agriculture

Article

Benchmarking Machine Learning Approaches to Evaluate the Cultivar Differentiation of Plum (*Prunus domestica* L.) Kernels

Ewa Ropelewska [1,*], Xiang Cai [2,3], Zhan Zhang [4], Kadir Sabanci [5] and Muhammet Fatih Aslan [5]

1 Fruit and Vegetable Storage and Processing Department, The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland
2 Faculty of Biology, Lomonosov Moscow State University, Leninskie Gory 1, 119991 Moscow, Russia; caixiang@smbu.edu.cn
3 Department of Ecology, School of Biology, Shenzhen MSU-BIT University, Dayun Newtown 1, Shenzhen 518172, China
4 University Office, Shenzhen MSU-BIT University, Dayun Newtown 1, Shenzhen 518172, China; zhangz@smbu.edu.cn
5 Department of Electrical and Electronics Engineering, Karamanoglu Mehmetbey University, Karaman 70100, Turkey; kadirsabanci@kmu.edu.tr (K.S.); mfatihaslan@kmu.edu.tr (M.F.A.)
* Correspondence: ewa.ropelewska@inhort.pl

Abstract: Plum fruit and kernels offer bioactive material for industrial production. The promising procedure for distinguishing plum kernel cultivars used in this study comprised two stages: image analysis to compute the texture parameters of plum kernels belonging to three cultivars 'Emper', 'Kalipso', and 'Polinka', and discriminant analysis using machine learning algorithms to classify plum kernel cultivars based on selected textures with the highest discriminative power. The discriminative models built separately for sets of textures selected from all color channels L, a, b, R, G, B, U, V, S, X, Y, Z, color space Lab and color channel b using the KStar (Lazy), PART (Rules), and LMT (Trees) classifiers provided the highest average accuracies reaching 98% in the case of the color space Lab and the KStar classifier. In this case, individual cultivars were discriminated with the accuracies of 97% for 'Emper' and 'Kalipso' to 99% for 'Polinka'. The values of other performance metrics were also satisfactory, higher than 0.95. The ROC curves were quite smooth and steady with the most satisfactory curve for the 'Kalipso' kernels. The present study sheds light on an objective, non-destructive, and inexpensive procedure for cultivar discrimination of plum kernels.

Keywords: plum kernel images; texture parameters; discrimination; algorithms; performance metrics

check for updates

Citation: Ropelewska, E.; Cai, X.; Zhang, Z.; Sabanci, K.; Aslan, M.F. Benchmarking Machine Learning Approaches to Evaluate the Cultivar Differentiation of Plum (*Prunus domestica* L.) Kernels. *Agriculture* **2022**, *12*, 285. https://doi.org/10.3390/agriculture12020285

Academic Editor: Yanbo Huang

Received: 23 January 2022
Accepted: 14 February 2022
Published: 17 February 2022

1. Introduction

Plum belongs to the genus *Prunus* and Rosaceae family. The hexaploid plum named *Prunus domestica* L. is the main cultivated plum in Europe and Asia. Plum cultivars can differ in many characteristics such as shape, size, color, weight, and chemical composition of the fruit, flesh adhesion and shape of stone, diameter and color of anthers of flower, dates of flowering and fruit maturity [1]. Different cultivars of plum can also be characterized by different resistance to diseases [2]. Individual plum cultivars can be cultivated under different conditions including temperature, water, light, nutrient [3]. Plum is a very important and healthy fruit that can be consumed in fresh and processed forms such as jellies, jams, dried products [4]. Mature plum contains about 84–90% (w/w) flesh and the pit (stone) with the kernel is the remaining 10–16% (w/w) [5]. The processing of plums results in the production of pits that should be removed from the fruit. Pits can be considered as unwanted waste material. However, the kernel contained in the pit can have great industrial potential as a source of oils, dietary proteins, vitamins, minerals, fibers, carbohydrates, as well as other bioactive components. Kernels also contain amygdalin that in the appropriate doses can have health-promoting properties. However, amygdalin can hydrolyze to hydrogen cyanide that can have harmful and toxic effects on human health.

Therefore, the processing of plum kernels should include a detoxification step to reduce the risk of hydrogen cyanide formation [4]. The physico-chemical properties of plum kernels and products of their processing depend on the cultivar [6–8].

It is well-documented that a plant cultivar can distinguish itself from other variants due to the inherent physicochemical properties formed and recorded during its growth [5,6]. Conventionally, the discrimination of plant cultivars is dependent on the phenotypic and genotypic traits such as morphological forms, pomological structure, fruit shape, and appearance, plant physiological indices, genetic information, etc. [9,10]. However, the phenotypic diversity of plant species could not be suitable for the cultivar identification, classification, and selection based on postharvest fruit kernels, nor would genetic information collection be cost-efficient in the broad application of the cultivar discrimination. In the present research, a machine learning technology for plum cultivar discrimination from images of the fruit kernel was developed and practiced. The machine learning method depends on (I) the comprehensive information collection from the images where the kernel textural traits are shown into a big dataset, and (II) the intelligent analysis of the quantitative data via artificial intelligence using algorithms in a computer program like MaZda and WEKA [11]. As a result, the technological machine combines a variety of skills into a network system, like identifying visual signals, converting them into digital data, processing data with different algorithms, modeling the results via analysis, and finally predicting a correlative trendline between plum cultivars and the textural traits of plum kernels.

In the available literature, there are reports on the application of image analysis for the discrimination of various kernels, seeds, or pits. Nine varieties of sweet maize seeds were discriminated using hyperspectral images and machine learning algorithms reaching an accuracy of 94.86% [12]. A very high accuracy equal to 99% was observed for the discrimination of 14 types of seeds using deep learning techniques [13], whereas, the CNN-ANN (convolutional neural network-artificial neural network) classifier was used to discriminate nine corn seed varieties with the correctness of 98.1% using images acquired by a digital camera [14]. The application machine learning classifiers and features extracted from images obtained using a flatbed scanner allowed for the cultivar discrimination of sour cherry pits with accuracies reaching 96.25% for four cultivars and 100% for two cultivars [15] and in the case of images acquired using a digital camera—for distinguishing two cultivars of sweet cherry pits in 100% of cases and three cultivars in 98% [16], and the discrimination of two cultivars of peach stones and seeds with the accuracy of up to 100% [17]. The application of linear dimensions and shape factors for the development of models allowed for the discrimination of the pits belonging to different sour cherry cultivars with an accuracy of up to 96% [11]. The morphometric features extracted from digital images acquired using a flatbed scanner and stepwise linear discriminant analysis were used to compare the modern and archaeological *Prunus* fruit stones. The archaeological stones were identified as *P. spinosa* and *P. domestica* and showed similarities with the modern samples [18].

The evaluation of the cultivar differentiation of plum kernels using image analysis can be of practical importance. Correct cultivar identification can be necessary for the processing industry to avoid mixing kernel cultivars with different compositions. The cultivar recognition may also allow avoiding falsification of kernel cultivars and reject kernels with undesirable properties for further processing [11,15,16]. Therefore, the objective of this study was to develop models for distinguishing the plum kernel cultivars based on selected image texture parameters using various algorithms (classifiers). The innovative nature of this study is related to the acquisition of new, not found in the literature, information on almost 2000 textures texture parameters of plum kernels belonging to different cultivars. The novelty is also the development of innovative models based on attributes selected from a set of computed textures using different machine learning algorithms and the comparison of their effectiveness. Plum kernels have been classified by using different machine learning approaches to provide strong discrimination with an objective, non-destructive and inexpensive procedure. The aim of the research was not to distinguish the plum fruit cultivars based on kernel properties. The fruits are not taken into account at all. Research focuses only on kernels that can be a waste product in plum processing and can be used in industry independently from fruit. The procedure refers to kernels that have already been extracted from fruits during processing. Non-destructiveness of the research refers to the fact that the kernels were not damaged during the analysis.

2. Materials and Methods

2.1. Materials

The experiment was carried out using the kernels belonging to three plum cultivars 'Emper', 'Kalipso', and 'Polinka'. The mature plums were harvested in August 2021 from the orchard located in Poland. First, the plum stones were extracted manually from each fruit. Then, the kernels were obtained by destroying the walls of stones. Fully developed kernels with no visible damage were used in this study. The kernels were subjected to imaging using a digital camera.

2.2. Image Analysis

2.2.1. Image Acquisition

Whole, undamaged kernels were subjected to image acquisition. This approach was very beneficial as it allowed obtaining objective results without damaging the kernel structure. The images of plum kernels positioned on a black background were acquired using a designed system consisting of a digital camera and LED (Light Emitting Diodes) illumination with stable parameters. A black background was obtained by placing the kernel samples in a box with black internal walls. Plum kernels were imaged after performing the color calibration of the digital camera. Twenty kernels were included in one image. For each cultivar, images of one hundred kernels were acquired. In total, a set consisting of three hundred digital color images of 'Emper', 'Kalipso', and 'Polinka' plum kernels was used in the study. The images were saved in a TIFF format. The exemplary plum kernel images are shown in Figure 1.

2.2.2. Image Processing

Image processing was carried out using the MaZda software (Łódź University of Technology, Institute of Electronics, Poland) [19]. Before processing, kernel images were converted to BMP format. The regions of interest (ROIs) were determined as a single kernel separated from the background for each ROI. Each kernel image was converted to individual color channels L, a, b, R, G, B, U, V, S, X, Y, Z. In the case of each kernel (ROI), about 2200 texture parameters of the outer surface (external structure) of images were extracted. The exemplary results of computed texture parameters for the 'Emper', 'Kalipso', and 'Polinka' kernels are provided in Supplementary Table S1. The image textures were computed based on the co-occurrence matrix, run-length matrix, Haar wavelet transform, gradient map, autoregressive model, histogram. The textures were used to build the models for distinguishing the plum kernel cultivars.

2.3. Discriminant Analysis

2.3.1. Cultivar Discrimination of Plum Kernels

To discriminate the plum kernels belonging to cultivars 'Emper', 'Kalipso', and 'Polinka', the models developed based on selected textures were applied. The discriminant analysis was carried out using the WEKA (Waikato Environment for Knowledge Analysis) machine learning software (University of Waikato, New Zealand) [20,21]. The analysis was performed for a set including textures from all color channels, as well as for sets of textures determined for individual color spaces and color channels. For each set, the selection of textures with the highest discriminative power was carried out using the Best First search algorithm. Twenty features were the optimal number to obtain high correctness and a short analysis time. The different algorithms (classifiers) from the groups of Rules, Functions, Trees, Bayes, Lazy, and Meta were examined [22].

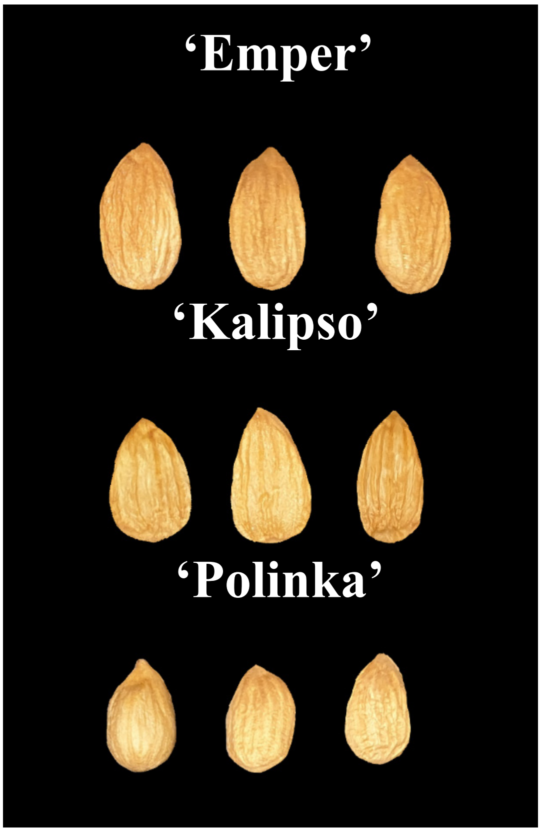

Figure 1. The exemplary images of 'Emper', 'Kalipso', and 'Polinka' plum kernels.

2.3.2. Performance Metrics

The confusion matrices including the accuracies (%) for predicted kernel classes 'Emper', 'Kalipso', and 'Polinka', average accuracies for three cultivars, and the values of performance metrics such as Precision, F-Measure, MCC (Matthews Correlation Coefficient), ROC (Receiver Operating Characteristic) Area, and PRC (Precision-Recall) Area were computed [20–22]. For the models providing the best results, the ROC (Receiver Operating Characteristic) curves were also determined. The Equations (1)–(8) were used to compute the performance metrics:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \times 100 \tag{1}$$

$$\text{Precision} = (\text{TP}/\text{TP} + \text{FP}) \tag{2}$$

$$\text{F-Measure} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \tag{3}$$

$$\text{MCC} = ((\text{TP} * \text{TN}) - (\text{FN} * \text{FP}))/\sqrt{((\text{TP} + \text{FN}) * (\text{TN} + \text{FP}) * (\text{TP} + \text{FP}) * (\text{TN} + \text{FN}))} \tag{4}$$

$$\text{ROC Area} = \text{ Area Under TPR vs. FPR Curve} \tag{5}$$

$$\text{PRC Area} = \text{ Area Under Precision vs. Recall Curve} \tag{6}$$

$$\text{TPR (Recall)} = \text{TP}/(\text{TP} + \text{FN}) \tag{7}$$

$$\text{FPR} = \text{FP}/(\text{FP} + \text{TN}) \tag{8}$$

TP: True Positive
TN: True Negative
FP: False Positive
FN: False Negative
TPR: True Positive Rate
FPR: False Positive Rate

The results for models built using selected classifiers based on a set of textures selected from all color channels as well as a texture set from one color space and a set of textures from one color channel providing the highest discrimination accuracies are presented in this paper.

2.3.3. Machine Learning Algorithms

Among the applied algorithms, the KStar from a group of Lazy, the PART from a group of Rules and the LMT from a group of Trees provided the highest results of discrimination. Therefore, the discrimination accuracies and values of Precision, F-Measure, MCC, ROC Area, and PRC Area are presented for these classifiers. The parameter settings of the algorithms were based on the WEKA:

- KStar—batchSize: 100, debug: False, doNotCheckCapabilities: False, entropicAutoBlend: False, globalBlend: 20,
- PART—batchSize: 100, binarySplits: False, confidenceFactor: 0.25, debug: False, doNotCheckCapabilities: False, minNumObj: 2, numFolds: 3, reducedErrorPruning: False, seed: 1, unpruned: False, useMDLcorrection: False,
- LMT—batchSize: 100, convertNominal: False, debug: False, doNotCheckCapabilities: False, fastRegression: True, minNumInstances: 15, numBoostingIterations: −1, splitOnResiduals: False, useAIC: False, weightTrimBeta: 0.0.

At the first step of the analysis, the models were developed for a set of textures selected from all color channels L, a, b, R, G, B, U, V, S, X, Y, Z of plum kernel images. When developing models for textures selected separately from the individual color spaces and color channels, it was found that discrimination performance metrics were the highest for models built based on textures selected from color space Lab and color channel b, respectively. Thus, the results obtained for these data sets were chosen to be presented.

3. Results

In the case of models built using the KStar classifier (Table 1), high discrimination performance metrics were acquired for a set of textures selected from all color channels, as well as color space Lab and color channel b. The 'Emper', 'Kalipso', and 'Polinka' plum kernels were correctly classified with an average accuracy reaching 98% in the case of color space Lab. Individual plum kernel cultivars were distinguished with accuracies of 97% for 'Emper' and 'Kalipso' to 99% for 'Polinka'. One kernel (1% of cases) belonging to plum 'Polinka' was incorrectly classified as 'Emper' and three kernels of 'Emper' were incorrectly included in class 'Polinka'. Whereas among a hundred cases of 'Kalipso', two cases were incorrectly classified as 'Polinka' and one case—as 'Emper'. The values of other metrics were very satisfactory. Precision and ROC Area reached 1.000 for 'Kalipso'. The values of F-Measure, MCC, and PRC Area were also the highest for the 'Kalipso' plum kernels and were equal to 0.985, 0.978, and 0.999, respectively. The ROC (Receiver Operating Characteristic) curves for the 'Emper', 'Kalipso', and 'Polinka' plum kernels proved high values of ROC Area for all the cultivars (Figure 2). The most satisfactory ROC curve was obtained for the 'Kalipso' kernels (Figure 2b).

Table 1. The results of cultivar discrimination of plum kernels performed using models built based on textures selected from all color channels *L*, *a*, *b*, *R*, *G*, *B*, *U*, *V*, *S*, *X*, *Y*, *Z*, color space Lab, and color channel *b* using the KStar (Lazy) classifier.

Set of Selected Textures	Predicted Class (%) 'Emper'	'Kalipso'	'Polinka'	Actual Class	Average Accuracy (%)	Precision	F-Measure	MCC	ROC Area	PRC Area
all color channels	94	2	4	'Emper'	95	0.959	0.949	0.925	0.998	0.995
	1	98	1	'Kalipso'		0.933	0.956	0.934	0.999	0.998
	3	5	92	'Polinka'		0.948	0.934	0.902	0.996	0.993
color space Lab	97	0	3	'Emper'	98	0.980	0.975	0.962	0.998	0.996
	1	97	2	'Kalipso'		1.000	0.985	0.978	1.000	0.999
	1	0	99	'Polinka'		0.952	0.971	0.956	0.998	0.997
color channel *b*	94	3	3	'Emper'	95	0.949	0.945	0.917	0.996	0.992
	1	97	2	'Kalipso'		0.942	0.956	0.933	0.998	0.996
	4	3	93	'Polinka'		0.949	0.939	0.910	0.993	0.986

MCC—Matthews Correlation Coefficient; ROC Area—Receiver Operating Characteristic Area; PRC Area—Precision-Recall Area.

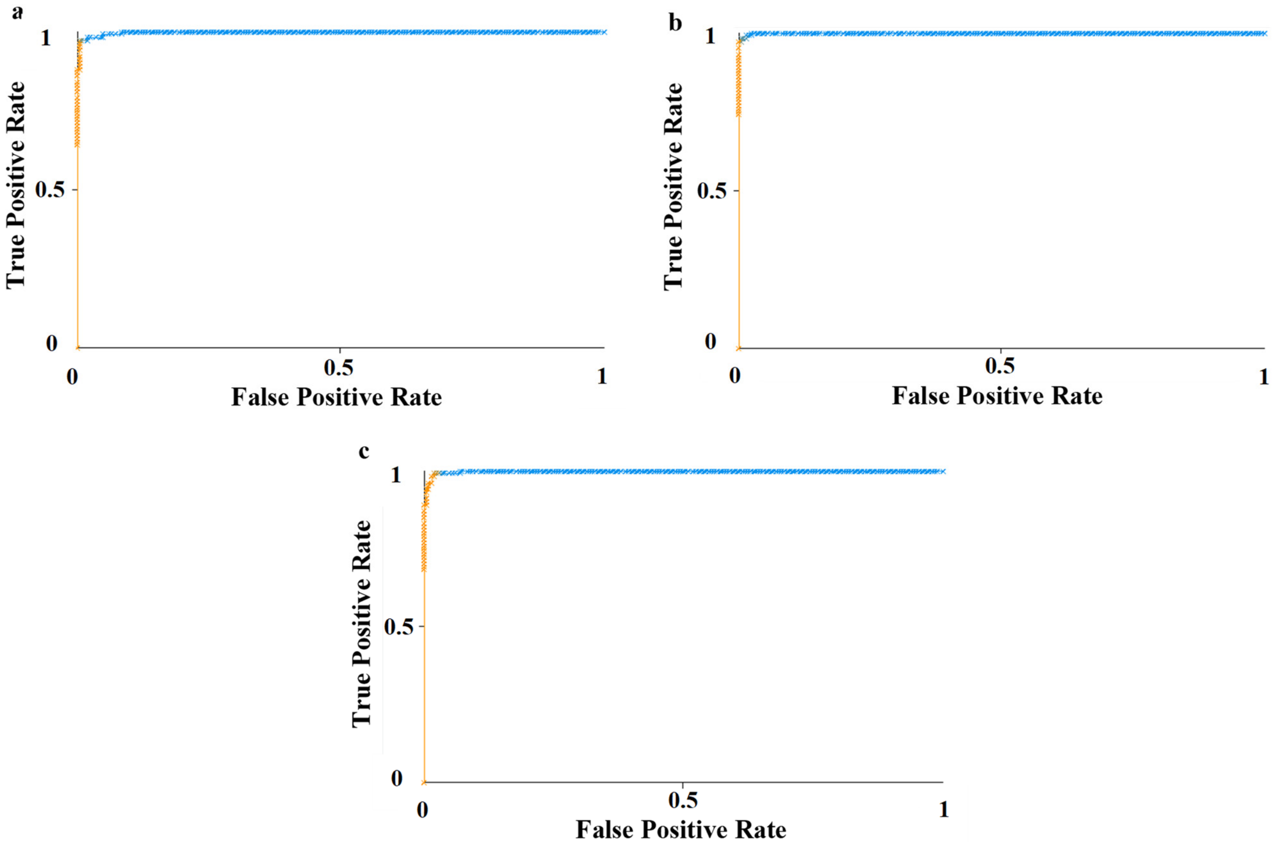

Figure 2. The ROC (Receiver Operating Characteristic) curves for the cultivar discrimination of plum kernels for a model built using the KStar classifier based on textures selected from color space Lab, (**a**) 'Emper', (**b**) 'Kalipso', (**c**) 'Polinka'.

In the case of the model built for all color channels and model developed for a color channel *b* using the KStar classifier (Table 1), the average accuracies of discrimination of the 'Emper', 'Kalipso', and 'Polinka' plum kernels were equal to 95%. For both models, the 'Emper' plum kernels were characterized by accuracies equal to 94%. The model built based on textures selected from all color channels provided an accuracy of 98% for the 'Kalipso' kernels and 92% for the 'Polinka' kernels. Whereas in the case of the model including textures selected from images converted to color channel *b*, the 'Kalipso' plum kernels were correctly classified in 97%, and the kernels of 'Polinka' were classified with correctness equal to 93%. Comparing other metrics for the model developed using textures selected from images from all color channels and color channel *b*, the highest values of Precision (0.959, 'Emper'), MCC (0.934, 'Kalipso'), ROC Area (0.999, 'Kalipso'), PRC Area (0.998, 'Kalipso') were observed for the model built for textures selected from combined all color channels. The highest F-Measure equal to 0.956 were determined for 'Kalipso' for both the model built for all color channels and color channel *b*.

The models built using the PART classifier (Table 2) produced a very satisfactory average accuracy of discrimination of plum kernels of 'Emper', 'Kalipso', and 'Polinka' equal to 95% for a set of textures selected from all color channels. The accuracy for the predicted class 'Kalipso' reached 97%. The 'Emper' plum kernels were correctly distinguished from other classes in 95% and kernels of 'Polinka'—in 94%. The values of Precision (0.979), F-Measure (0.964), MCC (0.947), and PRC Area (0.941) were the highest for the kernels of 'Emper', whereas ROC Area equal to 0.969 was the highest for the plum kernels of 'Kalipso'. The kernels of 'Kalipso' were also characterized by the smoothest ROC curve (Figure 3). In the case of color space Lab and color channel *b*, the average accuracies for models built using the PART classifier were slightly lower, equal to 88 and 87%, respectively (Table 2). Also, the accuracies for individual predicted classes were lower. In the case of the model developed based on a set of textures selected from the color space Lab, the 'Emper' and 'Polinka' plum kernels were correctly discriminated in 87% and 'Kalipso' in 91%. The other performance metrics were also the highest for the kernels 'Kalipso'. The model built based on textures selected from images converted to color channel *b* provided accuracies of 83% for the kernels of 'Polinka', 87% for 'Emper', and 92% for 'Kalipso'. The values of Precision, F-Measure, MCC, ROC Area, and PRC Area were the highest for the 'Kalipso' plum kernels.

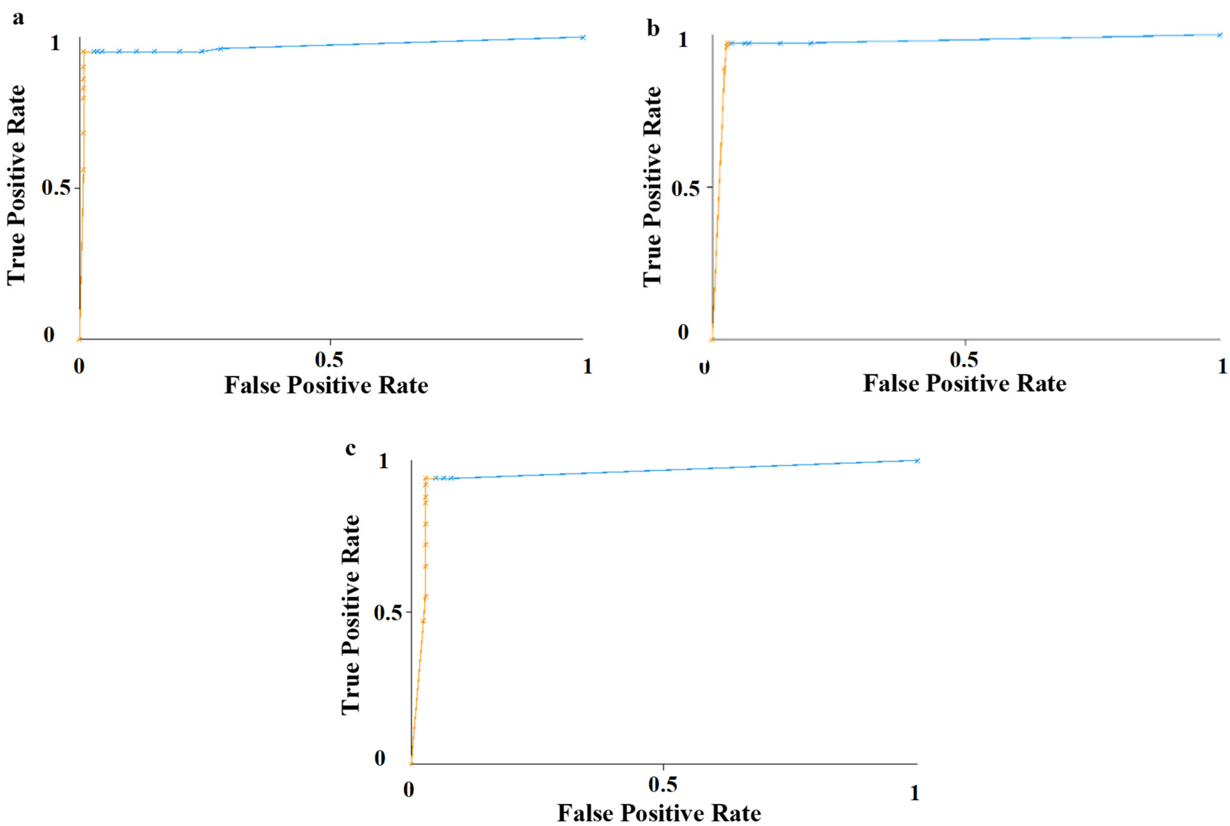

Figure 3. The ROC (Receiver Operating Characteristic) curves for the cultivar discrimination of plum kernels for a model built using the PART classifier based on textures selected from all color channels, (**a**) 'Emper', (**b**) 'Kalipso', (**c**) 'Polinka'.

Table 2. The results of cultivar discrimination of plum kernels performed using models built based on textures selected from all color channels *L*, *a*, *b*, *R*, *G*, *B*, *U*, *V*, *S*, *X*, *Y*, *Z*, color space Lab, and color channel *b* using the PART (Rules) classifier.

Set of Selected Textures	Predicted Class (%) 'Emper'	'Kalipso'	'Polinka'	Actual Class	Average Accuracy (%)	Precision	F-Measure	MCC	ROC Area	PRC Area
all color channels	95	1	4	'Emper'	95	0.979	0.964	0.947	0.965	0.941
	1	97	2	'Kalipso'		0.942	0.956	0.933	0.969	0.928
	1	5	94	'Polinka'		0.940	0.940	0.910	0.948	0.879
color space Lab	87	1	12	'Emper'	88	0.870	0.870	0.805	0.943	0.912
	4	91	5	'Kalipso'		0.948	0.929	0.894	0.963	0.945
	9	4	87	'Polinka'		0.837	0.853	0.778	0.935	0.845
color channel *b*	87	4	9	'Emper'	87	0.879	0.874	0.812	0.921	0.827
	3	92	5	'Kalipso'		0.885	0.902	0.852	0.952	0.900
	9	8	83	'Polinka'		0.856	0.843	0.766	0.899	0.810

MCC—Matthews Correlation Coefficient; ROC Area—Receiver Operating Characteristic Area; PRC Area—Precision-Recall Area.

The average accuracies of discrimination of 'Emper', 'Kalipso', and 'Polinka' plum kernels for models built using the LMT classifier were very high for a set of textures selected from all color channels (95%) as well as color space Lab (94%) and color channel *b* (92%) (Table 3). The individual cultivars were discriminated with the accuracies from 92 ('Emper') to 96% ('Kalipso', 'Polinka') for the model developed for textures from all color channels, 92 ('Emper') to 97% ('Polinka') for color space Lab and 90 ('Kalipso') to 94% ('Emper') for color channel *b*. The Precision, F-Measure, MCC, ROC Area, PRC Area, reached 0.979 ('Emper', all color channels and 'Kalipso', color space Lab), 0.959 ('Kalipso', color space Lab), 0.940 ('Kalipso', color space Lab), 0.985 ('Kalipso', color space Lab, color channel *b*), 0.979 ('Kalipso', color space Lab), respectively. The course of ROC curves was the smoothest in the case of the 'Kalipso' plum kernels (Figure 4).

Table 3. The results of cultivar discrimination of plum kernels performed using models built based on textures selected from all color channels *L*, *a*, *b*, *R*, *G*, *B*, *U*, *V*, *S*, *X*, *Y*, *Z*, color space Lab, and color channel *b* using the LMT (Trees) classifier.

Set of Selected Textures	Predicted Class (%) 'Emper'	'Kalipso'	'Polinka'	Actual Class	Average Accuracy (%)	Precision	F-Measure	MCC	ROC Area	PRC Area
all color channels	92	3	5	'Emper'	95	0.979	0.948	0.925	0.969	0.945
	1	96	3	'Kalipso'		0.941	0.950	0.925	0.979	0.953
	1	3	96	'Polinka'		0.923	0.941	0.911	0.978	0.961
color space Lab	92	1	7	'Emper'	94	0.948	0.934	0.902	0.975	0.969
	3	94	3	'Kalipso'		0.979	0.959	0.940	0.985	0.979
	2	1	97	'Polinka'		0.907	0.937	0.905	0.968	0.896
color channel *b*	94	1	5	'Emper'	92	0.940	0.940	0.910	0.977	0.969
	3	90	7	'Kalipso'		0.938	0.918	0.879	0.985	0.978
	3	5	92	'Polinka'		0.885	0.902	0.852	0.971	0.899

MCC—Matthews Correlation Coefficient; ROC Area—Receiver Operating Characteristic Area; PRC Area—Precision-Recall Area.

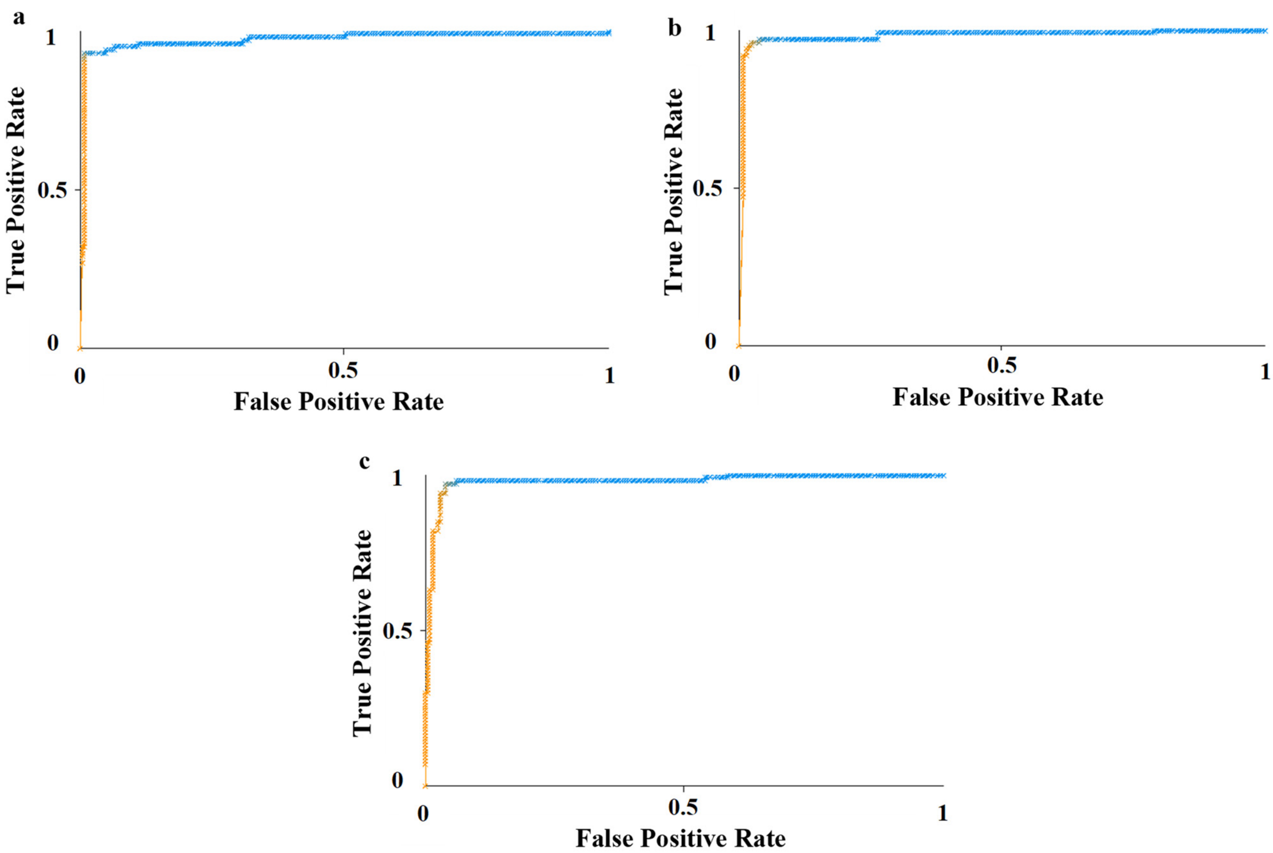

Figure 4. The ROC (Receiver Operating Characteristic) curves for the cultivar discrimination of plum kernels for a model built using the LMT classifier based on textures selected from all color channels, (**a**) 'Emper', (**b**) 'Kalipso', (**c**) 'Polinka'.

4. Discussion

This study presents machine learning approaches to cultivar discrimination of plum kernels. The obtained results objectively distinguished the kernels with high accuracy. This discrimination, which was carried out using only computers, can easily be used instead of manual techniques in laboratories since it is non-destructive and inexpensive. In this way, the production speed can be increased, and the producer and the consumer can be informed about the kernel cultivar and the quality of the product. Artificial intelligence-based methods produce completely data-driven results. Therefore, for successful discrimination, the dataset must be created before learning-based methods can be used, and the data and classes must have a regular distribution. Incorrect creation of datasets, images containing different features depending on lighting conditions, differences in the background, etc. reduce the performance of artificial intelligence studies. It is also very important which features to use in classification for machine learning methods. Appropriate feature selection should be decided together with an expert in that field. The texture features used in this study successfully represented plum kernels belonging to different classes. However, extracting different features can improve or decrease the current classification performance. Therefore, which features are appropriate for the existing data is a problem for machine learning. Extracting too many different features may also not improve discrimination ability. Due to such problems, deep learning-based methods have started to be applied quite a lot lately. Because deep learning extracts features from data hierarchically and learns high-level features in the last layers. However, in general, deep learning methods also require more data than machine learning [23,24].

The present study is an extension of the current directions of the application of machine learning in agricultural research. Computer vision solutions and artificial intelligence algorithms can be useful to recognize patterns in images, reduce subjectivity, and optimize the analysis process. The interactive and traditional machine learning approach was successfully applied, e.g., to classify soybean seeds and seedlings based on their appearance and physiological potential [25]. Machine learning tools may also be useful to predict the seed yield, 1000 seed weight, protein and oil yield and content based on the genotype and production year that is an important agricultural challenge necessary for stakeholders, producers, and the global trade market [26]. In the case of watermelon seeds, literature data indicated that the application of deep learning may increase the discrimination accuracy. The comparison of the performances of models developed using conventional machine learning and deep learning for the classification of watermelon seeds showed higher results for ResNet-50 (87.3%) than the LDA (83.6%) [27]. Machine learning proved to be useful to speed up the evaluation of germination of seeds belonging to different cultivars and to achieve higher results and performance than manual and conventional methods [28]. LDA-based machine learning models allowed for quick and robust discrimination of *Jatropha curcas* seed into classes related to germination capacity, speed, viability, and seedling vigor [29].

Robust, precise, high-throughput, and nondestructive analyses using machine learning algorithms can be very important for cultivar detection and seed quality evaluation [30,31]. Machine learning including classification, prediction, and clustering can be used, e.g., in the food industry, seed industry, or to forecast crop production in the field. Machine learning algorithms can improve the decision-support system. Due to the many possibilities of application, machine learning can be used even more extensively in the future [32].

5. Conclusions

The two-staged procedure including the image processing and discriminant analysis proved to be an objective, non-destructive, and inexpensive technique for the cultivar discrimination of plum kernels ('Emper', 'Kalipso', and 'Polinka'). The MaZda software was oriented toward image analysis involving calculations of the texture features on kernel surface, expecting to convert the image signals to feature parameters. A variety of mathematical methods, such as co-occurrence matrix, run-length matrix, Haar wavelet transform, gradient map, autoregressive model, and histogram, were successfully used to extract the texture parameters for well-matched models of cultivar discrimination.

The discriminative models based on selected textures were developed using machine learning classifiers to distinguish plum kernel cultivars. The selected texture features of the images of three plum cultivars from color space Lab (one of the individual color spaces) and color channel *b* (one of color channels) showed the highest discriminative performance with great average accuracies in the case of the KStar machine learning algorithm. This algorithm provided also smooth and steady ROC curves for all the cultivars. The study substantiated that the KStar classifier is the optimal algorithm to discriminate the three cultivars of plum kernels with high performance metrics including accuracies.

The developed models can be applied in practice for the plum kernel cultivar recognition. It may be useful in the processing industry to avoid cultivar mixing and falsification. Due to the usefulness of machine learning for the discrimination of kernels, future studies may be performed for other species and cultivars. The correctness of discrimination could be improved by including also color and geometric features in the models. The small number of kernels may be a limitation of this study and a threat to the development of accurate models. The application of deep learning would be possible if more seeds were available.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture12020285/s1, Table S1: Textures for 10 sample kernels of each cultivar.

Author Contributions: Conceptualization, E.R.; Methodology, E.R.; Software, E.R.; Validation, E.R., X.C., Z.Z., K.S. and M.F.A.; Formal analysis, E.R.; Investigation, E.R.; Resources, E.R.; Data curation, E.R.; Writing—original draft preparation, E.R., X.C., K.S. and M.F.A.; Writing—review and editing, E.R., X.C., Z.Z., K.S. and M.F.A.; Visualization, E.R.; Supervision, E.R. and K.S.; Project administration, Z.Z.; Funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Youth Innovative Talents Project of Guangdong Province, China, the funder: Shenzhen MSU-BIT University, China, the funding number: 2021WQNCX084.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Horvath, A.; Balsemin, E.; Barbot, J.; Christmann, H.; Manzano, G.; Reynet, P.; Laigret, F.; Mariette, S. Phenotypic variability and genetic structure in plum (*Prunus domestica* L.), cherry plum (*P. cerasifera* Ehrh.) and sloe (*P. spinosa* L.). *Sci. Hortic.* **2011**, *129*, 283–293. [CrossRef]
2. Polák, J.; Kundu, J.K.; Krška, B.; Beoni, E.; Komínek, P.; Pívalova, J.; Jarošová, J. Transgenic plum *Prunus domestica* L., clone C5 (cv. HoneySweet) for protection against sharka disease. *J. Integr. Agric.* **2017**, *16*, 516–522. [CrossRef]
3. Szalay, L.; Molnár, Á.; Kovács, S. Frost hardiness of flower buds of three plum (*Prunus domestica* L.) cultivars. *Sci. Hortic.* **2017**, *214*, 228–232. [CrossRef]
4. Sheikh, M.A.; Saini, C.S.; Sharma, H.K. Analyzing the effects of hydrothermal treatment on antinutritional factor content of plum kernel grits by using response surface methodology. *Appl. Food Biotechnol.* **2021**, *1*, 100010. [CrossRef]
5. Ali, M.; Cha, J.; Park, S.J.; Akhter, T.; Kim, G.S.; Yang, K.; Seonwoo, H.; Kim, H. Development and performance evaluation of automatic plum seed remover. *Eng. Agric. Environ. Food.* **2019**, *12*, 250–255. [CrossRef]
6. Anwar, F.; Manzoor, M.; Bukhari, I.H.; Aladedunye, F. Physico-chemical attributes of fruit seed oils from different varieties of peach and plum. *J. Adv. Biol.* **2014**, *4*, 384–392.
7. Górnaś, P.; Rudzińska, M.; Raczyk, M.; Mišina, I.; Soliven, A.; Lacis, G.; Segliņa, D. Impact of Species and Variety on Concentrations of Minor Lipophilic Bioactive Compounds in Oils Recovered from Plum Kernels. *J. Agric. Food Chem.* **2016**, *64*, 898–905. [CrossRef]
8. Savic, I.; Gajic, I.S.; Gajic, D. Physico-chemical properties and oxidative stability of fixed oil from plum seeds (*Prunus domestica* Linn.). *Biomolecules* **2020**, *10*, 294. [CrossRef]
9. Korir, N.K.; Han, J.; Shangguan, L.; Wang, C.; Kayesh, E.; Zhang, Y.; Fang, J. Plant variety and cultivar identification: Advances and prospects. *Crit. Rev. Biotechnol.* **2013**, *33*, 1–15. [CrossRef]
10. Osako, Y.; Yamane, H.; Lin, S.Y.; Chen, P.A.; Tao, R. Cultivar discrimination of litchi fruit images using deep learning. *Sci. Hortic.* **2020**, *269*, 10936. [CrossRef]
11. Ropelewska, E.; Sabanci, K.; Aslan, M.F. Discriminative Power of Geometric Parameters of Different Cultivars of Sour Cherry Pits Determined Using Machine Learning. *Agriculture* **2021**, *11*, 1212. [CrossRef]
12. Zhou, Q.; Huang, W.; Fan, S.; Zhao, F.; Liang, D.; Tian, X. Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm. *Infrared Phys. Technol.* **2020**, *109*, 103418. [CrossRef]
13. Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. *Symmetry* **2020**, *12*, 2018. [CrossRef]
14. Javanmardi, S.; Miraei Ashtiani, S.-H.; Verbeek, F.J.; Martynenko, A. Computer-vision classification of corn seed varieties using deep convolutional neural network. *J. Stored Prod. Res.* **2021**, *92*, 101800. [CrossRef]
15. Ropelewska, E. Classification of the pits of different sour cherry cultivars based on the surface textural features. *J. Saudi Soc. Agric. Sci.* **2020**, *20*, 52–57. [CrossRef]
16. Ropelewska, E. The Application of Machine Learning for Cultivar Discrimination of Sweet Cherry Endocarp. *Agriculture* **2021**, *11*, 6. [CrossRef]
17. Ropelewska, E.; Rutkowski, K.P. Differentiation of peach cultivars by image analysis based on the skin, flesh, stone and seed textures. *Eur. Food Res. Technol.* **2021**, *247*, 2371–2377. [CrossRef]
18. Ucchesu, M.; Sarigu, M.; Del Vais, C.; Sanna, I.; d'Hallewin, G.; Grillo, O.; Bacchetta, G. First finds of *Prunus domestica* L. in Italy from the Phoenician and Punic periods (6th–2nd centuries bc). *Veget. Hist. Archaeobot.* **2017**, *26*, 539–549. [CrossRef]
19. Szczypiński, P.M.; Strzelecki, M.; Materka, A.; Klepaczko, A. MaZda-A software package for image texture analysis. *Comput. Methods Programs Biomed.* **2009**, *94*, 66–76. [CrossRef]
20. Bouckaert, R.R.; Frank, E.; Hall, M.; Kirkby, R.; Reutemann, P.; Seewald, A.; Scuse, D. *WEKA Manual for Version 3-9-1*; University of Waikato: Hamilton, New Zealand, 2016; pp. 1–341.

21. Frank, E.; Hall, M.A.; Witten, I.H. The WEKA Workbench. In *Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann*, 4th ed.; Elsevier: San Francisco, CA, USA, 2016; pp. 1–128.
22. Witten, I.H.; Frank, E. *Data mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Elsevier: San Francisco, CA, USA, 2005; pp. 1–558.
23. Aslan, M.F.; Sabanci, K.; Durdu, A. A CNN-based novel solution for determining the survival status of heart failure patients with clinical record data: Numeric to image. *Biomed. Signal Process. Control* **2021**, *68*, 102716. [CrossRef]
24. Aslan, M.F.; Unlersen, M.F.; Sabanci, K.; Durdu, A. CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. *Appl. Soft Comput.* **2021**, *98*, 106912. [CrossRef] [PubMed]
25. de Medeiros, A.D.; Capobiango, N.P.; da Silva, J.M.; da Silva, L.J.; da Silva, C.B.; dos Santos Dias, D.C.F. Interactive machine learning for soybean seed and seedling quality classification. *Sci. Rep.* **2020**, *10*, 11267. [CrossRef] [PubMed]
26. Rajković, D.; Marjanović Jeromela, A.; Pezo, L.; Lončar, B.; Zanetti, F.; Monti, A.; Kondić Špika, A. Yield and Quality Prediction of Winter Rapeseed—Artificial Neural Network and Random Forest Models. *Agronomy* **2022**, *12*, 58. [CrossRef]
27. Ahmed, M.R.; Yasmin, J.; Park, E.; Kim, G.; Kim, M.S.; Wakholi, C.; Mo, C.; Cho, B.-K. Classification of Watermelon Seeds Using Morphological Patterns of X-ray Imaging: A Comparison of Conventional Machine Learning and Deep Learning. *Sensors* **2020**, *20*, 6753. [CrossRef]
28. Genze, N.; Bharti, R.; Grieb, M.; Schultheiss, S.J.; Grimm, D.G. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods* **2020**, *16*, 157. [CrossRef]
29. de Medeiros, A.D.; Pinheiro, D.T.; Xavier, W.A.; da Silva, L.J.; dos Santos Dias, D.C.F. Quality classification of *Jatropha curcas* seeds using radiographic images and machine learning. *Ind. Crops Prod.* **2020**, *146*, 112162. [CrossRef]
30. Wei, Y.; Li, X.; Pan, X.; Li, L. Nondestructive Classification of Soybean Seed Varieties by Hyperspectral Imaging and Ensemble Machine Learning Algorithms. *Sensors* **2020**, *20*, 6980. [CrossRef]
31. Gao, T.; Chandran, A.K.N.; Paul, P.; Walia, H.; Yu, H. HyperSeed: An End-to-End Method to Process Hyperspectral Images of Seeds. *Sensors* **2021**, *21*, 8184. [CrossRef]
32. Sharma, M.; Kaushik, P.; Chawade, A. Frontiers in the Solicitation of Machine Learning Approaches in Vegetable Science Research. *Sustainability* **2021**, *13*, 8600. [CrossRef]