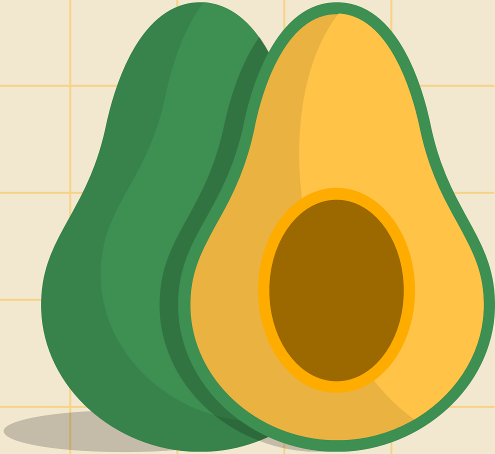
Detecting Avocado Ripeness

Group 3: Maansi Taori (Leader), Connor Overbay, Bianca Linares

DS 4002

December 4, 2024



Background

5.5 million

tons of avocados produced per year

40%

of food becomes consumer waste

Impact

identifying proper ripeness can inform healthier decisions and quality management

Project Details

Hypothesis

The **ResNet-50 model**, trained on a large dataset of avocado images at different ripeness levels, will achieve statistically significant classification performance across the **five ripeness stages**, with $p < 0.05$ and significant stage-specific metrics such as **precision, recall, and F1-score.**

Research Question

Can the ripeness of a fruit be effectively predicted using an image classification model?

Modeling Approach

By training **ResNet-50** on a labeled dataset representing the five ripeness stages, it will learn to **distinguish subtle visual differences**. Its robust capabilities make it well-suited for this task, with potential applications in areas such as retail and agriculture.

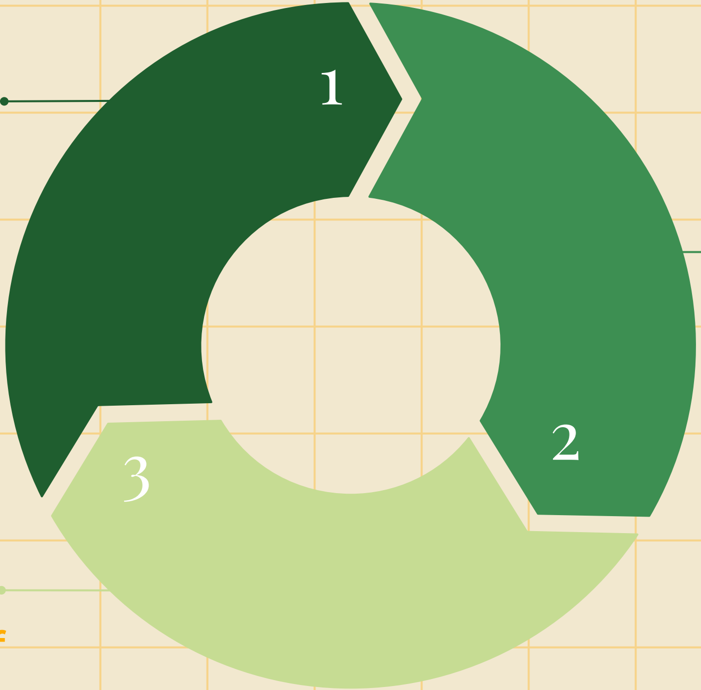
Data Collection

Data Acquisition

The Centro de Biotecnologia e Quimica Fina

Mendeley Data

No ethical or licensing concerns



Data Cleaning

Removed NA and duplicate images

Brightness normalization adjustments

Data Organization

14,710 labeled images

Divided into subsets of training (11,768) and testing (2,942)

Data Dictionary

- File Name
- Time Stamp
- Group
- Sample
- Day of Experiment
- Ripening Index Classification
- Image

Pipeline

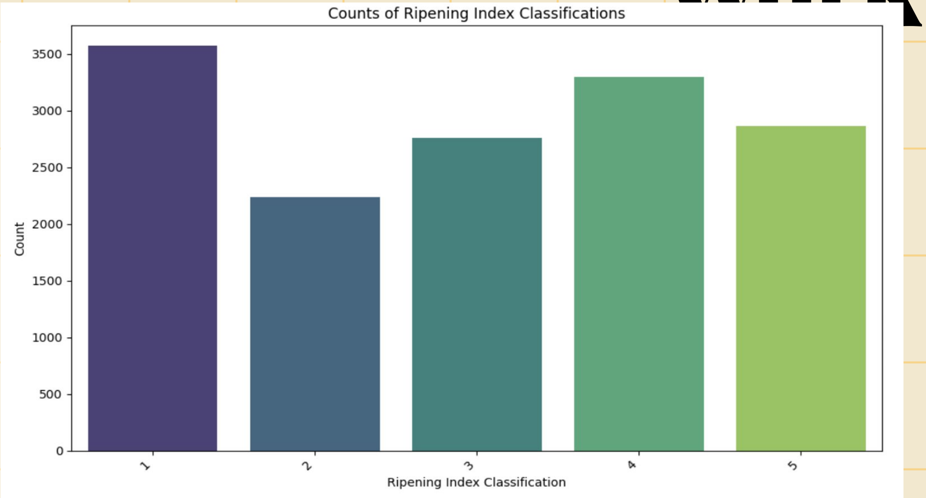
Establish Dataset

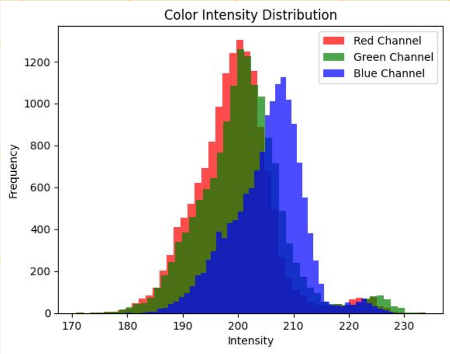
Initial Exploratory and Data Analysis

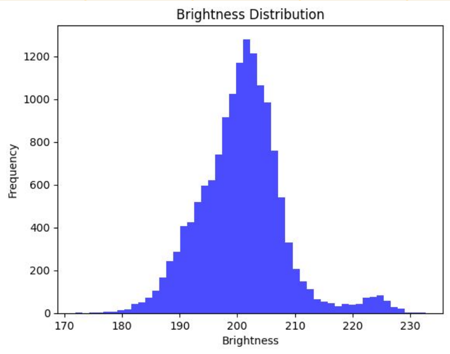
ResNet-50 CNN Modeling and Cross-Validation

Evaluation and One-Sample T-Test

Diagnostics and Validation Work







- After organizing the data, ran a simple count analysis ripening classification and brightness analysis to understand general trends
  - What do different ripening indices visually show?
  - Which ripening indices are most popular?
- Groups were then used to analyze trends and train dataset for accurate classification

Challenges to Analysis

Difficulties	Solutions
1. Size of dataset and computing capabilities	1. Randomizing a smaller sample from the data
2. Differing ripening periods from study due to temperatures	2. Taking random samples from each test group
3. Pictures not previously annotated with ripeness classification	3. Generated code to assign ripeness based on research findings and image file name

Bias and Uncertainty

Phase Imbalance	Labeling Errors	Image Quality
The dataset contains an unequal number of samples for each ripeness phase, leading the model to overfit to the majority class and perform poorly on underrepresented phases.	Ripeness labels may be inconsistently assigned due to subjective human judgment, introducing noise into the training data and reducing model accuracy.	The resizing of images to lower resolutions may obscure critical visual details needed for distinguishing between ripeness phases, particularly for subtle differences.

Results

Performace

Discovered bias toward majority stage (1)

Cross-Validation

Average accuracy (58.82%)

Accuracy

Highest ripeness classification was 66.48% accuracy

Statistical Test

Performance significantly better than random guessing (p-value of 0.0142)

—Conclusion—

- The null hypothesis proposed that the model's accuracy would not exceed random guessing (20% for 5 classes), while the alternative hypothesis suggested that the model would perform better than random.
  - The model achieved a cross-validation accuracy of 58.82% on average, with the highest accuracy reaching 66.48%, significantly outperforming random guessing as confirmed by a one-sample t-test ($p<0.05$).
- Precision, recall, and F1-scores demonstrated strong performance for the dominant class and moderate results for others, indicating the model’s ability to classify ripeness stages effectively in many cases.
- Results highlight promising performance while underscoring opportunities to improve predictions for underrepresented ripeness stages.

Further Research

Other forms of Produce	Larger Image Samples	Prediction of days until expiration
Using similar methods, the project could be expanded to include other forms of produce	The possibility of developing more a more accurate model is entirely achievable with more time and computing power	Using average ripeness windows and ripening periods, could predict time until produce goes bad

References

[1] X. Jing, Y. Wang, D. Li, et al., "Melon ripeness detection by an improved object detection algorithm for resource constrained environments," *Plant Methods*, vol. 20, p. 127, Oct. 2024. Available: https://doi.org/10.1186/s13007-024-01259-3

[2] C. Sun, Y. Chen, X. Qiu, R. Li, and L. You, "MRD-YOLO: A multispectral object detection algorithm for complex road scenes," *Sensors*, vol. 24, no. 10, p. 3222, May 2024. Available: https://doi.org/10.3390/s24103222

[3] A. Zewe, "Forestalling food waste: Student-developed device predicts when an avocado will be ripe," *Harvard John A. Paulson School of Engineering and Applied Sciences*, Jul. 20, 2020. Available: https://seas.harvard.edu/news/2020/07/forestalling-food-waste

[4] Xavier, Pedro; Rodrigues, Pedro; L. M. Silva, Cristina (2024), "'Hass' Avocado Ripening Photographic Dataset", Mendeley Data, V1, doi: 10.17632/3xd9n945v8.1

Questions?