ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Estimation of Proximate Composition in Rice Using ATR-FTIR Spectroscopy and Chemometrics

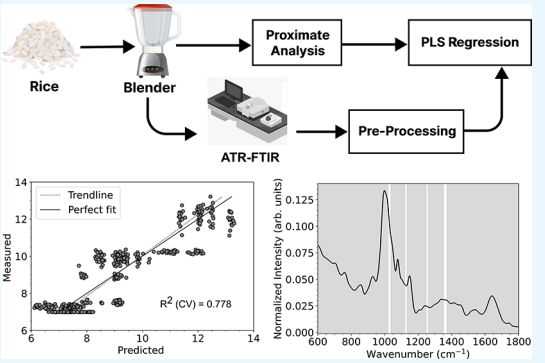
Syahril Siregar, Asep Nurhikmat,* Rima Zuriah Amdani, Retno Utami Hatmi, Mahargono Kobarsih, Annisa Kusumaningrum, Mirwan Ardiansyah Karim, Amarilia Harsanti Dameswari, Nugroho Siswanto, Siswoprayogi Siswoprayogi, and Ponco Yuliyanto

Cite This: *ACS Omega* 2024, 9, 32760−32768

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: This study presents an innovative approach for estimating the proximate composition of diverse rice varieties using attenuated total reflectance fourier transform infrared (ATR-FTIR) spectroscopy and chemometric techniques. Principal component analysis (PCA) reveals distinct separations among the seven rice varieties based on their FTIR spectra. Robust partial least squares (PLS) regression models, developed with high calibration (R^2) values from 0.778 for protein up to 0.941 for moisture, demonstrate high accuracy in predicting proximate composition. The root mean squared error (RMSE) in percentage values, indicative of prediction accuracy, were low across all proximate components. To ensure the response variable of regression, proximate composition measurements were taken five times, while FTIR spectra were scanned tens of times, employing random numbers around the average with the same standard deviation as the measurement. Notably, the study emphasizes the pivotal role of the amide-III band in protein determination, alongside specific wavenumber regions associated with molecular changes in proximate components. This research underscores the potential of ATR-FTIR spectroscopy and chemometrics for rapid and accurate proximate assessment in food science and agriculture.



INTRODUCTION

Rice serves as a foundational dietary staple for more than half of the global population, particularly prevalent in regions such as Asia, the Caribbean, and South America.[1−3] The nutritional profile of rice, determined by its macronutrients, comprising carbohydrates, proteins, and fats, plays a crucial role in sustaining essential bodily functions, necessitating substantial daily intake.[4] Understanding the macronutrient content in rice is vital, offering essential insights to guide dietary practices, especially among populations heavily reliant on rice as their primary food source.[5] Moreover, the analysis of macronutrient levels lays the foundation for a deeper exploration of proximate composition, a crucial aspect of nutritional assessment encompassing key components such protein, fat, carbohydrate, moisture, and ash. Additionally, analyzing macronutrient levels helps identify rice varieties with enhanced nutritional value, which is essential for regions grappling with food security challenges.[6]

Recent studies have shed light on potential health risks tied to rice consumption, especially its link to type 2 diabetes. Some research has highlighted a significant link between consuming substantial amounts of white rice and elevated blood sugar levels, increasing the vulnerability to type 2 diabetes in certain individuals.[7−9] These findings emphasize the need to look beyond rice as just a staple food and to explore its specific nutritional components. By accurately determining the amounts of macronutrients in rice, we can better understand how different types might affect metabolic health. This detailed understanding paves the way for more specific interventions and dietary advice, particularly for those at risk of metabolic conditions. Therefore, it is crucial to adopt advanced methods to precisely assess the nutritional content of rice, shaping more informed public health and dietary recommendation.

Macronutrients interact uniquely with infrared radiation, rendering them infrared-active molecules detectable by Fourier Transform Infrared (FTIR) spectroscopy. This technique extends beyond merely identifying functional groups; when calibrated against quantitative chemical analyses and employing multivariate statistical methods, FTIR spectroscopy emerges as a versatile tool essential for quantitative analysis

Received: March 23, 2024
Revised: June 26, 2024
Accepted: June 28, 2024
Published: July 16, 2024

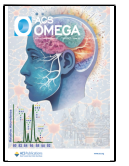


Table 1. Proximate Composition in Diverse Rice Varieties[a]

ID	Total FTIR spectra	proximate analysis results (%)[b]				
		Protein	Lipid	Moisture	Ash	Carbohydrate[c]
BC	120	9.899 ± 0.194	0.596 ± 0.077	12.589 ± 0.068	0.505 ± 0.050	76.416 ± 0.225
BKH	75	12.197 ± 0.404	3.692 ± 0.422	11.560 ± 0.532	0.944 ± 0.029	71.607 ± 0.790
BKP	45	7.509 ± 0.099	1.001 ± 0.147	11.811 ± 0.196	0.258 ± 0.006	79.420 ± 0.264
BM	45	10.245 ± 0.065	2.051 ± 0.183	12.504 ± 0.016	1.034 ± 0.041	74.165 ± 0.199
BP	120	6.992 ± 0.005	0.250 ± 0.096	11.984 ± 0.037	0.369 ± 0.018	80.405 ± 0.104
BS	45	8.911 ± 0.081	0.592 ± 0.141	11.447 ± 0.185	0.461 ± 0.023	78.859 ± 0.247
BW	120	7.273 ± 0.061	0.296 ± 0.067	14.248 ± 0.093	0.371 ± 0.080	77.812 ± 0.152

[a]Standard deviation (SD) is given after ±. [b]Based on the means of five times measurements. [c]SD of carbohydrate based on the error propagation from protein, lipid, moisture, and ash.

across diverse domains, including food science and chemistry.[10−12]

Recently, the combination of FTIR spectroscopy with chemometrics has revolutionized food analysis by facilitating rapid, nondestructive evaluations of chemical composition and structure. Notably, FTIR spectroscopy has accurately quantified sugars in fruits and mango juice, shedding light on mango ripening dynamics.[13−15] Moreover, integrating FTIR spectroscopy with machine learning algorithms has shown potential in identifying the geographical origin of rice samples, as demonstrated in studies focused on Thailand.[16]

Within the domain of chemometrics, partial least square (PLS) regression stands out for its efficacy in spectroscopy analysis across various food science applications, ranging from identifying aged rice to detecting adulteration.[17−19] Moreover, its ability to filter out irrelevant features enhances regression performance.[20,21]

Despite the widespread use of FTIR in rice research, there's a pronounced research gap concerning its application alongside chemometrics for analyzing proximate composition. Conventional methods, such as proximate analysis, although accurate, demand extensive time and labor, making them less feasible for large-scale or rapid analyses. Such limitations underscore the urgency for innovative, efficient techniques that can offer both accuracy and speed.

Addressing this, our study aims to devise a rapid and reliable method for estimating rice proximate contents using attenuated total reflection-fourier transform infrared (ATR-FTIR) spectroscopy integrated with PLS regression. Additionally, we extract important bands corresponding to proximate content. In the future, we will focus solely on these crucial bands for the rapid identification and quantification of macronutrients. We also employed principal component analysis (PCA) to observe the chemical composition difference among the rice samples. This innovative approach for quantifying proximate composition of rice using ATR-FTIR not only holds promise for versatile applications in agriculture, dietary guidance, and the advancement of nutritional research but also contributes to enhancing our understanding of FTIR spectra in the context of macronutrients.

MATERIALS AND METHODS

Sample Preparation. We gathered seven varieties of rice samples from various regions within the Yogyakarta province of Indonesia. These varieties include Beras Curah (BC) or bulk rice, Beras Ketan Hitam (BKH) or black sticky rice, Beras Ketan Putih (BKP) or white sticky rice, Beras Merah (BM) or brown rice, Beras Premium (BP) or premium rice, Beras Basmati (BS) or basmati rice, and Beras Wangi (BW) or fragrant rice. These selected varieties were subjected to the experiment due to its commercial availability in the Indonesian market. Samples collections were taken across several regions since each area produced different best products. For example, the Sleman region which is in the north area of Yogyakarta is the best for yielding brown rice. We would like to have fresh samples after being directly milled in the station. Thus, the samples are obtained directly from the manufacturer which are packed and branded differently after the milling process. However, some types of rice which is rarely consumed by the locals, such as basmati rice and black and white sticky rice, were obtained from the rice specialty store in the city center of Yogyakarta. To prepare the samples for analysis, we subjected them to grinding and sieving processes using a test sieve with a mesh size of 100 μm.

Proximate Analysis. To assess the macronutrient composition as target variables in our regression model, we utilized the proximate analysis technique. The components analyzed included protein, lipid, moisture, ash, and carbohydrate. Each type of rice underwent five times measurements for each parameter as outlined in Table 1. Thus, a total of 35 data were obtained for all seven varieties in one analyzed parameter (e.g., protein). For the sampling technique, as many as five portions of the samples were taken from the same packaging in each type of rice to minimize the contribution of differences results in repetition. The proximate analysis involved determining moisture and ash contents using the gravimetric method (AOAC 952.08, 930.30, 2016),[22] fat content using Soxhlet method (AOAC 948.15, 2016),[22] protein content using Kjeldahl method (AOAC 992.23, 2016),[22] and carbohydrate content through the carbohydrate by difference method (AOAC, 2005).[23] Detailed methodologies for measuring each parameter are presented in the subsequent subsections.

Protein Content Measurement. We employed the Kjeldahl method to determine the protein content, involving stages of destruction, distillation, and titration.[22] Initially, 0.5 g of the sample and a 0.5 g blank test were mixed with 10 mL of sulfuric acid (H_2SO_4) and a quarter of a Kjeldahl tablet. The destruction phase occurred in a KjelDigester K-446 Buchi at 400 °C for 90 min. For this process, 3 L of 10% sodium hydroxide (NaOH) was used in a Scrubber K-415 Buchi to capture the sulfuric acid vapor. Subsequent distillation and titration were conducted in an automatic KjelMaster K-375 Buchi, utilizing 4% boric acid (H_3BO_4), 32% NaOH, distilled water, and 0.2 N hydrochloric acid (HCl). The Kjeldahl method determines the nitrogen content in the sample, which is then converted to protein content using a general nitrogen factor (NF) of 6.25, derived from the assumption that proteins

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768

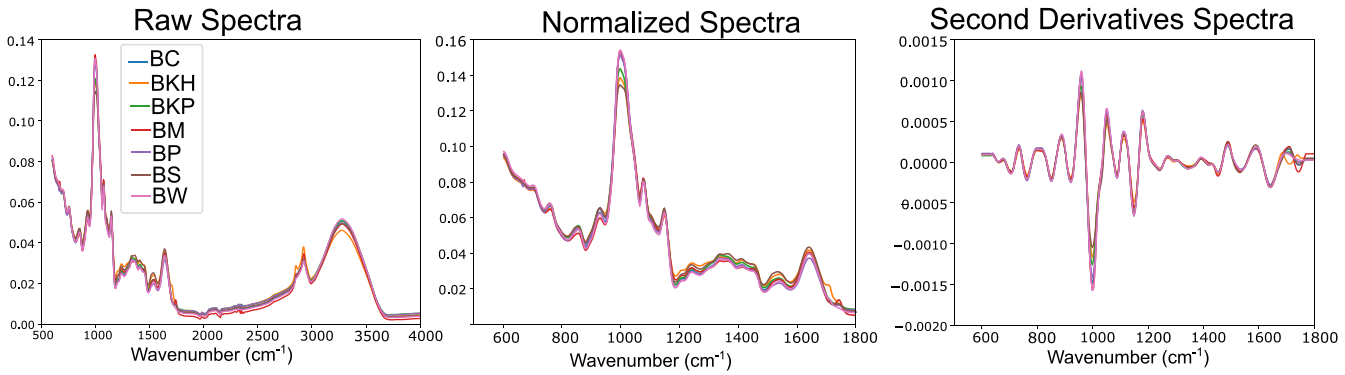


Figure 1. Mean spectra for each rice variety. Left: raw data; middle: normalized spectra; and right: second derivative spectra.

contain 16% (w/v) nitrogen. The derived protein content is termed "crude protein" due to the presence of various nitrogen-containing derivatives. The protein content calculation is represented by

$$\%\mathrm{N} = \frac{(v_{\mathrm{blank}} - v_{\mathrm{sample}}) \times N_{\mathrm{HCl}} \times \mathrm{ArN}}{w_{\mathrm{sample}} \times 1000} \times 100\% \quad (1)$$

$$\text{protein content } (\%) = \%N \times \mathrm{NF} \quad (2)$$

where v_{blank} represents the blank volume post-titration, v_{sample} is the sample volume post-titration, N_{HCl} indicates the molarity of HCl, w_{sample} is the sample weight prechemical treatment (in grams), and ArN is the atomic mass of nitrogen.

Lipid Content Measurement. The fat content was determined using the Soxhlet method.[22] Specifically, 2 g samples were subjected to extraction using a fat extractor Buchi for 20 cycles with petroleum benzene as the solvent. The weights of the samples pre- and postextraction were determined after drying them in an oven at 105 °C overnight. Since the extracted fat content includes not only pure oil but also other compounds like organic acids, alcohols, essential oils, pigments, and fat-soluble vitamins in trace amounts, it is referred to as "crude fat". The fat content lost from the samples is calculated as

$$\text{crude fat content } (\%) = \frac{(w_1 - w_2)}{w_3} \times 100\% \quad (3)$$

where w_1, w_2, and w_3, respectively, denote the weight of the sample pre-extraction and postdrying, the weight postextraction and postdrying, and the weight of the sample pre-extraction without drying.

Moisture Content Measurement. Moisture content was determined gravimetrically.[22,24] Samples weighing 2 g were dried in an oven set at 105 °C for 24 h, and their weights were monitored until a constant weight was attained. Following the drying phase, samples were allowed to equilibrate in a desiccator for 15 min before measurement. The moisture content was calculated using the formula:

$$\text{moisture content } (\%) = \frac{(w_{\mathrm{pre}} - w_{\mathrm{post}})}{w_{\mathrm{pre}}} \times 100\% \quad (4)$$

where w_{pre} is the weight of the sample before drying, and w_{post} is the weight of the sample after drying.

Ash Content Measurement. Ash content represents the inorganic minerals present in a sample and includes elements such as calcium, magnesium, potassium, phosphorus, sulfur, and trace minerals. It is an essential parameter in food and agricultural analysis as it provides insights into the mineral composition of the material. Ash content determination followed a gravimetric approach similar to that for moisture content.[22] Samples (2 g each) were ashed in a cabolite furnace at 550 °C for 5−6 h. Subsequently, they were dried in an oven at 105 °C for an additional 24 h until reaching a constant weight. After removal from the oven, a 15 min desiccator equilibration preceded the weighing process. The ash content is determined by the formula:

$$\text{ash content } (\%) = \frac{w_{\mathrm{a}}}{w_{\mathrm{b}}} \times 100\% \quad (5)$$

where w_{a} represents the weight of the sample after ashing and drying, and w_{b} is the weight of the sample before ashing and drying.

Carbohydrate Content Measurement. Total carbohydrate contents were estimated by difference.[23] In this method, carbohydrate content was determined indirectly by subtracting the values of other components, namely crude protein, crude fat, moisture content, and ash content, from 100%. Since carbohydrate levels are derived indirectly from these measurements, there is associated error (standard deviation) propagated through this approach. The standard deviation of carbohydrate was calculated using principles of error propagation.[25]

FTIR Spectra Preprocessing. The ATR FTIR spectrophotometer Bruker vertex 60 system was utilized to capture the spectral data. Each data acquisition was performed in reflectance mode, encompassing 32 scans per sample. The wavenumber resolution was set at 4 cm^{-1}, covering a spectral range from 600 to 4000 cm^{-1} with 882 points along the wavenumber.

For every rice variety, we conducted multiple random acquisitions, resulting in distinct spectra. Overall, we examined a total of 570 spectra. Specific details on the spectrum count for each sample are provided in Table 1.

Prior to any analysis, the raw spectra underwent a series of preprocessing steps. Initially, the spectra were truncated to focus solely on the fingerprint region, specifically ranging from 600 to 1800 cm^{-1}. Following this, vector normalization was applied to standardize the data. Further refinement involved calculating second derivatives spectra using Savitzky-Golay (SG) algorithm with specific parameters: a window length of 17 and a polynomial order of 2. This procedure yielded the second derivative spectra, which served as the basis for subsequent PCA and PLS regression analyses. Ultimately, the

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768

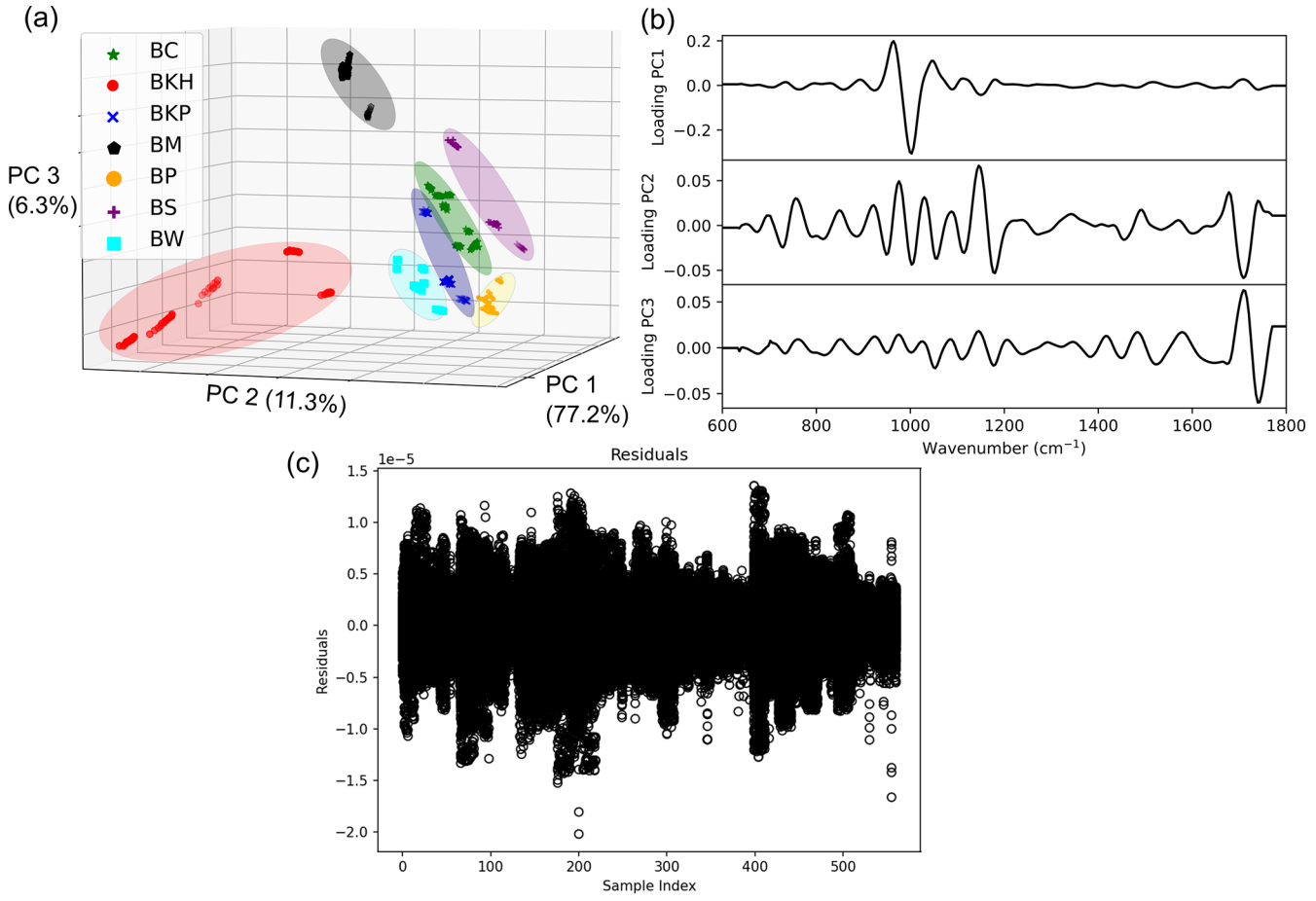


Figure 2. (a) 3 Dimensional PC scores, (b) PC loadings, and (c) PCA residual graph.

number of points in the wavenumber was reduced to 312. Please refer to Figure 1 to observe the impact of preprocessing on the spectra.

Principal Component Analysis. To discern the separation among the samples, PCA was employed, and the resultant separation will be elucidated through the PC scores plot. In mathematical terms, PCA is a matrix decomposition defined as follows:

$$\mathbf{A}(n, p) = \mathbf{U}(n, k)\mathbf{V}^{\mathrm{T}}_{(k,p)} + \mathbf{E} \tag{6}$$

Here, **A** represents the matrix slated for reduction consisting of all spectra, **U** denotes scores, **V** is designated as loadings, and **E** signifies the error. k is number of the principal components and can be estimated by evaluating the cumulative explained variance. Scores can be interpreted as new coordinates resulting from PCA reduction, while loadings can be understood as the most significant magnitudes or patterns that delineate the data.

Chemometric Analyses. Chemometric analyses were conducted to quantify the FTIR spectra using custom-built software. Given our goal of quantifying the FTIR spectra, we employed multivariate regression, specifically using PLS regression. PLS regression was chosen for its straightforwardness, rapidity, comparative effectiveness, user-friendliness, and widespread use in spectroscopy.[26−28] The construction of our PLS models adhered to guidelines from Daniel Pelliccia's tutorial on variable selection for PLS regression.[29] Additional details on specific algorithms referenced in this study can be found in papers by Mehmood et al.[20] and Cai et al.[30]

All computations were performed on a PC equipped with Ubuntu Linux, an Intel Core i7−12700K processor, an NVIDIA GeForce RTX 4060 GPU with 8 GB, and 64 GB of RAM. The software was developed using Python 3.10.

The number of collected FTIR spectra ranged from 45 to 120 samples for each rice variety, as detailed in Table 1. However, in the proximate analysis, each rice variety underwent five measurements. To align the number of targets with the number of spectra, Gaussian random numbers were employed. Each value's mean was associated with its respective target, and its standard deviation mirrored the measurement's precision.

The optimum number of components in PLS regression was determined by looping through numbers from 2 to 35. Simultaneously, a wavenumber-by-wavenumber loop was performed to exclude unimportant wavenumbers in the PLS regression. Finally, we employed another PLS regression with the optimum number of components and only important wavenumber data.

To assess the reliability of our model, a leave-one-group-out cross-validation technique was utilized, evaluating its performance based on RMSE and coefficient of determination (R^2). In this cross-validation technique, data set is divided into groups, and the validation process involves leaving out an entire group of data points for testing while training the model on the remaining groups.

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768

Table 2. PLS Regression Performance Metrics

	R^2		RMSE (%)		Optimum	Discharged
	Calibration	Cross validation	Calibration	Cross validation	Number of components	Wavenumber
Protein	0.881	0.778	7.23	9.92	5	305
Lipid	0.975	0.931	37.02	50.75	23	240
Carbohydrate	0.941	0.848	0.91	1.46	11	298
Moisture	0.963	0.941	1.61	2.08	29	247
Ash	0.968	0.939	14.23	17.47	24	209

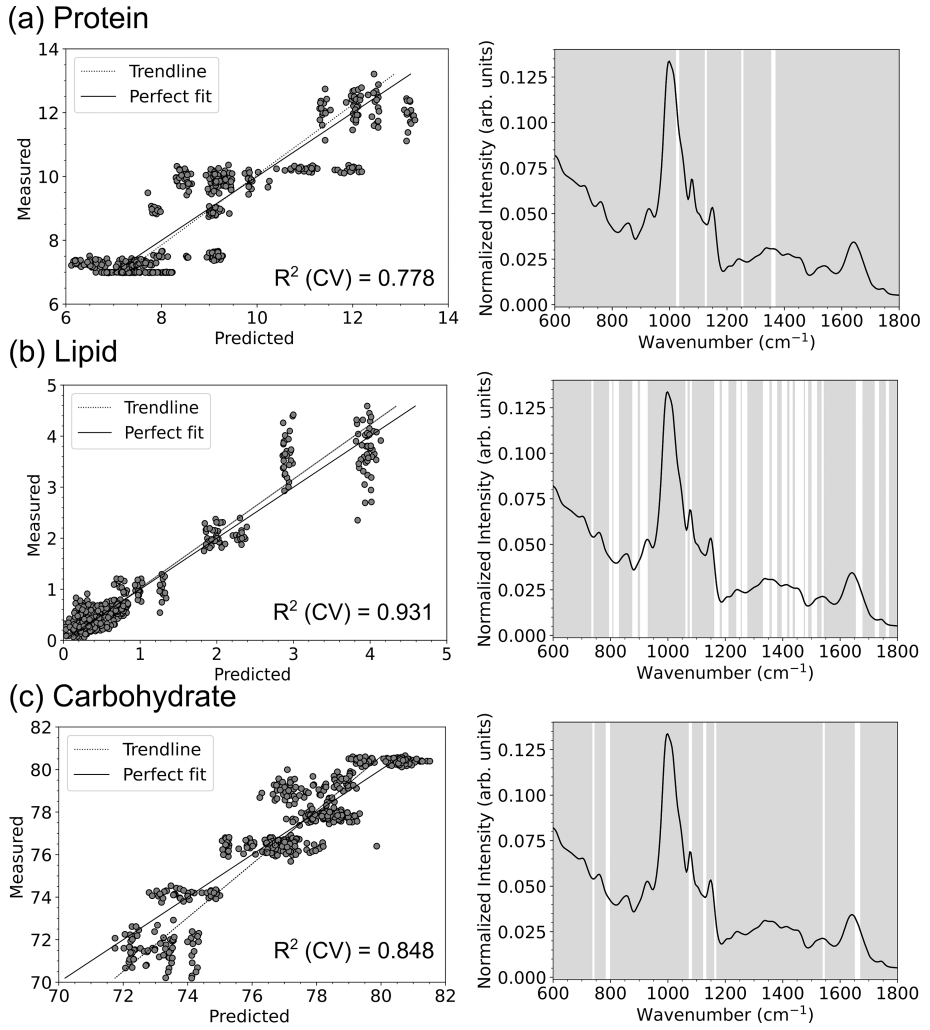


Figure 3. Regression plots on the left depict the predictions of the PLS model for the respective targets, whereas the FTIR spectra of rice on the right showcase key features. In these plots, intensity is represented in absorbance. The white lines mark the highlighted features, and the brown shaded areas indicate regions with diminished wavenumbers.

RESULTS AND DISCUSSION

PCA Separation. In our analysis, we utilized 5 principal components, which, according to the cumulative explained variance plot, already represent close to 98.5%. Therefore, further analysis focused on the first three PC scores (PC1, PC2, and PC3) due to their significant information content regarding the spectra of the rice samples. Figure 2a displays a 3D plot of PC scores, revealing a clear separation for each rice variety. The distinct clusters formed by the red, blue, light blue, and purple samples arise from the random selection of rice samples for FTIR measurements, reflecting the inherent variability in the samples and manifested in their clustering patterns. Despite these separate clusters, it is important to note that they still exist within the same spatial plane, underscoring the complexity of the sample distribution.

The loadings plot provides insight into the wavenumbers responsible for the observed separation. Figure 2b illustrates the loadings plot for PC1, revealing a pronounced magnitude in the region of 900−1100 cm^{-1}. This region exhibits notably stronger intensity compared to others across all PC components, indicating significant differences among rice varieties in this specific wavenumber range. These wavenumbers are associated with protein, total carbohydrate, and their overlap.[31,32] Additionally, the loadings for PC2 and PC3 display peaks around 1680−1770 cm^{-1}, this region belonging to the C═O band can be lipid and carbohydrate.[33] The PC loading also shows that some low to medium intensity peaks lie

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768

in the region of 600−1800 cm^{-1}, and this region is the figerprint region of the macronutrient.

The observed PCA residuals plot indicates that the majority of residuals fluctuate around small values close to zero, suggesting a consistent pattern within the data set as shown in Figure 2c. This pattern aligns with the concept of homoscedasticity, indicating that the variability in the residuals remains relatively constant across different values of the independent variables. While a small subset of samples, specifically only two out of 540, display slightly higher residual amplitudes, these deviations are minimal in comparison to the overall data set. Given the predominance of residuals clustering near zero and the limited occurrence of outliers, the PCA appears to be robust and effective in capturing the underlying structure of the data.

Chemometric Results. The metrics R^2 and RMSE calibration evaluate the model's performance during training. In contrast, cross-validation assesses how well the model generalizes to new data, specifically using the leave-one-group-out cross-validation technique.

From Table 2, our calibration R^2 values are commendable: 0.881 for protein and over 0.94 for other proximate composition, indicating effective model training. The R^2 values cross-validation range is between 0.778 and 0.941, affirming the model's robustness in predicting the desired targets. This underscores the efficacy of the PLS regression approach we adopted.

The RMSE values for protein, carbohydrate, moisture, and ash were relatively low, indicating the accuracy of the model in predicting these components. However, for lipids, the RMSE was quite large, at 37.02% for calibration and 50.75% for cross-validation. This noticeable discrepancy is attributed to the low lipid concentration in rice samples, which was even less than 1% for BC, BP, BS, and BW, as shown in Table 1. Due to this low concentration, it is challenging to estimate lipid content with high accuracy.

Despite the subpar performance of our model in predicting lipid concentration, the typically low lipid content in rice mitigates its impact on indirect estimations reliant on lipid concentration, such as carbohydrate concentration.

For determining the optimal number of PLS components, we iterated from 5 to 35 components. Correspondingly, setting 35 as the upper limit proved adequate. Our analysis, detailed in Table 2, revealed the optimal component range to be between 5 and 29.

Despite focusing on the fingerprint region in PLS regression, not every wavenumber significantly contributes to the regression models. Our calculations suggest that for each target, a subset of wavenumbers had a small contribution to regression. Consequently, 209 to 305 points of wavenumber need to be excluded from calculation as outlined in Table 2 as discharged wavenumber.

Protein. Our regression model for protein achieved R^2 values of 0.881 for calibration and 0.778 for cross-validation, which were the lowest among the various targets evaluated. However, the RMSE scores were low 7.23% for calibration and 9.92% for cross-validation. Despite this, the performance was deemed satisfactory, as illustrated in Figure 3a.

In the analysis, 305 wavenumbers were excluded due to their negligible impact on the protein regression. Consequently, only seven wavenumbers significantly contributed to the calculation. Specifically, these wavenumbers are associated with four distinct regions related to protein changes in rice samples: 1020, 1130, 1255, and 1360 cm^{-1}. Notably, the bands observed at 1250 and 1360 cm^{-1} correspond to the amide-III protein band, as referenced in previous studies,[34,35] while 1020 and 1030 cm^{-1} belongs to the C−N stretching mode.[36] Intriguingly, the amide-I (1650 cm^{-1}) and amide-II (1500 cm^{-1}) bands did not exhibit significant contributions to this regression analysis. This observation suggests that the amide-III band predominantly influences the determination of protein content in rice samples. While historically the amide-III region has not been utilized as extensively for determining protein secondary structure when compared to the amide-I region,[37−39] our findings are in line with recent studies indicating the accurate assignment of protein secondary structure changes in silk during deformation or processing using the amide III.[39] The absence of amide-I and amide-II in this regression model may also be caused by overlap with lipid and carbohydrate bands. However, further analysis is also required to confirm that amide-III is sensitive to the amount of protein content in rice.

Lipid. The regression performance of lipids outperformed other targets in this study, with R^2 values of 0.975 for calibration and 0.931 for cross-validation, as indicated in Table 2. The data closely followed the trendline, as illustrated in Figure 3b. However, the RMSE for carbohydrate estimation was also high. These results suggest that our model explains a large proportion of the variability in the dependent variable but may not provide precise predictions for individual data points.

In this analysis, 240 wavenumbers were excluded due to their minimal contribution to the regression. Consequently, 74 wavenumbers played a significant role in the regression. The bands associated with lipids include the $-PO_2$ bond in phospholipids, the vibration of $-CH_3$, the deformation of $-CH_2$, and the C═O bond. For a more detailed list of bands, please refer to Table 3.

Carbohydrate. Our regression model for carbohydrate achieved R^2 values of 0.941 for calibration and 0.848 for cross-validation. These values indicate good regression performance, as illustrated in Figure 3c. Additionally, the RMSE for

Table 3. Preliminary Assignment of FTIR Absorption Bands Related to Macronutrients

Targets	Wavenumber (cm^{-1})	Assignment	Ref
Protein	1020	Primary amine, C−N stretching	36
	1130	Secondary amine, C−N stretching	36
	1255, 1360	Amide-III	3435
Lipid	870−890		
	910−940		
	1120−1140		
	1230−1280	$-PO_2$ bond in phospholipid	40
	1340−1370	CH_3 bending	31
	1390−1410	Fatty acids, CH_3 bending	31
	1500−1530	Deformation of $-CH_2$ and $-CH_3$	40
	1660−1680	C═O	33
Carbohydrate	740		
	780		
	1070, 1130, 1160	Total carbohydrates	41
	1540		
	1660	C═O	33

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768

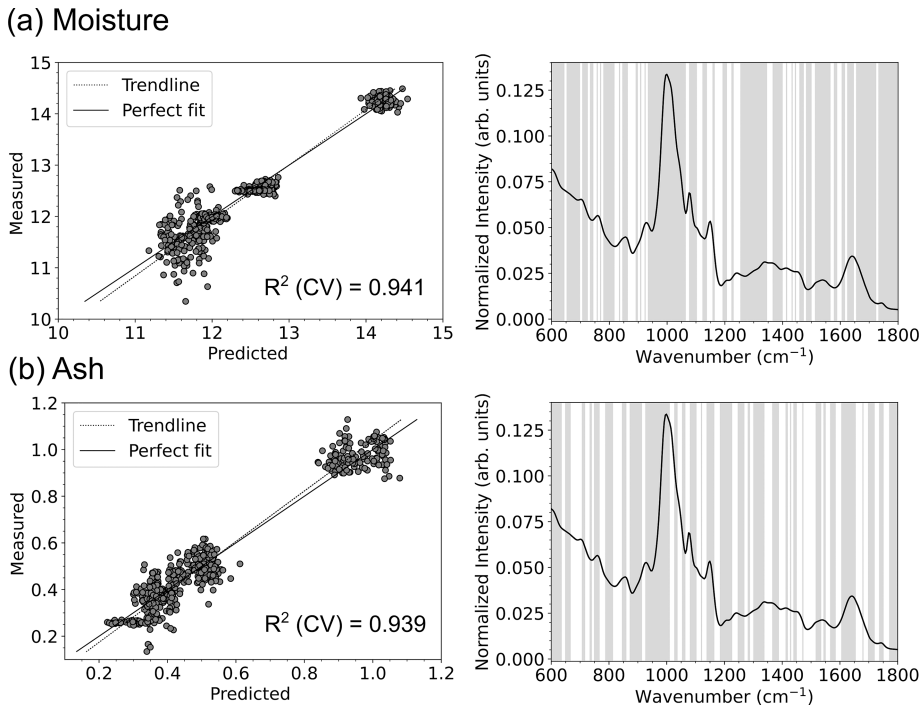


Figure 4. Left-side regression plots illustrate the PLS model's predictions for targets, while on the right-side, FTIR spectra of rice reveal key features associated with the targets. In these plots, intensity is represented in absorbance. The white lines represent highlighted features, and the brown regions indicate wavenumbers that were excluded.

carbohydrate was the lowest among other proximate compositions. Specifically, the RMSE for calibration was 0.91%, and for cross-validation, it was 1.46%. A low RMSE suggests robust model performance.

In the analysis, 298 wavenumbers were excluded due to their negligible impact on the carbohydrate regression. Consequently, only 14 wavenumbers significantly contributed to the calculation. Specifically, these wavenumbers relate to the total carbohydrate band and the C═O bond as listed in Table 3.

Moisture and Ash. The regression model demonstrates strong predictive performance for the moisture (water) and ash contents of rice, achieving R^2 (CV) values of 0.941 and 0.939, respectively, as illustrated in Figure 4. Water molecules, being infrared-active, exhibit absorption in two distinct regions, approximately 1300−2000 cm^{-1} and 3500−4000 cm^{-1}.[42] Our regression analysis pinpointed several regions associated with moisture content, including 640, 710, 860, 940, and numerous regions around 1200−1800 cm^{-1} (particularly 1630 cm^{-1}). Notably, only 255 wavenumber points were excluded in this regression. The results for 1630 cm^{-1} align well with Nesakumar et al. that assigned 1637 cm^{-1} to the moisture.[43]

In terms of ash content, our analysis revealed notable regions spanning 600−1800 cm^{-1}. This finding aligns with expectations, as ash content represents the inorganic residue and mineral composition in rice samples.

CONCLUSIONS

In conclusion, our comprehensive analysis of different rice varieties using FTIR spectroscopy and chemometric techniques has provided valuable insights into the distinctive molecular fingerprints associated with key nutritional components. The 3D PCA scores plot demonstrated a clear separation among rice varieties, indicating significant differences in their FTIR spectra. The subclusters within the same rice variety were attributed to the variability in the samples. Additionally, the residual plot suggests homoscedasticity, signifying that the variability in the residuals remains relatively constant across different values of the independent variables. The subsequent PLS regression models exhibited robust performance, achieving high calibration R^2 values and demonstrating the models' capacity to generalize effectively to new data during cross-validation. The identification of optimal PLS components and selective inclusion of critical wavenumber subsets underscored the importance of targeted spectral features in predicting nutritional components, enhancing the precision and efficiency of our models.

Notably, our findings challenged conventional norms in nutritional analysis by highlighting the pivotal role of the amide-III band in protein determination, deviating from the traditional emphasis on amide-I and amide-II bands. Furthermore, the regression analyses for lipid and carbohydrate revealed distinct wavenumber regions associated with their molecular changes in rice samples, providing specific insights into the chemical composition. This study not only contributes to the understanding of rice diversity but also showcases the potential of FTIR spectroscopy coupled with chemometric analysis as a robust tool for rapid and accurate nutritional assessment in various agricultural and food science applications. Future research endeavors can leverage this approach for a broader spectrum of food samples and nutritional components, expanding its applicability and impact in the field of food quality control and nutritional studies.

AUTHOR INFORMATION

Corresponding Author

Asep Nurhikmat − *Research Center for Food Technology and Processing-National Research and Innovation Agency of*

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768

Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia; Email: asep020@brin.go.id

Authors

Syahril Siregar − *Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia*; orcid.org/0000-0002-8041-8575

Rima Zuriah Amdani − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Retno Utami Hatmi − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Mahargono Kobarsih − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Annisa Kusumaningrum − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Mirwan Ardiansyah Karim − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Amarilia Harsanti Dameswari − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Nugroho Siswanto − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Siswoprayogi Siswoprayogi − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Ponco Yuliyanto − *Research Center for Food Technology and Processing-National Research and Innovation Agency of Indonesia, Gunungkidul, Yogyakarta 55861, Indonesia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c02816

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by Indonesia Endowment Fund for Education Agency, Ministry of Finance, Republic of Indonesia, under the scheme of Riset dan Inovasi untuk Indonesia Maju (RIIM) Gelombang I Tahun 2022, Grant KEP-5/LPDP/LPDP.4/2022. The computational work in this study was supported by the Directorate of Research and Development, Universitas Indonesia, under the Hibah Riset FMIPA UI Q3 2023 (Grant PKS-029/UN2.F3.D/PPM.00.02/2023) awarded to the first author, S.S.

■ REFERENCES

(1) Muthayya, S.; Sugimoto, J. D.; Montgomery, S.; Maberly, G. F. An overview of global rice production, supply, trade, and consumption. *Annals of the new york Academy of Sciences* **2014**, *1324*, 7−14.

(2) Fukagawa, N. K.; Ziska, L. H. Rice: Importance for global nutrition. *Journal of nutritional science and vitaminology* **2019**, *65*, S2−S3.

(3) Bandumula, N. Rice production in Asia: Key to global food security. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences* **2018**, *88*, 1323−1328.

(4) Venn, B. J. Macronutrients and human health for the 21st century. *Nutrients* **2020**, *12*, 2363.

(5) Dietary protein quality evaluation in human nutrition. *Report of an FAO Expert Consultation*; FAO Food and Nutrition Paper 92; FAO, 2011; pp 1−66.

(6) Khush, G. S. Origin, dispersal, cultivation and variation of rice. *Plant molecular biology* **1997**, *35*, 25−34.

(7) Hu, E. A.; Pan, A.; Malik, V.; Sun, Q. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *Bmj* **2012**, *344*.e1454

(8) Bhavadharini, B.; Mohan, V.; Dehghan, M.; Rangarajan, S.; Swaminathan, S.; Rosengren, A.; Wielgosz, A.; Avezum, A.; Lopez-Jaramillo, P.; Lanas, F.; Dans, A. L.; Yeates, K.; Poirier, P.; Chifamba, J.; Alhabib, K. F.; Mohammadifard, N.; Zatonska, K.; Khatib, R.; Vural Keskinler, M.; Wei, L.; Wang, C.; Liu, X.; Iqbal, R.; Yusuf, R.; Wentzel-Viljoen, E.; Yusufali, A.; Diaz, R.; Keat, N. K.; Lakshmi, P.V.M.; Ismail, N.; Gupta, R.; Palileo-Villanueva, L. M.; Sheridan, P.; Mente, A.; Yusuf, S. White rice intake and incident diabetes: a study of 132,373 participants in 21 countries. *Diabetes Care* **2020**, *43*, 2643−2650.

(9) Sun, Q.; Spiegelman, D.; van Dam, R. M.; Holmes, M. D.; Malik, V. S.; Willett, W. C.; Hu, F. B. White rice, brown rice, and risk of type 2 diabetes in US men and women. *Archives of internal medicine* **2010**, *170*, 961−969.

(10) Barth, A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2007**, *1767*, 1073−1101.

(11) Zhbankov, R.; Andrianov, V.; Marchewka, M. Fourier transform IR and Raman spectroscopy and structure of carbohydrates. *J. Mol. Struct.* **1997**, *436*, 637−654.

(12) Li, Q.; Chen, J.; Huyan, Z.; Kou, Y.; Xu, L.; Yu, X.; Gao, J.-M. Application of Fourier transform infrared spectroscopy for the quality and safety analysis of fats and oils: A review. *Critical reviews in food science and nutrition* **2019**, *59*, 3597−3611.

(13) Leopold, L. F.; Leopold, N.; Diehl, H.-A.; Socaciu, C. Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. *Spectroscopy* **2011**, *26*, 93−104.

(14) Duarte, I. F.; Barros, A.; Delgadillo, I.; Almeida, C.; Gil, A. M. Application of FTIR spectroscopy for the quantification of sugars in mango juice as a function of ripening. *Journal of agricultural and food chemistry* **2002**, *50*, 3104−3111.

(15) Anjos, O.; Campos, M. G.; Ruiz, P. C.; Antunes, P. Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. *Food chemistry* **2015**, *169*, 218−223.

(16) Srinuttrakul, W.; Mihailova, A.; Islam, M. D.; Liebisch, B.; Maxwell, F.; Kelly, S. D.; Cannavan, A. Geographical differentiation of Hom Mali rice cultivated in different regions of Thailand using FTIR-ATR and NIR spectroscopy. *Foods* **2021**, *10*, 1951.

(17) Li, Z.; Song, J.; Ma, Y.; Yu, Y.; He, X.; Guo, Y.; Dou, J.; Dong, H. Identification of aged-rice adulteration based on near-infrared spectroscopy combined with partial least squares regression and characteristic wavelength variables. *Food Chemistry: X* **2023**, *17*, 100539.

(18) Chen, H.; Lin, B.; Cai, K.; Chen, A.; Hong, S. Quantitative analysis of organic acids in pomelo fruit using FT-NIR spectroscopy coupled with network kernel PLS regression. *Infrared Physics & Technology* **2021**, *112*, 103582.

(19) Wongsaipun, S.; Theanjumpol, P.; Kittiwachana, S. Development of a universal calibration model for quantification of adulteration in Thai jasmine rice using near-infrared spectroscopy. *Food Analytical Methods* **2021**, *14*, 997−1010.

(20) Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. A review of variable selection methods in partial least squares regression. *Chemometrics and intelligent laboratory systems* **2012**, *118*, 62−69.

(21) Fernandez Pierna, J. A.; Abbas, O.; Baeten, V.; Dardenne, P. A Backward Variable Selection method for PLS regression (BVSPLS). *Analytica chimica acta* **2009**, *642*, 89−93.
(22) *Official Methods of Analysis of AOAC International*, 20th ed.; AOAC International: Gaithersburg, Maryland, 2016.
(23) *Official Methods of Analysis of AOAC International*, 17th ed.; AOAC International: Washington DC, 2005.
(24) Vera Zambrano, M.; Dutta, B.; Mercer, D. G.; MacLean, H. L.; Touchie, M. F. Assessment of moisture content measurement methods of dried food products in small-scale operations in developing countries: A review. *Trends in Food Science & Technology* **2019**, *88*, 484−496.
(25) Taylor, J. R.; Thompson, W. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*; Springer, 1982; Vol. *2*.
(26) Witjes, H.; Van den Brink, M.; Melssen, W.; Buydens, L. Automatic correction of peak shifts in Raman spectra before PLS regression. *Chemometrics and Intelligent Laboratory Systems* **2000**, *52*, 105−116.
(27) Wülfert, F.; Kok, W. T.; Smilde, A. K. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Analytical chemistry* **1998**, *70*, 1761−1767.
(28) Van, M.; Brink, D.; Hansen, J.-F.; De Peinder, P.; Van Herk, A. M.; German, A. L. Measurement of Partial Conversions During the Solution Copolymerization of Styrene and Butyl Acrylate Using On-Line Raman Spectroscopy Published online 20 November 2000. *J. Appl. Polym. Sci.* **2001**, *79*, 426−436.
(29) Pelliccia, D. A variable selection method for PLS in Python. https://nirpyresearch.com/variable-selection-method-pls-python/ (accessed on 2023-11-10).
(30) Cai, W.; Li, Y.; Shao, X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemometrics and intelligent laboratory systems* **2008**, *90*, 188−194.
(31) Coates, J. *Encyclopedia of Analytical Chemistry*; John Wiley Sons, Ltd, 2006.
(32) RP, S.; Sankar, M.; Sukumaran, R. K.; Savithri, S. Rapid estimation of the chemical composition of rice straw using FTIR spectroscopy: A chemometric investigation. *Biomass Conv. Bioref.* **2024**, *14*, 11829−11847.
(33) Hong, T.; Yin, J.-Y.; Nie, S.-P.; Xie, M.-Y. Applications of infrared spectroscopy in polysaccharide structural analysis: Progress, challenge and perspective. *Food chemistry: X* **2021**, *12*, 100168.
(34) Ji, Y.; Yang, X.; Ji, Z.; Zhu, L.; Ma, N.; Chen, D.; Jia, X.; Tang, J.; Cao, Y. DFT-calculated IR spectrum amide I, II, and III band contributions of N-methylacetamide fine components. *ACS omega* **2020**, *5*, 8572−8578.
(35) Wei, L.; Ma, F.; Du, C. Application of FTIR-PAS in rapid assessment of rice quality under climate change conditions. *Foods* **2021**, *10*, 159.
(36) Coates, J. Interpretation of infrared spectra, a practical approach. In *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*; Wiley, 2000, DOI: 10.1002/9780470027318.a5606.
(37) Krimm, S.; Bandekar, J. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Advances in protein chemistry* **1986**, *38*, 181−364.
(38) Moore, W. H.; Krimm, S. Vibrational analysis of peptides, polypeptides, and proteins. II. *β*-Poly (L-alanine) and *β*-poly (L-alanylglycine). *Biopolymers: Original Research on Biomolecules* **1976**, *15*, 2465−2483.
(39) Ridgley, D. M.; Claunch, E. C.; Barone, J. R. Characterization of large amyloid fibers and tapes with Fourier transform infrared (FT-IR) and Raman spectroscopy. *Applied spectroscopy* **2013**, *67*, 1417−1426.
(40) Casal, H. L.; Mantsch, H. H. Polymorphic phase behaviour of phospholipid membranes studied by infrared spectroscopy. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1984**, *779*, 381−401.
(41) Xin, H.; Zhang, Y.; Wang, M.; Li, Z.; Wang, Z.; Yu, P. Characterization of protein and carbohydrate mid-IR spectral features in crop residues. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2014**, *129*, 565−571.
(42) Troein, C.; Siregar, S.; Op De Beeck, M.; Peterson, C.; Tunlid, A.; Persson, P. OCTAVVS: a graphical toolbox for high-throughput preprocessing and analysis of vibrational spectroscopy imaging data. *Methods and Protocols* **2020**, *3*, 34.
(43) Nesakumar, N.; Baskar, C.; Kesavan, S.; Rayappan, J. B. B.; Alwarappan, S. Analysis of moisture content in beetroot using Fourier transform infrared spectroscopy and by principal component analysis. *Sci. Rep.* **2018**, *8*, 7996.

https://doi.org/10.1021/acsomega.4c02816
ACS Omega 2024, 9, 32760−32768