Evaluation of Cultivar Identification Performance Using Feature Expressions and Classification Algorithms on Optical Images of Sweet Corn Seeds

Yu Tang 1,2, Zhishang Cheng 1, Aimin Miao 1,*, Jiajun Zhuang 1, Chaojun Hou 1, Yong He 3, and Shaoming Luo 2

1 Academy of Contemporary Agriculture Engineering Innovations, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China; ty2008@zhku.edu.cn (Y.T.); chengzhichang@zhku.edu.cn (Z.C.); zhuangjiajun@zhku.edu.cn (J.Z.); houchaojun@zhku.edu.cn (C.H.); cauchx0105@163.com (X.C.)
2 College of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China;
3 College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China; yhe@zju.edu.cn
* Correspondence: am_miao@zju.edu.cn or miaoaimin@zhku.edu.cn

Received: 6 July 2020; Accepted: 22 August 2020; Published: 27 August 2020

Abstract: Cultivar identification of seeds is important for crop yield and quality. To study the impact of different features expressions and classification methods on cultivar identification, the performance of the feature expressions and classification algorithms affecting the accuracy of cultivar identification was evaluated by image processing techniques. A total of 448 samples of seeds from seven cultivars of sweet corn, namely, Orlando, Beiyasi, Jingketian 183, Jingtian 218, Suitian 1, CT76 and Lilixiangtian, were evaluated. The color, shape and texture features of the seeds were extracted from the images, and the class separability criterion was adopted to evaluate the separability of the features of the embryo side, nonembryo side and both of them combined. The results indicate that the class separability based on the features of the embryo side was higher than that based on the nonembryo side and both of them combined. Based on the embryo-side optical feature data, dimensionality reduction was conducted by two feature selection methods (stepwise discriminant analysis (SDA) and genetic algorithm (GA)) and two feature extraction methods (principal component analysis (PCA) and kernel principal component analysis (KPCA)). Performance evaluation of the feature reductions was conducted by constructing k-nearest neighbor (K-NN), naïve Bayes (NB), linear discriminant analysis (LDA) and support vector machine (SVM) classifiers. Compared to the PCA and KPCA algorithms, the SDA and GA algorithms were more conducive to the cultivar classification of sweet corn seeds; the critical features selected specifically by the SDA, K-NN, NB, LDA and SVM classifiers achieved the best classification accuracies (81.43%, 82.86%, 90%, and 87.14%, respectively). Analysis of variance (ANOVA) revealed that the approach for optical feature selection had a more significant effect on the identification of sweet corn seed cultivars than did the classifiers. Therefore, based on the optical images of the embryo side and the key features obtained by the feature selection method, a classification model was constructed for the accurate and nondestructive classification of different sweet corn seed cultivars.

Keywords: sweet corn seed; cultivar classification; optical image; performance evaluation; separability criterion
1. Introduction

Sweet corn (Zea mays var. saccharata) is a subspecies of maize whose milky stage is rich in sugar, various amino acids, vitamins, minerals and dietary fiber. Based on its high nutritional and edible value [1–3], the economic benefit of sweet corn is twice that of ordinary corn. It has been reported that the planting area of sweet corn in China has gradually expanded, which accounted for approximately 25% of the world’s planting area in 2018 [4]. To meet the yield and quality requirements for crop production, the safety, high quality and reliability of seeds are important for planting. However, the mixing of different cultivars during the cultivation, harvest, transportation and storage of seeds can occur, especially with the widespread adoption of hybrid seed techniques. The optimal harvest period of sweet corn occurs in an extremely short time, and the corn quality changes rapidly after harvest [5]. In particular, the sweet corns are generally harvested at milk-ripe stage, about 20–22 days after pollination, and the immature sweet kernel of ear endosperm is the main product; thus, pure seed is essential for uniform harvesting time, uniform maturity, appropriate shelf-life and timely consumption. Moreover, the economic value, nutritional value and pest resistance of sweet corn are related to the properties of the cultivars. To control seed quality and avoid repeat cultivation, rapid and accurate methods to measure the purity of sweet corn seed are highly important for industrial production of sweet corn.

Morphological identification, physical and chemical analyses, and molecular identification are the conventional methods for identifying plant cultivars [6–8]. However, these methods generally require the use of protein electrophoresis or DNA molecular markers, both of which are time consuming, expensive and destructive [9]. Therefore, these methods are generally used to measure a small number of sampled seeds. To develop accurate and nondestructive classification of a large number of seed samples, a series of research methods have been proposed by scholars, among which spectral imaging and optical imaging technology are widely used [10–13].

Spectral technology has been applied to seed feature analysis and cultivar classification by many scholars [14–16]. For example, in Qiu et al. (2019), a feature wavelength was selected for two sweet corn seed cultivars via a genetic algorithm (GA), and the classification models based on full spectral wavelengths were compared with models based on feature wavelengths; it was indicated that the model complexity could be reduced and the classification accuracy was still high after the feature wavelength selection [17]. In Zhao et al. (2018), 12,900 seeds of three maize cultivars were studied, and a radial basis function neural network based on the optimal wavelength selected by principal component analysis (PCA) was established. The experiment showed that the results of the small-size calibration model based on feature wavelengths were similar to those of the large-sample-size calibration model [18]. To enable an increased number of feature combinations, in Xia et al. (2019), spectral features and texture features were extracted from 1632 seeds of 17 different maize cultivars, and the optimal features were selected via uninformative variable elimination (UVE), the successive projections algorithm (SPA) and multilinear discriminant analysis (MLDA). The results showed that the least squares-support vector machine (LSSVM) classification model based on the features selected via MLDA gained the best performance, achieving the highest classification accuracy (99.13%) [19]. Miao et al. (2018) introduced t-distributed stochastic neighborhood embedding (t-SNE) for the classification of seeds from eight different waxy maize cultivars and found that the classification accuracy of the t-SNE models was improved by procrustes analysis (PA) preprocessing, and the models using the nonembryo-side data were more accurate than those using the embryo-side data [20]. In Liu et al. (2014), multispectral imaging technique was used to discriminate the non-transgenic seeds and transgenic rice seeds by combining with four chemometrics methods. By comparing the discrimination performance of different chemometrics methods, the best model for classifying the rice seeds was obtained [21]. In Shrestha et al. (2016), the classification model was investigated by multispectral imaging technique with the wavelengths ranging from 375 nm to 970 nm for five cultivars of tomato seeds, and a good classification accuracy for two independent test sets was obtained for all tomato cultivars irrespective of chemometric methods [22]. Hu et al. (2020) used multispectral imaging technology to separate
sweet clover seeds from alfalfa seed. The performance of multispectral imaging with object-wise multivariate image analysis was evaluated, and the results demonstrated that the linear discriminant analysis (LDA) model based on a combination of spectral and morphological data showed the best classification performance, with an accuracy of up to 99% [23]. Seed cultivar classification can be achieved via spectral technology, but the spectral system requires stringent experimental conditions, and the computational complexity of the data processing is high. As economical, convenient and easily adoptable methods, many optical imaging technique-based seed classification methods have been proposed. These methods have been successfully applied to the seed cultivar identification of rice, wheat, corn and other crop species [24–26]. For example, to study the influence of different optical properties on the classification of rice seed cultivars, a neural network-based classification model based on the texture, shape and the combination of the two properties was proposed by Chaugule et al. (2014) [27], and a positive classification result based on seed shape features was reported. Wu et al. (2018) used digital image processing techniques to extract six typical kinds of optical shape features, such as the area, perimeter and rectangularity of ordinary corn seeds, and developed many kinds of classification models. Their results showed that, based on shape features, a support vector machine (SVM) classification model combining GA and particle swarm optimization (PSO) could effectively classify different cultivars of maize seeds [28]. Kiratiratanapruk et al. (2011) investigated the extraction problems associated with key optical attribute features and compared the performance of color (those based on red-green-blue (RGB) and hue-saturation-value (HSV) color histograms) and texture features (those based on a gray level co-occurrence matrix and local binary patterns) to classify maize seeds, and the results showed that the color and texture features combined gained the best classification performance [29]. With respect to the problem of adhesion between optical images of seeds, the foreground segmentation of a single typical corn seed is achieved by using the line contour segmentation algorithm in Li et al. (2019) [30]. Moreover, on the basis of their color and shape, normal and damaged corn seeds are classified by the maximum likelihood estimation. In Abbaspourgilandeh et al. (2020), the color, shape and texture features from the optical images of different rice cultivars with nonlinear relationship were extracted, and the rice cultivar classification model was established by discriminant analysis (DA) and artificial neural network (ANN); the results indicated that ANN achieved a better identification accuracy than that of DA [31].

The above classification models were generally established through feature extraction and classification algorithms; thus, features reduction algorithms and classification methods have a strong effect on identification accuracy, and improving these model components has been studied previously [15–18,28]. As the feature expressions and classification methods have different effects on the accuracy of cultivar identification, a performance evaluation on these two aspects should be performed first to determine the feature space that is most beneficial to cultivar identification; however, this concept has not been studied in the above research. In addition, most current research focuses on the features from a random section or part of the seed (they do not distinguish between the embryo side and nonembryo side) [12,14–16,18,28–30] or only the embryo side [17,19] of maize seeds. Because of the presence of the germ, the characteristic information contained in the front (containing the embryo) and the back (not containing embryo) of corn seeds and its influence on the performance of cultivar classification differs. To improve the accuracy of cultivar classification and improve the stability of the model, the performance of feature information from the embryo side of seeds, the nonembryo side and both of them combined should be evaluated, with the optimal side subsequently used for feature analysis and classification modeling; however, this has not been investigated. To ensure the systematic nature and integrity of cultivar identification, it is especially important to establish a performance evaluation in feature analysis and processing methods for sweet corn seeds. Thus, in this paper, in view of the above problems, seed cultivar classification via image analysis was studied. Due to it being economical, convenient and easily adoptable, a charge-coupled device (CCD) camera (model H1600Cam) with 16 million pixel was used for image acquisition in this study. The optical image of seven cultivars of sweet corn seeds were collected and different optical features were generated to
evaluate the performance of cultivar identification. Through optical property features, such as the color, shape and texture features of the embryo side and nonembryo side of seeds, cultivar separability of the different seed sides was achieved, and the optimal side was determined. The key features of seed images were obtained by different dimension reduction methods, and the performance of the different feature spaces was evaluated by four classification algorithms. The key optical feature expressions and classification methods that affect the identification model of sweet corn seed cultivars were ultimately determined.

2. Materials and Methods

2.1. Sample Preparation and Image Acquisition

Seven cultivars of sweet corn seeds (Orlando, Beiyasi, Jingketian 183, Jingtian 218, Suitian 1, CT76 and Lilixiangtian, which are recorded as V1, V2, V3, V4, V5, V6 and V7, respectively) were purchased from a seed company (FMYS Technology Ltd., Beijing, China). They were packaged in plastic bags and stored in a refrigerator at 4 °C with their moisture content was 7% to 8%. Figure 1 shows the true appearance of samples of seeds of the seven cultivars, where each column shows the samples belong to the same cultivar. For each column, the top two shows the embryo side and the bottom two shows the nonembryo side of the seeds. All of the samples’ mass-tone attune was yellow, the average length of the seeds was 1 cm, and the average width was 0.8 cm.

A schematic of the optical imaging system designed for this research is shown in Figure 2. The system was composed of a white cube (60 cm per side), a charge-coupled device (CCD) camera (model H1600Cam, Ruishi Instrument Equipment Co., Ltd., Shenzhen, China) and a ring-shaped light source. The CCD camera was fixed onto the top of the cube. The lens of the CCD camera around which the ring-shaped light source was mounted was situated 50 cm above the center of the bottom. The maximum brightness of the light source was 35,000 lux, and the brightness was adjusted to 80% during the imagery acquisition.

The embryo side and nonembryo side of the seeds of each cultivar were placed in 8 × 8 arrays, and all seeds were oriented uniformly. Firstly, the images of embryo side were collected, then the seeds were turned over and the seed images of nonembryo side were collected. Blue gauze was used as a background to provide good contrast between the background and the yellow color of the seed samples. For 64 samples of each cultivar, 54 seeds were randomly selected as the training samples, and the remaining 10 seeds were used as the testing samples. In the following procedure, the classification model construction and variety discrimination were implemented by the training data and testing data, respectively.

![Figure 1. Images acquired from 7 cultivars of sweet corn seeds: (a) Orlando, (b) Beiyasi, (c) Jingketian 183, (d) Jingtian 218, (e) Suitian 1, (f) CT76, (g) Lilixiangtian.](image-url)
2.2. Methods

The flow diagram in Figure 3 shows the performance evaluation of feature expressions and classifications for optical image-based cultivar identification of sweet corn seeds. The figure illustrates the main procedures for all five steps, which are as follows: foreground region segmentation and feature generation, separability evaluation of the embryo side and nonembryo side, performance evaluation of feature expressions, performance evaluation of classifiers and importance evaluation of feature reduction algorithms and the classification model.

Figure 3. Flow diagram of the performance evaluation for cultivar classifications of sweet corn seeds based on optical images. SDA, KPCA, NB and K-NN represent stepwise discriminant analysis, kernel principal component analysis, naïve Bayes and K-nearest neighbor, respectively.
2.2.1. Foreground Region Segmentation

The segmentation of the foreground region of corn seed mainly includes the processes of grayscale image conversion, threshold segmentation and morphology reprocessing. First, the RGB color image was converted into a grayscale image, as shown in Figure 4c. The grayscale histogram was calculated in Figure 4d, from which we can see that there was a significant grayscale difference between the seed area and the background area. Otsu is an algorithm to determine the threshold of binary image segmentation based on the maximum variance between foreground and background images. The foreground region of the corn seed image is segmented by the Otsu threshold segmentation algorithm in MATLAB software [32], and the result is shown in Figure 4e. The epidermis has a small grayscale distribution feature in the local region, and there is no significant difference between the local region and the background region; thus, there is some noise and over segmentation in Figure 4e. Therefore, the hole-filling operation, morphology open and close operation were used to perform morphological postprocessing for the binary image of Figure 4e [33], and the results are shown in Figure 4f.

![Figure 4](image_url)

**Figure 4.** Foreground region segmentation of sweet corn seed: (a) image of sweet corn seeds, (b) magnification of a single seed, (c) the grayscale image of one seed, (d) the grayscale histogram, (e) binary image of one seed, (f) the image obtained after morphological postprocessing.

2.2.2. Feature Generation

It can be seen from Figure 1 that the apparent features of the sweet corn seeds, such as the color, shape and texture, are different between different cultivars. Hence, the color features of seeds were extracted through YC\(_b\)C\(_r\), the hue-saturation-value (HSV) and the International Commission on Illumination (CIE) L*\(a^*\)\(b^*\) space transformation. The binary image after background segmentation was calculated to obtain the geometric shape features. To obtain complete descriptive features of the seed region, two methods, a gray level co-occurrence matrix (GLCM) and local binary patterns (LBP), were used.

To study the image features of sweet corn seeds in different color spaces, the sweet corn seed images were transformed from an RGB color space to YC\(_b\)C\(_r\), HSV and CIE L*\(a^*\)\(b^*\) color spaces. The YC\(_b\)C\(_r\) space transformation was realized via Equation (1), and the transformation to HSV and
CIE L*a*b* color space were achieved according to references [34,35]. Since the brightness component does not contain color information, the information of the Y, V and L* components were eliminated. Moreover, since not every pixel had the same color component in the seed region, this paper used the mean value and standard deviation (std) of the different color components (including R, G, B, C_b, C_r, H, S, a* and b*) to determine the color features of the seeds of the different cultivars. A total of 18 color features were generated.

\[
\begin{bmatrix}
Y \\
C_b \\
C_r
\end{bmatrix} = 
\begin{bmatrix}
16 & 65.481 & -37.797 \\
128 & 128.553 & -74.203 \\
128 & 24.966 & 112
\end{bmatrix}
+ 
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix} 
\tag{1}
\]

To obtain the shape features of the sweet corn seeds, the regionprops function was used to analyze the binarization image after foreground segmentation to determine the perimeter, area, long axis and short axis of individual seeds. The degree of rectangularity was obtained by the ratio of seed area to the smallest circumscribed rectangle, and the degree of extension was obtained by the ratio of the long axis to short axis. The circularity of the seeds calculated via Equation (2) was used to describe the extent of the similarity of the seed shape to that of a circle. The shape complexity obtained by Equation (3) was then used to describe the relative perimeter per unit area [28].

\[
C = 4\pi A / P^2 
\tag{2}
\]

\[
S_c = P^2 / A 
\tag{3}
\]

where \(C\) is the circularity of the seeds, \(A\) is the area of an individual seed, \(P\) is the perimeter and \(S_c\) is the shape complexity.

In this paper, a total of 26 texture features, including the number of pixel pairs with specified positions and grayscale data, and the local spatial structure were extracted from the images of the seeds. Among these structures, the minimum bounding rectangle of each seed was calculated via a GLCM, with the step size set as 1; after the GLCM was obtained, four texture statistical parameters (contrast, correlation, energy and entropy) were calculated by a GLCM based on four different directional angles: 0°, 45°, 90° and 135° [36]. In addition, LBPs were used to encode the seed gray images [37], and the pixel distribution of 10 LBP feature values in the seed region was obtained. The distribution probability of each LBP feature value was ultimately calculated to characterize the local texture features.

Based on the above methods, a total of 52 optical features, including color, shape and texture features for sweet corn seeds, were extracted from the images. The variables are shown in Table 1.

2.2.3. Separability Criterion

To evaluate the class separability of the embryo-side and nonembryo-side features of sweet corn seeds, the class separability criterion of Equation (6) was used to measure the separability as follows:

\[
S_b = \sum_{i=1}^{c} P_i (M_i - M_0)(M_i - M_0)^T 
\tag{4}
\]

\[
S_w = \sum_{i=1}^{c} \frac{1}{n_i} \sum_{k=1}^{n_i} (X_i^k - M_i)(X_i^k - M_i)^T 
\tag{5}
\]

\[
J = \frac{T_r[S_b]}{T_r[S_w]} 
\tag{6}
\]

where \(c\) is the number of classes, \(P_i\) is the prior probability of class \(i\), \(M_i\) is the mean vector of class \(i\), \(M_0\) is the overall average of all the classes, \(n_i\) is the number of samples in class \(i\) and \(X_i^k\) is the \(k\)th sample data of class \(i\). In Equation (6), \(T_r[S_b]\) and \(T_r[S_w]\) are traces of \(S_b\) and \(S_w\), respectively, while \(S_b\) and \(S_w\) represent the between-class scatter matrix and within-class scatter matrix and are obtained by
Equations (4) and (5), respectively. $J$ represents the feature distance between the different cultivars and the feature tightness within the same cultivar; the larger the $J$ value is, the better the separability.

2.2.4. Feature Reduction Methods

Stepwise discriminant analysis (SDA), GA, PCA and kernel principal component analysis (KPCA) were applied to extract the key feature variables, and the classification performance for sweet corn seed cultivars based on these methods was evaluated.

Based on the Wilks criterion, the key features for seed cultivar classification are selected from the optical feature variables through an iterative process. To avoid selecting the variables that are linearly correlated with the early selected variables, the F criterion is used to statistically test for the selected variables, and these correlated features are eliminated until no variables can be selected or removed. For the GA algorithm, seeds’ optical features are used as genes to randomly generate an initial population, and cultivar classification accuracy is used as the fitness function. The maximum number of iterations is set to 100, and the feature variables corresponding to the highest fitness function are selected to achieve feature selection. With a cumulative contribution rate of 85% for the standard, the PCA algorithm was used to extract principal component features of the optical images of the corn seeds. The Gaussian radial basis function was selected as the kernel function for KPCA, namely, $k(x, y) = \exp(-||x - y||^2/\sigma)$ and $\sigma$ was set to 50 [38].

Table 1. Optical feature variables of sweet corn seeds.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Variable</th>
<th>Symbol</th>
<th>Variable</th>
<th>Symbol</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>Mean of R</td>
<td>$c_{12}$</td>
<td>Std of B</td>
<td>$s_4$</td>
<td>Short axis</td>
</tr>
<tr>
<td>$c_2$</td>
<td>Mean of G</td>
<td>$c_{13}$</td>
<td>Std of $C_b$</td>
<td>$s_5$</td>
<td>Aspect ratio</td>
</tr>
<tr>
<td>$c_3$</td>
<td>Mean of B</td>
<td>$c_{14}$</td>
<td>Std of $C_r$</td>
<td>$s_6$</td>
<td>Rectangularity</td>
</tr>
<tr>
<td>$c_4$</td>
<td>Mean of $C_b$</td>
<td>$c_{15}$</td>
<td>Std of $H$</td>
<td>$s_7$</td>
<td>Circularity</td>
</tr>
<tr>
<td>$c_5$</td>
<td>Mean of $C_r$</td>
<td>$c_{16}$</td>
<td>Std of $S$</td>
<td>$s_8$</td>
<td>Shape complexity</td>
</tr>
<tr>
<td>$c_6$</td>
<td>Mean of $H$</td>
<td>$c_{17}$</td>
<td>Std of $a^*$</td>
<td>$t_1$~$t_4$</td>
<td>Contrast $^1$</td>
</tr>
<tr>
<td>$c_7$</td>
<td>Mean of $S$</td>
<td>$c_{18}$</td>
<td>Std of $b^*$</td>
<td>$t_5$~$t_8$</td>
<td>Correlation $^2$</td>
</tr>
<tr>
<td>$c_8$</td>
<td>Mean of $a^*$</td>
<td>$s_1$</td>
<td>Perimeter</td>
<td>$t_9$~$t_{12}$</td>
<td>Energy $^3$</td>
</tr>
<tr>
<td>$c_9$</td>
<td>Mean of $b^*$</td>
<td>$s_2$</td>
<td>Area</td>
<td>$t_{13}$~$t_{16}$</td>
<td>Entropy $^4$</td>
</tr>
<tr>
<td>$c_{10}$</td>
<td>Std of R</td>
<td>$s_3$</td>
<td>Long axis</td>
<td>$t_{17}$~$t_{26}$</td>
<td>LBP feature $^5$</td>
</tr>
</tbody>
</table>

$^1$: $t_1$, $t_2$, $t_3$ and $t_4$ respectively represent the contrast obtained by the GLCM at four different directional angles: $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$. $^2$: $t_5$~$t_8$ respectively represent the correlations obtained by the GLCM at four different directional angles: $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$. $^3$: $t_9$~$t_{12}$ respectively represent the energy obtained by the GLCM at four different directional angles: $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$. $^4$: $t_{13}$~$t_{16}$ respectively represent the entropy obtained by the GLCM at four different directional angles: $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$. $^5$: $t_{17}$~$t_{26}$ represent the distribution probability of the 10 LBP coded features.

2.2.5. Classification Models

$K$-nearest neighbor (K-NN), naïve Bayes (NB), linear discriminant analysis (LDA) and SVM were applied for cultivar classification and model evaluation. Parameter optimization was carried out for the classification models by multiple cross-validation experiments [39].

3. Results and Discussion

3.1. Separability Evaluation of the Embryo Side and Nonembryo Side

To evaluate the class separability of the embryo side and nonembryo side of sweet corn seed, based on the color, shape and texture features, the classification performance of the embryo side, nonembryo side and both of them combined was evaluated by the classification separability criterion in Equation (6). The class separability values of the different feature combinations acquired from
the embryo side, nonembryo side and both of them combined are shown in Table 2. In this table, C represents color features, S represents shape features and T represents texture features.

From Table 2, it can be seen that, regardless of the combination of color, shape and texture features, the value for embryo side of seeds was significantly higher than the nonembryo side and both of them combined. Based on the separability criterion, it can be concluded that the separability based on the embryo side of seeds showed the best than the other two situations. These results indicated that more identification features are associated with the embryo side than with the nonembryo side of the sweet corn seed. The result was consistent with the research in Yang et al. (2015), in which the classifiers based on near-infrared (VIS/NIR) hyperspectral feature of corn seed was developed to recognize different cultivars, and the classification results of embryo side also performed better than that of the nonembryo side for waxy corn seed [10]. According to Cheng et al. (2014), the seed features of the nonembryo side change greatly across a corncob, but the features of the embryo side tend to be identical for the same cultivar, as white embryos are less affected by the position on the corncob [40]. Thus, the features of seed embryos constitute the key factor in identifying seed cultivars. As a result, the optical features of the embryo side were selected in this paper for the classification of sweet corn seeds.

Table 2. Comparison of class separability based on optical features of sweet corn seeds.

<table>
<thead>
<tr>
<th>Side</th>
<th>Feature</th>
<th>C∪S∪T 1</th>
<th>C∪S 2</th>
<th>C∪T 3</th>
<th>S∪T 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>embryo</td>
<td></td>
<td>0.651</td>
<td>0.854</td>
<td>0.618</td>
<td>0.562</td>
</tr>
<tr>
<td>nonembryo</td>
<td></td>
<td>0.476</td>
<td>0.750</td>
<td>0.421</td>
<td>0.381</td>
</tr>
<tr>
<td>combination</td>
<td></td>
<td>0.599</td>
<td>0.800</td>
<td>0.513</td>
<td>0.466</td>
</tr>
</tbody>
</table>

1. C∪S∪T represents the combination of all 52 optical feature variables. 2. C∪S represents a total of 26 features of color and shape combinations. 3. C∪T represents a total of 44 features of color and texture combinations. 4. S∪T represents a total of 34 features of shape and texture combinations.

Table 2 shows that the class separability based on the combination of color and shape features gained the highest value, while the combinations that contain texture features gained the worst value. To further analyze the influence of the color, texture and shape features on cultivar classification, the data set of seed samples with 18 color features and 8 shape features was defined as dataset 1, and the data set of seed samples with 26 texture features was defined as dataset 2. The low-dimensional feature representations of the 7 cultivars from LDA based on these two datasets are given in Figure 5, in which the 7 cultivars are marked with different colors.

Based on the features in dataset 1 in Figure 5a, it can be seen that the data from the same cultivar are more concentrated and that the data from different cultivars are scattered across a large distance, which is beneficial for cultivar classification. In Figure 5b, it can be seen that there are many overlaps among the different cultivars based on texture features, and it is difficult to distinguish the seed cultivars. It can therefore be concluded that the color and shape features play a key role in the classification of sweet corn seed cultivars. Hence, as shown in Table 2, the class separability-based color and shape features was better than that based on texture features, the result was verified in Chaugule et al. (2014), as it was reported that texture feature of seed has less discriminating power than shape feature [27].

3.2. Feature Reductions and Cultivar Classifications

It can be seen from Table 2 that regardless of whether the embryo side, nonembryo side or both of them combined is used, the class separability based on all of the variables did not gain the best results, which indicated that there is noise or disturbance among the 52 optical feature variables. There are even variables that are not conducive to cultivar classification. To determine the key features that affect the classification performance of sweet corn seed cultivars, feature reduction based on two aspects, the feature selection method and feature extraction method, was implemented. The feature reduction performance for the embryo-side optical feature data of the sweet corn seeds by different classification methods was evaluated.
3.2.1. Results of Feature Reductions

In this paper, the key optical features of the embryo side were selected by SDA and GA algorithms. In each step of SDA, the entry and removal of variables in the model depend on the threshold of entry and the threshold of removal based on the F criteria, respectively. The variable outside the model can enter the model when its F value is larger than the threshold of entry, and the variable in the model will be removed when its F value is smaller than the threshold of removal. Referring to Muhameed et al. (2014), the threshold of entry and threshold of removal in the SDA model were set as the default parameter values 3.84 and 2.71, respectively [41]. A total of 24 feature variables were selected based on SDA. According to the classification accuracy of the K-NN benchmark classification algorithm, the model parameters of the GA algorithm were optimized 20 times by independent cross-validation. Figure 6 shows the change in the average K-NN classification accuracy rate as the number of features selected increased from 2 to 26. From Figure 6, it can be seen that classification accuracy improved with the number of selected variables, and the best performance was achieved when the number of features was 11. To achieve cultivar classification based on the optimal features, the number of GA algorithms for feature selection was set to 11 in this paper.
Figure 6. Average accuracy of the $k$-nearest neighbor (K-NN) classifier based on features selected by genetic algorithm (GA) (std of 0.03–0.04).

The key features selected by the SDA and GA algorithms are presented in Table 3. It can be seen from Table 3 that 12, 5, and 7 color, shape and texture features, respectively, were selected by SDA, while the numbers were 7, 3 and 1, respectively, by GA. A large number of color features that reflect the external color difference and range of depth of the different cultivars were selected both two algorithms, and some identical variables were selected, such as $c_1$ (mean of the R component), $c_9$ (mean of the b* component) and $c_{16}$ (std of S). Three common shape features ($s_4$, $s_5$ and $s_8$) were selected by both algorithms. Among them, $s_5$ represents morphological characteristics; the larger $s_5$ is, the narrower the seed shape. The $s_8$ feature represents shape complexity; the larger $s_8$ is, the more complex the shape. Because of the varied features of the color and shape of the different cultivars, as shown in Figure 1, color and shape were considered main features and constituted the key information for cultivar identification. These results concerning the selected features are consistent with those of the study on the importance of color and shape in Section 3.1, which verified the effectiveness of the two algorithms in terms of feature selection. It can be seen from Table 3 that only one texture feature was selected by the GA algorithm, which probably occurred because the irregular shrinkage of the corn seed epidermis made it difficult to obtain consistent texture information. Figure 1 shows that there are no obvious differences in the texture structure of the seeds of the different cultivars, which translates to a relatively small contribution of texture features to the classification of sweet corn seed cultivars.

Table 3. Selected features of sweet corn seeds based on stepwise discriminant analysis (SDA) and GA algorithms.

<table>
<thead>
<tr>
<th>Property\Algorithms</th>
<th>SDA</th>
<th>GA</th>
<th>SDA\GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>$c_1$, $c_3$, $c_4$, $c_7$, $c_9$, $c_{10}$, $c_{11}$, $c_{12}$, $c_{16}$, $c_{17}$, $c_{18}$</td>
<td>$c_1$, $c_6$, $c_9$, $c_{14}$, $c_{16}$, $c_{17}$, $c_{18}$</td>
<td>$c_1$, $c_9$, $c_{16}$, $c_{17}$, $c_{18}$</td>
</tr>
<tr>
<td>Shape</td>
<td>$s_1$, $s_4$, $s_5$, $s_7$, $s_9$</td>
<td>$s_4$, $s_8$</td>
<td>$s_4$, $s_8$</td>
</tr>
<tr>
<td>Texture</td>
<td>$t_2$, $t_3$, $t_5$, $t_6$, $t_{14}$, $t_{17}$, $t_{21}$</td>
<td>$t_5$</td>
<td>$t_5$</td>
</tr>
</tbody>
</table>

1 SDA\GA represents common features selected by the SDA and GA algorithms.

To extract the key feature information, PCA was adopted for all 52 optical feature variables, and the scatter plot of the first three principal components (PC1, PC2, PC3) with the variance contribution rates 31.67%, 17.47% and 16.53%, respectively, for the embryo side were given in Figure 7. It can be noticed in Figure 7 that there are some overlaps among the samples from different cultivars as the data information of the first three principal components may not guarantee the significant data variance. Referring to Zhuang et al. (2019), the first six principal components which gain cumulative contribution rate of 86.76% were considered [39]. KPCA was also adopted for all 52 optical feature variables, and the first ten principal components with the cumulative contribution rate of 85.04% were extracted.
3.2.2. Performance Evaluation of Classifiers

Based on the 52 optical features and the reduced features obtained by the SDA, GA, PCA and KPCA, a classification model for sweet corn seed cultivars was constructed via K-NN, NB, LDA and SVM methodology. The classification accuracy of the seven varieties of sweet corn seeds based on optical images of the embryo side is shown in Table 4.

Table 4 shows that the accuracies of the classification models based on all the variables were lower than those based on feature selection methods (SDA and GA), which verifies the occurrence of certain correlations and redundant information among the feature data of all the variables. The classification models based on the features selected by the SDA and GA algorithms were excellent, and the average classification performance was higher than that of all the variables. Therefore, the SDA+LDA model reached the highest classification accuracy (90%), indicating that feature reduction could effectively improve the classification performance of sweet corn seed cultivars. In addition, the accuracy of the NB and LDA classification models based on SDA-selected features was higher than that of the GA classification model. The feature selection results in Table 3 show that SDA selected more features than did GA, and the shape feature combination selected by SDA includes $s_1$ (perimeter) and $s_7$ (circularity), each of which the GA did not select; circularity represents the extent of the similarity of the seed shape to that of a circle. Figure 1 shows that significant differences in perimeter and circularity exist among the different cultivars of corn seeds, but the GA algorithm did not select these features. This was probably because the GA algorithm is based on an optimization strategy that easily becomes associated with local optimization, while the SDA algorithm involves statistical tests on all the selected features at each iteration and selects the key feature according to the threshold value, which depends on the importance of the variables; thus, the acquired features are relatively consistent. Therefore, the features selected by SDA are more conducive to the cultivar classification of seven different varieties of the sweet corn seeds.

<table>
<thead>
<tr>
<th>Total Variables</th>
<th>SDA</th>
<th>GA</th>
<th>PCA</th>
<th>Kernel KPCA (KPCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN</td>
<td>80.00%</td>
<td>81.43%</td>
<td>81.43%</td>
<td>74.29%</td>
</tr>
<tr>
<td>Naive Bayes (NB)</td>
<td>74.29%</td>
<td>82.86%</td>
<td>77.14%</td>
<td>67.14%</td>
</tr>
<tr>
<td>LDA</td>
<td>88.57%</td>
<td>90.00%</td>
<td>88.57%</td>
<td>67.14%</td>
</tr>
<tr>
<td>SVM</td>
<td>85.71%</td>
<td>87.14%</td>
<td>87.14%</td>
<td>68.57%</td>
</tr>
</tbody>
</table>

It can be seen from Table 4 that in the cultivar classification of sweet corn seeds, the accuracies based on feature selection algorithms (SDA and GA) are mostly greater than 80%, while those based on feature extraction algorithms (PCA and KPCA) are generally approximately 70%. This is because PCA and KPCA are unsupervised algorithms whose goal is to obtain variable information through the
maximum population variance without considering the class difference of the data; however, the goals of SDA and GA are to construct a discriminant function based on the class difference and obtain the key discriminant variables of different cultivars through continuous iteration and optimization. To verify the features of the feature selection method in terms of preserving feature data containing classification information, the feature data after PCA projection of the embryo side of sweet corn seeds was used to design the experiment. The cumulative contribution of the first six principal components reached 86.76% (the contribution rates of PC1, PC2, PC3, PC4, PC5 and PC6 were 31.67%, 17.47%, 16.53%, 8.68%, 6.59% and 5.82%, respectively); the SDA algorithm was used to select principal component features, and the key features selected for use in the discriminant model were as follows (in order of importance): PC2, PC1, PC3, PC5, PC4, PC6, PC9, PC8, PC12 and PC7. These results show that the first six principal component variables with high degrees of contribution are preferentially selected for inclusion in the discriminant model by SDA, which indicates that the feature selection method based on SDA can obtain the key features from the data. However, the principal components with high contribution rates were not included in the model in advance, and PC2 was selected before PC1 and PC5 before PC4. Therefore, the later principal components (PC2, PC5) were more conducive to cultivar identification than the first ones (PC1, PC4) were, which demonstrates that SDA could obtain more key classification information than could PCA.

Table 4 shows that the accuracies of the four classification models established under the same feature data somewhat differ; for example, the accuracies of the four classification models based on the feature data selected by SDA vary slightly: from 81.43% to 90%. However, the accuracies of the same classification algorithm based on different feature data vary greatly; for example, the accuracy range of the LDA classification model based on different feature data ranges from 67.14% to 90%. These results indicate that the classification algorithm and feature reduction algorithms have different effects on classification accuracy. To compare the extent of influence of feature reduction algorithms and classification algorithms in terms of classification accuracy, analysis of variance (ANOVA) was performed. By decomposing the research objectives (classification rate) into different variable factors (feature reduction and classification algorithms) in ANOVA, the variation performance of different components (Adj SS) is measured after the data estimation to determine the contribution rate of different variable factors to the research objectives. Finally, the \( p \)-value of the mismatch test was compared with the significance level to evaluate the original hypothesis. The accuracies are shown in Table 4 as dependent variables, and the results are shown in Table 5. The feature reduction algorithms factor \( p \)-value of 0.001 is much lower than 0.05, and the classification algorithm factor \( p \)-value of 0.361 is much greater than 0.05 (a significance level of \( \alpha = 0.05 \) determines the influence of the control factors; \( p < 0.05 \) is significant); hence, five kinds of feature reduction algorithms exhibit significant differences in classification accuracy, but four kinds of classification algorithms do not. The last column in Table 5 lists the percentage contribution of the different factors to classification accuracy. The influence of the feature reduction algorithms (contribution of 70.89%) on the classification accuracy of sweet corn seed cultivars was significantly higher than that of the classification algorithms (contribution of 6.59%).

Table 5. ANOVA results for the classification accuracy of sweet corn seed cultivars.

<table>
<thead>
<tr>
<th>Source</th>
<th>DOF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>( F )</th>
<th>( p )</th>
<th>Contribution/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature reduction alg.</td>
<td>4</td>
<td>808.53</td>
<td>202.13</td>
<td>9.44</td>
<td>0.001</td>
<td>70.89</td>
</tr>
<tr>
<td>Classification alg.</td>
<td>3</td>
<td>75.22</td>
<td>25.07</td>
<td>1.17</td>
<td>0.361</td>
<td>6.59</td>
</tr>
<tr>
<td>Residual error</td>
<td>12</td>
<td>256.86</td>
<td>21.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>1140.6</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

To further analyze the influence of the classification algorithm on cultivar classification, different classification algorithms based on the features selected by SDA were compared by a discriminant confusion matrix. Tables 6–9 present the discrimination confusion matrices of the sweet corn seed cultivars based on four classification models: SDA+K-NN, SDA+NB, SDA+LDA and SDA+SVM.
It can be seen from the confusion matrices of the different classification methods (Tables 6 and 7) that the classification effect of K-NN and NB is relatively poor, with overall accuracies of 81.43% and 82.86%, respectively. This occurred mainly because V6 was misidentified as V1. In particular, the accuracy of the K-NN classification for V6 is only 30%, probably because K-NN is sensitive to noise disturbance and is better suited for pattern classification of large sample sizes. In addition, a relatively large sample size will better reflect the real distribution of different cultivars, but the sample size researched in this article was relatively small, which is conducive to misjudgment when using the K-NN algorithm. When the sample size was large, such as 380 samples per cultivar in Qiu et al. (2019), the high accuracy based on K-NN classifier may be obtained; however, the SVM model also performed better than K-NN model, which was similar to the results in this research [17]. With respect to the NB algorithm, the classification performance depends largely on whether the features used for modeling are independent; the stronger the data independence is, the greater the classification accuracy. However, the features extracted from the optical images of sweet corn seed are not completely independent; for example, the features of the different color components obtained by color space transformation could have a certain relevance. Therefore, the NB classification model established on the basis of these features gained poor discriminant results. The results shown in Tables 8 and 9 indicate that the best performance occurred from the classification model based on LDA and SVM algorithms, with overall accuracies of 90% and 87.14%, respectively, which made accurate discrimination for each sweet corn seed class; this is because LDA was a linear algorithm and the SVM model uses a linear kernel, which demonstrates that the experimental data from the seven sweet corn seed cultivars in this paper are linearly separable. The variety classification for the maize seed and sweet clover seeds in Xia et al. (2019) and Hu et al. (2020) [19,23], respectively, also verified the efficiency of LDA in the linear case. However, the performance was not guaranteed when the rice cultivars contain nonlinear relationship [31]. In addition, LDA classification aims to determine the projection direction in which samples from different cultivars acquire the largest ratio of between-class scatter to within-class scatter, which is consistent with the separability criterion in Section 3.1, in which the feature data based on the seed embryo side with better class separability make it easy for LDA to determine the optimal projection direction. Therefore, among the four kinds of classification models, especially the SDA+LDA model, the LDA classification model achieved the best results, where each sweet corn seed cultivar had a classification accuracy greater than 80%.

### Table 6. Confusion matrix for sweet corn seed classification based on the SDA+K-NN model.

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>80%</td>
</tr>
<tr>
<td>V2</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V3</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>80%</td>
</tr>
<tr>
<td>V5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>V6</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>V7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>100%</td>
</tr>
</tbody>
</table>

### Table 7. Confusion matrix for sweet corn seed classification based on the SDA+NB model.

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>80%</td>
</tr>
<tr>
<td>V2</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V3</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>V5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>V6</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>V7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 8. Confusion matrix for sweet corn seed classification based on the SDA+LDA model.

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>80%</td>
</tr>
<tr>
<td>V2</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V3</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>80%</td>
</tr>
<tr>
<td>V5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>90%</td>
</tr>
<tr>
<td>V6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>80%</td>
</tr>
<tr>
<td>V7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 9. Confusion matrix for sweet corn seed classification based on the SDA+SVM model.

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
<th>V7</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>70%</td>
</tr>
<tr>
<td>V2</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V3</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>V4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>90%</td>
</tr>
<tr>
<td>V5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>90%</td>
</tr>
<tr>
<td>V6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>70%</td>
</tr>
<tr>
<td>V7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>90%</td>
</tr>
</tbody>
</table>

4. Conclusions

In this paper, the performance of optical image feature expressions and cultivar classifications for seven sweet corn seed cultivars was evaluated. Thus, cultivar identification was performed by optical image feature generation, feature reduction and classification modeling. The main conclusions are as follows:

1. The separability of the optical features of the embryo side and nonembryo side of sweet corn seeds was evaluated by a class separability criterion, and the results indicated that the class separability of the embryo side was higher than that of the nonembryo side and both of them combined. Further, the class separability was compared among the color, shape and texture features of the embryo side, and the separability reached the highest (0.854) from the combination of color and shape features.

2. Dimensionality reduction was conducted by two feature selection methods (SDA and GA) and two feature extraction methods (PCA and KPCA), and their classification performance was evaluated by K-NN, NB, LDA and SVM classifiers. The results indicated that the key features obtained by the feature selection methods provided better classification accuracy than did those obtained by feature extraction methods. On the basis of the key features selected by SDA, the K-NN, NB, LDA and SVM classifiers obtained the best classification accuracies: 81.43%, 82.86%, 90% and 87.14%, respectively.

3. ANOVA was applied to characterize the impact of the feature reduction algorithms and classification algorithms on cultivar identification. The results showed that the factor of feature reduction algorithms achieved a maximum contribution of 70.89%, which had a more significant effect on the cultivar classification than the classification algorithm factor, whose contribution was 6.59%.

Author Contributions: Conceptualization, Y.T. and Z.C.; Data curation and formal analysis, C.H. and X.C.; Methodology, A.M., J.Z. and X.C.; Supervision, Y.H.; Writing—original draft, Z.C.; Writing—review & editing, Y.T., A.M., Y.H. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Project of Universities in Guangdong Province, China (No. 2017KZDKX047); the Rural Revitalization Strategy Project of Guangdong Province, China (No. 2019KJ138); the Planned Science and Technology Project of Guangdong Province, China (Nos. 2019A050510045 and
Conflicts of Interest: The authors declare no conflict of interest.

References


22. Shrestha, S.; Deleuran, L.C.; Gislum, R. Classification of different tomato seed cultivars by multispectral visible-near infrared spectroscopy and chemometrics. *J. Spectr. Imaging* 2016, 5, a1. [CrossRef]


30. Li, X.M.; Dai, B.S.; Sun, H.G.; Li, W.N. Corn classification system based on computer vision. *Symmetry* 2019, 11, 591. [CrossRef]


35. Wang, F.; Liu, F.; Zhang, Y.Y.; Zhang, A.; Cao, Y.Z.; Li, J.Q.; Zhang, S.B. The study of applying CIE 1976 (L*a*b*) colour space in the measurement of colour of wine. *Sino-Overseas Grapevine Wine* 2015, 4, 6–11. [CrossRef]


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).