

ניתוח אותות EEG לא-לינאריים

לסיווג מחלת פרקינסון באמצעות סטטיסטיקה מסדר גבוה

אסף רפאל דאבוש, אביחי שיאחי

Full Feature Extraction & Classification Summary

תקציר

בעבודה זו בוצע מימוש חישובי מלא ונאמן למאמר מחקרי העוסק באבחון מחלת פרקינסון באמצעות ניתוח לא-לינארי וסטטיסטיקה מסדר גבוה של אותות EEG. העבודה מתמקדת בשחזור מדויק של מתודולוגיית המאמר, החל משלב עיבוד הנתונים הגולמיים, דרך חילוף מאפיינים מתקדם, ועד שלב הסיווג והערכת הביצועים.

אותות ה-EEG עברו שרשרת עיבוד מקדים קפדנית, שכללה סינון, הסרת ארטיפקטים פיזיולוגיים, ניתוח ICA וחלוקה למקטעים לא-חופפים באורך 2 שניות בתדר דגימה של 500 Hz. מכל מקטע חולצו 23 מאפיינים לכל ערוץ, ללא ביצוע ממוצע בין ערוצים, בהתאם למאמר המקורי.

הסיווג בוצע באמצעות מסווגי Support Vector Machine ו-Bagged Trees, תוך שימוש באימות צולב מסוג K-fold וברמת המקטע. נבחנו שלושה תחומי תדר: האות המלא, תחום האלפא ותחום הבטא.

התוצאות מצביעות על יתרון מובהק לניתוח האות המלא בשילוב מסווג Bagged Trees, אשר השיג דיוק מרבי של 93.82% ורגשות מלאה (Sensitivity של 100%), לעומת ביצועים נמוכים יותר בעת ניתוח תחומי תדר מבודדים. ממצאים אלו מדגישים את תרומת האינטראקציות הבין-תדריות ואת התאמתם של מסווגי אנסמבל לעבודה עם מערכי מאפיינים מרובי-ממד במשימת אבחון מחלת פרקינסון.

1 מאגר הנתונים

מאגר הנתונים ששימש בעבודה זו כולל הקלטות EEG של חולי פרקינסון ושל נבדקים בריאים, אשר נמדדו בתנאי מנוחה. המאגר נאסף במסגרת מחקר קליני, והוא משמש במאמר המקורי כבסיס לבחינת השפעת מחלת פרקינסון על דינמיקת הפעילות העצבית.

בהתאם לפרוטוקול המחקרי ולמתודולוגיה המתוארת במאמר, נותחו אך ורק הקלטות שנמדדו במצב OFF-medication, כלומר ללא השפעת טיפול תרופתי. בחירה זו נועדה למנוע הטיה הנובעת מהשפעת תרופות על פעילות המוח, ולאפשר זיהוי מאפיינים עצביים הקשורים ישירות למחלה עצמה.

לאחר סינון ההקלטות, נכללו בניתוח 56 נבדקים. לכל נבדק הופקו 20 מקטעי EEG לא חופפים באורך 2 שניות, ובתדר דגימה של 500 Hz. כתוצאה מכך התקבלו 1120 מקטעים לכל תחום תדר, ובסך הכול 3360 מקטעים בשלושת תחומי התדר שנבחנו.

מבנה זה של המאגר מאפשר ניתוח סטטיסטי מבוסס מקטעים, תוך שמירה על אחידות בין הנבדקים ועל איזון בין קבוצות המחקר.

2 עיבוד מקדים של האות

שלב העיבוד המקדים מהווה רכיב מרכזי בעבודה עם אותות EEG, שכן האותות הגולמיים רגישים במיוחד לרעשים ולארטיפקטים פיזיולוגיים וסביבתיים. לפיכך, יושמה שרשרת עיבוד מקדים קפדנית, הנאמנה למתודולוגיית המאמר המקורי.

בשלב הראשון הוסרו ערוצים שאינם רלוונטיים לניתוח EEG, על מנת לצמצם השפעות לא-עצביות. בהמשך בוצעה דגימה מחדש של האותות לתדר דגימה אחיד של 500 Hz, בהתאם לדרישות המאמר.

לאחר מכן יושם סינון תחום תדרים כללי (0.1–100 Hz) באמצעות מסנן IIR, שמטרתו הסרת רכיבי תדר נמוכים מאוד ורכיבי רעש בתדרים גבוהים. בהמשך בוצעה הסרת ארטיפקטים פיזיולוגיים, ובפרט רעשי עיניים (EOG) ורעשי שרירים (EMG), באמצעות אלגוריתמים אוטומטיים.

בנוסף, בוצע ניתוח Independent Component Analysis (ICA), שנועד להפריד בין מקורות עצמאיים באות ולהפחית השפעת רעשים שנותרו לאחר שלבי הסינון. לבסוף חולקו האותות למקטעים לא חופפים באורך 2 שניות, כך שכל מקטע כלל 1000 דגימות.

לצורך בחינת תרומת תחומי תדר שונים, נוצרו שלושה מאגרי נתונים נפרדים: האות המלא, תחום אלפא (8–12 Hz), ותחום בטא (13–30 Hz).

3 חילוף מאפיינים

חילוף המאפיינים בוצע בהתאם לגישת המאמר, תוך דגש על ניתוח לא-ליניארי וסטטיסטיקה מסדר גבוה, אשר נועדו לחשוף מאפיינים דינמיים מורכבים של אותות EEG. מכל מקטע חולצו 23 מאפיינים לכל ערוץ, ללא ביצוע ממוצע בין ערוצים, על מנת לשמר מידע מרחבי מלא.

הקבוצה הראשונה כוללת מאפיינים לא-ליניאריים, וביניהם אקספוננט ליאפונוב המרבי, המודד את רגישות המערכת לתנאי התחלה, מעריך הרסט, המשקף זיכרון ארוך-טווח באות, אנטרופיה מקורבת, המשמשת למדידת מורכבות האות, וממד הקורלציה, המאפיין את ממדיות האטרקטור הדינמי.

הקבוצה השנייה כוללת מומנטים סטטיסטיים מסדר גבוה, מהדרגה השלישית ועד השמינית, המחושבים על אות מנורמל. מאפיינים אלו מאפשרים תיאור של סטיות מהתפלגות גאוסית, ומשקפים אסימטריה ומידת

קיצוניות בהתפלגות האות.

הקבוצה השלישית מבוססת על סטטיסטיקה מסדר גבוה בתחום התדר, ונשענת על ניתוח הביספקטרום והביספסטרום. מאפיינים אלו מאפשרים זיהוי אינטראקציות לא-ליניאריות בין רכיבי תדר שונים, וכן חישוב מדדי אנטרופיה המבטאים את מורכבות המבנה הספקטרלי של האות. בנוסף, חושבו שלוש אנטרופיות ביספקטרליות: normalized bispectral entropy (BE1), normalized bispectral squared entropy (BE2), ו-bispectral phase entropy (BEp), בהתאם להגדרות המופיעות במאמר.

בסך הכול התקבל וקטור מאפיינים באורך 1472 לכל מקטע (64 ערוצים כפול 23 מאפיינים), המהווה ייצוג עשיר ומפורט של הפעילות העצבית.

3.1 ניסוח מתמטי של חילוף המאפיינים

בחלק זה מוצג הניסוח המתמטי של המאפיינים שחולצו מאותות ה-EEG, בהתאם למתודולוגיה המתוארת במאמר. החילוף מבוסס על ספקטרומים מסדרים גבוהים (Higher-Order Spectra) ונגזרותיהם.

3.1.1 הגדרות בסיסיות

יהי $x[n]$ אות EEG בדיד באורך N , לאחר הסרת הממוצע. התמרת פורייה של האות מוגדרת כ-

$$X(f) = \mathcal{F}\{x[n]\}$$

כאשר f מייצג תדר בדיד, ו- $(\cdot)^*$ מציינ צמוד מרוכב.

3.1.2 ספקטרום ההספק (Power Spectrum)

ספקטרום ההספק מסדר שני מוגדר כ-

$$P(f) = X(f) X^*(f) \quad (1)$$

3.1.3 ביספקטרום (Bispectrum)

הביספקטרום, המתאר אינטראקציות לא-ליניאריות מסדר שלישי בין רכיבי תדר, מוגדר כ-

$$B(f_1, f_2) = X(f_1) X(f_2) X^*(f_1 + f_2) \quad (2)$$

3.1.4 טריספקטרום (Trispectrum)

הטריספקטרום מסדר רביעי מוגדר כ-

$$T(f_1, f_2, f_3) = X(f_1) X(f_2) X(f_3) X^*(f_1 + f_2 + f_3) \quad (3)$$

3.1.5 ביצפסטרום (Bicepstrum)

הביצפסטרום מוגדר כהתמרת פורייה הפוכה דו-ממדית של הלוגריתם של הערך המוחלט של הביספקטרום:

$$C_b(q_1, q_2) = \mathcal{F}_2^{-1} \{ \log(|B(f_1, f_2)|) \} \quad (4)$$

כאשר q_1, q_2 הם אינדקסי biquency.

3.1.6 מומנטים מרכזיים מתוקננים מסדר גבוה

יהי x משתנה מקרי המייצג דגימות של אות, עם תוחלת $\mathbb{E}[x]$ וסטיית תקן

$$\sigma = \sqrt{\mathbb{E}[(x - \mathbb{E}[x])^2]}$$

המומנט המרכזי המתוקנן מסדר k מוגדר כ-

$$\mu_k = \frac{\mathbb{E}[(x - \mathbb{E}[x])^k]}{\sigma^k}, \quad k = 3, \dots, 8 \quad (5)$$

3.1.7 מאפייני ביספקטרום (H1-H5)

בהתאם למאמר, הוגדרו חמישה מאפיינים המבוססים על הביספקטרום:

$$H1 = \sum_{f_1, f_2} |B(f_1, f_2)| \quad (6)$$

$$H2 = \sum_{f_1, f_2} |B(f_1, f_2)|^2 \quad (7)$$

$$H3 = \sum_{f_1, f_2} \log(|B(f_1, f_2)|) \quad (8)$$

$$H4 = \frac{\sum |B(f_1, f_2)|^2}{\sum |B(f_1, f_2)|} \quad (9)$$

$$H5 = \frac{\sum |B(f_1, f_2)|}{\max(|B(f_1, f_2)|)} \quad (10)$$

3.1.8 מאפייני ביצפסטרום (BH1-BH5)

באופן דומה, חמישה מאפיינים חושבו על בסיס הביצפסטרום:

$$BH1 = \sum_{q_1, q_2} |C_b(q_1, q_2)| \quad (11)$$

$$BH2 = \sum_{q_1, q_2} |C_b(q_1, q_2)|^2 \quad (12)$$

$$BH3 = \sum_{q_1, q_2} \log(|C_b(q_1, q_2)|) \quad (13)$$

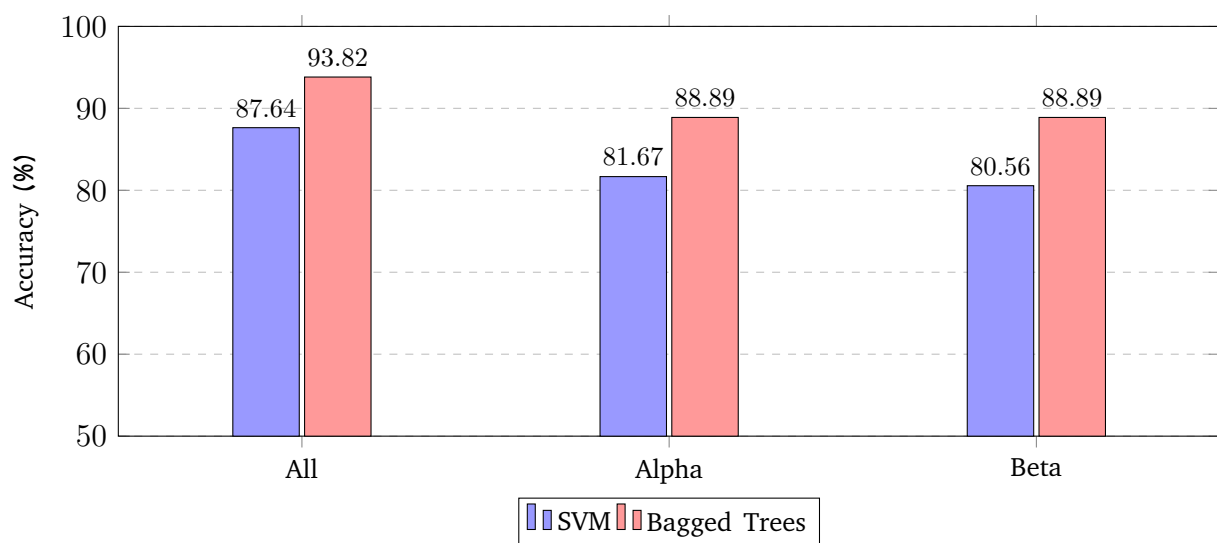
$$BH4 = \frac{\sum |C_b(q_1, q_2)|^2}{\sum |C_b(q_1, q_2)|} \quad (14)$$

$$BH5 = \frac{\sum |C_b(q_1, q_2)|}{\max(|C_b(q_1, q_2)|)} \quad (15)$$

יש לציין כי הנוסחאות המוצגות מהוות ייצוג כללי של המאפיינים, בעוד שהמימוש החישובי בוצע בהתאם להגדרות המדויקות והאזורים הייעודיים (אלכסון ו-principal domain) כמתואר במאמר המקורי.

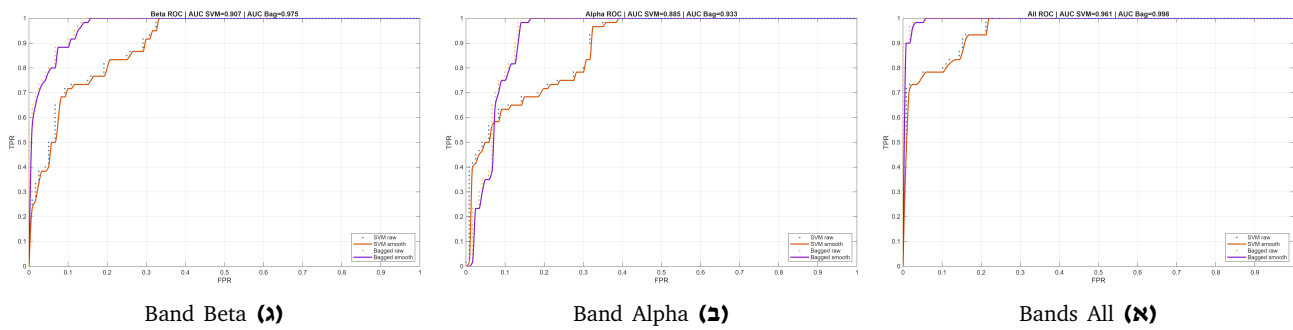
טבלה 1: סיכום ביצועי הסיווג לפי תחום תדר ומודל

Band Frequency	Classifier	(%) Accuracy	(%) Sensitivity	(%) Specificity
Bands All	Bagged Trees	93.82	100.00	90.68
	SVM	87.64	78.33	92.37
Alpha	Bagged Trees	88.89	100.00	83.33
	SVM	81.67	58.33	93.33
Beta	Bagged Trees	88.89	100.00	83.33
	SVM	80.56	55.00	93.33



איור 1: השוואה גרפית של דיוק הסיווג עבור שני המסווגים ושלושת תחומי התדר.

5 עקומות ROC



איור 2: עקומות ROC המתקבלות עבור שלושת תחומי התדר.

6 דיון

ניתוח תוצאות הסיווג מצביע על יתרון מובהק לשימוש באות EEG המלא בהשוואה לניתוח תחומי תדר מבודדים. בתצורת All-bands הושגו רמות הביצוע הגבוהות ביותר, כאשר מסווג Bagged Trees השיג דיוק של 93.82%, לעומת 87.64% במסווג SVM.

ממצא זה תואם את ההנחה כי מאפיינים לא-ליניאריים וסטטיסטיקה מסדר גבוה לוכדים אינטראקציות בין-תדריות מורכבות, אשר אינן נשמרות במלואן בעת סינון לתחום תדר צר. בהתאם לכך, ברצועות Alpha ו-Beta נרשמה ירידה מתונה בדיוק הכולל, עם ערכים זהים של 88.89% עבור מסווג Bagged Trees, וירידה חדה יותר בביצועי מסווג ה-SVM.

בנוסף, נצפתה רגישות מלאה (100%) במסווג Bagged Trees בכל תחומי התדר, לעומת ירידה משמעותית ברגישות מסווג ה-SVM, בפרט ברצועות האלפא והבטא. פער זה מדגיש את יתרונם של מסווגי אנסמבל בעת עבודה עם מערך מאפיינים רחב ורב-ממדי.

7 סיכום ומסקנות

בעבודה זו בוצע מימוש חישובי מלא ונאמן למאמר מחקרי העוסק באבחון מחלת פרקינסון באמצעות ניתוח לא-ליניארי וסטטיסטיקה מסדר גבוה של אותות EEG. מערך הנתונים כלל עד 180 מקטעי אות, עם התפלגות של 60 נבדקי PD ו-120 נבדקי ביקורת, כאשר מכל מקטע חולצו 1104 מאפיינים ללא ממוצע בין ערוצים.

התוצאות מצביעות על יתרון ברור לניתוח האות המלא (All-bands), בו הושג הדיוק הגבוה ביותר, Accuracy של 93.82% באמצעות מסווג Bagged Trees. ברצועות האלפא והבטא התקבל דיוק זהה של 88.89%, אך עם ירידה בספציפיות ביחס לניתוח האות המלא.

בנוסף, הודגשה השפעת בחירת המסווג על ביצועי הסיווג: מסווגי אנסמבל הציגו רגישות מלאה ויציבות גבוהה בעת עבודה עם מאפיינים מרובי-ממד, בעוד שמסווג SVM הציג ספציפיות גבוהה אך רגישות נמוכה יותר, בעיקר בתחומי תדר צרים.

ממצאים אלו מחזקים את הטענה כי שילוב של ניתוח לא-ליניארי מתקדם עם מסווגי אנסמבל מהווה גישה יעילה ואמינה לאבחון מחלת פרקינסון מתוך אותות EEG, ומספק בסיס מתודולוגי מוצק להמשך פיתוח מערכות סיוע קליניות.