

## פרוייקט סיום קורס - יסודות למידה עמוקה - תשפ"ד

מטרה: סיווג תמונות של צילומי רנטגן חזה (Chest X-ray)

הדטהסט לאימון:

<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>

בדטהסט ישנן 5863 צילומי רנטגן המחולקות לשתי קטגוריות מרכזיות:

א. בריא

ב. עם דלקת ריאות (pneumonia) - מחולק גם הוא לשתי תתי קטגוריות:

(1 דלקת ריאות חיידקית bacterial

(2 דלקת ריאות ויראלית viral

(את חלוקת תתי הקטגוריות ניתן להסיק מתוך שם התמונה)

הדטהסט מחולק לקבוצת אימון/תיקוף/בחינה.

את קבוצת הבחינה עליכם להשאיר כפי שהיא.

מותר לכם לאחד את קבוצת התיקוף והאימון ולבצע חלוקה שונה ביניהם.

### משימות הפרוייקט

א. עליכם לפתור את הבעיה באמצעות רשת נוירונים עמוקה:

(1 סיווג בריא/חולה

(2 סיווג בריא/דלקת ריאות ויראלית/דלקת ריאות חיידקית (רשת עם שלוש יציאות.

לא בצורה היררכית)

ב. לאחר שבחרתם רשת מוצלחת בסעיפים הקודמים, הציגו כיצד ניתן לסווג תמונה

חדשה ע"י שימוש בווקטור Embedding שנוצר מרשת הסיווג ושימוש ב-KNN.

הציגו ויזואליזציה של המחלקות השונות באמצעות T-sne (ניתן להשתמש בספריה

t-sne של מוכנה

ג. כעת יש לפתור את אותה הבעיה כאשר החליפו את קבוצת הבחינה והאימון/תיקוף כלומר בעיה של מיעוט דוגמאות לאימון (כלומר לטובת האימון יש לנו את קבוצת הבחינה נקרא לה כעת קבוצת האימון החדשה)

ד. בסעיף זה יש לנו רק את דטה של תמונות "בריא". עליכם לנסות לזהות "חולים" באמצעות שיטות של Anomaly Detection

זיכרו שהפרוייקט נועד לשקף את הידע שצברתם במהלך הקורס. לפתרון הסעיפים השונים עליכם להשתמש ולשקול שיקולים שדנו בהם במהלך הקורס כדוגמת מקדם למידה, אופטימיזרים, Transfer learning, ארכיטקטורות רשת שונות, פונקציות לוס ועוד. אין להשתמש בכלים שלא נלמדו בקורס

## הגשה:

הגשה בזוגות

תאריך הגשה: 14/02/2024

יש להגיש מחברת קולב + דוח מסודר הכולל את השיקולים השונים/גרפים/ביצועים ועוד

בדוח עצמו, מאוד חשוב לראות את ההבנה שלכם והיכולת שלכם הן בבניין מערכת למידה עמוקה על ההיבטים השונים והן בדיווח תוצאות מול הניסויים השונים

המלצה: ע"מ להיות יעילים יותר מבחינת מאבי חישוב, מומלץ להקים סביבת ניסויים לכל מטלה ולהריץ אותה על כמות מצומצמת של דטה. רק לאחר שתראו שהדברים "מתנהגים" כפי שציפיתם, הרחיבו לסט האימון הגדול יותר.

מחברת ה- Colab Google תהיה מחולקת בצורה מסודרת, תכיל תאי קוד נפרדים ותאי טקסט המסבירים על הפעולות שנעשו. מחברת עם תאי טקסט מפורטים לא מחליפים את הדוח ולהפך.

בעת ההגשה, המחברת תכיל את כל הפלטים הרלוונטיים לתוצאות האימון לאחר תאריך ההגשה, אין לגעת במחברת או להריץ שום תא קוד. פעולה כזאת תגרור הורדת ניקוד משמעותית\*\*

המחברת תכיל את שמות הסטודנטים המגישים ואת מספרי תעודת הזהות שלהם  
בתא טקסט שימוקם בחלק העליון של המחברת.

הוראות טכניות: את רשת הנוירונים יש לבנות על ידי שימוש בספרייה של Keras  
בלבד, עבור שאר הפרויקט ניתן להשתמש בספריות עזר. וכו', NumPy, Pandas,  
Seaborn, scikit-learn, Matplotlib  
שמרו את המשקולות המאומנות של ה-"ארכיטקטורה המנצחת" ותבנו סביבת מבחן  
סביבת מבחן אחת לכל סעיף (. סביבת המבחן תאפשר את טעינת המשקולות/המודל  
המאומן ותיתן אפשרות למשתמש להעלות למחברת קלט של חדשה וסיווגה (לפי  
הסעיפים השונים).  
יש לצרף הנחיות מדוייקות כיצד להריץ את המחברות אותן צירפתם.

## בהצלחה