

פרויקט סיום

ניתוח נתונים של דמיון תנועה

אור נחמני : 203533815

נינה פרטוש : 342472560

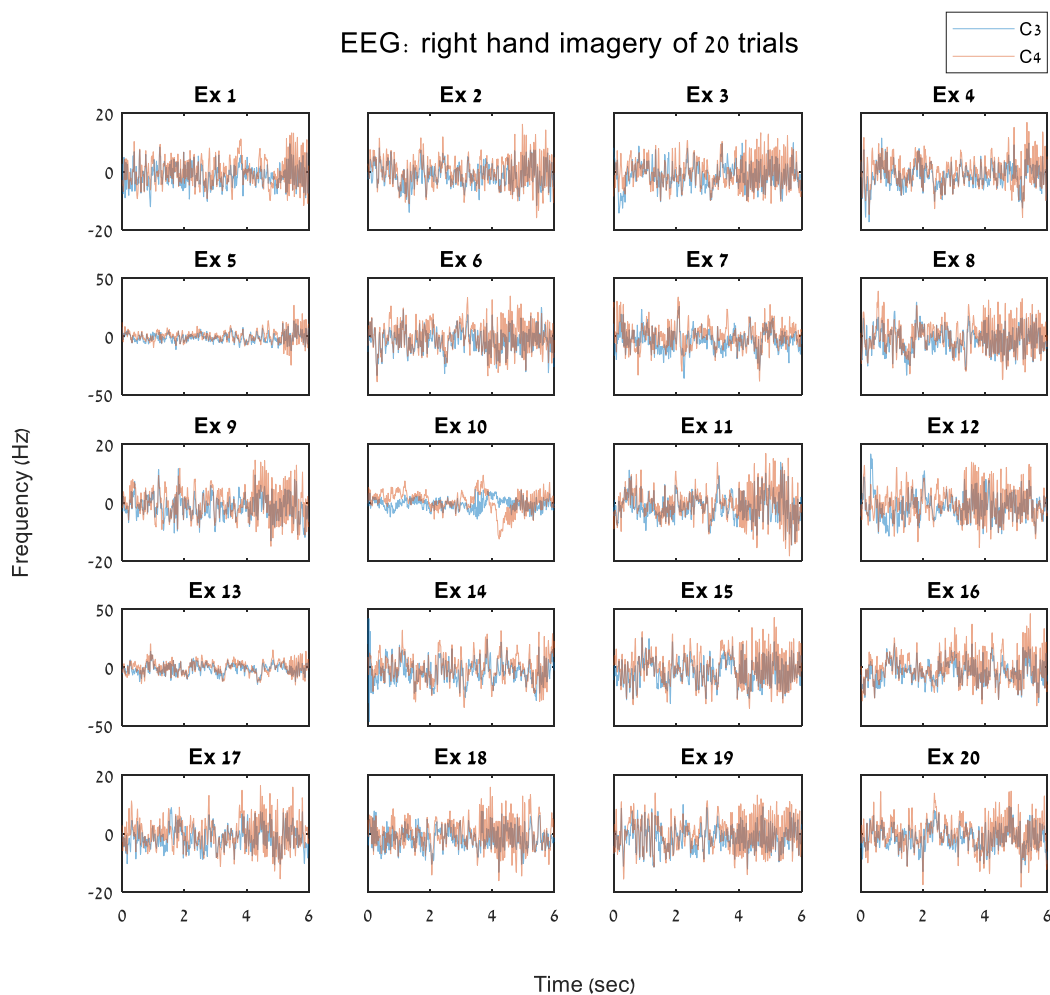
מטרת התרגיל

מטרת התרגיל הינה לנתח נתוני EEG של דימיון (imagery) תנועה של נסיינים. קיבלנו נתונים של 160 ניסויים (trials) בהם התבקשו לדמיין במשך כמה שניות את ידם הימנית והשמאלית (80 ימין ו80 שמאל בסדר רנדומלי) כאשר אלקטרודות C3, C4 מקליטות את פעילות המוחית.

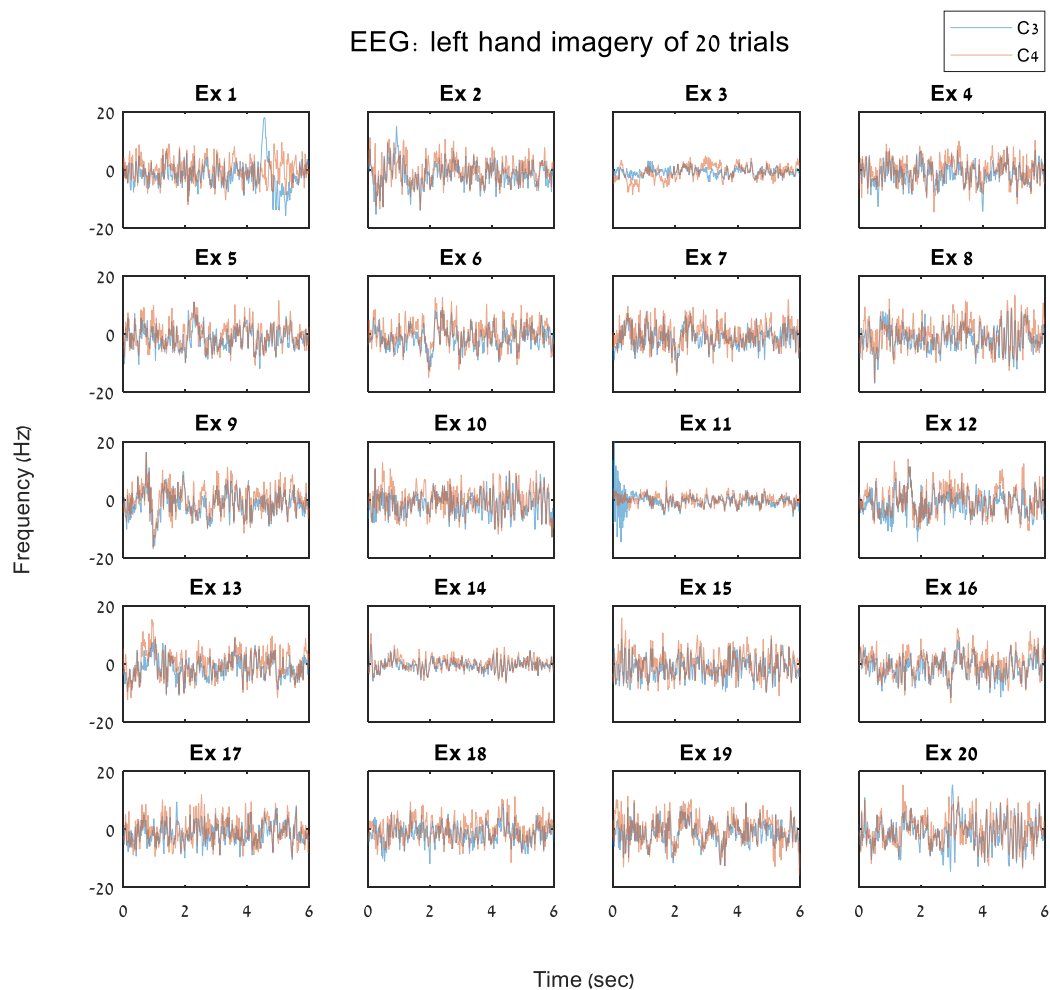
מטרתנו היא לחלץ מנתוני EEG תכונות (features) ולאמן אלגוריתם של קלסיפיקציה עליהן כך שנוכל לנבא את היד (ימנית או שמאלית) שהנסיין דמיין מתוך סיגל EEG.

חלק 1 : הצגת הנתונים הגולמיים

דגמנו רנדומלית 20 טריילים לדוגמא, עבור כל צד- ימין ושמאל, והצגנו בגרף את האותות משני הערוצים (C3 וC4) על מנת לאבחן ויזואלית הבדלים בין הצדדים השונים:



(Ex עבור Example)



מהתבוננות בגרפים ניתן להבין שיש שוני רב בין האותות ואי אפשר להבחין במאפיינים מסוימים מתוך הנתונים בצורתם הגולמית.

תיאור הקוד:

חילצנו ממבנה הנתונים הנתון את האינדקסים של הטריילים בהם דמיינו את יד שמאל ויד ימין, ויצרנו את הפונקציה `plot_EEG_sample` שמקבלת את הדאטה ומספר הטריילים הרצוי להצגה, דוגמת את הכמות הנדרשת (בעזרת הפונקציה `datasample` של מטלב) ומציגה את הגרפים- כשעבור כל ניסוי מוצגות שתי האלקטרודות: C3 ו-C4.

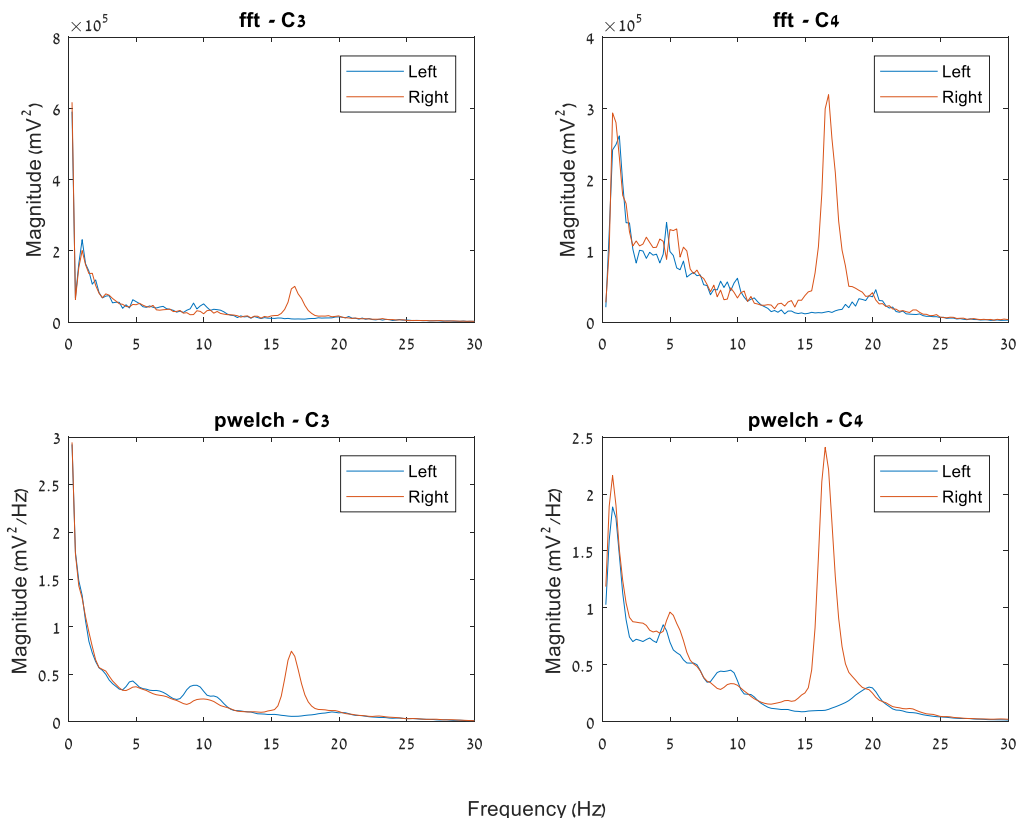
חלק 2 : Power Spectrum Analysis

1) אפיון על ידי FFT - PWELCH

על מנת לאבחן הבדלים בין הצדדים (יד ימין/שמאל) שונים בצורה יותר אפקטיבית, חישבנו והצגנו את ממוצעי ה power spectrum של הניסויים. את החישוב עשינו בשתי שיטות: pwelch ו-fft, כדי להחליט איזו מהן יעילה יותר. חישבנו את PS של האותות רק בזמן הדמיון- כ-2 שניות אחרי תחילת ההקלטה, כדי לקבל רק את התדרים הרלוונטיים (לא אמור להיות הבדל לפני הדמיון והמידע מזמן זה עשוי "ללכלך" את התמונה).

עבור כל שיטה (pwelch ו-fft) חישבנו את הספקטרום של הטריילים, מיצענו עבור כל יד, והצגנו לכל שיטה ואלקטרודה את הספקטרה הממוצעת עבור הידיים השונות. בצורה זו קל לזהות הבדלים באותות בכל אלקטרודה בין דמיון הצדדים השונים. בנוסף, בחרנו להציג את הגרפים בטווח תדרים שמצאנו שהכי אינפורמטיבי (0 - 30 Hz).

Power Spectra Comparison



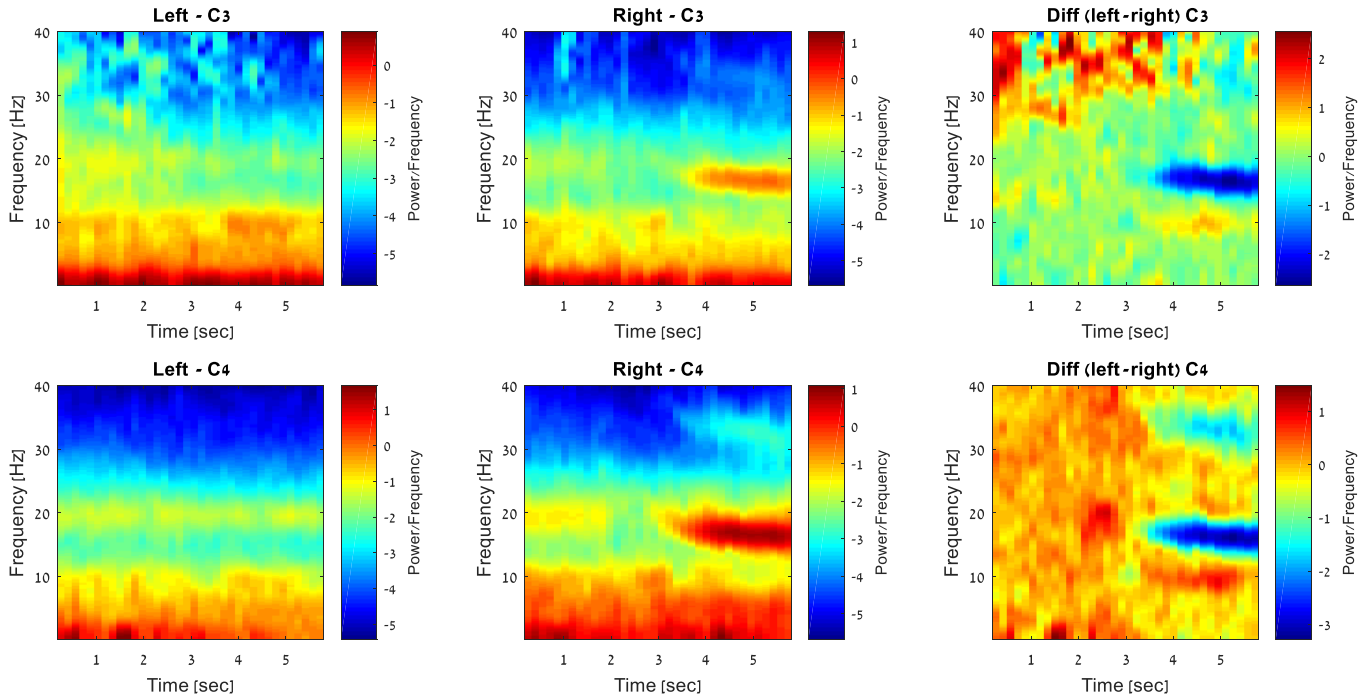
סך הכל, עקומות power spectrum בשתי שיטות החישוב דומות, אך גרף pwelch חלק יותר וברור יותר (כיוון ששיטה זו מחלקת את הסיגנל לחלונות זמן ועליהן מחשבת את הfft). *להסתכל על עבודה 5 ומבחינתי להעתיק את ההסבר שכתבנו. לדוגמא אם נתבונן בתדר 8-10 הרץ נראה שהfft המתקבל "רועש" לעומת ה- pwelch אשר מראה הבדל ברור יותר בין העקומות.

בתדרים 15-20 אפשר לראות הבדלים משמעותיים בין ימין ושמאל בשתי האלקטרודות, כאשר הפעילות בתדרים אלו חזקה יותר עבור דמיון יד ימין. תדרים בולטים יותר בעת דמיון יד שמאל הם 8-11, אך לא באופן משמעותי. הסיבה לכך שנצפו תדרים בולטים אך ורק ליד ימין עשויה להיות שהאלקטרודות שנבדקו לא ממצות, אפשרות נוספת היא שיש קשר ליד הדומיננטית של הנבדק (ושלניסוי נלקחו רק נבדקים בעלי יד ימין דומיננטית).

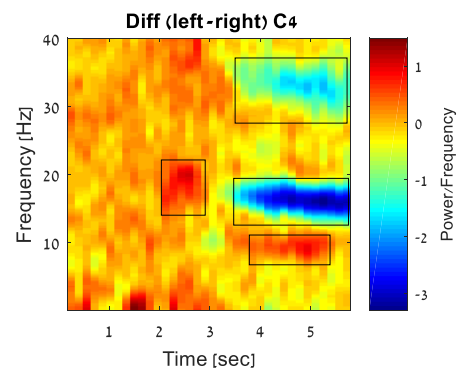
2) אפיון על ידי Spectrograms

כדי לבחון מזווית אחרת את הנתונים ולאתר תכונות נוספות, חישבנו והצגנו את הספקטרוגרמה הממוצעת לכל יד ואלקטרודה. החישוב נעשה בעזרת הפונקציה המובנית spectrogram, על כל טרייל בנפרד ולאחר מכן חישוב הממוצע. בשונה מה-pwelch, חישבנו את הספקטרוגרמה על כל חלון הזמן ולא רק הדמיון, משום ששיטה זו מציגה את העוצמה בתדר זמן ולא רק בתדר כמו pwelch. כך הפעילות השונה בזמנים השונים לא מתערבבת ומשפיעה על התצוגה.

הצגנו את הספקטרוגרמות של כל אלקטרודה וכל יד, וכדי להקל על זיהוי ההבדלים הצגנו גם את חיסור הספקטרוגרמות בין הידיים בכל אלקטרודה:



בספקטרוגרמות החיסור- הבדלים בין דמיון יד ימין ויד שמאל יוצגו בצבעים הרחוקים יותר מ-0 (על פי ציר הצבע). בחנו את הספקטרוגרמות וחילצנו תכונות (המאופיינות על ידי טווח תדרים וזמנים) בהם נראו הבדלים בין יד ימין לשמאל: ניתן לראות שוני בולט בשני הערוצים בטווח התדרים 14-18 הרץ, בשניות 3.5-6: בדומה למה שראינו בpower spectra באמצעות pwelch. חלונות נוספים לדוגמה הם תדרים 16-21 בשניות 2-3 באלקטרודה C4- כלומר פעילות שונה בדיוק לפני/בתחילת הדמיון, ותדרים 30-38 הרץ בזמן הדמיון:



בהמשך שיחקנו מעט עם הטווחים עד שהגענו לתכונות המדויקות ביותר לקליסיפיקציה של הצדדים השונים (פירוט בשלב הבא).

תיאור הקוד:

יצרנו struct שמכיל את המידע של כל הניסויים לפי יד ואלקטרודה: המידע הגולמי, ו-power spectrum ממוצע של האות בזמן הדמיון (שניות 2-6) לפי שתי השיטות- fft, pwelch. לכל אלקטרודה, יד, טרייל ושיטה, חישבנו את power spectrum וסכמנו- כיוון שאנו מעוניינים אך ורק בממוצע שלהם. כך חסכנו במשאבי מקום ועיבוד עבור התוכנית. את הסכום חילקנו בכמות הטריילים כדי להגיע לממוצע, והצגנו בגרף המוצג מעלה. את fft חישבנו על ידי הפונקציה calcFftPS בדומה לעבודה 5. Pwelch חושב על ידי הפונקציה המובנית של מטלב.

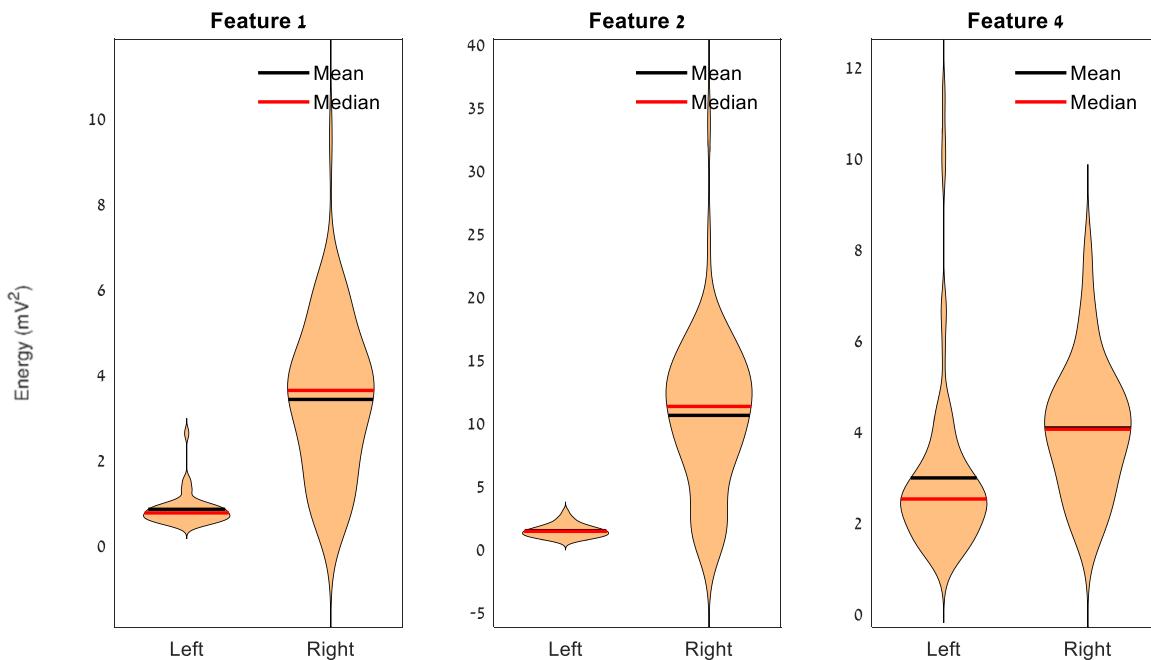
ספקטרוגרמות- עברנו שוב על הטריילים והפעם חישבנו את הספקטרוגרמות על כל חלון הזמן, סכמנו באותו אופן ושמרנו את הממוצע בstruct שלנו. לאחר מכן הצגנו את הספקטרוגרמות של כל יד ואלקטרודה, והחיסור בין הידיים. הגרפים מוצגים על ידי פונקציית עזר- plotSpectrogram.

3. ניתוח תכונות (features)

לאחר שהגדרנו את הפיצ'רים על סמך הספקטרוגרמות, חישבנו את ה"אנרגיה" של כל פיצ'ר בכל טרייל: bandpower עבור טווח התדרים והזמנים של הפיצ'ר.

כדי לאמוד את איכות התכונות שלנו- כלומר מידת יכולת ההפרדה שלהן בין יד ימין לשמאל, בחנו את התפלגות האנרגיה של שתי הידיים עבור הפיצ'רים השונים.

הצגנו התפלגות של פיצ'רים נבחרים:



כפי שניתן לראות, בתכונות 1,2 ההתפלגות מראה הבדלים בולטים בהתפלגות האנרגיה בין הידיים השונות, ונצפה שתכונות אלו יהיו יעילות בקלסיפיקציה של האותות. לעומת זאת בתכונת 4 ההבדל קטן יותר בין הממוצע וצורת ההתפלגות, וכנראה שתניב אחוזי הצלחה נמוכים יותר.

שיחקנו עם הפיצ'רים עד שקיבלנו חפיפה מינימלית בין ההתפלגויות.

תיאור הקוד:

חישוב האנרגיה נעשה על ידי הוצאת ה power spectra של הטרייל בחלון הזמן של הפיצ'ר (בעזרת pwelch), וחישוב העוצמה שלו בטווח התדרים של הפיצ'ר עם הפונקציה Bandpower.

התקבלה מטריצה של 160 הטריילים וערך האנרגיה שלהם בכל אחד מהפיצ'רים, איתה נבצע את תהליך הלמידה וקלסיפיקציה בהמשך.

את התפלגות ה violin plot הצגנו על ידי הפונקציה violin מתוך file exchange של מטלב.

* לבחינה מעמיקה של התפלגויות האנרגיה שהפיצ'רים מניבים- ניתן להשתמש בפונקציה features_analysis שמציגה את ה violin plots והיסטוגרמות של כל פיצ'ר, וגם- scattter plots להשוואה בין כל צמד פיצ'רים. הסירו את ההערה משורת BONUS (213)

4. קלסיפיקציה

בשלב זה אימנו אלגוריתם למידה לינארי להבדיל בין דמיון ימין ושמאל באמצעות הפיצ'רים שלנו. דגמנו רנדומלית 70% מהטריילים והשתמשנו באלגוריתם LDA1 לקבלת המשקולות לסיווג. לאחר האימון, בדקנו את אחוז ההצלחה של סיווג 30% הטריילים הנותרים.

אחוזי חלוקת המידע לאימון ולאיתמות צריכה להיות מאוזנת: אימון על הרבה דוגמאות יעזור לאלגוריתם לדייק את המשקולות, אך אימון על אחוז גדול מדי של הנתונים יגרום ל overfitting- האלגוריתם יסווג במדויק את הדוגמאות שניתנו אך לא ידע להכליל- ואחוז השגיאה בנתונים חדשים תהיה גבוהה (כלומר התאמת יתר לדוגמאות). המטרה היא לבצע קלסיפיקציה מוצלחת של מידע חדש, ולכן ההכללה של האלגוריתם היא קריטית. משקולות מוצלחות הן כאלו שיסווג נכונה מידע שטרם נצפה ע"י האלגוריתם, ולכן נבדוק את אחוז ההצלחה של סיווג 30% הניסויים עליהם לא התאמנו. זו סיבה נוספת לאימון על אחוז לא גבוה מדי של הנתונים- כיוון שככל שניקח טריילים לשלב האימון, יהיה לנו פחות נתונים לשלב הבדיקה כדי לאמת את יכולת הסיווג של הפיצ'רים שלנו.

את תהליך הדגימה, אימון ואיתמות ביצענו 70 פעמים וחישבנו את אחוז ההצלחה הממוצע וסטיית התקן (עבור כל הפיצ'רים יחד, וכל אחד בנפרד), כדי לוודא שהתכונות שלנו אכן קונסיסטנטיות.

אחוזי ההצלחה נעים סביב 89 עבור ששת התכונות יחד, עם סטיית תקן של ~4, כאשר כל תכונות בנפרד מניבה בין 72-91 אחוזי הצלחה (עם סטיית תקן 3-6 אחוז).

Results for 70 realizations
Accuracy: 89.345, std: 4.34

Features:

freq_range		time_range		channel	accuracy	std
14	18.5	3.8	6	1	87.44	4.6513
14	18.5	3.8	6	2	90.952	3.8017
16.5	20	2	3	2	72.619	5.5663
8	13	4.5	6	2	72.173	5.7634
19	21	4	6	1	88.839	4.4021
30	38	3.5	6	2	87.738	4.8574

*אלקטרודה 1- C3, אלקטרודה 2- C4.

תיאור הקוד:

הרצנו את אלגוריתם LDA1 על כמות הטריילים הנבחרת לאימון: ממטריצת האנרגיה של התכונות (וכמובן וקטור הלייבלים האמיתיים – ימין או שמאל). באמצעות המשקולות שהתקבלו בדקנו את הטריילים הנותרים, וחישבנו את אחוז ההצלחה על ידי השוואת את התוצאות שקיבלנו עם הלייבלים המקוריים.

בדקנו את הצלחת האלגוריתם על כל תכונות בנפרד ועל כולן יחד, והדפסנו את התוצאות. שמרנו את הנתונים: טבלת הפיצ'רים, האנרגיה, ותוצאות ה realizations.

דיון בתוצאות וסיכום:

על מנת להיות יכולים לסווג אותות EEG מאלקטרודות C3,C4 בזמן דמיון הזזת יד (ימין או שמאל), ניתחנו את הנתונים ב-3 שלבים:

- 1- Pre processing: חילוץ הנתונים הרלוונטים ועיבוד האותות (pwelch, spectrograms).
- 2- Feature extraction: בחינת ממוצעי הספקטרוגרמות ומציאת טווחי תדר/זמן/ערוץ רלוונטיים להבחנה בין יד ימין לשמאל.
- 3- Classifier training: מציאת הפרדה לינארית בין הנתונים של ימין/שמאל, על ידי למידה מפוקחת עם הפיצ'רים שלנו.

התוצאות:

בדקנו את אחוזי ההצלחה של הפיצ'רים שחילצנו – בנפרד ושל קומבינציות שלהם. אחוז ההצלחה האופטימלי שהצלחנו להשיג נמצא באזור 90% בממוצע עם סטיית תקן של כ-3.5%, באמצעות ארבעת הפיצ'רים הבאים:

Results for 70 realizations
Accuracy: 90.476, std: 3.45

Features:

freq_range		time_range		channel	accuracy	std
14	18.5	3.8	6	1	86.756	3.7876
14	18.5	3.8	6	2	90.714	3.4514
16.5	20	2	3	2	71.25	6.0568
19	21	4	6	1	88.661	3.7139

נשים לב שהפיצ'ר השני הינו בעל אחוז הצלחה גבוה יותר מכלל הפיצ'רים יחדיין- כלומר שימוש בפיצ'ר בודד על הנתונים הניב הצלחה גדולה יותר בסיווג. הסבר אפשרי הוא overfitting על הנתונים באמצעות פיצ'רים מרובים- ייתכן והפיצ'רים הנוספים היו עוזרים יותר עבור מאגר נתונים גדול ומגוון יותר.

מצאנו כי טווח התדרים היעיל ביותר לאבחון הבדלים בין דמיון יד ימין ויד שמאל בנתונים שלנו, הוא 14-21 Hz. טווח זה הוא חלק מטווח התדרים של גלי בטא.

התוצאה תואמת את ממצאי המאמר Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas (Neuper, Pfurtscheller) קורלציה בין פעילות גלי בטא בסיגנל EEG לאזורים סנסומוטורים באזורים קורטיקליים במח. בנוסף, טווח תדרים זה נותן תוצאות יחסית טובות גם בחלון הזמן של שניות 2-3 (זמן תחילת/לפני הדמיון). ייתכן שזה מייצג סוג של תכנון או הכנה לדמיון.

על אף ההבדל הבולט בpower spectra והכתם בספקטרוגרמה עבור תדרים 8-13 הרץ- פיצ'ר זה התגלה כפחות יעיל (ביחס לאחרים). טווח תדרים זה הינו של גלי אלפא, אשר קשורים לפעילות של אזורים תלמו-קורטיקליים המעורבים בהעברת סיגנלים מוטורים לאזורים קורטיקליים. (על פי המאמר Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates (Neuper, Pfurtscheller)).

כיום, התחום של ממשקי מח-מחשב מספק פתרונות חדשניים לשיפור אורח החיים של בני אדם ובמיוחד בעלי מוגבלויות, לדוגמה: שליטה ביד רובוטית ע"י מחשבה עבור חולי ALS. לכן, יצירת אלגוריתמי קלסיפיקציה כדוגמת זה שבתרגיל מהווה אתגר חשוב ביותר ☺



The End.