

# **ניתוח אוטות EEG לא-לינאריים**

## **לסיווג מחלת פרקינסון באמצעות סטטיסטיקה מסדר גבוה**

## אסף רפאל דאבורש, אביהו שייחי

Full Feature Extraction & Classification Summary

תקאייר

בעבודה זו בוצע מימוש חישובי מלא ונאמן למאמר מחקרי העוסק באבחון מחלת פרקיינסן באמצעות ניתוח לא-לינארי וסטטיסטיקה מסדר גבוהה של אוטות EEG. הבדיקה מתמקדת בשחזר מדויק של מתודולוגיות המאמר, החל משלב עיבוד הנתונים הגולמיים, דרך חילוץ מאפיינים מתקדם, ועד שלב הסיווג והערכת הביצועים.

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

הסיווג בוצע באמצעות מסוגי Bagged Trees ו-Support Vector Machine, תוך שימוש באימוט צולב K-fold וברמת המקטע. נבחנו שלושה תחומי תדר: האות המלא, תחום האלפא ותחום הבטא. התוצאות מצביעות על יתרון מובהק לניתוח האות המלא בשילוב מסוג 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 מסוגי אנסמבל לעובדה עם מערכיו מאפייני מרובי-ממך במשימות אבחון מחלת פרקיינסן.

# 1 מאגר הנתונים

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

בהתאם לפרטוקול המחקרי ולמתודולוגיה המתוארת במאמר, נתחו אץ וرك הקלטות שנמדדו במצב-OFF, ככלומר ללא השפעת טיפול רפואי. בחירה זו נועדה למגוון הטיה הנובעת מהשפעת תרופות על פעילות המוח, ולאפשר זיהוי מאפיינים עצביים הקשורים ישירות למחלת עצמה.

לאחר סינון הקלטות, נכללו בניתוח 56 נבדקים. לכל נבדק הופקו 20 מקטיעי EEG לא חופפים באורך 2 שניות, ובתדר דגימה של Hz 500. כתוצאה לכך התקבלו 1120 מקטיעים לכל תחום תדר, ובסך הכל 3360 מקטיעים בשלושת תחומי התדר שנבחנו.

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

## 2 עיבוד מקדים של האות

שלב העיבוד המקדים מהווה רכיב מרכזי בעבודה עם אותן EEG, שכן האותות הגולמיים רגיסרים במיוחד לרעשים ולארטיפקטים פיזיולוגיים וסבירתיים. לפיכך, יושמה שרשרת עיבוד מקדים קפדיית, הנאמנה למетодולוגיית המאמר המקורי.

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

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

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

לצורך בוחנת תרומות תחומי תדר שונים, נוצרו שלושה מאגרי נתונים נפרדים: האות המלא, תחום אלף Hz (12–8), ותחום בטא (30–12).

## 3 חילוץ מאפיינים

חילוץ המאפיינים בוצע בהתאם לגישת המאמר, תוך דגש על ניתוח לא-لينארי וסטטיסטי מסדר גובה, אשר נועד לחושף מאפיינים דינמיים מורכבים של אותן EEG. מכל מקטע חולצו 23 מאפיינים לכל ערז, ללא ביצוע ממוצע בין ערזים, על מנת לשמר מידע מרחבית מלא.

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

הקובוצה השנייה כוללת מומנטים סטטיסטיים מסדר גובה, מהדרגה השלישי ועד השמינית, המוחשבים על אות מנורמל. מאפיינים אלואפשרים תיאור של סטיות מהתפלגות גausית, ומשקפים אסימטריה ומידת

קייזוניות בהתפלגות האות.

הקבוצה השלישית מבוססת על סטטיסטיקה מסדר גבוהה בתחום התדר, ונשענת על ניתוח הביספקטרום והביספsectorום. מאפיינים אלו מאפשרים זיהוי אינטראקציות לא-ליניאריות בין רכיבי תדר שונים, וכן חישוב מדדי אנטרופיה המבטים את מרכיבות המבנה הספקטורי של האות. בנוסף, חשובו שלוש אנטרופיות bispectral, normalized bispectral squared entropy (BE1), normalized bispectral entropy (BE2) ו-phase entropy (BEP), בהתאם להגדרות המופיעות במאמר.

בsek הכל התקבל וקטור מאפיינים באורך 1472 לכל מקטע (64 ערכאים כפול 23 מאפיינים), המהווה ייצוג עשיר ומפורט של הפעולות העצביות.

### 3.1 ניסוח מתמטי של חילוץ המאפיינים

בחלק זה מוצג הניסוח המתמטי של המאפיינים שחולצו מאותות ה-EEG, בהתאם למתודולוגיה המתוארת במאמר. החילוץ מבוסס על ספקטרומים מסדריים 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- $\cdot^*(\cdot)$  מצין 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)

הbisפקטרום, 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)

הביביצפסטרום מוגדר כהתמרת פורייה הפוכה דומדמית 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 מתוקנים מסדר גובה

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

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

המומנט המרכזי המתוקן 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 ביספקטורים (H1–H5)

בהתאם למאמר, הוגדרו חמישה מאפיינים המבוססים על הביספקטורים:

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

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

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

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

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

### 3.1.8 מאפייני ביצפסטרום (BH1–BH5)

באופן דומה, חמישה מאפיינים חושבו על בסיס הביצפסטרום:

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

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

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

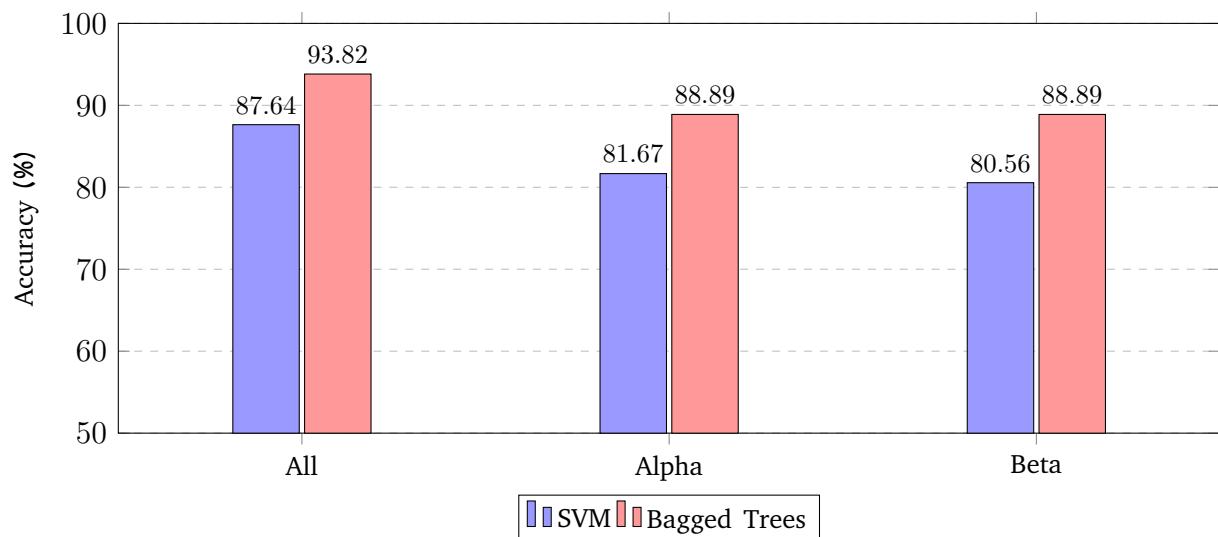
$$BH4 = \frac{\sum |C_b(q_1, q_2)|^2}{\sum |C_b(q_1, q_2)|} \quad (14)$$

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

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

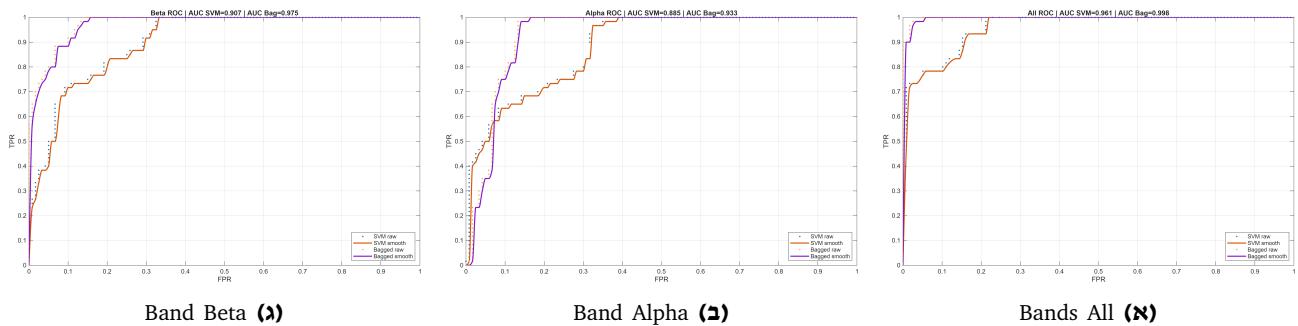
טבלה 1: סיכום ביצועי הסיווג לפי תחומי תדר ומודל

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



איור 1: השוואת גרף של דיקט הסיווג עבור שני המסוחים ושלושת תחומי התדר.

## 5 ROC עקומות



איור 2: עקומות ROC המתקבלות עבור שלושת תחומי התדר.

## 6 דיוון

ניתוח תוצאות הסיווג מצבי על יתרון מובהק לשימוש EEG המלא בהשוואה לנитוח תחומי תדר מבודדים. בתצורת All-the-bands הושגו רמות הביצוע הגבוהות ביותר, כאשר מסובג Bagged Trees השיג דיוק של 93.82%, לעומת 87.64% במסובג SVM.

ממצא זה תואם את ההנחה כי מאפיינים לא-לינאריים וסטטיטיסטיקה מסדר גבוה יותר לוכדים אינטראקטיביות בין-תדריות מורכבות, אשר אין נשמרות במלואן בעת סינון לתהום תדר צר. בהתאם לכך, ביציאות Alpha ו-Beta נרשמה ירידיה מתונה בדיק ה כולל, עם ערכיהם זהים של 88.89% עבור מסובג Bagged Trees, וירידיה חדה יותר ביצועי מסובג SVM.

בנוסף, נצפתה רגישות מלאה (100%) במסובג Bagged Trees בכל תחומי התדר, לעומת ירידיה משמעותית ברגישות מסובג SVM, בפרט ביציאות האלפא והבטא. פער זה מדגיש את יתרונות של מסובי אנסמבל בעת עבודה עם מערכ מאפיינים רחב ורב-מדדי.

## 7 סיכום ומסקנות

בעבודה זו בוצע מימוש חשוב מלא ונאמן למאמר מחקרי העוסק באבחון מחלת פרקינסן באמצעות ניתוח לא-לינארי וסטטיטיסטיקה מסדר גבוה של אותן EEG. מערכת הנתונים כלל עד 180 מקטיעות אחרות, 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 ביותר, אשר מכל מקטע חולצו 1104 מאפיינים ללא ממוצע בין ערוצים. ירידיה בספציפיות ביחס לניתוח האות המלא.

בנוסף, הודגשה השפעת בחירת המסובג על ביצועי הסיווג: מסובי אנסמבל הציגו רגישות מלאה ויציבות גבוהה בעת עבודה עם מאפיינים מרובי-מדדי, בעוד שמסובג SVM הציג ספציפיות גבוהה אך רגישות נמוכה יותר, בעיקר בתחום תדר צרים.

ממצאים אלו מוכיחים את הטענה כי שילוב של ניתוח לא-לינארי מתקדם עם מסובי אנסמבל מהוות גישה יעילה ואמינה לאבחן מחלת פרקינסן מתוך אותן EEG, וספק בסיס מתודולוגי מוצק להמשך פיתוח מערכות סיוע קליניות.