



Article

Development of a Non-Destructive Tool Based on E-Eye and Agro-Morphological Descriptors for the Characterization and Classification of Different Brassicaceae Landraces

Alessandra Biancolillo, Rossella Ferretti, Claudia Scappaticci, Martina Foschi, Angelo Antonio D'Archivio, Marco Di Santo and Luciano Di Martino

Special Issue

Innovative Technologies in Food Detection

Edited by

Dr. Kaiqiang Wang and Dr. Weiwei Cheng



Article

Development of a Non-Destructive Tool Based on E-Eye and Agro-Morphological Descriptors for the Characterization and Classification of Different Brassicaceae Landraces

Alessandra Biancolillo ¹, Rossella Ferretti ², Claudia Scappaticci ¹, Martina Foschi ¹,
Angelo Antonio D'Archivio ^{1,*}, Marco Di Santo ² and Luciano Di Martino ^{2,*}

- ¹ Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell'Aquila, Via Vetoio, 67100 L'Aquila, Italy; alessandra.biancolillo@univaq.it (A.B.); claudia.scappaticci@graduate.univaq.it (C.S.); martina.foschi@univaq.it (M.F.)
- ² Majella Seed Bank-Parco Nazionale della Majella, Via Badia 28, 67039 Sulmona, Italy; rossella.ferretti93@gmail.com (R.F.); marco.disanto@parcomajella.it (M.D.S.)
- * Correspondence: angeloantonio.darchivio@univaq.it (A.A.D.); luciano.dimartino@parcomajella.it (L.D.M.)

Featured Application: Non-destructive characterization and classification of Brassicaceae landraces.

Abstract: In recent years, Brassicaceae have piqued the interest of researchers due to their extremely rich chemical composition, particularly the abundance of antioxidants and anti-inflammatory compounds, as well as because of their antimutagenic and potential anticarcinogenic activity. Vegetables in this family can be found practically everywhere on the planet. In Italy, numerous varieties of Brassicaceae, as well as a diverse pool of local variants, are regularly cultivated. These landraces, which have a variety of peculiar features, have recently sparked increased interest, and the need to safeguard them to preserve genetic biodiversity has become a relevant topic. In the present study, eight distinct Brassicaceae folk varieties were studied using non-destructive tools (Multivariate Image analysis and agro-morphological descriptors). Eventually, the data were handled using explorative analysis (EA) and Soft Independent Modeling by Class Analogy (SIMCA). EA pointed out similarities/dissimilarities among the diverse investigated populations. SIMCA led to high sensitivity (>70%) in prediction (on the external test set) for seven (over eight) investigated classes. Although the investigated plants belong to different landraces, they bear strong similarities. This is mainly linked to the ability of Brassicaceae to hybridize. Despite this, the combination of colorgrams and SIMCA allowed for classifying samples with excellent accuracy.

Keywords: Brassicaceae; local varieties; biodiversity; agro-morphological descriptors; classification; SIMCA; multivariate image analysis; e-eye analysis; explorative analysis



Citation: Biancolillo, A.; Ferretti, R.; Scappaticci, C.; Foschi, M.; D'Archivio, A.A.; Di Santo, M.; Di Martino, L. Development of a Non-Destructive Tool Based on E-Eye and Agro-Morphological Descriptors for the Characterization and Classification of Different Brassicaceae Landraces. *Appl. Sci.* **2023**, *13*, 6591. <https://doi.org/10.3390/app13116591>

Academic Editors: Kaiqiang Wang, Weiwei Cheng and Monica Gallo

Received: 27 April 2023

Revised: 18 May 2023

Accepted: 27 May 2023

Published: 29 May 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, some plants, particularly those in the Brassicaceae family, have sparked considerable interest owing to their extremely rich chemical composition, particularly the abundance of antioxidant, anti-inflammatory, and antimicrobial substances, as well as their antimutagenic activity and potential anticarcinogenic effects [1–9]. In fact, this family has about 4000 distinct species that include vital antioxidants, such as vitamin C and flavonoids, that have been shown to benefit human health [10–13]. Furthermore, they are rich in glucosinolates [14], which have anti-inflammatory and cardioprotective properties [15]. Additionally, these compounds are efficient chemopreventive agents both in vitro and in vivo [16–20].

In general, plants in this family can be found almost all over the world. Numerous types of Brassicaceae are commonly cultivated in Italy, and there is also a wide pool of local variants. Recently, these landraces, which have various and peculiar characteristics,

have attracted increasing interest, and the necessity to protect them to preserve genetic biodiversity has become a more pressing concern. Despite the fact that the concept of safeguarding botanical genetic heritage applies to all plants, it is especially important for all those folk varieties that were previously discarded by seed companies to select more marketable plants (for example, those that are less prone to hybridization or allow for simultaneous harvesting rather than a scalar one). As a result, the institutions have recently provided assistance to marginal agriculture in order to conserve local plant varieties in situ and ex situ. In this context, the current study intends to investigate various folk varieties of Brassicaceae growing in Abruzzo (Central Italy). These landraces are Mugnoli broccoli, Guardiagrele turnip, curly kale from Lama dei Peligni, Rapa senza testa, Cima dell'Osento, Cima 90° San Marzano, and Cima Grande. The importance of characterizing and authenticating these local varieties lies in the fact that this can help protect and enhance these niche varieties.

Mugnoli is a plant that most likely evolved from a cross between cabbage (*Brassica oleracea* L.) and turnip (*Brassica rapa* L.) [21]. Indeed, the mugnoli exhibit traits of both parental species, with a predominance of cabbage characteristics. However, among their populations, some accessions present more pronounced traits of the other parental species (turnip). Furthermore, individuals generated via hybridization with other brassica species growing in the area are also common. Like other Brassicaceae, this plant is noticeable from a nutritional point of view [22–24]; among the other benefits, it is particularly rich in glucosinolates [25]. The Guardiagrele turnip is an old variety of *Brassica napus* grown in the homonymous town that has the accentuated characteristics of turnip (*Brassica rapa*). Similarly, the curly kale of Lama dei Peligni is a *Brassica oleracea* var. *sabellica* grown at the foot of Majella Mountain. Rapa senza testa is a *Brassica rapa* that does not have a main shoot but many leaves. Cima dell'Osento is a variety of *Brassica napus* grown in the plain of the Osento River. Eventually, Cima 90° San Marzano and Cima Grande will be the two commercial varieties investigated together with the local varieties for comparison.

In the present work, the above-indicated landraces were investigated using two non-destructive approaches: the estimation of agro-morphological descriptors (in accordance with the Agricultural Biodiversity Working Group) [26] and multivariate image analysis (MIA) [27] through e-eye analysis [28]. A similar attempt to exploit different analytical techniques has already been discussed in the literature; it is the case of the work of Sohn and collaborators [29], who have exploited visible-near infrared spectroscopy coupled with chemometric approaches for the discrimination of eight *Brassica napus* cultivars from Korea. The proposed strategies provided high accuracy. In the present work, the data acquired by estimating agro-morphological descriptors and MIA were then chemometrically analyzed utilizing exploratory analysis [30,31] and class-modeling approaches [32]. These strategies have been exploited because they prove to be effective in comparable situations [33–35]. The goal is to emphasize the traits and examine the similarities and differences of these peculiar plants that have not yet been thoroughly studied.

2. Materials and Methods

2.1. Samples

In this work, several Brassicaceae landraces from the Abruzzo Region (Central Italy) were grown on four plots of land in the same experimental field located in Pettorano sul Gizio (Abruzzo, Central Italy). These local varieties (and the related denomination of class) are: Mugnoli broccoli A (Class A), Mugnoli broccoli B (Class B), Mugnoli broccoli C ©, Guardiagrele turnip (Class D), curly kale from Lama dei Peligni (Class E), Rapa senza testa (Class F), Cima dell'Osento (Class F), Cima 90° San Marzano (Class G), and Cima Grande (Class H). All samples are folk varieties, except for the last two, which are commercial. As can be noticed, three categories of Mugnoli have been taken into consideration because three different accessions (from different producers) were available.

As can be appreciated by Figure 1, the leaves of all plants show a strong resemblance, which does not allow a straightforward distinction, which requires high botanical expertise.

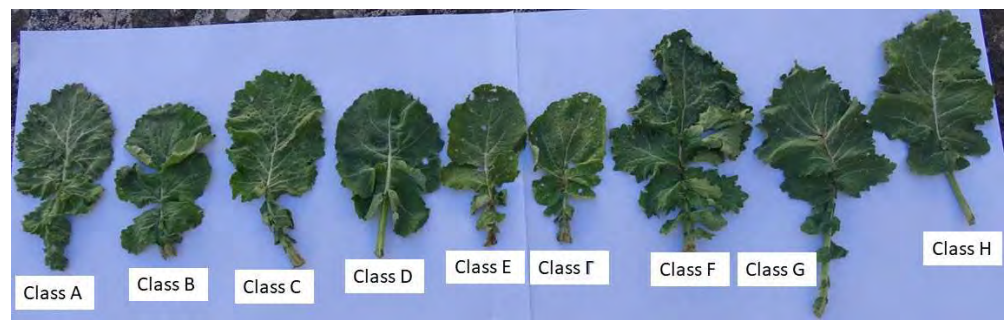


Figure 1. Pictures of leaves of the investigated landraces. Legend: Mugnoli Broccoli A (Class A), Mugnoli Broccoli B (Class B), Mugnoli Broccoli C (C), Guardiagrele Turnip (Class D), curly kale from Lama dei Peligni (Class E), Rapa senza testa (Class F), Cima dell’Osento (Class F), Cima 90° San Marzano (Class G), and Cima Grande (Class H).

As mentioned, the investigated plants were grown in an experimental field divided into four sub-plots, which were organized as shown in Figure 2. Each sub-field consisted of nine rows (one for each class), including 10 plants. All seeds were sown on 8th September 2022, and the plants were harvested on 6th February 2023. The sowing was carried out in such a way as to randomize the landrace order in the rows in the different sub-plots.

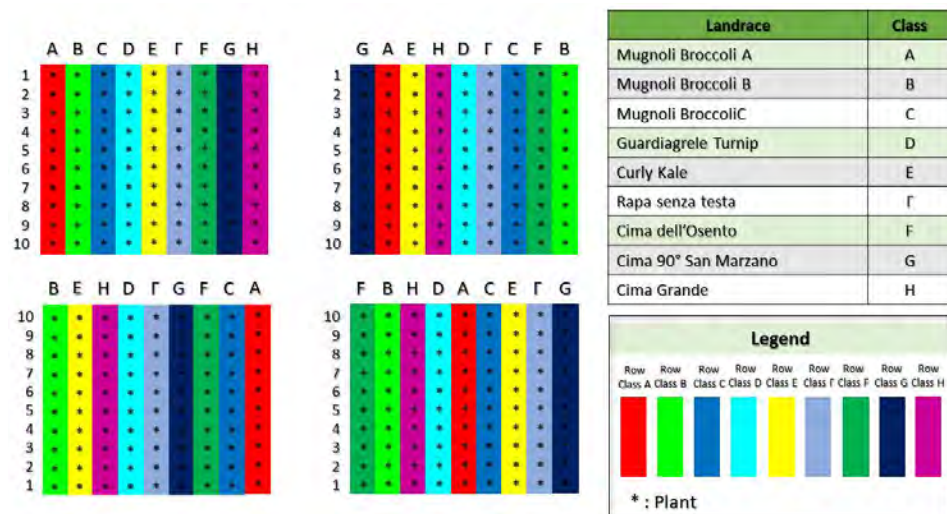


Figure 2. Organization of the experimental field. Plants were sown in rows according to their class. Each row consisted of 10 plants, and the order of the rows was randomized in the 4 sub-plots.

It is important to note that the plants have been sown and grown in the same experimental field in order to remove the variability related to the soil and climatic conditions. Consequently, the modeled variability is mainly related to inter- and intra-landrace differences.

Unfortunately, the growth of the Rapa senza testa plants failed, and it was not possible to obtain a statistically significant number of samples. As a result, Class F was removed from the analysis and will not be mentioned further.

2.2. Morphological Analysis

To characterize the investigated Brassicaceae landraces, the 27 morpho-agronomic features of plants and leaves listed and described in Table 1 were used. These descriptors are those suggested by the Gruppo di Lavoro Biodiversità in Agricoltura (GIBA) [26].

Table 1. Morpho-agronomic descriptors defined by GIBA.

GLBA	Plant Part	Descriptors	Expression
2	Leaf	Plant habit	Erect/semi-erect/horizontal
3	Leaf	curvature	Absent/weak/medium/strong/very strong
4	Leaf	Green color	Very light/light/medium/dark/very dark
5	Leaf	Leaf type	Integer/Lobed
6	Leaf	Number of lobes	Low/medium/high
7	Leaf	Incisions	Very superficial/superficial/medium/deep/very deep
8	Leaf	Waviness of the leaf margin	Absent/weak/medium/strong/very strong
9	Leaf	Serrated margin	Absent/weak/medium/strong/very strong
10	Leaf	Length	Short/medium/long
11	Leaf	Width	Small/medium/large
12	Leaf	Length of the terminal lobe	Narrow/medium/wide
13	Leaf	Width of the terminal lobe	Narrow/medium/wide
14	Leaf	Villous surface upper leaf	Absent/weak/medium/strong/very strong
15	Leaf	Anthocyanin pigmentation	Absent/weak/medium/strong/very strong
16	Root	Position	Very shallow/shallow/medium/deep/very deep
17	Root	Suborous layer of the epidermis	Absent/present
18	Root	Color of the epidermis outside the soil	White/green/yellow/orange/bronze/scarlet/red/purple red/purple blue
19	Root	Intensity of the color of the epidermis outside the soil	Light/medium/dark
20	Root	Color of the epidermis inside the soil	White/yellow/red/purple
21	Root	Color of flesh	White/yellow
24	Root	Shape in longitudinal section	narrow transverse elliptical/transverse elliptical/rounded/oval/squared/wide oblong/narrow oblong/obtriangular
25	Root	Length	Very short/Short/medium/long/very long
26	Root	Diameter	Small/medium/large
27	Root	Position of the widest part of the root	above the central part/in the center/below the central part
28	Root	Curvature of the main axis	Absent/present
29	Root	collar shape	Very depressed/Depressed/Flattened/Prominent/Very Prominent
30	Root	Shape of the base	Depressed/truncated/rounded/obtuse/pointed

Dataset Used for the Morphological Analysis

The organization of the plots in the experimental field is shown in Figure 2. As described, in the four plots, each row is dedicated to a category of Brassicaceae. The morphological descriptors described above were estimated for all plants (except for class I) by trained experts. Thus, the values of the descriptors were averaged by row; consequently, for each investigated class, four samples of Brassicaceae were obtained (one for each sub-plot of the field). This led to a data matrix of dimensions 32×27 .

2.3. E-Eye Analysis

The images of leaves were taken using the RS Pro Wi-Fi USB Microscope (RS Components S.r.l., Milan, Italy). This instrument exploits 1280×1024 -pixel resolution and a magnification power from 10 to 160, and the light is provided by an emitting diode lighting. To collect images, the microscope was held at a constant distance from the leaves. As an example, one of the images is shown in Figure A1 (Appendix A).

Dataset Used for the E-Eye Analysis

The e-eye analysis was conducted by collecting images of the leaves as described in Section 2.3. For each investigated Brassicaceae landrace, 30 photos per class were collected, for a total of 240 pictures. The colorgrams were created using MATLAB (R2015b; The Mathworks, Natick, MA, USA), following the technique outlined in [28]. As a result, a total of 240 colorgrams were obtained, which were then divided into training and test sets as described in Section 3.2. The average colorgram is shown in Figure A2 (Appendix A).

2.4. Chemometric Modeling

Given the different nature of the two datasets, different chemometric workflows were applied. Indeed, due to the limited number of analyzed samples, morphological descriptors

were investigated by exploratory analysis, in particular by principal component analysis (PCA) [36] and hierarchical cluster analysis (HCA) using average linkage [37].

PCA is one of the most common exploratory analysis (EA) methods. It is based on the concept that a data matrix \mathbf{X} can be decomposed into a set of scores (\mathbf{T} , called *principal components*) and one of the loadings (\mathbf{P}), according to the expression:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (1)$$

This allows the compression of the information present in the data into a very small sub-space and creates very useful graphs for the interpretation of similarities/differences among the analyzed samples [30].

HCA is another EA approach, often used in data mining. It allows for the creation of trees that express the degree of similarity between objects.

On the other hand, e-eye analysis was applied to a greater number of samples; therefore, it was possible to apply a predictive classification method and validate models using an external set of objects. In particular, Soft Independent Modeling by Class Analogy (SIMCA) [38] was used to model each investigated category.

SIMCA has been designed for the analysis of individual categories. It assumes that samples belonging to the same class have similarities that can be captured by a PCA model. Consequently, the building of a SIMCA model begins with the creation of a PCA model for each category of interest. Subsequently, for each i – th object, a distance d_i is estimated as follows:

$$d_i = \sqrt{T_{red,i}^2 + Q_{red,i}^2} \quad (2)$$

where T_{red}^2 and Q_{red}^2 represent the normalized (by the 95th percentiles of their distributions) Mahalanobis distance from the center of the scores' space and the orthogonal distance, respectively. Eventually, if $d_i < \sqrt{2}$ the i – th object is accepted by the modeled class (and predicted as belonging to it); otherwise, it is rejected (and predicted as not appertaining to this category).

3. Results

As mentioned, the chemometric workflows applied to agro-morphological descriptors and e-eye profiles are different. Nevertheless, disregarding the approach used, data were auto-scaled before modeling.

Agro-morphological descriptors were analyzed by EA, whereas colorgrams were divided into a training and a test set (in order to carry out an external validation of the models), and SIMCA was used to classify all the different investigated landraces. More details on the used procedures are provided in the dedicated sections.

3.1. Explorative Analysis of Morphological Descriptors

Figure 3 shows the hierarchical cluster tree obtained by applying the average linkage method on morphological descriptors.

The cluster analysis shows a good degree of similarity among samples belonging to the same landrace. The dendrogram indicates that all mugnoli (Class A, B, and C) are similar to each other from an agro-morphological point of view and form clusters among one another (light blue and green clusters). All the other landraces except Class F form distinct clusters (yellow for Class D, purple for Class G, and magenta for Class H). Samples belonging to Class F are divided into two sub-clusters, one of each mixed with Class E. This is not unexpected and is ascribable to the high internal variability present in the classes (given by the strong tendency of the Brassicaceae to hybridize with each other).

Principal component analysis was then used in order to maximize the extraction of multivariate information from the data. The model, which required a total of 8 PCs, allowed for the collection of 88% of the explained variance.

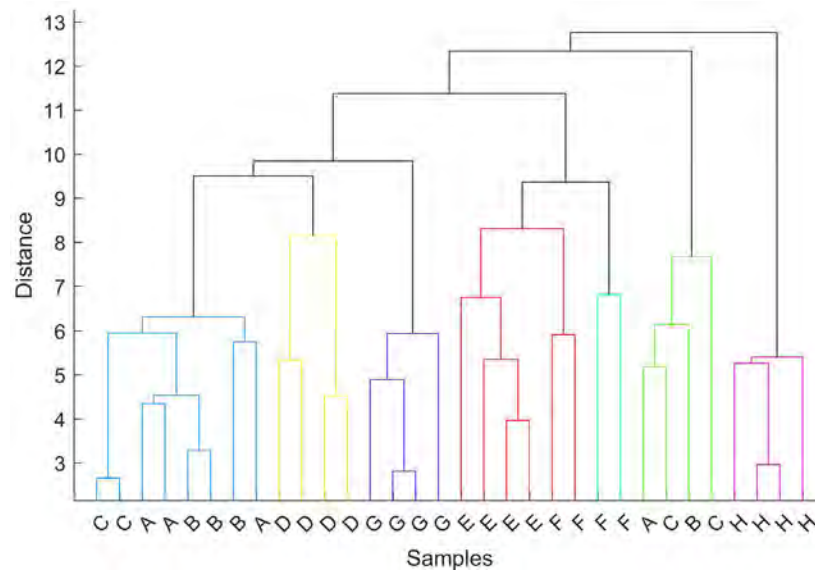


Figure 3. Hierarchical clustering using average linkage of the morphological descriptors. Legend: Mugnoli Broccoli A (Class A), Mugnoli Broccoli B (Class B), Mugnoli Broccoli C (C), Guardiagrele Turnip (Class D), curly kale from Lama dei Peligni (Class E), Cima dell’Osento (Class F), Cima 90° San Marzano (Class G), and Cima Grande (Class H).

The score plot (Figure 4) demonstrates how the samples are distributed in the space of the first two principal components (PC1 and PC2). Although this representation takes into account only part of the variance explained by the model, it is feasible to spot distinct grouping tendencies among samples belonging to the same population. Indeed, the mugnoli fall all together at negative PC1, showing a relevant degree of overlapping. Moreover, they are very close to samples belonging to Class D, in agreement with the hierarchical cluster analysis. Plants belonging to Classes G and H present positive or slightly negative values of PC1 but can be distinguished from all the other individuals along PC2, where they fall at negative values. Samples belonging to Classes E and F fall at positive values of both PCs, showing similar behavior, reflecting the similarity already pointed out by cluster analysis.

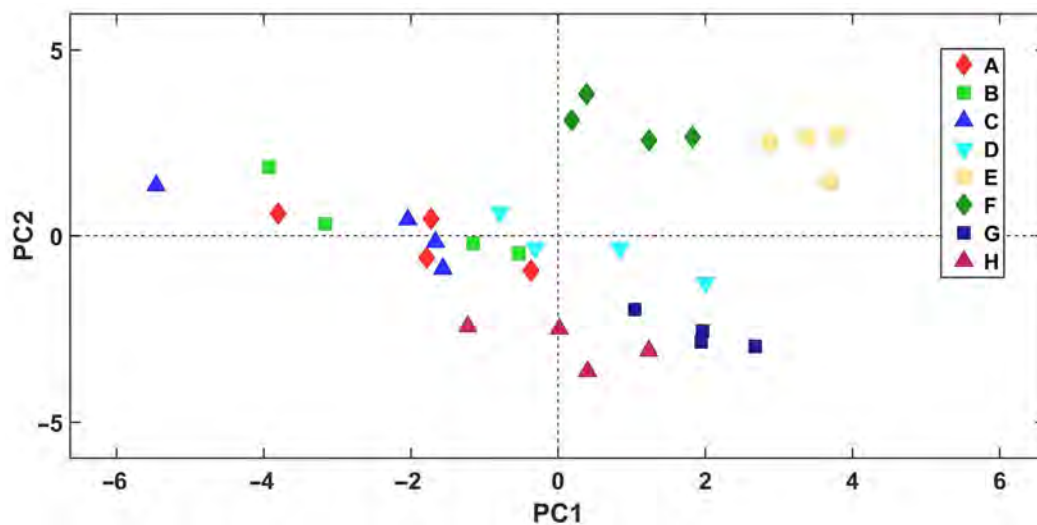


Figure 4. PCA of the agro-morphological descriptors. Legend: Mugnoli Broccoli A (Class A), Mugnoli Broccoli B (Class B), Mugnoli Broccoli C (C), Guardiagrele Turnip (Class D), curly kale from Lama dei Peligni (Class E), Cima dell’Osento (Class F), Cima 90° San Marzano (Class G), and Cima Grande (Class H).

3.2. SIMCA Analysis of the E-Eye Profiles

Being a class-modeling approach, SIMCA allows the classification of individual categories of interest. Due to the peculiarity and uniqueness of the investigated landraces, all eight available categories of Brassicaceae were individually modeled. To allow external validation of the models, samples were divided into a training set of 160 samples (20 for each class) and a test set of 80 objects (10 for each category). In order to ensure representativeness, the sample splitting was carried out by applying the Duplex algorithm [4].

For calculating SIMCA models, it is necessary to define the optimal number of PCs to be extracted. In order to avoid over-optimistic estimations, the optimal complexity of the model was defined through a seven-fold cross-validation procedure. In order to do so, for each class, 10 different calibration models (built using an increasing number of LVs from one up to 10) were calculated, and three cross-validated figures of merit were retained: sensibility ($Sens_{cv}$), sensitivity ($Spec_{cv}$), and efficiency (Eff_{cv}). These entities represent the percentage of samples properly accepted by the model, the percentage of individuals properly rejected by the class model, and their geometric average, respectively. Eventually, the examination of the Eff_{cv} revealed the suitable number of PCs to be retained; in particular, the number of PCs that provided the most accurate Eff_{cv} was taken. The number of PCs selected for every model, cross-validated sensibilities, specificities, and efficiencies are reported in Table 2.

Table 2. SIMCA models for each class. Number of selected PCs, $Sens_{cv}$, $Spec_{cv}$, and Eff_{cv} .

Class	PCs	$Sens_{cv}$	$Spec_{cv}$	Eff_{cv}	$Sens_{pred}$	$Spec_{pred}$
A	6	55.0	45.0	39.7	30.0	37.1
B	4	90.0	86.9	88.4	100.0	92.8
C	4	75.0	75.4	75.2	70.0	77.1
D	6	55.0	39.2	46.4	70.0	32.8
E	4	80.0	69.1	74.3	90.0	68.6
F	6	75.0	61.3	63.1	80.0	68.6
G	4	75.0	60.8	67.5	90.0	67.1
H	4	80.0	68.2	73.8	80.0	65.7

From the table, it can be noted that all models, except for those dedicated to Class A and Class D, provide high sensitivities and good specificities, which lead to satisfying efficiencies. All the calibration models have been used to predict the test set, leading to the predicted sensitivities ($Sens_{pred}$) and specificities ($Spec_{pred}$) reported in Table 2. From this, it is possible to confirm the trend already observed in cross-validated calibration models. Indeed, all predictive models provide high sensitivity and good specificity except for Class A and, to a lesser extent, Class D.

In Table 3, the specificities with respect to all the categories ($SpecwrtA$, $SpecwrtB$, $SpecwrtC$, $SpecwrtD$, $SpecwrtE$, $SpecwrtF$, $SpecwrtG$, $SpecwrtH$) are shown. These have been reported to deeply investigate which classes are confused with one another. In fact, the specificity with respect to another category represents the percentage of samples (belonging to the other class) that have been properly rejected. This means that if, after modeling one class, all samples belonging to another category are rejected, the specificity with respect to the modeled class will be 100.0. As a consequence, Table 3 reveals which classes are the most similar from the e-eye point of view.

Table 3. Prediction of SIMCA models on the test set. Specificities with respect to all the investigated classes (Spec_{wrtA}, Spec_{wrtB}, Spec_{wrtC}, Spec_{wrtD}, Spec_{wrtE}, Spec_{wrtF}, Spec_{wrtG}, Spec_{wrtH}) are displayed.

	Spec _{wrtA}	Spec _{wrtB}	Spec _{wrtC}	Spec _{wrtD}	Spec _{wrtE}	Spec _{wrtF}	Spec _{wrtG}	Spec _{wrtH}
A	–	0.0	30.0	80.0	50.0	20.0	30.0	50.0
B	100.0	–	70.0	100.0	100.0	90.0	90.0	100.0
C	90.0	0.0	–	90.0	100.0	70.0	100.0	90.0
D	80.0	30.0	50.0	–	20.0	20.0	10.0	20.0
E	100.0	80.0	80.0	90.0	–	70.0	40.0	20.0
F	100.0	10.0	60.0	80.0	70.0	–	80.0	80.0
G	90.0	90.0	90.0	70.0	30.0	70.0	–	30.0
H	100.0	70.0	80.0	80.0	20.0	60.0	50.0	–

This kind of inspection highlights that the low accuracy provided by the model of Class A is mainly because it erroneously accepts all samples belonging to Class B, 80% of Class F, and 70% of individuals from Classes C and G. This is completely expected for Class B, as it is an accession of mugnoli. A further inspection of similarities/dissimilarities among these categories is provided in Appendix B. The confusion with the other mentioned classes could be attributed to the fact that mugnoli and Class C, Class G, and Class F present the darkest leaves and, at the same time, are the richest in incisions.

Eventually, the possibility of creating SIMCA models merging the three different accessions of mugnoli into a single class was tested (not shown). This led to relatively low sensitivity and specificity for the mugnoli class in prediction (53.3% and 58.0%, respectively), indicating that the variability among the samples belonging to Class A, Class B, and Class C is relevant and that the unification of these three categories into a single class is not reasonable.

4. Conclusions

The current study has analyzed some folk varieties of Brassicaceae cultivated in Italy. In this regard, a fundamental aspect that should be noted is that all the analyzed varieties have been cultivated in the same experimental field; consequently, the modeled variability is attributable to the actual inter- and intra-population differences.

The analysis of the agro-morphological descriptors showed that, despite the similarities among plants, the analysis of these features with EA allows the observation of clusters associated with the different local varieties. Subsequently, SIMCA made it possible to create classification models based on colorgrams. This strategy has demonstrated that the combination of MIA and a class-modeling approach allows classifying the different categories of plants with high accuracy. Indeed, the models of all classes correctly accepted a relevant percentage of samples (>70%) except for accession of mugnoli, which is confused with the other mugnoli or with the other two categories. This is not surprising and is to be considered related to the strong ability that Brassicaceae have to hybridize.

This suggests that an important future perspective is an investigation of the DNA of these plants. In conclusion, these promising results indicate that it is possible to develop non-destructive approaches for the characterization and classification of Brassicaceae varieties. From a future perspective, destructive strategies such as DNA analysis or chromatographic analysis of the extracts could be investigated. On the other hand, non-destructive analyses could be pursued using ATR-FT-IR spectroscopy or hyperspectral analysis.

Author Contributions: Conceptualization, A.A.D. and L.D.M.; methodology, A.B. and R.F.; software, M.F. and A.B.; validation, R.F., C.S. and M.F.; investigation, M.F., R.F., C.S. and M.D.S.; resources, M.D.S. and L.D.M.; data curation, C.S. and M.F.; writing—original draft preparation, A.B.; writing—review and editing, A.B. and A.A.D.; supervision, A.A.D. and L.D.M.; project administration, M.D.S.; funding acquisition, L.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Majella National Park within the Project “Tipizzazione di specie vegetali endemiche, crop wild relatives e varietà agricole autoctone del Parco della Majella mediante metodi analitici ed approcci statistici multivariati” with the Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell’Aquila.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available on request.

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

Appendix A

By way of example, Figure A1 shows one of the pictures collected on one of the leaves. A transparent ruler has been superimposed on the leaf in order to provide an idea of the scale. The unit used by the ruler is cm.

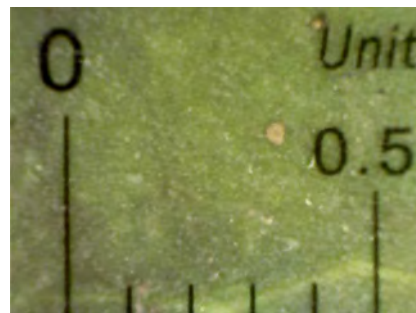


Figure A1. Picture collected on one leaf using the RS Pro Microscope. A transparent ruler has been superimposed on the leaf; the unit used is cm.

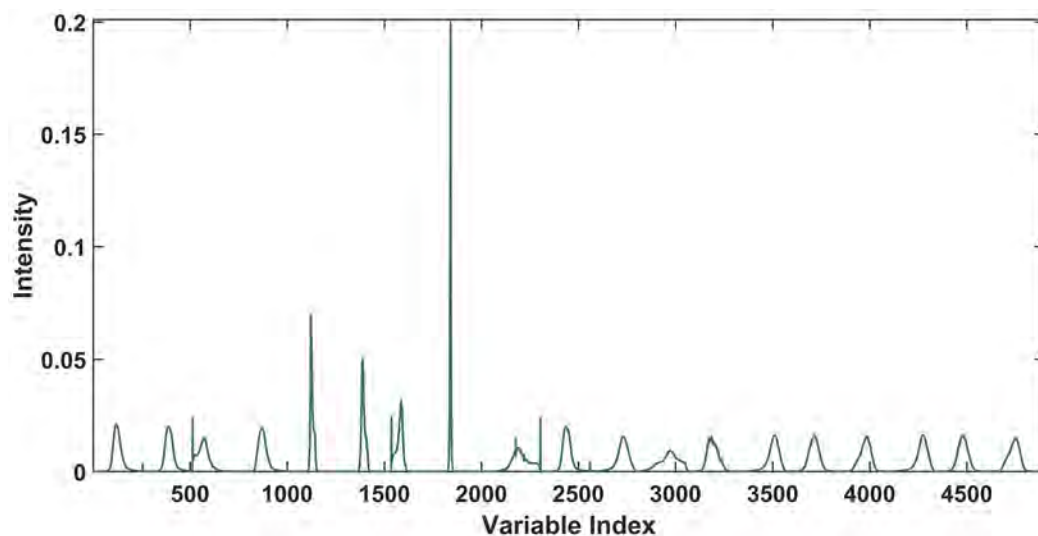


Figure A2. Mean colorgram.

Figure A2 shows the average colorgram (calculated on the signals of all samples).

As depicted in [28], the different peaks of the colorgram can be associated with specific characteristics of the image. Variables from 1 to 2560 are associated with RGB canals, hue, and saturation; in particular, the first three peaks represent the distribution curves of the red (R), green (G), and blue (B) values; the following signal is the distribution curve of the lightness values ($L = R + G + B$); then, the relative values of the R, G, and B values and the distribution curve of the hue values follow. Features from 2561 to the end are principal components extracted from the mean-centered and auto-scaled unfolded RGB matrix. The average colorgram shows the mean RGB color of the leaves given by the code [120 136 0], which is dark green. Together with the expected peaks associated with the RGB values, this shows a high peak at variable 1840 associated with the hue values.

Appendix B

In order to inspect the overlapping of the three investigated accessions of mugnoli, a PCA model has been calculated on the colorgrams of these samples. As shown in the PCA plot in Figure A3, there is quite a strong overlap among these classes. Mainly, samples belonging to Class A fall at positive values of PC1; those pertaining to Class B present negative PC1 values; and mugnoli from Class C spread along this component. Only two samples belonging to Class C present highly positive PC2 values, but the inspection of the T^2 -Q plot did not allow to consider them outliers; consequently, they were retained.

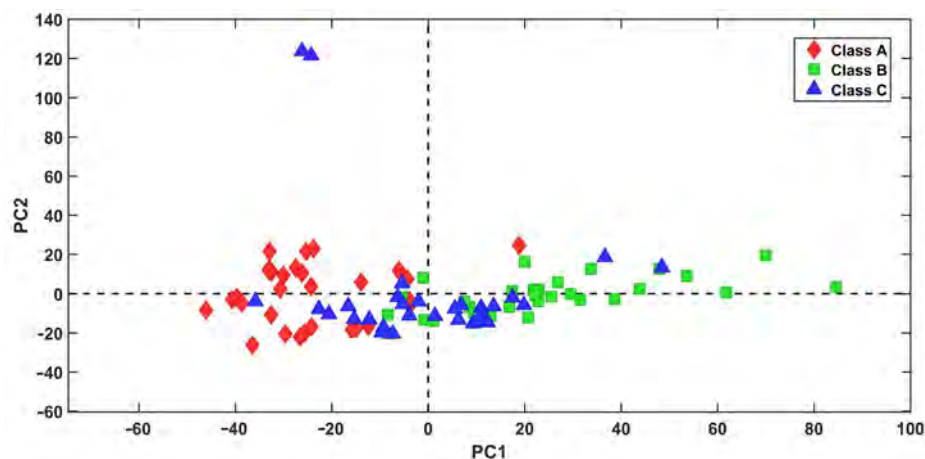


Figure A3. PCA plot of samples belonging to Class A, Class B, and Class C.

References

1. Aly, S.H.; Kandil, N.H.; Hemdan, R.M.; Kotb, S.S.; Zaki, S.S.; Abdelaziz, O.M.; AbdelRazek, M.M.M.; Almahli, H.; El Hassab, M.A.; Al-Rashood, S.T.; et al. GC/MS Profiling of the Essential Oil and Lipophilic Extract of *Moricandia sinaica* Boiss. and Evaluation of Their Cytotoxic and Antioxidant Activities. *Molecules* **2023**, *28*, 2193. [CrossRef]
2. Davì, F.; Taviano, M.F.; Acquaviva, R.; Malfa, G.A.; Cavò, E.; Arena, P.; Ragusa, S.; Cacciola, F.; El Majdoub, Y.O.; Mondello, L.; et al. Chemical Profile, Antioxidant and Cytotoxic Activity of a Phenolic-Rich Fraction from the Leaves of *Brassica fruticulosa* subsp. *fruticulosa* (Brassicaceae) Growing Wild in Sicily (Italy). *Molecules* **2023**, *28*, 2281. [CrossRef]
3. Lučić, D.; Pavlović, I.; Brkljačić, L.; Bogdanović, S.; Farkaš, V.; Cedilak, A.; Nanić, L.; Rubelj, I.; Salopek-Sondi, B. Antioxidant and Antiproliferative Activities of Kale (*Brassica oleracea* L. Var. *acephala* DC.) and Wild Cabbage (*Brassica incana* Ten.) Polyphenolic Extracts. *Molecules* **2023**, *28*, 1840. [CrossRef]
4. Malfa, G.A.; Pappalardo, F.; Miceli, N.; Taviano, M.F.; Ronsisvalle, S.; Tomasello, B.; Bianchi, S.; Davì, F.; Spadaro, V.; Acquaviva, R. Chemical, Antioxidant and Biological Studies of *Brassica incana* subsp. *raimondoi* (Brassicaceae) Leaf Extract. *Molecules* **2023**, *28*, 1254. [CrossRef]
5. Montaner, C.; Mallor, C.; Laguna, S.; Zufiaurre, R. Bioactive compounds, antioxidant activity, and mineral content of bróquil: A traditional crop of *Brassica oleracea* var. *italica*. *Front. Nutr.* **2022**, *9*, 1006012. [CrossRef] [PubMed]
6. Basit, A.; Ahmad, S.; Khan, K.U.R.; Aati, H.Y.; Sherif, A.E.; Ovatlarnporn, C.; Khan, S.; Rao, H.; Arshad, M.A.; Shahzad, M.N.; et al. Evaluation of the anti-inflammatory, antioxidant, and cytotoxic potential of *Cardamine amara* L. (Brassicaceae): A comprehensive biochemical, toxicological, and in silico computational study. *Front. Chem.* **2022**, *10*, 1077581. [CrossRef] [PubMed]
7. Hip Kam, A.; Li, W.-W.; Bahorun, T.; Neergheen, V.S. Traditional processing techniques impacted the bioactivities of selected local consumed foods. *Sci. Afr.* **2023**, *19*, e01558. [CrossRef]

8. Salami, M.; Heidari, B.; Tan, H. Comparative profiling of polyphenols and antioxidants and analysis of antiglycation activities in rapeseed (*Brassica napus* L.) under different moisture regimes. *Food Chem.* **2023**, *399*, 133946. [CrossRef]
9. Tan, J.; Jiang, H.; Li, Y.; He, R.; Liu, K.; Chen, Y.; He, X.; Liu, X.; Liu, H. Growth, Phytochemicals, and Antioxidant Activity of Kale Grown under Different Nutrient-Solution Depths in Hydroponic. *Horticulturae* **2023**, *9*, 53. [CrossRef]
10. Peña, M.; Guzmán, A.; Martínez, R.; Mesas, C.; Prados, J.; Porres, J.M.; Melguizo, C. Preventive effects of Brassicaceae family for colon cancer prevention: A focus on in vitro studies. *Biomed. Pharmacother.* **2022**, *151*, 113145. [CrossRef]
11. Bouranis, J.A.; Beaver, L.M.; Jiang, D.; Choi, J.; Wong, C.P.; Davis, E.W.; Williams, D.E.; Sharpton, T.J.; Stevens, J.F.; Ho, E. Interplay between Cruciferous Vegetables and the Gut Microbiome: A Multi-Omic Approach. *Nutrients* **2023**, *15*, 42. [CrossRef] [PubMed]
12. Khalil, H.E.; Abdelwahab, M.F.; Emeka, P.M.; Badger-Emeka, L.I.; Abdel Hafez, S.M.N.; AlYahya, K.A.; Ahmed, A.-S.F.; Anter, A.F.; Abdel-Wahab, N.M.; Matsunami, K.; et al. Chemical Composition and Valorization of Broccoli Leaf By-Products (*Brassica oleracea* L. Variety: *Italica*) to Ameliorate Reno-Hepatic Toxicity Induced by Gentamicin in Rats. *Appl. Sci.* **2022**, *12*, 6903. [CrossRef]
13. Koksai, E.; Gode, F.; Ozaltin, K.; Karakurt, I.; Suly, P.; Saha, P. Controlled Release of Vitamin U from Microencapsulated *Brassica oleracea* L. var. *capitata* Extract for Peptic Ulcer Treatment. *Food Bioprocess Technol.* **2023**, *16*, 677–689. [CrossRef]
14. Jo, J.S.; Bhandari, S.R.; Kang, G.H.; Shin, Y.K.; Lee, J.G. Selection of broccoli (*Brassica oleracea* var. *italica*) on composition and content of glucosinolates and hydrolysates. *Sci. Hortic. Amst.* **2022**, *298*, 110984. [CrossRef]
15. Favela-González, K.M.; Hernández-Almanza, A.Y.; la Fuente-Salcido, N.M. The value of bioactive compounds of cruciferous vegetables (*Brassica*) as antimicrobials and antioxidants: A review. *J. Food Biochem.* **2020**, *44*, e13414. [CrossRef] [PubMed]
16. El-Daly, S.M.; Gamal-Eldeen, A.M.; Gouhar, S.A.; Abo-elfadl, M.T.; El-Saeed, G. Modulatory Effect of Indoles on the Expression of miRNAs Regulating G1/S Cell Cycle Phase in Breast Cancer Cells. *Appl. Biochem. Biotechnol.* **2020**, *192*, 1208–1223. [CrossRef] [PubMed]
17. Wang, X.; Chan, Y.S.; Wong, K.; Yoshitake, R.; Sadava, D.; Synold, T.W.; Frankel, P.; Twardowski, P.W.; Lau, C.; Chen, S. Mechanism-Driven and Clinically Focused Development of Botanical Foods as Multitarget Anticancer Medicine: Collective Perspectives and Insights from Preclinical Studies, IND Applications and Early-Phase Clinical Trials. *Cancers* **2023**, *15*, 701. [CrossRef]
18. Zhou, T.; Zhou, M.; Tong, C.; Zhuo, M. Cauliflower bioactive compound sulforaphane inhibits breast cancer development by suppressing NF- κ B/MMP-9 signaling pathway expression. *Cell. Mol. Biol.* **2022**, *68*, 134–143. [CrossRef]
19. Khalid, W.; Iqra, Afzal, F.; Rahim, M.A.; Abdul Rehman, A.; ul Rasul, H.; Arshad, M.S.; Ambreen, S.; Zubair, M.; Safdar, S.; et al. Industrial applications of kale (*Brassica oleracea* var. *sabellica*) as a functional ingredient: A review. *Int. J. Food Prop.* **2023**, *26*, 489–501. [CrossRef]
20. Bozic, D.; Živančević, K.; Baralić, K.; Miljaković, E.A.; Djordjević, A.B.; Ćurčić, M.; Bulat, Z.; Antonijević, B.; Đukić-Ćosić, D. Conducting bioinformatics analysis to predict sulforaphane-triggered adverse outcome pathways in healthy human cells. *Biomed. Pharmacother.* **2023**, *160*, 114316. [CrossRef]
21. Spoor, W.; Zohary, D.; Hopf, M. *Domestication of Plants in the Old World*, 3rd ed.; Oxford University Press: New York, NY, USA, 2000; 316p. [CrossRef]
22. Laghetti, G.; Martignano, F.; Falco, V.; Cifarelli, S.; Gladis, T.; Hammer, K. “Mugnoli”: A Neglected Race of *Brassica oleracea* L. from Salento (Italy). *Genet. Resour. Crop Evol.* **2005**, *52*, 635–639. [CrossRef]
23. Palmitessa, O.D.; Gadaleta, A.; Leoni, B.; Renna, M.; Signore, A.; Paradiso, V.M.; Santamaria, P. Effects of Greenhouse vs. Growth Chamber and Different Blue-Light Percentages on the Growth Performance and Quality of Broccoli Microgreens. *Agronomy* **2022**, *12*, 1161. [CrossRef]
24. Hammer, K.; Montesano, V.; Drenzo, P.; Laghetti, G. Conservation of crop genetic resources in Italy with a focus on vegetables and a case study of a neglected race of brassica oleracea. *Agriculture* **2018**, *8*, 105. [CrossRef]
25. Argentieri, M.P.; Accogli, R.; Fanizzi, F.P.; Avato, P. Glucosinolates profile of “mugnolo”, a variety of *Brassica oleracea* L. native to southern Italy (Salento). *Planta Med.* **2011**, *77*, 287–292. [CrossRef]
26. GIBA. Gruppo di Lavoro Biodiversità in Agricoltura. Available online: <https://www.reterurale.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/9580> (accessed on 19 April 2023).
27. Geladi, P.; Grahn, H.F. Multivariate Image Analysis. In *Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation*; Wiley-Blackwell: Hoboken, NJ, USA, 2006.
28. 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]
29. Sohn, S.-I.; Pandian, S.; Zaukuu, J.-L.Z.; Oh, Y.-J.; Lee, Y.-H.; Shin, E.-K.; Thamilarasan, S.K.; Kang, H.-J.; Ryu, T.-H.; Cho, W.-S. Rapid discrimination of *Brassica napus* varieties using visible and Near-infrared (Vis-NIR) spectroscopy. *J. King Saud Univ. Sci.* **2023**, *35*, 102495. [CrossRef]
30. 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.
31. Biancolillo, A.; Marini, F.; Ruckebusch, C.; Vitale, R. Chemometric strategies for spectroscopy-based food authentication. *Appl. Sci.* **2020**, *10*, 6544. [CrossRef]
32. Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Data Analysis for Omic Sciences: Methods and Applications, Comprehensive Analytical Chemistry*; Jaumot, J., Bedia, C., Tauler, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; Volume 82, pp. 265–299. ISBN 9780444640444.

33. Di Donato, F.; Di Cecco, V.; Torricelli, R.; D'Archivio, A.A.; Di Santo, M.; Albertini, E.; Veronesi, F.; Garramone, R.; Aversano, R.; Marcantonio, G.; et al. Discrimination of potato (*Solanum tuberosum* L.) accessions collected in majella national park (Abruzzo, Italy) using mid-infrared spectroscopy and chemometrics combined with morphological and molecular analysis. *Appl. Sci.* **2020**, *10*, 1630. [[CrossRef](#)]
34. Calvini, R.; Orlandi, G.; Foca, G.; Ulrici, A. Colourgrams GUI: A graphical user-friendly interface for the analysis of large datasets of RGB images. *Chemom. Intell. Lab. Syst.* **2020**, *196*, 103915. [[CrossRef](#)]
35. Foschi, M.; Di Maria, V.; D'Archivio, A.A.; Marini, F.; Biancolillo, A. E-Eye-Based Approach for Traceability and Annuality Compliance of Lentils. *Appl. Sci.* **2023**, *13*, 1433. [[CrossRef](#)]
36. Jolliffe, I.T. A Note on the Use of Principal Components in Regression. *J. R. Stat. Soc. Ser. C Applied Stat.* **1982**, *31*, 300–303. [[CrossRef](#)]
37. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*, 1st ed.; John Wiley: New York, NY, USA, 1990; ISBN 0-471-87876-6.
38. 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.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.