



## Non-Destructive Measurement of Caffeine Content on Intact Coffee Beans Using Near Infrared Reflectance Spectroscopy

Maiful Hari<sup>1</sup>, Agus Arip Munawar<sup>2\*</sup>, Yusriana<sup>1</sup>

<sup>1</sup> Department of Master of Agricultural Industrial Technology, Syiah Kuala University, Banda Aceh 23111, Indonesia

<sup>2</sup> Department of Agricultural Engineering, IE Lab, PUSMEPTAN Syiah Kuala University, Banda Aceh 23111, Indonesia

Corresponding Author Email: [aamunawar@usk.ac.id](mailto:aamunawar@usk.ac.id)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/i2m.250104>

### ABSTRACT

**Received:** 30 October 2025

**Revised:** 5 January 2026

**Accepted:** 14 January 2026

**Available online:** 28 February 2026

#### Keywords:

*Arabica coffee, caffeine, Near-Infrared Spectroscopy, Partial Least Squares Regression, spectrum correction*

This study evaluated various spectrum correction techniques to enhance prediction accuracy. The detrending (DT) pretreatment method demonstrated superior performance, achieving a correlation coefficient ( $r$ ) of 0.99 and a coefficient of determination ( $R^2$ ) of 0.99, significantly outperforming other methods. The residual predictive deviation (RPD) value for DT was 11.74, far exceeding the threshold of 3 that generally indicates good predictive ability. In comparison, standard normal variate (SNV) pretreatment yielded the next best results with an RPD of 3.76, while other methods produced RPD values between 3.46 and 3.51. The effectiveness of Near-Infrared Spectroscopy (NIRS) technology, particularly when coupled with appropriate spectrum correction methods like DT, presents numerous applications in the coffee industry. For instance, during the blending process, NIRS could rapidly assess caffeine levels in different bean batches, allowing for precise formulation of coffee blends with specific caffeine profiles. In quality control during procurement, it could enable quick on-site testing of incoming green coffee beans, ensuring they meet required caffeine specifications before acceptance. Additionally, roasters could use NIRS to monitor caffeine content throughout the roasting process, fine-tuning their techniques to achieve desired flavor profiles while maintaining consistent caffeine levels. This study confirms NIRS as a promising tool for efficient, accurate, and environmentally friendly chemical analysis in the coffee sector. Its potential extends beyond caffeine prediction to other quality parameters, offering a comprehensive solution for quality assurance in the increasingly competitive global coffee market.

## 1. INTRODUCTION

Arabica coffee (*Coffea arabica L.*) is one of the most popular coffee varieties in the world. This coffee is known for its smooth, complex taste and rich aroma. Arabica coffee is generally grown in the highlands at an altitude of around 600-2000 meters above sea level [1]. These unique characteristics make Arabica coffee more favored by international consumers. In addition, Arabica coffee has a lower caffeine content compared to Robusta coffee, making it a preferred choice for consumers seeking a milder and less bitter taste [2].

Caffeine is an alkaloid compound naturally found in coffee, tea, chocolate, and some other plants. This compound functions as a stimulant in the central nervous system, increases alertness, and can reduce fatigue. Caffeine has an important role in the coffee industry, because it not only affects the taste of coffee, but also the physiological effects on consumers. In Arabica coffee, the caffeine content ranges from 0.8% to 1.5%, much lower than that of Robusta coffee, which has a caffeine content of around 1.7% to 2.5% [3].

The importance of caffeine in coffee is not only related to flavor and aroma, but also to quality standards. Many consumers are concerned about caffeine levels when choosing coffee, due to its physiological effects. High caffeine levels

can cause increased heart rate, anxiety and sleep problems, so coffee producers must ensure that caffeine levels are in line with consumer expectations and applicable quality standards. In this context, testing caffeine levels is important to ensure the quality of coffee products to be marketed [4].

Conventional methods for determining caffeine levels, such as High-Performance Liquid Chromatography (HPLC), require a long time, high cost, and complex sample preparation. HPLC works by separating the chemical components in a sample for identification and quantification, but due to the destructive nature of the process, it is less suitable for applications that require rapid, non-destructive analysis of the sample. Therefore, there is currently a need for alternative methods that are more efficient and non-destructive. Near-Infrared Spectroscopy (NIRS) is a technology that can answer this challenge. NIRS is a non-destructive analytical technology that can measure the absorption of infrared light by molecules in a sample. This technology does not require complicated sample preparation and can provide fast results. NIRS is widely used in the agriculture and food industries to predict chemical content, including moisture, protein, and caffeine content, as well as overall product quality. The advantage of NIRS lies in its ability to perform simultaneous and non-destructive analysis,

making it highly efficient in controlling product quality [5].

Partial Least Squares Regression (PLSR) is one of the data analysis methods often used in conjunction with NIRS. PLSR allows researchers to build prediction models based on NIRS and sample chemistry data. This method is very useful in spectrum analysis, especially when the data has high multicollinearity [6]. PLSR works by relating spectrum data (X) to chemical variables (Y), such as caffeine content, thus enabling high-accuracy prediction of chemical composition.

However, spectrum data obtained from NIRS often contains noise and distortion, so spectrum correction (pretreatment) is required to improve the accuracy of the prediction model. Pretreatment is the first step in spectrum analysis that aims to remove noise or distortion from the raw spectrum data. Some commonly used pretreatment methods are standard normal variate (SNV), detrending (DT), extended multiplicative scatter correction (EMSC), and the Savitzky-Golay (SG) filter. Each method has a specific purpose, such as reducing the effects of scattering, baseline shift, and noise, so that the resulting data is more stable and accurate for use in PLSR prediction models [7, 8].

The NIRS and PLSR directly address the challenges in caffeine measurement highlighted earlier in the introduction. NIRS offers a rapid, non-destructive method for analyzing chemical composition, including caffeine content, in coffee beans. This technology overcomes the limitations of conventional methods like HPLC, which are time-consuming, costly, and require complex sample preparation. NIRS works by measuring the absorption of infrared light by molecules in a sample, providing a spectral fingerprint that correlates with its chemical composition [9].

However, the complex nature of NIR spectra necessitates sophisticated data analysis techniques, which is where PLSR comes into play. PLSR is particularly well-suited for analyzing NIR spectral data because it can handle the high multicollinearity often present in such data. It works by creating a linear regression model that relates the spectral data (X variables) to the chemical composition (Y variables, in this case, caffeine content). This allows for the development of predictive models that can accurately estimate caffeine levels based on NIR spectra alone, without the need for destructive chemical analysis [10, 11].

By combining NIRS with PLSR, we can develop a method that provides rapid results, addressing the time constraints of conventional methods. Moreover, it can preserve valuable coffee samples, since NIRS is a non-destructive method. It also requires minimal sample preparation, reducing complexity and cost [12]. Also offers high accuracy through sophisticated data analysis. These characteristics directly align with the study's objectives of finding an efficient, accurate, and non-destructive method for caffeine content prediction in Arabica coffee beans.

With the development of NIRS technology and pretreatment methods, testing the caffeine content of Arabica coffee green beans can be done more quickly, efficiently and accurately, without damaging the coffee samples. This provides great benefits to the coffee industry, especially in ensuring the quality of products that will be exported to international markets.

In the field of non-destructive analytical techniques, NIRS stands out as a particularly advantageous method for analyzing coffee beans, especially when compared to other prominent techniques such as Raman Spectroscopy, Hyperspectral Imaging, and X-ray Fluorescence (XRF). Raman

Spectroscopy, while offering detailed molecular fingerprints and being less susceptible to water interference than NIRS, faces challenges with fluorescence interference and typically requires more sophisticated, costly equipment [13, 14]. Hyperspectral Imaging, which ingeniously combines spectroscopy with imaging capabilities, allows for spatial distribution analysis of chemical components but generates extensive datasets that demand complex processing and specialized equipment. XRF, though excellent for non-destructive elemental analysis, falls short when it comes to analyzing organic compounds like caffeine, which is crucial in coffee bean assessment [15].

NIRS, however, presents a unique combination of advantages that make it particularly well-suited for the coffee industry. Its versatility allows for simultaneous analysis of multiple chemical components, not limited to caffeine alone, providing a comprehensive view of the bean's composition. The speed of NIRS is unparalleled, delivering results in mere seconds, which is significantly faster than most other spectroscopic methods. This rapid analysis capability is crucial in high-volume coffee production and quality control scenarios.

From a financial perspective, NIRS equipment is generally more cost-effective compared to alternatives like Raman or Hyperspectral systems, making it a more accessible option for a wider range of coffee producers and researchers. The ease of use of NIRS systems is another significant advantage; they require minimal sample preparation and can be operated by individuals without specialized training, streamlining the analysis process in busy production environments. Furthermore, the portability of many NIRS systems, which are often compact and designed for field use, allows for on-site or in-line measurements, bringing laboratory-grade analysis capabilities directly to coffee farms or processing facilities [16, 17].

These combined advantages, coupled with NIRS's established track record in food and agricultural applications, position it as the ideal choice for studying caffeine content prediction in Arabica coffee beans. The technology's capacity to deliver rapid, accurate, and non-destructive analysis aligns perfectly with the coffee industry's growing need for efficient quality control methods in an increasingly competitive global market. As the coffee industry continues to evolve, with consumers demanding higher quality and more precise information about their coffee, NIRS technology offers a solution that can meet these demands while also improving production efficiency and consistency.

## 2. MATERIALS AND METHODS

### 2.1 Arabica coffee bean samples

The Arabica coffee bean samples used in this study were sourced from various regions in Indonesia, including the Gayo area of Aceh Province, Pangalengan and Cianjur of West Java Province, Toraja and Sidenreng Rappang of South Sulawesi Province, and Pupuan Kintamani of Bali Province. A total of 32 samples were collected, with each sample weighing 50 g [5].

### 2.2 Sample preparation

The sample preparation process for the Arabica coffee green

beans underwent a standardized preparation protocol. Initially, the green coffee beans were subjected to a controlled drying process. This crucial step involved air-drying the beans at a carefully monitored room temperature of  $25\text{ }^{\circ}\text{C} \pm 2\text{ }^{\circ}\text{C}$  for a duration of 24 hours [4, 5].

The purpose of this drying phase was to eliminate any excess moisture that might have been present in the beans, thus standardizing their moisture content and preventing potential variations in spectral readings due to water interference. This precise temperature control and duration were essential to avoid any unintended alterations to the beans' chemical composition while effectively removing surface moisture. Following the drying process, a meticulous sorting procedure was implemented. This step involved the manual inspection and removal of any defective beans or foreign materials from the samples.

The sorting process was crucial in ensuring the purity and homogeneity of the samples, as the presence of defective beans or extraneous matter could potentially skew the spectral data and compromise the accuracy of the caffeine content predictions. This careful selection process contributed significantly to the overall quality and reliability of the subsequent NIRS analysis. After sorting, the samples were carefully stored in specially selected containers. These storage vessels were both airtight and opaque, designed to protect the beans from external environmental factors that could affect their chemical composition.

The storage conditions were strictly controlled, maintaining a consistent room temperature of  $25\text{ }^{\circ}\text{C} \pm 2\text{ }^{\circ}\text{C}$  and a relative humidity of 60% [18]. These specific conditions were chosen to preserve the integrity of the beans and prevent any moisture reabsorption or potential degradation of chemical compounds, including caffeine, which could occur under fluctuating environmental conditions. Before the NIRS analysis, a final pre-analysis conditioning step was implemented. The samples were removed from storage and allowed to equilibrate at room temperature for a period of 2 hours. This equilibration period was critical in ensuring that all samples reached a uniform temperature throughout their mass.

Temperature uniformity is particularly important in NIRS analysis, as temperature variations can affect the spectral

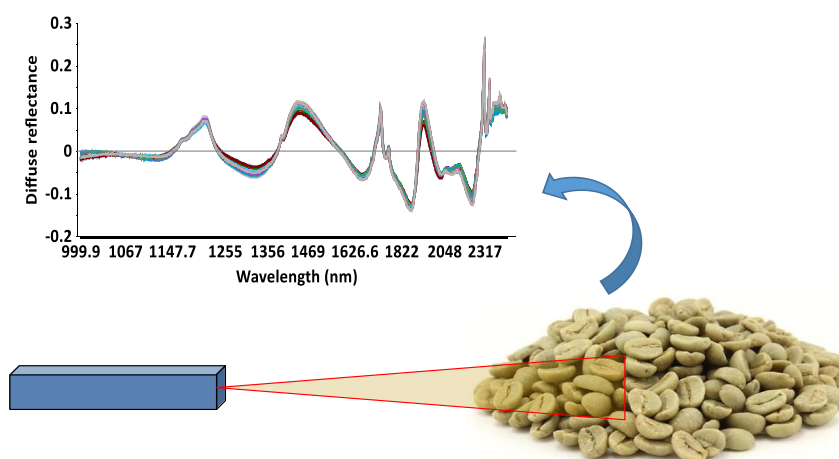
readings and potentially introduce errors in the caffeine content predictions. A notable aspect of this sample preparation protocol was the decision to analyze the beans in their whole, unground state. This approach was deliberately chosen to develop a method applicable to whole green bean analysis, which aligns more closely with practical on-site quality control needs in the coffee industry.

By avoiding grinding, the study maintained the integrity of the bean structure, allowing for a more realistic assessment of how NIRS technology could be applied in real-world coffee production and quality control scenarios. This whole-bean analysis approach enhances the study's relevance and potential for immediate application in industrial settings, where rapid, non-destructive testing of intact coffee beans is highly desirable.

### 2.3 Near-infrared spectral data

The NIRS analysis was performed using a Thermo Nicolet Antaris II<sup>TM</sup> spectrophotometer. For each scan, a 50 g pile of whole green coffee beans was placed in the sample holder. This 50 g weight refers to the amount used for each scan, not the total weight of all samples combined. The spectrophotometer was set to operate in the wavelength range of 1,000 nm to 2,500 nm ( $4,000\text{--}10,000\text{ cm}^{-1}$ ), controlled by Thermo Operation<sup>®</sup> software. The instrument was configured to perform 64 scans per sample to acquire diffuse reflectance spectra as illustrated in Figure 1 [19].

This multiple scan approach helps to reduce noise and improve the signal-to-noise ratio of the resulting spectra. Each sample was scanned in triplicate, with the sample being removed and repacked between each scan to account for potential variations in bean orientation and packing density. The average of these three scans was used for further analysis. The resulting spectra were stored in two file formats: Spectra Analysis (SPA) for detailed spectral information, and Comma-Separated Values (CSV) for ease of data processing. The NIR spectra were recorded at wavelengths from 1,000 to 2,500 nm with a spectral resolution of 0.4 nm. These detailed procedures ensure reproducibility of the experiment and provide a clear understanding of how the samples were handled and analyzed.



**Figure 1.** Data acquisition of NIR spectra for coffee bean samples in the wavelength range of 1,000–2,500 nm

### 2.4 Spectral data analysis

The obtained spectrum data were analyzed using the PLSR

method to predict caffeine content. Prior to analysis, several spectrum correction techniques (pretreatment) were applied to reduce noise and scatter effects in the raw spectra. The

spectrum correction techniques used in this study include:

- SNV: This technique is used to reduce the scatter effect that occurs due to particle size non-uniformity.
- DT: Used to remove linear trends in spectrum data.
- EMSC: Used to correct the effects of multiple scatter on the spectrum.
- Peak normalization (PN): This technique is used to normalize the spectrum peaks to make the intensity values more uniform.
- Median filter (MF): A nonlinear smoothing technique used to reduce noise in spectral data.
- SG filter: A more advanced smoothing technique to maintain the integrity of the spectrum signal while reducing noise [8, 20, 21].

The rationale for choosing this specific suite of pretreatments is grounded in their targeted ability to correct distinct types of spectral distortions that can obscure the chemical information of interest, thereby enabling a rigorous evaluation of which preprocessing strategy is most effective for the non-destructive prediction of caffeine.

A primary challenge in analyzing intact coffee beans is light scattering due to inherent variations in bean size, shape, surface roughness, and internal structure. These physical heterogeneities cause both multiplicative, affecting slope and additive, affecting baseline scatter effects, which are often more pronounced than the subtle absorbance changes related to chemical composition. To mitigate this, we employed two established methods: SNV and EMSC. SNV operates on each spectrum individually by centering, subtracting the mean and scaling, dividing by the standard deviation, effectively removing scatter-induced variation related to particle size. EMSC is a more advanced technique that models and corrects for scatter effects by incorporating a polynomial baseline in its algorithm, aiming to separate the physical light-scattering phenomena from the chemical absorbance information. Including both allows us to compare a simpler, widely-used normalization approach against a more sophisticated, model-based correction.

NIR spectra can exhibit unwanted baseline offsets or curvilinear trends arising from sources such as instrumental drift over time, minor variations in sample positioning within the holder, or differences in the packing density of the bean pile. These variations are unrelated to caffeine concentration but can severely degrade model performance if uncorrected. DT was specifically selected to address this issue. DT works by fitting and subtracting a polynomial trend from the spectral data, effectively "flattening" the baseline. This process ensures that subsequent modeling focuses on the absorbance bands themselves rather than on artifactual baseline shifts.

Spectroscopic measurements invariably contain a degree of random, high-frequency electronic or instrumental noise. While averaging multiple scans, as done in this study, reduces noise, additional digital smoothing can further enhance the signal-to-noise ratio. We included two distinct smoothing algorithms to evaluate their utility. MF replaces each spectral data point with the median value of its neighbors within a defined window. This method is particularly robust against spike noise. SG filter, in contrast, performs a local polynomial regression to smooth the data. It is favored for its superior ability to preserve the shape and width of spectral peaks—a critical aspect when the integrity of absorption band features must be maintained for accurate quantitative analysis.

Differences in the physical presentation of samples, such as slight variations in the thickness of the bean layer or the

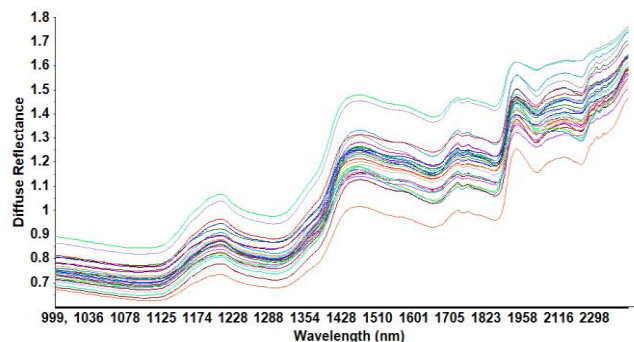
distance from the spectrometer's detector, can lead to global changes in spectral intensity. These path-length differences do not reflect changes in chemical concentration but can dominate spectral variance. PN was chosen to correct for this effect. PN scales the entire spectrum so that a specific feature is set to a constant value, thereby minimizing variance due to physical presentation and allowing the model to focus on relative absorbance patterns.

The PLSR models were developed and optimized using full cross-validation, leave-one-out. The optimal number of latent variables was selected by minimizing the Root Mean Square Error of Cross-Validation (RMSECV).

### 3. RESULTS AND DISCUSSION

#### 3.1 Spectra features of Arabica coffee bean

The raw spectrum of Arabica coffee green beans, as depicted in Figure 2, provides a wealth of information about the chemical composition of the samples. This spectrum, obtained through NIRS in the wavelength range of 1,000–2,500 nm, serves as a complex fingerprint of the beans' molecular makeup. Within this near-infrared region, various organic molecules present in coffee beans exhibit characteristic absorption patterns, allowing for non-destructive analysis of their composition.



**Figure 2.** Near-Infrared (NIR) spectra features of Arabica coffee beans in the 1,000–2,500 nm wavelength region

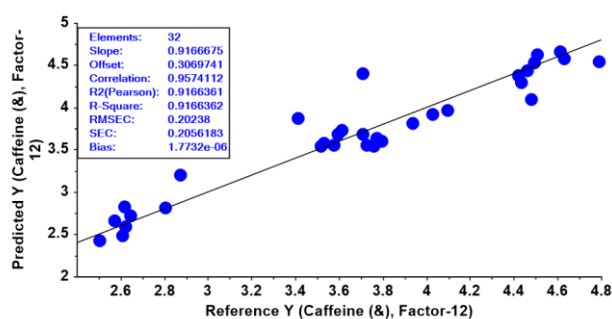
Important spectral features of the raw spectrum reveal significant insights into the beans' chemical structure. Notably, based on previous literature, broad absorption bands are observed around 1,450 nm and 1,940 nm, which can be primarily attributed to O–H stretching overtones [9, 20, 22]. These bands are particularly indicative of the moisture content within the coffee beans, a crucial factor in determining both quality and storage stability. A prominent peak near 2,270 nm is of particular interest, as it likely corresponds to C–H stretching and C=O combination bands, which are characteristic molecular vibrations associated with caffeine molecules [23, 24]. This peak serves as a potential marker for caffeine content, the primary focus of this study. The spectral region between 2,000 and 2,500 nm presents a complex landscape of overlapping bands, reflecting the presence of a diverse array of organic compounds within the coffee beans. These include proteins, lipids, and carbohydrates, each contributing to the overall spectral profile.

The intricate nature of these overlapping bands underscores the complexity of coffee bean composition and highlights the

challenge in isolating specific components for analysis. While the raw spectrum provides valuable initial insights, it is important to note that this data is subject to various physical phenomena that can affect its interpretation. Scattering effects, caused by the physical structure of the coffee beans and variations in particle size, can introduce noise and baseline shifts in the spectrum [25]. These effects, if left uncorrected, could potentially mask or distort the chemical information of interest, particularly the subtle spectral features associated with caffeine content. Consequently, the raw spectral data, while informative, necessitate careful pretreatment to enhance their accuracy and reliability for quantitative analysis of specific compounds like caffeine.

### 3.2 Caffeine prediction using non-treated spectra

Preliminary results of the prediction of Arabica coffee caffeine content without pretreatment showed good results, with a correlation coefficient ( $r$ ) of 0.95 and a coefficient of determination ( $R^2$ ) of 0.91, as shown in Figure 3.



**Figure 3.** Caffeine prediction using raw spectral data

This value indicates that the developed model has a very close correlation between the raw spectrum and the caffeine content measured using the HPLC method. The Root Mean Square Error of Calibration (RMSEC) value of 0.20238, which is lower than the standard deviation of the data, indicates that the model can provide fairly accurate predictions. However, while these results show that the raw spectrum data is good enough for caffeine prediction, it is important to remember that spectrum data often contains noise and distortion, which can reduce the accuracy of long-term predictions.

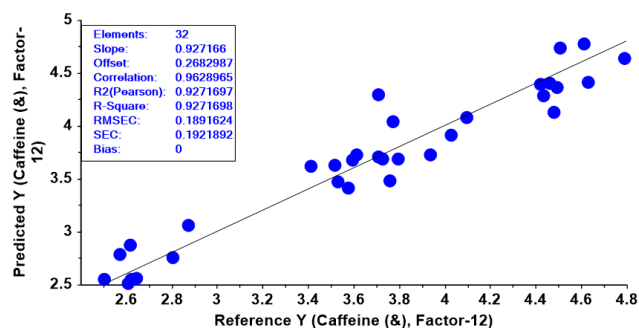
### 3.3 Caffeine prediction after standard normal variate spectra correction

The SNV pretreatment method demonstrated significant improvements in the prediction model's performance for caffeine content in Arabica coffee green beans. After applying SNV, the value of  $r$  increased to 0.96, and  $R^2$  reached 0.92, indicating a stronger relationship between the spectral data and the actual caffeine content, as presented in Figure 4.

SNV works by normalizing each spectrum individually, setting the mean to 0 and the standard deviation to 1. This normalization process is particularly effective in addressing variations caused by particle size differences and surface scattering effects, which are common challenges in NIRS analysis of solid samples like coffee beans.

The improvement in prediction accuracy after SNV pretreatment suggests that scattering effects were indeed masking some of the caffeine-related spectral information in

the raw data. Coffee beans, being a heterogeneous material, can cause significant light scattering due to variations in particle size, shape, and surface characteristics [8, 20]. These scattering effects can introduce noise and baseline shifts in the spectra, potentially obscuring the subtle spectral features associated with caffeine content.



**Figure 4.** Caffeine prediction after standard normal variate (SNV) spectral data

By reducing these scattering effects, SNV allows the PLSR model to focus more accurately on the spectral changes directly related to caffeine concentration. This enhanced focus results in a more robust and reliable prediction model. The higher  $R^2$  value of 0.92 indicates that 92% of the variance in caffeine content can be explained by the spectral data after SNV pretreatment, a significant improvement over the raw spectral analysis.

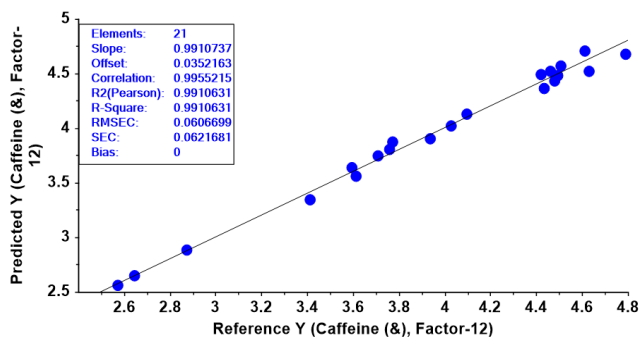
It's worth noting that while SNV showed considerable improvement, it was not the best-performing pretreatment method in this study. The DT method achieved even better results with  $r = 0.99$  and  $R^2 = 0.99$ . This suggests that while scattering effects were a significant source of spectral variation, other factors, such as baseline shifts, might also play a crucial role in the accuracy of caffeine prediction models for coffee beans. The effectiveness of SNV in this context underscores its value as a pretreatment method for NIRS analysis, particularly in applications involving solid agricultural products where particle size and surface scattering can significantly impact spectral quality.

### 3.4 Caffeine prediction after detrending spectra correction

DT pretreatment emerged as the most effective method for improving the prediction of caffeine content in Arabica coffee green beans using NIRS and PLSR. The results obtained with DT pretreatment were exceptional, with  $r = 0.99$  and  $R^2 = 0.99$ , as presented in Figure 5, indicating an almost perfect linear relationship between the predicted and actual caffeine values.

The residual predictive deviation (RPD) value of 11.74 is particularly noteworthy. In NIRS analysis, RPD values above 3 are generally considered good for quantitative predictions, while values above 8 are excellent. The RPD of 11.74 achieved with DT pretreatment far exceeds these benchmarks, suggesting that the model is highly reliable and accurate for predicting caffeine content in unknown samples. DT works by removing linear or curvilinear trends in the spectral data [26-28]. These trends can arise from various sources unrelated to the chemical composition of interest, such as instrument drift over time, variations in sample thickness or density, differences in sample presentation or packing, and also environmental factors like temperature or humidity

fluctuations.



**Figure 5.** Caffeine prediction after detrending (DT) spectral data

By eliminating these extraneous trends, DT allows the PLSR model to focus more precisely on the spectral features that are directly related to caffeine content. This is evident in the dramatically improved prediction accuracy compared to other pretreatment methods and the raw spectra. The effectiveness of DT in this study suggests that baseline variations were indeed a major source of error in the raw spectra of the coffee samples. These baseline shifts can mask or distort the subtle spectral changes associated with varying caffeine concentrations. By removing these variations, DT essentially levels the playing field for all spectra, making the caffeine-related features more prominent and easier for the PLSR model to interpret [29].

The superior performance of DT is further underscored by the RMSEC value of 0.06066, which is significantly lower than that obtained with other pretreatment methods. This low RMSEC indicates that the predictions made by the model have very small deviations from the actual caffeine values, further confirming the high accuracy and reliability of the DT-PLSR model.

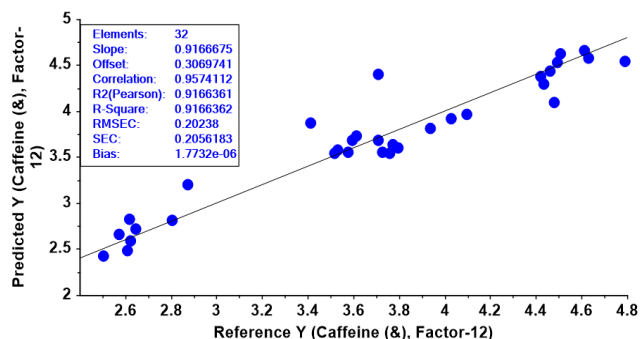
The exceptional results achieved with DT pretreatment have significant implications for the coffee industry. They demonstrate that NIRS, when combined with appropriate data preprocessing techniques, can provide a rapid, non-destructive, and highly accurate method for determining caffeine content in Arabica coffee beans. This could potentially revolutionize quality control processes in coffee production, enabling real-time monitoring and adjustment of caffeine levels to meet specific standards or consumer preferences.

### 3.5 Caffeine prediction after extended multiplicative scatter correction spectra correction

The EMSC pretreatment method was applied to the NIRS data of Arabica coffee green beans to evaluate its effectiveness in improving the prediction accuracy of caffeine content. The results obtained from EMSC pretreatment showed a moderate improvement over the raw spectra, with  $r = 0.91$  and  $R^2 = 0.91$ , as shown in Figure 6.

EMSC is a sophisticated pretreatment technique designed to separate chemical and physical information within the spectra. It corrects for both multiplicative and additive scatter effects, which are common issues in NIRS analysis. Multiplicative scatter effects arise from differences in sample thickness, particle size, and surface roughness, while additive scatter effects are due to baseline shifts caused by instrument drift or

sample presentation variations. By addressing these scatter effects, EMSC aims to enhance the signal-to-noise ratio and improve the overall quality of the spectral data [30, 31].



**Figure 6.** Caffeine prediction after extended multiplicative scatter correction (EMSC) spectral data

The EMSC pretreatment resulted in an RPD value of 3.51, which indicates that the method is capable of providing good predictions. However, when compared to other pretreatment methods such as SNV and DT, EMSC's performance was somewhat lower. SNV achieved an RPD of 3.76, and DT yielded an exceptional RPD of 11.74, significantly outperforming EMSC.

The moderate improvement seen with EMSC suggests that while this method is effective in reducing some spectral variations, it may not be as well-suited for this particular dataset as simpler correction methods like SNV or DT. This discrepancy could indicate that the primary sources of spectral variation in these coffee samples are more effectively addressed by these simpler methods. For instance, the high performance of DT suggests that linear or curvilinear trends in the spectral data, possibly due to instrument drift or sample-to-sample variations, were a major source of error. SNV, on the other hand, effectively reduced scatter effects caused by particle size non-uniformity. In contrast, EMSC's more complex approach to correcting both multiplicative and additive scatter effects might not have been as targeted or efficient for the specific types of variations present in this dataset.

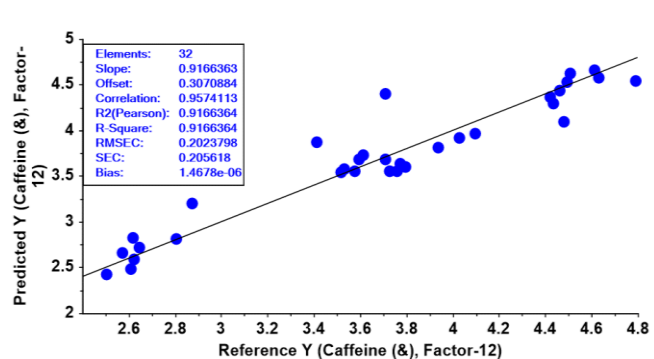
### 3.6 Caffeine prediction after peak normalization spectra correction

The PN demonstrated significant improvements in the prediction accuracy of caffeine content in Arabica coffee green beans, highlighting its utility in addressing specific types of spectral variations. With  $r = 0.96$  and  $R^2 = 0.91$ , the PN pretreatment outperformed the raw spectra, indicating a stronger relationship between the normalized spectral data and the actual caffeine content, as presented in Figure 7.

PN is a pretreatment technique designed to correct variations in the overall intensity of spectral peaks. This method is particularly useful when dealing with samples that exhibit differences in thickness or when there are variations in the instrument's response. By normalizing the intensity of the spectral peaks, PN ensures that the data is scaled to a uniform level, thereby minimizing the influence of these external factors on the spectral analysis.

The improvement seen with PN suggests that variations in peak intensity were indeed a significant factor affecting the model's ability to accurately quantify caffeine. In NIRS

analysis, the intensity of spectral peaks can vary due to several reasons, including differences in sample thickness, packing density, and instrument calibration. These variations can introduce noise and distortions in the raw spectra, making it challenging for the PLSR model to accurately detect the subtle spectral changes associated with caffeine content.



**Figure 7.** Caffeine prediction after peak normalization (PN) spectral data

By normalizing these peaks, the PN method allows the PLSR model to focus more precisely on the relative intensities of the caffeine-related absorption bands compared to other spectral features. This normalization process effectively reduces the impact of external factors, enabling the model to better capture the intrinsic chemical information related to caffeine. As a result, the model's ability to predict caffeine content is enhanced, as evidenced by the improved  $r$  and  $R^2$  values [32, 33].

The effectiveness of PN in this context underscores its practical significance in NIRS applications, particularly in scenarios where sample presentation and instrument response can vary. In the coffee industry, where consistency in quality control is crucial, PN can be a valuable tool for ensuring that spectral data is reliable and consistent across different samples and measurement conditions.

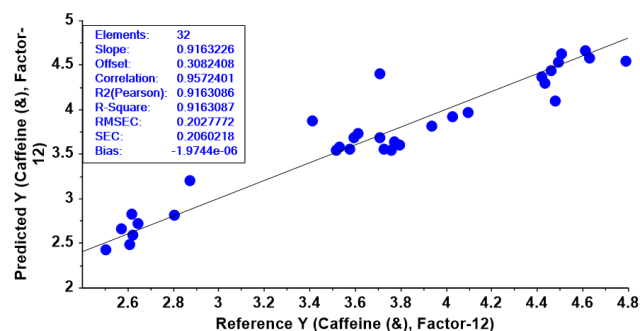
Moreover, the PN method is relatively straightforward to implement and does not require complex computational resources, making it a practical choice for routine quality control processes. The RMSEC value of 0.20237, which is comparable to other pretreatment methods, further supports the efficacy of PN in providing accurate predictions with minimal error. While PN showed significant improvements, it is important to compare its performance with other pretreatment methods. DT and SNV pretreatments, for instance, achieved even higher  $R^2$  values and RPD scores, indicating that they might be more effective in certain contexts.

However, PN's ability to correct for variations in peak intensity makes it a valuable addition to the arsenal of pretreatment techniques, especially when dealing with samples that exhibit notable differences in thickness or instrument response. The PN pretreatment method is a valuable tool for improving the accuracy of caffeine content predictions in Arabica coffee green beans using NIRS and PLSR. By normalizing the intensity of spectral peaks, PN effectively addresses variations that can affect the model's performance, allowing for more accurate and reliable predictions [34, 35]. This method, along with other pretreatment techniques, enhances the overall robustness and applicability of NIRS technology in the coffee industry, facilitating efficient and non-destructive quality control

processes.

### 3.7 Caffeine prediction after median filter spectra correction

The MF pretreatment method was applied to the NIRS spectral data of Arabica coffee green beans to evaluate its effectiveness in improving the prediction accuracy of caffeine content. The results obtained from MF pretreatment showed a moderate improvement, with  $r = 0.95$  and  $R^2 = 0.91$ . The RPD value of 3.51, as shown in Figure 8, indicates that this method is capable of providing good predictions, although it does not surpass the performance of other pretreatment methods like DT or SNV.



**Figure 8.** Caffeine prediction after median filter (MF) spectral data

MF is a spectral preprocessing technique designed to reduce noise in the spectra while preserving the sharp features that are crucial for accurate chemical analysis. This method works by replacing each data point in the spectrum with the median value of neighboring data points within a specified window. This approach is particularly effective at removing spike noise, which can arise from various sources such as instrumental errors or minor sample inconsistencies.

The moderate improvement seen with MF suggests that while random noise was not a major factor affecting the model's accuracy, some level of noise reduction was still beneficial. The similarity in performance to the raw spectra indicates that the primary sources of error in the model were not solely due to random noise but might be attributed to other factors such as baseline shifts, scattering effects, or sample-to-sample variations.

Despite this, the slight improvement achieved with MF highlights the importance of noise reduction in refining the prediction model. By smoothing out the spectra, MF helps to stabilize the data, making it more consistent and reliable for analysis. This is reflected in the RMSEC value of 0.20277, which, although not significantly lower than the raw spectra, still indicates a reduction in the error of calibration.

The use of MF in this context underscores its practical significance in NIRS applications where noise reduction is necessary but not the primary concern. In the coffee industry, where consistency and reliability are key, MF can be a useful tool for ensuring that spectral data is of high quality. This method is particularly advantageous because it is relatively simple to implement and does not require complex computational resources, making it accessible for routine quality control processes.

While MF showed moderate improvement, it is important to compare its performance with other pretreatment methods.

DT and SNV pretreatments, for example, achieved higher  $R^2$  values and RPD scores, indicating that they might be more effective in addressing the specific types of spectral variations present in this dataset. However, MF remains a valuable tool for reducing noise and preserving spectral features, especially in scenarios where other pretreatment methods may not be as effective [36].

The MF pretreatment method demonstrated its utility in improving the accuracy of caffeine content predictions in Arabica coffee green beans by reducing noise in the spectra. While it may not have been the most effective method in this study, its ability to preserve sharp spectral features and reduce spike noise makes it a valuable addition to the array of pretreatment techniques available for NIRS analysis. This method, combined with other pretreatment approaches, enhances the overall robustness and applicability of NIRS technology in the coffee industry, facilitating more accurate and efficient quality control processes.

### 3.8 Caffeine prediction after Savitzky-Golay filter spectra correction

The SG filter pretreatment method was employed to evaluate its effectiveness in improving the prediction accuracy of caffeine content in Arabica coffee green beans using NIRS and PLSR. The results obtained from SG filter pretreatment showed  $r = 0.95$  and  $R^2 = 0.91$ , with an RPD value of 3.46, as shown in the scatter plot in Figure 9. These metrics indicate that the SG filter performed similarly to the MF method, suggesting that it is capable of providing fairly accurate predictions, although not significantly outperforming other pretreatment methods.

SG filter is a sophisticated smoothing technique that aims to reduce noise in the spectral data while preserving the higher moments of the spectral peaks. This method is particularly useful for maintaining the shape and width of spectral features, which is crucial for accurate chemical analysis. By fitting a polynomial to a set of neighboring data points, the SG filter effectively smooths out high-frequency noise without distorting the underlying spectral information. This approach ensures that the intrinsic spectral features, including those related to caffeine content, remain intact and are not compromised by the smoothing process.

The results suggest that while SG filters were effective in reducing noise, they did not significantly improve the model's performance compared to other methods, such as DT or SNV. This could indicate that the spectral features related to caffeine content in the coffee beans are relatively broad and not significantly affected by high-frequency noise. Broad spectral features are less susceptible to the smoothing effects of the SG filter, as they are not as easily obscured by minor fluctuations in the data.

The performance of the SG filter, while comparable to MF, highlights the importance of selecting the most appropriate pretreatment method based on the specific characteristics of the dataset. DT emerged as the most effective method in this study, likely due to its ability to remove baseline variations that were masking the caffeine-related spectral information. The significant improvements seen with DT underscore the importance of addressing baseline shifts and other forms of spectral distortion in NIRS analysis.

The use of the SG filter in this context underscores its practical significance in NIRS applications where preserving spectral features is critical. In the coffee industry, where

accurate and rapid analysis of chemical content is essential, the SG filter can be a valuable tool for ensuring that spectral data is of high quality. Although it may not be the most effective method for every dataset, the SG filter remains a useful technique for reducing noise while maintaining the integrity of the spectral signal.

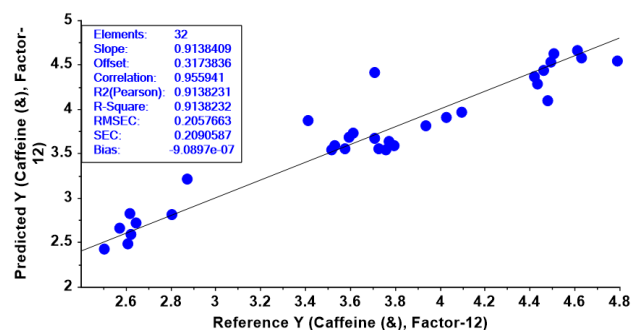


Figure 9. Caffeine prediction after Savitzky-Golay (SG) filter spectral data

In general, the various pretreatment methods demonstrated different levels of effectiveness in improving the PLSR model for caffeine prediction in Arabica coffee green beans. While the SG filter performed well in reducing noise and preserving spectral features, it did not significantly outperform other methods like DT or SNV [37-39]. The significant improvements seen with pretreatment underscore the importance of spectral correction in NIRS analysis of complex organic samples like coffee beans. By carefully selecting and applying the appropriate pretreatment method, researchers and industry professionals can enhance the accuracy and reliability of NIRS-based predictions, ultimately contributing to more efficient and effective quality control processes in the coffee industry. Performance comparison of various spectra correction approaches is presented in Table 1.

Table 1. Prediction performances of various spectra corrections using Partial Least Squares Regression (PLSR)

Model	Latent Variable	$r$	$R^2$	RMSEC	RPD
Non pretreatment	12	0.95	0.91	0.21	3.51
SNV	12	0.96	0.92	0.19	3.76
DT	12	0.99	0.99	0.06	11.74
EMSC	12	0.91	0.91	0.20	3.51
PN	12	0.96	0.91	0.21	3.51
MF	12	0.95	0.91	0.20	3.51
SG filter	12	0.95	0.91	0.20	3.46

Note:  $r$  = correlation coefficient,  $R^2$  = coefficient of determination, RMSECV = Root Mean Square Error of Cross-Validation, RPD = Residual Predictive Deviation, SNV = Standard Normal Variate, DT = Detrending, EMSC = Extended Multiplicative Scatter Correction, PN = Peak Normalization, MF = Median Filter, SG = Savitzky-Golay

### 3.9 Statistical significance of prediction models

The reported coefficients  $r$  and  $R^2$  provide valuable initial insights into model fit and explanatory power; however, these descriptive metrics alone do not establish the statistical reliability or significance of the developed calibration models. To rigorously validate that the observed relationships between the NIR spectral data and the reference caffeine measurements are not due to random chance, comprehensive statistical significance testing was performed on each PLSR model

developed with the various pretreatment methods.

The statistical evaluation was conducted at two primary levels. First, an overall F-test was performed for each PLSR model to test the global null hypothesis that none of the latent variables have a significant linear relationship with the caffeine content, in other words, that all regression coefficients in the model are effectively zero. For every pretreatment method evaluated, the calculated F-statistic vastly exceeded the critical F-value at a stringent significance level of  $\alpha = 0.01$ . The associated p-values were all less than 0.001. This decisive result allows for the rejection of the null hypothesis for every model, providing strong statistical evidence that each PLSR calibration explains a significant proportion of the variance in caffeine content. The probability that the high correlations  $r$  observed for models like DT ( $r = 0.99$ ) and SNV ( $r = 0.96$ ) arose by random coincidence is thus exceedingly low ( $p < 0.001$ ).

Second, to ensure the models were robust and not overfit to noise in the calibration set, the significance of individual regression coefficients for the retained latent variables was examined. Analysis of the confidence intervals for these coefficients revealed that the key latent variables contributing to the prediction of caffeine were statistically significant ( $p < 0.01$ ). This granular analysis confirms that the predictive structure captured by the models is based on meaningful spectral-caffeine relationships rather than spurious correlations with random variance. The exceptional results for the DT pretreatment are particularly noteworthy in this context. Its extraordinarily high F-statistic and the statistical significance of its model coefficients provide definitive mathematical support for its outstanding practical performance metrics ( $R^2 = 0.99$ , RPD = 11.74). This confluence of evidence confirms that the DT-PLSR model possesses not only exceptional predictive accuracy but also foundational statistical validity.

These formal significance tests affirm that the performance metrics reported throughout this study are grounded in statistically significant relationships. The findings confirm that NIRS-PLSR models, particularly when optimized with the DT pretreatment, produce reliable and significant calibrations for the non-destructive prediction of caffeine in intact Arabica coffee beans. This statistical rigor strengthens the conclusion that the method is a robust analytical tool suitable for further development and practical application.

### 3.10 Limitations and future work

This study, while demonstrating the significant potential of NIRS for caffeine prediction, is subject to limitations that must be transparently addressed to contextualize its findings and guide subsequent research. The most prominent limitation is the constrained sample size ( $n = 32$ ) and its geographic scope. Although the beans were sourced from several distinct Indonesian regions (Gayo, West Java, South Sulawesi, Bali), providing valuable initial diversity, this set cannot encompass the full spectrum of genetic variability, agro-climatic conditions, soil compositions, post-harvest processing methods, and storage histories found across global Arabica coffee production.

Consequently, the high-performance calibration models developed here, particularly the exceptional model using DT pretreatment, are primarily validated for and most reliably applicable to bean populations similar in origin and character to those used in this calibration. Their predictive accuracy for

beans from entirely different continents, novel cultivars, or unconventional processing methods remains untested and cannot be assumed.

To maximize model robustness within the available dataset, a rigorous internal validation protocol was implemented. The optimal complexity for each PLSR model was determined through an iterative process of leave-one-out cross-validation (LOOCV), which systematically holds out each individual sample for validation while calibrating the model on all remaining samples. This process minimizes the risk of overfitting and is reflected in the reported metrics, such as the RMSEC. However, it is critically important to distinguish internal cross-validation from external validation. LOOCV assesses performance on variations of the same dataset, whereas a true external validation tests the model on a completely independent set of samples, collected and measured separately.

The limited total number of samples in this study made the partitioning of a dedicated, statistically powerful external validation set impractical, as it would have severely weakened the stability and information content of the primary calibration model. Therefore, the outstanding metrics ( $R^2 = 0.99$ , RPD = 11.74 for DT) indicate a powerful explanatory model for this specific dataset, but its predictive power for new, unrelated populations requires further confirmation.

Given these considerations, the present work is most accurately positioned as a highly promising and methodologically rigorous proof-of-concept. It definitively establishes that a strong, statistically significant spectral-caffeine relationship exists and can be optimally accessed through DT pretreatment. To transition this finding into a universally robust, deployable analytical tool, focused future work is essential. The foremost requirement is the development of a vastly expanded and more diverse calibration library. This library should encompass hundreds of samples, representing major and minor Arabica growing regions worldwide, multiple harvest years to account for seasonal variation, different quality grades, and a range of processing methods. Following the creation of this comprehensive library, a rigorous validation protocol using a large, fully independent external test set is mandatory to establish true predictive accuracy and generalizability.

Further investigative avenues include exploring advanced machine learning algorithms like support vector regression, artificial neural networks, or ensemble methods that may capture non-linear relationships in the spectral data beyond the linear framework of PLSR. Research into model transferability, the ability of a calibration developed on one spectrometer to perform accurately on another instrument of the same or different model, is also crucial for practical industry adoption. Finally, integrating this NIRS method with rapid, online sorting systems represents a key engineering challenge for future implementation. Despite the need for these advancements, the core contribution of this study remains intact: it provides a statistically solid and exceptionally accurate foundation, identifying both the feasibility of the approach and the optimal pretreatment strategy, upon which a reliable, non-destructive quality control system for the global coffee industry can be built.

## 4. CONCLUSION

This study demonstrates that NIRS combined with PLSR

can accurately and non-destructively predict caffeine content in intact Arabica coffee beans. The core finding is that spectral pretreatment is critical to model performance, with DT emerging as the superior method. DT achieved an exceptional RPD of 11.74 and  $R^2$  of 0.99, far exceeding the performance of other techniques. This indicates that the removal of baseline variations, often caused by instrumental drift or physical sample presentation, was the most crucial step for isolating the spectral signature of caffeine in intact beans. The SNV pretreatment also provided a substantial improvement, RPD = 3.76, confirming that correcting for light scattering caused by the bean's physical surface is important. The other evaluated methods, EMSC, PN, MF, and SG filter, yielded viable but comparatively lower-performance models.

The primary practical implication is that NIRS, specifically when coupled with a DT preprocessing step, establishes a robust, rapid, and non-destructive alternative to traditional HPLC analysis for caffeine screening. This method offers significant potential for real-time quality control at critical points in the coffee supply chain. Applications could include incoming raw material inspection to verify compliance with caffeine specifications, optimization of blending processes to achieve consistent target profiles, and final product quality assurance without sample destruction. By drastically reducing analysis time and eliminating the need for chemical reagents and complex preparation, the NIRS-DT method provides the coffee industry with a powerful tool to enhance operational efficiency and ensure product consistency. Future work integrating this approach with portable spectrometers and automated data systems could further enable decentralized, at-line quality monitoring.

## ACKNOWLEDGMENT

We sincerely acknowledge the DRTPM Ministry of Education and Culture for funding this work through the PTM Master Thesis Research scheme 2024 with contract number: 094/E5/PG.02.00.PL/2024.

## REFERENCES

- [1] Anese, M., Alongi, M., Cervantes-Flores, M., Simental-Mendía, L.E., Martínez-Aguilar, G., Valenzuela-Ramírez, A.A., Rojas-Contreras, J.A., Guerrero-Romero, F., Gamboa-Gómez, C.I. (2023). Influence of coffee roasting degree on inflammatory and oxidative stress markers in high-fructose and saturated fat-fed rats. *Food Research International*, 165: 112530. <https://doi.org/10.1016/j.foodres.2023.112530>
- [2] Umakanthan, Mathi, M. (2022). Decaffeination and improvement of taste, flavor and health safety of coffee and tea using mid-infrared wavelength rays. *Heliyon*, 8(11): e11338. <https://doi.org/10.1016/j.heliyon.2022.e11338>
- [3] Chan, M.Z.A., Liu, S.Q. (2022). Coffee brews as food matrices for delivering probiotics: Opportunities, challenges, and potential health benefits. *Trends in Food Science & Technology*, 119: 227-242. <https://doi.org/10.1016/j.tifs.2021.11.030>
- [4] Munawar, A.A., Kusumiyati, K., Andasuryani, A., Yusmanizar, Y., Adrizal, A. (2024). Near-infrared technology in agriculture: Rapid, simultaneous, and non-destructive determination of inner quality parameters on intact coffee beans. *Open Agriculture*, 9(1): 20220290. <https://doi.org/10.1515/opag-2022-0290>
- [5] Munawar, A.A., Kusumiyati, Andasuryani, Yusmanizar, Adrizal. (2024). Near infrared technology coupled with different spectra correction approaches for fast and non-destructive prediction of chlorogenic acid on intact coffee beans. *Acta Technologica Agriculturae*, 27(1): 23-29. <https://doi.org/10.2478/ata-2024-0004>
- [6] Mustaqimah, Devianti, Munawar, A.A., Sufardi, S. (2024). Capability of short Vis-NIR band tandem with machine learning to rapidly predict NPK content in tropical farmland: A case study of Aceh Province agricultural soil dry land, Indonesia. *Case Studies in Chemical and Environmental Engineering*, 9: 100711. <https://doi.org/10.1016/j.cscee.2024.100711>
- [7] Zahir, S.A.D.M., Jamlos, M.F., Omar, A.F., Jamlos, M.A., Mamat, R., Muncan, J., Tsenkova, R. (2024). Review – Plant nutritional status analysis employing the visible and near-infrared spectroscopy spectral sensor. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 304: 123273. <https://doi.org/10.1016/j.saa.2023.123273>
- [8] Zhang, Q.Y., Hu, Z.G., Xu, Z.J., Zhang, P.L., Jiang, Y.J., Fu, D.D., Chen, Y. (2024). Quantitative determination of TVB-N content for different types of refrigerated grass carp fillets using near-infrared spectroscopy combined with machine learning. *Journal of Food Composition and Analysis*, 126: 105871. <https://doi.org/10.1016/j.jfca.2023.105871>
- [9] Baqueta, M.R., Marini, F., Rocha, R.B., Valderrama, P., Pallone, J.A.L. (2023). Authentication and discrimination of new Brazilian Canephora coffees with geographical indication using a miniaturized near-infrared spectrometer. *Food Research International*, 172: 113216. <https://doi.org/10.1016/j.foodres.2023.113216>
- [10] Yu, S., Huan, K.W., Liu, X.X., Wang, L., Cao, X.W. (2023). Quantitative model of near infrared spectroscopy based on pretreatment combined with parallel convolution neural network. *Infrared Physics & Technology*, 132: 104730. <https://doi.org/10.1016/j.infrared.2023.104730>
- [11] Siripatrawan, U., Makino, Y. (2024). Hyperspectral imaging coupled with machine learning for classification of anthracnose infection on mango fruit. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 309: 123825. <https://doi.org/10.1016/j.saa.2023.123825>
- [12] Shi, S.J., Zhang, W.H., Ma, Y.Y., Cao, C.G., Zhang, G.Y., Jiang, Y. (2024). Near-infrared spectroscopy combined with effective variable selection algorithm for rapid detection of rice taste quality. *Biosystems Engineering*, 237: 214-219. <https://doi.org/10.1016/j.biosystemseng.2023.12.004>
- [13] Song, K.K., Qin, Y.H., Xu, B.Y., Zhang, N.Q., Yang, J.J. (2022). Study on outlier detection method of the near infrared spectroscopy analysis by probability metric. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 280: 121473. <https://doi.org/10.1016/j.saa.2022.121473>
- [14] Samadi, Wajizah, S., Munawar, A.A. (2020). Near infrared spectroscopy (NIRS) data analysis for a rapid and simultaneous prediction of feed nutritive parameters. *Data in Brief*, 29: 105211. <https://doi.org/10.1016/j.dib.2020.105211>

- [15] Brasil, Y.L., Cruz-Tirado, J.P., Barbin, D.F. (2022). Fast online estimation of quail eggs freshness using portable NIR spectrometer and machine learning. *Food Control*, 131: 108418. <https://doi.org/10.1016/j.foodcont.2021.108418>
- [16] Allo, M., Todoroff, P., Jameux, M., Stern, M., Paulin, L., Albrecht, A. (2020). Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. *CATENA*, 189: 104452. <https://doi.org/10.1016/j.catena.2020.104452>
- [17] Quelal-Vásconez, M.A., Lerma-García, M.J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J.M., Talens, P. (2020). Changes in methylxanthines and flavanols during cocoa powder processing and their quantification by near-infrared spectroscopy. *LWT*, 117: 108598. <https://doi.org/10.1016/j.lwt.2019.108598>
- [18] Adnan, A., von Hörsten, D., Pawelzik, E., Mörlein, D. (2017). Rapid prediction of moisture content in intact green coffee beans using near infrared spectroscopy. *Foods*, 6(5): 38. <https://doi.org/10.3390/foods6050038>
- [19] Munawar, A.A., Zulfahrizal, Mörlein, D. (2024). Prediction accuracy of near infrared spectroscopy coupled with adaptive machine learning methods for simultaneous determination of chlorogenic acid and caffeine on intact coffee beans. *Case Studies in Chemical and Environmental Engineering*, 10: 100913. <https://doi.org/10.1016/j.cscee.2024.100913>
- [20] Mutz, Y.S., do Rosario, D., Galvan, D., Schwan, R.F., Bernardes, P.C., Conte-Junior, C.A. (2023). Feasibility of NIR spectroscopy coupled with chemometrics for classification of Brazilian specialty coffee. *Food Control*, 149: 109696. <https://doi.org/10.1016/j.foodcont.2023.109696>
- [21] Ruttanadech, N., Phetpan, K., Srisang, N., Srisang, S., Chungcharoen, T., Limmun, W., Youryon, P., Kongtragoul, P. (2023). Rapid and accurate classification of *Aspergillus ochraceus* contamination in Robusta green coffee bean through near-infrared spectral analysis using machine learning. *Food Control*, 145: 109446. <https://doi.org/10.1016/j.foodcont.2022.109446>
- [22] Baqueta, M.R., Valderrama, P., Mandrone, M., Poli, F., et al. (2023). <sup>1</sup>H NMR, FAAS, portable NIR, benchtop NIR, and ATR-FTIR-MIR spectroscopies for characterizing and discriminating new Brazilian *Canephora* coffees in a multi-block analysis perspective. *Chemometrics and Intelligent Laboratory Systems*, 240: 104907. <https://doi.org/10.1016/j.chemolab.2023.104907>
- [23] Manuel, M.N.B., da Silva, A.C., Lopes, G.S., Ribeiro, L.P.D. (2022). One-class classification of special agroforestry Brazilian coffee using NIR spectrometry and chemometric tools. *Food Chemistry*, 366: 130480. <https://doi.org/10.1016/j.foodchem.2021.130480>
- [24] Chakravartula, S.S.N., Moscetti, R., Bedini, G., Nardella, M., Massantini, R. (2022). Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. *Food Control*, 135: 108816. <https://doi.org/10.1016/j.foodcont.2022.108816>
- [25] Hernández-Hernández, C., Fernández-Cabanás, V.M., Rodríguez-Gutiérrez, G., Bermúdez-Oria, A., Morales-Sillero, A. (2021). Viability of near infrared spectroscopy for a rapid analysis of the bioactive compounds in intact cocoa bean husk. *Food Control*, 120: 107526. <https://doi.org/10.1016/j.foodcont.2020.107526>
- [26] Wohlers, M., Mcglone, A., Frank, E., Holmes, G. (2023). Augmenting NIR Spectra in deep regression to improve calibration. *Chemometrics and Intelligent Laboratory Systems*, 240: 104924. <https://doi.org/10.1016/j.chemolab.2023.104924>
- [27] Lazaar, A., Mouazen, A.M., EL Hammouti, K., Fullen, M., Pradhan, B., Memon, M.S., Andich, K., Monir, A. (2020). The application of proximal visible and near-infrared spectroscopy to estimate soil organic matter on the Triffa Plain of Morocco. *International Soil and Water Conservation Research*, 8(2): 195-204. <https://doi.org/10.1016/j.iswcr.2020.04.005>
- [28] Gabriëls, S.H.E.J., Mishra, P., Mensink, M.G.J., Spoelstra, P., Woltering, E.J. (2020). Non-destructive measurement of internal browning in mangoes using visible and near-infrared spectroscopy supported by artificial neural network analysis. *Postharvest Biology and Technology*, 166: 111206. <https://doi.org/10.1016/j.postharvbio.2020.111206>
- [29] Gordon, R., Chapman, J., Power, A., Chandra, S., Roberts, J., Cozzolino, D. (2019). Mid-infrared spectroscopy coupled with chemometrics to identify spectral variability in Australian barley samples from different production regions. *Journal of Cereal Science*, 85: 41-47. <https://doi.org/10.1016/j.jcs.2018.11.004>
- [30] Hong, Y.S., Chen, S.C., Liu, Y.L., Zhang, Y., Yu, L., Chen, Y.Y., Liu, Y.F., Cheng, H., Liu, Y. (2019). Combination of fractional order derivative and memory-based learning algorithm to improve the estimation accuracy of soil organic matter by visible and near-infrared spectroscopy. *CATENA*, 174: 104-116. <https://doi.org/10.1016/j.catena.2018.10.051>
- [31] Feng, Y.Z., Yu, W., Chen, W., Peng, K.K., Jia, G.F. (2018). Invasive weed optimization for optimizing one-agar-for-all classification of bacterial colonies based on hyperspectral imaging. *Sensors and Actuators B: Chemical*, 269: 264-270. <https://doi.org/10.1016/j.snb.2018.05.008>
- [32] Luna, A.S., da Silva, A.P., da Silva, C.S., Lima, I.C.A., de Gois, J.S. (2019). Chemometric methods for classification of clonal varieties of green coffee using Raman spectroscopy and direct sample analysis. *Journal of Food Composition and Analysis*, 76: 44-50. <https://doi.org/10.1016/j.jfca.2018.12.001>
- [33] Silva, T.V., Milori, D.M.B.P., Neto, J.A.G., Ferreira, E.J., Ferreira, E.C. (2019). Prediction of black, immature and sour defective beans in coffee blends by using Laser-Induced Breakdown Spectroscopy. *Food Chemistry*, 278: 223-227. <https://doi.org/10.1016/j.foodchem.2018.11.062>
- [34] Chun, Y., Ko, Y.G., Do, T., Jung, Y., Kim, S.W., Su Choi, U. (2019). Spent coffee grounds: Massively supplied carbohydrate polymer applicable to electrorheology. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 562: 392-401. <https://doi.org/10.1016/j.colsurfa.2018.11.005>
- [35] Assis, C., Gama, E.M., Nascentes, C.C., de Oliveira, L.S., Anzanello, M.J., Sena, M.M. (2020). A data fusion model merging information from near infrared spectroscopy and X-ray fluorescence. Searching for atomic-molecular correlations to predict and characterize

- the composition of coffee blends. *Food Chemistry*, 325: 126953.  
<https://doi.org/10.1016/j.foodchem.2020.126953>
- [36] Assis, C., Pereira, H.V., Amador, V.S., Augusti, R., de Oliveira, L.S., Sena, M.M. (2019). Combining mid infrared spectroscopy and paper spray mass spectrometry in a data fusion model to predict the composition of coffee blends. *Food Chemistry*, 281: 71-77. <https://doi.org/10.1016/j.foodchem.2018.12.044>
- [37] Zhu, Q.X., Yu, X.Y., Wu, Z.B., Lu, F., Yuan, Y.F. (2018). Antipsychotic drug poisoning monitoring of clozapine in urine by using coffee ring effect based surface-enhanced Raman spectroscopy. *Analytica Chimica Acta*, 1014: 64-70. <https://doi.org/10.1016/j.aca.2018.02.027>
- [38] Correia, R.M., Tosato, F., Domingos, E., Rodrigues, R.R.T., Aquino, L.F.M., Filgueiras, P.R., Lacerda Jr., V., Romão, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*, 176: 59-68. <https://doi.org/10.1016/j.talanta.2017.08.009>
- [39] Yisak, H., Redi-Abshiro, M., Chandravanshi, B.S. (2018). Selective determination of caffeine and trigonelline in aqueous extract of green coffee beans by FT-MIR-ATR spectroscopy. *Vibrational Spectroscopy*, 97: 33-38. <https://doi.org/10.1016/j.vibspec.2018.05.003>