Department of
Computer Science, System and Communication

PhD program in Computer Science   Cycle XXXV

Process of analyzing organic materials, based on processing of near infrared spectra through advanced methods

Tegegn Dagmawi Delelegn

772827

Tutor: Prof. Vizzari Giuseppe

Tutor: Dott. Lotti Edoardo

Supervisor: Prof. Zoppis Italo Francesco

Co-supervisor: Prof. Manzoni Sara Lucia

Coordinator: Prof. Leonardo Mariani

ACADEMIC YEAR   2022/23

TEGEGN DAGMAWI DELELEGN

APPLICATIONS OF HANDHELD NEAR-INFRARED SENSOR IN PLANT, ANIMAL AND HUMAN WELLNESS.

No research without action, no action without research.

— Kurt Lewin

If we knew what we were doing, it would not be called research,
would it?

— Albert Einstein

ABSTRACT

Near-infrared spectroscopy is a mature technique continuing to demonstrate steady progress. This is thanks to cutting-edge developments of new handheld spectrometers. These spectrometers can quickly generate a high volume of spectral data, requiring advanced methodologies such as artificial intelligence, big data, and deep learning to decipher the hidden content of the spectral data. This experimental research focuses on a handheld spectrometer based on MEMS technology and its application in various fields. We show that developing NIR models can be challenging upon application. We show the use case of this technology in the healthcare area and the industry (food). Plants and Fruits have been subjected to our research, such as Pothos and orange fruit. Organic powders have also found an application in our research. Furthermore, we extend our experimental research into animal wellness, particularly in dairy cow farms. We used advanced methodologies in specific use cases. We applied 1D-CNN when predicting quantities of content in mixtures of organic powder where a high volume of data was collected or a VAE model to predict water conditions as a health anomaly detection in the Pothos plant. We show spectra analysis and preprocessing approaches to remove signal artifacts. We implement different effective preprocessing methods: scaling methods, spectral derivatives, and the Savitsky Golay filter. We present the results of different data scaling and transformation methods using a convolutional neural network.

PUBLICATIONS

These are some publications related to this research work:

[1] Dagmawi Delelegn Tegegn Leonardo Badia Sara L. Manzoni Ivan Reguzzoni Edoardo Lotti Italo Zoppis Alberto Zancanaro Giulia Cisotto†. «Variational Autoencoder for Early Stress Detection in Smart Agriculture: A Pilot Study.» In: *Accepted: IEEE metroagrifor*.

[2] Manzoni Sara Reguzzoni Ivan Lotti Edoardo Dagmawi Delelegn Tegegn Zoppis Italo Francesco. «Improving Prediction of Organic Mixture Powders Using Near-Infrared Spectra.» In: *In submission, Frontiers* (2023).

[3] Tegegn Dagmawi Delelegn, Italo Francesco Zoppis, Sara Manzoni, Alessio Mognato, Ivan Reguzzoni, and Edoardo Lotti. «Rapid Analysis of Powders Based on Deep Learning, Near-Infrared and Derivative Spectroscopy.» In: *CEUR Workshop Proceedings*. Vol. 3102. 2021.

[4] Dagmawi Delelegn Tegegn., Italo Zoppis., Sara Manzoni., Cezar Sas., and Edoardo Lotti. «Convolutional Neural Networks for Quantitative Prediction of Different Organic Materials using Near-Infrared Spectrum.» In: *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOSIGNALS*, INSTICC. SciTePress, 2021, pp. 169–176. ISBN: 978-989-758-490-9. DOI: 10.5220/0010244101690176.

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— Donald E. Knuth [48]

ACKNOWLEDGMENTS

"It is not the destination that matters, but the journey." As I reflect on all the individuals and experiences that have brought me to this point, I am filled with immense gratitude. Despite the challenges, laughter, and long nights, I was never alone on this journey.

I am deeply grateful to Professor Italo Francesco Zoppis and his group for their invaluable support and feedback. I would like to extend a special thank you to Ivan Reguzzoni, who has been a constant source of encouragement and provided me with the foundation for my research. I also want to express my appreciation to Dr. Edoardo Lotti for his unwavering support, including actively shaping my research plans and providing financial assistance. I am grateful to have had the opportunity to work with Giulia Cisotto as well.

Most importantly, my sincerest gratitude goes to my mother, Nigist, who has always been by my side, no matter the distance. I am also grateful to my sister, Mariamawit, and brother, Mikael, for their unwavering support.

I would like to extend my gratitude to my second family, Massimo, Michele, and Aziz, for their love, kindness, and friendship throughout this journey.

Finally, I want to thank all my dear friends, who have been with me every step of the way. Your positive energy, companionship, and laughter have been a source of comfort during both the good times and the tough. And let's not forget Bill Clinton, my fat cat, who always provided a stress-relieving hug.

Dedicated to all the individuals who have helped me during my years of study. I love you all.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

INTRODUCTION

INTRODUCTION

1.1 BACKGROUND

Conventional chemical analysis has long been the only reference procedure for the analytical determinations of specific macro and micro markers of a particular product. Nowadays, several techniques have been developed to be real-time but also non-invasive, non-ionizing, portable, and low-cost, providing quantitative information quickly. Near Infrared Spectroscopy (NIRS) presents these advantages and is used in many fields. NIRS is a spectroscopic method that uses the near-infrared region of the electromagnetic spectrum (from 700 nm to 2500 nm). It is used in various fields, including chemistry, physics, and engineering. NIRS is based on the principle that molecules absorb different wavelengths of light to different degrees. By measuring light absorption at different wavelengths, it is possible to identify the molecules present and determine their concentrations. NIRS is a non-destructive and non-invasive technique that can measure the concentrations of various molecules in various samples, including liquids, gases, and solids. NIRS has several advantages over other spectroscopic methods, including its ability to measure a wide range of molecules, high sensitivity, and low cost. NIRS is used in various applications, including food and beverage analysis, pharmaceutical analysis, and environmental monitoring (knowing how to design chemical and chemical manufacturing processes with little or no risk to human health or the environment) spectroscopy. NIRS is used in various applications, including medical diagnosis, food analysis, and industrial process control.

Medical applications of NIRS include the detection of cancer, the monitoring of blood oxygenation, and the assessment of tissue viability. It is a non-invasive optical imaging technique to monitor tissue oxygen status [35]. This technique is low-cost and user-friendly compared to other neuroimaging methodologies, so it is ideally suited for this area.

The industrial applications of NIRS include monitoring process variables such as temperature, pressure, and composition. According to various literature studies, NIR spectroscopy can be used to determine the main components of milk, meat, fish, eggs, dairy products, and more. It is also used to detect possible adulterations and food fraud, discriminate products obtained in particular production areas, and discriminate fish and mollusks raised or fished. One of the main advantages of this technology is its ability to deliver results quickly

and accurately. However, NIR technology requires an updated and representative calibration based on usual laboratory analyses as a secondary analytical system. The potential of this technology is truly vast, many of which are still unexplored.

Food analysis applications of NIRS include the determination of the moisture content of food, the analysis of fat and protein content, and the detection of adulterants. A recent study carried out by a group of Italian researchers [7] illustrates and discusses the main applications of this technology in the food sector.

Today, the main objective of sensor development is to develop increasingly less bulky instruments (miniaturized, portable) equipped with probes capable of analyzing a wide range of analytes and designing instruments that can be easily incorporated into production lines and processing machines.

1.2   MOTIVATIONS

This thesis aims to bring to light the application of a MEMS-based portable infrared spectrometer in the analytical field and, more particularly, organic chemistry by combining a portable handheld spectrometer and deep learning and machine learning methods. The electromagnetic spectrum is divided into three zones: shortwave infrared (NIR), which spans from 700 to 2000 nm; medium wave infrared (MIR), which spans from 2000 to 4000 nm (MIR); and long-wave infrared (FIR), which spans from 4000 to 10000000 nm. An infrared emitter (light) spreads infrared energy to target surfaces that readily absorb heat. Infrared irradiation involves direct heating. Therefore its effectiveness depends on the source-target line of the light, which is a tungsten filament sealed in a quartz envelope. It is important to note that the thin filament, due to its small mass, is very reactive to the voltage applied. Consequently, a high correlation between the NIR heat output and the corresponding applied voltage and the interaction of NIR consequently, the matter can be considered from the classical mechanical model for a diatomic molecule. Spectrometers that employ the NIR technology have long been expensive, complicated, and extensive, but the availability of a new generation of micro-electromechanical system (MEMS) based spectrometers changed the game. These new sensors have lower costs and come in miniaturized sizes. Now, it should not come as a surprise that the NIR technology is used in a vast field of application. Many industries now use NIR technology for qualitative and quantitative purposes in agricultural products, industrial food products, precision agriculture, polymer quality classification, fuel quality control, fuel production process, petroleum characterization, environmental products, industrial products, petroleum characterization, and polymer. This thesis aims to offer an overview of works where we have employed our MEMS-based NIR spectrometer. When available, we compared our use of the spectrometer with other works in our field. Thus, one of the goals of this thesis is to offer a comprehensive overview of the potential application of our MEMS-based NIR sensory combined with machine learning methods in agriculture, healthcare, and industry fields.

1.3   CONTRIBUTIONS

Miniaturized NIR spectrometers are now available at more affordable prices than conventional spectrometers, but their performances have yet to be studied. The field of application on infrared sensors is vast, as seen in the literature. Especially for analyzing organic matter, which can be food, human skin, or other items, the near-infrared range offers substantial advantages, most of all, a suitable penetration depth and relevant spectral information. However, the spectrom-

eters are usually very sophisticated and laboratory use only, unlike the one used for this study, a portable miniaturized sensor with a growing trend. The next generations of mobile phones will contain spectral analyzers. Different concepts and system designs compete in this ultra-high volume market. The present work explored the applicability of our MEMS-based NIR infrared spectrometer and spectral data analysis and prediction models. This study concerns how the NIR technology of MEMS bases sensors functions combined with machine learning and deep learning techniques, how the measurement system works, what can be achieved with it, and how the development should be continued. The measurements are easy enough to perform without the expertise of laboratory personnel, creating a massive amount of data in a short time, making prediction models a necessity in this industry. The study aimed to analyze the measurement system's performance from different points of view: data, results, usability, and how the performance could be enhanced. The study also presents a methodological approach and findings in the industry and medical fields.

1.4 ORGANIZATION OF THIS THESIS

This thesis work is organized as follows. Chapter 2 provides a general overview of the miniaturized spectrometer that is used to assess the possible application of near-infrared in various fields. Chapter 3 introduces the field of spectroscopy, and the main characteristics are described. Furthermore, it includes state-of-the-art near-infrared spectroscopy applications. Chapter 4 presents approaches made in the healthcare area, with a particular emphasis on issues analyzing dynamic organic matters. It is shown the use case in plant leaf water stress detections. Chapter 5 includes the application of the spectrometer in the industrial field, such as determining the composition percentage of mixed powders using convolutional neural networks (CNN). A Spectra processing pipeline is also introduced, which includes signal pre-processing and normalization. Each chapter provides a brief literature review of the topic-related works, describes the proposed strategies, and reports some experimental results.

Part II

THE SPECTROMETER

INTRODUCTION

2.1 MICROELECTROMECHANICAL SYSTEMS (MEMS)

Microelectromechanical system (MEMS)-based sensors are essential in many modern applications, from consumer electronics to the automobile sector. Technology enables the creation of movable structures that may be employed as actuators and transducers. MEMS emerged from the silicon-based microelectronics sector.

HISTORY   Few imagined that Charles Smith's 1954 Physical Review publication on stress-sensitive phenomena in silicon and germanium would launch a multi-billion dollar industry. After micro machining and silicon processing technologies improved in the early 1970s, a handful of forward-thinking, essentially American companies produced pressure sensors with non-planar diaphragms, the first actual MEMS sensors. Since then, MEMS technology has produced a variety of compact, durable, and often inexpensive, high-performance sensors that respond to numerous physical variables (Table 1) and some gases, chemical species, and biological quantities [9].

Variable	Type of sensor
Pressure	Piezoresistive and capacitive pressure sensors
Vibration, shock, motion	Accelerometers
Position, rotation	Gyroscopes, yaw-rate sensors
Inclination, an	Inclinometers
Strain	Semiconductor strain gauges
Infrared-radiation	Microbolometers
Sound	Microphones
Flow	Thermal mass flow sensors

Table 1: MEMS Sensors for physical variables

Table 2 shows three waves of MEMS sensor commercialization [9], which is an oversimplification. They are used in process, petrochemicals, building services, power generation, defense, aerospace, and healthcare, as well as in computers, phones, digital cameras, game consoles, and autos. This article describes contemporary MEMS sensor applications for detecting physical variables and gases .

Most pressure sensors today are MEMS-based, unlike most other physical variables, where many technologies coexist. This is due to several factors:

	Dates	User-industries/early products
Wave 1	1970s onwards	Low volume uses of high cost products in industry and aerospace
		Pressure sensors, strain gauges, accelerometers
Wave 2	1980s onwards	High volume uses of low cost products by the automotive industry
		Pressure sensors, accelerometers, yaw sensors, air flow sensors
Wave 3	Early 2000s onwards	Ultra-high volume uses of very low cost products in consumer electronics
		Accelerometers, gyroscopes, microphones

Table 2: Commercialisation of MEMS sensors

- inherent technical features that allow outstanding performance flexibility to accept a wide range of pressures by adjusting diaphragm diameters

- the high-volume batch fabrication that provides cheap unit prices, as requested by several industries

MEMS pressure sensors had not profited from the consumer electronics sector's sensor consumption expansion until Samsung's use of them in the Galaxy S4 [79] and other smartphone models and Sony Mobile's use in two 2012 models. In 2012, smartphones used 82 million MEMS pressure sensors. Robotic surgery and medical implants demand even smaller sensors, although these sensors are based on traditional MEMS technology. Singapore and South Korean researchers developed a sensor that uses a combination of materials to miniaturize silicon diaphragm devices.

MEMS microphones have also profited from consumer electronics. These were once rare but are now found in smartphones, laptops, tablets, video cameras, and other devices. Apple introduced MEMS microphones in the iPod nano 5 and iPhone 4 in 2009. Smartphones now have two MEMS microphones. Handset manufacturers are exploring using three or more microphones for noise reduction and video audio recording. As cell phones have become more multifunctional, being used for music, video, and other applications, the emphasis on crisper sound has increased. Handset manufacturers can differentiate their phones with acoustics. Apple launched Siri voice command with the iPhone 4S, which was later included in the iPhone 5 and other Apple products like the fifth-generation iPod touch and the fourth-generation tablet.

Accelerometers have been used in MEMS sensor commercialization since 1970, as seen in Table 2. Since the first iPhone made accelerometers famous for auto-screen rotation and gesture-based command capabilities, the consumer electronics industry has deployed millions to activate airbags and signal stability control systems, among other purposes. MEMS accelerometer technology has evolved as users want smaller packages, lower power consumption, and more functionality.

MEMS has had a modest impact on gas sensors compared to physical sensing, and few basic gas sensing approaches have profited. Semiconducting metal oxide (MO) sensors used in millions of industrial

applications are the most important use [44]. MEMS substrates with a thin silicon nitride membrane support a sensing material film in the latest design. This method produces miniaturized sensors with lower power consumption, faster response times, and lower unit costs in high volumes. Micro-hotplates often have a temperature sensor.

MOEMS gas sensing are micro engineered silicon emitters for NDIR gas sensors are a growing market [106]. Most NDIR products use miniaturized filament bulbs as optical sources, which are cheap but limit output to 5 μm. Many essential molecules, such as ammonia, alcohol vapor, refrigerants, nitrogen oxides, and sulfur dioxide, have absorption peaks in the 6-14 μm region, limiting the gases that can be detected. Cambridge CMOS Sensors, a UK company pioneering MOEMS technology for this application, produces silicon mid IR sources that operate over the wavelength range 2-14 μm and feature integrated FET drivers to control heater temperature by adjusting the gate or supply voltage, eliminating the need for an off-chip driver. The company makes MO gas sensor MEMS micro-hotplates. MOEMS IR technology can detect methane, as has conventional NDIR. Two MEMS-based methane detection methods and a novel catalytic sensing technology are now available. Since the 1960s, MEMS technology has produced high-performance sensors for various physical variables at cheap unit prices, enabling high-volume applications in price-sensitive markets. Automotive and consumer electronics applications will grow. MEMS is progressing in gas sensing, but high-volume applications will likely spur its use.

2.2 SPECTROMETERS: DEFINITIONS AND CLASSIFICATIONS

Spectrometers are specialized laboratory equipment used by professional spectral analysts and researchers. However, many spectroscopic applications, such as mineral exploration, safety screening, and drug testing, can also benefit from field measurements that are taken outside of the laboratory. To accommodate this, compact and portable spectrometers have been developed that are battery-powered and equipped with light sources. These handheld spectrometers also have advanced data processing capabilities that can assist users without spectroscopic expertise to perform measurements and analyze relevant information about the composition and content of materials.

Industrial and civilian applications often require high accuracy and reliability, which allows manufacturers to sell their spectrometers at high prices (typically $\geqslant$ 1000\$). However, applications such as food quality and adulteration monitoring, anti-counterfeiting, and personal health tracking require lower unit prices ($\leqslant$ 100\$), smaller size, and ease of use.

Spectrometers and hyperspectral imaging devices are a novel type of sensor technology that have the potential to enable new appli-

cations beyond the capabilities of traditional MEMS sensors. Spectroscopy involves analyzing emitted, transmitted, or reflected visible and infrared light to determine the atomic and vibrational structures of matter. This technique is widely used in the chemical industry for quality control and in research laboratories where high spectral resolution is required. However, these types of spectrometers can be large and expensive.

The late century saw the development of miniaturized spectroscopic devices [21]. However, during the last decade, research efforts have increased significantly, intending to open new spectroscopic applications outside laboratories, where both size and cost rather than high-end precision matter as shown in Figure 1.
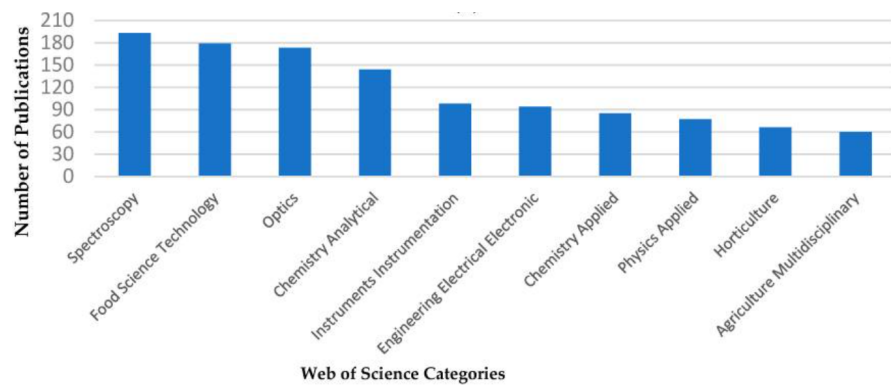
Figure 1: Publication records from the Web of Science database, using the search words *NIR, portable* or *NIR, handheld* or *NIR, hand-held* from 2000 to 2020 (Data accessed on 29 October 2020). Number of publications for the first ten Web of Science categories

[80]

The application of miniaturized spectrometers appears in popular scientific journals and online PR articles. Except for a few exceptions, these research devices are not mature enough for market entry, so spectroscopy has yet to be commercialized for everyone. They need to be small and affordable, to begin with, but most importantly, they have to be able to get customers that need to be more professional in this field. The most used NIR sensor that can be for this use case is the MEMS-based Fabry-Pérot interferometers (FPIs) can be used as tunable filters for single-channel detectors [91]. At least one of two highly reflective mirrors can change the optical gap and filter transmittance wavelength in such devices. FPI can address different wavelength ranges by choosing nonabsorptive mirror materials and adjusting device dimensions, in addition to its small size and low cost. In the near-infrared (NIR) spectral range, for example, a filter resolution below has already been demonstrated. This resolves fundamental vibrational mode overtones that govern material reflectance spectra [2], and allows material classification.

Market penetration rates for compact spectrometers - 2019
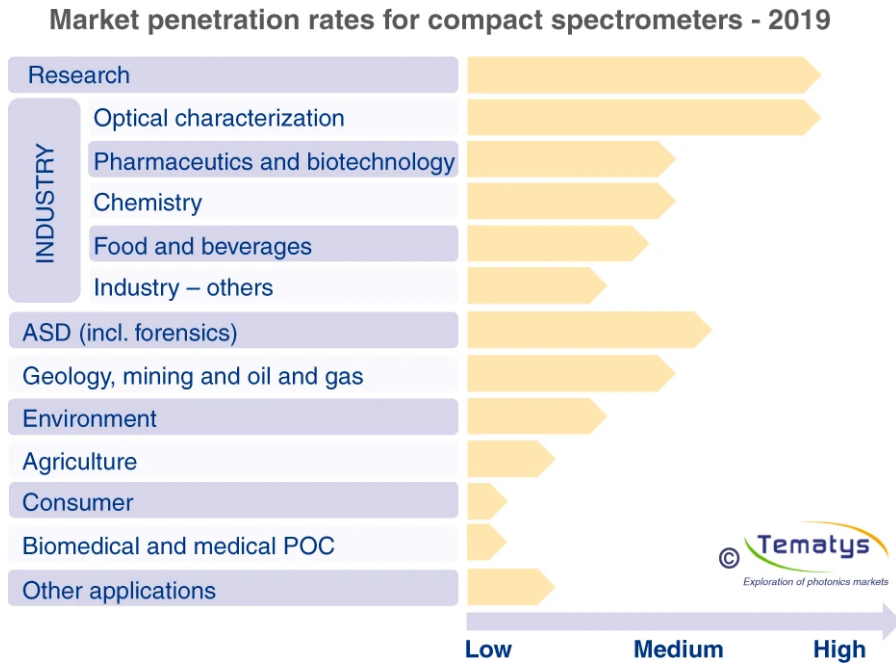
Figure 2: Market penetration rates for compact spectrometers [51].

These spectrometers are essential instruments in many industries, such as chemical, medical, materials, etc... The projected market for mini- and micro-spectrometers has gone up to about 900 million dollars, which indicates there are efforts from academia and industry to use these devices [53]. The predicted penetration rates and market values of compact spectrometers in different applications and segments are shown in Figure 2.

Portable and handheld spectrometers provide the ability to make informed decisions on the spot. The time saved in not having to send samples back to a laboratory is precious, and rapidly delivering those results at the point of need changes or transforms how the user gathers information. Portable instruments are designed for non-scientists in the field, which may be experts in other areas, such as emergency response. Combined with algorithms and libraries for identification and quantification, portability enables on-site analyses, followed by informed decision-making. Economic, security factors and ease of use have driven the historic growth of the commercial handheld spectroscopy market. These transportable instruments are typically packaged in a small suitcase or mounted in some structure for stability and brought to the point of interest. We can combine them with machine learning model for prediction and make them a dedicated field analyzers that are configured to visualize answers as their primary output, not spectra (e.g if a certain leaf has hydrated enough or not) or to provide quantitative or semi-quantitative information to a process controller (e.g can identify the quantity of a certain ingredient in a mixture).

2.3   MINIATURIZATION CHARACTERISTICS

Miniaturization has reduced costs, yet devices are still too expensive and big for market acceptance. The low-cost devices which rely on a static architecture are confined to the Si detector sensitivity range below 1100 nm. Tunable spectrometers with single detectors are the only affordable options for this spectral range. Tunable FPI spectrometers offer various benefits:

- proven, scalable MEMS technology for production

- high potential for downsizing due to the incorporation of flat FPI filter elements adjacent to the detector

- comparatively large SNR [114]

- expansion to HSI devices with the same imaging sensor filter technology Tunable FPI-based spectrometers have a low SWR compared to other approaches, which is a drawback.

Indeed, numerous handheld NIR spectrometers (Viavi LVF-,Texas Instruments DLP-, Si-Ware MEMS FT-, and Spectral Engines FPI) were tested for quantifying pharmaceutical formulation components[114]. In the investigation, the tiny SWR from 1550 nm to 1950 nm spanning only the initial C-H overtone range and a few O-H pairings lowered FPI prediction accuracy.

The potential advantages of miniaturized spectrometers can be established by comparing them and analyzing the limitation of conventional systems. The essential advantage is that they can be integrated into consumer electronics goods with low power consumption. The possibility of building a low-cost mobile NIR spectrometer (the total hardware effort well below 100€ in addition to the spectrum sensor) capable of fast distinguishing between different sorts of materials.

2.4   CONCLUSIONS

A few years ago, commercial portable spectrometers were much larger, heavier, and less powerful than the latest generation of these instruments, which are now significantly more compact, lighter, more powerful, and significantly more capable of performing qualitative and quantitative analyses.

This field has seen rapid development in a self-perpetuating circular fashion: the desire to perform analyses in the field has driven the development of portable instruments, while their availability has driven the development of new applications. Specifically, the desire to perform analyses in the field has driven the development of portable instruments. As a result, more applications will go from the laboratory to the loading dock and then out into the field, providing answers that can be implemented where needed.

Spectrometers were predominantly limited to labs, tests, and metrology, while they are now sharply penetrating the consumer market in healthcare monitoring and food safety. Optical spectroscopy has proven to be an efficient technique to non-invasively analyze compositions of chemical materials, gases, biological tissues, food, and much more.

Part III

SPECTROSCOPY

INTRODUCTION

The Near-Infrared Reflectance Spectroscopy (NIRS) method enables quick and direct measurements of materials with minimal or even no sample preparation. This method typically produces complicated spectra, which chemometric models can clearly comprehend. These are developed with the use of samples that have been carefully chosen. These samples are qualified by making use of analytical reference standards. Consequently, NIRS is not typically utilized as a "primary" method for chemometric analysis. NIRS procedures used in general cases need to be developed and validated in conjunction with the reference methods. As NIRS procedures cannot be repeated easily by official control laboratories, the reference methods and corresponding specifications remain in the valuable specifications. Internet-of-Things has promoted a wide variety of emerging applications that require compact, lightweight, and low-cost optical spectrometers. While substantial progress has been made in the miniaturization of the spectrometer, most have a significant focus on the technical side but tend to feature a lower technology readiness level for manufacturability and interpretability. This thesis discusses the applicability of portable spectroscopy and gives a general idea of this technology and how to use it.

3.1 GENERAL REQUIREMENTS

Spectroscopic analysis has been crucial in developing some of the most fundamental theories in physics. Spectroscopy studies the absorption and emission of light and other radiation by matter. The definition has been expanded to include interactions between particles such as electrons, protons, and ions. Spectroscopic techniques have been applied in virtually all science and technology technical fields. Optical spectroscopes are used routinely to identify the chemical composition of matter. Radio-frequency spectroscopy of nuclei in a magnetic field has been employed in MRI to visualize the body's internal soft tissue.

NIR SPECTROSCOPY   Infrared spectroscopy is used to investigate the vibrational properties of a sample. Molecular vibration gives rise to absorption bands generally located in the mid-infrared, where they are the most intense and straightforward. Adjacent to the mid-infrared, the NIR region covers the internal absorption bands corresponding to overtones and a combination of fundamental vibrations.

Vibrational spectroscopy in the NIR region differs from that in the IR region because NIR deals only with bands arising from overtones and combination modes. At the same time, IR mainly involves bands due to fundamentals. A schematic view of the absorption band is shown in Figure 3

Infrared radiation absorbed by a molecule causes individual bonds to vibrate in a manner similar to that of a diatomic oscillator [92]. The absorption bands that characterize the NIR region are weak or very weak due to the weak electric transition originating from vibrational. The NIR region is valuable from the point of application because only the NIR region offers a highly transmitting window to radiation [74].

The bands that are found in the NIR region are all due to overtones and combinations. Fundamental vibrational frequencies of a molecule correspond to the transition from $v = 0$ to $v = 1$. For a non-linear molecule, there will be $3N - 6$ (where N is the number of atoms) number of vibrations. Overtones occur when a vibrational mode is excited from $v = 0$ to $v = 2$, which is called the first overtone, or $v = 0$ to $v = 3$, the second overtone. Combination bands are observed when more than two or more fundamental vibrations are excited simultaneously. Overlapping bands are also present in this region.

Absorption of a molecule's photon requires a molecule's transition from an initial state into an exciting final state to obey energy conversation. An overview of the NIR spectral region where there is an interaction between the photons and bond is seen in Figure 3.
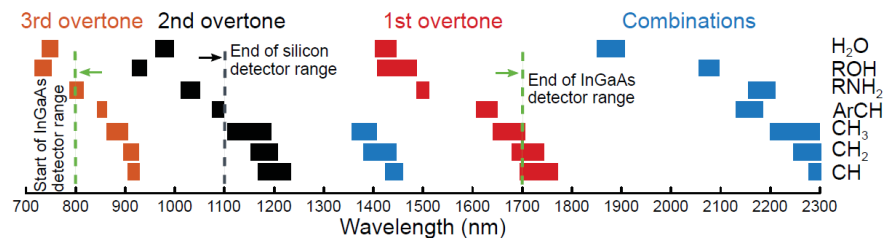
Figure 3: Selected overtones and combination absorption bands of functional groups in the near infrared spectral range[41]. The end of the sensitivity range for silicon and non-extended InGaAs detectors is marked by dashed lines.

Since each overtone should include identical information, determining the NIR spectral range needed for a material classification is difficult.

3.2 APPLICATIONS

NIR radiation penetrates materials deep without altering the target, which is why it can be a powerful non-destructive analysis method.

NIR measurements do not need extensive sample preparation to identify their biochemical composition, simplifying time and accuracy. NIR spectrometers require modest resolution due to the small aspect, the resolution is sufficient for many natural product tests, and moisture detection can be done much lower [2]. Retrieving information from NIR spectra is usually challenging to analyze due to the presence of profoundly convoluted and strongly overlapped peaks. This characteristic makes NIRS suitable for qualitative and quantitative analysis, thanks to chemometric data processing. Another major factor for poor spectra besides the device is how the acquisition setup is done. A different setup scenario can bring different performances among handheld spectrometers. For example, there is a substantial difference when analyzing the protein and fat content of a sample, as shown in Table 3 and 4; we can see that using the same devices but different scenarios have the different performance of the spectrometers.

Spectrometer	Sample	R2	RMSECV
NIRFlex N-500	Protein	0.952	0.365
SCiO	Protein	0.876	0.601

Table 3: Millet analysis with PLS-R model for protein content [105].

Spectrometer	Sample	R2	RMSECV
NIRFlex N-500	Fat	0.9726	1.5711
SCiO	Fat	0.9801	1.2466

Table 4: Cheese analysis with PLS-R model for protein content [108].

These results demonstrate the application of handheld spectrometers that can estimate the variability of complex constituents of samples. These high performances given by handheld spectrometers gain interest in many fields, such as the use of adulteration identification [113] and even in the pharmaceutical formulation [115]. Quantitative analysis can be very challenging; e.g., researchers are still working to have clinically available non-invasive glucose-sensing spectrometers that could benefit many people. Mainly because there is a trade-off between the miniaturization of these devices against the performance factor

3.3 DEEP LEARNING APPLIED TO SPECTROSCOPY

By entering the terms *NIRS*, *machine learning*, and *deep learning* into the Elsevier database, we can observe that there is a growing trend to-

ward the study of near-infrared spectroscopy with cutting-edge techniques. The growing trend can be seen in Figure 4.
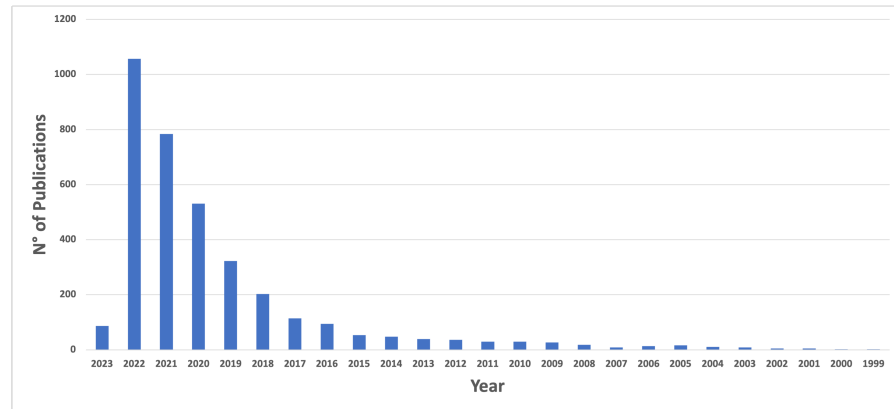
Figure 4: The number of published articles in the field of NIRS and machine learning.

Most often, the analysis of NIR spectra by classical methods requires that the number of observations is higher or equal to the number of variables. The variables usually correspond to the reflectance or absorbance values measures at each wavelength of the NIR region. So, most of the time, the typical approach is to reduce the number of variables limited to the assignment bands of the target and implement linear and statistical models. Since it is rare to know which function group a band arose to determine to classify or quantify a specific target, particularly when we have to distinguish between very similar spectra, we have to use sophisticated models.

The most common method used to analyze the NIR spectra are principal component analysis for classification and partial least-squares regression for quantification. Since NIR spectroscopy is ideal for the quick analysis of inhomogeneous samples without considerable sample preparation, it gives space for spectra analysis. Moreover, the possibility to make in-field analyses that miniature spectrometers gave to industries contributes more to massive data generation. Therefore, one of the industries that benefited is the agriculture and food business (where the subjects studied are typically organic, thus containing abundant C-H, N-H, and O-H bonds). NIR spectroscopy has been used to measure milk fat, detect tea species, forecast nitrogen and pH values in soil and fertilizer for precision agriculture, and detect melamine in milk powder [87], [116]. NIR spectroscopy can detect food adulteration in raw materials.

The high volume of data generated by NIR sensors in every aspect of our life can require more efficient analytical tools. Therefore, chemometrics has often been employed to extract qualitative or quantitative information from the spectra. The significant models in multivariate analysis are partial least square regression (PLSR), principal component analysis or regression (PCA/PCR) [[78], [43]]. However,

recently there has been a growing trend in using neural networks, AI, and machine learning for NIR spectra analysis. This is due to the large amount of data generated by NIR sensors in various fields, e.g., in agriculture, where robots can equip NIR sensors, NIRS in hyperspectral imaging, NIRS in mobile phones, and NIRS, where a continuous acquisition mode is required.

Deep learning models are increasingly exploited to facilitate information extraction from large data collections [39]. These non-linear models are suitable for the NIR dataset that contain a non-linear effect. Nevertheless, they come with a drawback: the ability to interpret and practice these models (black box). These models are considered mainly for NIR imaging fields where they can decode hyperspectral images e.g., convolutional neural networks compared to the use of PCA on deconvoluted images Mahesh et al. A contribution to the use of deep learning in NIRS has been made by [24] that employed 1D-CNN to predict quantities of mixed organic powders. However, the primary use of deep learning with NIR data is an industry with continuous streams of data where these models can learn specific patterns and detect anomalous sections of data [99]. Nevertheless, to make a good prediction AI-model, we have to generate good-quality spectra, which in many cases is difficult. In particular, the application of handheld spectrometers out of the control room can generate many unwanted characteristics in the spectra that can affect the model prediction.

The use of deep learning methods comes in handy when we combine NIR sensors data with signals coming from other analytical platforms and metadata fusion [[11], [70]], this also can lead to prediction improvement. For example, we can combine data of skin temperature and skin spectra, respectively, from temperature and NIR sensors and try to predict moisture of the skin [111].

Part IV

MEDICAL APPLICATION

HEALHCARE APPLICATION

This chapter describes the viability of applying our handheld near-infrared sensor to determine a macro market in the health care of plants and animals. Specifically, we have investigated the viability of this field with our MEMS-based spectrometer. The data from the spectrometer are then processed for prediction models suitable for this kind of real-world application.

In particular, for the growing field of precision agriculture, we investigated our spectrometer on plants and fruits. Measurements performed on the plant leaf and fruit external and/or sections, e.g., regarding fruit acquisition, were taken in at multiple locations around the fruit equator, in the sunlight and with moderate temperature control; We also investigated the applications of our NIRS humans and animals, and made a pilot study to detect biomarkers for diabetic subjects. We assess the applicability of NIRS in these fields by comparing literature found for our area of investigation and plots derived from the comparison.

4.1   APPLICATIONS

4.1.1   *Orange Fruit: pilot study*

Each study that uses Vis-NIRS to evaluate fruit quality begins by collecting reflectance spectra (R), which are usually transformed into the appropriate absorbance log(1/R) spectra. The visible region, specifically the 400-750 nm range, contains most of the spectral variations seen in a wide variety of fruits, and it is used to distinguish fruit attributes [103]. The spectrum data on the pigments' absorption range may offer indirect supplemental correlations of the fruit's internal attributes (e.g., such as sweetness, bitterness), including hardness, as seen in some fruits whose color changes from green to yellow/orange/red throughout ripening [15]. Citrus fruit's color shift, which depends on the environment of the orchards' location and does not correspond to their age, makes this less pronounced [6]. Other than that, the pattern of the NIR absorption spectra is very similar across the range of fruit species. Even among citrus fruit kinds, the peaks' location and size vary depending on the fruit [64]. The acquisition mode also influences the magnitude and minima of the peaks. However, in general, the same characteristics are present, and the spectral landscape among the same fruit is identical [104].

Given these premises, we investigated our NIR sensors output to identify and distinguish some internal and external characteristics of citrus fruit and compared our result to that described here [14]. In Figure 5 it is shown

We were able to replicate the spectral information of [14]. Figure 6 shows our orange fruit's vibrational bands and peaks of the O-H, C-H, C-O, and C=O functional groups due to the stretching and bending of chemical bonds. All organic molecules include these functional groups that can vary based on the internal attributes of the fruit [14]. When the absorption peaks are distinguishable, they can be used for better quantification of vitamin C, citric acid, or sucrose that contain specific compounds that correlate directly to peaks [10]. On the other hand, the overtone bands relate to transitions to higher excited states, with many falling in the NIR range. A general rule is that overtones have higher frequencies and lower amplitudes than the fundamental. However, the overtones are within our instrumental range: *2v2v* at 1420 nm (first overtone, strong intensity band). The vibrations depend on the chemical environment, resulting in a frequency spread of the bands. The NIR spectra of fruit contain mainly overtones and combination bands of stretching and bending vibrations of the main functional organic groups of relevant organic compounds regarding the fruit's internal composition, such as O–H and C–H. The fruit tissue is composed of many different organic molecules, so it is easy to understand the spectral footprint [10]. A large number of possible vi-
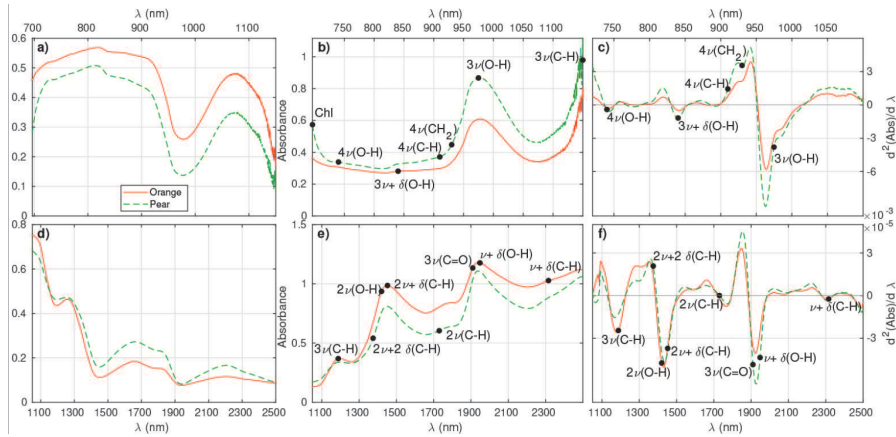
Figure 5: (a) Average reflectance spectra of a set of 255 'Valencia Late' oranges and 239 'Rocha' pears acquired in the Vis/NIR; (b) Average absorbance spectra of the same set of fruit. The nominal positions of the most important absorption bands are indicated in the curves. The number is the order of the transition, ν stands for stretching vibration, δ for bending vibrations, and the sum indicates combination bands (for example, $3\nu + \delta$(O–H) represents the combination band of the second overtone of stretching with the fundamental bending in O–H); (c) Savitzky–Golay [90] filter of second derivative order applied to the absorbance. The bands are again indicated; (d) to (f), same as in (a) to (c) but in the NIR range, with the spectra acquired in reflectance mode. [14]

brations and corresponding bands originate a spectral landscape with very broad and unspecific features, from which it is nevertheless possible to retrieve valuable information. For instance,[10] obtained better prediction of fructose and reducing sugars when using NIRS than MIRS.

In Figure 6, the fruit spectra show firm water absorption peaks around 1450 and 1940 nm. However, the C-H bands may distort the water peaks slightly, and this distortion can lead to more information. The patterns associated with the OH and CH vibrations can be used to retrieve information about sugars. Using the water absorption bands are correlated to that of the sugars [[64], [15]]. Moreover, due to a variety of reasons, the various peaks, even when they correspond among different fruits, may have varying degrees of significance, as indicated by their infrared intensities. This pilot study on citrus fruit investigates if these broad peaks differ in different parts of the same fruit and give us a better understanding of the deeper section of fruits.

4.1.1.1 *methods*

We used our mems-based spectrometer, within the range from 1350 nm to 2150 nm. The orange fruit used for this experiment is cultivated in Italy and bought from a grocery store. The spectra acquisition began with the whole orange placed on top of our spectrometer,
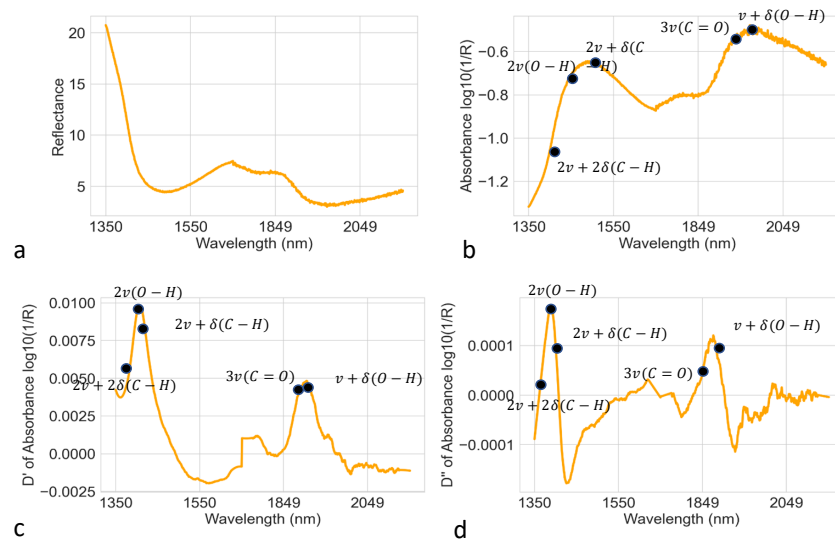
Figure 6: (a) Average reflectance spectra of whole-orange acquired in the NIR region; (b) Average absorbance spectra of the same set of spectra. The nominal positions of the most important absorption bands are indicated in the curves. The number is the order of the transition, ν stands for stretching vibration, δ for bending vibrations, and the sum indicates combination bands (for example, $3\nu + \delta$(O–H) represents the combination band of the second overtone of stretching with the fundamental bending in O–H); (c) (d) Savitzky–Golay [90] filter of first and second derivative order applied to the absorbance. The bands are again indicated;

as shown in Figure 8 - (a), and NIR spectra are acquired from three different points on the equatorial line of the fruit. Each spectrum representing the point is acquired with different lamp power. The lamp power values are integers, from 100 to 250, that correspond to the light bulb intensity. In the next step, the orange has been peeled. Two peels from the equatorial line of the orange and one from one of the orange poles. We labeled *orangeDown_whiteUP* spectra corresponding to acquisitions made by placing the orange part of the peel facing our sensor, as shown in Figure 8 - (b). Subsequently, we separated the white part of the peel from the orange part and made NIR acquisition from the obtained sections, labeling them "white" and *orange* peels.

We calibrated the spectrometer using Formula 2. We did not include the distance and temperature factor when calibrating. The experimental setup can be seen in Figure 7. Different colors represent the different sections of the orange. Each line is the average spectrum of 10 single acquisitions at a chosen lamp power. The dataset contained 1920 total spectra. Each part of the fruit contained a total of 480 spectra.
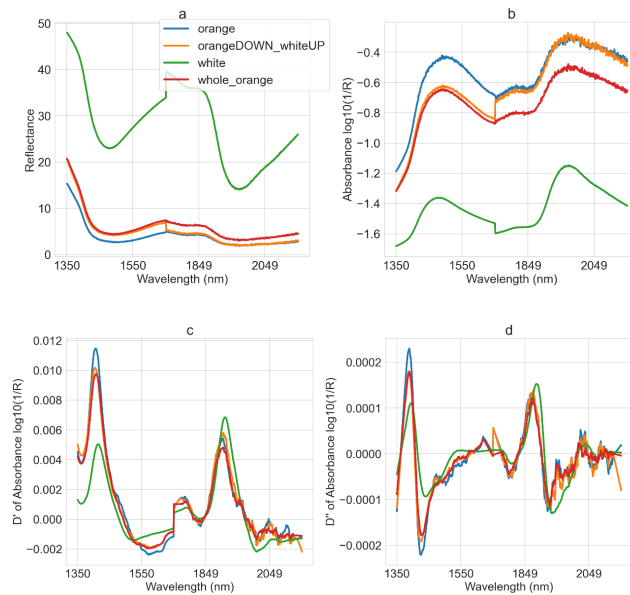
Figure 7: (a) Average reflectance spectra of different parts of the same fruits acquired in the NIR region; The *orange* is the exocarp part of the fruit, "white" is the mesocarp part, *orangeDOWN_whiteUP* is the exocarp part facing the device and lastly the "whole_orange" which is the whole orange placed on top of our acquisition device (b) Average absorbance spectra of the same set of spectra.(c) (d) Savitzky–Golay [90] filter of first and second derivative order applied to the absorbance.

In this study, we use a Support Vector Machine (SVM), Random forest (RF) and K-nearest neighbor (KNN) models to classify the differ-
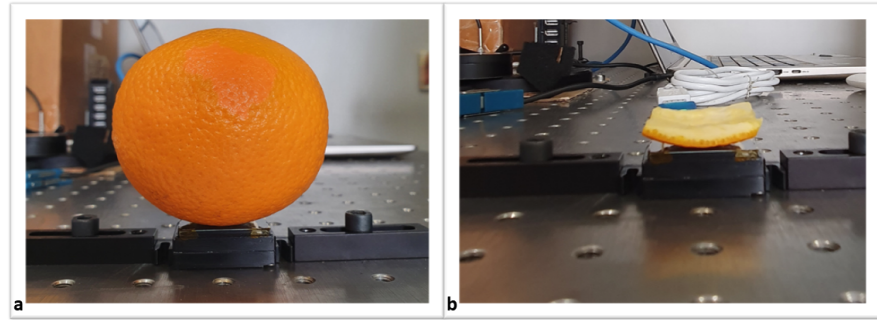
Figure 8: Experimental setup for the acquisition of orange fruit spectra.(a) whole orange acquisition setup. (b) orangeDown_whiteUp peel acquisition.
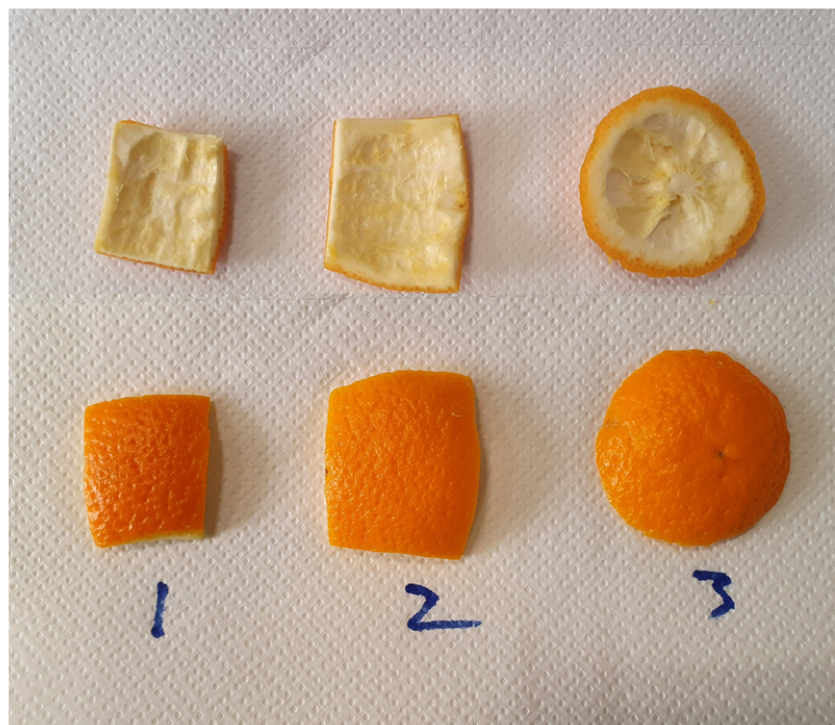
Figure 9: Three different peels from the same orange fruit. Peel 1 & 2 are carved from the equatorial line of the fruit white Peel 3 from the bottom.

ent selected parts of the citrus fruit. Support Vector Machines (SVM) is one of the most widely used non-linear models in Chemometrics. This model seeks class boundaries in the dataset; it optimizes class boundaries by maximizing neighbor sample distance margins. It can use kernel tricks to transform data points into a mathematical space with easier boundaries. SVMs are now used for regression as well as classification (SVR). Many datasets have used SVR successfully, but its optimization task is complex [17, 26, 52, 55]
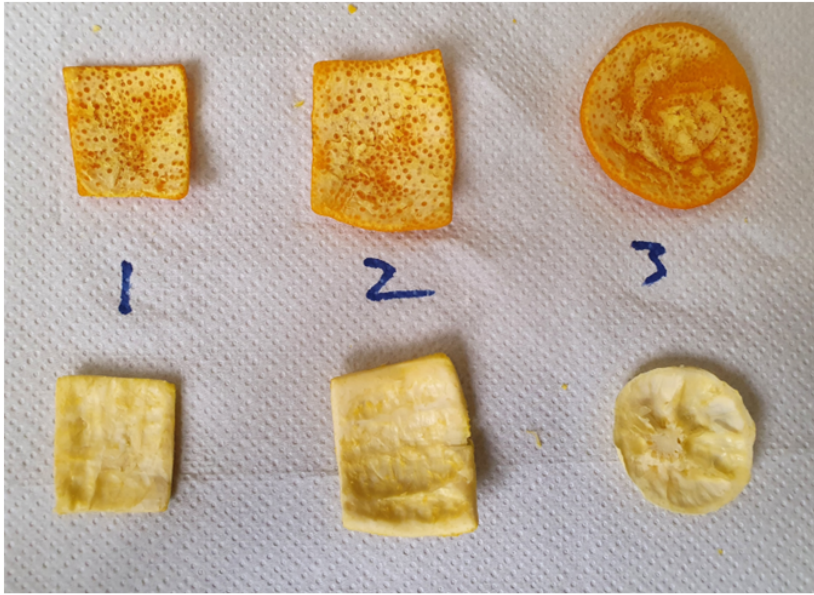
Figure 10: The exocarp (orange part) and mesocarp (white part) part of the
same peel.

Random Forests is another advanced method of machine learn-
ing. The classification is achieved by constructing an ensemble of
randomised classification and regression tress (CART) [12]. For a
given training dataset, $A = (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, where $x_i = 1, 2, ..., n$, is a variable or vector and y_i is its corresponding property
or class label.

Another powerful classification method is the K-nearest neighbor
(KNN), a popular and powerful classification method that uses a non-
parametric approach to categorize data into different classes. In the
context of spectral data, the goal of KNN is to group similar spectral
data into the same class.

The algorithm operates by identifying the k nearest neighbors from
the training data set that are closest to the target or unlabeled spectra
based on a selected distance measure. The most commonly used dis-
tance measure is the Euclidean distance, which calculates the straight-
line distance between two points in a multi-dimensional space.

Once the k nearest neighbors have been identified, the algorithm
assigns the target spectra to the class that is most commonly repre-
sented among these neighbors. This approach is considered simple
and effective, making KNN a widely used classification method for
spectral data.

We used the AUC, CA, F1, Precision, and Recall metrics to evalu-
ate the performance of our classifier for the classification of the near
infrared (NIR) spectra profiles.

AUC (Area Under the ROC Curve) measures the overall performance of a binary classifier by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various classification thresholds. AUC ranges from 0 to 1, with a value of 1 indicating perfect performance, and a value of 0.5 indicating random performance. The formula for AUC can be represented as follows:

$$AUC = \int_0^1 TPR(FPR) dFPR$$

CA (Confusion Matrix Accuracy) measures the overall accuracy of a classifier by comparing the number of true positive and true negative predictions to the total number of predictions made. The formula for CA can be represented as follows:

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

F1 (F1 Score) is the harmonic mean of precision and recall and provides a balance between the two metrics. It can be considered as a single metric that summarizes both precision and recall into a single value. The formula for F1 can be represented as follows:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Precision measures the number of true positive predictions among all positive predictions made by the classifier. The formula for precision can be represented as follows:

$$Precision = \frac{TP}{TP + FP}$$

Recall (also known as Sensitivity or TPR) measures the number of true positive predictions among all actual positive observations. The formula for recall can be represented as follows:

$$Recall = \frac{TP}{TP + FN}$$

where TP (True Positives) represents the number of correctly classified positive samples, TN (True Negatives) represents the number of correctly classified negative samples, FP (False Positives) represents the number of samples incorrectly classified as positive, and FN (False Negatives) represents the number of samples incorrectly classified as negative.

4.1.1.2   *Result and Discussion*

All NIR citrus spectra in Figure 12 of the different parts show similar trends and are difficult to distinguish. After applying the 1st Savitzky-Golay derivatives preprocessing method to the original spectra, several peaks of spectra can be observed, which may be related to the

main chemical constituents of sugar (sucrose, fructose, and glucose) and water. These constituents may differ among the different parts of the fruit that are chosen. The wavelength range used was 1350-2150 nm. The results of the acquisition of the NIRS spectra are then carried out by the Principle Component Analysis (PCA).
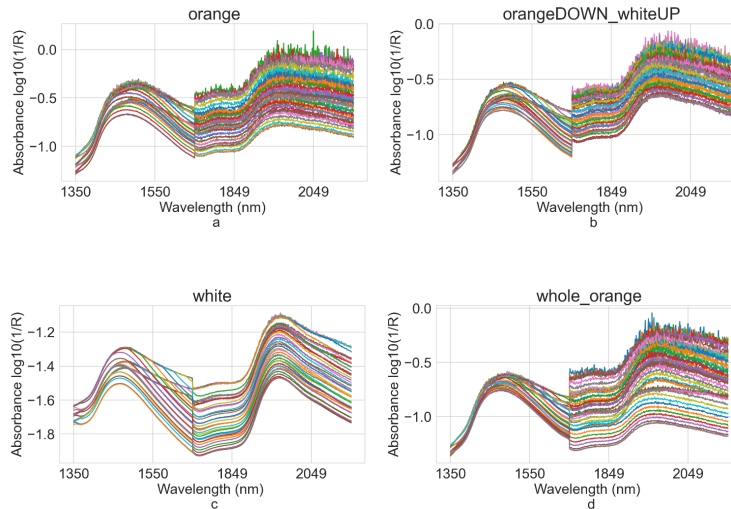
Figure 11: The absorbance NIR spectra of the different parts of Citrus fruit. a) The spectra from only the exocarp (the orange part of the peel) placed on top of the spectrometer. b) The exocarp of the citrus placed on top of the spectrometer with the mesocarp facing upwards c) The mesocarp of the fruit placed on top of the spectrometer. d) The spectra from the whole orange placed on top of the spectrometer.

Data preprocessing is done to improve the classification of the different spectra representing the different parts of the fruit. We used the most common NIR data pretreatment method, which is the Savitsky Golay first derivative and second derivative. This method is able to remove the background noises and increase the resolution of the spectra, and can clarify the peaks and valleys of absorbance spectra of NIRS data. We applied the Standard Normal Variate (SNV) to eliminate multiplicative interference from scattering effects on spectral data. The effect of SNV is on a vertical spectrum scale centered at zero.

As an exploratory analysis, we used the Principal Component Analysis (PCA). The principle of PCA is to group data into new data that are no longer correlated. This method is usually used in multivariable analysis to avoid multicollinearity, and it allows us to classify by looking at differences and similarities between the different parts of citrus fruit. Figure 13 it is shown the explained variance by two components of the PCA algorithm, and Figure 14 shows how the score plots of the
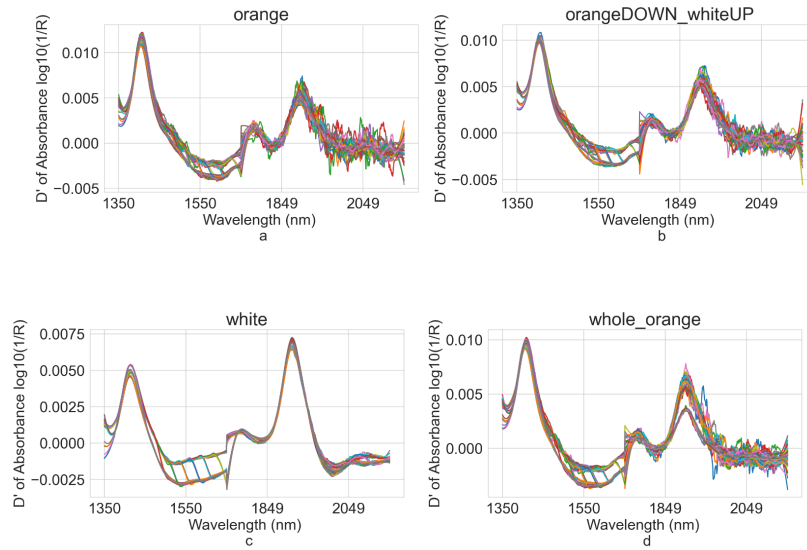
Figure 12: The First derivative of the absorbance NIR spectra of the different parts of Citrus fruit. a) The spectra from only the exocarp (the orange part of the peel) placed on top of the spectrometer. b) The exocarp of the citrus placed on top of the spectrometer with the mesocarp facing upwards c) The mesocarp of the fruit placed on top of the spectrometer. d) The spectra from the whole orange placed on top of the spectrometer.

two principal components that separate our reflectance spectral data into groups. We can see that when no pretreating is applied to our spectral data, the score plot of the two principal components clusters the data into two major groups that is the "white" part and the "orange", "orangeDOWN_whiteUP" and "whole_orange" into one group. This is because the spectra are acquired with the exocarp side facing the spectrometer and thus contain mostly information from it. With a further preprocessing step, we can distinguish each acquisition sufficiently, as seen in Figure 14(b)(c). We can assume from these results that each part of the citrus fruit has its unique spectra and that we can separate the inner parts as well.

The result of our classification model are grouped into three modalities: using data with no pretreatment, using the SNV scaler combined with the first derivative, and using the SNV scaler with the second derivative. We used the stratified 10-fold Cross validation strategy to have robust results and applied a local outlier factor for the detection of spectra that were not in line with the others.

The classification result using just the raw data (no pretreatment) is shown in Table 5. We can see that the KNN and the Random Forest models performed really adequately in classifying the different spectra in their specific classes. Generally, the models performed satisfactorily with not treated spectral data. Nevertheless, preprocessing
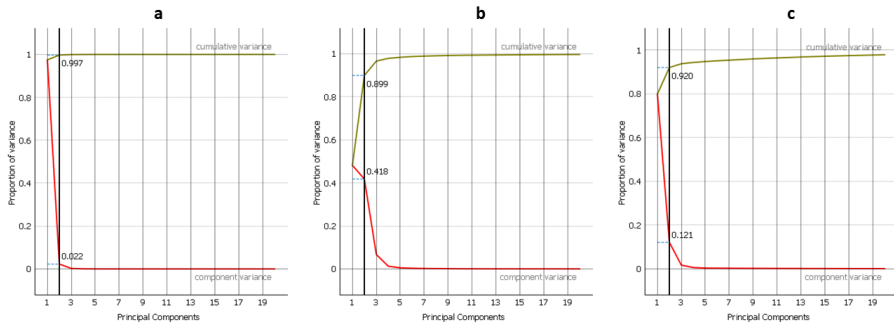
Figure 13: The explained variance of our spectral data by two principal components (vertical black line). (a) Reflectance data with no pretreatment. (b) Reflectance data with Savitsky Golay first derivative. (c) Reflectance data with Savitsky Golay second derivative.
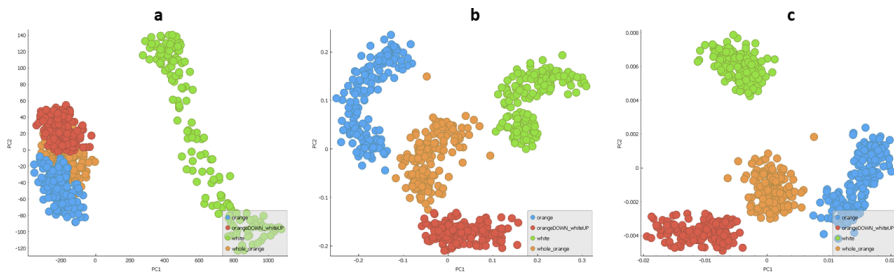
Figure 14: The scatter plot of two principal components. (a) Reflectance data with no pretreatment. (b) Reflectance data with Savitsky Golay first Derivative. (c) Reflectance data with Savitsky Golay second Derivative.

the data with SNV and the Savitsky Golay derivatives gives a more robust classification and higher accuracy values as shown in Tables 6, 7.

Model	AUC	CA	F1	Precision	Recall
kNN	0.9993	0.9970	0.9970	0.9970	0.9970
SVM	0.9210	0.7787	0.7804	0.7828	0.7787
Random Forest	0.9989	0.9940	0.9940	0.9940	0.9940

Table 5: Classification models comparison using raw data (no pretreatment).

4.1.1.3 *Conclusions*

We investigated the application of our MEMS-based NIR Sensor for detecting the internal and external characteristics of citrus fruit. This will allow for future industrial applications in the detection of sev-

Model	AUC	CA	F1	Precision	Recall
kNN	0.9993	0.9990	0.9990	0.9990	0.9990
SVM	0.9997	0.9840	0.9840	0.9841	0.9840
Random Forest	1.0000	0.9970	0.9970	0.9970	0.9970

Table 6: Classification models comparison using preprocessed data (SNV + Savitsky Golay 1st derivative)

Model	AUC	CA	F1	Precision	Recall
kNN	0.9993	0.9952	0.9952	0.9952	0.9952
SVM	0.9962	0.9654	0.9654	0.9660	0.9654
Random Forest	0.9999	0.9962	0.9962	0.9962	0.9962

Table 7: Classification models comparison using preprocessed data (SNV + Savitsky Golay 1st derivative)

eral citrus fruit classifications based on their internal features. We compared our citrus spectra profile on our range or wavelength in the literature to confirm the reliability of our sensor. Furthermore, we show that we are able to distinguish different parts of citrus fruit with regular classification models. Further work is needed to assess the different subsections of the citrus fruit just by decomposing the spectra of the exocarp. However, several topics regarding this technology's full potential and limitations need attention and further research to provide the consistency required by the citrus supply chain's daily routines when assessing fruit quality and ripening. Future work need to be done to verify the robustness of our findings using multiple citrus fruits.

4.1.2 *Plant*

The rising digitization of the agrifood market is driving an increased need for innovative technology to facilitate the sector's evolution toward intelligent agriculture, a more environmentally friendly food business, and practical greenhouses and crop breeding management.

In this experimental approach, we applied our NIR senor on the reflectance spectra of pothos called *Epipremnum aureum* (Pothos)

Data from Pothos leaf was used to assess stressful circumstances in plants (water stress) using two cutting-edge technologies that show great promise in terms of their potential to assist in early identification.

The reflectance spectra were obtained using our portable MEMS-based near-infrared spectrometer covering a range from 1350 nm to 2150 nm. This spectrometer is used to acquire spectra from the leaf of the plant and from the surface of the fruits.

We evaluated two distinct hydration circumstances for the Pothos plant, which were referred to as normal and anomalous. After that, we used a beta-variational autoencoder (beta-VAE) to recognize the irregularities in the plant's hydration over three months after the first purchase. We can demonstrate that our suggested combination of near-infrared spectrometry and the beta-VAE can correctly identify abnormalities or detect stressed circumstances in plants. This was accomplished by demonstrating that the combination is feasible.

This part of the work contributes to the recent and promising advancements in smart agriculture by utilizing a new generation of high-resolution, portable, and non-destructive near-infrared sensing technology and powerful machine learning data analytics.

INTRODUCTION   The knowledge of one's surrounding environment is an essential part of the remote sensing process, which, in turn, makes for more effective digital administration and interaction within a cyber-physical system. This makes it possible to improve the scheduling of activities and interventions, as well as to harmonize them, and to extract performance evaluations in a timely manner [6, 85, 93]. The combination of technological progress and predictive model advancement makes the interaction within a cyber-physical system more beneficial. This interaction is especially pertinent for situations involving smart agriculture, where monitoring for abnormalities and unfortunate plant occurrences is essential for ensuring the long-term viability of agriculture, the food business, and crop breeding. [28, 29, 83, 89, 119].

The word "anomalies" refers to dangerous conditions, strange occurrences, or failures that have the potential to harm the crop, influence it as an epidemic, or generally make greenhouse management less effective, which can result in the waste of resources like light, water, or electricity. As a result, it is necessary to recognize and investigate them in advance, not only to stop them from occurring but also to determine the factors that may have contributed to their development and the qualities they share. Several studies have reported advancements in modern farming and smart agriculture, particularly with the employment of new high-resolution optical sensors, i.e.,[63, 67, 83] implementing NIRS in conjunction with advanced real-time predictive techniques based on machine learning. These kinds of advancements have been made possible, in large part, by the employment of new high-resolution optical sensors and machine learning [117, 118]. The latter has, in recent times, been shown to be particularly useful in solving challenges like classification, feature extraction,

or clustering when applied to the agricultural setting. On the other hand, the most cutting-edge inference techniques. [117, 118].

We employ a betavae, a deep learning model that has been shown to offer good accuracy performances for a variety of anomaly detection applications. This allows us to fill in this gap that previously existed [1, 75, 95].

We focus on an artificial anomaly detection problem regarding the hydration state of the plants as a means of demonstrating the viability of this technique in the context of this application situation.

To be more specific, we establish a normal state (that is, dry) and an abnormal one (that is, wet, that is, following water delivery), and then we train the suggested beta-VAE model on a collection of normal samples. Then, we use the trained model to identify new unseen samples, and we show that it is able to successfully recognize anomalous samples under the condition that smaller reconstruction errors are associated with normal samples and larger reconstruction errors are associated with anomalous samples. This is done so that we can demonstrate that the model is able to successfully recognize anomalous samples.

This work represents a feasibility study to prove the potentiality of the beta-VAE model to identify anomalies in plants. We did this by using a dataset that was generated by SeleTech Engineering Srl (Milan, Italy) using the handheld spectrometer.

In further depth, the following is a list of the primary contributions that we make:

1. deploy a novel method of NIRS that is portable, high-resolution, and low-cost for smart agriculture

2. propose and effectively verify a new deep learning-model for anomaly detection in plants. Both of these goals will be accomplished.

4.1.2.1   *Material and Methods*

The current development of portable spectrometers is a direct result of the tremendous advancements that have been made in the miniaturization of processes[22]. They are in high demand in mobile and on-site spectroscopy because of their multiple uses [67]. In this study, the reflectance spectra of Pothos (Epipremnum aureum) were obtained using a spectrometer constructed out of two NIR microelectromechanical sensors (MEMS). The spectrometer was positioned and secured under the leaf's surface.

The experiment was conducted at Seletech Engineering Srl laboratory based in Italy, Milan. One vase of plant Pothos was bought at Effe garden (45.526456162639725, 9.310908391160392). The setting was very carefully monitored during the process of spectra acquisition. The pothos plant was chosen because of its large leaves (that
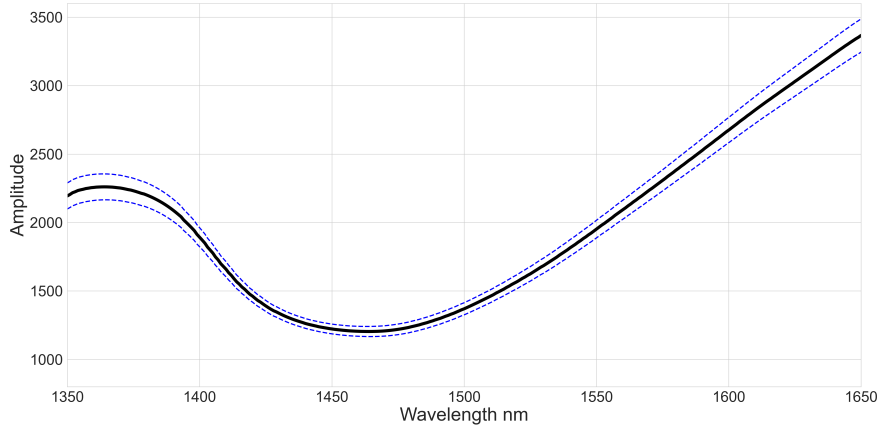
(a) First MEMS $(1350, 1650)$nm.
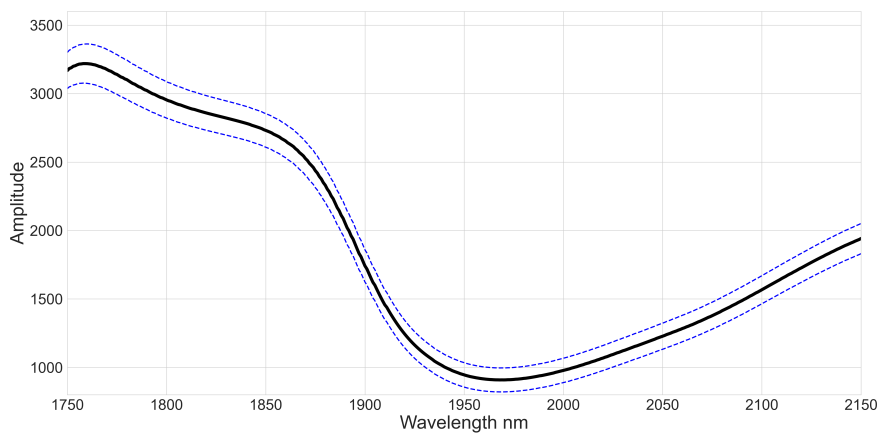
(b) Second MEMS $(1750, 2150)$nm.

Figure 15: Spectra of the normal (dry) period of acquisition. The solid line represents the mean across all "dry" spectra; dashed lines represent mean$\pm$standard deviation.

is, leaves that are big enough in comparison to the sensor), its capacity to adapt to its surroundings (that is, it does not require any particular treatments), and its growth characteristics (i.e., slow growth compared to the acquisition duration). The entire experiment was carried out using just a single leaf. We selected one leaf that is large, and that is not young for this experiment. Under the selected leaf, we placed a humidity and temperature sensor to capture these parameters around the spectra acquisition area. A MI-flora sensor was inserted in the vase to control the moisture and conductivity of the soil, and it also has an integrated temperature sensor. A webcam was also used to detect leaf movements.

The spectrum was collected continuously day and night, every minute for the whole period of acquisition. The acquisition was restarted only seven times to validate data. The first data acquisition began on the 5th of august 2021 and continued through the 26th of November 2021.
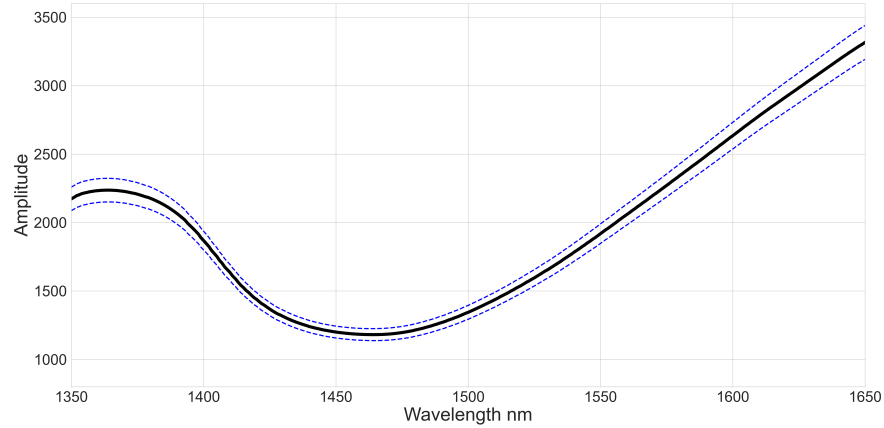
(a) First MEMS $(1350, 1650)$nm.
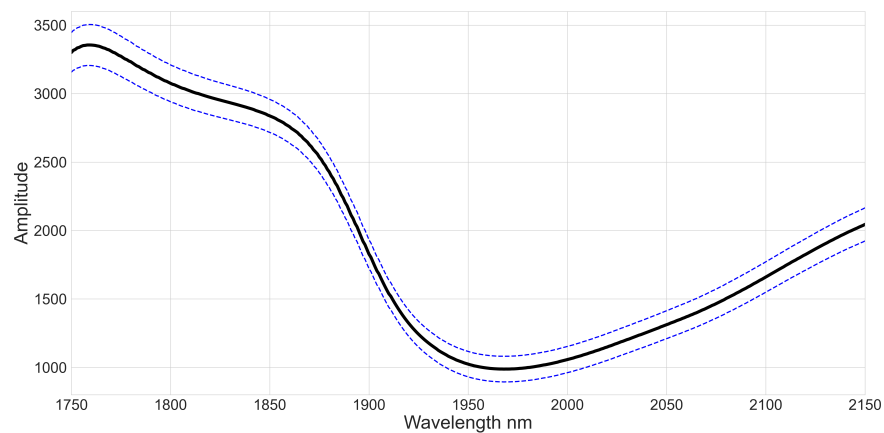
(b) Second MEMS $(1750, 2150)$nm.

Figure 16: Spectra of the *anomalous* (wet) period of acquisition. The solid line represents the mean across all "wet" spectra; dashed lines represent mean±standard deviation.

During the acquisition process of the spectra, water was added at particular (but not regular) intervals, namely, whenever it was determined that the soil in the pot containing the pothos was dry. The periods when water was made available are depicted in the figure referred to as 17. In addition, the quantity of water that was used varied from time to time (it was somewhere between 50 and 150 ml). While we were examining the efficacy of our suggested anomaly detection model, this further presented us with a variable component to take into account.

In literature, it is well known that certain regions are responsible for water absorption, as shown in Figure 18. These are at 1400 nm and 1950 nm, which correspond to our sensor range. Making our MEMS base sensor reliable in terms of the waveform that is output.

As a preliminary analysis, we conducted a water absorption trend at specific wavelengths for spectra acquired after the 5th of October
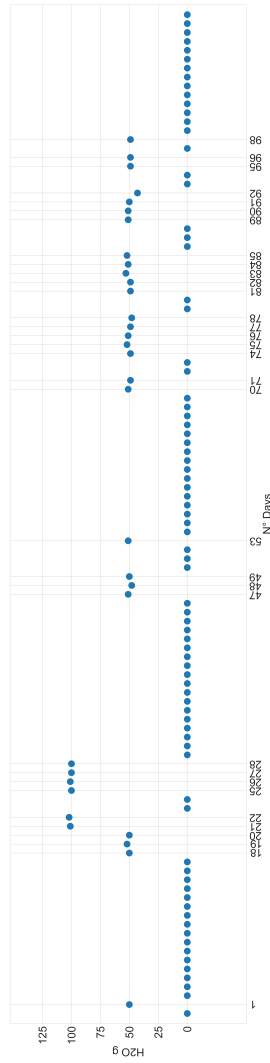
Figure 17: Time-course of water supply.
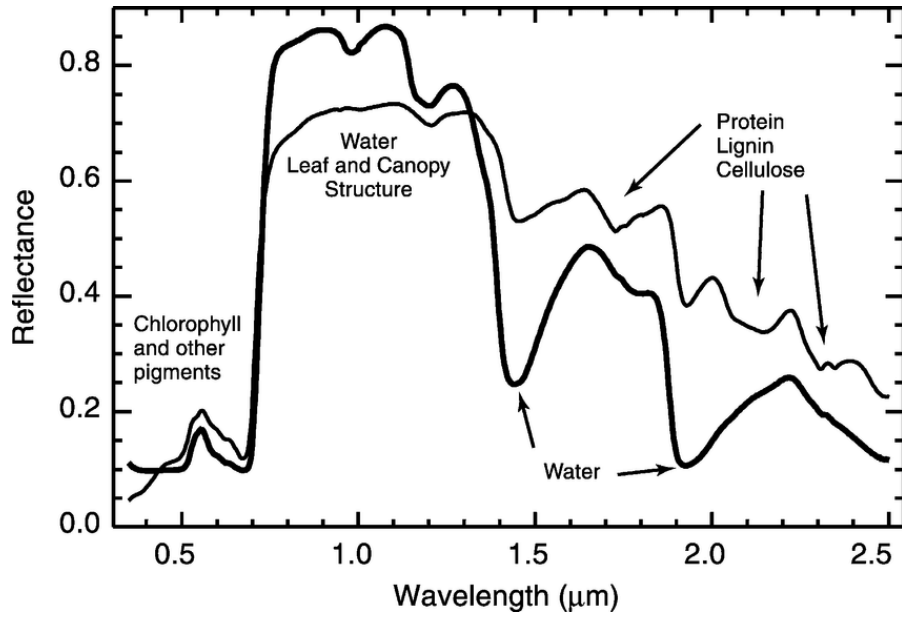
Figure 18: Laboratory reflectance spectra of an oak leaf in the fresh (thick line) and dry (thin line) states. The causes of major plant absorption features are indicated [50]

2021. In particular, we selected the wavelength peaks (1365 nm, 1468 nm, 1600 nm, 1780 nm, 1840 nm, 1860 nm, 1970 nm, and 2080 nm) that correspond to peaks in our spectra. As shown in Figure 19 (water trend), these peaks are influenced by water input and water absence.
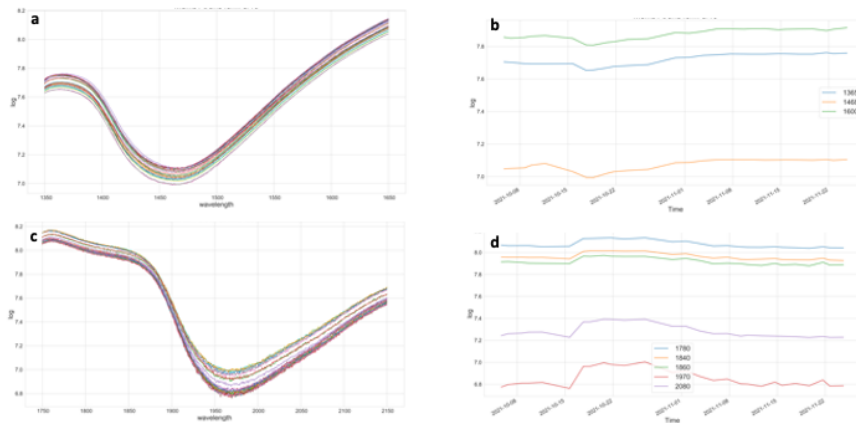
Figure 19: Spectra and selected wavelength in time range from 5 Oct to 26 Nov. (a)(c), Spectra corresponding to the two memes of our sensor. The first one from 1350 nm to 1650 nm and the second one from 1750 nm to 2150 nm). (b)(d), The trend of the selected wavelength from both memes.

Also, the ratio of the values of the selected wavelength shows a good correlation between the moments of water given to the plant

and the moments where there was no water. In particular, in the range of 1350-1650 nm, water absorption is dominant and shows a clear correlation on the days water was given to the plant. In Figure 20 we can see the upward and downward trends of the water in correspondence with water input.

The weight difference between fresh and dried leaves is used today as the basis for the traditional measurement of water content. It takes a lot of time and is undoubtedly destructive. A primary objective of field-based remote sensing is the detection of drought-related plant water stress.

We defined a *normal* and an *anomalous* condition depending on the hydration level of the pothos plant. Particularly, we labeled all spectra in the time range between 1 hour and 5 hours after water supply as *anomalous*, while all remaining spectra (when the plant is dry) as *normal*.

VAE   The objective function of a VAE is a variational lower-bound of the marginal likelihood of the data, since the computation of the marginal likelihood is intractable [8]. In this way, the training consists of the maximization w.r.t. parameters $\mathbf{ffi}$ and „ of

$$\mathcal{L}(\text{„}, \mathbf{ffi}; \mathbf{x}) =_{q_\phi(\mathbf{z}|\mathbf{x})} \left(\log \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x}))} \right)$$

that provides an evidence lower bound (ELBO) for the (evidence) probability $p(\mathbf{x})$. The gap between $p(\mathbf{x})$ and $\mathcal{L}(\text{„}, \mathbf{ffi}; \mathbf{x})$ can be best expressed by considering the Kullback-Leibler divergence ($\mathcal{KL}$) between the variational $q_\phi(\mathbf{z}|\mathbf{x})$ and the posterior $p_\theta(\mathbf{z}|\mathbf{x})$ distributions, which turns out to be

$$\mathcal{KL}[q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z}|\mathbf{x})] = -\mathcal{L}(\text{„}, \mathbf{ffi}; \mathbf{x}) + p(\mathbf{x})$$

Since $\mathcal{KL}[q_\phi(\mathbf{z}|\mathbf{x})\|p_\theta(\mathbf{z}|\mathbf{x})] \geqslant 0$, one arrives at the lower bound $\mathcal{L}(\text{„}, \mathbf{ffi}; \mathbf{x}) \leqslant p(\mathbf{x})$. Similarly, the ELBO can be also formulated as

$$\mathcal{L}(\text{„}, \mathbf{ffi}; \mathbf{x}) =_q (p_\theta(\mathbf{x}|\mathbf{z})) - \mathcal{KL}[q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})]$$

The first term, $_q(p_\theta(\mathbf{x}|\mathbf{z}))$ gives the reconstruction error, which forces the decoded samples to match the initial inputs. This makes VAEs extremely suited for modelling normal behaviors in an anomaly detection task. As introduced above, if the model comes across anomalous samples, it will generate a reconstructed object X which is significantly different from the distribution of the original data, thus considerably increasing the reconstruction error. The second term, $\mathcal{KL}[q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})]$ acts as a regularizer, that is, it penalizes *surrogate* distributions, $q_\phi(z|x)$ that are too far away from the predefined $p(\mathbf{z})$.

Figure 20: The ratio of the selected wavelength in time range from 5 Oct to 26 Nov, the dotted vertical lines correspond to the days of water input. (a) The ratio of selected wavelengths from the first MEMS (1350 - 1650 nm): 1375/1475, 1365/1468, 1600/1468. (b) The ratio of selected wavelengths from the second MEMS (1750 - 2150 nm): 1780/1975, 1830/1975. (c) The ratio of selected wavelengths from the first and second MEMS: 1475/1950.

More specifically, we employed a β-VAE model [13], which slightly differs from the standard VAE in that the loss function contains a hyper-parameter β > 1 as follows

$$\mathcal{L}(\textbf{\textit{,,}}, \textbf{ffi}; \mathbf{x}) =_q (p_\theta(\mathbf{x}|\mathbf{z})) - \beta \cdot \mathcal{KL}[q_\phi(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})] \tag{1}$$

This choice has been made to further help the model regularizing the latent space.

The β-VAE was implemented with two feed-forward neural networks, one for the encoder and the other for the decoder. For the latent distribution, we empirically chose a Gaussian distribution with dimension $d = 2$.

The encoder's network is shown in Fig. 21(a). The network consists of two separate branches, each processing the spectra acquired by one of the MEMS sensors, separately. Their outputs are concatenated in a vector and given as input to 3 additional layers. The encoder produces 4 outputs, where two of them are interpreted as mean values, and the remaining two as variance values. For each pair of spectra acquired by the two MEMS, i.e., spectra x_1 and x_2, a latent variable z is sampled from the latent distribution. Then, z is fed to the decoder that aims to reconstruct the spectra to match the distribution of the input samples.

The VAE decoder was implemented with the same architecture of the encoder, by mirroring its layers, as displayed in Fig. 21(b). The first layer consists of 2 neurons, only (the dimension of the Gaussian distribution). The sample is passed through two feed-forward layers. The output of the second layer is split in two and each part gives input to two additional layers that reconstruct the original signal. It is worth to note that corresponding layers in the encoder and in the decoder share the same number of neurons, with the exception of the last layer of the encoder and the first layer of the decoder (which have 4 and 2 neurons, respectively).

We trained the *betavae* with spectra belonging to the *normal* condition, only. The training was implemented using 50 epochs, AdamW optimizer [62], a learning rate of 0.01 and a batch size of 32. The training set (*normal* or dry hydration condition) consists of 99982 spectra, the validation set (*normal* or dry hydration condition) of 21424 and the test set (*anomalous* or wet hydration condition) of 15607 spectra.

4.1.2.2 *Results and Discussion*

We used the proposed *betavae* model to identify spectra corresponding to the periods of time when water was supplied. First, we conducted a visual inspection analysis. As it can be seen from Figs.15 and 16, both MEMS sensors are able to show the well-established water-related wavelength range, i.e., the pronounced troughs around 1450 nm (visible in the MEMS1's spectrum) and around 1950 nm (visible in the MEMS2's spectrum). Second, we trained the betavae model and
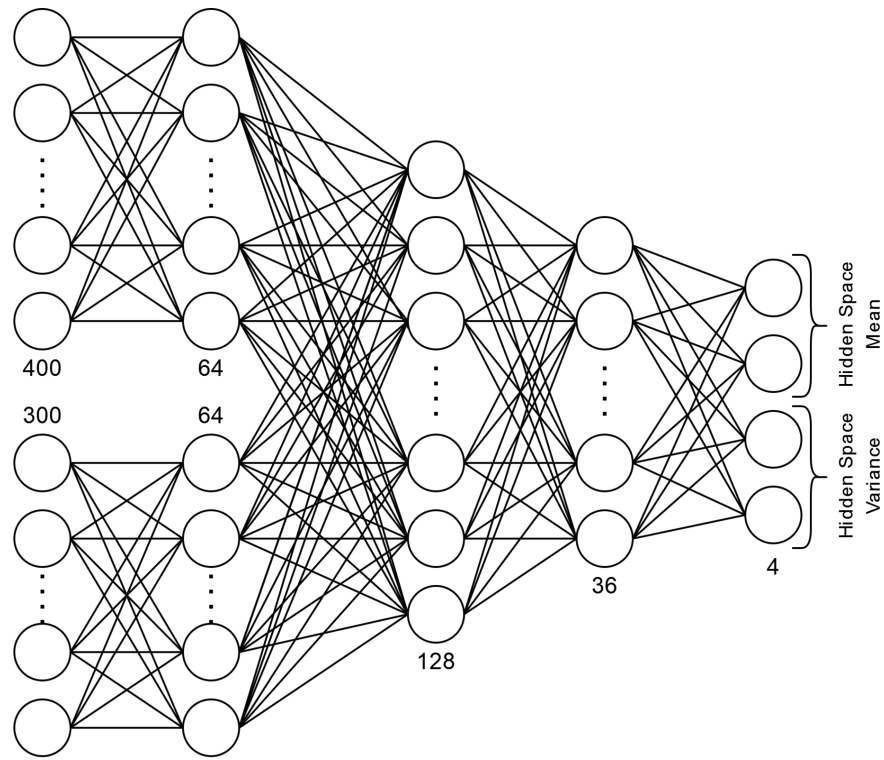
(a) VAE Encoder
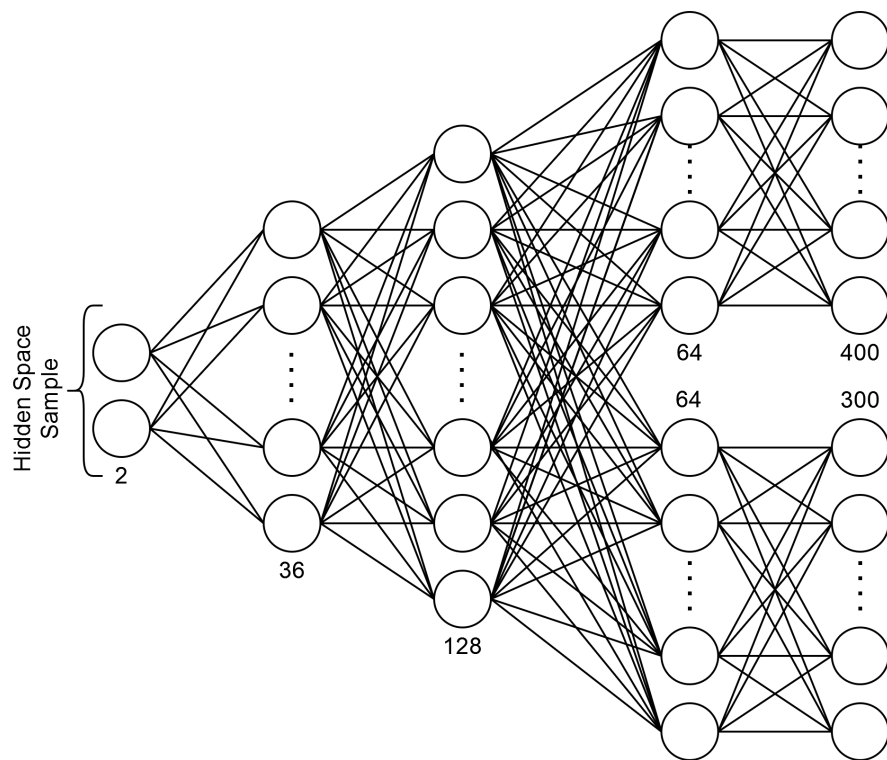
(b) VAE Decoder

Figure 21: Our proposed VAE architecture. The numbers below each layer represent the number of neurons for that specific layer.

investigated its reconstruction ability, both in case of *normal*, previously unseen, spectra (i.e., in the validation set) and *anomalous* spectra (i.e., in the test set). In Table 8, we report the average reconstruction errors of the *betavae* model. The error made by the model in the case of spectra belonging to the *anomalous* condition (i.e., the test set) is significantly larger w.r.t to the error for the spectra in the *normal* condition. More precisely, the error made in reconstructing a spectrum of the *anomalous* condition is 67,2% larger w.r.t. the error in the training phase. On the other hand, a previously unseen "dry" spectrum (belonging to the validation set) is reconstructed with an average error of 0.6% w.r.t. to the training phase. Thus, we can conclude that the proposed *betavae* model is able to recognize those spectra corresponding to the times in which water has been supplied, while the plant was in a water stress condition.

	Reconstruction Error (Mean±std)	Reconstruction Error (% Mean w.r.t.Training error)
Training ("Dry" spectra)	$(25.69 \pm 4.96) \cdot 10^{-4}$	-
Validation (unseen "Dry" spectra)	$(25.85 \pm 4.99) \cdot 10^{-4}$	0.6%
Test ("Wet" spectra)	$(42.96 \pm 7.02) \cdot 10^{-4}$	67.2%

Table 8: Reconstruction errors of the betavae model. The errors are reported in terms of average and standard deviation (second column), as well as as the average percentage w.r.t. the error made during the training phase.

In this work, we face the challenge of identifying stressful events in plants monitoring, with the perspective of advancing smart agriculture. We took advantage from the promising ability, recently reported, of *dl*-based inference methods to detect anomalies, and we proposed a *betavae* model to identify them in the hydration level of a pothos plant, in controlled laboratory settings. As a further novelty, we employed a low-cost, portable, yet high-resolution NIR spectrometer, exploiting MEMS technology, engineered by SeleTech Engineering Srl.The spectrometer was able to provide quantitative information in two broad NIR ranges, i.e., $(1350, 1650)$ nm and $(1750, 2150)$ nm. To validate the model, we exploited the well-known water-related effect on the wavelength ranges around 1450 nm and 1950 nm, and we purposely defined a *normal* (i.e., dry) and an *anomalous* (i.e., wet or water supplied) hydration condition. Although preliminary, this work showed the possibility for a betavae model to identify different conditions in the hydration of the plant, i.e., when water was supplied by the experimenter. Thus, this work contributes to the recent and promising advancements in smart and precision agriculture, made it possible by new generations of high-resolution optical technologies and deep learning -based data analytics. The work can be extended with further investigations. For example, the data can be calibrated to allow for a fair comparison with other studies, and a more precise quantification of the amount of water supplied could be ensured by using the proposed *betavae* architecture. Nevertheless, this study proves it feasible

to use the proposed model, in conjunction with a portable NIR spectrometer, to identify different conditions in the hydration, i.e., water stress, in plants. In the future, the same architecture could be tested to assess its generalizability, i.e., to identify other (i.e., more relevant and unknown) stress conditions. In that case, the trained model could be also used to detect real-time anomalies (i.e., using the model as in the test phase of our analysis) and provide an alert to domain experts and on-field operators.

4.1.3 *Diary Cow*

The digital revolution is having a greater and greater impact on dairy farming and leading to the third agricultural 'green' revolution of the mid-20th century. Precision livestock farming is a new technology that is employed to address contemporary issues, including environmental, economic, and social sustainability, supported by technologies that collect data useful for farm and supply chain performance improvement, along with task automation and compliance. This has been made feasible by the creation of numerous sensors that can be used in the barn. Near-infrared spectroscopy (NIRS), which is adaptable and can be used online/inline to evaluate and regulate the crucial stages of the production process, is undoubtedly the technology that is having an impact on numerous elements of dairy cattle breeding. NIRS can now collect data and keep an eye on the dairy cow's general health inside the barn. By removing the expense and analytical response waiting periods, all of this can be done quickly, improving livestock management. Numerous research supports the use of NIRS as a trustworthy and prognostic tool. This evaluation emphasizes the value of our handheld NIR sensor in dairy farms.

4.1.3.1 *Introduction*

New technologies strongly influence farm management, reducing manual labor, costs, and waste, consequently increasing income. Precision agriculture has evolved in agriculture 4.0 through the collection, integration, and automatic analysis of data from the environment, sensors, and any other third source [77].

Precision livestock farming (PLF) has concerned the whole sector: from cattle to pigs to poultry, achieving the best results in cattle farms, especially dairy farms. The adoption of PLF techniques allows for safe food production with a reduced environmental impact [5]. Improving animal welfare is an effective tool to increase profits by reducing costs related to poor animal health. Consumers are becoming more focused on purchasing and spending more on animal products obtained through the satisfaction of animal welfare and protection of the environment [58].

PLF aims to have increasingly automatic, precise, and accurate farm management. The main technological innovations have to consider all aspects of dairy farming. Currently, the challenge is to get the most considerable amount of data automatically, quickly, and accurately by employing machine learning and deep learning models.

Near-infrared spectroscopy (NIRS) could be used on the farm in a PF system. It could be an economically viable system to provide the right amount of daily nutrients with fewer metabolic alterations. It can be used to get an insight into the physiological status of the cow. Nutritional imbalances, dietary deficiencies, or improper management can generate a range of health disorders in dairy cows. These stress disorders are generally categorized as metabolic diseases such as ketosis, hepatic lipidosis, hypocalcemia, and hypomagnesemia [32]. NIRS can be defined as an innovative process/product analytical technology that can positively affect the production process.

Blood biochemistry is commonly used as part of a diagnostic evaluation to confirm the suspected disease, assess the prognosis, control the progression of the disease, and appraise the effectiveness of treatments. Infrared spectroscopy can be used as a non-invasive method to detect the presence of specific chemical bonds that correspond to the cause of stress in cows. We haven't yet found articles that implement the use of NIR spectroscopy to gauge the cow's stress response in a non-invasive manner. A short list of articles that use handheld spectrometers is shown in Table 9; this list intends to inform dairy processors, researchers, technologists, and engineers about portable and miniaturized NIR analyzers and their dairy industry applications.

This section investigates a preliminary study on the practical application of our handheld sensor in a dairy farm. Our first objective is to assess the feasibility of using NIR spectra acquired from specific body parts of the cow to predict a broad set of blood parameters, such as metabolites related to energy metabolism, liver function/hepatic damage, oxidative stress, inflammation/innate immunity, and minerals in twelve dairy cows. Second, we explored a few prediction models for the NIR spectra predictions, identifying the one suitable for our goal.

Commercial Name	Wavelength (nm)	Measurement Mode	Wavelength Selector	Weight and Size	Dairy Applications	Reference
MicroPhazir	1600–2400	Reflectance,Transmittance	MEMS	Weight: 1.2 kg	Liquid Milk	[110]
MicroNIR 2200	1128–2162	Reflectance	LVF	Weight: <60 g	Milk powder	[123]
MicroNIR 1700/1700ES	950–1650	Reflectance;Transmittance;Transflectance	LVF	Weight: 64 g Size: 45 × 50 mm	Liquid milk , Cheese	[37],[59], [57]
Phazir 1624	1600–2400	Reflectance	MEMS	Weight: 1.7 kg	Liquid Milk	[27]
NIRONE	1100–2500	Transmittance	MEMS	Weight: 15 g; Size: 25 × 25 × 17.5 mm	Liquid milk	[61]
X-NIR	950–1800	Reflectance		Weight: 1.6 kg;	Cheese	[101]
NeoSpectra	1350–2500	Reflectance	MEMS	Spectral resolution:16 nm;Weight: 17 g; Size: 32 × 32 × 22 mm	Liquid Milk	[66]
NIR-S-G1	750–1700	Reflectance	DLP	Weight: 87 g; Size: 76 × 82 × 27 mm	Milk powder	[86]
SCiO	740–1070	Reflectance	-	Weight: 35 g; Size: 3.15 × 9.5 × 27.5 mm	Cheese, Liquid milk	[122],[107]

Table 9: The state of the art of the application of NIR spectroscopy in the field of dairy cow

4.1.3.2 *Material and Methods*

The measurements took place in the Carlazzo area on a dairy farm. For this pilot study, we selected 12 bovines, ten adult female bovines, and two calves (one male and one female, respectively, aged 6 and 8 months). The four adult bovines were fasting for the spectra acquisition, and the other six had just returned from grazing. Figure 22 it is shown the 12 subjects.

Figure 22: Bovine subjects used to test our NIR sensor.

We selected two major acquisition points on the bovine's body: the udder and the ear 23. A large amount of blood flows into the bovine udder, enabling the mammary epithelial cells to synthesize and secrete milk or milk components [73]. In cows, more than 500 l of blood flow is necessary to yield 1l of milk ([25]). Monitoring blood supply and hemodynamics in the udder tissue would provide information potentially helpful in interpreting the individual dairy cows' stress markers. The ear is another possible acquisition area due to its ease of acquisition, as shown in Figure 24.
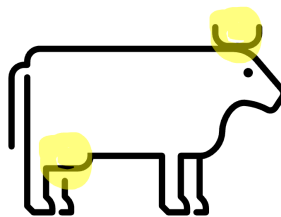
Figure 23: The selected acquisition points from the body: the ear and udder.

During the acquisition, a significant issue was holding steady the bovines, which resulted in a different total number of spectra acquired for each of them, as shown in Figure 25. Six hundred sixty-six spectra were acquired: 468 from the ear and 198 from the udder area. After inspecting the acquired spectra, some samples were removed

Figure 24: Spectra acquisition from ear point. Spectrometer placed and hold against the inside of the ear.

from the dataset. Their spectral waveforms were not in line with the others because they were acquired while the subject was moving. For the final analysis, only 82 good spectra from the udder and 264 from the ear totaled 346 good spectra. In Figure 26, the spectra selected for analysis for each subject are shown. The spectra were acquired using our spectrometer's different lamp intensities and other parameters.
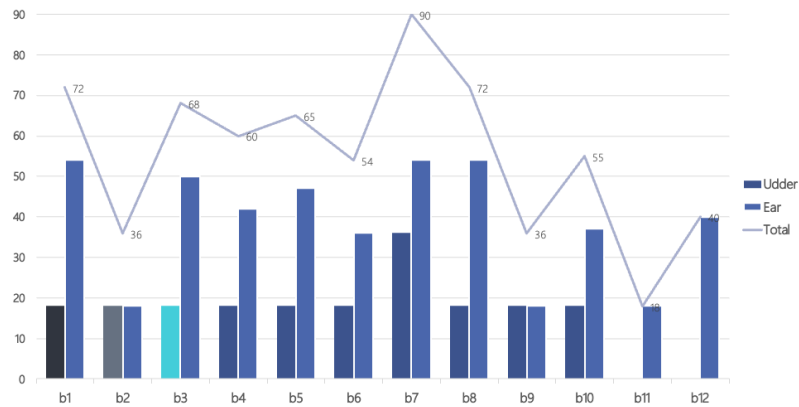
Figure 25: Bar plot of the total number of spectra acquired from bovine body parts. The straight line shows the total number of spectra.

In conjunction with the spectra acquisition, blood was drawn for analysis in the laboratory. In Table 10, we can see the blood test results for each bovine. As all the cows involved in the study were clinically healthy, the range of variability in the data is representative of a physiological condition, as shown in Table 11. We separated them into the subjects in the fasting period, the subjects in regular
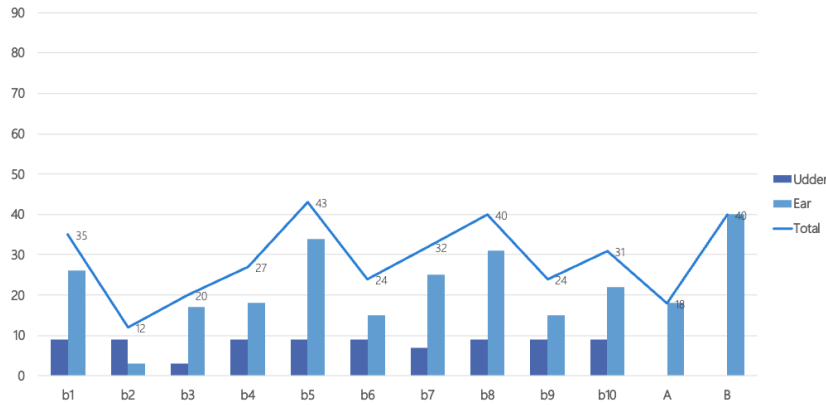
Figure 26: Bar plot of the number of remaining spectra after eliminating the bad spectra.

pasturing, and the two calves. We observed a significant variation in cows' total protein levels; most had elevated globulin levels. Although the cows did not manifest clinical disease, the high variability in some blood biomarkers means we cannot exclude the presence in specific individuals of subclinical conditions, an expected finding in a large population. Regarding urea concentration was found to be steady among the subject. The blood urea level reflects the effects of dietary intake of crude protein, its digestive utilization, milk protein secretion, body protein turnover, and nitrogen (N) urinary excretion [45]. High-producing dairy cattle are commonly fed diets with crude protein levels exceeding 16% to ensure maximum milk output [20].

	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	Reference values (min-max)	
Total proteins (g/l)	75.5	91.3	88.8	97.8	77.5	78.1	79.5	90.9	93.6	73.7	80.8	88.9	74	96
Albumine (g/l)	34.7	34.6	33.1	38.3	37.9	31.1	32.6	36.5	37	35.1	36.5	41.2	33	38
Globuline (g/l)	40.8	56.7	55.7	59.5	39.6	47	46.9	54.4	56.6	38.6	44.3	47.7	38	51
Urea (BUN) (mmol/l)	3.6	3.2	3.7	4.1	3	3.6	3.9	3.2	3.4	3.8	4.8	5.1	3.5	5.5
Creatinine (umol/l)	66	78	75	64	66	71	79	77	78	71	59	62	60	80
Glucose (mmol/l)	2.9	2.77	2.91	3.05	3.12	4.11	3.13	3	2.97	2.81	2.71	2.83	2.82	3.56
Alkaline phosphatase - ALP (UI/L)	97	101	92	88	112	133	146	107	98	141	141	137	83	138
Aspartate aminotransferase - AST (UI/L)	69	79	63	81	76	72	95	92	68	99	99	91	67	97
Alanine transaminase - ALT (UI/L)	43	38	33	31	29	41	45	27	31	30	25	36	29	47
Total Bilirubin (umol/l)	1.54	1.56	1.67	1.69	1.46	1.76	1.7	1.54	1.58	1.65	1.46	1.5	1.52	1.72

Table 10: Blood reference table of each bovine. In columns [b1-b12] represent the twelve considered bovines. The reference values show the ideal minimum and maximum values of the specific blood component.

The spectra acquired from the ear part of the subject where taken by placing our spectrometer on top of a visible vein and part of the ear were there is no vein. Interestingly, we are able to distinguish visibly the these two spectra regions as shown in Figure 27. At the moment we are not able to distinguish the primary factor that is the

	Fasting (m)	Regular (m)	b11	b12
Total proteins (g/l)	88.35 ± 8.11	82.21 ± 7.34	80.8	88.9
Albumine (g/l)	35.17 ± 1.91	35.03 ± 2.44	36.5	41.2
Globuline (g/l)	53.17 ± 7.28	47.18 ± 6.73	44.3	47.7
Urea (BUN) (mmol/l)	3.65 ± 0.32	3.48 ± 0.39	4.8	5.1
Creatinine (umol/l)	70.75 ± 5.89	73.66 ± 4.69	59	62
Glucose (mmol/l)	2.9075 ± 0.1	3.19 ± 0.42	2.71	2.83
Alkaline phosphatase - ALP (UI/L)	94.5 ± 4.92	122.83 ± 18.05	141	137
Aspartate aminotransferase - AST (UI/L)	73 ± 7.35	83.66 ± 12.06	99	91
Alanine transaminase - ALT (UI/L)	36.25 ± 4.66	33.83 ± 6.7	25	36
Total Bilirubin (umol/l)	1.615 ± 0.07	1.61 ± 0.10	1.46	1.5

Table 11: Mean and standard deviation of blood component grouped into the fasting group and regular group of the bovines. Without including the two calves.

cause of these separation due to various issues during the acquiring further investigation is needed.
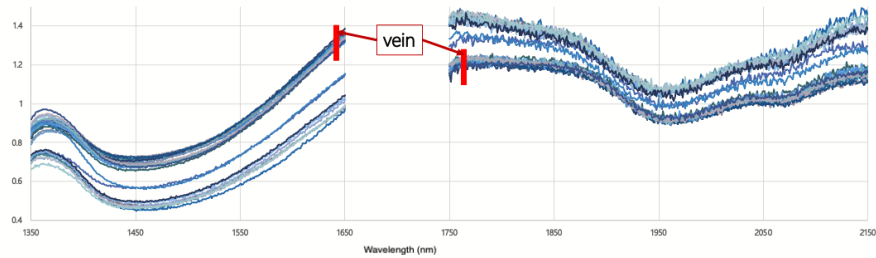
Figure 27: Bovine b8 spectra of the ear. The acquisitions from the vein area are labelled with *vein* and they are aggregated together.

In this study we compare different models to predict the different blood components.

- Linear regression model (LS) with L2 Ridge regularization

- Gradient boosting on decision trees,which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees (for this test we used 100 trees). The boosting method selected is the XGboost.

- Random Forest (RF). It is an ensemble learning method that we used for regression. Random Forest builds a set of decision trees. Each tree is developed from a bootstrap sample from the training data. When developing individual trees, an arbitrary subset of attributes is drawn (hence the term *Random*), from which the best attribute for the split is selected. The final model is based on the majority vote from individually developed trees in the forest.

- Partial Least Squares (PLS) is a regressor for data with numeric target variable. It uses latent variables, which are also called score vectors, to model the relationship between input and response variables. In the case of regression problems, PLS first generates the latent variables from the given data and uses them as new predictor variables. Here, regularization is performed with the choice of the components - the more components, the lesser the effect of regularization.

- Stack, which is an ensemble method that computes a meta model from several base models. With this method, we used as our aggregating model the linear regression.

The spectral data is preprocessed using the Standard Normal Variate normalization method. After that we used the Savitsky Golay filter to produce the first and the second derivative as well as the raw smoothed spectra.

4.1.3.3 *Results and Discussion*

We used machine learning techniques (random forest (RF), gradient boosting (GB), linear regression, PLSR, and stack) in order to predict blood parameters from NIR spectra captured from dairy cows using our spectrometer. The stacking ensemble had the highest predictive ability for the majority of the blood metabolite-related traits among the various machine learning techniques tested, as demonstrated. The result of the model predictions is shown in Table 12. Regarding the performances of the machine learning models, the prediction accuracy is obtained through random 10-fold cross-validation.

The use of the first and second derivatives gave the most performing result among the models due to the ability of these methods to emphasize regions of spectra that are relevant. Since there are no previous studies on the prediction of blood parameters from the spectra acquired on top of a cow's skin, this work gives a first look at the application of handheld spectrometers that can be applied directly on a cows body as a non-invasive tool to retrieve the status of the cow's health. For the first time, we showed the potential of a tailored NIR sensor that can be used to check stress markers on a daily basis at the farm level without the need for invasive approaches and lab testing.

In this study, we predicted blood metabolites from dairy cow skin spectra overall and developed models that have moderate predictive ability. NIR spectrometry uses light scattering principles to measure particle size, which makes it challenging to capture the nonlinear diffusive scattering produced by the skin and hair even if our spectrometer has benefits, including rapidity of acquisition, cost-effectiveness, and continuous recording. Another major issue is that near-infrared spectra are highly influenced by the presence or absence of water, and the status of the cow's skin can influence the prediction of the

blood parameters. Despite all this, we have managed to acquire spectra from dairy cows as a preliminary study and, with this study, saw that we can predict the blood components fairly. We can improve these results by using more advanced methods that require the use of a huge amount of data. For future work, we are hopeful of acquiring more data to make robust prediction models that make it possible to check the metabolic conditions of cows multiple times daily, allowing for real-time intervention to improve health status. The availability of real-time metabolic indicators is a significant advancement in the management of subclinical diseases from a managerial standpoint. Improvements to our prediction models and data acquisition could be the answer to the problem of monitoring the health of individual cows.

	Model	Total Proteins		Albumin		Globuline		AST		UREA		Creatine		Glucose		ALP		ALT		Bilirubin	
		MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2
First Derivative	Stack	11.41075	0.837447	1.113181	0.761099	10.3511	0.81557	11.66848	0.899781	0.01594	0.873745	5.982034	0.825474	0.005082	0.67548	106.7674	0.753235	9.897562	0.779975	0.002871	0.630958
	Random Forest	16.52832	0.764544	1.206555	0.74106	14.22712	0.74651	14.49262	0.875525	0.018025	0.857223	7.080124	0.793437	0.005771	0.631454	142.627	0.670354	11.83933	0.736809	0.003059	0.666817
	PLS	34.98576	0.501607	2.156701	0.537148	27.10327	0.51709	37.89134	0.674557	0.044898	0.644527	17.5902	0.486806	0.011033	0.295415	151.538	0.649759	21.25841	0.527421	0.004218	0.457814
	Linear Regression	33.64094	0.520765	1.896232	0.593047	27.13646	0.516499	34.90549	0.700202	0.027974	0.778423	11.09432	0.676323	0.00913	0.416973	126.4193	0.707814	15.20174	0.662062	0.003027	0.610918
	Gradient Boosting	20.81618	0.703461	1.624928	0.651272	23.3488	0.583985	28.08377	0.758793	0.029993	0.762427	11.96698	0.650776	0.01063	0.321156	159.5553	0.631229	20.84107	0.536698	0.004649	0.402326
Second Derivative	Stack	25.17172	0.654436	2.072874	0.549491	26.23603	0.553857	42.06047	0.637322	0.01477778	0.87663109	9.331277	0.712051	0.008265	0.470523	194.9617	0.501255	11.73171	0.713329	0.003541	0.551649
	Random Forest	38.72629	0.468356	2.788217	0.394022	29.25954	0.502442	38.2022	0.670591	0.03534684	0.70491499	10.61692	0.672378	0.008431	0.459923	239.3954	0.387587	18.00397	0.560062	0.004474	0.433487
	PLS	62.98653	0.135305	3.763894	0.181973	50.39892	0.142967	72.8281	0.372019	0.03638133	0.69627885	11.93935	0.63157	0.015746	-0.00865	267.0683	0.316795	21.44601	0.475954	0.004709	0.403847
	Linear Regression (1)	63.74701	0.124865	4.575038	0.005683	49.82557	0.152717	69.42297	0.401381	0.08246899	0.31152659	18.61164	0.425673	0.013751	0.119133	357.8329	0.084604	41.6767	-0.0184	0.007258	0.081011
	Gradient Boosting (1)	28.80208	0.604598	2.252229	0.510511	29.35187	0.500872	46.20627	0.601573	0.01693592	0.85861434	14.8947	0.540372	0.010566	0.323157	329.9598	0.155908	18.46851	0.548711	0.003267	0.586426
Raw spectra	Stack	14.53463	0.759043	2.235482	0.554221	12.1907	0.759849	20.80497	0.840599	0.015867	0.868576	6.621225	0.807955	0.005284	0.681289	151.5652	0.67083	11.37082	0.764191	0.002656	0.64841
	Random Forest	15.97484	0.735167	2.150433	0.57118	14.15496	0.721153	21.2154	0.837454	0.016565	0.862795	8.373374	0.757135	0.004996	0.698685	139.6788	0.696645	11.92426	0.752714	0.002772	0.653003
	PLS	22.96123	0.619346	2.547773	0.491947	19.51766	0.615511	40.5434	0.689369	0.038211	0.6835	8.530806	0.752569	0.009347	0.436278	152.3406	0.669146	17.04173	0.646587	0.002981	0.605319
	Linear Regression	65.51192	-0.08607	2.425784	0.516272	51.99239	-0.02423	67.31822	0.484229	0.051969	0.569543	14.69055	0.573909	0.020079	-0.21104	301.5158	0.345168	34.19725	0.290815	0.003963	0.475294
	Gradient Boosting	24.4137	0.595266	2.589289	0.483668	14.88754	0.7706722	31.32303	0.760013	0.02988	0.752504	11.30817	0.672013	0.007182	0.566836	248.0684	0.461245	10.24422	0.787554	0.005193	0.312492

Table 12: Result of the different regression models for the prediction of each blood components grouped by the prepossessing step.

4.2 GLUCOSE MONITORING

In this section, we have shown the application of our MEMS-based NIR spectrometer in the healthcare field for plants and animals. We have also considered using our miniaturized spectrometer in the human wellness area to detect blood glucose levels as a non-invasive method. Why? The glucose level in the blood is generally measured by a skin-pricking tool that reaches the capillaries in the dermal section of our skin tissue leading to the risk of infection [36]. People with diabetes monitor their glucose levels on their own. Diabetes mellitus has been steadily increasing in the recent decade. In the WHO European Region, nearly 62 million people live with diabetes. The prevalence of this disease is growing throughout the Region, reaching rates of 10-14% in some states. This growth, partly due to the general aging of the population, is mainly secondary to the diffusion of conditions at risk, such as overweight and obesity, incorrect diet, physical inactivity, and socioeconomic inequalities (type 2 diabetes). In 2021, over 1.1 million deaths in Europe were caused by diabetes, the fourth leading cause of death in the European Union.

According to ISTAT 2020 data, the prevalence of diagnosed diabetes in Italy is approximately 5.9% (5.9% in men, 5.9% in women), equal to over 3.5 million people, with a trend in a slow increase in recent years. The prevalence increases with age, up to 21% in people 75 years or older [88]. The development of a reliable method for non-invasive blood glucose monitoring has a tremendous impact on the wellness of diabetic people, and it is a growing topic. An overview of the blood glucose monitoring method available is shown in Figure 28.
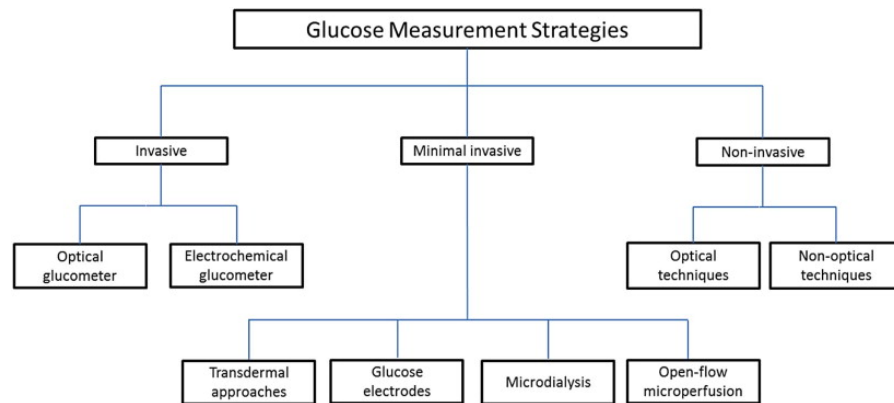
Figure 28: Overview of blood glucose monitoring methods. Adopted from [33]

One of the non-invasive techniques for blood glucose monitoring is the optical one, including near-infrared technology. Since NIR waves can reach deep in the dermal part of the skin tissues where capillaries are present. NIR sensors are able to detect glucose, thanks to the in-

teraction of light with glucose molecules in the skin tissue, as shown in Figure 29 from [71].
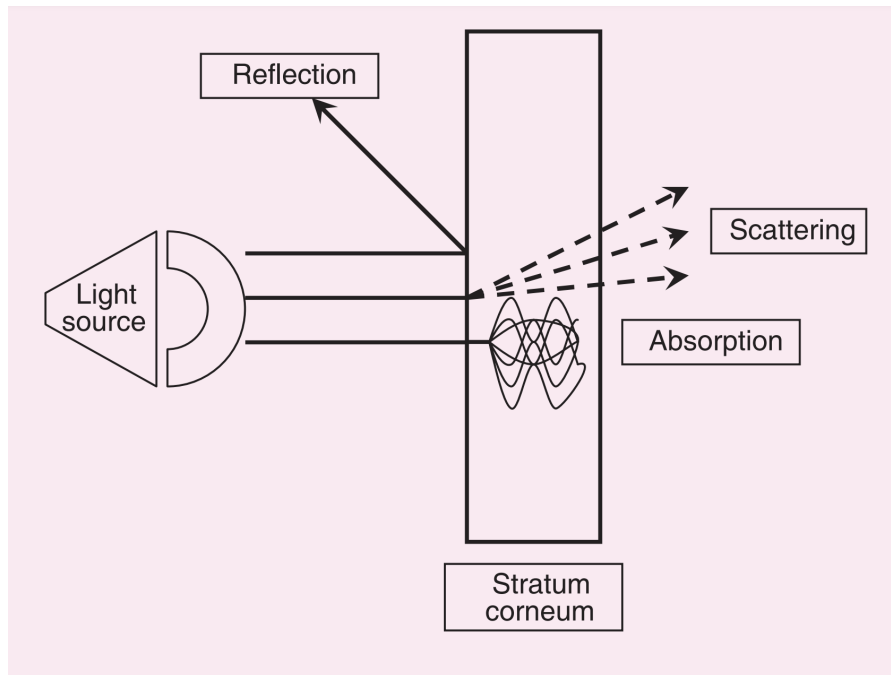
Figure 29: Light propagation through skin tissue. Adopted from [71]

Nevertheless, this technology suffers from scattering, making it hard to estimate glucose levels [38], and the presence of water absorption in the NIR band makes it very challenging.

For instance, we made some basic skin spectra analyses. We carried out some basic exploratory tests. For instance, we made measurements with our spectrometer near the wrist area to try and distinguish areas with visible blood vessels and areas smooth areas, positioning the device accordingly. In Figure 30, we can see the visible vessel blood area with the layout of our device pressure-marked on the skin.

In Figure 31, the result of our test also wants to highlight the behaviors of spectra obtained when pointing the device in veins ad no vein area of the wrist. We can see that from 1700 nm to 1850, there is evidence of the spectra of the two main groups. Most probably, it confirms that the NIR light has reached the dermis.

There are many studies that use advanced methods to predict glucose levels in the blood [94] [76] [46] [81]. But it is a big challenge to reach clinical-level accuracy to make these technologies available for the market. This area of research is very challenging and open to be exploited. With that being said, in the future, there will certainly be safer and simpler methods for noninvasive blood glucose detection, which will result in as accurate detection method as invasive methods nowadays.

Figure 30: Application of our spectrometer in wrist area to detect directly blood vessels. Here is also show the mark on the skin where outlined by our spectrometer.
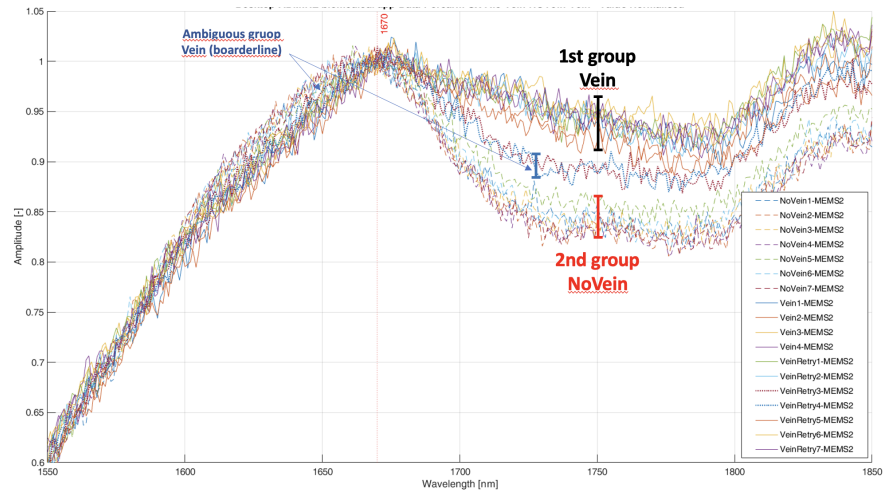
Figure 31: Difference of spectra between vein area and no-vein of a single subject. Spectra measured from the internal wrist area. The dashed are the case with no visible external veins; the continuous line is the veins visible on the outside, while the dotted case is an area where the external veins were slightly less dense than in the previous cases. The *VeinRetry5* case is also slightly ambiguous and remains very much on the edge of the block.

Part V

INDUSTRIAL APPLICATION

INDUSTRIAL APPLICATION

<div style="text-align: right">5</div>

Since its discovery and application, Near-infrared ($12{,}500$–$4{,}000\,\mathrm{cm}^{-1}$; 800–$2{,}500$ nm), commonly abbreviated as NIR, has distinguished itself as one of the most modern analytical techniques with solid prospects for further expansion. It has been around for decades, and much research has been done in this field [7, 72, 97, 100, 109]. Its universality, vast applicability, uncomplicated instrumentation, low time-to-result, and low-cost factors are prominent advantages of qualitative and quantitative analysis. Daily, spectroscopy is performing analyses impossible by any other method. More common analyses are completed in a few minutes, which previously required hours. Numerous applications of this methodology have been eminently successful and have become familiar to many chemometricians.

NIR spectra of biological materials are signals composed of peaks because of molecular vibrations of mostly O-H, C-H, and N-H groups [18, 54] caused by their interaction with infrared light within the NIR wavelength region. The spectral data measured in this region are generally composed of high noises and overlapping peaks associated with the sample's chemical composition. In recent years, it has been employed in many industries to analyze the chemical composition of organic samples, drugs, food, and other compounds. In particular, the food industry is used for the quantitative and qualitative analysis of foods such as meat, fruit, grain, dairy products, and beverages [16, 40, 120].

However, a significant challenge with NIRS is the mixed physico-chemical phenomena captured by the interaction of light with matter. The interaction often results in both absorption and scattering of the light. A mixture of materials can be analyzed quickly and accurately so long as the components present in the mix are known. From a study of the spectra of the known compounds, it is usually possible to find a frequency at which only one component possesses strong absorption and thus find its quantity in a mixture. This rapid method, combined suitably with deep learning methods, has shown a high accuracy detecting the combination percentages of a pair of organic mixtures [98]. The overall NIRS signal contains information related to the two phenomena mixed. Therefore, when the data analysis aims to predict chemical components, it is necessary to remove the scattering effects from the spectra as much as possible. There are several approaches to solving the prediction of chemical composition. Figure 32 shows a schematic view of our signal processing pipeline. The

dataset containing all the spectra is treated with filtering, scaling, and (or) data transformation method.
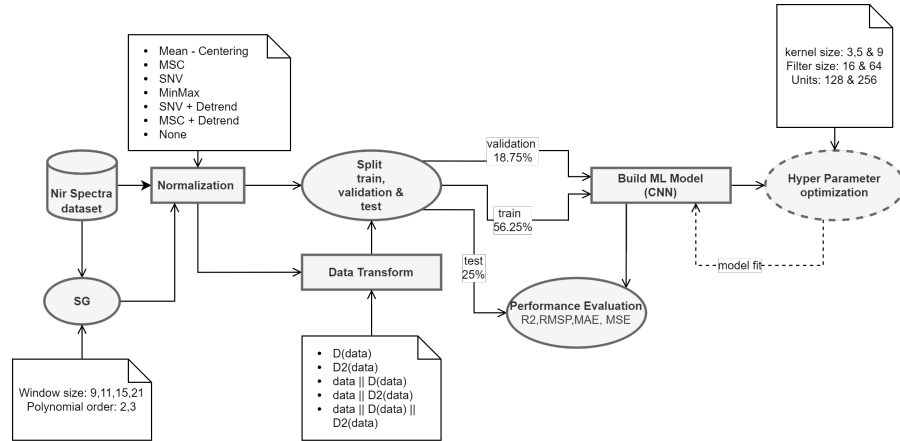
Figure 32: Pipeline of our processing steps

5.1 MATERIALS AND METHODS

This section presents the data collection process and the neural network model architecture used to predict the quantitative measure for the mixed organic materials. The data collection is made of several steps that include the sample preparation procedure of the six organic powders, the data acquisition that describes the mechanism of acquiring the spectral data. Furthermore, we describe the 1D-CNN architecture and its parameters. Furthermore, we describe the convolutional neural network architecture and its parameters, the Savitsky Golay filter, and the use of derivatives.

5.1.1 *Sample Preparation*

Each sample was prepared by carefully mixing a given fraction in weight of two base materials and placing the mixture in a container of the Petri dish type as described here [98]. This work it is also included the new mixtures of three base materials, as also mentioned in [23]. However, because of the unique characteristics of the powders used, such as grain size and tendency to form lumps, it is impossible to guarantee that the mixture is homogeneous.

We made fifteen pairwise combinations using cocoa (Cocoa), ice sugar (IceSugar), baby milk powder (BabyMilk), potato starch (Potato), rice starch (Rice), and baking soda (NaHCO3). We added other pair mixtures to the first dataset totaling 69 samples. Their composition percentage is made up from set: $P = \{15, 25, 33, 35, 40, 45, 50, 65, 75, 85\}$. The composition percentage of a given mixture of two materials adds up to 100%, e.g., (A=15, B=85), where A and B are the mixed mate-

rials. Moreover, we added six other mixtures using three materials in different compositions, and their percentages were retrieved from set P. Their composition adds up also to 100%.

E.g., (A=33, B=33, C=33) or (A=45, B=40, C=15), where A,B and C are the three mixed materials.

Table 13: The pairwise mixtures overview. Value 1 indicates the presence of a powder mixture, while 0 means that the powders are not mixed. The diagonal values correspond to the base materials at 100%.

	BabyMilk	IceSugar	NaHCO3	Cocoa	Potato	Rice
BabyMilk	1					
IceSugar	1	1				
NaHCO3	1	1	1			
Cocoa	1	1	1	1		
Potato	1	1	0	1	1	
Rice	1	1	1	1	0	1

5.1.2 *Data Acquisition*

We took the measurements using an automatized mechanical setup. Figure 33 illustrates the tools used to collect spectral profiles. We put the NIR sensor and the Petri dish into a dark box to avoid external interference. The data is then viewed using custom-made software for the sensor.
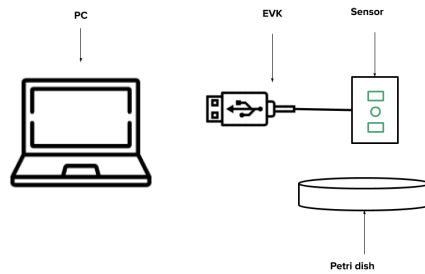
Figure 33: The figure shows the tools used for collecting NIR spectrum data. The setup comprises a Petri dish where the powders are mixed and put inside, the sensor that collects NIR spectra, and an Evaluation kit (EVK) to transfer data from the sensor. Custom-made PC software to visualize the spectrum.

5.1.2.1 *Sensor Used*

We calibrated the NIR sensor with SRS-99-020 Reflectance Standard (a white diffuse reflectance sample) by collecting the spectra of the

"white" reference at the minimum distance allowed by the scanner and with the maximum level of light bulb ignition. We used a device that captures two ranges of wavelength points: $[1350 - 1650]$nm and $[1750 - 2150]$nm. The total number of wavelength points captured is 702.

5.1.2.2 *Capture Mode*

We set the Petri dish and the NIR sensor inside a box and inside a dark room along with the automatic acquiring mechanical setup to avoid any outside interference. The NIR sensor captured the spectra of the samples at different sensor-sample distances, moving the vertical axis of the scanner with 1mm pitch along the entire 20mm useful range. The minimum distance between the outer surface of the Petri dish window and the detector is 5mm. For each distance, we examined three different areas of the sample's exposed surface. We have acquired three spectra for each zone, hence there are nine spectra acquired at the same distance from each sample. Assuming the sample is homogeneous, the spectra of the same mixture should be the same with each other, with minor differences due to measurement noise. Instead, we found that there are some differences, particularly between zones. In isolated cases, we also found differences between spectra acquired at the same location, probably because of measurement errors related to electrical disturbances or mechanical vibrations. We measured all samples by varying the level of lamp intensity (parameter varied between 200 and 250 with step 1).

5.1.2.3 *Reflectance Values*

The measured reflectance R of a generic sample is calculated by the sensor at each wavelength as follows:

$$R = \frac{I_C - I_F}{I_{SRS} - I_F} R_{SRS} \tag{2}$$

where I_C is the intensity of light received from the sample, I_{SRS} the intensity of light received from the reference SRS-99-020 placed at a given distance during calibration, I_F the intensity of background light (cross-talk), measured during calibration with the lamp on but without a target, and R_{SRS} the reflectance of the reference SRS-99-020. In practice, since cross-talk levels generally have small values, the measured reflectance is proportional to the ratio of the light intensity received by the sample in question to the one measured during calibration using the SRS-99-020 reference. For this reason, although reflectance is an intrinsic property of the sample, the reflectance measured using the same sample at different distances from that used during calibration is different. A similar effect occurs if the measurement uses a different light bulb ignition level than when calibrating.

5.1.2.4 *Dataset*

We took the measurements using an automatized mechanical setup. We put the NIR sensor and the Petri dish, filled with specific powder, into a dark box to avoid external interference. Following the material preparation and the acquisition of the NIR spectra, we collected 506160 samples[1]. Each material has 702 features corresponding to the captured wavelengths, and for each composition percentage of a mixture, we have ~ 7000 samples. The target variable of each sample is a percentage distribution over the six base materials describing the quantity of that material in the spectral sample. Given that each spectral sample represents only the mixture of two or three materials, only two or three elements in the target vector contain the value of the individual materials described in the spectral sample. At the same time, we set the remaining four base materials to 0. Whereas for the mixtures containing only one material, we set the five remaining target variables to 0 and assigned the value 100% to the material represented by the pure material spectra.s containing only one powder, we set the five remaining target variables to 0 and assigned the value 100% to the material represented by the spectra.
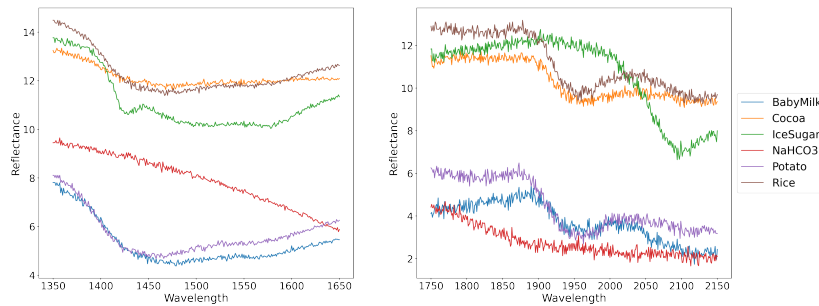
Figure 34: Reflectance values of each base material in the range of 1350 nm to 2150 nm.

5.2 PREPROCESSING

When analyzing near-infrared spectra, a fundamental step is its preprocessing, which is integral to building a predictive model. The purpose of preprocessing spectra is to remove unwanted physical phenomena in their wavelength ranges to improve the subsequent multivariate regression, classification model, or exploratory analysis. The most widely used preprocessing techniques can be divided into scaling and data transformation methods. We compare current preprocessing methods' theoretical and algorithmic foundations and their application's qualitative and quantitative consequences.

---

1 The dataset is available upon request.

Several preprocessing techniques are available to do this, but it is often difficult to decide which one to choose. In this work, we present the use of a recently developed preprocessing approach to improving the predictive power of multivariate models based on NIR spectra of food materials.

5.2.1 *Savitzky Golay*

One of chemometrics most commonly used and frequently cited filters is the Savitzky Golay smoothing, and differentiation filter [90]. The filter is often used to preprocess spectroscopy and signal processing. The filter can reduce high-frequency noise in a signal due to its smoothing properties and reduce low-frequency signal (e.g., due to offsets and slopes) using differentiation. After a brief description of the filter, we show the smoothing aspects of the filter and the differentiation filtering.

Unlike simple digital filtering methods, the Savitzky-Golay filter preserves higher-order moments around inflection points that a simple digital FIR filter cannot. It can better preserve features - like local maxima and minima - through a least-squares polynomial fit around each point. Also, unlike a moving average, in estimating the value of the fit at a certain point, the Savitsky - Golay filter does not introduce a bias at inflection points [3].

Many applications implemented the proposed polynomial smoothing and differentiating functions by Savitzky and Golay, primarily as smoothing filters [4] and performing numerical differentiation. Smoothing and differentiation are essential in various fields, such as signal processing, imaging processing, analytical chemistry, and spectroscopic analysis. To accommodate these, Savitsky and Golay developed a digital filter.

For a given signal measured at N points and a width filter, *w*, *savgol*, the method provided by scipy calculates a polynomial fit in each filter window as the filter is moved across the signal. In Figure 35, we can see a single spectrum of baby milk powder in grey and its filtered signals with Savitsky-Golay, with a window size of 9 and 21, respectively, in blue and green lines.

The polynomial fit of each window *w* is done using the Least of Squares (LS) estimate between the X matrix and the y vector:

$$y = Xb \tag{3}$$

where the matrix X is the so-called Vandermonde matrix. From the equation above, we calculate the coefficients:

$$b = (X^T X)^{-1} X^T y. \tag{4}$$
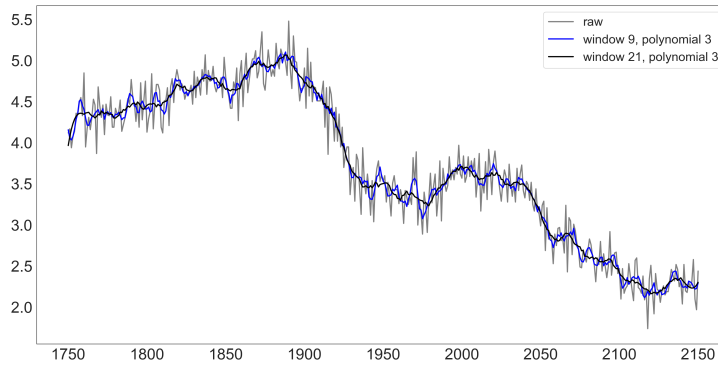
Figure 35: Effect of Savitsky Golay filter on Baby milk powder. The grey line represents the raw spectra of the powders. The lines in color blue and black are, respectively, the application of the Savitsky Golay filter with a window size of 9 and 21 points.

Replacing the Equation 3 in Equation 4 we get the estimated values of the polynomial fit:

$$\hat{y} = Xb = X(X^TX)^{-1}X^Ty = Hy \tag{5}$$

The product $H = X(X^TX)^{-1}X^T$ is also called the hat matrix and is the same for any y for a given polynomial. So it can be calculated once and stored for a latter application. This is what Savitzky and Golay have done for polynomials of various orders and pieces of different lengths n. The $(n+1)$th row of the H matrix gives the tabulated coefficients for the Savitzky-Golay filters. We only use the estimate for the middle point of the moving window for smoothing. The other rows are used only for smoothing the signal's endpoint when fewer data points are left than the window size $2n + 1$.

The fitted polynomial coefficients can be used to calculate smoothed first and second derivatives of the signal. In Figure 36 it is shown first derivative of one spectra for each base materials, and in Figure 37 the second derivative of the same spectra.

In the field of spectra preprocessing, derivatives play a valuable part. Using them can increase the performance of discrimination tasks, enhance significant spectra peaks, and have a detrending effect on the spectra [84].

5.2.1.1 *Scaling Methods*

Scaling data is frequently used to compensate for differences in the sample surface optical characteristics, such as density, scatter, or varying smoothness/roughness. While looking at the differences between
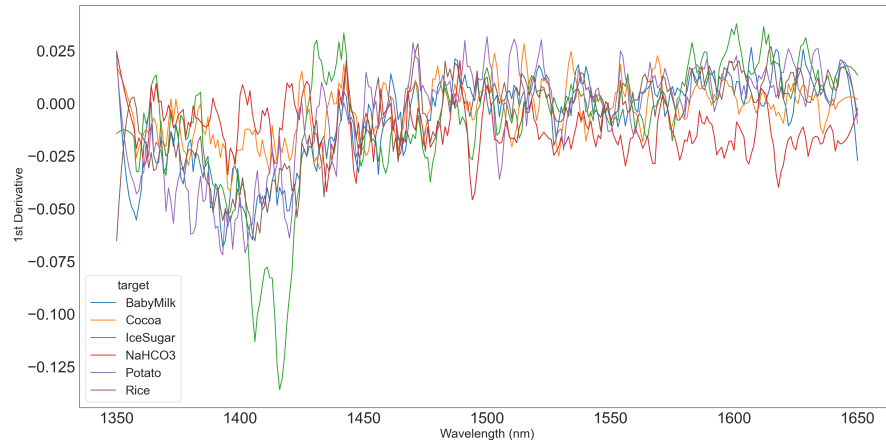
Figure 36: First Derivative of one spectra for each base material.
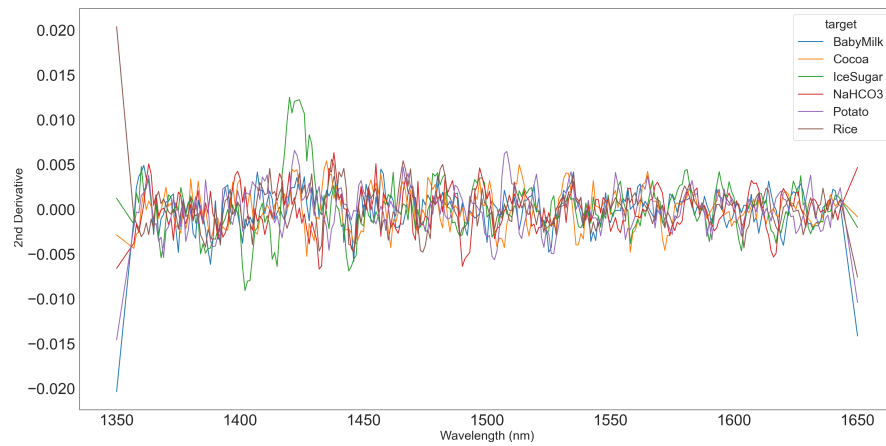
Figure 37: Second Derivative of one spectra for each base material.

the spectra of the different organic pure powders, we can see how the different scaling methods affect and help to differentiate each spectrum, as shown in Figure 39. The different scaling methods proposed each affect the signal in a particular way that can benefit the prediction model, as shown in Figure 38.

5.2.2 *Multiplicative scatter correction (MSC)*

Multiplicative scatter correction (MSC) is a signal processing algorithm that is particularly useful for reducing the non-linear scatter present in both transmission and reflectance spectra. MSC performs a linear transformation of each spectrum to best match each signal to the mean of all the spectra in the data set. This method is often used for spectra measured in diffuse reflection and is considered the best technique for removing scatter signals from the chemical absorbance. In our case, the data is reasonably well-behaved, which allows us to
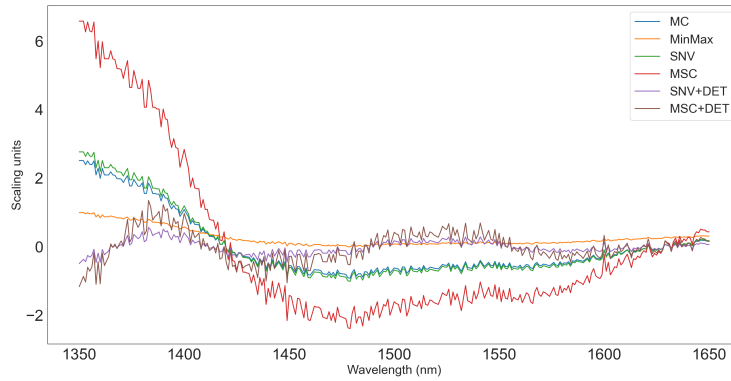
Figure 38: The different scaling methods applied on BabyMilk powder.
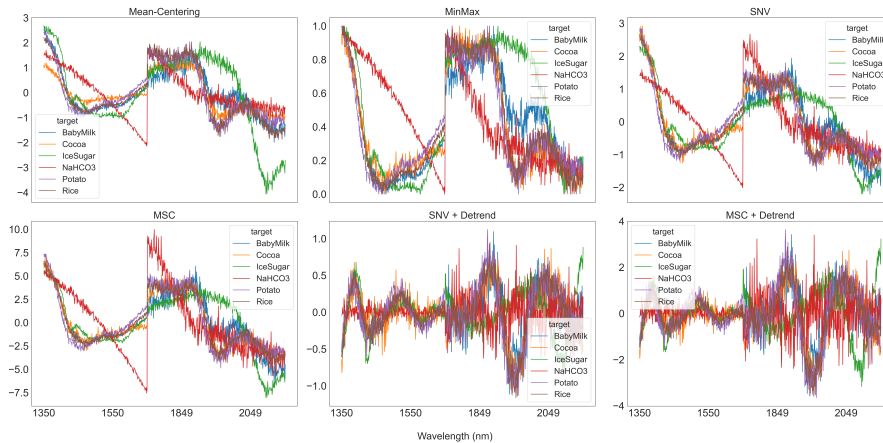
Figure 39: The different scaling methods applied on each powder show a visual representation of spectra difference.

take the average spectrum so that it can be a close approximation to the ideal spectrum we are after. The particle size of the powder mixtures and path length effects should vary randomly from sample to sample. Therefore the average should reasonably reduce these effects, at least in the approximations that these effects are genuinely random. This is the primary assumption behind MSC.

Mathematically, if we call X_m the mean spectrum, the multiplicative scatter correction is done in two steps. We first regress each spectrum X_i against the mean spectrum. This is done by ordinary least squares: $X_i \approx a_i + b_i X_m X$. Then, We calculate the corrected spectrum $X_i^{msc} = (X_i - a_i)/b_i$. All these spectra generally have a non-zero mean, so we can optionally mean-center the spectra beforehand.

5.2.3 *SNV*

Light scattering due to the interactions between IR radiation and sample particles often creates a shift of absorbance levels that could make spectral interpretation and linear calibration of NIR diffuse reflectance spectra more difficult. Light scattering results in path-length variations that lead to a background signal level that varies with the wavelength producing baseline shift and curvature, which may vary significantly within and among samples [49]. The SNV transformation was introduced by Barnes et al.22 to reduce the multiplicative effects of scattering and particle size and also to reduce differences in the global intensities of the signals. Normalizes a spectrum by calculating the average intensity value and subsequently subtracting this value from each spectrum. Then, the sum of the squared intensities is calculated, and the spectrum is divided by the square root of this sum (the standard deviation). SNV correction is done on each spectrum, and a reference spectrum is not required.

Mean center each spectrum X_i by taking away it's mean $\bar{X}_i$. Then Divide each mean-centered spectrum by its standard deviation:

$$X_i^{snv} = \frac{(X_i - \bar{X}_i)}{\sigma_i} \tag{6}$$

Mathematically, it is identical to autoscaling (i.e., normalization) of the rows instead of the columns of the matrix. The scattering is removed by normalizing each spectrum with the standard deviation of the responses across the entire spectral range [31]. The benefit of SNV is that the normalization scheme is based on a sample spectrum alone. On the other hand, SNV is often influenced by a low signal-to-noise ratio to enhance unwanted spectral features arising from noise or interference in samples because of the standard deviation in Equation 6.

5.2.4 *Min Max normalization*

Min–Max normalization is by far the most simple normalization method. In Min–Max normalization, spectra are first offset-corrected by setting the minimum intensity of the whole spectrum or a defined spectral region to zero. Spectra are then scaled with the maximum intensity value equaling one.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{7}$$

5.2.5 *Mean Centering*

Mean-centering is one of the most common pre-treatment procedures involving subtracting the variable averages from the data. Since multivariate data is typically handled in table format (i.e., matrix) with columns as variables and rows as samples, we can consider mean-centering row centering. The outcome of this method is the transformation of each row so that the resulting spectra will have a zero mean. The generic equation for mean centering is:

$$X_c = X - \overline{X} \tag{8}$$

5.3 MODEL

The spectral information extracted from these broad peaks for the quantitative determination of the chemical composition is often analysed using chemometry and other linear based methods (i.e. partial least square, multivariate regression) to capture the various possible infrared spectra patterns of a single material.

There are many applications in the field of NIR spectroscopy and mixture analysis. However, this work focuses on their use in the food analysis industry. Initial works in this field used Multivariate Analysis (MVA) like Principal Component Analysis (PCA), Partial Least Square (PLS), and Support Vector Machine (SVM). For instance, [112] used the least squares support vector machine (LS - VM) to analyze milk powder's NIR spectra and determine fat, protein, and carbohydrate contents. [109] and [82] assessed the potential of NIR to determine the quality of rice using respectively PLS and a multi-linear regression (MLR). [102] used rapid methods, like NIR technology combined with multivariate analysis (PCA and partial least squares discriminant analysis (PLS-DA)), to detect fraud of cocoa powder.

Other methods improved the multivariate analysis by using kernel-based methods like Support Vector Machines (SVM): [96] investigated the feasibility of NIR spectroscopy combined with kernel PLS regression algorithm for the quantitative determination of reducing sugar content in potato flours.

Further improvements are revealed by models that use machine learning. Machine learning approaches for spectral profiles analysis [**galli2016support**, 30, 42], and in particular, Convolutional Neural Networks (CNN) for spectroscopy signal classification have reported promising results in the literature. [121] proposed a one-dimensional CNN (1D-CNN) to classify the origin of tobacco using their NIR spectrum and concluded that the performance of 1D-CNN and 2D-CNN was better than traditional PLS models. Similarly, [69] use a 1D-CNN to perform a regression task, instead of a classification task as [121], to find the amount of nitrogen in the Masson pine seedling leaves

using NIR spectrum. We can view a summary of works that use the NIR spectrum with their relative task and methods in Table 14.

Reference	Sample	Range	Target	Task	Method	Performance
[82]	Indica Rice	540,640,970 nm	Indica Rice	Reg	MVA	R^2 = 0.71
[112]	Infant Milk Powder	800-1025 nm	Fat, Protein, Carbohydrates	Reg, Class	ICA-LS-SVM	R^2 = 0.985, 0.231, 0.982 Pred = 98%
[34]	Olive Oil	569,835 nm	Peroxide, Phenol	Reg	Undefined (Excel)	R^2 = 0.885 (Peroxide), 0.895 (Phenol)
[102]	Cocoa Powder	1100-2500 nm	Adulterant	Reg,Class	PLS-DA	R^2 = 0.974, Acc = 98%
[96]	Tubers	1000-2500 nm	Sugars	Reg	PLS, LS-SVM	R^2 = 0.950
[105]	Wheat and Potato Flour	1000-2500 nm	Potato Flour	Reg	PLS	R^2 = 0.8865
[100]	Rice	740-1070 nm	Rice	Class	KNN, SVM	ACC = 91.88%
[68]	Soil	350-25000 nm	Sand, Clay, Organic Content	Reg	PLSR,Cubist-tree, CNN	R^2 = 0.85~0.95, 0.91~0.97, 0.95~0.98
[66]	Masson Pine Seedlings	780-2500 nm	Nitrogen	Reg	1D-CNN	R^2 = 0.984
[60]	Rice	1000-2500 nm	Rice	Reg	PLS-DA, SVM	R^2 = 0.96
[56]	Cumin, Black Pepper	1100-2500 nm	Adulterant	Reg	MLR,PLS	R^2 = 0.90
[121]	Tobacco	1000-2500 nm	Tobacco	Class	1D-CNN,2D-CNN	Acc = 93.15%, 93.05%

Table 14: Relevant works in the field of Near-infrared for the qualitative and quantitative analysis of organic materials

5.3.1 *Convolutional Neural Network (CNN)*

In this work, we propose a modified version of the 1D-CNN model proposed by [68]. Compared to our model, [69] performs regression on a single variable (the nitrogen content). Their model is used on data coming from the visible/near-infrared, mid-infrared, and a combination of both, while ours uses information only on NIR data. Moreover, they do not perform experiments with unseen combinations and unseen percentages.

We proposed a 1D-CNN network with some modifications that was used in [19]. The 1D-CNN in Figure 40 model proposed is mainly inspired by the philosophy of 1D-CNN [47]. It consists of seven trainable layers - five convolutional layers and two fully connected layers. Our model's primary objective in implementing a grid search approach is hyperparameter optimization. Hyperparameter optimization is applied to find the optimal values of the kernel and filter size from a discrete set. These are values that can control the learning process, and tuning them ensures that the model can optimally solve a problem by minimizing the loss function used and giving accurate results. The input of the 1D-CNN is a one-dimensional spectral vector containing values of the 702 wavelength points, and the target is also a one-dimensional vector containing the percentage distributions of the six materials. Every convolutional layer is there to capture patterns. For example, the first layer captures patterns like edges, corners, dots, etc. Subsequent layers combine those patterns to make more extensive patterns (like combining edges to make squares, circles, etc.). Now, the deeper the network gets, the more complex the patterns become; hence, we applied hyperparameter optimization to capture as many combinations as possible. The output of our model is a percentage distribution of the six materials that represents precisely the content amount of the materials in the spectra.
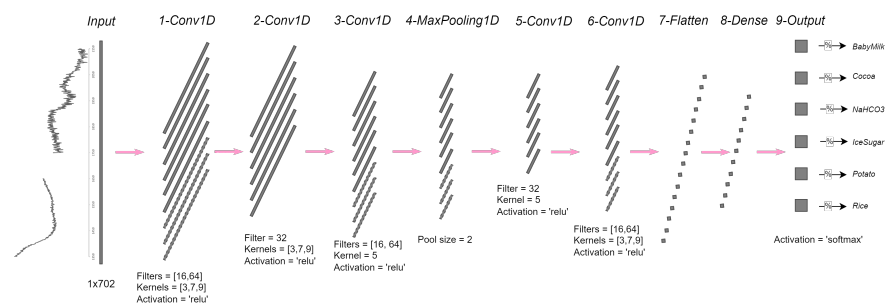
Figure 40: 1D-CNN architecture. Boxes with dashes represent whether a hyper-parameter search of filter size is made.

5.4 EVALUATION INDICES

Training, validation, and testing sets were created to assess the model's predictive qualities. Model training was performed based on 56% of the total spectra. Model validation and testing to value the performance of the pre-processing procedures and our 1d-CNN model were evaluated by evaluating the differences between observed and predicted values after fitting the model using respectively 18% and 25% of the entire data set. The applied indices were the coefficient of determination (R2), the root means square error (RMSE), the mean square error (MSE), and the mean absolute error (MAE). R2 (the coefficient of determination) is a commonly used measure for the goodness of fit. The RMSE is also commonly used to measure differences between the calculated and observed values from the measured model and indicates the accuracy of the calculation.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_1 - \overline{y_1})^2}{\sum_{i=1}^{n}(\hat{y}_i - \overline{y_1})^2} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=t}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{10}$$

.

5.5 RESULTS SCALING AND DATA TRANSFORMATION METHODS

The results are grouped into two. First, we present the results obtained from the different scaling (normalization) methods adopted and how these compare with the use of the Savitsky Golay filter. Secondly, we show for each normalization method the result of applying data transformation. The data transformation is done by applying the Savitsky Golay filter to obtain the derivative of each spectral data. As mentioned before, these techniques were divided into the data treatments of data scaling and transformation.

5.5.1 *Scaling*

Pre-processing techniques were used to improve the spectral characteristics and optimize the relationships with the intrinsic spectra properties and the distribution of percentages contained in the spectra. Table 15 shows the result of the normalization methods adopted and their relative parameter used for the model. For each scaling method, a hyperparameter optimization process was applied to obtain the optimal parameter for the specific scaling method.

For each scaling method used in Table 15, we compared the results with the use of the Savitsky Golay filter. Conducting a grid search on

Scaling	Num of Filters	Kernel Size	Dense Layers Units	MAE - Train	MAE - Validation	MAE - Test
Mean Centering	64	9	256	0.0062	0.0064	0.0055
SNV	64	9	128	0.0067	0.0067	0.0070
MSC	64	7	256	0.0083	0.0086	0.0096
None	16	9	256	0.0098	0.0090	0.0086
MinMax	16	3	256	0.0098	0.0098	0.0093
Detrend	64	3	256	0.0098	0.0107	0.0109
MSC + DET	64	3	128	0.0107	0.0108	0.0104
SNV + DET	64	7	256	0.0103	0.0109	0.0100

Table 15: Result of hyperparameter tuning performed on the different scaling methods, ordered by the smallest mean absolute error on the test set.

the window size of the filter and its polynomial degree. The results are seen in Table 16

Normalization	Polynomial Order	Window Size	MAE - Train	MAE - Validation	MAE - Test
Mean Centering	3	15	0.0055	0.0050	0.0046
SNV	2	11	0.0069	0.0073	0.0061
MSC	3	9	0.0073	0.0068	0.0065
MinMax	3	9	0.0097	0.0097	0.0087
None	2	11	0.0101	0.0100	0.0090
SNV + DET	3	9	0.0091	0.0090	0.0093
MSC + DET	2	9	0.0111	0.0114	0.0107

Table 16: Results of grid search on the polynomial order and window size of the Savitsky Golay Filter. Applied on each scaling method.

5.5.2 *Data Transformation*

Baseline shift is a very common effect in spectroscopy that is caused either from the instrument (lamp or detector instabilities) or sample handling (cuvette repositioning) effects. Because the first derivative of a constant absorbance offset is zero, using the first derivative spectra always eliminates such baseline shifts and improves the accuracy of quantification. In Table 17 is shows the result of applying derivatives on our spectral data and combining them with the relative raw data. This is done for each scaling method.

5.6 RESULTS: SCENARIO BASED

The quantitative analysis of the NIR spectra based on the multi-modal 1D-CNN is studied in three scenarios. The first scenario, Whole Prediction (*WP*), where we test the model ability to predict the same composition percentages of mixtures seen in the training set. In this scenario, given two materials A, B, and their combination (A, B), the model sees the spectral profiles with the same percentage and combi-

	D1	D2	D1 ‖ D2	Raw ‖ D1	Raw ‖ D2	Raw ‖ D1 ‖ D2
Mean Centering	0.0084	0.0213	0.0082	0.0064	0.0066	0.0070
SNV	0.0080	0.0159	0.0084	0.0073	0.0080	0.0058
MinMax	0.0135	0.0212	0.0140	0.0084	0.0081	0.0084
MSC	0.0082	0.0229	0.0084	0.0073	0.0078	0.0087
SNV + DET	0.0193	0.0281	0.0196	0.0101	0.0102	0.0114
MSC + DET	0.0230	0.0322	0.0198	0.0107	0.0108	0.0117

Table 17: Result of the application of derivatives on spectral data and its combination with the relative raw data. The results are ordered by the smallest mean square error on the test set.

nation (A=25, B=75)[2] in training and in testing set. The WP is used as a baseline for the other scenarios.

We used the second scenario, Unseen Percentages (*UP*), to test the model's ability to predict the unseen composition percentage of the same mixtures. In this scenario, the model sees the same mixtures of materials in the training set, but with different composition percentages. For example, with the combination set {(A=15, B=85),(A=35, B=65), (C=85, D=15)} in training set, the model is tested using {(A=25, B=75),(A=50, B=50), (C=75, D=25)}.

The last scenario, Unseen Percentage and Mixture (UPM), used to test the model's ability to predict the unseen composition percentage of unseen mixtures. UPM is similar to UP, but here the primary goal is to train and test the model with different composition percentage and mixtures. This means that the model will see the following pairs and percentages in the training set {(A=25, B=75),(A=50, B=50), (C=75, D=25)}, and in the testing set it will only see {(A=15, C=85), (A=35, D=65), (B=85, C=15)}.

For each scenario is reported a table that summarises the model's overall performance for each material, this is done by filtering the model's output for each of the six materials for the corresponding scenario. The full output of each mixture for each scenario can be retrieved from the GitHub repository here or upon request [3] [4].

5.6.1 *Whole Prediction (WP)*

The first scenario (WP) we used all the composition percentages of all mixtures. We use this scenario as a baseline result for the next experiments, as this one shows the ideal case where both the mixtures and the percentages have been seen during training.

Given the 454896 samples, we split them into train, validation, and test set, respectively with the ratio of 45%, 22%, and 33% which is a

---

[2] The mixture of two materials A and B, where A is at 25% of the total composition and B at 75%.

[3] https://github.com/dtegegn/CNN-NIR-Spectra

[4] https://www.seletech.com

good amount of data to train and test the model's performance. All the three dataset have the same type and balanced number of the 62 different composition percentage of all mixtures including the six base materials of 100%. In Table 18 we report the metrics of the overall performance of our model for this experiment. We can see that most of the materials have good performances except for Babymilk, which has a higher Mean Absolute Error (MAE).

Table 18: The overall performance for the WP scenario.

Material	MAE	MSE	RMSE	R^2
BabyMilk	0.0161	0.0009	0.0298	0.9910
IceSugar	0.0065	0.0002	0.0143	0.9976
NaHCO3	0.0046	0.0001	0.0114	0.9983
Cocoa	0.0089	0.0005	0.0228	0.9941
Potato	0.0060	0.0003	0.0167	0.9961
Rice	0.0073	0.0004	0.0189	0.9958

5.6.2 *Unseen Percentage (UP)*

The WP experiment settled the baseline result for the prediction for all the mixtures and their quantities, and we use it to see the best case scenario for our model. Therefore, we created specific subsets from the whole dataset used in WP, by filtering specific set of composition percentage of the mixtures, therefore we can test the model's performance on predicting unseen composition percentage of the same mixture and compare it with the baseline scenario. Thus, in the second experiment UP we created two subsets of $P = \{15, 25, 35, 50, 65, 75, 85\}$: $P1 = \{15, 35, 65, 85\}$ and $P2 = P - P1 = \{25, 50, 75\}$, then defined the set MP1 as the mixtures of materials belonging in set P1 and used it to train the model. We also defined the set MP2 that comprises of the mixtures of materials belonging to set P2 and the six base materials at 100%, and used it for testing. We use 33% of the training set MP1 for the validation set, totaling 118380 samples for the training, 58308 for the validation and 278208 for the testing sets.

Table 19: The overall performance for the UP scenario.

Material	MAE	MSE	RMSE	R^2
BabyMilk	0.166	0.013	0.115	0.865
IceSugar	0.060	0.003	0.052	0.969
NaHCO3	0.046	0.001	0.031	0.989
Cocoa	0.069	0.003	0.058	0.950
Potato	0.041	0.002	0.046	0.969
Rice	0.062	0.004	0.063	0.941

With the UP experiment, the number of mixtures in the training (13 mixtures) set outnumbered the ones in the test set (6 mixtures), therefore, the set of mixtures in the training set is a subset of the mixtures in the test set but in different quantities. This setup allows us to test the model's predicting ability only on the composition percentage. We can see from Table 19 the model's overall performance for each material.

5.6.3 *Unseen Percentage and Mixture (UPM)*

The goal of the UPM experiment is to evaluate the model performance on predicting the unseen composition percentage of the unseen mixtures, thus the model's ability to extract from the mixture spectral data the single component's features and its ability to use these to generalize on the unseen percentage with unseen mixtures.

The experiment UP had already two sets for the training and testing, MP1 and MP2 respectively. In the UP experiment, the set MP1 contained a greater number of different mixtures than those found in MP2. While in the UPM experiment, we used the MP2 set as the training set and the MP1 set as the testing set. Consequently, this procedure had the number of mixtures for the test set outnumbering the ones found in the training set, unlike for the UP experiment. This allowed us to test the different portions of the test set, as we created subsets of the testing set. For the first testing subset ($MP1_{S1}$), we used mixtures that are found also in the training set, while for the second testing subset ($MP1_{S2}$), we used the mixtures that are not found in the training set. Reminding that the composition percentage of the mixtures in the training and test sets are totally different.

Finally, we obtained three results, the whole test set for this scenarios (MP1), the first subset ($MP1_{S1}$), and the second subset ($MP1_{S2}$). The subset $MP1_{S1}$ contained all mixtures that are also included in the training set except the base materials at 100%. The $MP1_{S2}$ subset contained all mixtures that not found in the training set, which is the experiment for the UPM scores. The total test set MP1 contained $MP1_{S1}$ and $MP1_{S2}$. Our focus here is to see the model prediction on the subset $MP1_{S2}$ for the composition percentage of unseen mixtures, and we can see the model's performance for each material in Table 22. In Table 20 and 21 we can see the overall performance of the model for each material respectively using the whole testing set MP1 and the subset $MP1_{S1}$.

5.7 DISCUSSION

The results achieved from the WP experiment are encouraging. The 1D-CNN predicted all the composition percentages of the mixtures with a very low error as seen from Table 18, with an average of

Table 20: The overall performance for the UPM scenario. The results are for the whole test set (MP1) of the UPM scenario

	MAE	MSE	RMSE	R^2
BabyMilk	0.123	0.0426	0.2063	0.5135
IceSugar	0.0566	0.0136	0.1165	0.8361
NaHCO3	0.0434	0.0146	0.1209	0.791
Cocoa	0.0703	0.0201	0.1419	0.7933
Potato	0.0425	0.0168	0.1297	0.7765
Rice	0.0593	0.0175	0.1322	0.813

Table 21: The overall performance for the UPM scenario. The table shows the results for the $MP1_{S1}$ subset.

	MAE	MSE	RMSE	R^2
BabyMilk	0.0489	0.0048	0.0696	0.9535
IceSugar	0.0115	0.0007	0.0257	0.991
NaHCO3	0.014	0.0011	0.0336	0.9883
Cocoa	0.0135	0.0017	0.0411	0.97
Potato	0.0078	0.0006	0.0236	0.9901
Rice	0.0144	0.0016	0.0396	0.9723

$R^2 = 0.99$. This result is promising since the model is able to extract the features of the specific composition percentages of mixtures. The outcome of the WP experiment encouraged us for the much harder tasks that are the UP an the UPM experiments.

The UP experiment is created to see how well we can predict the unseen composition percentages by training the model with the same mixtures as in the testing set but different composition percentage of the same mixtures. Training the model with MP1 and testing it with MP2 gave good results in terms of the determination coefficient, $R^2 = 0.9471$, with a 5% decrease in respect to the WP average R_2 score.

The UPM experiment have fewer variation of mixtures and quantities in the training set than in the test set and scored an average of $R^2 = 0.7539$ using all the test set MP1, with 25% decrease in respect to WP average R^2 score. Using the $MP1_{S1}$ subset led to better result since the model had to predict only the unseen composition percentage but same mixture, just as the UP experiment, the average determination coefficient for this testing subset is $R^2 = 0.9775$ with less that 2% decrease in respect to the WP experiment. The results for the $MP1_{S1}$ set can be compared also to the UP experiment: the $MP1_{S1}$ results showed a 3% improvement in respect to UP experiment. This is because the UPM training set contained uniformly distributed composition percentage of each material that are in $\{25, 50, 75\}$ and the six base materials at 100%, while the UPM missed the mixtures that con-

Table 22: The overall performance for the UPM scenario. The table shows the main results for the UPM experiment using $MP1_{S2}$ testing subset.

	MAE	MSE	RMSE	R^2
IceSugar	0.0972	0.0252	0.1587	0.718
NaHCO3	0.0699	0.0267	0.1635	0.3653
Cocoa	0.1213	0.0367	0.1916	0.6769
Potato	0.0737	0.0314	0.1773	0.6486
Rice	0.0996	0.0318	0.1783	0.7266

tains 65% of one material and 35% of the other leading to unevenly distributed composition percentage for one mixture.

While using the $MP1_{S2}$ subset as the test set, we have an average determination coefficient of $R^2 = 0.627$, with a significant 36% loss in respect the average R_2 of the WP experiment. This experiment is set up to to see if the model learned the representation of the single materials in different compositions so that it can predict unseen composition percentage of unseen mixtures. We must take into account also the fact that the materials quantities are prepared by weight rather than volume, this mean that we can have powders like BabyMilk that have a greater volume for a small amount. This characteristic can affect the spectral acquisition since the material with higher volumes tend to occupy most of the Petri dish causing little signal for the other materials mixed with them. In this specific test case scenario of UPM the model is trained with mixtures containing only BabyMilk mixed with the other 5 materials in different compositions and one other mixture of IceSugar and NaHCO3, in particular the subset $MP1_{S2}$ of the test set for this experiment doesn't contain any combination of BabyMilk, but it is trained with mixtures that contain BabyMilk. The worst results in terms of the R^2 score NaHCO3 gave the worst results because we have only two combination of this materials in the test set, with Cocoa and Rice, while in the training set there is no such mixture, adding also the fact that the NaHCO3 have higher density in terms of g/cm^3 in respect to Cocoa and Rice. Therefore, the spectra of of NaHCO3 mixed with Cocoa and then with Rice can be very difficult to interpolate without being trained.

In WP experiment the model was able to overcome the errors during mixtures and the weight-volume ratio and gave good result this is thanks to the huge amount of sample it is trained on. In the UP and especially the UPM experiments the errors due to the preparation of the materials and the the models ability to overcome them became very clear.

The objective of selecting the data preprocessing method is to alter the spectra such that our model can improve on predicting the content percentage of the powder mixtures, and this is done by establishing a good correlation between spectra and concentration data.

The mean-centering method, which is mainly used to eliminate the linear baseline shifts and results as the best scaling method for our spectra data and model in every experiment block as shown in Tables 15, 16

Also, the SNV scaling method was very robust in various experimental blocks. It performed -20% less than the mean-centering methods in Table 15 and 24% less when using the Savitsky Golay filter....

The MSC method is often used for measurements in diffuse reflection, performed -42% worse than the mean-centering method from Table 15 and -30% worse when using the Savitsky Golay filter.

In the case of the min-max, the normalization method that is mainly used to eliminate the influence of different optical path lengths, showed -40% performance on the test set compared with the mean-centering method from Table 15 and -40% when using the Savitsky Golay filter from Table 16.

As for the first derivative method, signals with steep edges emphasize more than relatively flat bands by calculating the first derivative. In the second derivative, even the highly flat structures can be evaluated. The results achieved by comparing the normalization methods show that mean centering. The 1D-CNN predicted all the composition percentages of the mixtures with a deficient error, as seen from Table 15, with an MAE on the test set of $MAE = 0.0055$. This result is promising since the model can extract the features of the specific composition percentages of mixtures. The outcome of the normalization methods was also compared against the Savitsky Golay filter. Applying this filter improved most of the results shown in Table 15 that were obtained without using the filter. In table 23 we can see the improvement in percentage for each normalization method.

	MAE - Train	MAE - Validation	MAE - Test
Mean Centering	13%	28%	20%
SNV	-3%	-8%	15%
MSC	14%	26%	48%
None	-3%	-10%	-4%
MinMax	1%	1%	7%
Detrend	14%	10%	17%
MSC + DET	-4%	-5%	-3%
SNV + DET	13%	21%	8%

Table 23: Improvement of Savitsky Golay on spectral data for every normalization method used

In regard with the data transformation process we can see that concatenating the derivative information to the various normalization methods we can see that concatenating the first derivative information to raw spectra generally gives the best performance on our test set. whilst using the SNV normalization method with the concatenation of the derivatives gives the best performance. Thus, when con-

sidering using the derivative information is best to use the following modalities: $Raw \| D1$, $Raw \| D2$, $Raw \| D1 \| D2$.

We must take into account also the fact that the materials quantities are prepared by weight rather than volume, this mean that we can have powders like BabyMilk that have a greater volume for a small amount. This characteristic can affect the spectral acquisition since the material with higher volumes tend to occupy most of the Petri dish causing little signal for the other materials mixed with them. Nevertheless the prediction for baby milk with MC scaling and with Savitsky Golay filter, are better than those presented in [98]. These improvements are shown in table 24

Material	MAE	MSE	RMSE	R2
BabyMilk	0.0073 (54.47%)	0.0003 (67.09%)	0.0172 (42.25%)	0.996 (0.5%)
IceSugar	0.0045 (31.53%)	0.0002 (17.66%)	0.0128 (10.26%)	0.998 (0.04%)
NaHCO3	0.0027 (40.66%)	0 (58.56%)	0.0064 (43.53%)	0.9995 (0.12%)
Cocoa	0.0031 (64.61%)	0.0001 (85.53%)	0.0085 (62.7%)	0.9992 (0.51%)
Potato	0.0049 (17.6%)	0.0002 (26.46%)	0.0149 (11.06%)	0.9973 (0.12%)
Rice	0.0053 (27.76%)	0.0002 (40.85%)	0.0154 (18.62%)	0.9973 (0.15%)

Table 24: Result for MC scaling method with the Savitsky Golay filter and their improvement in percentage compared to the result presented in [98]. The percentage in parenthesis shows the amount of improvement.

In table 25,is shown the performance of prediction of each percentage that is a given material is mixed with.

	BabyMilk	IceSugar	NaHCO3	Cocoa	Potato	Rice
0%	0.09% (0.0089)	0.02% (0.0021)	0% (0.0005)	0% (0.0003)	0.02% (0.0032)	0.03% (0.0041)
15%	15.26% (0.0237)	14.22% (0.0274)	14.89% (0.0091)	15.26% (0.0103)	14.15% (0.0308)	14.23% (0.0208)
25%	23.93% (0.0304)	25.88% (0.0154)	25.03% (0.01)	25.37% (0.0099)	25.33% (0.0183)	24.03% (0.0244)
33%	32.9% (0.0217)	32.77% (0.0167)	33.75% (0.0146)	34.38% (0.0176)	33.33% (0.0266)	32.72% (0.0207)
35%	34.37% (0.0232)	35.26% (0.0302)	35.37% (0.0127)	35.83% (0.0141)	33.61% (0.0332)	
40%	-	-	-	41.5% (0.0122)	-	-
45%	44.32% (0.0157)		-	-	-	-
50%	48.65% (0.0284)	50.83% (0.0206)	49.76% (0.0119)	49.8% (0.0319)	51.36% (0.0225)	51.16% (0.025)
65%	64.25% (0.0114)	64.8% (0.0163)	-	65.85% (0.0091)	64.77% (0.0227)	65.09% (0.031)
75%	74.4% (0.0137)	73.33% (0.0236)	75.71% (0.0145)	75.72% (0.0062)	77.75% (0.0236)	76.34% (0.0415)
85%	84.61% (0.0098)	85.2% (0.0239)	85.17% (0.0099)	85.44% (0.0071)	85.27% (0.0251)	85.18% (0.0301)
100%	100% (0.0001)	99.92% (0.007)	99.98% (0.0028)	100% (0.0001)	99.99% (0.001)	99.65% (0.0125)

Table 25: Performance of prediction of various percentages for each material using the Mean-centering scaling method combined with Savitsky Golay filter

5.8 CONCLUSION

In this work, we analysed the problem of predicting composition percentage of organic material mixtures. The NIR spectra of organic materials holds intrinsic information on the analyte, including its quantity. To uncover these intrinsic characteristics of the 1D-CNN showed

great performance, in the WP experiment, by extracting directly relevant wavelength (feature) from the NIR spectrum that described the quantity of the analyte. This led to better performance of the model avoiding the accumulation of errors caused by manual wavelength selection.

The NIR spectra of the mixtures are most probably affected by the density, in terms of g/cm^3, of each material in the mixture. Thus, affecting the result of each experiment especially in the UPM experiment.

The research leads to future developments of this work. The results of the UPM using the $MP1_{S2}$ testing subset can be improved by taking into account the weight-volume ratio and by modeling new 1D-CNN architecture for the unseen composition of unseen mixtures. It also interesting to extend the WP experiment's model by testing it on different composition percentage of more than two mixtures to see if the model is acquire a good generalizing ability.

The NIR spectra of organic materials hold intrinsic properties of the composition, including their quantity. A specific preprocessing method can add another characteristic to the spectra, making them easy to analyze. In this work, we evaluated different scaling and transformation methods in the context of spectra preprocessing when predicting the composition percentage of organic material mixtures.

The filtering method introduced by Savitzky Golay has long been used in the absorption spectroscopy community for its ability to simultaneously smooth and differentiate absorption spectra. In this investigation, the Savitzky Golay method applied to our near-infrared range (1350nm - 2150nm) significantly improved the material content percentage prediction in a mixture. Combining the scaled NIR spectra of a material or mixtures with their relative derivatives can uncover further information and result in a classification or regression task, making the model robust.

Our dataset also contained the pure powders spectra and their mixtures of up to three different powders. We saw a growing trend in the complexity of preparing such mixtures and their analysis when combining more than three powders since we cannot guarantee the homogeneity of the mixture.

Additional work is needed to tackle the high number of attributes when concatenating the derivative data to the original spectra. So, models that reduce the dimensionality of the single sample without losing too much information are good starting points to improve the overall performance of near-infrared spectra analysis.

BIBLIOGRAPHY

[1] Jinwon An and Sungzoon Cho. «Variational autoencoder based anomaly detection using reconstruction probability.» In: *Special Lecture on IE* 2.1 (2015), pp. 1–18.

[2] Jarkko Antila, Mikko Tuohiniemi, Anna Rissanen, Uula Kantojärvi, Markku Lahti, Kai Viherkanto, Marko Kaarre, Jouko Malinen, and Antti Näsilä. «MEMS- and MOEMS-Based Visible and Near-Infrared Spectrometers.» In: *Encyclopedia of Analytical Chemistry*. John Wiley Sons, Ltd, 2021, pp. 1–41. ISBN: 9780470027318. DOI: https://doi.org/10.1002/9780470027318.a9376.pub2. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470027318.a9376.pub2. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470027318.a9376.pub2.

[3] Mersedeh Beitollahi and S. Abolfazl Hosseini. «Using Savitsky-Golay filter and interval curve fitting in order to hyperspectral data compression.» In: *2017 Iranian Conference on Electrical Engineering (ICEE)*. 2017, pp. 1967–1972. DOI: 10.1109/IranianCEE.2017.7985378.

[4] Mersedeh Beitollahi and S. Abolfazl Hosseini. «Using Savitsky-Golay Smoothing Filter in Hyperspectral Data Compression by Curve Fitting.» In: *Electrical Engineering (ICEE), Iranian Conference on*. 2018, pp. 452–457. DOI: 10.1109/ICEE.2018.8472702.

[5] D. Berckmans. «General introduction to precision livestock farming.» In: *Animal Frontiers* 7.1 (Jan. 2017), pp. 6–11. ISSN: 2160-6056. DOI: 10.2527/af.2017.0102. eprint: https://academic.oup.com/af/article-pdf/7/1/6/32411623/6.pdf. URL: https://doi.org/10.2527/af.2017.0102.

[6] F. Berggren, A. Bria, L. Badia, I. Karla, R. Litjens, P. Magnusson, F. Meago, H. Tang, and R. Veronesi. «Multi-radio resource management for ambient networks.» In: *Proc. IEEE Pimrc*. Vol. 2. 2005, pp. 942–946.

[7] Paolo Berzaghi and Roberto Riovanto. «Near infrared spectroscopy in animal science production: principles and applications.» In: *Italian Journal of Animal Science* 8.sup3 (2009), pp. 39–62. DOI: 10.4081/ijas.2009.s3.39.

[8] Christopher M Bishop. «Variational principal components.» In: *Proc. ICANN*. IET, 1999.

[9] Robert Bogue. «Recent developments in MEMS sensors: a review of applications, markets and technologies.» In: *Sensor Review* 33.4 (2013), pp. 300–304. ISSN: 0260-2288. DOI: 10.1108/SR-05-2013-678. URL: https://doi.org/10.1108/SR-05-2013-678.

[10] Karla Rodrigues Borba, Poliana Cristina Spricigo, Didem Peren Aykas, Milene Corso Mitsuyuki, Luiz Alberto Colnago, and Marcos David Ferreira. «Non-invasive quantification of vitamin C, citric acid, and sugar in 'Valência' oranges using infrared spectroscopies.» In: *Journal of Food Science and Technology* 58.2 (2021), pp. 731–738. ISSN: 0975-8402. DOI: 10.1007/s13197-020-04589-x. URL: https://doi.org/10.1007/s13197-020-04589-x.

[11] Eva Borràs, Joan Ferré, Ricard Boqué, Montserrat Mestres, Laura Aceña, and Olga Busto. «Data fusion methodologies for food and beverage authentication and quality assessment - a review.» en. In: *Anal Chim Acta* 891 (Apr. 2015), pp. 1–14.

[12] Leo Breiman. «Random Forests.» In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

[13] Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. «Understanding disentangling in β-VAE.» In: *arXiv preprint arXiv:1804.03599* (2018).

[14] Ana M. Cavaco, Dário Passos, Rosa M. Pires, Maria D. Antunes, and Rui Guerra. «Nondestructive Assessment of Citrus Fruit Quality and Ripening by Visible–Near Infrared Reflectance Spectroscopy.» In: *Citrus*. Ed. by Muhammad Sarwar Khan and Iqrar Ahmad Khan. Rijeka: IntechOpen, 2021. Chap. 13. DOI: 10.5772/intechopen.95970. URL: https://doi.org/10.5772/intechopen.95970.

[15] Ana M. Cavaco, Pedro Pinto, M. Dulce Antunes, Jorge Marques da Silva, and Rui Guerra. «'Rocha' pear firmness predicted by a Vis/NIR segmented model.» In: *Postharvest Biology and Technology* 51.3 (2009), pp. 311–319. ISSN: 0925-5214. DOI: https://doi.org/10.1016/j.postharvbio.2008.08.013. URL: https://www.sciencedirect.com/science/article/pii/S0925521408002433.

[16] Haiyan Cen and Yong He. «Theory and application of near infrared reflectance spectroscopy in determination of food quality.» In: *Trends in Food Science and Technology* 18.2 (2007), pp. 72–83. ISSN: 0924-2244. DOI: https://doi.org/10.1016/j.tifs.2006.09.003. URL: http://www.sciencedirect.com/science/article/pii/S0924224406002524.

[17] Chen Chen, Hongyi Li, Xiaoyi Lv, Jun Tang, Cheng Chen, and Xiangxiang Zheng. «Application of near infrared spectroscopy combined with SVR algorithm in rapid detection of cAMP content in red jujube.» In: *Optik* 194 (2019), p. 163063. ISSN: 0030-4026. DOI: https://doi.org/10.1016/j.ijleo.2019.163063. URL: https://www.sciencedirect.com/science/article/pii/S0030402619309404.

[18] Quansheng Chen, Dongliang Zhang, Wenxiu Pan, Qin Ouyang, Huanhuan Li, Khulal Urmila, and Jiewen Zhao. «Recent developments of green analytical techniques in analysis of tea's quality and nutrition.» In: *Trends in Food Science & Technology* 43.1 (2015), pp. 63–82. ISSN: 0924-2244. DOI: https://doi.org/10.1016/j.tifs.2015.01.009. URL: http://www.sciencedirect.com/science/article/pii/S0924224415000291.

[19] Xiaoyi Chen, Qinqin Chai, Ni Lin, Xianghui Li, and Wu Wang. «1D convolutional neural network for the discrimination of aristolochic acids and their analogues based on near-infrared spectroscopy.» In: *Anal. Methods* 11 (40 2019), pp. 5118–5125. DOI: 10.1039/C9AY01531K. URL: http://dx.doi.org/10.1039/C9AY01531K.

[20] J.J. Olmos Colmenero and G.A. Broderick. «Effect of Dietary Crude Protein Concentration on Milk Production and Nitrogen Utilization in Lactating Dairy Cows1.» In: *Journal of Dairy Science* 89.5 (2006), pp. 1704–1712. ISSN: 0022-0302. DOI: https://doi.org/10.3168/jds.S0022-0302(06)72238-X. URL: https://www.sciencedirect.com/science/article/pii/S002203020672238X.

[21] Richard A. Crocombe. *Miniature Optical Spectrometers, Part III: Conventional and Laboratory Near-Infrared Spectrometers — spectroscopyonline.com*. https://www.spectroscopyonline.com/view/miniature-optical-spectrometers-part-iii-conventional-and-laboratory-near-infrared-spectrometers. [Accessed 14-Nov-2022].

[22] Richard A Crocombe. «Portable spectroscopy.» In: *Appl. Spectrosc.* 72.12 (2018), pp. 1701–1751.

[23] Italo Francesco Zoppis Dagmawi D. Tegegn, Sara Manzoni, Alessio Mognato, Ivan Reguzzoni, and Edoardo Lotti. «Rapid Analysis of Powders Based on Deep Learning, Near-Infrared and Derivative Spectroscopy.» In: *Artificial Intelligence and Applications for Business and Industries 2021*. 2021. URL: http://ceur-ws.org/Vol-3102/paper12.pdf.

[24] Tegegn Dagmawi Delelegn, Italo Francesco Zoppis, Sara Manzoni, Alessio Mognato, Ivan Reguzzoni, and Edoardo Lotti. «Rapid Analysis of Powders Based on Deep Learning, Near-

Infrared and Derivative Spectroscopy.» In: *CEUR Workshop Proceedings*. Vol. 3102. 2021.

[25] S.R. Davis and R.J. Collier. «Mammary Blood Flow and Regulation of Substrate Supply for Milk Synthesis.» In: *Journal of Dairy Science* 68.4 (1985), pp. 1041–1058. ISSN: 0022-0302. DOI: https://doi.org/10.3168/jds.S0022-0302(85)80926-7. URL: https://www.sciencedirect.com/science/article/pii/S0022030285809267.

[26] Diego A. Delgadillo-Duran, Cesar A. Vargas-García, Viviana M. Varón-Ramírez, Francisco Calderón, Andrea C. Montenegro, and Paula H. Reyes-Herrera. «Vis–NIR spectroscopy and machine learning methods to diagnose chemical properties in Colombian sugarcane soils.» In: *Geoderma Regional* 31 (2022), e00588. ISSN: 2352-0094. DOI: https://doi.org/10.1016/j.geodrs.2022.e00588. URL: https://www.sciencedirect.com/science/article/pii/S2352009422001080.

[27] Carl Emil Eskildsen, Karen Wahlstrøm Sanden, Sileshi Gizachew Wubshet, Petter Vejle Andersen, Jorun Øyaas, and Jens Petter Wold. «Estimating dry matter and fat content in blocks of Swiss cheese during production using on-line near infrared spectroscopy.» In: *Journal of Near Infrared Spectroscopy* 27.4 (2019), pp. 293–301. DOI: 10.1177/0967033519855436. eprint: https://doi.org/10.1177/0967033519855436. URL: https://doi.org/10.1177/0967033519855436.

[28] Charles Farber and Dmitry Kurouski. «Detection and identification of plant pathogens on maize kernels with a hand-held Raman spectrometer.» In: *Anal. Chem.* 90.5 (2018), pp. 3009–3012.

[29] Angelica Galieni, Nicola D'Ascenzo, Fabio Stagnari, Giancarlo Pagnani, Qingguo Xie, and Michele Pisante. «Past and future of plant stress detection: an overview from remote sensing to positron emission tomography.» In: *Front. Plant Sc.* 11 (2021), p. 1975.

[30] Manuel Galli, Fabio Pagni, Gabriele De Sio, Andrew Smith, Clizia Chinello, Martina Stella, Vincenzo L'Imperio, Marco Manzoni, Mattia Garancini, Diego Massimini, et al. «Proteomic profiles of thyroid tumors by mass spectrometry-imaging on tissue microarrays.» In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1865.7 (2017), pp. 817–827.

[31] Takuma Genkawa, Hideyuki Shinzawa, Hideaki Kato, Daitaro Ishikawa, Kodai Murayama, Makoto Komiyama, and Yukihiro Ozaki. «Baseline Correction of Diffuse Reflection Near-Infrared Spectra Using Searching Region Standard Normal Variate (SRSNV).» In: *Applied Spectroscopy* 69.12 (2015), pp. 1432–1441. ISSN: 19433530. DOI: 10.1366/15-07905/ASSET/IMAGES/LARGE/10.1366_15-

07905 - FIG2 . JPEG. URL: https : / / journals - sagepub - com . unimib.idm.oclc.org/doi/10.1366/15-07905.

[32]   Diana Giannuzzi et al. «In-line near-infrared analysis of milk coupled with machine learning methods for the daily prediction of blood metabolic profile in dairy cattle.» In: *Scientific Reports* 12.1 (2022), p. 8058. ISSN: 2045-2322. DOI: 10 . 1038 / s41598 - 022 - 11799 - 0. URL: https : / / doi . org / 10 . 1038 / s41598-022-11799-0.

[33]   Mohammad Goodarzi, Sandeep Sharma, Herman Ramon, and Wouter Saeys. «Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring.» In: *TrAC Trends in Analytical Chemistry* 67 (2015), pp. 147–158. ISSN: 0165-9936. DOI: https://doi.org/10.1016/j.trac.2014.12.005. URL: https : / / www . sciencedirect . com / science / article / pii / S0165993615000266.

[34]   Marco Grossi, Giuseppe Di Lecce, Marco Arru, Tullia Gallina Toschi, and Bruno Riccò. «An opto-electronic system for in-situ determination of peroxide value and total phenol content in olive oil.» In: *J. of Food Engineering* 146 (2015), pp. 1–7. ISSN: 0260-8774. DOI: https : / / doi . org / 10 . 1016 / j . jfoodeng . 2014.08.015. URL: http://www.sciencedirect.com/science/ article/pii/S0260877414003537.

[35]   Matthew W. Harer and Valerie Y. Chock. «Renal Tissue Oxygenation Monitoring—An Opportunity to Improve Kidney Outcomes in the Vulnerable Neonatal Population.» In: *Frontiers in Pediatrics* 8 (2020). ISSN: 2296-2360. DOI: 10.3389/fped.2020. 00241. URL: https : / / www . frontiersin . org / articles / 10 . 3389/fped.2020.00241.

[36]   A. Haupt, B. Berg, P. Paschen, M. Dreyer, H.-U. Häring, J. Smedegaard, and S. Matthaei. «The Effects of Skin Temperature and Testing Site on Blood Glucose Measurements Taken by a Modern Blood Glucose Monitoring Device.» In: *Diabetes Technology & Therapeutics* 7.4 (2005). PMID: 16120031, pp. 597–601. DOI: 10.1089/dia.2005.7.597. eprint: https://doi.org/ 10.1089/dia.2005.7.597. URL: https://doi.org/10.1089/ dia.2005.7.597.

[37]   Raphael Henn, Christian G. Kirchler, Maria-Elisabeth Grossgut, and Christian W. Huck. «Comparison of sensitivity to artificial spectral errors and multivariate LOD in NIR spectroscopy – Determining the performance of miniaturizations on melamine in milk powder.» In: *Talanta* 166 (2017), pp. 109–118. ISSN: 0039-9140. DOI: https : / / doi . org / 10 . 1016 / j . talanta . 2017 . 01 . 035. URL: https : / / www . sciencedirect . com / science / article/pii/S003991401730139X.

[38] Aminah Hina and Wala Saadeh. «Noninvasive Blood Glucose Monitoring Systems Using Near-Infrared Technology-A Review.» eng. In: *Sensors (Basel, Switzerland)* 22.13 (2022). 35808352[pmid], p. 4855. ISSN: 1424-8220. DOI: 10.3390/s22134855. URL: https://pubmed.ncbi.nlm.nih.gov/35808352.

[39] Gao Huang, Guang-Bin Huang, Shiji Song, and Keyou You. «Trends in extreme learning machines: A review.» In: *Neural Networks* 61 (2015), pp. 32–48. ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2014.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0893608014002214.

[40] Haibo Huang, Haiyan Yu, Huirong Xu, and Yibin Ying. «Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review.» In: *J. of Food Engineering* 87.3 (2008), pp. 303–313. ISSN: 0260-8774. DOI: https://doi.org/10.1016/j.jfoodeng.2007.12.022. URL: http://www.sciencedirect.com/science/article/pii/S0260877408000071.

[41] Christian Huber, Christoph Krammer, Benedikt Stein, and Heinz Kalt. «Micromechanical tunable Fabry-Pérot interferometers with membrane Bragg mirrors based on silicon/silicon carbonitride.» In: (2019).

[42] Zoppis Italo, Gianazza Erica, Borsani Massimiliano, Chinello Clizia, Mainini Veronica, Galbusera Carmen, Ferrarese Carlo Galimberti, Sandro Gloria Sorbi, Barbara Borroni, et al. «Mutual information optimization for mass spectra data alignment.» In: *IEEE/ACM transactions on computational biology and bioinformatics* 9.3 (2011), pp. 934–939.

[43] D.S. Jayas, J. Paliwal, C. Erkinbaev, P.K. Ghosh, and C. Karunakaran. «Chapter 16 - Wheat Quality Evaluation.» In: *Computer Vision Technology for Food Quality Evaluation (Second Edition)*. Ed. by Da-Wen Sun. Second Edition. San Diego: Academic Press, 2016, pp. 385–412. ISBN: 978-0-12-802232-0. DOI: https://doi.org/10.1016/B978-0-12-802232-0.00016-5. URL: https://www.sciencedirect.com/science/article/pii/B9780128022320000165.

[44] Sofian Kanan, Oussama El-Kadri, Imad Abu-Yousef, and Marsha Kanan. «Semiconducting Metal Oxide Based Sensors for Selective Gas Pollutant Detection.» In: *Sensors (Basel, Switzerland)* 9 (Oct. 2009), pp. 8158–96. DOI: 10.3390/s91008158.

[45] Evelyne C. Kessler, Rupert M. Bruckmaier, and Josef J. Gross. «Milk urea nitrogen concentration is higher in Brown Swiss than in Holstein dairy cows despite identical feeding.» In: *Journal of Animal Physiology and Animal Nutrition* 104.6 (2020), pp. 1671–1677. DOI: https://doi.org/10.1111/jpn.13408. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jpn.13408. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jpn.13408.

[46]   H. Khadem, H. Nemat, J. Elliott, and M. Benaissa. «Signal fragmentation based feature vector generation in a model agnostic framework with application to glucose quantification using absorption spectroscopy.» English. In: *Talanta* 243 (2022). cited By 0. ISSN: 00399140. DOI: 10.1016/j.talanta.2022.123379. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126560772&doi=10.1016\%2fj.talanta.2022.123379&partnerID=40&md5=110c1775cd159b7b5744508629b6a0cf.

[47]   Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. «1D convolutional neural networks and applications: A survey.» In: *Mechanical Systems and Signal Processing* 151 (2021), p. 107398. ISSN: 0888-3270. DOI: https://doi.org/10.1016/j.ymssp.2020.107398. URL: https://www.sciencedirect.com/science/article/pii/S0888327020307846.

[48]   Donald E. Knuth. «Computer Programming as an Art.» In: *Communications of the ACM* 17.12 (1974), pp. 667–673.

[49]   A. Kohler, M. Zimonja, V. Segtnan, and H. Martens. «Standard Normal Variate, Multiplicative Signal Correction and Extended Multiplicative Signal Correction Preprocessing in Biospectroscopy.» In: *Comprehensive Chemometrics* 2 (2009), pp. 139–162. DOI: 10.1016/B978-044452701-1.00102-2.

[50]   Raymond Kokaly, Don Despain, Roger Clark, and K.Eric Livo. «Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data.» In: *Remote Sensing of Environment* 84 (Mar. 2003), pp. 437–456. DOI: 10.1016/S0034-4257(02)00133-5.

[51]   Jacek Kulakowski and Benoît d'Humières. «Chip-size spectrometers drive spectroscopy towards consumer and medical applications.» In: *Photonic Instrumentation Engineering VIII*. Ed. by Yakov Soskind and Lynda E. Busse. Vol. 11693. International Society for Optics and Photonics. SPIE, 2021, 116931A. DOI: 10.1117/12.2591048. URL: https://doi.org/10.1117/12.2591048.

[52]   Tuo Leng, Feng Li, Yi Chen, Lijun Tang, Jianhua Xie, and Qiang Yu. «Fast quantification of total volatile basic nitrogen (TVB-N) content in beef and pork by near-infrared spectroscopy: Comparison of SVR and PLS model.» In: *Meat Science* 180 (2021), p. 108559. ISSN: 0309-1740. DOI: https://doi.org/10.1016/j.meatsci.2021.108559. URL: https://www.sciencedirect.com/science/article/pii/S0309174021001352.

[53]   Ang Li, Chunhui Yao, Junfei Xia, Huijie Wang, Qixiang Cheng, Richard Penty, Yeshaiahu Fainman, and Shilong Pan. «Advances in cost-effective integrated spectrometers.» In: *Light: Science & Applications* 11.1 (2022), p. 174. ISSN: 2047-7538. DOI:

10.1038/s41377-022-00853-1. URL: https://doi.org/10.1038/s41377-022-00853-1.

[54] Zhengxuan Li, Xiuying Tang, Zhixiong Shen, Kefei Yang, Lingjuan Zhao, and Yanlei Li. «Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for Aspergillus flavus contamination detection in peanut.» In: *J. of the Science of Food and Agriculture* 99.13 (2019), pp. 5671–5679. DOI: 10.1002/jsfa.9828. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jsfa.9828. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jsfa.9828.

[55] Haibo Liang, Gang Liu, Li Zhongbing, and Wang Ren. «Improved SVR Based on CARS and BAS For Hydrocarbon Concentration Detection.» In: *Vibrational Spectroscopy* (2023), p. 103494. ISSN: 0924-2031. DOI: https://doi.org/10.1016/j.vibspec.2023.103494. URL: https://www.sciencedirect.com/science/article/pii/S0924203123000012.

[56] Amanda Beatriz Sales de Lima, Acsa Santos Batista, Josane Cardim de Jesus, Jaqueline de Jesus Silva, Antônia Cardoso Mendes de Araújo, and Leandro Soares Santos. «Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling.» In: *Food Control* 107 (2020), p. 106802. ISSN: 0956-7135. DOI: https://doi.org/10.1016/j.foodcont.2019.106802. URL: http://www.sciencedirect.com/science/article/pii/S0956713519303913.

[57] Glaucia Ferreira de Lima, Samara Alvachian Cardoso Andrade, Vitor Hugo da Silva, and Fernanda Araújo Honorato. «Multivariate Classification of UHT Milk as to the Presence of Lactose Using Benchtop and Portable NIR Spectrometers.» In: *Food Analytical Methods* 11.10 (2018), pp. 2699–2706. ISSN: 1936-976X. DOI: 10.1007/s12161-018-1253-7. URL: https://doi.org/10.1007/s12161-018-1253-7.

[58] T. Lindena and S. Hess. «Is animal welfare better on smaller dairy farms? Evidence from 3,085 dairy farms in Germany.» In: *Journal of Dairy Science* 105.11 (2022), pp. 8924–8945. ISSN: 0022-0302. DOI: https://doi.org/10.3168/jds.2022-21906. URL: https://www.sciencedirect.com/science/article/pii/S0022030222005458.

[59] Ningjing Liu, Hector Aya Parra, Annemieke Pustjens, Kasper Hettinga, Philippe Mongondry, and Saskia M. van Ruth. «Evaluation of portable near-infrared spectroscopy for organic milk authentication.» In: *Talanta* 184 (2018), pp. 128–135. ISSN: 0039-9140. DOI: https://doi.org/10.1016/j.talanta.2018.02.097. URL: https://www.sciencedirect.com/science/article/pii/S0039914018302157.

[60] Yachao Liu, Yongyu Li, Yankun Peng, Yanming Yang, and Qi Wang. «Detection of fraud in high-quality rice by near-infrared spectroscopy.» In: *Journal of Food Science* (2020). DOI: 10.1111/1750-3841.15314. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1750-3841.15314.

[61] P. Llano Suárez, A. Soldado, A. González-Arrojo, F. Vicente, and B. de la Roza-Delgado. «Rapid on-site monitoring of fatty acid profile in raw milk using a handheld near infrared sensor.» In: *Journal of Food Composition and Analysis* 70 (2018), pp. 1–8. ISSN: 0889-1575. DOI: https://doi.org/10.1016/j.jfca.2018.03.003. URL: https://www.sciencedirect.com/science/article/pii/S0889157518300693.

[62] Ilya Loshchilov and Frank Hutter. «Decoupled Weight Decay Regularization.» In: *Proc. ICLR*. 2017.

[63] R Madhumathi, T Arumuganathan, and R Shruthi. «A Survey on Wireless Sensor Networks and Instrumentation Techniques for Smart Agriculture.» In: *Mob. Comput. Sustain. Inf.* Springer, 2022, pp. 453–467.

[64] Lembe S. Magwaza, Umezuruike Linus Opara, Hélène Nieuwoudt, Paul J. R. Cronje, Wouter Saeys, and Bart Nicolaï. «NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit—A Review.» In: *Food and Bioprocess Technology* 5.2 (2012), pp. 425–444. ISSN: 1935-5149. DOI: 10.1007/s11947-011-0697-1. URL: https://doi.org/10.1007/s11947-011-0697-1.

[65] S. Mahesh, A. Manickavasagan, D.S. Jayas, J. Paliwal, and N.D.G. White. «Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes.» In: *Biosystems Engineering* 101.1 (2008), pp. 50–57. ISSN: 1537-5110. DOI: https://doi.org/10.1016/j.biosystemseng.2008.05.017. URL: https://www.sciencedirect.com/science/article/pii/S1537511008001797.

[66] Laura Marinoni, Angelo Stroppa, Stefania Barzaghi, Katia Cremonesi, Nicolò Pricca, Aurora Meucci, Giulia Pedrolini, Andrea Galli, and Giovanni Cabassi. «On site monitoring of Grana Padano cheese production using portable spectrometers.» In: Feb. 2019, pp. 85–90. ISBN: 9781906715274. DOI: 10.1255/nir2017.085.

[67] Puneet Mishra, Ronald Klont, Theo Verkleij, and Sjaak Wisse. «Translating near-infrared spectroscopy from laboratory to commercial slaughterhouse: Existing challenges and solutions.» In: *Infrar. Phys. Techn.* 119 (2021), p. 103918.

[68] Wartini Ng, Budiman Minasny, Maryam Montazerolghaem, Jose Padarian, Richard Ferguson, Scarlett Bailey, and Alex B. McBratney. «Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra.» In: *Geoderma* 352 (2019), pp. 251–267. ISSN: 0016-7061. DOI: https://doi.org/10.1016/j.geoderma.2019.06.016. URL: http://www.sciencedirect.com/science/article/pii/S0016706119300588.

[69] Chao Ni, Dongyi Wang, and Yang Tao. «Variable weighted convolutional neural network for the nitrogen content quantization of Masson pine seedling leaves with near-infrared spectroscopy.» In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 209 (2019), pp. 32–39. ISSN: 1386-1425. DOI: https://doi.org/10.1016/j.saa.2018.10.028. URL: http://www.sciencedirect.com/science/article/pii/S1386142518309557.

[70] I. Noda. «Generalized Two-Dimensional Correlation Method Applicable to Infrared, Raman, and other Types of Spectroscopy.» In: *Applied Spectroscopy* 47.9 (1993), pp. 1329–1336. DOI: 10.1366/0003702934067694. eprint: https://doi.org/10.1366/0003702934067694. URL: https://doi.org/10.1366/0003702934067694.

[71] N. S. Oliver, C. Toumazou, A. E. G. Cass, and D. G. Johnston. «Glucose sensors: a review of current and emerging technology.» In: *Diabetic Medicine* 26.3 (2009), pp. 197–210. DOI: https://doi.org/10.1111/j.1464-5491.2008.02642.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1464-5491.2008.02642.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1464-5491.2008.02642.x.

[72] Brian G Osborne. «Near-infrared spectroscopy in food analysis.» In: *Enc. of analytical chemistry: applications, theory and instrumentation* (2006).

[73] S. Oya, H. Inoue, T. Nakade, A. Ogata, M. Tamura, and S. Kato. «Near-infrared Spectroscopy Evaluated as a Technique for Estimating Udder Haemodynamics in the Lactating Cow.» In: *Journal of Veterinary Medicine Series A* 50.5 (2003), pp. 230–234. DOI: https://doi.org/10.1046/j.1439-0442.2003.00524.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1439-0442.2003.00524.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1439-0442.2003.00524.x.

[74] Yukihiro Ozaki and Yusuke Morisawa. «Principles and Characteristics of NIR Spectroscopy.» In: *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*. Ed. by Yukihiro Ozaki, Christian Huck, Satoru Tsuchikawa, and Søren Balling Engelsen. Singapore: Springer Singapore, 2021,

pp. 11–35. ISBN: 978-981-15-8648-4. DOI: 10.1007/978-981-15-8648-4_2. URL: https://doi.org/10.1007/978-981-15-8648-4_2.

[75] Daehyung Park, Yuuna Hoshi, and Charles C Kemp. «A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder.» In: *IEEE Access* 3.3 (2018), pp. 1544–1551.

[76] W. Peng, G. Beggio, A. Pivato, H. Zhang, F. Lü, and P. He. «Applications of near infrared spectroscopy and hyperspectral imaging techniques in anaerobic digestion of bio-wastes: A review.» English. In: *Renewable and Sustainable Energy Reviews* 165 (2022). cited By 1. ISSN: 13640321. DOI: 10.1016/j.rser.2022.112608.

[77] Francis J. Pierce and Peter Nowak. «Aspects of Precision Agriculture.» In: ed. by Donald L. Sparks. Vol. 67. Advances in Agronomy. Academic Press, 1999, pp. 1–85. DOI: https://doi.org/10.1016/S0065-2113(08)60513-1. URL: https://www.sciencedirect.com/science/article/pii/S0065211308605131.

[78] Alexey L. Pomerantsev and Oxana Ye. Rodionova. «Process analytical technology: a critical view of the chemometricians.» In: *Journal of Chemometrics* 26.6 (2012), pp. 299–310. DOI: https://doi.org/10.1002/cem.2445. eprint: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/cem.2445. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/cem.2445.

[79] Electronic Products. *Samsung turns up the pressure on competition with pressure sensor in Galaxy S4*. 2013. URL: https://www.electronicproducts.com/samsung-turns-up-the-pressure-on-competition-with-pressure-sensor-in-galaxy-s4/.

[80] Yuanyuan Pu, Dolores Pérez-Marín, Norah O'Shea, and Ana Garrido-Varo. «Recent Advances in Portable and Handheld NIR Spectrometers and Applications in Milk, Cheese and Dairy Powders.» en. In: *Foods* 10.10 (Oct. 2021).

[81] S.A. Pullano, M. Greco, M.G. Bianco, D. Foti, A. Brunetti, and A.S. Fiorillo. «Glucose biosensors in clinical practice: Principles, limits and perspectives of currently used devices.» English. In: *Theranostics* 12.2 (2022). cited By 9, pp. 493–511. ISSN: 18387640. DOI: 10.7150/thno.64035. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85120487647&doi=10.7150\%2fthno.64035&partnerID=40&md5=ce3932f411892359f36f9f0ad0881b80.

[82] Lu Qingyun, Chen Yeming, Takashi Mikami, Motonobu Kawano, and Li Zaigui. «Adaptability of four-samples sensory tests and prediction of visual and near-infrared reflectance spectroscopy for Chinese indica rice.» In: *J. of Food Engineering*

79.4 (2007), pp. 1445–1451. ISSN: 0260-8774. DOI: https://doi.org/10.1016/j.jfoodeng.2006.04.046. URL: http://www.sciencedirect.com/science/article/pii/S0260877406003670.

[83] Dorijan Radočaj, Mladen Jurišić, and Mateo Gašparović. «The Role of Remote Sensing Data and Methods in a Modern Approach to Fertilization in Precision Agriculture.» In: *Rem. Sens.* 14.3 (2022), p. 778.

[84] Timothy W Randolph. «Scale-based normalization of spectral data.» In: *Cancer Biomarkers* (2006), pp. 135–144.

[85] Partha Pratim Ray. «Internet of things for smart agriculture: Technologies, practices and future direction.» In: *J. Amb. Intell. Smart Env.* 9.4 (2017), pp. 395–420.

[86] Jordi Riu, Giulia Gorla, Dib Chakif, Ricard Boqué, and Barbara Giussani. «Rapid Analysis of Milk Using Low-Cost Pocket-Size NIR Spectrometers and Multivariate Analysis.» In: *Foods* 9.8 (2020). ISSN: 2304-8158. DOI: 10.3390/foods9081090. URL: https://www.mdpi.com/2304-8158/9/8/1090.

[87] Mai Said, Ayman Wahba, and Diaa Khalil. «Semi-supervised deep learning framework for milk analysis using NIR spectrometers.» In: *Chemometrics and Intelligent Laboratory Systems* 228 (2022), p. 104619. ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2022.104619. URL: https://www.sciencedirect.com/science/article/pii/S0169743922001307.

[88] Ministero della Salute. *14 novembre 2022, Giornata mondiale del diabete — salute.gov.it.* https://www.salute.gov.it/portale/nutrizione/dettaglioNotizieNutrizione.jsp?lingua=italiano&menu=notizie&p=dalministero&id=6064. [Accessed 16-Nov-2022]. 2022.

[89] Lee Sanchez, Charles Farber, Jiaxin Lei, Keyan Zhu-Salzman, and Dmitry Kurouski. «Noninvasive and nondestructive detection of cowpea bruchid within cowpea seeds with a handheld Raman spectrometer.» In: *Anal. Chem.* 91.3 (2019), pp. 1733–1737.

[90] Abraham. Savitzky and M. J. E. Golay. «Smoothing and Differentiation of Data by Simplified Least Squares Procedures.» In: *Analytical Chemistry* 36.8 (1964), pp. 1627–1639. DOI: 10.1021/ac60214a047. eprint: https://doi.org/10.1021/ac60214a047. URL: https://doi.org/10.1021/ac60214a047.

[91] Leo Schuler, Jason Milne, John Dell, and Lorenzo Faraone. «MEMS-based microspectrometer technologies for NIR and MIR wavelengths.» English. In: *Journal of Physics D-Applied Physics* 42.13 (2009), 13 pages. ISSN: 0022-3727. DOI: 10.1088/0022-3727/42/13/133001.

[92]   Heinz Siesler, Yukihiro Ozaki, Satoshi Kawata, and Herbert Heise. *Near-infrared spectroscopy: Principles, instruments, applications*. Feb. 2002. ISBN: 3527301496.

[93]   Bam Bahadur Sinha and R Dhanalakshmi. «Recent advancements and challenges of Internet of Things in smart agriculture: A survey.» In: *Fut. Gen. Comp. Syst.* 126 (2022), pp. 169–184.

[94]   C. Srichan, W. Srichan, P. Danvirutai, C. Ritsongmuang, A. Sharma, and S. Anutrakulchai. «Non-invasively accuracy enhanced blood glucose sensor using shallow dense neural networks with NIR monitoring and medical features.» English. In: *Scientific Reports* 12.1 (2022). cited By 2. ISSN: 20452322. DOI: 10.1038/s41598-022-05570-8.

[95]   Jiayu Sun, Xinzhou Wang, Naixue Xiong, and Jie Shao. «Learning sparse representation with variational auto-encoder for anomaly detection.» In: *IEEE Access* 6 (2018), pp. 33353–33361.

[96]   Xudong Sun, Ke Zhu, and Junbin Liu. «Nondestructive detection of reducing sugar of potato flours by near infrared spectroscopy and kernel partial least square algorithm.» In: *Journal of Food Measurement and Characterization* 13.1 (2019), pp. 231–237.

[97]   Éva Szabó, Szilveszter Gergely, Tamás Spaits, Tamás Simon, and András Salgó. «Near-infrared spectroscopy-based methods for quantitative determination of active pharmaceutical ingredient in transdermal gel formulations.» In: *Spectroscopy Letters* 52.10 (2019), pp. 599–611. DOI: 10.1080/00387010.2019.1681459. eprint: https://doi.org/10.1080/00387010.2019.1681459. URL: https://doi.org/10.1080/00387010.2019.1681459.

[98]   Dagmawi Delelegn Tegegn, Italo Zoppis, Sara Manzoni, Cezar Sas, and Edoardo Lotti. «Convolutional Neural Networks for Quantitative Prediction of Different Organic Materials using Near-Infrared Spectrum.» In: *The title of the conference proceedingsProceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOSIGNALS*. SciTePress, 2021, pp. 169–176.

[99]   A P Teixeira, R Oliveira, P M Alves, and M J T Carrondo. «Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative.» en. In: *Biotechnol Adv* 27.6 (2009), pp. 726–732.

[100]   Ernest Teye, Charles L.Y. Amuah, Terry McGrath, and Christopher Elliott. «Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics.» In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spec-*

troscopy 217 (2019), pp. 147–154. ISSN: 1386-1425. DOI: https://doi.org/10.1016/j.saa.2019.03.085. URL: http://www.sciencedirect.com/science/article/pii/S1386142519303324.

[101] Sanna Uusitalo, Ben Aernouts, Juha Sumen, Eero Hietala, Mikko Utriainen, Lilli Frondelius, Sari Kajava, and Matti Pastell. *Comparison of milk analysis performance between NIR laboratory analyser and miniaturised NIR MEMS sensors*. 2019-12-12.

[102] Maribel Vásconez, Édgar Pérez-Esteve, Alberto Arnau-Bonachera, José Barat, and Pau Talens. «Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy.» In: *Food Control* 92 (2018), pp. 183–189. ISSN: 0956-7135. DOI: https://doi.org/10.1016/j.foodcont.2018.05.001. URL: http://www.sciencedirect.com/science/article/pii/S0956713518302263.

[103] Kerry B. Walsh, José Blasco, Manuela Zude-Sasse, and Xudong Sun. «Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use.» In: *Postharvest Biology and Technology* 168 (2020), p. 111246. ISSN: 0925-5214. DOI: https://doi.org/10.1016/j.postharvbio.2020.111246. URL: https://www.sciencedirect.com/science/article/pii/S0925521419303230.

[104] Aichen Wang, Dong Hu, and Lijuan Xie. «Comparison of detection modes in terms of the necessity of visible region (VIS) and influence of the peel on soluble solids content (SSC) determination of navel orange using VIS–SWNIR spectroscopy.» In: *Journal of Food Engineering* 126 (2014), pp. 126–132. ISSN: 0260-8774. DOI: https://doi.org/10.1016/j.jfoodeng.2013.11.011. URL: https://www.sciencedirect.com/science/article/pii/S0260877413005876.

[105] Hui Wang, Du Lv, Nan Dong, Sijie Wang, and Jia Liu. «Application of near-infrared spectroscopy for screening the potato flour content in Chinese steamed bread.» In: *Food science and biotechnology* 28.4 (2019), pp. 955–963.

[106] Mial E. Warren. *Micro-opto-electro-mechanical systems (MOEMS)*. en. New York, 2020. DOI: 10.1036/1097-8542.757728. URL: https://doi.org/10.1036/1097-8542.757728.

[107] Verena Wiedemair, Dominik Langore, Roman Garsleitner, Klaus Dillinger, and Christian Huck. «Investigations into the Performance of a Novel Pocket-Sized Near-Infrared Spectrometer for Cheese Analysis.» In: *Molecules* 24.3 (2019). ISSN: 1420-3049. DOI: 10.3390/molecules24030428. URL: https://www.mdpi.com/1420-3049/24/3/428.

[108] Verena Wiedemair, Dominik Mair, Carina Held, and Christian W. Huck. «Investigations into the use of handheld near-infrared spectrometer and novel semi-automated data analysis for the determination of protein content in different cultivars of Panicum miliaceum L.» In: *Talanta* 205 (2019), p. 120115. ISSN: 0039-9140. DOI: https://doi.org/10.1016/j.talanta.2019.120115. URL: https://www.sciencedirect.com/science/article/pii/S0039914019307416.

[109] William R Windham, Brenda G Lyon, Elaine T Champagne, Franklin E Barton, Bill D Webb, Anna M McClung, Karen A Moldenhauer, Steve Linscombe, and Kent S McKenzie. «Prediction of cooked rice texture quality using near-infrared reflectance analysis of whole-grain milled samples.» In: *Cereal Chemistry* 74.5 (1997), pp. 626–632.

[110] R F Wolffenbuttel. «MEMS-based optical mini- and microspectrometers for the visible and infrared spectral range.» In: *Journal of Micromechanics and Microengineering* 15.7 (2005), S145. DOI: 10.1088/0960-1317/15/7/021. URL: https://dx.doi.org/10.1088/0960-1317/15/7/021.

[111] Young-Ah Woo, Jhii-Weon Ahn, In-Koo Chun, and Hyo-Jin Kim. «Development of a Method for the Determination of Human Skin Moisture Using a Portable Near-Infrared System.» In: *Analytical Chemistry* 73.20 (2001), pp. 4964–4971. ISSN: 0003-2700. DOI: 10.1021/ac0102563. URL: https://doi.org/10.1021/ac0102563.

[112] D. Wu, S. Feng, and Y. He. «Short-Wave Near-Infrared Spectroscopy of Milk Powder for Brand Identification and Component Analysis.» In: *Journal of Dairy Science* 91.3 (2008), pp. 939–949. ISSN: 0022-0302. DOI: https://doi.org/10.3168/jds.2007-0640. URL: http://www.sciencedirect.com/science/article/pii/S0022030208713493.

[113] Hui Yan and Heinz W Siesler. «Identification of textiles by handheld near infrared spectroscopy: Protecting customers against product counterfeiting.» In: *J. Near Infrared Spectrosc.* 26.5 (2018), pp. 311–321. URL: https://opg.optica.org/jnirs/abstract.cfm?URI=jnirs-26-5-311.

[114] Hui Yan and Heinz W. Siesler. «Quantitative analysis of a pharmaceutical formulation: Performance comparison of different handheld near-infrared spectrometers.» In: *Journal of Pharmaceutical and Biomedical Analysis* 160 (2018), pp. 179–186. ISSN: 0731-7085. DOI: https://doi.org/10.1016/j.jpba.2018.07.048. URL: https://www.sciencedirect.com/science/article/pii/S0731708518313645.

[115] Hui Yan and Heinz W. Siesler. «Quantitative analysis of a pharmaceutical formulation: Performance comparison of different handheld near-infrared spectrometers.» In: *Journal of Pharmaceutical and Biomedical Analysis* 160 (2018), pp. 179–186. ISSN: 0731-7085. DOI: https://doi.org/10.1016/j.jpba.2018.07.048. URL: https://www.sciencedirect.com/science/article/pii/S0731708518313645.

[116] Jingru Yang, Jin Wang, Guodong Lu, Shaomei Fei, Ting Yan, Cheng Zhang, Xiaohui Lu, Zhiyong Yu, Wencui Li, and Xiaolin Tang. «TeaNet: Deep learning on Near-Infrared Spectroscopy (NIR) data for the assurance of tea quality.» In: *Computers and Electronics in Agriculture* 190 (2021), p. 106431. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2021.106431. URL: https://www.sciencedirect.com/science/article/pii/S0168169921004488.

[117] Lei Yin and Yu Zhang. «Village precision poverty alleviation and smart agriculture based on FPGA and machine learning.» In: *Microproc. Microsyst.* (2020), p. 103469.

[118] A. Zancanaro, G. Cisotto, and L. Badia. «Challenges of the Age of Information Paradigm for Metrology in Cyberphysical Ecosystems.» In: *Proc. MetroLivEnv.* 2022.

[119] Alberto Zancanaro, Giulia Cisotto, and Leonardo Badia. «Modeling value of information in remote sensing from correlated sources.» In: *Proc. IEEE MedComNet.* 2022, pp. 47–53.

[120] Muhammad Zareef, Quansheng Chen, Md Mehedi Hassan, Muhammad Arslan, Malik Muhammad Hashim, Waqas Ahmad, Felix YH Kutsanedzie, and Akwasi A Agyekum. «An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis.» In: *Food Engineering Reviews* (2020), pp. 1–18.

[121] Lei Zhang, Xiangqian Ding, and Ruichun Hou. «Classification Modeling Method for Near-Infrared Spectroscopy of Tobacco Based on Multimodal Convolution Neural Networks.» In: *Journal of Analytical Methods in Chemistry* 2020 (2020).

[122] John-Lewis Zinia Zaukuu, Balkis Aouadi, Mátyás Lukács, Zsanett Bodor, Flóra Vitális, Biborka Gillay, Zoltan Gillay, László Friedrich, and Zoltan Kovacs. «Detecting Low Concentrations of Nitrogen-Based Adulterants in Whey Protein Powder Using Benchtop and Handheld NIR Spectrometers and the Feasibility of Scanning through Plastic Bag.» In: *Molecules* 25.11 (2020). ISSN: 1420-3049. DOI: 10.3390/molecules25112522. URL: https://www.mdpi.com/1420-3049/25/11/2522.

[123] Begoña de la Roza-Delgado, Ana Garrido-Varo, Ana Soldado, Amelia González Arrojo, María Cuevas Valdés, Francisco Maroto, and Dolores Pérez-Marín. «Matching portable NIRS instruments for in situ monitoring indicators of milk composition.» In: *Food Control* 76 (2017), pp. 74–81. ISSN: 0956-7135. DOI: https://doi.org/10.1016/j.foodcont.2017.01.004. URL: https://www.sciencedirect.com/science/article/pii/S095671351730004X.