

דוח פרויקט גמר – נוי מיכאלי (326640208), שירה זלנפרוינד (212641021)**הקדמה**

בפרויקט זה, מטרתנו הייתה לחקור נתוני EEG שנאספו מאלכוהוליסטים ומשתתפים המשמשים כבקרה על מנת למצוא האם קיימים הבדלים בפעילות המוח. השאלות שהנחנו אותנו הן כיצד מגיבים אזורים במוח לגירויים ויזואליים חוזרים בקבוצות הללו לאורך זמן, ואילו אזורי מוח (על פי מנחי חיישנים) ספציפיים מראים את הבדלי הפעילות הגדולים ביותר.

נתוני ה-data כוללים: EEG של 122 נבדקים, שעברו 120 מבחנים בהם הוצגו להם אחד משני גירויים (S1 כלל גירוי ויזואלי אחד, בעוד S2 כלל 2 גירויים: הגירוי השני יכול להיות דומה לראשון/שונה ממנו). המאגר הינו directory שמכיל תיקיות נסיינים: בכל תיקיית נסיין נמצאים – מס' המבחן, מיקום החיישן, גובה אות חשמלי במיקרו וולט, מס' דגימה, קבוצת הנסיין, גירוי תואם/שונה, זמן תגובה.

לאחר שלקחנו נתונים אלה וסידרנו אותם בדרכים שנפרט בהמשך, זיהינו מיקומי חיישן ספציפיים המראים הבדלים משמעותיים בין הקבוצות. מחקר זה תורם להבנת המאפיינים הנוירולוגיים של אנשים אלכוהוליסטים בהשוואה ללא אלכוהוליסטים.

שיטות

כפי שצינו, נתוני ה-EEG בהם השתמשנו מחולקים ל 120 נבדקים, כאשר לכל נבדק היה קובץ משל עצמו. עקב כך היה צורך ביצירת קובץ מאוחד + נקי מערכים ריקים למקרה ויש בו.

לשם כך התחלנו ביצירת קובץ **data cleaning** הכוללת 2 פונק':

csv_combined - מאתרת את כל הקבצים עם הסיימות csv. בתיקיה שניתנה כקלט בעזרת ספריית glob ויוצרת רשימה ריקה שתשמש לאחסון ה-DataFrame ים של כל הקבצים. בלולאה, הפונקציה קוראת כל קובץ CSV ל-DataFrame תוך כדי טיפול בשגיאות, מזהה אם הקובץ ריק ומדפיסה אזהרות במידת הצורך. אם נמצאו DataFrame ים תקינים, היא משלבת את כולם ל-DataFrame יחיד באמצעות pd.concat ומנקה את הנתונים על ידי הסרת שורות כפולות ושורות עם ערכים חסרים. בסיום, היא מחזירה את ה-DataFrame המשולב והנקי.

convert_numeric_val על מנת להמיר את הערכים בעמודה מסוימת של DataFrame לסוג מספרי (int, float).

לאחר מכן יצרנו את **data analysis** הכוללת 5 פונק':

compute_group_differences מחשבת את ההבדלים הממוצעים בערכי חיישני EEG בין שתי קבוצות (למשל, "Alcoholic" ו-"Control") עבור כל מיקום חיישן, ומחזירה DataFrame מסודר לפי גודל ההבדלים. תחילה, היא מחשבת את הערכים הממוצעים של עמודת המדידה (value) עבור כל קבוצה בכל מיקום חיישן, תוך שימוש בקיבוץ לפי עמודות מיקום החיישן (position) וזיהוי הקבוצה (subject_id). לאחר מכן, היא מוסיפה עמודה חדשה בשם "difference", שבה מחושב ההפרש המוחלט בין הממוצעים של שתי הקבוצות בכל מיקום. הנתונים ממויינים לפי עמודת ההבדל בסדר יורד, כך שהחיישנים עם ההבדלים הבולטים ביותר מופיעים בראש הרשימה. הפונקציה מחזירה DataFrame הכולל את הממוצעים של כל קבוצה, את ההבדלים המוחלטים, ואת הנתונים המסודרים לפי גודל ההבדל.

*לאחר מכן- הקוד מגדיר רשימה של שיוך אזורי מוח בהתבסס על קידומות שמות החיישנים (region_mapping) ומשתמש בו בפונקציה **assign_brain_region** כדי לשייך כל חיישן לאזור מוחי מסוים.

assign_brain_region מקבלת כקלט את שם החיישן של ה-EEG ומחזירה את האזור במוח שמתאים לחיישן זה, בהתבסס על תחיליות מסוימות בשמות החיישנים. הפונקציה מתמודדת עם מקרים בהם יש סתירה בין תחיליות קצרות וארוכות, כמו לדוגמה 'P' ו-'PO'. בראש ובראשונה, הפונקציה בודקת אם שם החיישן מתחיל באחת מהתחיליות 'F' או 'AF', במקרה כזה היא מחזירה את האזור "Frontal Lobe" (החלק הקדמי של המוח). לאחר מכן, הפונקציה נותנת עדיפות לתחיליות ארוכות יותר, כמו 'PO' (הממופה ל-"Parietal-Occipital Lobe") ו-'TP' (הממופה ל-"Temporal-Parietal Lobe"), מתוך מטרה להימנע מטעויות שמקורן בהתאמה לשמות חיישנים עם תחיליות קצרות יותר. אם שם החיישן מתחיל ב-'FT', הפונקציה מחזירה את האזור "Frontal-Temporal Lobe", ואם הוא מתחיל ב-'CP' או 'FC', היא מחזירה את האזור "Sensory-Motor Cortex" ו-"Motor Cortex" בהתאמה. אם החיישן לא מתאים לאף אחת מהתחיליות הנ"ל, הפונקציה פונה לרשימה חיצונית של מיופיים (region_mapping) ומחפשת שם את האזור המתאים לפי התחילית של שני תווים. אם לא נמצא אף אזור שמתאים לחיישן, הפונקציה מחזירה את המילה "Unknown Region". בצורה כזו, הפונקציה מספקת מיפוי של אזורי מוח שונים לפי שמות החיישנים, תוך שמירה על סדר עדיפויות ברור במקרים של חפיפות בין תחיליות.

map_sensors_to_regions מקבלת כקלט נתונים בפורמט של DataFrame של ספרייט pandas, יחד עם שם של עמודה שכוללת את מזהי החיישנים של ה-EEG. הפונקציה ממפה כל חיישן לאזור במוח המתאים ומוסיפה עמודה חדשה בשם 'region' ל-DataFrame. בשלב הראשון, הפונקציה משתמשת בפונקציה assign_brain_region (שכבר דיברנו עליה) כדי למפות את כל שם חיישן לעבר האזור המתאים במוח. זה נעשה על ידי קריאה ל-apply על העמודה שכוללת את שמות החיישנים, כך שלכל שם חיישן נקבע האזור המתאים במוח. התוצאה נשמרת בעמודה חדשה בשם 'region'. לאחר מכן, הפונקציה מבצעת סינון של הנתונים: היא מסירה שורות שבהן לא נמצא אזור מוח מתאים (כלומר, כאשר הערך בעמודה 'region' הוא NaN). בנוסף, הפונקציה מסירה את השורות שבהן הערך בעמודת 'region' הוא "Unknown Region", כלומר השורות שבהן לא נמצאה התאמה לאזור מוח כלשהו. לבסוף, הפונקציה מחזירה את ה-DataFrame המעודכן, הכולל את העמודה החדשה 'region' שמכילה את המידע על האזורים במוח שממופים לשמות החיישנים.

analyze_responses_by_condition_and_group מבצעת ניתוח סטטיסטי על נתונים מתוך DataFrame ולחשב ממוצע וסטיית תקן עבור ערכים מספריים, תוך התמקדות בקיבוץ הנתונים לפי שני פרמטרים: תנאי הניסוי (למשל, האם מדובר בתנאי אלכוהול או ביקורת) וקבוצת הנבדקים (למשל, קבוצת נבדקים עם אלכוהול מול קבוצת ביקורת). הפונקציה מקבלת ארבעה ארגומנטים: DataFrame שמכיל את הנתונים, שם העמודה שמכילה את הערכים המספריים (כמו תוצאות ניסוי), שם העמודה שמכילה את התנאים (למשל, סוג הניסוי - s1, s2, ...), ושם העמודה שמכילה את מזהי הנבדקים או קבוצות הניסוי. הפונקציה מקבצת את הנתונים לפי תנאים ולקבוצות, ואז מחשבת את הממוצע ואת סטיית התקן עבור כל קבוצת תנאי, ומשיבה את התוצאות כ-DataFrame חדש שמראה את הממוצע וסטיית התקן לכל שילוב של תנאי וקבוצה.

perform_t_tests מבצעת סדרת מבחני t-test בלתי תלויים (independent t-tests) על נתוני EEG, תוך השוואת שתי קבוצות של נבדקים לפי מיקומי סנסורים שונים, במטרה לזהות אם יש הבדל מובהק סטטיסטי בין הקבוצות עבור כל סנסור. הפונקציה מתחילה ביצירת מילון לאחסון אתרי סנסורים שהראו הבדל מובהק, ומבצעת לולאה על כל מיקום סנסור ייחודי. עבור כל סנסור, היא בודקת אם הוא נמצא ברשימת האזורים הלא ידועים (unknown regions) או אם הוא שייך לאזור מוח בלתי מוגדר. אם הסנסור ייחודי, היא בוחרת את הערכים עבור כל קבוצה (group1 ו-group2) עבור הסנסור הנוכחי ומבצעת את מבחן ה-t על מנת לבדוק אם יש הבדל מובהק בין הקבוצות. אם תוצאת המבחן מראה הבדל מובהק (p-value קטן מהסף שהוגדר - 0.05), הפונקציה מזהה את אזור המוח שבו נמצא הסנסור באמצעות

פונקציה חיצונית (assign_brain_region) ומוסיפה את הסנסור למילון המייצג אזורים עם הבדל מובהק. בסופו של דבר, הפונקציה מדפיסה את האזורים שבהם נמצאו הבדלים מובהקים, עם רשימת הסנסורים שזוהו בכל אזור. אם לא נמצאו תוצאות מובהקות, היא מדפיסה הודעה המתארת שאין הבדל מובהק בסנסורים הנבדקים.

ולבסוף יצרנו את data visualization הכוללת 3 פונק':

time_series_visualization EEG, כדי להראות כיצד ערכי החיישן משתנים לאורך הזמן עבור נושאים שונים.

plot_brain_region_analysis נועדה להציג גרף עמודות הממחיש את ממוצעי ערכי ה-EEG עבור אזורי המוח השונים, מחולקים לפי קבוצות נבדקים. הפונקציה מקבלת כקלט DataFrame שנקרא grouped_data, שבו המידע כבר מקובץ לפי אזורי המוח והקבוצות השונות (למשל, קבוצת ניסוי וקבוצת ביקורת). הפונקציה מתחילה בקיבוץ מחדש של הנתונים לפי שני פרמטרים: אזור המוח (בעמודות region) ומזהה הקבוצה (בעמודות subject_identifier). לאחר מכן, היא מחשבת את ממוצע ערכי ה-EEG לכל קבוצה בכל אזור מוח, ולאחר מכן ממירה את הערכים למוחלטים (על מנת להימנע ממספרים שליליים). הגרף שמתקבל מציג את ממוצע הערכים עבור כל אזור מוח, עם עמודות שונות לכל קבוצה, כך שניתן להשוות בין הקבוצות עבור כל אזור. הגרף כולל כותרת שניתנת על ידי המשתמש (או כותרת ברירת המחדל), ציר ה-X מייצג את אזורי המוח, וציר ה-Y מייצג את ערך הממוצע של ה-EEG.

visualize_all_conditions נועדה להציג גרף עמודות (bar plot) שממחיש את ממוצעי ערכי התגובה עבור כל תנאי ניסוי, תוך השוואה בין קבוצות שונות (כגון קבוצת ניסוי וקבוצת ביקורת). הפונקציה מתחילה בקיבוץ הנתונים לפי התנאים השונים (condition) ומזהה הקבוצה (subject_identifier), ומחשב את ממוצע ערכי התגובה (עמודות value) עבור כל שילוב של תנאי וקבוצה. לאחר חישוב הממוצעים, הפונקציה הופכת את הערכים למוחלטים כדי להימנע מערכים שליליים, אם יש כאלו. הגרף המתקבל מציג את ממוצעי התגובה עבור כל תנאי, כאשר כל תנאי מייצג עמודה נפרדת בציר ה-X, והממוצע של ערכי התגובה מוצג על ציר ה-Y. לכל תנאי יוצגו עמודות נפרדות עבור כל קבוצה, והצבעים השונים בגרף מציינים את הקבוצות (באמצעות הפרמטר hue). בסופו של תהליך, הפונקציה מוסיפה כותרת לגרף, כותרות לצירים ומסדרת את הגרף בצורה שתהיה ברורה וקריאה. לבסוף, הפונקציה מציגה את הגרף, מה שמאפשר השוואה ויזואלית בין תגובות הקבוצות תחת התנאים השונים בניסוי.

לאחר מכן, ייבאנו את הפונק' מקבצים אלו אל ה-MAIN, בו הוספנו הגדרות ספציפיות לפרויקט שלנו (שמות כותרות לגרפים, למשל).

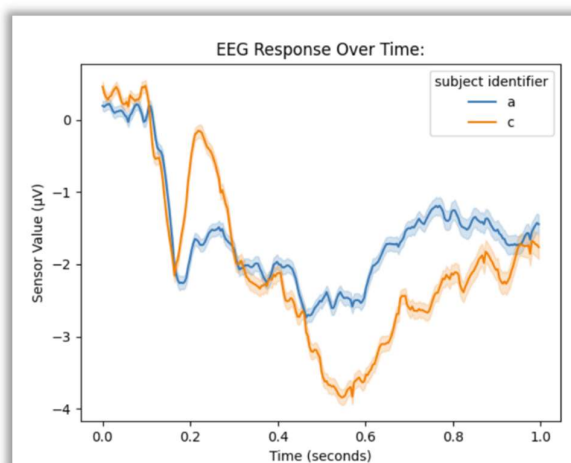
*בנוסף נציין כי עשינו קובץ test-cleaning ו-visualisation. בקבצים הללו יש test לכל פונק' שפירטנו מעלה. בנוסף - בתוך הפונק' המפורטות מעלה, יש הערות שגיאה המועלות בעת הצורך למען ביקורת נוספת.

תוצאות

Top 10 Sensor Positions with Most Pronounced Differences:

subject identifier	a	c	difference
sensor position			
X	0.496181	-1.729530	2.225711
FPZ	-0.209381	-1.771118	1.561737
AFZ	-0.331138	-1.856284	1.525146
FP2	-0.238817	-1.574624	1.335808
AF2	-0.401475	-1.692856	1.291381
FP1	-0.326518	-1.526354	1.199836
Y	-3.080066	-4.240884	1.160818
AF1	-0.528066	-1.636546	1.108479
AF7	-0.605110	-1.707868	1.102758
AF8	-0.379403	-1.472781	1.093378

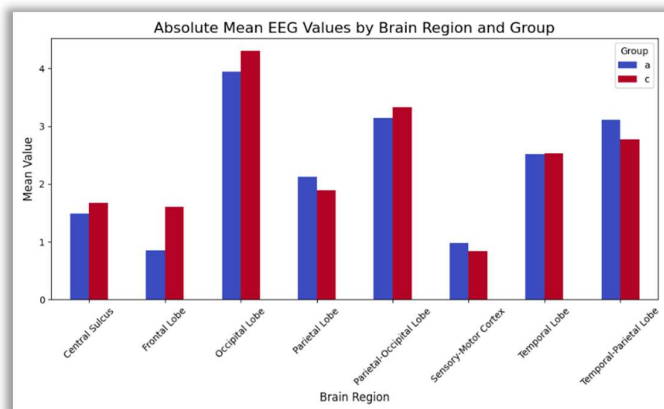
ניתן לראות את תוצאות הניתוח הסטטיסטי שעשינו על מנת לראות באיזה מיקומי חיישנים יראו את ההבדלים הגדולים ביותר. (פירוש יוסבר בגרפים הבאים)



תוצאה: קבוצת האלכוהליסטים מציגה תנודות פחות דרמטיות בערכי ה-EEG לאורך זמן, מה שעשוי לשקף ירידה ביכולת המוח להגיב באופן דינמי לגירויים חיצוניים.

מסקנה/פירוש אפשרי: דפוס זה עשוי להעיד על נזק מבני או שינויים נוירופלסטיים הנגרמים משימוש כרוני באלכוהול, בנוסף עשוי לשקף ירידה בגמישות הקוגניטיבית וביכולת המוח להסתגל לשינויים סביבתיים.

השונויות הגבוהה בקבוצת הביקורת מצביעה על רגישות תקינה לגירויים חיצוניים ועל יכולת להגיב באופן דינמי. לעומת זאת, קבוצת האלכוהוליסטים עשויה להציג עיכוב בתגובה או יכולת מופחתת לעיבוד מידע בזמן אמת.



תוצאה: קבוצת האלכוהוליסטים (a) מציגה ערכי EEG ממוצעים גבוהים יותר ברוב אזורי המוח בהשוואה לקבוצת הביקורת (c), במיוחד באזורים כמו האונה הפריאטלית והאזורים הלא ידועים.

מסקנה/פירוש: הבדלים אלו עשויים להעיד על שינויים נוירופיזיולוגיים הנובעים משימוש כרוני באלכוהול. העלייה בערכים הממוצעים עשויה לשקף פעילות לא סדירה או תגובה מוגברת של חלק מהמוח, ייתכן כתוצאה מפגיעה או שינויי נוירופלסטיות.

האונה הפרונטלית: הערכים הנמוכים יחסית בקבוצת האלכוהוליסטים עשויים לרמז על פגיעה בתפקודי האונה הפרונטלית, הקשורים לתהליכים כמו ויסות תגובות, שליטה עצמית, ותכנון.

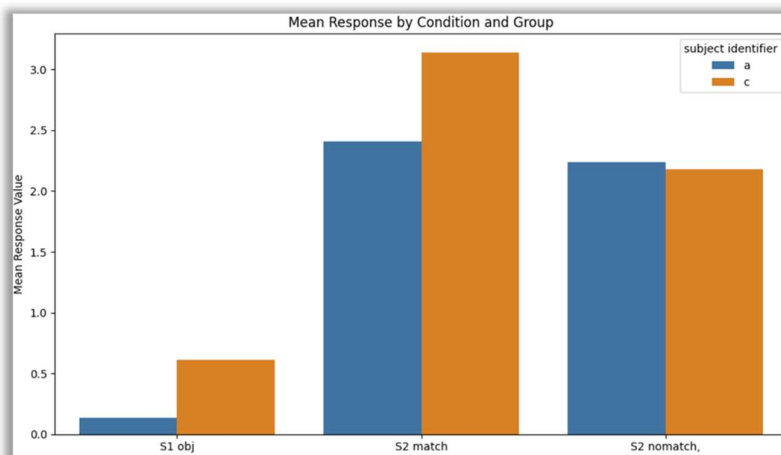
לאחר שיצא לנו הגרף הקודם, רצינו לראות את מובהקות התוצאות. ניתן לראות את החיישנים בהם ההבדלים בין הקבוצות היו מובהקים סטטיסטית. רבים מהם הראו הבדלים משמעותיים, שניתן לראות כתוצאות מוחשיות.

The following regions have shown a significant difference ($p < 0.05$):
 Frontal Lobe: FP1, FP2, F7, F8, AF1, AF2, FZ, F4, F3, FC6, FC5, FC2, FC1, AF7, AF8, F5, F6, FT7, FPZ, FC3, F2, F1, AFZ, FCZ
 Temporal Lobe: T8, T7
 Central Sulcus: C2, C3, C4, C6, C5, C1
 Sensory-Motor Cortex: CP5, CP6, CP1, CP2, CP4, CPZ
 Parietal Lobe: P4, P2, P8, P5, P6, P2, P1
 Parietal-Occipital Lobe: PO2, PO1, PO7, PO8, POZ
 Occipital Lobe: O1, OZ
 Temporal-Parietal Lobe: TP8, TP7

בטבלה ניתן לראות את ממוצע וס"ת של הבדלי התגובות בכל אחת מהתנאים (S1,S2) בין קבוצת האלכוהוליסטים והביקורת.

	matching condition	subject identifier	mean	std
0	S1 obj	a	-0.136911	8.943033
1	S1 obj	c	-0.613380	9.230380
2	S2 match	a	-2.410830	9.311058
3	S2 match	c	-3.136941	10.503128
4	S2 nomatch,	a	-2.235719	9.094835
5	S2 nomatch,	c	-2.179646	9.997804

לאחר מכן, רצינו להמחיש זאת בגרף שמצורף מטה.



כמו שניתן לראות, התגובה המשמעותית ביותר הופקה מתנאי S2 - שבו הגירויים הויזואלים התאימו אחד לשני. תגובה זאת הגיונית כפי שניתן לראות באנליזת הזמן, בגלל שתגובות האלכוהוליסטים היו בעלות אמפליטודה יותר נמוכה.

דיון

השימוש הכרוני באלכוהול גורם להפחתה בשונות ובדינמיות של הפעילות המוחית, דבר שעשוי להצביע על פגיעות בתפקודים קוגניטיביים כמו קשב, זיכרון עבודה ושליטה ניהולית. השינויים הפחות דרמטיים שנראים בקבוצת האלכוהוליסטים עשויים לשקף ירידה ברגישות המוח לגירויים חיצוניים, כתוצאה מנוירואדפטציה או דה-סנטיזציה. בנוסף, השפעות אלו עשויות להיות קשורות לנזק באונה הפרונטלית, המעורב בוויסות התגובות לגירויים. במקביל, קבוצת הביקורת מציגה תגובתיות מוחית גמישה ותקינה, המעידה על יכולת עיבוד קוגניטיבי ותחושת בריאה. בסך הכל, השפעות השימוש הכרוני באלכוהול על פעילות המוח מעידות על פגיעות ארוכות טווח, כולל ירידה בתפקוד קוגניטיבי ופוטנציאל לנזק נוירולוגי.