Bulgarian Journal of Agricultural Science, 32 (No 1) 2026, 219–226

The usefulness of approaches combining fluorescence spectroscopic data and machine learning algorithms to distinguish between different hybrid varieties of tomatoes grown in a greenhouse

Vanya Slavova[1*], Ewa Ropelewska[2], Muhammet Fatih Aslan[3], Kadir Sabanci[3] and Todorka L. Dimitrova[4]

[1] *Agricultural Academy, Maritsa Vegetable Crops Research Institute, Department of Plant Breeding, 4003 Plovdiv, Bulgaria*
[2] *The National Institute of Horticultural Research, Konstytucji, Fruit and Vegetable Storage and Processing Department, Skierniewice, Poland*
[3] *Karamanoglu Mehmetbey University, Department of Electrical and Electronics Engineering, Karaman, Turkey*
[4] *University of Plovdiv "Paisii Hilendarski", 4000 Plovdiv, Bulgaria*
[*] *Corresponding author:* vanya_plachkova@abv.bg

Abstract

Slavova, V., Ropelewska, E., Aslan, M. F., Sabanci, K. & Dimitrova, T. L. (2026). The usefulness of approaches combining fluorescence spectroscopic data and machine learning algorithms to distinguish between different hybrid varieties of tomatoes grown in a greenhouse. *Bulg. J. Agric. Sci., 32*(1), 219–226

The objective of this study was to discriminate different hybrid varieties of tomatoes, grown in a greenhouse, based on fluorescence spectroscopic data using models, developed with the use of machine learning algorithms belonging to groups of Meta, Trees, Bayes, Functions, Lazy, and Rules.

The combination of the fluorescence spectroscopic data and various machine learning algorithms is a great novelty in distinguishing the tomatoes grown in a greenhouse. The discriminant analysis including selected spectroscopic data was performed for all three samples and pairs of samples. In the case of discrimination of three samples, an average accuracy of 100%, TP (True Positive) Rate, Precision, MCC (Matthews Correlation Coefficient), ROC (Receiver Operating Characteristic) Area, PRC (Precision-Recall) Area, and F-Measure equal to 1.000, and FP (False Positive) Rate of 0.000 was determined for the Multi Class Classifier from the group of Meta, Bayes Net from the group of Bayes and Logistic from the group of Functions.

These results proved the complete differentiation of the samples in terms of selected spectroscopic data. Completely correct discriminations were also observed in the case of comparing some pairs of samples for the Multi Class Classifier, Bayes Net, Logistic, and kStar (Lazy) machine learning algorithms. The developed procedures can be used in practice to distinguish tomato samples in a non-destructive and objective way.

Keywords: Tomato fruit; fluorescence spectroscopy; discrimination; performance metrics

Introduction

Tomato (*Solanum lycopersicum* L.), originated from the Andes, and has been cultivated in Europe since the 16th century. The tomato is a fruit in botanical terms. However, it contains a lower sugar content than other fruits (Gerszberg et al., 2015). Tomato is one of the most economical crops around the world. New tomato cultivars, including those resistant to pathogen infestation, are constantly under development and used in breeding programs (El-Sitiny et al., 2022). Tomato taste and aroma are very important attributes for consumers. Tomato taste depends mainly on the content of sugars. Aroma results from volatiles. Tomato is rich in fiber, lycopene, and vitamins A and C. Its consumption may reduce the risk of some types of cancers and cardiovascular disease. Tomato can be consumed as fresh fruit, in salads and sandwiches, as salsa and in processed forms as dried tomato, juices or preserves, sauces, pastes, and soups. The selection of tomato cultivars and the postharvest practices are performed to reduce crop loss, as well as lengthen fruit shelf-life (Beckles, 2012).

Vegetables and fruits are not packaged directly after harvest. Sorting is made between crop varieties beforehand, and crops with similar physical characteristics are grouped. These groups are then packaged. In general, each package contains crops that are similar in shape, color and size. This packaging makes fruit and vegetables preferable for the consumer. These sorting and packaging processes are also very important for the tomato vegetable producer and consumer (Nassiri et al., 2021). The sorting process before packaging should be non-destructive, non-biased and fast. However, the discrimination of tomatoes by traditional methods is usually based on the practical experience of the farmers. This is very time-consuming and erroneous discriminations can be made. In addition, molecular marker and gene expression-based techniques are time-consuming and costly. Therefore, unlike traditional methods, automatic sorting studies have increased in the last decade for faster and more accurate differentiation of tomatoes (Huynh et al., 2021; Ropelewska et al., 2022).

In general, various sensors are used to automatically distinguish tomato varieties, and this sensor data can be analyzed to discriminate varieties. Analyzing this data by a computer instead of an expert provides faster and more accurate discrimination. For this reason, it is very popular to develop artificial intelligence-based methods that pave the way for Agriculture 4.0 (De Clercq et al., 2018). In this context, various machine learning methods such as Support Vector Machine (SVM), Decision Tree, Naive Bayes, etc. are in high demand by researchers (Kwak and Park, 2019). These methods can automatically distinguish crops by analyzing sensor data. Data generation is the most important step in machine learning, since the decision is completely based on data, and the obtained data should have a strong discriminating ability. So far, a large number of studies (Kaur et al., 2018; Pavithra et al., 2015; Wan et al., 2018) have been carried out to differentiate tomato hybrid varieties by traditional techniques, i.e. by features related to shape, color and size. However, unlike these features, spectral features have been identified by Zhong et al. (2014), as the primary factor that should be used to differentiate crops (Csendes and Mucsi, 2016). The spectral analysis provides important information for the determination of the physiological conditions and invisible defects of the crops, as well as providing species discrimination of crops (Ropelewska and Piecko, 2022). In the case of tomatoes, spectroscopic techniques were used for various purposes (de Oliveira et al., 2023). Among others, visible and near-infrared spectroscopy was applied to determine the sensory attributes of intact tomatoes (Li et al., 2021). Near-infrared spectroscopy was also used by Huang et al. (2018a) and de Brito et al. (2021), to evaluate the soluble solid content which is a very important quality parameter of tomatoes. Furthermore, Sheng et al. (2019) applied near-infrared spectroscopy to estimate the lycopene contents and soluble solids of cherry tomatoes. Laser-Induced Breakdown Spectroscopy (LIBS) and Chlorophyll Fluorescence Imaging (CFI) and analytical techniques can be useful to determine the chlorophyll amount in tomato plants (Senesi et al., 2022; Zhou et al., 2023). Huang et al. (2018b) reported that visible and near-infrared spectroscopy can be useful to predict the firmness parameters of tomatoes. Whereas Huang et al. (2018) predicted the firmness of tomatoes by spatially-resolved spectroscopy. Trebolazabala et al. (2017) monitored the tomato ripening by Raman spectroscopy. Spectroscopy can also be efficient at monitoring the tomato processing chain (Bureau et al., 2020). Despite the use of spectroscopic techniques for various studies of tomatoes, there is no literature data on the use of fluorescence spectroscopy combined with machine learning to distinguish different hybrid varieties of tomatoes grown in a greenhouse.

This study aimed at distinguishing differen thybrid varieties of tomatoes grown in a greenhouse based on fluorescence spectroscopic data. The models were developed using machine learning algorithms belonging to groups of Meta, Trees, Bayes, Functions, Lazy and Rules. The results were evaluated using confusion matrices, average accuracies, True Positive Rate, False Positive Rate, Precision, MCC (Matthews Correlation Coefficient), ROC (Receiver Operating Characteristic) Area, PRC (Precision-Recall) Area, and F-Measure. The combination of fluorescence spectroscopy and machine learning to discriminate the tomatoes grown in a greenhouse is a great novelty of the present study.

Materials and Methods

Experimental design

Ten samples of three different hybrid varieties of tomatoes grown in a greenhouse were tested. The samples are solid biological material, with a weight of the order of a gram taken by biopsy method from tomato fruits. The installation used in the experiment is compact enough to perform field analyses. It is aligned on an area of 38 cm long and 42 cm wide locally at the research sites. The developed installation was applied for practical studies of tomatoes locally in the greenhouse.

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 3V. It is housed in a hermetically sealed TO39 metal housing. The emitter has a voltage drop of 1.9 to 2.4V 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 is defined mainly for its refractive, dispersive and thermo-optical properties, as well as for its transparency in the UV range [240-280 nm].
- Quartz glass area 4 cm2. 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 nm 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.

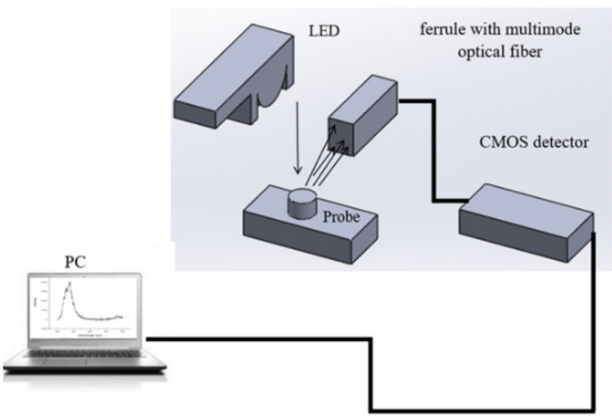

Fig. 1. Mobile experimental installation used by fluorescence spectroscopy
Source: Authors' own elaboration

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.

Statistical analysis

The obtained fluorescence spectroscopic data were used to discriminate hybrid varieties of tomatoes using the WEKA application (Machine Learning Group, University of Waikato) (Bouckaert et al., 2016; Eibe et al., 2016; Witten et al., 2005). The differentiation of fluorescence spectroscopic data of all three samples of 1, 2 and 3 were determined. Additionally, the discriminant analysis was carried out for pairs of samples 1 vs. 2, 1 vs. 3, 2 vs. 3. In the first step of the discriminant analysis, the attribute selection was performed using the Best First search method to select the data with the highest discriminative power to distinguish the tomato samples. The attribute selection step was performed for all three samples and each pair of samples separately. A 10-fold cross-validation mode was used to develop models based on the selected features using the machine learning algorithms from the Meta, Trees, Bayes, Functions, Lazy and Rules groups were used. In the case of each group, one algorithm providing the most satisfactory discrimination performance metrics was selected. The results were determined as confusion matrices including accuracy for each sample, average accuracies, TP (True Positive) Rate, FP (False Positive) Rate, Precision, MCC (Matthews Correlation Coefficient), ROC (Receiver Operating Characteristic) Area, PRC (Precision-Recall) Area, and F-Measure (Ropelewska, 2022; Ropelewska et al., 2022a; Ropelewska and Szwejda-Grzybowska, 2021).

Result and Discussions

The results of discrimination accuracies of three tomato classes are presented in Table 1. All three classes of tomatoes were discriminated with an average accuracy reaching 100% for the Multi Class Classifier (group of Meta), Bayes Net (group of Bayes) and Logistic (group of Functions). It means that all cases belonging to the actual classes of 1, 2 and 3 were correctly included in the predicted classes of 1, 2

Table 1. The confusion matrices and average accuracies of discrimination of three classes of tomatoes based on fluorescence spectroscopic data

Classifier	Predicted class (%)			Actual class	Average accuracy (%)
	1	2	3		
meta. Multi Class Classifier	100	0	0	1	100
	0	100	0	2	
	0	0	100	3	
trees. J48	100	0	0	1	96.67
	0	100	0	2	
	0	10	90	3	
bayes. Bayes Net	100	0	0	1	100
	0	100	0	2	
	0	0	100	3	
functions. Logistic	100	0	0	1	100
	0	100	0	2	
	0	0	100	3	
lazy. kStar	100	0	0	1	93.33
	0	90	10	2	
	0	10	90	3	
rules. JRip	100	0	0	1	96.67
	0	100	0	2	
	0	10	90	3	

Source: Authors' own elaboration

and 3. Other machine learning algorithms also distinguished three analyzed classes with very high accuracies equal to 96.67% for J48 (group of Trees) and JRip (group of Rules) and 93.33% for kStar (group of Lazy). In the case of J48 and JRip algorithms, 1 and 2 tomato samples were discriminated from each other with an accuracy of 100%. Whereas the cases of 3 were distinguished from others in 90% and the remaining 10% of cases belonging to the 3 class were incorrectly included in the predicted class 2. For kStar, only all cases of 1 were completely correctly classified as 1. The cases belonging to 2 and 3 classes were correctly classified in 90% and the mixing of cases between the two classes was noticed. 10 % of cases from the actual class of 2 were incorrectly classified as 3 and 10 % of cases belonging to the actual class of 3 were incorrectly included in the predicted class of 2.

Other performance metrics of discrimination of three classes of tomatoes are shown in Table 2. In the case of the Multi Class Classifier, Bayes Net and Logistic algorithms, the values of TP Rate, Precision, MCC, ROC Area, PRC Area, and F-Measure were equal to 1.000, and FP Rate was 0.000 for all three samples. These results confirmed the complete differentiation of the samples in terms of selected fluorescence spectroscopic data. The least satisfactory results were obtained in the case of the kStar machine learning algorithm. TP Rate, Precision, MCC, ROC Area, PRC Area, and

Table 2. The performance metrics of discrimination of three classes of tomatoes based on fluorescence spectroscopic data

Classifier	Class	TP Rate	FP Rate	Precision	MCC	ROC Area	PRC Area	F-Measure
meta. MultiClassClassifier	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	0.000	1.000	1.000	1.000	1.000	1.000
trees. J48	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.050	0.909	0.929	0.975	0.909	0.952
	3	0.900	0.000	1.000	0.926	0.928	0.933	0.947
bayes. Bayes Net	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	0.000	1.000	1.000	1.000	1.000	1.000
functions. Logistic	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	0.000	1.000	1.000	1.000	1.000	1.000
lazy.kStar	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	0.900	0.050	0.900	0.850	0.995	0.991	0.900
	3	0.900	0.050	0.900	0.850	0.995	0.991	0.900
rules.JRip	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.050	0.909	0.929	0.965	0.878	0.952
	3	0.900	0.000	1.000	0.926	0.935	0.933	0.947

TP Rate – True Positive Rate; FP Rate – False Positive Rate; MCC – Matthews Correlation Coefficient; ROC Area - Receiver Operating Characteristic Area; PRC Area – Precision-Recall Area

Source: Authors' own elaboration

F-Measure equal to 1.000, and FP Rate equal to 0.000 were observed only for 1 sample. Samples 2 and 3 were characterized by the lowest TP Rate, Precision and F-Measure equal to 0.900, MCC of 0.850 and the highest FP Rate of 0.050.

In the next stages, the discriminant analysis was performed for pairs of samples. The results for distinguishing 1 and 2 are shown in Table 3, including confusion matrices and average accuracies and Table 4, including the values of other discrimination performance metrics. Three (Multi Class Classifier, Bayes Net, Logistic) of the six algorithms distinguished the samples with an average accuracy equal to 100%. Models built using the other three algorithms (J48, kStar, JRip) correctly distinguished 1 and 2 in 95%. For each algorithm, the accuracy 1 was 100% and for 2 – 90% (Table 3).

Table 3. The discrimination accuracies of 1 and 2 tomatoes based on fluorescence spectroscopic data

Classifier	Predicted class (%)		Actual class	Average accuracy (%)
	1	2		
meta. Multi Class Classifier	100	0	1	100
	0	100	2	
trees. J48	100	0	1	95
	10	90	2	
bayes. Bayes Net	100	0	1	100
	0	100	2	
functions. Logistic	100	0	1	100
	0	100	2	
lazy.kStar	100	0	1	95
	10	90	2	
rules.JRip	100	0	1	95
	10	90	2	

Source: Authors' own elaboration

In the case of average accuracies reaching 100%, TP Rate, Precision, MCC, ROC Area, PRC Area, and F-Measure were 1.000. Whereas FP Rate was equal to 0.000 for both samples. For models developed using the J48, kStar and JRip machine learning algorithms, providing 95% average accuracy, higher values of TP Rate (1.000) and F-Measure (0.952) were obtained for the 1 sample and higher Precision (1.000) and PRC Area (0.950) were determined for 2. Both samples were characterized by the same MCC (0.905) and ROC Area (0.950). It was found that FP Rate was 0.100 for 1 and 0.000 for 2 (Table 4).

In the case of the discrimination of 1 vs. 3, slightly lower results were obtained for some of the used algorithms. Only models developed using Bayes Net and kStar produced accuracies of 100% (Table 5), FP Rate of 0.000 and the other metrics equal to 1.000 for both samples (Table 6). The average accuracy of 95% was determined for the Multi Class Classifier and Logistic for which 100% accuracy was obtained for 1 and 90% for 3. The models built using J48 and JRip algorithms provided an average discrimination accuracy equal to 90%. In the case of these algorithms, 1 sample was correctly discriminated in 100% and 3 – in 80% (Table 5).

For Bayes Net and kStar algorithms, the complete discrimination of the examined samples was confirmed by the FP Rate equal to 0.000 and the values of other metrics (TP Rate, Precision, MCC, ROC Area, PRC Area, F-Measure) equal to 1.000. The highest FP Rate of 0.200 (for 1) and the lowest results of the other performance metrics (TP Rate of 0.800 for 3, Precision of 0.833 for 1, MCC of 0.816 for 1 and 3, ROC Area of 0.870 for 3, PRC Area of 0.825 for 1, F-Measure of 0.889 for 3) were observed in the case of models built using J48 and JRip (Table 6).

Table 4. The performance metrics of discrimination of 1 and 2 tomatoes based on fluorescence spectroscopic data

Classifier	Actual class	TP Rate	FP Rate	Precision	MCC	ROC Area	PRC Area	F-Measure
meta. Multi Class Classifier	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
trees. J48	1	1.000	0.100	0.909	0.905	0.950	0.909	0.952
	2	0.900	0.000	1.000	0.905	0.950	0.950	0.947
bayes. Bayes Net	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
functions. Logistic	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
lazy.kStar	1	1.000	0.100	0.909	0.905	0.950	0.909	0.952
	2	0.900	0.000	1.000	0.905	0.950	0.950	0.947
rules.JRip	1	1.000	0.100	0.909	0.905	0.950	0.909	0.952
	2	0.900	0.000	1.000	0.905	0.950	0.950	0.947

TP Rate – True Positive Rate; FP Rate – False Positive Rate; MCC – Matthews Correlation Coefficient; ROC Area – Receiver Operating Characteristic Area; PRC Area – Precision-Recall Area

Source: Authors' own elaboration

Discriminant analysis performed for 2 vs. 3 revealed 100% accuracy (Table 7), as well as FP Rate of 0.000 and TP Rate, Precision, MCC, ROC Area, PRC Area, and F-Measure equal to 1.000 for both samples (Table 8) only in the case of the model, built using the Bayes Net algorithm. Whereas the lowest average accuracy of 90% (100% for 2 and 80% for 3) was observed for the model developed using kStar. In the case of other machine learning algorithms used, the average accuracies reached 95% and the 2 and 3 samples were correctly discriminated in 100% and 90%, respectively (Table 7). The kStar algorithm produced the highest value of FP Rate (0.200 for 2) and the lowest values of the other metrics (Precision: 0.833, PRC Area: 0.825 for 2, TP Rate: 0.800, ROC Area: 0.870, F-Measure: 0.889 for 3, and MCC: 0.816 for 2 and 3) (Table 8).

The result of our study confirmed the usefulness of spectroscopy and machine learning to discriminate different hybrid varieties of tomato. Spectroscopy can be used very successfully to differentiate crop varieties and detect plant pathological conditions based on spectral data (Ropelewska et al., 2022b). The spectroscopic techniques could be considered non-destructive, rapid, easy and require no sample preparation or using chemical reagents. These techniques were used in tomato research, among others, to direct estimate the sensory quality or determine physicochemical parameters (de Brito et al., 2021; Li et al., 2021).

Table 5. Discrimination accuracies of 1 vs. 3 tomatoes based on fluorescence spectroscopic data

Classifier	Predicted class (%)		Actual class	Average accuracy (%)
	1	3		
meta. Multi Class Classifier	100	0	1	95
	10	90	3	
trees. J48	100	0	1	90
	20	80	3	
bayes. Bayes Net	100	0	1	100
	0	100	3	
functions. Logistic	100	0	1	95
	10	90	3	
lazy.kStar	100	0	1	100
	0	100	3	
rules.JRip	100	0	1	90
	20	80	3	

Source: Authors' own elaboration

Table 7. The accuracies of the discrimination of 2 vs. 3 tomato samples based on fluorescence spectroscopic data

Classifier	Predicted class (%)		Actual class	Average accuracy (%)
	2	3		
meta. Multi Class Classifier	100	0	2	95
	10	90	3	
trees. J48	100	0	2	95
	10	90	3	
bayes. Bayes Net	100	0	2	100
	0	100	3	
functions. Logistic	100	0	2	95
	10	90	3	
lazy.kStar	100	0	2	90
	20	80	3	
rules.JRip	100	0	2	95
	10	90	3	

Source: Authors' own elaboration

Table 6. The results of discrimination of 1 vs. 3 tomatoes based on fluorescence spectroscopic data

Classifier	Actual class	TP Rate	FP Rate	Precision	MCC	ROC Area	PRC Area	F-Measure
meta. Multi Class Classifier	1	1.000	0.100	0.909	0.905	1.000	1.000	0.952
	3	0.900	0.000	1.000	0.905	1.000	1.000	0.947
trees. J48	1	1.000	0.200	0.833	0.816	0.890	0.825	0.909
	3	0.800	0.000	1.000	0.816	0.870	0.922	0.889
bayes. Bayes Net	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	0.000	1.000	1.000	1.000	1.000	1.000
functions. Logistic	1	1.000	0.100	0.909	0.905	1.000	1.000	0.952
	3	0.900	0.000	1.000	0.905	1.000	1.000	0.947
lazy.kStar	1	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	0.000	1.000	1.000	1.000	1.000	1.000
rules.JRip	1	1.000	0.200	0.833	0.816	0.890	0.825	0.909
	3	0.800	0.000	1.000	0.816	0.870	0.922	0.889

TP Rate – True Positive Rate; FP Rate – False Positive Rate; MCC – Matthews Correlation Coefficient; ROC Area – Receiver Operating Characteristic Area; PRC Area – Precision-Recall Area

Source: Authors' own elaboration

Table 8. The discrimination performance metrics of 2 vs. 3 tomato samples based on fluorescence spectroscopic

Classifier	Actual class	TP Rate	FP Rate	Precision	MCC	ROC Area	PRC Area	F-Measure
meta. Multi Class Classifier	2	1.000	0.100	0.909	0.905	1.000	1.000	0.952
	3	0.900	0.000	1.000	0.905	1.000	1.000	0.947
trees. J48	2	1.000	0.100	0.909	0.905	0.950	0.909	0.952
	3	0.900	0.000	1.000	0.905	0.950	0.950	0.947
bayes. Bayes Net	2	1.000	0.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	0.000	1.000	1.000	1.000	1.000	1.000
functions. Logistic	2	1.000	0.100	0.909	0.905	1.000	1.000	0.952
	3	0.900	0.000	1.000	0.905	1.000	1.000	0.947
lazy.kStar	2	1.000	0.200	0.833	0.816	0.890	0.825	0.909
	3	0.800	0.000	1.000	0.816	0.870	0.922	0.889
rules.JRip	2	1.000	0.100	0.909	0.905	0.945	0.901	0.952
	3	0.900	0.000	1.000	0.905	0.945	0.950	0.947

TP Rate – True Positive Rate; FP Rate – False Positive Rate; MCC – Matthews Correlation Coefficient; ROC Area – Receiver Operating Characteristic Area; PRC Area – Precision-Recall Area

Source: Authors' own elaboration

Conclusion

The non-destructive and robust differentiation of highly sensitive tomato species is an important challenge. In this study, an application was carried out for fast, reliable and non-destructive differentiation of tomatoes of different hybrid varieties grown in the greenhouse. A large number of machine learning algorithms belonging to Meta, Trees, Bayes, Functions, Lazy and Rules groups were used to perform the discrimination process automatically. These algorithms were fed with fluorescence spectroscopic features extracted from tomato hybrid varieties with three different classes. The discrimination accuracy rates obtained at the end of the study showed the ability and objectivity of machine learning algorithms fed with spectroscopic data to distinguish tomato hybrid varieties. In the future, it is planned to develop more successful discrimination techniques using different visual features in addition to spectroscopic features. In addition, these experiments are planned to be applied on more than one agricultural product.

References

Beckles, D. M. (2012). Factors affecting the postharvest soluble solids and sugar content of tomato (*Solanum lycopersicum* L.) fruit. *Postharvest Biology Technology, 63*(1), 129 – 140.

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

Bureau, S., Boas, A. A. d. C. V., Giovinazzo, R., Jaillais, B. & Page, D. (2020). Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. *LWT, 130*, 109518.

Csendes, B. & Mucsi, L. (2016). Identification and spectral evaluation of agricultural crops on hyperspectral airborne data. *Journal of Environmental Geography, 9*(3-4), 49 – 53.

de Brito, A. A., Campos, F., dos Reis Nascimento, A., de Carvalho Corrêa, G., da Silva, F. A., de Almeida Teixeira, G. H. & Júnior, L. C. C. (2021). Determination of soluble solid content in market tomatoes using near-infrared spectroscopy. *Food Control, 126*, 108068.

De Clercq, M., Vats, A. & Biel, A. (2018). Agriculture 4.0: The future of farming technology. *Proceedings of the World Government Summit*, Dubai, UAE, 11 – 13.

de Oliveira, A. N., Bolognini, S. R. F., Navarro, L. C., Delafiori, J., Sales, G. M., de Oliveira, D. N. & Catharino, R. R. (2023). Tomato classification using mass spectrometry-machine learning technique: A food safety-enhancing platform. *Food Chemistry, 398*, 133870.

Eibe, F., Hall, M. A. & Witten, I. H. (2016). The WEKA workbench. Online appendix for data mining: practical machine learning tools and techniques. In: *Morgan Kaufmann: Morgan Kaufmann Publishers,* 24 – 93.

El-Sitiny, M. F., Omar, H. M., El-Shehawi, A. M., Elseehy, M. M., El-Tahan, A. M., El-Saadony, M. T. & Selem, G. S. (2022). Biochemical and molecular diagnosis of different tomato cultivars susceptible and resistant to Tuta absoluta (Meyrick) infestation. *Saudi Journal of Biological Sciences, 29*(4), 2904 – 2910.

Gerszberg, A., Hnatuszko-Konka, K., Kowalczyk, T. & Kononowicz, A. K. (2015). Tomato (*Solanum lycopersicum* L.) in the service of biotechnology. *Plant Cell., Tissue Organ. Culture, 120*(3), 881 – 902.

Huang, Y., Lu, R. & Chen, K. (2018a). Assessment of tomato soluble solids content and pH by spatially-resolved and conventional Vis/NIR spectroscopy. *Journal of Food Engineering, 236*, 19 – 28.

Huang, Y., Lu, R. & Chen, K. (2018b). Prediction of firmness parameters of tomatoes by portable visible and near-infrared spectroscopy. *Journal of Food Engineering, 222*, 185 – 198.

Huang, Y., Lu, R., Xu, Y. & Chen, K. (2018). Prediction of tomato firmness using spatially-resolved spectroscopy. *Postharvest Biology and Technology, 140*, 18 – 26.

Huynh, D. P., Van Vo, M., Van Dang, N. & Truong, T. Q. (2021). Classifying maturity of cherry tomatoes using Deep Transfer Learning techniques. In: *IOP Conference Series: Materials Science and Engineering, 1109*, 012058. IOP Publishing.

Kaur, S., Girdhar, A. & Gill, J. (2018). Computer vision-based tomato grading and sorting. In: *Springer, Advances in data and information sciences*, 75 – 84.

Kwak, G.-H. & Park, N.-W. (2019). Impact of texture information on crop classification with machine learning and UAV images. *Applied Sciences, 9*(4), 643.

Li, X., Tsuta, M., Hayakawa, F., Nakano, Y., Kazami, Y. & Ikehata, A. (2021). Estimating the sensory qualities of tomatoes using visible and near-infrared spectroscopy and interpretation based on gas chromatography–mass spectrometry metabolomics. *Food Chemistry, 343*, 128470.

Nassiri, S. M., Tahavoor, A. & Jafari, A. (2021). Fuzzy logic classification of mature tomatoes based on physical properties fusion. *Information Processing in Agriculture, 9*(4), 547 – 555.

Pavithra, V., Pounroja, R. & Bama, B. S. (2015). Machine vision based automatic sorting of cherry tomatoes. In: *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, 271 – 275, IEEE.

Ropelewska, E. (2022). Distinguishing lacto-fermented and fresh carrot slice images using the Multilayer Perceptron neural network and other machine learning algorithms from the groups of Functions, Meta, Trees, Lazy, Bayes and Rules. *European Food Research and Technology*, 1-9.

Ropelewska, E. & Piecko, J. (2022). Discrimination of tomato seeds belonging to different cultivars using machine learning. *European Food Research Technology, 248*(3), 685 – 705.

Ropelewska, E., Sabanci, K. & Aslan, M. F. (2022). Authentication of tomato (*Solanum lycopersicum* L.) cultivars using discriminative models based on texture parameters of flesh and skin images. *European Food Research Technology*, 1-18.

Ropelewska, E., Slavova, V., Sabanci, K., Aslan, M. F., Cai, X. & Genova, S. (2022a). Discrimination of onion subjected to drought and normal watering mode based on fluorescence spectroscopic data. *Computers and Electronics in Agriculture, 196*, 106916.

Ropelewska, E., Slavova, V., Sabanci, K., Aslan, M. F., Cai, X. & Genova, S. (2022b). Discrimination of onion subjected to drought and normal watering mode based on fluorescence spectroscopic data. *Computers Electronics in Agriculture, 196*, 106916.

Ropelewska, E. & Szwejda-Grzybowska, J. (2021). A comparative analysis of the discrimination of pepper (Capsicum annuum L.) based on the cross-section and seed textures determined using image processing. *Journal of Food Process Engineering, 44*(6), e13694.

Senesi, G. S., De Pascale, O., Marangoni, B. S., Caires, A. R. L., Nicolodelli, G., Pantaleo, V. & Leonetti, P. (2022). Chlorophyll fluorescence imaging (CFI) and laser-induced breakdown spectroscopy (LIBS) applied to investigate tomato plants infected by the root knot nematode (RKN) Meloidogyne incognita and tobacco plants infected by Cymbidium ringspot virus. In: *Photonics, 9,* 627. MDPI.

Sheng, R., Cheng, W., Li, H., Ali, S., Agyekum, A. A. & Chen, Q. (2019). Model development for soluble solids and lycopene contents of cherry tomato at different temperatures using near-infrared spectroscopy. *Postharvest Biology and Technology, 156*, 110952.

Trebolazabala, J., Maguregui, M., Morillas, H., de Diego, A. & Madariaga, J. M. (2017). Portable Raman spectroscopy for an in-situ monitoring the ripening of tomato (Solanum lycopersicum) fruits. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 180*, 138 – 143.

Wan, P., Toudeshki, A., Tan, H. & Ehsani, R. (2018). A methodology for fresh tomato maturity detection using computer vision. *Computers electronics in agriculture, 146*, 43 – 50.

Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. (2005). Practical machine learning tools and techniques. In: *Data Mining, 2*(4), 403 – 413.

Zhong, L., Gong, P. & Biging, G. S. (2014). Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment, 140*, 1 – 13.

Zhou, X., Zhao, C., Sun, J., Cao, Y., Yao, K. & Xu, M. (2023). A deep learning method for predicting lead content in oilseed rape leaves using fluorescence hyperspectral imaging. *Food Chemistry, 409*, 135251.

Received: **May, 24, 2024**; *Approved*: **October, 02, 2024**; *Published*: **February, 2026**