

## Article

# Rootstock's and Cover-Crops' Influence on Grape: A NIR-Based ANN Classification Model

Teodora Basile \*, Antonio Maria Amendolagine and Luigi Tarricone 

Consiglio per la Ricerca in Agricoltura e L'analisi dell'Economia Agraria—Centro di Ricerca Viticoltura ed Enologia (CREA—VE), 70010 Turi, BA, Italy

\* Correspondence: teodora.basile@crea.gov.it; Tel.: +39-080-8915711; Fax: +39-080-4512925

**Abstract:** In this study, a multivariate analysis combined with near-infrared (NIR) spectroscopy was employed to classify intact grape berries based on the rootstock x cover crops combination. NIR spectra were collected in diffuse reflection mode using a TANGO FT-NIR spectrometer (Bruker, Germany) with  $8\text{ cm}^{-1}$  resolution and 64 scans in the wave number range of  $4000\text{--}10,000\text{ cm}^{-1}$ . The chemometric analyses were performed with the statistical software R version 4.2.0 (2022-04-22). Elimination of uninformative variables was accomplished with a PCA and a genetic algorithm (GA). The discrimination performance of a linear discriminant analysis (LDA) model was not enhanced with either a PCA- or a GA-based selection. A multiclass classification model was built with an artificial neural network (ANN). The best fit multiclass classification model on test data was obtained with the GA-ANN model that gave a classification accuracy of close to 80% for samples belonging to the four classes. These results demonstrate that NIR spectroscopy could be used as a rapid method for the classification of berries based on their rootstock x cover-crops combination.

**Keywords:** NIR; genetic algorithm; LDA; PCA; ANN; grape



**Citation:** Basile, T.; Amendolagine, A.M.; Tarricone, L. Rootstock's and Cover-Crops' Influence on Grape: A NIR-Based ANN Classification Model. *Agriculture* **2023**, *13*, 5. <https://doi.org/10.3390/agriculture13010005>

Academic Editors: Martin Weih and Vitale Nuzzo

Received: 11 November 2022

Revised: 4 December 2022

Accepted: 15 December 2022

Published: 20 December 2022



**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/>).

## 1. Introduction

Near-infrared (NIR) spectroscopy is successfully applied for the authentication and traceability of food and beverages [1–3]. In this study, we tested the ability of this technique to discriminate among *Vitis vinifera* berries of the same variety, grown in the same vineyard and subjected to the same environmental conditions, which differ only for rootstock and cover crop combinations. Since the arrival of phylloxera in Europe, all *V. vinifera* varieties have been grafted onto resistant rootstocks (*Vitis* spp.). The rootstocks provide not only resistance but also strongly influence scion growth, development, productivity, and phenology in a specific terroir [4]. Rootstock has been shown to affect berry quality traits such as weight at harvest and the secondary metabolism leading to the production of sugars and polyphenolic compounds [5,6]. Among sustainable soil management practices, cover crops are widely used in vineyards. Several studies have been carried out to evaluate the influence of different cover crops on grapevines' vegetative growth, yield, berry, and wine quality [7,8]. Weeds and cover crops compete with grapevines for water and nutrients during the growing season, resulting in a decrease in vegetative growth (vigor) and yield in the short period or after several years, especially for vines growing in semi-arid conditions [9]. This competition is beneficial for the producer since it allows for a reduction of excessive grape vigor and crop yield while improving grape quality [10]. The impact of this technique on the yield and quality parameters of grapes strongly depends on the selection and management of the plants used as cover crops [11]. The choice of rootstock and the cover crop species is crucial since it influences the adaptability of grapevines to the increasing environmental challenges, such as climate change [12]. This study aimed to investigate the extent to which two different rootstocks (*Vitis berlandieri* × *Vitis riparia* 140 Ruggeri and 1103 Paulsen) in two grassing conditions (legumes or grass mixtures) modified the metabolic response of the Autumn Pearl table grape variety. The samples

were separated into four classes based on the rootstock (140Ruggeri or 1103Paulsen) x cover crop (only grass or grass–legume mixture) combination. In this work, we labeled “A” the 140Ruggeri x only grass, “B” 1103Paulsen x only grass, “C” 140Ruggeri x grass plus legumes, and “D” 1103Paulsen x grass plus legumes samples. In this article, NIR spectroscopy was employed in combination with chemometric techniques to classify the samples into four classes defined by the rootstock x cover crop combination. In this work, both supervised and unsupervised recognition procedures were applied. Linear discriminant analysis (LDA) and artificial neural networks (ANNs) were used to classify the samples. To achieve significant differentiations among the berries, a selection of the most significant wavelengths was made with a PCA and a genetic algorithm (GA). This allowed the elimination of unnecessary information and amplified relevant variations in the spectra. The use of intact berries poses some analytical issues due to the lack of homogeneity of this type of sample. That results in high coefficients of variation in the NIR spectra when samples are scanned at different points relative to the source [13]. Despite the analytical difficulties, there is a great advantage in having an intact sample after the NIR analysis. A NIR-based analysis of intact berries can be used for real-time monitoring of the berries in vivo on the vine. This could be particularly useful in the context of precision agriculture, which requires real-time data acquisition to make the best decision for digital farming management [14].

## 2. Materials and Methods

### 2.1. Grape Samples

The research was carried out in an organic vineyard located in Southern Italy (Apulia region, Ginosa, Taranto, 40°27′41.0″ N; 16°50′27.8″ E) in the 2021 season. The vineyard consisted of two blocks of the same variety with a surface of 1.2 hectares each. The two-year-old grapevines were spaced at 3.0 × 2.20 m apart (1515 vines ha<sup>−1</sup>). Vines were trained to a “tendone system” (Apulia type) and drip irrigated. Each vineyard block was divided into four blocks of 12 rows of 36 vines, alternately sown with the two cover crop mixtures according to randomized blocks. The table grape variety analyzed was Autumn Pearl, a new medium-ripening (from September to October) red seedless cultivar characterized by high productivity, large, round, and crunchy berries with a sweet and slightly fruity taste. The two different rootstocks were 140 Ruggeri and 1130 Paulsen. The two commercially available cover crops were called San Martino (*Festuca arundinacea* cv Sitka 50%, *Festuca rubra* cv Maxima 1 40%, and *Poa pratense* cv Sunbeam 10%) and Elena, a mixture of grass and legume cover crops (*Lolium perenniale* cv Mathilde 50%, *Festuca rubra* cv Maxima 1 47%, and *Trifolium repens* cv Rivendell 3%). Figure 1 shows the fresh samples upon arrival in the laboratory and the berries prepared for the analyses. Each berry was shortly washed with distilled water and gently tapped with paper before the NIR measurement. The 360 berry samples were measured over several days. Therefore, to ensure the reproducibility of the results, each day, the spectra of a few randomly chosen berries among those already analyzed were collected again. Maturity parameters were measured on the berries. Total soluble solids content (TSS, °Brix) and total acidity (TA, g/L as tartaric acid) were measured in triplicate using an Atago PR1 digital refractometer (Atago Co., Tokyo, Japan). Total polyphenolic content (TP, mg/Kg of grape) was measured following the Folin–Ciocalteu method in triplicate for each berry [15].



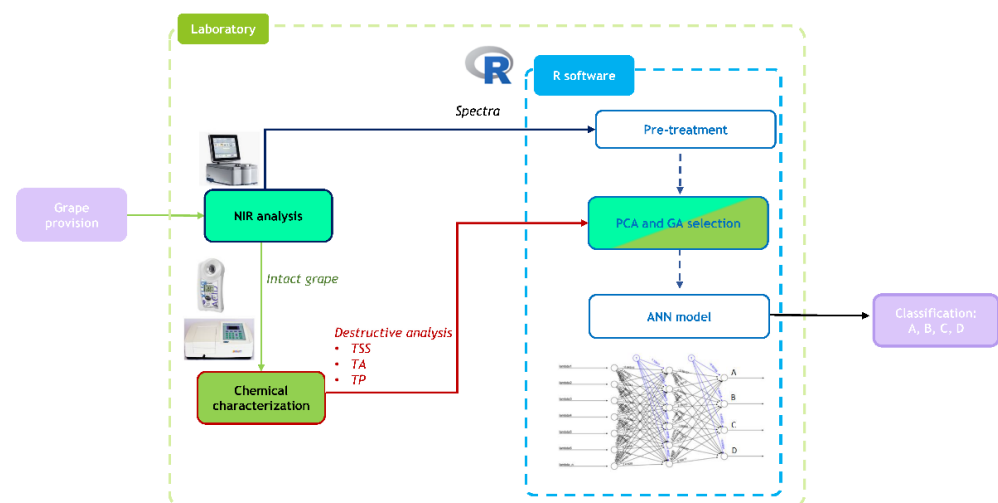
**Figure 1.** Grapes belonging to A and C (a), D and B classes (b) upon arrival in the laboratory; (c) slot of samples prepared for NIR analysis (from bottom to top: A, B, C, and D).

## 2.2. NIR Spectroscopy

NIR absorption measurements were carried on in diffuse reflection mode using a TANGO FT-NIR spectrometer (Bruker, Germany). NIR spectra were collected by data acquisition software OPUS/QUANT software version 2.0 (Bruker Optik GmbH, Ettlingen, Germany) between  $12,000\text{--}4000\text{ cm}^{-1}$  ( $833\text{--}2500\text{ nm}$ ), with  $8\text{ cm}^{-1}$  resolution and 64 scans. The spectrum of each sample was the average of three successive scans on three different berry faces. A background spectrum was automatically recorded, before each sample. Both temperature and relative humidity of the room were kept constant with an air conditioning system.

## 2.3. Statistical Analysis

The statistical procedures described in detail in the following paragraphs, including pre-treatments of the original spectra, GA algorithm, PCA, calibration, cross-validation, and external validation of classification models built with LDA and multilayer perceptron artificial ANN, were performed using R Statistical Software (v4.2.0; R Core Team 2022) [16]. The R packages used are listed in alphabetical order: caret [17], funModeling [18], GA [19], genalg [20], ggplot2 [21], keras [22], MASS [23], and mdatools [24]. A process flow block diagram is reported in Figure 2.



**Figure 2.** Process flow block diagram.

### 3. Results

#### 3.1. Maturity Parameters and Data Preparation

The samples were labeled as “A” 140Ruggeri x grass, “B” 1103Paulsen x grass, “C” 140Ruggeri x legumes, and “D” 1103Paulsen x legumes. In terms of plant behavior, higher rates of photosynthetic assimilation, transpiration, and increased water use efficiency (WUE) were measured for grapevines growing on grass cover crops (A and B) while the plant vigor, expressed by the weight of the pruned wood of the year, did not show any difference [25]. WUE is especially important in semiarid regions where water availability is limited. Grape grows and matures during the driest months, making irrigation scheduling and timing critical. This situation is worsened by climate change predictions that indicate increases in temperatures and more frequent episodes of climatic anomalies, such as droughts and heat waves [26]. In our case, a grass cover crop was able to increase WUE in the vineyard without increasing the vigor of the plant, even if both the 1103P and 140R rootstocks usually confer a high vigor to the scion. These results are consistent with previous findings, which show a containment of vigor due to plant-cover crops competition toward nutrients [10]. Increasing the WUE for a crop such as table grape, which usually requires frequent irrigation, especially during the summer months, is important to ensure the environmental sustainability of food production. Table 1 shows the basic parameters measured on the berries after NIR analysis. The polyphenolic content is significantly different between samples grown on different rootstock. The TSS values were not significantly different among the classes while the acidity content shows a small difference only between two groups which are characterized by different rootstock but the same grass cover crop. The evaluation of grape composition indicates a significantly higher amount of polyphenolic compounds in 140R over 1103P, despite the cover crop type. In terms of maturity parameters, no difference was found for sugars while a lower acidity was found for 140R compared to 1130, but only when vines were grown on grass. Previous findings show that grasses promoted higher content of sugars and phenols in berries [9]. Instead, we did not find significant differences in sugar content, and the differences in polyphenolics were linked to the rootstock more than to the cover crop.

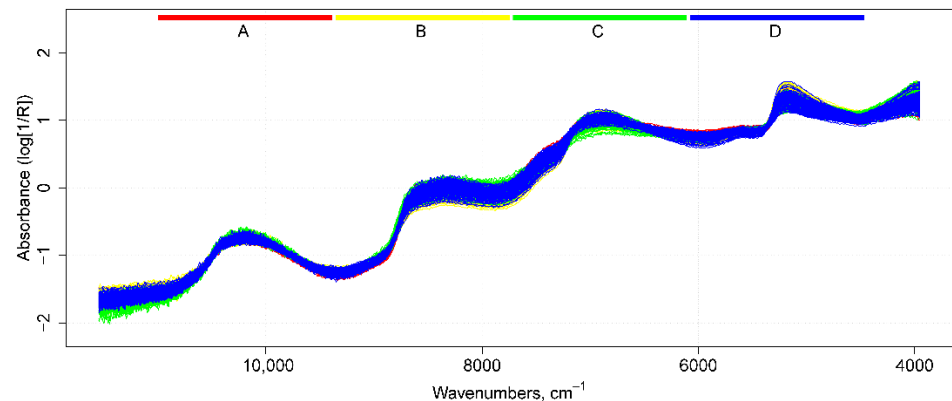
**Table 1.** Basic maturity parameters.

| Class          | Berry Weight (g) | TP (mg/Kg, as Catechin) | TSS (°Brix) | TA (g/L, as Tartaric Acid) |
|----------------|------------------|-------------------------|-------------|----------------------------|
| A              | 6.59 ± 1.70 bc   | 399.0 ± 91.6 a          | 18.6 ± 4.0  | 60.1 ± 27.7 b              |
| B              | 8.46 ± 1.86 a    | 310.0 ± 110.0 b         | 20.1 ± 1.9  | 81.0 ± 16.9 a              |
| C              | 7.31 ± 1.77 b    | 413.0 ± 95.5 a          | 19.0 ± 1.0  | 60.1 ± 7.5 b               |
| D              | 6.09 ± 1.24 c    | 305.0 ± 60.2 b          | 18.2 ± 1.7  | 59.5 ± 6.0 b               |
| <i>p</i> value | ***              | ***                     | n.s.        | *                          |

All values are reported as mean ± standard deviation. For each variety, values in the same column bearing different letters are significantly different. n.s. = not significant; \*  $p < 0.05$ ; \*\*\*  $p < 0.001$ . Data have been analyzed with an ANOVA test followed by a Duncan post-hoc test.

The spectra are dominated by the intense absorption bands of water from various overtones and combinations of water’s three fundamental vibrational transitions: symmetric stretching, bending, and asymmetric stretching. The water absorption signals are found around 1950 nm ( $5128\text{ cm}^{-1}$ ), 1450 nm ( $6896\text{ cm}^{-1}$ ), and two minor centered near 1200 nm ( $8333\text{ cm}^{-1}$ ) and 970 nm ( $10,300\text{ cm}^{-1}$ ) [27]. The preprocessing step is of great importance to properly remove noise and perform background and baseline correction of NIR spectra. Among the methods conventionally applied to spectroscopic data, we applied multiplicative scatter correction (MSC), standard normal variate (SNV), smoothing (e.g., Savitsky–Goley), derivatives, and a combination of them [28]. A PCA was performed on each pre-treated spectral dataset. Unfortunately, none of the PCAs showed a clear separation of the samples by classes. Therefore, we could not base our pretreatment choice on the higher amount of cumulative variance explained by the first two PCs and the better ability to cluster the samples based on the classes. Since the selection of the optimal preprocessing

method follows a trial-and-error approach and depends to a large extent on the nature of the data, the search for the optimal preprocesses for our data started from previous findings [29]. The SNV pretreatment was selected, and the spectra are shown in Figure 3.



**Figure 3.** SNV pre-treated spectra of the four classes.

For a proportional distribution of the four classes (81 A, 87 B, 90 C, and 102 D samples) in both test and training sets, a class-balanced random 80/20 split of the dataset was performed. The validation of the training set was performed with a 5-fold cross-validation repeated 10 times. The test set was employed for the external validation to ensure the robustness of the model. The selection of characteristic wavelengths (holding sample-specific or component-specific information) can improve the performance of prediction or classification algorithms since it eliminates the uninformative variables and improves the performance of the models. Several methods have been developed and can be found in the literature [30]. The two selection procedures performed for the selection of the independent variables on our data were a PCA and a GA.

### 3.2. GA-Based Feature Selection Done

A GA procedure is based on Darwin's theory of biological evolution. In a natural selection, starting from a random population of individuals the ones who are "most fit" for the environment have a greater chance to survive and reproduce to generate a better offspring [31]. In a GA the individuals are called chromosomes. A chromosome is a bit vector of binary values where every gene (bit) represents one of the independent variables (i.e., the NIR wavelengths). The binary values represent the inclusion (1) or exclusion (0) of that gene (variable). The set of genes of each chromosome represents a possible solution. The process starts with a random population evaluated based on a "fitness" function. The fittest chromosomes are selected for reproduction and generate offspring by crossovers of two parents' chromosomes and mutations of individual chromosomes. The population has a fixed size, therefore, the least fit individuals must "die" to be replaced by the new "fittest" offspring. The algorithm is performed iteratively over several generations. It terminates if the population has converged (does not produce offspring which are significantly different from the previous generation) [32].

### 3.3. Development of Supervised Classification Models

Highly correlated variables (overfitting issues) or complex or non-linear class boundaries (underfitting issues) could affect LDA discrimination performance [33]. Correlation-reducing methods such as PCA are employed to improve the classification process [34]. After the PCA, only wave numbers with loadings  $>0.03$  on the first seven PCs (cumulative variance explained 89.71%) were retained for a total amount of 2167 selected predictors. However, for our dataset the removal of correlated variables only affected the accuracy of the outcome to a small extent (LDA models in Table 2).

**Table 2.** Accuracy and kappa values for the LDA and ANN classification models.

| Algorithm | Dataset | Accuracy | K    |
|-----------|---------|----------|------|
| LDAwhole  | Train   | 0.64     | 0.52 |
|           | Test    | 0.66     | 0.55 |
| PCA-LDA   | Train   | 0.61     | 0.48 |
|           | Test    | 0.61     | 0.47 |
| GA-LDA    | Train   | 0.61     | 0.48 |
|           | Test    | 0.58     | 0.44 |
| ANN       |         | 0.76     | 0.68 |

The genetic algorithms (GAs) are efficient optimization techniques for interrogating a large search space in which many combinations of wavelengths are possible. GAs have already been used in variable selection problems and seem to be a solution to the multivariate selection of variables [35]. In our case, each chromosome contained 1899 genes. The binary values of the genes indicated if the corresponding wave number was included in the classification (value 1) or not (value 0). The GA was performed on an initial randomly set population, with a 0.03 probability mutation rate, a number of best individuals to pass to the next iteration (elitism) of 3, and a uniform crossover over 100 generations. A maximum number of iterations without improvement (stopping criteria) was set at seven and a maximum number of runs (generations) was set at 50. For the fitness calculation, we applied a custom fitness function for multi-class classification returning Cohen’s kappa statistic as the fitness function value. We chose kappa in place of accuracy to measure the classifier performance since this coefficient evaluates the difference between the accuracy and the null error rate, thus accounting for the possibility of a correct classification occurring by chance [36]. The script employed to select the best variables for classification using genetic algorithms based on the “GA” library with a custom fitness function was based on the structure reported in the following public GitHub repository: <https://github.com/pablo14/genetic-algorithm-feature-selection> (accessed on 1 March 2022) and adapted for multi-class classification. The “best” subset of variables based on genetic algorithm selection (958 predictors) was used as input to build both an LDA and an ANN model. The discrimination performance of the LDA was not enhanced with a GA-based selection (GA-LDA model in Table 2).

### 3.4. ANN Structure

ANNs are a powerful method for the extraction of quantitative information from large spectroscopic databases. Their pattern recognition abilities are important for datasets which show inherent non-linearity due to complex biological, environmental, and instrumental variations. However, the ANN network implementation, method setup, training, and estimation of parameters are relatively complex compared to linear regression methods [29,37]. A dimension reduction of the dataset performed with a variable selection method is necessary to increase the predictive efficiency of the final ANN model [38]. Min–max data normalization was applied prior to the training of the neural network since it generally increases the learning rate and leads to faster convergence. The input variables were scaled in the interval. The structure of our feed-forward fully connected neural network consisted of three layers. We found that increasing the number of hidden layers resulted in a worsening of the prediction. The number of neurons in each layer was: number of predictors for the input, half of the input data for the hidden, and four neurons for the output layer since we were performing a multiclass classification analysis. In summary, the ANN configuration was input:hidden:output,  $n:(n + 1)/2:4$ , where  $n$  is the numeric vector representing the selected wave numbers for each NIR spectrum. Two activation functions were used: the rectified linear activation (ReLU) function was used in the input and hidden layers, while for the output layer we used the Softmax function, which is commonly used for multiclass classification problems. Following the choice of the ReLU activation function in the input layer of our neural network, we used a He normal weight initialization (samples from a

truncated normal distribution centered on 0 with  $\text{stddev} = \sqrt{2/\text{fan\_in}}$  where  $\text{fan\_in}$  is the number of input units in the weight tensor) [39]. In the training procedure of the ANN model, we used the Adam optimizer, the categorical cross entropy as a loss function (the function to minimize during optimization), and the accuracy to monitor the training. The training was structured into 1000 epochs, with a batch size of 32 and a validation split of 0.2 (80% of the data was used to train and 20% to validate the model). The ANN model structure used is available in the Supplementary Material.

#### 4. Discussion

The evaluation of accuracy and kappa statistics of these machine-learning classifiers on our dataset using the predicted classes for the whole and the selected spectral ranges is reported in Table 2. The LDA models built using the whole wavelength spectrum were not improved by using the PCA or GA selected values. For both the LDA models the accuracy and Cohen's kappa values indicate a 60% of overall accuracy and a moderate agreement between the model predictions and the actual class values non-happening by chance. For the ANN model with GA selection, both accuracy and K values are higher than all the other models. K is close to 0.70, that indicates a substantial agreement [35]. The predicted classes for the LDA models on the test set are reported in Tables 3–6.

**Table 3.** Confusion matrix for LDA using all the wave numbers.

| Prediction |    | Reference |   |    |
|------------|----|-----------|---|----|
|            | A  | B         | C | D  |
| A          | 10 | 1         | 2 | 1  |
| B          | 2  | 11        | 2 | 2  |
| C          | 1  | 3         | 9 | 0  |
| D          | 3  | 2         | 5 | 17 |

**Table 4.** Confusion matrix for LDA with the GA selected wave numbers.

| Prediction |   | Reference |   |    |
|------------|---|-----------|---|----|
|            | A | B         | C | D  |
| A          | 9 | 2         | 1 | 3  |
| B          | 3 | 11        | 4 | 4  |
| C          | 1 | 4         | 9 | 1  |
| D          | 3 | 0         | 4 | 12 |

**Table 5.** Confusion matrix for LDA with the PCA selected wave numbers.

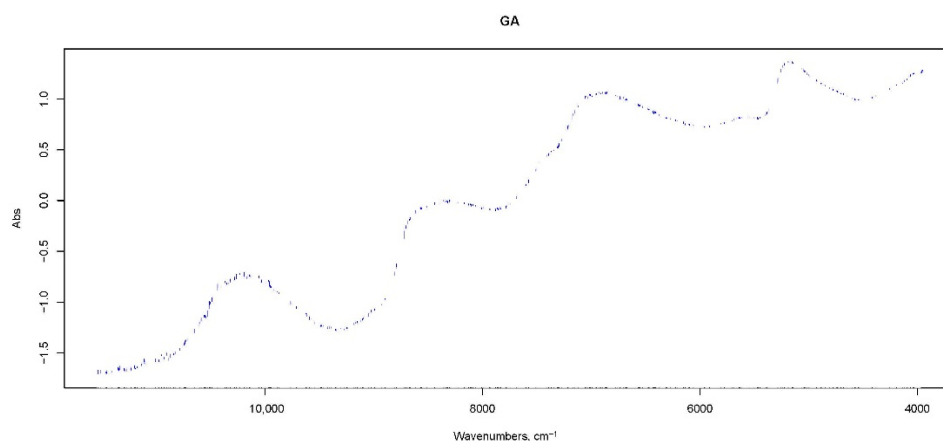
| Prediction |   | Reference |    |    |
|------------|---|-----------|----|----|
|            | A | B         | C  | D  |
| A          | 8 | 0         | 1  | 3  |
| B          | 2 | 13        | 3  | 6  |
| C          | 1 | 2         | 11 | 0  |
| D          | 5 | 2         | 3  | 11 |

**Table 6.** Confusion matrix for ANN with the GA selected wave numbers.

| Prediction |    | Reference |    |    |
|------------|----|-----------|----|----|
|            | A  | B         | C  | D  |
| A          | 13 | 0         | 3  | 5  |
| B          | 0  | 15        | 0  | 2  |
| C          | 2  | 2         | 15 | 2  |
| D          | 1  | 0         | 0  | 11 |

Each column of the matrix represents the actual class while each row represents the class predicted by the model. The number of correctly classified samples (true positives) in the confusion matrix for LDA whole and GA-LDA models is higher than the misclassified (false negatives) for each class, except for class C, which shows a 50% of samples correctly classified (Tables 3 and 4). Instead, in the PCA-LDA samples belonging to class A show the worst classification (50% correctly classified) (Table 5). The percentage of correct classification improves drastically with the ANN algorithm. The ANN model built with GA selected wavelengths lead to a better classification of samples compared to all the LDA models (correct classification: 81% A, 88% B, 83% C of the test sets), with still some misclassifications for the D group (55% D correctly classified) as shown in Table 6.

For our samples, linear regression models were not able to effectively predict the classes. The use of a non-linear model built with an artificial neural network created an efficient multi-classification model. Some conventional parameters have been measured on samples belonging to the four classes but not analyzed with the NIR spectrometer. Among the parameters analyzed, significant differences were found concerning berry weight, sugar content, number of clusters per vine, and texture parameters. Both sugar content and berry weight were higher for samples A and B, followed by C and D. The texture parameters showed the same trend of berry weight and sugar content. Moreover, A and C samples had a higher number of clusters per vine, followed by B and D [26]. Even if these results are mean values of parameters measured on berries different from those employed in the NIR analysis, they allowed us to understand that generally the samples A and D or B and C have a different chemical composition. If the attribution to a different class could be understandable for berries sharing the same rootstock or cover crop system, the misclassification of A as D or B as C poses some doubts since those classes have in common neither the rootstock nor the cover crop. The GA-selected wavenumbers are spread on the whole wave number dataset therefore, it is not possible to attribute the differentiation to a specific spectral region. Figure 4 shows a spectrum with only the wavelengths selected by the GA in blue.



**Figure 4.** GA-selected regions of NIR spectrum.

Concerning the misclassified samples in the ANN model, due to the inherent differences found with primary methods, we hypothesize that for those samples, sugar and water content (the latter linked to the berry weight) could have played a role; however, it is necessary to perform further analysis to confirm this hypothesis. Further studies involving a larger number of samples with a chemical characterization are ongoing to improve the classification model. Moreover, since the composition of grape, in addition to being strongly dependent on the nature of the cultivar, is also influenced by several other factors such as climate, agricultural practices and technological factors, the study will be repeated in several years with climate data detection. Striking differences could appear in several years of analysis, smoothing out the discrepancies found here. However, to our knowledge, it is the first time that NIR spectroscopy is applied for the classification of grape berries based

on the rootstock–cover crop combination. This discrimination ability indicates another potentiality of the NIR technique that could be further exploited.

## 5. Conclusions

For more sustainable viticulture oriented towards high-quality production, an appropriate combination of cover crop and rootstock represents a strategy to optimize yield, improve grape quality, and preserve the environment. Choosing the best combination of cover crops and rootstock requires costly untargeted metabolomic screenings. As an alternative to conventional techniques, which require large investments in terms of both time and money, we applied an economical and fast procedure, based on an open-source software for the chemometric analysis. In this work, NIR spectroscopy with machine learning methods was used to investigate the differences in the metabolic composition of the samples. A genetic algorithm-based method for selecting wavelengths as independent variables for an ANN classification model was applied. A model able to discern the influence of rootstock and cover crop combinations on the same variety sharing the geographical origin could be useful for a NIR-based characterization of grapes. Even if these results are limited to one harvest year on a limited number of samples, the good ability of the model to differentiate NIR spectra of berries matrix samples by rootstock and grassing indicates the presence of a correlation that is captured by the NIR analysis. The same procedure could be followed by others on their spectra, even recorded on a different spectrometer in different field conditions to obtain a classification of four or even more classes.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agriculture13010005/s1>.

**Author Contributions:** Conceptualization, T.B. and L.T.; methodology, T.B.; software, T.B.; formal analysis, T.B. and A.M.A.; data curation, T.B.; writing—original draft preparation, T.B.; writing—review and editing, T.B., L.T. and A.M.A.; project administration, T.B. and L.T.; funding acquisition, L.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Italian Ministry of University and Research (MUR), Project AGRIFOOD ARS01\_00640, “POFACS—Conservabilità, qualità sicurezza dei prodotti ortofrutticoli ad alto contenuto di servizio”, PON R&I 2014-2020.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Raw data are available from the authors upon request.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript.

## References

1. Cozzolino, D.; Roumeliotis, S.; Eglinton, J. Exploring the use of near infrared (NIR) reflectance spectroscopy to predict starch pasting properties in whole grain barley. *Food Biophys.* **2013**, *8*, 256–261. [\[CrossRef\]](#)
2. Ali Shah, S.S.; Zeb, A.; Qureshi, W.S.; Arslan, M.; Malik, A.U.; Alasmay, W.; Alanazi, E. Towards fruit maturity estimation using NIR spectroscopy. *Infrared Phys. Technol.* **2020**, *11*, 103479. [\[CrossRef\]](#)
3. Basile, B.; DeJong, T.M. La Riduzione del Vigore Delle Pianta Arboree da Frutto Mediante il Portainnesto. *Italus Hortus* **2007**, *14*, 58–70. Available online: <https://www.soihs.it/ItalusHortus/Review/Review%2006/Basile%20e%20De%20Jong.pdf> (accessed on 1 January 2022).
4. Habran, A.; Commisso, M.; Helwi, P.; Hilbert, G.; Negri, S.; Ollat, N.; Gomès, E.; van Leeuwen, C.; Guzzo, F.; Delrot, S. Rootstocks/Scion/Nitrogen Interactions Affect Secondary Metabolism in the Grape Berry. *Front. Plant Sci.* **2016**, *9*, 1134. [\[CrossRef\]](#)
5. Dong, D.; Shi, Y.N.; Mou, Z.M.; Chen, S.Y.; Zhao, D.K. Grafting: A potential method to reveal the differential accumulation mechanism of secondary metabolites. *Hortic. Res.* **2022**, *28*, uhac050. [\[CrossRef\]](#)
6. Mercenaro, L.; Nieddu, G.; Pulina, P.; Porqueddu, C. Sustainable management of an intercropped Mediterranean vineyard. *Agric. Ecosyst. Environ.* **2014**, *192*, 95–104. [\[CrossRef\]](#)
7. Guerra, B.; Steenwerth, K. Influence of floor management technique on grapevine growth, disease pressure, and juice and wine composition: A review. *Am. J. Enol. Vitic.* **2012**, *63*, 149–164. [\[CrossRef\]](#)

8. Muscas, E.; Cocco, A.; Mercenaro, L.; Cabras, M.; Lentini, A.; Porqueddu, C.; Nieddu, G. Effects of vineyard floor cover crops on grapevine vigor, yield, and fruit quality, and the development of the vine mealybug under a Mediterranean climate. *Agric. Ecosyst. Environ.* **2017**, *237*, 203–212. [CrossRef]
9. Tesic, D.; Keller, M.; Hutton, R.J. Influence of vineyard floor management practices on grapevine vegetative growth, yield, and fruit composition. *Am. J. Enol. Vitic.* **2007**, *58*, 1–11. [CrossRef]
10. Ferrara, G.; Nigro, D.; Torres, R.; Gadaleta, A.; Fidelibus, M.W.; Mazzeo, A. Cover crops in the inter-row of a table grape vineyard managed with irrigation sensors: Effects on yield, quality and glutamine synthetase activity in leaves. *Sci. Hortic.* **2021**, *281*, 109963. [CrossRef]
11. Naulleau, A.; Gary, C.; Prévot, L.; Hossard, L. Evaluating strategies for adaptation to climate change in grapevine production—a systematic review. *Front. Plant Sci.* **2021**, *11*, 607859. [CrossRef] [PubMed]
12. Basile, T.; Marsico, A.D.; Perniola, R. NIR analysis of intact grape berries: Chemical and physical properties prediction using multivariate analysis. *Foods* **2021**, *10*, 113. [CrossRef] [PubMed]
13. Nasirahmadi, A.; Hensel, O. Toward the next generation of digitalization in agriculture based on digital twin paradigm. *Sensors* **2022**, *22*, 498. [CrossRef] [PubMed]
14. OIV (Organisation Internationale de la Vigne et du Vin). *Compendium of International Methods of Analysis—OIV, Folin-Ciocalteu Index, Method OIV-MA-AS2-10: R2009*; OIV: Paris, France, 2009.
15. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: <https://www.R-project.org/> (accessed on 1 January 2022).
16. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
17. Casas, P. Package “funModeling”. Exploratory Data Analysis and Data Preparation Tool-Box. 2020. Available online: <https://cran.r-project.org/web/packages/funModeling/funModeling.pdf> (accessed on 1 January 2022).
18. Scrucca, L. GA: A Package for Genetic Algorithms in R. *J. Stat. Softw.* **2013**, *53*, 1–37. [CrossRef]
19. Willighagen, E.; Ballings, M. R Based Genetic Algorithm. 2022. Available online: <https://cran.r-project.org/web/packages/genalg/genalg.pdf> (accessed on 1 January 2022).
20. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
21. Kalinowski, T.; Falbel, D.; Allaire, J.J.; Chollet, F.; Tang, Y.; Van Der Bijl, W.; Studer, M.; Keydana, S. R Interface to ‘Keras’. 2022. Available online: <https://cran.r-project.org/web/packages/keras/keras.pdf> (accessed on 1 January 2022).
22. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002.
23. Kucheryavskiy, S. mdatools—R package for chemometrics. *Chemom. Intell. Lab. Syst.* **2020**, *198*, 103937. [CrossRef]
24. Masi, G.; Gentile, G.; Amendolagine, A.M.; Mazzone, F.; Rocotelli, S.; Caputo, A.R.; Tarricone, L. Effetti della gestione del suolo sulle performance della varietà ad uva da tavola Autumn Pearl in coltivazione bio: Primi risultati del progetto Oltre.bio (P.S.R. Regione Puglia 16.2). In Proceedings of the IX Convegno Nazionale di Viticoltura, Conegliano, Italy, 13–15 June 2022.
25. Ullah, H.; Santiago-Arenas, R.; Ferdous, Z.; Attia, A.; Datta, A. Chapter Two—Improving water use efficiency, nitrogen use efficiency, and radiation use efficiency in field crops under drought stress: A review. In *Advances in Agronomy*; Sparks, D.L., Ed.; Elsevier: Amsterdam, The Netherlands; Academic Press: Cambridge, MA, USA, 2019; Volume 156, pp. 109–157. [CrossRef]
26. Louw, E.D.; Theron, K.I. Robust prediction models for quality parameters in Japanese plums (*Prunus salicina* L.) using NIR spectroscopy. *Postharvest Biol. Technol.* **2010**, *58*, 176–184. [CrossRef]
27. Huang, J.; Romero-Torres, S.; Moshgbar, M. Practical Considerations in Data Pre-Treatment for NIR and Raman Spectroscopy. *Am. Pharm. Rev.* **2010**, *13*, 116–127. Available online: <https://www.americanpharmaceuticalreview.com/Featured-Articles/116330-Practical-Considerations-in-Data-Pre-treatment-for-NIR-and-Raman-Spectroscopy/> (accessed on 1 January 2022).
28. Basile, T.; Marsico, A.D.; Perniola, R. Use of Artificial Neural Networks and NIR Spectroscopy for Non-Destructive Grape Texture Prediction. *Foods* **2022**, *11*, 281. [CrossRef]
29. Cai, W.; Li, Y.; Shao, X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 188–194. [CrossRef]
30. Wutzl, B.; Leibnitz, K.; Rattay, F.; Kronbichler, M.; Murata, M.; Golaszewski, S.M. Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness. *PLoS ONE* **2019**, *14*, e0219683. [CrossRef] [PubMed]
31. Koljonen, J.; Nordling, T.E.M.; Alander, J.T. A review of genetic algorithms in near Infrared spectroscopy and chemometrics: Past and future. *J. Near Infrared Spectrosc.* **2008**, *16*, 189–197. [CrossRef]
32. Kim, H.C.; Kim, D.; Bang, S.Y. Extensions of LDA by PCA mixture model and class-wise features. *Pattern Recognit.* **2003**, *36*, 1095–1105. [CrossRef]
33. Li, Y.; Yang, H. Honey discrimination using visible and near-infrared spectroscopy. *Int. Sch. Res. Not.* **2012**, *2012*, 487040. [CrossRef]
34. Trejos, J.; Villalobos-Arias, M.A.; Espinoza, J.L. Chapter 5: Variable Selection in Multiple Linear Regression Using a Genetic Algorithm. In *Handbook of Research on Modern Optimization Algorithms and Applications in Engineering and Economics*; Vasant, P., Weber, G.-W., Dieu, V.N., Eds.; IGI Global Publisher: Hershey, PA, USA, 2016; pp. 133–159.
35. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]
36. Büchmann, N.B.; Josefsson, H.; Cowe, I.A. Performance of European Artificial Neural Network (ANN) calibrations for moisture and protein in cereals using the Danish Near-Infrared Transmission (NIT) network. *Cereal Chem. J.* **2001**, *78*, 572–577. [CrossRef]

37. Kumar, D.; Pandey, A.; Sharma, N.; Flügel, W. Modeling Suspended Sediment Using Artificial Neural Networks and TRMM-3B42 Version 7 Rainfall Dataset. *J. Hydrol. Eng.* **2015**, *20*, C4014007. [[CrossRef](#)]
38. Brownlee, J. *Deep Learning with Python: Develop Deep Learning Models on Theano and TensorFlow using Keras, v1.9*; Machine Learning Mastery: Melbourne, Australia, 2016.
39. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016; ISBN-13: 978-0262035613.

**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.