**פרויקט סופי בתכנות מתקדם בפיתון למדעי המוח**

**“Sleep Analysis - AI vs Experts”**

**מגישות**:

שי ימיני 322994344

שילת אלגאלי 212739692

**רקע**

רבים יסכימו כי העידן המודרני, המביא עימו התפתחויות טכנולוגיות חדשות ומחקרים מרתקים, משפיע באופן חיובי על החיים של כולנו ויוצר פיתוחים חדשים המקלים על חיי היום יום של כל אחת ואחד מאיתנו. עם זאת, התפתחויות טכנולוגיות אלו מכילות בתוכם גם צדדים שאינם חיוביים, לרבות חשיפה מרובה למסכים הפוגעים בהפרשות חלבונים קריטיים למחזורי שינה תיקניים בנוסף לחשיפה לגורמים מעוררי לחץ המופיעים ברשתות החברתיות ובאתרי החדשות. דברים אלו עלולים לפגוע קשות באיכות השינה של אנשים רבים, ובגלל החשיבות של שינה לתפקוד הכללי של הגוף, הפרעות השינה הללו יכולות להשפיע גם על כלל תחומי חייהם של האנשים הסובלים מהם. כיום קיימת שיטה סטנדרטית לזיהוי הפרעות שינה בקרב מטופלים, אך סדיקה זו דורשת הגעה למעבדת שינה מיוחדת, בה מחברים את המטופלים במהלך הלילה למכשירים רבים על מנת לעקוב אחר מדדים כמו קצב לב, תנועות עיניים, תנועות גוף אחרות והרבה פעמים גם כולל חיבור לאלקטרודות. שיטה זו אינה יעילה מספיק, שכן היא מערבת מכונות רבות ואנשי מקצוע שתפקידם לעבור על תוצאות הבדיקה, מה שעולה כסף רב, היא דורשת קביעת תור למעבדת שינה, מה שמבזבז את זמנו של המטופל שכן לפעמים לוקח חודשים עד שמתפנה תור למעבדה, והיא דורשת מהמטופל לבלות את הלילה בחדר לא מוכר בעודו מחובר למכשירים רבים, מה שמוביל לחוויה לא נעימה עבור המטופלים.

במהלך החיפוש שלנו אחר דאטא לשימוש בפרויקט, נתקלנו בדאטא מסקרן מאוד הבוחן טכנולוגיה חדשה שמערבת שימוש בבינה מלאכותית (AI). הטכנולוגיה המדוברת היא סרט ראש (headband), אותו הנבדקים שמים לפני שהם הולכים לישון. אותו סרט משתמש ב-EEG על מנת לקלוט גלי מוח במהלך השינה והבינה המלאכותית מסווגת לפי הנתונים שהיא מקבלת מה-EEG באיזה שלב שינה המטופלים נמצאים בהפרשים של 30 שניות; N1,N2,N3,REM sleep.

הדאטא מכילים מידע שנאסף מנבדק אחד במשך 30 לילות בהם הוא היה מחובר גם למכשירים בהם משתמשים במהלך הבדיקה במעבדת שינה (PSG-Polysomnography) וגם לאותו סרט-הראש. המידע הועבר לשני קבצי TSV נפרדים, כאשר כל קובץ מכיל טבלה בה כל שורה מייצגת משך זמן של 30 שניות מתוך הלילה, וקיימות עמודות המתארות מהו מצב השינה של הנבדק בכל שלב. בקובץ ה-headband, העמודה הרלוונטית שעניינה אותנו היא עמודה בה מוצגים שלבי השינה שהגדיר ה-AI לפי הנתונים שהתקבלו מסרט הראש. לעומת זאת, בקובץ ה-PSG הופיעו שתי עמודות רלוונטיות שסיקרנו אותנו, כאשר העמודה הראשונה הציגה את הקטלוג שהעניק ה-AI לשלבי השינה השונים על פי נתונים שהתקבלו ממכשירי מעבדת השינה והעמודה השניה הציגה את הקטלוג לשלבי השינה השונים שהעניקו שלושה מומחים בתחום חקר השינה בהתאם לתוצאות מכשירי מעבדת השינה.

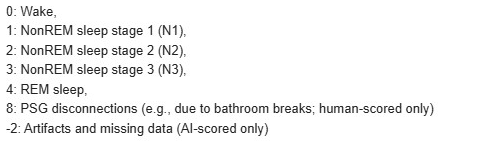
הקוד שכתבנו מנסה להקל על החוקרים שפרסמו את המאמר לנתח את הנתונים שהם אספו ממחקר זה, במטרה להבין האם ניתן להחליף את מעבדת השינה בטכנולוגיה פשוטה, זולה ונוחה יותר לשימוש ביתי המתבססת על AI ועל סריקות EEG ביתיות.

**שיטות**

הקוד בנוי מ-6 פונקציות, ועוד פונקציה אחת ליצירת גרף המתאר את ההבדלים בין שלושת הקביעות השונות בלילה ספציפי כלשהו.

פונקצייה ראשונה- מקבלת את התיקייה הכללית המכילה את התיקיות לפי לילות, מושכת את כל הקבצים הרלוונטים מהתיקיות השונות ומחלקת אותם ל-2 רשימות: רשימת קבצי הנתונים מהheadband ורשימת קבצי הנתונים מה-PSG, הרשימות זה מה שהפונקציה מחזירה.

תפקידן של שתי הפונקציות הבאות הן לקבוע את אחוז ההתאמה בין הAI למומחים, אחת הפונקציות משווה את קביעת המומחים לקביעת הAI שמקבלת נתונים מהheadbend והפונקצייה השניה משווה את קביעת המומחים לקביעת הAI שמקבלת נתונים מהPSG. שתי הפונקציות פועלות דומה, כל אחת מהן קודם כל מנקה את הנתונים. הפונקציות הללו מחזירות גם לוג בו נעשית השוואה בין קביעת המומחים לבין קביעת ה-AI על פי הנתונים מה-PSG, בנוסף להשוואה בין קביעת המומחים לבין קביעת ה-AI על פי הנתונים מה-headband עבור כל קובץ בנפרד. מצבי השינה מוצגים בעזרת מספרים 0-4, ולפי מייל שקיבלנו מעורכי המחקר זוהי הדרך לקרוא את הנתונים:



הפונקציות מנקות את הנתונים, כל אחת בהתאמה לקבצים שהיא מקבלת.

הפונקציה שמשווה בין דעת המומחים ל-headband מסננת החוצה את כל השורות (משני הקבצים) שקוטלגו כמצב שינה "2-" או "8", מאחר וזו מציינת שגיאה, אך קודם לכן היא בודקת האם שגיאה זו מכילה 40% או יותר מהמסמך, מכיוון שבמקרה כזה כנראה הייתה איזו שגיאה עם סרט הראש וכל הנתונים המתקבלים לא שמישים, במקרה זה הפונקציה תחזיר None.

הפונקציה שמשווה בין המומחים לAI של PSG מסננת גם היא את מה שמקוטלג כ "8" וכאמור, שתי הפונקציות מחזירות את אחוז ההתאמה בין קביעת המומחים לAI.

הפונקציה הרביעית שכתבתנו מחשבת כמה שעות שינה בלילה היו בעלות נתונים חסרים מה-headband, והפונקציה החמישית מחשבת בסך הכל כמה שעות שינה היו בלילה הנתון (מטרת הפונקציה הזאת היא לראות האם מכשיר ה-EEG הביתי בכלל שמיש או שמה היא מחזיר תוצאות חלקיות ולא תקינות לעיתים קרובות). בנוסף, קיימת פונקציה שמטרתה לאפשר למשתמש לחפש את הנתונים שמעניינים אותו ולשלוף גרףים הרלוונטיים ללילות ספציפים בהם המשתמשים מעוניינים להתעמק. הפונקציה עושה זאת על ידי הרצת לולאת פור שמצוותת בין כל שני קבצים המתייחסים לאותו יום הניסוי. לאחר מכן, הפונקציה נכנסת ללולאת ווייל בה היא מבקשת אינפוט מהמשתמש באשר למספר לילה ספציפי בו הוא רוצה להתעמק ומציגה לו גרף הבנוי על הנתונים שהתקבלו מזוג הקבצים הרלוונטיים למספר הלילה שחיפש המשתמש. לאחר מכן, הפונקציה שואלת את המשתמש האם יהיה מעוניין לבחון קבצים מלילות נוספים, והיא תאפשר לו לבחור קבצים שונים עד שיענה כי אינו מעוניין לבחון קבצים נוספים.

בנוסף, קיימת פונקציה שמטרתה להמחיש בצורה נוחה למשתמש את שאירע במהלך הלילה. הפונקציה עושה זאת על ידי יצירת שני גרףים: הראשון מקבל את שלבי השינה שקבעו המומחים ומציג את שלבי השינה השונים שחווה הנבדק במהלך הלילה באמצעות scatter plot, בזמן שהשני מציג בו זמנית את שלבי השינה שקבעו המומחים בהתבסס על ה-PSG, את שלבי השינה שקבע ה-AI בהתבסס על ה-PSG ואת שלבי השינה שקבע ה-AI בהתבסס על ה-headband באמצעות שלושה line plots שונים.

בסופו של דבר שש הפונקציות האלה נכנסות בקוד הראשי, שם הרצנו את רשימות הקבצים דרך לולאת for שכל פעם לוקחת את שני הקבצים המתאימים לכל לילה ומכניסה אותם לפונקציות. אחוזי ההתאמה שמתקבלים חזרה נכנסים לרשימות המתאימות להם וחישוב שעות השינה החסרות או הכלליות מצטרפות לסכימה כללית של הערכים היוצאים.

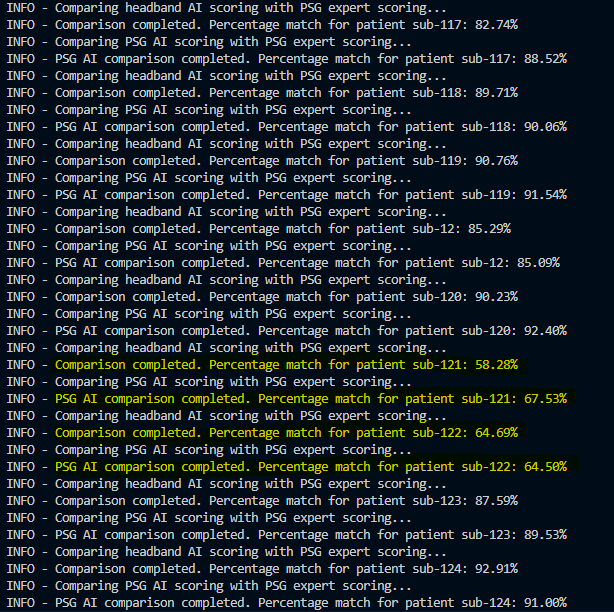
בסוף מחשבים ממוצע של כל רשימה סופית (2 בסך הכל- מומחים מול headbend וAI PSG מול מומחים) וכמה שעות שינה חסרות היו מתוך כלל שעות השינה.

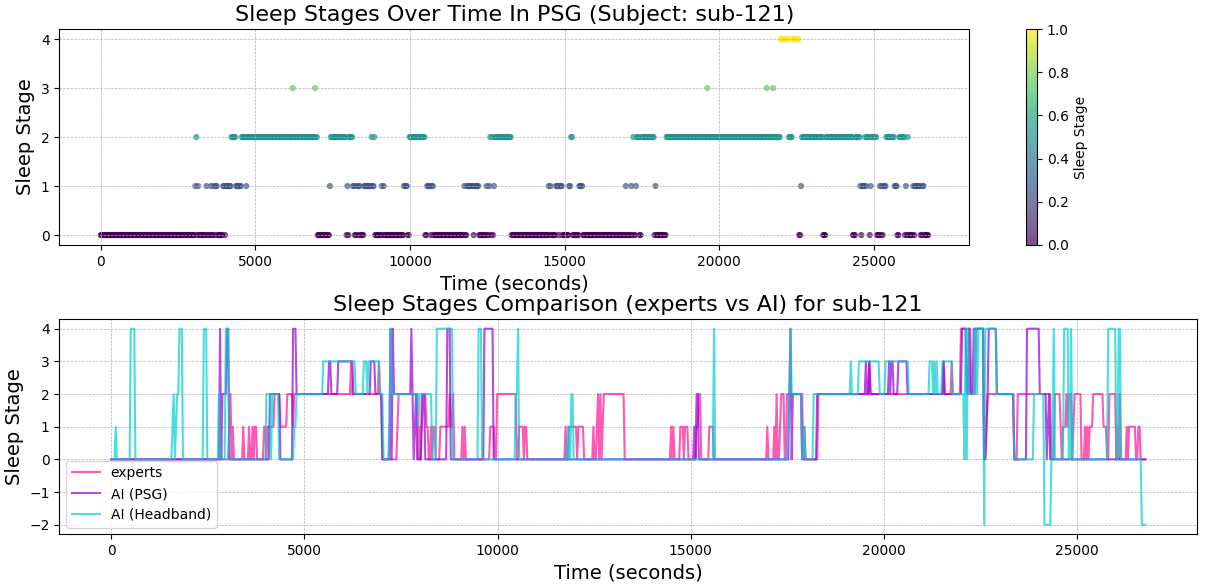
**תוצאות ודיון**

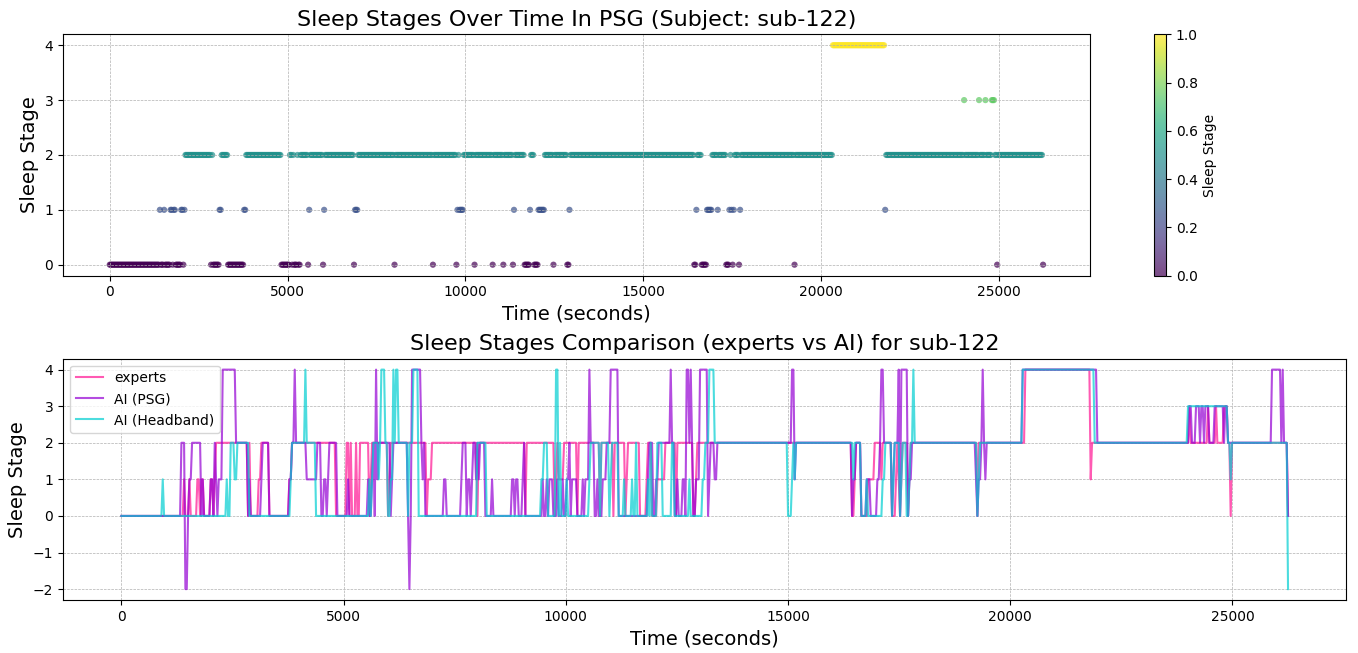
התוצאות הסופיות שלנו:

****

תוצאות ממוצע כל הניסויים הראו 84.89% אחוזי התאמה בין קביעת ה-AI בהתבססות על נתונים שהגיעו מה-headband לבין קביעת המומחים, ו-85.83% התאמה בין קביעת ה-AI שהתבססה על נתונים מה-PSG לבין קביעת המומחים. עם זאת, מעבר על אחוזי ההתאמה עבור כל לילה באופן נפרד מראה כי למרות שברוב הלילות אחוזי ההתאמה היו גבוהים מ-80%, היו מספר לילות בהן אחוזי ההתאמה היו נמוכים יחסית והיו רק בסביבות ה-55%-70%.



כאשר התעמקנו בלילות אלו על מנת לראות מה מקור חוסר ההתאמה שהתגלה, גילינו שלא היה מדובר בחוסר קבלה של נתונים כפי שציפינו, אלא בקטלוג לא נכון של ה-AI על בסיס הנתונים שנאספו (לרוב היה מדובר בפרקי זמן מאוד קצרים אותם ה-AI קטלג כשלבי שינה גבוהים משלבי השינה האמיתיים שחווה הנבדק):  




בגלל שמרבית הלילות הראו אחוזי התאמה גבוהים יחסית (80%+) בין הקטלוג של ה-AI מה-headband לבין הקטלוג של המומחים ממעבדת השינה הסטנדרטית, נראה כי ניתן לקבל ממכשיר ה-EEG הביתי נתונים מספקים לקביעת שלבי השינה השונים של נבדקים, זאת בלי הצורך בחיבור הנבדק למכשירים נוספים המודדים ממדדים אחרים. הבנה זו לבד יכולה להקל מאוד על חיי המטופלים שצריכים לעבור בדיקות להפרעות שינה, ותאפשר להם לעבור בדיקה נוחה יותר בביתם מבלי להיות מחוברים למכשירים שיפריעו לשינתם אפילו יותר.

כמו כן, נראה שתוכנת ה-AI שפיתחו החוקרים כדי לזהות את שלבי השינה פועלת כראוי עם אחוזי הצלחה גבוהים ועם כמות זניחה יחסית של דאטא שלא נאסף כראוי. עם זאת, משום שעדיין היו מספר ימים בהם היו הרבה אי התאמות בין ניתוח הנתונים שהראה ה-AI לבין ניתוח הנתונים שהראו החוקרים, נראה שלא ניתן עדיין להחליף את ניתוח הנתונים שנאספים בבדיקות מעבדות שינה בתוכנות AI. עם זאת, משום שנראה שמרבית בעיות ההתאמה הסתכמו לשגיאה של מקטעי זמן ספציפיים של 30 שניות במהלכם ה-AI קטלג את שלב השינה באופן שאינו נכון, ייתכן שניתן לפתור את הבעיה על ידי הוספת הגדרה ל-AI שתגרום לו להשוות את הקטלוג שהוא נתן למקטע זמן ספציפי לתוצאות שניתנו במהלך 180 השניות לפני ואחרי המקטע. אם במהלך השוואה זו ה-AI רואה שהקטלוג שהוא נתן לא ממשיך רצף של לפחות 3 קטלוגים זהים במקטעים שלפני, או שהקטלוג שמגיע המקטע אחרי אינו זהה לקטלוג שהוענק למקטע, ה-AI יעניק למקטע זה קטלוג זהה לשלב השינה שניתן למקטע הקודם.