Mapping the variability to develop a decision for end quality of fresh horticultural product upon arrival in the retail by machine learning based techniques

TURAN D. (1), KNOTT M. (2), DEFRAEYE T. (2), FOGLIANO V. (1)
1 Food Quality and Design Group, Department of Agrotechnology and Food Sciences, Wageningen University, P.O. Box 17, 6700 AA , Wageningen, Netherlands
2 Empa, Swiss Federal Laboratories for Material Science and Technology, Laboratory for Biomimetic Membranes and Textiles, Lerchenfeldstrasse 5, CH-9014, St. Gallen, Switzerland

Food supply chains are confronted with increased consumer demands on food quality, safety and sustainability. On the other hand, one third of the globally produced food for human (more than 42% accounts for fruit and vegetables) gets lost or is waste every year. The current strategy to optimize postharvest supply chains is intensive in-transit time-temperature monitoring with quality evaluations when the shipments are transferred between stakeholders. That way, extensive amounts of hygrothermal sensor data and metadata (e.g. production area, variety, transit duration etc.) are gathered for each shipment. These data are however not used to their full potential to predict quality upon arrival at retail which is essential for several stakeholders since the quality evolution of the fresh horticultural produce is unique and differs between individual fruit due to pre-harvest biological variability between individual fruit at harvest and postharvest variations in hygrothermal conditions between refrigerated shipments. This drives our mission towards achieving higher fresh-produce quality with less variability, to reduce food losses and the associated environmental impact. Artificial intelligence techniques, such as machine learning (ML) models can be used as a valuable tool for aiding decision makers due to its ability to learn and find interesting patterns in data. However, the predictive accuracy of machine learning strongly depends on the quality and quantity of the respective training data and the choice of specific features. Here, our study aims to identify multiple correlations and derive the feature importance quantitatively on citrus shipment metadata and temperature sensor data to develop ML pipeline predicting fruit quality. For this, multiple ML algorithms are examined and their applications for the citrus supply chain processes are explored. To this end, we investigated supervised classifiers, logistic regression (LR) and random forest (RF), which are used to classify good or bad quality of products upon arrival at the market. LR classifier showed better classification accuracy with 69% than RF. Citrus end quality decision was mostly correlated with chilling injury defect (28%, the highest among other postharvest defects). Variability among product variety and production area have the largest effect on the model to predict good or bad arrival.