



מכון טכנולוגי חולון
Holon Institute of Technology

חיזוי מדדים קוגניטיביים מתוך סריקות fMRI ברזולוציה גבוהה באמצעות למידת-מכונה.

Prediction of cognitive measures from high resolution fMRI scans using machine learning

דוח ביניים

מיכאל טולוצ'ינסקי

312272685

יואב פלדמן

300394426

בהנחיית ד"ר ארז סימוני

יולי 2019

תקציר:

המוח האנושי הוא איבר מורכב ומתוחכם, שנים רבות מדענים מתחומים שונים ומגוונים פועלים במטרה להרחיב את הידע וההבנה על פעילותו, ומפתחים ומשפרים את היכולת האנושית לאסוף ולנתח מידע.

fMRI – בזכות ה- fMRI אנו כיום מסוגלים לנתח פעילות מוחית בצורה מרחבית טובה ולנסות להבין כיצד חלקים שונים עובדים ביחד ולחוד בעקבות גירויים חיצוניים לאורך זמן.

מטרת הפרויקט שלנו היא לקדם את הבנה זו ולנסות להמשיך בחקר. עבודה זו מתבססת על נתוני נבדקים שנאספו לאחרונה ובאיכות יוצאת דופן ע"י פרויקט מיוחד של ארגון הבריאות בינלאומי בשם HCP. בעזרת טכניקות סטטיסטיות אנו ננסה להבין מנתונים אלו את תפקוד המוח טוב יותר. בחלק הראשון של העבודה ננסה להבין האם ניתן ליצור טביעת אצבע של נבדק מתוך סריקת fMRI בזמן מנוחה, ובחלק השני נחקור האם סריקת fMRI בזמן צפייה בסרט יכולה לנבא מדדי אינטליגנציה שונים.

Summery:

The human brain is a complicated and sophisticated organ, for many years a variety of researches from different fields, are working in aim of expending our understanding of the way the brain works by developing and improving the technology needed to collect and analysis the required data.

The fMRI is letting us go further and study the brain activity, today it is possible to analyze spatial brain activity and understand how different parts of the brain works separately and together as a response to outside stimulation.

Our project goal is to advance these brain activates understanding. This work is based on high quality data that collected by a project of the international health organization called HCP.

Using statistical techniques, we will try to understand from these data the brain functionality.

In the first part of the work, we will try to understand whether it is possible to create “fingerprint” of subjects from their brain by fMRI scans in resting state, and in the second part we will investigate whether movie task fMRI scans can predict intelligence.

תוכן עניינים

5.....	רציונאל ומטרה	.1
6.....	מבוא	.2
7.....	מילון מונחים	.3
8.....	רקע תיאורטי	.4
8.....	אנטומיה של המוח	.4.1
9.....	MRI	.4.2
9.....	fMRI	.4.3
11.....	רקע מתמטי	.4.4
16.....	סוגי קישוריות	.4.5
16.....	Functional Connectivity	.4.5.1
17.....	Inter-Subject Correlation	.4.5.2
17.....	Inter-Subject Functional Connectivity	.4.5.3
18.....	למידת מכונה	.4.6
20.....	K-nearest neighbors	.4.6.1
21.....	support vector machines	.4.6.2
22.....	מאגר המידע	.5
22.....	HCP	.5.1
23.....	סריקות 7T	.5.2
23.....	מהלך הסריקה	.5.2.1
23.....	rfMRI	.5.2.2
23.....	tfMRI	.5.2.3
23.....	פרוטוקול הסריקה	.5.2.4
24.....	עיבוד מקדים (HCP Processing Pipeline)	.5.3
24.....	פרסלציה לאזורים ורשתות	.5.4
25.....	המחקר :	.6
25.....	פיתוח תשתיות	.6.1
26.....	חלק א' : חיזוי תאומים מתוך סריקות מצב מנוחה	.6.2
26.....	חיזוי מתוך מטריצות קורלציה	.6.2.1
28.....	מהלך המחקר והאופטימיזציות	.6.2.2
28.....	FC network based correlations	.6.2.2.1
28.....	Prediction by multiple matrices	.6.2.2.2
29.....	Monozygotic twins predictions	.6.2.2.3
29.....	7T vs 3T data	.6.2.2.4
31.....	K-Nearest neighbor correlations	.6.2.2.5
31.....	סיכום תוצאות	.6.2.3
31.....	חלק ב' : סיווג רמת אינטליגנציה לפי סריקות בזמן צפייה בסרט - יושלם בהמשך	.6.3
32.....	כלי עבודה	.7
32.....	תוכנות	.7.1
32.....	שפת תכנות	.7.2
32.....	ספריות קוד פתוח	.7.3
33.....	ביבליוגרפיה	.8
33.....	נספחים	.9
33.....	מקורות מידע :	.9.1

רשימת איורים:

8	1 ארבע האונות של המוח
9	2 רשתות המוח
10	3 דוגמה לחלוקת ווקסלים
10	4 יחס בין TR & TE
13	5 פונקציית Sigmoid
16	6 Functional Connectivity
16	7 דוגמה לקורלציה בין אזורים שונים במוח
17	8 Inter-Subject Correlation
17	9 Inter-Subject functional connectivity
18	10 אות ה-BOLD מורכב מאות הנובע מגירוי חיצוני ורעשים
20	11 דוגמה לחשיבות בחירת K מתאים באלגוריתם KNN
21	12 דוגמה ל-SVM
23	13 פרוטוקול הסריקה
25	14 המוח מחולק לרשתות
26	15 מטריצת קורלציה בין תאומים עם סף רגישות של 0.7
26	16 מטריצת קורלציה בין תאומים בלי סף רגישות
28	17 מספר ההצלחות ברשתות השונות במוח
30	18 תוצאות ההצלחות השונות ב-7T
30	19 תוצאות ההצלחות ברשתות בשונות ב-3T
31	20 סיכום תוצאות חיזוי תאומים באמצעות FC

1. רציונאל ומטרה

בפרויקט זה נבדוק את הקישוריות בין אזורים במוח הצפויים להיות שונים בין אדם לאדם, ונחפש מדדים קורלטיביים בין קישוריות זו לטביעת האצבע המוחית של נבדקים שתאפשר לזהות קשרים גנטיים ולנבא מדדים קוגניטיביים אצל נבדקים שביצעו סריקת fMRI. לצורך כך נבחן ונפתח שיטות סטטיסטיות שיאפשרו לפתור את הבעיה. בפרויקט זה נעבוד על מידע סריקות ברזולוציית T7 שנאסף בשנים האחרונות במסגרת פרויקט Human Connectome ונחשב למידע האיכותי ביותר שקיים כיום.

מטרות עיקריות:

1. פיתוח יכולת זיהוי האם אדם מסוים הוא תאום של אדם אחר או לא בהינתן סריקות fmri בזמן מנוחה באמצעות חישוב קישוריות פונקציונאלית (FC) במוח.
2. פיתוח מסווג רמות אינטליגנציה נזילה באמצעות חישוב קישוריות פונקציונאלית בין מוחית (ISFC).

2. מבוא

מיפוי מוח משמש למדידת פעילותם של אזורים במוח לצרכים שונים, כגון דיאגנוסטיקה ומחקר. fMRI הוא תחום דימות ספציפי לדימות פעילות המוח, שבו בעזרת שדות אלקטרומגנטיים חזקים אנו מצליחים להבדיל בין רמות פעילות שונות בדרך לא פולשנית. ההבדלה בין רמות הפעילות השונות נעשית ע"י הבדלה בין רמות החמצון השונות, אלו עוזרות לנו לזהות אזורים בהם הדם מחומצן יותר במוח ואזורים בעלי רמת חמצון נמוכה יותר. כך בצורה עקיפה מתקבלת אינדיקציה על פעילות נוירולוגית באותו אזור.

ככל שהאזור פחות מחומצן משמעות הדבר שכרגע התאים באותו אזור צורכים יותר אנרגיה ולכן נדרש יותר חמצן. כך אנו למעשה מבצעים מיפוי שלם של אזורים פעילים במוח. מיפוי זה מאפשר לבדוק האם יש קשר בין אזורים שונים והאם ניתן למצוא קורלציה בין ערכי רמת הפעילות שלהם (BOLD), מתאם זה מצביע על קישור פעיל ביניהם (Connectivity Functional). קישוריות זאת היא בעלת משמעות עמוקה, כיוון שממנה ניתן ללמוד כיצד המוח מחובר מבחינה תפקודית.

בשנים האחרונות נעשים מאמצים רבים בחקר והבנה של רשתות המוח השונות. בעבודה זו אנו נתמקד בעיקר ברשת הנקראת רשת ברירת המחדל - Default Mode Network (DMN). ממחקרים שנעשו על רשת זו עולה שהיא קשורה בעיקר למצבי מנוחה, בהם אין פעילות ספציפית ואדם מהרהר, עם זאת מחקרים אחרונים מראים שגם לרשת זו יש קשר לפעילויות ספציפיות שאנו עושים כגון, גישה לרגשות ומחשבות.

בפרויקט זה נפתח יכולות לימוד מסריקות שונות של פעילות מוחית. ננסה לחזות מתוך הסריקות דמיון גנטי בין נבדקים ונשווה את תוצאות החיזוי לאמת בפועל.

עיקר העבודה בפרויקט נעשית על ידי שימוש באלגוריתמים של לימוד מכונה וכלים הסתברותיים. במסגרת פרויקט זה קיימים אתגרים (אשר חלקם שכיחים בתחום מחקר זה), עיקרי האתגרים: התמודדות עם כמויות של מידע רב, מידע מצומצם מבחינת מספר הנבדקים יחסית לנדרש באלגוריתמים של למידת מכונה.

פרויקט זה מבוסס על מידע דימות מסוג fMRI שנלקחו במסגרת פרויקט בשם human connectome. המידע מכיל סריקות FMRI של כ-1200 נבדקים שעברו שני סוגים של סריקות: באחת הם נמצאים במצב מנוחה ללא כל פעילות ובשניה הם צופים בסרט תוך כדי הסריקה. בין הנבדקים, מצויים כ-150 זוגות של תאומים בהם אנו נתמקד במהלך הפרויקט.

את המידע הזה אנו ננתח באמצעות שיטות מוכרות לחישוב קישוריות מוחית: FC, ISC, ISFC. על המידע המנותח אנו נפעיל אלגוריתמי למידת מכונה שונים על מנת להצליח לחזות בדיוק גבוהה עבור נבדק מסויים מיהו התאום הזהה שלו.

3. מילון מונחים

SNR – Signal to Noise Ratio

MRI –Magnetic Resonance Imaging

fMRI – Functional Magnetic Resonance Imaging

rfMRI - Resting state fMRI

tfMRI - Task based fMRI (Movie Task)

SR- Spatial Resolution

TR - Temporal Resolution

BOLD - Blood-Oxygen-Level Dependent

Voxel – Volume Element

DMN – Default Mode Network FC- Functional Correlation

FC - Functional connectivity

ISC - Inter-Subject Correlation

ISFC- Inter-Subject Functional Correlation

HCP - Human Connectome Project

Grayordinates - voxels of the gray-matter component in the brain

CIFTI - (Connectivity Sub-committee of the Neuroimaging Informatics Technology Initiative) - fMRI processed data file with grayordinates voxels

ספין - דרגת חופש פנימית של חלקיקי האטום

חומר אפור - סוג חומר במוח שתפקידו לנתב מידע על גירויים חושיים או תנועתיים אל תאים שונים של מערכת העצבים

חומר לבן - סוג חומר במוח האחראי לקשר תאים מסוג חומר אפור

פרסלציה - מיצוע מידע ממספר ווקסלים המהווים אזור מסוים ומוכר במוח.

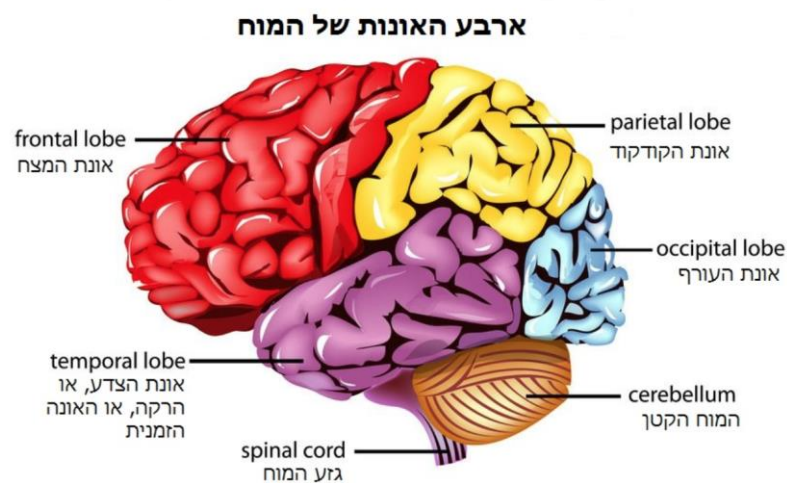
מטריצת קורלציות - מטריצה המכילה קורלציות בין כל זוג נבדקים הקיימים במאגר. על כן, זוהי מטריצה סימטרית שהאלכסון בה מתאר את קורלציות בין כל נבדק לעצמו ולכן שווה 1.

4. רקע תיאורטי

4.1. אנטומיה של המוח

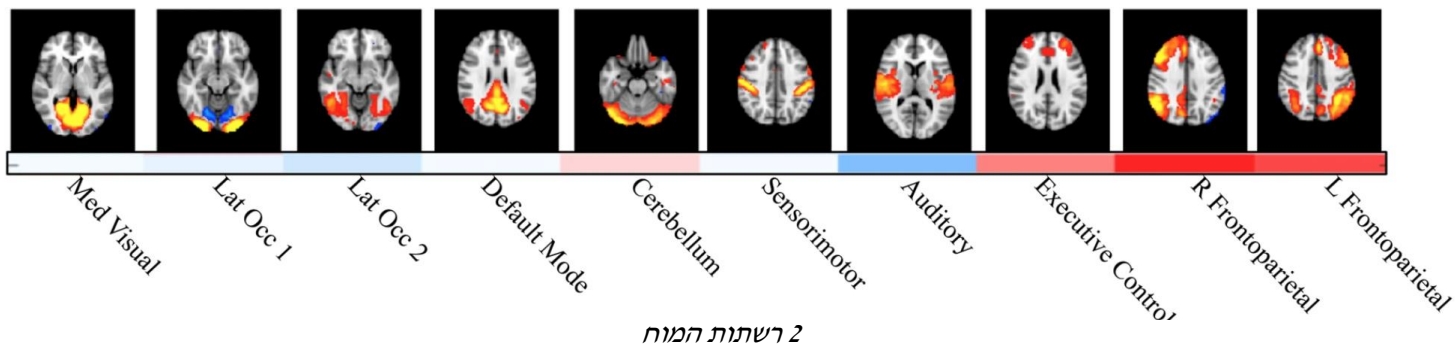
מוח הוא האיבר הראשי במערכת העצבים האחראי לתאם את הפעילות של מערכות הגוף השונות. תפקידו של המוח הוא לקבל מידע ממערכות החושים, לעבד אותו ולהשתמש בו על מנת לנהל את התנהגותו של בעל החיים המפותח. המוח האנושי מורכב משלושה מבנים עיקריים: המוח הגדול (Cerebrum), התופס את מרבית נפח הגולגולת והוא מחולק לשני המיספרות, הימנית והשמאלית. ביניהם מקשר כפיס המוח (Corpus Callosum). גזע המוח (Brainstem), הנמצא בחלקו התחתון של המוח ומקושר לחוט השדרה. והמוח הקטן (Cerebellum), היושב מאחורי גזע המוח ומתחת למוח הגדול. מבחינה תפקודית, גזע המוח שולט על תפקודים בסיסיים של הגוף, כמו נשימה ועוררות. והוא מתפקד כנתיב להעברת מידע חושי ופקודות מוטוריות בין המוח לחוט השדרה. המוח הקטן אחראי בעיקר על תיאום מדויק של פעולות מוטוריות. עיקר העיבוד של המידע מהחושים וביצוע הפעולות הרצוניות של הגוף נמצאים תחת שליטתו של המוח הגדול, על שלל מרכיביו. בחלוקה גסה, רקמת המוח נחלקת לשני סוגים עיקריים - החומר האפור המכיל ריכוזים של גופי תאי עצב. והחומר הלבן המורכב ברובו מאקסונים המחוברים בין האזורים השונים של חומר אפור במוח. מלבד הרקמה התאית, המוח מכיל גם מערכת של חללים הנקראים חדרי המוח. חללים אלו מלאים בנוזל מוחי שדרתי (CSF) אשר נמצא גם בין קרומי המוח ובתעלת חוט השדרה.

המוח הגדול מתחלק לארבע אונות. אונת המצח (פרונטלית), אונת הרקה (הטמפורלית) אונת העורף (אוקסיפיטלית) ואונת הקודקוד (פריאטלית).



1 ארבע האונות של המוח

בשנים האחרונות נעשה ניסיון רב לחקור את רשתות הקישור השונות במוח. רשתות קישור הן רשת של מספר אזורים במוח שיוצרים קשר אחד עם השני בזמנים מסוימים על מנת לבצע פעולה. את רשתות קישור אלה מוצאים, בין השאר, בעזרת סריקת fMRI.



MRI.4.2

דימות תהודה מגנטית - **Magnetic Resonance Imaging**, הוא סוג של סריקה לא-פולשנית המשמשת להמחשת איברים פנימיים בגוף ללא חשיפה לקרינה מייננת, למטרות אבחון רפואי, חקר המוח, פסיכיאטריה, מחקר ביולוגי ועוד. בשיטה זו משתמשים בשילוב של קרינת RF ושדה מגנטי חזק (גבוה מ-1 טסלה) כדי להשפיע על כיוון הספין של הפרוטונים בגרעין אטומי המימן שנמצאים במולקולות המים שבגוף. במהלך הבדיקה הנבדק מצוי בתוך שדה מגנטי קבוע שמיישר את כיוון הספין בין כל הפרוטונים. בעזרת שליחת פולס של גלי רדיו נוצרת תנועת נקיפה (פרסציה) של ספין הפרוטון (יישור הפאזה והטיית כיוון הספין). הדבר גורם לפרוטון ליצור תהודה (רזוננס) הניתנת למדידה בעזרת סלילי הקלט של המכונה וע"י כך ניתן לקבל הפרדה ברורה בין הרקמות השונות. שינוי תכונות הספין תלוי ברקמה בה הוא נמצא (למשל רקמת חומר הלבן או רקמת חומר אפור שבמוח).

fMRI.4.3

דימות תהודה מגנטית תפקודי או **Imaging Resonance Magnetic Functional**. היא תת שיטה של MRI שבה מודדים פעילות מוחית. בניגוד ל-MRI ה-fMRI אינו מראה תמונה סטטית של מבנה הרקמות אלא תמונות רבות הנלקחות בתדירות גבוהה המשקפות פונקציונליות של הרקמות. שיטה זו מתבססת על התכונות המגנטיות של החלבון - המוגלובין, שתפקידו להוביל מולקולות של חמצן/פחמן דו חמצני. המוגלובין מחומצן מגיב שונה מהמוגלובין לא מחומצן, לכן בדרך זו אפשר להבחין בין מצבים שונים של המוגלובין וכך לדעת אם אזור מסוים במוח צורך יותר חמצן, מדד זה נקרא **Blood-Oxygen-Level Dependent** או BOLD בקיצור. תוצאת סריקת fMRI אינה מראה פעילות חשמלית באופן ישיר אלא בצורה עקיפה כפי שבאה לידי ביטוי על ידי ה BOLD, אשר נמדד ברזולוציה של כ-2 שניות.

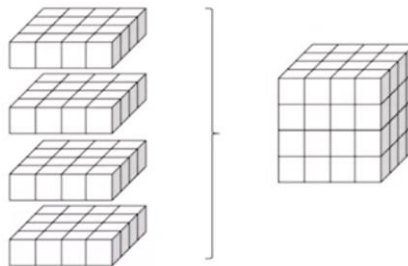
פרמטרים חשובים של סריקת fMRI :

Resolution Temporal - רזולוציה במרחב הזמן, מתארת את היכולת להפריד בין אירועים במוח המשתנים בזמן.

Resolution Spatial - רזולוציה מרחבית, מתארת את היכולת להבדיל בשינויים בין אזורים מרחביים בתמונה.

Slice - חלוקת המוח למספר פרוסות אופקיות המרכיבות תמונה תלת ממדית מפוקסלת של המוח. פרוסות אלו נסרקות ע"י הסורק אחת אחרי השנייה.

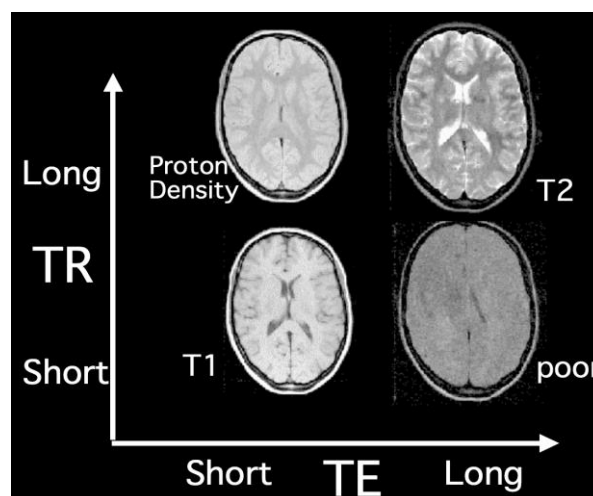
Voxel חלוקת כל slice לתת יחידות שקרויות ווקסלים (איור 3). זהו למעשה פיקסל במרחב תלת-מימדי. כל תמונה של fMRI מכילה בערך כ 100,000 ווקסלים.



3 דוגמה לחלוקת ווקסלים

TR - Repetition Time - זהו הזמן בין פולסי RF. זמן TR ארוך מאפשר לפרוטונים ברקמות להתיישר בחזרה עם השדה המגנטי הראשי. TR קצר יגרום לחלק מהפרוטונים ברקמות לא להספיק להתיישר עם השדה המגנטי לפני שתבצע המדידה הבאה ולכן מקטין את הסיגנל מאותה רקמה.

TE - Echo Time - זמן שהיה בין רגע הפולס RF לבין רגע מדידת הסיגנל הנוצר מהתיישרות הספין של הפרוטון לכיוון אנכי. לאחר הפולס, הפרוטונים מאבדים פאזה עם הזמן, ולכן TE ארוך גורם להנחתת האות מהרקמות בעקבות שינוי הפאזה של חלק מהפרוטונים. TE קצר מוריד את הסבירות לשינוי פאזה של הפרוטונים.



4 יחס בין TE & TR

לסורקי fMRI ישנן עוצמות מגנט שונות. כיום נפוצים סורקים עם מגנט של T3 או T7. מחקרים רבים משווים את התוצאות המושגות מכל גודל מגנט והוכח שבסריקות עם T7 ניתן לסרוק ברזולוציה גבוהה יותר (חלוקה ליוצר voxels) ולקבל SNR גבוה יותר מבT3. מצד שני, סריקה בT7 רגישה יותר לתופעות לוואי בלתי רצויות במהלך הסריקה כגון תזוזות ראש של הנבדק והפרעות של תדרים נמוכים. עם תופעה זו ניתן להתמודד באמצעות עיבוד מקדים של המידע. בפרויקט זה אנו נתמקד בעיקר בסריקות של T7 מתוך ההנחה שניתן למצות מהן יותר מאפיינים של הנבדק.

כיום נפוצים שני סוגים עיקריים של קואורדינטות למיפוי המוח המקובלים בתחום סריקות fMRI :

1. MNI - מערכת קואורדינטות ארבעה מימדית (x,y,z,t) כאשר 3 המימדים הראשונים מייצגים מיקום במוח (נקודה במרחב תלת מימדי) וt מייצג נקודת זמן. לשיטה זו כמה חסרונות בולטים :

- a. המיפוי מכיל גם את אזורי החומר הלבן כאשר לרוב הם אינם רלוונטיים לאנליזה.
- b. בעקבות כך קבצי המידע גדולים מאוד עבור כל סריקה, ולכן קשה לעבוד איתם.

2. Grayordinates - מיפוי שפותח בשנים האחרונות. מיפוי זה מכיל את אזורי החומר האפור בלבד ומכיל מידע של surface and volume בקובץ אחד. פורמט קבצי המידע נקרא CIFTI הוא מכיל מטריצה דו מימדית (x,v) כאשר מימד אחד מייצג ווקסל מסוים והמימד השני מייצג נקודת זמן. קבצי הCIFTI תופסים משמעותית פחות מקום בזיכרון המחשב מאשר קבצי MNI. בעבודה זו נשתמש רק בקבצים בפורמט זה.

4.4. רקע מתמטי

1. תוחלת

$$E[x] = \sum_{i=1}^{\infty} x_i \cdot P_i$$

התוחלת מייצגת את הערך אליו שואפת התוצאה הממוצעת של ניסוי כשמספר הניסויים שואף לאינסוף פעמים.

2. סטיית תקו

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$$

סטיית התקן היא מדד סטטיסטי לתיאור הפיזור של ערכי קבוצת נתונים סביב הממוצע שלהם.

3. שונות

$$\text{var}(x) = E((x - \mu)^2) = E(x)^2 - \mu^2$$

שונות היא מדד לפיזור ערכים באוכלוסייה הנתונה ביחס לתוחלת שלה

4. קרוס קורלציה

$$\text{Corr}(x, y) = E[x \cdot y]$$

קרוס קורלציה התוחלת המשותפת של שני משתנים שונים. והיא מדד לדמיון ביניהם.

5. שונות משותפות

$$\text{Cov}(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

שונות משותפת היא מדד לקשר בין שני משתנים מקריים. השונות המשותפת חיובית כאשר המשתנים נוטים לסטות באותו כיוון, מעל או מתחת לממוצע, ושלילית כאשר הם משתנים בכיוונים מנוגדים זה לזה.

עבור משתנים חסרי קורלציה מתקיים $\text{Cov}(y, x) = 0$

6. מתאם פירסון

$$\rho_{x,y} = \frac{E((x - \mu_x)(y - \mu_y))}{\sigma_x \sigma_y}$$

מתאם פירסון (קורלוציה) הוא מדד למתאם לינארי בין שתי קבוצות של מספרים. כאשר מדובר בעיבוד נתונים סטטיסטי, ההתייחסות היא בדרך כלל לקשר סימטרי בין שני משתנים. ערכי המדדים נעים בין (-1) לבין (+1) והם מסומנים באות R או ב- ρ .

7. חוק בייס

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

חוק בייס הוא תוצאה בתורת ההסתברות המאפשרת לחשב הסתברות מותנית של מאורע כאשר יודעים את ההסתברויות המותנות ההפוכות.

8. נוסחא להסתברות שלמה

$$P(B) = P(B|A) \cdot P(A) + P(B|A^C) \cdot P(A^C)$$

היא אחת הנוסחאות האלמנטריות בתורת ההסתברות. היא מאפשרת לחשב הסתברות של מאורעות מסובכים, על ידי פירוק מרחב ההסתברות למרכיבים זרים, וחישוב ההסתברות בכל אחד מהם בפני עצמו.

9. גרדיאנט

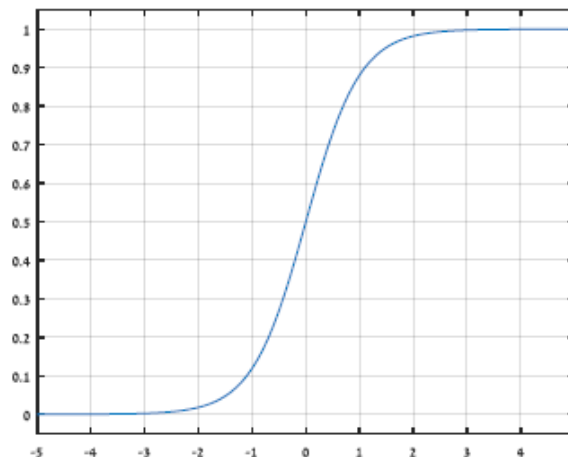
$$\nabla \stackrel{\text{def}}{=} \hat{x} \frac{\partial}{\partial x} + \hat{y} \frac{\partial}{\partial y} + \hat{z} \frac{\partial}{\partial z} \stackrel{\text{def}}{=} \partial_x + \partial_y + \partial_z$$

גרדיאנט הוא הכללה של מושג הנגזרת בעבור חשבון אינפיניטסימלי של מספר משתנים. הגרדיאנט הוא אופרטור וקטורי המופעל על שדה סקלרי. הגרדיאנט של שדה סקלרי הוא שדה וקטורי המשייך לכל נקודה במרחב וקטור. כיוון ווקטור הגרדיאנט מצביע אל הכיוון בו השינוי בשדה הסקלרי מקסימלי (חיובי) גודל ווקטור הגרדיאנט כשיעור השינוי המקסימלי.

10. פונקציית Sigmoid

$$S(t) = \frac{1}{1 + e^{-t}}$$

פונקציית ה Sigmoid מעבירה כל מספר ממשי לתחום שבין 0 ל-1. ניתן לפרש זה כהסתברות.



5 פונקציית Sigmoid

11. פונקציית Softmax

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

פונקציה זו היא הכללה רב מימדית של פונקציית Sigmoid.

12. mean square error

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

טעות ריבועית ממוצעת של אומד מודדת את ממוצע ריבועי ה"טעויות", כלומר ההבדל בין האומד לבין מה שנאמד. ההבדל נובע בגלל האקראיות או בגלל שהאומד לא לוקח בחשבון מידע שיכול לייצר אומד מדויק יותר.

13. מרחק אוקלידי

$$d(x, y) = \|x, y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

14. התפלגות נורמאלית

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

התפלגות גאוסית היא פונקציה סימטרית סביב התוחלת, ובעלת שתי נקודות פיתול במרחק סטיית תקן אחת מן הממוצע, כלומר בנקודות $\mu \pm \sigma$ את העובדה שמשתנה X הוא בעל התפלגות כזו, מקובל לציין בסימון $X \sim N(\mu, \sigma^2)$

15. התפלגות נורמאלית רב מימדית

$$f_x(x_1, \dots, x_k) = \frac{e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}}{\sqrt{(2\pi)^k |\Sigma|}}$$

16. התמרת הילברט

$$\hat{u}(t) = H(u)(t) = \int_{-\infty}^{+\infty} u(\tau)h(t - \tau)d\tau = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{u(\tau)}{t - \tau} d\tau$$

התמרת הילברט לוקחת פונקציה במרחב הזמן, ומשאירה אותה במרחב הזמן, בערך המשובב ב 90° בתדר כמו כן שימוש בתמרת הילברט משמש למציאת מעטפת של אות.

17. התמרת פורייה

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t} dt$$

התמרת פורייה היא כלי מרכזי באנליזה הרמונית שאפשר לתארו כפירוק של פונקציה לרכיבים מחזוריים (סינוסים וקוסינוסים או לחלופין אקספוננטים מרוכבים) וביצוע אנליזה מתמטית לפונקציה על ידי ניתוח רכיביה. כמו כן היא מעבירה למישור התדר.

18. התמרת פורייה בדידה

$$X_k = \sum_{n=0}^{N-1} X_n e^{-\frac{2\pi i k n}{N}}, K \in \mathbb{Z}$$

19. בלתי תלוי סטטי

$$P(A \cap B) = P(A) \cdot P(B)$$

שני מאורעות הם בלתי תלויים אם ההסתברות ששניהם יקרו שווה למכפלת ההסתברויות של כל מאורע בנפרד.

20. פונקציית מחיר

$$cost = J = \frac{1}{2m} \sum_{i=1}^m \left(h_{w(x^{(i)})} - y^{(i)} \right)^2$$

פונקציה זו מייצגת את "מחיר" הטעות בקלסיפקציה, כלומר כמה האלגוריתם צדק בחיזוי. וגורם הלמידה מנסים להוריד אותה למינימום.

21. פונקציית היפותזה

$$h_w(x^{(i)}) = \frac{1}{1 + e^{-w^T \cdot x^{(i)}}}$$

בתחום למידת המכונה פונקציה זו משמשת לקישור בין משתנים כאשר הקשר ביניהם לא ידוע ונדרש להעריך על מנת לקבל את הפונקציה המתאימה ביותר.

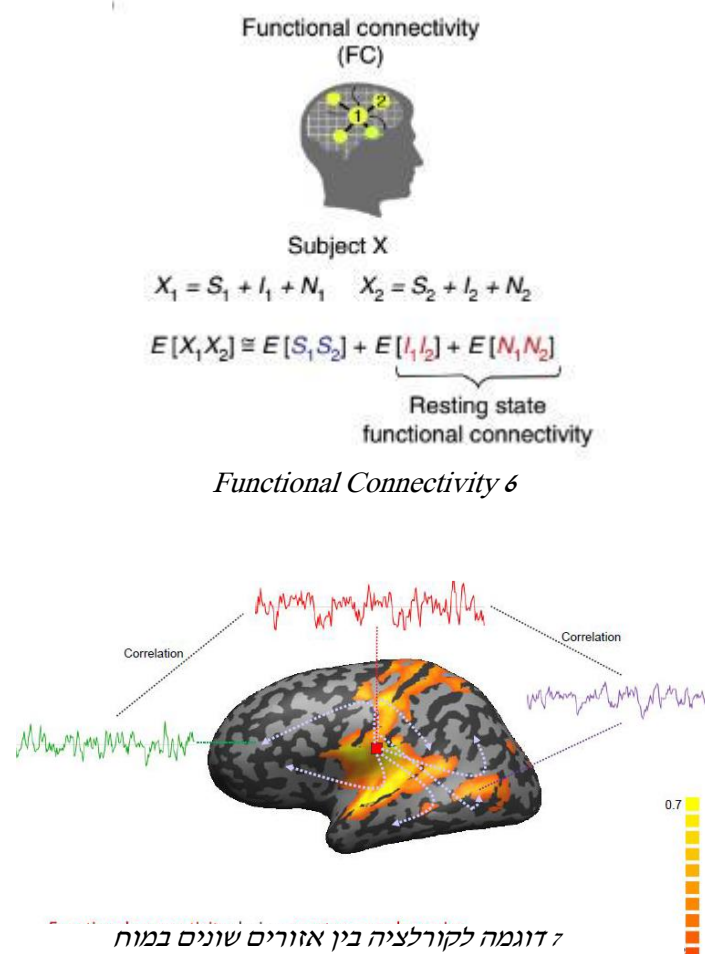
4.5. סוגי קישוריות

Functional Connectivity.4.5.1

קישוריות תפקודית (איור 6) – היא תחום שהתפתח באמצע שנות ה-2000 שבו בודקים את הקורלציה בזמן בין ערכי ה-BOLD של ה seed מסוים במוח של נבדק בודד, לבין ערכי ה-BOLD של כל האזורים במוחו של אותו נבדק.

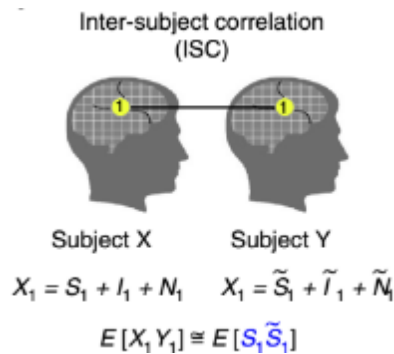
כתוצאה מקבלים מטריצה המתארת מפה של קישוריות בין האזורים השונים במוח.

החיסרון המרכזי של שיטה זו היא שיש תלות רבה במחשבות שאינן קשורות בגירוי החיצוני שאליו נחשף הנבדק. כגון הרהורים יום-יומיים, אולם בשיטה זו אנו מרווחים פתח אל תוך הרהורים אלו.



Inter-Subject Correlation.4.5.2

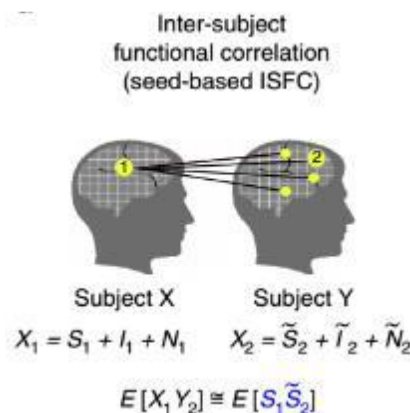
בשיטה זו מחשבים קורלציה של ערכי BOLD בזמן בין seed ספציפי במוח לאותו ה-seed במוחו של אדם אחר(איור 8). וכך חוזרים על פעולה זו עבור שאר ה-seeds. גם בשיטה זו אנו מקבלים מטריצה המתארת מפה של קישוריות בין אזורים שונים במוח, יחסית לגירוי החיצוני. יתרון של שיטה זו על פני שיטת ה-FC, היא סינון הרעש האינטרניזי ורעשי המדידה השונים וכתוצאה שיפור יחס אות לרעש. אולם בשיטה זו אנו מאבדים מידע המאפיין את הנבדק הספציפי.



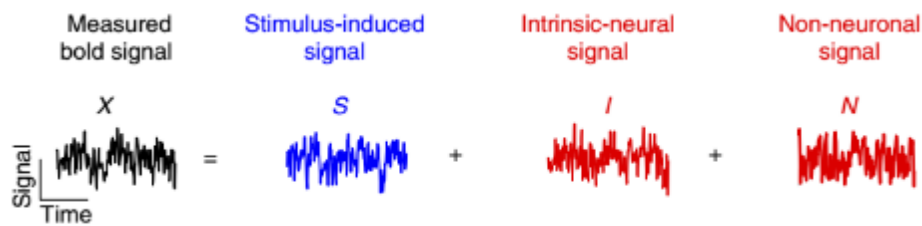
Inter-Subject Correlation 8

Inter-Subject Functional Connectivity.4.5.3

טכניקה זו היא שילוב של שתי השיטות הקודמות(איור 9). בשיטה זו מחשבים קורלציה של ערכי BOLD בזמן בין seed ספציפי במוח לשאר האזורים במוחו של אדם אחר. גם בשיטה זו אנו מקבלים מטריצה המתארת מפה של קישוריות בין אזורים שונים במוח יחסית לגירוי חיצוני. יתרון של שיטה זו על פני שיטת ה-FC היא סינון הרעש האינטרניזי ורעשי המדידה השונים וכתוצאה שיפור יחס אות לרעש ויתרונות על ISC היא קבלת מידע נוסף. היתרונות של ISC ו-ISFC על פני FC הן סינון רעשים אינטרזיים מחשבות פנימיות שלא קשורות לגירוי וכתוצאה מכך יחס האות לרעש (SNR) גבוהה יותר.



Inter-Subject functional connectivity 9



10 אות ה-BOLD מורכב מאות הנובע מגירוי חיצוני ורעשים

4.6. למידת מכונה

למידת מכונה היא תת-תחום ממדעי המחשב ובינה מלאכותית אשר התפתח והפך להיות מולטידיסציפלינרי. הוא הוגדר על ידי ארתור סמואל: "תחום מחקר המאפשר למחשבים את היכולת ללמוד ללא להיות מתוכנתים באופן ספציפי".

באמצעות אלגוריתמי למידת מכונה אנו יכולים "להציג" למחשב נתונים שונים ולתת לו לנתח אותם להגיע לתוצאות. במקרה שלנו לסווג אותם למקומות הנכונים.

ישנם מספר מונחים חשובים לענף זה.

מסווג – אלגוריתם אשר "לומד" את הנתונים השונים ו"מחליט" לאיזה קבוצה הנתון הספציפי שייך. מסווג הינו פונקציה המעבירה ממרחב הקלט למרחב ההחלטות.

מאפיין – לכל דוגמה יש מספר מאפיינים אשר לפיה היא מזוהה, לדוגמה: בית מזוהה בין היתר לפי מספר חדרים ולפי גודל השטח שלו.

סדרת לימוד – זוהי סדרת הדוגמאות המתויגות/המסווגות. $\{x_k, y_k\}$ כאשר y_k הוא הסיווג הנכון של תבנית הקלט x_k . בתחום זה חשוב שיהיה מספר דוגמאות נרחב בכדי שיתאפשר בניית מסווג אשר יסווג כל קלט חדש למחלקה המתאימה עם שגיאה קטנה ככל האפשר.

כלל אצבע הוא לפחות עשר דוגמאות על מאפיין אחד.

סוגי האלגוריתמים שאנו נתמקד בהם כאן נקראים למידה מונחת - כל דוגמה מגיעה ביחד עם תווית סיווג. מטרת האלגוריתם היא לחזות את הסיווג של דוגמאות חדשות שאותן לא פגש בתהליך הלמידה. לסוגי אלגוריתם זה יש שני שלבים ראשיים:

לימוד – שלב בו המכונה רק לומדת לא מחליטה/מסווגת שום דבר.

מסקנות – שלב בו למכונה כבר יש ידע מספק על מנת לסווג את הדוגמאות החדשות.

אולם, קיימות מספר בעיות בסיסיות הקשורות ללמידה המונחית.

בחירת אופי המודל המתאים לבעיה מסוימת הינה בעיה ללא פתרון חד משמעי. ובחירת המודל בדרך כלל מסתמך על ידי השוואה לבעיות דומות, ניסיון עבר וכדומה. בחירת סדר המודל גם היא בעיה שאין לה פתרון והיא מתייחסת למספר הפרמטרים שהמודל ישתמש. מודל פשוט מדי עלול לא לתאר את הקשר בין הקלט לפלט בצורה טובה. בעוד שמודל מסדר גבוה מדי מגדיל מאוד את סיבוכיות המודל ואת מספר של סדרת הלימוד הנדרש.

הניגוד בין שני קטבים אלה קרוי הדילמה בין הטיה לבין שונות (Tradeoff Variance Bias).

ההטיה מתארת את המרחק המינימלי האפשרי בין פונקציה כלשהי מתוך המודל לפונקציה הנלמדת האמתית. ההטיה תקטן ככל שסדר המודל יעלה.

השונות מתארת את המרחק בין הפונקציה האופטימאלית מתוך המודל לבין זו שהתקבלה בפועל לאחר התאמה לסדרת הלימוד. השונות תגדל ככל שסדר המודל גבוה יותר.

בכדי לנסות ולהתמודד עם בעיה זו משתמשים בגישת הרגולציה הפרמטרית. אשר מוסיפה לשגיאה

האמפירית איבר נוסף אשר מגדיל את שגיאה כאשר הפרמטרים של המודל גדולים מדי ובכך מאלץ את המודל להיות בעל פרמטרים קטנים ככל האפשר. וכתוצאה מתאפשרת שליטה בפשרה שבין גודל

הפרמטרים להתאמה לסדרת הלימוד.

בעיה נוספת הקשורה במודלים בעלי סדר גבוה היא התאמת יתר (overfitting) כלומר התאמה טובה מדי של המודל לסדרת הלימוד, מה שגורם לחוסר כלליות של המודל ולחוסר התאמה לדוגמאות חדשות.

בכדי להימנע מבעיות אלו אנו מחלקים את המידע שברשותנו לתת הקבוצות הבאות :

סדרת הלימוד (Set Training) - משמשת לכיוון הפרמטרים, בדרך כלל על ידי מזעור שגיאה אמפירית.

סדרת האימות (Set Validation) - משמשת לבחירת מודל וסדר המודל.

סדרת הבוחן (Set Test) - משמשת להערכת ביצועים סופית.

החלוקה לקבוצות אלו מתבצעת באופן אקראי.

כאשר מספר הדוגמאות הוא קטן ואין מספיק מידע עבור חלוקה נכונה לתת קבוצות, נהוג להשתמש בשיטת אימות צולב (Validation-Cross).

בגישה זו קבוצת הדוגמאות מחולקת ל-K קבוצות זרות. עבור כל אחת מקבוצות אלו מתבצע שלב לימוד, פרט לקבוצה אחת. ולאחר מכן מחושבת השגיאה האמפירית על קבוצה זו. חוזרים

על תהליך זה K פעמים ובכל פעם משנים את הקבוצה שעליה מחשבים את השגיאה האמפירית. לבסוף מחושבת השגיאה הכוללת כסכום של K

השגיאות האמפיריות. כאשר K שווה למספר הדוגמאות הכולל, שיטה זו מכונה "אשאר אחד בחוץ" (Leave One Out).

במסגרת פרויקט זה, התמקדנו בכמה אלגוריתמי סיווג מתחום הלמידה המונחית, אשר מותאמים יותר ליחס הדוגמאות למאפיין של המידע שברשותנו. נתאר אותם בקצרה.

K-nearest neighbors.4.6.1

הוא אלגוריתם סיווג "עצלן" שבו הפונקציה מקורבת באופן מקומי בלבד וכל החישובים נדחים עד סיווגה. אלגוריתם זה פורס את המאפיינים על פי מרחק אוקלידי (איור 11) (אופציה אחת מבין רבות ומחליט האם הם שייכים לאותה קבוצה או לא על סמך K הדגימות "הקרובות" ביותר).

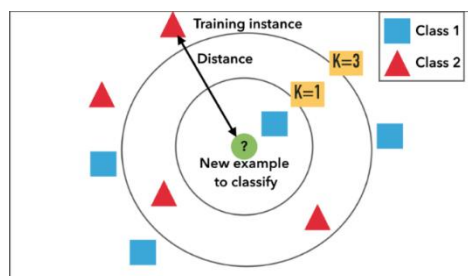
כיצד יבחר ה- K ? הבחירה הטובה ביותר של K תלויה בנתונים. בדרך כלל, ערכים גבוהים יותר של K גורמים לצמצום ההשפעה של הרעש על סיווג, אבל גורמים לגבולות בין המחלקות להיות פחות מובהקים. K טוב יכול להיבחר באמצעות מספר שיטות. במקרה המיוחד בו נחזתה מראש סוג המחלקה כמחלקה הקרובה ביותר לנקודות האימון ($1=K$) נקרא אלגוריתם השכן הקרוב ביותר. הדיוק של אלגוריתם KNN יכול להיפגע קשות על ידי נוכחות של רעש או תכונות לא רלוונטיות, או במידה וסקלת התכונה אינה עקבית עם חשיבותה. מאמצי מחקר רבים הושקעו עבור בחירת תכונות או דירוג תכונות לשיפור סיווגם.

במצב בינארי (כאשר יש שתי מחלקות) בעיות סיווג, כדאי לבחור את K להיות מספר אי-זוגי כדי להימנע ממצבי תיקו. דרך אחת פופולרית לבחירת K אופטימלי באופן אמפירי למצב זה היא באמצעות שיטת אתחול.

לדוגמא ישנם מספר צורות: מרובע ומשולש. אני מעוניין לסווג את הצורה.

האלגוריתם יפרוס את הצורות השונות במרחב. המספר K קובע כמה דוגמאות יכנסו למעגל ההחלטה.

כאשר $K=1$ הוא יבחר את הדוגמאות הקרובות ביותר ויחליט שהוא הכי דומה להן ולכן יסווג לקבוצה שלה.



11 דוגמה לחשיבות בחירת K מתאים באלגוריתם KNN

support vector machines.4.6.2

מכונת וקטורים תומכים (SVM) הוא אחד מן הכלים הפופולריים כיום בתחום למידת המכונה, הודות לביצועיו בפתרון בעיות באמצעים סטטיסטיים (איור 12).

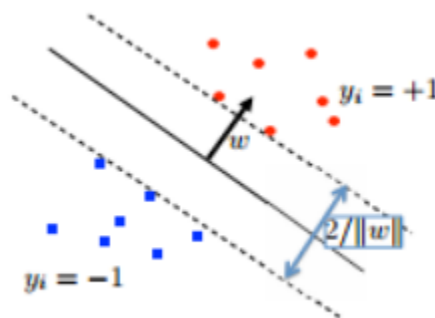
דוגמאות האימון מיוצגות כווקטורים במרחב לינארי והוא עצמו מסווג לינארי.

תחילה נתייחס אל המקרה הבינארי שבו יש רק 2 מחלקות ולאחריו נרחיב למקרה הכללי.

ה-SVM מוצא מישור-על (hyperplane) לינארי שאמור להפריד בין 2 סוגי הדוגמאות. כך שבכל צד של המישור יהיו כמה שיותר דוגמאות בעלי אותו התיוג (label) ועוד שני מישורי על המקבילים לו, הנמצאים אחד מכל צד.

המרחק בין שני מישורי העל המקבילים מכונה השוליים (margin) או מרחק הביטחון.

דוגמאות האימון המתלכדות עם מישורי השוליים נקראות וקטורים תומכים, ומכאן גם נגזר שמו של האלגוריתם.



12 דוגמה ל-SVM

אלגוריתם הלמידה מנסה לספק שתי מטרות: האחד, מקסום השוליים והשני סיווג נכון של דוגמאות האימון.

שתי מטרות אלו נמצאות בסתירה האחת לשנייה. לצורך שליטה על החשיבות של כל מטרה, מגדירים קבוע ענישה על סוג לא נכון של דוגמאות אימון. קבוע זה מסומן בד"כ ב-C.

HCP .5.1

HCP (Human Connectome Project) - הוא פרויקט בן חמש שנים שהוקם ע"י המכוניס הלאומיים לבריאות (NIH). הפרויקט הושק ביולי 2009 כראשון מבין שלושת האתגרים הגדולים של תכנית המחקר של NIH לחקר המוח.

מטרתו של HCP היא לבנות "מפת רשת" מוחית (connectome) אשר תשפוך אור על הקישוריות האנטומית והפונקציונלית בתוך המוח האנושי הבריא, כמו גם כדי לייצר גוף של נתונים שיאפשר מחקר על תפקודים קוגניטיביים שונים של האדם ע"פ מפה מוחית.

באמצעות שיטות משופרות מאוד לרכישת נתונים, ניתוח ושיתוף, ה-HCP סיפק לקהילה המדעית נתונים ונתונים אשר מגבירים מאוד את הבנתנו את מבנה המוח האנושי, את הפונקציונליות והקישוריות ואת היחסים שלהם להתנהגות.

בHCP נבנה מאגר מידע המונגש באינטרנט, המכיל סריקות fMRI במצב מנוחה ובזמן צפיה בסרט, ברזולוציה T3 של כ-1200 נבדקים וברזולוציה T7 של כ-180 נבדקים (מתוך אותו מאגר של 1200 נבדקים). בנוסף מונגשים גם נתונים אישיים ודמוגרפיים (כגון גיל, מין, מוצא, קשרי משפחה ועוד) ותוצאות של מבחנים קוגניטיביים שונים שביצעו אותם נבדקים. לאורך השנים שוחררו כמה גרסאות של מידע. אנו נעבוד בפרויקט זה על סדרת מידע HCP S 1200 ששוחררה בשנת 2016 ועברה עיבוד מקדים נרחב.

5.2. סריקות 7T

5.2.1. מהלך הסריקה

סריקות באמצעות מגנט T7 התבצעו ב-4 מפגשים. כל מפגש נערך כשעה ורבע. כל מפגש התחיל בסריקה במצב מנוחה (rfMRI), לאחר מכן 2 סריקות בזמן צפייה בסרט ועוד 2 סריקות במצבים אחרים בהם לא נעסוק בפרויקט זה.

5.2.2. rfMRI

נתוני סריקות במצב מנוחה נרכשו בארבעה ריצות של כ-16 דקות כל אחת, כל ריצה התחילה בחדר חשוך עם עיניים פקוחות ועם פיקוס הנבדק בתמונה של צלב בהיר על רקע כהה המוקרן על הקיר.

5.2.3. tfMRI

בריצות הראשונה והרביעית, לאחר סריקה במצב מנוחה, התבצעו 2 סריקות בזמן צפייה בסרט. בכל אחת מהסריקות הוצגו סרטים באורך 15 דקות המורכבים מ-4 תתי סרטונים שונים ולא קשורים אחד לשני. בסה"כ בוצעו 4 סריקות במצב צפייה בסרט.

5.2.4. פרוטוקול הסריקה

כל סריקות fmri ב-T7 התבצעו עם זמן בין פולסים (TR) של 1000 ms וזמן מדידה של 22.2 ms. עובי slice וגודל כל ווקסל הוא 1.6 mm ומספר slices הוא 85. באיור x מופיע פרוטוקול הסריקה המלא.

Parameter	Value
Sequence	Gradient-echo EPI
TR	1000 ms
TE	22.2 ms
flip angle	45 deg
FOV	208 x 208 mm (RO x PE)
Matrix	130 x 130 (RO x PE)
Slice thickness	1.6 mm; 85 slices; 1.6 mm isotropic voxels
Multiband factor	5
Image Acceleration factor (iPAT)	2
Partial Fourier (pF) sampling	7/8
Echo spacing	0.64 ms
BW	1924 Hz/Px

Condition	Runs	Frames (TRs) per run	Run Duration (min:sec)
REST (Resting-state)	4	900	16:00

5.3. עיבוד מקדים (HCP Processing Pipeline)

לאחר ביצוע סריקות fMRI יש לבצע עיבוד מקדים למידע המתקבל לפני שניתן להתחיל לבצע עליו אנליזות.

בHCP מנגישים גם את המידע הלא מעובד וגם את המידע לאחר שהועבר שרשרת של שלבי עיבוד מקדים (HCP Processing Pipeline). תהליכי עיבוד המידע של HCP נחקרו רבות ונחשבים כיום כפורצי דרך בתחום ניתוח סריקות fMRI. על כן בחרנו להשתמש במידע המעובד. שלבי העיבוד הם כדלקמן:

1. Spatial Artifacts Removal:

1.1. תיקון זמנים - כיוון שה slices נדגמים אחד לאחר השני, יש לנרמל את כל הדגימות לאותו זמן.

1.2. תזוזות ראש - יש ליישר את תזוזות הראש של הנבדק במהלך הבדיקה.

1.3. הורדת תדרים נמוכים - הורדת רעשים שהתווספו במהלך הבדיקה.

2. Co-Registration:

כיוון שצורת המוח ומיקום הקפלים בו משתנה בין נבדק לנבדק יש לנרמל את הסריקות לתבנית מסוים (group template) על מנת להגיע לקורלציות גבוהות יותר של ערכי BOLD בין נבדקים. בHCP פיתחו תהליך רגיסטרציה שנקרא MSMall הנעזר במפות מאילין כדי לבצע התאמת כל סריקה לתבנית.

3. Alignment to standard space:

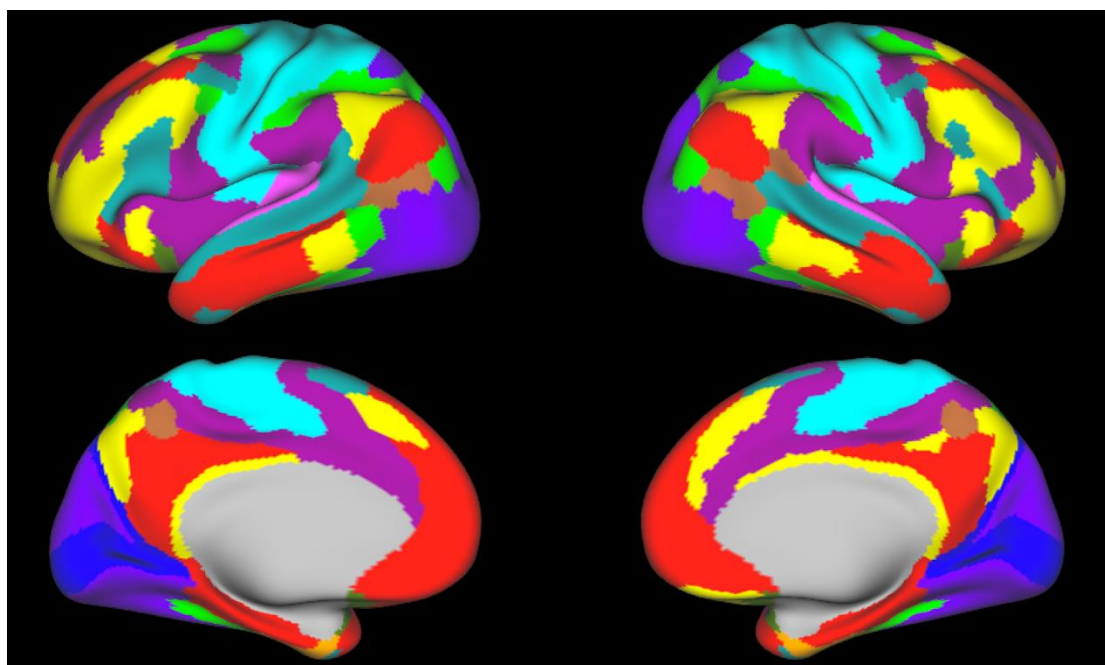
המרת המידע לקבצי CIFTI עם 91,282 grayordinates המכילים את אזורי קליפת המוח ופנים המוח ברזולוציה מרחבית של 1.6 מ"מ.

5.4. פרסלציה לאזורים ורשתות

בנוסף לעיבוד המקדים הנ"ל, ב HCP פיתחו שיטות מתקדמות למיצוע המידע לפי אזורים תפקודיים במוח המוכרים למדע מלפני. מיצוע זה נקרא פרסלציה (parcellation). בעבודה זו נשתמש בפרסלציה הממצעת את המוח ל718 אזורים.

6.1. פיתוח תשתיות

כיוון שפרויקט זה הוא הראשון במכון שעובד על המידע של HCP והראשון שעובד עם מידע בפורמט CIFTI היה עלינו ללמוד איך לעבוד עם המידע. כשלב ראשון, פיתחנו סט כלים לחישוב סוגי הקישוריות במוח (Network based ISFC & FC, ISC, Seed based) מתוך קבצי CIFTI. בנוסף, למדנו איך למצות מידע מרשתות המוח, ובפרט מרשת הDMN ע"י פרסלציה. תוך התייעצות עם בכיר מפרויקט HCP, בחרנו להשתמש בפרסלציה הממצעת את 91,282 התאים השונים במוח ל718 ממוצעים המייצגים אזורים שונים בקליפת המוח ופנים המוח (Cortex & Subcortex Regions). כל עבודת הפרסלציה נעשתה בעזרת כלים שפותחו ומונגשים ע"י HCP. בנוסף, השתמשנו בקובץ הממפה כל אחד מהאזורים לרשת מסוימת במוח.



14 המוח מחולק לרשתות

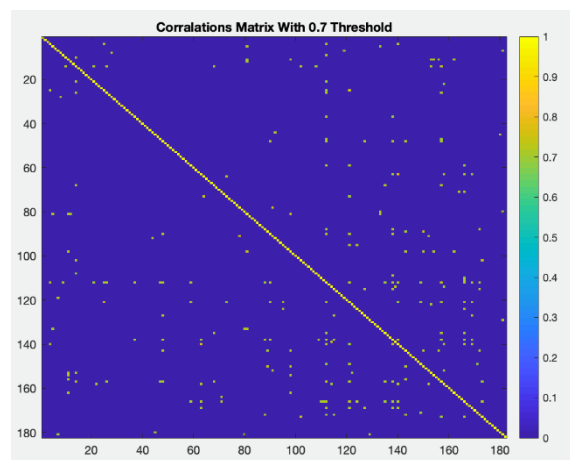
באיור 14 ניתן לראות את המוח מחולק לרשתות, כל צבע מייצג רשת כאשר רשת הDMN היא באדום. ניתן לראות שהDMN היא רשת גדולה יחסית ואכן היא מונה 109 מתוך 718 האזורים. שני החלקים העליונים בתמונה הם קליפת המוח ושני התחתונים הם פנים המוח.

6.2. חלק א': חיזוי תאומים מתוך סריקות מצב מנוחה

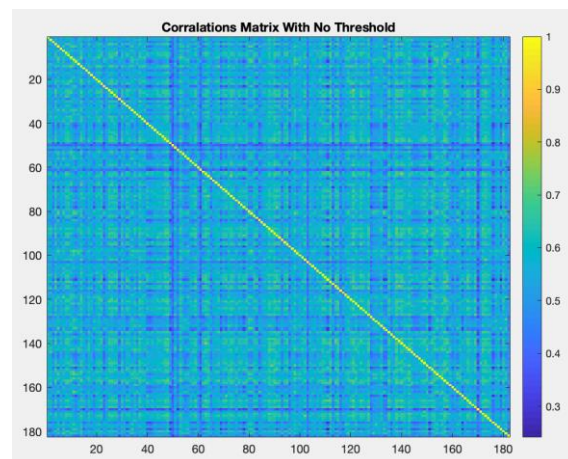
במאגר המידע של ה-T7 קיימים 85 זוגות תאומים, מתוכם 54 זוגות תאומים זהים. בחלק זה נתאר שיטות שונות שפיתחנו לצורך מציאת אחיו התאום של הנבדק מתוך מאגר של 182 הנבדקים שביצעו סריקת T7.

6.2.1. חיזוי מתוך מטריצות קורלציה

בחלק זה פיתחנו אלגוריתם לחיזוי תאומים ע"י הפקת מטריצות קורלציות בגודל 182×182 המכילות קורלציות בין כל זוג נבדקים. ההנחה - עבור שורה של נבדק מסוים, ערך הקורלציה הגבוהה ביותר נמצא בטור של אחיו התאום.



15 מטריצת קורלציה בין תאומים עם סף רגישות של 0.7



16 מטריצת קורלציה בין תאומים בלי סף רגישות

הערה: ערכי האלכסון של המטריצה מייצגים קורלציות של נבדקים בינם לבין עצמם. על מנת שיהיה קל לבצע שינויים, האלגוריתם נכתב בצורה מודולרית ומחולק ל-4 תתי תוכניות עצמאיות כאשר כל אחת מקבלת את הפלט של הקודמת לה:

חלק א'

ביצוע פרסלציה על המידע כפי שהוא התקבל - תוכנית זו מקבלת קבצי CIFTI של מסר נבדקים (קובץ אחד עבור נבדק), וקובץ הגדרות של פרסלציה מסוימת ובעזרת כלי של HCP מייצרת מטריצת ממוצעים עבור כל נבדק.

חלק ב'

חישוב מטריצת הקורלציות - תוכנית זו מקבלת את הקבצי הממוצעים של הנבדקים ומייצרת את מטריצת הקורלציות. חישוב קורלציה בין נבדקים מחושב באופן הבא: עבור כל מטריצת ממוצעים של נבדק:

a. חישוב וקטור ערכי FC עבור כל ערך ממוצע לשאר ערכי הממוצע (מתוך 718 הממוצעים).

b. הוספת הוקטור לשורה במטריצה גלובלית המכילה את תוצאות FC של כל הנבדקים לבסוף חישוב Pearson Correlation בין כל שורות המטריצה הגלובלית.

חלק ג'

התאמת תאומים - תכנית המקבלת את מטריצת הקורלציות, ומתאימה לכל נבדק תאום (לרוב על פי ערך הקורלציה הגבוהה ביותר).

חלק ד'

חישוב הצלחות - תכנית המקבלת את זוגות התאומים שחזתה המטריצה וקובץ המכיל את קשרי המשפחה בין הנבדקים ובודקת כמה הצלחות וכמה כישלונות חזתה מטריצת הקורלציה.

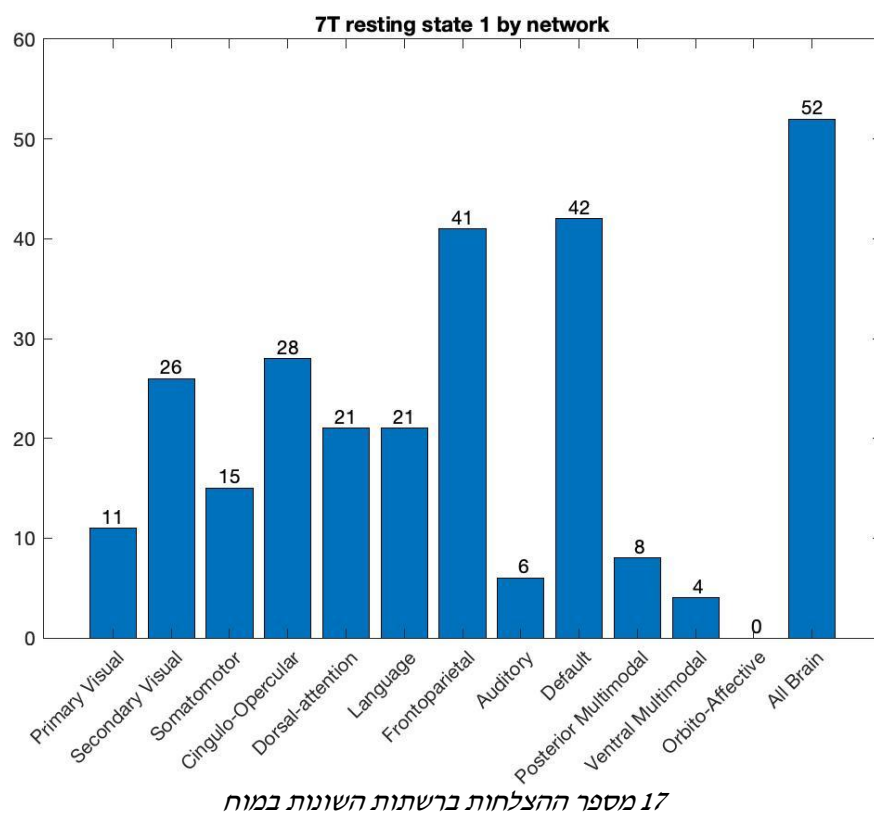
אלגוריתם זה ניבא 52 הצלחות מתוך 182 חיזויים. כלומר, מתוך 182 נבדקים, שמתוכם 170 קיים אח תאום בתוך הקבוצה, האלגוריתם מצא עבור 52 נבדקים את האח התאום שלהם. אחוזי ההצלחה הם כאמור 52/182. הסיכוי לבצע חיזוי כזה נכון בצורה רנדומלית הוא 1/182 ולכן תוצאה זו מראה לנו כי בהחלט קיימת קישוריות במוח הדומה יותר בין שני נבדקים שהם תאומים מאשר בין שני נבדקים שאינם תאומים.

שיעור ההצלחה: 0.29

6.2.2. מהלך המחקר ואופטימיזציות

6.2.2.1. FC network based correlations

בשלב זה ניסינו להבין את השפעת כל רשת במוח על הקורלציה בין נבדקים (איור 17). לצורך כך חישובי ה-FC ובניית מטריצת הקורלציה נעשו על רקע ממוצעי אזוריים (מתוך 718 האזורים) השייכים לרשת מסוימת. התוצאה: רשתות מסוימות מנבאות בדיוק רב יותר מאחרות.



באיור x ניתן לראות את כמות ההצלחות בהתאמת תאום לנבדק (מתוך 182) של כל אחת מ-12 רשתות המוח ושל המוח כולו (עמודה אחרונה). ניתן לראות שהמטריצה שמכילה את כל אזורי המוח היא זו שניבאה כמות הכי גדולה של תאומים ואחריה זו המטריצה שמכילה את אזורי ה-DMN בלבד.

6.2.2.2. Prediction by multiple matrices

מבדיקה שביצענו ראינו שרוב ההצלחות של כל הרשתות הן זהות בין הרשתות (זיהוי של אותם זוגות תאומים), אך לא כולן. כלומר אם נאחד את ההצלחות של כמה רשתות נוכל לקבל מספר גבוהה יותר של הצלחות. לכן הוספנו לאלגוריתם את היכולות הבאות:

1. חלוקה גסה יותר של המוח. בנוסף לחלוקה לרשתות חילקנו כל רשת למספר תתי רשתות, בצורה רנדומלית, על מנת לקבל מספר גבוהה יותר מ-12 של מטריצות

קורלציה. ההנחה - חישוב FC על חלקים קטנים יותר של המוח אולי ינבאו מידע מדויק יותר

2. ביצוע החיזוי ע"פ מספר מטריצות קורלציה ע"י לקיחת, עבור כל נבדק, את ערך הקורלציה הגבוהה ביותר אשר התקבל ממספר הכי גבוהה של מטריצות קורלציה.

לדוגמא:

נניח שהקלט הוא 3 מטריצות, ואנו מחפשים את התאום של נבדק מספר 1. נניח כי שתי מטריצות מכילות ערך קורלציה מקסימאלי בין נבדק 1 לנבדק 5 והמטריצה השלישית מכילה ערך קורלציה מקסימאלי בין נבדק 1 לנבדק 70. אנו נסיק מכך שהתאום של 1 הוא 5.

בדרך זו לאחר כיוון מספר החלוקות של כל רשת ובחירת רשתות מסויימות, הגענו ל70 הצלחות ע"י 32 מטריצות של 8 רשתות (1,3,5,6,2,4,7,9) + 4 מטריצות של המוח כולו.

שיעור ההצלחה: 0.38

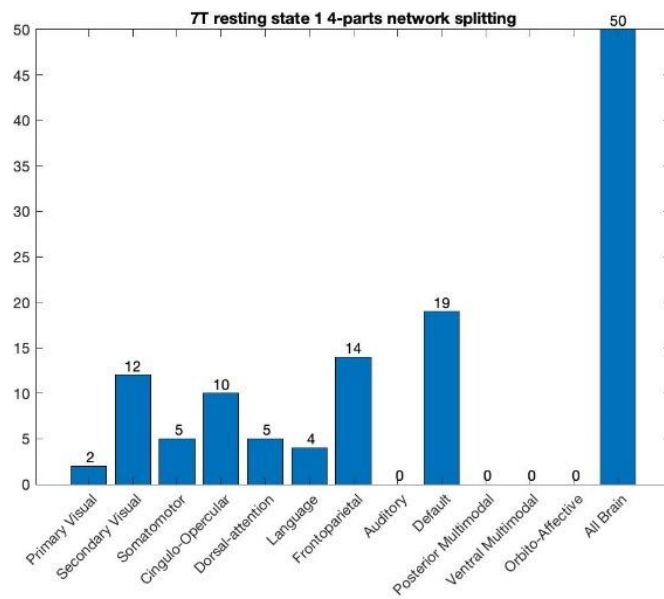
6.2.2.3 Monozygotic twins predictions

מתוך הידיעה שקיים דמיון גנטי גדול יותר בין תאומים זהים לתאומים שאינם זהים, ניסינו למקד את החיזוי לתאומים זהים בלבד. בתוך המאגר קיימים 54 זוגות תאומים זהים. חיזוי זה הניב 64 חיזויים נכונים מתוך 108 נבדקים עם תאום זהה.

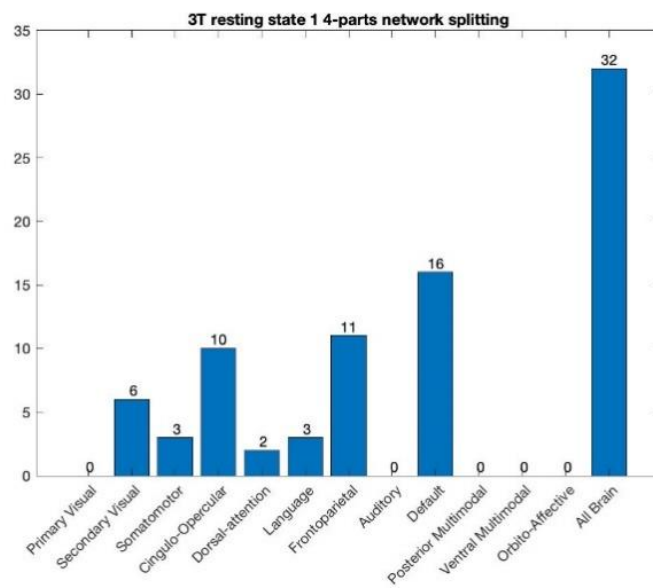
שיעור ההצלחה: 0.59

6.2.2.4 7T vs 3T data

בחלק זה השווינו שתי הרצות של אותו האלגוריתם, פעם אחת על מטריצות שחושבו מתוך סריקות T7 ופעם שניה על מטריצות שחושבו מתוך סריקות T3. ניתן לראות שכמות ההצלחות של האלגוריתם עבור רשת גבוהה יותר בקצת כשהוא מוזן בסריקות T7:



18 תוצאות ההצלחות השונות ב-7T



19 תוצאות ההצלחות ברשתות בשונות ב-3T

6.2.2.5 K-Nearest neighbor correlations

בחלק זה ניסינו לשנות את דרך חישוב ערכי הקורצליה. במקום לחשב Pearson correlation, מתייחסים לכל וקטור ערכי FC של נבדק כאל נקודה במרחב ומחשבים מרחק בין כל אחת מהנקודות. הנקודה הקרובה ביותר היא וקטור ערכי FC של האח התאום. בשיטה זו האלגוריתם חזה 50 הצלחות עבור 36 מטריצות של 8 רשתות + כל המוח.

שיעור ההצלחה: 0.27

6.2.3 סיכום תוצאות

כלל התוצאות מופיעות בטבלה (איור 20). הדיוק הגבוהה ביותר התקבל ע"י קלט של 36 מטריצות (בחירת 8 רשתות המשפיעות ביותר פיצול כל אחת מהן 4 תתי רשתות ובנוסף פיצול 4 של המוח כולו) המכילות ערכי Pearson correlation בין חישוב FC על סריקות T7. בנוסף כשמגבילים את החיזוי לתאומים זהים בלבד ניתן לראות קפיצה גדולה בדיוק.

dataset	min same predicts	network whitelist	num of cov matrixes	num of networks	num of subjects	trues	falses	success rate	error rate	comments
7T rest1(182_subjects)	1		13	13	182	58	64	0.3186813187	0.6813186813	
7T rest1(182_subjects)	2		13	13	182	51	28	0.2802197802	0.7197802198	
7T rest1(182_subjects)	1		12	12	182	50	131	0.2747252747	0.7252747253	
7T rest1(182_subjects)	0		12	12	182	52	129	0.2857142857	0.7142857143	
7T rest1(182_subjects)	1		13	13	182	58	94	0.3186813187	0.6813186813	
7T rest1(182_subjects)	1		24	12	182	57	120	0.3131868132	0.6868131868	
7T rest1(182_subjects)	0		24	12	182	53	124	0.2912087912	0.7087912088	
7T rest1(182_subjects)	1		26	13	182	65	115	0.3571428571	0.6428571429	
7T rest1(182_subjects)	1		48	12	182	59	122	0.3241758242	0.6758241758	
7T rest1(182_subjects)	1		50	13	182	67	114	0.3681318681	0.6318681319	
7T rest1(182_subjects)	1		95	13	182	62	119	0.3406593407	0.6593406593	
7T rest1(182_subjects)	1		159	13	182	52	129	0.2857142857	0.7142857143	
7T rest1(182_subjects)	1	[1,3,5,6,2,4,7,9,13]	36	9	182	69	112	0.3791208791	0.6208791209	
3T rest1(181_subjects)	1		50	13	182	51	114	0.2802197802	0.7197802198	
7T rest1(182_subjects) - nearest neighbor	1	[1,3,5,6,2,4,7,9,13]	36	9	182	50	131	0.2747252747	0.7252747253	
7T rest1(182_subjects) - nearest neighbor	1		13	13	182	42	124	0.2307692308	0.7692307692	
7T_rest_1234_merged (174 subjects)	1	[1,3,5,6,2,4,7,9,13]	36	9	182	46	127	0.2527472527	0.7472527473	
7T rest2(182_subjects)	1	[1,3,5,6,2,4,7,9,13]	36	9	182	70	111	0.3846153846	0.6153846154	
7T rest1(182_subjects)	1	[1,3,5,6,2,4,7,9,13]	36	9	182	64	44	0.3516483516	0.6483516484	MZ only

20 סיכום תוצאות חיזוי תאומים באמצעות FC

6.3 חלק ב: סיווג רמת אינטליגנציה לפי סריקות בזמן צפייה בסרט - יושלם בהמשך

7. כלי עבודה

7.1. תוכנות

בפרויקט זה סביבת הפיתוח שעמה עבדנו הייתה מטלב- MATLAB (matrix laboratory). התוכנה מאפשרת טיפול קל ונוח במטריצות, שימוש בפונקציות ובנתונים, מימוש אלגוריתמים על נתונים.

בנוסף עבדנו עם wb_commd ו wb_view- תוכנה מטעם HCP שאפשרה לנו לפתוח ולקרוא את כל קבצי הCIFTI.

7.2. שפת תכנות

בפרויקט זה השתמשנו בשפת תכנות MATLAB.

7.3. ספריות קוד פתוח

<https://www.artefact.tk/software/matlab/gifti/> - Matlab lib for working with CIFTI file.

<https://github.com/kakearney/foodwebgraph-pkg/tree/master/InterPointDistanceMatrix>
- Calculate distance between points in matrix.

8. ביבליוגרפיה

1. Simony, Erez, et al. "Dynamic reconfiguration of the default mode network during narrative comprehension." Nature communications 7.(2016)
2. Matthew F.Glass. "The minimal preprocessing pipelines for the Human Connectome Project". NeuroImage.(2013)
3. Matthew F.Glass. "A multi-modal parcellation of human cerebral cortex". Nature Research Journa.(2016)

9. נספחים

9.1. מקורות מידע :

- a. <https://github.com/Washington-University/HCPpipelines> - HCP Preprocess pipeline.
- b. <https://github.com/Washington-University/HCPpipelines/tree/master/MSMAll> - MSMAll Registration.
- c. <https://scholar.google.com/citations?user=hqA9AugAAAAJ&hl=en> – Matthew F. Glasser, An HCP senior researcher.
- d. https://github.com/ColeLab/ColeAnticevicNetPartition/blob/master/CortexSubcortex_ColeAnticevic_NetPartition_wSubcorGSR_parcel_LR.dlabel.nii - 718 Regions parcellation file.
- e. https://github.com/ColeLab/ColeAnticevicNetPartition/blob/master/cortex_subcortex_parcel_network_assignments.txt - 718 Regions to 12 networks assignments file.