Strawberry Spectroscopy: Calibration Based on the Picked Date

Strawberry Field Trip

Phynx Yamaguchi, Qi Zheng

StellarNet, Inc.

Abstract

Strawberry quality is a major concern for commercial stores, yet determining freshness after transportation is challenging. This study employs UV-VIS spectra to determine the days elapsed since harvest. The methodology involves collecting spectra from multiple strawberries of various cultivars over nine days and training a K-Nearest Neighbors (KNN) machine learning algorithm. Various classification models are evaluated for freshness, considering their accuracy and applicability. While two-class models attain 95.19% accuracy, nine-class and three-class variants offer significant value. The nine-class version furnishes detailed age estimates crucial for scientific research or shelf-life determination, whereas coarser models, like the three-class or two-class versions, simplify decision-making, ideal for applications such as quality control or logistics. This quantitative approach complements qualitative evaluations, meeting diverse marketing needs in the strawberry industry.

Introduction

California and Florida produce the most strawberries in America. California produces over 91% of the entire strawberry supply, and Florida produces most of the strawberry supply in the Winter (Agricultural Marketing Resource Center, 2021). In 2019, Florida produced most of the strawberries from December to March, with its peak being about 80 million pounds produced in February. California still produces some strawberries from December to March, but this production is slightly less than Florida's production. In the months of April to November, California is the main supplier of strawberries with a peak production of 275 million pounds in May, while Florida has no local strawberry harvest from April to November (USDA, 2021). In the months where Florida does not produce their own strawberries, these strawberries are imported from California or Mexico. By the time these strawberries reach the market, they are a couple of days old and can be of poor quality for consumers even if the physical presentation of the strawberry appears normal. This study aims to create a quantitative method to measure the quality of a strawberry harvest without destroying the fruits themselves.

Methods

The most popular strawberry varieties in Florida are FL-127, Florida Medallion, and Florida Brilliance; hence, these cultivars are the focus of this model. The University of Florida Gulf Coast Research Center provided samples for this study, harvesting strawberries directly from their fields. A total of 28 Brilliance, 28 Medallion, and 22 FL-127 strawberries were measured.

A UVN spectrometer was used to measure how the absorbance of the strawberries changed over nine days, indicating the time passed since harvest. The spectra of each strawberry were taken before being lined up 7 cm apart on a metal rack on top of a sheet pan. The strawberries were then refrigerated until the spectra were taken again the next day. This procedure was repeated for all 78 strawberries over nine days. The K-Nearest Neighbors (KNN) algorithm is employed in classification mode, with a 30-70% data split utilized for cross-validation. KNN is a straightforward, non-parametric algorithm used for classification and regression. It works by identifying the 'k' closest data points in the training set to a new data point and assigning the most common class (in classification) or averaging the values (in regression) of these neighbors to the new point.

Figure 1

a. Brilliance strawberry patch   b. Medallion strawberry patch   c. Florida-127 strawberry patch

Results

Several classification calibration models are being created using different combinations of processing steps and wavelengths of the collected spectra. The processing steps experimented with are SDG (Savitzky-Golay Derivative), MC (Mean Center), MSC (Multiplicative Scattering Correction), and SNV (Standard Normal Variate). The experimental results indicate that the freshness of strawberries is determined by the absorbance spectra measured between wavelengths of 740-900 nm, particularly with SDG and MSC processing steps.

Additionally, various calibrations are being created by splitting the picked days of the strawberries into different classes as listed below:

- Nine-class: One day per class
- Three-class: Day 1 to 3, Day 4 to 6, Day 7 to 9
- Two-class 2: Day 1 to 3, Day 4 to 9
- Two-class 1: Day 1 to 4, Day 5 to 9

Creating classification models with different classes is essential for achieving precise predictions and flexible applications. Fine-grained models, like the nine-class version, offer detailed age estimates, beneficial for scientific research or shelf-life determination. Coarser models, like the three-class or two-class versions, simplify decision-making for practical uses such as quality control or logistics. Balancing accuracy and complexity, these models cater to various industry needs, optimizing performance and resource efficiency. In developing these models, we unlock a realm of precision and adaptability essential for today's dynamic markets. Broad classifications aid in quick operational decisions and are more user-friendly, ensuring both accuracy and practical usability. By striking this balance, these classifications not only facilitate swift operational decisions but also resonate with consumers, providing them with clear, intuitive guidance.

The model's accuracy is summarized below:

Figure 2

	Nine-Class classifications Accuracy (%)	Three-class classifications Accuracy (%)	Two-class classifications (Before/at and after Day 3) Accuracy (%)	Two-class classifications (Before/at and after Day 4) Accuracy (%)
All cultivars	**65.95%**	**77.92%**	**87.61%**	**89.25%**
Brilliance	**55.73%**	**84.52%**	**92.26%**	**94.83%**
Medallion	**60.1%**	**88.49%**	**92.86%**	**93.45%**
FL-127	**61.71%**	**85.35%**	**92.93%**	**95.19%**

The table compares the accuracy of classification models using different class splits for strawberry freshness across various cultivars (see Appendix for individual graphs). Nine-class models, predicting the exact day from 1 to 9, have lower accuracy, with all cultivars combined at 65.95% and individual cultivars ranging from 55.73% (Brilliance) to 61.71% (FL-127). Despite their lower accuracy, these models still outperform random guessing, which would be about 11% accurate, thus providing more informed predictions. Three-class models, grouping days into 1-3, 4-6, and 7-9, show higher accuracy, with all cultivars combined at 77.92%, and individual cultivars ranging from 84.52% (Brilliance) to 88.49% (Medallion).

Two-class models, distinguishing between broader day ranges (before/at day 4 and after day 4, and before/at day 3 and after day 3), offer the highest accuracy. For the split before/at day 4 and after day 4, accuracy is 89.25% for all cultivars combined, reaching up to 95.19% for FL-127. For the split before/at day 3 and after day 3, accuracy is 87.61% for all cultivars combined, up to 92.93% for FL-127. Overall, two-class models provide the most accurate predictions, making them particularly effective for practical applications due to their balance between simplicity and accuracy. Models combining all cultivars tend to have lower accuracy than those for individual cultivars, likely due to increased variability and differences in characteristics among the different cultivars.

References

International Fresh Produce Association, et al. “U.S. Strawberry Market Report: 2021-2022.” freshproduce.com, 2022, pp. 1–6. www.qa.freshproduce.com/siteassets/files/reports/retail/commodities/23-commodityreport-strawberry.pdf.

“Strawberries.” Copyright 2024 Agricultural Marketing Resource Center. All Rights Reserved., www.agmrc.org/commodities-products/fruits/strawberries.

“Strawberries - Food Source Information.” Food Source Information, 23 May 2024, www.chhs.colostate.edu/fsi/food-articles/produce-2/strawberries-2.

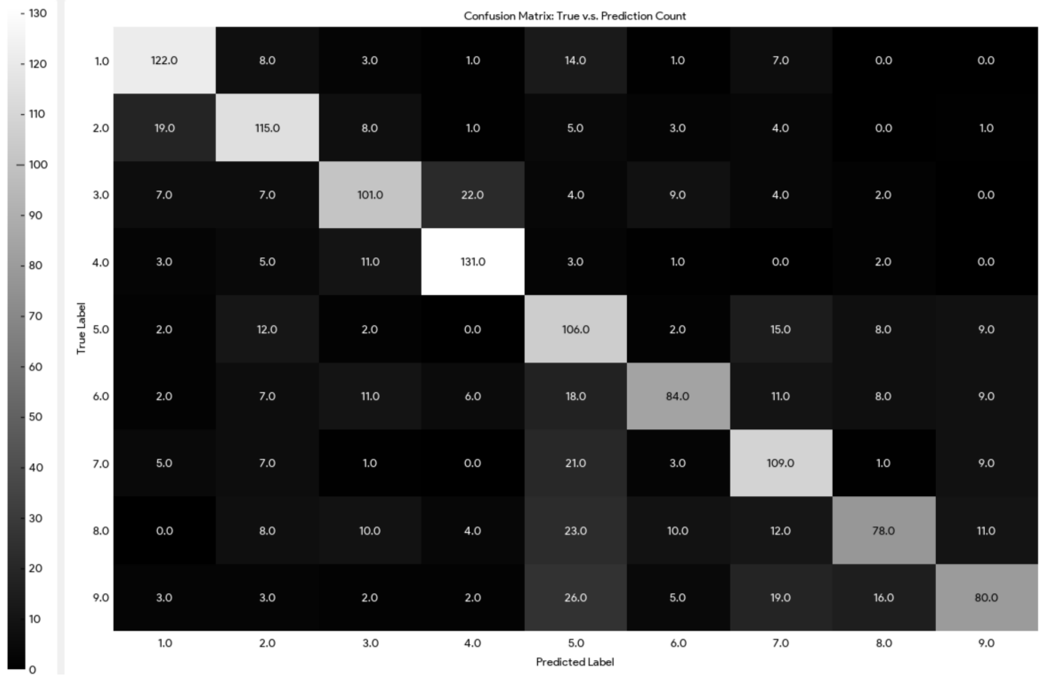
U.S. Fresh Strawberry Production Expands With Newer Varieties. www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=101156.

Appendix

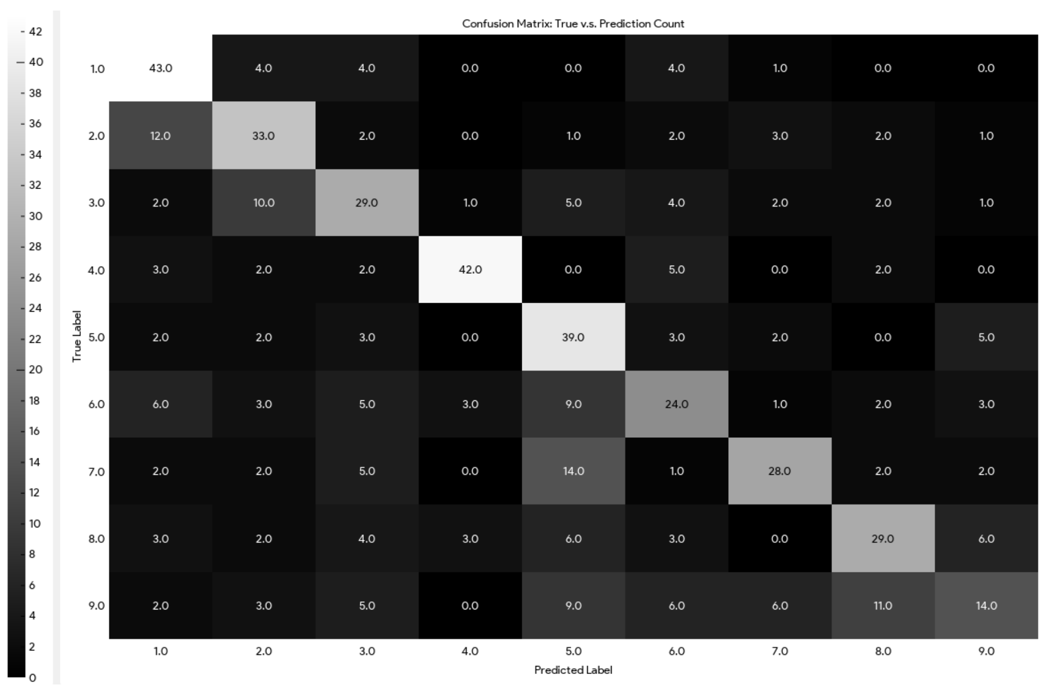
This Appendix Consists of Supplementary Data Relevant to Figure 2.

The graphs below show the Nine-Class classifications Accuracy (%) column in Figure 2.

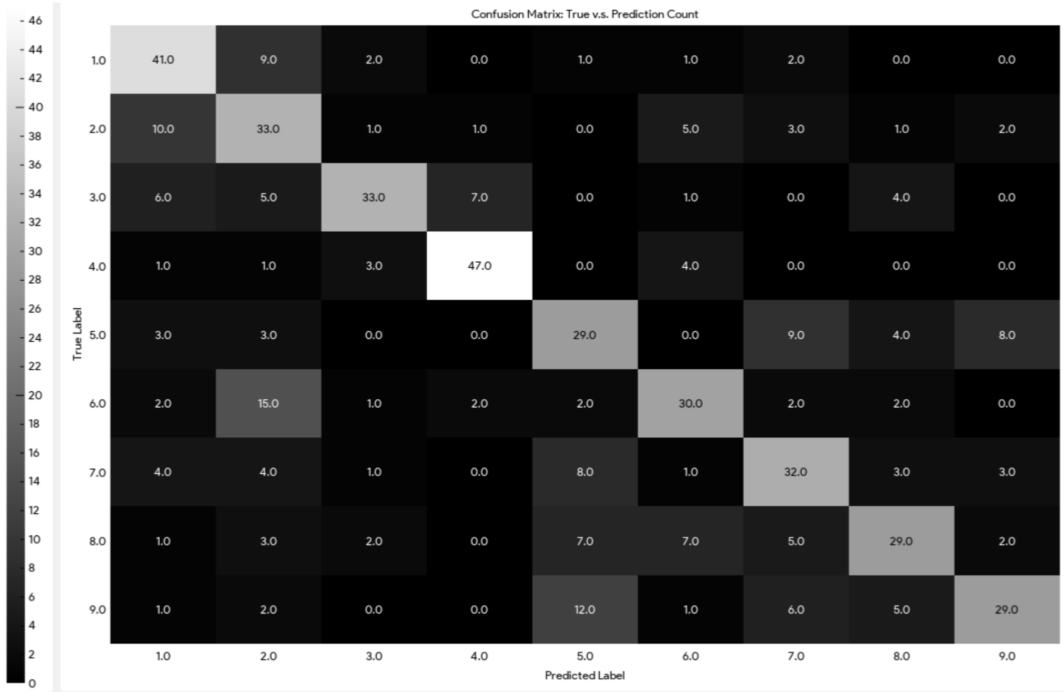
All cultivars:



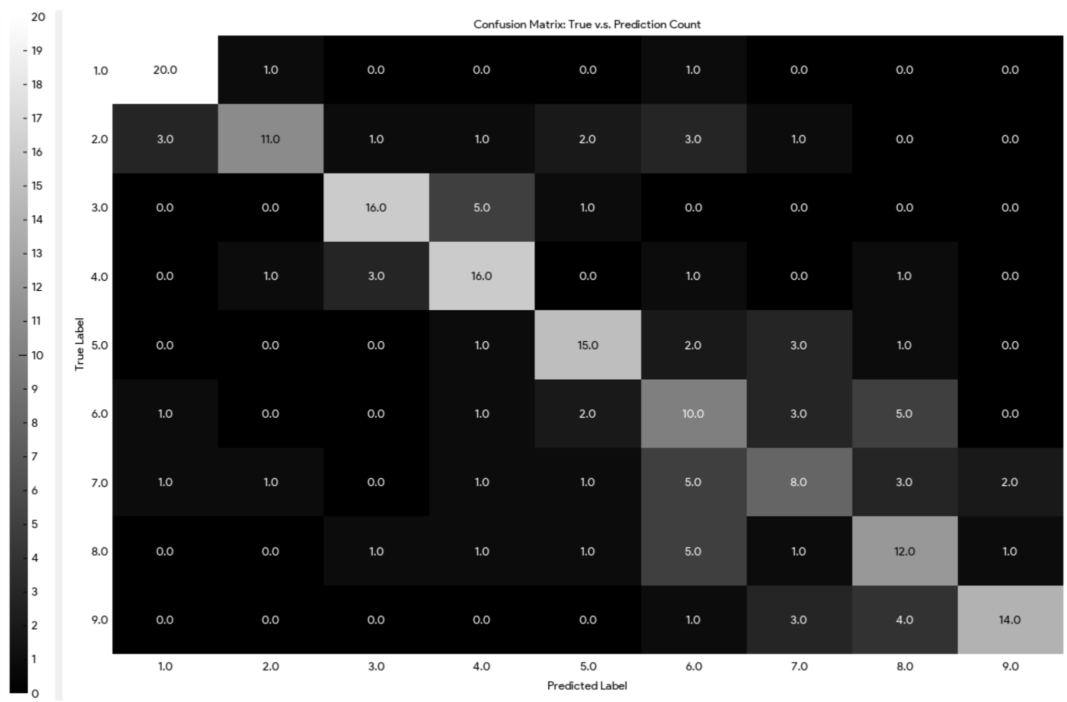
Brilliance:



Medallion:

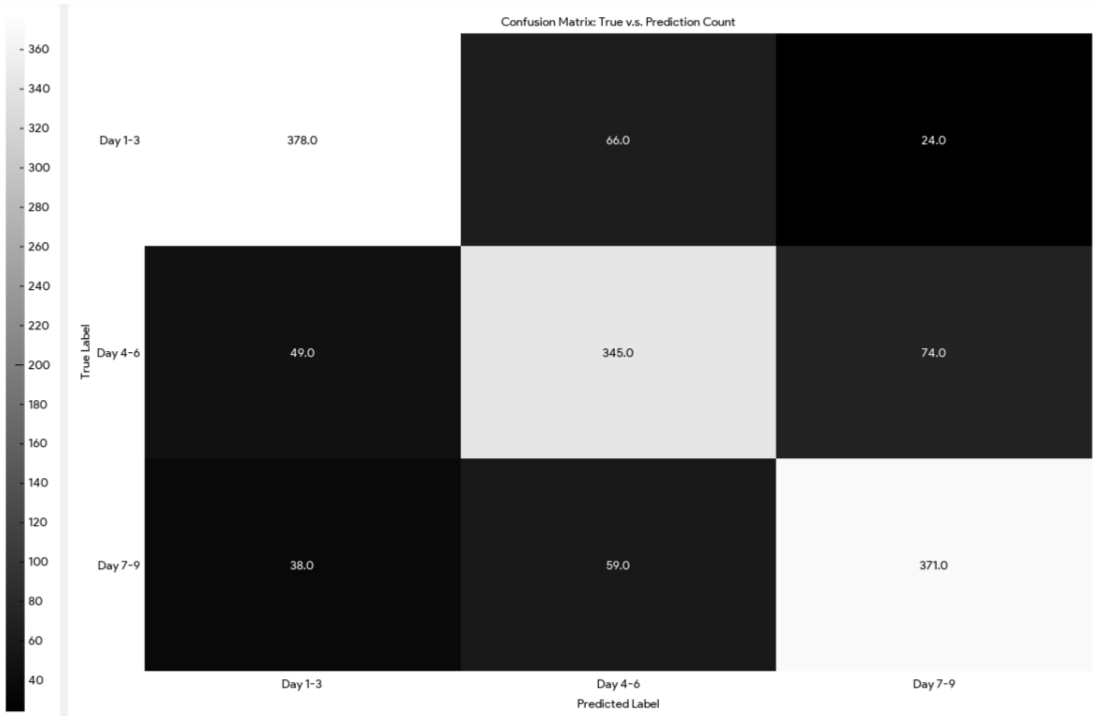


FL-127:

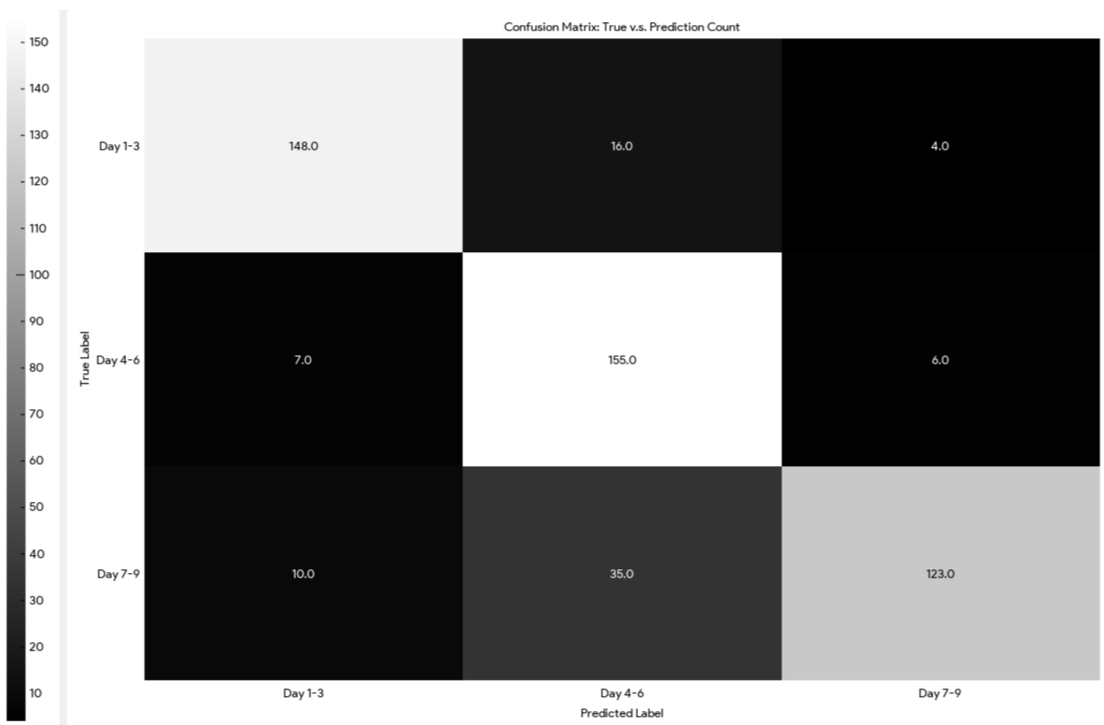


The graphs below show the Three-Class classifications Accuracy (%) column in Figure 2.

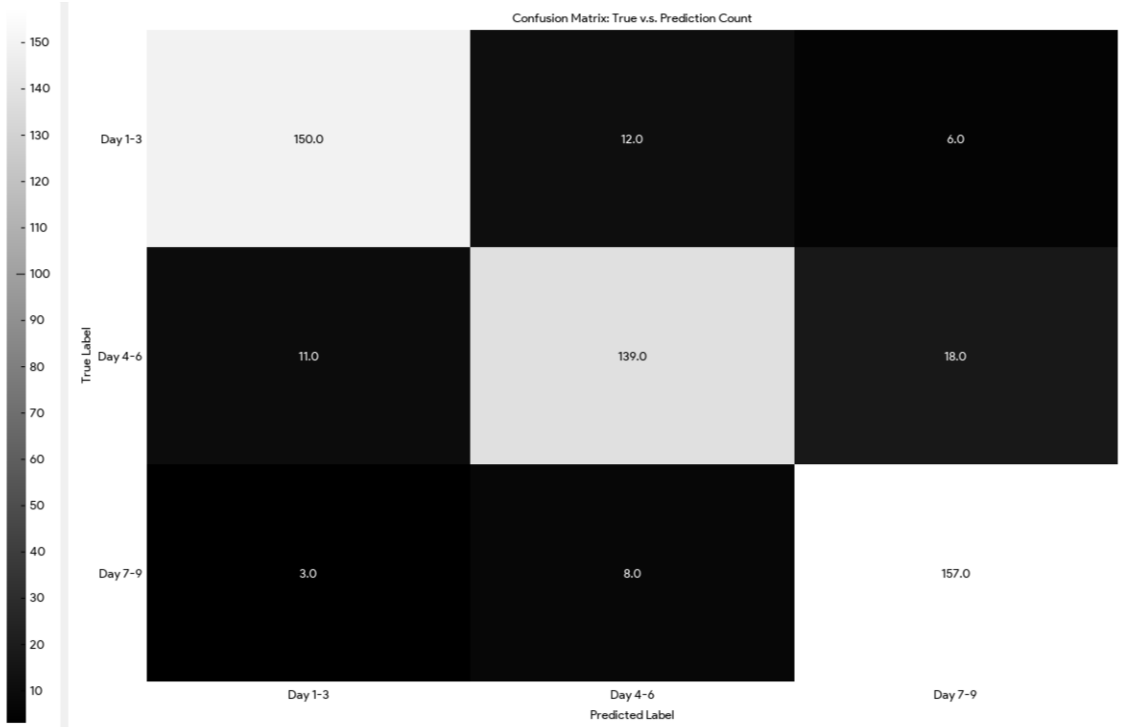
All cultivars:



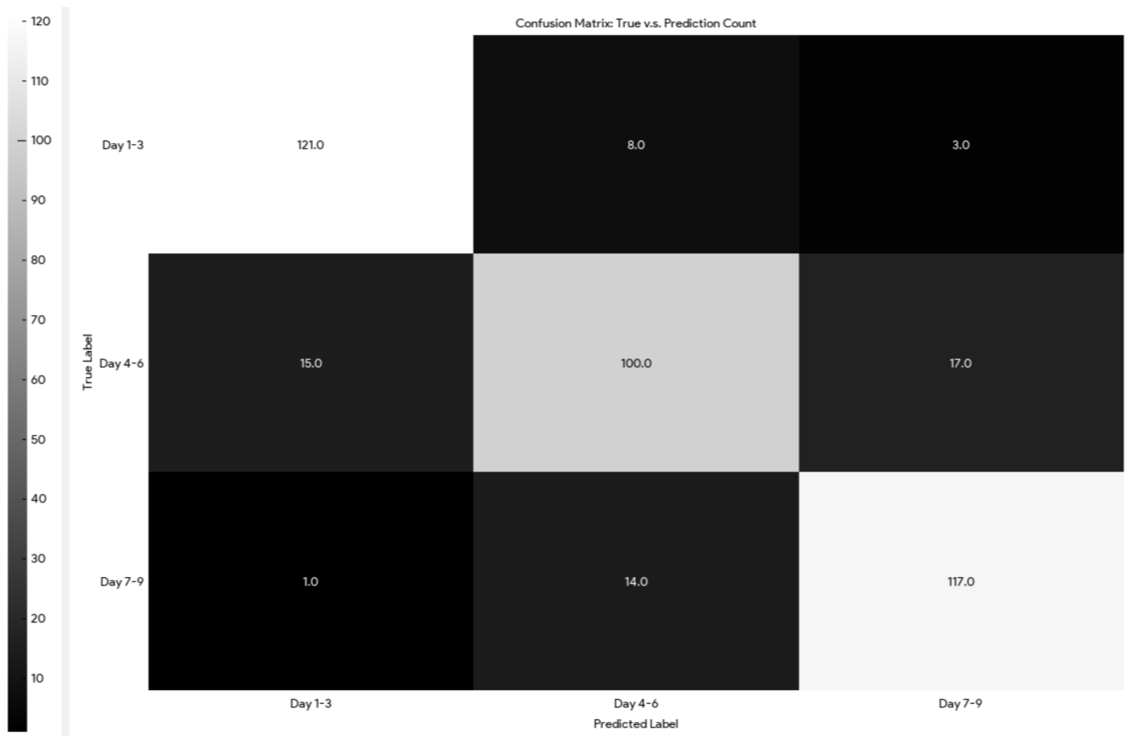
Brilliance:



Medallion:

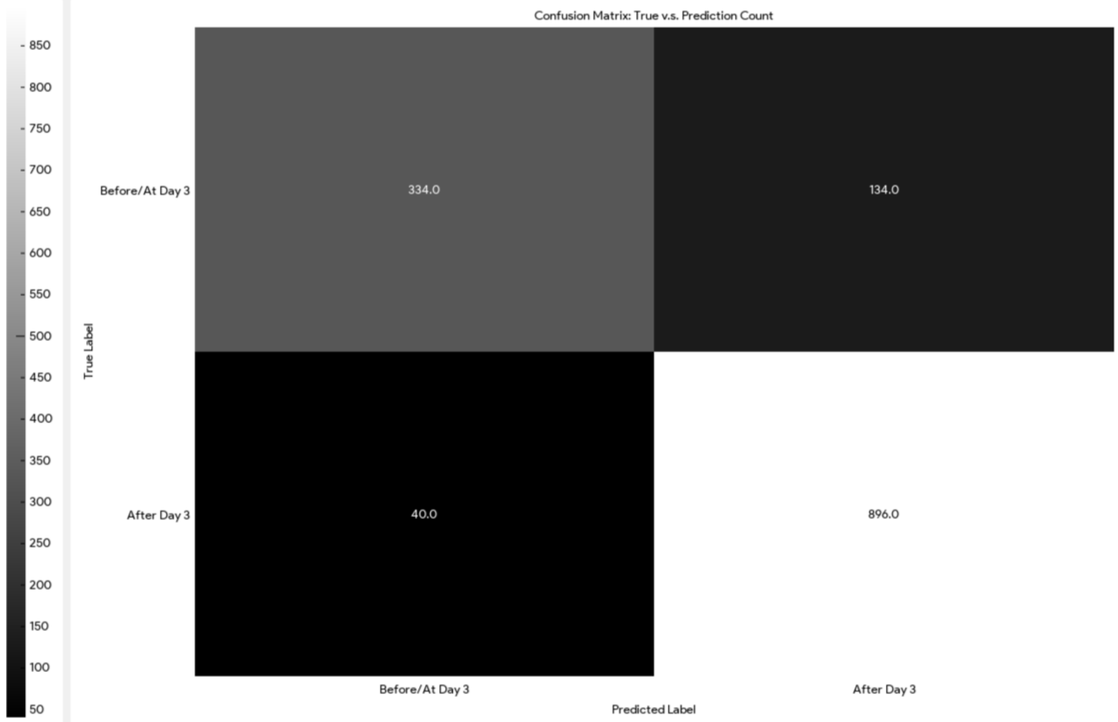


FL-127:

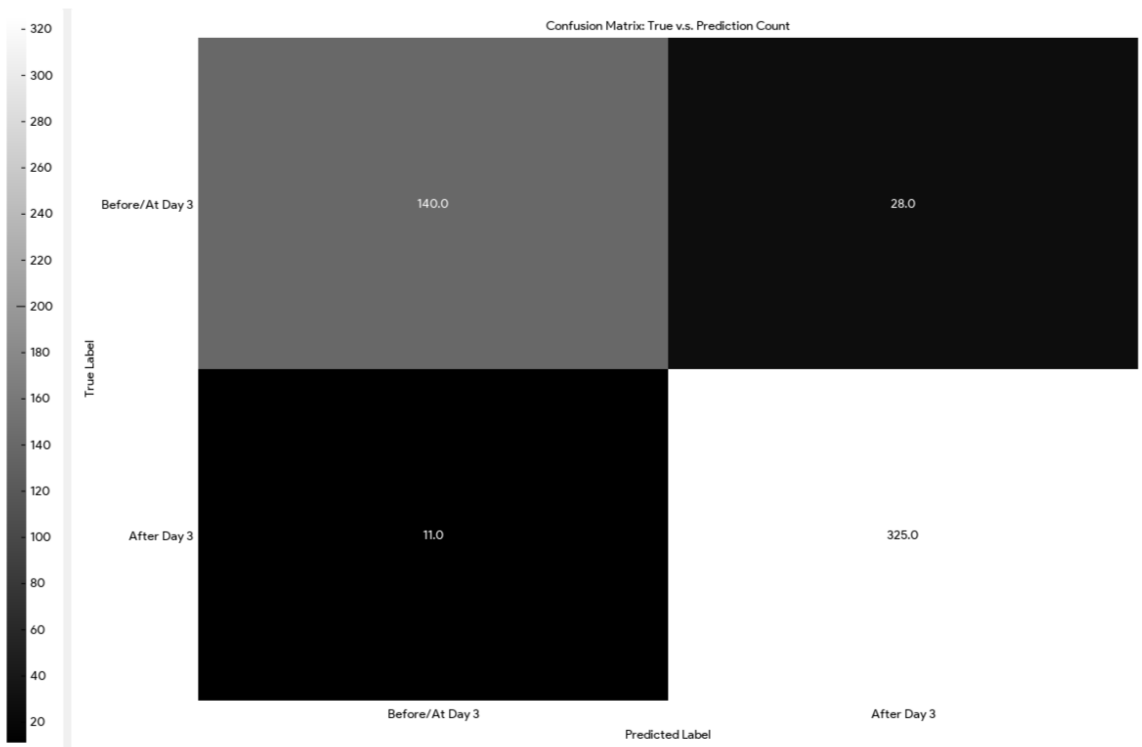


The graphs below show the Two-class classifications (Before/at and after Day 3) Accuracy (%) column in Figure 2.

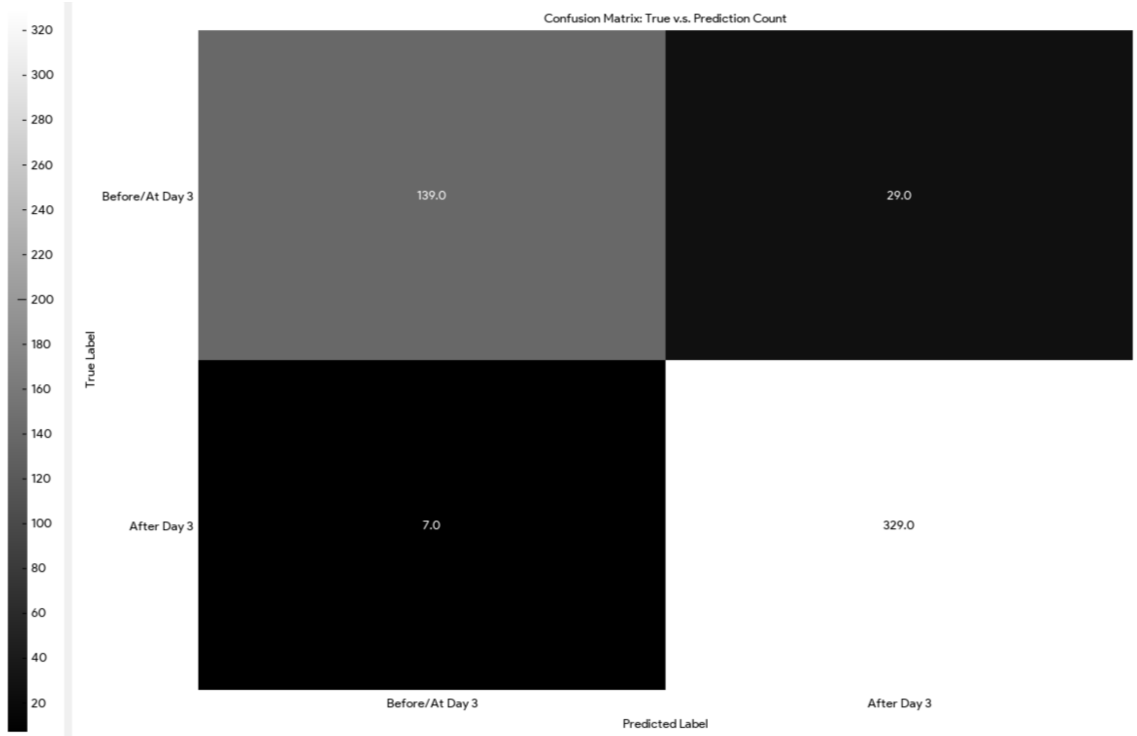
All cultivars:



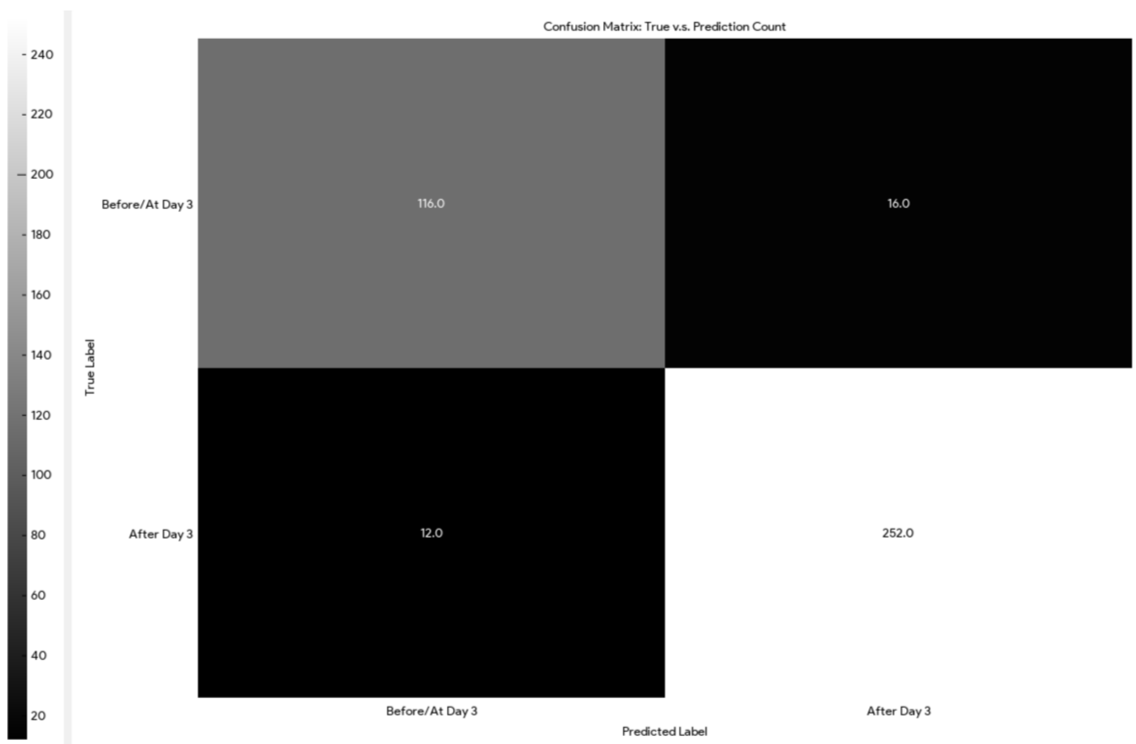
Brilliance:



Medallion:

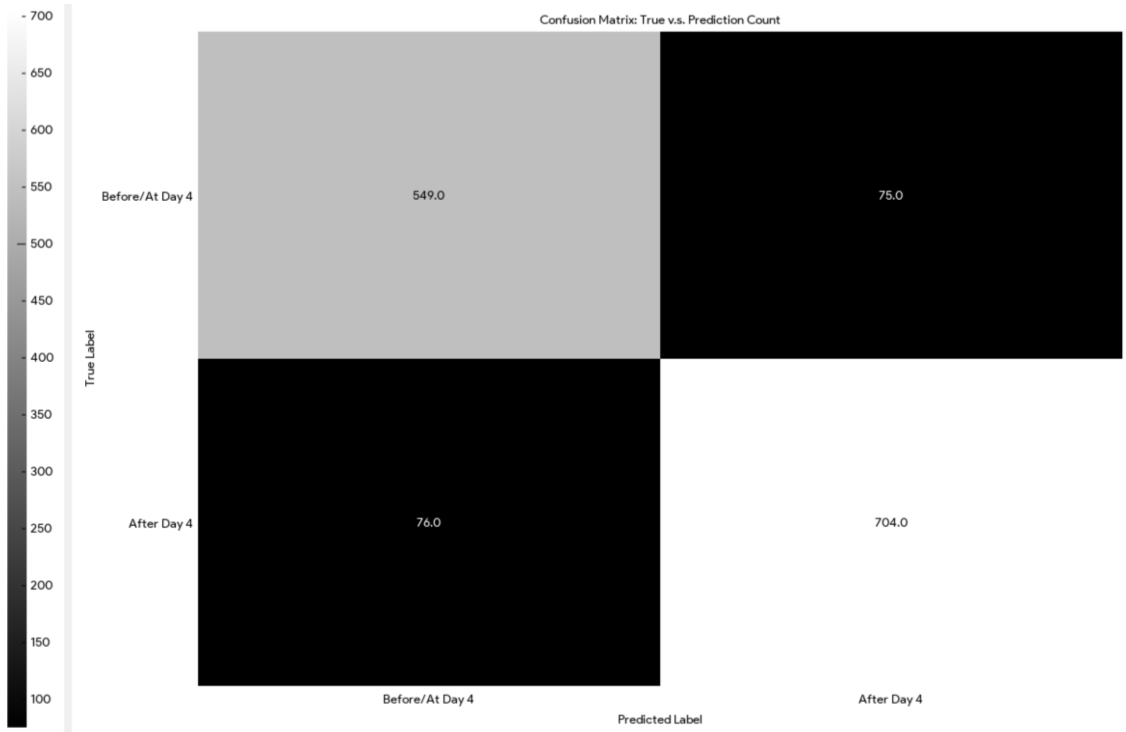


FL-127:

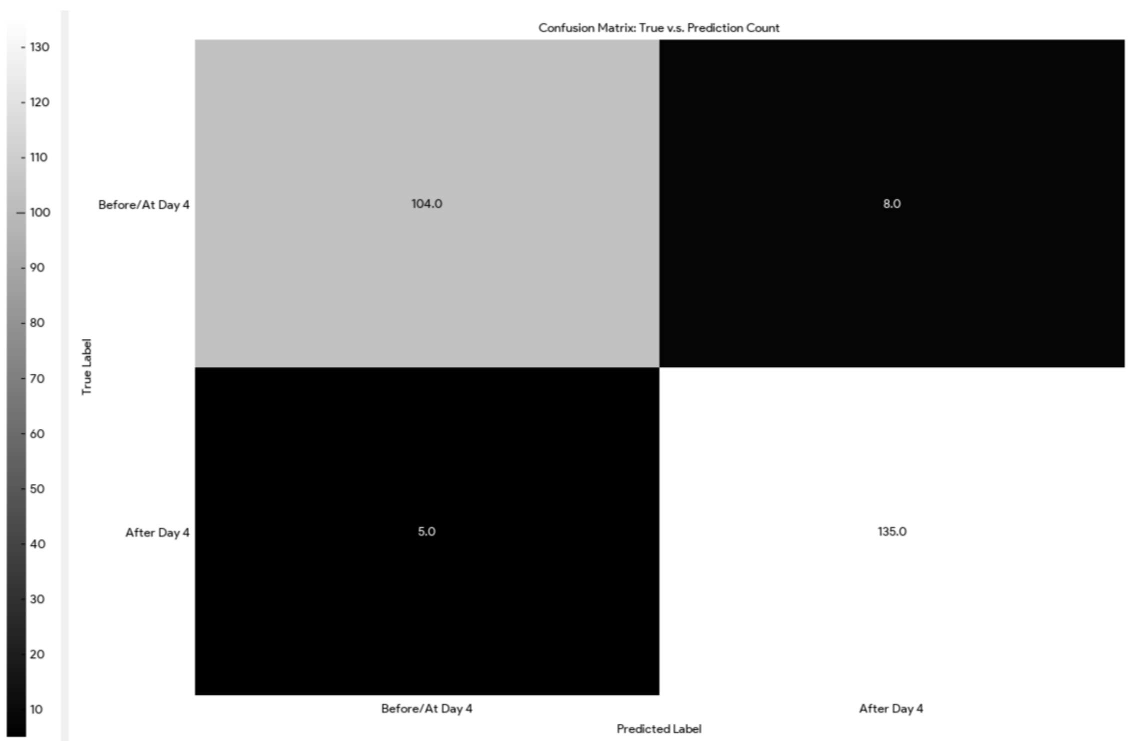


The graphs below show the Two-class classifications (Before/at and after Day 4) Accuracy (%) column in Figure 2.

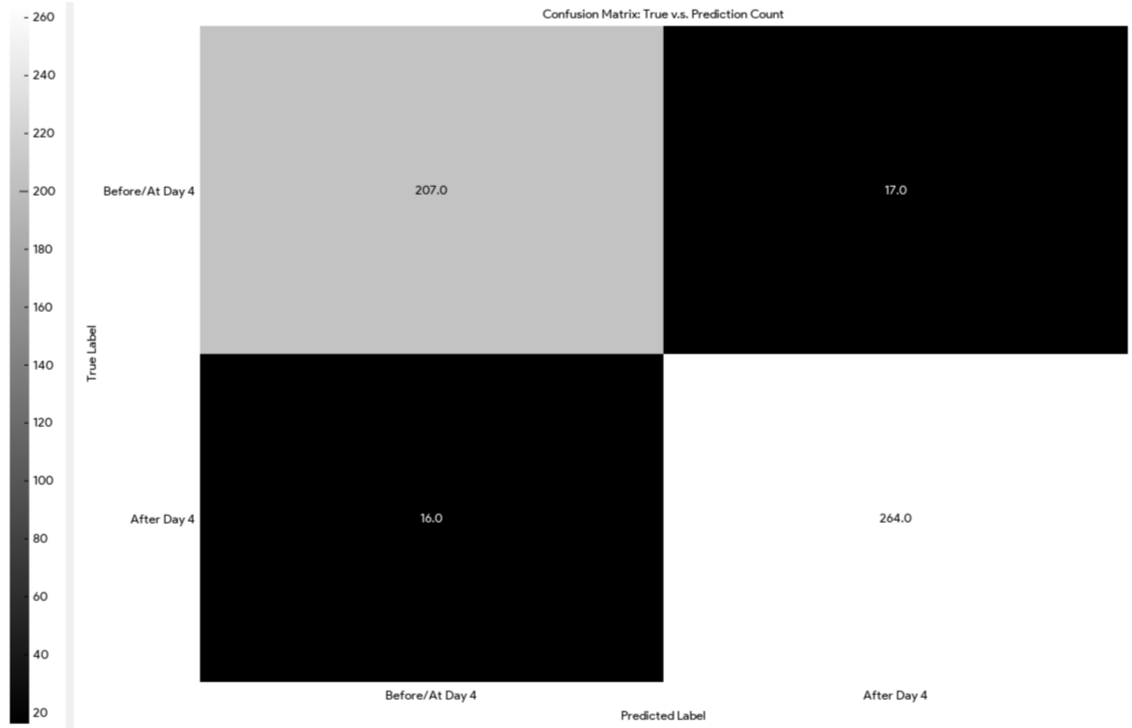
All cultivars:



Brilliance:



Medallion:



FL-127: