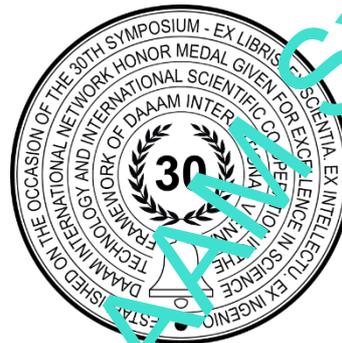# PREDICTING PEACH FRUIT RIPENESS USING EXPLAINABLE MACHINE LEARNING

Dejan Ljubobratović, Guoxiang Zhang, Marija Brkić Bakarić, Tomislav Jemrić & Maja Matetić

Abstract

Predicting fruit ripeness allows us to choose the optimal time to harvest. The parameter by which peach ripeness is commonly represented is its firmness. As traditional methods for determining firmness of peaches are destructive, this paper uses an alternative method for determining peach ripeness which is based on peach impedance, as recommended by the domain expert.
The data set on which the data analysis is performed contains measurements obtained from a couple of hundred fruit measurements, which also include peach impedance. In our data analysis, we use one of the high accuracy machine learning models, which are called black box models and which are characterized by low interpretability. The paper presents the results of applying a black box type machine learning method, as well as methods for interpreting black box models which facilitate understanding of the model behavior for domain experts, i.e. Variable importance, Tree Surrogate, Local Interpretable Model-Agnostic Explanations and Break Down.

Keywords: interpretability; explainable machine learning; predicting fruit ripeness; peach impedance

1. Introduction

Using machine learning in agriculture, especially in fruit growing, is not a novelty. Large and complex agricultural systems that include the crop management, harvesting, packaging and distribution of fruit represent a great area for upgrading and refining the process using machine learning.

There are several different areas in agriculture in which machine learning can be used for production improvement: crop management, livestock management, water management and soil management [1]. Reducing food losses in the fruit and vegetable supply chain [2] is also one of the challenging tasks for machine learning. This paper plays a part in it.

Determining fruit ripeness in fast and simple way can improve food production and decrease transportation costs associated with fruit decay. This is of significant importance for perishable crops such as peach which is characterized by rapid change of post-harvest quality due to its intensive climacteric respiration regulated by ethylene [3] [4]. The most important change occurring during post- harvest is rapid softening which makes fruit prone to decay and complete loss of quality pm short period of time. Therefore, firmness is one of the most important quality parameters for peaches. A

couple of hundred of peaches are analysed and the impedance is measured with the aim of building a model for predicting fruit ripeness based on measured parameters [5]. The parameters obtained for each fruit are the following: height, width, volume, mass, density, firmness, soluble solids concentration (SSC), titratable acidity (TA), and Zs and theta which represent the measure of impedance of each fruit. The aim is to find a model that will predict the ripeness of the fruit from the measured values. Firmness of the peach is taken to represent its ripeness, and it is obtained as the average of four opposite peach firmness measurements taken at equatorial side of the fruit, using penetrometer with 8 mm probe.

Measuring the electrical impedance of fruits is a known non-invasive method of determining fruit ripeness [6] [7] and we use it in this research. The measurement process can be performed a number of times on the same fruits. Fruit rot can also be prevented by timely detection of changes in environmental parameters during transport [8][9]. Predictive models also play an important role here. The implementation of a functional model would be greatly beneficial in cold chain fruit distribution because the described method represents a non-invasive way of determining ripeness, as opposed to the classic firmness measurement where the peach is destroyed by the measuring process.

In the first part of the paper, a brief overview of the papers presenting approaches which deal with predicting fruit ripeness using machine learning is given. Dataset description and data exploratory are represented afterwards. In the second part of the paper, a Random Forest model (RF model) for predicting peach ripeness is built. Several methods, such as Variable importance, Tree Surrogate, Local Interpretable Model-Agnostic Explanations, and Break Down are used in order to explain and interpret the model.

In this research, R language (ver 4.0) is used for all the calculations, data exploration, predictions and explanations.

2. Explainable Machine Learning

Although the emphasis of the paper is on measuring and predicting peach ripeness, it is extremely important to explain the obtained model prediction. The algorithm used in creating the prediction model is the so-called *black box* algorithm. Black box models are usually reliable and precise in prediction, but their internal mechanisms are unknown, or at best very difficult to explain. To override the lack of transparency behind their behaviour, various techniques are used to help explaining the results of such models [10], and in this paper we use several of them.

In order to achieve complete reliability and assessment of the ethical and moral standards of the machine decisions, detailed "explanations" of AI decisions are needed [11]. Explanation, according to some authors [12], is a prerequisite to ensure the scientific value of the results. Explainable Machine Learning (Explanatory Artificial Intelligence, XAI) provides explanations of black box model decisions with a certain level of detail, making systems more understandable and transparent [13]. Such techniques are often referred to as interpretable machine learning.

Although many authors equate Explainable Machine Learning and Interpretable Machine Learning, some authors clearly separate these two concepts [14] [13]. According to [13], explainability in models is able to summarize the reasons for black-box behaviour, gain the trust of users, and produce insights about the causes of their decisions. Interpretability, by the same authors, is just a first step in a wider concept called explainability. They claim that all explainable models are interpretable by default, while the reverse is not always true.

3. Related work

According to some authors, almost one third of fresh produce harvested worldwide is lost in the distribution between production and consumption [15]. Due to world population growth, it is necessary to increase food production and decrease food loss. To reduce postharvest food losses, it is important to apply all current knowledge to improve the food handling systems, especially packaging and cold chain maintenance [16]. One of the important factors of fruit rot are oscillations in temperatures during transport [17]. To assess fruit ripeness, researchers often use destructive techniques which destroy fruit during measurement. Therefore, in order to avoid losses during measurement, non-destructive methods of assessing fruit ripeness have been investigated. Near infrared spectroscopy (NIRS) model is used to predict mango quality in a non-destructive way [18][19][20]. Fruit ripeness can in some cases be determined by its visual appearance, i.e. red apples, strawberries or cherries. In that case, computer vision is used in the evaluation of the ripening stages [21] [22].

One of the non-invasive methods of determining fruit ripeness is also measuring the electrical impedance of fruits [6] [7]. Researchers in [23] [24] use electrical impedance to measure peach ripeness and to compare chilling-injured fruits with the fruits ripened after harvesting. In both cases peach impedance decreases. They consider the impedance to be associated with cytoplasmic or membrane electrical resistance [23] [24]. Electrical impedance is also used in assessing the quality of other fruits and vegetables [25].

Although the emphasis is put on non-destructive methods, an important component of such research is the automation of the whole process. While ripe fruits once used to be selected by hand, nowadays it is important to use measurements to automate the process, and with the help of machine learning, to speed up and improve the process of prediction and selection. The main idea is not to eliminate people from the work process, but to speed up simple but time consuming tasks [26][27].

4. Data set description

For the purpose of this research *height*, *width*, *volume*, *mass*, *firmness*, *SSC*, *TA*, *Zs* and *theta* are measured on over two hundred peaches. The pre-processing of the data is guided by the expert, which results in the exclusion of *height*, *width*, *volume* and *mass* variables from the data set since they are used in the calculation of *density*.

This research uses variables: *density*, *SSC* (soluble solids concentration), *TA* (titratable acidity), *Zs*, *theta* and *firmness* which represents *fruit ripeness* (Table 1).

Variable	Description
Density	peach density (calculated from mass and volume)
SSC	soluble solids concentration
TA	titratable acidity
Zs	magnitude of Z (real part of impedance)
Theta (θ)	phase angle between the current and the voltage
Firmness	pach firmness

Table 1. Variables used in this research

Soluble solids concentration (SSC%, oBrix) can be determined from a small sample of fruit juice using refractometer. Sugars are the major soluble solids in fruit juice, so SSC is used as a measure of fruit sweetness. Titratable acidity (TA) represents an acidity measurement using titration of fruit juice with NaOH and good SSC/TA ratio is important for consumer acceptance of peach fruit.

We also measure the impedance of the fruits. Impedance extends the concept of resistance to AC circuits. It possesses magnitude and phase, unlike resistance, which has only magnitude. In case of direct current (DC), there is no distinction between impedance and resistance. Resistance can be thought of as impedance with zero phase angle (shift). Impedance is a complex number, with the SI unit ohm (Ω) same as resistance. Its symbol is usually Z, and it may be represented by writing its magnitude (Zs) and phase (theta) in the form |Z|∠θ [20]. Firmness is obtained as the average of four opposite peach firmness measurements taken at equatorial side of the fruit using penetrometer with 8 mm probe.

5. Data exploration

Our data set consists of 200 observations with altogether six variables predicting peach firmness based on five predictors (*density*, *SSC*, *TA*, *Zs* and *theta*).

By looking at the correlation graph (Fig. 1), one can notice high correlation between firmness and two impedance parameters (Zs and Theta). The correlation between *Zs* and *Theta* variables is expected, since these are two different values of the same measurement. Figure 2 shows correlation between *firmness* (that represents peach ripeness) and *Theta* (that represents peach electrical impedance). We will not take into account the mutual correlation of the variables *Zs* and *Theta* since these are two impedance features and are thus related to each other.

The distribution of variable *firmness* (Fig. 3) is bimodal, so we use K-Means Cluster Analysis to split it into two classes and make our prediction binary (Fig. 4).
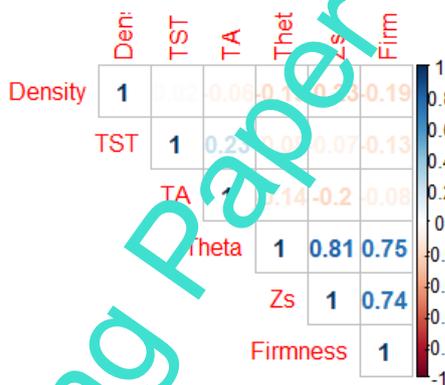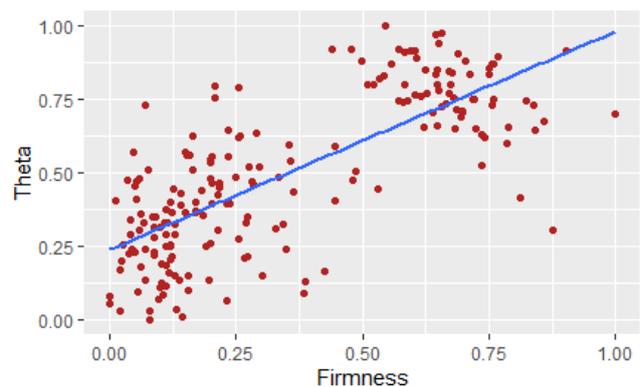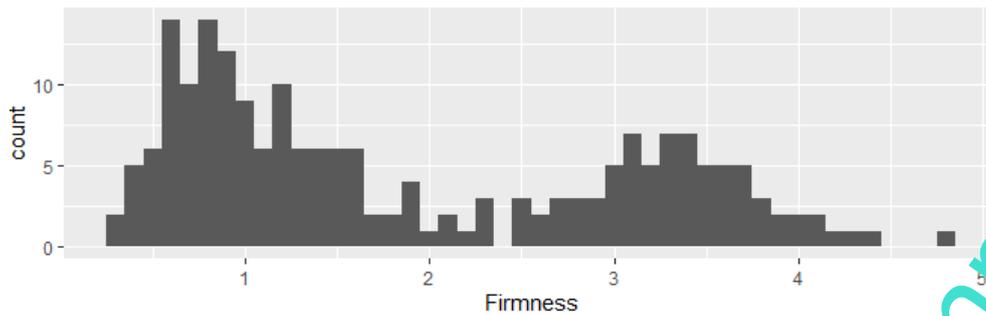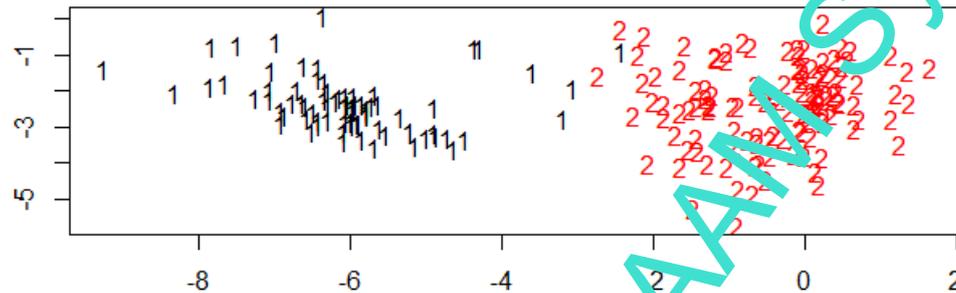

Fig. 1. Correlation between variables

Fig. 2. Correlation between *Theta* and *Firmness*

Fig. 3. Distribution of variable *Firmness*

Fig. 4. K-Means Cluster Analysis split of variable *Firmness*

6. Building a prediction model using Random forest algorithm

Random Forest is a machine learning algorithm based on decision trees, which is used for supervised learning for both classification and regression tasks [29]. After the set is divided into two subsets (training set 80%, testing set 20%) and the data is normalised, a random forest model is built on the training data. As the training control, three-fold cross-validation repeated five times is used.

Model accuracy on the test data is 97.5% and, as evident from the confusion matrix (Fig. 5), only one "false positive" prediction is made.

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 13  0
         1  1 26

              Accuracy : 0.975
                95% CI : (0.8684, 0.9994)
    No Information Rate : 0.65
    P-Value [Acc > NIR] : 7.404e-07
                 Kappa : 0.9441
```

Fig. 5. Confusion matrix

So far, we built the model which is based on several predictors and which predicts fruit ripening with great accuracy. However, the mechanism behind these results is unknown due to the very nature of *black box* algorithms such as random forest.

7. Interpreting the results

To override the lack of transparency in our *black box* model behaviour, we use various techniques to explain the results of the model. We are most interested in the contribution of variables in the decision-making process.

7.1 Variable importance

Variables of greater importance have a significant impact on outcome values, while, on the other hand, variables of low importance can be omitted from the model and make it simpler and faster to set up and predict.

Using variable importance technique on our model reveals that the variables *Theta* and *Zs* have the greatest influence on predicting peach firmness (ripeness) (Fig. 6).
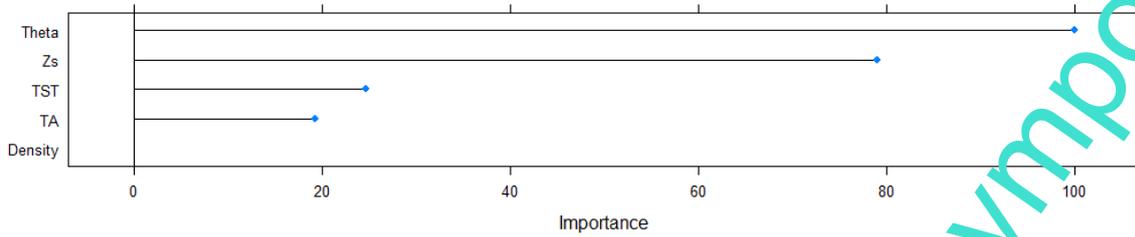

Fig. 6. Variable importance of Random forest model

7.2 Tree surrogate

A surrogate model is a technique in which a generally simpler model is used to explain another usually more complex model or, in our case, the *black box* model [30].

In this research, the tree surrogate method uses decision trees on the predictions where conditional inference tree is fitted on the predictions from the machine learning model and data.

The R-squared value (variance explained) of the surrogate model is 0.73, which means it approximates the underlying black box model behaviour quite well, but not perfectly.

Again, the variables *Theta* and *Zs* prove to be most significant (Fig. 7).
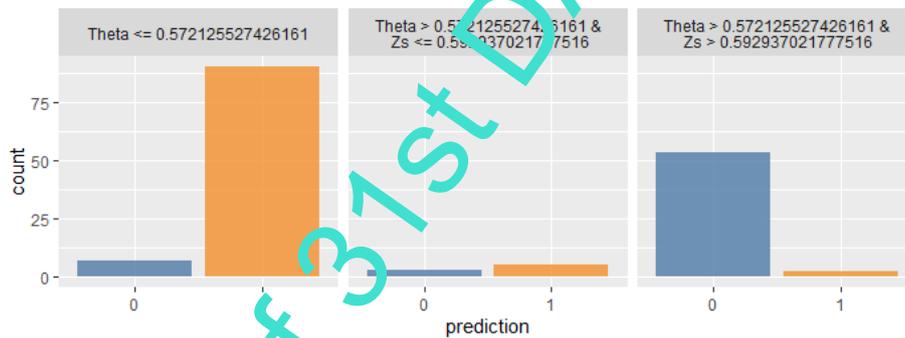

Fig. 7. Tree surrogate results

7.3 Break Down

The *Break down* algorithm is a model agnostic tool that decomposes predictions of *black box* models. *Break down* shows the contribution of each variable to a final prediction. It is like *Varible importance* but for every single prediction.

If we go through a few randomly selected predictions and look at the variables' contributions, we can see that the results are the same as with the *Variable importance* method revealing that *Theta* and *Zs* are the most important predictors (Table 2).

	contribution		contribution		contribution
(Intercept)	0.607	(Intercept)	0.607	(Intercept)	0.607
- Zs=0.826957763549147	-0.235	- Theta=0.920556434599156	-0.326	- Theta=0.364451476793249	0.225
- Theta=0.75450481940928	-0.281	- Zs=0.672395526780459	-0.202	- Zs=0.332195409064155	0.128
- TA=0.434968017751569	-0.032	- Density=0.12784957387128	0.091	- Density=0.263613479777446	0.020
- Density=0.324409721468172	-0.019	- TA=0.264392324093817	-0.060	- TA=0.398720682302772	0.009
- TST=0.333508771929824	-0.01	- TST=0.543859649122807	0.028	- TST=0.333333333333333	0.001
(22. row)		(76. row)		(135. row)	

Table 2. *Break down* results for 3 randomly selected predictions

7.4 Local Interpretable Model-Agnostic Explanations (LIME)

Using *LIME* as an interpretation tool gives even more detailed explanations to a single prediction than *Break down*. Although they are seemingly similar, LIME, in addition to the importance of the variable, also gives a value interval for each individual predictor. A random selection of cases again proves *Theta* and *Zs* to be the most important predictors (Fig. 8).
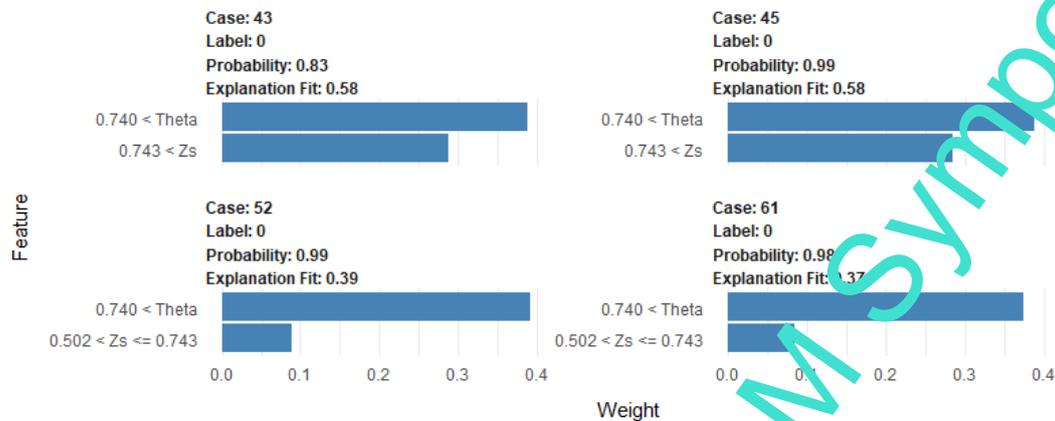

Fig. 8. Couple of randomly chosen cases from LIME algorithm

8. Conclusion

In this paper, we built a Random forest model by which we predict the ripeness of peaches based on several measured parameters. The model shows high prediction accuracy on the test data. Later in the paper, we use explainable machine learning methods to show that the strongest predictor for predicting the ripeness of peaches is electrical impedance (variables Zs and Theta). Thus, the firmness of the peach, which is chosen by the expert as the indicator of its ripeness, can be predicted quite accurately from the electrical impedance of the fruit.

Acceptance and confirmation of this idea, opens the possibility of building a precise, non-destructive and fast model for determining the maturity of the fruit using electrical impedance.

Such a model can be used in automated fruit picking where a mechanical hand could measure the peach impedance before picking the fruit and decide whether the fruit needs to be picked or not. The whole process would be further improved by combining several methods such as the use of computer vision combined with measuring fruit impedance.

During transport, the loss of fruit by rotting can be significantly reduced by adjusting parameters such as temperature, humidity and the presence of certain gases in the cargo area in order to create ideal conditions for transport. Predicting the ripeness of peaches during transport plays an important part.

Of course, to confirm our main idea, it is necessary to do more extensive research or controlled prototyping. If such a more extensive study would show that the electrical impedance cannot accurately predict fruit ripeness, then a new non-destructive measurement method should be introduced and few of them we have already mentioned in the related work chapter. A combination of several such methods should also be tested, which could increase the accuracy of the model.

In future work, another machine learning prediction model will be built and the results will be compared with those presented in this paper. Also, in agreement with the expert, multiple fruit ripeness classes will be designated indicating the degree of peach ripeness, as opposed to the binary model presented which has only two classes: *ripe* and *unripe*. Such a model would give much more sensitive and precise results, but at the cost of prediction accuracy. More measurements would be needed for improving it.

9. Acknowledgment

10. References

[1]   Kelley, K. M.; Primrose, R.; Crassweller, R.; Hayes, J. E.; Marini, R. (2016). Consumer peach preferences and purchasing behavior: A mixed methods study. J Sci Food Agric 96:2451–2461 . https://doi.org/10.1002/jsfa.7365
[2]   Balaji, R. D.; Gardas, B. B.; Narwane, V. S.; Narkhede, B. E. (2019). Improvement in the food losses in fruits and vegetable supply chain - a perspective of cold third-party logistics approach. Oper Res Perspect 6:100117 . https://doi.org/10.1016/j.orp.2019.100117
[3]   Kao, M. W. S.; Brecht, J. K.; Williamson, J. G.; Huber, D. J. (2012). Ripening development and quality of melting and non-melting flesh peach cultivars. HortScience 47:879–885 . https://doi.org/10.21273/hortsci.47.7.879

[4] Hayama, H.; Shimada, T.; Fujii, H.; Ito, A.; Kashimura, Y. (2006). Ethylene-regulation of fruit softening and softening-related genes in peach. J Exp Bot 57:4071–4077 . https://doi.org/10.1093/jxb/erl178

[5] Wang, X.; Matetić, M.; Zhou, H.; Zhang, X.; Jemrić, T. (2017). Postharvest quality monitoring and variance analysis of peach and nectarine cold chain with multi-sensors technology. Appl Sci 7: . https://doi.org/10.3390/app7020135

[6] Jamaludin, D.; Aziz, S. A.; Ibrahim, N. U. A. (2014). Dielectric Based Sensing System for Banana Ripeness Assessment. Int J Environ Sci Dev 5:286–289 . https://doi.org/10.7763/ijesd.2014.v5.493

[7] Li, J.; Xu, Y.; Zhu, W.; Wei, X.; Sun, H. (2019). Maturity assessment of tomato fruit based on electrical impedance spectroscopy. Int J Agric Biol Eng 12:154–161 . https://doi.org/10.25165/j.ijabe.20191204.4664

[8] Zhang, G.; Fu, Q.; Fu, Z.; Li, X.; Matetić, M.; Bakaric, M. B.; Jemrić, T. (2020). A comprehensive peach fruit quality evaluation method for grading and consumption. Appl Sci 10: . https://doi.org/10.3390/app10041348

[9] Juric, P.; Bakaric, M. B.; Wang, X.; Zhang, X.; Matetic, M. (2016). Mining data streams for the analysis of parameter fluctuations in iot-aided fruit cold-chain. Ann DAAAM Proc Int DAAAM Symp 27:756–761 . https://doi.org/10.2507/27th.daaam.proceedings.109

[10] Du, M.; Liu, N.; Hu, X. (2020). Techniques for interpretable machine learning. Commun ACM 63:68–77 . https://doi.org/10.1145/3359786

[11] Doran, D.; Schulz, S.; Besold, T. R. (2017). What Does Explainable AI Really Mean? A New Conceptualization of Perspectives Applications of Automated Theory-Formation using Meta-Interpretive Machine Learning View project Westwood Partnership to Prevent Juvenile Repeat Violent Offenders View project W

[12] Roscher, R.; Bohn, B.; Duarte, M. F.; Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access 8:42200–42216 . https://doi.org/10.1109/ACCESS.2020.2976199

[13] Gilpin, L. H.; Bau, D.; Yuan, B. Z.; Bajwa, A.; Specter, M.; Kagal, L. (2019). Explaining explanations: An overview of interpretability of machine learning. Proc - 2018 IEEE 5th Int Conf Data Sci Adv Anal DSAA 2018 80–89 . https://doi.org/10.1109/DSAA.2018.00018

[14] Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L. K.; Muller, K.-R. (2019). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning

[15] Kader, A. A. (1999). Future research needs in postharvest biology and technology of fruits. Acta Hortic 485:209–213 . https://doi.org/10.17660/actahortic.1999.485.28

[16] Kader, A. A. (2013). Postharvest Technology of Horticultural Crops - An Overview from Farm to Fork. J Appl Sci Technol 1:1–8

[17] Mercier, S.; Villeneuve, S.; Mondor, M.; Uysal, I. (2017). Time–Temperature Management Along the Food Cold Chain: A Review of Recent Developments. Compr Rev Food Sci Food Saf 16:647–667 . https://doi.org/10.1111/1541-4337.12269

[18] Nordey, T.; Joas, J.; Davrieux, F.; Chillet, M.; Léchaudel, M. (2017). Robust NIRS models for non-destructive prediction of mango internal quality. Sci Hortic (Amsterdam) 216:51–57 . https://doi.org/10.1016/j.scienta.2016.12.023

[19] Jha, S. N.; Narsaiah, K.; Jaiswal, P.; Bhardwaj, R.; Gupta, M.; Kumar, R.; Sharma, R. (2014). Nondestructive prediction of maturity of mango using near infrared spectroscopy. J Food Eng 124:152–157 . https://doi.org/10.1016/j.jfoodeng.2013.10.012

[20] Saranwong, S.; Sornsrivichai, J.; Kawano, S. (2004). Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy. Postharvest Biol Technol 31:137–145 . https://doi.org/10.1016/j.postharvbio.2003.08.007

[21] Cárdenas-Pérez, S.; Chanona-Pérez, J.; Méndez-Méndez, J. V.; Calderón-Domínguez, G.; López-Santiago, R.; Perea-Flores, M. J.; Arzate-Vázquez, I. (2017). Evaluation of the ripening stages of apple (Golden Delicious) by means of computer vision system. Biosyst Eng 159:46–58 . https://doi.org/10.1016/j.biosystemseng.2017.04.009

[22] Makino, Y.; Goto, K.; Oshita, S.; Sato, A.; Tsukada, M. (2016). A Grading Method for Mangoes on the Basis of Peel Color Measurement Using a Computer Vision System. Agric Sci 07:327–334 . https://doi.org/10.4236/as.2016.76033

[23] Inaba, A.; Manabe, T.; Tsuji, H.; Iwamoto, T. (1995). Electrical impedance analysis of tissue properties associated with ethylene induction by electric currents in cucumber (Cucumis sativus L.) fruit. Plant Physiol 107:199–205 . https://doi.org/10.1104/pp.107.1.199

[24] Harker, F. R.; Maindonald, J. H. (1994). Ripening of nectarine fruit. Changes in the cell wall, vacuole, and membranes detected using electrical impedance measurements. Plant Physiol 106:165–171 . https://doi.org/10.1104/pp.106.1.165

[25] Rehman, M.; Abu Izneid, B. A. J. A.; Abdullah, M. Z.; Arshad, M. R. (2011). Assessment of quality of fruits using impedance spectroscopy. Int J Food Sci Technol 46:1303–1309 . https://doi.org/10.1111/j.1365-2621.2011.02636.x

[26] Tekic, Z.; Cosic, I.; Katalinic, B. (2019). Manufacturing and the rise of artificial intelligence: Innovation challenges. Ann DAAAM Proc Int DAAAM Symp 30:192–196 . https://doi.org/10.2507/30th.daaam.proceedings.025

[27] Liakos, K. G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors (Switzerland) 18:1–29 . https://doi.org/10.3390/s18082674

[28] Rajesh K Sani; Navanietha Krishnaraj Rathinam. (2019). Bioelectrochemical Interface Engineering. John Wiley & Sons, Inc., South Dakota

[29] Breiman, L. (2001). Random Forests. Mach Learn 45:5–32 . https://doi.org/10.1023/A:1010933404324

[30] Hall, P.; Gill, N.; Kurka, M.; Phan, W.; Bartz, A. (2019). Machine Learning Interpretability with H2O Driverless AI: First Edition Machine Learning Interpretability with H2O Driverless AI