

פרויקט עיבוד תמונה רפואי

סיווג תמונות הרשתית שנלקחו בשיטת optical coherence tomography לזיהוי מחלות.

קישור לגיט שמכיל את הקוד:

https://github.com/LitalBarak96/image_proc

קישור למקור בו מופיע הדאטה:

<https://data.mendeley.com/datasets/rscbjbr9sj/2>

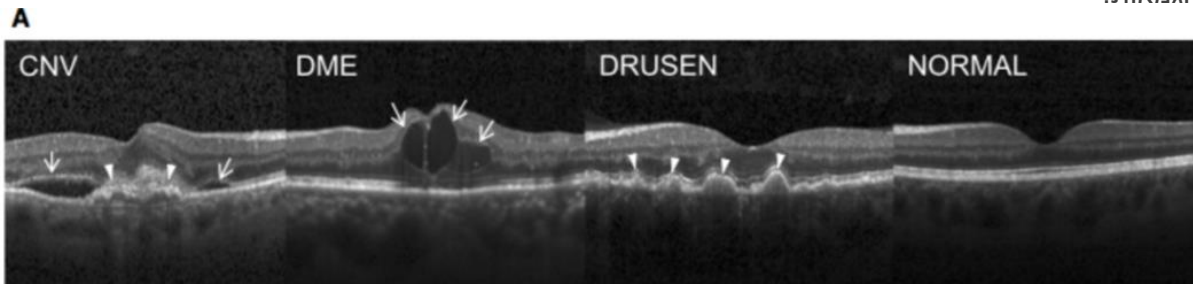
מגשים: אבישי ויזל וליטל ברק

מבוא

הרשתית (RETINA)) היא שכבת תאים דקה באחורי גלגל העין. זהו החלק בעין שממיר אור לאותות עצביים.

בדיקת Optical Coherence Tomography (ובקיצור: OCT) היא צילום לא פולשני של העין באמצעות קרני אור היוצר תמונות חתך רוחביות של חלקי העין השונים כמו הרשתית ועצב הראייה. בדיקה זו מאפשרת לרופאי עיניים מעקב אחרי הרקמות שאינן מתאפשר בבדיקות קליניות ומסייע לאבחון מחלות עיניים שונות. בפרויקט שלנו קיבלנו סריקות OCT ואנו מעונינים למצוא אלגוריתם שקובע האם קיימת מחלת רשתית, ואם כן איזו מחלה, מתוך 3 מחלות אפשריות. המחלות הם:

1. **CNV** - צמיחת כלי דם פתולוגיים ברשתית במטופלים הסובלים מקוצר ראייה חמור. ניתן לראות בחצים כמות נוזלים נכבדת.
2. **DME** - בצקת מקולרית על רקע סוכרת. מחלת רשתית זו נגרמת בשל זרימת דם מוגברת בכלי הדם ברשתית, עד כדי פגיעה בדופנותיהם. ניתן לראות בחצים התעבות נוזלים.
3. **DRUSEN** - משקעים דמויי חלבון המצטברים במקולה (מרכז רשתית העין) ועשויים לגרום לאיבוד חדות הראייה ולבעיות עיניים נוספות. החצים מייצגים כמה דורזנים.
4. **NORMAL** - רטינה תקינה. דוגמאות:



במאגר קיימות 84495 תמונות בפורמט JPEG שמחולקות ל test train ו-validation.

בכל שנה מבוצעות 30 מיליון סריקות OCT והאנליזה והפענוח של כל אחת מהן לוקחת זמן משמעותי לכן יש חשיבות לכלי אשר יוכל לסווג מחלות באופן לתת אבחון מדויק ולטיפול אפקטיבי לדוגמה - מטופלים עם DME או CNV יצטרכו התערבות רפואית מידית לעומת מטופלים עם DRUSEN יצטרכו מעקב.

תיאור מסד הנתונים:

כמות התמונות ב- train היא 84,495 תמונות, כאשר החלוקה היא:

0.59 CNV
0.18 DME
0.14 DRUSEN
0.42 NORMAL

זוהי חלוקה לא שווה שעלולה להשפיע על התוצאות אך בחרנו לא לשנות את היחסים. השתמשנו בחבילות Scikit learn לצורך מימוש אלגוריתם ה-KNN, ו-Pytorch ו-torchvision למימוש רשת ה-CNN.

גודל התמונות:

אחת הבעיות שנתקלנו בה היא שהתמונות בדאטה אינם בגדלים שווים. לכן, עבור אלגוריתם ה-KNN השתמשנו בפונקצית `resize` מחבילת `cv2`, כדי לשנות את התמונות לגדלים שווים. בנוסף, השתמשנו בפונקציה `CenterCrop` מחבילת `torchvision` עבור רשת ה-CNN. האלגוריתם חותך את התמונות באמצע בהתאם לגודל רצוי (יכולנו להשתמש ב-PAD בשביל להגדיל את התמונות הקטנות אך בחרנו ב-Crop שעלול להוביל לאיבוד מידע אך בהמשך גילינו שזה מספק מבחינת דיוק).

טעינת הדאטה:

עבור רשת ה-CNN השתמשנו ב-`Tensor` כדי להפוך את התמונות ל-`Tensor`, בשביל שהחבילה תוכל להשתמש בה (זהו חלק מהדרישות שלה) בנוסף השתמשנו גם ב-`datasets.ImageFolder` לצורך יצירת דאטה לודר גנרי על מנת שנוכל להשתמש בתמונות בצורה קלה יותר. ב-KNN הפכנו כל תמונה לוקטור, בעזרת פונקציה `flatten()` של חבילת `cv2`.

עבודה קשורה

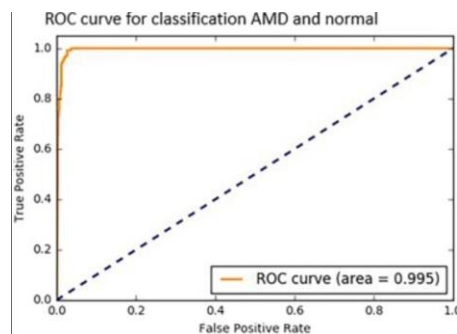
מצאנו עבודות דומות שמנסות לאבחן מחלות הקשורות לרישתית, אחת מהן מנסה לאבחן AMD (ניוון מוקולרי) והשניה מנסה לאבחן `central serous chorioretinopathy` (מרכזית) בעזרת הדמיות OCT.

בשני המודלים נעשה שימוש ב-CNN בארכיטקטורות שונות. בנוסף ניתן היה לראות שבשניהם נעשה שימוש בדאטה סט קטן בהרבה מהדאטה סט שקיים לנו. באחד המודלים נעשה שימוש ב-16-VGG ו-50-Resnet כלומר הרשת שלהם כללה

13 CNN layers with ReLU activation functions, 4 max pooling layers, 2 dropout layers, and 4 FC layers

מאחר ובמאמרים הללו נעשה שימוש בכמות נמוכה בהרבה של תמונות לא יכולנו לממש באופן ישיר את הרשתות שהוצעו במאמר לכן לקחנו דברים נבחרים שחשבנו שישפיעו: פונקציית האקטיבציה `Relu` שכבות `maxpool` ושכבות FC.

היה מעניין לראות שבאחד המאמרים קיבלו ROC 0.995% של כלומר הם קיבלו הרבה מאוד ערכים של TRUE POSITIVE מה שאנחנו מעוניינים לקבל לצורך אבחון נכון.



שיטות

ביצענו עבור הלמידה הלא עמוקה את אלגוריתם KNN (k-nearest neighbors). בדקנו עבור K שונים: בין 1 ל-9 עם ערכים אי-זוגיים 1 3 5 7 9 (כידוע מספרים אי-זוגיים ב-KNN מביאים לתוצאות החלטיות יותר) ובחרנו למדוד את ההצלחה על פי כמות הפעמים בהם יש התאמה (accuracy).

האלגוריתם מקבל תמונה, משנה אותה לגודל של 32*32 פיקסלים, והופך אותה לוקטור באמצעות פונקציית `flatten`. לאחר מכן אנו משתמשים בחבילת `sklearn` בשביל להפעיל את האלגוריתם:

בהתחלה מקודדים את הלייבלים ללייבלים מספריים, לאחר מכן מאמנים על סט האימון שנתון לנו, ולבסוף בודקים על סט ה-test.

עבור $K=1$ קיבלנו אחוז הצלחה של 89.97
עבור $K=3$ קיבלנו אחוז הצלחה של 76.34
עבור $K=5$ קיבלנו אחוז הצלחה של 69.93
עבור $K=7$ קיבלנו אחוז הצלחה של 66.52
עבור $K=9$ קיבלנו אחוז הצלחה של 64.04



לכן ה K הנבחר הוא $K=1$, כלומר נבחר על פי השכן היחיד שלו.

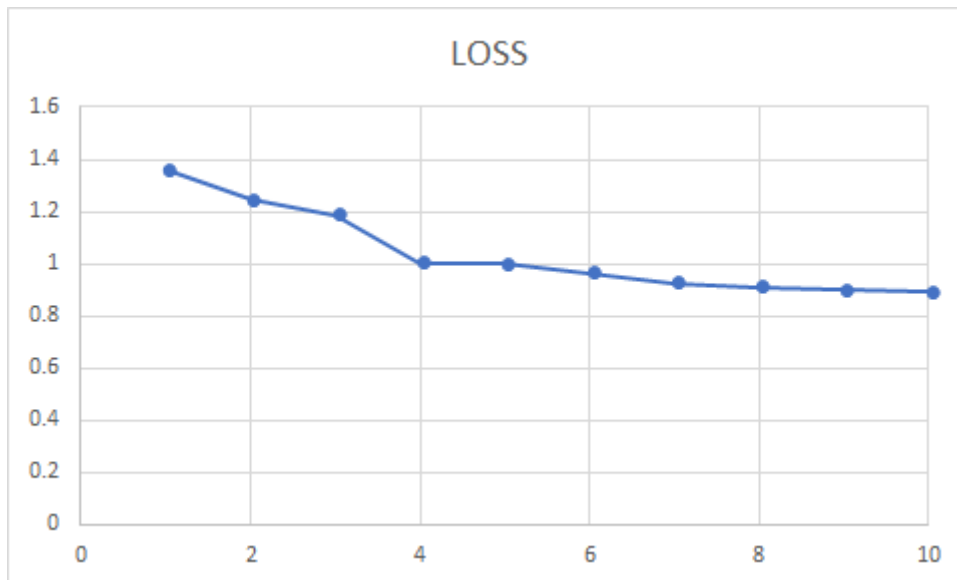
```
accuracy of 1 = 0.8997933884297521
accuracy of 3 = 0.7634297520661157
accuracy of 5 = 0.6993801652892562
accuracy of 7 = 0.6652892561983471
accuracy of 9 = 0.640495867768595
```

עבור הלמידה העמוקה בחרנו ברשת CNN עם 4 שכבות שכל אחת מהן מכילה D2 קונבולוציה, Maxpool, batch normalization, ופונקציית אקטיבציה RELU. לאחר מכן ישנם 3 שכבות fully connected ליניאריות, ולבסוף פונקציית log_softmax.

פונקציית המחיר (Loss) היא negative log-likelihood (NLL), האופטימיזר הוא SGD (Stochastic gradient descent) קצב למידה: 0.05
10 אפוקים

ביצענו כיוון של היפר פרמטרים ע"י הרצה של מספר פעמים, ובחירת התוצאה שמביאה לאחוז דיוק הגבוה ביותר.

הכיוון של היפר פרמטרים היה באופן בסיסי, מכיון שכל הרצה לקחה מספר שעות. בתחילה, התלבטנו האם להוריד את כמות התמונות שב-train, אבל לאחר מכן ראינו שאנחנו לא מקבלים תוצאה באופן עקבי ולכן השארנו את ה-dataset כמו שהוא.
גרף הירידה ב-LOSS:



ניתוח ואנליזה

לכל שיטה יש את ההבטים שלה: באלגוריתם ה-KNN הבדיקה היא פשוטה יותר, ומודדת מרחק אוקלידי בין שני וקטורים. לעומת זאת, רשת ה-CNN לומדת את התמונות בצורה עמוקה יותר, ע"י כך שהיא לומדת לזהות אזורים ספציפים בתמונה. ממבט מקדים נראה שאלגוריתם ה-CNN יהיה מוצלח יותר למשימה שלנו- זיהוי שוני באזור מסוים בתמונה.

השוני השני הוא במציאת הפרמטרים: ב KNN לא היינו צריכים להריץ את האלגוריתם מספר רב פעמים באפוקים שונים כדי למצוא את ההיפר פרמטרים, זאת מכיון של-KNN יש היפר פרמטר בודד (K). לעומת זאת, ב CNN יש מספר רב של היפר פרמטרים. לדוגמה- מספר השכבות, פונקציית האקטיבציה, קבוע הלמידה ועוד. עקב כך, היינו צריכים לבצע מספר רב יותר של הרצות כדי למצוא את הפרמטרים שיהיו יעילים.

הבדל נוסף בהרצה הוא שרשת ה-CNN מצריכה לימוד ע"י מספר מעברים על כל ה dataset (אפוקים), בעוד שאלגוריתם ה-KNN הוא אפוק יחיד. בגלל סיבות אלו, הרצת ה-KNN הייתה קלה יותר, והתקשנו למצוא את ההיפר פרמטרים המתאימים.

תוצאות המודלים:

KNN:

confiusion matrix: (rows: true labels. columns: predicted labels)

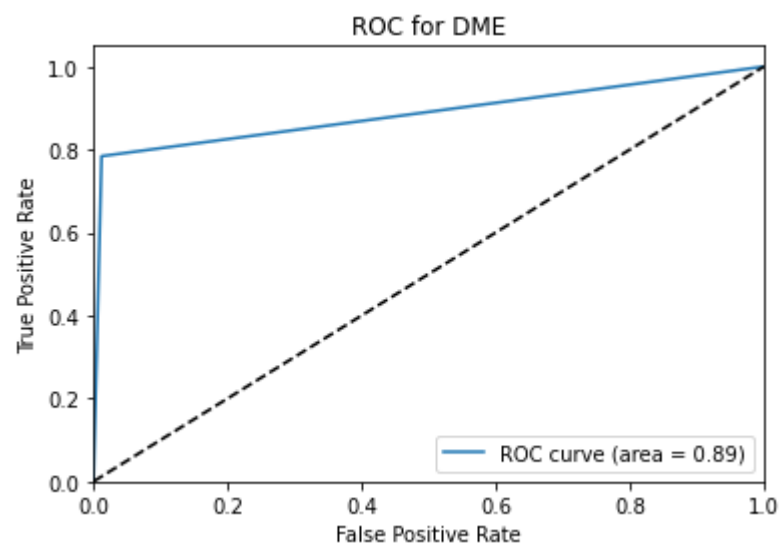
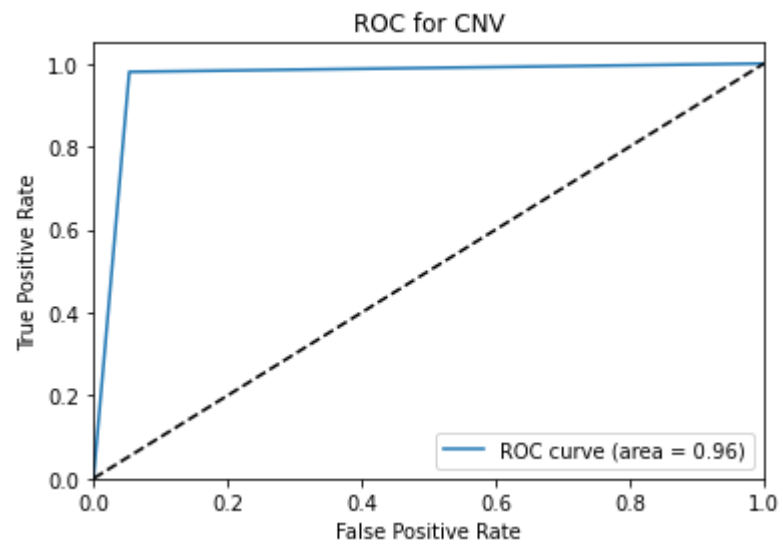
[0: CNV, 1: DME, 2: DRUSEN, 3: NORMAL]

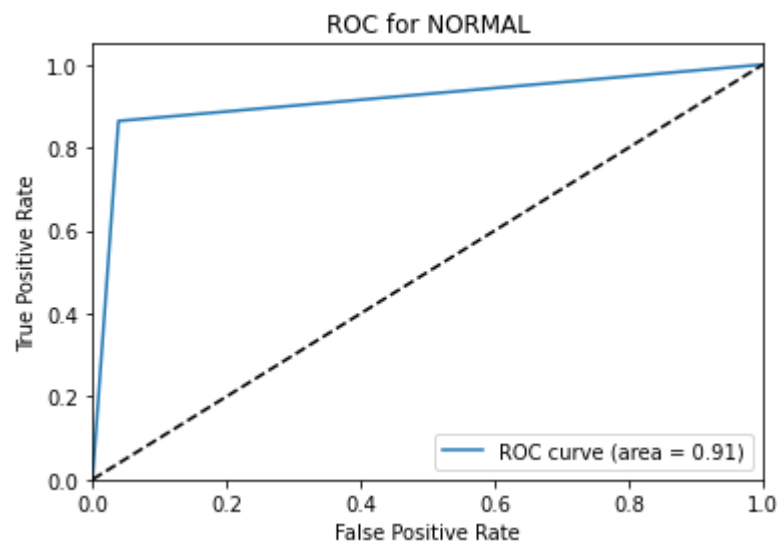
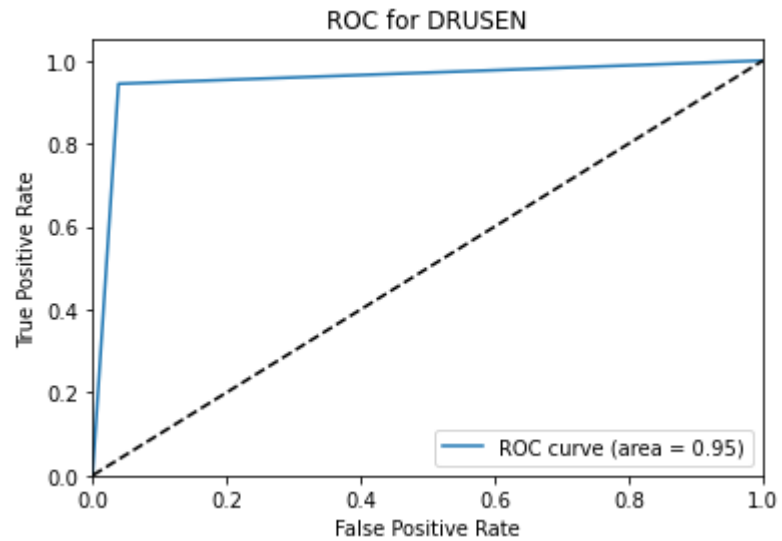
	0	1	2	3
0	245	1	2	2
1	22	196	9	23
2	10	0	236	4
3	8	8	18	216

f1 score:

`f1_score(testY ,y_pred, average = 'weighted') = 0.891`

ROC and AUC:





CNN:

confiusion matrix for CNN (rows: true labels. columns: predicted labels)

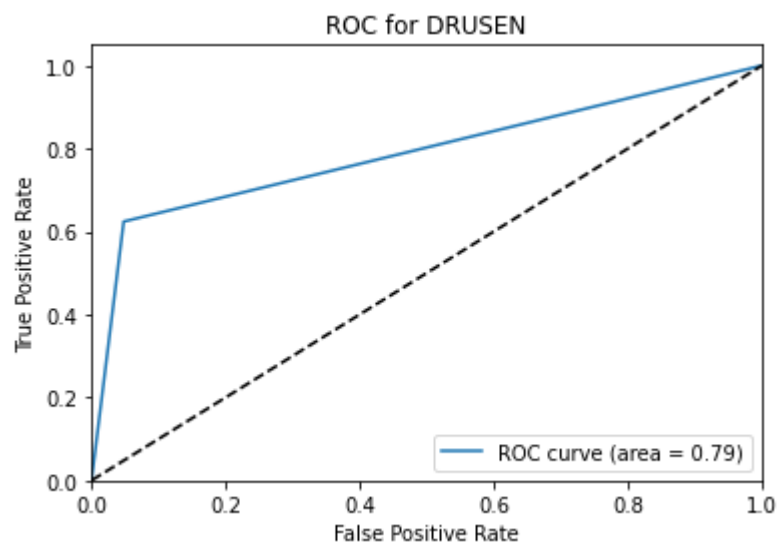
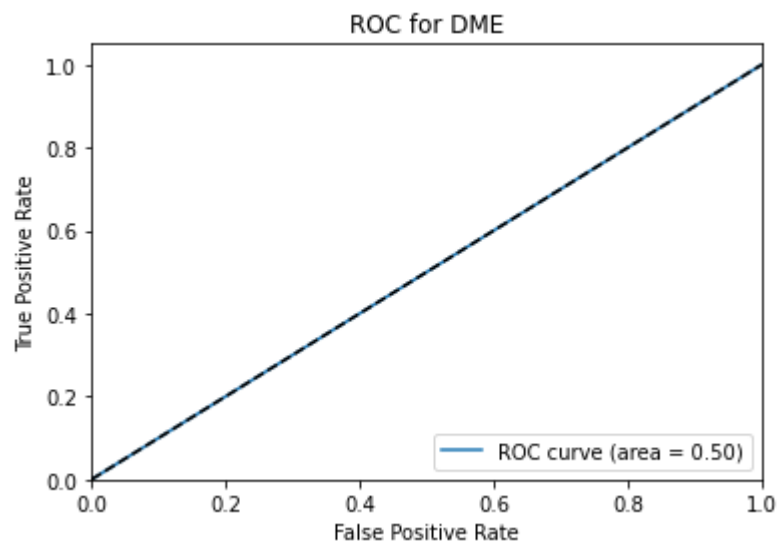
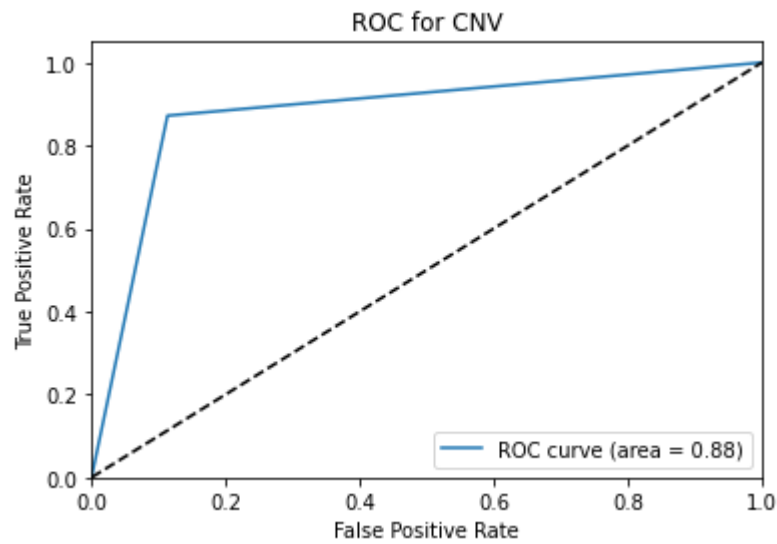
[0: CNV, 1: DME, 2: DRUSEN, 3: NORMAL]

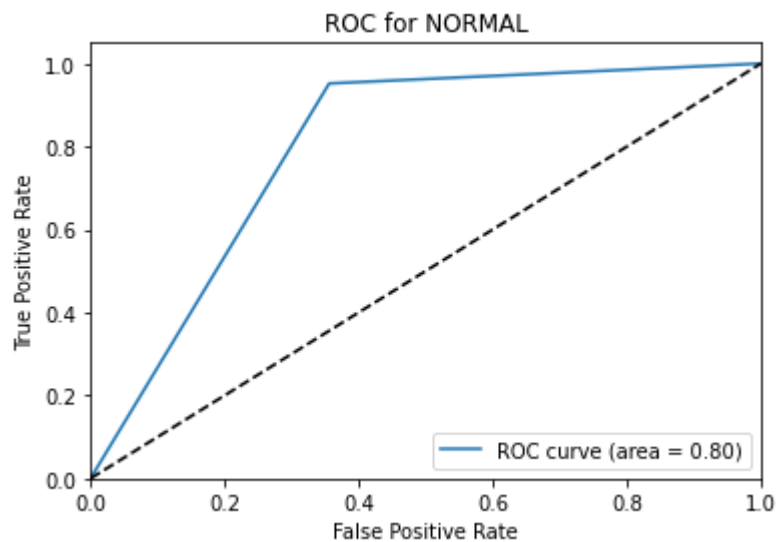
	0	1	2	3
0	218	0	13	19
1	45	0	20	185
2	31	0	156	63
3	9	0	3	238

f1 score:

`f1_score(dataset_test.targets,y_pred,average = 'weighted') = 0.531`

ROC and AUC:





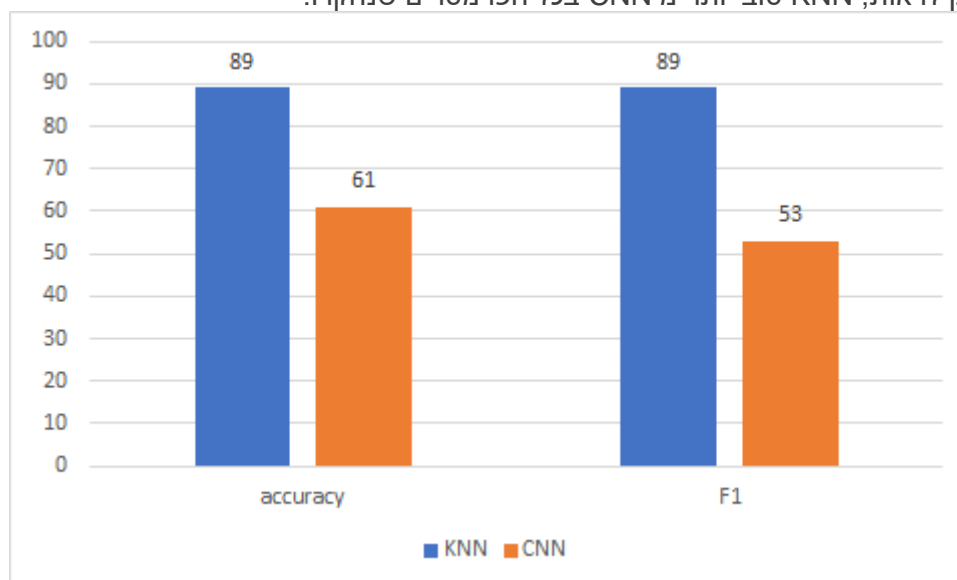
אחוזי דיוק:

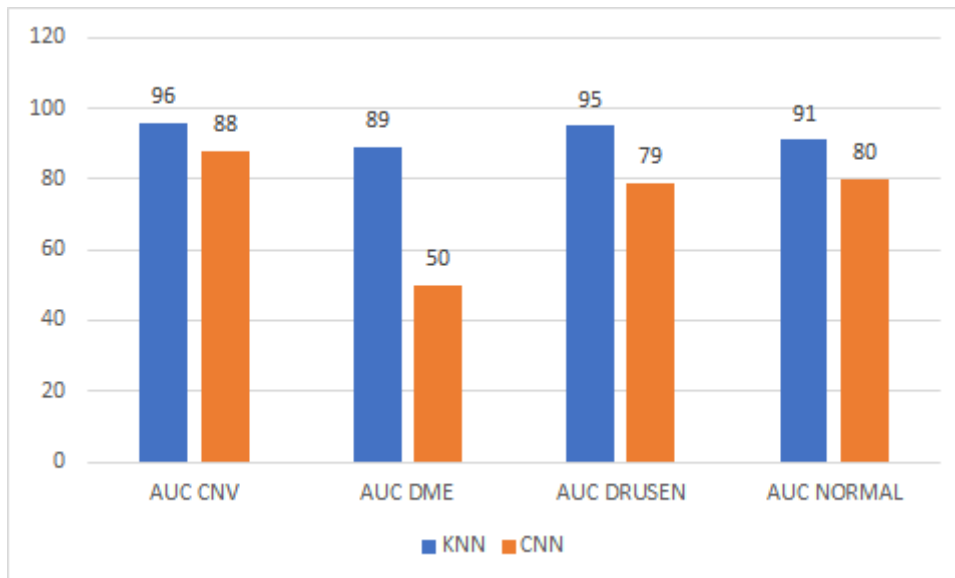
כפי שהראנו, אחוז הדיוק ב-KNN הוא 89%. לעומת זאת אחוז הדיוק ב-CNN הוא 61% בלבד:

```
model.score(testX, testY) = 0.89
```

Test validation set: Average loss: 0.8270, Accuracy: 612/1000 (61%)

כפי שניתן לראות, KNN טוב יותר מ-CNN בכל הפרמטרים שנחקרו:





הסיבות לכך יכולות להיות:

1. לא הצלחנו לכוון את ההיפר פרמטרים של ה-CNN, עקב זמן ההרצה הארוך. ככל הנראה יש היפר פרמטרים של הרשת שאיתם החיזוי היה הרבה יותר מדויק כפי שראינו בספרות.
2. חוסר האיזון בין הקלאסים משפיע לרעה על אלגוריתם ה-CNN: כפי שניתן לראות, ה-AUC על קלאס ה-DME הוא נמוך במיוחד. לעומת זאת, ה-AUC של CNV היה גבוה מאוד. יתכן שהדבר נובע מכך שבסט האימון יש ייצוג גדול יותר לקלאס ה-CNV.
3. בניגוד לכך, לקלאס ה-DRUSEN ישנו גם כן ייצוג נמוך, ובכל זאת ה-AUC שלו יחסית גבוה. הקלאס DME לא נחזה בכלל ב-CNN (כפי שניתן לראות ב-confiusion matrix). אפשר להסביר זאת עקב צורתו השונה של המחלה: שני המחלות האחרות מתבטאות ברשתית (הקו הלבן) בעוד ש-DME מתבטא באזור אחר. לדעתנו, כנראה שרשת ה-CNN לא מצליחה להבחין בשינויים אלו.

דיון

היינו מנסים להריץ מודלים נוספים, למשל לנסות לעשות AUTOENCODER ללמידה עמוקה או לחילופין להריץ CNN עם פרמטרים נוספים מאחר וכל ריצה לוקחת הרבה זמן גם עם GPU. היינו מנסים לבצע PAD לתמונות במקום CROP ובנוסף היינו מנסים לאזן את כמות התמונות בכל תיוג כך שנקבל כמות שווה של תמונות למניעת הטייה בלמידה כמו כן היינו מפחיתים את כמות התמונות לאימון מכיוון שראינו במאמרים שניתן לקבל תוצאות גם עם דאטה סט קטן יותר.

המסקנה שלנו שניתן לאבחן באופן יחסי את המחלות בעזרת KNN עם אחוז TRUE POSITIVE גבוה. אך עם זאת יש בעייתיות בסיווג כזה, מאחר ויכול להיות אבחון "בריא" שבפועל הסיווג האמיתי הוא "חולה", ולכן המודל שאנו בנינו לא יכול להיות כלי לאבחון מחלות מוחלט.

ביבליוגרפיה

<https://pubmed.ncbi.nlm.nih.gov/31407214/>
<https://www.nature.com/articles/s41598-020-75816-w>
[https://www.cell.com/cell/fulltext/S0092-8674\(18\)30154-5](https://www.cell.com/cell/fulltext/S0092-8674(18)30154-5)