

## מערכות למדות ולמידה عمוקה – סמסטר ב – תשפ"ה – פרויקט סיום

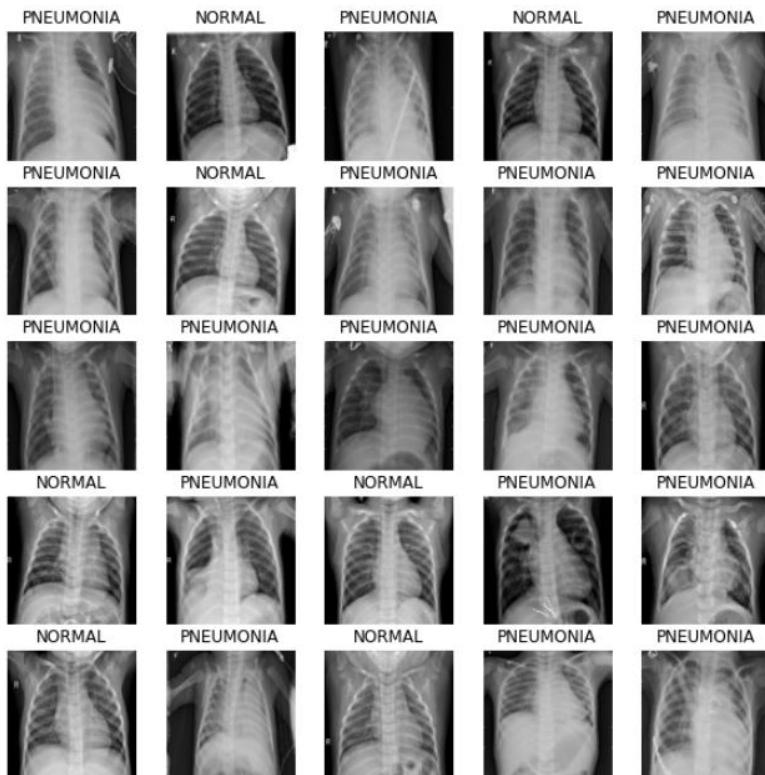
הנדסת חשמל ואלקטרוניקה – פרופ' אמיר אדר

הנחיות כלויות

- 1) הפרויקט יבוצע בקבוצות של 4-3 סטודנטים. ההגנות יתקיימו בסוף תקופת הבוחינות.
- 2) כל קבוצה תציג דוח מסכם בפורמט PDF ובנוסף קבצי קוד ב- PYTHON בהתאם למשימות השונות. יש לכלול בדוח המסכם אך ורק תשובות בהתאם למשימות. אין לכלול בדוח המסכם הסברים ו/או סיכומים של חומר הלימוד (ירדו נקודות על דוח שיבולו סיכומים של החומר).
- 3) בהגנה על כל סטודנט/ית לדעת להסביר את הפתרונות בכל המטלות, וכן להזכיר את מבנה כל רכיביו הרשותות העמוקות שהופיעו בפרויקט, וכן את אלגוריתמי אימון הרשות שביהם עסקנו בקורס.
- 4) משקל מרכיבי הפרויקט בציון הסופי בקורס: 55% (דו"ח הפרויקט) + 35% (תשובות אישיות לשאלות בע"פ בהגנה).

### משימות הפרויקט: גליי דלקת ריאות בצלומי רנטגן

בפרויקט ימומשו פתרונות מבוססי CNN לגליי דלקת ריאות מ- 5863 צילומים אמיתיים הכוללים אבחון של רופאים. התמונות נמצאות בקישור הבא: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>. להלן מספר דוגמאות משלטי הקטגוריות:



בנוסף, קיימת הפרדה לשני סוגי דלקת ריאות: חיידקית (Viral) ונגיפית (Bacterial) , להלן דוגמאות:



מטרת הפרויקט היא תכנון רשתות למידה عمוקה לגילוי דלקת ריאות, אשר הביצועים ימדו על סט תמונות בדיקה, אשר יופרד באופן מוחלט מסט האימון. ביצועים טובים יחויבו לסיכון שגיאה נמוך מ-7% (בלומר Accuracy גבוהה מ-93%). יש להשתמש בסט תמונות בדיקה שככל לפחות 200 תמונות עם דלקת ריאות (100 מכל סוג) ו-200 תמונות ללא דלקת ריאות. יש ליצור סט אימון ובדיקה חדשים מתוך כלל התמונות הקיימות באתר, ולא לפי החלוקה הקיימת באתר. טוים חדשניים אלה יושמו את כל סעיפים הפרויקט.

#### משימה 1

עליכם לתכנן שתי רשתותعمוקות שונות לפתרון הבעיה: כניסה כל רשת عمוקה היא תמונה והמוצא הינו הסתברות שהתמונה מייצגת מקרה חיובי של דלקת ריאות (בסעיף זה אין זה משנה מה סוג הדלקת-חיידקית או נגיפית). ההסתברות היא בין 0 ל-1 כאשר 1 מייצג מקרה וודאי של דלקת ריאות.

רשת ראשונה: רשת CNN ללא שימוש ב- TRANSFER LEARNING .

רשת שנייה: רשת CNN מבוססת TRANSFER LEARNING .

モותר להיעזר בהצעות לפתרונות המופיעים בקישור הבא (מדובר על הצעות של אנשים פרטיים ולא פתרון "רשמי"):

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

יש לפרט בדוח הפרויקט את מבנה שתי הרשתות שתכונთם, כולל כל השכבות. לגבי הרשת מבוססת TRANSFER LEARNING יש לפרט את שם ומבנה רשת הבסיס (שכבר אומנה בעבר), איזה שכבות נלקחו ממנה, ואיזה שכבות הוספთם להשלמת הפתרון הנדרש.

#### משימה 2

1. יש לאמן את שתי הרשתות, ולציין בדוח את אלגוריתם האימון, מספר ה- EPOCHS , גודל MINI-BATCH . לגבי הרשת מבוססת TRANSFER LEARNING , יש להציג שני אופני אימון: א) פרמטרי רשת הבסיס מוקפאים ורק השכבות שהוספთ מटעדכנות באימון. ב) FINE-TUNING בו מעודכנים גם

- חלק – לבחירתכם – של שכבות רשת הבסיס. איזה מהשיטות הביאה לbijoux טוביים יותר מבחינה?  
ACCURACY המשיכו בסעיפים הבאים רק עם השיטה הטובה מבין השתיים.
2. עבור שתי הרשותות שתכוננתם ציירו את גраф הביצועים , RECALL - PRECISION כאשר כל נקודה בגרף תוחשב עבור רמת סף שונה (ביחס להסתברות שטפיקה הרשות) להחלטה על דוגמא חיובית (כלומר עם דלקת ריאות). הסך יהיה בטוחה 0.1 עד 0.9 בקפיצות של 0.05 . סמןו גם על הграф את נקודות SCORE - F שיחסבו מכל זוג ערכים של PRECISION ו- RECALL .
  3. עבור איזה סף התקבל ערך F - SCORE הגבוה ביותר?

הסבר: שימושו לב שהחלה האם יש או אין בתמונה דלקת ריאות, מתקבלת ע"י הפעלת סף על ההסתברות שהרשות מוציאה (זכרו שהרשות מפיקה הסתברות רציפה בין 0 ל-1). מזאת הרשות מתקבל באמצעות הפונקציה predict(). במשימה 2 יש לשנות את הסף בתחום 0.1 עד 0.9 (בקפיצות של 0.05) ולהשאיב את ההחלטות בהתאם לסף: אם ההסתברות במוצא הרשות גבוהה מהסף אז ההחלטה חיובית אחרת ההחלטה שלילית. לדוגמא: נניח שהסף הוא 0.3 והרשות מוציאה הסתברות 0.4, אז יהיה ש:  $0.3 < 0.4$  ההחלטה היא שיש דלקת ריאות. לכל ערך סף תקבלו ביצועים שונים של הרשות, אשר באמצעותם תיחסו צמד אחד של ערכי SCORE. כל נקודה בgraf RECALL-PRECISION תתקבל מערך שונה של הסף.

### משימה 3

1. בדקו את ביצועי הרשותות עם אלגוריתמי האימון הבאים (בדקו השפעת מספר ה EPOCHS ו- LEARNING RATE לכל אלגוריתם):

  - א. אלגוריתם SGD
  - ב. אלגוריתם SGD עם MOMENTUM.
  - ג. אלגוריתם ADAM
  - ד. אלגוריתם RMSPROP

- ה. בחרו את האלגוריתם שהביא לתוצאות הטובות ביותר בסעיפים א-ד והפעילו מנגנון EARLY STOPPING בהתאם לאלגוריתם שנלמד בהרצאה (קובץ שקפים בנושא רגוליזציה). האם הושג שיפור ביצועים באמצעות מנגנון זה?

במשימה זאת יש להציג את גраф התוכניות של תהליכי האימון לכל אחד מהסעיפים כולל LOSS TRAIN ו- VALIDATION LOSS, וכן TRAIN ACCURACY ו- VALIDATION ACCURACY. זאת באשר נעשה שימוש בסט ולידציה שכולל לפחות 50 תמונות עם דלקת ריאות (25 מכל סוג) ו- 50 תמונות ללא דלקת ריאות. סט הולידייז ילקח מהתוך סט האימון (בר שיוו קבוצות זרות של תמונות).

### משימה 4

- ממשו באמצעות שינוי הרשות לא TRANSFER LEARNING (משימה 1) פתרון לסיווג 3 קטגוריות: (1) אין דלקת (2) יש דלקת חידקית (3) יש דלקת נגיפית. שימוש לב שיש להשתמש במקרה זה ב- SOFTMAX במודול CONFUSION MATRIX על סט הבדיקה.



בצלחה!