DOES A MACHINE DO BETTER AT PREDICTING STARCH FROM GENE EXPRESSION IN APPLES THAN A HUMAN?

Aden Athar[1], Huiting Zhang[1], Zach Hall[1], Heidi Hargarten[2], Loren Honaas[2], Stephen Ficklin[1]
[1]Washington State University Department of Horticulture,
[2]Physiology and Pathology of Tree Fruits Research Unit, USDA Agricultural Research Service, Wenatchee, WA , USA

WASHINGTON STATE UNIVERSITY

INTRODUCTION

Background

- An apple is "mature" and ready for harvest when appearance, flavor, storability, marketability, etc. are optimized.
- Growers use a few ways to test when an apple is ready for harvest such as starch clearing and visual assessment.
- These tests are often inconstant at determining maturity from year to year.
- Measurements such as starch clearing are biased by human raters who give a rating from 1-8 (Figure 1) and actual values vary from person to person.
- Apple cultivars are genetic clones of another so maturity differences from year to year are caused by environmental differences.
- Even though all apples in a cultivar are genetically identical, changes in expression of their genes is the result of the changes in environmental conditions.

Hypotheses

Because gene expression is affected by the environment, there exists a set of genes whose expression profile can be used to predict maturity.

Preliminary Results

- Our group develop a software tool Granny (https://granny.readthedocs.io/) that uses deep learning and image processing to provide a percentage of starch clearing for all apples on a tray (see Figure 3)
- We have shown that a supervised machine learning model (Random Forest) using gene expression profiles can predict human-rated starch clearing (see Figure 4B).

Project Goal

Answer the question: Can we improve the predictability of our models by using a machine-learning/image processing approach instead of human ratings and remove human bias.

METHODS

- 364 apple samples collected from multiple varieties (2021 season).
  - Granny Smith, Red Delicious, Cosmic Crisp ®, HoneyCrisp
- 324 apples were used for this study
  - Apples were placed on trays, imaged and rated by a human.
  - Tray images were segmented and rated using Granny.
- RNA was extracted and sequenced
  - Gene counts were normalized and filtered.
- Random forest models were trained to predict starch ratings
  - Model 1: gene expression is used to predict human ratings
  - Model 2: gene expression is sued to predict Granny ratings
  - Model stability assessed via repeated k-fold cross-validation; feature selection via Boruta.

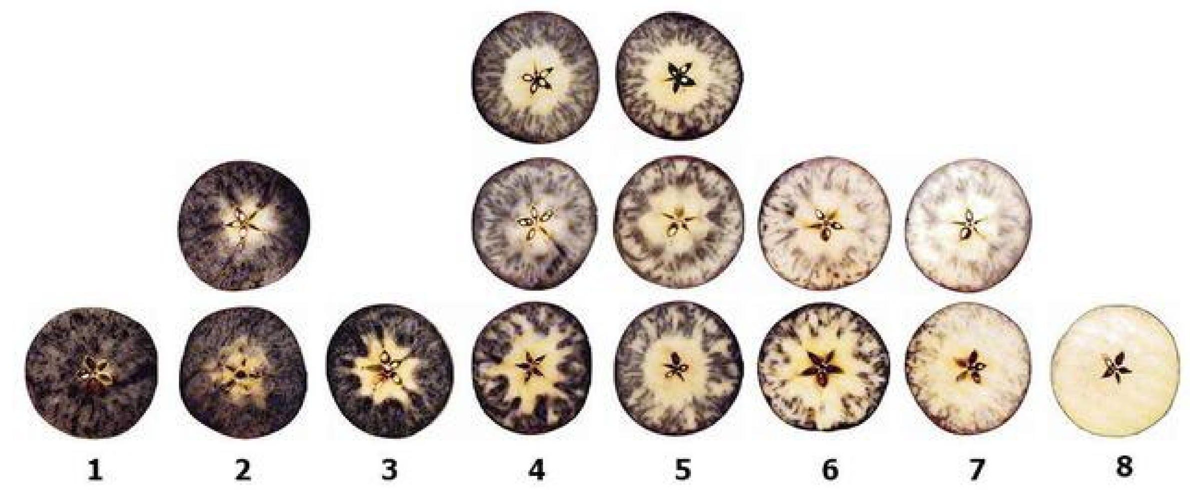

Figure 1. Example Starch Card Used by Human Raters. This image is the Cornell Starch-Iodine index card. It shows a variety of iodine-stained cross-sections of apple fruit. The numbers 1 to 8 are the rating provided by the human-rating if the picture above the number looks similar.

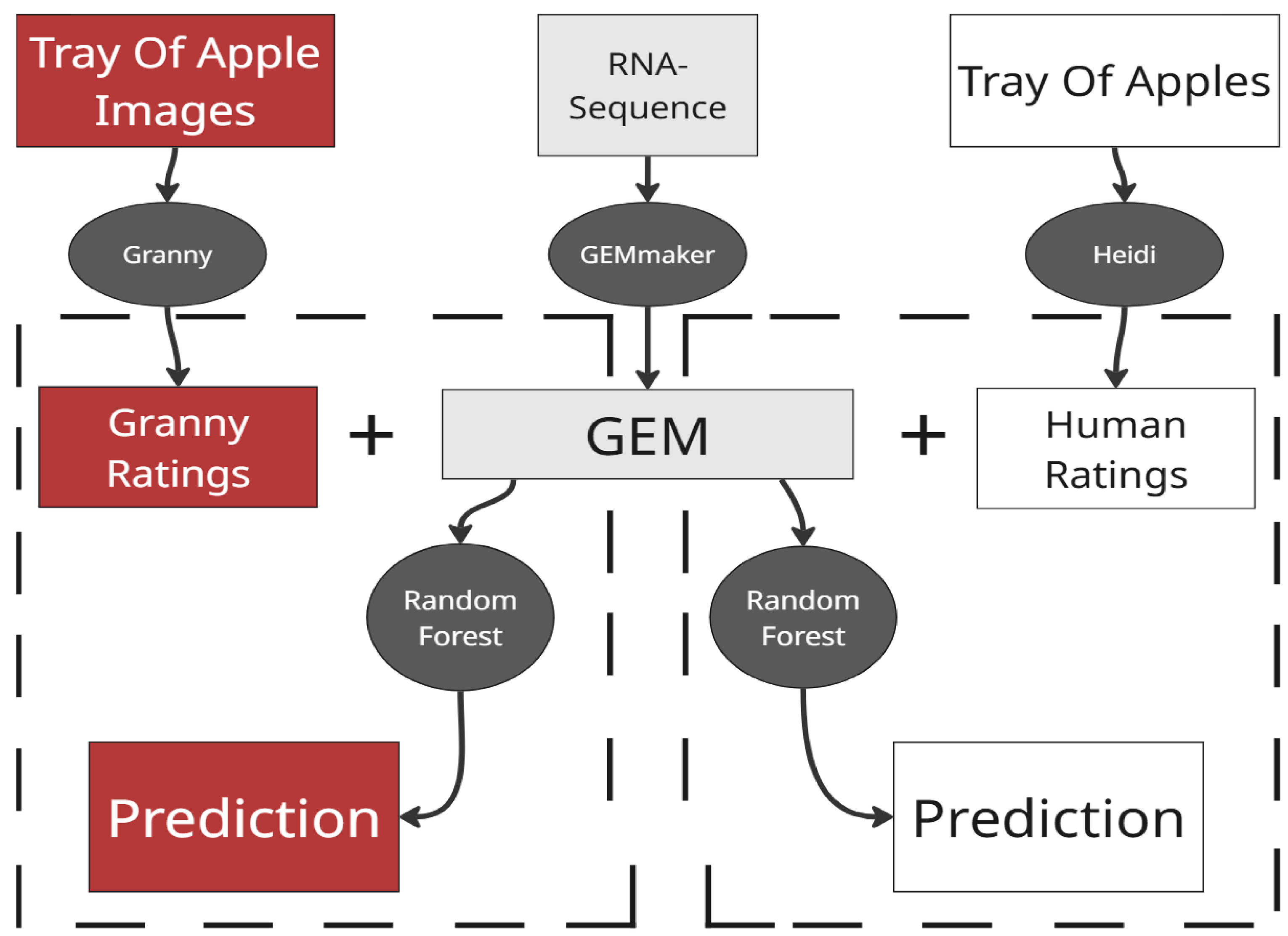

Figure 2. Project Workflow. Flowchart of the entire analytical pipeline, apple imaging, starch rating, RNA extraction through RNA-seq processing (GEMmaker), data preparation (merging and cleaning), Granny image analysis,. Colored blocks and icons represent data sources, analytical tools, and outputs.

Figure 3. Segmented Tray of Apples. Cross-section of iodine-stained apples on the tray are identified on the tray using our Granny software. Individual apples are colored differently, with bounding boxes highlighting their coordinates in the image.

RESULTS

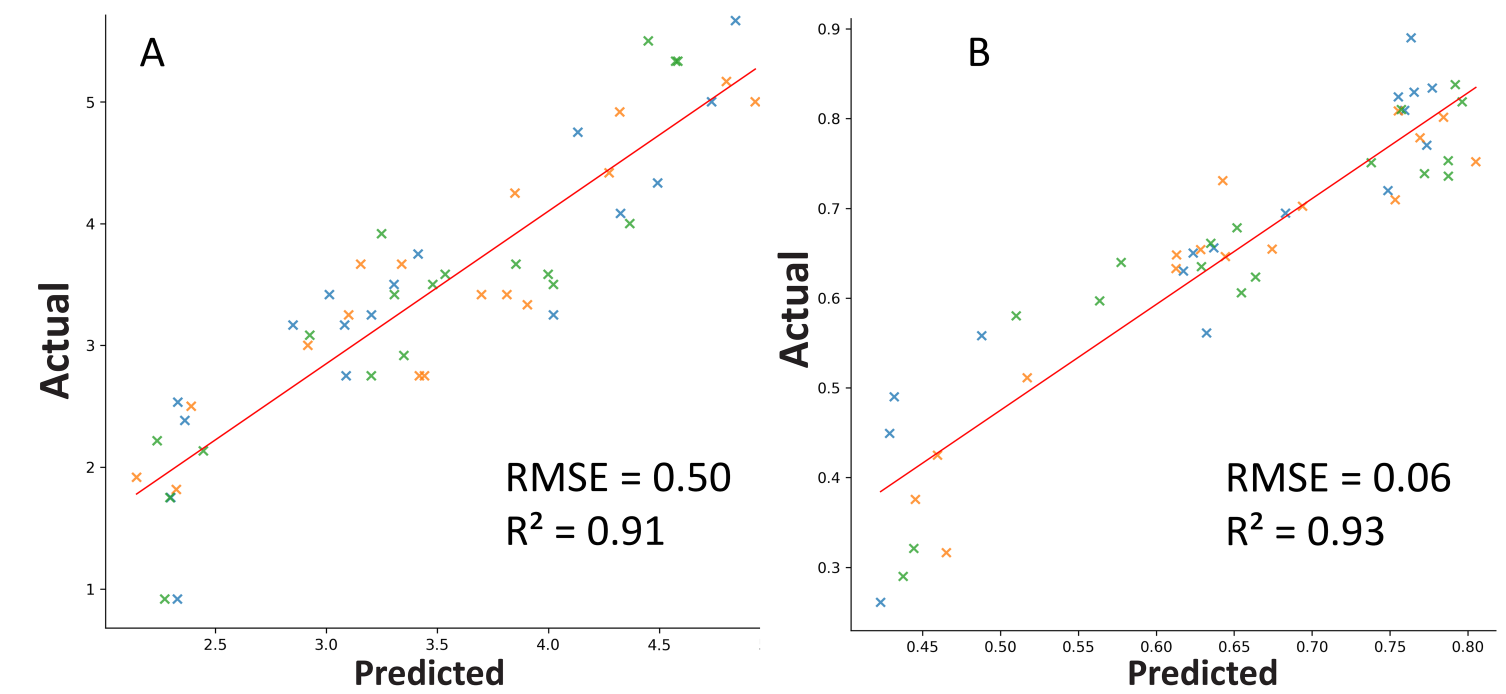

Figure 4. Model Performance Scatterplots. Each point is an individual apple from 18 trays of apples. In panel A, the y-axis is the human starch rating from 1 to 8. In panel B the y-axis in the Granny percent. The x-axis in both is the result from the random forest model for predicting starch.

INTEPRETATION

Overall, the random forest model trained on the Granny ratings did outperform the human ratings with an R^2 of 0.93 vs 0.91. However, the difference may not be significant, but there are some steps that may need to be addressed to make a conclusive statement if the Granny ratings are better than human ratings. First, we should be using differences in RMSE and not R^2 for measuring performance, but due to time constraints mapping of Granny ratings to the scale was not possible and may not be precise. Second, human ratings for different apple varieties have different scales, and these were not taken into account when training. Third, the human-ratings were provided by a single rater, to determine if Granny ratings are more consistent, we would need ratings from multiple humans.

CONCLUSIONS

Granny's automated image analysis delivers continuous (0–100) starch estimates that are objective, reproducible, and free from scorer fatigue or bias. In our study, models trained on Granny ratings achieved lower prediction error and higher explained variance than those trained on human scores, but it may not be meaningful. More work is needed with larger, well-annotated datasets.

ACKNOWLEDGMENTS

USDA NIFA
United States Department of Agriculture
National Institute of Food and Agriculture

WASHINGTON TREE FRUIT RESEARCH COMMISSION

Funding for this project was provided by the Washington State Tree Fruit Research Commission and the USDA NIFA.