Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 313–318
DOI: 10.15439/2024F5111
ISSN 2300-5963 ACSIS, Vol. 39

# An AI-empowered energy-efficient portable NIRS solution for precision agriculture: A pilot study on a citrus fruit

Giulia Cisotto
0000-0002-9554-9367
University of Milano-Bicocca,
Department of Informatics, Systems and Communication,
Viale Sarca, 336, 20126 Milan, Italy
Email: giulia.cisotto@unimib.it

Dagmawi Delelegn Tegegn
0000-0002-5031-7589
SeleTech Engineering Srl
Via Volturno, 37, 20861 Brugherio (Monza Brianza), Italy
Email: dag.tegegn@seletech.com

Alberto Zancanaro
0000-0002-5276-7030
University of Padova,
Department of Information Engineering,
Via Gradenigo, 6, 35136 Padova, Italy
Email: alberto.zancanaro.1@phd.unipd.it

Ivan Reguzzoni, Edoardo Lotti
SeleTech Engineering Srl
Via Volturno, 37, 20861 Brugherio (Monza Brianza), Italy
Email: {i.reguzzoni, edoardo.lotti}@seletech.com

Sara L. Manzoni
0000-0002-6406-536X
University of Milano-Bicocca,
Department of Informatics, Systems and Communication,
Viale Sarca, 336, 20126 Milan, Italy
Email: sara.manzoni@unimib.it

Italo F. Zoppis
0000-0001-7312-7123
University of Milano-Bicocca,
Department of Informatics, Systems and Communication,
Viale Sarca, 336, 20126 Milan, Italy
Email: italo.zoppis@unimib.it

***Abstract*—Smart agriculture has seen impressive progresses in monitoring the quality of the crop and early detecting the onset of pathogens. However, this is typically achieved through smart, expensive, and energy-demanding robots and autonomous systems. We propose an AI-empowered portable low-cost short-wave near-infrared spectroscopy (sw-NIRS) solution that allows non-destructive measurements from plants and vegetables. In this pilot study, we specifically targeted an orange fruit and showed that it is possible to classify its different parts through sw-NIRS in the range 1350-2150 nm by using AI models, exceeding 97% accuracy. Also, we explored the minimum amount of energy needed to reach such high classification performance. In the future, we aim to extend this investigation to other targets (e.g., bean plants), to develop AI architectures to more accurately model the physiological conditions of the target, and to create a network of sw-NIRS sensors to simultaneously monitor a large-scale crop.**

***Index Terms*—Near-infrared spectroscopy, machine learning, AI, chemometrics, energy efficient, green technology, smart agrifood, precision agriculture.**

## I. Introduction

SMART agri-food has recently seen tremendous developments thanks to new generation sensing, networking, and data analytics, i.e., ICT and artificial intelligence (AI), technologies.

This is allowing experts in the domain to quantitatively, continuously, and precisely monitor the conditions of the crops [1]. Near-infrared spectroscopy (NIRS) is one of the most popular techniques employed in the field, as it has already shown great potential in analyzing the quality and composition of foods [1], the maturity of fruits [2], [3] and crops [4], as well as the stress conditions of plants [5]. Another important advantage of NIRS is that this technology is available as portable devices (a very good review on the most recent handheld spectrometers can be found in [6]) with very fast scanning times (in the range of a few seconds) [7]. However, current portable NIRS devices suffer from some important limits, e.g., relatively large mass (over a few kilograms) [8], [9], relatively large sizes (e.g., a few tens of centimeters), no continuous acquisition modality, and a spectral resolution rarely below 2 nm. Here, we present a pilot study where a new handheld extremely lightweight but accurate NIRS spectrometer is used to acquire spectra from an orange fruit. With the complementary application of AI and machine learning (ML) modeling techniques [5], [10]–[12], we were able to classify different parts of the fruit with very high precision (classification accuracy over 97%), using very low energy.

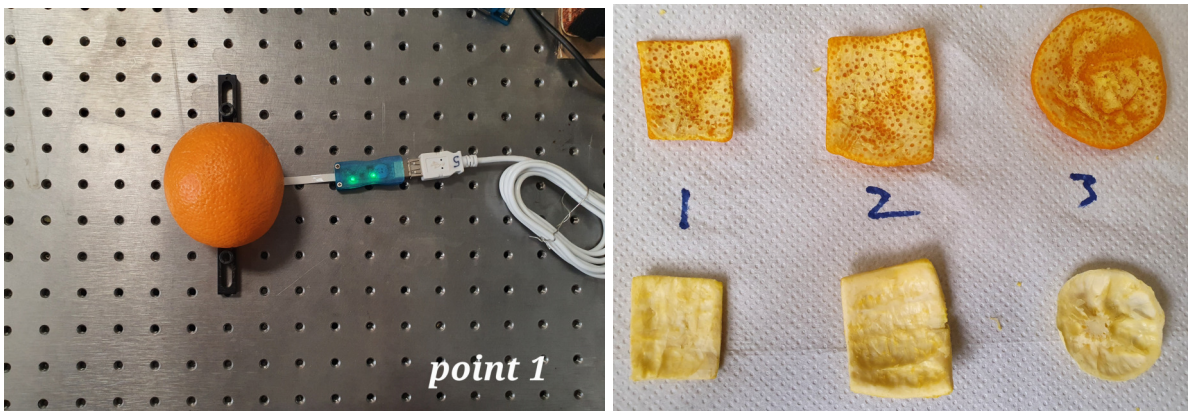


Fig. 1. Experimental setup and some representative targets. (a) *Point 1* for *Part 1* (whole orange). (b) *Points* 1 to 3 for *Part 3* (white layer) and *Part 4* (orange layer).

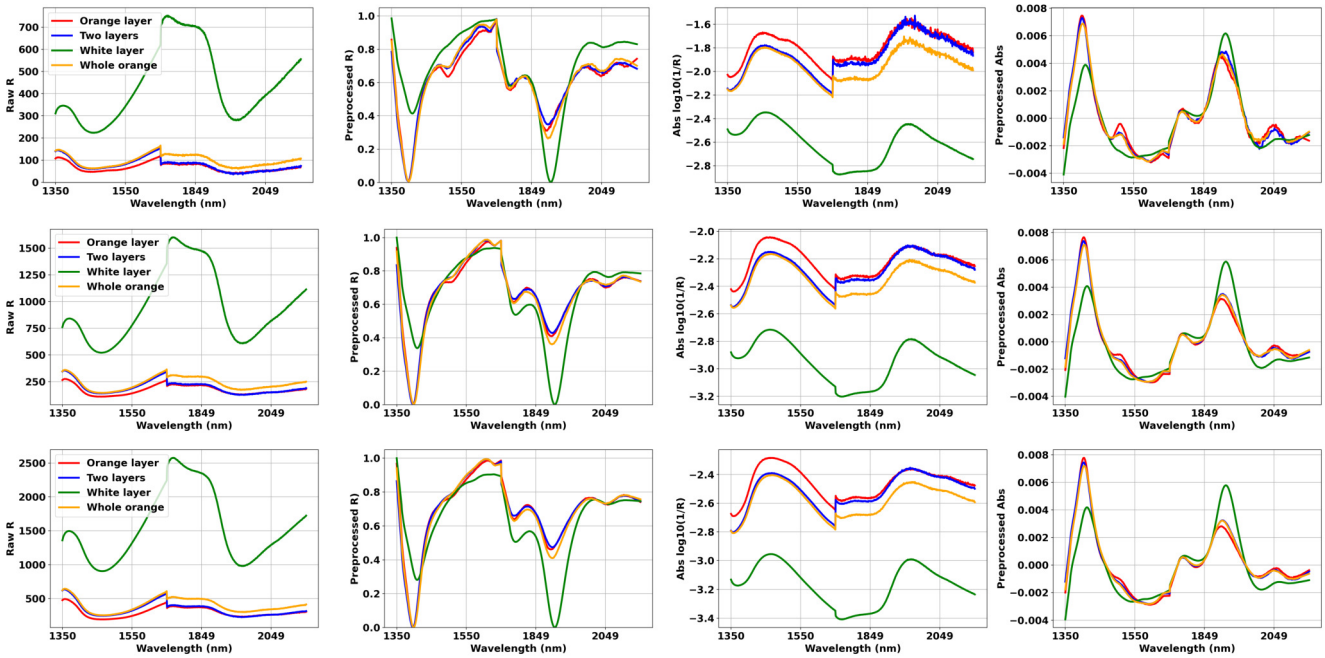


Fig. 2. Dataset in three single-lamp power settings at 100, 150, and 200.

## II. Materials and Methods

### A. Portable NIRS

In this study, we employed a portable near-IR spectrometer that is able to capture the reflected light components from the target in a wide range of wavelengths, i.e., from 1350 nm to 2150 nm (technically defined as *short-wave* NIRS), with very high resolution of 1 nm. It is extremely lightweight (20 g) and small (45x25x13 mm).

Our spectrometer utilizes two microelectromechanical systems (MEMS) spectral sensors by Hamamatsu[1]. These sensors generate two independent spectra: one captures the spectrum from 1350 nm to 1650 nm, while the second captures from 1750 nm to 2150 nm. Although the sensors operate independently, the data from both of them are processed together in our analysis pipeline. This dual-sensor approach allows us to cover a wide spectral range (1350-2150 nm). Its power consumption is primarily determined by the lamp's power. An Oshino lamp, integrated into the device, consumes approximately 60 mW at lamp power 100 and 270 mW at lamp power 250. Additionally, the SoC system and other operations consume 120 mW. This results in a total power consumption falling within the range of approximately 180 mW to 390 mW, which is exceptionally low for these handheld spectrometers [9].

This energy-efficient profile enables users to perform precise

[1]The datasheets of these sensors (C14272, C14273) can be found at https://www.hamamatsu.com/eu/en/product/optical-sensors/spectrometers/mems-fpi-spectrum-sensor.html

spectral analysis without significant power demands, enhancing the spectrometer's practicality and versatility for a wide range of field and laboratory applications. The cost for such a hardware setup is very low (less than $500$ $ in the current device configuration), compared to many other available NIRS (e.g., hyperspectral) solutions (more than $1000$ $, depending on the specific features). Moreover, the device is prepared to embed a Bluetooth Low Energy (BLE) chipset for wireless transmissions (available in a later version). The acquisition time is very short, i.e., less than 2 s is enough to scan the entire range of wavelengths, making this device suitable for on-the-fly measurements of different types of targets, from plants to fruit, to animals [1].

### B. Experimental protocol

Four different *parts* of an orange fruit were targeted for the NIRS measurement: (1) the whole orange fruit, (2) a two-layer target formed by the two outermost layers of the orange fruit, i.e., the orange and the white layers, (3) the white layer, and (4) the orange outermost layer. Spectra were collected from three different *points* on each part of the fruit. Fig. 1 shows the experimental setup and some representative targets.

Sixteen different lamp powers were available for the acquisition: from the value of $100$ to the value of $250$, with a $10$-width step. Ten repeated measures were collected, in a continuous modality, on the same experimental condition, i.e., the combination of *part-point-lamp power*. The overall dataset finally included $1920$ samples, with $702$ features each (i.e., each feature representing one wavelength).

It is worth noting that we selected the three different *points* of every *part* with no rigorous localization criterion. This might have led to an increased *intra-part* variability but, at the same time, allowed us to prove that a ML model trained on these NIRS data is still able to distinguish across different parts, making our contribution closer to real-world applications.

### C. Data preparation and pre-processing

From our previous investigations [1], we decided to use the standard normal variate (SNV) which consists in normalizing every spectrum by removing its own mean and dividing by its own standard deviation. This method aims to reduce the multiplicative effects of scattering and particle size, and allows to reduce the differences in the global intensities of the signals [13].

The Savitsky-Golay filter (SGF) is one of the most commonly used pre-processing steps in spectrometry and it consists of a 1-D filter that fits a polynomial function with degree $p$ to a piece of data of length $w$. Often, first or second-order derivative is computed on the data before applying the filter. Based on our previous empirical investigations, we applied SGF on the reflectance values, setting $p = 2$, and $w = 30$ for MEMS1 while $w = 50$ for MEMS2 (to cope with the higher noise level), and using the first-order derivative. Thus, the first-order derivative emphasizes the dynamic changes in the reflectance spectrum, while the SGF smoothes up irrelevant small peaks. Then, min-max normalization was applied to the filtered reflectance data to limit their values in the $[0, 1]$ range.

In this study, we investigated both the performance of our system when using all lamp powers (namely, *multi-lamp power setting*), or selecting one specific lamp power (namely, *single-lamp power setting*). In the latter case, during pre-processing, we extracted values obtained from the selected lamp power and performed the other operations on the reduced dataset, in the same way as in the case of the multi-lamp power setting. In both cases, we also investigated the possibility to reduce the dimensionality of the dataset (i.e., from the 702 dimensions) using principal component analysis (PCA). We decomposed the entire dataset in order to obtain a minimum number of principal components that explain the most variance in the dataset.

For visual inspection purposes, we also computed the absorbance spectra from the reflectance ones by computing $A = \log_{10}(1/R)$. The absorbance values were further filtered with a SGF with the same parameters values as the reflectance. All processing steps were implemented in Python $3.9$, as available in the Jupyter environment in Google Colab.

### D. Numerical experiments

Both for the *multi-lamp power setting* and the *single-lamp power setting*, we investigated the possibility of classifying the fruit parts, e.g., its different layers, from sw-NIRS data. Also, by visually inspecting the results of PCA, we observed the amount of separability among different classes while reducing the dataset dimensionality (originally set to 702) to the first 2 principal components.

Following the literature mainstream [4], we selected support vector machine (SVM) for the $4$-class classification task, as it represents one of the most successful ML models in NIRS analysis. The Python class $sklearn.svm.SVC$ was employed in the implementation of the classification task. This library is based on the $libsvm$ package [14] that finds the best classification solution using a one-vs-one approach (default solver: C-SVC) in multi-class classification problems. During training, we applied a grid-search parameter optimization with the following values: kernel = [linear, radial basis function (rbf), polynomial], $C = [0.01, 1, 10, 100]$, $\gamma = [0.01, 0.1, 1, 10]$, where $C$ is the regularization parameter that allows having a certain number of mis-classified samples, while $\gamma$ represents the influence of far away samples in the computation of the separation hyperplane. We left the $degree$ parameter for the polynomial kernel to the default value of 3. To train and validate the SVM model, we randomly selected $70\%$ of the dataset, while keeping the remaining $30\%$ for the test phase. During training, $5$-fold cross-validation was applied.

We built the SVM models using pre-processed reflectance $R$ values or, alternatively, using the first 2 principal components. However, in most cases, the models relying on the $R$ values returned the best results. Thus, in the following, we show the classification performance obtained from the pre-processed R values, while the PCA results are used for visualization purposes, only.

To support the global effort of the scientific community in the direction of full reproducibility of research [15], [16], we share our Colab file at [17] and the dataset at [18].

## III. Results and Discussion

*Visual inspection:* Fig. 2 shows the dataset obtained by averaging 10 repeated measurements from three different points of every orange's part in three single-lamp power settings. It includes the raw reflectance spectrum, the pre-processed reflectance spectrum (i.e., as used for the subsequent analysis), the raw absorbance spectrum, and the pre-processed absorbance spectrum (as described in Section II-C).

We could observe that all spectra are consistent with previous literature [4], [8], [19]. Then, it can be noted that the white layer shows the most reflective intensity (both MEMS ranges), while the other three parts are much more similar to each other. Nevertheless, as one might expect, the whole orange and the two-layer targets produced spectra in between the white layer and the orange layer, with some slight differences in the two MEMS ranges.

*Multi-lamp power setting study:* When all measurements performed with any value of lamp power are included in the analysis, we found that the most explained variance is accounted for by the first 2 principal components, which represent 75% and 18% total variance in the dataset, respectively.

Furthermore, the best SVM classifier model (built on the pre-processed R values) was obtained with $C = 10$, $\gamma = 10$, and $rbf$ kernel and reached 99.3% accuracy during the test.

Fig. 3 shows the normalized confusion matrix for the 4 classes in the multi-lamp power setting.

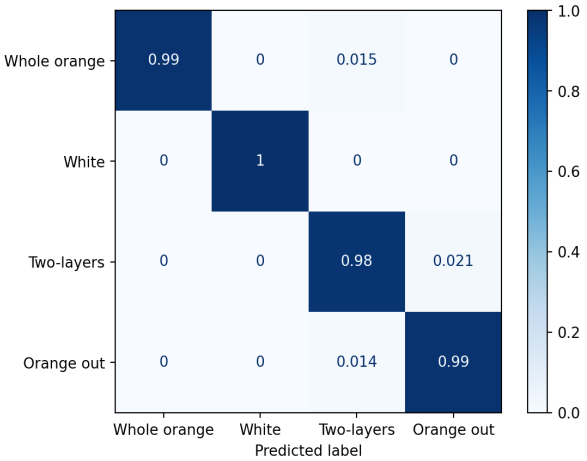


Fig. 3. Multi-lamp power setting: normalized confusion matrix for the 4 classes.

However, this setting leads to a high energy consumption that could reduce the battery life of the device. Thus, we performed a second stage of experiments while limiting the lamp power, i.e., the energy consumption of the device. We extensively tested the classification performance with every single lamp power, in order to find the best trade-off between low energy usage and satisfactory classification accuracy values.


Fig. 4. Single-lamp power setting: normalized confusion matrix for the 4 classes for three different values of the lamp power (100, 150, 200, respectively).

*Single-lamp power setting study:* Fig. 4 shows the normalized confusion matrices for three different single-lamp powers, i.e., 100, 150, and 200. We can observe that, as the lamp power increases, the accuracy in the classification of the four orange's parts improves.

On the other hand, we also applied PCA to the three above configurations and noticed that for higher lamp powers, more variance is explained by the first principal components, only. Table I shows the explained variances for the first five principal components for the three different lamp powers.

TABLE I
EXPLAINED VARIANCE FOR THE FIRST 5 PRINCIPAL COMPONENTS FOR THREE DIFFERENT LAMP POWERS.

| Lamp power | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| 100 | 0.73 | 0.13 | 0.04 | 0.02 | 0.01 |
| 150 | 0.83 | 0.11 | 0.03 | 0.01 | 0 |
| 200 | 0.89 | 0.07 | 0.02 | 0.01 | 0 |

One might conclude that the lamp power acts as a relevant factor that increases the data variance. Including all measurements, regardless the lamp power, makes the variance spread more onto two principal components. However, the classification results are very satisfactory. On the other hand, it seems also possible that using a low-power lamp setting (i.e., 100) leads variance to be spread over more than one principal component. However, in this case, the classification performance decrease, possibly due to an increase of the variance that is not related to the target's characteristics.

Finally, Fig. 5 reports all classification accuracies, found during the test phase (over the pre-processed R values), for every lamp power value, i.e., from 100 to 250.

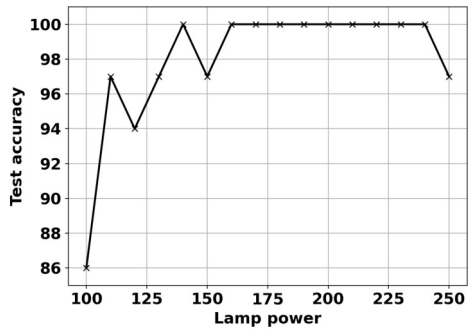


Fig. 5. Classification accuracy for every single-lamp power.

We could observe that, even with a single lamp power, it is possible to obtain very high classification accuracy values in the separation of the four orange fruit parts (chance level is 25%). Our portable spectrometer includes power-saving features when not in use, and switches to a higher power mode when acquiring spectra continuously. However, with this work we suggest that by empowering NIRS with AI (specifically, ML modeling), we could fasten the acquisition times and lower the energy consumption during the operating phase, while obtaining very high classification performance. Specifically, from Fig. 5, one can choose to set the lamp power, for example 150, to obtain consistently good classification, consistently reduce the energy in the acquisition step and, at the same time, reduce the redundant variance not related to the target itself, as discussed above.

## IV. CONCLUSIONS AND FUTURE PERSPECTIVES

The results presented in this study are still preliminary but made us prove the reliability and the advantages of an AI-empowered NIRS solution for the smart agri-food domain. In the future, we will extend this investigation to more complex scenarios: we will use the NIRS technology to monitor other target plants (e.g., bean and pothos plants), fruits, or to analyze organic compounds. Also, we can leverage the Bluetooth connection available on our spectrometer to deploy a network of wireless NIRS sensors [20] to precisely and promptly extract relevant information about the physiological conditions of plants in large-scale crops and new generation smart greenhouses. Furthermore, the current study was conducted in a semi-controlled lab environment as our aim was to perform a pilot study with our prototype. For real-world outdoor deployments, we will operate ad-hoc calibration based on the surrounding ambient light and we will perform a systematic feasibility study. Finally, we will the SVM model performance with other ML models and AI architectures (e.g., convolutional neural networks) [12]. Although there are other well-established solutions for precision agricolture in the market (e.g., RGB cameras), they often operate in different wavelength ranges, making direct comparisons with our system difficult. Therefore, future additional efforts will be dedicated to compare our system with more similar NIRS devices, currently not available in the market. Additionally, our system could be reviewed as a complimentary solution for satellite-based crop monitoring techniques [21] or drones-based weed mapping [22]. Then, to increase the sustainability of our system, we will investigate the actual amount of power consumption in the final NIRS product. However, with this first prototype, we were able to show the combination of hardware energy-saving features (e.g., battery size, lmap type, and other sensor configurations) with AI-based methods can significantly reduce the overall need of energy, while ensuring an effective monitoring of crops and plants for long periods of time.

## ACKNOWLEDGMENT

This work was partially supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy. AZ is also supported by PON 2014-2020 action IV.4 and GC by action IV.6 funded by the MUR.

## REFERENCES

[1] D. D. Tegegn, "Process of analyzing organic materials, based on processing of near-nfrared spectra through advanced methods (PhD Thesis)," 2023.

[2] P. Rodríguez, J. Villamizar, L. Londoño, T. Tran, and F. Davrieux, "Quantification of dry matter content in hass avocado by near-infrared spectroscopy (NIRS) scanning different fruit zones," *Plants*, vol. 12, no. 17, p. 3135, 2023.

[3] K. Ncama, L. S. Magwaza, A. Mditshwa, and S. Z. Tesfay, "Application of visible to near-infrared spectroscopy for non-destructive assessment of quality parameters of fruit," *Infrared Spectroscopy-Principles, Advances, and Applications*, 2018.

[4] A. M. Cavaco, D. Passos, R. M. Pires, M. D. Antunes, and R. Guerra, "Nondestructive assessment of citrus fruit quality and ripening by visible-near infrared reflectance spectroscopy," *Citrus-Research, Development and Biotechnology*, p. 95970, 2021.

[5] A. Zancanaro, G. Cisotto, D. D. Tegegn, S. L. Manzoni, I. Reguzzoni, E. Lotti, and I. Zoppis, "Variational autoencoder for early stress detection in smart agriculture: A pilot study," in *2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. IEEE, 2022, pp. 126–130.

[6] J. Müller-Maatsch and S. M. van Ruth, "Handheld devices for food authentication and their applications: A review," *Foods*, vol. 10, no. 12, p. 2901, 2021.

[7] K. R. Borba, P. C. Spricigo, D. P. Aykas, M. C. Mitsuyuki, L. A. Colnago, and M. D. Ferreira, "Non-invasive quantification of vitamin C, citric acid, and sugar in Valência oranges using infrared spectroscopies," *Journal of Food Science and Technology*, vol. 58, pp. 731–738, 2021.

[8] J. A. Cayuela and C. Weiland, "Intact orange quality prediction with two portable NIR spectrometers," *Postharvest Biology and Technology*, vol. 58, no. 2, pp. 113–120, 2010.

[9] "Spectral Evolution NaturaSpec portable spectroradiometer," https://spectralevolution.com/products/hardware/field-portable-spectroradiometers-for-remote-sensing/naturaspec-portable-spectroradiometer/, accessed: 2023-09-25.

[10] J. Martins, R. Guerra, R. Pires, M. Antunes, T. Panagopoulos, A. Brázio, A. Afonso, L. Silva, M. Lucas, and A. Cavaco, "SpectraNet–53: A deep residual learning architecture for predicting soluble solids content with VIS-NIR spectroscopy," *Computers and Electronics in Agriculture*, vol. 197, p. 106945, 2022.

[11] Seletech Engineering Srl, "AI-empowered monitoring systems," https://lnx.seletech.com/index.php/it/prodotti-e-progetti/intelligenza-artificiale/, accessed: 2023-09-25.

[12] W. Suphamitmongkol, G. Nie, R. Liu, S. Kasemsumran, and Y. Shi, "An alternative approach for the classification of orange varieties based on near infrared spectroscopy," *Computers and electronics in agriculture*, vol. 91, pp. 87–93, 2013.

[13] J. A. Prananto, B. Minasny, and T. Weaver, "Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues," *Advances in agronomy*, vol. 164, pp. 1–49, 2020.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

[15] F. Cabitza and A. Campagner, "The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical ai studies," p. 104510, 2021.

[16] K. Choudhary, D. Wines, K. Li, K. F. Garrity, V. Gupta, A. H. Romero, J. T. Krogel, K. Saritas, A. Fuhr, P. Ganesh *et al.*, "Jarvis-leaderboard: a large scale benchmark of materials design methods," *npj Computational Materials*, vol. 10, no. 1, p. 93, 2024.

[17] G. Cisotto, "Python code associated with this publication at colab," https://colab.research.google.com/drive/14xU8u1Ao3smfnc_epTcIP9AS8etUjG8V?usp=sharing, last access: 2024-07-18.

[18] G. Cisotto, D. D. Tegegn, I. Reguzzoni, and E. Lotti, "NIRS dataset associated with this publication," https://github.com/CisottoGiulia/PON22-AI-NIRS-AgriFood.

[19] A. M. Cavaco, R. Pires, M. D. Antunes, T. Panagopoulos, A. Brázio, A. M. Afonso, L. Silva, M. R. Lucas, B. Cadeiras, S. P. Cruz *et al.*, "Validation of short wave near infrared calibration models for the quality and ripening of Newhall orange on tree across years and orchards," *Postharvest Biology and Technology*, vol. 141, pp. 86–97, 2018.

[20] A. Zancanaro, G. Cisotto, and L. Badia, "Modeling value of information in remote sensing from correlated sources," *Computer Communications*, vol. 203, pp. 289–297, 2023.

[21] P. Bertellini, G. D'Addese, G. Franchini, S. Parisi, C. Scribano, D. Zanirato, and M. Bertogna, "Binary classification of agricultural crops using sentinel satellite data and machine learning techniques," in *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE, 2023, pp. 859–864.

[22] G. Castellano, P. De Marinis, and G. Vessio, "Applying knowledge distillation to improve weed mapping with drones," in *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE, 2023, pp. 393–400.