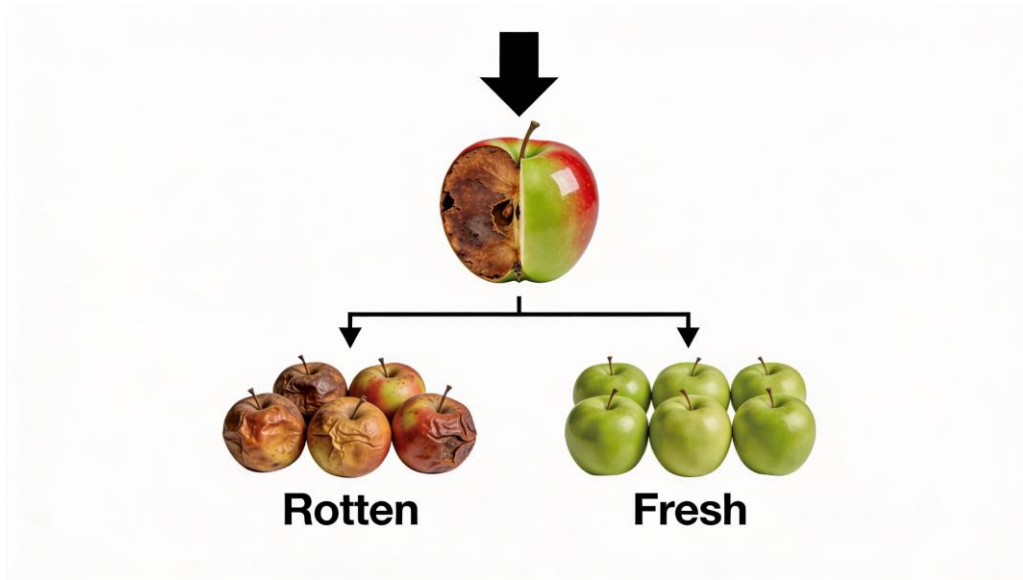


פרויקט גמר עיבוד תמונה

Traditional and Deep Learning-Based Image Processing for Fruit Quality and Rot Detection Using Public Datasets



מגשים: עמית דרורי ועמרי בוטון

תאריך הגשה:

קורס מספר 71254

מבוא

במהלך שרשרת האספקה החקלאית, חלק גדול מאובדן היבול מתרחש דווקא אחרי הקטיף, בעקבות פגמי איכות ובראשם ריקבון של הפירות. בבתי אריזה רבים, המיון בין פירות טריים לפירות פגומים עדיין מתבסס בחלקו על בדיקה חזותית של עובדים, שנדרשים לעבור על אלפי יחידות בשעה. זהו תהליך איטי, יקר ותלוי מאוד בגורם האנושי – עייפות, שונות בין עובדים ותנאי תאורה משתנים גורמים לכך שפירות רקובים "בורחים" למגש הצרכן, ופירות טובים נזרקים לחינם. מערכות ראייה ממוחשבת מציעות חלופה אוטומטית, עקבית ולא פולשנית, שמסוגלת לעבוד בקצב פס הייצור ולהפחית אובדן ותלונות איכות.

בשנים האחרונות מודלים של למידה עמוקה שינו את התחום: רשתות קונבולוציה (CNN) מצליחות ללמוד בעצמן מאפייני צורה, צבע ומרקם ישירות מהתמונות, והובילו לאוטומציה של משימות רבות כמו זיהוי מחלות בעלים, סיווג פירות ודרוג איכות. לצד זאת, שיטות קלאסיות יותר – המבוססות על סגמנטציה, חילוך מאפיינים ידני וספים שהוגדרו על ידי מומחים – עדיין נפוצות בגלל הפשטות, השקיפות והצורך המופחת בנתוני אימון.

בפרויקט זה השונו בין שתי הגישות. בגישה הראשונה אימצנו מודל קלאסי ללא למידה עמוקה: חילוך ידני של מאפייני צבע (במרחב HSV) ומרקם (מדדי GLCM כגון קונטרסט, הומוגניות ואנרגיה), ולאחר מכן בניית ציון משוקלל המבוסס על ספים שהוגדרו באופן ידני. זהו מודל ידני שנשען על הידע שלנו לגבי איך פירות רקובים נראים: כהים יותר, פחות רוויים ובעלי טקסטורה מחוספסת, אך דורש כוונן ידני רב ואינו בהכרח כללי למקרים חדשים. כתוספת לגישה הבסיסית, במקום להשתמש בערכי סף קשיחים בודדים לכל מאפיין, הגדרנו עבור כל סף רשימה של מספר ערכים אפשריים, הנמצאים סביב הסף שמצאנו (גרף 1) ואפשרנו למודל לבחור את הצירוף האופטימלי ביניהם.

בגישה השנייה השתמשנו במודל למידה עמוקה מוג 50-ResNet עם העברת למידה (transfer learning). במקום להגדיר בעצמנו אילו מאפיינים חשובים, נתנו לרשת קונבולוציה ללמוד אותם ישירות מן הפיקסלים, על בסיס משקלים שאומנו מראש על מאגר תמונות כללי. לצורך כך החלפנו את שכבת הסיווג האחרונה בשכבה דו מחלקתית (fresh/rotten) וביצענו כיוונון (fine-tuning) מדורג של חלק מהשכבות העמוקות על סט התמונות שלנו, עד לקבלת מודל שמסוגל להבחין בין פרי טרי לרקוב ללא הגדרה ידנית של פיצ'רים.

השוואת שתי הגישות מאפשרת לבדוק עד כמה ניתן לשפר את הדיוק והיציבות של המיון, ומה המחיר מבחינת מורכבות חישובית וכמות נתונים נדרשת.

בסיס נתנים

לצורך הפרויקט נעשה שימוש במאגר התמונות "Fruits fresh and rotten for classification" מאתר Kaggle. המאגר כולל כ-10,900 תמונות צבעוניות של שלושה סוגי פירות – תפוחים, בננות ותפוזים – כאשר לכל פרי שתי קטגוריות: fresh ו-rotten.

התמונות מסודרות בתיקיות בהתאם לסוג הפרי ולמצב הטריות שלו, ומופרדות לתיקיות train ו-test.

הצילומים הם בתצורת RGB, באיכות טובה ובפורמט PNG, עם חדות מספקת לזיהוי פרטים על פני הפרי. הפירות מצולמים מטווח קרוב, על רקע לבן ונקי, בתאורה אחידה ובזווית צילום פשוטה – כך שהאובייקט המרכזי בולט ומופרד היטב מהרקע. תנאים אלה מתאימים במיוחד למשימות סיווג חזותי.

מכיוון שהמאגר הראשון יצר תנאים פשוטים יחסית למודל, ביקשנו לבחון את ביצועיו גם על מאגר מורכב יותר. לשם כך הורדנו מדאטה סט נוסף לכונן (Drive) כ-200 תמונות לכל סגנון (1200 סה"כ), שעליו הרצנו את שני המודלים. מאגר זה ייקרא בהמשך הפרויקט Drive.

תמונה מספר 1- דוגמה מהדאטה סט של קאגל



תמונה מספר 2- דוגמה מהדאטה סט Drive



שיטות

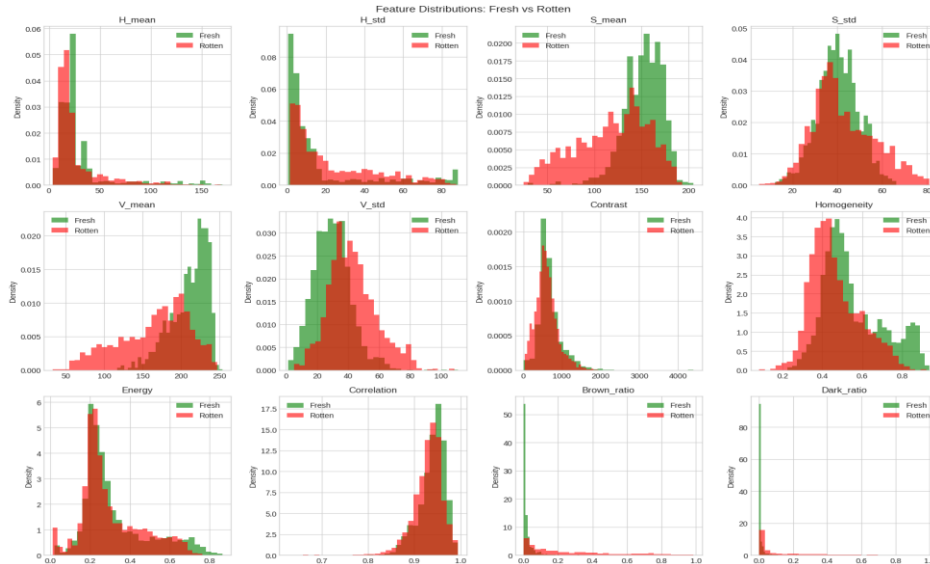
מודל ידני:

בשלב הראשון ביצענו ניתוח של הפיצ'רים על ידי שרטוט התפלגויות (היסטוגרמות) לכל אחד מאפיין בנפרד מהתמונות (kaggle), עבור דוגמאות של פרי טרי ופרי רקוב. מטרת השלב הזה הייתה לזהות אילו פיצ'רים מראים הפרדה משמעותית בין שתי הקבוצות, ואילו פחות תורמים, מתוך הבנה שלא נוכל להשתמש בכל המאפיינים במודל הסופי. שימוש בכל הפיצ'רים היה מגדיל מאוד את המורכבות החישובית, מעלה סיכון ל-overfitting ומקשה על כוונן ספי החלטה, ולכן בחרנו להתמקד בתת קבוצה מצומצמת של מאפיינים שנראים מבדילים היטב בין פרי טרי לרקוב.

על בסיס גרף ההתפלגויות המצורף בחרנו מספר פיצ'רים בולטים והגדרנו עבורם ספי החלטה ידניים, שישמשו בכללי הסיווג הפשוטים שלנו:

- `v_mean_thr`: בהירות ממוצעת במרחב HSV, המשקף את רמת הבהירות הכללית של הפרי.
- `s_mean_thr`: רמת הרוויה הממוצעת, המתארת עד כמה הצבע "חי" או דהוי.
- `contrast_thr`: ערך ה-contrast מתוך תכונות ה-GLCM, שמודד שונות מקומיות בטקסטורה ומאפשר לזהות אזורים מחוספסים יותר האופייניים לרקבון.
- `homogeneity_thr`: מידת ההומוגניות של הטקסטורה (גם היא מ-GLCM), כאשר ערכים נמוכים מעידים על טקסטורה לא אחידה יותר.
- `brown_ratio_thr`: יחס הפיקסלים החומים בתמונה, המשמש אינדיקציה חזותית לרקבון.
- `dark_ratio_thr`: יחס הפיקסלים הכהים מאוד, המאפשר לקלוט אזורים חשוכים/פגומים בפרי.
- `H_std`: סטיית התקן של הצבע במרחב HSV, מתאר את השונות בצבע בתמונה. עבור פיצ'ר זה לא בחיפוש רשת והגדרנו סף ידני.

פיצ'רים אלו נבחרו משום שהתפלגויותיהם הראו הבדל ברור יחסית בין פירות טריים לרקובים, ובמקביל מספרם המצומצם מאפשר לשמור על מודל קל יחסית להבנה, לחישוב ולכיוון.



גרף 1 היסטוגרמות של פיצ'רים שונים עבור דוגמאות פרי טרי (מסומן בירוק) ורקוב (מסומן באדום).

לאחר מכן עברנו לשיטת Grid Search: במקום לקבע ערך סף יחיד לכל פיצ'ר, הגדרנו עבור כל אחד מהם טווח קטן של ערכי סף אפשריים סביב הערכים שנמצאו בנינוח הראשוני (מגרף 1), והרצנו לולאה שמנסה את כל הקומבינציות האפשריות ובוחרת את שילוב הספים שמביא ל-accuracy הגבוה ביותר על סט האימון.

מודל CNN

השתמשנו במודל 50-ResNet מאומן-מראש על ImageNet כלמידת העברה. התמונות הותאמו לגודל 224×224 פיקסלים, ושכבת הסיווג הסופית של המודל הוחלפה בשכבה דו מחלקתית (/ fresh / rotten), בעוד ששאר שכבות הרשת נשארו ללא שינוי ושימשו לחילוץ מאפיינים חזותיים כלליים מהתמונות.

חלוקת הנתונים

תמונות ה-train מדאטהסט Kaggle שימשו לאימון, וסט ה-test של Kaggle חולק אקראית (עם seed קבוע) ל-50% ולידציה ו-50% בדיקה.

תהליך האימון – שתי פאזות

האימון בוצע בשתי פאזות:

1. Phase 1 – אימון ראש המודל בלבד

כל שכבות 50-ResNet הוקפאו, ורק שכבת ה-FC החדשה. שלב זה כיוון את הסיווג הסופי לבעיה הדו מחלקתית שלנו על גבי תכונות כלליות שהמודל כבר מכיר.

2. Phase 2 – Fine-tuning של 4layer + ראש המודל בשלב זה שוחררו גם פרמטרי 4layer יחד עם שכבת ה-FC, וקצב הלמידה הוקטן לצורך כיוון עדן ללא פגיעה במשקולות שנלמדו בעבר על ידי המודל.

בסיום האימון טענו את המשקלים הטובים ביותר והערכנו את המודל על שני הדאטה סטים.

תוצאות

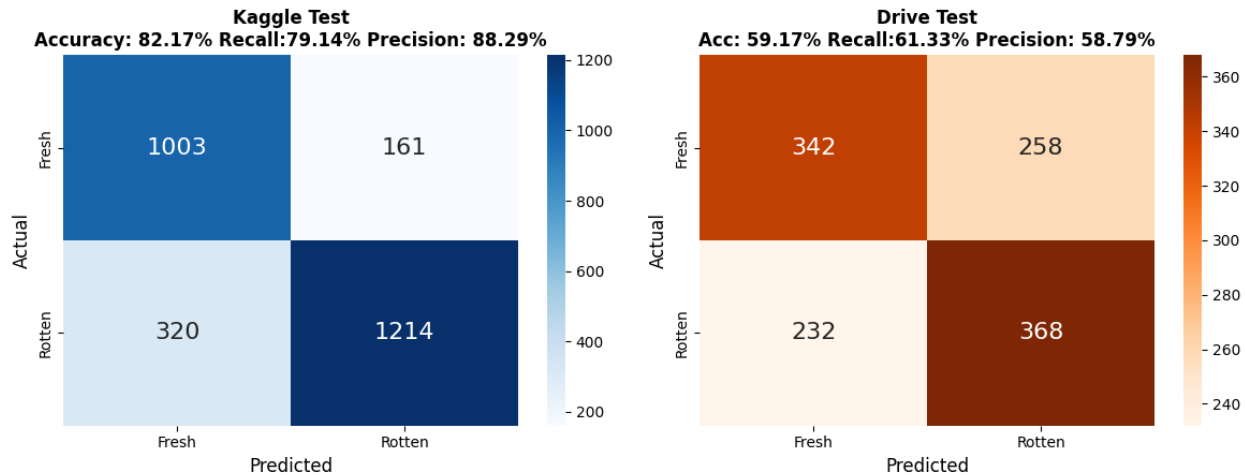
ביצועי המודל הידני על דאטהסט Kaggle

בשלב הראשון הערכנו את המודל הקלאסי על סט התמונות מ-Kaggle, שהוא מאורגן ומצולם בתנאים מבוקרים (רקע לבן, תאורה אחידה, זווית קבועה). במבט כולל, המודל השיג ביצועים טובים למדי גם על סט האימון וגם על סט הבדיקה: דיוק (Accuracy) של כ-82%, עם Precision ו-Recall סביב 80–88%, מה שמצביע על כך שהוא מצליח גם לזהות רוב הפירות הרקובים, וגם לא “להאשים” יותר מדי פירות רקובים כטריים.

המודל נוטה לבצע יותר טעויות מהסוג False Negative (פירות רקובים שמסווגים כטריים) מאשר False Positive, אך הפער אינו קיצוני. חלוקה לפי סוג פרי מגלה שבננות הן הקטגוריה הקלה ביותר לסיווג (דיוק ~90%, 1F ~92%), תפוחים באמצע (דיוק ~80%), ותפוזים הם האתגר המרכזי, עם דיוק סביב 75% ו-1F נמוך יותר (~71%).

ביצועי המודל הידני על דאטהסט DRIVE

בבחינת המודל על דאטהסט DRIVE, נרשמה ירידה משמעותית בביצועים: ה-Accuracy הכללי ירד לכ-59%, וה-1F לכ-60%, עם עלייה ברורה במספר השגיאות בשתי הקטגוריות (Recall, precision).

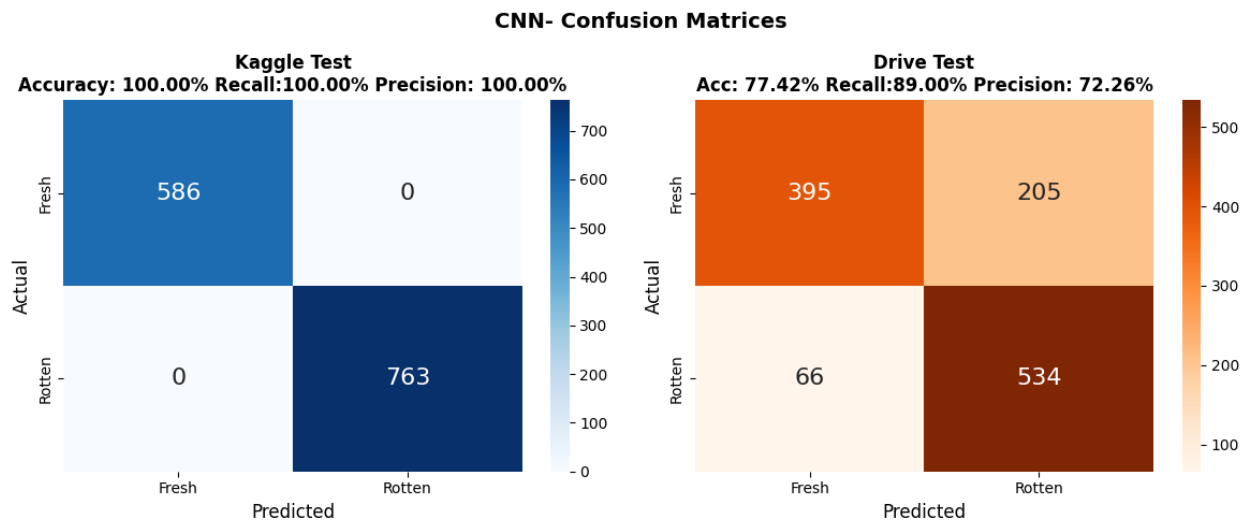


גרף 2 השוואת ביצועי המודל על פי פירות בין דאטה סט קאגל לדאטה סט Drive

ביצועי מודל ה CNN

המודל מבוסס 50-ResNet הציג ביצועים כמעט מושלמים על סט הבדיקה של Kaggle: דיוק, רגישות (Recall) ו-F1-Score של 100%.

בהרצת המודל על דאטהסט DRIVE, הביצועים נחלשו אך נותרו יציבים: הדיוק ירד לכ-77% וה-F1-Score לכ-80%. הרגישות נותרה גבוהה יחסית (89% Recall), כלומר רוב הפירות הרקובים עדיין מזוהים נכון), אך חלה עלייה ניכרת ב-False Positives – 205 דוגמאות של פירות טריים שסווגו כרקובים – מה שמסביר את הירידה ב-Precision לכ-72%. התוצאה היא מודל רגיש מאוד לרקבון, שמכליל לא רע לתנאי צילום קשים יותר, אך במחיר של יותר אזהקות שווא על פירות טריים בדאטהסט DRIVE.



גרף 3- השוואת ביצועי המודל דאטה סט קאגל לדאטה סט Drive

סיכום ההשוואה

Fruit	Dataset	Samples	Classical Accuracy	CNN Accuracy	Classical Precision	CNN Precision	Classical Recall	CNN Recall
Apples	Kaggle	996	80.02%	100.00%	89.88%	100.00%	75.37%	100.00%
Apples	Drive	400	75.50%	82.50%	78.02%	83.51%	71.00%	81.00%
Bananas	Kaggle	911	90.12%	100.00%	86.30%	100.00%	98.68%	100.00%
Bananas	Drive	400	68.00%	70.50%	61.18%	62.89%	98.50%	100.00%
Oranges	Kaggle	791	75.73%	100.00%	89.81%	100.00%	59.06%	100.00%
Oranges	Drive	400	34.00%	79.25%	23.77%	75.77%	14.50%	86.00%

טבלה 1- השוואת ביצועי המודלים תוך חלוקה לפירות

דיון ומסקנות

השוואת המודל הקלאסי למודל ה-CNN מבוסס 50-ResNet מראה הפרש עקבי בביצועים לטובת ה-CNN, בשני מאגרי התמונות. בהרצה על Kaggle, שני המודלים מצליחים לסווג פירות ברמת דיוק טובה, אך ה-CNN כמעט ואינו טועה, בעוד המודל הקלאסי עדיין מפספס חלק מהדוגמאות. בהרצה על DRIVE, שבו תנאי הצילום פחות אחידים, הפער בין המודלים גדל: המודל הקלאסי נפגע בצורה משמעותית, בעוד שה-CNN שומר על רמת ביצועים גבוהה בהרבה, גם אם נמוכה יותר מהמאגר הראשוני. הדבר מעיד שה-CNN מצליח לסווג את מאפייני הפירות בצורה טובה יותר, גם כאשר יש שינויים בתאורה, רקע וזווית צילום, בעוד שהספים הידניים רגישים מאוד להפרעות כאלה.

כאשר בוחנים את התוצאות לפי סוג פרי, התמונה מתחדדת עוד יותר. עבור תפוחים ובננות, המודל הקלאסי מתפקד לא רע, אך ה-CNN מוסיף עליו שיפור עקבי הן בדיוק והן ב-F1 (מדד המשקלל את ה-Recall וה-Precision) בשני מאגרי הנתונים. עבור תפוזים ההבדל הוא המשמעותי ביותר: המודל הקלאסי מתקשה מאוד, במיוחד ב-DRIVE, בעוד שה-CNN מצליח לייצב את הדיוק וה-F1 לרמה טובה, כך שגם עבור סוג הפרי הבעייתי ביותר מתקבלת הבחנה משמעותית טובה יותר בין טרי לרקוב. כלומר, היתרון של ה-CNN בולט במיוחד בדיוק באותם מצבים שבהם הפיצ'רים הידניים מפספסים, סוג פרי מורכב יותר ותנאי צילום פחות נקיים. הסבר אפשרי לכך הוא שספי המדדים בתפוז שונים משאר הפירות, מדדים שוני עלולים ליצור מצב זה: מדדי המרקם (GLCM), או מדדי היחס החום/שחור. קושי זה מעלה את הצורך ביצירת ספים ידניים עבור כל פרי בנפרד ומדגיש את היכולות של מודל ה-CNN.

מן ההשוואה עולה שהמודל הקלאסי מבוסס הספים מתאים כבסיס פשוט, אינטואיטיבי ומוסבר, במיוחד כאשר תנאי התמונה קרובים לתנאים עליהם נבנה המודל, וכאשר רוצים להבין במפורש אילו תכונות משפיעות על ההחלטה. עם זאת, הוא מוגבל ביכולת ההכללה שלו לתנאי צילום אחרים ולפירות שונים. ה-CNN, לעומת זאת, מצריך יותר משאבים חישוביים ופחות "שקיפות" אנושית, אבל מספק רמת דיוק ויציבות גבוהה בהרבה, הבדל זה מתבטא במיוחד בתמונות בעלות רעש, רקע מורכב או שונות גדולה בין פירות.

במישור המעשי, מודל קלאסי יכול לשמש ככלי ראשוני או כפתרון קל, בתנאים מבוקרים. לעומת זאת, במערכת שנועדה להתמודד עם מגוון רחב של מקורות תמונה וסוגי פירות, בפרט כאשר איכות הצילום אינה אחידה – מודל CNN כדוגמת 50-ResNet יציג ביצועים טובים יותר משמעותית. זמן המיון הקצר מאפשר, באופן תיאורטי, להטמיע את הפתרון בצד הלקוח עבור מכונות המצלמות בתנאים קבועים, וכך לשמור על רמות ההצלחה שנמדדו על הדאטה סט המקורי (שואף ל-100%)

שילוב שתי הגישות יכול להיות מעניין, שכן בשתייהן קיים בשלב כלשהו וקטור פיצ'רים: במודל הקלאסי אלו פיצ'רים בעלי משמעות פיזיקלית ברורה, וב-CNN אלו ייצוגים עמוקים ומורכבים יותר של אותה תמונה. יתכן וניתן לנצל זאת כדי לבנות מערכת היברידית שבה הפיצ'רים הידניים משמשים כשכבת שניתן להסביר ותאפשר בקרה על תהליך הסיווג, בעוד שהפיצ'רים הלמידים של ה-CNN מספקים את כוח ההפרדה הגבוה בפועל, כך שההחלטה מתקבלת מתוך שילוב של ידע אגרונומי מפורש וייצוג חזותי שנלמד מן הנתונים.

מקורות

- דאטה סט קאגל
<https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification>
- מקור הדאטה סט Drive
<https://data.mendeley.com/datasets/bdd69gyhv8/1>
- קישור לתיקיה עם דאטה סט Drive
https://drive.google.com/drive/folders/1z9_h3TB-RAvwEdgLgtkmKXdcNybRt_nl?usp=drive_link
- מאמר ששימש כמקור השראה לגישה הידנית ולהשוואת המודל הקלאסי למודל למידה עמוקה:
<https://pmc.ncbi.nlm.nih.gov/articles/PMC11452848/>
- תיאור מודל 50-resnet ב pytorch
<https://docs.pytorch.org/vision/stable/models/generated/torchvision.models.resnet50.html>
!
- קישור לגיטהאב:
https://github.com/omribooton/fruits_rotten_ripen_project.git
- קישור לקוד בקולאב
https://drive.google.com/file/d/1tz43sFWvYHeeicdzLQx8FkLHaFmrMPI4/view?usp=drive_link