AppliedChem

Article

E-Eye Solution for the Discrimination of Common and Niche Celery Ecotypes

Alessandra Biancolillo *, Martina Foschi and Angelo Antonio D'Archivio

Department of Physical and Chemical Sciences, University of L'Aquila, 67100 L'Aquila, Italy
* Correspondence: alessandra.biancolillo@univaq.it

Abstract: Celery (*Apium graveolens* L.) is a well- known plant and at the basis of the culinary tradition of different populations. In Italy, several celery ecotypes, presenting unique peculiarities, are grown by small local producers, and they need to be characterized, in order to be protected and safeguarded. The present work aims at developing a fast and non-destructive method for the discrimination of a common celery (the "Elne" celery) from a typical celery of Abruzzo (Central Italy). The proposed strategy is based on the use of an e-eye tool which allows the collection of images used to infer *colorgrams*. Initially, a principal component analysis model was used to investigate the trends and outliers in the data. Then, the classification between the common celery (Elne class) and celery from Torricella Peligna (Torricella class) was achieved by a discriminant analysis, conducted by sequential preprocessing through orthogonalization (SPORT) and sequential and orthogonalized covariance selection (SO-CovSel) and by a class-modelling method called soft independent modelling of class analogies (SIMCAs). Among these, the highest accuracy was provided by the strategies, based on the discriminant classifiers, both of which provided a total accuracy of 82% in the external validation.

Keywords: celery; e-eye; image analysis; colorgrams; classification; botanical origin; SPORT; discriminant analysis; PCA

check for updates

Citation: Biancolillo, A.; Foschi, M.; D'Archivio, A.A. E-Eye Solution for the Discrimination of Common and Niche Celery Ecotypes. *AppliedChem* **2023**, 3, 1–10. https://doi.org/10.3390/appliedchem3010001

Academic Editor: Jason Love

Received: 3 November 2022
Revised: 19 December 2022
Accepted: 21 December 2022
Published: 22 December 2022

1. Introduction

Celery (*Apium graveolens* L.) is a fundamental plant for the traditional cuisine of many populations in different parts of the world. Its use is reported as far back as in antiquity, when its curative and digestive properties had already been appreciated [1]. From the nutritional point of view, this plant, in addition to being rich in water, allows for an excellent supply of vitamins (A, C, E, K, 33 B1, B2, B6), minerals (calcium, phosphorus, iron, sodium, zinc, and copper) phytocompounds and fibers [2–4]. On average, concerning the macronutrients, it contains 46% of proteins, 45% of carbohydrates, and 9% of lipids. The *Apium graveolens* is mainly grown for gastronomic purposes. However, the presence of phenolic compounds and terpenoids makes celery attractive for antioxidant, anti-inflammatory, antimicrobial, and cardiovascular protective activities [5]. Currently, celery is one of the most appreciated and widespread vegetable crops in the world, and for this reason, abundant germplasm resources are available; in Europe, several cultivars are recognized, but the common celery, easily available in supermarkets, belongs to the "Elne" or the "Magnum" type. However, there are some small and less widespread local varieties, that are important from a genetic and organoleptic point of view. In Italy, there are different ecotypes of celery, generally grown by small local producers, and whose name is directly linked to their origin; an example of this can be celery from "Torricella Peligna" (Abruzzo, Central Italy), from "Trevi" (Umbria, Central Italy) or "Sperlonga" (Lazio, Central Italy), just to name a few. Despite the interesting characteristics, and the need to protect these typical foods, few studies have been conducted on these niche Italian celeries. For instance, Torricelli and collaborators have investigated the molecular and morpho-physiological characterization of celery from Trevi, supporting the request for a quality mark on this endangered species [1]. A very comprehensive study in this field is the one carried out by Ingallina et al. [6], where white celery from Sperlonga PGI was analyzed by exploiting a multi-platform approach, based on nuclear magnetic resonance (NMR), high-performance liquid chromatography-photodiode array detection, gas chromatography-mass spectrometry (GC-MS), and spectrophotometric analyses for the characterization of the leaves and petioles. Finally, one of the most studied aspects of celeries is the volatilome. Two studies concerning the volatile profiles of the typical Italian celery are available in the literature. The first, published in 2004, involved celery from Trevi, Elne, "Pascal" celery, and "Dorato d′Asti" celery. Plant samples were analyzed by the head-space solid-phase microextraction, combined with gas-chromatography/mass spectrometry (HS-SPME/GC-MS). The authors concluded that the most characterizing volatiles are limonene and γ-terpinene [7]. A more recent study, published by Reale et al., is focused on the volatile composition of a wild-celery ecotype and celeries from Sperlonga, and Torricella [8]. The authors concluded that, among the investigated celeries, celery from Torricella is the most different, because of its high content in β-myrcene, p-cymene, and γ-terpinene, and its lower amount of (Z)-β-ocimene and α-pinene, and β-pinene. Studies on these niche varieties are promoted by consumers′ growing interest in buying local products and the increased awareness of the genetic wealth that these landraces represent. Thus, in addition to being highly valued products for their sensory characteristics, they represent the Italian tradition of on-farming conservation. Although the aforementioned studies were found to be reliable and adequate for the chemical characterization of these interesting celery ecotypes, they all require a long time of analysis and complex sample preparation. These features make the methods unsuitable both for the needs of small producers and for a rapid routine screening, useful in revealing the dishonest practices of mislabeling and counterfeiting. Starting from these considerations, the present work stems from the need to safeguard the quality and biodiversity of specific endangered landraces by developing a method that is rapid and non-destructive, useful to authenticate and protect the peculiar Torricella ecotype, and to discriminate it from a common quality (celery Elne). The study was promoted considering that this particular variety is about to be recognized as a typical Abruzzo landrace, to be protected because it is at risk of extinction. For this purpose, an e-eye device-based method was investigated. A multivariate image analysis was applied to very different food matrices, proving effective in dealing with different issues; from the olive quality assessment [9], to the discrimination of different suppliers for vegetable oil [10] to quantify the chlorophyll content and assess the leaf health status [11] and, eventually, to discriminate cultivars. In this context, a flatbed scanner and an image processing program, combined with the mean color information and the linear discriminant analysis, were used to identify five Sicilian landraces and three common Canadian accessions of lentils [12]. An image analysis provided reproducible, objective, and accurate results for the tomato fruit cultivar discrimination from the images of tomato flesh and skin [13]. However, to the best of our knowledge, there is no example in the literature that reports the MIA applied to the varietal discrimination from the leaves. In detail, from the image analysis alone, the study ambitiously aims to distinguish the celery samples of different varieties grown in the same experimental field, thus, subjected to the same soil and climate conditions. In detail, the images were converted into colorgrams [14], and then the samples were classified, according to their ecotype. The distinction between one category and another took place using two discriminant classifiers: sequential preprocessing through orthogonalization (SPORT) [15] and sequential and orthogonalized covariance selection (SO-CovSel) which have demonstrated to be suitable in similar contexts [16–22]. Furthermore, one of the most well-known class-modelling methods, the soft independent modelling of class analogies (SIMCAs), was also exploited for the same goal, since it has proved to be an appropriate solution for this purpose [23].

2. Materials and Methods

2.1. Samples

Two varieties of celery (*Apium graveolens* L.) were investigated: *"Elne"* and "*Torricella Peligna*". The former is a common type of celery, frequently found in supermarkets, whereas the other one is a celery typical of the Abruzzo region (Central Italy). To be protected and safeguarded, the celery of Torricella Peligna is currently undergoing the process of being registered in the regional Abruzzo registry of plant biodiversity.

The investigated plants were grown in an experimental field, located in Torricella Peligna (Abruzzo); the seeds were planted in May 2021 and harvested in December 2021. It has to be pointed out that all of the celeries were grown in the same field in the Majella National Park; i.e., they were exposed to the same soil and climatic conditions. The petioles were then frozen (−18°C) until the analysis.

2.2. E-Eye Analysis

Prior to the collection of the images, the celery leaves were thawed, patted dry with absorbent paper, and placed on a flat surface.

The pictures were collected using the RS Pro Wi-Fi USB Microscope (RS Components S.r.l., Milan, Italy) (1280 × 1024 pixel resolution, magnification power from 10 × to 160 ×, 36 mm diameter, 142 mm length, and light-emitting diode lighting) kept at the same constant distance from the surface upon which the leaves were laid. An example of pictures collected on both typologies of celery is displayed in Figure 1.

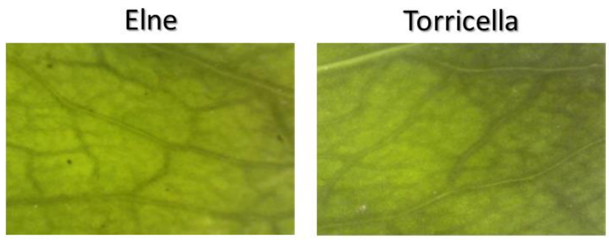

Figure 1. Examples of the images collected of the Elne (left picture) and Torricella (right picture) celeries.

A total of 219 images, 99 belonging to the Elne class, and 120 to the Torricella class, were collected. The colorgrams were obtained using MATLAB (R2015b; The Mathworks, Natick, MA, USA), exploiting the in-house functions, and following the procedure described in [14].

2.3. Chemometric Analysis

At first, the principal component analysis (PCA) was used to give an insight in the data and detect the suspicious samples or outliers. This method is one of the most widely-used exploratory data analysis approaches in chemometrics. Briefly, the PCA is based on the decomposition of the data matrix **X** into two further matrices: the scores matrix **T** and the loadings matrix **P**. The inspection of these two allows for the realization of highly informative plots, that are suitable for the interpretation of the system under study and the visualization of the trends in the data. For more detail about the algorithm and its application, the reader is addressed to [24–26].

Following the detection of the outliers and the interpretation of the possible trends, it is possible to continue further with the creation of the classification model [27]. In particular, in this work, the classification problem faced is a two-class problem, i.e., the model aims at distinguishing the celeries belonging to the Elne class from those appertaining to the Torricella class.

The classification has been carried out using three different strategies, two employing the discriminant approaches and one exploiting a class-modelling method. The discriminant approaches used were the sequential preprocessing through orthogonalization (SPORT) [15] and the sequential and orthogonalized covariance selection, coupled with a linear discriminant analysis (SO-CovSel-LDA) [28].

The SPORT is an ensemble preprocessing method that originates from the sequential and orthogonalized partial least squares (SO-PLS) [29]. Briefly, its algorithm requires that the **X** data matrix is preprocessed by different pretreatments, obtaining a set of preprocessed matrices (which will be as many as the tested pretreatments). Then, each block is sequentially modeled extracting information in the same fashion as the SO-PLS. As described in [15], once the response **Y** has been predicted by the SPORT, the linear discriminant analysis (LDA) [30] can be applied on the **Y**, as discussed in [31]. The reader can refer to [15,29,32] for details on the algorithms and the pros and cons of these methods.

Similarly to the SPORT, the SO-CovSel is an approach derived by the SO-PLS. Basically, it follows the same algorithm, but the feature reduction operated by the PLS is achieved by a covariance selection [33]. Furthermore, this approach can be exploited for the classification by applying the LDA on the predicted **Y**. A wider discussion on the SO-CovSel can be found in [34].

Moreover, the class-modelling method used is the soft independent modelling of class analogies (SIMCAs) [35]. This approach, due to its nature, models each class individually. Basically, in order to create a SIMCA model for a class of interest, the algorithm starts with the calculation of a PCA model on the calibration samples belonging to the category that needs to be modelled. Then, the individuals are accepted or rejected by the class-model, depending on their position in the PCA-space. For each i-th sample, a distance d is calculated, according to Equation (1):

$$d_i = \sqrt{\left(T^2_{i,red}\right)^2 + \left(Q_{i,red}\right)^2} \tag{1}$$

where $T^2_{i,red}$ represents the distance in the scores-space for the i-th object (normalized by the 95th percentiles of its distribution) and $Q_{i,red}$ the orthogonal distance for the same individual (normalized by the 95th percentiles of its distribution). Customarily, when $d_i < \sqrt{2}$, the object is accepted by the model and predicted as belonging to the modelled category, otherwise ($d_i > \sqrt{2}$), it is rejected, and considered not appertaining to the investigated class [27].

3. Results

The workflow of the study is depicted in Figure 2. Basically, the diverse ecotypes of celery were cultivated in the same experimental field and harvested on the same day. Then, the images were collected of the leaves, imported into MATLAB (R2015b; The Mathworks, Natick, MA, USA), and the colorgrams generated. A chemometric analysis comprehended an initial exploratory analysis by means of the PCA, and then classification models were built. The discriminant and class-modelling approaches were exploited. The details on the models and the outcome of the analysis is provided in the related sections below.

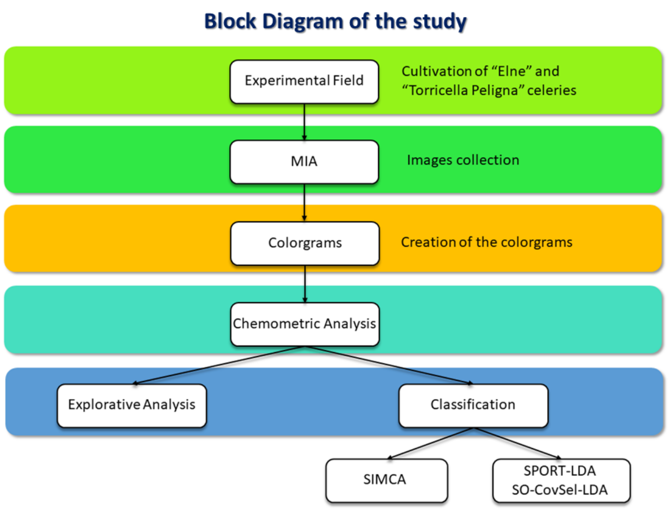

Figure 2. Block diagram of the research work.

3.1. Exploratory Analysis

As above-mentioned, at first, all of the (mean-centered) data were processed by the PCA. This allowed for obtaining the PCA-plot shown in Figure 3. From the figure, it is immediately evident that it is not possible to distinguish two clear groups between the two varieties of celery. What is observed is that part of the samples is superimposed at the positive values of PC2, while a partial distinction between the two types of celery is moderately visible for the objects presenting the negative scores of PC2. In particular, among those, the samples belonging to the Torricella class (black triangles) fall at negative values of PC1, while the others have positive scores of this component.

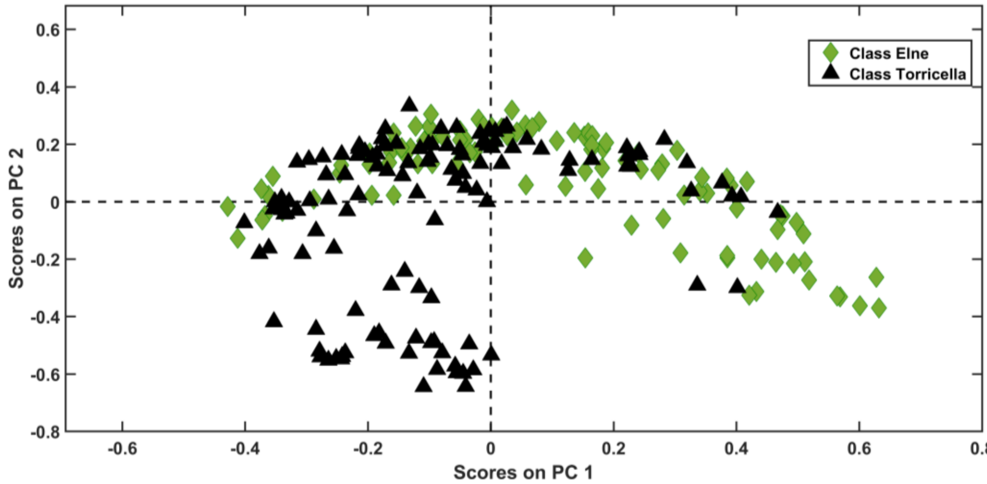

Figure 3. PCA-plot. Projection of the samples onto the first two PCs. Legend: dark green diamonds: Elne class; black triangles: Torricella class.

The interpretation of the loadings associated with the component along which it is possible to see a (albeit partial) distinction between the groups can be made to understand the characteristics of the samples. Consequently, the loading plot associated with PC1 (shown in Figure 4) has been inspected.

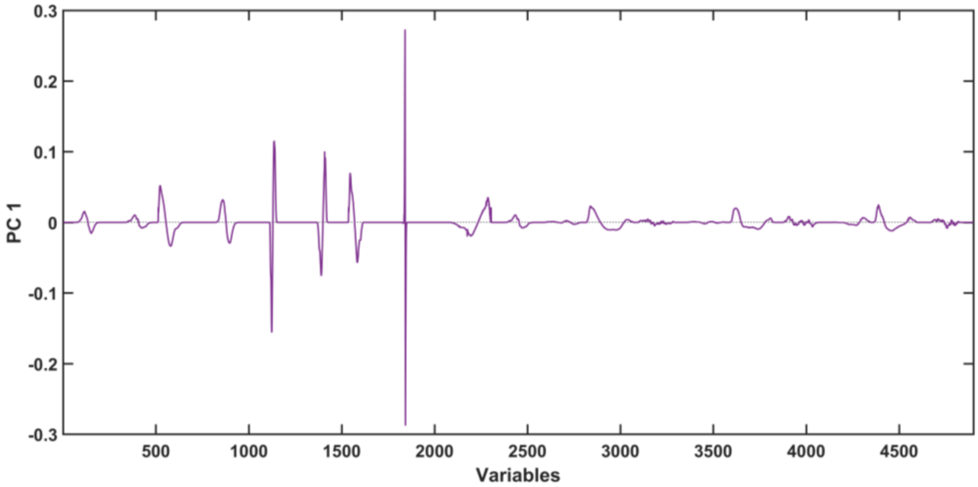

Figure 4. Loading plot associated with the first component of the PCA model.

This shows that the most significant variables are those around 1123, 1135, 1181, and 1843, which are associated with the green channel (variables of the colorgrams between 257 and 512), and the relative green (variables of the colorgrams between 1281 and 1536). Looking at the pictures (Figure 1), the fact that the green color could play a key role, was somehow expected. The PCA analysis offers an indication that the different abundance of green allows for a partial distinction between the two classes (the reader is addressed to [14] for a better understanding of the features constituting the colorgram).

The PCA analysis did not reveal any suspicious or outlier samples, so all objects were kept for the classification.

3.2. Discriminant Classification

3.2.1. SPORT Analysis

Two preprocessing approaches were tested on the colorgrams: mean-centering and autoscaling. Two different data blocks, $\mathbf{X}_1$, constituted the mean-centered spectra, and $\mathbf{X}_2$, made of the auto-scaled signals, were modeled (in this order) by the SPORT algorithm. The optimal number of LVs to be extracted from each block has been defined into a 7-fold cross-validation (CV) procedure. All of the possible combinations of LVs (under a fixed maximum of 10) have been tested and the models developed. The one leading to the lowest classification error in the CV has been retained as the optimal one and used as the calibration model. The choice fell on the model built extracting nine LVs from $\mathbf{X}_1$ and one from $\mathbf{X}_2$, which led to a correct classification rate in the CV of 72%. The application on the test set provided a correct classification rate of 77% on both the Elne class and the Torricella class. A graphical representation of the results is displayed in Figure 5.

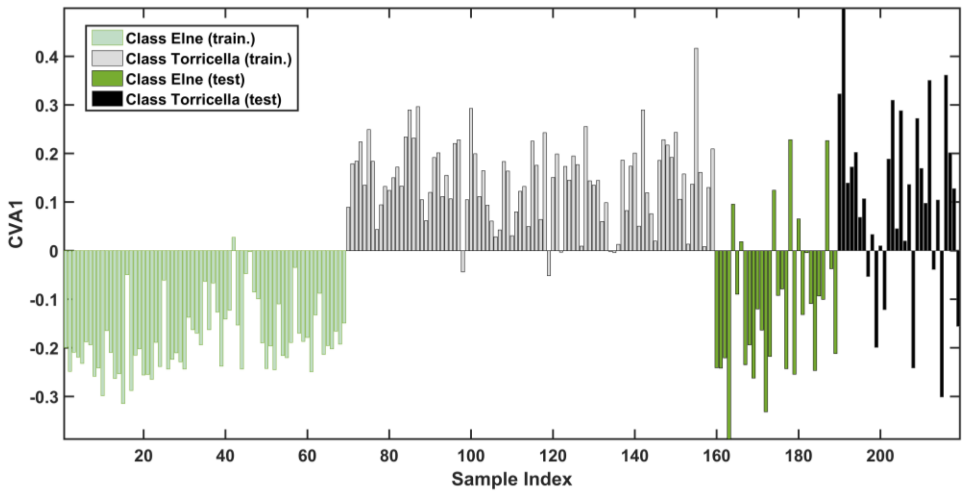

Figure 5. SPORT Analysis: distribution of the canonical variate scores (CVA1). Legend: light green: the Elne class, training samples; gray: the Torricella class, training samples; dark green: the Elne class, test samples; black: the Torricella class, test samples.

In the plot, the light green and gray bars represent the calibration samples for the Elne class and the Torricella class, respectively, while the dark green and black bars depict the validation samples for the same categories, in the same order. From this, it is straightforward that the objects belonging to the Elne class present negative values for the CVA1, whereas samples belonging to the Torricella class show the opposite tendency. The validation samples that do not follow this trend are the misclassified ones. In fact, seven samples per class are wrongly predicted; these are the six Elne objects at the positive CVA1, together with the one presenting a slightly negative score (sample index 183 in Figure 4), and the seven individuals from the Torricella class at the negative CVA1.

Eventually, to evaluate whether the feature selection allows for improving the predictions, the analysis was restricted to the variables highlighted as influential by the PCA. Therefore, of the 4900 variables, only the 512 variables related to the green and relative green channels were kept. The models were recalculated on this circumscribed set of features under the same condition described above (same pretreatments tested in a 7-fold cross-validation procedure). In this case, the model that gave rise to the greatest accuracy in the CV (80%) is the one calculated by extracting eight and one LVs from the blocks. The application of this model to the test set allowed for the misclassification of only six samples belonging to the Elne class and five objects from the Torricella class, providing a slight improvement of the predictions.

3.2.2. SO-CovSel Analysis

The SO-CovSel model was calculated, exploiting the same preprocessing used for the SPORT: mean-centering and autoscaling. The first input block ($\mathbf{X}_1$) was the one containing the mean-centered data, whereas the second one ($\mathbf{X}_2$) was constituted of the auto-scaled colorgrams. The optimal number of variables to be retained was defined into a 7-fold cross-validation procedure. All of the possible combinations of variables (the maximum was set to 30) have been tested and cross-validated models developed, in order to choose the calibration model by the inspection of the classification error in the CV. Eventually, the model was built using 16 and 13 variables from $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, which led to a classification error in the CV of 35%. The application on the test set provided a correct classification rate of 77% and 87% on the Elne class and the Torricella class, respectively, corresponding to a total accuracy of 82%.

The selected $\mathbf{X}_1$-features refer to the red and green channels (1117, 1121, 1123, 1130, 1378, 1381, 1384), hue (1840, 1841, 1842, 1844, 1845) and PCA-scores (3171, 4766, 4770, 4801); the $\mathbf{X}_2$-features are associated with the red and green channels (251, 295, 488), relative green (1410), saturation and intensity (2252, 2297, 2408) and PCA-scores (2910, 3148, 3208, 3437, 3672, 4852). The reader is addressed to Table 1 in [14] for more details on the meaning of the features.

Table 1. SIMCA results.

Pretreatment	Modelled Class	PCs	Sensitivity (%CV)	Specificity (%CV)	Efficiency (%CV)
Mean-Centering	Elne	10	79.7	40.9	57.1
Autoscaling	Elne	10	81.1	24.6	44.7
Mean-Centering	Torricella	9	50.7	57.3	53.9
Autoscaling	Torricella	10	62.2	44.0	52.3

From the predictive point of view, the results obtained by the SO-CovSel are comparable to the outcome of the SPORT+VIP. The total accuracy (82%) is equal; nevertheless, the misclassified test samples are not the same. In fact, the SPORT misclassifies six Elne samples and five Torricella celeries, whereas the SO-CovSel erroneously predicts seven Elne objects and four Torricella individuals.

3.3. Class-Modelling Analysis

Contrarily to the SPORT, the SIMCA does not allow for ensemble preprocessing. Consequently, the different models were built exploiting a 7-fold cross-validation procedure to define the optimal data pretreatment. In the SIMCA, the results are provided, in terms of sensitivity, i.e., the percentage of samples correctly accepted by the model, the specificity, i.e., the percentage of objects correctly rejected, and the efficiency, the geometrical mean of the two. Two different preprocessings were tested, the autoscaling and mean centering. Consequently, for each class, two diverse SIMCA models were calculated, and then, the pretreatment was chosen with the aim of maximizing the efficiency in the cross-validation. The results for the modelling of the two categories (Elne and Torricella) and the pretreatment are reported in Table 1.

As appreciable from Table 1, for both classes, the most suitable preprocessing was the mean-centering; in fact, the models calculated on the mean-centered data are those providing the highest efficiencies in the CV. These models were applied to the test set; for the Elne class, the 73% of both the sensitivity and specificity was achieved, whereas, for the Torricella class, the percentage of correctly accepted individuals was 63.3%, while 36.7% were properly rejected. This indicated that the SIMCA provides satisfying results for the Elne class, but it is not very specific for the Torricella class, accepting ~64% of individuals not belonging to this class.

4. Conclusions

An e-eye tool, based on the modelling of the colorgrams using different natured classifiers, finalized to the discrimination of two ecotypes of celeries, has been developed. The proposed methodologies provided different outcomes. At first, the calculation of a PCA model revealed that, as expected, the signals were characterized by the variables associated with the green channels. The class-modelling strategy appeared not very suitable for facing the investigated problem. In fact, despite being efficient on the modelling of the Elne class, it was not highly specific for the Torricella class. Moreover, the discriminant analyses achieved more than satisfying accuracies, resulting in the most appropriate strategies among the three. In conclusion, the proposed methodologies (e-eye + SPORT + VIP or e-eye + SO-CovSel), demonstrated to be rapid, non-destructive, and accurate solutions for discriminating the Elne and Torricella celeries. In fact, both the SPORT coupled with the VIP analysis and the SO-CovSel provided accurate models, properly predicting 49 (out of 60) test samples.

This represents a good starting point for the creation of fast, non-destructive tools for the recognition of celery ecotypes, useful for the protection of specific species of interest. A predictive error of about 18% is still relatively high and represents a limit for the definitive application of the tool, but predictive improvements are possible. A future challenge to overcome this issue could be the development of a multi-platform strategy including other analytical, non-destructive approaches (for example, the attenuated total reflection (ATR) FT-IR spectroscopy) and the subsequent data analysis through multi-block classifiers. A further future prospective and application of this work is the development of smart solutions for agriculture, and the quality assessment of agro-foods products.

Author Contributions: Conceptualization, A.B. and A.A.D.; Methodology, A.B.; Software, A.B. and M.F.; Validation, A.B. and A.A.D.; Formal Analysis, A.B. and M.F.; Investigation, A.B., M.F. and A.A.D.; Resources, A.A.D.; Data Curation, A.B. and M.F.; Writing-Original Draft Preparation, A.B. and M.F.; Writing-Review & Editing, A.B. and A.A.D.; Visualization, A.B. and M.F.; Supervision, A.A.D.; Project Administration, A.A.D.; Funding Acquisition, A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Torricelli, R.; Tiranti, B.; Spataro, G.; Castellini, G.; Albertini, E.; Falcinelli, M.; Negri, V. Differentiation and structure of an Italian landrace of celery (*Apium graveolens* L.): Inferences for on farm conservation. *Genet. Resour. Crop Evol.* **2013**, *60*, 995–1006. [CrossRef]
2. Raffo, A.; Sinesio, F.; Moneta, E.; Nardo, N.; Peparaio, M.; Paoletti, F. Internal quality of fresh and cold stored celery petioles described by sensory profile, chemical and instrumental measurements. *Eur. Food Res. Technol.* **2006**, *222*, 590–599. [CrossRef]
3. Yao, Y.; Sang, W.; Zhou, M.; Ren, G. Phenolic Composition and Antioxidant Activities of 11 Celery Cultivars. *J. Food Sci.* **2010**, *75*, C9–C13. [CrossRef] [PubMed]
4. Liu, G.; Zhuang, L.; Song, D.; Lu, C.; Xu, X. Isolation, purification, and identification of the main phenolic compounds from leaves of celery (*Apium graveolens* L. var. dulce Mill./Pers.). *J. Sep. Sci.* **2017**, *40*, 472–479. [CrossRef]
5. Wang, X.J.; Luo, Q.; Li, T.; Meng, P.H.; Pu, Y.T.; Liu, J.X.; Zhang, J.; Liu, H.; Tan, G.F.; Xiong, A.S. Origin, evolution, breeding, and omics of Apiaceae: A family of vegetables and medicinal plants. *Hortic. Res.* **2022**, *9*, uhac076. [CrossRef]
6. Ingallina, C.; Capitani, D.; Mannina, L.; Carradori, S.; Locatelli, M.; Di Sotto, A.; Di Giacomo, S.; Toniolo, C.; Pasqua, G.; Valletta, A.; et al. Phytochemical and biological characterization of Italian "sedano bianco di Sperlonga" Protected Geographical Indication celery ecotype: A multimethodological approach. *Food Chem.* **2020**, *309*, 125649. [CrossRef]

7. Tirillini, B.; Pellegrino, R.; Pagiotti, R.; Pocceschi, N.; Menghini, L. Volatile compounds in different cultivars of *Apium graveolens* L. *Ital. J. Food Sci.* **2004**, *16*, 477–482.
8. Reale, S.; Di Cecco, V.; Di Donato, F.; Di Martino, L.; Manzi, A.; Di Santo, M.; D'Archivio, A.A. Characterization of the Volatile Profile of Cultivated and Wild-Type Italian Celery (*Apium graveolens* L.) Varieties by HS-SPME/GC-MS. *Appl. Sci.* **2021**, *11*, 5855. [CrossRef]
9. Salvucci, G.; Pallottino, F.; De Laurentiis, L.; Del Frate, F.; Manganiello, R.; Tocci, F.; Vasta, S.; Figorilli, S.; Bassotti, B.; Violino, S.; et al. Fast olive quality assessment through RGB images and advanced convolutional neural network modelling. *Eur. Food Res. Technol.* **2022**, *248*, 1395–1405. [CrossRef]
10. Godoy, A.C.; dos Santos, P.D.S.; Nakano, A.Y.; Bini, R.A.; Siepmann, D.A.B.; Schneider, R.; Gaspar, P.A.; Pfrimer, F.W.D.; da Paz, R.F.; Santos, O.O. Analysis of Vegetable Oil from Different Suppliers by Chemometric Techniques to Ensure Correct Classification of Oil Sources to Deal with Counterfeiting. *Food Anal. Methods* **2020**, *13*, 1138–1147. [CrossRef]
11. Agarwal, A.; Dutta Gupta, S. Assessment of spinach seedling health status and chlorophyll content by multivariate data analysis and multiple linear regression of leaf image features. *Comput. Electron. Agric.* **2018**, *152*, 281–289. [CrossRef]
12. Venora, G.; Grillo, O.; Shahin, M.A.; Symons, S.J. Identification of Sicilian landraces and Canadian cultivars of lentil using an image analysis system. *Food Res. Int.* **2007**, *40*, 161–166. [CrossRef]
13. Ropelewska, E.; Sabanci, K.; Aslan, M.F. Authentication of tomato (*Solanum lycopersicum* L.) cultivars using discriminative models based on texture parameters of flesh and skin images. *Eur. Food Res. Technol.* **2022**, *248*, 1959–1976. [CrossRef]
14. Antonelli, A.; Cocchi, M.; Fava, P.; Foca, G.; Franchini, G.C.; Manzini, D.; Ulrici, A. Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm. *Anal. Chim. Acta* **2004**, *515*, 3–13. [CrossRef]
15. Roger, J.-M.; Biancolillo, A.; Marini, F. Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy. *Chemom. Intell. Lab. Syst.* **2020**, *199*, 103975. [CrossRef]
16. Foschi, M.; D'Addario, A.; Antonio D'Archivio, A.; Biancolillo, A. Future foods protection: Supervised chemometric approaches for the determination of adulterated insects' flours for human consumption by means of ATR-FTIR spectroscopy. *Microchem. J.* **2022**, *183*, 108021. [CrossRef]
17. Biancolillo, A.; Di Donato, F.; Merola, F.; Marini, F.; D'Archivio, A.A. Sequential data fusion techniques for the authentication of the P.G.I. senise ("crusco") bell pepper. *Appl. Sci.* **2021**, *11*, 1709. [CrossRef]
18. Maraphum, K.; Saengprachatanarug, K.; Wongpichet, S.; Phuphuphud, A.; Posom, J. Achieving robustness across different ages and cultivars for an NIRS-PLSR model of fresh cassava root starch and dry matter content. *Comput. Electron. Agric.* **2022**, *196*, 106872. [CrossRef]
19. Giannetti, V.; Mariani, M.B.; Marini, F.; Torrelli, P.; Biancolillo, A. Grappa and Italian spirits: Multi-platform investigation based on GC–MS, MIR and NIR spectroscopies for the authentication of the Geographical Indication. *Microchem. J.* **2020**, *157*, 104896. [CrossRef]
20. Schiavone, S.; Marchionni, B.; Bucci, R.; Marini, F.; Biancolillo, A. Authentication of Grappa (Italian grape marc spirit) by Mid and Near Infrared spectroscopies coupled with chemometrics. *Vib. Spectrosc.* **2020**, *107*, 103040. [CrossRef]
21. Biancolillo, A.; Foschi, M.; D'Archivio, A.A. Geographical Classification of Italian Saffron (*Crocus sativus* L.) by Multi-Block Treatments of UV-Vis and IR Spectroscopic Data. *Molecules* **2020**, *25*, 2332. [CrossRef] [PubMed]
22. Liu, Z.; Yang, S.; Wang, Y.; Zhang, J. Multi-platform integration based on NIR and UV–Vis spectroscopies for the geographical traceability of the fruits of Amomum tsao-ko. *Spectrochim. Acta-Part A Mol. Biomol. Spectrosc.* **2021**, *258*, 119872. [CrossRef] [PubMed]
23. Jiménez-Carvelo, A.M.; González-Casado, A.; Bagur-González, M.G.; Cuadros-Rodríguez, L. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity—A review. *Food Res. Int.* **2019**, *122*, 25–39. [CrossRef] [PubMed]
24. Jolliffe, I. Principal Component Analysis. In *Encyclopedia of Statistics in Behavioral Science*; American Cancer Society: Atlanta, GA, USA, 2005; ISBN 9780470013199.
25. Biancolillo, A.; Marini, F.; Ruckebusch, C.; Vitale, R. Chemometric strategies for spectroscopy-based food authentication. *Appl. Sci.* **2020**, *10*, 6544. [CrossRef]
26. Li Vigni, M.; Durante, C.; Cocchi, M. Exploratory Data Analysis. In *Data Handling in Science and Technology*; Marini, F., Ed.; Elsevier: Amsterdam, The Netherlands, 2013; Volume 28, pp. 55–126.
27. Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Comprehensive Analytical Chemistry*; Jaumot, J., Bedia, C., Tauler, R., Eds.; Elsevier B.V.: Amsterdam, The Netherlands, 2018; Volume 82, pp. 265–299. ISBN 9780444640444.
28. Picca, A.; Ponziani, F.R.; Calvani, R.; Marini, F.; Biancolillo, A.; Coelho-Junior, H.J.; Gervasoni, J.; Primiano, A.; Putignani, L.; Del Chierico, F.; et al. Gut microbial, inflammatory and metabolic signatures in older people with physical frailty and sarcopenia: Results from the BIOSPHERE study. *Nutrients* **2020**, *12*, 65. [CrossRef]
29. Næs, T.; Tomic, O.; Mevik, B.-H.; Martens, H. Path modelling by sequential PLS regression. *J. Chemom.* **2011**, *25*, 28–40. [CrossRef]
30. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]
31. Indahl, U.G.; Martens, H.; Næs, T. From dummy regression to prior probabilities in PLS-DA. *J. Chemom.* **2007**, *21*, 529–536. [CrossRef]

32. Biancolillo, A.; Næs, T. The Sequential and Orthogonalized PLS Regression for Multiblock Regression: Theory, Examples, and Extensions. *Data Handl. Sci. Technol.* **2019**, *31*, 157–177. [CrossRef]
33. Roger, J.M.; Palagos, B.; Bertrand, D.; Fernandez-Ahumada, E. CovSel: Variable selection for highly multivariate and multi-response calibration. Application to IR spectroscopy. *Chemom. Intell. Lab. Syst.* **2011**, *106*, 216–223. [CrossRef]
34. Biancolillo, A.; Marini, F.; Roger, J.-M. SO-CovSel: A novel method for variable selection in a multiblock framework. *J. Chemom.* **2020**, *34*, e3120. [CrossRef]
35. Wold, S. Sjöström, M. SIMCA: A method for analysing chemical data in terms of similarity and analogy. In *Chemometrics, Theory and Application*; Kowalski, B.R., Ed.; American Chemical Society: Washington, DC, USA, 1977; pp. 243–282.