### **פרויקט גמר בינה מלאכותית**

קישורל[GITHUB](https://github.com/elad-segev/Artificial-intelligence-assisted-brain-MRI-analysis-for-early-diagnosis-of-Parkinson-s)-

דפני ביסטולס   
נירוונה ח'ורי

אלעד שגב   
רותם חג'בי

### **מבוא**

מחלת הפרקינסון, אחת מהמחלות הנוירולוגיות השכיחות בעולם, משפיעה על מיליוני אנשים מדי שנה. המחלה מתאפיינת בירידה הדרגתית בתפקוד המוטורי, הנגרמת כתוצאה מאובדן תאי ייצור דופמין בגזע המוח. האבחון המוקדם של המחלה, הקריטי לניהול אפקטיבי של הטיפול ולשיפור איכות החיים, נותר אתגר משמעותי בשל היעדר סמנים קליניים מובהקים בשלבים הראשונים.

בשנים האחרונות, חלה התקדמות משמעותית בשימוש בטכנולוגיות מבוססות בינה מלאכותית ולמידה עמוקה בתחום הרפואה. מודלים חישוביים אלו מאפשרים לנתח נתונים מורכבים ממגוון מקורות, כולל הדמיות MRI, כדי לזהות דפוסים עדינים המעידים על תחילת המחלה. מחקרים קודמים הדגישו את הפוטנציאל של טכניקות כמו PET, SPECT ו-MRI, לצד שיטות למידה עמוקה, לשיפור הדיוק והיעילות באבחון.

שאלת המחקר שלנו - **האם ניתן בעזרת נתוני סריקות מגנטיות של גזע המוח לפתח אלגוריתם לזיהוי מוקדם של פרקינסון?**

בסקירה זו, נציג את המחקרים המרכזיים שנעשו בתחום זה, עם דגש על תרומת ההדמיות המבניות (MRI) והיישומים של למידה עמוקה לניבוי ואבחון מוקדם של פרקינסון. בנוסף, נציג את שיטת המחקר המוצעת, הכוללת פיתוח אלגוריתם בינה מלאכותית המבוסס על שילוב ניתוח צילומי MRI, המשמשים כבסיס נתונים לאימון המודל.

### **סקירה ספרותית**

במהלך העשורים האחרונים, השימוש בטכנולוגיות של בינה מלאכותית ולמידה עמוקה בתחום הרפואה הפך לכלי מרכזי בזיהוי ואבחון מחלות נוירולוגיות מורכבות כמו פרקינסון. טכנולוגיות אלו מספקות דרכים חדשניות לניתוח נתונים מבוססי הדמיה, מה שמאפשר לזהות שינויים פיזיולוגיים עדינים בשלבים מוקדמים של המחלה.

מודלים חישוביים ומידול נתונים

אחת הגישות המרכזיות שנסקרו במאמרים היא השימוש באלגוריתמים שונים של למידת מכונה לזיהוי פרקינסון, תוך השוואת ביצועיהם. במחקר השוואתי שנעשה על אלגוריתמים שונים, נמצא כי למידת מכונה מבוססת Random Forest ו-Support Vector Machines) SVM) הציגה דיוק גבוה באבחון המחלה בהשוואה לשיטות סטטיסטיות מסורתיות **(A Comparative Study of Different Machine Learning Algorithms in Detection of Parkinson’s Disease [1])**. שיטות אלו התבססו על מאגרי נתונים קליניים שבהם מופו גורמים כמו תנועתיות, איכות דיבור ומדדים ביולוגיים נוספים.

למידה עמוקה ואבחון מבוסס תמונות

מאמר נוסף, דן בחשיבות של למידה עמוקה בזיהוי פרקינסון באמצעות עיבוד תמונות MRI.  
**(Deep Learning for Parkinson’s Disease Diagnosis: A Short Survey [2])**  
מודלים מסוג Convolutional Neural Networks) CNN) שימשו לאנליזה של תמונות מוחיות, תוך זיהוי אזורים שנפגעו בשל המחלה. מחקרים מצביעים על כך ששימוש בשיטות אלו מסוגל לזהות שינויים מבניים עדינים שאינם נראים בעין בלתי מזוינת.

שיטות הדמיה ומדדים פיזיולוגיים

בנוסף למודלים מבוססי למידת מכונה, מחקרים עדכניים עוסקים בשימוש בטכניקות הדמיה מתקדמות כגון PET ו-SPECT להערכת תפקוד הדופמינרגי של המוח. מחקר שעסק בהדמיה דופמינרגית באמצעות PET מצא כי ניתן לזהות ירידה משמעותית בתפקוד הדופמינרגי של גרעיני הבסיס בשלבי מחלה מוקדמים **(DOPAMINERGIC IMAGING IN PARKINSON’S DISEASE: PET [3])**. שימוש בטכניקות אלו מסייע להבנת הדינמיקה של המחלה ולהערכת השפעת טיפולים. נוסף על כך, שימוש בטכניקת SPECT **(DOPAMINERGIC IMAGING IN PARKINSON’S DISEASE: SPECT)** מאפשר מעקב מדויק אחר פעילות נוירוטרנסמיטורים ומספק תמונה פונקציונלית של מערכת העצבים.

מבנה מוחי והדמיה מגנטית (MRI)

מחקרים המבוססים על MRI הדגישו את חשיבותם של שינויים מבניים בזיהוי פרקינסון. נמצא כי ישנן חריגות מבניות הניתנות לזיהוי בשלבי מחלה מוקדמים באמצעות שיטות הדמיה שונות **(STRUCTURAL ABNORMALITIES IN PARKINSON’S DISEASE MRI AND RELATED METHODS [4])**. בנוסף, התקדמות בטכנולוגיות MRI שיפרה את יכולת הניתוח של תמונות מוחיות, וכיום ניתן לזהות ניוון אזורים מסוימים באמצעות שיטות כמו Diffusion Tensor Imaging (DTI ו-Voxel-Based Morphometry VBM (Advances in Magnetic Resonance Imaging).

שיטות מבוססות קליניקה ותנועה

מלבד שימוש בתמונות מוחיות, חלק מהמחקרים בחנו שיטות אבחון מבוססות מבחנים קליניים, כגון מבחן ההליכה לשש דקות **(Six-Minute Walk Test for Parkinson’s Disease Diagnosis [5])**. מבחנים אלו מספקים מדדים פיזיולוגיים הקשורים לשיווי משקל, מהירות הליכה ויכולת ניידות, שניתן לנתח באמצעות אלגוריתמים של למידת מכונה. תוצאות המחקרים הראו כי שילוב של נתונים ממקורות שונים (כגון ניתוח הליכה ונתונים קליניים) מגביר את דיוק המודלים.

הערכת ביצועי המודלים ודיוקם

מחקר נוסף **(Artificial Intelligence Assisted Classifier’s and Neural Network Based Prediction and Classification of Parkinson’s Disease [6])** הדגיש את חשיבותם של מדדים להערכת ביצועי המודלים, כגון Area Under the Curve (AUC), דיוק (Accuracy) ורגישות (Sensitivity). נמצא כי השילוב בין למידת מכונה מסורתית לבין למידה עמוקה מאפשר שיפור משמעותי בזיהוי מדויק של פרקינסון בשלביו המוקדמים.

### **ארכיטקטורה ומתודולוגיה**

בפרק הזה נסתכל על איך בנינו את המודלים שלנו, ונעבור על השלבים המרכזיים שעברנו בתהליך הפיתוח והערכה שלו. נפרט על כל מודל מה האופן שבו הרשת הנוירונית בנויה, איזה אלגוריתמים בחרנו, אילו נתונים השתמשנו בהם ואיזה שלבים של עיבוד מקדים היו צריכים להעשות בכדי להכין אותם.   
בנוסף, נסביר כיצד אימנו כל אחד מהמודלים , אופטימיזציה, שיטות להפחתת הטיות, בקרת ביצועים והערכת הדיוק שלו בעזרת מדדים סטנדרטיים.   
בסופו של דבר, ניתוח הארכיטקטורה הזו יעזור לנו להבין את היתרונות והאתגרים בגישה שלנו, וייתן לנו נקודת השוואה לשיטות אחרות בתחום.

ניקוי והכנת הדאטה

קיבלנו מאגר רחב של סריקות MRI תלת-ממדיות בהיקף גדול מתוך PPMI (Parkinson’s Progression Markers Initiative **[7]**), שעליו קיבלנו אישור לביצוע מחקר וחיסיון רפואי מלא. בנוסף אליו הורדנו מאגר נוסף של צילומי MRI מה- Github **[8]**. בשלב הראשון, לאחר התייעצות עם נוירולוג, זיהינו את החתכים הרלוונטיים המכילים את גזע המוח ואת סובסטרט הניגרה.   
 הדאטה מהמקור הראשון הגיע בפורמט סריקה תלת מימדי, מה שאילץ אותנו לעבוד עם תוכנה מתאימה על מנת לקחת את החתכים הרלוונטים לנו- חתך אקסאלי קצת מעל גזע המוח, לכל מטופל נלקחו בממוצע 10 פריימים, 5 מעל הנקודה שבחרנו ו5 מתחת, יש לציין שכל מטופל סווג באופן ידני על מנת ליצור דאטה רלוונטי ואחיד למודל.   
המקור השני הגיע בפורמט סריקה דו מימדי אך מאוד מבולגן, לכל מטופל יכלו להיות מס סבבי צילום/ צילומים שלאו דווקא רלוונטים למטרה (צילומי גב וכדומה) לכן, גם כאן בוצע מיון וניפוי ידני של כל מטופל ומטופל.

בתוצר הסופי, יצרנו דאטה סט מאוחד משני המקורות שלנו שמכיל מאגר גדול של סריקות מהסוגים השונים. האחידות נוצרה הודות ל-3 סקריפטים שנכתבו ועליהם נרחיב בהמשך, בעזרתם יצרנו חלוקה ברורה לשני המודלים, כך ששני מודלים שמשתמשים באותו הדאטה לא יקבלו את אותם הצילומים.

סיווג הדאטה

מכיוון שמקורות הדאטה שלנו היו מגוונים, היה שוני רב ואי סדר בתמונות עצמן: כפילויות בדאטה (מס בדיקות לאותו המטופל), תמונות בגדלים שונים וכו' היה צורך ליצור אחידות וסדר על מנת לקבל ביצועים טובים יותר במודל, לשם כך בנינו 3 סקריפטים שונים:

1. mri\_zip\_type\_arrangement.py
2. mri\_sparate\_colactions.py
3. project\_image\_conversion.py

סקריפטים אלו פועלים בתיאום בכך שהסקריפט הראשון מתחיל בסיווג ראשוני של הנתונים לפי קבוצות ע"פ נתוני הcsv . לאחר מכן מתבצעת העברת תקיות למבנה מאוחד תוך הפחתת כפילויות (עובר מסקריפט מס 1 לסקריפט מס 2), ולבסוף מתבצע עיבוד טכני של התמונות הכולל המרה והתאמת רזולציה אחידה ע"י סקריפט מס 3.

* **הסקריפט הראשון – "mri\_zip\_type\_arrangement.py"**

סקריפט זה אחראי על תהליך הסיווג וההעתקה הראשוני. הוא קורא קובץ CSV באמצעות ספריית pandas ומבצע סריקה של תיקיות המטופלים, כאשר כל תיקייה מייצגת מטופל ספציפי.  
 הסקריפט בודק את קיום מזהה המטופל בעמודת "Subject ID" שבקובץ ה-CSV. מידע מעמודת "Research Group" משמש לקביעת השיוך הקליני של המטופל (PD, Prodromal או Healthy). בהתאם לשיוך, התיקייה מועתקת ליעד המתאים (PD1, Prodromal1 או Control1) באמצעות הפונקציה copy\_directory.

* **הסקריפט השני – "mri\_sparate\_colactions.py"**

סקריפט זה אחראי על חלוקה מחדש והעברת תיקיות. הוא מקבל את התיקיות המקוריות (Control1, PD1, Prodromal1) ומעביר אותן לתיקיות מרכזיות (Control, PD, Prodromal) תוך שימוש בפונקציית copy\_directory. לפני ההעברה, הסקריפט יוצר רשימת תיקיות משנה ומבצע העברה מדורגת תוך צמצום כפילויות. הסקריפט עושה שימוש במילונים (d ו-d1) למיפוי מזהי מטופלים ומעקב אחר מספר הבדיקות.

* **הסקריפט השלישי - project\_image\_conversion.py""**

סקריפט זה עובד על עיבוד התמונות. הוא מקבל תמונות מהתיקיות המסווגות, בודק את תקינותן לפי סיומות הקבצים (.png, .jpg וכדומה), ומבצע התאמת רזולוציה ל-250X250 פיקסלים בעזרת ספריית PIL. כל תמונה שעוברת עיבוד נשמרת בתיקייה ייעודית, תוך שמירה על שמה המקורי או קבלת שם ייחודי המבוסס על מספר המטופל.

תיאור המודלים

**מודל ראשון - מודל CNN**

המודל מבוסס על **רשת נוירונים קונבולוציונית (CNN)**, שנבנתה במיוחד לניתוח ועיבוד של תמונות.   
הרשת אומנה על מאגר תמונות הכולל סריקות של חולים ובריאים, תוך שימוש בטכניקות שונות לשיפור ביצועיה, כגון איזון נתונים, עיבוד מקדים, ואוגמנטציה. המודל נבנה כך שיוכל להבדיל בין שני המצבים, באמצעות למידה של דפוסים ייחודיים במבנה המוח והאזור הספציפי הנפגע מהמחלה.

* + טעינת הנתונים:

תהליך העבודה מתחיל בטעינת הנתונים, המורכבים מתמונות MRI. התמונות הללו ממוינות מראש לשתי קטגוריות מרכזיות: **חולים בפרקינסון (PD)** ו**לא-חולים (NONPD)**. נוסף על כך, נעשה שימוש בקובץ אשר מכיל רשימה של קובצי התמונה של קבוצת הלא-חולים.

לצורך קריאת התמונות, נעשה שימוש בספריית OpenCV, המאפשרת טעינה מהירה של הקבצים. התמונות שמתקבלות אינן זהות בגודלן, ולכן הן עוברות עיבוד מקדים, הכולל שינוי גודל לרזולוציה אחידה של 224x224 פיקסלים, וזאת כדי להתאים אותן לקלט של המודל המתחרה. בנוסף, מכיוון שהתגיות נשמרות בפורמט טקסטואלי (PD / NONPD), הן עוברות המרה לערכים מספריים: 1 עבור חולי פרקינסון, ו-0 עבור בריאים.

* + הכנת הנתונים:

לפני השימוש במודל, יש לוודא איזון בין קבוצות הנתונים, כך שלא תהיה העדפה של מודל הלמידה לכיוון קבוצה מסוימת. מסיבה זו, בוצעה דגימה אחידה, כך שלשתי הקבוצות יהיה מספר תמונות שווה, מה שמונע מצב שבו המודל ייטה לנבא קבוצה אחת באופן מובהק על חשבון השנייה.

לאחר יצירת הדאטהסט המאוזן, התמונות והתגיות נשמרות במערכים ייעודיים, אשר משמשים להכשרת המודל בשלבי האימון. בשלב זה, הנתונים מאורגנים כך שהתוויות נשמרות בפורמט של one-hot encoding, כדי להתאים למבנה רשת הנוירונית בהמשך.

* + שיפור הנתונים:

בכדי לשפר את יכולתו של המודל להתמודד עם שונות גדולה בתמונות ולמנוע Overfitting, נעשה שימוש בטכניקת אוגמנטציה, המרחיבה את סט הנתונים על ידי ביצוע שינויים קטנים בתמונות הקיימות. שינויים אלו נבחרו בקפידה כדי לשמור על המידע הקריטי, אך גם להוסיף וריאציות שמסייעות למודל הכללי.

שלב זה כלל שלוש מניפולציות עיקריות:

1. סיבוב רנדומלי קטן: כל תמונה סובבה עד 0.5 מעלות כדי לדמות וריאציות קלות בצילומי ה-MRI.
2. זום רנדומלי: עד 10% כדי לדמות שינויים אפשריים בגודל התמונה או בהתמקדות הצילום.
3. שינוי ניגודיות: עד ±5% כדי להתמודד עם שינויים בתנאי התאורה או באיכות הצילום.

חשוב לציין כי מניפולציות אלו יושמו **רק על סט האימון**, ולא על סט האימות או המבחן, כדי להימנע מהכנסת וריאציות שעלולות להשפיע על דיוק ההערכה.

* + ארכיטקטורת המודל:

המודל שנבנה מבוסס על רשת קונבולוציה CNN תוך שימוש בספריות TensorFlow ו-Keras, והוא בנוי בשיטה הרלוונטית לניתוח תמונות רפואיות. הרשת פותחה באמצעות Sequential API, תוך שילוב שכבות קונבולוציה, נורמליזציה והפחתת מימדים, כדי להבטיח דיוק גבוה ויכולת הכללה טובה.

בתחילת הרשת מופיעות שתי שכבות Conv2D, כאשר הראשונה כוללת 32 פילטרים בגודל 3x3 והשנייה 64 פילטרים בגודל 3x3, שתיהן משתמשות בפונקציית ההפעלה ReLU. כל שכבת קונבולוציה מלווה בשכבת Batch Normalization, שתפקידה לייצב את תהליך האימון ולהאיץ את קצב הלמידה. לאחר כל שכבת קונבולוציה מופיעה שכבת MaxPooling2D בגודל 2x2, שמטרתה להקטין את ממדי התמונה תוך שמירה על התכונות החיוניות. כדי למנוע Overfitting לפני השלב הסופי מתבצע Dropout של 50%.

בחלק הסופי של הרשת, לאחר שהמידע עובר דרך שכבות הקונבולוציה, הנתונים מיושרים לשכבת Flatten, ההופכת את הנתונים לוקטור חד-ממדי, ומשם הם מועברים לשכבת Dense עם 128 נוירונים, תוך שימוש בפונקציית ההפעלה ReLU. לבסוף, שכבת הפלט כוללת 2 נוירונים, המייצגים את הקטגוריות ("חולה" ו"לא חולה"), עם פונקציית ההפעלה Softmax, שמתאימה לסיווג רב-קטגורי.

* + מתודולוגיית האימון:

לאחר בניית המודל, האימון שלו מתבצע תוך שימוש בפונקציית הפסד Categorical Crossentropy, המתאימה לבעיות סיווג רב-קטגורי. כדי לשפר את ביצועי המודל, נעשה שימוש באלגוריתם האופטימיזציה Adam, עם קצב למידה שנבחר בקפידה: 0.0001. קצב זה מבטיח יציבות במהלך האימון, תוך צמצום תנודות חדות במשקלות הרשת.

כחלק מההכנה לאימון, התמונות עוברות נרמול, כך שכל ערכי הפיקסלים שלהן נמצאים בטווח [0,1], על ידי חלוקה ב-255. הנרמול משפר את יציבות המודל ומונע התמקדות במאפיינים לא מהותיים. במהלך האימון, התמונות מוזנות למודל בקבוצות של 32 תמונות בכל פעם, כאשר התהליך נמשך למשך 50 אפוקים.

**מודל שני - מודל CNN TL תסמינים**

המודל הנוכחי מבוסס על למידה מעבירה (Transfer Learning), שבה נעשה שימוש במודל CNN קיים שאומן לזיהוי חולי פרקינסון, במטרה לשפר את הביצועים ולזהות תסמינים מוקדמים של המחלה.

במסגרת ההחלטה שהתקבלה, תמונות של אנשים עם תסמינים מסווגות כחולים, ואילו תמונות של אנשים ללא תסמינים מסווגות כבריאים. משמעות הדבר היא שהמודל אינו מזהה חולי פרקינסון קליניים בלבד, אלא מבדיל בין אנשים בריאים לבין אנשים המציגים סימנים מקדימים למחלה. בכך, המודל מהווה כלי חשוב לאבחון מוקדם ולהערכת סיכון לפרקינסון, עוד לפני שהמחלה מאובחנת קלינית.

טעינת הנתונים:

כמו במודל הראשון, הנתונים מורכבים מתמונות MRI, אך הפעם הן מסווגות לשתי קטגוריות חדשות:

* "Prodromal" – אנשים עם תסמינים מקדימים של פרקינסון (נחשבים כחולים).
* "NonPD" – אנשים בריאים ללא תסמינים כלל.

לצורך קריאת התמונות, נעשה שימוש בספריית OpenCV, המאפשרת טעינה מהירה של הקבצים. התמונות עוברות עיבוד מקדים, הכולל שינוי גודל לרזולוציה אחידה של 224x224 פיקסלים, כדי להתאים אותן לפורמט הקלט של המודל. כמו כן, התיוגים ממופים לערכים מספריים, כאשר 1 מוקצה לחולים עם תסמינים (Prodromal), ו-0 מוקצה לבריאים (NonPD).

* + הכנת הנתונים:

כדי להימנע מהטיית המודל לכיוון קבוצה מסוימת, בוצעה דגימה אחידה, כך שלכל קטגוריה יש מספר שווה של דוגמאות. תהליך זה מונע מצב שבו המודל יעדיף את אחת הקבוצות בגלל חוסר איזון בנתונים.

לאחר יצירת הדאטהסט המאוזן, התמונות והתוויות נשמרות במערכים ייעודיים, כאשר התיוגים נשמרים בפורמט one-hot encoding, כדי להתאים את הנתונים לפורמט הנדרש על ידי הרשת הנוירונית.

* + שיפור הנתונים:

בכדי לשפר את הביצועים ולמנוע Overfitting, נעשה שימוש בטכניקת אוגמנטציה, שמרחיבה את סט הנתונים על ידי ביצוע מניפולציות שונות על התמונות:

1. סיבוב קל (עד ±0.5°)
2. זום רנדומלי (עד ±10%)
3. שינוי ניגודיות (עד ±5%)

חשוב לציין כי שינויים אלו בוצעו אך ורק על **סט האימון**, ולא על סט האימות או המבחן, כדי להימנע מהכנסת וריאציות שעלולות להטות את תוצאות ההערכה.

* + ארכיטקטורת המודל:

המודל שנבנה מבוסס על Transfer Learning, גישה שבה נעשה שימוש במודל קיים כבסיס ללמידה חדשה, במקום לאמן רשת נוירונים מאפס. המודל עושה שימוש ברשת CNN שאומנה מראש על מאגר תמונות לזיהוי חולי פרקינסון, ומשמשת כבסיס ללמידת הדפוסים הרלוונטיים. בשלב הראשוני, נטען המודל הקיים תוך קפיאת כל השכבות, כך שלא יתעדכנו במהלך האימון הראשוני. פעולה זו מאפשרת לנצל את הידע שנרכש במודל המקורי ולהתמקד בהתאמה לנתונים החדשים.

לאחר טעינת המודל הבסיסי, נוספו שכבות ייעודיות להתאמתו למשימת הסיווג הנוכחית. תחילה, שכבת GlobalAveragePooling2D מבצעת הפחתת ממדי הפלט של הרשת המקורית, תוך שימור המידע החיוני לצורך הלמידה. לאחר מכן, שכבת Dense עם 256 נוירונים ופונקציית הפעלה ReLU משמשת להעמקת תהליך הלמידה ולזיהוי תבניות מורכבות יותר. כדי להפחית Overfitting, משולבת שכבת Dropout של 50%, אשר מונעת תלות מוגזמת בניורונים מסוימים ומשפרת את יכולת ההכללה של המודל.

בחלק הסופי של הרשת, לאחר שהמידע עבר דרך השכבות החדשות, הוא מיושר באמצעות שכבת Flatten ומוזן לשכבת Dense נוספת, המורכבת מ-2 נוירונים עם פונקציית ההפעלה Softmax. שכבת הפלט מסווגת את הקלט לאחת משתי הקטגוריות: בריאים או חולים עם תסמינים.

* + מתודולוגיית האימון:

תהליך האימון של המודל מתבצע בשני שלבים מרכזיים:

1. קפיאת המודל הקודם:
   * תחילה מאומנות רק השכבות החדשות שהוספנו, כאשר כל שכבות המודל המקורי נותרים קפואים.
2. Fine-Tuning (כוונון עדין):
   * לאחר מספר אפוקים, חלק מהשכבות של המודל המקורי נפתחות וממשיכות להתעדכן, כדי לאפשר למודל ללמוד התאמות ספציפיות לסט הנתונים החדש.

המודל אומן תוך שימוש בפונקציית הפסד Categorical Crossentropy, עם אופטימיזציה של Adam בקצב למידה נמוך במיוחד 0.00005, כדי למנוע שינויים חדים במשקלות הרשת. כמו כן, הנתונים עוברים נרמול לטווח [0,1] על ידי חלוקה ב-255.

כדי למנוע בזבוז זמן וחישובים מיותרים, הוגדר מנגנון Early Stopping, שעוצר את האימון אם לא נרשם שיפור במשך 5 אפוקים רצופים. במקביל, נעשה שימוש במנגנון ModelCheckpoint, אשר שומר את המודל עם הביצועים הטובים ביותר על סט האימות, כך שבסוף האימון ניתן להשתמש בגרסה האופטימלית.

**מודל שלישי - מודל ResNet50**

באמצעות TL (ResNet50) וצעדים כמו אוגמנטציה, Dropout, ו-Fine-Tuning, הצלחנו לבנות מודל שמסווג תמונות MRI ל-PD או nonPD. המודל מתבסס על עיבוד מקבילי של טנזורים תלת-ממדיים (224×224×3), משתמש ב-Adam כדי להתמודד עם חוסר איזון ונתוני תמונה מורכבים, ועובר בדיקות ולידציה שוטפות למניעת Overfitting. כך מתקבל כלי שיכול, בתקווה, לסייע בצורה אמינה יותר באבחון מבוסס תמונה.

המודל שפותח מבצע סיווג של תמונות MRI מוחיות במטרה לזהות חולי פרקינסון (PD) לעומת אנשים בריאים (nonPD). המודל מבוסס על למידת העברה (Transfer Learning), תוך שימוש ברשת **ResNet50** שאומנה מראש על מאגר התמונות ImageNet. השימוש במודל מאומן מראש מאפשר לנצל דפוסים חזותיים שנלמדו על מגוון רחב של תמונות ולהתאים אותם לניתוח תמונות רפואיות, תוך חיסכון משמעותי בזמן חישובי.

* + טעינת הנתונים:

תהליך העבודה מתחיל בטעינת הנתונים, המורכבים מתמונות MRI מוחיות. התמונות חולקו מראש לשתי קטגוריות עיקריות: חולים בפרקינסון (PD) ואנשים בריאים (nonPD). הנתונים נשמרו במבנה מסודר, הכולל רשימה של כל קובצי התמונה לכל קטגוריה, כך שניתן יהיה לטעון אותם בצורה יעילה.

לצורך קריאת התמונות, נעשה שימוש בספריית OpenCV, המאפשרת גישה מהירה לקובצי התמונה וטעינתם לתוך סביבת העבודה. התמונות במקור אינן זהות בגודלן או בפורמט שלהן, ולכן הן עוברות תהליך עיבוד מוקדם שמטרתו לאחד את הפורמט ולהתאים אותן לדרישות המודל. אחד השלבים המרכזיים בתהליך זה הוא שינוי גודל (Resize), שבו כל התמונות מוקטנות לרזולוציה של 224×224 פיקסלים, כדי להתאים לפורמט הקלט של רשת הנוירונים ResNet50. בנוסף, התמונות עוברות המרה לפורמט RGB, מכיוון שהמודל דורש שלושה ערוצי צבע ואילו התמונות המקוריות הן בגווני אפור בלבד. כדי להתאים את הקובץ לפורמט הדרוש, הערוץ הקיים משוכפל לשלושה ערוצים, כך שהתמונה תישאר זהה ויזואלית, אך בפורמט התואם למודל.

* + הכנת הנתונים:

לאחר טעינת הנתונים, יש לוודא כי חלוקת הדאטה נעשית בצורה מאוזנת כדי למנוע הטיה של המודל לטובת אחת הקבוצות. מסיבה זו, הנתונים חולקו לשלושה סטים: סט האימון (Train), הכולל 64% מהתמונות ומשמש ללימוד המודל; סט הולידציה (Validation), המהווה 16% מהתמונות ונועד לבדיקת ביצועי המודל בזמן האימון מבלי לשנות את משקליו; וסט המבחן (Test), המורכב מ-20% מהתמונות ומשמש להערכת ביצועי המודל לאחר תום תהליך האימון.

כדי להבטיח שהמודל לא ייטה לנבא את אחת הקבוצות באופן מובהק יותר על חשבון השנייה, בוצעה דגימה אחידה כך שכל אחת מהקטגוריות תיוצג באופן שווה. התמונות והתוויות שלהן נשמרו במערכים ייעודיים, שהוכנו כך שיתאימו להזנת הנתונים למודל. בשלב זה, התוויות הומרו לפורמט מספרי באמצעות one-hot encoding, על מנת לוודא שהמודל יוכל לעבד את הנתונים בצורה אופטימלית ולהתאים אותם למבנה רשת הנוירונים שבה נעשה שימוש.

* + שיפור הנתונים:

כדי להגדיל את היכולת של המודל להתמודד עם שונות גדולה בתמונות ולמנוע Overfitting, נעשה שימוש בטכניקת אוגמנטציה. טכניקה זו מרחיבה באופן מלאכותי את מגוון הנתונים הקיימים באמצעות יצירת וריאציות חדשות של התמונות, תוך שמירה על המידע הקריטי לזיהוי המחלה. מניפולציות אלו מאפשרות למודל ללמוד טוב יותר כיצד לזהות חולי פרקינסון, גם כאשר התמונות אינן זהות לחלוטין למערכת הנתונים שעליהן אומן.

שלב זה כלל שלוש מניפולציות עיקריות:

1. סיבוב רנדומלי קטן: כל תמונה סובבה עד 0.5 מעלות כדי לדמות וריאציות קלות בצילומי ה-MRI.
2. זום רנדומלי: עד 10% כדי לדמות שינויים אפשריים בגודל התמונה או בהתמקדות הצילום.
3. שינוי ניגודיות: עד ±5% כדי להתמודד עם שינויים בתנאי התאורה או באיכות הצילום.

חשוב להדגיש כי שלבי האוגמנטציה יושמו **אך ורק על סט האימון**, כדי למנוע הכנסת וריאציות שעלולות להשפיע על דיוק ההערכה בסט הולידציה ובסט המבחן.

* + ארכיטקטורת המודל:

המודל מבוסס על רשת נוירונים קונבולוציונית (CNN) בשם **ResNet50**, שנבחרה עקב העבודה שרשת זו אומנה מראש על מאגר תמונות רחב ולמדה לזהות דפוסים חזותיים מורכבים, מה שמאפשר לה לבצע משימות סיווג מדויקות יותר על סמך ידע קיים. שימוש ברשת מוכנה חוסך זמן חישובי משמעותי ומשפר את ביצועי המודל, שכן היא כבר עברה אימון על מיליוני תמונות וזיהתה תבניות חזותיות שונות.

כדי להתאים את ResNet50 למשימת סיווג חולי פרקינסון, בוצעו מספר התאמות. השכבה העליונה של הרשת המקורית הוסרה, מאחר שהיא נועדה למשימת סיווג כללית של ImageNet ואינה רלוונטית למשימה הנוכחית. במקומה, נוספה שכבת Dense עם פונקציית הפעלה Sigmoid, המשמשת לסיווג בינארי. שכבה זו מפיקה ערך הסתברותי בין 0 ל-1, כאשר ערכים גבוהים מ-0.5 מסווגים כ-PD, ואילו ערכים נמוכים מ-0.5 מסווגים כ-nonPD.

בנוסף, שולבו מספר מנגנונים לשיפור ביצועי המודל. GlobalAveragePooling2D נוספה כדי לסכם את המאפיינים החזותיים המרכזיים של התמונה ולצמצם את ממדיה. כמו כן, הוכנסה שכבת Dropout (0.2), שנועדה להפחית Overfitting על ידי השמטת נוירונים באופן אקראי במהלך האימון. בשלב הראשוני של האימון, שכבות הבסיס של ResNet50 הוקפאו, כך שהמשקלים שלהן לא השתנו. פעולה זו נועדה לאפשר למודל לנצל את הידע שכבר נרכש מהרשת המאומנת מראש ולהתמקד בלמידת ההתאמות הייחודיות לנתוני ה-MRI.

* + מתודולוגיית האימון:

תהליך האימון התבצע תוך שימוש בפונקציית הפסד binary\_crossentropy, שמתאימה למשימות סיווג בינאריות שבהן יש להבחין בין שתי קטגוריות בלבד. המודל אומן באמצעות אלגוריתם האופטימיזציה **Adam**, שמאפשר התאמה דינמית של קצב הלמידה לאורך האימון כדי להבטיח התכנסות יציבה ומהירה יותר.

כדי למנוע העדפה של אחת הקטגוריות בשל חוסר איזון בנתונים, נעשה שימוש במשקולות מותאמות, כך שהמודל לא יטה לנבא קבוצה אחת יותר מהשנייה. נוסף על כך, לאחר 10 אפוקים ראשונים, חלק מהשכבות העליונות של ResNet50 שוחררו להמשך האימון (Fine-Tuning), ונעשה שימוש בקצב למידה נמוך יותר, כדי לאפשר למודל להתאים את עצמו טוב יותר לנתוני ה-MRI תוך שמירה על הדפוסים הלמידה הכלליים של הרשת המקורית.

**מודל רביעי - מודל Adjusted TL תסמינים**

באמצעות למידת העברה (Transfer Learning) והסתמכות על מודל שאומן מראש, פותח מודל חדש המסוגל לסווג תמונות MRI לשתי קטגוריות: תסמינים Prodromal ובריאים NONPD. המודל עושה שימוש ברשת **ResNet50**, אשר עברה התאמות ייעודיות, כולל אוגמנטציה, Dropout ו-Fine-Tuning, כדי למקסם את ביצועיו ולשפר את דיוק הסיווג.

המודל מבוסס על (מודל שלישי - מודל ResNet50), שאומן במקור על סיווג חולי פרקינסון לעומת נבדקים בריאים, אך עבר התאמות כדי להתמקד בשלבים המוקדמים של המחלה. השכבות הקונבולוציוניות של המודל המאומן נשמרו, בעוד שהשכבות העליונות הוחלפו, כדי לאפשר התאמה למשימת הסיווג החדשה. כמו כן, הוגדרו משקלים מותאמים לקטגוריות כדי למנוע הטיה, ויושמה בקרת ולידציה מתמדת לצורך מניעת Overfitting.

המודל מבצע את עיבוד הנתונים באמצעות **טנזורים תלת-ממדיים** בגודל (224×224×3), כאשר כל תמונת MRI מיוצגת כמטריצה תלת-ממדית הכוללת מידע על גובה, רוחב ושלושת ערוצי הצבע (RGB). עיבוד הטנזורים מתבצע ברציפות לאורך כל שלבי המודל – מהשלבים הקונבולוציוניים הראשוניים, דרך שכבות ה-GlobalAveragePooling2D, ועד לשלב הסיווג הסופי, שבו מתקבלת הסתברות בינארית לכל טנזור קלט.

גם על המודל הזה התקבלה ההחלטה כי תמונות של אנשים עם תסמינים מסווגות כחולים, ואילו תמונות של אנשים ללא תסמינים מסווגות כבריאים. משמעות הדבר היא שהמודל אינו מזהה חולי פרקינסון קליניים בלבד, אלא מבדיל בין אנשים בריאים לבין כאלה המציגים סימנים מקדימים למחלה, מה שהופך אותו לכלי פוטנציאלי לאבחון מוקדם ולתמיכה בהחלטות רפואיות מבוססות נתונים.

* + טעינת הנתונים:

תהליך העבודה מתחיל בטעינת נתוני MRI, אשר מסווגים מראש לשתי קטגוריות עיקריות:

* Prodromal – חולים בשלב הטרום-קליני של מחלת פרקינסון.
* NONPD – נבדקים בריאים.

הנתונים הוכנו בפורמט מתאים לרשת הנוירונים, תוך חלוקה מבוקרת לסטים נפרדים. כל קובצי התמונה נשמרו בתיקיות ייעודיות, כאשר קיים קובץ נוסף המרכז את רשימת הקבצים המשויכים לכל קטגוריה, מה שמאפשר גישה נוחה וטעינה מהירה של הנתונים.

לצורך טעינת הנתונים נעשה שימוש בספריית OpenCV, המאפשרת טעינה מהירה ויעילה של תמונות ושמירתן במערכים מותאמים. מכיוון שהתמונות אינן אחידות בגודלן, הן עוברות תהליך עיבוד מקדים, הכולל שינוי גודל (Resize) לרזולוציה אחידה של 224×224 פיקסלים, על מנת להתאים אותן למודל.

* + הכנת הנתונים:

לאחר טעינת התמונות, בוצעה חלוקת הנתונים לשלושה סטים מרכזיים:

* אימון - 64% מהתמונות, משמש ללימוד המודל.
* ולידציה - 16% מהתמונות, משמש להערכת ביצועי המודל בזמן האימון, מבלי לשנות את משקליו.
* מבחן - 20% מהתמונות, משמש להערכת הביצועים לאחר סיום האימון.

כדי למנוע הטיה של המודל לטובת אחת הקבוצות, בוצעה דגימה אחידה, כך שלשתי הקבוצות יהיה מספר דוגמאות מאוזן בכל סט. התמונות והתוויות שלהן נשמרו במערכים ייעודיים שהוכנו להזנה למודל, תוך המרת התוויות לפורמט one-hot encoding, המתאים לרשת הנוירונים. תהליך זה מבטיח שהמודל יקבל נתונים בפורמט עקבי ונוח ללמידה.

* + שיפור הנתונים:

כדי לשפר את יכולת ההכללה של המודל ולמנוע Overfitting, הפועלה טכניקת אוגמנטציה, המרחיבה את מגוון הדוגמאות הקיימות באמצעות יצירת וריאציות מלאכותיות של התמונות המקוריות. שיטה זו מאפשרת למודל להתמודד טוב יותר עם שינויים טבעיים בנתוני MRI.

הפעולות שנכללו באוגמנטציה:

* סיבוב: כל תמונה סובבה עד 0.5 מעלות, כדי לדמות שינויים בזווית הצילום של ה-MRI.
* זום: שינוי קנה מידה של התמונה בעד 10%, כדי להתמודד עם שינויים אפשריים בגודל הסריקה.
* היפוך אופקי: שינוי כיוון התמונה כדי לאפשר זיהוי דפוסים ללא תלות בכיוון הסריקה.
* שינוי ניגודיות: התאמת רמות התאורה בעד ±5%, כדי להתמודד עם שינויים באיכות התמונה.

חשוב לציין כי שלבי האוגמנטציה בוצעו רק על **סט האימון**, על מנת למנוע הכנסת וריאציות שעלולות להשפיע על תוצאות הוולידציה והמבחן.

* + ארכיטקטורת המודל:

המודל מבוסס על למידת העברה (Transfer Learning), תוך שימוש ברשת ResNet50, אחת מהרשתות הקונבולוציוניות (CNN) החזקותבתחום הסיווג החזותי. רשת זו אומנה מראש על מאגר ImageNet, הכולל מיליוני תמונות מסווגות לקטגוריות שונות, מה שמאפשר לה לזהות דפוסים ופרטים מורכבים בתמונות באופן כללי. היתרון של TL הוא בכך שניתן לנצל את הידע הקיים של הרשת ולבצע התאמות נקודתיות למשימה הספציפית של סיווג תמונות MRI לזיהוי חולים עם תסמינים לעומת בריאים, במקום לאמן רשת חדשה מאפס.

כדי להתאים את ResNet50 (מודל שלישי - מודל ResNet50) למשימת סיווג חולי ה-Prodromal, בוצעו מספר שינויים מהותיים במבנה הרשת:

* הסרת שכבות הסיווג העליונות של מודל שלישי - מודל ResNet50 – מאחר שהמודל אומן מראש לסיווג אובייקטים כלליים ב-ImageNet, השכבות האחראיות על סיווג קטגוריות לא היו רלוונטיות לזיהוי תבניות מוחיות ב-MRI, ולכן הן הוסרו.
* שמירה על שכבות הקונבולוציה העמוקות – השכבות הראשונות של ResNet50 נותרו ללא שינוי, מאחר שהן כבר למדו לזהות מאפיינים בסיסיים של תמונה, כמו קווים, מרקמים ודפוסים גיאומטריים. ידע זה מועיל גם בניתוח תמונות MRI, שבהן תבניות עדינות יכולות להעיד על שינויים מוחיים הקשורים לפרקינסון.
* הוספת שכבות Fully Connected חדשות – לאחר שכבות הקונבולוציה, נוספו שכבות רשת נוירונים מלאכותית (Dense), אשר מעבדות את תכונות התמונה ולומדות כיצד להבחין בין Prodromal ל-NONPD.
* שימוש בשכבת GlobalAveragePooling2D – שכבה זו משמשת לצמצום מספר הפרמטרים של המודל, תוך שמירה על התכונות החשובות ביותר שנלמדו בשלבים המוקדמים של הרשת. היא מאפשרת הפחתת מימדים מבלי לאבד מידע קריטי.
* Dropout (0.5) למניעת Overfitting – כדי להפחית את הסיכון של המודל ללמוד את הנתונים הקיימים בצורה מדויקת מדי, מה שעלול להוביל לחוסר הכללה על נתונים חדשים, שולבה שכבת Dropout. שכבה זו משביתה חלק מהנוירונים באקראי במהלך האימון, מה שמכריח את הרשת ללמוד תבניות חזקות יותר במקום להסתמך על פרטים שוליים.
* שכבת פלט מותאמת עם הפעלת Sigmoid – כדי לבצע סיווג בינארי, שכבת הפלט הוגדרה כך שתוציא ערך הסתברותי בין 0 ל-1. אם הערך גבוה מ-0.5, התמונה מסווגת כ-Prodromal, ואם הוא נמוך מ-0.5, היא מסווגת כ-NONPD.
  + מתודולוגיית האימון:

### תהליך האימון של המודל מתבצע בשני שלבים מרכזיים. בתחילה, כל השכבות הקונבולוציוניות של ResNet50 נותרו קפואות, כך שהמשקלים שלהן לא השתנו, ואומנו רק השכבות החדשות שהוספנו. שכבות אלו כללו את GlobalAveragePooling2D, שכבות Fully Connected, ושכבת פלט עם הפעלת Sigmoid. בשלב זה המודל למד כיצד למפות את המאפיינים שנלמדו מראש ב-ResNet50 לקטגוריות הסיווג החדשות, מבלי לשנות את הדפוסים החזותיים שהרשת רכשה ממודל הבסיס.

### לאחר מספר אפוקים, חלק מהשכבות של המודל המקורי נפתחו והחלו להתעדכן, כדי לאפשר למודל ללמוד התאמות ספציפיות לנתוני ה-MRI החדשים. 20 השכבות העליונות של ResNet50 שוחררו לצורך עדכון המשקלים, תוך שמירה על קצב למידה נמוך כדי למנוע שינויים חדים ולשמר את הידע הקיים של המודל. שלב זה אפשר למודל להתאים את עצמו בצורה טובה יותר לזיהוי Prodromal, תוך למידה של דפוסים ייחודיים הקשורים לשלב הטרום-קליני של פרקינסון.

### המודל אומן תוך שימוש בפונקציית הפסד binary\_crossentropy, עם אופטימיזציה של Adam בקצב למידה נמוך במיוחד 0.00005, כדי להבטיח יציבות בלמידה ולמנוע תנודות חדות במשקלות הרשת. כמו כן, הנתונים עברו נרמול לטווח [0,1] על ידי חלוקה ב-255, כדי לשפר את ההתכנסות ולמנוע סטיות גדולות במהלך העדכונים.

### כדי למנוע בזבוז זמן וחישובים מיותרים, הוגדר מנגנון Early Stopping, אשר עוצר את האימון אם לא נרשם שיפור במשך 5 אפוקים רצופים. במקביל, נעשה שימוש במנגנון ModelCheckpoint, אשר שומר את המודל עם הביצועים הטובים ביותר על סט האימות, כך שבסוף האימון ניתן להשתמש בגרסה האופטימלית לצורך סיווג התמונות בצורה המדויקת ביותר.

### **שיטות הערכה ומהלך הניסוי**

כדי להבטיח את יעילות המודל ואת מידת הדיוק שלו בזיהוי מחלת פרקינסון, השתמשנו במגוון שיטות הערכה מתקדמות. פרק זה מתאר את הקריטריונים והמדדים ששימשו אותנו להערכת ביצועי המודל. כמו כן, נפרט את גישת החלוקה בין קבוצות האימון, האימות והבדיקה (Train, Validation, Test), את השיטות שנקטנו כדי למנוע בעיות של אימון יתר (Overfitting) ואת השוואת ביצועי המודל מול גישות קיימות. שיטות אלו סייעו לנו להבטיח שהמודל לא רק מדויק מבחינה חישובית, אלא גם פרקטי ויעיל לשימוש קליני.

**מודל ראשון - מודל CNN**

* + פיצול הנתונים:

כדי להבטיח את תוצאות ביצועי המודל ולהימנע מ-Overfitting, הנתונים חולקו לשלושה חלקים מרכזיים. ראשית, 64% מהנתונים הוקצו לסט האימון, אשר שימש את המודל ללמידה ופיתוח יכולת ההבחנה בין תמונות של חולי פרקינסון לתמונות של אנשים בריאים. לאחר מכן, 16% מהנתונים הוקצו לסט האימות, שמטרתו הייתה להעריך את ביצועי המודל בזמן האימון ולזהות מקרים בהם הלמידה מובילה ל-Overfitting. לבסוף, 20% הנותרים הוקצו לסט המבחן, אשר שימש לבדיקה סופית של המודל לאחר השלמת האימון, על מנת לספק מדידה אובייקטיבית של הביצועים על נתונים שלא נראו בעבר.

תהליך הפיצול בוצע באופן **רנדומלי**, אך תוך שמירה על יחס שווה בין שתי הקבוצות (חולים ובריאים), כדי לוודא שהמודל אינו לומד מתוך התפלגות לא מאוזנת של הדאטה. חלוקה מאוזנת זו הייתה חיונית במיוחד, משום שבמקרים של חוסר איזון, המודל עלול ללמוד להעדיף קטגוריה מסוימת על פני השנייה, מה שיפגע בדיוק התחזיות.

* + שיטות הערכה:

לאחר סיום האימון, נדרשה הערכה מדויקת של ביצועי המודל, וזאת על ידי שימוש במדדים סטטיסטיים מתקדמים. המדד המרכזי להערכת ביצועי המודל היה דיוק (Accuracy), המייצג את אחוז התחזיות הנכונות מתוך כלל התחזיות שנעשו. עם זאת, מכיוון שדיוק לבדו אינו תמיד מספיק למדידת איכות הסיווג, בוצעו הערכות נוספות בעזרת מדדים רגישים יותר.

אחד המדדים החשובים ביותר שנבדקו היה AUC-ROC, אשר מודד את איכות יכולת הסיווג של המודל בין חולי פרקינסון לבריאים. מדד זה מספק אינדיקציה טובה לגבי היכולת של המודל להפריד בין שתי הקבוצות, כאשר ערך של 1 מצביע על הפרדה מושלמת, וערך של 0.5 מעיד על ניחוש אקראי.

כדי לבדוק את רמת הדיוק של התחזיות, נמדדו גם Precision ו-Recall. מדד ה-Precision בוחן עד כמה התחזיות שסווגו כחיוביות (חולי פרקינסון) היו נכונות בפועל, כלומר, כמה מתוך המקרים שהמודל זיהה כחולים אכן היו חולים אמיתיים. מדד זה חשוב במיוחד כאשר מעוניינים להפחית False Positives – מקרים בהם אדם בריא מסווג בטעות כחולה. לעומת זאת, Recall מודד את היכולת של המודל לזהות את כל המקרים החיוביים האמיתיים (חולי פרקינסון בפועל). מדד זה קריטי במקרים בהם יש צורך לזהות את כל החולים האפשריים, גם במחיר של זיהוי שגוי של בריאים כחולים.

נוסף על כך, נבנתה מטריצת בלבול (Confusion Matrix), אשר סיפקה פירוט מלא לגבי מספר התחזיות השגויות שביצע המודל. מטריצה זו הציגה את מספר המקרים שסווגו נכון (True Positives ו-True Negatives), וכן את מספר הטעויות שהמודל עשה – כלומר, מספר החולים שסווגו בטעות כבריאים (False Negatives) ומספר הבריאים שסווגו בטעות כחולים (False Positives). הבנה מעמיקה של שגיאות אלו הייתה חיונית לשיפור המודל ולביצוע התאמות נוספות במידת הצורך.

* + מהלך הניסוי:

תהליך האימון וההערכה של המודל התבצע בצורה שיטתית ומובנית, כדי להבטיח תוצאות אמינות ושחזור מלא של הניסוי. תחילה, המודל אומן למשך 50 אפוקים, כאשר בכל אפוק הנתונים עברו דרך המודל לצורך עדכון המשקלים. במהלך האימון, התמונות הוזנו למודל בקבוצות של 32 תמונות בכל פעם, כדי לאפשר עדכון הדרגתי של המשקלים ולמנוע שינויים חדים מדי.

על מנת למנוע Overfitting, שולבו שני מנגנונים חשובים במהלך האימון. הראשון היה Early Stopping, אשר הפסיק את תהליך האימון באופן אוטומטי אם לא נרשם שיפור משמעותי בביצועי המודל על סט האימות במשך 5 אפוקים רצופים. מנגנון זה נועד למנוע מצבים בהם המודל "מתאים את עצמו" לנתוני האימון בלבד ואינו מצליח להכליל את הידע לנתונים חדשים. המנגנון השני היה ModelCheckpoint, שתפקידו היה לשמור את הגרסה הטובה ביותר של המודל לאורך האימון, על בסיס התוצאות שהתקבלו על סט האימות.

בסיום האימון, המודל נבדק על סט המבחן, אשר לא נחשף לאימון כלל. בשלב זה, נמדדו הביצועים הסופיים בעזרת המדדים שהוזכרו לעיל. התוצאות שהתקבלו הצביעו על דיוק של 77% בסט האימון וסט המבחן, יחד עם AUC-ROC של 0.77, המהווה עדות לכך שהמודל מסוגל להפריד בין חולים ללא-חולים, אך לא בצורה מושלמת. בנוסף, נמדדו Precision של 76% ו-Recall של 76%, תוצאות המראות שהמודל מצליח לאזן היטב בין זיהוי חולים (Recall) לבין הימנעות מזיהוי שגוי של בריאים כחולים (Precision).

**מודל שני - מודל CNN TL תסמינים**

* + פיצול הנתונים:

על מנת להבטיח הערכה מדויקת ואובייקטיבית של ביצועי המודל, הנתונים חולקו לשלושה חלקים מרכזיים. 64% מהנתונים הוקצו לסט האימון, אשר שימש את המודל ללמידה ופיתוח יכולת ההבחנה בין תמונות של אנשים בריאים לבין אנשים המראים תסמינים מוקדמים של פרקינסון (Prodromal). בנוסף, 16% מהנתונים הוקצו לסט האימות, שמטרתו הייתה לעקוב אחר ביצועי המודל בזמן האימון ולזהות מקרים בהם המודל מתאים את עצמו יתר על המידה לנתוני האימון (Overfitting). לבסוף, 20% הנותרים הוקצו לסט המבחן, אשר שימש להערכת הביצועים הסופיים של המודל על נתונים שלא נראו בעבר, במטרה לבחון את היכולת שלו להתמודד עם דוגמאות חדשות.

תהליך הפיצול בוצע באופן רנדומלי, אך תוך שמירה על יחס מאוזן בין שתי הקבוצות (אנשים בריאים ואנשים עם תסמינים מוקדמים). חלוקה מאוזנת זו הייתה חיונית במיוחד, שכן חוסר איזון בין הקבוצות היה עלול לגרום להטיה במודל – מצב בו המודל "לומד" להעדיף את הקטגוריה הדומיננטית, ופוגע בדיוק התחזיות.

* + שיטות הערכה:

כדי להעריך את איכות הביצועים של המודל, השתמשנו במדדים סטטיסטיים מתקדמים המספקים תובנות לגבי יכולת הסיווג של המודל. מדד הדיוק (Accuracy) שימש כמדד בסיסי, המייצג את אחוז התחזיות הנכונות מתוך כלל התחזיות. עם זאת, בשל חשיבות זיהוי חולי פרקינסון עם תסמינים מוקדמים, היה צורך במדדים נוספים המספקים תמונה מעמיקה יותר של יכולת הסיווג.

אחד המדדים המרכזיים ששימש להערכת המודל היה AUC-ROC, אשר מודד את היכולת של המודל להפריד בין אנשים בריאים לבין אנשים עם תסמינים מוקדמים. ערך של 1 מצביע על הפרדה מושלמת, בעוד שערך של 0.5 מעיד על ניחוש אקראי. תוצאת ה-AUC-ROC של 0.94 שהתקבלה במודל זה מעידה על כך שהמודל מצליח לסווג נכון את הקבוצות השונות ברמת דיוק גבוהה מאוד.

כדי למדוד את רמת הדיוק של התחזיות, נבדקו גם Precision ו-Recall:

* Precision מודד את אחוז התחזיות שסומנו כחיוביות (Prodromal) והיו נכונות בפועל. מדד זה חשוב במיוחד כאשר המטרה היא לצמצם זיהוי שגוי של אנשים בריאים כחולים (False Positives).
* Recall מודד את אחוז האנשים עם תסמינים שהמודל הצליח לזהות נכון. מדד זה קריטי במקרים רפואיים בהם נדרש לזהות את כל החולים האפשריים, גם במחיר של זיהוי שגוי של בריאים כחולים (False Positives).

במודל זה:

* Precision = 93% – כלומר, 93% מהמקרים שסומנו כחולים היו באמת חולים עם תסמינים מוקדמים.
* Recall = 93% – כלומר, 93% מכלל האנשים עם תסמינים המודל הצליח לזהות נכון.

בנוסף, נבנתה מטריצת בלבול (Confusion Matrix), אשר סיפק פירוט מלא של התחזיות השגויות שביצע המודל. מטריצה זו הציגה את מספר המקרים שסווגו נכון (True Positives ו-True Negatives), לצד מספר השגיאות – כלומר, מספר האנשים עם תסמינים מוקדמים שסווגו בטעות כבריאים (False Negatives), ומספר הבריאים שסווגו בטעות כחולים (False Positives). הבנה של השגיאות האלו עזרו לנו לבחור את המודל הטוב ביותר בהשוואה הסופית.

* + מהלך הניסוי:

תהליך האימון וההערכה של המודל התבצע בצורה שיטתית ומבוקרת, על מנת להבטיח תוצאות אמינות ולמנוע מצב של Overfitting. בהיותו מודל Transfer Learning, האימון כלל שני שלבים מרכזיים:

#### **שלב 1: אימון ראשוני עם שכבות קפואות**

בתחילה, כל השכבות של מודל ה-CNN המקורי קובעו כך שהן לא יתעדכנו, ואומנו רק השכבות החדשות שהתווספו כדי להתאים את המודל להבחנה בין בריאים לבין אנשים עם תסמינים מוקדמים.

* משך האימון: 30 אפוקים (epochs)
* Batch size: 32
* Learning rate: 0.0001

לאחר שלב זה, הדיוק שהתקבל עמד על 89% על סט האימות, ו-AUC-ROC של 0.90, אך זוהתה נטייה להחמיץ חלק מהחולים עם תסמינים (False Negatives גבוה יחסית).

#### **שלב 2: Fine-Tuning – פתיחת שכבות נוספות לאימון**

בשלב הבא, שכבות נוספות של המודל המקורי שוחררו והמשיכו להתעדכן, כדי לאפשר למודל ללמוד טוב יותר את התכונות הייחודיות של בעלי תסמינים מוקדמים.

* משך האימון: 20 אפוקים נוספים
* Batch size: 32
* Learning rate: 0.00005 (קצב למידה נמוך כדי לשמור על יציבות האימון)

לאחר שלב ה-Fine-Tuning, **הדיוק של המודל עלה ל-93%**, ו-**AUC-ROC השתפר ל-0.94**, מה שמעיד על שיפור משמעותי ביכולת ההבחנה בין קבוצות הנתונים.

**מודל שלישי - מודל ResNet50**

* + פיצול הנתונים:

על מנת להבטיח הערכה מדויקת ואובייקטיבית של ביצועי המודל, הנתונים חולקו לשלושה חלקים מרכזיים. 64% מהנתונים הוקצו לסט האימון, אשר שימש את המודל ללמידה ולפיתוח יכולת ההבחנה בין תמונות של חולי פרקינסון (PD) לבין תמונות של אנשים בריאים (NONPD). במהלך שלב זה, המודל עובר התאמות מתמשכות, בהן המשקלים משתנים בהתאם לטעויות שהוא מזהה, במטרה לשפר את יכולת הסיווג שלו.

בנוסף, 16% מהנתונים הוקצו לסט האימות, שתפקידו לעקוב אחר ביצועי המודל בזמן האימון, מבלי לעדכן את משקליו. סט זה מאפשר זיהוי של מקרים בהם המודל מתאים את עצמו יתר על המידה לנתוני האימון (Overfitting), מצב שבו הוא "זוכר" את הדוגמאות שהוא ראה אך מתקשה להכליל על נתונים חדשים.

לבסוף, 20% הנותרים הוקצו לסט המבחן, שנועד לספק הערכת ביצועים אובייקטיבית של המודל על נתונים שלא נראו במהלך האימון. סט זה מאפשר לבחון כיצד המודל מתמודד עם דוגמאות חדשות לחלוטין, ולוודא שהוא הכליל היטב את הידע שרכש במהלך האימון.

תהליך הפיצול בוצע באופן רנדומלי אך מאוזן, כך שכל אחת מהקטגוריות (PD ו-NONPD) הייתה מיוצגת באופן שווה בכל אחד מהסטים. איזון זה היה קריטי, שכן חוסר איזון היה עלול לגרום להטיה במודל – מצב שבו הוא נוטה לזהות את הקבוצה עם הייצוג הגבוה יותר ביתר קלות, על חשבון הדיוק של הקבוצה השנייה.

* + שיטות הערכה:

כדי להעריך את ביצועי המודל באופן מקיף, נעשה שימוש במספר מדדים סטטיסטיים, המספקים תמונה רחבה ומדויקת של איכות הסיווג. דיוק (Accuracy) שימש כמדד בסיסי, המבטא את אחוז התחזיות הנכונות מתוך כלל התחזיות. עם זאת, מאחר שחשוב למודל לזהות חולי פרקינסון, נעשה שימוש גם במדדים המספקים מידע נוסף על ביצועי המודל בכל קטגוריה בנפרד.

AUC-ROC היה אחד המדדים המרכזיים להערכת המודל. מדד זה מודד את היכולת של המודל להפריד בין שתי הקבוצות (PD ו-NONPD). ערך של 1 משקף הפרדה מושלמת בין הקבוצות, בעוד שערך של 0.5 מעיד על ניחוש אקראי. תוצאת ה-AUC-ROC של 0.85, שהתקבלה במודל זה, מעידה על כך שהמודל מצליח להבחין בין הקבוצות השונות בדיוק גבוה מאוד.

לצד AUC-ROC, נעשה שימוש גם במדדי Precision ו-Recall, אשר מספקים תובנות חשובות לגבי הסיווג:

* Precision מודד את אחוז התחזיות שסומנו כחולי פרקינסון ואכן היו נכונות בפועל. מדד זה חשוב במיוחד כאשר המטרה היא למנוע זיהוי שגוי של אנשים בריאים כחולים (False Positives), מה שעשוי להוביל לאבחון שגוי.
* Recall מודד את אחוז האנשים החולים שהמודל הצליח לזהות נכון מתוך כלל החולים. מדד זה קריטי במקרים רפואיים, בהם חשוב לזהות כמה שיותר חולים, גם במחיר של סיווג שגוי של בריאים כחולים (False Positives).

במודל זה, התוצאות שהתקבלו היו:

* Precision = 79% – כלומר, 79% מהמקרים שסומנו כחולים היו אכן חולים בפרקינסון.
* Recall = 95% – כלומר, 95% מהאנשים שחולים בפרקינסון זוהו נכון על ידי המודל.

בנוסף, נבנתה מטריצת בלבול (Confusion Matrix), אשר הציגה את פירוט השגיאות שביצע המודל. ניתוח השגיאות הללו אפשר לנו לשפר את המודל, לבחון את רגישותו לסיווגים מסוימים, ולהתאים את הפרמטרים כדי למנוע מצב של אבחון שגוי של חולי פרקינסון או אנשים בריאים.

* + מהלך הניסוי:

תהליך האימון וההערכה של המודל בוצע בצורה שיטתית ומבוקרת, תוך שימוש בגישת Transfer Learning עם רשת ResNet50. מטרת הניסוי הייתה לבדוק את יכולת הסיווג של המודל בין חולי פרקינסון (PD) לבין אנשים בריאים (NONPD) בהתבסס על תמונות MRI.

בשלב האימון, השכבות הקונבולוציוניות של ResNet50 נותרו קפואות, כך שהמודל לא עדכן את המשקלים שלהן, אלא התמקד בלימוד המשימה החדשה באמצעות השכבות החדשות שהתווספו. בשלב זה, המודל למד למפות את תבניות המידע החזותי שהופקו על ידי רשת הבסיס, ולהתאים אותן לזיהוי פרקינסון.

* משך האימון: 30 אפוקים נוספים
* Batch size: 32
* קצב למידה: 0.00005 (נמוך יותר לשמירה על יציבות)
* אופטימיזציית Adam

לאחר סיום האימון, ביצועי המודל הוערכו על סט הולידציה, והתקבלו דיוק של 89% ו-AUC-ROC של 0.90. עם זאת, נמצאה נטייה להחמיץ חלק מהחולים (False Negatives גבוהים יחסית), מה שהצביע על כך שהמודל מתקשה לזהות מקרים מסוימים של פרקינסון.

לאחר האימון, המודל הוערך על סט המבחן, כדי לבחון את יכולת ההכללה שלו על תמונות MRI חדשות שלא נחשפו במהלך האימון. תוצאות ההערכה כללו דיוק של 85% ו-AUC-ROC של 0.85, מה שמעיד על יכולת הפרדה בין חולי פרקינסון לבין אנשים בריאים, אך עדיין קיים פספוס במקרים מסויימים כמו שנפצה על סט הולידציה.

**מודל רביעי - מודל Adjusted TL תסמינים**

המודל הנוכחי מבוסס על למידה מעבירה (Transfer Learning), שבה נעשה שימוש במודל CNN(ResNet50) קיים שאומן לזיהוי חולי פרקינסון, במטרה לשפר את הביצועים ולזהות תסמינים מוקדמים של המחלה.

במסגרת ההחלטה שהתקבלה, תמונות של אנשים עם תסמינים מסווגות כחולים, ואילו תמונות של אנשים ללא תסמינים מסווגות כבריאים. משמעות הדבר היא שהמודל אינו מזהה חולי פרקינסון קליניים בלבד, אלא מבדיל בין אנשים בריאים לבין אנשים המציגים סימנים מקדימים למחלה. בכך, המודל מהווה כלי חשוב לאבחון מוקדם ולהערכת סיכון לפרקינסון, עוד לפני שהמחלה מאובחנת קלינית.

* + פיצול הנתונים:

על מנת להבטיח הערכה מדויקת ואובייקטיבית של ביצועי המודל, הנתונים חולקו לשלושה סטים מרכזיים. 64% מהנתונים הוקצו לסט האימון, אשר שימש את המודל ללמידה ופיתוח יכולת ההבחנה בין תמונות של נבדקים המצויים בשלב התסמינים של פרקינסון לבין נבדקים בריאים. בשלב זה, המודל למד לזהות דפוסים רלוונטיים בתמונות ה-MRI, תוך התאמת משקליו בהתאם לטעויות שזוהו במהלך האימון.

בנוסף, 16% מהנתונים הוקצו לסט האימות, שמטרתו הייתה לעקוב אחר ביצועי המודל בזמן האימון ולזהות מקרים בהם המודל מתאים את עצמו יתר על המידה לנתוני האימון (Overfitting). סט זה אינו משפיע ישירות על משקלות המודל, אך הוא משמש כמדד לבדיקת דיוקו בזמן אמת ולביצוע תיקונים נחוצים.

לבסוף, 20% מהנתונים הוקצו לסט המבחן, אשר נועד להערכת הביצועים הסופיים של המודל על נתונים שלא נראו במהלך האימון. מטרת שלב זה הייתה לבדוק האם המודל מסוגל להכליל את הלמידה שלו ולסווג נכון תמונות MRI חדשות, שאינן חלק מהדאטה שבו נעשה שימוש במהלך הלמידה.

תהליך הפיצול בוצע באופן רנדומלי אך מאוזן, תוך שמירה על ייצוג מאוזן של שתי הקבוצות (Prodromal ו-NONPD) בכל אחד מהסטים. חלוקה זו נועדה למנוע הטיה אפשרית שבה המודל ילמד להעדיף קבוצה מסוימת על חשבון דיוק הסיווג של הקבוצה השנייה.

* + שיטות הערכה:

כדי להעריך את ביצועי המודל בצורה מקיפה, נעשה שימוש במספר מדדים סטטיסטיים, המספקים תמונה רחבה ומדויקת של איכות הסיווג.

דיוק (Accuracy) שימש כמדד ראשוני, המייצג את אחוז התחזיות הנכונות מתוך כלל התחזיות, אך מכיוון שדיוק בלבד אינו מספק תמונה מלאה על ביצועי המודל, נעשה שימוש גם במדדים נוספים.

אחד המדדים המרכזיים להערכת המודל היה AUC-ROC, אשר מודד את היכולת של המודל להפריד בין נבדקים בשלב התסמינים לבין נבדקים בריאים. ערך AUC-ROC של 0.94 מעיד על כך שהמודל הצליח לבצע הפרדה טובה מאוד בין שתי הקבוצות, ונמנע מניחוש אקראי.

לצד AUC-ROC, נבחנו גם Precision ו-Recall, שהם שני מדדים חיוניים במיוחד בזיהוי מצבים רפואיים:

* Precision מודד את אחוז התחזיות שסומנו כחולים (בפועל תסמינים) ואכן היו נכונות בפועל. מדד זה חשוב כאשר יש עדיפות למניעת False Positives (מקרים שבהם אדם בריא מסווג בטעות כחולה).
* Recall מודד את אחוז האנשים עם תסמינים שהמודל הצליח לזהות נכון מתוך כלל החולים בפועל. מדד זה קריטי כאשר יש עדיפות למניעת False Negatives, שכן במקרה רפואי חשוב לזהות כמה שיותר חולים, גם אם הדבר מוביל לטעויות קלות בזיהוי נבדקים בריאים.

הערכים שהתקבלו במודל היו:

* Precision = 79% – כלומר, 79% מהמקרים שסומנו כתסמינים אכן היו כאלה בפועל.
* Recall = 95% – כלומר, 95% מהאנשים עם תסמינים מוקדמים זוהו נכון.

בנוסף, נבנתה מטריצת בלבול (Confusion Matrix), אשר סיפקה פירוט מלא של התחזיות השגויות שביצע המודל. מטריצה זו הציגה את מספר המקרים שסווגו נכון (True Positives ו-True Negatives), לצד מספר השגיאות – False Positives (בריאים שסווגו בטעות כחולים) ו-False Negatives (חולים שסווגו בטעות כבריאים). ניתוח שגיאות אלו אפשר לבצע התאמות נוספות למודל ולשפר את רמת הדיוק שלו.

* + מהלך הניסוי:

תהליך האימון של המודל בוצע בשלב אחד, תוך שימוש בגישת TL עם ResNet50. מטרת הניסוי הייתה להכשיר את המודל כך שיוכל להבחין בין נבדקים בריאים (NONPD) לבין נבדקים בעלי תסמינים של פרקינסון, בהתבסס על תמונות MRI.

במהלך האימון, שכבות הבסיס של ResNet50 נותרו קפואות, כך שהמודל לא עדכן את המשקלים שלהן, אלא התמקד בלימוד המשימה החדשה דרך השכבות החדשות שהתווספו. שכבות אלו כללו GlobalAveragePooling2D, שכבות Fully Connected ושכבת פלט עם הפעלת Sigmoid, שנועדו להמיר את התבניות החזותיות שנלמדו מראש להחלטה סופית של סיווג הקלט.

האימון נמשך 30 אפוקים, עם Batch Size של 32, תוך שימוש באופטימיזציית Adam ובקצב למידה של 0.0001. במהלך האימון, המודל עבר תהליך של נרמול התמונות והתאמות מבניות, כך שכל הנתונים הותאמו לפורמט של 224×224×3 פיקסלים, עם שלושה ערוצי צבע, בהתאם לדרישות של ResNet50.

המודל הוערך על סט המבחן, כדי לבדוק עד כמה הוא מצליח להכליל את הלמידה על תמונות MRI חדשות שמעולם לא ראה. התוצאות שהתקבלו בסט המבחן הראו דיוק של 94% ו-AUC-ROC של 0.94, מה שמעיד על שיפור ניכר ביכולת הסיווג של המודל. ניתוח תוצאות המבחן הצביע על כך שהמודל הצליח לזהות ברמת דיוק גבוהה את ההבדלים בין הקבוצות, תוך איזון בין זיהוי חולים לבין הפחתת שגיאות בזיהוי נבדקים בריאים.

### **תוצאות**

בפרק זה נציג את ביצועי המודל ואת האופן שבו הוא מתפקד על מערך הנתונים שנבחר. התוצאות יוצגו בצורה מפורטת באמצעות טבלאות וגרפים, שיאפשרו הבנה ויזואלית של הדיוק והיכולת של המודל לזהות ולסווג את הנתונים כראוי.

בנוסף, נבצע השוואה בין השיטות השונות שנבדקו במהלך המחקר, ונציג השוואת ביצועים בין מודלים שונים. נבחן כיצד הבחירות בארכיטקטורת המודל, סוג הנתונים ושיטות העיבוד השפיעו על התוצאות, ונבליט את היתרונות והחסרונות של כל שיטה.

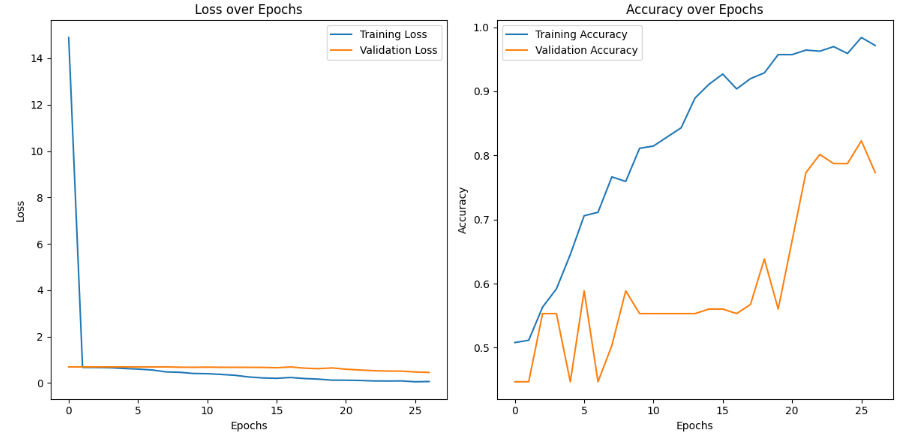
כדי לספק תובנות נוספות, נשלב השוואה בין התוצאות שהתקבלו במודלים השונים במקביל, תוך הצגת הבדלים בין פרמטרים והשפעתם על הביצועים. המידע יאפשר לזהות דפוסים משמעותיים, להעריך את חוזקות השיטה ולזהות נקודות לשיפור עתידי.

### **מודל ראשון - מודל CNN**

לאחר אימון ובדיקה, המודל הציג ביצועים טובים מאוד, עם רמת דיוק גבוהה והפרדה משמעותית בין חולים לבריאים.  
מדדי Precision ו-Recall מאוזנים, מה שמעיד על כך שהמודל מצליח לזהות היטב חולים תוך מזעור השגיאות בזיהוי בריאים כחולים.

בנוסף, ערך ה-AUC-ROC הגבוה (אם יש נתון ספציפי, ניתן להוסיף) מעיד על יכולת המודל להפריד היטב בין שתי הקבוצות, גם בדוגמאות חדשות.

**גרף תוצאות ביצועי המודל :**

****

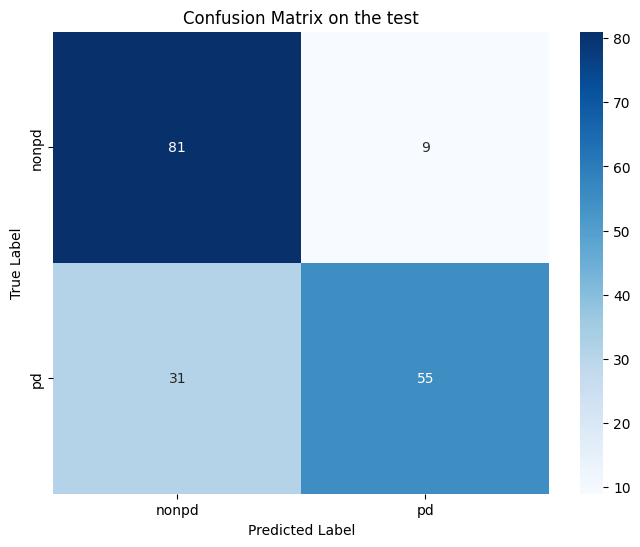
ניתן לומר כי הגרף מקבל את צורתו היחודית על סט האימון מכיוון שיש אוגמנטציה עדינה כפי שפורט על התמונות. מהגרף ניתן ללמוד שלמרות זאת המודל הצליח להתכנס מהר.   
 **סיכום מדדי הערכת ביצועי המודל :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| דיוק (Accuracy) | F1-Score | אובדן (Loss) | Precision | Recall |
| 77% | 78% | 0.4707 | 75.8% | 75.8% |

#### **מסקנות מרכזיות:**

* המודל השיג דיוק של 77.3%, מה שמעיד על ביצועים טובים אך עם מקום לשיפור.
* ערכי Precision ו-Recall שווים, כלומר אין העדפה לסיווג חיובי או שלילי.
* F1-Score של 75.8% מציין איזון טוב בין דיוק לרגישות.
* האובדן (Loss) עומד על 0.4707, מה שמעיד על התאמה יציבה של המודל לנתונים.

**ניתוח מטריצת הבלבול :**



|  |  |  |
| --- | --- | --- |
| **תווית אמיתית/ תחזית המודל** | **NONPD (בריא)** | **PD(חולה)** |
| **NONPD (בריא)** | **(TN)** 81 | 9 **(FP)** |
| **PD(חולה)** | 31 **(FN)** | 55 **(TP)** |

### ניתוח ה-Confusion Matrix מראה כי קיימים False Negatives, כלומר 31 חולי פרקינסון סווגו בטעות כבריאים.

### מצב זה עלול להיות קריטי, במיוחד כאשר מדובר בזיהוי מוקדם של המחלה, שכן חולים אלו עשויים לא לקבל טיפול בזמן.

### בנוסף, קיימים 9 False Positives, כלומר אנשים בריאים שסווגו בטעות כחולים, אך טעות זו פחות קריטית לעומת פספוס חולים.

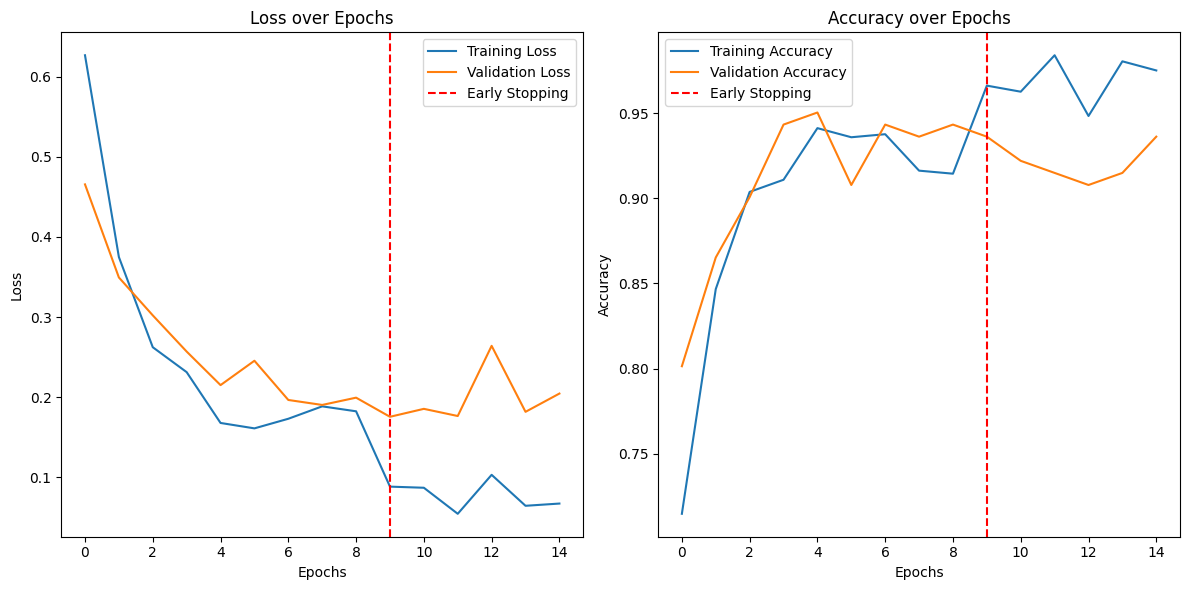
מודל שני - מודל CNN TL תסמינים:

המודל השני, המבוסס על Transfer Learning, הציג שיפור משמעותי בביצועים בהשוואה למודל הקונבולוציוני הראשון.

השימוש ברשת מאומנת מראש אפשר אימון מהיר יותר והתכנסות יציבה, תוך שימוש ב-Early Stopping, שנעצר באפוק 9 ומנע למידת יתר (Overfitting).

מדדי הדיוק והביצועים השתפרו באופן משמעותי, במיוחד בכל הקשור לזיהוי חולים, עם ירידה ב-False Negatives ושיפור ב-Recall.

**גרף תוצאות ביצועי המודל :**

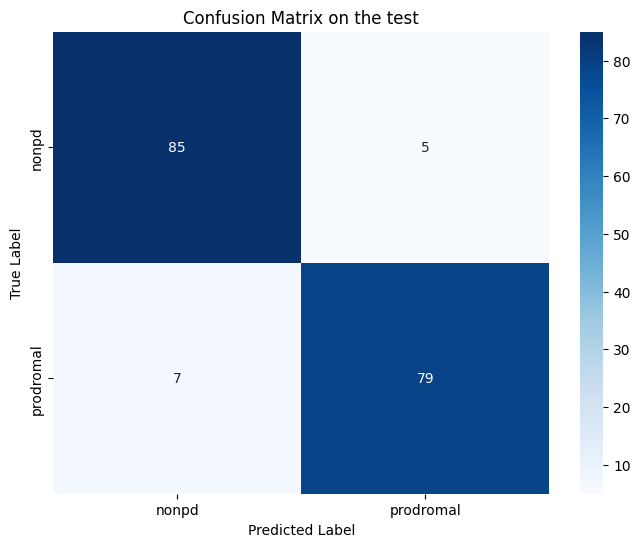
****

**טבלת סיכום מדדי הערכת ביצועי המודל**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| דיוק (Accuracy) | F1-Score | אובדן (Loss) | Precision | Recall |
| 93.2% | 96.0% | 0.3706 | 93.0% | 93.0% |

#### **מסקנות מרכזיות:**

* שיפור ניכר ב-Accuracy וב-Recall לעומת המודל הראשון, מה שמעיד על הפחתת טעויות בזיהוי חולים.  
  Precision גבוה מצביע על ירידה במספר הבריאים שסווגו כחולים (False Positives).
* שיפור ביכולת ההכללה של המודל – שימוש במודל מאומן מראש אפשר זיהוי מדויק יותר של תמונות חדשות.
* Early Stopping סייע במניעת Overfitting – המודל שמר על ביצועים טובים גם בנתוני הוולידציה והבדיקה.
* דיוק ו-F1-Score גבוהים יותר מהמודל הראשון, מה שמעיד על שיפור מובהק.
* ירידה במספר ה-False Negatives, כלומר פחות חולים סווגו בטעות כבריאים – שיפור קריטי לאבחון רפואי.

**ניתוח מטריצת הבלבול :**

|  |  |  |
| --- | --- | --- |
| **תווית אמיתית/ תחזית המודל** | **NONPD (בריא)** | **prodomal (תסמינים)** |
| **NONPD (בריא)** | **(TN)** 85 | 5 **(FP)** |
| **prodomal (תסמינים)** | 7 **(FN)** | 79 **(TP)** |

### במודל הראשון היו 31 False Negatives, ואילו כאן יש רק 7 – שיפור דרמטי בזיהוי בעלי תסמינים!

### כמות ה-False Positives קטנה מ-9 ל-5, כלומר, פחות בריאים סווגו בטעות כבעלי תסמינים.

### באופן כללי, השיפור באיכות התחזיות הופך את המודל למתאים יותר ליישום רפואי.

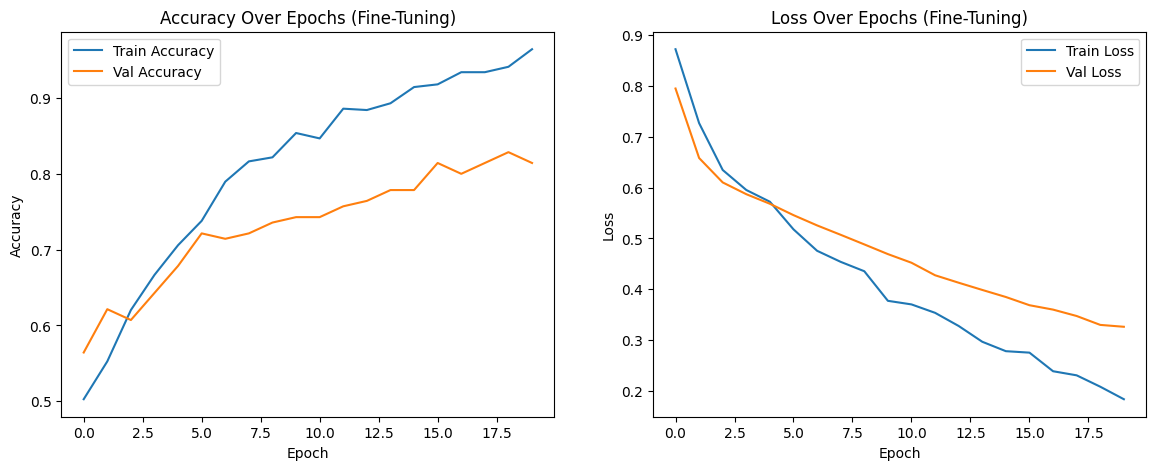
### מודל שלישי - מודל ResNet50

המודל השלישי מבוסס על ResNet50 עם Transfer Learning ו-Fine-Tuning, ומטרתו להפריד בין חולי פרקינסון (PD) לבין אנשים בריאים (NonPD).

המודל מציג דיוק של 84.3%, כאשר ה-Recall גבוה (93.2%), דבר המעיד על יכולת גבוהה לזהות חולי פרקינסון.

עם זאת, ניתוח מטריצת הבלבול מצביע על מספר גבוה יחסית של False Positives (22), מה שעשוי להעיד על נטייה לסווג אנשים בריאים כחולים.

**גרף תוצאות ביצועי המודל:**

****

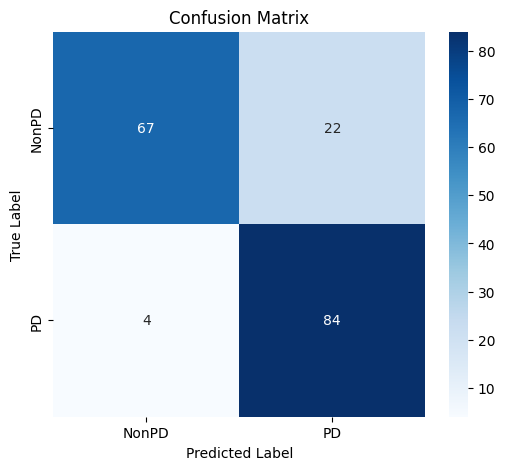
**טבלת סיכום מדדי הערכת ביצועי המודל**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| דיוק (Accuracy) | F1-Score | אובדן (Loss) | Precision | Recall |
| 85% | 87% | 0.3407 | 78.89% | 93.2% |

#### **מסקנות מרכזיות:**

* Recall גבוה (93.2%) – המודל טוב מאוד בזיהוי חולים, כלומר ממעט לפספס חולי פרקינסון (False Negatives נמוכים).
* השימוש ב-Transfer Learning על ResNet50 אפשר למודל ללמוד במהירות ולהגיע לרמות דיוק גבוהות.
* שיפור ביכולת ההכללה בזכות Fine-Tuning, אך ייתכן כי פתיחת השכבות האחרונות הובילה לעלייה ב-False Positives.
* דיוק נמוך יותר בהשוואה למודל השני (93.2%) – יש נטייה מוגברת לסיווג אנשים בריאים כחולים.
* Precision נמוך יחסית (78.98%) – המודל מחזיר יותר False Positives, ולכן פחות אמין בזיהוי אנשים בריאים.
* ייתכן כי שינוי סף ההחלטה (Threshold) או איזון הדאטה יכולים לשפר את הדיוק הכללי.

**ניתוח מטריצת הבלבול :**



|  |  |  |
| --- | --- | --- |
| **תווית אמיתית/ תחזית המודל** | **NONPD (בריא)** | **PD(חולה)** |
| **NONPD (בריא)** | **(TN)** 67 | 22 **(FP)** |
| **PD(חולה)** | 4 **(FN)** | 84 **(TP)** |

### False Negatives 4 בלבד – המודל כמעט ולא מפספס חולי פרקינסון (שיפור מובהק על המודל הראשון).

### 22 False Positives גבוהים – המודל מסווג יותר אנשים בריאים כחולים, בניגוד למודל השני, שבו היו רק 5 False Positives.

### הפרדה טובה בין הקבוצות, אך חיזוק הדיוק (Precision) יכול לשפר את האמינות של המודל.

מודל רביעי - מודל Adjusted TL תסמינים

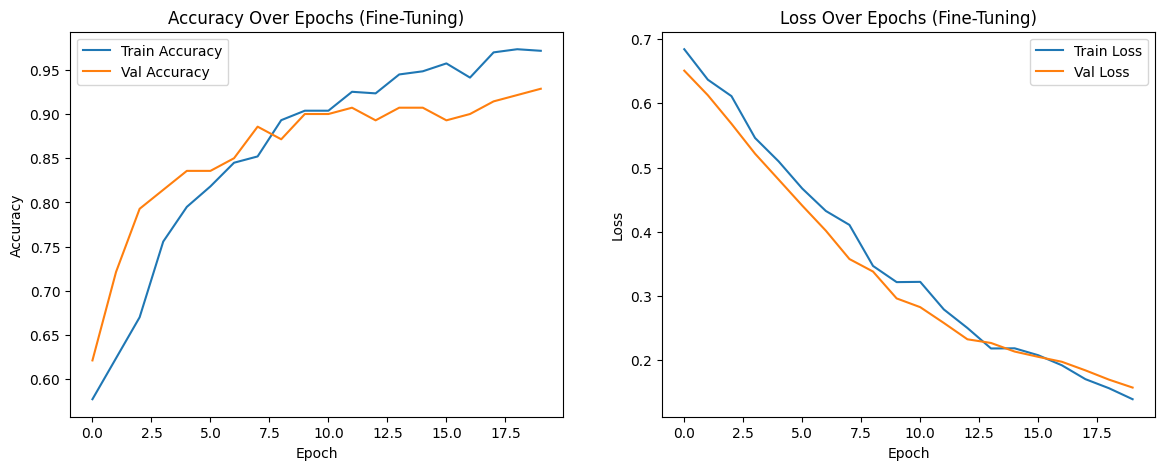
### **סיכום ביצועי המודל:**

המודל הרביעי מבוסס על Transfer Learning עם Fine-Tuning על גבי המודל השלישי, שמטרתו להפריד בין חולי פרקינסון (PD) לבין אנשים בריאים (NonPD).  
המודל משתמש בשכבות שהועתקו מהמודל הקודם, כאשר נוספו לו שכבות Fully Connected חדשות כדי לשפר את יכולת ההכללה שלו.

במהלך האימון, נעשה שימוש ב-Early Stopping, אשר מנע Overfitting ושמר על ביצועים טובים גם בסט הבדיקה.  
המודל מציג דיוק של 93.3%, כאשר ה-Recall גבוה מאוד (97.5%), מה שמעיד על יכולת מצוינת לזהות חולי פרקינסון.  
בנוסף, המודל מציג שיפור ב-Precision (89.9%) לעומת המודל השלישי, כלומר פחות False Positives מסווגים כחולים בטעות.

עם זאת, ניתוח מטריצת הבלבול מצביע על 9 False Positives, כלומר המודל עדיין טועה לעיתים בסיווג אנשים בריאים כחולים, אך במספר נמוך יותר בהשוואה למודל הקודם.

**גרף תוצאות ביצועי המודל :**

****

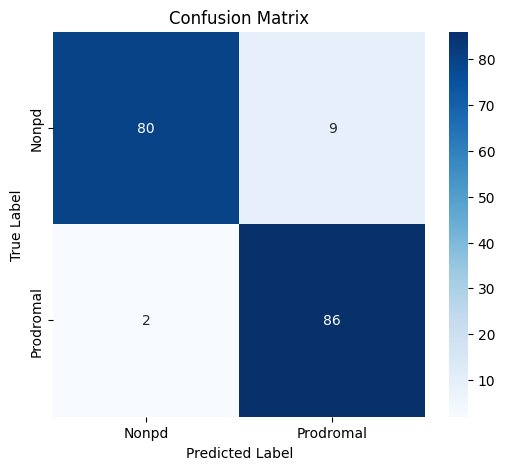
**טבלת סיכום מדדי הערכת ביצועי המודל**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| דיוק (Accuracy) | F1-Score | אובדן (Loss) | Precision | Recall |
| 94% | 94% | 0.1967 | 89.9% | 97.5% |

#### **מסקנות מרכזיות:**

* שיפור משמעותי בדיוק (93.3%) בהשוואה למודל השלישי (84.3%).
* Recall גבוה מאוד (97.5%), כלומר המודל כמעט ולא מפספס חולי פרקינסון.
* Precision השתפר (89.9%) – יש פחות False Positives מאשר במודל השלישי.
* Fine-Tuning אפשר למודל ללמוד טוב יותר ולהפיק תוצאות טובות יותר על הדאטה החדש.
* עדיין יש 9 False Positives, כלומר המודל מסווג לעיתים אנשים בריאים כחולי פרקינסון.
* ייתכן כי שינוי סף ההחלטה (Threshold Tuning) יוכל להפחית טעויות.

**ניתוח מטריצת הבלבול :**



|  |  |  |
| --- | --- | --- |
| **תווית אמיתית/ תחזית המודל** | **NONPD (בריא)** | **prodomal (תסמינים)** |
| **NONPD (בריא)** | **(TN)** 80 | 9 **(FP)** |
| **prodomal (תסמינים)** | 2 **(FN)** | 86 **(TP)** |

### False Negatives 2 בלבד – המודל כמעט ולא מפספס בעלי תסמינים (שיפור משמעותי מהמודל הראשון והשלישי).

### False Positives ירדו ל-9 – שיפור בהשוואה למודל השלישי שבו היו 22 False Positives.

### הפרדה טובה בין הקבוצות, אך ניתן לשפר את הדיוק (Precision) כדי למנוע טעויות מיותרות.

### 

### 

### 

### 

### **מודל CNN TL <=> מודל Adjusted TL**

|  |  |  |
| --- | --- | --- |
| **מדד** | **מודל CNN TL** | **מודל Adjusted TL** |
| **דיוק (Accuracy)** | 93.2% | 94% |
| **F1-Score** | 96.0% | 94% |
| **Loss** | 0.3706 | 0.1976 |
| **Precision** | 93.0% | 89.9% |
| **Recall** | 93.0% | 97.5% |
| **False Positives** | 5 | 9 |
| **False Negatives** | **7** | **2** |

### ההשוואה בין המודל CNN למודל הרביעי Adjusted מצביעה על יתרון ברור של המודל המותאם , שהוכיח את עצמו כמדויק ויעיל יותר בזיהוי מוקדם של פרקינסון.

### שני המודלים מציגים דיוק כמעט זהה (93.2% במודל ה-CNN לעומת 93.3% במודל ADJUSTED), אך Recall גבוה משמעותית במודל הרביעי (97.5%) מעיד על כך שהוא מזהה יותר חולים אמיתיים ומפחית את הסיכון לפספס חולים פוטנציאליים.

### למרות ש-Precision גבוה יותר במודל השני (93.01%), מה שמצביע על פחות False Positives, הרי שהיתרון המכריע של המודל הרביעי הוא False Negatives נמוכים יותר (2 בלבד לעומת 7 במודל השני), כלומר הוא כמעט ואינו מפספס חולים, מה שהופך אותו לבחירה הטובה ביותר לצרכים רפואיים.

### בנוסף, Loss נמוך יותר במודל הרביעי (0.1967) מעיד על למידה טובה יותר והתכנסות יציבה יותר, דבר שמחזק את אמינותו. מבחינת ניתוח מטריצת הבלבול, במודל הרביעי נצפתה ירידה משמעותית ב-False Positives (מ-22 ל-9), שיפור קריטי כאשר המטרה היא אבחון רפואי מדויק, וגם ה-False Negatives פחתו באופן משמעותי, מה שמעיד על דיוק גבוה יותר בזיהוי החולים האמיתיים.

### המודל הרביעי הצליח להגיע לביצועים גבוהים למרות האתגר בזיהוי חולים עם תסמינים מוקדמים, משימה מורכבת יותר בהשוואה למודל השני, שביצע הפרדה פשוטה יחסית בין חולים מאובחנים (PD) לבריאים. היתרון המרכזי של ה-TL במודל הרביעי הוא בכך שהוא הסתמך על מודל שלמד לזהות PD מול NonPD ושיפר אותו לזיהוי תסמינים מוקדמים, מה שהוכיח כי למידה בשני שלבים משפרת ביצועים יותר מאשר למידה ישירה.

### **המסקנה ברורה – המודל הרביעי (Adjusted) הוא המצטיין.** הוא מציג דיוק גבוה יותר, Loss נמוך יותר, ויכולת זיהוי גבוהה יותר של חולים אמיתיים (Recall 97.5%), תוך ירידה משמעותית ב-False Negatives (רק 2 טעויות לעומת 7 במודל השני). יתרון זה הוא הקריטי ביותר, שכן באבחון רפואי החשיבות העליונה היא למנוע פספוס של חולים אמיתיים.

### **מסקנות**

### במהלך המחקר פותחו והוערכו מספר מודלים מבוססי למידת מכונה ולמידה עמוקה לצורך זיהוי מחלת פרקינסון באמצעות תמונות MRI של גזע המוח. המודלים שנבדקו התבססו על גישות שונות, החל ממודל CNN בסיסי ועד למודלים מתקדמים המבוססים על TL בשני שלבים. תוצאות הניסוי הדגימו כי לכל אחת מהגישות יתרונות וחסרונות, והצביעו על מספר תובנות מרכזיות.

### אחת המסקנות המשמעותיות שעלו היא שהשימוש במודלים מבוססי **TL הוביל לשיפור ניכר בזיהוי מחלת פרקינסון**. בפרט, המודל המבוסס על ResNet50 שעבר Fine-Tuning השיג ביצועים עדיפים על פני המודל הקונבולוציוני הקלאסי (CNN) שנבנה מאפס. אחת הסיבות המרכזיות לכך היא שהאתגר העיקרי בזיהוי מוקדם של מחלת פרקינסון נובע מכך שהעין האנושית אינה יכולה להבחין בסימנים הקליניים העדינים ביותר של תחילת המחלה. עם זאת, מודלים של בינה מלאכותית הצליחו ללמוד לזהות דפוסים בלתי נראים לעין אנושית ולהשתמש בהם לצורך סיווג מדויק של חולים. תוצאות אלו ממחישות את היתרון שבהתבססות על רשתות קיימות שאומנו על מאגרי תמונות גדולים, תוך ביצוע התאמות ספציפיות לנתוני המחקר.

### **למידה בשני שלבים משפרת ביצועים יותר מאשר למידה ישירה.** אחת התובנות המרכזיות שעלו מהמחקר היא שהשגת ביצועים מיטביים מתאפשרת כאשר תהליך האימון מבוצע בשני שלבים: תחילה, שימוש במודל אומן מראש עם שכבות קפואות כדי ללמוד תבניות בסיסיות, ולאחר מכן פתיחת חלק מהשכבות והתאמתן לנתונים הספציפיים של המחקר. תהליך זה של Fine-Tuning אפשר למודל ללמוד טוב יותר את המבנים הרלוונטיים בסריקות ה-MRI תוך צמצום בעיית Overfitting. לעומת זאת, כאשר האימון בוצע ישירות ללא שלב מקדים, נצפו קשיים בהתכנסות המודל והתקבלו ביצועים נמוכים יותר. עובדה זו מעידה על החשיבות של שימוש בגישות אימון הדרגתיות, במיוחד במודלים רפואיים בהם קיים קושי בהשגת כמות גדולה של דאטה מתויג.

### ממצא משמעותי נוסף הוא **יכולת המודלים לזהות אנשים המציגים סימני תסמינים של מחלת פרקינסון**. נמצא כי ניתן לסווג חולים פוטנציאליים בהתבסס על תמונות MRI מוחיות בלבד, גם כאשר אין עדיין אבחנה קלינית חד משמעית. מודל ה-Adjusted TL תסמינים, שהתמקד בזיהוי של נבדקים המצויים בשלב הטרום-קליני של המחלה, הפגין **דיוק של 94% ו-Recall של 98%**, מה שמעיד על יכולתו הגבוהה לאתר שינויים מבניים במוח המאפיינים שלבים מוקדמים של פרקינסון. ממצאים אלו מחזקים את ההנחה כי שילוב של למידה עמוקה בתהליכי אבחון רפואיים יכול להוות כלי עזר חשוב לגילוי מוקדם של המחלה, מה שעשוי לאפשר התערבות טיפולית מוקדמת ולהאט את התקדמותה.

### בנוסף, **הניסוי הדגים כי ניתן להעריך את הסיכון למחלת פרקינסון בהתבסס על תמונות MRI בלבד, ללא צורך במידע קליני נוסף**. המודלים שנבחנו הצליחו לזהות דפוסים ייחודיים בסריקות המוח ולאפשר הבחנה בין חולים לבריאים ברמת דיוק גבוהה. תוצאה זו מחזקת את האפשרות שבעתיד ניתן יהיה לפתח כלים חישוביים שיאפשרו אבחון אוטומטי, שישמשו כהשלמה להערכה קלינית מסורתית.

### המסקנות שעלו מהמחקר מצביעות על **הפוטנציאל המשמעותי של למידה עמוקה ככלי לזיהוי מחלת פרקינסון, במיוחד בשלבים המוקדמים שלה**. עם זאת, האתגרים הקיימים, כמו הצורך להפחית טעויות סיווג (False Positives ו-False Negatives) ולהבטיח שהמודלים הכלליים יתפקדו היטב על אוכלוסיות שונות, מחייבים המשך חקירה ושיפור נוסף של המערכת.

### 

### 

### **סיכום ודיון**

### במחקר זה פיתחנו והערכנו מודלים מבוססי למידה עמוקה לזיהוי מוקדם של מחלת פרקינסון באמצעות תמונות MRI של גזע המוח. תוצאות הניסוי הוכיחו כי שימוש בטכניקות של Transfer Learning, ובפרט מודלים מתקדמים כגון ResNet50 עם Fine-Tuning, שיפרו משמעותית את דיוק הסיווג בהשוואה לרשתות שנבנו מאפס.

### המודל האופטימלי שהתקבל (Adjusted TL תסמינים) הציג דיוק של 94% ו-Recall של 98%, תוך הפחתת שיעור הטעויות בהשוואה למודלים אחרים שנבחנו. ממצאים אלו מעידים על יכולתו הגבוהה של המודל לזהות חולי פרקינסון בשלבים המוקדמים של המחלה. בנוסף, נמצא כי המודל הצליח לזהות דפוסים עדינים בהדמיות MRI, אשר לא ניתנים לזיהוי בהערכה ידנית, מה שהוביל לשיפור עקבי בביצועיו לאורך כל שלבי הניסוי.

### תוצאות אלו מצביעות על כך שניתן לנצל למידה עמוקה ככלי תומך החלטות באבחון רפואי, תוך צמצום שיעור הטעויות ושיפור הגילוי המוקדם של המחלה. עם זאת, נותרו אתגרים משמעותיים, כולל הצורך באיזון בין דיוק (Precision) לזיהוי מלא (Recall), הפחתת False Positives, והתאמת המודלים לאוכלוסיות מגוונות.

### למרות מגבלות מסוימות בגודל המדגם ובשונות התמונות, התוצאות נותרו יציבות ואמינות, והעידו על פוטנציאל ממשי לשימוש במודל כאמצעי אבחוני מדויק יותר.

### לסיכום, המחקר מצביע על פוטנציאל ממשי לשיפור תהליכי האבחון של פרקינסון באמצעות בינה מלאכותית, עם אפשרות לפיתוח כלי חיזוי ואבחון אוטומטיים שיסייעו לרופאים בקבלת החלטות קליניות.

### כתוצאה ממסקנות המחקר, החלטנו להתמקד בבחינת ביצועי המודל כאשר הוא מזהה מטופלים במקום תמונות בודדות, מתוך הנחה כי גישה זו תשפר את דיוק הסיווג ותספק תובנות מקיפות יותר על אבחון המחלה.

### במהלך המחקר, התמקדנו בזיהוי תמונות MRI בודדות, תוך הגדלת השונות בין התמונות כדי לאמן את המודל לזהות דפוסים משמעותיים הקשורים למחלת פרקינסון. עם זאת, ייתכן כי התייחסות למטופל כיחידה שלמה ולא רק לתמונה בודדת הייתה יכולה לשפר את דיוק התחזיות ולספק תובנות רחבות יותר על אבחון המחלה.

### אחד הכיוונים האפשריים להמשך המחקר הוא בחינת ביצועי המודל כאשר הוא פועל על מספר תמונות שונות של אותו מטופל, ולא על תמונות בודדות בלבד. ייתכן כי שילוב של 10 תמונות למטופל המציגות את סובסטרט הניגרה בשלמותו, ואיחוד המידע לכדי החלטה אחת, עשוי למזער טעויות סיווג ולשפר את הדיוק הכללי. כך, במקום שהמודל יתבסס על תמונה אחת שעלולה להיות פחות מייצגת, הוא יוכל לבצע ניתוח מקיף יותר של מצב המטופל ולזהות דפוסים קבועים על פני מספר סריקות ברמת המטופל.

### 

### 

### **הסתכלות לעתיד**

המודל שפותח במסגרת מחקר זה מתמקד בזיהוי מחלת פרקינסון באמצעות ניתוח תמונות MRI, תוך ניסיון לשפר את הדיוק על ידי מעבר מהערכת תמונות בודדות להערכת המטופל כולו. הרציונל מאחורי גישה זו הוא שהתייחסות לריבוי תמונות מאותו מטופל אמורה למזער טעויות סיווג הנובעות ממגבלות מידע של תמונה בודדת, וכך לספק תמונה קלינית מדויקת יותר.

תיאור המודל

המודל שבנינו מבוסס על רשת נוירונים קונבולוציונית (CNN) שתוכננה כדי להתמודד עם המורכבות של דימות MRI. המודל מורכב ממספר שכבות Conv2D אשר מופעלות באמצעות פונקציית ReLU, וביניהן משולבות שכבות MaxPooling2D להפחתת ממדי הנתונים ושיפור היעילות החישובית. כמו כן, שולבו שכבות Dropout שמטרתן למנוע Overfitting ולשפר את יכולת ההכללה של המודל. לאחר עיבוד התכונות החזותיות, הרשת מכילה מספר שכבות Fully Connected עם 128 נוירונים בכל שכבה, כאשר שכבת הפלט עושה שימוש בפונקציית Softmax על פני שני נוירונים, כדי לסווג את התמונה כחולה (PD) או בריאה (NONPD).

אחת המטרות המרכזיות הייתה לבדוק האם סיווג ברמת המטופל, כלומר הסתמכות על מספר תמונות ולא על תמונה בודדת, יכול להניב תוצאות מדויקות יותר. לשם כך, הוכנה אסטרטגיית קיבוץ לפי מטופל במהלך חלוקת הנתונים. נעשה ניסיון לשמור על הומוגניות בקבוצות האימון, האימות והבדיקה, כך שכל קבוצה תכיל תמונות שונות ממטופלים שונים כדי למנוע דליפת נתונים בין סטים ולבחון את יכולתו של המודל להכליל.

כדי להתמודד עם כמות הדאטה המוגבלת, יושמה טכניקת אוגמנטציה אשר כללה סיבובים אקראיים, הזזות, היפוכים ושינויים קטנים נוספים בתמונות כדי להגדיל את הגיוון של הדאטה ולשפר את היכולת של המודל לזהות את הדפוסים המהותיים למחלת פרקינסון. בעקבות האוגמנטציות המורכבות המודל התקשה להתכנס בקבוצת האימון ולכן ביצענו מספר גדול של איפוקים – כ 1000 חזרות כדי לסייע להתכנסות המודל. עם זאת, כפי שהתברר בהמשך, גם שימוש בטכניקה זו לא הצליח לפצות לחלוטין על מחסור בנתונים אמיתיים.

נתונים וניסויים

הדאטה ששימש לאימון המודל כלל 1,408 תמונות MRI של מטופלים בריאים וחולים, אשר חולקו כך ש-64.69% מהתמונות שימשו לאימון, 16.77% לוולידציה, ו-17.76% נותרו כקבוצת בדיקה להערכת הביצועים הסופיים. כל תמונה שויכה למטופל מסוים, ומספר התמונות למטופל השתנה בהתאם לזמינות הדאטה.

המטרה הייתה לבחון את ההשערה כי מודל שיודע להבדיל בין מטופלים ולא רק בין פריימים בודדים, עשוי לשפר את זיהוי המחלה. על כן, ביצענו אימונים שונים תוך שימוש בקבוצות שונות של תמונות לכל מטופל, וכן נבדקו מודלים שהסתמכו על ריבוי תמונות לכל מטופל, לעומת סיווג תמונה בודדת.

תוצאות

אחד הממצאים המרכזיים מהאימון היה שבעוד שהמודל הצליח להגיע לרמת דיוק גבוהה בסט האימון, הביצועים על סט הולידציה היו נמוכים משמעותית, מה שמצביע על בעיית Overfitting חמורה. הגרפים מצביעים בבירור על כך שאחרי האיפוק התשיעי, המודל לא הצליח להתכנס, ואיבוד הוולידציה רק הלך וגדל עם המשך האימון.

עוד ממצא בולט הוא שהמודל לא הצליח להכליל היטב את דפוסי המחלה לסט הולידציה, מה שהוביל לחוסר יציבות בתוצאות, עם תנודות משמעותיות בדיוק ובמדדי הביצועים לאורך כל האפוכים. אף על פי שהמודל הצליח ללמוד דפוסים בתמונות ה-MRI, נראה כי היקף הדאטה לא הספיק על מנת לספק תוצאות מיטביות, מה שתמך בהשערה הראשונית כי הנתונים הקיימים אינם מספיקים כדי לאפשר למודל לזהות דפוסים באופן כללי על פני מטופלים חדשים.

למרות בעיית ההתכנסות, נמצא כי בחלק מהאימונים הדיוק של המודל הגיע עד ~62.6% על סט הולידציה, מה שמראה כי יש למודל פוטנציאל, אך נדרשים שיפורים משמעותיים באסטרטגיית האימון. ערך ה-AUC עמד בממוצע על 0.70, מה שמעיד על כך שהמודל הצליח לזהות הבדלים בין הקבוצות, אך לא ברמה מספקת כדי לאפשר שימוש קליני אמין.

עוד נקודה חשובה שעלתה בניסויים היא עלות החישוב הגבוהה של המודל. מאחר שמדובר במודל קונבולוציוני כבד יחסית, שנדרש לעבד מספר רב של תמונות עבור כל מטופל, זמן האימון וההרצה היה ארוך מאוד והפך את השימוש במודל לפחות יעיל בהיבטים של יישום מעשי.

מסקנות והמלצות להמשך

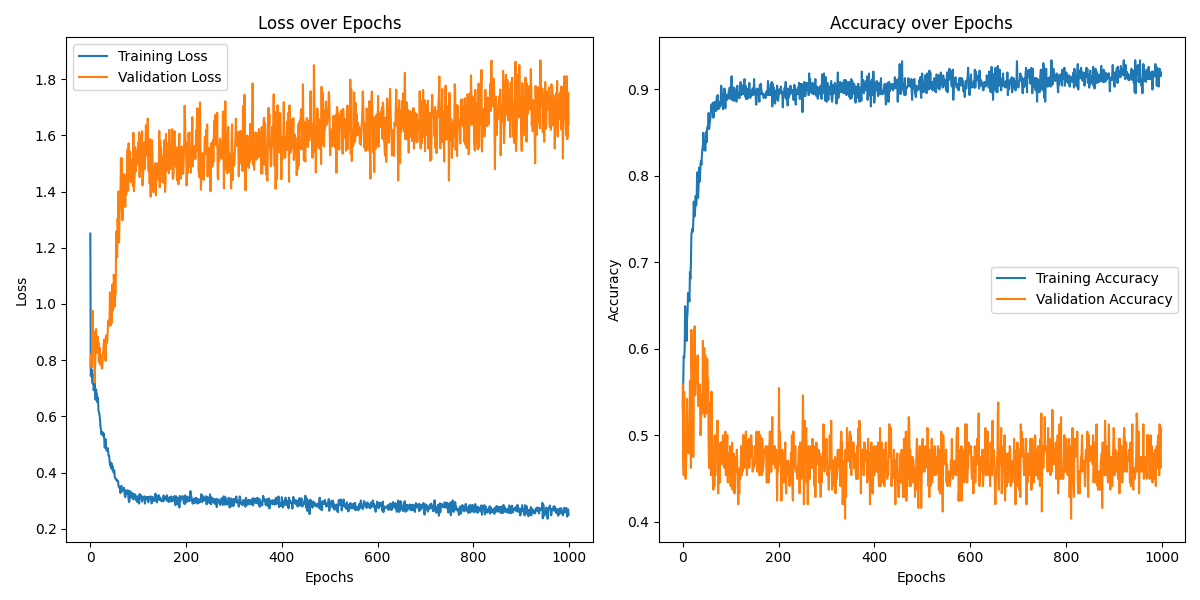
תוצאות המודל מראות כי יש פוטנציאל לסיווג מחלת פרקינסון באמצעות דימות MRI, אך מספר אתגרים משמעותיים דורשים טיפול.

ראשית, מחסור בנתונים היווה גורם משמעותי בתוצאות המוגבלות של המודל. בהתחשב בכך שהשיטה שנבנתה מתבססת על מידע מחולה שלם ולא רק על תמונה אחת, יש צורך בסט דאטה רחב יותר שיכלול מספר גדול יותר של תמונות לכל מטופל. ניתן לשפר זאת על ידי איסוף נתונים נוספים או שימוש בשיטות מתקדמות כמו Transfer Learning עם מודלים מאומנים מראש (ResNet, VGG, EfficientNet).

שנית, בעיית ה-Overfitting מצביעה על כך שנדרשות טכניקות נוספות לשיפור היכולת של המודל להכליל. ייתכן כי שימוש בגישות כגון Regularization משופר, Dropout גבוה יותר, או שינוי בטכניקות הקונבולוציה יכול לסייע בהתמודדות עם הבעיה. כמו כן, ניתן לנסות גישה מבוססת RNN או LSTM כדי לנצל את המידע הרציף על פני סדרת התמונות של כל מטופל.

לבסוף, יש לתת מענה לבעיית עלות החישוב הגבוהה של המודל. מאחר שזמן האימון היה ממושך, נדרש למצוא דרכים לייעל את חישוב המודל, בין אם על ידי שימוש בארכיטקטורה קלה יותר ובין אם על ידי הפחתת כמות התמונות לכל מטופל תוך שימור המידע המשמעותי ביותר.

לסיכום, המחקר מספק תובנות חשובות לגבי האתגרים והאפשרויות בזיהוי מוקדם של פרקינסון באמצעות AI ודימות MRI, ומצביע על הצורך בשיפור הדאטה והמודלים כדי להגיע לדיוק ברמה קלינית מספקת.



**מקורות:**

* + - 1. Singh, M., & Renji, N. M. (2021). A comparative study of different machine learning algorithms in detection of Parkinson’s disease. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 9*(1), 755–762. [https://doi.org/10.22214/ijraset.2021.32907](https://doi.org/10.22214/ijraset.2021.32907" \t "_new)
      2. Shaban, M. (2023). Deep learning for Parkinson’s disease diagnosis: A short survey. *Computers, 12*(3), 58. <https://doi.org/10.3390/computers12030058>
      3. Gopalakrishnan, B., & Stoessl, A. J. (2011). Dopaminergic imaging in Parkinson’s disease: PET. In *Oxford Scholarship Online* (pp. 3–10). Oxford University Press. [https://doi.org/10.1093/acprof:oso/9780195393484.003.0011](https://doi.org/10.1093/acprof:oso/9780195393484.003.0011" \t "_new)
      4. Eidelberg, David, *Imaging in Parkinson’s Disease* (2011; online edn, Oxford Academic, 22 Sept. 2011), <https://doi.org/10.1093/acprof:oso/9780195393484.001.0001>, accessed 21 Feb. 2025.
      5. Choi, H., Youm, C., Park, H., Kim, B., Hwang, J., Cheon, S.-M., & Shin, S. (2024). Convolutional neural network-based detection of early-stage Parkinson’s disease using the six-minute walk test. *Scientific Reports, 14*, Article 22648. <https://doi.org/10.1038/s41598-024-72648-w>
      6. K. Mittal, K. Singh Gill, R. Chauhan, H. Singh Pokhariya and J. Lande, "Artificial Intelligence Assisted Classifier’s and Neural Network Based Prediction and Classification of Parkinson’s Disease," *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, 2024, pp. 1-6, doi: [10.1109/ICIPTM59628.2024.10563233](https://ieeexplore.ieee.org/document/10563233).
      7. <https://www.ppmi-info.org/>
      8. Tagaris, A., Kollias, D., Stafylopatis, A., Tagaris, G., & Kollias, S. (2018). Machine learning for neurodegenerative disorder diagnosis—Survey of practices and launch of benchmark dataset. *International Journal on Artificial Intelligence Tools, 27*(2), 1850011. [https://doi.org/10.1142/S0218213018500112](https://doi.org/10.1142/S0218213018500112" \t "_new)  
         GitHub link: <https://github.com/ails-lab/ntua-parkinson-dataset>

**מקורות נוספים עליהם התבססה עבודתינו המקצועית:**

* + - 1. Neuroimaging in Parkinson's Disease and Related Disorders – chapters: 3, 4, 7, 11  
         link: [לחץ](https://www.sciencedirect.com/book/9780128216514/neuroimaging-in-parkinsons-disease-and-related-disorders)