ORIGINAL PAPER

The application of fluorescence spectroscopy and machine learning as non-destructive approach to distinguish two different varieties of greenhouse tomatoes

Vanya Slavova[1] · Ewa Ropelewska[2] · Kadir Sabanci[3]

Abstract

The application of interdisciplinary non-invasive diagnostic methods combining fluorescence spectroscopy with multiple machine learning algorithms as tools for rapid application in tomato breeding programs is essential when crossing specific genotypes or parental samples to obtain representatives with better performance. Non-destructive distinguishing tomato species is of great importance for the preservation of product quality. This study aimed at combining fluorescence spectroscopic data and machine learning algorithms for distinguishing greenhouse tomatoes. The models for the discrimination of greenhouse tomato samples were built based on selected spectroscopic data using different machine learning algorithms from the groups of Meta, Functions, Bayes, Trees, Rules, and Lazy. The confusion matrices with accuracy for each sample, average accuracy, time taken to build the model, Kappa statistic, mean absolute error, root mean squared error and relative absolute error were determined. The greenhouse tomato samples were discriminated with an accuracy reaching 100% for the models built using Multi-Class Classifier (Meta), Logistic (Function), Bayes Net (Bayes), PART (Rules), and J48 (Trees). In the case of these algorithms, Kappa statistic was 1.0 and mean absolute error, root mean squared error and relative absolute error were equal to 0.

Introduction

Tomato (*Solanum lycopersicum* L.) is a garden plant consumed by millions of people worldwide every season. Since tomatoes are grown in moderately dry soil, production is very high and easy. 90% of farmers grow tomatoes on their farms [1]. Tomato, one of the most popular vegetables in people's daily life, has great economic importance for countries [2, 3]. According to the data from the Food and Agriculture Organization of the United Nations (FAO), tomato production is constantly increasing worldwide, and tomato production was 182 million tons in 2017. In 2019, the total production reached 243.6 million tons. The countries that play the most roles in this production are China, India, Turkey and the USA, respectively [4, 5]. Tomato has become one of the main food products in the world, due to its high daily consumption and its rich content of fiber, vitamins, minerals and antioxidants [6]. In addition to their rich nutritional content, tomatoes provide protection against diseases such as hepatitis, hypertension, inflammation and cancer [7]. However, the nutrient content and composition data of tomatoes differ according to the species, genetic and environmental factors [8]. Tomato types are available in a wide variety of sizes, colors and shapes. But in general, tomato varieties are expressed by four species, these are cherry, Italian, salad, and Santa Cruz. Among these species, cherry tomato cultivation is more profitable than others [9]. In this context, studies are being developed to increase the productivity of cherry tomato cultivation [10].

Due to its high production and increasing demand, it has become important to distinguish, package and transport

✉ Ewa Ropelewska
ewa.ropelewska@inhort.pl

1   Department of Plant Breeding, Maritsa Vegetable Crops Research Institute, Agricultural Academy Bulgaria, 32, Brezovsko Shosse St., 4003 Plovdiv, Bulgaria

2   Fruit and Vegetable Storage and Processing Department, The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100, Skierniewice, Poland

3   Department of Electrical and Electronics Engineering, Karamanoglu Mehmetbey University, Karaman, Turkey

tomato species. Tomatoes have a very sensitive structure to different production, transportation and packaging conditions and if they are damaged, the quality of the product decreases. In addition, the fertilization and inoculation practices during the growing process, as well as the growing region, affect the yield and content of the tomato. In the presence of such different production and growing conditions, nutritional security and successful discrimination of crops have gained importance today [3]. Providing both fast and contactless product discrimination can only be achieved with computerized systems today. Manual sorting of tomatoes according to their physiological ripeness is difficult with human control. This process is time-consuming and expensive, crops can be damaged by impact and erroneous sorting can occur. An automation system to be designed for this will both reduce the cost, increase the speed and accuracy of discrimination and increase the yield and productivity of the crop [11]. In addition, the development of smart agricultural applications depending on computerized systems both enables the development of non-destructive methods and accelerates the economic growth of many countries [12].

Applications based on computer vision and artificial intelligence for automatic discrimination of agricultural products have increased recently. In this context, shape, color and texture features are frequently used for automatic differentiation from crops. In order to perform classification according to these distinctive features with computers, various machine learning algorithms that learn these features are used [12]. Nyalala et al. [6] performed an application based on computer vision and machine learning methods for estimating tomato mass and volume. They made predictions with five different regression methods using 2D and 3D features obtained from depth images of tomatoes. As a result, the Radial Basis Function (RBF)-Support vector machine (SVM) model provided the most successful prediction. El-Bendary et al. [13] proposed a study based on color features for the automatic classification of tomato ripeness stages. They used Principal Components Analysis (PCA) for feature extraction and SVM and Linear Discriminant Analysis (LDA) for classification. In that study, which was applied with tenfold cross-validation, up to 90.80% accuracy was achieved with SVM. Semary et al. [14] performed an application that classifies infected/uninfected tomato fruits according to their color and texture (Gray Level Co-occurrence Matrix (GLCM)) features. The authors used PCA for feature reduction and SVM for classification. At the end of that study, tomatoes were classified as infected and uninfected with 92% accuracy depending on the external surface. Dhakshina Kumar et al. [15] developed a system based on texture (GLCM), shape and color characteristics to classify tomatoes according to their maturity. They also segmented the

defects in tomatoes with Gabor wavelet transform. Later, these defective regions were divided into three classes according to their color and geometrical characteristics. SVM was used in the classification stages. Ropelewska et al. [3] proposed a texture-based application for the discrimination of tomatoes based on flesh and skin images. Images from six different tomato species were then converted to R, G, B, L, a, b, X, Y, and Z color channels. Texture features extracted from different color channels were classified by various machine learning algorithms. Finally, Ireri et al. [16] classified tomatoes according to color, texture, and shape. LAB color space was preferred for color features and GLCM was preferred for texture properties. SVM with different kernel functions was used for grading recognition and RBF-SVM was used for defect detection.

The objective of this study was to combine fluorescence spectroscopic data and machine learning algorithms for distinguishing greenhouse tomato samples. According to the selected spectroscopic data, different machine learning algorithms from Meta, Functions, Bayes, Trees, Rules and Lazy groups distinguished greenhouse tomatoes with high success.

The salient contributions of this manuscript are summarized below.

- Use of fluorescent spectroscopy data for distinguishing tomato cultivars,
- Using machine learning algorithms for classification,
- Distinguishing tomato species with high accuracy.

By applying an author-designed mobile fiber-optic configuration using the phenomenon of fluorescence of light, it is possible to create non-invasive methods for field evaluation of tomatoes. So far, there is no data on their characterization by the proposed method. The aim is to validate fluorescence spectroscopy in the proposed configuration as a non-invasive method for the evaluation of two different varieties of greenhouse tomatoes. As a result of the successfully applied research in this study, it is expected that the creation of an interdisciplinary method for tomato analysis will be initiated.

A literature survey was conducted to conduct similar research. It turned out that until now the described experimental approach for tomato analysis has not been applied nationally and internationally. This gives us reason to claim that it is the first time that fluorescence spectroscopy in combination with machine learning has been applied to the analysis of tomatoes in field conditions. This study marks the beginning of these studies and will be of benefit to scientists who are developing their scientific directions in the field of optoelectronics or machine learning in the analysis of vegetable crops.

Materials and methods

Experimental design

For processing from the fluorescence study, ten averaged graphs from two different varieties of greenhouse tomatoes are presented. Graphs are averaged after the 15th measurement of each sample. There were over a thousand spectral data at different wavelengths for a single sample. The samples were measured on site at the farm where they were grown, as the fluorescence signal acquisition scheme is mobile. In this way, the effect of damaging the sample is avoided. The samples were measured immediately after recultivation. The mobile spectral installation (Fig. 1) for the study of fluorescence signals was designed specifically for the rapid analysis of plant biological samples.

The mobile experimental installation used by fluorescence spectroscopy contains the following blocks:

- Laser diode (LED) with an emission radiation of 245 nm with a supply voltage in the range of 3 V. It is housed in a hermetically sealed TO39 metal housing. The emitter has a voltage drop of 1.9 to 2.4 V and a current consumption of 0.02A. The minimum value of their reverse voltage is—6 V.
- Forming optic, which is a hemispherical lens made of N-BAK2 glass. The post-LED forming optics can defined mainly for the refractive, dispersive and thermo-optical properties, as well as for the transparency in the UV range [240–280 nm].
- Quartz glass area 4 cm^2. Its optical properties are to be transparent to visible light and to ultraviolet rays. This allows it to be free of inhomogeneities that scatter light. Its optical and thermal properties exceed those of other types of glass due to its purity. Light absorption in quartz glasses is weak.
- CMOS detector with photosensitive area 1.9968×1.9968 mm. Its sensitivity ranges from 200 to 1100 nm. Its resolution is $\delta\lambda = 5$. The profile of the detector sensor projections along the X and Y axes is also designed for very small amounts of data, unlike widely used sensors.

The radiation is led from the LED through the forming optics block by means of a quartz fiber. The secondary radiation from the illuminated sample (visible spectrum)—illuminated by the impacting UV radiation is coupled to the CMOS detector by means of light-guide optics. The quartz multimode fiber has a step index of refraction and a numerical aperture of 0.22. In the CMOS detector, the light signal is converted into an electrical–digital signal and, by means of a USB 2.0 wire, it is taken for analysis and downloading of the data to a laptop. The obtained fluorescence spectroscopic data were subjected to statistical analysis involving discriminant analysis to distinguish two different varieties of greenhouse tomato.
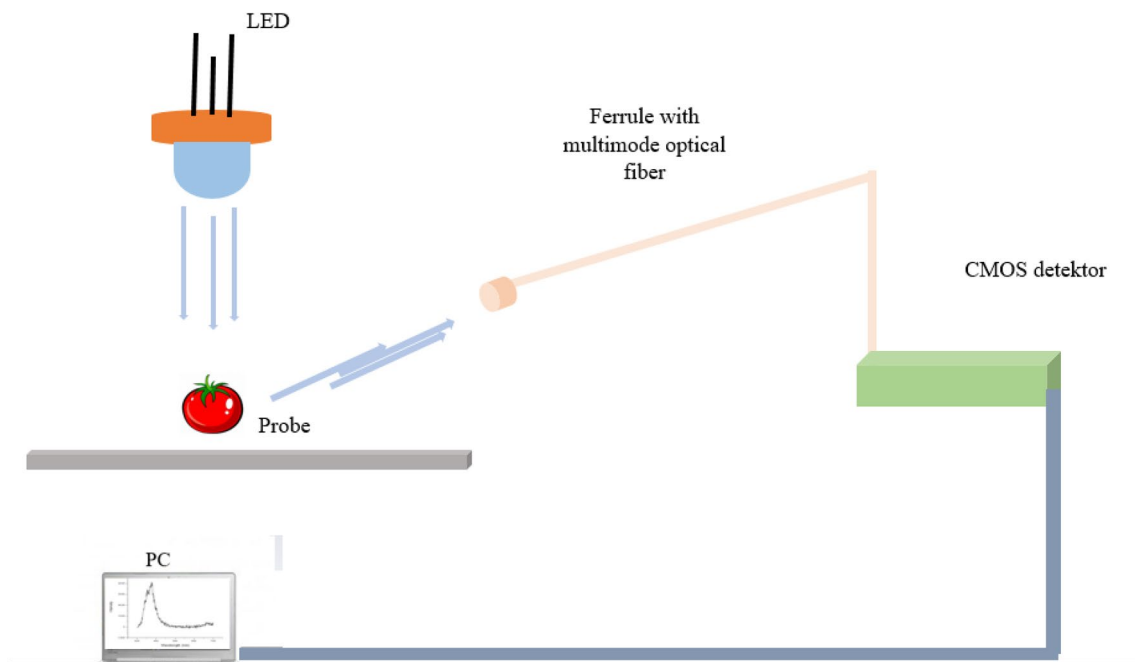

Fig. 1 Mobile experimental installation used by fluorescence spectroscopy

Statistical analysis

The samples of greenhouse tomatoes were discriminated with the use of the WEKA machine learning application (Machine Learning Group, University of Waikato) [17–19]. The differences in spectroscopic data of greenhouse tomato 1 and greenhouse tomato 2 varieties were analyzed. The flowchart presenting the applied procedure is shown in Fig. 2. After obtaining fluorescence spectroscopic data, the first step of the analysis included the attribute selection performed using the Ranker search method with the OneR Attribute Evaluator. The spectroscopic data with the highest
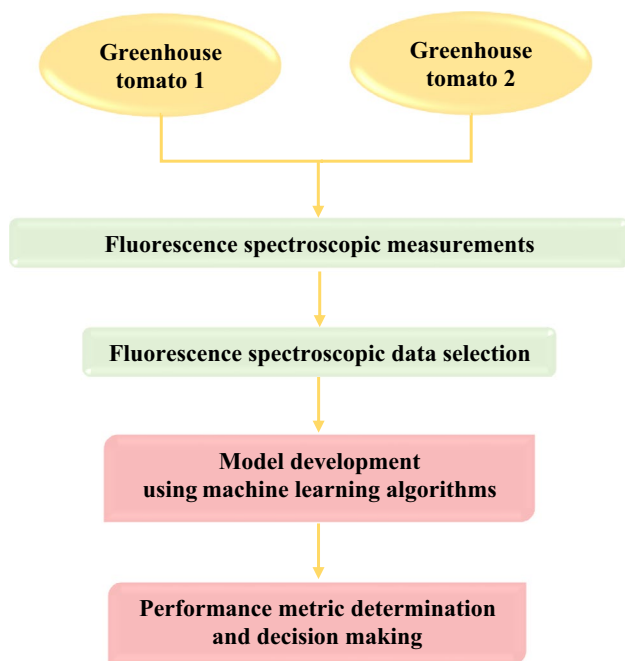
power to discriminate the tomato samples were selected. The discriminative models were built based on selected features using a tenfold cross-validation mode. The machine learning algorithms from the Meta, Functions, Bayes, Trees, Rules and Lazy groups were used. In the case of each group, algorithms providing the most satisfactory discrimination performance metrics were selected. The results were determined as confusion matrices including an accuracy for each sample, average accuracy, time taken to build the model, Kappa statistic, mean absolute error, root mean squared error, and relative absolute error. These performance metrics were computed using the WEKA application.

Results and discussion

The greenhouse tomato samples were completely correctly discriminated for the models developed based on fluorescence spectroscopic data using the following algorithms: Multi-Class Classifier from the group of Meta, Logistic (group of Function), Bayes Net (group of Bayes), PART (group of Rules), and J48 (group of Trees) (Table 1). The average accuracy, as well as accuracies for both greenhouse tomato 1 and greenhouse tomato 2 equal to 100%, were obtained. It meant that all cases belonging to the actual class of greenhouse tomato 1 were correctly classified as greenhouse tomato 1 and all cases from the class of greenhouse tomato 2 were correctly included in the predicted class of greenhouse tomato 2. The values of Kappa statistic equal to 1.0 and mean absolute error, root mean squared error and relative absolute error equal to 0 also indicate a completely correct classification. In the case of Bayes Net, time taken to build the model of 0.02 s was the shortest. Also, models built using other algorithms were characterized by the short time to build them, the longest for Logistic equal to 0.24 s.

For some algorithms from different groups, greenhouse tomato samples were distinguished with an average accuracy

Fig. 2 The flowchart of the stages of distinguishing greenhouse tomato samples using fluorescence spectroscopic data and machine learning algorithms

Table 1 The results of discrimination of greenhouse tomatoes for models built based on fluorescence spectroscopic data using selected algorithms providing an average accuracy of 100%

Algorithm	Time taken to build model (s)	Predicted class (%)		Actual class	Average accuracy (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)
		Greenhouse tomato 1	Greenhouse tomato 2						
Meta.Mult-classclas-sifier	0.10 0.24 0.02	100	0	Greenhouse tomato 1	100	1.0	0	0	0
Functions. Logistic	0.03 0.03	0	100	Greenhouse tomato 2					
Bayes.Bayes-Net									
Rules.PART									
Trees.J48									

of 95% (Table 2). The cases of greenhouse tomato 1 were classified with an accuracy of 100%. Whereas greenhouse tomato 2 samples were correctly discriminated in 90% and the remaining 10% were incorrectly classified as greenhouse tomato 1. These results were obtained for LDA and QDA (Quadratic Discriminant Analysis) from the group of Functions, Naive Bayes from the group of Bayes, Hoeffding Tree from the group of Trees, Filtered Classifier, Logit Boost and Random Committee from the group of Meta, and LWL from the group of Lazy. The high value of Kappa statistic of 0.9 was observed and low values of errors including the mean absolute error of 0.05, root mean squared error of 0.22 and the relative absolute error of 10% were found. The time taken to build the model was in the range of 0.00 s (Naive Bayes, LWL) to 6.42 s (QDA).

Slightly lower accuracies of discrimination of greenhouse tomato samples were determined for the models developed using other machine learning algorithms. For example, an average accuracy of 90% was obtained for JRip from the group of Rules and 85% for FLDA (Fisher Linear Discriminant Analysis) from the group of Functions (Table 3). In the case of a model built using the JRip algorithm, both classes were correctly discriminated with an accuracy of 90%. Whereas for the model developed using FLDA, the samples were correctly distinguished from each other in 80% for greenhouse tomato 1 and 90% for greenhouse tomato 2. In the case of using the FLDA algorithm, the value of Kappa statistic of 0.7 was the lowest and mean absolute error of 0.15, root mean squared error of 0.39, and relative absolute error of 30% were the highest.

Table 2 The performance metrics of discrimination of greenhouse tomatoes for models developed based on fluorescence spectroscopic data using selected algorithms providing an average accuracy of 95%

Algorithm	Time taken to build model (s)	Predicted class (%)		Actual class	Average accuracy (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)
		Greenhouse tomato 1	Greenhouse tomato 2						
Functions. LDA	3.24 6.42	100	0	Greenhouse tomato 1	95	0.9	0.05	0.22	10
Functions. QDA	0.00 0.03	10	90	Greenhouse tomato 2					
Bayes.Naive-Bayes	0.01 0.02								
Trees.Hoef-fdingTree	0.02 0.00								
Meta.Filtere-dClassifier									
Meta.Logit-Boost									
Meta.Ran-domCom-mittee									
Lazy.LWL									

Table 3 The results of discrimination of greenhouse tomatoes for models built based on fluorescence spectroscopic data using selected algorithms providing an average accuracy of 90 and 85%

Algorithm	Time taken to build model (s)	Predicted class (%)		Actual class	Average accuracy (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)
		Greenhouse tomato 1	Greenhouse tomato 2						
Rules.JRip	0.06	90	10	Greenhouse tomato 2	90	0.8	0.10	0.30	20
		10	90	Greenhouse tomato 1					
Functions. FLDA	0.51	80	20	Greenhouse tomato 2	85	0.7	0.15	0.39	30
		10	90	Greenhouse tomato 1					

The obtained results confirmed the effectiveness of the approach combining fluorescence spectroscopy and machine learning to distinguish greenhouse tomato varieties. The literature data also reported the usefulness of spectroscopy for the classification of tomatoes. Tomatoes belonging to different genotypes were classified using visible and short-wave spectroscopy, least-squares support vector machines (LS-SVM), soft independent modeling of class analogy (SIMCA), discriminant analysis (DA) and discriminant partial least-squares (DPLS) [20]. Additionally, spectroscopy was used to diagnose tomato diseases [21]. Furthermore, spectroscopy, i.e., spatially offset Raman spectroscopy (SORS) or fluorescence spectroscopy can be used for the evaluation of tomato maturity and postharvest ripening during storage [22–25]. Further studies may focus on the use of deep learning to discriminate tomatoes with a high probability.

Conclusions

Fluorescent spectroscopic data have proven to be highly effective for distinguishing greenhouse tomatoes. Numerous machine learning algorithms distinguished two different tomato varieties with high accuracy according to these data. The most successful discrimination was achieved with the Multi-Class Classifier, Logistic, Bayes Net, PART and J48 models in the Meta, Functions, Bayes, Rules and Trees groups, and all greenhouse tomato species were correctly classified. With other learning algorithms, discrimination accuracies of 95%, 90% and 85% were obtained. These results are quite satisfactory in terms of successful non-destructive and automatic discrimination of greenhouse tomato species. The performed research can be expanded to include more varieties and apply deep learning to discriminant analysis. In addition, the variety of data can be increased by taking images of tomatoes with a camera and adding color features to the fluorescence spectroscopy data. The successful conduct of this research allows for the formulation of interdisciplinary non-invasive diagnostic methods combining fluorescence spectroscopy with multiple machine learning algorithms as rapid application tools in tomato breeding programs. By monitoring the signal intensity, it will be possible to monitor the stability of a breeding line and its common blacks with an established cultivar of the same species. This will allow the crossing of specific genotypes or parental samples, with the aim of obtaining representatives with better indicators.

Declarations

References

1. Chen H-C, Widodo AM, Wisnujati A, Rahaman M, Lin JC-W, Chen L, Weng C-E (2022) AlexNet convolutional neural network for disease detection and classification of tomato leaf. Electronics 11:951
2. Zhang L, Jia J, Gui G, Hao X, Gao W, Wang M (2018) Deep learning based improved classification system for designing tomato harvesting robot. IEEE Access 6:67940–67950. https://doi.org/10.1109/ACCESS.2018.2879324
3. Ropelewska E, Sabanci K, Aslan MF (2022) Authentication of tomato (*Solanum lycopersicum* L.) cultivars using discriminative models based on texture parameters of flesh and skin images. Eur Food Res Technol 248(8):1959–1976
4. Arslan Ş, Arısoy H, Karakayacı Z (2022) The situation of regional concentration of tomato foreign trade in Turkey. Turkish J Agric Food Sci Technol 10:280–289
5. Ropelewska E, Piecko J (2022) Discrimination of tomato seeds belonging to different cultivars using machine learning. Eur Food Res Technol 248:685–705. https://doi.org/10.1007/s00217-021-03920-w
6. Nyalala I et al (2019) Tomato volume and mass estimation using computer vision and machine learning algorithms: cherry tomato model. J Food Eng 263:288–298. https://doi.org/10.1016/j.jfoodeng.2019.07.012
7. Trivedi NK et al (2021) Early detection and classification of tomato leaf disease using high-performance deep neural network. Sensors 21:7987
8. Slimestad R, Verheul M (2009) Review of flavonoids and other phenolics from fruits of different tomato (*Lycopersicon esculentum* Mill.) cultivars. J Sci Food Agric 89:1255–1270
9. Oziel FP, Edmilson ES (2021) Cherry tomato production with different doses of organic compost. Afr J Agric Res 17:1192–1197
10. Guo X-X, Zhao D, Zhuang M-H, Wang C, Zhang F-S (2021) Fertilizer and pesticide reduction in cherry tomato production to achieve multiple environmental benefits in Guangxi China. Sci Total Environ 793:148527. https://doi.org/10.1016/j.scitotenv.2021.148527
11. Tamakuwala S, Lavji J, Patel R (2018) Quality identification of tomato using image processing technique. Int J Electr Electron Data Commun 6:67–70
12. Sabanci K, Aslan MF, Durdu A (2020) Bread and durum wheat classification using wavelet based image fusion. J Sci Food Agric 100:5577–5585

13. El-Bendary N, El Hariri E, Hassanien AE, Badr A (2015) Using machine learning techniques for evaluating tomato ripeness. Expert Syst Appl 42:1892–1905. https://doi.org/10.1016/j.eswa.2014.09.057

14. Semary NA, Tharwat A, Elhariri E, Hassanien AE (2015) Fruit-based tomato grading system using features fusion and support vector machine. In: Filev D et al (eds) Intelligent systems 2014. Springer International Publishing, Cham, pp 401–410

15. Dhakshina Kumar S, Esakkirajan S, Bama S, Keerthiveena B (2020) A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier. Microprocessors Microsyst 76:103090. https://doi.org/10.1016/j.micpro.2020.103090

16. Ireri D, Belal E, Okinda C, Makange N, Ji C (2019) A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artif Intell Agric 2:28–37

17. Bouckaert RR, Frank E, Hall M, Kirkby R, Reutemann P, Seewald A, Scuse D (2016) WEKA manual for version 3-9-1. University Of Waikato, Hamilton

18. Witten I, Frank E, Hall MA, Pal CJ (2005). Data mining: practical machine learning tools and techniques, 4th edn. p 654

19. Witten I, Frank E, Hall M, Pal C (2016). In: Kaufmann M (ed) Data mining: practical machine learning tools and techniques. University of Waikato, Hamilton

20. Xie L, Ying Y, Ying T (2009) Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-squares support vector machines and other chemometrics. J Food Eng 94(1):34–39

21. Cordon G, Andrade C, Barbara L, Romero AM (2022) Early detection of tomato bacterial canker by reflectance indices. Inf Process Agric 9:184–194

22. Qin J, Chao K, Kim MS (2012) Nondestructive evaluation of internal maturity of tomatoes using spatially offset Raman spectroscopy. Postharvest Biol Technol 71:21–31

23. Kim DS, Lee DU, Choi JH, Kim S, Lim JH (2019) Prediction of carotenoid content in tomato fruit using a fluorescence screening method. Postharvest Biol Technol 156:110917

24. Fatchurrahman D, Amodio ML, de Chiara MLV, Chaudhry MMA, Colelli G (2020) Early discrimination of mature-and immature-green tomatoes (*Solanum lycopersicum* L.) using fluorescence imaging method. Postharvest Biol Technol 169:111287

25. Kasampalis DS, Tsouvaltzis P, Siomos AS (2020) Chlorophyll fluorescence, non-photochemical quenching and light harvesting complex as alternatives to color measurement, in classifying tomato fruit according to their maturity stage at harvest and in monitoring postharvest ripening during storage. Postharvest Biol Technol 161:111036