

Article

Vis-NIR Spectroscopy and Machine Learning Methods for the Discrimination of Transgenic *Brassica napus* L. and Their Hybrids with *B. juncea*

Soo-In Sohn ^{1,*}, Subramani Pandian ^{1,†}, Young-Ju Oh ^{2,†}, John-Lewis Zinia Zaukuu ^{3,†}, Chae-Sun Na ⁴, Yong-Ho Lee ^{5,6}, Eun-Kyoung Shin ¹, Hyeon-Jung Kang ¹, Tae-Hun Ryu ¹, Woo-Suk Cho ¹ and Youn-Sung Cho ¹

¹ Department of Agricultural Biotechnology, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Korea; pandiannsp7@gmail.com (S.P.); novis7@korea.kr (E.-K.S.); happykorean@korea.kr (H.-J.K.); thryu@korea.kr (T.-H.R.); phyto@korea.kr (W.-S.C.); younsung@korea.kr (Y.-S.C.)

² Institute for Future Environmental Ecology Co., Ltd., Jeonju 54883, Korea; 50joo@hanmail.net

³ Department of Food Science and Technology, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi AK-039-5028, Ghana; zaukuu.jz@knust.edu.gh

⁴ Seed Conservation Research Division, Baekdudewgan National Arboretum, Bonghwa 36209, Korea; chaesun.na@kiam.or.kr

⁵ Institute of Ecological Phytochemistry, Hankyong National University, Anseong 17579, Korea; yongho@korea.ac.kr

⁶ OJeong Resilience Institute, Korea University, Seoul 02841, Korea

* Correspondence: sisohn@korea.kr; Tel.: +82-063-238-4712

† These authors contributed equally to this work.



Citation: Sohn, S.-I.; Pandian, S.; Oh, Y.-J.; Zaukuu, J.-L.Z.; Na, C.-S.; Lee, Y.-H.; Shin, E.-K.; Kang, H.-J.; Ryu, T.-H.; Cho, W.-S.; et al. Vis-NIR Spectroscopy and Machine Learning Methods for the Discrimination of Transgenic *Brassica napus* L. and Their Hybrids with *B. juncea*. *Processes* **2022**, *10*, 240. <https://doi.org/10.3390/pr10020240>

Academic Editor: Juan Francisco García Martín

Received: 3 December 2021

Accepted: 21 January 2022

Published: 26 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Abstract: The rapid advancement of genetically modified (GM) technology over the years has raised concerns about the safety of GM crops and foods for human health and the environment. Gene flow from GM crops may be a threat to the environment. Therefore, it is critical to develop reliable, rapid, and low-cost technologies for detecting and monitoring the presence of GM crops and crop products. Here, we used visible near-infrared (Vis-NIR) spectroscopy to distinguish between GM and non-GM *Brassica napus*, *B. juncea*, and F₁ hybrids (*B. juncea* X GM *B. napus*). The Vis-NIR spectra were preprocessed with different preprocessing methods, namely normalization, standard normal variate, and Savitzky–Golay. Both raw and preprocessed spectra were used in combination with eight different chemometric methods for the effective discrimination of GM and non-GM plants. The standard normal variate and support vector machine combination was determined to be the most accurate model in the discrimination of GM, non-GM, and hybrid plants among the many combinations (99.4%). The use of deep learning in combination with Savitzky–Golay resulted in 99.1% classification accuracy. According to the findings, it is concluded that handheld Vis-NIR spectroscopy combined with chemometric analyses could be used to distinguish between GM and non-GM *B. napus*, *B. juncea*, and F₁ hybrids.

Keywords: *Brassica napus*; *Brassica juncea*; genetically modified crops; F₁ hybrid; Vis-NIR spectroscopy; chemometrics; deep learning

1. Introduction

Brassica juncea L. Czern (Brown Mustard) is an important annual crop and is an outcome of hybridization between the diploid *Brassica* species *B. rapa* (AA, 2n = 20) and *B. nigra* (BB, 2n = 16) followed by spontaneous hybridization with chromosome doubling [1]. In China and Korea, wild *B. juncea* is a natural weedy species widely found along roadsides or empty lands [2,3]. It is known to have the highest potential for gene transfer from *B. napus* after *B. rapa* [4]. It has previously been reported that conventional and transgenic *B. napus* hybridize with *B. juncea* spontaneously or by hand pollination [5–8]. Recently, Tang

et al. [9] found the estimated frequencies of natural gene flow from the genetically modified (GM) *B. napus* to 10 different *B. juncea* cultivars in the field experiment varied from 0.08 to 0.93%. The transgenic hybrids' ability to persist is determined by their fitness as crop–wild hybrids [10]. Little is known about the fitness of the F₁ hybrid between *B. juncea* and *B. napus* in the environment. According to Lim et al. [2], seeds from a hybrid of *B. juncea* and GM *B. napus* have shown an increase in dormancy and overwintering traits, suggesting that they could become soil seed banks. Seeds in such a seed bank can germinate again if they meet a favorable environment, leading to the formation of a feral population. As a result, the transgene may spread across the ecosystem. If the flowering period of *B. juncea* overlaps with that of *B. napus*, there is a possibility of forming hybrids with GM *B. napus* and releasing them into the environment. If GM *B. napus* and hybrids (*B. juncea* X GM *B. napus*) can be quickly identified and removed, it will be useful to avoid the unintentional environmental release of transgenes and promote the safe management of GM *B. napus*.

Various methods have been used to detect genetically modified organisms (GMOs), including enzyme-linked immunosorbent assays (ELISA), lateral flow strips, biosensors, Western blots, real-time PCR, qualitative polymerase chain reaction (qPCR), microarrays, electrophoresis, Southern blots, liquid chromatography, and gas chromatography [11]. Nowadays, spectroscopy is one of the rapid, accurate, and nondestructive methods for distinguishing between GM and non-GM crops that does not require complex sample processing [11]. Spectroscopy-based GMO identification is not to detect changes in DNA or single proteins but to detect unknown structural changes due to genotype alterations generated by the introduction of transgenes for specific traits [12]. Generally, a vast number of spectroscopy methods are available for detecting structural changes in different samples, including absorption spectroscopy, photoacoustic spectroscopy, light-induced thermoelastic spectroscopy, and photothermal spectroscopy [13–15]. Among them, near-infrared (NIR) spectroscopy working with the principle of absorption spectroscopy is the most common for the detection of GMOs [11]. NIR spectroscopy coupled with chemometric analyses was found to be effective in discriminating various types of GM and non-GM crops with very high accuracy [11,16,17]. To distinguish transgenic soybean oils from non-transgenic ones, Luna et al. [18] used NIR and support vector machine discriminant analysis (SVM-DA). Later, Garcia-Molina et al. [19] used NIR spectroscopy in combination with partial least square (PLS) analysis to successfully distinguish low gliadin wheat grain from non-transgenic wheat lines with 96% of classification accuracy. It has been shown that using spectroscopic and machine learning algorithms makes it possible to distinguish not only GM and non-GM plants, but also plant species [11,20] and even varieties [21]. However, there is no study that discriminates the GM and non-GM plants with their interspecific hybrids. Therefore, in this study, we used visible near-infrared (Vis-NIR) spectroscopy coupled with different preprocessing and machine learning methods for effective discrimination of *B. juncea*, GM *B. napus*, and their hybrids (*B. juncea* X GM *B. napus*).

2. Results and Discussion

2.1. Spectral Analysis and Preprocessing

The averaged raw spectra of the *B. napus*, GM *B. napus*, *B. juncea*, and F₁ hybrids collected in the green house are depicted in Figure 1A. The original unprocessed raw spectra were ones that had not been altered in any manner. The Savitzky–Golay preprocessed spectra are shown in Figure 1B. Standard normal variate (SNV) (Figure 1C) and normalization (Figure 1D) procedures were used to preprocess the spectra acquired from these plants. The spectra were preprocessed to remove systemic noise and highlight the variations across the samples [17]. The majority of the spectra acquired from four plants followed a similar pattern, despite variances in spectral reflectance. The difference in average reflectance between GM and non-GM *B. napus*, *B. rapa*, and F₁ hybrids is thought to reflect the changes in hundreds of physicochemical constituents in the plant leaves. In general, NIR spectra disclose the information about a material's chemical composition and physical state. This

provides structural data on the chemical functional groups of the elements that constitute the molecular fingerprints of the sample [22,23]. The spectral data were preprocessed to remove systemic noise and emphasize variations across samples. Using a variety of preprocessing methods at the same time will help us achieve a higher level of classification accuracy and provide us with the opportunity to choose the optimal preprocessing method for a specific sample [11]. The average spectra for all the plants, raw and preprocessed, were effectively visualized using three different methods: the Savitzky–Golay smoothing filter (21 points), normalization, and standard normal variate (Figure 1). In general, the normalization is the process of regularizing the data with respect to variations in sample preparation, sample thickness, absorber concentration, etc. Derivatives are mainly used to resolve peak overlap and eliminate constant and linear baseline shifts between samples. SNV is often used on spectra where baseline and path length changes cause differences between otherwise identical spectra [11,24].

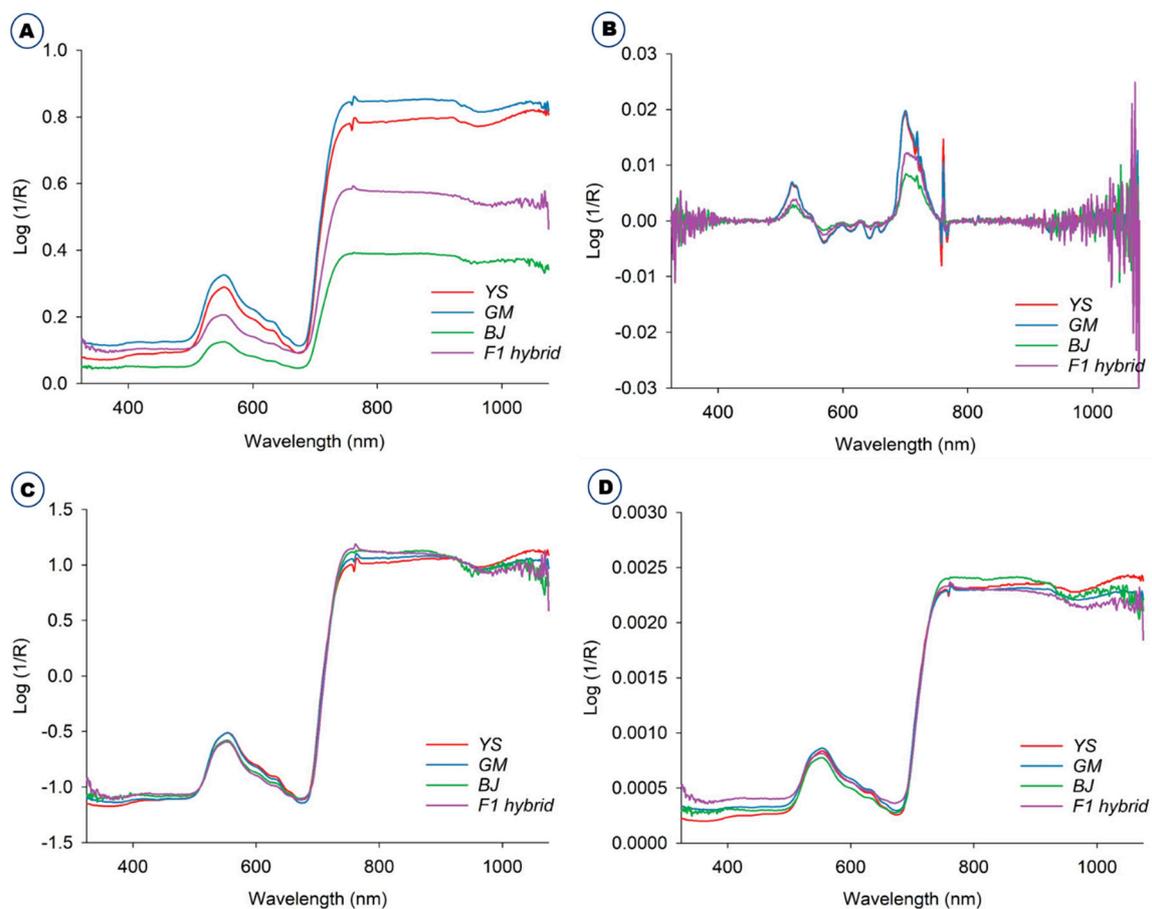


Figure 1. Average raw and preprocessed spectra of *B. napus* (YS), GM *B. napus* (GM), *B. juncea* (BJ), and F₁ hybrids (*B. juncea* X GM *B. napus*). (A) Average raw and preprocessed spectra with different methods, including (B) Savitzky–Golay, (C) standard normal variate, and (D) normalization.

Some typical peaks can be seen in this figure, especially around 500–600 nm, which is the spectral range for chlorophyll [25], and also around 800 nm. However, it is difficult to differentiate these samples solely on the basis of spectral reflectance. Thus, Vis-NIR spectroscopy coupled with various models and machine learning methods such as discriminant analysis and principal component analysis (PCA) was used for effective discrimination [11]. All of the different PCs showed the same slight pattern of separation for the different samples in the PCA paired plot from PC1 to PC6 (Figure 2A), but PC1 vs. PC2 showed the most visual differences, as shown in Figure 2B, so outlier detection was performed using these two PCs before starting preprocessing for the machine learning methods.

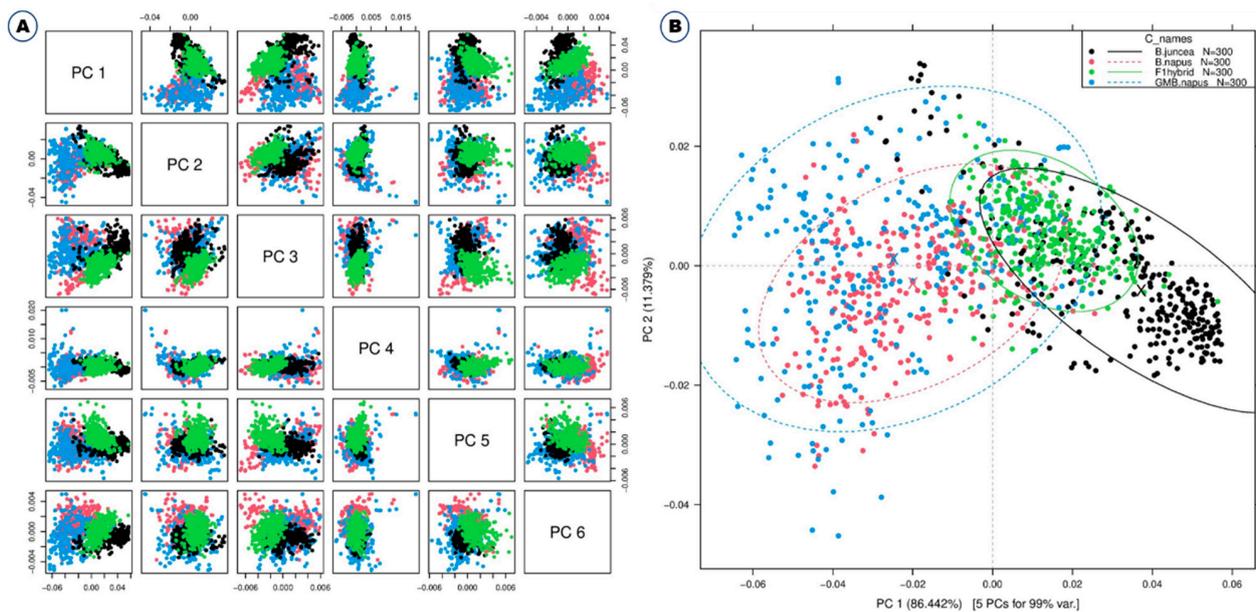


Figure 2. Principal component analyses based on the Vis-NIR spectra of *B. napus*, GM *B. napus*, *B. juncea*, and F_1 hybrids. Raw spectra have been used. (A) Paired blot; (B) axes are first and second principal components.

2.2. Chemometric Analysis for Discrimination of *B. napus*, GM *B. napus*, *B. juncea*, and F_1 Hybrids

The classification accuracy of different chemometric methods combined with various preprocessing methods was calculated in order to determine the most exact way of distinguishing between GM and non-GM *B. napus*, *B. juncea*, and F_1 hybrids. A summary of the classification accuracy for the different methods can be found in Table 1. Both original raw spectra and preprocessed spectra assessed with chemometric analyses resulted in effective discrimination with different classification accuracies. However, preprocessed spectra were found to have comparatively higher classification accuracy than raw spectra in most of the chemometric analyses. The classification accuracies of the different methods generally ranged from 62.6 to 99.4% (Table 1).

From Table 1, the Savitzky–Golay pretreatment proved to be the most efficient preprocessing method for classifying the different plant species with all the tested classification methods except for the support vector machine (SVM) classification technique, where SNV proved to be more effective. Using the Savitzky–Golay, classification accuracies were always higher than when only raw spectra were used. Classification accuracies for the Savitzky–Golay ranged from 80.1 to 99.1%.

Among the different classification methods, support vector machine, linear discriminant analysis, deep learning, and fast large margin were found to have higher classification accuracies in combination with different preprocessing methods (SNV/SVM, 99.4%; Savitzky–Golay/Deep Learning, 99.1%; Savitzky–Golay/SVM, 98.8%) (Table 1). The support vector machine model showed a high accuracy of 97.1% even when using the raw spectrum without preprocessing the data. The support vector machine is especially suitable for high-dimensional data, and the value of each attribute has no limit [26]. When comparing the average value of accuracy according to each model application for each of the four preprocessing methods, Savitzky–Golay showed the highest accuracy, followed by standard normal variate, raw spectrum, and normalization (Table 1). Similar studies have already been performed by various researchers on various crops. Feng et al. [17] used NIR in combination with support vector machine and partial least squares discriminant analysis (PLS–DA) for the effective discrimination of GM and non-GM maize. Similarly, VNIR multispectral imaging and PLS–DA were used for discrimination of GM and non-GM rice using least squares support vector machines (LS-SVM) and PCA backpropagation

neural network (PCA-BPNN) [27]; Fourier transform Infrared (FT-IR) was also used for discrimination of GM and non-GM soybeans with Kth nearest neighbors (KNN) [28]. NIR and support vector machine discriminant analysis (SVM-DA) and PLS-DA were used for discrimination of GM and non-GM soybean [18], and NIR and PLS-DA were used for identification of herbicide-resistant GM soybean seeds [16]. The use of Vis-NIR for discrimination of transgenic tomato using DA and PLS-DA [23] and the use of Vis-NIR for discrimination of RNAi transgenic wheat using NIR and PLS [19] are examples of effective discrimination of GM and non-GM crops using spectroscopy and chemometric analyses. Linear discriminant analysis also yields higher accuracy of 96.5% even when no preprocessing is performed. Figure 3 also shows the linear discriminant analysis plot for discriminating the four different plant varieties. GM *B. napus* slightly overlapped with *B. napus*, but *B. juncea* and F₁ hybrids were completely separated from each other and all the other plant varieties. This suggests that GM *B. napus* and non-GM *B. napus* may share similar biological composition compared to *B. juncea* and F₁ hybrids. Similar studies also reported higher classification accuracy using NIR spectroscopy and linear discriminant analysis to monitor mung bean sprouts [29], classify different melon varieties [30], and detect pea protein powder containing adulterants [31].

Table 1. Classification accuracy of the combinations of preprocessing and model for reflectance spectra from *B. napus*, GM *B. napus*, *B. juncea*, and F₁ hybrids.

S. No.	Model	Preprocessing	Average Accuracy (%)	Run Time (ms)
1	Linear Discriminant Analysis	Raw Spectra	98.5	-
		Normalization (Area)	96.7	-
		Standard Normal Variate	96.5	-
		Savitzky–Golay	98.9	-
2	Deep Learning	Raw Spectra	89.3	5285.4
		Normalization (Area)	93.3	4902.0
		Standard Normal Variate	97.2	3439.5
		Savitzky–Golay	99.1	3287.5
3	Support Vector Machine	Raw Spectra	97.1	6883.3
		Normalization (Area)	87.9	23,700.0
		Standard Normal Variate	99.4	7341.6
		Savitzky–Golay	98.8	7933.3
4	Generalized Linear Model	Raw Spectra	82.9	3691.6
		Normalization (Area)	93.2	3364.5
		Standard Normal Variate	93.6	5212.5
		Savitzky–Golay	91.5	3231.2
5	Decision Tree	Raw Spectra	76.7	3014.5
		Normalization (Area)	79.4	2977.0
		Standard Normal Variate	67.1	2995.8
		Savitzky–Golay	80.1	2785.4
6	Naive Bayes	Raw Spectra	63.5	7727.0
		Normalization (Area)	62.6	3614.5
		Standard Normal Variate	84.8	3691.6
		Savitzky–Golay	92.4	3575.0
7	Fast Large Margin	Raw Spectra	94.6	9466.6
		Normalization (Area)	74.8	10,341.6
		Standard Normal Variate	98.8	9095.8
		Savitzky–Golay	96.9	9552.0
8	Random Forest	Raw Spectra	79.9	3612.5
		Normalization (Area)	85.9	4618.7
		Standard Normal Variate	92.4	4583.3
		Savitzky–Golay	92.6	4062.5

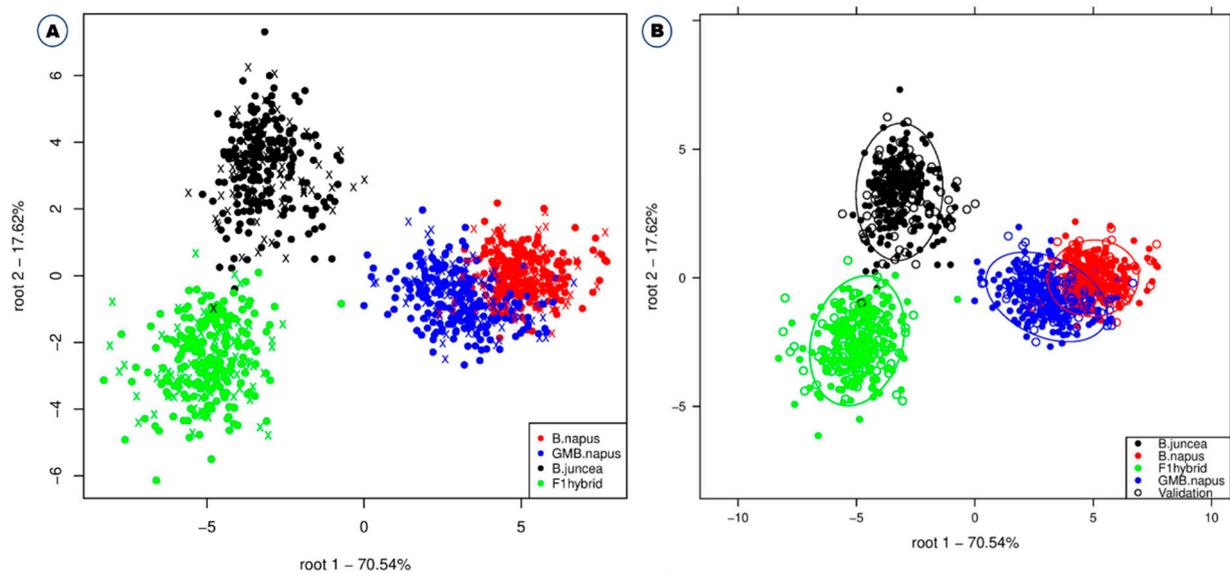


Figure 3. Linear discriminant analysis for the effective discrimination of *B. napus*, GM *B. napus*, *B. juncea*, and F₁ hybrids without confidence circles (A) and with confidence circles (B).

2.3. Significance of Preprocessing and Selection of Optimal Classification Model

The efficiency of preprocessing and machine learning methods used in the study was statistically analyzed (Table 2). After cross-validation, the mean percentage of classification accuracy of each chemometric method combined with various preprocessing methods indicated significant modeling for the discrimination of GM and non-GM *B. napus*, *B. juncea*, and F₁ hybrids (Table 2).

Table 2. Means of percentage of *B. napus*, GM *B. napus*, *B. juncea*, and F₁ hybrids correctly classified using four different preprocessing methods and four different classification models using reflectance spectra.

Model	Species Accuracy (% ± SE)				Significance
	Raw Spectra	Normalization (Area)	Savitzky–Golay	SNV	
Naive Bayes	63.5 ± 3.2 Cb	62.6 ± 5.6 Cb	92.4 ± 3.3 a	84.8 ± 2.1 ABa	***
Generalized Linear Model	82.9 ± 3 AB	93.2 ± 3.2 A	91.5 ± 6	93.6 ± 3.5 A	ns
Fast Large Margin	94.6 ± 3 ABa	74.8 ± 2.5 BCb	96.9 ± 3.1 a	98.8 ± 0.8 Aa	***
Deep Learning	89.3 ± 6.1 AB	93.3 ± 5 A	99.1 ± 0.6	97.2 ± 2 A	ns
Decision Tree	76.7 ± 11.7 BC	79.4 ± 10.2 AB	80.1 ± 15.4	67.1 ± 14.7 B	ns
Random Forest	79.9 ± 4.4 ABC	85.9 ± 3.7 AB	92.4 ± 4.6	92.6 ± 4.6 A	ns
Support Vector Machine	97.1 ± 2.9 Aa	87.9 ± 2.5 ABb	98.8 ± 0.8 a	99.4 ± 0.3 Aa	**
Significance	**	**	ns	*	

ns, not significant; * significant with $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$. Different alphabetical small and capital letters show the significance of the value in the order of column (machine learning) and row (preprocessing), respectively. Same letters are not significantly different at $p \leq 0.05$ based on Tukey's range test.

The statistical analysis with analysis of variance (ANOVA) (Table 3) showed the sum of square and mean sum of square values of various preprocessing and machine learning approaches used with statistical significance at $p \leq 0.005$. However, there was no significance with $p \geq 0.005$ when using a combination of preprocessing and different machine learning methods together (p value of 0.0005).

Table 3. Analysis of variance of percentage of correctly classified *B. napus*, GM *B. napus*, *B. juncea*, and F₁ hybrids from four different preprocessing methods and four different classification models using reflectance spectra.

Source	df	SS	MS	f Value	p Value
Preprocessing (P)	3	2289.98041	763.326803	5.35	0.002
Model (M)	7	6677.677368	1112.946228	7.8	<0001
P × M	21	3664.723846	203.595769	1.43	0.0005
Error	84	11,992.48392	142.76767		
Total	115	24,624.86555			

df: degree of freedom. SS: sum of squares. MS: mean sum of squares.

The confusion matrix depicts the degree of error in the classification of the evaluated plants, indicating that Savitzky–Golay smoothing in combination with support vector machine was the most effective classification approach (Table 4).

Table 4. Confusion matrix from the execution with the best accuracy (Savitzky–Golay and support vector machine).

SNV/SVM	Classified as				Classification Accuracy (%)
	<i>B. napus</i>	GM <i>B. napus</i>	<i>B. juncea</i>	F ₁ Hybrid	
<i>B. napus</i>	86	1	0	0	98.85
GM <i>B. napus</i>	0	84	1	0	98.82
<i>B. juncea</i>	0	0	85	0	100
F ₁ hybrid	0	0	0	86	100
Class recall (%)	100	98.82	98.84	100	-

Savitzky–Golay/Deep Learning	Classified as				Classification Accuracy (%)
	<i>B. napus</i>	GM <i>B. napus</i>	<i>B. juncea</i>	F ₁ Hybrid	
<i>B. napus</i>	84	1	0	0	98.82
GM <i>B. napus</i>	2	84	0	0	97.67
<i>B. juncea</i>	0	0	86	0	100
F ₁ hybrid	0	0	0	86	100
Class recall (%)	97.67	98.82	100	100	-

Savitzky–Golay/SVM	Classified as				Classification Accuracy (%)
	<i>B. napus</i>	GM <i>B. napus</i>	<i>B. juncea</i>	F ₁ Hybrid	
<i>B. napus</i>	84	3	1	0	96.65
GM <i>B. napus</i>	1	83	2	0	98.81
<i>B. juncea</i>	0	0	85	0	100
F ₁ hybrid	0	0	0	86	100
Class recall (%)	98.82	96.51	100	100	-

3. Materials and Methods

3.1. Plant Materials

The seeds used in the study, namely *B. napus* L. ‘Youngsan’ and *B. juncea* var. integrifolia and GM *B. napus* seeds with CAMV 35S-regulated *bar* and early flowering gene (*BrAGL20*), were procured from the National Agrobiodiversity Center, Jeonju, Republic of Korea. For the hybrid preparation, artificial hand pollination was performed with *B. juncea* and GM *B. napus*, and the seeds of F₁ hybrids (*B. juncea* × GM *B. napus*) were used for further research. The hybrids were confirmed through the survival assay after 0.3% Basta treatment; the phenotype of the hybrids; and polymerase chain reaction with 35S ribosomal DNA, *BrAGL20* gene partial region, *bar* gene, and chloroplast marker. All of the seeds were grown in soil pots (Figure 4) and kept in a controlled environment. This research was carried out in the greenhouse of the National Institute of Agricultural Sciences, Jeonju, Republic of Korea, during May–July 2019.

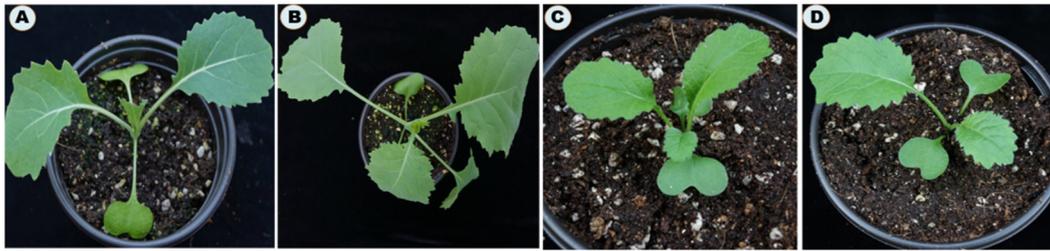


Figure 4. Representative figures for the plants selected for the spectral analysis. (A) *B. napus*; (B) GM *B. napus*; (C) *B. juncea*; (D) F₁ hybrids.

3.2. Spectral Data Collection

A handheld integrated portable spectrum analyzer (FieldSpec HandHeld 2, ASD Inc., Longmont, CO, USA) was used to collect Vis-NIR diffuse reflectance spectra in the range of 325–1075 nm with a stepping of 1.5 nm in reflectance mode (log/R). The spectra were collected on the adaxial surface of the fully expanded leaves, which can easily capture light. Three spectra were obtained from various parts of the leaf blade of 100 plants in each group. A total of 300 ($3 \times 100 = 300$) spectra were collected from each group and used for further analysis. To avoid unnecessary noise, the optical window of the Vis-NIR device was placed in direct contact with the leaf's surface throughout each spectrum capture, ensuring that the sensor window was completely covered [32,33].

3.3. Preprocessing and Machine Learning Methods

Due to system parameters and environmental noise, background signals appeared in the raw spectra of samples. Different preprocessing methods, such as raw spectra assessment, normalization (area), standard normal variate (SNV), and derivatives (Savitzky–Golay (first differentiation)) were used, which can reduce the spectral noise and improve the accuracy of modeling approaches. The computations on preprocessing were performed with Unscrambler X software, version 10.5.1 (CAMO ASA, Oslo, Norway).

For the effective visualization and discrimination of spectral data, several machine learning methods were used and compared. The modeling was performed with the software package RapidMiner studios Version 9.0.002 (Rapidminer, Inc., Boston, MA, USA). In the study, eight classification methods were used to find the best modeling approach with the highest classification accuracy, namely deep learning, decision tree, support vector machine, random forest, generalized linear model, fast large margin, naive Bayes, and linear discriminant analysis. Linear discriminant analysis was performed in R-studio using the Aquap2 package developed by Kovacs and Pollner [34]. For each of the algorithms, the inputs were provided as the data points of the spectra and the classes were the identification labels of *B. napus*, GM *B. napus*, *B. juncea*, and F₁ hybrids (*B. juncea* X GM *B. napus*). Cross-validation was performed to assess the robustness of the models in predicting the different sample types. For this, the data were divided into a training set and a validation set. The training set was made up of two-thirds of the data; thus, the spectra from the first and second replicates of each sample were included, while the validation set was made up of spectra from the third replicate. The data splitting was done three times, such that each sample was used at least once in the calibration and validation set. The classification results are displayed as score plots or confusion matrix, which illustrates the percentages of classification accuracy. One-way analysis of variance (ANOVA) was used to compare means for determining the influence of (1) the scatter correction method, (2) the eight machine learning methods, and (3) the interaction of preprocessing and machine learning methods. As a mean comparison method, Tukey's range test was used at a significance level of $p \leq 0.05$.

4. Conclusions

In conclusion, Vis-NIR spectroscopy in combination with machine learning methods could effectively discriminate between GM and non-GM *B. napus*, *B. juncea*, and the F₁ hybrids (*B. juncea* X GM *B. napus*). The utilization of Vis-NIR spectroscopy and chemometric analyses for the discrimination of GM and non-GM crops is quick and accurate. It can also deliver information for monitoring and safety management of agro-food market products in which GMOs are introduced. Among the different combinations of preprocessing and machine learning methods, the combination of standard normal variate and support vector machine was found to be the most effective method, with 99.6% classification accuracy, but Savitzky–Golay smoothing also yields good classification accuracy when other classification methods are used. Thus, it is proposed that this nondestructive method be employed in the field for the rapid detection and management of unintended releases of GM *Brassicaceae* crops into the environment. It is suggested to create a database with broad-spectrum results on GM and non-GM *Brassicaceae* crops for the effective utilization of the technology in the field.

Author Contributions: Conceptualization: S.-I.S., Y.-J.O. and C.-S.N.; methodology: S.-I.S., S.P., Y.-J.O., Y.-H.L., E.-K.S., H.-J.K., T.-H.R. and Y.-S.C.; formal analysis: S.-I.S., S.P., Y.-J.O. and W.-S.C.; data curation: S.-I.S., Y.-J.O. and C.-S.N.; writing—original draft preparation: S.-I.S., S.P. and J.-L.Z.Z.; visualization: S.P. and J.-L.Z.Z.; project administration: S.-I.S.; funding acquisition: S.-I.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out with the support of “Research Program for Agricultural Science & Technology Development and 2022 Post-doctoral Fellowship Program (Project No. PJ014943012022)”, National Institute of Agricultural Sciences, Rural Development Administration, Korea.

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. Yang, J.; Liu, D.; Wang, X.; Ji, C.; Cheng, F.; Liu, B.; Hu, Z.; Chen, S.; Pental, D.; Ju, Y.; et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* **2016**, *48*, 1225–1232. [[CrossRef](#)] [[PubMed](#)]
2. Lim, Y.; Yook, M.J.; Zhang, C.J.; Nah, G.; Park, S.; Kim, D.S. Dormancy associated weedy risk of the F₁ hybrid resulted from gene flow from oilseed rape to mustard. *Weed Turfgrass Sci.* **2015**, *4*, 35–43. [[CrossRef](#)]
3. Zhang, C.J.; Yook, M.J.; Park, H.R.; Lim, S.H.; Kim, J.W.; Song, J.S.; Nah, G.; Song, H.R.; Jo, B.H.; Roh, K.H.; et al. Evaluation of maximum potential gene flow from herbicide resistant *Brassica napus* to its male sterile relatives under open and wind pollination conditions. *Sci. Total Environ.* **2018**, *634*, 821–830. [[CrossRef](#)] [[PubMed](#)]
4. Devos, Y.; De Schrijver, A.; Reheul, D. Quantifying the introgressive hybridisation propensity between transgenic oilseed rape and its wild/weedy relatives. *Environ. Monit. Assess.* **2009**, *149*, 303–322. [[CrossRef](#)] [[PubMed](#)]
5. Scheffler, J.A.; Dale, P.J. Opportunities for gene transfer from transgenic oilseed rape (*Brassica napus*) to related species. *Transgenic Res.* **1994**, *3*, 263–278. [[CrossRef](#)]
6. Song, X.L.; Huangfu, C.H.; Qiang, S. Gene flow from transgenic glufosinate-or glyphosate-tolerant oilseed rape to wild rape. *Chin. J. Plant Ecol.* **2007**, *31*, 729–737.
7. Cao, D.; Stewart Jr, C.N.; Zheng, M.; Guan, Z.; Tang, Z.X.; Wei, W.; Ma, K.P. Stable *Bacillus thuringiensis* transgene introgression from *Brassica napus* to wild mustard *B. juncea*. *Plant Sci.* **2014**, *227*, 45–50. [[CrossRef](#)]
8. Liu, Y.; Neal Stewart Jr, C.; Li, J.; Wei, W. One species to another: Sympatric Bt transgene gene flow from *Brassica napus* alters the reproductive strategy of wild relative *Brassica juncea* under herbivore treatment. *Ann. Bot.* **2018**, *122*, 617–625. [[CrossRef](#)]
9. Tang, T.; Chen, G.; Bu, C.; Liu, F.; Liu, L.; Zhao, X. Transgene introgression from *Brassica napus* to different varieties of *Brassica juncea*. *Plant Breed.* **2018**, *137*, 171–180. [[CrossRef](#)]
10. Di, K.; Stewart Jr, C.N.; Wei, W.; Shen, B.C.; Tang, Z.X.; Ma, K.P. Fitness and maternal effects in hybrids formed between transgenic oilseed rape (*Brassica napus* L.) and wild brown mustard [*B. juncea* (L.) Czern et Coss.] in the field. *Pest Manag. Sci.* **2009**, *65*, 753–760. [[CrossRef](#)]

11. Sohn, S.-I.; Pandian, S.; Oh, Y.-J.; Zaukuu, J.-L.Z.; Kang, H.-J.; Ryu, T.-H.; Cho, W.-S.; Cho, Y.-S.; Shin, E.-K.; Cho, B.-K. An Overview of Near Infrared Spectroscopy and Its Applications in the Detection of Genetically Modified Organisms. *Int. J. Mol. Sci.* **2021**, *22*, 9940. [[CrossRef](#)] [[PubMed](#)]
12. Ahmed, F.E. Detection of genetically modified organisms in foods. *Trends Biotechnol.* **2002**, *20*, 215–223. [[CrossRef](#)]
13. Ma, Y.; Lewicki, R.; Razeghi, M.; Tittel, F.K. QEPAS based ppb-level detection of CO and N₂O using a high power CW DFB-QCL. *Opt. Express* **2013**, *21*, 1008–1019. [[CrossRef](#)]
14. Ma, Y.; He, Y.; Tong, Y.; Yu, X.; Tittel, F.K. Quartz-tuning-fork enhanced photothermal spectroscopy for ultra-high sensitive trace gas detection. *Opt. Express* **2018**, *26*, 32103–32110. [[CrossRef](#)] [[PubMed](#)]
15. Qiao, S.; Ma, Y.; He, Y.; Patimisco, P.; Sampaolo, A.; Spagnolo, V. Ppt level carbon monoxide detection based on light-induced thermoelastic spectroscopy exploring custom quartz tuning forks and a mid-infrared QCL. *Opt. Express* **2021**, *9*, 25100–25108. [[CrossRef](#)] [[PubMed](#)]
16. Lee, J.H.; Choung, M.G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. *Food Chem.* **2011**, *126*, 368–373. [[CrossRef](#)]
17. Feng, X.; Peng, C.; Chen, Y.; Liu, X.; Feng, X.; He, Y. Discrimination of CRISPR/Cas9-induced mutants of rice seeds using near-infrared hyperspectral imaging. *Sci. Rep.* **2017**, *7*, 15934. [[CrossRef](#)]
18. Luna, A.S.; da Silva, A.P.; Pinho, J.S.; Ferré, J.; Boqué, R. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2013**, *100*, 115–119. [[CrossRef](#)]
19. Garcia-Molina, M.D.; Garcia-Olmo, J.; Barro, F. Effective identification of low-gliadin wheat lines by near infrared spectroscopy (NIRS): Implications for the development and analysis of foodstuffs suitable for celiac patients. *PLoS ONE* **2016**, *11*, e0152292. [[CrossRef](#)]
20. Li, T.; Su, C. Authenticity identification and classification of *Rhodiola* species in traditional Tibetan medicine based on Fourier transform near-infrared spectroscopy and chemometrics analysis. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018**, *204*, 131–140. [[CrossRef](#)]
21. Carvalho, L.C.; Morais, C.L.; Lima, K.M.; Leite, G.W.; Oliveira, G.S.; Casagrande, I.P.; Neto, J.P.S.; Teixeira, G.H. Using intact nuts and near infrared spectroscopy to classify Macadamia cultivars. *Food Anal. Met.* **2018**, *11*, 1857–1866. [[CrossRef](#)]
22. Cordella, C.; Moussa, I.; Martel, A.C.; Sbirrazzuoli, N.; Lizzani-Cuvelier, L. Recent developments in food characterization and adulteration detection: Technique-oriented perspectives. *J. Agric. Food Chem.* **2002**, *50*, 1751–1764. [[CrossRef](#)] [[PubMed](#)]
23. Xie, L.; Ying, Y.; Ying, T.; Yu, H.; Fu, X. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. *Anal. Chim. Acta* **2007**, *584*, 379–384. [[CrossRef](#)] [[PubMed](#)]
24. Rinnan, Å.; Van Den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [[CrossRef](#)]
25. Smith, H.L.; McAusland, L.; Murchie, E.H. Don't ignore the green light: Exploring diverse roles in plant processes. *J. Exp. Bot.* **2017**, *68*, 2099–2110. [[CrossRef](#)]
26. Gaye, B.; Zhang, D.; Wulamu, A. Improvement of support vector machine algorithm in big data background. *Mat. Prob. Eng.* **2021**, *2021*, 5594899. [[CrossRef](#)]
27. Liu, C.; Liu, W.; Lu, X.; Chen, W.; Yang, J.; Zheng, L. Nondestructive determination of transgenic *Bacillus thuringiensis* rice seeds (*Oryza sativa* L.) using multispectral imaging and chemometric methods. *Food Chem.* **2014**, *153*, 87–93. [[CrossRef](#)]
28. Alcantara, G.B.; Barison, A.; Santos, M.D.S.; Santos, L.P.; de Toledo, J.F.; Ferreira, A.G. Assessment of genetically modified soybean crops and different cultivars by Fourier transform infrared spectroscopy and chemometric analysis. *Orbital Electr. J. Chem.* **2010**, *2*, 41–52.
29. Nugraha, D.T.; Zaukuu, J.L.Z.; Bósquez, J.P.A.; Bodor, Z.; Vitalis, F.; Kovacs, Z. Near-infrared spectroscopy and aquaphotomics for monitoring mung bean (*Vigna radiata*) sprout growth and validation of ascorbic acid content. *Sensors* **2021**, *21*, 611. [[CrossRef](#)]
30. Zaukuu, J.L.Z.; Gillay, Z.; Kovacs, Z. Standardized extraction techniques for meat analysis with the electronic tongue: A case study of poultry and red meat adulteration. *Sensors* **2021**, *21*, 481. [[CrossRef](#)]
31. Zaukuu, J.L.Z.; Aouadi, B.; Lukács, M.; Bodor, Z.; Vitalis, F.; Gillay, B.; Gillay, Z.; Friedrich, L.; Kovacs, Z. Detecting low concentrations of nitrogen-based adulterants in whey protein powder using benchtop and handheld NIR spectrometers and the feasibility of scanning through plastic bag. *Molecules* **2020**, *25*, 2522. [[CrossRef](#)] [[PubMed](#)]
32. Sohn, S.I.; Oh, Y.J.; Pandian, S.; Lee, Y.H.; Zaukuu, J.L.Z.; Kang, H.J.; Ryu, T.H.; Cho, W.S.; Cho, Y.S.; Shin, E.K. Identification of *Amaranthus* Species using Visible-Near-Infrared (Vis-NIR) spectroscopy and machine learning methods. *Remote Sens.* **2021**, *13*, 4149. [[CrossRef](#)]
33. Sohn, S.I.; Pandian, S.; Zaukuu, J.L.Z.; Oh, Y.J.; Park, S.-Y.; Na, C.S.; Shin, E.K.; Kang, H.J.; Ryu, T.H.; Cho, W.S.; et al. Discrimination of transgenic canola (*Brassica napus* L.) and their hybrids with *B. rapa* using Vis-NIR spectroscopy and machine learning methods. *Int. J. Mol. Sci.* **2022**, *23*, 220. [[CrossRef](#)] [[PubMed](#)]
34. Pollner, B.; Kovacs, Z. Dedicated Aquaphotomics-Software R-Package „aquap2“ General Introduction and Workshop. Aquaphotomics: Understanding Water in the Biological World. In Proceedings of the 5th Kobe University Brussels European Centre Symposium Innovation, Environment and Globalization—Latest EU-Japan Research Collaboration, Bruxelles, Belgium, 14 October 2014.