The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan

Apple Weight Prediction based on Lifecycle Growth Information

Chang-Hyeok Park[1], Gang-In Je[1], Ho-Jun Kwon[1], Ye-Seong Kang[2], Chan-Seok Ryu[1*]

[1] Department of Bio-system Engineering, Institute of Agriculture & Life Science, GyeongSang National University, Jinju 52828, Rep. Korea
[2] Department of Smart Agro-industry, Institute of Agriculture & Life Science, GyeongSang National University, Jinju 52725, Rep. Korea

*Corresponding Author: ryucs@gnu.ac.kr

Abstract

It is essential for determining optimal harvest timing, the grade of quality, and fresh maintenance, all of which directly impact growers' economic returns, to accurately predict individual apple fruit weight.
This study aims to predict the fruit weight of Fuji apples at the main branch level (n = 126) using growth data collected throughout the growing season.
Fuji apples were monitored at 23 orchards in 2022 and 2023, and at 24 orchards in 2024. Growth data were collected across the entire growing season, including tree height, canopy width, trunk diameter, number of main(primary) branches, shoot length, main(primary) branch length, main branch diameter, and fruit weight. Training and validation datasets were based on the growth data in 2022 and 2023. The test dataset based on it in 2024 was reserved exclusively for testing.
A random forest regression (RFR) model with five-fold cross-validation was selected as the best machine learning algorithm among the other models, such as RASSO, MLR, and KNR. The performance of a model was 0.852, 19.05 g, and 4.72% in calibration, and 0.685, 42.91 g, and 15.53% in validation in the order of R2, RMSE, and MAPE.
Based on the testing, the performance was 42.20 g of RMSE and 11.86% of MAPE, indicating that the prediction error remained within the validation range. Tree height was selected as the most influential predictor in the main branch-level fruit weight estimation model in the feature importance analysis. It might be necessary to improve the performance of the main branch-level fruit weight estimation model based on different conditions, such as weather conditions, sea level height, management factors, and so on.

INTRODUCTION

Accurate growth information collected throughout the entire lifecycle of apples, from flowering to harvest, is essential for determining fruit quality, harvest timing, and storability. Machine learning offers a powerful approach for capturing complex linear and nonlinear growth relationships, thereby improving prediction accuracy. This study aims to develop and validate a predictive model for average fruit weight of 'Fuji' apples at the main branch level using multi-year growth and fruit weight data.

MATERIALS AND METHODS

Data acquisition

This study was conducted over a three-year period from 2022 to 2024, involving branch-level growth surveys across 47 orchards. The survey encompassed major growth stages from flowering to harvest, measuring tree height, canopy width, trunk diameter, number of primary branches, shoot length, primary branch length, and primary branch diameter. At harvest, the average fruit weight per main branch was recorded.

Statistical analysis

Prior to regression analysis, outliers were removed using the interquartile range (IQR) method. Variable selection was performed using the stepwise selection procedure to identify predictors with significant contributions to the model's explanatory power. To ensure the generalization performance of the model, K-fold cross-validation (k = 5) was conducted.

Regression analysis

Machine learning analyses were conducted by comparing four models—multiple linear regression (MLR), k-nearest neighbor regression (KNR), Lasso regression, and random forest regression (RFR). The datasets from 2022 and 2023 were used for model calibration and validation, while the 2024 dataset was employed as an independent test set.

RESULTS & DISCUSSION

The comparison of models revealed that the RFR achieved the highest predictive performance and was therefore selected as the final model. In the calibration stage, the model achieved an R^2 of 0.852, with an MAPE of 4.72% and an RMSE of 19.05 g. During validation, the performance decreased to an R^2 of 0.685, with an MAPE of 15.53% and an RMSE of 42.91 g. When applied to the independent test dataset from 2024, the model yielded an R^2 of 0.328, an MAPE of 11.86%, and an RMSE of 42.20 g. Although the RMSE and MAPE values were comparable between validation and test datasets, indicating stable error levels, the R^2 dropped markedly in the test set relative to validation. This decline suggests that environmental or management conditions in 2024 differed from those represented in the training data, thereby reducing the explanatory power of the model. Feature importance analysis indicated that tree height was the most influential predictor, which aligns with previous studies reporting that tree height strongly affects tree growth and fruit quality.

CONCLUSIONS

This study developed and validated a machine learning model to predict the average fruit weight of 'Fuji' apples at the main branch level using multi-year growth data. The random forest regression (RFR) model showed the best performance, and tree height was identified as the most influential predictor. However, the test R^2 in 2024 was relatively low, and it is considered that future studies could improve the model's generalization performance by expanding the dataset and incorporating environmental variables.

ACKNOWLEDGEMENTS

Acknowledgements: This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. RS-2024-00398963)" Rural Development Administration, Republic of Korea.