Alfalfa quality detection by means of VIS-NIR optical fiber reflection spectroscopy

C. R. Zamarreño[1,2], A. Gracia-Moises[3], I. Vitoria[1,2], J. J. Imas[1,2], L. Castaño[1], A. Avedillo[1], Ignacio R. Matías[1,2].
[1]Public University of Navarra, Electrical, Electronic and Communications Engineering Department, 31006 Pamplona, Spain
[2] Institute of Smart Cities, Jeronimo de Ayanz Building, 31006 Pamplona, Spain
[3] Pyroistech S.L., Tajonar 22, 31006 Pamplona, Spain
carlos.ruiz@unavarra.es
ORCID: 0000-0001-6601-5449

A first approach study for the classification of alfalfa (medicago sativa) quality has been performed by means of VIS-NIR optical fiber reflection spectroscopy. Reflection spectral data has been obtained from alfalfa samples comprising six different qualities. Obtained data has been classified and organized to feed supervised self-learning algorithms. Neural networks have been used in order to differentiate the quality level of the samples. Obtained results permit to validate the proposed approach with 72% of the samples properly classified. In addition, proposed solution was implemented in a low cost automated detection prototype suitable to be used by non-qualified operators. Obtained equipment consist of a first step towards its utilization in quality monitoring and classification of many other products in the agri-food field.

Keywords—optical fiber, alfalfa, optical spectroscopy, reflection, neural networks

I. Introduction

Raw products quality in the agri-food sector can be generally associated to their visual aspect. In this sense, a rough estimation of their maturity, freshness, or the amount of imperfections can be obtained by a simple visual inspection. Traditionally, the quality sorting process was made by visual inspection. However, visual inspection does not guarantee a proper identification and classification of the products on a day by day basis. Human visual inspection is not reliable or as objective as it is commonly believed. Variables like changes in the lighting, the operator who is inspecting the product or the mood of the operator himself/herself can severely influence the process quality determination, which also impacts the final price of a product and the benefits of the company.

The mentioned problem affects almost every industry in the agri-food industry that demands and objective and repeatable messurment solution for quality monitoring and product classification. Optical spectroscopy in the visible (VIS) and near infrared (NIR) regions can provide rapid, objective and accurate information for quality sorting competing the conventional time-consuming, expensive and sometimes complex quality processes performed in laboratories, such as Fourier transform NIR (FTNIR) or mass spectroscopy. In particular, VIS-NIR optical spectroscopy has been described in literature with positive results for the classification and characterization of many different agri-food products, such as olive oil [1,2], beer [3], wine [4], meat [5] or tomatoes [6].

While visible spectroscopy is commonly used for quality inspections [7] there are many applications that also include the NIR region [8-9]. NIR region of the electromagnetic spectrum can be used for detecting organic compounds due to the molecular overtone and combination vibrations [10].

Alfalfa (medicago sativa) also called lucerne is an important forage crop, most often stored as hay. The quality of the hay varies between the harvested alfalfa lots and along the harvesting season. Consequently, the price also varies, which makes essential the utilization of a fine classification system for the collection of alfalfa at the reception locations. Expensive approaches have already addressed this problem using expensive FTNIR equipment that provides a high resolution in a wide region of the NIR spectrum and data-treatment methods like support vector machines, or partial least squares regression [11-12]. However, current quality determination procedure generally consists of a visual inspection based on the color of the product made by an operator at the treatment plant where the best quality is given to a green color and the decreased qualities are either given to yellow and/or yellowish-brown colors. This quality determination method depends on the light and the subjectivity of the operator. Therefore, there is still room for improvement in order to obtain a uniform and repeatable test that enable to classify the product in the different categories of qualities. Current approach enables the utilization of cost-effective VIS-NIR equipment to acquire reflected spectral data from alfalfa samples. Obtained data is used here to train supervised self-learning algorithms that validate the solution and enable the implementation of an automated and affordable solution for the industry as it is described in the following sections.

II. Materials and Methods

A. Measurement setup

A typical reflection setup (see Figure 1) has been used to obtain the spectral information. The setup consists of a halogen broadband source (TAKHI from Pyroistech S.L.) that couples light into the reflectance probe (QR400-7- VIS-NIR from Oceaninsight™). The probe is fixed to a holder that allows maintain a fixed distance to the samples while they are moved below the probe. The light passes through the probe and illuminates the sample, part of it is reflected from the sample, enters the probe and is guided back to the detector. The spectrum of the reflected light is measured with the spectrometers USB2000 from Ocean Insight™ for visible spectrum and NIRQUEST512 from Ocean Insight™ for the near infrared spectrum. Here it is important to remark that the reflection probe is fixed in a perpendicular position to the

This work was supported by the Spanish Ministry through FPU18/03087 and PID2019-106231RB-I00 grants, Government of Navarra 0011-1365-2021-000065 and 0011-1408-2021-000007 grants and UPNA-ISC PhD grant.

samples with a gap between the tip of the probe and the top surface of the sample. Finally, the spectral data is stored in a computer for a later processing and analysis using Python.

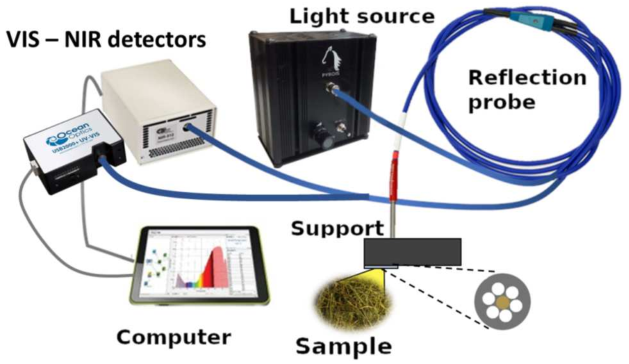

Fig. 1. Experimental setup used for diffuse reflection measurements.

B. Sample data collection

The samples used in this study were obtained and previously classified by the forage company NAFOSA S.A. in 6 different qualities (C1 – C6). Prior to the measurements the samples were pressed to homogenize the measurement surface. Two different methods were used in order to acquire the VIS-NIR reflected spectral data from the samples and compare the performance. The first method (discrete) consisted of obtaining discrete measurements from different points of the sample while the second method (continuous) consisted of obtaining continuous measurements while the sample moves under the probe.

TABLE I. NUMBER OF SAMPLES OBTAINED FOR THE STUDY IN THE VIS-NIR REGION USING CONTINUOUS AND DISCRETE ACQUISITION METHODS

		VIS	NIR	TOTAL
C1	*Continuous*	21095	21672	42767
	Discrete	1631	1694	3325
C2	*Continuous*	23992	28969	52961
	Discrete	1892	2143	4035
C3	*Continuous*	22000	18600	40600
	Discrete	1874	1613	3487
C4	*Continuous*	14502	14702	29204
	Discrete	1357	1378	2735
C5	*Continuous*	6800	5600	12400
	Discrete	243	427	670
C6	*Continuous*	8845	11138	19983
	Discrete	856	474	1330

After the measurement of the samples the acquisition data was organized in 4 different matrices attending to the spectral range (VIS/NIR) and method (discrete/continuous).

C. Data analysis

After grouping all the obtained spectral information, these data are used to fed a neural network. In total, the network is trained 4 times, corresponding to each matrix, in order to check which of the two ranges, visible or near infrared, and which type of analysis, discrete or continuous, offers better results.

The neural network used consisted of an intermediate layer with a total of 100 neurons, a dropout (parameter that reduces the overfitting of the network) of 0.1 and a batch size (parameter that indicates the number of samples processed before updating the model) of 12. 80% of the total data supplied to the network was dedicated to train the network while the remaining 20% were used to check the proper performance of the obtained model.

Results of the efficiency of the system are presented in the form of confusion matrices after the training and verification of the correct operation of the neural network.

III. RESULTS AND DISCUSSION

One of the main challenges in this work is the heterogeneous nature of the samples. Therefore, there is a need for a large number of measurements of each class in order to have a representative sampling. Since the spectral data obtained from the diffuse reflection measurements do not have a direct correlation with the quality parameter it is required to use supervised learning algorithms to train the IA (neural network in this case).

The network was trained and validated with the 4 groups of samples defined in section 2b (VIS-discrete, NIR-discrete, VIS-continuous and NIR-continuous), obtaining the confusion matrices for each of them. The confusion matrices show the successes and errors of the model when classifying the data in different categories or in other words, what percentage of the samples of a certain input class (rows) are correctly classified in the same output class (columns). For the correct interpretation of the matrices, it must be taken into account that the input qualities are represented in rows while the qualities identified by the network are associated with the columns. The number contained in each box represents the probability (recall) of the input class to be identified in each of the output classes and, therefore, the sum of all the columns in a single row is equal to 1. The confusion matrices obtained using the VIS and NIR spectrometer with the samples taken continuously and discretely are shown in Fig. 2 and Fig. 3 respectively.

VIS

	C1	C2	C3	C4	C5	C6
C1	0.97	0.022	0.005	0.003	0	0
C2	0.043	0.76	0.089	0.057	0.023	0.027
C3	0.027	0.1	0.74	0.12	0	0.016
C4	0.009	0.052	0.034	0.84	0.016	0.044
C5	0.002	0.048	0.054	0.088	0.77	0.033
C6	0.001	0.04	0.006	0.014	0.02	0.92

NIR

	C1	C2	C3	C4	C5	C6
C1	0.71	0.12	0.11	0.014	0.053	0
C2	0.2	0.42	0.24	0.05	0.088	0.007
C3	0.14	0.063	0.56	0.13	0.078	0.035
C4	0.068	0.048	0.15	0.56	0.15	0.027
C5	0.049	0.057	0.11	0.027	0.64	0.12
C6	0.014	0.012	0.05	0.021	0.21	0.69

Fig. 2. Confusion matrix obtained with VIS (left) and NIR (right) spectral information using continuous acquisition method.

VIS

	C1	C2	C3	C4	C5	C6
C1	0.97	0.016	0	0.013	0	0
C2	0.028	0.78	0.13	0.023	0.003	0.036
C3	0.033	0.11	0.77	0.074	0	0.01
C4	0.024	0.076	0.088	0.76	0.02	0.032
C5	0	0.087	0.11	0.087	0.72	0
C6	0	0.052	0.046	0.029	0	0.87

NIR

	C1	C2	C3	C4	C5	C6
C1	0.7	0.15	0.11	0.035	0.003	0.003
C2	0.16	0.52	0.23	0.087	0.004	0.007
C3	0.15	0.17	0.53	0.14	0.003	0.02
C4	0.07	0.11	0.16	0.65	0.004	0.008
C5	0.11	0.26	0.086	0.21	0.16	0.17
C6	0.062	0.1	0.12	0.073	0.031	0.6

Fig. 3. Confusion matrix obtained with VIS (left) and NIR (right) spectral information using discrete acquisition method.

The results of the confusion matrices in Figs. 2 and 3 show how the classification obtained through neural networks permits to classify most of the input samples correctly in the same output class. Deviations from the input class are generally towards a lower or higher class, which is associated to the fact that the Classification of input samples is performed by visual inspection assigning a discrete value between 1 and 6 without intermediate values. However, it is also observed that the deviation of the output class with respect to the input class is somewhat higher in the classification of the data obtained in the NIR compared with the results obtained with the samples measured in the VIS region for both discrete and continuous measurements. The improved performance of the neural network associated to the samples acquired the VIS region can be attributed to two different facts. First, the samples are classified using the naked eye where also operates the VIS spectrometer. Second, the larger amount of spectral information obtained with the VIS spectrometer (2048 points in the range 200-1100nm) compared with the reduced resolution of the NIR spectrometer (512 points in the range 850-1700nm). The better performance of the data in the VIS region also reduces the implementation costs of the prototype (more expensive in NIR) as it is detailed in the next paragraphs. Finally, it is important to observe how in the VIS region, the classification based on neural networks shows similar results for the case of discrete and continuous samples. In some cases, the results shown with the samples obtained continuously are even better, which indicates that it is possible to automate the data collection process for the classification of the samples, thus greatly reducing the acquisition time (around 10 times less).

The main objective of the prototype to be developed here is to reproduce the laboratory set-up in a compact unit that allows the user to use the device in the simplest way as well as avoiding possible human errors. Likewise, it has to guarantee a fast and precise analysis.

The proposed prototype consisted of a wide tray that enables to allocate the alfalfa sample inside. The tray is mounted on a rotating system in a dark chamber. Here, it is mounted the reflection probe that acquires the samples. In the upper part of the instrument is located the electronic control system based on a Raspberry Pi 4 as well as a touch screen and the VIS-NIR spectrometers. The halogen broadband light source is installed at the top rear part including two fans, placing the power electronics below it. A front view of the developed system is shown in Fig. 4.

Fig. 4. Front view of the prototype developed for spectral data acquisition from alfalfa samples.

Finally, software control of the entire device that includes the measurement protocol and the AI implementation was developed using Python. It was also required the implementation of an easy-to-use interface via the touchscreen including alerts and reminders, which guarantees correct use of the equipment by the operators. Obtained data is also stored in the system in order to improve the classification model in the future. The system is being tested now in-field with new samples acquired by the operators maintaining the success rate obtained during the training.

IV. CONCLUSIONS

A total of 426,994 alfalfa samples including continuous and discrete samples taken in visible and infrared have been obtained and analyzed. The classification of the samples based on supervised learning algorithm (neural network) permitted to successfully classify the alfalfa samples effectively in their quality (C1-C6) using both VIS and NIR information. It was observed that the deviation produced in the output classification was greater in the NIR, being the results obtained for the classification using the VIS data the ones that achieved the best results (greater than 72% in all cases).

Regarding the measurement methods, it has been observed that the results obtained for both continuous and discrete methods are similar or even higher in some cases for the continuous method (>74% versus >72% in continuous and discrete acquisition respectively), which facilitates the implementation of an automated prototype and speeds up the data acquisition.

The processing of the data and subsequent classification of the samples obtained through a system based on neural networks permitted to select the optimal configuration for the development and fabrication of a prototype. Developed system is currently used at the alfalfa reception point of the company NAFOSA S.A. The new system overrides the former human visual inspection system and removes the subjectivity of the operator in the decision at the same time that reduces the errors associated to the light conditions maintaining the same classification system scale.

Finally, it is important to remark that this work shows the potential of reflection spectroscopy aided by supervised neural networks in the agrifood field. This work consists of a first step towards future applications for product classification or quality determination in this field.

References

[1] N. Abu-Khalaf and M. Hmidat, "Visible/Near Infrared (VIS/NIR) spectroscopy as an optical sensor for evaluating olive oil quality," Comput. Electron. Agric., vol. 173, no. April, p. 105445, 2020.

[2] A. G. Mignani et al., "EAT-by-LIGHT: Fiber-optic and micro-optic devices for food quality and safety assessment," IEEE Sens. J., vol. 8, no. 7, pp. 1342–1354, 2008.

[3] A. G. Mignani and L. Ciaccheri, "Belgian beer mapping and digital fingerprinting using color and turbidity assessment," Opt. Sens. II, vol. 6189, p. 61892E, 2006.

[4] R. Ríos-Reina, S. M. Azcarate, J. M. Camiña, and R. M. Callejón, "Sensory and spectroscopic characterization of Argentinean wine and balsamic vinegars: A comparative study with European vinegars," Food Chem., vol. 323, no. April, p. 126791, 2020.

[5] S. Weng et al., "Rapid detection of adulteration of minced beef using Vis/NIR reflectance spectroscopy with multivariate methods," Spectrochim. Acta – Part A Mol. Biomol. Spectrosc., vol. 230, p. 118005, 2020.

[6] L. Xie, Y. Ying, T. Ying, H. Yu, and X. Fu, "Discrimination of transgenic tomatoes based on visible/near-infrared spectra," Anal. Chim. Acta, vol. 584, no. 2, pp. 379–384, 2007.

[7] O. Maria Porcu, "The Importance of UV-Vis Spectroscopy: Application in Food Products Characterization," Sch. J. Food Nutr., 2018.

[8] H. Cen and Y. He, "Theory and application of near infrared reflectance spectroscopy in determination of food quality," Trends Food Sci. Technol., 2007.

[9] H. Huang, H. Yu, H. Xu, and Y. Ying, "Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review," Journal of Food Engineering. 2008.

[10] B. G. Osborne, "Near-Infrared Spectroscopy in Food Analysis," Encycl. Anal. Chem., pp. 1–14, 2000.

[11] C. Petisco, B. García-Criado, L. García-Criado, B. R. Vázquez-de-Aldana, and A. García-Ciudad, "Quantitative analysis of chlorophyll and protein in alfalfa leaves using fiber-optic near-infrared spectroscopy," Commun. Soil Sci. Plant Anal., 2009.

[12] T. A. Liu, X. Y. Zhai, Q. Zhang, W. Lv, and W. C. Lu, "Using NIR with support vector regression to predict the crude protein of alfalfa," in Proceedings - 2016 International Conference on Information System and Artificial Intelligence, ISAI 2016, 2017.