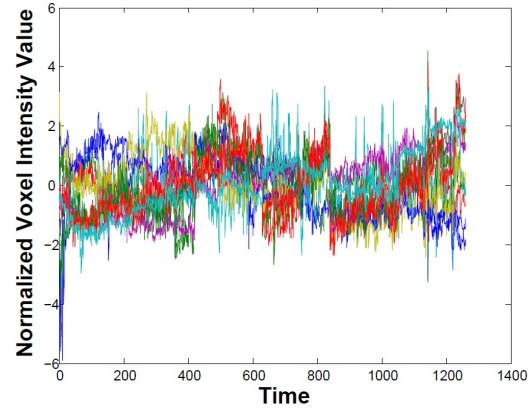
P	FLM Corr	FLM Parcorr	FLM Granger
1	91.42	88.33	86.67
2	83.81	87.50	78.24
3	92.86	87.03	87.96
4	89.05	79.63	77.78
5	78.09	81.95	74.07
Avg.	87.04	84.89	80.94

6.1.4 Analysis of Local Meshes

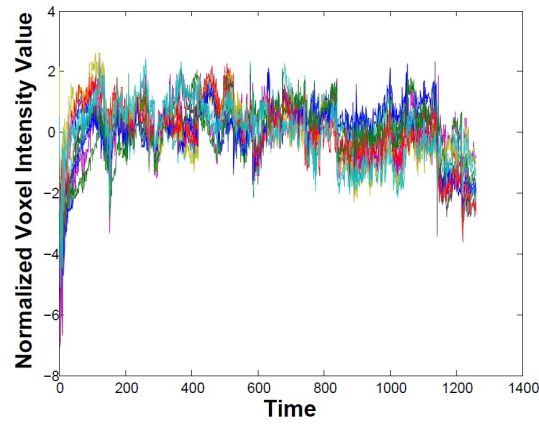
The major assumption of the proposed mesh representation is that the node BOLD responses are similar to each other at a predefined locality, so that a BOLD response of a node can be represented by a linear combination of the BOLD responses of its nearest neighbors. In other words, we assume that the error of a regularized linear regression model is small enough such that a cognitive stimulus can be represented by a set of local meshes, where the arc descriptors of the meshes can be estimated by minimizing the expected square error. Let us now, empirically observe the behavior of voxel time series in a certain locality in the following subsections.

6.1.4.1 Statistical Analysis of Correlations of Voxels

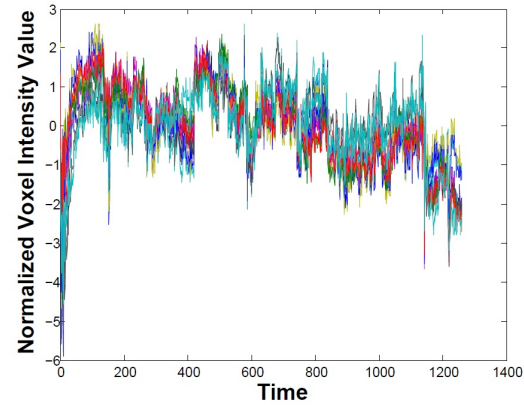
In order to understand how the nodes of a local mesh, especially voxels within a predefined neighborhood behave, and how their behavior differs from the behavior of random voxel sets, we plot the time series a seed voxel with *i*) a set of randomly selected voxels (Figure 6.4a), *ii*) its ten spatially (Figure 6.4b) and *iii*) ten functionally (Figure 6.4c) nearest neighbors. We observe that, sample voxels located within spatial and functional neighborhoods perform similar under presentation of the same stimuli.



(a) Randomly selected voxels.



(b) A random seed voxel with its spatial neighbors.



(c) A random seed voxel with its functional neighbors.

Figure 6.4: BOLD responses of a voxel set consisting of 10 voxels, each of which is plotted with a different color.

Although we select spatially or functionally closest voxels as neighbors, we need to further analyze the distribution of correlation values between the seed voxels and their neighbors to understand if the closest voxels are also highly correlated. Therefore, we compute pairwise relationship (i.e. correlation) between the seed voxels and the surrounding voxels using Pearson correlation. In our analysis, surrounding voxels are selected as (a) spatially nearest neighbors of seed voxels, and functionally nearest neighbors of seed voxels selected using (b) Granger causality (c) partial correlation (d) Pearson correlation. First, we compute correlations between ten surrounding voxels and the seed voxel. Then, we depict the histograms of all correlations computed for all voxels with their surrounding voxels in meshes.

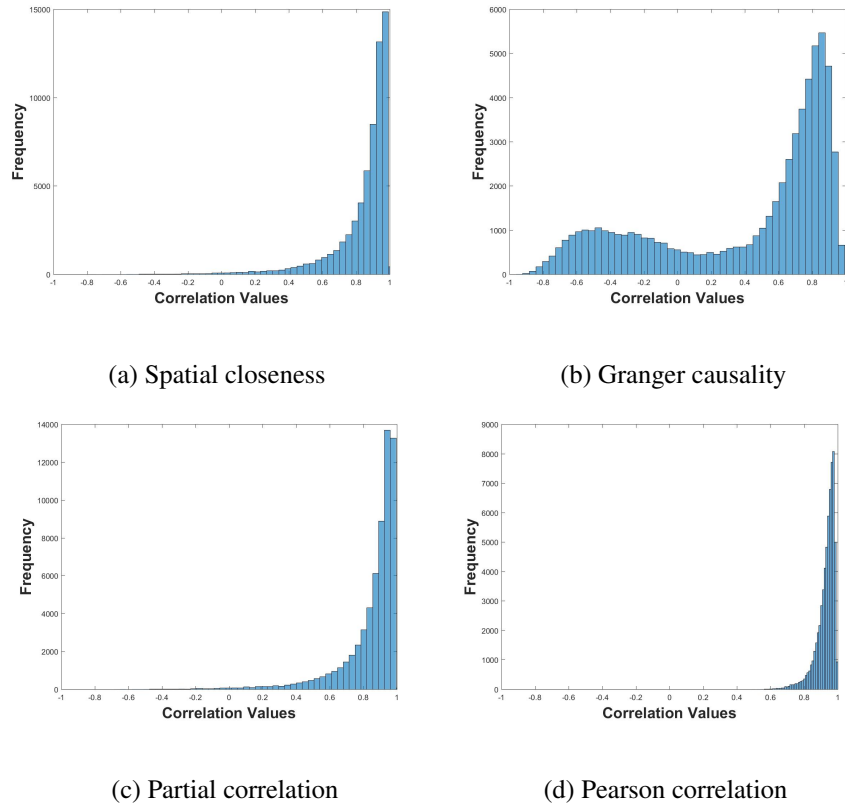


Figure 6.5: Histograms of correlations between seed voxels and surrounding voxels selected using (a) Spatial closeness, (b) Granger causality, (c) Partial correlation and (d) Pearson correlation.

The normalized Granger causality histogram in Figure 6.5b reflects that neighboring voxels with maximum Granger causality value may have negative correlations with the seed voxel. On the other hand, the results show that correlation values of the

voxels observed with the highest frequency are close to 0.9 when the histograms are computed for meshes formed using spatially (see Figure 6.5a) and functionally close voxels selected using partial correlation (Figure 6.5c) and Pearson correlation (Figure 6.5d). These observations indicate that the voxels modeled in the same mesh have statistical relationship with each other. Therefore, one may expect that the linear relationship among the BOLD responses at a locality, defined by a neighborhood system, provides a reasonable fit to the data.

6.1.4.2 Statistical Analysis of Models

Recall that, we represent the intensity values of each node as a regularized linear combination of those values of its p -nearest neighbors in our models. Yet, we obtain a regression error for each mesh for the estimation of seed nodes considering its neighbors. In the next set of experiments, we analyze how well the seed nodes are represented in terms of their nearest neighbors by exploring the model estimation error of the proposed representation.

For this purpose, we employ a goodness of fit measure, called R^2 , defined as one minus the residual variation divided by total variation, where residual variation is the summation of square errors obtained from regression model, and total variation is the variance of the distribution of the actual data.

More precisely, let error $\epsilon_{w,i}$ be the ridge regression error vector for window w and node i computed as:

$$\epsilon_{w,i} = \hat{x}_{w,i} - x_{w,i}, \quad (6.1)$$

where $\epsilon_{w,i} = [\epsilon_{w,i}(t)]_{t=1}^{D_w}$ and D_w represents the number of measurements recorded within window w . Then, we compute the residual variation as,

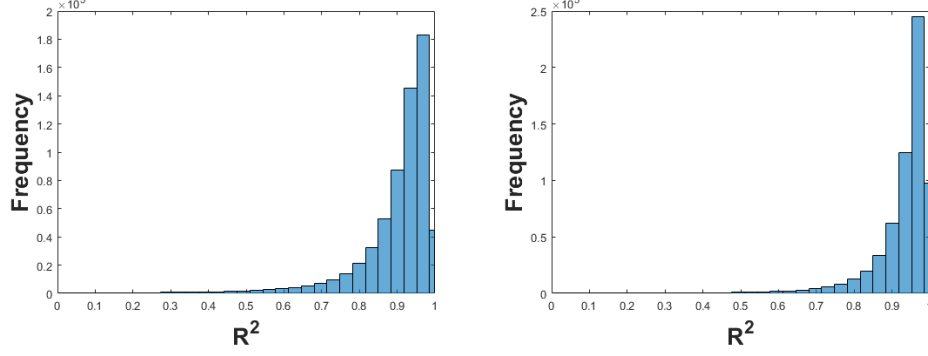
$$SS_{w,i}^r = \sum_{t=1}^{D_w} \epsilon_{w,i}(t)^2, \quad (6.2)$$

On the other hand, total variance is computed as,

$$SS_{w,i}^t = \sum_{t=1}^{D_w} x_{w,i}(t)^2. \quad (6.3)$$

Then, the R^2 is computed as

$$R_{w,i}^2 = 1 - \left(\frac{SS_{w,i}^r}{SS_{w,i}^t} \right). \quad (6.4)$$



(a) Spatial neighbors.

(b) Functional neighbors.

Figure 6.6: Histograms of R^2 values computed when the meshes are formed using (a) Spatial neighbors and (b) Functional neighbors.

Notice that the R^2 measure takes values between 0 and 1, such that the values closer to one represent better fit of models. In the analysis, we computed R^2 values for all meshes of a participant formed for all samples and around all voxels. In Figure 6.6, we plot the histograms for R^2 values computed when the meshes are formed using (a) spatially nearest voxels, and (b) functionally nearest voxels.

In the experiments, if the seed voxels of meshes formed using spatial or functional neighbors are represented in terms of their neighbors, then we observe that the mean values of histograms for R^2 are 0.90 and 0.93, respectively (Figure 6.6a and 6.6b). Therefore, employing spatial or functional neighbors during the construction of meshes results in good model fit, where the latter leads to slightly better fit compared to the former.

6.2 Experiments and Results of Encoding the Mesh Arc Descriptors

Recall that, we encode mesh arc descriptors using Fisher Vector (FV), Vector of Locally Aggregated Descriptors (VLAD) and Bag-of-Words (BoW) methods. In this subsection, we share the results of encoding the MADs.

Table 6.6: Classification performance (%) obtained using MADs

p	HCP	EMR
10	88.14	71.17
20	91.14	71.10
30	93.00	70.15
40	93.86	71.10

We perform two sets of experiments. First, we classify the cognitive tasks of HCP fMRI dataset and emotional memory retrieval (EMR) dataset using the raw and encoded MADs to see the effect of encoding on MADs and compare the performance of various encoding methods. In order to examine the power of MADs compared to popular fMRI representation methods, we also classify the cognitive tasks using state-of-the-art fMRI data representations. Finally, we visualize and analyze the codewords of the proposed *brain connectivity dictionary*. Note that, we need a large amount of samples to learn the codewords of the dictionary. Therefore, in this subsection, we perform inter-subject experiments using HCP and EMR datasets.

6.2.1 Classification Results

In order to perform classification, we first compute FV ($\mathbf{G}_{\beta}^A$), VLAD ($\mathbf{V}^A$) and BoW ($\mathbf{B}^A$) encoding of local MADs for all tasks and all participants. Then, we use raw and encoded MAD vectors to train and test linear SVM classifiers. We measure the classification performance of HCP dataset using a 10-fold cross validation (CV) scheme by randomly splitting the data into 10 subsets according to a uniform distribution, training each model on 9 splits, and testing the learned model on the remaining split. Note that, training and test splits contain data associated to different participants for each fold. For EMR dataset, we perform 13-fold cross-validation, where each split contains data collected from a participant. We train each model on 12 splits and test the learned model on the 13th split. In our experiments, we used a subset of randomly selected 100 subjects of HCP dataset. We provide classification accuracies of raw MADs in Table 6.6.

Table 6.7: Classification performance (%) computed for different encoding methods obtained using HCP and EMR datasets.

Encoding	p	HCP			EMR		
		PCA	ICA	None	PCA	ICA	None
FV	10	91.71	87.29	86.00	75.97	73.88	73.81
	20	94.71	85.29	89.29	76.08	73.41	72.27
	30	95.29	87.71	91.71	75.05	71.61	71.76
	40	95.71	87.86	92.14	75.53	71.87	71.79
VLAD	10	89.86	84.29	86.71	74.91	71.10	72.09
	20	91.43	88.29	89.00	74.47	71.14	71.76
	30	92.57	89.86	90.43	74.62	71.90	72.09
	40	93.43	90.00	91.14	73.70	71.25	72.34
BoW	10	62.86	41.57	54.57	64.76	59.67	60.62
	20	68.57	42.14	57.86	64.91	57.00	58.63
	30	67.29	46.14	60.43	63.88	58.42	61.61
	40	68.71	45.00	60.29	63.55	57.25	61.72

In the proposed framework, we use only training data for implementation of all methods that are employed in learning phase, such as computation of PCA and ICA using MADs, estimation of Gaussian mixture models, k-means clustering and training SVM classifiers. We examine the proposed methods for various number of neighbors $p \in \mathcal{P}^4 = \{10, 20, 30, 40\}$, in each mesh to observe the effect of degree of locality of the meshes. Considering the fact that MADs are N -dimensional vectors, employment of smaller p values provides increased locality and sparser MAD vectors. Before employment of FV encoding, we apply PCA or ICA to MADs for decorrelation purpose and did not change their dimension ($D = N$). Moreover, we select the number of Gaussians in GMM and the number of clusters in k-means using $k \in \{20, 40, 60, 80, 100, 120\}$. We compute average performance of cross validation for all κ . Accuracy values denoted in Table 6.7 show maximum of these performances with respect to the values of k .

Results given in Table 6.7 show that, FV encoding of MADs computed after employment of PCA gives the best performance to classify cognitive tasks compared to VLAD and BoW encoding. Unlike PCA, applying ICA on MADs before encoding with FV, VLAD and BoW decreases the performance of MADs. We observe that VLAD encoding of MADs does not increase the discrimination power of MADs. Although accuracy values of VLAD-MAD are comparable with that of MAD, BoW representations of MADs yield substantial decrease in accuracy.

We also share the running times of each step of our framework for $p = 40$ on HCP dataset in Table 6.8. It is observed that computation of ICA both takes a large amount of time and decreases the encoding performance. On the other hand, computation of PCA takes much shorter time, while contributing a lot to the classification accuracy. Computation of FV, VLAD and BoW takes much shorter compared to computation of GMM and k-means.

Table 6.8: Running times of each step of the framework measured in seconds.

MAD	PCA / ICA	GMM / k-means	FV / VLAD / BoW
105.6	8.4 / 373.6	211.3 / 62.8	14.3 / 7.7 / 11.9

In order to investigate the representation power of our encoding method, we compare the performance of the encoded MADs to that of state-of-the-art fMRI representations. These representations include average region BOLD responses, Pearson and partial correlations, mutual information and Granger causality between BOLD responses. We further apply PCA and ICA on average BOLD responses to change our feature space. When the average region BOLD responses are employed for classification, we obtain almost chance accuracy with 7-class HCP dataset (14.29%) and 2-class EMR dataset (51.24%). Since average BOLD signals do not carry significant information, we also get poor results when we apply PCA and ICA on them. On the other hand, when pairwise correlations between all pairs of region BOLD responses are used as features for HCP dataset, we obtain a classification accuracy of 77.42% with Pearson correlation and 67.86% with partial correlation. Since EMR dataset is event-related, partial correlation cannot be computed for this dataset yet using Pearson correlation gives 66.19% accuracy. Employing Granger causality computed be-

tween all pairs of node time series as a feature vector gives chance results for both datasets while employing mutual information gives 27.85% on HCP data and 64.32% on EMR dataset. In Table 6.9, we observe a substantial increase (10 – 15%) obtained MADs. Finally, by encoding MADs with FV, we obtain a (2 – 5%) increase.

Table 6.9: Classification accuracy (%) measured with state-of-the-art fMRI representations.

Methods	Dataset	
	HCP	EMR
BOLD Response	14.29	51.24
BOLD + ICA	14.29	52.27
BOLD + PCA	14.14	51.32
Pearson Corr.	77.42	66.19
Partial Corr.	67.86	-
Granger Causality	19.61	50.29
Mutual Information	27.85	64.32
MAD	93.86	71.17
FV+MAD+PCA	95.71	76.08

In order to study the significance of the proposed FV encoding of MADs compared to just using MADs, we further employ McNemar’s test with Yates’ correction, which is a first order check on the statistical significance of an observed difference in recognition performance [38]. It uses the times of success/failure trials of the compared algorithms. Let N_{ss} , N_{sf} , N_{fs} , N_{ff} denote the times of both algorithms succeed, the first algorithm succeeds but the second fails, the second algorithm succeeds but the first fails and both algorithms fail. Then, we compute χ^2 as follows:

$$\chi^2 = \frac{(|N_{sf} - N_{fs}| - 1)^2}{\sqrt{N_{sf} + N_{fs}}} \quad (6.5)$$

After χ^2 value is computed, corresponding p -value is found from χ^2 distribution table. If the p -value is below the significance level 0.05, the performance difference between two compared methods is considered to be statistically significant. Results in Table 6.10 reflect that encoding MADs with Fisher vector significantly improves

the accuracy of MADs for all neighborhood sizes (p) of both datasets.

Table 6.10: Results of McNemar’s test with Yates’ correction to compare significance of using Fisher vector encoding of MADs over using MADs

Dataset	p	N_{ss}	N_{ff}	N_{sf}	N_{fs}	χ^2	p -value
HCP	10	586	27	56	31	6.621	0.01008
	20	626	25	37	12	11.76	0.00061
	30	638	20	29	13	5.36	0.02064
	40	647	20	23	10	4.36	0.03671
EMR	10	1724	437	350	219	29.70	0.00001
	20	1740	452	337	201	33.88	0.00001
	30	1712	478	337	203	32.76	0.00001
	40	1733	461	328	208	26.42	0.00001

Next, we examine the success of encoding for each cognitive task. We observe in Figure 6.7 that the accuracy value obtained for Gambling task is lower compared to the other tasks performed using FV-MAD and VLAD-MAD. However, accuracy values obtained from BoW-MAD for all classes are low, meaning that BoW is not successful to encode MADs. When we compare the results shown in Figure 6.7a, Figure 6.7b and Figure 6.7c, we observe that we can boost the accuracy for Emotion, Relational processing and WM tasks compared to concatenated MADs by encoding MADs with FV and VLAD. Yet, the accuracy is decreased for Gambling task, especially for the lower values of p .

6.2.2 Analysis of Codewords

Recall that, estimation of a Gaussian Mixture Model (GMM) for mesh arc descriptors provides a set of codewords corresponding to the center values of the individual Gaussians in the mixture. In this subsection, we explore the contribution of a particular codeword (component of GMM) to classify the cognitive tasks by measuring the energy (Euclidean norm) of the corresponding columns of FVs. Recall that, the columns of the k^{th} Gaussian in a FV encoding is $[\mathcal{G}_{\mu_K}^A, \mathcal{G}_{\sigma_K}^A]$ for a set A of MADs. If

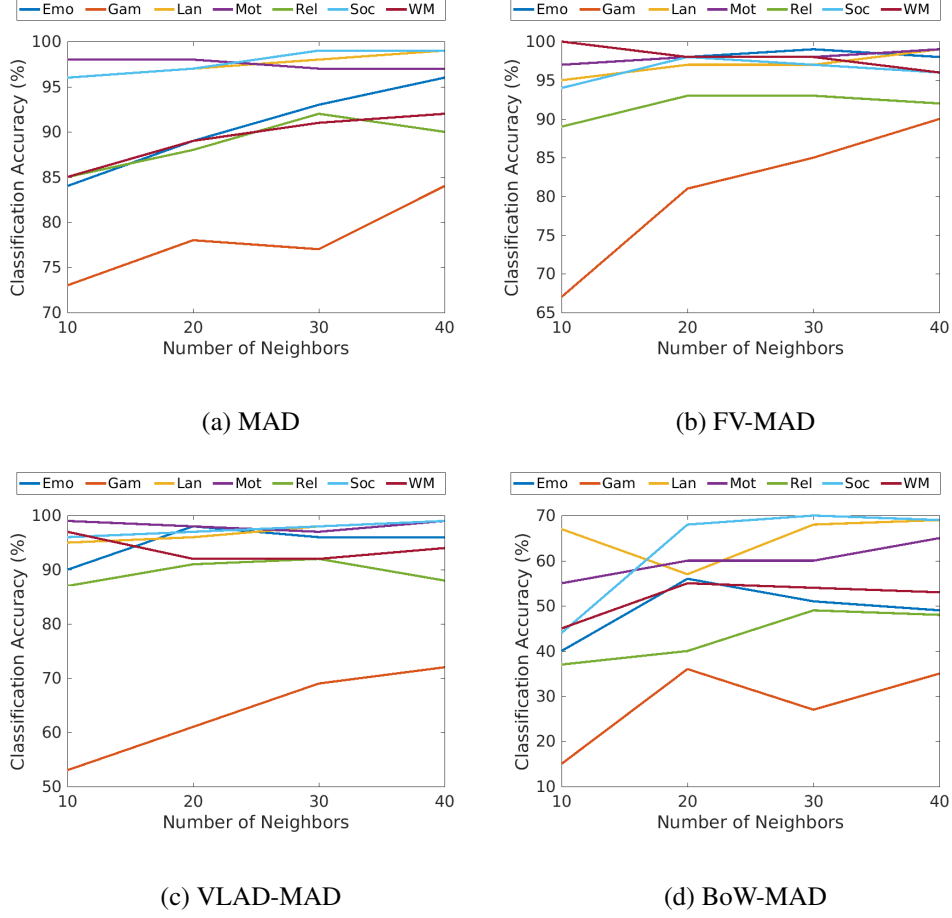


Figure 6.7: Classification accuracy (%) measured for each task of HCP dataset.

we denote the corresponding columns of the k^{th} Gaussian in the FV encoding using all training data by $\mathcal{G}_k = [\mathcal{G}_{\mu_K}, \mathcal{G}_{\sigma_K}]$, then the energy is measured by the Euclidean norm $\|\mathcal{G}_k\|_2$. We denote a Gaussian whose FV columns have the lowest energy by GLE, and the one whose FV columns have the highest energy by GHE.

Next, we visualize N -dimensional codewords obtained from MADs without using PCA or ICA. Each codeword corresponds to a mesh pattern formed around a seed node, particularly a seed region. Large values of codeword elements represent that, the MAD formed between the region corresponding to the codeword element and its seed region has large value. Similarly, small values of codeword elements indicate small values of MADs between seed region and the corresponding region. As an example, we plot codewords for $p = 40$ (we form local meshes around seed nodes with their 40 nearest neighbors) and $k = 20$ (the number of Gaussians in the mixture equals

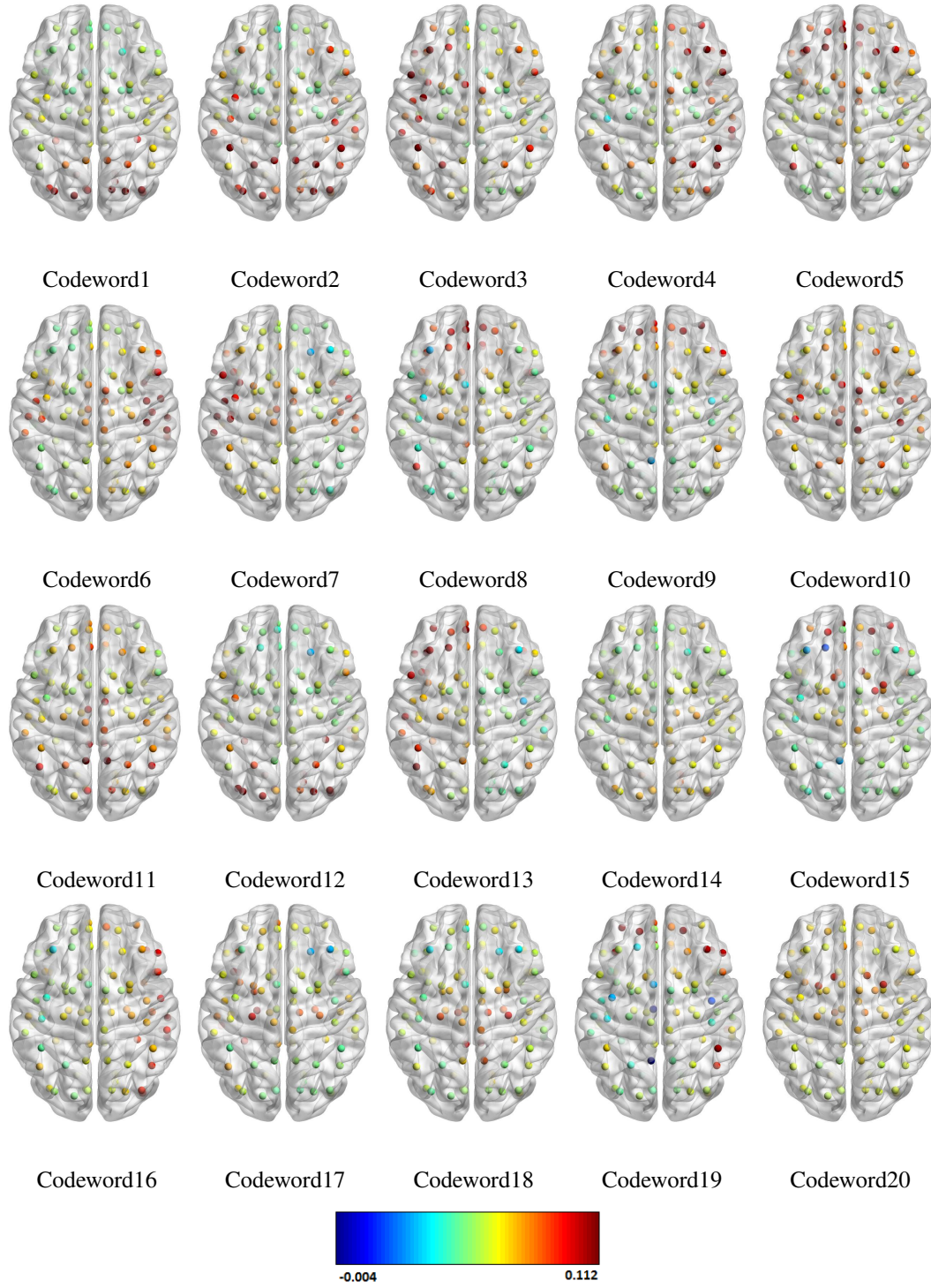


Figure 6.8: Codewords of brain connectivity dictionary obtained by FVs for encoding of MADs with $p = 40$. Codewords are visualized using BrainNet Viewer [155], and depicted in an order such that Codeword1 and Codeword20 represent the codeword of GHE, and GLE, respectively.

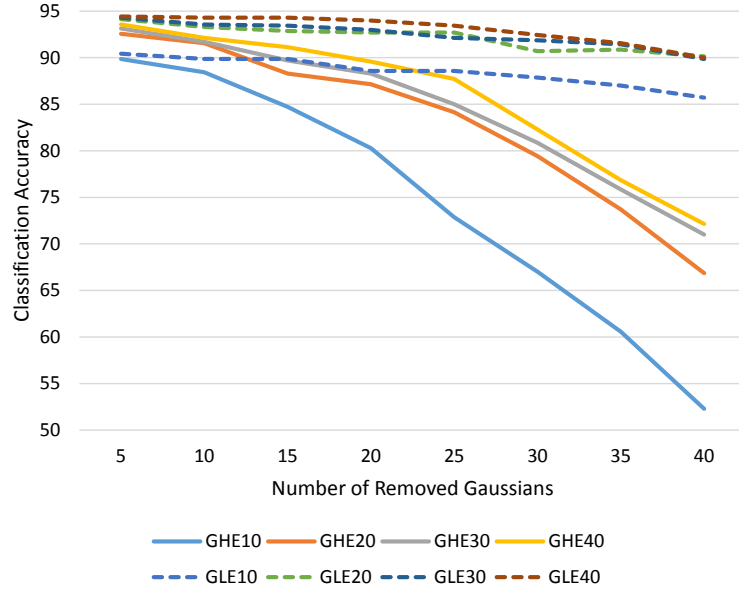


Figure 6.9: Effect of removing Gaussian mixture components from representation on classification accuracy. Results of removal of Gaussians, whose FV entries have high energy (GHE) are plotted with solid lines while results of removal of Gaussians, whose FV entries have low energy (GLE) are plotted with dashed lines. The numbers next to GHE and GLE correspond to the number of neighboring nodes in mesh (p). For example, GHE10 represents removal of Gaussians, whose FV entries have high energy such that GMMs are estimated for MADs formed with $p = 10$. It can be observed that removal of GHE decreases the accuracy much more compared to removal of GLE.

to 20), in Figure 6.8. Notice from Figure 6.8 that, codewords represent different mesh patterns. For example, Codeword1 represents a mesh pattern in which MADs formed between a seed region and regions residing in the occipital lobe have large values while MADs formed between a seed region and the other regions have smaller values. Recall that, each GMM component obtained without using PCA corresponds to $2N$ columns of FVs. In Figure 6.8, we provide visualization of the Gaussians in an order from GHE to GLE by plotting the codewords of the corresponding Gaussians. The results show that codeword of GHE has large values mainly in the occipital lobe. On the other hand, codeword of GLE has large values in central structures including Caudate, Putamen and Thalamus.

Simonyan et al. [135] showed that GHEs represent the facial features whereas the GLEs cover the background areas. In other words, they stated that GHEs are significant for comparing human face images. Following our results, we conjecture that GHEs convey task discriminative information. In order to validate this assumption, and investigate the relationship between the energy of FV columns of Gaussians and the classification accuracy, we employ an energy based feature selection scheme for FV encoding. For this purpose, we select and remove columns from FVs, and perform classification. When we remove FV columns with minimum and maximum energy, we observe that the accuracy decreases slightly and significantly, respectively (see Figure 6.9). We conclude that, GHEs provide more task discrimination power to classifiers compared to GLEs.

6.3 Experiments on Performances of Multi-resolution Mesh Networks for Cognitive State Classification

Recall that, in Section 5.1 we propose a framework called, hierarchical multi-resolution mesh networks (HMMNS), in which we fuse the decisions of classifiers trained with multi-resolution mesh networks under a hierarchical ensembling architecture.

In this section, first, we test the representation power of HMMNs for cognitive state classification, in different subbands. Then, we test the boosting effect of fusion of all subbands. After that, we investigate the significance of meta-layer and base-layer features. Finally, we explore the diversity of base-layer classifiers. Since visual object recognition and emotional memory retrieval datasets do not contain enough measurements collected during stimulus presentation for multi-resolution analysis, in this subsection we provide classification results of only HCP dataset.

6.3.1 Performances of Cognitive State Classification

In order to decode the cognitive tasks of HCP dataset, first, we decompose the original fMRI signals into $L = 6$ levels, and obtain $2L + 1$ subbands to compute multi-resolution representations of fMRI data. Next, for each subband, we estimate the mesh arc descriptors, at each resolution to obtain a total of $2L+1=13$ coarse-to-fine

brain networks for each task.

In order to optimize the mesh size p and the regularization parameter λ , we perform cross validation on the mesh arc descriptors obtained from the original signal. We searched for the optimal parameters for $p \in \{10, 20, 30, 40\}$ and $\lambda \in \{32, 64, 128, 256, 512\}$. We optimize the parameters λ and p for each band separately using cross validation. After we estimate these parameters, we use classifiers trained with the corresponding mesh arc descriptors in the ensembles. We perform between-subject classification such that, we train classifiers using optimal parameters with data collected from 707 subjects and test classifiers with data collected from 101 subjects at each of the 8 folds.

For comparison, we also compute the connectivity graphs where arc descriptors correspond to Pearson correlation coefficients computed between pairs of nodes (anatomical regions). Finally, as a baseline, we train and test the classifiers using vectors of representative time series, which are computed for each node (anatomic region) and for each subband. In summary, we train logistic regression classifiers by three feature sets, namely, mesh arc descriptors (MAD), pairwise correlations (PC) and vector of representative time series of regions (RTS).

Recall that in Section 5.4.2, we fuse the posterior probabilities obtained at the output of the classifiers to form the decision space by concatenating them under the same vector space. The vectors in the decision space are then used as the input to the meta-layer of FSG, where logistic regression classifiers are trained and tested. For comparison, we also fuse decisions of classifiers using majority voting (**MV**) and weighted majority voting (**WMV**) approaches. In majority voting method, all of the classifiers have equal confidence, whereas in weighted majority voting, the classification performances of each classifier obtained in training step are used as descriptors.

Table 6.11 shows classification accuracy obtained for features extracted from HCP dataset using approximation and detail parts of various levels of decomposition. Note that A_0 corresponds to the original fMRI signal. It is observed that feature vectors formed by mesh arc descriptors, suggested in this thesis perform substantially better compared to Pearson's pairwise correlation and the representative time series of regions for each subband. We also observe that, when mesh arc descriptors and pair-

Table 6.11: Classification performances (%) of approximation and detail parts of decomposition levels $l \in [0, 6]$ at the output of base-layer classifiers of FSG, trained and tested by three sets of feature vectors, namely, mesh arc descriptors (MAD), pairwise correlation (PC) and representative time series (RTS). The performances are obtained on HCP dataset.

Level	MAD		PC		RTS	
	A_l	D_l	A_l	D_l	A_l	D_l
0	97.15	-	89.97	-	16.11	-
1	97.35	47.38	90.06	19.78	17.93	14.25
2	97.49	63.60	90.19	37.62	27.67	16.80
3	97.49	87.69	89.29	74.08	37.16	16.39
4	96.98	93.79	86.90	82.25	40.45	29.03
5	94.98	95.24	78.06	87.94	43.28	48.20
6	81.45	93.41	49.88	86.81	44.38	51.11

wise correlations are used, the best classification performances of approximation part is obtained at levels 2 and 3, whereas the best performance of the detailed part is observed at level 5. On the other hand, representative time series, although performs substantially low compared to mesh arc descriptors, give the best classification performances at level 6. It is interesting to note that for all of the features in $l \in [5, 6]$, the information content obtained in the detail part (high resolution part) is more discriminative compared to approximation part.

Table 6.12 shows the performance of fusion of the ensemble of classifiers using mesh arc descriptors, pairwise correlation and region time series as the input of logistic regression classifiers, at the base layer. The classifier ensemble is formed by fusing the subbands from 2 to 2L base-layer classifiers. In addition to FSG, two popular ensemble learning algorithms, namely, majority voting (MV) and weighted majority voting (WMV) are used for comparison purpose. Among the feature sets, mesh arc descriptors give substantially higher brain decoding performances at the output of all of the ensemble learning methods. We observe that the performances increase as the number of subbands increases when mesh arc descriptors (**HMMNs**), pairwise cor-

Table 6.12: Classification performances (%) of fuzzy stacked generalization (FSG), majority voting (MV) and weighted majority voting (WMV) methods, obtained using mesh arc descriptors, pairwise correlation (PC) and representative time series (RTS). Numbers with + or - sign show the standard deviation of the performances obtained during the cross validation.

Level	Mesh-FSG (HMMNs)	Mesh-MV	Mesh-WMV
2	97.9 $\pm$ 0.8	90.7 $\pm$ 0.8	97.5 $\pm$ 0.8
3	98.2 $\pm$ 0.6	96.9 $\pm$ 0.8	97.7 $\pm$ 0.7
4	98.7 $\pm$ 0.6	97.9 $\pm$ 0.8	98.0 $\pm$ 0.8
5	98.9 $\pm$ 0.7	98.4 $\pm$ 0.7	98.3 $\pm$ 0.7
6	99.0 $\pm$ 0.6	98.3 $\pm$ 0.7	98.3 $\pm$ 0.7
Level	PC-FSG	PC-MV	PC-WMV
2	90.6 $\pm$ 2.5	85.3 $\pm$ 2.4	90.4 $\pm$ 2.1
3	92.3 $\pm$ 2.0	90.7 $\pm$ 2.1	92.5 $\pm$ 1.9
4	93.5 $\pm$ 1.5	91.5 $\pm$ 1.9	91.1 $\pm$ 1.9
5	94.9 $\pm$ 1.3	92.0 $\pm$ 1.9	91.7 $\pm$ 2.1
6	95.8 $\pm$ 0.8	92.6 $\pm$ 1.8	92.4 $\pm$ 1.9
Level	RTS-FSG	RTS-MV	RTS-WMV
2	42.3 $\pm$ 0.9	18.2 $\pm$ 0.7	20.9 $\pm$ 1.1
3	47.7 $\pm$ 1.2	18.6 $\pm$ 1.0	27.0 $\pm$ 1.2
4	61.9 $\pm$ 0.9	20.5 $\pm$ 1.0	32.5 $\pm$ 2.1
5	82.8 $\pm$ 0.9	26.7 $\pm$ 1.6	40.34 $\pm$ 1.7
6	92.1 $\pm$ 1.3	35.2 $\pm$ 1.6	49.8 $\pm$ 1.6

relation (**PC-FSG**) or representative time series of regions (**RTS-FSG**) are fed at the input of the base layer classifiers of FSG. The proposed HMMN method gives the best performance as 99%. This results show the power of the suggested HMMN model for representing and decoding the cognitive tasks from the fMRI data recorded during these tasks. Specifically, mesh network, which models the multiple local connectivities among the anatomic regions, have much better representation power compared to

pairwise correlation or representative time series. Also, fusing the network information extracted from each subband boosts the performance of the individual subbands which may indicate that the subbands, indeed, carry complementary information. Finally, the suggested FSG architecture fuses the diverse information coming from the subbands with a better performance compared to majority voting and weighted majority voting.

6.3.2 Analysis and Interpretation of Topology of Mesh Network

In order to analyze the topology of mesh network at various subbands, suggested in Section 5.3 we compute the out-degree of nodes of mesh network, strength of nodes of mesh network, betweenness centrality of nodes of mesh network and global efficiency of mesh network for different approximation and detail subbands separately.

Figure 6.10 shows the average node out-degrees on a human brain template for all subbands. In these plots, average node out-degrees are quite low and similar to each other at high decomposition levels of approximation parts. Note that, for subbands $A_1 - A_4$, regions mainly in frontal lobe, especially in supplementary motor area, precentral regions, bilateral superior frontal regions and bilateral medial frontal regions (see Figure 6.11) have large out-degrees. We, also, observe large out-degrees in precuneus. On the contrary, olfactory gyrus, and bilateral pallidum and rectus regions have relatively smaller out-degrees. In lower subbands, we observe approximately same number of node out-degrees for all regions.

Similarly, we observe large out-degrees mainly in frontal lobe for the detail parts at subbands $D_1 - D_4$. Also, superior frontal gyri, medial frontal gyri precentral and precuneus regions have large out-degrees (see Figure 6.12). Nodes with small out-degrees for these subbands correspond to bilateral olfactory, pallidum, rectus regions together with amygdala and bilateral medial temporal poles. Comparison of Figure 6.11 and Figure 6.12 reveals that the behavior of node degree distribution, specifically, the location of peaks are quite similar for the approximation and detail parts. This fact indicates that the node degrees of brain regions are represented in both approximation and detail parts, consistently.

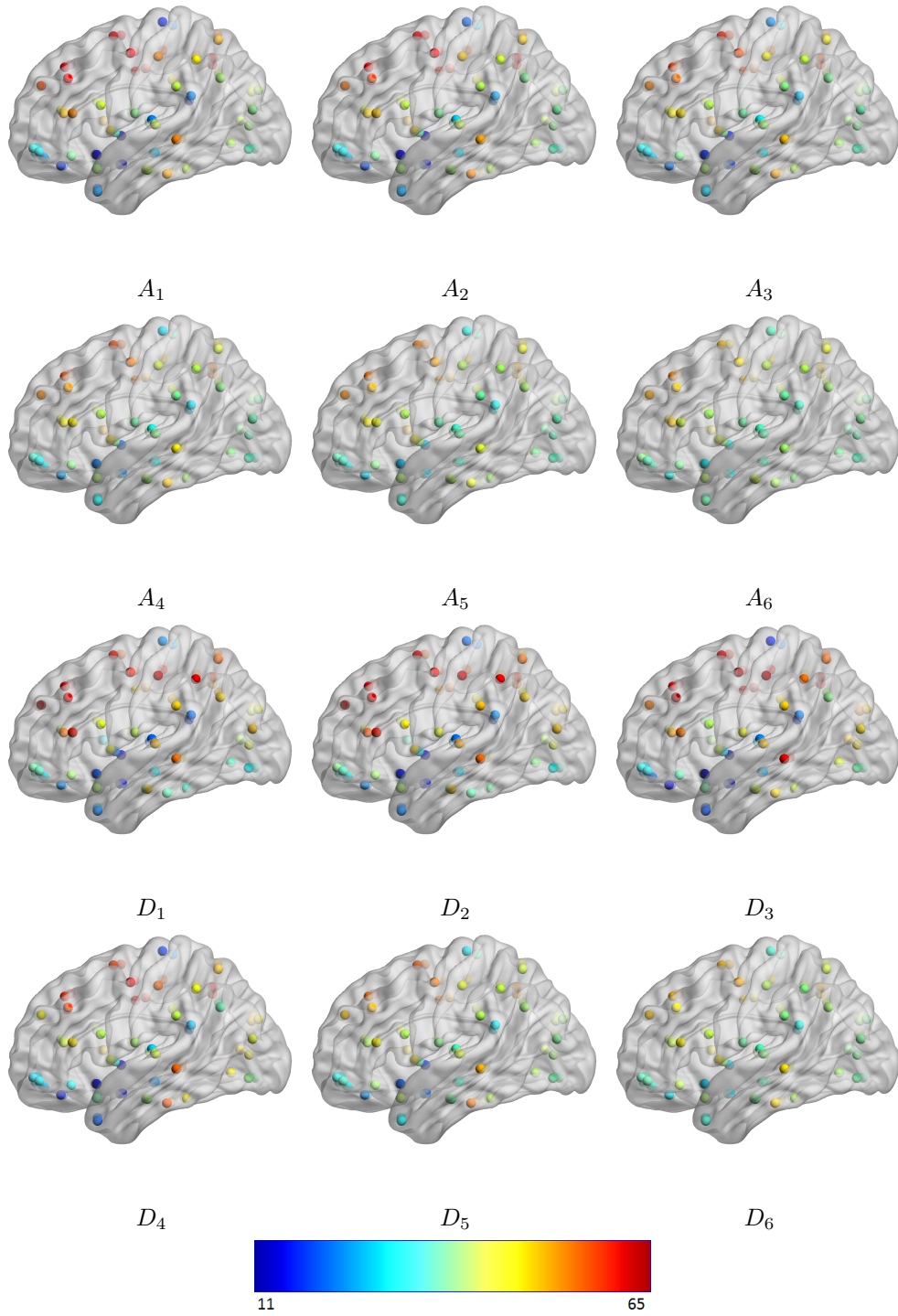


Figure 6.10: Overlay of the average out-degrees of nodes of mesh networks formed for HCP dataset on human brain templates for approximation ($[A_1 - A_6]$) and detail ($[D_1 - D_6]$) subbands. Darker blue colors denote smaller out-degrees while darker red colors denote larger out-degrees.

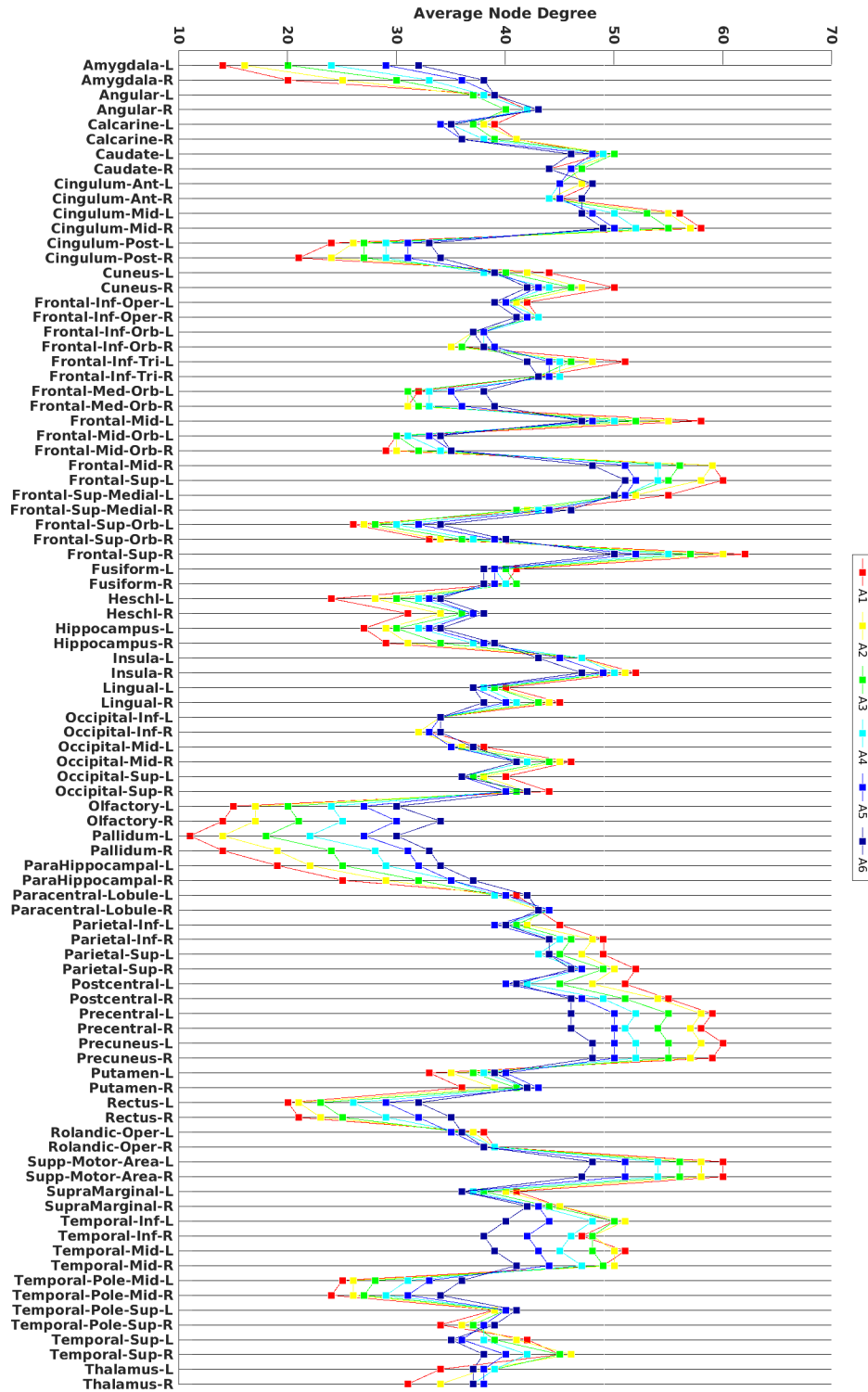


Figure 6.11: Average out-degrees of nodes (anatomical regions) of mesh networks formed at subbands $[A_1, A_6]$. Maximum values show the massively connected anatomic regions whereas the minimum values show the rarely connected regions.

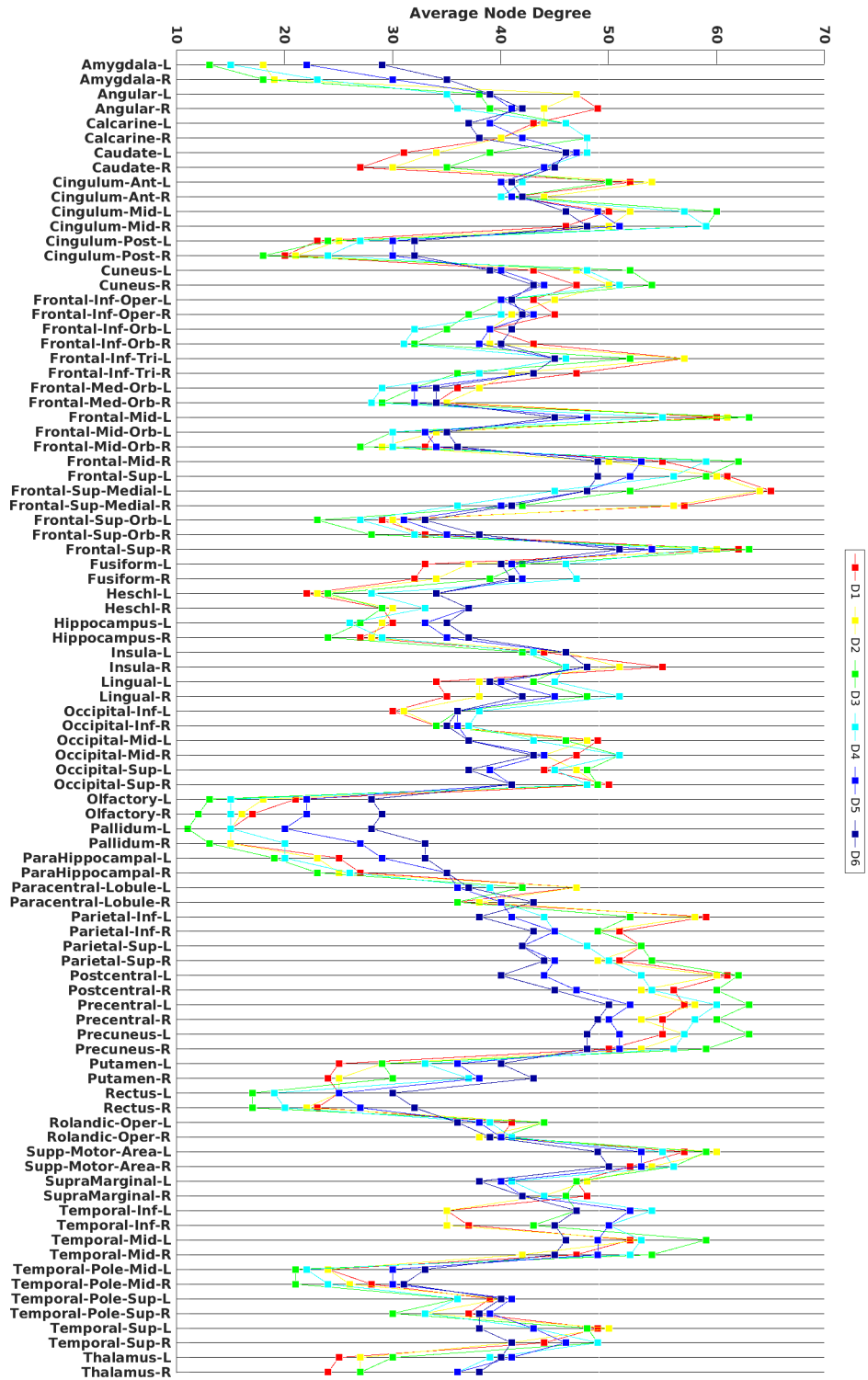


Figure 6.12: Average out-degrees of nodes (anatomical regions) of mesh networks formed at subbands $[D_1, D_6]$. Maximum values show the massively connected anatomic regions whereas the minimum values show the rarely connected regions.

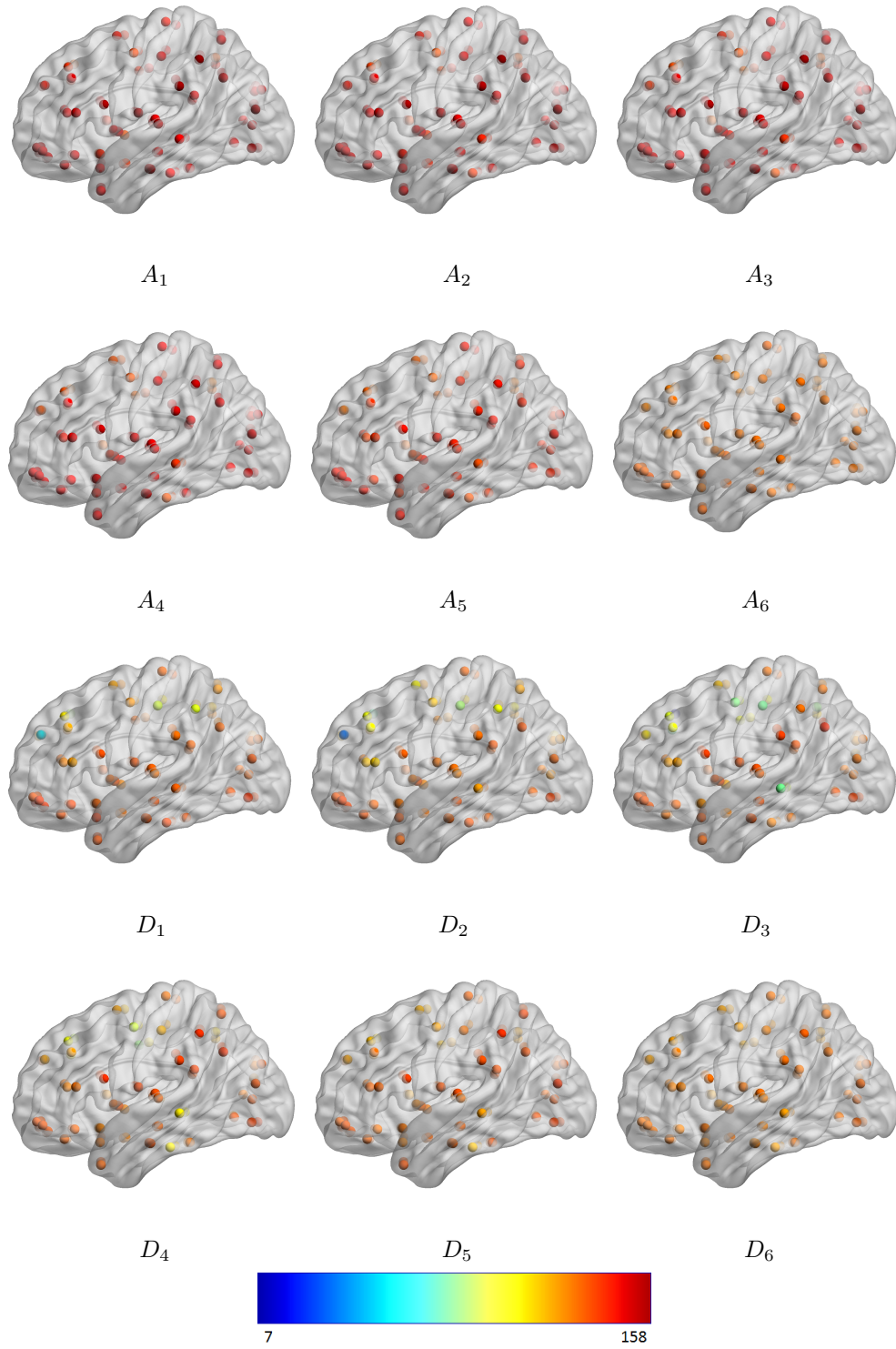


Figure 6.13: Overlay of the average out-strengths of nodes of mesh networks formed for HCP dataset on human brain templates for approximation ($[A_1 - A_6]$) and detail ($[D_1 - D_6]$) subbands. Darker blue colors denote smaller out-strengths while darker red colors denote larger out-strengths.

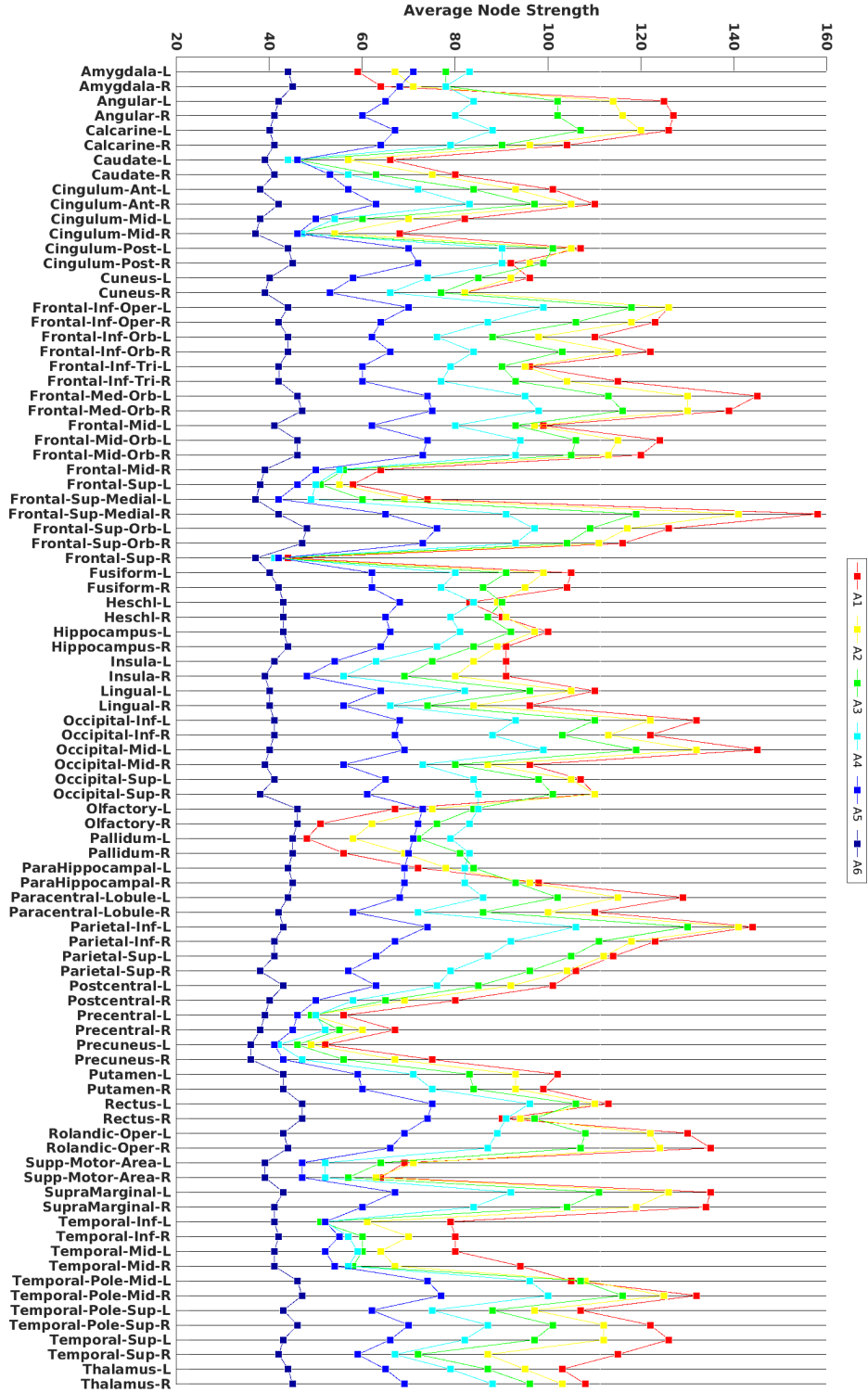


Figure 6.14: Average strength of nodes (anatomical regions) of mesh networks formed at subbands $[A_1, A_6]$. Note that the strength and variations among the anatomical regions decrease as the level increases.

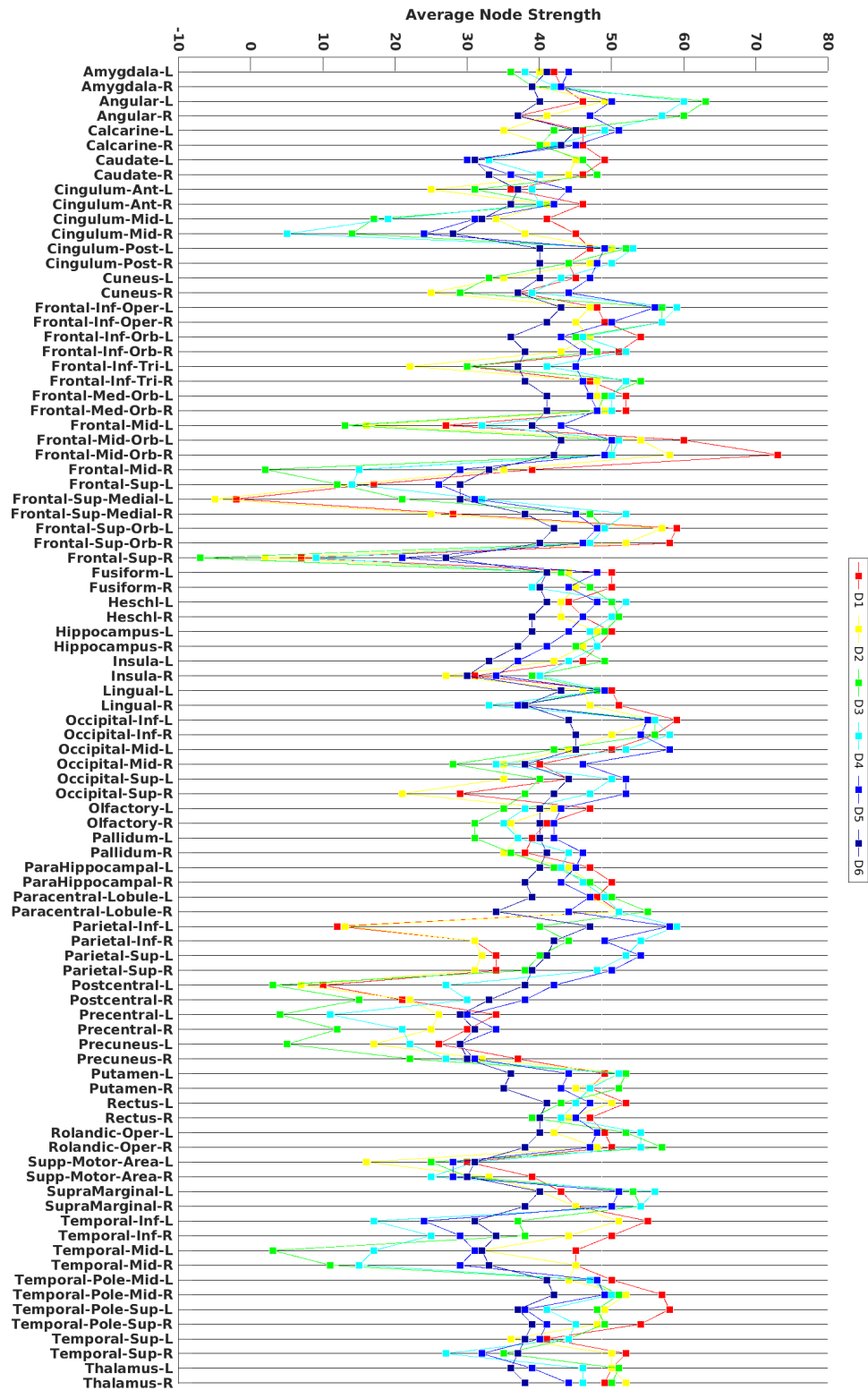


Figure 6.15: Average strength of nodes (anatomical regions) of mesh networks formed at subbands $[D_1, D_6]$.

Figure 6.13 represents node out-strengths for each subband. It can be observed that as the decomposition level increases, the values of node out-strength decreases and approaches to each other within the subband. This behavior is observed in both approximation and detail parts of fMRI signal.

In order to further analyze the strength values of anatomical regions, we plot them for approximation (Figure 6.14) and detail (Figure 6.15) parts separately. Regions with maximum node strength in subbands $[A_1, A_5]$ are right medial frontal, inferior and medial occipital (left) and bilateral middle frontal orbital gyri while regions with minimum node-strength are bilateral superior frontal gyri, precuneus (left), middle frontal (left). Although there is a great resemblance at the location of the peaks, relative node out-strength values of regions change within the subbands. For example, the highest node out-strength value in A_1 is observed at right medial frontal, inferior region; whereas the highest peak in A_4 is in left parietal inferior region. Note that for subbands $[A_6, A_{11}]$, all regions have similar out-strength values, yielding almost a flat distribution. This observation may be attributed to the negligible amount of information conveyed in subbands with higher levels than $l = 7$.

Figure 6.15 shows node out-strength behavior for detailed parts of mesh networks. In this case, regions with maximum node strength values in subbands $[D_1, D_6]$ are bilateral middle frontal orbital, bilateral superior frontal orbital and bilateral angular gyri. On the other hand, regions with minimum node strength are superior frontal medial (left), superior frontal (right), precentral (left), precuneus (left) and postcentral (left) regions. For subbands $[D_6, D_{11}]$, all regions have similar and low strength values. Note that, the detail parts distribute the node out-strength across the bands more than the approximation parts, where most of the node out-strength is observed in subbands $D_1 - D_{11}$.

We also analyze the total out-strength of mesh networks by summing out-strengths over all nodes. The results obtained for total strength in Figure 6.16 show that as the decomposition level increases, total strength of approximation part decreases monotonically. On the other hand, total strength of detail parts shows a slight increase up to level 4, and then decreases with increasing level of decomposition.

We computed betweenness centrality of each node and for each subband, separately

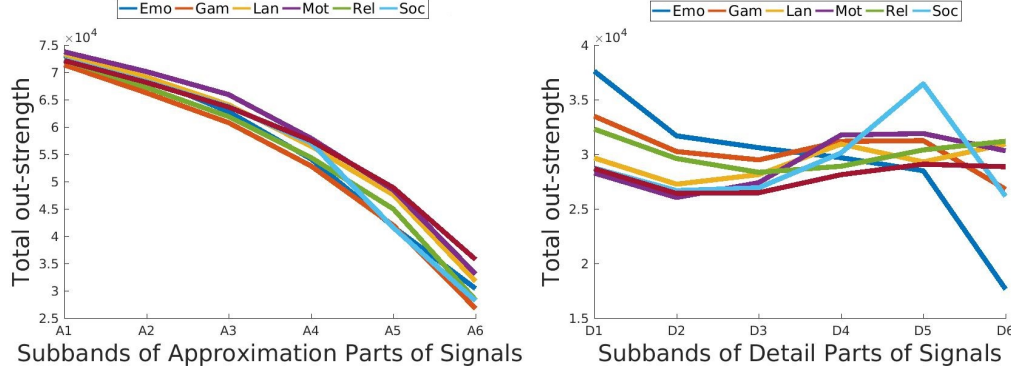


Figure 6.16: Average total strength of all mesh networks formed for each task category, over the subbands, for approximation and detail parts.

(see Figure 6.17). Results reflect that, as the decomposition level increases, betweenness centrality of nodes decreases, and approaches to zero for majority of the nodes. In other words, we have more hub candidates at lower decomposition levels. Notice that, although betweenness centrality values are large for subbands $A_1 - A_4$, we can still observe that different regions have different centrality values within a subband. In these subbands, angular gyrus and inferior occipital gyrus are the main hub candidates.

In order to analyze hub candidates for each subband, we plot node betweenness centrality values for approximation (Figure 6.18) and detail (Figure 6.19) parts. For subbands $[A_1, A_5]$ hub candidates are bilateral middle temporal pole, inferior occipital (right), angular (right), Heschl gyri and hippocampus (see Figure 6.18). For subbands $[A_6, A_8]$, we have similar hubs including amygdala (left), bilateral middle frontal orbital, bilateral rectus. For $[A_9, A_{11}]$ all nodes have similar centrality values.

On the other hand, hub candidates for subbands $[D_1, D_2]$ are bilateral parahippocampal gyri, bilateral frontal medial orbital, while for subbands with larger classification accuracy ($[D_3, D_5]$), hub candidates are middle temporal pole (left), bilateral frontal medial orbital, bilateral angular. For larger levels of decomposition, all regions have similar centrality values.

We analyze the global efficiency of all graphs formed for each task in a subband. Note that, we used only the positive arc descriptors in our graphs to find global efficiency.

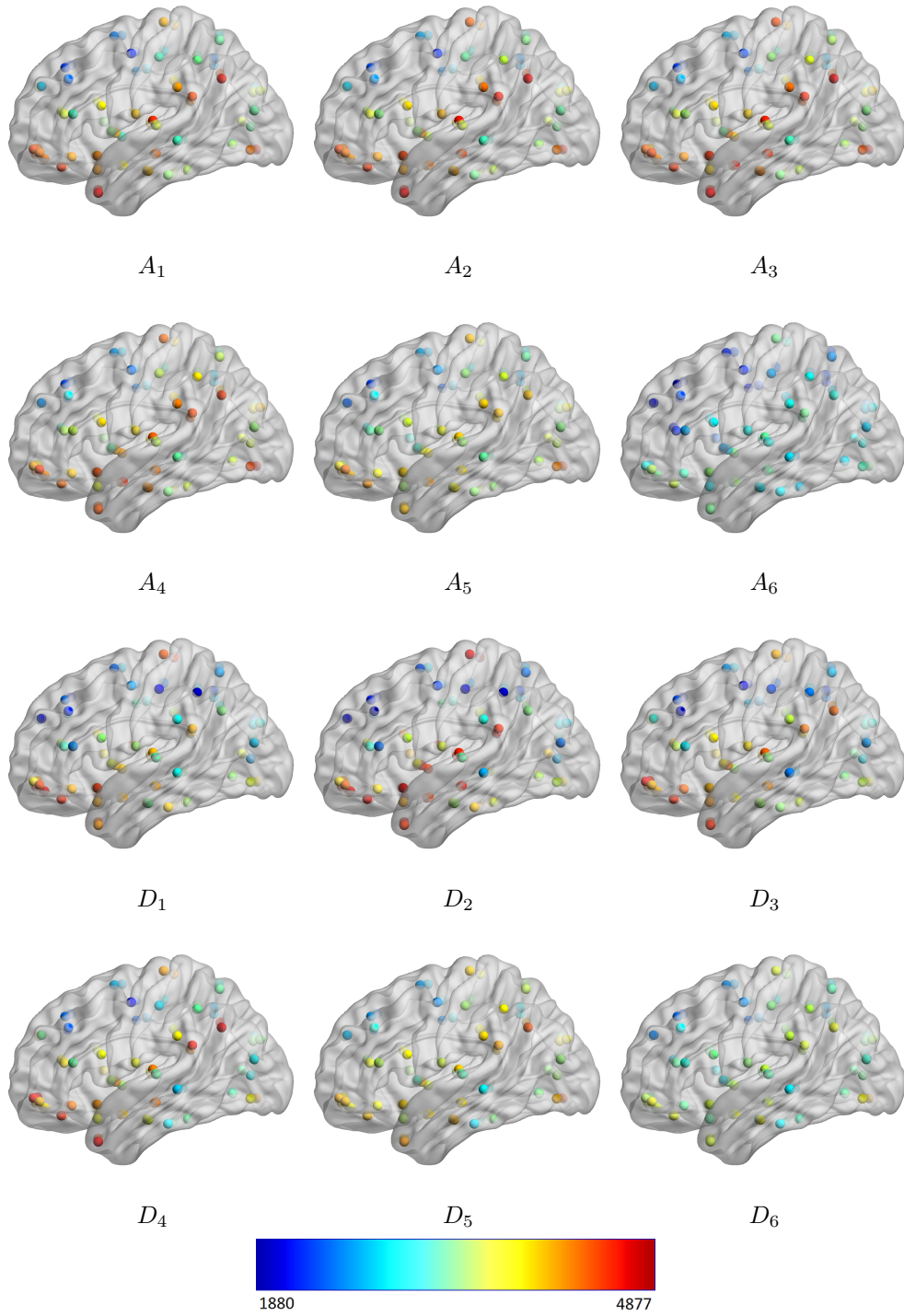


Figure 6.17: Overlay of the average betweenness centrality of nodes of mesh networks formed for HCP dataset on human brain templates for approximation ($[A_1 - A_6]$) and detail ($[D_1 - D_6]$) subbands. Darker blue colors denote smaller betweenness centrality while darker red colors denote larger betweenness centrality.

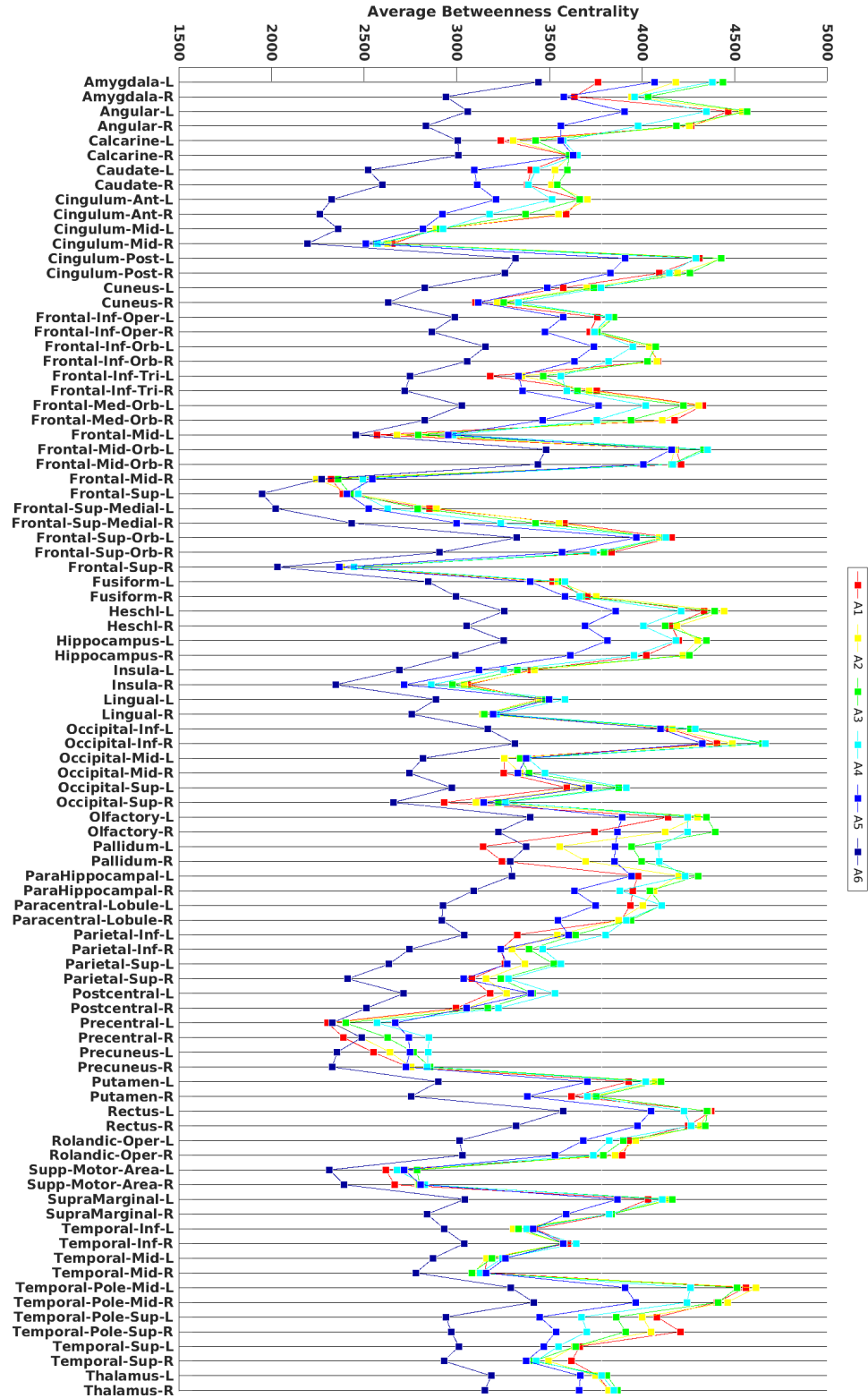


Figure 6.18: Average betweenness centrality of nodes (anatomical regions) of mesh networks formed at subbands $[A_1, A_6]$.

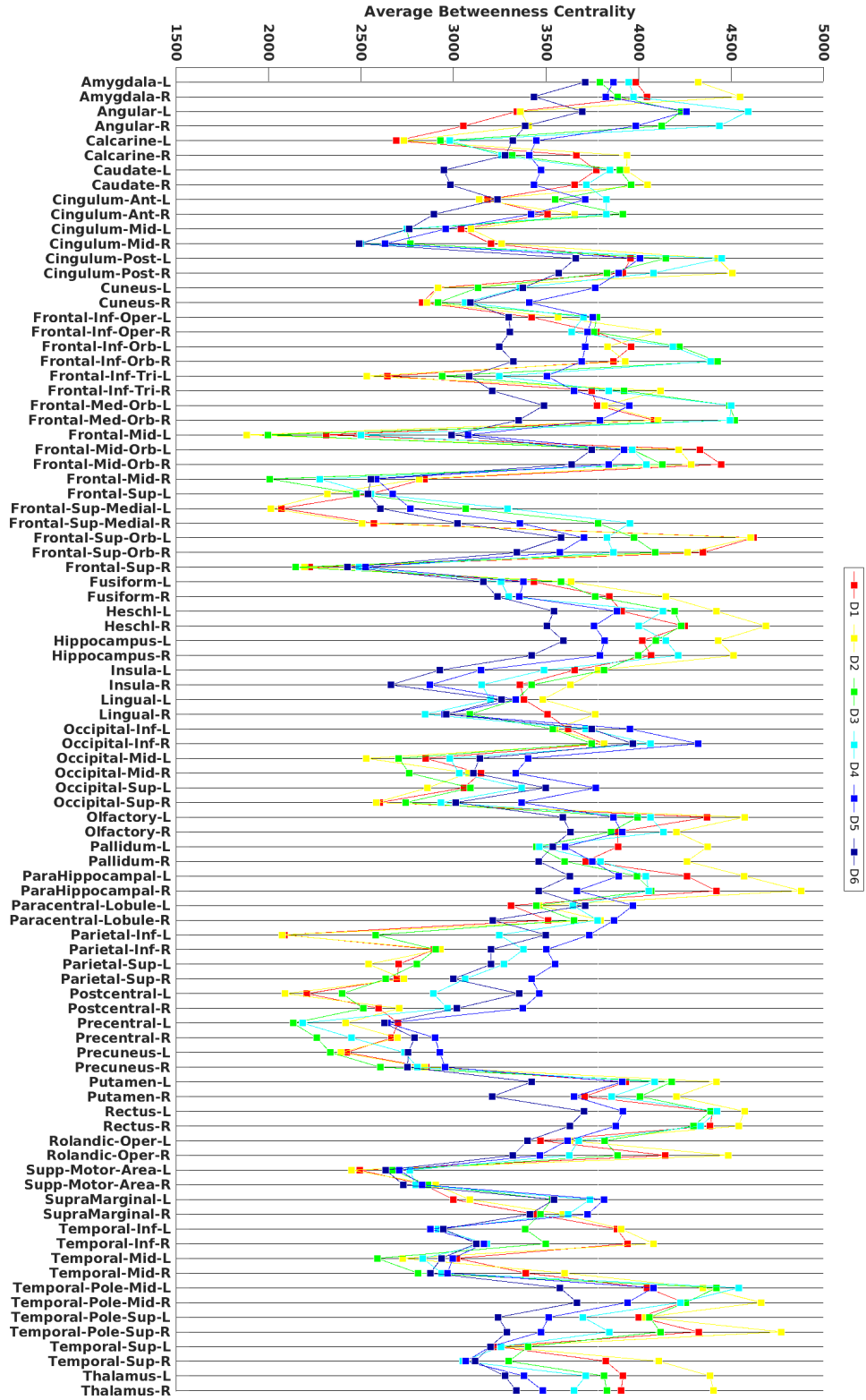


Figure 6.19: Average betweenness centrality of nodes (anatomical regions) of mesh networks formed at subbands $[D_1, D_6]$.

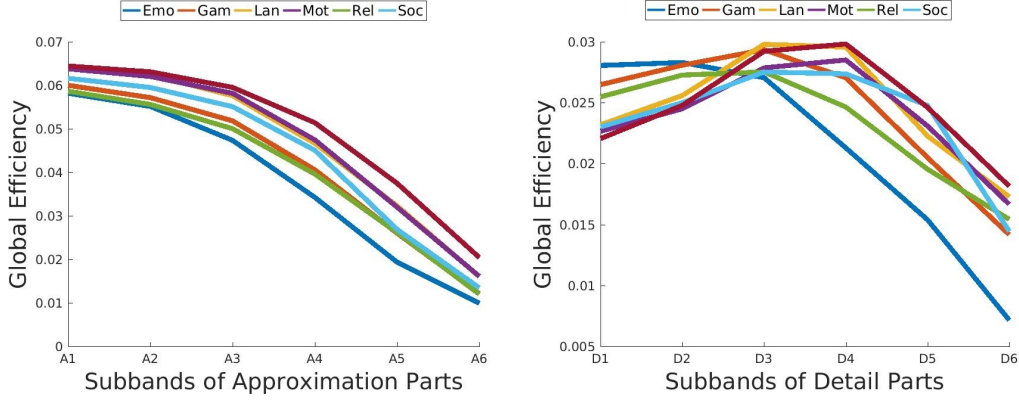


Figure 6.20: Average global efficiency of all mesh networks formed for each task category, over the subbands, for approximation and detail parts.

Our results in Figure 6.20 show that, as the decomposition level increases, global efficiency monotonically decreases in the approximation parts of data. We observe a similar behavior for the detail parts. However, global efficiency slightly increases in subbands $[D_1, D_3]$ then start to decrease monotonically. Recall that, efficiency is related to the small-world property of the networks. Based on our results, we can say that our graphs formed at lower decomposition levels are more likely to exhibit small-world property than the ones obtained at higher levels.

6.3.3 Significance Analysis of Meta-layer Features

Recall that, in Section 5.4.3 we estimate the significance of class membership values using absolute value two-sample t-test with pooled variance estimate and assign z -values to MADs. In this subsection, we investigate the significance of subbands with respect to the cognitive tasks by analyzing the significance of class membership values. Figure 6.21 denotes the significance of meta-layer features (class membership values) to classify the cognitive tasks in terms of maximum z -values among all the task categories obtained at each subband. High z -values indicate that the corresponding class memberships have high discriminative power for related tasks in a given subband, whereas the low z -values correspond to less significant class memberships. The general trend of approximation parts in Figure 6.21 shows relatively high

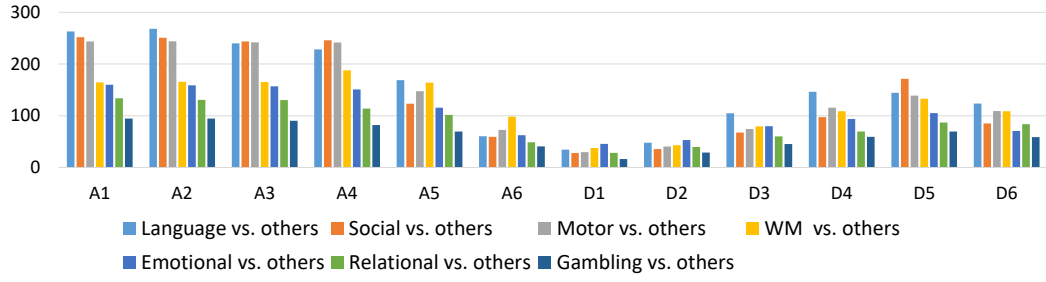


Figure 6.21: Significance of each task at each subband. Maximum z values of class memberships obtained using absolute value two-sample t-test with pooled variance estimate for subbands $[A_1, A_6]$ and $[D_1, D_6]$. z values are computed under one versus all condition. The bars show the maximum of z -values obtained from the membership values during one-versus-all task classification for each task category.

z -values in subbands $[A_1 - A_4]$, which means that class memberships obtained for these subbands are powerful to discriminate all the categories.

Z -values for task category *Language*, *Motor* and *Social* are relatively high (above 200) compared to those of the *Emotion*, *Gambling*, *Relational* and *Working Memory* categories, in subbands $[A_1 - A_4]$.

Behavior of detail parts in Figure 6.21 are rather different compared to approximation parts. z -values for task categories *Emotion*, *Motor*, *Relational* and *Social* task categories reach to maximum z -values at D_5 and then decrease at D_6 . Task categories *Language* and *Social* have the highest z -values for detail parts at D_4 and D_5 , respectively. Subbands D_1 and D_2 have very low z -values compared to the rest of the subbands, which may be an indication of the noise embedded in the original signal.

In summary, variations of z -values with respect to the subbands, show the discriminative power of each subband for a specific task. This observation indicates that the tasks are best represented in different subband combinations.

6.3.4 Significance Analysis of Base-layer Features

Recall that, features of HMMN to train the base-layer classifiers are mesh arc descriptors (MADs) obtained for each task at each subbands. Figure 6.22 and Figure 6.23

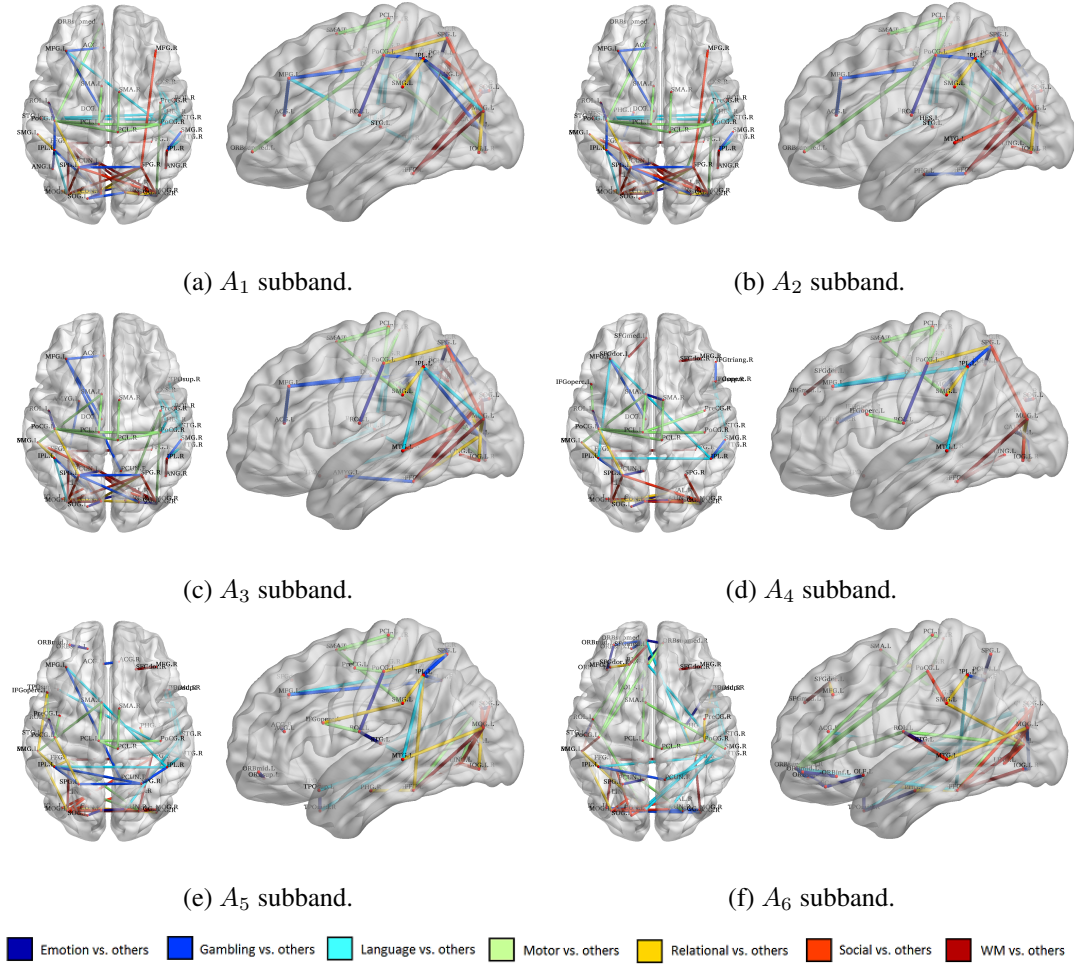


Figure 6.22: The most significant mesh arc descriptors that discriminate a specific task from others for different approximation subbands. Color of the mesh arc indicates the task that is discriminated from the others by the corresponding mesh arc descriptor.

represent the most significant mesh arc descriptors which discriminate a specific task from others for subbands $[A_1 - A_6]$ and $[D_1 - D_6]$, respectively. We observe that, the significant mesh arc descriptors are formed between distinct anatomical region pairs for different tasks within a subband.

Mesh arcs between the seed region rolandic operculum and its neighbors, namely, postcentral gyrus and superior temporal gyrus are significant to discriminate *Emotion* task, for subbands $[A_1 - A_5, D_6]$ and $[A_5, A_6]$, respectively. Moreover, the mesh arc formed between precuneus and superior parietal gyrus is discriminative for sub-

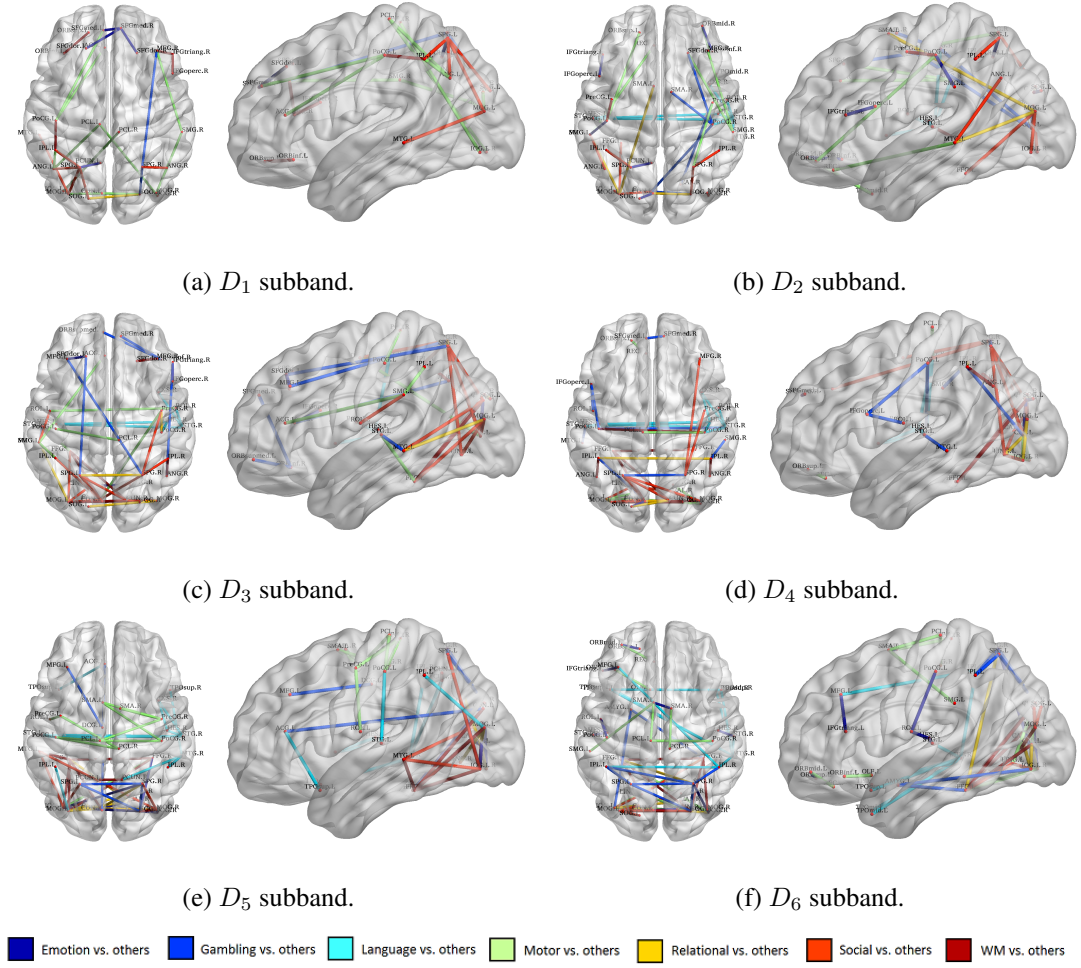


Figure 6.23: The most significant mesh arc descriptors that discriminate a specific task from others for different detail subbands. Color of the mesh arc indicates the task that is discriminated from the others by the corresponding mesh arc descriptor.

bands $[A_1 - A_4, A_6, D_1, D_2]$ while the one formed between pars triangularis and pars opercularis is significant for D_1 and D_2 .

Relationships that discriminate *Gambling* task from others are formed around middle frontal gyrus with its neighbors, namely, dorsal middle cingulum and anterior middle cingulum. The former is significant for subbands $[A_1 - A_4]$ while the latter is significant for subbands $[A_1 - A_3]$. In addition, interhemispheric relation between bilateral superior parietal gyri carry discriminative information for subbands $[A_1 - A_3, A_5, D_4 - D_6]$. The mesh arc between superior parietal gyrus and superior occipital gyrus is also significant for subbands A_2, A_3, D_5 and D_6 .

We observe that the significant mesh arcs for *Language Processing* task are mainly formed around right heschl gyrus with its neighbors, namely, right postcentral gyrus, bilateral superior temporal gyri and right insula regions, for subbands $[A_1 - A_3]$ and $D_4 - D_5$. Moreover, interhemispheric relations between left and right heschl gyri and left and right superior temporal gyri are observed for A_1, A_2 and $[D_2 - D_4]$.

We also identified significant mesh arcs formed around the seed region inferior parietal lobule between medial temporal gyrus and medial occipital gyrus, for subbands $[A_3 - A_5]$ and $[A_1 - A_3, D_5]$, respectively. For D_1 , there is no significant mesh arc descriptor for *Language processing* task. We can conclude that, relationships of regions heschl gyrus, superior temporal gyrus and inferior parietal lobule with their neighbors are significant for *Language processing*.

Mesh arcs, formed between seed region paracentral lobule and its neighbors, namely, postcentral gyrus and supplementary motor area carry significant information for *Motor* tasks. While the former is significant for subbands $[A_1 - A_6]$ and $[D_3 - D_6]$, the latter is significant for $[A_1 - A_5]$ and $[D_4 - D_6]$. Moreover, mesh arc descriptor formed between supramarginal gyrus and supplementary motor area carry discriminative information in subbands $[A_3 - A_5]$ and D_6 . Relationship between postcentral and frontal medial orbital is discriminative for subbands A_1, A_2 and A_6 . For subband D_1 , the significant mesh arc descriptors are formed around seed region paracentral lobule with inferior occipital gyrus and inferior temporal gyrus. For subband D_2 , we obtain a significant mesh descriptor formed between frontal medial orbital and postcentral gyrus. We can infer that, meshes formed around paracentral lobule, postcentral gyrus, supplementary motor area and frontal medial orbital with their neighbors represent information to discriminate *Motor* task from others.

The most significant mesh arc descriptors to discriminate *Relational Processing* task from other tasks are formed between calcarine region and its neighbors, middle occipital gyrus for subbands $[A_1 - A_5, D_5, D_6]$, fusiform gyrus for subbands $[A_1 - A_3, A_5, D_5, D_6]$ and superior parietal gyrus for subbands $[A_1 - A_3]$. Moreover, mesh arcs formed between the region pairs (middle occipital gyrus - inferior occipital gyrus), (supramarginal gyrus - inferior parietal gyrus), (superior parietal gyrus - postcentral gyrus) and (middle temporal gyrus - middle occipital gyrus) are significant for

subbands $[A_1, A_2]$, $[A_1 - A_6]$, $[A_1 - A_3]$ and $[A_5, A_6, D_2, D_3]$ respectively.

Mesh arcs between superior parietal gyrus and its neighboring regions, namely middle occipital gyrus, middle frontal gyrus represent information to discriminate *Social Processing* task, for subbands $[A_1 - A_4, D_1]$, $[A_1, A_2, D_4]$, respectively. Moreover, for subbands D_3 and D_4 , arc descriptors of mesh formed around superior parietal gyrus with its neighbors, namely, cuneus and calcarine are significant. Additionally, the mesh arc descriptor formed between middle temporal gyrus and angular gyrus is significant for subbands A_3 and D_1 . It can be observed that, relationships between superior parietal gyrus and its neighbors represent information to discriminate *Social Processing* task from others.

Mesh arcs, that discriminate *Working Memory* task from others are mainly around fusiform gyrus. Mesh arcs between seed region fusiform gyrus and neighboring regions, namely, middle occipital gyrus, superior occipital gyrus and cuneus are significant for subbands $[A_1 - A_6, D_5, D_6]$, $[A_1 - A_3, D_6]$ and $[A_1 - A_3, D_3 - D_5]$. On the other hand, the most significant mesh arc descriptor for subband D_1 is the one between superior frontal orbital and superior inferior frontal regions.

Notice that, the mesh arcs denoted in Figure 6.22 and Figure 6.23 are the ones that carry information to discriminate only a specific task from others. Although majority of the significant arc descriptors carry discriminative power for only one task, here exists a few of them which are important for multiple tasks. Among them, the mesh arc between (left inferior parietal - left angular gyrus), (precentral - postcentral), (left middle occipital - right middle occipital), (left inferior occipital - right inferior occipital) and (middle temporal gyrus - middle occipital gyrus) pairs are significant for multiple tasks for subbands $[A_1 - A_3, A_6, D_2 - D_4]$, $[A_1 - A_5, D_2, D_4, D_5]$, $[A_1 - A_5, D_5, D_6]$ and $[A_1 - A_5, D_4, D_6]$, $[A_1 - A_4, D_2]$, respectively. Our results reflect that, there does not exist a mesh arc descriptor, which is significant for all tasks and all subbands.

We can conclude that within the same subband, generally the significant mesh arc descriptors differ for each task. Moreover, the same task is represented best by different mesh arc descriptors for different subbands. Therefore, we can say that base-layer classifiers trained with mesh arc descriptors obtained at different subbands model di-

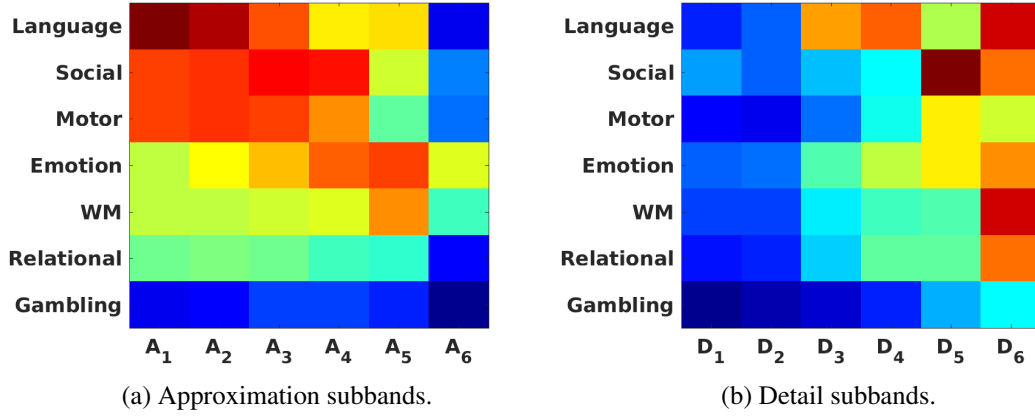


Figure 6.24: Average z -values of top-20 mesh arc descriptors that discriminate a specific task from others for different subbands.

verse information.

Recall from Section 5.4.4 that, we take the most discriminative 20 mesh arc descriptors with maximum z -values and we identify the ones that are significant to discriminate only a specific class from others in this set. The reason why we choose the best 20 MADs is to make sure that we obtain MADs (significant for only one task) for all of the 7 tasks. In order to understand the power of top-20 mesh arc descriptors to discriminate a specific task from others, we plot average z -values of top-20 mesh arc descriptors for each subband in Figure 6.24. Dark red color represents the highest discriminative power whereas dark blue color represents the lowest discriminative power.

We observe from Figure 6.24a that, top-20 mesh arc descriptors that discriminate *Language*, *Motor* and *Social Processing* tasks from others are more powerful in subbands $[A_1 - A_4]$. Moreover, significant mesh arc descriptors to discriminate *Emotion* and *WM* tasks are observed in subbands $[A_3 - A_5]$ and A_5 respectively. Top-20 mesh arc descriptors that discriminate *Relational* and *Gambling* tasks from others are much less powerful compared to the other tasks. Average z -values obtained at subband A_6 are lower compared to other approximation subbands. The representative power of mesh arc descriptors is relatively low for *Gambling* task. This fact is observed from the last rows of matrices of the approximation and detail subbands, which show the average significance of MADs to discriminate *Gambling* from others in Figure 6.24.

Figure 6.24b shows that average z -values increase as the level increases. Note that, while at subbands D_1 and D_2 the top-20 mesh arc descriptors carry the least discriminative power, they carry the most discriminative power at D_6 .

6.3.5 Analysis of Diversity of Base-layer Classifiers

Remember that, in Section 5.4.5 we employ a number of classifier diversity measures to understand whether classifiers trained with multi-resolution mesh networks represent complementing information. In this subsection, we provide the diversity values obtained with pairwise and non-pairwise measures.

Table 6.13 shows the diversity measures for the classifier groups starting from combining subbands in the first two levels, A_1 , D_1 , A_2 and D_2 and then, adding pairs of approximation and detail parts at each level. Higher values of coincident failure diversity (**CFD**), generalized diversity (**GD**) and lower values of double fault measure (**DF**), mean correlation coefficient (**Mean ρ**), measure of difficulty (**θ**) correspond to more diversity of classifiers.

In Table 6.13, the arrows ($\uparrow$) or ($\downarrow$) denote that the diversity is greater if the measure is higher or lower, respectively. Note that the diversity measures consistently show a slight decay in diversity, then it gets better as we add more levels reaching to the best diversity, when we include the classifiers trained in all subbands. This observation shows that all of the subbands carry complementary information about the cognitive tasks.

Comparison of Table 6.12 and 6.13 shows the relationship between the diversity measures and classification accuracy. Note that, performances and diversities increase proportionally, as we add classifiers to the ensemble. These results show that the information about the task categories released in each subband complements each other. Therefore, fusing the base-layer classifiers results in boosted performances.

Table 6.13: Diversity measures obtained for ensemble of classifiers trained with mesh arc descriptors. The arrow specifies whether diversity is larger if the measure is smaller ($\downarrow$) or larger ($\uparrow$).

Level	Pairwise		Non-pairwise		
	DF($\downarrow$)	Mean $\rho(\downarrow)$	CFD($\uparrow$)	GD($\uparrow$)	$\theta(\downarrow)$
2	0.06	0.25	0.83	0.74	0.05
3	0.04	0.27	0.86	0.76	0.03
4	0.03	0.27	0.88	0.78	0.03
5	0.03	0.26	0.89	0.79	0.02
6	0.03	0.24	0.89	0.79	0.02

6.4 Experiments and Results of Multi-resolution Mesh Networks for Gender Classification

In order to test the power of the suggested HMMN model, we conduct another set of experiments, in addition to cognitive task classification. In this set of experiments, we add another level that aims to fuse the information obtained from multiple tasks into our architecture and we try to identify the genders from fMRI recordings.

In this subsection, we first share the performances of gender classification experiments obtained by fusing multi-resolution multi-task mesh networks under a 2-level hierarchical architecture. Then, we analyze the significance of MADs obtained for different tasks and subbands to discriminate gender.

6.4.1 Gender Classification Experiments and Results

In this subsection, we provide gender classification performances of base layer classifiers trained with mesh networks formed for different tasks and different subbands. We investigate the results obtained by fusing decisions of *task-classifiers*, *subband-classifiers* and both. We compare the results of using mesh arc descriptors (MAD), pairwise correlations (PC) between anatomical regions and representative time series (RTS) as features to train and test base layer classifiers.

Table 6.14: Classification performances (%) obtained with classifiers trained with mesh networks formed for different cognitive tasks. We compare the classification performances obtained using mesh arc descriptors (**MAD**), pairwise correlations (**PC**) and representative time series (**RTS**)

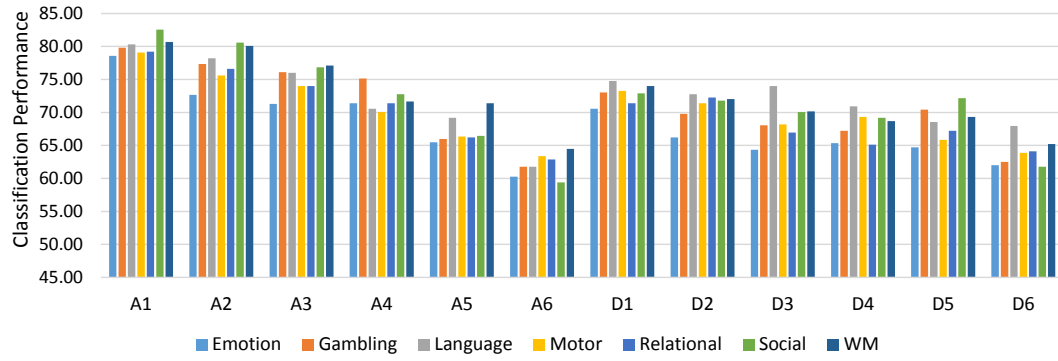
	Emotion	Gambling	Language	Motor	Relational	Social	WM
MAD	80.07	82.05	80.82	80.69	80.20	82.80	79.33
PC	74.50	74.88	77.60	73.27	74.13	75.87	71.66
RTS	84.41	82.92	84.53	83.04	84.16	83.42	83.29

6.4.1.1 Performances Obtained for Each Cognitive Task

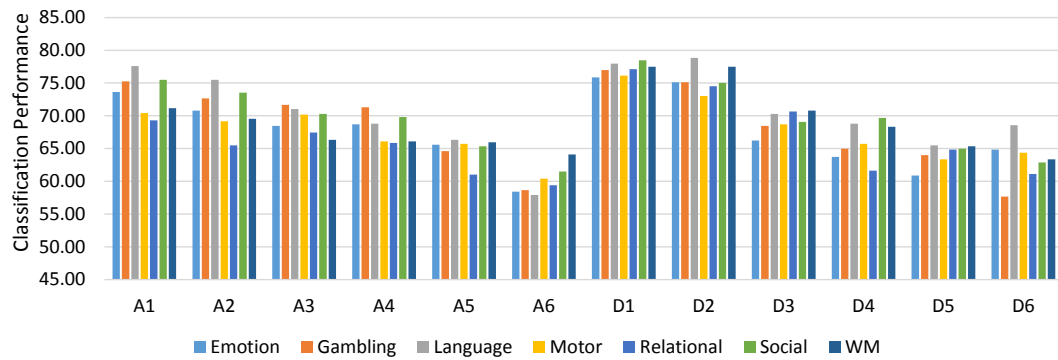
We first perform gender classification experiments for each cognitive task separately by collecting the MADs of mesh networks under feature matrices and using genders of individuals under label vectors. At this stage, we do not decompose the original signal A_0 into multiple subbands. Moreover, we do not ensemble the decisions of classifiers trained for different tasks. The results obtained for each task is presented in Table 6.14. We observe that, RTS discriminates gender better than MADs or PC between regions for all tasks. The results may be explained in such a way that, unlike classification of cognitive states, the long sequence of fMRI measurements collected under cognitive tasks carry gender discriminative information and modeling relationship among nodes lead to a slight decrease in the gender discriminative power when original fMRI signal is employed.

6.4.1.2 Performances Obtained for Each Subband of Each Cognitive Task

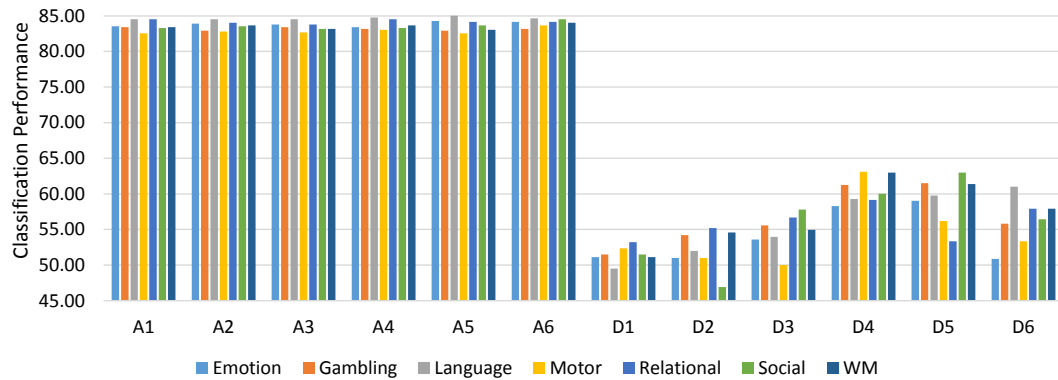
For this experiment, we present the gender classification results obtained for each subband of each task. We do not fuse decisions of classifiers and provide separate base layer classifier performances obtained for subbands $[A_1 - A_6]$ and $[D_1 - D_6]$. Results obtained for each subband of each task using mesh arc descriptors, Pearson correlation and representative time series are presented in Figure 6.25. We observe that, as the decomposition level increases, gender classification performances obtained for tasks decrease when mesh arc descriptors and Pearson correlation are used as features.



(a) Classification performances obtained using mesh arc descriptors as input to the logistic regression classifier at each subband for each cognitive task.

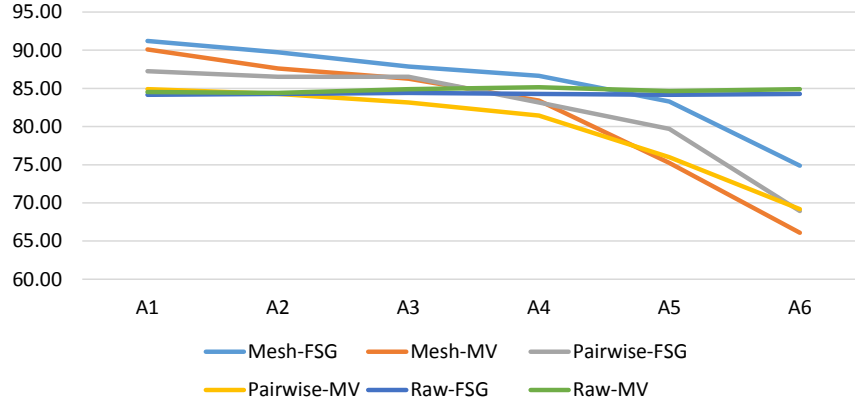


(b) Classification performances obtained using pairwise correlations (PC) as input to the logistic regression classifier at each subband for each cognitive task.

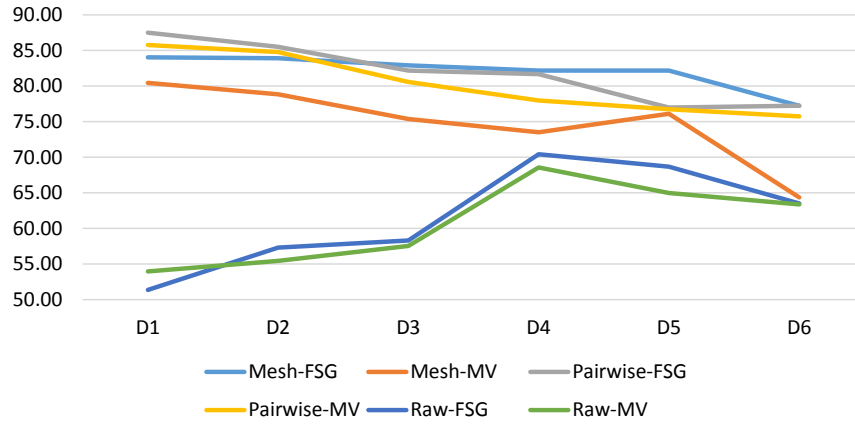


(c) Classification performances obtained using representative time series as input to the logistic regression classifier at each subband for each cognitive task.

Figure 6.25: Gender classification performances obtained for each subband and each task separately, using three different feature sets, namely, mesh arc descriptors (MAD), pairwise correlation (PC) and representative time series (RTS).



(a) Performances obtained for subbands $[A_1 - A_6]$



(b) Performances obtained for subbands $[D_1 - D_6]$

Figure 6.26: Gender classification results performances by the 1-level fusion of decisions of *task-classifiers*

On the other hand, gender classification performances obtained using representative time series for subbands $[A_1 - A_6]$ are high and do not differ significantly. This shows that all of the approximation subbands carry same amount and type of information contained in each task concerning the gender. Moreover, we observe slight changes in the classification performances obtained using representative time series for subbands $[D_1 - D_6]$, which are much lower than the performances of subbands $[A_1 - A_6]$. Notice that, without fusion of classifier decisions, maximum gender classification results are obtained with representative time series.

6.4.1.3 Performances Obtained by 1-level Fusion of Task-Classifiers

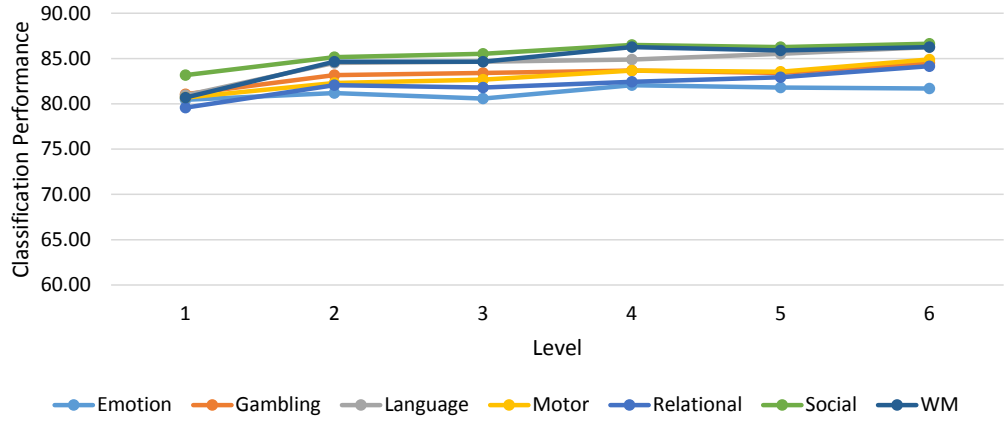
In order to observe the complementary information about gender at each cognitive task, we fuse the decisions of *task-classifiers*, which are trained with mesh networks formed while individuals perform different cognitive tasks. We expect that the classification performances are boosted when the *task-classifiers* collaborate with each other.

In this set of experiment, we fuse the decisions of classifiers trained for all tasks at each subband separately as shown in Figure 5.4. Besides computing the performances of meta layer classifiers, we also ensemble the decisions of base layer *task-classifiers* using majority voting (MV) for comparison. Figure 6.26 represents the classification performances obtained by fusing the decisions of *task-classifiers* with fuzzy-stacked-generalization (FSG) and majority voting (MV). We infer from Figure 6.26a that, fusion of *task-classifiers* at approximation subbands of small decomposition levels lead to better FSG and MV performances with mesh arc descriptors and Pearson correlation. On the other hand, fusion of *task-classifiers* with FSG and MV gives nearly the same performance (around 85%) when representative time series are used as the input of the base-layer classifiers. For subbands $A_0 - A_4$ raw data gives lower performances compared to mesh arc descriptors whereas for subbands $A_4 - A_6$ representative time series outperform mesh arc descriptors and pairwise correlation.

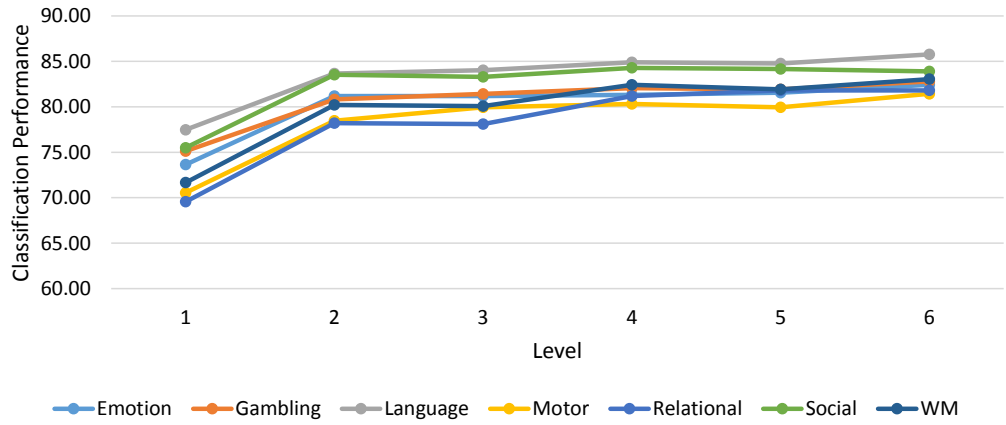
The performances obtained by the fusion of *task-classifiers* at detail subbands in Figure 6.26b reflect that as the decomposition level increases, FSG and MV performances decrease except for D_5 , when mesh arc descriptors and Pearson correlation are used as the features at the input of base-layer classifiers. Also, FSG and MV performances obtained with representative time series vary much less compared to mesh arc descriptors and Pearson correlation. Representative time series performs worse than mesh arc descriptors and Pairwise correlation for subbands $D_1 - D_6$.

6.4.1.4 Performances Obtained by 1-level Fusion of Subband-Classifiers

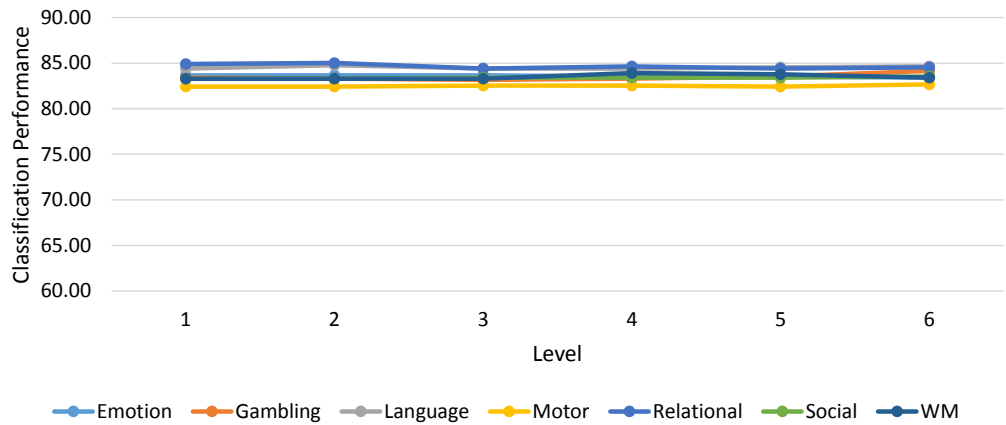
In Section 6.3, we show that fusion of decision of classifiers trained with multi-resolution mesh networks boosts the accuracy of cognitive state classification. In this



(a) Performances of FSG when the input of each base-layer classifier is the mesh arc descriptors.

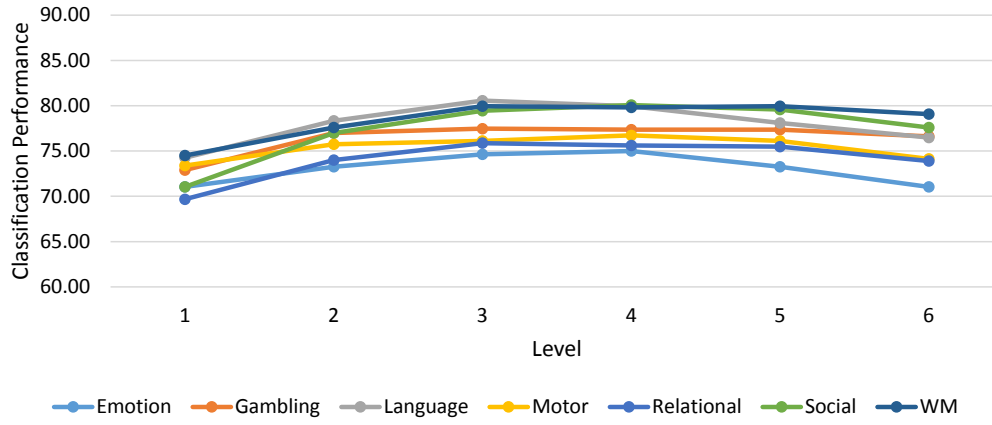


(b) Performances of FSG when the input of each base-layer classifier is the pairwise correlation

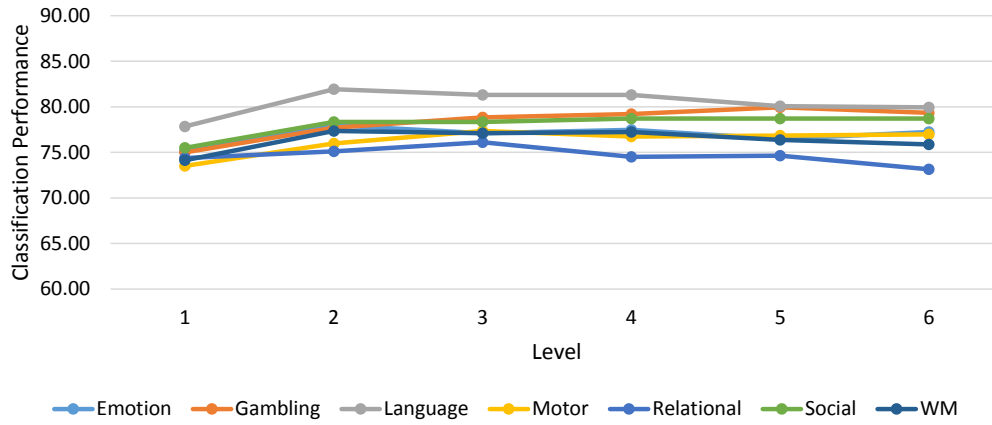


(c) Performances of FSG when the input of each base-layer classifier is the representative time series

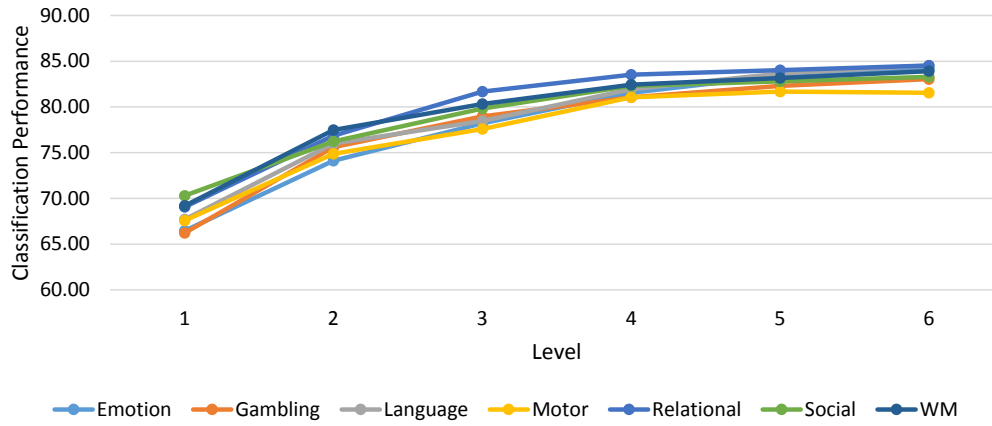
Figure 6.27: Gender classification performances obtained by the 1-level fusion of decisions of *subband-classifiers* using fuzzy stacked generalization (FSG).



(a) Performances of MV when the input of each base-layer classifier is the mesh arc descriptors.



(b) Performances of MV when the input of each base-layer classifier is the pairwise correlation.



(c) Performances of MV when the input of each base-layer classifier is the representative time series.

Figure 6.28: Gender classification performances obtained by the 1-level fusion of decisions of *subband-classifiers* using majority voting (MV).

section, we analyze whether combining information obtained from multi-resolution mesh networks also increases the gender classification performance.

In this set of experiments, we fuse the decisions of *subband-classifiers* for each task separately as shown in Figure 5.5. Note that, for different levels of decomposition, we fuse different numbers of *subband-classifiers*. Therefore, in Figure 6.27 and Figure 6.28, we share the classification performances obtained at different levels of decomposition. We observe that, performances obtained by fusing the decisions of classifiers with FSG using mesh arc descriptors (see Figure 6.27a) and pairwise correlation as inputs (Figure 6.27b) increase for all tasks as the decomposition level increases and more *subband-classifiers* are used in the ensemble. However, performances obtained by fusing the decisions of classifiers with FSG using representative time series shown in Figure 6.27c do not change for any of the tasks with decomposition level. Maximum results obtained for all tasks with mesh arc descriptors, pairwise correlation and representative time series vary in the ranges 81.19% – 87.62%, 83.79% – 86.26% and 83.54% – 85.02%, respectively.

On the other hand, performances obtained by fusing the decisions of classifiers trained with mesh arc descriptors with MV (see Figure 6.28a) first increase up to level 4 and then decrease up to level 6 for all tasks. Moreover, performances obtained with pairwise correlation shown in Figure 6.28b are similar for all levels. Finally, MV results of representative time series in Figure 6.28c increase as the decomposition level increases. When the decisions of classifiers are fused with MV, maximum results obtained for all tasks with mesh arc descriptors, pairwise correlation and representative time series vary in the ranges 75.00% – 80.57%, 76.11% – 81.93% and 81.68% – 84.53%, respectively.

6.4.1.5 Performances Obtained by 2-level Architecture Containing Fusion of Task-Classifiers Followed by Fusion of Subband-Classifiers

Recall from Figure 5.6 that, in order to understand the complementing information obtained from multi-resolution, multi-task mesh networks, we fuse the decisions of classifiers trained with these mesh networks in a 2-level architecture. In this subsection, we present the results of fusion of base layer *task-classifiers* followed by

fusion of *subband-classifiers* of first meta layer in Table 6.15. We fuse the decisions of *subband-classifiers* of first meta layer using FSG and MV. Experimental evidence show that fusing the information coming from different subbands and different tasks leads to an increase in the performance. Employing mesh arc descriptors or pairwise correlation results in a boost in the performance of 2-level architecture containing fusion of *task-classifiers* followed by fusion of *subband-classifiers* compared to single subband, single task or 1-level fusion results. Among them, mesh arc descriptors lead to the best accuracy (94.06%). When MV is used to fuse the decisions of *subband-classifiers* of first level meta layer, we obtain the best result with mesh arc descriptors (92.45%), compared to pairwise Correlation (91.58%) and representative time series (84.53%).

Table 6.15: Gender classification performances obtained by the 2-level fusion of *task-classifiers* followed by fusion of *subband-classifiers*

Level	MAD		PC		RTS	
	FSG	MV	FSG	MV	FSG	MV
1	91.34	88.74	87.25	86.76	84.90	56.44
2	92.82	90.72	92.45	90.10	84.53	64.36
3	93.19	92.45	92.57	91.21	84.41	73.27
4	93.32	91.34	93.19	91.58	84.65	79.46
5	93.07	91.09	93.07	90.59	84.53	83.54
6	94.06	90.59	93.56	90.59	84.53	84.53

We can infer from Table 6.15 that, employing FSG performs slightly better than MV and mesh arc descriptors lead to the best results compared to Pearson correlation and representative time series. Also, we can say that FSG works better for larger decomposition levels in which more number of classifiers are included in the ensemble when mesh arc descriptors and Pearson correlation are used.

Table 6.16: Gender classification results obtained by the 2-level fusion of *subband-classifiers* followed by fusion of *task-classifiers*

Level	MAD		PC		RTS	
	FSG	MV	FSG	MV	FSG	MV
1	90.97	91.58	87.13	85.15	84.65	84.90
2	92.95	91.46	92.20	89.85	84.65	84.90
3	93.19	91.96	92.33	89.60	84.41	84.41
4	93.32	92.45	91.71	90.97	84.65	84.78
5	93.44	91.96	92.08	90.47	84.28	84.90
6	93.69	92.57	93.07	91.34	84.90	84.90

6.4.1.6 Performances Obtained by 2-level Architecture Containing Fusion of Subband-Classifiers Followed by Fusion of Task-Classifiers

In order to test the effect of the order of fusion for *task-classifiers* and *subband-classifiers* on the gender classification performance, we first fuse the decisions of *subband-classifiers* and then those of *task-classifiers*. We compare the performances obtained by this architecture with the ones given in previous subsection.

We compute the classification performances of 2-level architecture, in which decisions of the base layer *subband-classifiers* are fused in the first meta layer and their decisions are further fused in the second meta layer as shown in Figure 5.7. Similar to previous architecture, we fuse the decisions of *subband-classifiers* of first level meta layer using FSG and MV. FSG results in Table 6.16 reflect that, mesh arc descriptors discriminate gender slightly better than pairwise correlation and much better than representative time series. We observe that the best performances of mesh arc descriptors and pairwise correlation are obtained when decomposition level is larger. On the other hand, performances do not increase with the level of decomposition for representative time series. We can infer that, classifiers trained with multi-resolution and multi-task representative time series do not carry complementing information. However, mesh arc descriptors and pairwise correlation features extracted for multi-resolution and multi-task contain complementing information such that we observe an increase in the classification accuracy as we fuse more *subband-classifiers*.

We also get the best MV performances using mesh arc descriptors compared to other features as shown in Table 6.16. We observe that performances obtained using FSG are slightly better than the ones obtained using MV when mesh arc descriptors and pairwise correlation are used as features.

By comparing the classification performances given in Table 6.15 and Table 6.16, we can conclude that the order of fusion of *subband-classifiers* or *task-classifiers* in our 2-level architecture do not change the gender classification accuracy significantly.

6.4.2 Analysis of Gender Discriminative Power of Mesh Arc Descriptors

Is there any relationship between anatomical regions, which carry information to discriminate gender independent from the cognitive task the individual is performing? Is there a specific region that is significant to discriminate gender only the individual is performing a specific task? In order to answer these questions, we analyze the gender discriminative power of MADs in this subsection as explained in Section 5.5.2.

We suffice to show the anatomical regions of the top-50 mesh arc descriptors with corresponding subbands in Figures 6.29 - 6.35 for each task separately, since we can understand the overall distribution of significant MADs to subbands by plotting only the best 50 MADs. We denote the number of tasks under which the corresponding mesh arc descriptor is discriminative for the corresponding task with the numbers in parenthesis following the names of region pairs. We observe that the mesh arc descriptor formed between right inferior parietal and right angular regions is the most gender discriminative descriptor for Emotion (see Figure 6.29), Gambling (see Figure 6.30) and Relational Processing (see Figure 6.33) tasks for subbands D_1 and D_2 . Similarly, this mesh arc descriptor takes place in top-50 most discriminative mesh arc descriptors of remaining tasks. Also, the mesh arc descriptors formed between right middle occipital and right inferior parietal region pairs for subband D_1 and right inferior parietal and right superior parietal region pairs for subbands D_1 and D_2 carry gender discriminative information under all of the 7 cognitive tasks. Notice that, the mesh arc descriptors formed between reverse order of these region pairs are discriminative for all tasks as well.

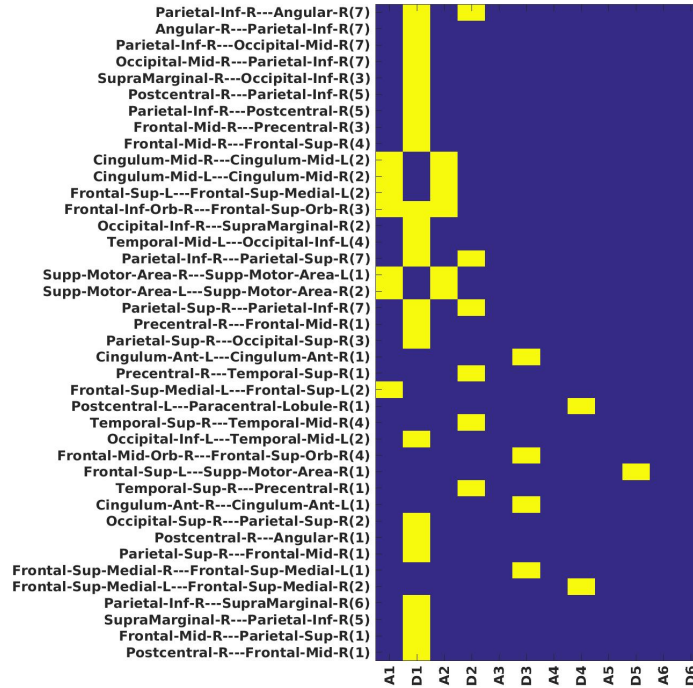


Figure 6.29: Top 50 gender discriminative MADs obtained for Emotion

In addition to the discriminative mesh arc descriptors for all tasks, there exist the ones which are discriminative for only a specific cognitive task. Notice that, mesh arc descriptors formed between the region pairs left inferior frontal orbital & left inferior temporal, right medial superior frontal & right middle temporal, left middle occipital & left inferior occipital, left supplementary motor area & left precentral and right middle frontal orbital & right inferior frontal orbital are very discriminative under only Gambling (see Figure 6.30), Language Processing (see Figure 6.31), Motor (see Figure 6.32), Relational Processing (see Figure 6.33) and Working Memory (see Figure 6.35) tasks, respectively. Also, mesh arc descriptor between left and right supplementary motor areas carry gender discriminative information under only Emotion task whereas the one between right middle temporal and right inferior occipital regions is discriminative under Social Processing task. All of the mesh arc descriptors which are discriminative under a specific task are denoted with (1) followed by the corresponding region pairs in Figures 6.29 - 6.35.

From Figures 6.29 - 6.35 we can also infer that, for large levels of decomposition, we rarely observe gender discriminative mesh arc descriptors. We observe a few dis-

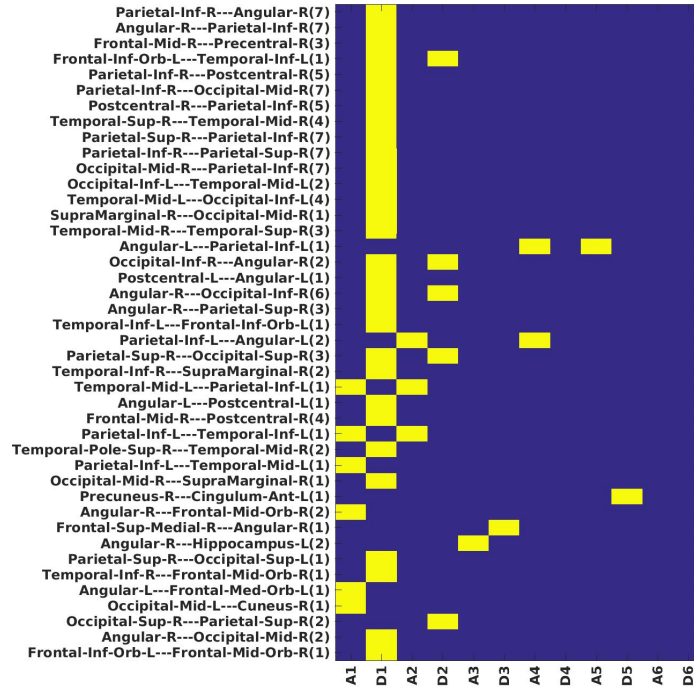


Figure 6.30: Top 50 gender discriminative MADs obtained for Gambling

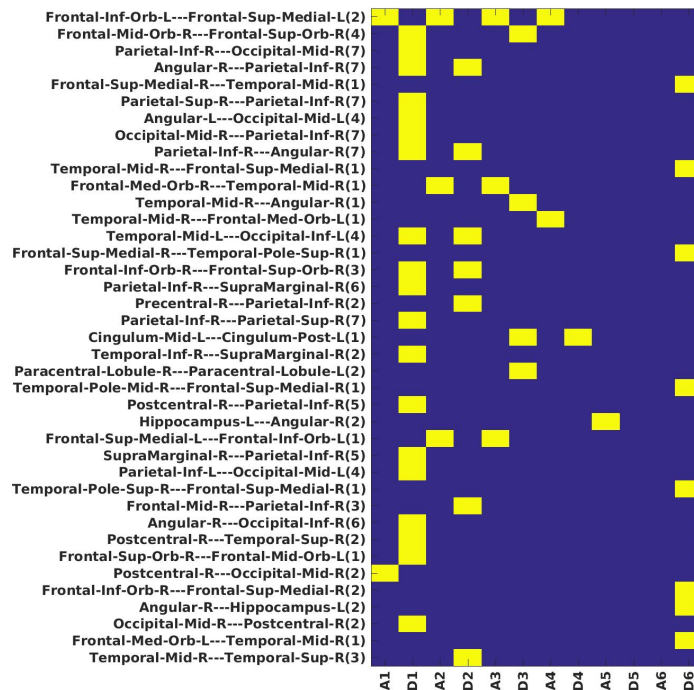


Figure 6.31: Top 50 gender discriminative MADs obtained for Language Processing

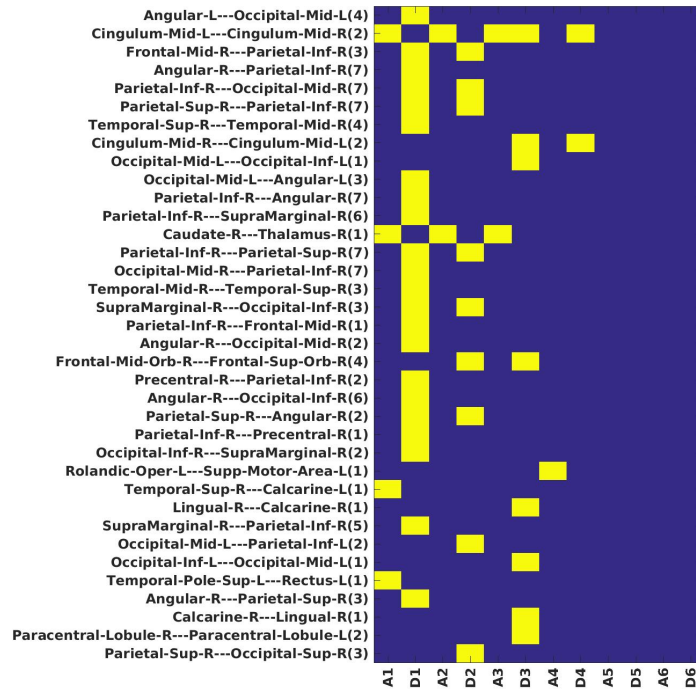


Figure 6.32: Top 50 gender discriminative MADs obtained for Motor

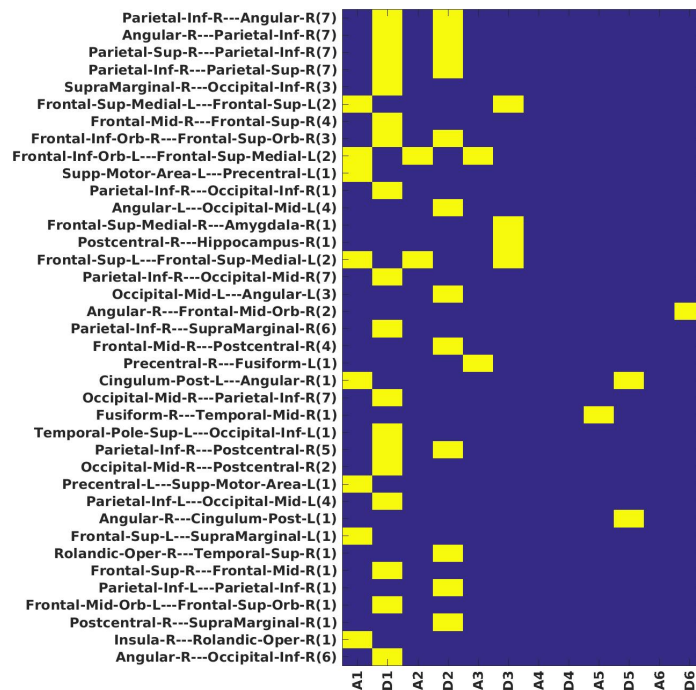


Figure 6.33: Top 50 gender discriminative MADs obtained for Relational Processing

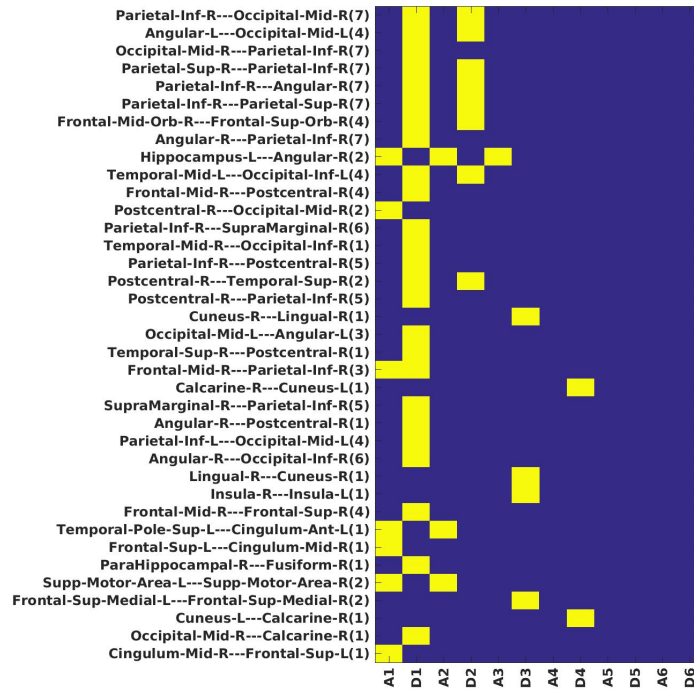


Figure 6.34: Top 50 gender discriminative MADs obtained for Social Processing

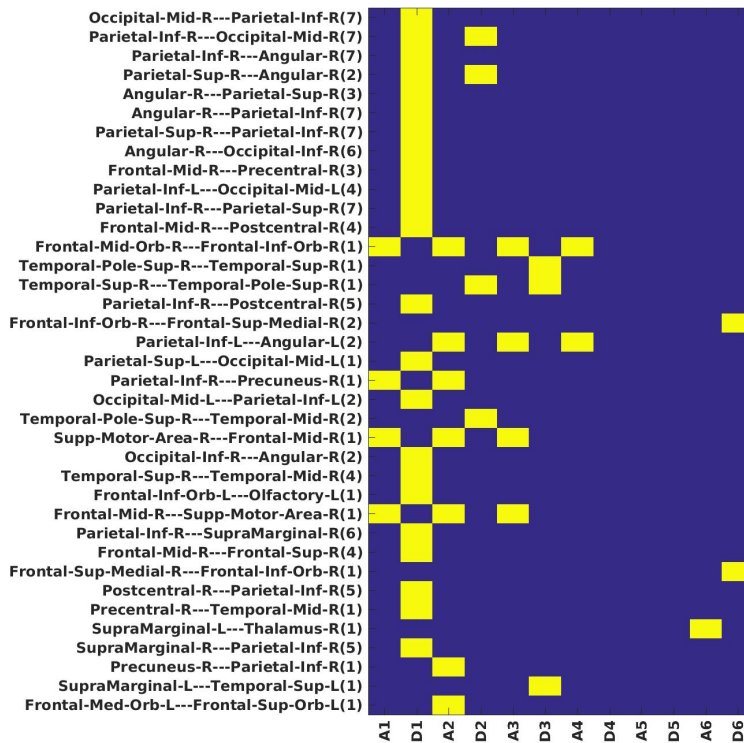


Figure 6.35: Top 50 gender discriminative MADs obtained for Working Memory

criminative mesh arc descriptors at level 6 under Language, Relational Processing and Working Memory tasks. On the other hand, D_1 is the subband that has the majority of the discriminative mesh arc descriptors for all tasks.

We observe that MAD formed between left inferior parietal and left angular regions occurs in the top-10 list of 21 feature matrices among 154. Moreover, MADs formed between right middle cingulum - left middle cingulum, left frontal superior medial - left frontal inferior orbital, right frontal superior orbital - right frontal middle orbital regions occur more in top-10 lists compared to other MADs.

MADs denoted in Figures 6.29 - 6.35 carry gender discriminative information for multiple tasks. Especially, the ones formed between the first six pairs are significant for more than or equal to 5 tasks, meaning that relationship between these regions have gender discriminative power independent from the task that the participant is performing. On the other hand, we obtain MADs which discriminate gender under a specific task. For example, the ones formed between right inferior frontal orbital - right middle frontal orbital and right cuneus - left temporal middle represent gender discriminative information only under working memory (WM) task. Moreover, MADs formed between left anterior cingulum - left superior temporal pole and right posterior cingulum - left middle cingulum discriminate gender only when participants perform social processing task. Similarly, MADs formed between the region pairs, left middle cingulum - left posterior cingulum and right middle temporal - right angular carry gender discriminative information under Language processing task. However, majority of the significant MADs discriminate gender under multiple tasks.

Figures 6.29 - 6.35 show that MADs formed between region pairs carry gender discriminative information for multiple subbands. We observe that, mesh descriptors formed between the first six region pairs of Figures 6.29 - 6.35 are discriminative for at least five distinct subbands. Note that, mesh arc descriptors formed between region pairs right angular - right inferior parietal, right middle occipital - right inferior parietal, right inferior parietal - right superior parietal and left inferior occipital - left middle temporal carry discriminative information for only subbands D_1 and D_2 , which are similar subbands.

In our analyses, we do not derive conclusions from the perspective of neuroscience.

Instead, we only mention the general distribution of significant MADs to different subbands and tasks. We can only infer from our results that, mesh arc descriptors formed between different region pairs may be significant for different subbands. In other words, it is not the case that a few subbands carry gender discriminative information while others do not. Therefore, we can conclude that taking decisions of multiple subbands into consideration during gender classification is crucial.

6.5 Chapter Summary

In this chapter we provided performances of our classification experiments and results of our further analyses. First of all, we showed that classification of cognitive states using MADs within functional (FLM) and spatial (SLM) neighborhoods gives better accuracy compared to state-of-the-art feature representations including mesh features discarding temporal information, pairwise correlations, representative time series and extracted features using PCA and ICA. We also validated our neighborhood assumption by analyzing the model error.

Then, we showed the superiority of Fisher Vector encoding of MADs over raw MADs, VLAD or BoW encoding of MADs and other state-of-the-art fMRI representations. Moreover, we identified a relationship between the energy of columns of Fisher vectors and classification accuracy and conclude that Gaussians with high energy lead to features with more discriminative power. Recall that, we also introduced a method to visualize *brain connectivity dictionary* on a human brain template.

Following the extraction of MADs from original fMRI signals, we provided results of multi-resolution experiments. We showed that fusion of multi-resolution mesh networks boosts the classification accuracy of MADs extracted from original resolution fMRI signal. Our analyses results reflected that classifiers trained with multi-resolution mesh networks represent diverse information related to cognitive states so that their fusion leads to the complementation of diverse information. We further showed that multi-resolution multi-task mesh networks represent diverse information related to gender. By fusing the multi-task, multi-resolution information in a 2-level hierarchical architecture, we obtained the best gender classification accuracy.

CHAPTER 7

CONCLUSION AND FUTURE WORK

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.

Carl Friedrich Gauss

In this thesis, we proposed novel representations and frameworks to discriminate cognitive states and gender from fMRI data. We first introduced mesh networks formed by ensembling a set of Functionally Local Meshes (FLM) or Spatially Local Meshes (SLM), in which nodes are represented as a weighted linear combination of its neighboring nodes. We estimated the temporal relationships among nodes, called Mesh Arc Descriptors, using ridge regression by employing all the fMRI measurements recorded during a state. We showed that representing fMRI signal in terms of estimated MADs of mesh networks discriminates cognitive states better compared to state-of-the-art fMRI representations, including raw fMRI time series or pairwise relationship between them.

Motivated by the success of encoding methods, especially Fisher Vector encoding, applied on local descriptors in computer vision domain, we encoded our local descriptors (MADs), using Fisher Vector (FV), Vector of Locally Aggregated Descriptors (VLAD) and Bag-of-Words (BoW) methods. Our cognitive state classification results reflect that MADs are good local descriptors for encoding approaches and FV encoding of MADs yield a boost in the classification accuracy. We also identified a relationship between the energy of FV columns of Gaussians in the GMM and classi-

fication accuracy, such that Gaussians, whose FV columns have higher energy represent more discriminative information. Finally, we suggested a visualization method to plot codewords (means of Gaussians) on a human brain template. This way, we could clearly visualize the relationship patterns revealed by codewords of Gaussians. Moreover, by sorting the codewords based on the energy their Gaussian columns have in FV, we could analyze the discriminative power of relationship patterns provided by the codewords.

It is postulated that human brain is assumed to process information in multiple frequency bands in [72, 141]. Based upon this observation, we argued whether multi-resolution fMRI data represent distinct information which can be combined under a fusion architecture. In order to verify our arguments, we proposed a multi-resolution framework, called Hierarchical Multi-resolution Mesh Networks (HMMNs), in which first we decompose the fMRI signal into multiple frequency subbands using Discrete Wavelet Transform and its inverse. Then, for each subband, we estimated MADs around all nodes and form multi-resolution mesh networks. Finally, we fused the information obtained from multi-resolution mesh networks under Fuzzy Stacked Generalization (FSG) architecture. We observed that, multi-resolution mesh networks represent complementing information and fusion of these networks under FSG architecture boosts the cognitive state classification accuracy. Results of both significance tests of meta-layer and base-layer features and classifier diversity measures validated our assumption.

Finally, we hypothesized that MADs of multi-task multi-resolution mesh networks carry diverse information to discriminate gender and fusion of these networks yields better classification accuracy. For that purpose, we suggested a 2-level FSG architecture, in which these two layers are responsible for the fusion of diverse information coming from distinct subbands and distinct cognitive tasks. We showed that, fusing the representations computed at multiple resolutions for multiple tasks provides the best gender classification accuracy compared to fusion of representations obtained using only multi-task or only multi-resolution data. Besides, mesh arc descriptors slightly outperformed pairwise correlations between regions, and significantly outperformed representative time series. In addition to classification accuracy, significance tests also validated our hypothesis.

7.1 Future Work

The proposed mesh network representations and multi-resolution frameworks perform well with fMRI data collected from healthy participants for cognitive task and gender classification. A future direction of this study can be the application of the suggested approaches to resting state fMRI data. This way, our approaches can be applied for the identification of diseases such as Alzheimer’s disease, Schizophrenia, etc. from resting state networks.

We showed that gender of a participant can be identified with a substantially high classification performance when participants perform multiple tasks. It is observed that gender information is complementary in different cognitive tasks. Another research direction would be to record fMRI measurements from patients while they perform multiple tasks and investigate whether the disease can be diagnosed more precisely while patients perform multiple tasks.

The proposed methods of this thesis can, also, be applied to the identification of demographic information, such as age or ethnicity of participants from their fMRI measurements. Ethnicity analysis may reveal significant connectivity patterns unique to a specific ethnicity. On the other hand, a machine learning approach to age analysis may be crucial for the discovery of change in connectivity patterns with age.

In this thesis, temporal relationships are extracted from fMRI data in a hand-crafted way. Yet, deep learning methods are successful in almost all areas today and types of Recurrent Neural Networks may be suitable to extract significant temporal relationships among region time series. The problem of task fMRI data is that it has limited number of measurements while deep architectures require a large amount of data. We believe that, when the number of data would increase, the success of modeling temporal relationships would increase.

REFERENCES

- [1] A. Abbas, N. Deligiannis, and Y. Andreopoulos. Vectors of locally aggregated centers for compact video representation. In *International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2015.
- [2] A. Abraham, K. Thybusch, K. Pieritz, and C. Hermann. Gender differences in creative thinking: behavioral and fmri findings. *Brain imaging and behavior*, 8(1):39–51, 2014.
- [3] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006.
- [4] S. Alkan and F. T. Yarman-Vural. Ensembling brain regions for brain decoding. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2948–2951, 2015.
- [5] M. R. Arbabshirani, K. Kiehl, G. Pearlson, and V. D. Calhoun. Classification of schizophrenia patients based on resting-state functional network connectivity. *Frontiers in neuroscience*, 7:133, 2013.
- [6] C. Baldassano, M. C. Iordan, D. M. Beck, and L. Fei-Fei. Voxel-level functional connectivity using spatial regularization. *NeuroImage*, 63(3):1099 – 1106, 2012.
- [7] L. Baldassarre, J. Mourao-Miranda, and M. Pontil. Structured sparsity models for brain decoding from fmri data. In *International Workshop on Pattern Recognition in NeuroImaging (PRNI)*, pages 5–8. IEEE, 2012.
- [8] 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, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg,

- A. Z. Snyder, and D. C. V. Essen. Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, 80:169 – 189, 2013.
- [9] A. T. Baria, M. N. Baliki, T. Parrish, and A. V. Apkarian. Anatomical and functional assemblies of brain bold oscillations. *Journal of Neuroscience*, 31(21):7910–7919, 2011.
- [10] L. C. Baxter, A. J. Saykin, L. A. Flashman, S. C. Johnson, S. J. Guerin, D. Babcock, and H. A. Wishart. Sex differences in semantic language processing: a functional MRI study. *Brain and language*, 84(2):264–272, 2003.
- [11] N. Bazargani and A. Nosratinia. Joint maximum likelihood estimation of activation and hemodynamic response function for fMRI. *Med Image Anal*, 18(5):711 – 724, 2014.
- [12] 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, K. Koenig, P. S. Morgan, J. G. Ojemann, C. Rorden, J. P. Szaflarski, M. E. Tivarus, and K. E. Weaver. Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage*, 54(2):1465 – 1475, 2011.
- [13] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.
- [14] A. Boghi, O. Rampado, M. Bergui, F. Avidano, C. Manzone, M. Coriasco, P. Mortara, L. Orsi, R. Ropolo, and G. Bradac. Functional MRI study of a motor task and the tower of London task at 1.0 T. *Neuroradiology*, 48(10):763–771, 2006.
- [15] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
- [16] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [17] M. Brett, J.-L. Anton, R. Valabregue, and J.-B. Poline. Region of interest analysis using an SPM toolbox. In *8th International Conference on Functional Mapping of the Human Brain*, Sendai, Japan, 2002.

- [18] K. H. Brodersen, F. Haiss, C. S. Ong, F. Jung, M. Tittgemeyer, J. M. Buhmann, B. Weber, and K. E. Stephan. Model-based feature construction for multivariate decoding. *NeuroImage*, 56(2):601–615, 2011.
- [19] K. H. Brodersen, T. M. Schofield, A. P. Leff, C. S. Ong, E. I. Lomakina, J. M. Buhmann, and K. E. Stephan. Generative embedding for model-based classification of fmri data. *PLoS Comput Biol*, 7(6):e1002079, 2011.
- [20] R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. T. Yeo. The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(5):2322–2345, 2011.
- [21] E. Bullmore, J. Fadili, V. Maxim, L. Şendur, B. Whitche, J. Suckling, M. Brammer, , and M. Breakspear. Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage*, 23, Supplement 1:S234 – S249, 2004.
- [22] T. Butler, J. Imperato-McGinley, H. Pan, D. Voyer, A. C. Cunningham-Bussell, L. Chang, Y.-S. Zhu, J. J. Cordero, E. Stern, and D. Silbersweig. Sex specificity of ventral anterior cingulate cortex suppression during a cognitive task. *Human Brain Mapping*, 28(11):1206–1212, 2007.
- [23] G. Buzsáki and A. Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, 2004.
- [24] C. Cabral, M. Silveira, and P. Figueiredo. Decoding visual brain states from fmri using an ensemble of classifiers. *Pattern Recognition*, 45(6):2064–2074, 2012.
- [25] L. Cahill. Why sex matters for neuroscience. *Nature Reviews Neuroscience*, 7(6):477–484, 2006.
- [26] J. Carvajal, C. McCool, B. Lovell, and C. Sanderson. Joint recognition and segmentation of actions via probabilistic integration of spatio-temporal fisher vectors. *arXiv preprint arXiv:1602.01601*, 2016.
- [27] 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.

- [28] M. S. Cetin, J. M. Houck, V. M. Vergara, R. L. Miller, and V. Calhoun. Multimodal based classification of schizophrenia patients. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 2629–2632. IEEE, 2015.
- [29] K. Chatfield, V. S. Lempitsky, A. Vedaldi, and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. In *BMVC*, volume 2, page 8, 2011.
- [30] K. Chen, E. M. Reiman, Z. Huan, R. J. Caselli, D. Bandy, N. Ayutyanont, and G. E. Alexander. Linking functional and structural brain images with multivariate network analyses: a novel application of the partial least square method. *Neuroimage*, 47(2):602–610, 2009.
- [31] L. Cossell, M. F. Iacaruso, D. R. Muir, R. Houlton, E. N. Sader, H. Ko, S. B. Hofer, and T. D. Mrsic-Flogel. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, 518(7539):399–403, 2015.
- [32] M. N. Coutanche, S. L. Thompson-Schill, and R. T. Schultz. Multi-voxel pattern analysis of fmri data predicts clinical symptom severity. *Neuroimage*, 57(1):113–123, 2011.
- [33] 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*, 19(2):261–270, 2003.
- [34] R. C. Craddock, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. Disease state prediction from resting state functional connectivity. *Magnetic resonance in Medicine*, 62(6):1619–1628, 2009.
- [35] C. Davatzikos, K. Ruparel, Y. Fan, D. Shen, M. Acharyya, J. Loughhead, R. 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.
- [36] M. R. Delgado, L. E. Nystrom, C. Fissell, D. C. Noll, and J. A. Fiez. Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, 84(6):3072–3077, 2000.

- [37] J. Delhumeau, P.-H. Gosselin, H. Jégou, and P. Pérez. Revisiting the vlad image representation. In *International Conference on Multimedia*, pages 653–656. ACM, 2013.
- [38] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.
- [39] I. D. Dinov, J. W. Boscardin, M. S. Mega, E. L. Sowell, and A. W. Toga. A wavelet-based statistical analysis of fmri data. *Neuroinformatics*, 3(4):319–342, 2005.
- [40] J. M. Duarte-Carvajalino, N. Jahanshad, C. Lenglet, K. L. McMahon, G. I. de Zubicaray, N. G. Martin, M. J. Wright, P. M. Thompson, and G. Sapiro. Hierarchical topological network analysis of anatomical human brain connectivity and differences related to sex and kinship. *Neuroimage*, 59(4):3784–3804, 2012.
- [41] I. O. Ertugrul, M. Ozay, and F. T. Y. Vural. Encoding the local connectivity patterns of fmri for cognitive state classification. *arXiv preprint arXiv:1610.05036*, 2016.
- [42] 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*, 9:1871–1874, 2008.
- [43] Y. Fan. On the approximate decorrelation property of the discrete wavelet transform for fractionally differenced processes. *IEEE Transactions on Information Theory*, 49(2):516–521, 2003.
- [44] S. Fine, J. Navratil, and R. A. Gopinath. A hybrid gmm/svm approach to speaker identification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 1, pages 417–420. IEEE, 2001.
- [45] O. Firat, I. Onal, E. Aksan, B. Velioglu, I. Oztekin, and F. T. Y. Vural. Large scale functional connectivity for brain decoding. In *11 th IASTED International Conference on Biomedical Engineering (BioMed)*, 2014.

- [46] O. Firat, M. Ozay, I. Onal, I. Oztekin, and F. Yarman Vural. Representation of cognitive processes using the minimum spanning tree of local meshes. In *EMBC*, pages 6780–6783, Osaka, Japan, July 2013.
- [47] O. Firat, M. Ozay, I. Onal, I. Oztekin, and F. Vural. Functional mesh learning for pattern analysis of cognitive processes. In *ICCI*CC*, pages 161–167, New York City, USA, July 2013.
- [48] A. Floren, B. Naylor, R. Miikkulainen, and D. Ress. Accurately decoding visual information from fmri data obtained in a realistic virtual environment. *Frontiers in human neuroscience*, 9, 2015.
- [49] Y. Freund, R. E. Schapire, et al. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.
- [50] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [51] G. Giacinto and F. Roli. Design of effective neural network ensembles for image classification purposes. *Image and Vision Computing*, 19(9):699–707, 2001.
- [52] 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.
- [53] G. Gong, Y. He, and A. C. Evans. Brain connectivity: gender makes a difference. *The Neuroscientist*, 17(5):575–591, 2011.
- [54] A. Gramfort, B. Thirion, and G. Varoquaux. Identifying predictive regions from fmri with tv-l1 prior. In *International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 17–20. IEEE, 2013.
- [55] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

- [56] L. Grosenick, B. Klingenberg, K. Katovich, B. Knutson, and J. E. Taylor. Interpretable whole-brain prediction analysis with graphnet. *NeuroImage*, 72:304–321, 2013.
- [57] L. K. Hansen and P. Salamon. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12(10):993–1001, 1990.
- [58] 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.
- [59] 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 – 2429, 2001.
- [60] J.-D. Haynes and G. Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience*, 8(5):686–691, 2005.
- [61] B. J. He, J. M. Zempel, A. Z. Snyder, and M. E. Raichle. The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66(3):353–369, 2010.
- [62] T. K. Ho. The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):832–844, 1998.
- [63] A. Hofer, C. M. Siedentopf, A. Ischebeck, M. A. Rettenbacher, M. Verius, S. Felber, and W. W. Fleischhacker. Sex differences in brain activation patterns during processing of positively and negatively valenced emotional words. *Psychological medicine*, 37(01):109–119, 2007.
- [64] X. Hu, L. Guo, J. Han, and T. Liu. Decoding power-spectral profiles from fmri brain activities during naturalistic auditory experience. *Brain imaging and behavior*, pages 1–11, 2016.
- [65] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, 10(3):626–634, 1999.

- [66] T. Jaakkola, M. Diekhans, and D. Haussler. A discriminative framework for detecting remote protein homologies. *Journal of Computational Biology*, 7(1-2):95–114, 2000.
- [67] T. S. Jaakkola, D. Haussler, et al. Exploiting generative models in discriminative classifiers. *Advances in neural information processing systems*, pages 487–493, 1999.
- [68] H. Jégou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3304–3311. IEEE, 2010.
- [69] Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5):679 – 685, 2005.
- [70] Y. Kamitani and F. Tong. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr Biol*, 16(11):1096 – 1102, 2006.
- [71] J.-P. Kauppi, I. P. Jääskeläinen, M. Sams, and J. Tohka. Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. *Frontiers in Neuroinformatics*, 4(5), 2010.
- [72] J.-P. Kauppi, J. Pajula, and J. Tohka. A versatile software package for inter-subject correlation based analyses of fmri. *Frontiers in Neuroinformatics*, 8(2), 2014.
- [73] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi. Automatic classification of alzheimer’s disease with resting-state fmri and graph theory. In *21th Iranian Conference on Biomedical Engineering (ICBME)*, pages 252–257. IEEE, 2014.
- [74] S. Kloppel, A. Abdulkadir, C. R. Jack, N. Koutsouleris, J. Mourao-Miranda, and P. Vemuri. Diagnostic neuroimaging across diseases. *NeuroImage*, 61(2):457 – 463, 2012.
- [75] S.-I. Ktena and R. Daniel. A topological graph kernel for gender classification of functional brain networks. In *OHBM*, 2016.

- [76] J. Kubilius, A. Baeck, J. Wagemans, and H. P. O. de Beeck. Brain-decoding fmri reveals how wholes relate to the sum of parts. *Cortex*, 72:5–14, 2015.
- [77] V. Kulkarni, J. S. Pudipeddi, L. Akoglu, J. T. Vogelstein, R. J. Vogelstein, S. Ryman, and R. E. Jung. Sex differences in the human connectome. In *International Conference on Brain and Health Informatics*, pages 82–91. Springer, 2013.
- [78] L. I. Kuncheva. "fuzzy" versus "nonfuzzy" in combining classifiers designed by boosting. *IEEE Transactions on Fuzzy Systems*, 11(6):729–741, 2003.
- [79] L. I. Kuncheva. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.
- [80] L. I. Kuncheva and J. J. Rodriguez. Classifier ensembles with a random linear oracle. *IEEE Transactions on Knowledge and Data Engineering*, 19(4), 2007.
- [81] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumptre, D. E. J. Linden, and S. J. Johnston. Random subspace ensembles for fmri classification. *IEEE Transactions on Medical Imaging*, 29(2):531 – 542, 2010.
- [82] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.
- [83] Y. Li, C.-Y. Wee, B. Jie, Z. Peng, and D. Shen. Sparse multivariate autoregressive modeling for mild cognitive impairment classification. *Neuroinformatics*, 12(3):455–469, 2014.
- [84] L. Liu, C. Shen, L. Wang, A. Van Den Hengel, and C. Wang. Encoding high dimensional local features by sparse coding based fisher vectors. In *Advances in Neural Information Processing Systems*, pages 1143–1151, 2014.
- [85] W. Liu, H. Chen, and Q. Lu. Discriminate brain states from fmri images using fuzzy support vector machines. In *International Symposium on Intelligent Information Technology Application*, volume 2, pages 567–571. IEEE, 2008.
- [86] B. B. Mandelbrot. *The Fractal Geometry of Nature*. Springer, New York, 1977.

- [87] A. Marchewka, K. Jednorog, M. Falkiewicz, W. Szeszkowski, A. Grabowska, and I. Szatkowska. Sex, lies and fmri—gender differences in neural basis of deception. *PloS one*, 7(8):e43076, 2012.
- [88] L. M. Meriño, J. Meng, S. Gordon, B. J. Lance, T. Johnson, V. Paul, K. Robbins, J. M. Vettel, and Y. Huang. A bag-of-words model for task-load prediction from eeg in complex environments. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1227–1231. IEEE, 2013.
- [89] 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.
- [90] E. Mizrak and I. Oztekin. Relationship between emotion and forgetting. *Emotion*, 2015.
- [91] G. Modinos, A. Mechelli, W. Pettersson-Yeo, P. Allen, P. McGuire, and A. Aleman. Pattern classification of brain activation during emotional processing in subclinical depression: psychosis proneness as potential confounding factor. *PeerJ*, 1:e42, 2013.
- [92] P. J. Moreno and R. Rifkin. Using the fisher kernel method for web audio classification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 6, pages 2417–2420. IEEE, 2000.
- [93] B. Ng and R. Abugharbieh. Generalized group sparse classifiers with application in fmri brain decoding. In *CVPR*, pages 1065–1071, Colorado Springs, USA, 2011.
- [94] B. Ng and R. Abugharbieh. Modeling spatiotemporal structure in fmri brain decoding using generalized sparse classifiers. In *International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 65–68, Korea, Seoul, 2011.
- [95] B. Ng, A. Vahdat, G. Hamarneh, and R. Abugharbieh. Generalized sparse classifiers for decoding cognitive states in fmri. In *MLMI*, volume 6357, pages 108–115. Beijing, China, 2010.

- [96] N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *PNAS*, 103(10):3863 – 3868, 2006.
- [97] L. Oliveira, C. D. Ladouceur, M. L. Phillips, M. Brammer, and J. Mourao-Miranda. What does brain response to neutral faces tell us about major depression? evidence from machine learning and fmri. *PloS one*, 8(4):e60121, 2013.
- [98] I. Önal. *An Information Theoretic Representation of Brain Connectivity for Cognitive State Classification Using Functional Magnetic Resonance Imaging*. PhD thesis, Middle East Technical University, 2013.
- [99] I. Onal, E. Aksan, B. Velioglu, O. Firat, M. Ozay, I. Oztekin, and F. T. Yarman Vural. Modeling the brain connectivity for pattern analysis. In *ICPR*, Stockholm, Sweden, 2014.
- [100] I. Onal, M. Ozay, O. Firat, I. Oztekin, and F. Yarman Vural. Analyzing the information distribution in the fmri measurements by estimating the degree of locality. In *EMBC*, pages 6772–6775, Osaka, Japan, July 2013.
- [101] I. Onal, M. Ozay, E. Mizrak, I. Oztekin, and F. Yarman-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*, 2017.
- [102] I. Onal, M. Ozay, and F. T. Y. Vural. Functional mesh model with temporal measurements for brain decoding. In *International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 2624–2628. IEEE, 2015.
- [103] I. Onal, M. Ozay, and F. T. Y. Vural. Hierarchical multi-resolution mesh networks for brain decoding. *arXiv preprint arXiv:1607.07695*, 2016.
- [104] I. Onal, M. Ozay, and F. Yarman Vural. Modeling voxel connectivity for brain decoding. In *International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 5 – 8, Stanford, CA, USA, 2015.
- [105] D. Oneata, J. Verbeek, and C. Schmid. Action and event recognition with fisher vectors on a compact feature set. In *International Conference on Computer Vision*, pages 1817–1824. IEEE, 2013.

- [106] G. Orru, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci. Biobehav. Rev.*, 36(4):1140 – 1152, 2012.
- [107] M. Özay. *Decision Fusion for Supervised, Unsupervised and Semi-Supervised Learning*. PhD thesis, Middle East Technical University, 2013.
- [108] M. Ozay, I. Oztekin, U. Oztekin, and F. T. Yarman-Vural. Mesh learning for classifying cognitive processes. *CoRR*, abs/1205.2382, 2012.
- [109] M. Ozay and F. T. Yarman-Vural. Hierarchical distance learning by stacking nearest neighbor classifiers. *Information Fusion*, 29:14 – 31, 2016.
- [110] I. Oztekin and D. Badre. Distributed patterns of brain activity that lead to forgetting. *Front Hum Neurosci*, 5(86), 2011.
- [111] 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 Comput Biol*, 8(3):e1002441, 2012.
- [112] S. Parida, S. Dehuri, and S.-B. Cho. Neuro-fuzzy ensembler for cognitive states classification. In *IEEE International Advance Computing Conference (IACC)*, pages 1243–1247, Feb 2014.
- [113] D. Partridge and W. Krzanowski. Software diversity: practical statistics for its measurement and exploitation. *Information and software technology*, 39(10):707–717, 1997.
- [114] F. Pereira, T. Mitchell, and M. Botvinick. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1):S199–S209, 2009.
- [115] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007.
- [116] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for

- large-scale image classification. *Computer Vision–ECCV 2010*, pages 143–156, 2010.
- [117] M. Piefke, P. H. Weiss, H. J. Markowitsch, and G. R. Fink. Gender differences in the functional neuroanatomy of emotional episodic autobiographical memory. *Human brain mapping*, 24(4):313–324, 2005.
 - [118] S. M. Polyn, V. S. Natu, J. D. Cohen, and K. A. Norman. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756):1963–1966, 2005.
 - [119] J. Richiardi, S. Achard, H. Bunke, and D. V. de Ville. Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Process. Mag*, 2013.
 - [120] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. V. D. Ville. Decoding brain states from fmri connectivity graphs. *NeuroImage*, 56(2):616 – 626, 2011.
 - [121] J. J. Rodríguez and L. I. Kuncheva. Naïve bayes ensembles with a random oracle. In *International Workshop on Multiple Classifier Systems*, pages 450–458. Springer, 2007.
 - [122] J. J. Rodríguez, L. I. Kuncheva, and C. J. Alonso. Rotation forest: A new classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1619–1630, 2006.
 - [123] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059 – 1069, 2010.
 - [124] R. Salvador, A. Martinez, E. Pomarol-Clotet, J. Gomar, F. Vila, S. Sarró, A. Capdevila, and E. Bullmore. A simple view of the brain through a frequency-specific functional connectivity measure. *Neuroimage*, 39(1):279–289, 2008.
 - [125] R. Salvador, J. Suckling, C. Schwarzbauer, and E. Bullmore. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1457):937–946, 2005.

- [126] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105(3):222–245, 2013.
- [127] J. Sánchez and J. Redolfi. Exponential family fisher vector for image classification. *Pattern Recognition Letters*, 59:26–32, 2015.
- [128] S. Sarraf and G. Tofghi. Classification of alzheimer’s disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*, 2016.
- [129] T. D. Satterthwaite, D. H. Wolf, D. R. Roalf, K. Ruparel, G. Erus, S. Vandekar, E. D. Gennatas, M. A. Elliott, A. Smith, H. Hakonarson, et al. Linked sex differences in cognition and functional connectivity in youth. *Cerebral cortex*, 25(9):2383–2394, 2014.
- [130] M. A. Savelonas, I. Pratikakis, and K. Sfikas. Fisher encoding of differential fast point feature histograms for partial 3d object retrieval. *Pattern Recognition*, 55:114–124, 2016.
- [131] V. J. Schmithorst and S. K. Holland. Functional mri evidence for disparate developmental processes underlying intelligence in boys and girls. *Neuroimage*, 31(3):1366–1379, 2006.
- [132] V. J. Schmithorst and S. K. Holland. Sex differences in the development of neuroanatomical functional connectivity underlying intelligence found using bayesian connectivity analysis. *Neuroimage*, 35(1):406–419, 2007.
- [133] M. Sekma, M. Mejdoub, and C. B. Amar. Human action recognition based on multi-layer fisher vector encoding method. *Pattern Recognition Letters*, 65:37–43, 2015.
- [134] H. Shen, L. Wang, Y. Liu, and D. Hu. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fmri. *Neuroimage*, 49(4):3110–3121, 2010.
- [135] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *British Machine Vision Conference*, volume 5, page 11, 2013.

- [136] R. Smith, K. Keramatian, and K. Christoff. Localizing the rostrolateral prefrontal cortex at the individual level. *NeuroImage*, 36(4):1387 – 1396, 2007.
- [137] G. W. Snedecor and W. G. Cochran. Statistical methods., 1967.
- [138] B. Solmaz, S. Dey, A. R. Rao, and M. Shah. Adhd classification using bag of words approach on network features. In *SPIE Medical Imaging*, pages 83144T–83144T, 2012.
- [139] G. Sucu, E. Akbaş, İ. Öztekin, E. Mızrak, and F. Y. Vural. Decoding cognitive states using the bag of words model on fmri time series. In *Signal Processing and Communication Application Conference (SIU)*, pages 2245–2248. IEEE, 2016.
- [140] F. T. Sun, L. M. Miller, and M. D’Esposito. Measuring interregional functional connectivity using coherence and partial coherence analyses of fmri data. *Neuroimage*, 21(2):647–658, 2004.
- [141] W. H. Thompson and P. Fransson. The frequency dimension of fmri dynamic connectivity: Network connectivity, functional hubs and integration in the resting brain. *NeuroImage*, 121:227 – 242, 2015.
- [142] L. Tian, J. Wang, C. Yan, and Y. He. Hemisphere-and gender-related differences in small-world brain networks: a resting-state functional mri study. *Neuroimage*, 54(1):191–202, 2011.
- [143] P. Toivainen, V. Alluri, E. Brattico, M. Wallentin, and P. Vuust. Capturing the musical brain with lasso: Dynamic decoding of musical features from fmri data. *Neuroimage*, 88:170–180, 2014.
- [144] 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*, 15:273–289, 2002.
- [145] D. Van De Ville, T. Blu, and M. Unser. Surfing the brain—An overview of wavelet-based techniques for fMRI data analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):65–78, March-April 2006.

- [146] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage*, 2016.
- [147] S. Vergun, W. Gaggl, V. A. Nair, J. I. Suhonen, R. M. Birn, A. S. Ahmed, M. E. Meyerand, J. Reuss, E. A. DeYoe, and V. Prabhakaran. Classification and extraction of resting state networks using healthy and epilepsy fmri data. *Frontiers in Neuroscience*, 10, 2016.
- [148] J. T. Vogelstein, W. G. Roncal, R. J. Vogelstein, and C. E. Priebe. Graph classification using signal-subgraphs: Applications in statistical connectomics. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1539–1551, 2013.
- [149] J. Wang, P. Liu, M. F. She, S. Nahavandi, and A. Kouzani. Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control*, 8(6):634–644, 2013.
- [150] J. Wang, X. You, W. Wu, M. R. Guillen, M. Cabrerizo, J. Sullivan, E. Donner, B. Bjornson, W. D. Gaillard, and M. Adjouadi. Classification of fmri patterns—a study of the language network segregation in pediatric localization related epilepsy. *Human brain mapping*, 35(4):1446–1460, 2014.
- [151] L. Wang, H. Shen, F. Tang, Y. Zang, and D. Hu. Combined structural and resting-state functional mri analysis of sexual dimorphism in the young adult human brain: an mvpa approach. *Neuroimage*, 61(4):931–940, 2012.
- [152] C.-Y. Wee, P.-T. Yap, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen. Resting-state multi-spectrum functional connectivity networks for identification of mci patients. *PloS one*, 7(5):e37828, 2012.
- [153] 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.
- [154] K. Wu, Y. Taki, K. Sato, H. Hashizume, Y. Sassa, H. Takeuchi, B. Thyreau, Y. He, A. C. Evans, X. Li, et al. Topological organization of functional brain

- networks in healthy children: differences in relation to age, sex, and intelligence. *PloS one*, 8(2):e55347, 2013.
- [155] M. Xia, J. Wang, and Y. He. Brainnet viewer: a network visualization tool for human brain connectomics. *PloS one*, 8(7):e68910, 2013.
 - [156] Z. Xu and A. K. Chan. Encoding with frames in mri and analysis of the signal-to-noise ratio. *IEEE Transactions on Medical Imaging*, 21(4):332–342, 2002.
 - [157] S. Yoshida. Decoding of emotional visual stimuli using fmri brain signal. In *International Conference on Computer and Information Science (ICIS)*, pages 1–4. IEEE, 2016.
 - [158] L. Zhou, L. Wang, L. Liu, P. Ogunbona, and D. Shen. Discriminative brain effective connectivity analysis for alzheimer’s disease: a kernel learning approach upon sparse gaussian bayesian network. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2243–2250. IEEE, 2013.
 - [159] L. Zhou, L. Wang, L. Liu, P. Ogunbona, and D. Shen. Learning discriminative bayesian networks from high-dimensional continuous neuroimaging data. *IEEE transactions on pattern analysis and machine intelligence*, 38(11):2269–2283, 2016.
 - [160] X.-N. Zuo, A. Di Martino, C. Kelly, Z. E. Shehzad, D. G. Gee, D. F. Klein, F. X. Castellanos, B. B. Biswal, and M. P. Milham. The oscillating brain: complex and reliable. *Neuroimage*, 49(2):1432–1445, 2010.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Önal Ertuğrul, İtir

Nationality: Turkish (TC)

Date and Place of Birth: 11.01.1989, Ankara

Marital Status: Married

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Dept. of Computer Engineering, METU, Turkey	2013
B.S.	Dept. of Computer Engineering, METU, Turkey	2011
High School	Ankara Atatürk Anatolian High School	2006

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
Jul2016 - Sep2016	TU Delft, Netherlands	Visiting Ph.D. Student
Sep2011 - Aug2017	Dept. of Computer Eng., METU, Turkey	Research Assistant

PUBLICATIONS

1. **İtir Onal Ertugrul**, Mete Ozay, Fatos T. Yarman Vural. Gender Classification using Mesh Networks on Multi-resolution Multi-task fMRI Data, *Neurocomputing*, 2017. (Under Revision)

2. **Itir Onal Ertugrul**, Mete Ozay, Fatos T. Yarman Vural. Hierarchical Multi-resolution Mesh Networks for Brain Decoding, *Brain Imaging and Behavior*, 2017. (Under Revision)
3. **Itir Onal Ertugrul**, Mete Ozay, Fatos T. Yarman Vural. Encoding the Local Connectivity Patterns of fMRI for Cognitive State Classification. *Pattern Recognition Letters*, 2017. (Under Revision)
4. **Itir Onal**, Mete Ozay, Eda Mizrak, Ilke Oztekin and Fatos Tunay Yarman 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*, 2017.
5. **Itir Onal Ertugrul**, Hamdi Dibeklioglu, What will Your Future Child Look Like? Modeling and Synthesis of Hereditary Patterns of Facial Dynamics. *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2017.
6. Ali Mert Ertugrul, **Itir Onal**, Cengiz Acarturk. Does the strength of sentiment matter? A regression based approach on Turkish social media. *International Conference on Natural Language & Information Systems (NLDB)*, 2017.
7. Arman Afrasiyabi, **Itir Onal**, Fatos T. Yarman Vural, A Sparse Temporal Mesh Model for Brain Decoding, *15th IEEE International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC)*, 2016.
8. **Itir Onal**, Mete Ozay, Fatos T. Yarman Vural, Functional Mesh Model with Temporal Measurements for Brain Decoding, *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
9. **Itir Onal**, Mete Ozay, Fatos T. Yarman Vural, Modeling Voxel Connectivity for Brain Decoding, *International Workshop on Pattern Recognition in Neuroimaging (PRNI)* 2015.
10. **Itir Onal**, Alptekin Temizel, Fatos T. Yarman Vural, Spatial and Temporal Feature Extraction for Brain Decoding using CUDA, GPU Technology Conference (GTC) 2015.

11. **Itir Onal**, Emre Aksan, Burak Velioglu, Orhan Firat, Mete Ozay, Ilke Oztekin, Fatos T. Yarman Vural, A Brain Network for Cognitive State Analysis, *IEEE 23rd Conference on Signal Processing and Communications Applications (SIU)*, 2015.
12. Orhan Firat, Mete Ozay, **Itir Onal**, Ilke Oztekin, Fatos T. Yarman Vural, Enhancing Local Linear Models Using Functional Connectivity for Brain State Decoding, *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 2014.
13. **Itir Onal**, Emre Aksan, Burak Velioglu, Orhan Firat, Mete Ozay, Ilke Oztekin, Fatos T. Yarman Vural, Modeling the Brain Connectivity for Pattern Analysis, *22nd International Conference on Pattern Recognition (ICPR)*, 2014
14. Orhan Firat, **Itir Onal**, Emre Aksan, Burak Velioglu, Ilke Oztekin, Fatos T. Yarman Vural, Large Scale Functional Connectivity For Brain Decoding, *11th IASTED International Conference on Biomedical Engineering (BioMed)*, 2014.
15. Burak Velioglu, Emre Aksan, **Itir Onal**, Orhan Firat, Mete Ozay, Fatos T. Yarman Vural, Functional Networks of Anatomic Brain Regions, *13th IEEE International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC)*, 2014.
16. **Itir Onal**, Ali Mert Ertugrul, Ruken Cakici, Effect of Using Regression on Class Confidence Scores in Sentiment Analysis of Twitter Data, *5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2014.
17. Ali Mert Ertugrul, **Itir Onal**, RemindMe: An Enhanced Mobile Location-Based Reminder Application, *International Conference on Future Internet of Things and Cloud (FiCloud)*, 2014.
18. **Itir Onal**, Emre Aksan, Burak Velioglu, Orhan Firat, Mete Ozay, Ilke Oztekin, Fatos T. Yarman Vural, Estimating Brain Connectivity for Pattern Analysis, *22nd IEEE Conference on Signal Processing and Communications Applications (SIU)*, 2014.

19. **Itir Onal**, Ali Mert Ertugrul, Effect of Using Regression in Sentiment Analysis, *22nd IEEE Conference on Signal Processing and Communications Applications (SIU)*, 2014.
20. Ali Mert Ertugrul, **Itir Onal**, Çeşitli Konum Etiketleme Opsiyonlarıyla Zenginleştirilmiş Yeni Bir Konum Bazlı Hatırlatma Uygulaması, 8. *Ulusal Yazılım Mühendisliği Sempozyumu (UYMS)*, 2014.
21. **Itir Onal**, Mete Ozay, Orhan Firat, Ilke Oztekin, Fatos T. Yarman Vural, An Information Theoretic Approach to Classify Cognitive States Using fMRI, *13th IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, 2013.
22. **Itir Onal**, Mete Ozay, Orhan Firat, Ilke Oztekin, Fatos T. Yarman Vural, Analyzing the Information Distribution in the fMRI measurements by estimating the degree of locality, *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013
23. Orhan Firat, Mete Ozay, **Itir Onal**, Ilke Oztekin, Fatos T. Yarman Vural, Functional Mesh Learning for Pattern Analysis of Cognitive Processes, *12th IEEE International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC)*, 2013. **(Best Paper Award)**.
24. Orhan Firat, Mete Ozay, **Itir Onal**, Ilke Oztekin, Fatos T. Yarman Vural, Representation of Cognitive Processes Using the Minimum Spanning Tree of Local Meshes, *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.
25. **Itir Onal**, Karani Kardas, Yousef Rezaeitabar, Ulya Bayram, Murat Bal, Ilkay Ulusoy, Nihan Kesim Cicekli, A framework for Detecting Complex Events in Surveillance Videos, *3rd IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety (AAMS-PS 2013)*.
26. **Itir Onal**, Mete Ozay, Orhan Firat, Ilke Oztekin, Fatos T. Yarman Vural, Information Distribution Analysis in the fMRI measurements with Degree of Locality Estimation, *IEEE 21th Conference on Signal Processing and Communications Applications (SIU)*, 2013.

27. Orhan Firat, Mete Ozay, **Itir Onal**, Ilke Oztekin, Fatos T. Yarman Vural, Cognitive Processes Representation Using Minimum Spanning Tree of Local Meshes, *IEEE 21th Conference on Signal Processing and Communications Applications (SIU)*, 2013.
28. Orhan Firat, Mete Ozay, **Itir Onal**, Ilke Oztekin, Fatos T. Yarman Vural, A Mesh Learning Approach for Brain Data Modeling, *IEEE 20th Conference on Signal Processing and Communications Applications (SIU)*, 2012.