Selcuk Journal of Agriculture and Food Sciences
SJAFS
http://sjafs.selcuk.edu.tr/sjafs/index
(2022) 36 (3), 476-481
Research Article
e-ISSN: 2458-8377
DOI:10.15316/SJAFS.2022.062

Determination of the ADF and IVOMD Content of Sugarcane Using Near Infrared Spectroscopy Coupled with Chemometrics

Ozcan CATALTAS[1,*], Kemal TUTUNCU[2]
[1,2]Selçuk University, Faculty of Technology, Department of Electrical Electronic Engineering, Konya, Türkiye

ARTICLE INFO

Article history:
Received date: 31.10.2022
Accepted date: 28.11.2022

Keywords:
Near-infrared spectroscopy
Chemometrics
Partial least squares regression
Food analysis

ABSTRACT

Sugarcane is a plant whose quality parameters are required to be determined both for being one of the substances used in sugar production and for being used as animal feed. Near-infrared spectroscopy is a technique that has already been used for predicting the parameters of various plants and has gained popularity in recent years. This study proposes a near-infrared spectroscopy-based model for the rapid and effortless analysis of acid detergent fiber fraction and vitro organic matter digestibility parameters of the sugarcane plant. Partial least squares regression was combined with common preprocessing methods for modeling. This model yielded an R^2_{CV} value of 0.935 and 0.953 for the acid detergent fiber fraction and vitro organic matter digestibility parameters, respectively. Then, the spectra from three handheld spectrometers were combined using a proposed combination method to generate new spectra with higher spectral resolution. New models were built using these generated spectra and compared to the previous result. Obtained results showed that combining spectra from different spectrometers can improve model performance.

1. Introduction

Near-Infrared Spectroscopy (NIRS) has become a popular method in recent years as it allows non-destructive and fast analysis. NIRS, which is based on measuring vibrational frequencies in the near-infrared (NIR) region (780-2500 nm) in the electromagnetic spectrum and is characterized by combination vibrations with molecular overtones, stands out as an advantageous method. Chemical bonds with high vibrational frequencies, such as C-H, O-H, and N-H form overtone and combination bands in the NIR region and their intensity can be measured in this region (Ciurczak et al. 2021). Thus, it can be easily used in qualitative and quantitative analysis in many areas, such as agriculture, food, and pharmacy. Although it has been used in many different areas, the most intensive studies have been carried out in the food sector. Several studies have been conducted to determine food quality (Mohamed et al. 2021; Teye et al. 2020; Yang et al. 2022), detect food adulteration (De Girolamo et al. 2020; Genis et al. 2021; Laborde et al. 2021), and classify food types (Fu et al. 2022).

Sugarcane is one of the primary materials for sugar production. Its production is demanding as it requires special growing conditions (Guo et al. 2014). For this reason, it is important to determine the quality of the products and to apply the necessary fertilization or insecticide spraying. For this purpose, several studies have been carried out to determine the mineral content (Steidle Neto et al. 2017), to predict the crude protein and sugar content (Ryckewaert et al. 2022), to determine the lignin content of sugarcane (Assis et al. 2017), to screen sugarcane breeding (Guo et al. 2014), and to determine soil organic carbon in sugarcane fields (Zhao et al. 2022). Although sugarcane is considered inferior as an animal feed, its high dry matter and organic matter digestibility means it can be a good feed, particularly for ruminants (So et al. 2020).

Spectra acquisition in NIRS has traditionally been performed with expensive laboratory-type instruments (Schuler et al. 2009). In recent years, however, mini/micro spectrometers have continued to be developed that work in a narrow spectrum or with low spectral resolution. The increase in the variety of these spectrometers, their widespread use, and the increase in resolution will facilitate more comprehensive studies in the NIRS.

The spectra obtained from spectrometers cannot be used directly as they may contain several distortions. Although various preprocessing methods have been developed to suppress these distortions, there is no global method. On the other hand, various methods are being developed to extract meaningful data from the

* Corresponding author email: *ozcancataltas@selcuk.edu.tr*

preprocessed spectrum. Application examples of artificial neural networks (Pérez-Marín et al. 2007), deep learning (Cui & Fearn 2018), or ensemble methods (Kim et al. 2023; Mishra et al. 2020) have been encountered frequently in recent years. However, partial least square regression (PLSR), a cult method for qualitative analysis, is still the most popular method.

The objective of our work is twofold: to develop a model for the prediction of acid detergent fiber fraction (ADF) and in vitro organic matter digestibility (IVOMD) content of sugarcane samples using the spectra of different handheld spectrometers and to investigate whether combining the spectra of different devices would be beneficial.

The paper is organized as follows: Section two provides brief information on preprocessing techniques and partial least squares regression. The properties of the used dataset and spectra combining procedure are also given in section two. Performance metrics and results are presented in section three. Section four contains the main conclusion of our study.

2. Materials and Methods

2.1. Dataset description

This study used the sugarcane dataset to build and evaluate a reliable model. This dataset contains spectra of 60 sugarcane samples which measured different regions of sugarcane. These spectra were obtained from eight different spectroscopy instruments; each has different ranges and resolutions. Of these, NIRScan Nano (Texas Instrument), MicroNIR1700 (Viavi), and TellSpec were included in this study. The main reason for our choice is that all three devices operate in a spectral range close to each other. Detailed information about the chosen instruments is shown in Table 1. The dataset also included four reference values: Total sugar content (TS), crude protein (CP), ADF, and IVOMD. Because (Ryckewaert et al. 2022) used TS and CP in their study, only ADF and IVOMD parameters were included. Statistical information of these parameters is given in Table 2. Absorbance spectra and their relation to the ADF and IVOMD parameters are shown in Figures 1 and 2. The dataset is available at (Zgouz et al. 2020).

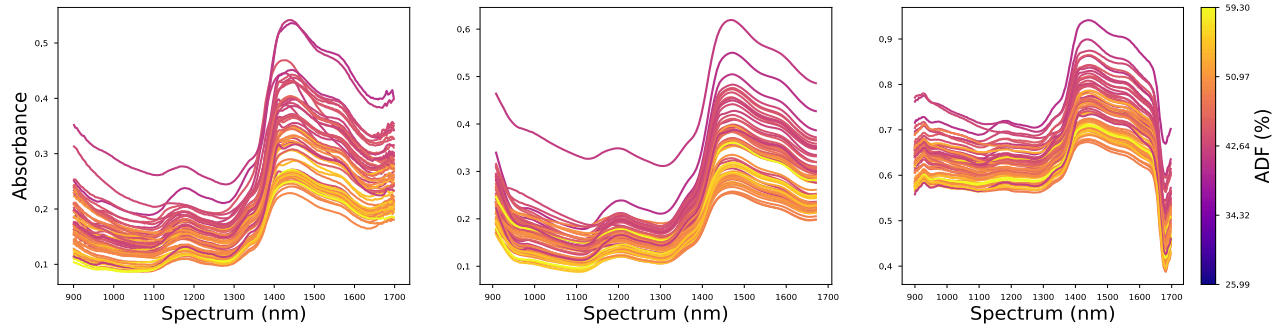
Table 1
Spectral range and resolution of the instruments used in the study

Instrument	Spectral Range (nm)	Resolution (nm)	Number of Features
NIRScan Nano	901-1701	3.9	228
MicroNIR1700	908-1676	6.1	125
TellSpec	900-1700	3.8	256

Table 2
Statistical information on ADF and IVOMD parameters

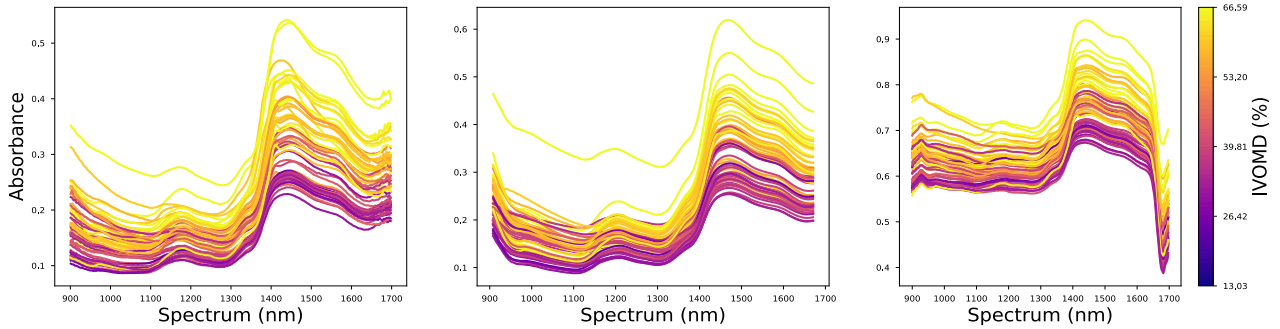
Parameter	Min	Max	Mean	Std
ADF (% DM)	25.99	59.30	39.17	8.61
IVOMD (% DM)	13.03	66.59	41.03	14.97

DM: Dry Matter



a) b) c)

Figure 1
The relation of used spectra and ADF content. These spectra belong to a) NIRScan Nano, b) MicroNIR1700, c) TellSpec



a) b) c)

Figure 2
The relation of used spectra and IVOMD content. These spectra belong to a) NIRScan Nano, b) MicroNIR1700, c) TellSpec

2.2. Preprocessing the spectra

Preprocessing the spectra is one of the crucial steps in NIRS that directly affects the performance of models. The sugarcane spectra were preprocessed before applying regression analysis using a combination of common preprocessing methods. Standard Normal Variate (SNV) (Barnes et al. 1989), Multiplicative Scatter Correction (MSC) (Geladi et al. 1985), Savitzky-Golay (SG) (Savitzky & Golay 1964), and a combination of these methods were used as preprocessing method.

2.3. Partial least squares regression

The main goal of a regression model is to find a relationship between independent variables and dependent variable(s). However, in some cases, as in spectroscopy data, many independent variables correlate with others. This situation is named as multicollinearity problem (Frank & Friedman 1993). To deal with this problem, (Wold et al. 1983) proposed partial least squares regression. PLSR proposes a solution by reducing the dimensionality of correlated variables.

2.4. Spectra combination procedure

In this study, a spectrum combining technique was proposed that combines spectra from three instruments. Since the spectra of different instruments contain baseline differences, each spectrum in the dataset was first normalized to a range between 0 and 1. After, these spectra were combined column by column and sorted by wavelength number. In this new spectrum, peaks were seen across the spectrum. A five-point moving average filter was used to eliminate these peaks. The combined dataset contains 609 features. The combination procedure is illustrated in Figure 3.

3. Results and Discussion

In this paper, preprocessing methods and regression models were implemented in Python (version 3.7.13) using scikit-learn library (version 1.0.2) (Lemaitre 2021). Two popular metrics were chosen to evaluate model performance. One of them, the coefficient of determination (R^2), is a measure that corresponds to the proportion of variation for a dependent variable that is explained by the independent variables. R^2 values close to 1 are preferable. The second metric, root mean squared error (RMSE), represents the square root of the average squared difference between the target value and the predicted value. Since RMSE is based on error, values closer to 0 indicate that the model performs better. The formulas of R^2 and RMSE are given in Equations 1 and 2.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (2)$$

In Equations 1 and 2, N is the number of samples, y_i, $\hat{y}_i$, and $\bar{y}_i$ are real, predicted, and mean values of the target, respectively.

This study proposes a chemometrics-based method for determining ADF and IVOMD parameters of sugarcane. The first experiment was carried out using classical PLSR. Different preprocessing methods were applied to the spectra taken from NIRScan, MicroNIR1700, and TellSpec instruments. At this point, popular preprocessing methods such as SNV, MSC, SG, and their binary combinations were preferred. In addition, a search space was created to determine the optimal values of window length, the order of the polynomial, and the order of derivative for SG. This search space was 5-29 for window length and 0-2 for the order of polynomial and derivative. Models were built to determine ADF and IVOMD parameters with preprocessed spectra and PLSR. At this stage, the dataset was randomly divided for calibration and testing at a rate of 66% to 33%. That is, 40 samples are used for the calibration of the model and 20 samples are reserved for the testing. Then, the optimal latent variables (LV) value for PLSR was determined using 5-fold cross-validation (CV) with the calibration set. In determining the LV, 1-15 values were selected as the search space. Table 3 and Table 4 show the performance results of the created CV model for the most appropriate LV value. Then, a new model was created using the calibration data set for the most appropriate LV value. The performance of this model was evaluated using the test set and is given in Table 3 and Table 4.

Table 3
Obtained results on prediction of the ADF parameter.

Instrument	Preprocessing	LV	R^2_{test}	R^2_{CV}	$RMSE_{test}$	$RMSE_{CV}$
Combined	SG(1,2,9)	7	0.885	0.953	3.281	1.743
NIRScan Nano	SG(1,2,13) + MSC	8	0.823	0.935	4.074	2.044
NIRScan Nano	MSC + SG(0,1,21)	13	0.912	0.899	2.868	2.553
MicroNIR1700	SG(2,2,5) + SNV	4	0.926	0.849	2.639	3.118
Combined	MSC + SG(2,2,19)	9	0.939	0.797	2.385	3.628
MicroNIR1700	SG(2,2,9) + SNV	3	0.951	0.797	2.138	3.624
TellSpec	SNV + SG(1,2,19)	5	0.845	0.671	3.813	4.610
TellSpec	SNV + SG(2,2,25)	4	0.945	0.606	2.276	5.047

The results are sorted based on R^2_{CV} in descending order. LV: Latent variables, SG: Savitzky-Golay (window length, the order of the polynomial, the order of the derivative)

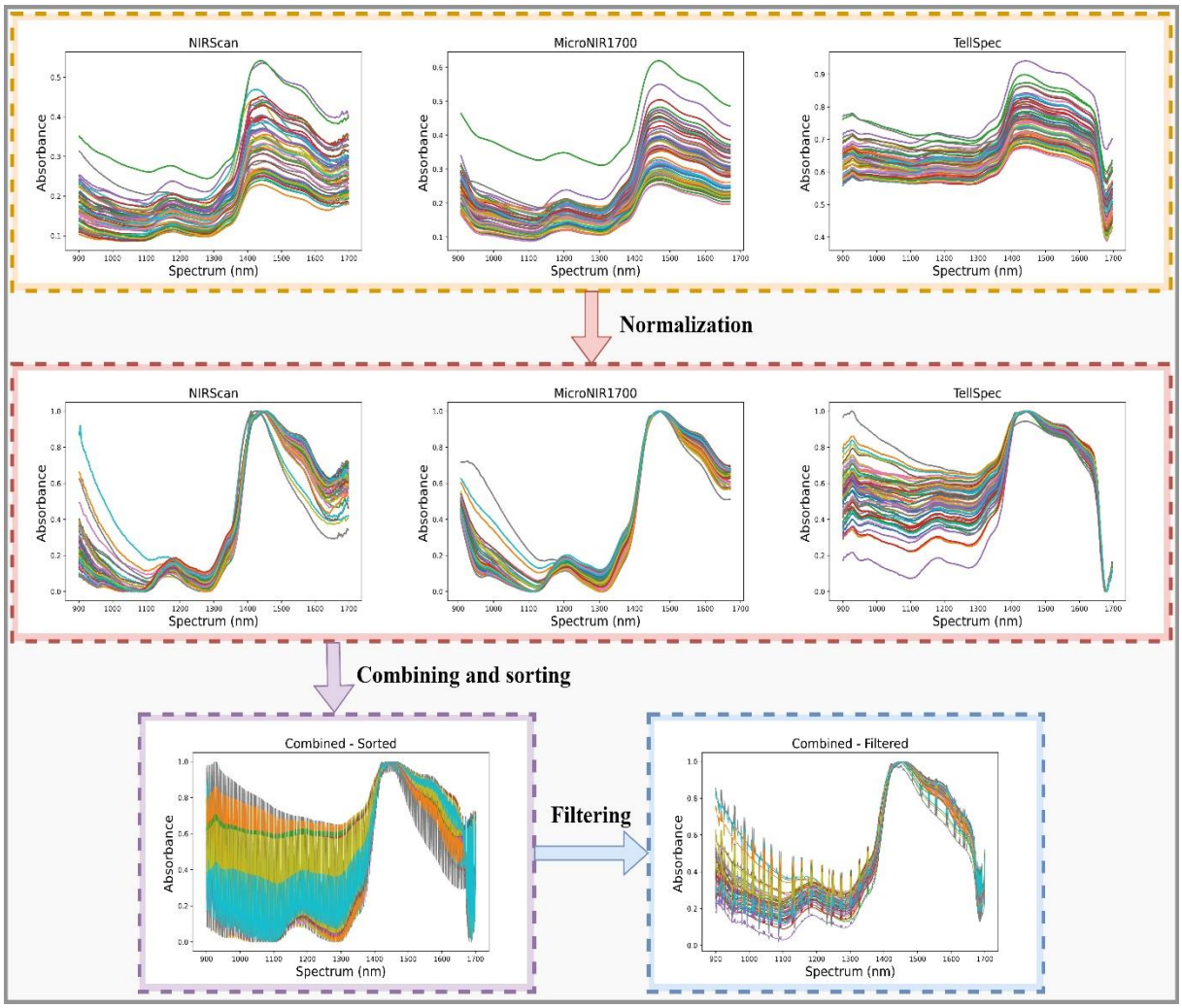


Figure 3
Application steps of the proposed spectrum combining algorithm

In the second experiment, the spectra of three instruments were combined, as explained in section 2.4. After the combination, the obtained spectra had a mean resolution value of 1.318 ± 0.899 nm. Similar preprocessing and regression process was done for the new spectra. Obtained results are shown in Table 3 and Table 4.

When we examine Table 3, the best performance according to R^2_{CV} and $RMSE_{CV}$ metrics belongs to the combined spectra. The higher R^2_{test} value was obtained with MicroNIR1700 spectra. For predicting the IVOMD parameter, the spectra of NIRScan Nano showed higher R^2_{CV} and lower $RMSE_{CV}$ values, while the combined spectra showed higher R^2_{test} and lower $RMSE_{test}$ values.

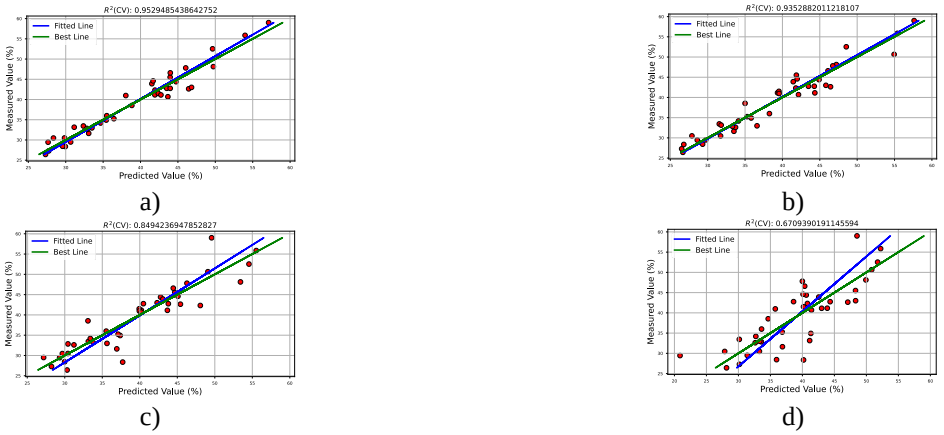


Figure 4
Predicted vs. Measured ADF values. Best values are given according to R^2_{CV} for instruments; a) Combined, b)NIRScan Nano, c) MicroNIR1700, d) TellSpec

Table 4
Obtained results on prediction of the IVOMD parameter.

Instrument	Preprocessing	LV	R^2_{test}	R^2_{CV}	$RMSE_{test}$	$RMSE_{CV}$
NIRScan Nano	SG(1,2,19) + MSC	12	0.746	0.953	8.064	3.128
Combined	SG(1,2,17)	10	0.858	0.945	6.037	3.375
NIRScan Nano	SG(0,0,5) + SNV	13	0.899	0.879	5.089	5.023
MicroNIR1700	SNV + SG(2,2,5)	4	0.893	0.867	5.228	5.257
Combined	SG(2,2,21)	5	0.930	0.861	4.244	5.373
MicroNIR1700	SG(2,2,13)	4	0.866	0.778	5.863	6.797
TellSpec	SG(1,2,21) + SNV	4	0.789	0.673	7.363	8.241
TellSpec	SNV + SG(2,2,13)	3	0.909	0.541	4.840	9.770

The results are sorted based on R^2_{CV} in descending order. LV: Latent variables, SG: Savitzky-Golay (window length, the order of the polynomial, the order of the derivative)

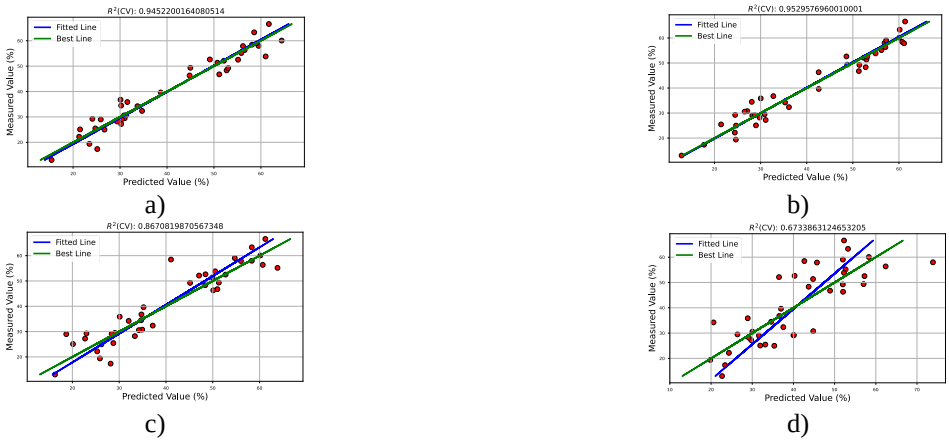


Figure 5
Predicted vs. Measured IVOMD values. Best values are given according to R^2_{CV} for instruments; a) Combined, b) NIRScan Nano, c) MicroNIR1700, d) TellSpec

Table 5
The mean value of obtained results for every preprocessing combination

Instrument	ADF		IVOMD	
	R^2_{test}	R^2_{CV}	R^2_{test}	R^2_{CV}
Combined	0.876	0.830	0.843	0.849
NIRScan Nano	0.848	0.869	0.808	0.879
MicroNIR1700	0.832	0.685	0.819	0.734
TellSpec	0.817	0.602	0.767	0.598

Each value in this table was calculated by taking the average of 390 combinations.

Figure 4 and Figure 5 show the predicted and measured output for each spectra. As expected, more successful results were obtained with the spectra from the NIRscan Nano, since the NIRscan Nano has a higher spectral resolution than the MicroNIR1700 device. On the contrary, the worst performance for both parameters was obtained with the spectra of TellSpec, although it has the highest spectral resolution among the three instruments. This may be due to the spectral sensitivity of the device.

One of the conclusions from this study is that the preprocessing method to be used varies according to the spectrum. All of the high-success models were obtained with different combinations of preprocessing methods or different parameters. This is the weakest point of classical chemometrics. Another conclusion is that combining the spectra of different instruments can yield better results. Table 5 shows the mean values of obtained results. It is essential to take the mean of all combinations to eliminate the parameter selection effect of preprocessing methods. According to Table 5, combined spectra give the highest scores according to the R^2_{test} metric for both targets and the second highest scores according to the R^2_{CV} metric.

4. Conclusion

In this study, the ADF and IVOMD parameters of sugarcane were determined noninvasively using near-infrared spectroscopy. PLSR, coupled with common preprocessing methods, was utilized for the building prediction model. The best results were obtained with NIRScan Nano according to R^2_{CV} for both parameters (0.928 for ADF and 0.947 for IVOMD). The spectra from three instruments were combined to increase spectral resolution. New PLSR models were developed using these spectra. Obtained results have shown that combining spectra from different spectrometers helps to improve model performance. Future studies can focus on developing machine learning-based spectra combination algorithms.

5. Acknowledgements

This study was partially supported Selcuk University Academic Staff Training Program (2018-OYP-041).

6. References

Assis C, Ramos RS, Silva LA, Kist V, Barbosa MHP, and Teófilo RF (2017). Prediction of Lignin Content in Different Parts of Sugarcane Using Near-Infrared Spectroscopy (NIR), Ordered Predictors Selection (OPS), and Partial Least Squares (PLS). *Applied Spectroscopy* 71:2001-2012.

Barnes RJ, Dhanoa MS, and Lister SJ (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy* 43:772-777. 10.1366/0003702894202201

Ciurczak EW, Igne B, Workman Jr J, and Burns DA. 2021. *Handbook of near-infrared analysis*: CRC press.

Cui C and Fearn T (2018). Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemometrics and Intelligent Laboratory Systems* 182:9-20. https://doi.org/10.1016/j.chemolab.2018.07.008

De Girolamo A, Arroyo MC, Cervellieri S, Cortese M, Pascale M, Logrieco AF, and Lippolis V (2020). Detection of durum wheat pasta adulteration with common wheat by infrared spectroscopy and chemometrics: A case study. *LWT* 127:109368. https://doi.org/10.1016/j.lwt.2020.109368

Frank IE, and Friedman JH (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics* 35:109-135. 10.2307/1269656

Fu L, Sun J, Wang S, Xu M, Yao K, Cao Y, and Tang N (2022). Identification of maize seed varieties based on stacked sparse autoencoder and near-infrared hyperspectral imaging technology. *Journal of Food Process Engineering* 45. 10.1111/jfpe.14120

Geladi P, MacDougall D, and Martens H (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy* 39:491-500. 10.1366/0003702854248656

Genis HE, Durna S, and Boyaci IH (2021). Determination of green pea and spinach adulteration in pistachio nuts using NIR spectroscopy. *LWT* 136:110008. https://doi.org/10.1016/j.lwt.2020.110008

Guo H, Chen J, Pan T, Wang J, and Cao G (2014). Vis-NIR wavelength selection for non-destructive discriminant analysis of breed screening of transgenic sugarcane. *Analytical Methods* 6:8810-8816. 10.1039/C4AY01833H

Kim S-Y, Hong S-J, Kim E, Lee C-H, and Kim G (2023). Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy. *Journal of Food Engineering* 338:111254. https://doi.org/10.1016/j.jfoodeng.2022.111254

Laborde A, Puig-Castellví F, Jouan-Rimbaud Bouveresse D, Eveleigh L, Cordella C, and Jaillais B (2021). Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and Multivariate Curve Resolution. *Food Control* 119:107454. https://doi.org/10.1016/j.foodcont.2020.107454

Lemaitre G (2021). scikit-learn. *Available at https://github.com/scikit-learn/scikit-learn*.

Mishra P, Biancolillo A, Roger JM, Marini F, and Rutledge DN (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry* 132:116045. https://doi.org/10.1016/j.trac.2020.116045

Mohamed H, Nagy P, Agbaba J, and Kamal-Eldin A (2021). Use of near and mid infra-red spectroscopy for analysis of protein, fat, lactose and total solids in raw cow and camel milk. *Food Chemistry* 334:127436. https://doi.org/10.1016/j.foodchem.2020.127436

Pérez-Marín D, Garrido-Varo A, and Guerrero JE (2007). Non-linear regression methods in NIRS quantitative analysis. *Talanta* 72:28-42. https://doi.org/10.1016/j.talanta.2006.10.036

Ryckewaert M, Chaix G, Héran D, Zgouz A, and Bendoula R (2022). Evaluation of a combination of NIR micro-spectrometers to predict chemical properties of sugarcane forage using a multi-block approach. *Biosystems Engineering* 217:18-25. https://doi.org/10.1016/j.biosystemseng.2022.02.019

Savitzky A, and Golay MJE (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36:1627-1639. 10.1021/ac60214a047

Schuler LP, Milne JS, Dell JM, and Faraone L (2009). MEMS-based microspectrometer technologies for NIR and MIR wavelengths. *Journal of Physics D: Applied Physics* 42:133001. 10.1088/0022-3727/42/13/133001

So S, Cherdthong A, and Wanapat M (2020). Improving sugarcane bagasse quality as ruminant feed with Lactobacillus, cellulase, and molasses. *Journal of Animal Science and Technology* 62:648.

Steidle Neto AJ, Toledo JV, Zolnier S, Lopes DdC, Pires CV, and Silva TGFd (2017). Prediction of mineral contents in sugarcane cultivated under saline conditions based on stalk scanning by Vis/NIR spectral reflectance. *Biosystems Engineering* 156:17-26. https://doi.org/10.1016/j.biosystemseng.2017.01.003

Teye E, Anyidoho E, Agbemafle R, Sam-Amoah LK, and Elliott C (2020). Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review. *Infrared Physics & Technology* 104:103127. https://doi.org/10.1016/j.infrared.2019.103127

Wold S, Martens H, and Wold H (1983). The multivariate calibration problem in chemistry solved by the PLS method. In: Kågström B, and Ruhe A, editors. Matrix Pencils. Berlin, Heidelberg: Springer Berlin Heidelberg. p 286-293.

Yang J, Luo X, Zhang X, Passos D, Xie L, Rao X, Xu H, Ting KC, Lin T, and Ying Y (2022). A deep learning approach to improving spectral analysis of fruit quality under interseason variation. *Food Control* 140:109108. https://doi.org/10.1016/j.foodcont.2022.109108

Zgouz A, Héran D, Barthès B, Bastianelli D, Bonnal L, Baeten V, Lurol S, Bonin M, Roger J-M, Bendoula R, and Chaix G (2020). Dataset of visible-near infrared handheld and micro-spectrometers – comparison of the prediction accuracy of sugarcane properties. *Data in Brief* 31:106013. https://doi.org/10.1016/j.dib.2020.106013

Zhao X, Zhao D, Wang J, and Triantafilis J (2022). Soil organic carbon (SOC) prediction in Australian sugarcane fields using Vis–NIR spectroscopy with different model setting approaches. *Geoderma Regional* 30:e00566. https://doi.org/10.1016/j.geodrs.2022.e00566