The Application of Machine Learning for Cultivar Discrimination of Sweet Cherry Endocarp

Ewa Ropelewska

Fruit and Vegetable Storage and Processing Department, Research Institute of Horticulture, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland; ewa.ropelewska@inhort.pl

Abstract: The aim of this study was to evaluate the usefulness of the texture and geometric parameters of endocarp (pit) for distinguishing different cultivars of sweet cherries using image analysis. The textures from images converted to color channels and the geometric parameters of the endocarp (pits) of sweet cherry ‘Kordia’, ‘Lapins’, and ‘Büttner’s Red’ were calculated. For the set combining the selected textures from all color channels, the accuracy reached 100% when comparing ‘Kordia’ vs. ‘Lapins’ and ‘Kordia’ vs. ‘Büttner’s Red’ for all classifiers. The pits of ‘Kordia’ and ‘Lapins’, as well as ‘Kordia’ and ‘Büttner’s Red’ were also 100% correctly discriminated for discriminative models built separately for RGB, Lab and XYZ color spaces, G, L and Y color channels and for models combining selected textural and geometric features. For discrimination ‘Lapins’ and ‘Büttner’s Red’ pits, slightly lower accuracies were determined—up to 93% for models built based on textures selected from all color channels, 91% for the RGB color space, 92% for the Lab and XYZ color spaces, 84% for the G and L color channels, 83% for the Y channel, 94% for geometric features, and 96% for combined textural and geometric features.

Keywords: sweet cherry cultivars; texture parameters; geometric features; discriminative classifiers

1. Introduction

Sweet cherry (Prunus avium L.) belongs to the Rosaceae family. Cherry fruits consist of the skin (exocarp), the flesh (mesocarp), and the pit (endocarp), including the seed [1]. The endocarp protects the seed and its function is to ensure a secure environment for the development of the seed. Endocarp hardening is a determinant of the maturation of drupe fruits [2]. Cherry pits can be considered as a waste product, often used in biomass production [3]. However, pits can be valuable products with industrial potential. The pits contain crude oil, proteins, crude fibers, phenols, carbohydrates, vitamins, and minerals [4]. The pits of cherries can be used in the food industry, for example, in the production of alcoholic beverages [4] and for the production of edible oil that can improve the quality of food products [5]. The cherry seed oil has a high fat content and unsaturated fatty acid composition. The oil obtained from the cherry seed is very important for human health and can be applied as a natural antioxidative additive. Unsaturated fatty acids can support the treatment of type 2 diabetes, as well as in the reduction of cancers, for example, colon cancer [5]. There is also a literature report on using the kernel oil from sweet cherry seeds, which contains more than 87% unsaturated fatty acids (oleic acid (43.7% by weight), linoleic acid (41.8% by weight), and linolenic acid) for producing biodiesel [6]. Sweet cherry kernel oil ingredients can also be used in the cosmetic industry [7]. Radenkovs and Feldmane [8] reported that the concentration of tocopherol and carotenoids in cherry kernels depends on the cultivar. Agil et al. [7] determined that the composition of sweet cherry kernel oil may differ depending on the cultivar. Due to the different chemical compositions of different cultivars of sweet cherry kernels, it may be desirable to select cultivars with suitable properties for further processing. It can be extremely important in practice, among others, in the food industry. The quality of the final food product...
depends on the chemical composition of the raw materials. So, raw materials with specific properties cannot be mixed. Therefore, to avoid the adulteration of pits with different chemical compositions, the pit cultivar may need to be identified.

Identification of the cultivars or subspecies of Prunus groups can be difficult due to the wide range of variation, as well as morphological transitional states. Detailed knowledge of the variations in endocarp dimension and shape are important. In some cases, the identification of diagnostic features to clearly distinguish the groups is difficult. The characteristics of the stones of Prunus taxa can be considered the most constant for identifying cultivars and species. The visual judgment of Prunus endocarps can be subjective and may result in the presence of classification errors [9]. An excellent alternative for the manual discrimination of endocarps of different Prunus sp. cultivars may be the use of an image analysis approach to eliminate the problems of subjectivity, labor and time consumption [10]. Image analysis can allow great accuracy, repeatability, and throughput at a relatively low cost [11,12]. Image processing can be used, for example, to determine the geometric parameters including linear dimensions, shape factors, and the moment of inertia [13]. In image processing, textural image analysis is also a very important issue, as it can allow the quantitative data for texture parameters to be determined [13]. Quantitative analysis of texture parameters can provide diagnostic information about product quality. Texture analysis provides numerical data computed from the image of the object. Texture parameters can even specify the changes that are difficult to relate to changes that are perceived visually. In their digital image form, objects can be characterized by different textures even if they are composed of the same number of pixels and the same color histograms, but have a dissimilar color distribution [14,15].

The aim of this study was to evaluate the usefulness of the texture and geometric parameters of endocarp (pit) for distinguishing different cultivars of sweet cherries using image analysis. The discriminative models developed using machine learning may be useful for identifying the cultivars of sweet cherry endocarps and, indirectly, the fruit cultivar based on the endocarp’s characteristics.

2. Materials and Methods

2.1. Materials

The research material comprised of pits of three sweet cherry cultivars ‘Kordia’, ‘Lapins’, and ‘Büttner’s Red’. These cultivars were selected considering the similarities and differences in the external appearance of the fruit. Two cultivars were characterized by a red skin (‘Kordia’—very dark red and ‘Lapins’—red) and one cultivar ‘Büttner’s Red’ have a light yellow skin covered by a red blush. Sweet cherries ‘Kordia’ and ‘Lapins’ were collected from the Experimental Orchard in Dąbrowice (central Poland). Sweet cherries ‘Büttner’s Red’ were harvested from a local orchard located in north-eastern Poland. The pits were manually extracted from sweet cherries. The pits were washed and cleaned to remove fruit flesh and air-dried.

2.2. Image Analysis

The pit (endocarp) image acquisition was performed with the use of an Epson Perfection flatbed scanner. Six hundred pits including 200 pits of ‘Kordia’, 200 pits of ‘Lapins’ and 200 pits of ‘Büttner’s Red’ were scanned on the dorsal and ventral sides. The pits were scanned on a black background at a resolution of 1200 dpi. The images were saved in TIFF format. The image processing was carried out using MaZda software (Łódź University of Technology, Institute of Electronics, Łódź, Poland) [13]. The image analysis steps are presented in Figure 1.
The regions of interest (ROIs) were identified using the segmentation of the images. Each ROI including a pit was a set of pixels separated from the black background using the manually determined brightness threshold. The used version of MaZda application allowed to convert the images to individual color channels. Exemplary images of sweet cherry pits ‘Kordia’, ‘Lapins’ and ‘Büttner’s Red’ from selected color channels are presented in Figure 2. Based on the converted images, the suitability of individual color channels for further discriminant analysis was visually examined. In the case of $R$, $G$, $B$, $L$, $a$, $b$, $X$, $Y$, and $Z$ color channels, the differences between the pits of individual cultivars allowing for their distinction, were observed. Therefore, images from these color channels were chosen for the texture calculation. Color channels $R$, $G$, $B$ were from RGB color space, color channels $L$, $a$, $b$-from Lab color space and $X$, $Y$ and $Z$-from XYZ color space. The RGB color space refers to red ($R$), blue ($B$) and green ($G$). The Lab color space depends on the position in a 3D color space. The $L^*$ is the lightness component from black to white, $a^*$ indicates green for negative values and red for positive values and $b^*$ indicates blue for negative values and yellow for positive values. In the case of XYZ color space, $Y$ indicates lightness and $X$ and $Z$ components are color information [16].

Figure 2. Images of sweet cherry pits from different color channels.
Almost 200 texture features based on the run-length matrix (20 texture parameters including 5 features computed for 4 various directions), co-occurrence matrix (132 texture parameters including 11 features computed for 4 various directions and 3 between-pixels distances), gradient map (5 texture parameters), histogram (9 texture parameters), Haar wavelet transform (20 texture parameters), and autoregressive model (5 texture parameters) were computed for each ROI in each color channel [13]. For the calculation of geometric parameters, a caliper was scanned and used to calibrate. The geometric parameters including linear dimensions, shape factors, and moments of inertia determined for each endocarp using MaZda software are presented in Table 1. The geometric parameters were described in detail by Zapotoczny [17].

Table 1. The geometric parameters including linear dimensions, shape factors, and moments of inertia.

| Parameters | Shape Factors | Moments of Inertia |
|------------|---------------|-------------------|
| Linear Dimensions | | |
| L-length | SigR-standard deviation of all radii | M2x-horizontal second order moment of inertia |
| S-width | RjF-Haralick ratio | M2y-vertical second order moment of inertia |
| Lsz-length of the skeletonized object | RjL-Blair—Bliss ratio | M2xy-second order moment of inertia |
| FE-area of circumscribing ellipse on the object | RjM-Malinowska ratio | |
| Lmax-maximal length of the ellipse axis on the object | RjF-Feret ratio (Fh/Fv) | |
| Lmin-minimal length of the ellipse axis on the object | Rjmax-Feret ratio (Fmax/Fmin) | |
| Fd2-area of circumscribing circle | Rj-circularity (Rj1/Rj2) | |
| D1-radius of circumscribing circle | Rj1-circularity (2F/πF) | |
| UL-profile specific perimeter | Rj2-circularity (ULj/π) | |
| Mmax-Martin’s maximal radius | W1-elliptic shape factor | |
| Mmin-Martin’s minimal radius | W1-circular shape factor | |
| Fv-vertical Feret diameter | W1-circularity | |
| Uc-convex perimeter | W1-folding factor | |
| Uc1-object boundary specific perimeter | W1g-mean thickness factor | |
| Snec-equivalent circular area diameter | W7-elongation and irregularity ratio | |
| Sxl-area of circumscribing rectangle (length x width) | W5-rectangular aspect ratio | |
| S2-minimal width | W5-area ratio | |
| Fs-total object specific area | W10-radius ratio | |
| Fh-horizontal Feret diameter | W11-radius range | |
| Fmax-maximal Feret diameter | W12-roundness ((4F)/πSmax2)) | |
| Fmin-minimal Feret diameter | W13-roundness (Smax/F) | |
| W14-roundness (F/Smax2) | W15-roundness (4F/(πSmin Smax)) | |

2.3. Statistical Analysis

A discriminant analysis using WEKA 3.9 machine learning software (Machine Learning Group, University of Waikato, Hamilton, New Zealand) [18] was carried out to classify the pits belong to different cultivars based on selected textural and geometric features. The following classes were compared together: ‘Kordia’ vs. ‘Lapins’, ‘Kordia’ vs. ‘Büttner’s Red’, ‘Lapins’ vs. ‘Büttner’s Red’, ‘Kordia’ vs. ‘Lapins’ vs. ‘Büttner’s Red’. Pairwise comparisons were performed to check whether there are differences between the pits of two individual cultivars that distinguish the pits of one cultivar from the pits of second cultivar. In the case of discrimination of pits of three cultivars, it was checked whether it is also possible to group mixed pits belonging to more than two cultivars into separate classes with high accuracy. The sets of textures or geometric features and their number were determined using the Best First with the CFS (Correlation-based Feature Selection algorithm) subset evaluator. The Best First is a search method for attribute selection performing greedy hill-climbing with backtracking. The CFS considers the predictive value of individual attributes and the degree of redundancy among these attributes [19]. For the discrimination of pits of different sweet cherry cultivars, the 10-fold cross-validation mode and the discriminative classifiers from groups of Bayes (NaiveBayes, BayesNet), Functions (Logistic, MultilayerPerceptron), Meta, Rules (JRip, PART) and Decision Trees (J48, LMT)
were applied. NaiveBayes is a probabilistic classifier, BayesNet learns Bayesian networks, Logistic uses linear logistic regression models, MultilayerPerceptron is based on a neural network with backpropagation, JRip uses RIPPER algorithm for the rule implementation, PART applies partial decision trees for rule induction, J48 generates decision trees using C4.5 algorithm, LMT uses logistic model trees [19]. In the first step, a discriminant analysis was performed for models built based on a set of textures selected from all color channels (R, G, B, L, a, b, X, Y, Z). In the following step, the discriminant models were built separately for RGB, Lab and XYZ color spaces. Then, the models were built based on selected textures for one color channel for each color space, for which the accuracy was the highest. In the next steps, the discriminant models included the selected geometric features. The next models were built based on sets combining the selected textural and geometric features.

3. Results

In the case of discrimination of the endocarp of different sweet cherry cultivars performed based on sets of textures selected from all color channels (R, G, B, L, a, b, X, Y, Z), several dozen texture parameters (about 50) were selected for each analysis. For discrimination, analyses were carried out using textures from individual color spaces, also several dozen features (20–30) were chosen. Several textures, e.g., HMean, HVariance, HPerc90, HPerc99, SGNonZeros repeated in each case. The predominant features with the highest discriminant power were from color channels G, L, and Y. In the case of the discrimination performed for individual color channels, the smallest number of textures (10–20) was selected. For discriminative models built based on geometric features, several to over a dozen parameters were selected.

The discriminant analysis was performed using two discriminative classifiers from groups of Bayes, Functions, Meta, Rules and Decision Trees and it was observed that the accuracies were highest for Bayes Net, Logistic, Multi Class Classifier, PART, and LMT (Logistic Model Tree), respectively. Therefore, only the results obtained for these classifiers were included in this paper. Other classifiers provide an accuracy of a few percent lower.

The discriminant analysis performed based on sets including the endocarp textures selected from all analyzed color channels (R, G, B, L, a, b, X, Y, Z) (Table 2) revealed the accuracy of up to 100% for comparison of ‘Kordia’ vs. ‘Lapins’ and ‘Kordia’ vs. ‘Büttner’s Red’ in the case of all classifiers. It indicates that in the case of discrimination of ‘Kordia’ and ‘Lapins’, all pits belong to sweet cherry ‘Kordia’ were correctly assigned to class “Kordia” and all pits ‘Lapins’ were correctly classified as “Lapins”. In the case of discriminant analysis of ‘Kordia’ vs. ‘Büttner’s Red’, all pits belong to ‘Kordia’ and ‘Büttner’s Red’ were correctly classified as ‘Kordia’ and ‘Büttner’s Red’, respectively. These results are highly satisfactory. For distinguishing the sweet cherry pits of ‘Lapins’ vs. ‘Büttner’s Red’, the correctness ranged from 87% (PART classifier) to 93% (LMT classifier) was observed. The pits of all three sweet cherry cultivars (‘Kordia’, ‘Lapins’, ‘Büttner’s Red’) were discriminated with an accuracy of 92% (Bayes Net classifier) to 95% (LMT classifier).

Table 2. The accuracy of discrimination of the endocarp (pit) of different sweet cherry cultivars based on sets of textures selected from all color channels (R, G, B, L, a, b, X, Y, Z).

| Classifier            | ‘Kordia’ vs. ‘Lapins’ | ‘Kordia’ vs. ‘Büttner’s Red’ | ‘Lapins’ vs. ‘Büttner’s Red’ | ‘Kordia’ vs. ‘Lapins’ vs. ‘Büttner’s Red’ |
|-----------------------|-----------------------|-----------------------------|-----------------------------|------------------------------------------|
| bayes.BayesNet        | 100                   | 100                         | 90                          | 92                                       |
| functions.Logistic    | 100                   | 100                         | 92                          | 94                                       |
| meta.Multi Class Classifier | 100                 | 100                         | 92                          | 94                                       |
| rules.PART            | 100                   | 100                         | 87                          | 93                                       |
| trees.LMT             | 100                   | 100                         | 93                          | 95                                       |
In the case of discrimination of sweet cherry endocarp carried out using models built separately for color spaces (RGB, Lab, and XYZ) (Table 3), 100% accuracy was achieved for analyses of ‘Kordia’ vs. ‘Lapins’ and ‘Kordia’ vs. ‘Büttner’s Red’, for all color spaces and all classifiers. For RGB color space, the accuracies of 86% (PART classifier) to 91% (Logistic and Multi Class Classifier) were reached for discrimination of pits ‘Lapins’ and ‘Büttner’s Red’. The pits belong to three cultivars were correctly classified in 90% (Bayes Net classifier) to 94% (Logistic, Multi Class Classifier, and LMT) for RGB color space. In the case of Lab color space, the pits ‘Lapins’ and ‘Büttner’s Red’ were distinguished with the correctness of 87% (Bayes Net) to 92% (LMT). All three cultivars ‘Kordia’, ‘Lapins’, and ‘Büttner’s Red’ were characterized by 91% (Bayes Net) to 94% (Logistic, Multi Class Classifier and LMT) accuracies of pit discrimination. In the case of XYZ color space, discrimination accuracies for ‘Lapins’ and ‘Büttner’s Red’ ranged from 83% (Bayes Net) to 92% (Multi Class Classifier), and for ‘Kordia’, ‘Lapins’, and ‘Büttner’s Red’, the accuracies were in the range of 89% (PART) to 92% (Multi Class Classifier, LMT). These tendencies are similar to discriminant analysis performed for models built based on textures selected from all analyzed color channels (R, G, B, L, a, b, X, Y, Z) (Table 2). The correctness was the highest for discriminations of ‘Kordia’ vs. ‘Lapins’ and ‘Kordia’ vs. ‘Büttner’s Red’, and the lowest accuracy was obtained for discrimination of ‘Lapins’ vs. ‘Büttner’s Red’.

Table 3. The accuracy of discrimination of the endocarp (pit) of different sweet cherry cultivars based on textures selected from individual color spaces.

| Classifier         | 'Kordia' vs. 'Lapins' | 'Kordia' vs. 'Büttner’s Red' | 'Lapins' vs. 'Büttner’s Red' | 'Kordia' vs. 'Lapins' vs. 'Büttner’s Red' |
|--------------------|-----------------------|-------------------------------|------------------------------|-----------------------------------------|
| RGB color space    |                       |                               |                              |                                         |
| bayes.BayesNet     | 100                   | 100                           | 87                           | 90                                      |
| functions.Logistic | 100                   | 100                           | 91                           | 94                                      |
| meta.Multi Class Classifier | 100                | 100                           | 91                           | 94                                      |
| rules.PART         | 100                   | 100                           | 86                           | 91                                      |
| trees.LMT          | 100                   | 100                           | 89                           | 94                                      |
| Lab color space    |                       |                               |                              |                                         |
| bayes.BayesNet     | 100                   | 100                           | 87                           | 91                                      |
| functions.Logistic | 100                   | 100                           | 90                           | 94                                      |
| meta.Multi Class Classifier | 100              | 100                           | 90                           | 94                                      |
| rules.PART         | 100                   | 100                           | 90                           | 92                                      |
| trees.LMT          | 100                   | 100                           | 92                           | 94                                      |
| XYZ color space    |                       |                               |                              |                                         |
| bayes.BayesNet     | 100                   | 100                           | 83                           | 90                                      |
| functions.Logistic | 100                   | 100                           | 89                           | 90                                      |
| meta.Multi Class Classifier | 100             | 100                           | 92                           | 92                                      |
| rules.PART         | 100                   | 100                           | 84                           | 89                                      |
| trees.LMT          | 100                   | 100                           | 87                           | 92                                      |

In the next stages of the study, the discriminations were performed for individual channels of each color space. The results for one classifier producing the highest accuracies for each color channel were presented in Table 4. In the case of individual color spaces, the following color channels were characterized by the highest results: color channel G from RGB color space, color channel L from Lab color space and color channel Y from XYZ color space. The pits of sweet cherry ‘Kordia’ vs. ‘Lapins’ and ‘Kordia’ vs. ‘Büttner’s
Red’ were fully correctly (100%) discriminated for all analyzed color channels G, L and Y for all classifiers. The lowest accuracy (79%) was noted in the comparison of pits ‘Lapins’ and ‘Büttner’s Red’ for color channel Y and PART classifier. In the case of these cultivars for discriminative models based on the textures from color channel Y, the discrimination correctness reached 83% for Logistic and Multi Class Classifier. The discrimination accuracy of pits of ‘Lapins’ and ‘Büttner’s Red’ for the models built based on selected textural features form channel L was in the range of 82% (Bayes Net, Logistic, Multi Class Classifier) to 84% (LMT), and for textures selected from color channel G, the accuracy ranged from 83% (PART, LMT) to 84% (Bayes Net, Logistic, Multi Class Classifier). In the case of discrimination of pits belong to three sweet cherry cultivars (‘Kordia’, ‘Lapins’, ‘Büttner’s Red’), the accuracy reached 90% in the case of textures selected from the color channel G for Multi Class Classifier. The lowest correctness equal to 88% was achieved for the Bayes Net classifier for color channel G, for the Bayes Net and Logistic classifiers for color channel L, and the Logistic classifier for color channel Y.

Table 4. The comparison of discrimination accuracies of the endocarp (pit) of different sweet cherry cultivars for separate color channels R, G, and B from RGB color space, color channels L, a, and b from Lab color space, color channels X, Y, and Z from XYZ color space for one the best classifier.

| Classifier              | Accuracy (%)                                                          |
|-------------------------|-----------------------------------------------------------------------|
|                         | ‘Kordia’ vs. ‘Lapins’ | ‘Kordia’ vs. ‘Büttner’s Red’ | ‘Lapins’ vs. ‘Büttner’s Red’ | ‘Kordia’ vs. ‘Lapins’ vs. ‘Büttner’s Red’ |
| Color channel R         | 100 | 100 | 81 | 88 |
| Color channel G         | 100 | 100 | 84 | 90 |
| Color channel B         | 100 | 100 | 82 | 86 |
| Color channel L         | 100 | 100 | 84 | 89 |
| Color channel a         | 100 | 100 | 83 | 89 |
| Color channel b         | 100 | 100 | 83 | 88 |
| Color channel X         | 100 | 100 | 83 | 88 |
| Color channel Y         | 100 | 100 | 83 | 89 |
| Color channel Z         | 100 | 100 | 82 | 87 |

For the discrimination of sweet cherry pits based on selected geometric features, 100% accuracy was not observed (Table 5). The highest accuracy of 99% was noted for comparison of pits of ‘Kordia’ vs. ‘Lapins’ in the case of the Bayes Net and LMT classifiers. The pits ‘Kordia’ and ‘Büttner’s Red’ were classified with an accuracy in the range of 93% (PART) to 95% (Bayes Net, Logistic, Multi Class Classifier). The accuracy reaching 94% (Logistic, Multi Class Classifier, LMT) was observed for analysis of ‘Lapins’ and ‘Büttner’s Red’ and reaching 93% (LMT) for discriminant analysis of ‘Kordia’, ‘Lapins’, and ‘Büttner’s Red’.
Table 5. The accuracy of discrimination of the endocarp (pit) of different sweet cherry cultivars based on selected geometric features.

| Classifier                  | ‘Kordia’ vs. ‘Lapins’ | ‘Kordia’ vs. ‘Büttner’s Red’ | ‘Lapins’ vs. ‘Büttner’s Red’ | ‘Kordia’ vs. ‘Lapins’ vs. ‘Büttner’s Red’ |
|-----------------------------|-----------------------|-------------------------------|-------------------------------|------------------------------------------|
| bayes.BayesNet              | 99                    | 95                            | 87                            | 87                                       |
| functions.Logistic          | 98                    | 95                            | 94                            | 92                                       |
| meta.Multi Class Classifier | 98                    | 95                            | 94                            | 92                                       |
| rules.PART                  | 98                    | 93                            | 89                            | 87                                       |
| trees.LMT                   | 99                    | 94                            | 94                            | 93                                       |

The pits ‘Kordia’ with ‘Lapins’, as well as ‘Kordia’ with ‘Büttner’s Red’ were discriminated with an accuracy of up to 100% based on sets of combined selected textural and geometric parameters in the case of all applied classifiers (Table 6). The results for discrimination of ‘Lapins’ vs. ‘Büttner’s Red’ and ‘Kordia’ vs. ‘Lapins’ vs. ‘Büttner’s Red’ were also highly satisfactory and were equal up to 96% (LMT) and 98% (Logistic), respectively. It proved the usefulness of both the textures and the geometric features for distinguishing the sweet cherry pits of different cultivars.

Table 6. The accuracy of discrimination of the endocarp (pit) of different sweet cherry cultivars based on sets comprised of selected textural and geometric features.

| Classifier                  | ‘Kordia’ vs. ‘Lapins’ | ‘Kordia’ vs. ‘Büttner’s Red’ | ‘Lapins’ vs. ‘Büttner’s Red’ | ‘Kordia’ vs. ‘Lapins’ vs. ‘Büttner’s Red’ |
|-----------------------------|-----------------------|-------------------------------|-------------------------------|------------------------------------------|
| bayes.BayesNet              | 100                   | 100                           | 95                            | 96                                       |
| functions.Logistic          | 100                   | 100                           | 95                            | 98                                       |
| meta.Multi Class Classifier | 100                   | 100                           | 95                            | 97                                       |
| rules.PART                  | 100                   | 100                           | 91                            | 94                                       |
| trees.LMT                   | 100                   | 100                           | 96                            | 97                                       |

The scatter plots for exemplary textures and geometric features of the endocarp of individual sweet cherry cultivars are presented in Figure 3. Each scatter plot shows two randomly selected parameters. The scatter plots indicated that the selected textures may have full discriminative power. The complete distinction between the endocarp of sweet cherry cultivars may be possible by using textures selected from one color space, textures selected from one color channel, as well as texture selected from different color spaces. The full discrimination of ‘Kordia’ and ‘Büttner’s Red’, as well as ‘Kordia’ and ‘Lapins’ for selected variables is confirmed. An exemplary scatter plot based on two textures from RGB color space (RHMean and BHMean) for endocarp ‘Kordia’ and ‘Büttner’s Red’ revealed the complete separation between these groups (Figure 3a). In the case of selected textures (YH Variance and YHMean) from one color channel (Y), the complete distinction of endocarp of two cultivars ‘Kordia’ and ‘Lapins’ was observed (Figure 3b). The scatter plot confirmed the possibility of distinguishing two sweet cherry cultivars (e.g., ‘Kordia’ and ‘Büttner’s Red’) using textures from two color spaces (texture XSGNonZeros from XYZ color space and texture GHPerc90 from RGB color space) (Figure 3c). In the case of scatter plot including the values of two geometric features, e.g., for endocarp ‘Kordia’ and ‘Lapins’, almost complete separation is shown, and only a few cases are mixed (Figure 3d).
4. Discussion

The results of the study proved the usefulness of machine learning for discrimination of different sweet cherry cultivars based on texture and geometric parameters of the endocarp. The research revealed that some cultivars may be completely different in the terms of endocarp characteristics and the determined discriminative models may allow to classify the endocarp with total accuracy of up to 100%. The feature set, both selecting specific textures and their number may prove to be crucial for the correctness of the discrimination of endocarp of different sweet cherry cultivars. It was observed that the larger the initial data set, the more features are selected. Almost 200 texture features were determined for one endocarp in each color channel. So, in the case of discrimination based on sets of textures selected from all color channels \((R, G, B, L, a, b, X, Y, Z)\), initially about 1800 textures were subjected to a selection. For each color space, initially were about 600...
textures. In the case of textures from all color channels, about 50 features were selected, for each color space about 20–30 features were chosen and for each color channel, only 10–20 textures were selected. However, the accuracies were the highest for discriminative models built based on sets of textures selected from all color channels \((R, G, B, L, a, b, X, Y, Z)\) and the lowest in the case of individual channels. For geometric features, several to over a dozen parameters were selected. Furthermore, research to include textures selected from other channels, such as \(V, H, S, I, U\) in the discriminative models were carried out. However, a set combining textures from color channels \((R, G, B, L, a, b, X, Y, Z)\) provided the highest accuracy. Therefore, the other results were not included in this paper. Considering the results obtained using individual classifiers, the Multi Class Classifier and LMT provided the highest accuracies and these classifiers are recommended for cultivar discrimination of sweet cherry endocarp. The obtained results reaching 100% are very high compared with literature data. Depypere et al. [9] investigated the usefulness of the dimensions and shape of \(Prunus\) \((P. domestica, P. insititia, P. x fruticans, P. spinosa, P. cerasifera)\) endocarps. The authors found that the following endocarp characteristics: ‘Perimeter’, ‘Triangle’, ‘Area’, ‘Ellipse’, ‘Circular’, ‘Rectangular’ and some index values may be very appropriate for taxonomic analysis. The values of parameters can be changed depending on endocarp maturity and levels of hydration. Frigau et al. [10] used size, shape, and texture parameters determined by image analysis for the discrimination of the \(Prunus\) sp. seeds to different plum cultivars \((Prunus salicina, Prunus domestica, Prunus cerasifera)\). The classification accuracy reached 0.9067 was observed for set combining textures, size and shape parameters. In the case of data set including only textures, the accuracy of up to 0.7730 was observed. For size parameters, the accuracy was equal up to 0.5732, and for shape features—up to 0.4390 [10]. Sarigu et al. [20] observed high accuracies of cultivars discrimination of plum endocarps based on morpho-colorimetric (shape, size, surface color) and texture parameters determined using an image analysis technique. The overall accuracy of 86.1% was obtained for distinguishing endocarps of different \(P. domestica\) cultivars. Discrimination between \(P. domestica\) and \(P. spinosa\) provided the accuracy of 99.3%. Thereby, Sarigu et al. [20] proved the usefulness of image processing in taxonomic studies, including the cultivar level. The developed procedures might be applied in germplasm banks, nurseries or other institutions for the conservation of biodiversity and enhancement of traditional plum cultivar for consumer satisfaction. Beyaz and Öztürk [21] successfully applied stone image analysis based on geometric parameters, such as length and width for the classification of olive cultivars. Stone characteristics were also applied by Milatović et al. [22] for the identification of apricot \((Prunus armeniaca)\) cultivars. The authors found that the size, shape, participation of stone in fruit mass were the most significant in apricot cultivar discrimination. In the case of jujube, stone shape, length, width and weight were useful for discrimination between different cultivars [23]. Furthermore, stone (endocarp) dimensions, such as, for example, length, width, and suture diameter can be used in archaeological studies in terms of cultivation and domestication fruit [24].

Own research has significantly expanded the scope of application of machine learning for sweet cherry research. Approximately 60 geometric parameters and about 200 textures for images from each color channel were calculated for one endocarp. Discrimination of sweet cherry endocarp based on features selected from such a large set is a great novelty of the performed research. The developed models for the cultivar discrimination, extended in the future with more detailed research of sweet cherry pits and even other species of fruit, can be extremely useful in industries, mainly in the food industry, the production of biodiesel and the cosmetic industry. Further research may include more cultivars. Additionally, a more detailed explanation of the relationship between the appearance of endocarp and fruit is needed. For different cultivars, the discrimination accuracy of the fruit with the accuracy of the discrimination of endocarp may be compared using the discriminative models built based on selected textures and geometric parameters. The relationship between texture or geometric features and chemical composition may be investigated. The usefulness of the obtained results and the need for further research are
confirmed by the literature data on the valuation of sweet cherry pits and differentiation of properties depending on the cultivar. Due to the increase in cherry production and fruit processing, among others, for juice, jelly, or jam, there are also large amounts of pits that can be used. There are many compounds relevant to humans, the content of which varies with the cultivar. Some compounds may have a positive effect on human health Therefore, the cultivars with a high content of these compounds may be desirable for the food industry. Different cultivars of sweet cherry kernels may be characterized by different antioxidant activity and bioactive content including β-Carotene, total flavonoid, total phenolic, saponins [25]. Aqil et al. [7] confirmed that the final food products obtained from sweet cherry kernels of different cultivars may have different properties. Aqil et al. [7] reported that an active ingredient included in sweet cherry kernel oil can be used in the food and cosmetic industries. Sweet cherry kernel oil may be characterized by different fatty acids and sterols composition, tocopherol content, depending on cultivars. Different cultivars may also differ in the oil yield. Therefore, the identification of cultivars of pits including kernels used in processing is important. Due to the risk of adulteration, methods or techniques to distinguish between sweet cherry pit cultivars can be used in processing plants, sorting lines or even during the purchase of pits. Image analysis is objective and may contribute to increase the correctness of classification and reduce costs, time and labor consumption of analyzes. The procedures developed in the present study may prove crucial to cultivar identification of sweet cherry pits and can also be adapted to other fruit including endocarp (pit or stone). The usefulness of image analysis based on texture parameters was also confirmed for cultivar discrimination of sour cherry pits [26]. Therefore, the results of this study have great potential and practical applications and research should be expanded and continued for different species of fruit.

5. Conclusions

The results of the research revealed the usefulness of machine learning for cultivar discrimination of sweet cherry endocarp. Discriminative models built based on selected textures allowed to classify the endocarp of different sweet cherry cultivars with total accuracy of up to 100%. The complete distinction between endocarp ‘Kordia’ and ‘Lapins’ and ‘Kordia’ and ‘Büttner’s Red’ achieved in the case of the models including combined selected textures from all color channels, separate models based on textures from RGB, Lab, XYZ color spaces, G, L, Y color channels and models combined selected textural and geometric features. The lowest accuracies reached 96% were obtained for discrimination of ‘Lapins’ and ‘Büttner’s Red’. All three cultivars, ‘Kordia’, ‘Lapins’, and ‘Büttner’s Red’, were correctly discriminated with an accuracy of up to 98% based on selected features of the endocarp. Both, in the case of discrimination ‘Lapins’ and ‘Büttner’s Red’, as well as ‘Kordia’, ‘Lapins’ and ‘Büttner’s Red’, the models comprised of combined textural and geometric parameters produced the most satisfactory results. The developed discriminative models may be used in practice for the identification of cultivar of sweet cherry endocarp in an inexpensive, fast, and objective way to avoid adulteration.

Funding: This research received no external funding.

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 author declares no conflict of interest.

References

1. Naderiboldaji, M.; Khadivi khub, A.; Tabatabaeefar, A.; Ghasemi Varnamkhasti, M.; Zamani, Z. Some Physical Properties of Sweet Cherry (Prunus avium L.) Fruit. Am. Eurasian J. Agric. Environ. Sci. 2008, 3, 513–520.
2. Zhang, X.; Zhang, L.; Zhang, Q.; Xu, J.; Liu, W.; Dong, W. Comparative transcriptome profiling and morphology provide insights into endocarp cleaving of apricot cultivar (Prunus armeniaca L.). BMC Plant Biol. 2017, 17, 1–14. [CrossRef] [PubMed]
3. Straccia, M.C.; Siano, F.; Coppola, R.; La Cara, F.; Volpe, M.G. Extraction and Characterization of Vegetable Oils from Cherry Seed by Different Extraction Processes. *Chem. Eng. Trans.* 2012, 27, 391–396. [CrossRef]

4. Senica, M.; Stampar, F.; Veberic, R.; Mikulic-Petkovsek, M. Transition of phenolics and cyanogenic glycosides from apricot and cherry fruit kernels into liqueur. *Food Chem.* 2016, 203, 483–490. [CrossRef]

5. Doğantürk, M.; Canbay, H.S. Oil ratio and fatty acid composition of cherry seed oil. *Turk. J. Health Sci. Life* 2019, 2, 21–24. [CrossRef]

6. Demirbas, A. Biodiesel from kernel oil of sweet cherry (*Prunus avium* L.) seed. *Energy Source. Part A* 2016, 38, 2503–2509. [CrossRef]

7. Ağil, Y.; Ouassor, I.; Belmaghraoui, W.; El Hajjaji, S. *Prunus avium* kernel oil characterization: A comparative study of four varieties from Setrou, Morocco. *OCL* 2020, 27, 1–5. [CrossRef]

8. Radenkovs, V.; Feldmane, D. Profile of lipophilic antioxidants in the by-products recovered from six cultivars of sour cherry (*Prunus cerasus* L.). *Nat. Prod. Res.* 2017, 31, 2549–2553. [CrossRef]

9. Depypere, L.; Chaerle, P.; Mijnsbrugge, K.V.; Goetghebeur, P. Stony endocarp dimension and shape variation in *Prunus* section *Prunus*. *Ann. Bot.* 2007, 100, 1585–1597. [CrossRef] [PubMed]

10. Frigau, L.; Antoch, J.; Bacchetta, G.; Sarigu, M.; Ucchesu, M.; Alves, C.h.Z.; Mola, F. A Statistical Approach to the Morphological Classification of *Prunus* sp. Seeds. *Plant Biosyst.* 2019. [CrossRef]

11. Abdullah, M.Z. Image Acquisition Systems. In *Computer Vision Technology for Food Quality Evaluation*, 2nd ed.; Sun, D.-W., Ed.; Academic Press/Elsevier Inc.: San Diego, CA, USA, 2016; pp. 3–43.

12. Ropelewska, E. Evaluation of wheat kernels infected by fungi of the genus Fusarium based on morphological features. *J. Food Saf.* 2019, 39, e12623. [CrossRef]

13. Szczypiński, P.M.; Strzelecki, M.; Materka, A.; Klepaczko, A. MaZda—A software package for image texture analysis. *Comput. Meth. Prog. Biomed.* 2009, 94, 66–76. [CrossRef] [PubMed]

14. Fernández, L.; Castillero, C.; Aguilera, J.M. An application of image analysis to dehydration of apple discs. *J. Food Eng.* 2005, 67, 185–193. [CrossRef]

15. Armì, F.; Fekri-Ershad, S. Texture image analysis and texture classification methods—A review. *Int. Online J. Image Process. Pattern Recogn.* 2019, 2, 1–29.

16. Ibraheem, N.A.; Hasan, M.M.; Khan, R.Z.; Mishra, P.K. Understanding Color Models: A Review. *ARPN J. Sci. Technol.* 2012, 2, 265–275.

17. Zapotoczny, P. Discrimination of wheat grain varieties using image analysis: Morphological features. *Eur. Food Res. Technol.* 2011, 233, 769–779. [CrossRef]

18. 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.

19. Witten, I.H.; Frank, E. Data mining. In *Practical Machine Learning Tools and Techniques*, 2nd ed.; Elsevier: San Francisco, CA, USA, 2005.

20. Sarigu, M.; Grillo, O.; Lo Bianco, M.; Ucchesu, M.; d’Hallewin, G.; Loi, M.C.; Venora, G.; Bacchetta, G. Phenotypic identification of plum varieties (*Prunus domestica* L.) by endocarps morpho-colorimetric and textural descriptors. *Comput. Electron. Agric.* 2017, 136, 25–30. [CrossRef]

21. Beyaz, A.; Öztürk, R. Identification of olive cultivars using image processing techniques. *Turk. J. Agric. For.* 2016, 40, 671–683. [CrossRef]

22. Milatović, D.; Durović, D.; Milivojević, J. Stone and kernel characteristics as elements in identification of apricot cultivars. *Voćarstvo* 2006, 40, 311–319.

23. Kim, S.H.; Nam, J.I.; Kim, C.W. Analysis of Qualitative and Quantitative Traits to Identify Different Chinese Jujube Cultivars. *Plant Breed. Biotechnol.* 2019, 7, 175–185. [CrossRef]

24. Zheng, Y.; Crawford, G.W.; Chen, X. Archaeological evidence for peach (*Prunus persica*) cultivation and domestication in China. *PLoS ONE* 2014, 9, e106595. [CrossRef] [PubMed]

25. Afonso, S.; Oliveira, I.V.; Meyer, A.S.; Aires, A.; Saavedra, M.J.; Gonçalves, B. Phenolic Profile and Bioactive Potential of Stems and Seed Kernels of Sweet Cherry Fruit. *Antioxidants* 2020, 9, 1295. [CrossRef] [PubMed]

26. Ropelewska, E. Classification of the pits of different sour cherry cultivars based on the surface textural features. *J. Saudi Soc. Agric. Sci.* 2020. [CrossRef]