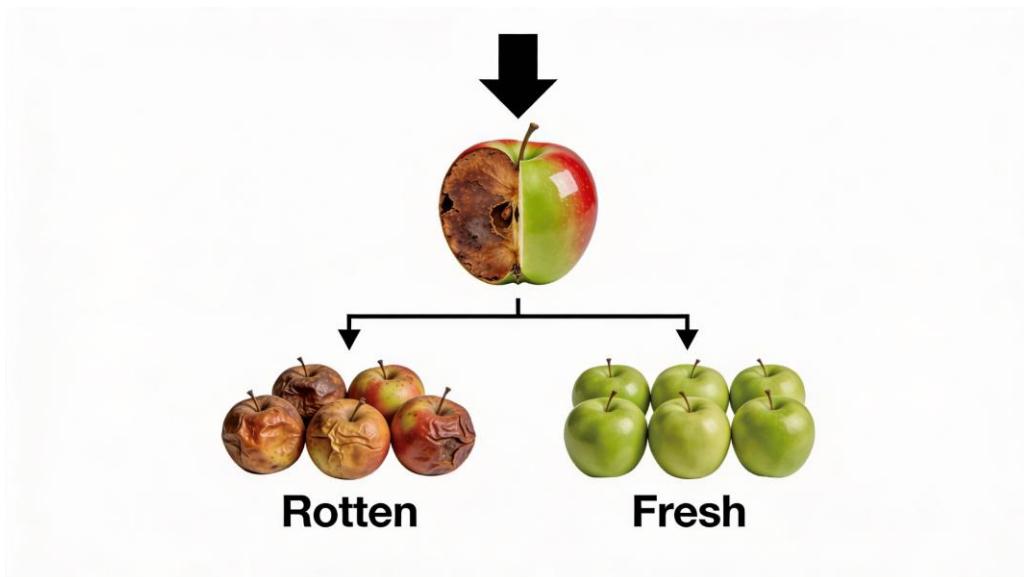


פרויקט גמר עיבוד תמונה

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



מגישיים: עמית דרורי ועמרי בוטון

תאריך הגשה:

קורס מס' 71254

מבוא

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

בשנים האחרונות מודלים של למידה عمוקה שינו את התחום: רשתות קונבולוציה (CNN) מצילחות ללמידה בעצמן מאפייני צורה, צבע ומרקם יישורות מהתמונה, והובילו לאוטומציה של משימות רבות כמו זיהוי מחלות בעליים, סיוג פירות ודרוג איכות. לצד זאת, שיטות קלאליסטיות יותר – המבוססות על סגמנטציה, חילוץ מאפיינים ידני וספיקים שהוגדרו על ידי מומחים – עדין נפוצות בגלל הפשטות, השקיפות והצורף המופחת בתנאי אימון.

בפרויקט זה השווינו בין שתי הגישות. בגישה הראשונה אימצנו מודל קלאלי ללא למידה عمוקה: חילוץ ידני של מאפייני צבע (במרחב HSV) ומרקם (מדדי GLCM כגון קונטרסט, הומוגניות ואנרגיה), ולאחר מכן בניה ציון ציון משוקל המבוסס על ספיקים שהוגדרו באופן ידני. זהו מודל ידני שנשען על הידע שלנו לגבי איך פירות רקובים נראים: כהים יותר, פחות רזויים ובReLU טקסטורה מחוספסת, אך דרוש כוונון ידני רב ואין בהכרח כללי למקרים חדשים. כתוספת לגישה הבסיסית, במקומות להשתמש בערכי סף קשייחים בודדים לכל מאפיין, הגדרנו עבור כל סף רשימה של מספר ערכים אפשריים, הנמצאים סביב הסף שמצוינו (גרף 1) ואפשרנו למודל לבחור את הצירוף האופטימלי ביניהם.

בגישה השנייה השתמשנו במודל למידה عمוקה מסוג ResNet-50 עם העברת למידה (transfer learning). במקומות להגדיר בעצמינו אילו מאפיינים חשובים, נתנו לרשת קונבולוציה ללמידה אותם יישורות מן הפיקסלים, על בסיס משקלים שאומנו מראש על מאגר תמונות כללי. לצורך כך החלפנו את שכבת הסיווג האחורונה בשכבה זו מחלקתית (fresh/rotten) וביצענו ציוןון (fine-tuning) מדווג של חלק מהשכבות העמוקות על סט התמונות שלנו, עד לקבלת מודל שמסוגל להבחין בין פרי טרי לroker לא הגדירה ידנית של פיצ'רים.

השוואת שתי הגישות אפשרית לבדוק עד כמה ניתן לשפר את הדיקוק והיציבות של המיוון, ומה המחיר מבחינה מורכבות חישובית וכמות נתונים נדרש.

בסיום נתנים

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

התמונות מסודרות בהתאם לסוג הפרי ולמצב הטריות שלו, ומופרדות לתיקיות *train*-*i-test*.

הצילומים הם בתצורת RGB, באיכות טובה ובפורמט PNG, עם חdots מספקת לזיהוי פרטיהם 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 לבדוק את ביצועיו גם על מאגר מורכב יותר. לשם כך הורדנו מדאטה סט נוסף לכונן (Drive) כ-200 תמונות לכל סגנון (1200 סה"כ), שעליינו הרצנו את שני המודלים. מאגר זה יקרא בהמשך הפרויקט *Drive*.

תמונה מספר 1 - דוגמה מהדאטה סט של Kaggle



תמונה מספר 2 - דוגמה מהדאטה סט Drive



שיטות

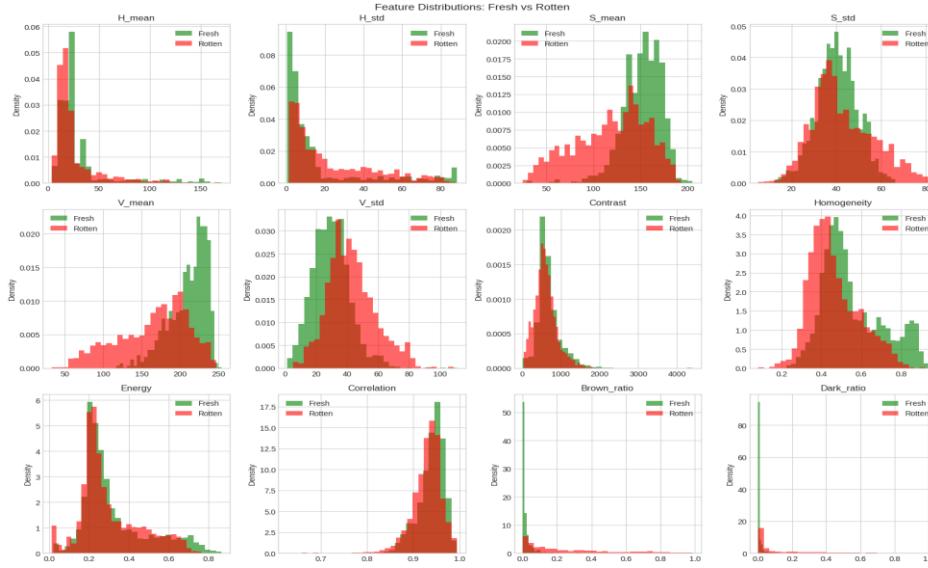
מודל ידני:

בשלב הראשון ביצענו ניתוח של הפיצרים על ידי שרטוט התפלגיות (היסטוגרמות) לכל אחד מאפיין בנפרד מהתמונה (kaggle), עברו דוגמאות של פרי טרי ופרי רקוב. מטרת השלב זהה הייתה לזהות אילו פיצרים מראים הפרדה משמעותית בין שתי הקבוצות, ואילו פחות תורמים, מתוך הבנה שלא יוכל להשתמש בכל המאפיינים במודל הסופי. שימוש בכל הפיצרים היה מגדיל מאוד את המורכבות החישובית, מעלה סיכון ל-overfitting ומקשה על כוונון ספי החלטה, ולכן בחרנו להתמקד בתת קבוצה מצומצמת של מאפיינים שנראים מבדילים היטב בין פרי טרי לركוב.

על בסיס גרפ' התפלגיות המצויר בחרנו מספר פיצריםבולטים והגדכנו עבורם ספי החלטה ידניים, שיישמו בכללי הסיווג הפשוטים שלנו:

- mean_thr: בהירות ממוצעת במרחב HSV, המשקף את רמת הבاهירות הכללית של הפרי.
- mean_s: רמת הרויה ממוצעת, 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 לרכיבן.
- homogeneity_thr: מידת ההומוגניות של הטקסטורה (גם היא מ-GLCM), כאשר ערכיהם נמוכים מעידים על טקסטורה לא אחידה יותר.
- brown_ratio_thr: ייחס הפיקסלים החומים בתמונה, המשמש אינדיקציה חזותית לרכיבן.
- dark_ratio_thr: ייחס הפיקסלים הכהים מאוד, המאפשר לקלוט אזורים חשוכים/פגומים בפרי.
- H_std: סטיית התקן של האבע במרחב HSV, מתאר את השונות בצבע בתמונה. עבור פיצר זה לא בחיפוש רשת והגדכנו סף ידני.

פיצרים אלו נבחרו מסווג שההתפלגויותיהם הראו הבדל ברור יחסית בין פירות טריים לרכיבים, ובמקביל מסווג המצוצם מאפשר לשומר על מודל קל יחסית לבניה, לחישוב ולכיזון.



גרף 1 היסטוגרמות של פיצרים שונים עבור דוגמאות פרי טרי (מסומן בירוק) ורקב (מסומן באדום).

לאחר מכך ערכנו לשיטת Grid Search: במקומם קבוע ערך ספ. יחיד לכל פיצר, הגדרנו עבור כל אחד מהם טווח קטן של ערכי ספ. אפשריים סבירים שנמצאו בניתוח הראשוני (מגרף 1), והרצינו לו לאלה שמנסה את כל הקומבינציות האפשריות ובוחרת את שילוב הספים שմביא ל-*accuracy* הגבוה ביותר על סט האימון.

מודל CNN

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

חלוקת הנתונים

תמונות ה-*train* מדאטסה Kaggle שימשו לאימון, וט ה-*test* של Kaggle חולק אקראיית (עם seed קבוע) ל-50% ולידציה ו-50% בדיקה.

תהליכי האימון – שתי פאזהות

האימון בוצע בשתי פאזהות:

1. Phase 1 – אימון ראש המודל בלבד כל שכבות ResNet-50 הוקפאו, ורק שכבת ה-FC החדשה. שלב זה כיוון את הסיום הסופי. לבעה הדו מחלוקתית שלנו על גבי תכונות כלליות שהמודל כבר מכיר.

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

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

תוצאות

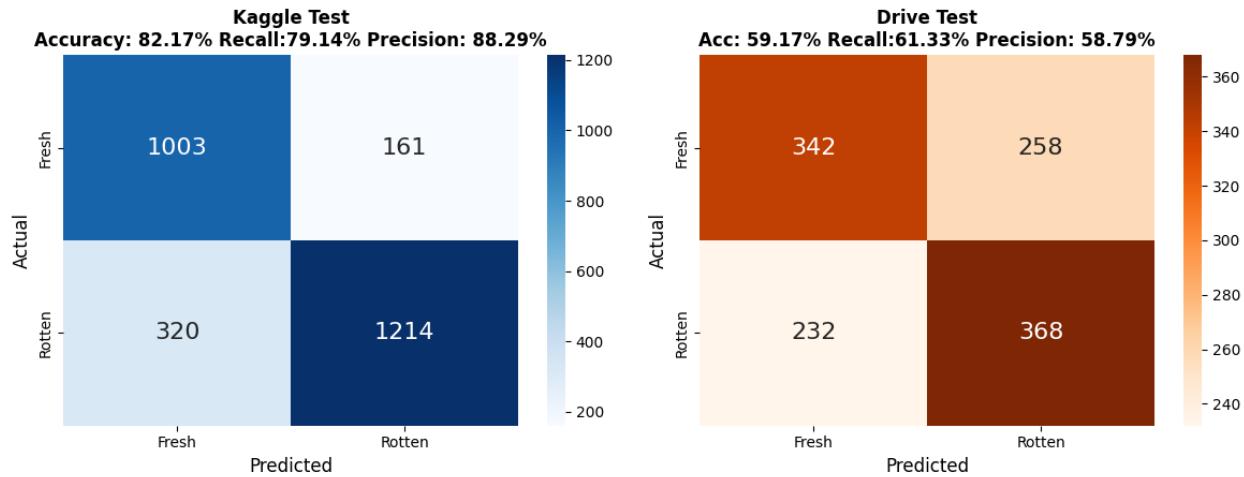
ביצועי המודל הידייני על דאטסהט Kaggle

בשלב הראשון הערכנו את המודל הקלטי על סט התמונות m-Kaggle, שהוא מאורגן ומצלום 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 סביבה 80–88%, מה שמצויב על כך שהוא מצליח גם לזהות רוב הפירות הרקופיים, וגם לא "להAsset" יותר מדי פירות רקובים כטריים.

המודל נוטה לבצע יותר טעויות מהסוג False Negative (פירוט רקובים שמסוגים כטריים) מאשר False Positive, אך הטעות אינה קיצונית. חלוקה לפי סוג פרי מגלת שבנון הנקטורי הקליה ביותר לשינוי (דיוק ~90%, 1F ~92%), תפוחים באמצע (דיוק ~80%), ותפוזים הם האתגר המרכזי, עם דיוק סביר 75% ו-1F נמוך יותר (~71%).

ביצועי המודל הידייני על דאטסהט DRIVE

בבחינת המודל על דאטסהט DRIVE, נרשמה ירידת משמעותית ביצועים: ה-accuracy הכללי ירד לכ-59%, וה-1F לכ-60%, עם עלייה ברורה במספר השגיאות בשתי הקטגוריות (Recall, precision).

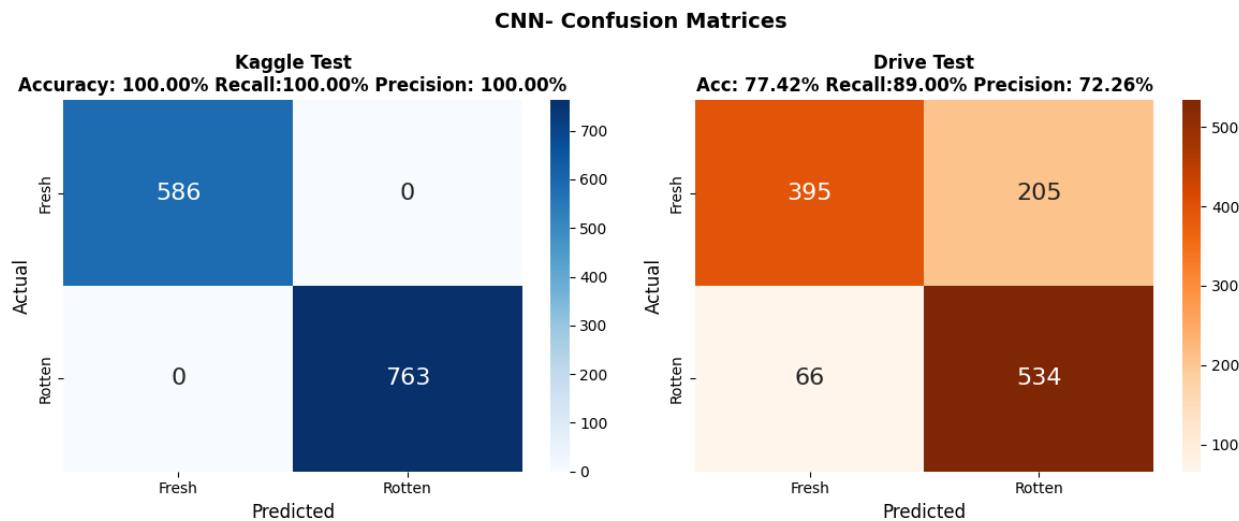


graf 2 השוואת ביצועי המודל על פי פירוט בין DATAה סט Kaggle לDATAה סט Drive

ביצועי מודל ה CNN

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

בהרצאת המודל על DATAה סט DRIVE, הביצועים נחלשו אך נותרו יציבים: הדיווק ירד לכ- 77% וה- F1-Score לכ- 80%. הרגישות נותרה גבוהה יחסית (Recall 89%), ככלומר רוב הפירות הרקובים עדין מזוהים נכון), אך חלה עלייה ניכרת ב- False Positives – 205 דוגמאות של פירות טריים שסואגו כרקבובים – מה שמסביר את הירידה ב- Precision לכ- 72%. התוצאה היא מודל רגיש מאוד לרקען, שמקליל לא רע לתנאי צילום קשים יותר, אך במחיר של יותר אזעקות שווא על פירות טריים בDATAה סט DRIVE.



גרף 3- השוואת ביצועי המודל דатаה סט קאג'ל לדטה סט דרייב

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

דין ומסקנות

השווות המודל הקלואטי למודל ה-CNN מבוסס ResNet-50 מראה הפרש עקבי בביטויים לטובת ה-CNN, שני מאגרי התמונות. בהערכתה על Kaggle, שני המודלים מצילחים לסוג פירות ברמת דיקוטובה, אך ה-CNN כמעט ואינו טועה, בעוד המודל הקלואטי עדין מפספס חלק מהדגימות. בהערכתה על DRIVE, שבו תנאי הצילום פחותים איחדים, הפער בין המודלים גדול: המודל הקלואטי נפגע בצורה משמעותית, בעודשה-CNN שומר על רמת ביצועים גבוהה בהרבה, גם אם נמוכה יותר מהמאגר הראשון. הדבר מעיד שה-CNN מצילח לסוג את מאפייני הפירות בצורה טוביה יותר, גם כאשר יש שינויים בתאורה, רקע וזווית צילום, בעוד שההסתפים הידניים רגילים מאוד להפרעות אלה.

כאשר בוחנים את התוצאות לפי סוג פרי, התמונה מתחדשת עוד יותר. עבור תפוחים ובננות, המודל הקלואטי מתפרק לא רע, אך ה-CNN מוסיף עליו שיפור עקבי הן בדיק והן ב-F1 (מדד המשקל את Precision והחומר Precision) בשני מאגרי הנתונים. עבור תפוזים ההבדל הוא המשמעותי ביותר: המודל הקלואטי מתקשה מאוד, במיוחד ב-DRIVE, בעודשה-CNN מצילח לייצב את הדיק וה-F1 לרמה טוביה, כך שגם עבור סוג הפרי הבויאתי ביותר מתקבלת הבדיקה משמעותית טוביה יותר בין טרי לרקוב. לעומת זאת, היתרון של ה-CNN בולט במיוחד בבדיקה מוצבים שונים שבתוכם היפיצרים הידניים מפספסים, סוג פרי מסוים יותר ותנאי צילום פחות נקיים. הסבר אפשרי לכך הוא שטפי המדדים בתפוז שונים מאשר היפירות, מדדים שונים עלולים ליצור מצב זה: מדדי המרכיב (GLCM), או מדדי היחס החום/שחור. קושי זה מעלה את הצורך ביצירת ספיקים ידניים עבור כל פרי בנפרד ומדגיש את היכולות של מודל CNN.

מן ההשוואה עולה שהמודל הקלואטי מבוסס בסיסי פשוט, אינטואיטיבי ומוסבר, במיוחד כאשר תנאי התמונה קרובים לתנאים עליהם נבנה המודל, וכך אשר רוצים להבין במפורש אילו תכונות משפיעות על ההחלטה. עם זאת, הוא מוגבל ביכולת ההכללה שלו לתנאי צילום אחרים ולפיריות שונים. ה-CNN, לעומת זאת, מצריך יותר משאים חישוביים ופחות "שקייפות" אונושית, אבל מספק רמת דיקוטיות גבוהה בהרבה, הבדל זה מתבטא במיוחד במקרים בעלות רעש, רקע מורכב או שונות גדולה בין פירות.

במשור המעשית, מודל קלואטי יכול לשמש ככלי ראשוני או כפתרון קל, בתנאים מבוקרים. לעומת זאת, המערכת שנועדה להתמודד עם מגוון רחב של מקורות תמונה וסוגי פירות, בפרט כאשר איכות הצילום אינה אחידה – מודל CNN כדוגמת ResNet-50 יציג ביצועים טובים יותר משמעותית. זמן המיכון קצר ככל, באופן תיאורטי, להטמע את הפתרון לצד הלקווח עבור מכונות המצלמות בתנאים קבועים, וכך לשמר על רמות הצלחה שנמדדו על הדטה סט המקורי (שואף ל-100%)

שילוב שתי הגישות יכול להיות מעניין, שכן בשתיهن קיימ בשלב כלשהו וקיטור פיצ'רים: במודל הקלאסי אלו פיצ'רים בעלי משמעות פיזיקלית ברורה, וב-CNN אלו ייצוגים עמוקים ומורכבים יותר של אותה תמונה. ניתן לנצל זאת כדי לבנות מערכת היברידית שבה הפיצ'רים הידניים משמשים כשבב שnitן להסביר ותאפשר בקרה על תהליכי הסיווג, בעוד שהפיצ'רים הלמידים של ה-CNN מספקים את כוח ההפרדה הגדולה בפועל, כך שההחלטה מתתקבלת מtower של ידע אגרונומי מפורש וייצוג חזותי שנלמד מן הנתונים.

מקורות

• דата סט Kaggle

<https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification>

• מקור הדטה סט Drive

<https://data.mendeley.com/datasets/bdd69gyhv8/1>

• קישור לתיקייה עם דטה סט Drive

https://drive.google.com/drive/folders/1z9_h3TB-RAvwEdgLgkrmKXdcNyBt_nI?usp=drive_link

• מאמר ששימושו כמקור השראה לגישה הידנית ולהשוואת המודל הקלאסי למודל למידה عمוקה:

<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 בקולאוב

https://drive.google.com/file/d/1tz43sFWvYHeeidzLQx8FkLHaFmrMPI4/view?usp=drive_link