LWT - Food Science and Technology 149 (2021) 111813

Contents lists available at ScienceDirect

LWT

journal homepage: www.elsevier.com/locate/lwt

Development of NIR spectroscopy based prediction models for nutritional profiling of pearl millet (*Pennisetum glaucum* (L.)) R.Br: A chemometrics approach

Maharishi Tomar [a,b], Rakesh Bhardwaj [f,**], Manoj Kumar [d], Sumer Pal Singh [g], Veda Krishnan [b], Rekha Kansal [e], Reetu Verma [c], Vijay Kumar Yadav [a], Anil dahuja [b], Sudhir Pal Ahlawat [f], Jai Chand Rana [i], C. Tara Satyavathi [h,g], Shelly Praveen [b,***], Archana Sachdev [b,*]

[a] *Division of Seed Technology, ICAR - Indian Grassland and Fodder Research Institute, Jhansi, India*
[b] *Division of Biochemistry, ICAR - Indian Agricultural Research Institute, New Delhi, 110012, India*
[c] *Division of Crop Improvement, ICAR -Indian Grassland and Fodder Research Institute, Jhansi, India*
[d] *Chemical and Biochemical Processing Division, ICAR - Central Institute for Research on Cotton Technology, Mumbai, 400019, India*
[e] *ICAR-National Institute for Plant Biotechnology, Pusa, New Delhi, 110 012, India*
[f] *Germplasm Evaluation Division, National Bureau of Plant Genetic Resources, New Delhi, 110012, India*
[g] *Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India*
[h] *All India Coordinated Research on Pearl Millet, Jodhpur, 342304, India*
[i] *The Alliance of Bioversity International and CIAT, NASC Complex, Pusa Campus, New Delhi, 110012, India*

ARTICLE INFO

Keywords:
NIRS (Near-infrared spectroscopy)
Quantification
Regression based-satistical modelling
Nutritional composition
MPLS (Modified partial least squares) regression

ABSTRACT

Pearl millet can be viably used for food diversification due to its balanced nutritional composition. Nutritional parameters are conventionally assessed using labour and time-intensive strenuous conventional methods for germplasm screening. Near-infrared reflectance spectroscopy (NIRS) uses near-infrared sections of the electromagnetic spectrum for precise and speedy determination of biochemical parameters for large germplasm. MPLS (Modified Partial Least Squares) regression based NIRS prediction models were developed to assess starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars, phytic acid for high throughput screening of pearl millet germplasm. Mathematical treatments executed by permutation and combinations for calibrating the model, where 2nd, 3rd, and 4th derivatives produced the best results. Treatments "4,5,4,1" was finalized for protein, oil, resistant starch, total dietary fibre, "3,4,4,1" for phenolics, "2,8,4,1" for amylose, "2,4,4,1" for phytic acid, "4,7,4,1" for total soluble sugars and "2,8,4,1" for starch. Treatments with the highest 1-Variance ratio, $RSQ_{internal}$ (coefficient of determination) values, lowest SEC(V) (standard error of cross-validation), SEP(C) (standard error of performance) were identified for subsequent validation. External validation determined the prediction accuracy based on $RSQ_{external}$, RPD (residual prediction deviation), SD (standard deviation), p-value ≥ 0.05 and low SEP(C).

1. Introduction

Millets endowed with high nutritional attributes and resilience to non-conducive effects of climate change can consequentially address nutritional and food security, chiefly in developing countries. Pearl millet (*Pennisetum glaucum* (L.)) R.Br is gradually being acclaimed for their superior nutritional properties and pliability to climate change (Jukanti et al., 2016). It can also be employed as a viable alternative for food diversification due to a balanced nutritional composition of carbohydrates (72.20 g/100 g), proteins (11.80 g/100 g), lipids (6.40 g/100 g), dietary fibres (7.80 g/100 g) and minerals (1.80 g/100 g) (Dias-Martins et al., 2018). Traditionally, pedigree information, morphological traits, and cytological characters are used to evaluate the genetic diversity of pearl millet (Kumar et al., 2020). The available

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: rakesh.bhardwaj1@icar.gov.in (R. Bhardwaj), shellypraveen@iari.res.in (S. Praveen), arcs_bio@yahoo.com (A. Sachdev).

https://doi.org/10.1016/j.lwt.2021.111813
Received 24 February 2021; Received in revised form 24 May 2021; Accepted 25 May 2021
Available online 27 May 2021

genetic and biochemical diversity preserved in the germplasm directly influences the success of crop improvement programs (Kumar et al., 2016).

Pearl millet germplasm is bestowed with a wide range of nutritional and biochemical traits. The exploitation, assessment, and availability of biochemical diversity along with genetic relationships among cultivars or accessions could help to develop novel cultivars with superior nutritional traits (Ramya et al., 2018). Biochemical parameters like starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars and phytic acid primarily determine the nutritional diversity and functionality of pearl millet germplasm. These parameters are conventionally determined through complex chemical methods which are time and labour-intensive and require expensive analytical instruments and technical expertise. This makes these methods intricate and strenuous to be used for screening a large number of samples with accuracy. NIRS ascertains high speed, accurate, non-destructive, quantitative and qualitative analysis and screening of biochemical parameters in a large germplasm collection for genetic analysis and breeding programs (Anne Frank Joe et al., 2020). This technique covers electromagnetic radiations from 780 to 2500 nm (wave numbers 12500–4000 cm^{-1}) in NIR regions (Masithoh et al., 2020), relies on the absorption of infrared radiation, transmitted or reflected by a sample due to the vibrations of hydrogen bonds (Bagchi et al., 2016).

Biological samples have numerous multiple overlapping near-infrared absorption bands which arise due to a combination of vibrations and overtones of S–H, O–H, N–H and C–H functional groups (Hacisalihoglu et al., 2010). These are useful for determining the molecular interaction between functional groups and deriving chemical information about the material (Shi et al., 2019). Spectrum calibration involves the process of developing a spectrochemical prediction model. The intricate diversity of organic constituents in biological materials generates a broad spectrum absorbance peaks and multivariate regression methods, also called chemometrics, are used to calibrate the NIR spectra to these organic constituents (Hacisalihoglu et al., 2010). In essence, calibration delineates the biochemical information encompassed in the spectral properties of a substance to the physical or chemical information indicated through reference laboratory techniques.

Chemometrics is defined as the discipline that connects the evaluated values of a chemical system to the state of a system through statistical and applied mathematics (Li et al., 2020). Statistics, applied mathematics and other methods are used for identifying the optimal measurement method and test design. This is followed by analyzing and processing the measurement data for maximizing the structure, composition, and other relevant information on pertinent variables (Gendrin et al., 2008). Chemometrics can be classified into two broad categories: unsupervised and supervised pattern recognition methods (Gad et al., 2013). It includes preprocessing of spectral data (multivariate scatter correction, smoothing, centralization, and derivation), calibration models for quantitative and qualitative analysis. Thus, chemometrics is necessary for extracting the relevant chemical information from the NIR spectra and construct calibration models for associating the spectral features with the parameter of choice. Chemometrics can be successfully used along with NIR for detecting quality and nutritional parameters in pearl millet like amino acid, oil, protein, starch, moisture contents. Since it is a rapid detection technology of specific nutritional trait, it can be used for rapid and high thruput screening of diverse pearl millet germplasm and the development of varieties. It can also further help in identifying the nutritional quality of pearl millet based food products.

The intrinsic correspondence between the quantity of a biochemical parameter and their related absorption spectra determines the analytical capacity of NIRS. The authenticity and validity of an NIRS prediction model rely on the accuracy with which the relationship is delineated through pre-processing the spectral data and multivariate statistical analysis. Common pre-treatment steps of data include multiplicative scatter corrections (MSC), standard normal variate (SNV) transformations and detrending (Wu et al., 2019). These normalize the spectral data by eliminating trivial information (noise), which cannot be managed by regression analysis, thereby counterbalancing the scatter (particle size), multiplicative (tilt) and additive (baseline shift) effects. Precise correlations between biochemical components and spectral data are characterized by multivariate regression techniques, which commonly include partial least squares regression (PLS), modified partial least squares (MPLS) regression and principal component regression (PCR). Predictions made through NIR calibrations are reproducible and can even attain the accuracy of the standard reference analytical techniques for individual components. Also, the analysis through NIRS is non-destructive, cost-effective and can precisely predict multiple nutritional and biochemical traits simultaneously.

Although the biochemical constituents like protein in wheat (Shi et al., 2019), starch and moisture in cumin (Thangavel & Dhivya, 2019), the starch in corn (Jiang & Lu, 2018), amylose, and proximate composition of rice (Bagchi et al., 2016), olive oil quality (Abu-Khalaf & Hmidat, 2020), total dietary fibre in soybean and rice (Bagchi et al., 2016; Ferreira et al., 2015), phenolics in wheat (Tian et al., 2020), resistant starch in pea (Zeng & Chen, 2018), phytic acid in common beans (Carbas et al., 2020), starch, fat, protein, and amino acids in foxtail millet (Yang et al., 2013) were assessed using NIRS prediction models, no reports are available on the evaluation of these parameters in pearl millet through this technique. Thus, the present study was undertaken to develop NIRS prediction models using various combinations of smoothening, gap, derivatives, and scatter correction techniques for nine biochemical traits viz. starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars, and phytic acid to access the nutritional diversity in pearl millet germplasm.

2. Materials and methods

2.1. Materials and sample preparation

Eighty-seven pearl millet germplasm consisting of released varieties, landraces and accessions representing different parts of India were used for NIRS calibration. Grains of these germplasms were harvested at maturity and further sun-dried to a grain moisture content of 8–10% and stored at 4 °C. The required quantity of these samples were ground, homogenized and sieved through 1 mm sieve on Foss Cyclotec mill. They were subsequently subjected to NIRS and wet lab analysis for biochemical parameters namely starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars, and phytic acid. To establish an effective calibration of each of the models, the sample number in the prediction and reference sets were kept consistent for all the models.

2.2. Quantification of biochemical parameters

2.2.1. Starch content

Starch content was determined through the Megazyme assay kit (K-TSTA-100 A, Wicklow, Ireland) using amyloglucosidase, thermostable α-amylase, and glucose oxidase/peroxidase. The assay was based on a spectrophotometric AOAC method 996.11, used by McCleary et al., 2019.

2.2.2. Resistant starch

Resistant starch was determined by Megazyme assay kit (K-RSTAR, Wicklow, Ireland) using amyloglucosidase, pancreatic α-amylase, and glucose oxidase/peroxidase for the assay-based on a spectrophotometric AOAC Method 2002.02, used by McCleary et al., 2020.

2.2.3. Amylose content

Amylose content was determined by the iodometric method, where

the formation of the iodine-starch complex was detected spectroscopically (Perez & Juliano, 1978).

2.2.4. Total protein content

Protein content was estimated through nitrogen % assessment by Foss Tecator 2300 Kjeltec nitrogen auto-analyzer. The % Nitrogen was changed to per cent protein by multiplying the result by a conversion factor of 5.95. The determination was based on AOAC method 978.02, 1990 by Sáez-Plaza, Michałowski, Navas, Asuero, & Wybraniec (2013).

2.2.5. Oil content

Oil content in dried pearl millet whole grains was analyzed by a non-destructive method using Newport NMR analyzer (Model-4000) from Oxford Analytical Instruments Ltd. U.K, having a 40 mL coil assembly (Shukla et al., 2018). The instrument was calibrated using pure pearl millet seed oil which was extracted using soxhlet apparatus. The analysis was based on the estimation of proton resonance energy under the external magnetic field. The NMR responses (signal/mass) of seed samples were matched with the NMR response of pure oil to evaluate the oil percentage of the samples.

2.2.6. Total dietary fibre

Total dietary fibre was determined using Megazyme assay kit (K-TDFR-100 A, Wicklow, Ireland) by gravimetry using thermostable α-amylase, purified protease, and amyloglucosidase The determination was based on AOAC method 985.29 by McCleary et al., 2015.

2.2.7. Total soluble sugars

For total soluble sugars, the spectrophotometric method using anthrone was used (Hansen & Møller, 1975).

2.2.8. Total phenolic content

Total phenolic content was analyzed using Folin Ciocalteau reagent, and absorption at 650 nm and expressed as gallic acid equivalents (Bray & Thorpe, 1954).

2.2.9. Phytic acid content

Phytic acid was estimated spectrophotometrically using the Megazyme assay kit (K-PHYT, Wicklow, Ireland) with phytase and alkaline phosphatase. The determination was based on AOAC method 986 by McKie & McCleary, 2016.

2.3. Method for obtaining the training and validation sets

A total of 87 pearl millet germplasm including landraces and commercial varieties was assessed for nutritionally relevant biochemical parameters through standard biochemical protocols. These samples were further subdivided into training (53 samples) and validation set (test set) (34 samples). The categorization into training and validation set was based on the variability in the biochemical parameters. The values were sorted through excel such that both the training and validation set contains samples with nearly equal variability and almost equal minimum and maximum values.

2.4. Spectroscopic analysis

The homogenized samples were kept at room temperature (25 °C) for a period of 6 h to equalize temperature and moisture since these factors can influence the absorbance and reflectance of NIR waves. Before scanning and after every 30 min, the NIR spectrometer was calibrated with scanning a reference tile (100% white). Approximately 5 g of homogenized flours were scanned on FOSS NIRS 6500 spectrometer equipped with Win ISI Project Manager Software version 1.50 to obtain the reflectance spectra. The spectra were obtained by loading the homogenized sample in a circular ring cup with a quartz window of (3.8 cm in diameter and 1 cm in thickness). The samples were slightly pressed with a circular cardboard backing to assure an even packing without any air pockets. Each spectrum represented an average of 32 scans at 400–2500 nm, and was registered as log (1/R) (R stands for relative reflectance) at the increments of 2 nm. All these operations were executed at room temperature (25 °C).

2.5. Parameters to measure the accuracy and robustness of the model

The accuracy and predictive capacity of the models were evaluated using global statistical values like RSQ, slope, bias, RPD and SEP (C) (Williams et al., 2017).

$$RSQ_{internal/external} = \frac{\sum \left(y_{calculated} - y_{predicted}\right)^2}{\sum \left(y_{calculated} - y_{mean}\right)^2} \quad (1)$$

$$Bias\ (b) = \frac{1}{n}\sqrt{\sum \left(y_{predicted} - y_{Calculated}\right)^2} \quad (2)$$

$$SEP\ (C)\sqrt{\frac{\sum \left(y_{predicted} - y_{calculated} - b\right)^2}{n}} \quad (3)$$

$$1 - VR = 1 - \frac{\sum \left(x_i - x_{mean}\right)^2}{n-1} \quad (4)$$

$$RPD = \frac{SD}{SEP(C)} \quad (5)$$

Here.

$RSQ_{internal/external}$: Coefficient of determination for calibration (internal) and validation (external) sample set.

$Y_{calculated}$: The actual evaluated value of the parameter.

$Y_{predicted}$: The predicted value of the parameter through a regression line.

Y_{mean}: Arithmetic mean of y values

n: Number of spectra

x_i: Value of the one observation

x_{mean}: Mean value of all observations.

2.6. Calibration and validation

After the acquisition of laboratory reference and spectral data, they were matched followed by statistical and mathematical procedures. The prediction equations were developed through a MPLS regression method along with cross-validation. For each biochemical parameter, various mathematical algorithms were applied for scatter correction and pre-processing the spectral data including the Standard Normal Variate (SNV) and Detrending (DT). Application of these algorithms and treatments enables smoothening of the spectral data before subjecting to regression analysis. This corrects the particle size effect (light scatter) and nullifies the variation caused by an alteration in light path length. Overlapping absorption bands and baseline shift effects were eliminated by calculating the spectral derivatives. Additionally, mathematical treatments with four derivatives were tested by trial and error to develop the NIRS calibrations for 400–2500 nm spectral region. For example, the mathematical treatment denoted as "2,5,4,1" means D = 2, G = 5, S1 = 4 and S2 = 1, where D represents the order of derivative, G represents the gap (data points calculated by the specific derivation), S1 depicts the number of data points in the first smoothing and S2 shows the number of data points in the second smoothing. Overfitting of the model was prevented by cross-validating the calibration under SNV and detrend scatter correction (Wu & Shi, 2004).

The developed calibrations were assessed by various statistical parameters. The extent of distribution and variability was evaluated by the range and standard deviation (SD). Win ISI Project Manager Software version 1.50, was used to automatically calculate the standard error of calibration (SEC) and coefficient of determination in internal validation

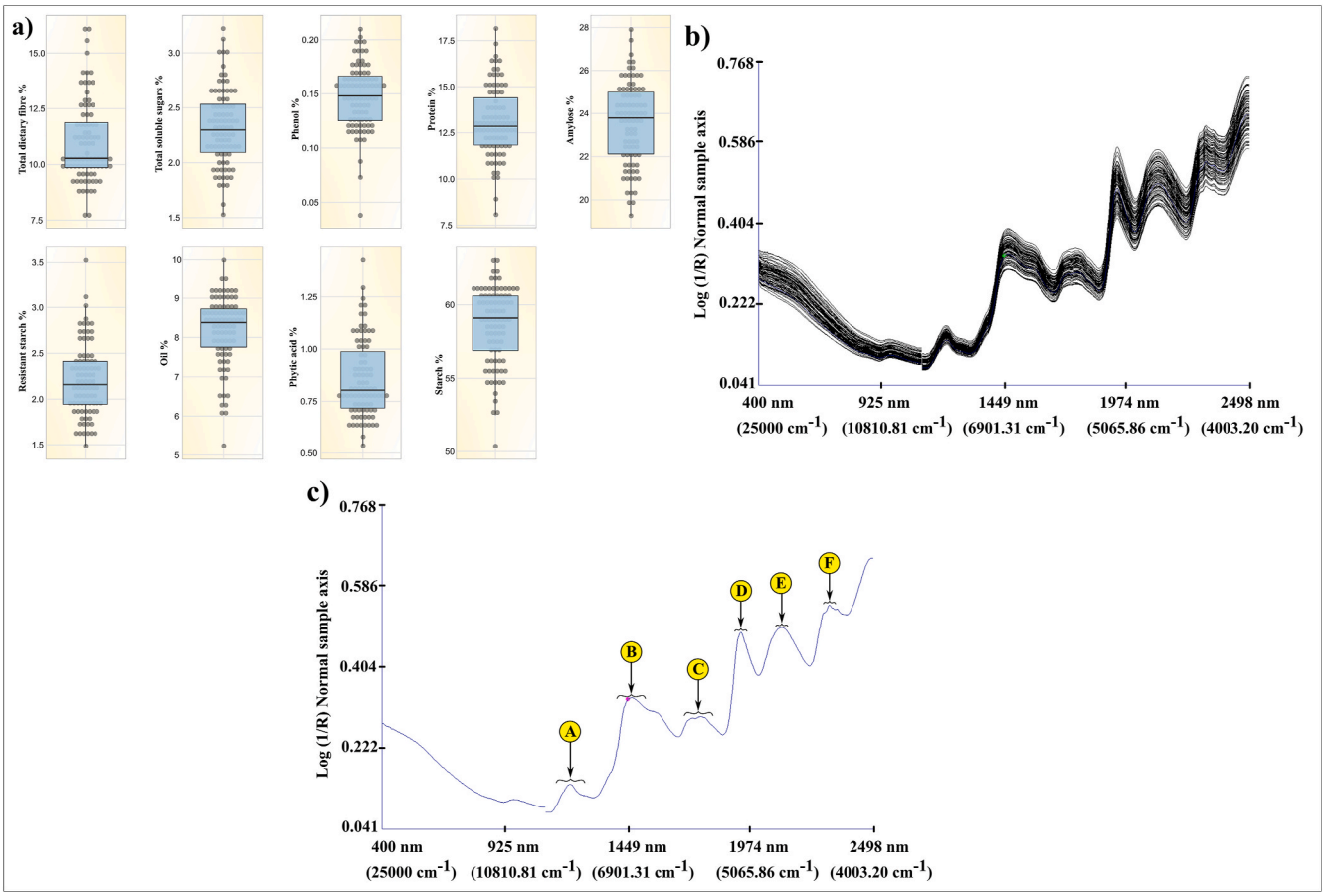

Fig. 1. a: Distributions of starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars, phytic acid in 87 pearl millet genotypes
Fig. 1b:A combined plot of the reflectance of all the entire pearl millet germplasm (87 samples)
Fig. 1c: An average reflectance spectrum of pearl millet homogenized flour with peaks (A) indicate weak absorption bands, which could arise by symmetric stretching of (-CH) in methyl groups (-CH_3) of biomolecules; (B) indicate O–H stretch first overtone of hydroxyl phenol groups and C–H combinations of aromatic compounds, combinations of symmetric stretching of O–H in amylose. O–H stretch and first overtone of starch and cellulose; (C) indicate asymmetric C–O–O stretches the third polysaccharide overtone, (C–O) of oils; (D) indicate O–H bending/stretching of polysaccharides, combinations of bending and stretching of O–H in amylose; (E) indicate combinations of C–O and N–H stretching, linked to protein, O–H groups of phytic acids, proteins and tannins. O–H bend/C–O stretch combination bands for starch. C–O–O stretch and third overtone for starch or cellulose; (F) indicate C–H bending, the second overtone of oils.

($RSQ_{internal}$). SEC indicates the standard error of variation between the reference values and values predicted by NIRS calibration models in the calibration set. RSQ displays the portion of the variance in reference data that can be defined by the variance in predicted data. Models having a lower SEC and a higher RSQ are superior to those with higher SEC and lower RSQ values. Two comparable parameters, standard error of cross-validation (SEC(V)) and 1 minus variance ratio (1-VR) [Eq. (4)] were computed as assessors for error and cross-validation coefficient, respectively. The components of mathematical processing treatments were sought through trial and error so is to decrease the SEC(V) and increase the 1-VR during cross-validation. Moreover, these models were individually authenticated by an external sample validation set since only cross-validation is inadequate for verifying the model. The prediction accuracy of every equation for the external validation was assessed based on $RSQ_{external}$ (coefficient of determination in the external validation), bias (the systematic difference between the reference and predicted value, and thus suggests the correctness of correlation between the two values), standard error of performance (SEP), and corrected standard error of performance (SEP(C)). Accuracy of MPLS models was evaluated by residual prediction deviation (RPD), which was calculated as the ratio of SD of the reference values of the samples in the validation set to the SEP corrected for bias (Cozzolino et al., 2006).

2.7. Statistical analysis and experimental design

Win ISI III Project Manager software version 3.1 was used to compute all the calibrations and predictions through various mathematical treatments based on spectral and analyzed data. MS excel and Veusz –a scientific plotting package was used to compute the coefficient of regression ($RSQ_{internal/external}$) Eq. (1) for reference vs. predicted values for various biochemical parameters. The prediction accuracy of the model was evaluated by performing a paired *t*-test using Jamovi at a 95% confidence interval. The result of the paired *t*-test was determined in form of a *p*-value. *p*-values higher than 0.05 indicates the rejection of the null hypothesis that the difference between the means of predicted and reference values are considerably different while *p*-value less than 0.05 indicates the acceptance of the null hypothesis.

A completely randomized design (CRD) was used for the experiment where the treatments (spectral acquisition and evaluation of nutritional traits) were completely random so that each sample unit (homogenous pearl millet flour) has the same probability of receiving any single treatment. The duplicates of each sample were scanned twice. The average spectrum of each sample was used in subsequent data analysis. The biochemical traits were evaluated in duplicates and their mean values were used for calibration and validation set.

Table 1
Model statistics for the calibration set.

TRAIT	N	RANGE (%)	MATH TREATMENT	MEAN	$RSQ_{internal}$	SLOPE	1-VR	SD	SEC (V)
Protein	53	8.91–18.15	4,5,4,1	12.98	0.933	0.989	0.9032	1.718	0.5472
Oil	53	5.24–9.99	4,5,4,1	8.26	0.946	1.035	0.6440	0.932	0.5286
Total dietary fibre	53	7.68–16.18	4,5,4,1	10.88	0.953	0.983	0.8987	1.809	0.5840
Starch	53	52.49–63.25	2,8,4,1	58.642	0.827	1.000	0.8073	2.493	1.0937
Amylose	53	19.88–26.50	2,8,4,1	23.38	0.819	1.000	0.7993	1.674	0.7470
Resistant Starch	53	1.48–3.52	4,5,4,1	2.19	0.762	1.000	0.6905	0.357	0.1981
Total Soluble Sugars	53	1.62–3.22	4,7,4,1	2.33	0.874	0.975	0.4133	0.323	0.2484
Phenolics	53	0.04–0.21	3,4,4,1	0.15	0.803	1.000	0.6343	0.033	0.0199
Phytic Acid	53	0.54–1.43	2,4,4,1	0.85	0.833	1.037	0.7097	0.1990	0.1014

$RSQ_{internal}$ = coefficient of determination for calibration; SD = standard deviation; SEC(V) = standard error of cross validation; 1-VR = 1 minus variance ratio; N=Number of samples; the values of traits are expressed as g/100 g.

3. Results and discussion

3.1. Biochemical parameters and NIRS spectra

Nutritionally relevant biochemical parameters for 87 diverse pearl millet germplasm represented as range, mean value, and the standard deviation is presented in Table S1 (Supplementary Document) and Fig. 1a. Fig. 1b represents the NIR spectra of homogenized flour of 87 diverse pearl millet germplasm in the wavelength range of (400–2490 nm/25000–4016 cm^{-1}). The average spectra obtained from all the tested samples is represented by the spectral line. It is very strenuous to visually differentiate between the NIR regions due to the presence of highly overlapping and broad combination bands of fundamental vibrations (Cozzolino, 2015). It is more difficult for biological materials like pearl millet due to the presence of complex structural matrix having hydrogen bonding between protein, fatty acids, carbohydrates, and other biomolecules. These absorption peaks result from overlapping absorptions that fundamentally correspond with combinations and overtones of vibrational modes including N–H, O–H and C–H, associated with proteins, fatty acids, and carbohydrates, respectively.

Peak found in the spectral region between 4500 and 5000 cm^{-1} can be designated to the combinations of C–O and N–H stretching, linked to protein contents (Plans et al., 2013). Regions between 4500 and 5000 cm^{-1} can be allocated to O–H groups, present in biological compounds like phytic acids, proteins, and tannins, while regions between 6100 and 6400 cm^{-1} are also associated to phytic acid content (Pande & Mishra, 2015). A relatively wide peak at 7000–6800 cm^{-1} can be assigned to O–H stretch first overtone of hydroxyl phenol groups and C–H combinations of aromatic compounds. Peak near 5200 cm^{-1} was identified with O–H bending/stretching of polysaccharides, converging with those of water. Peaks found near 4800 cm^{-1} could arise from asymmetric C–O–O stretches the third polysaccharide overtone (Zhang et al., 2017). Sharp absorption bands were noticed at around 5184 cm^{-1} representing the combinations of bending and stretching of O–H in amylose, also the peak near 6835 cm^{-1} is concerned with combinations of symmetric stretching of O–H in amylose. The wavelength of 8316 cm^{-1} causes weak absorption bands, which could arise by symmetric stretching of (-CH) in methyl groups ($-CH_3$) of biomolecules. Absorption bands near 8278 cm^{-1} were connected to (C–H) stretching, the second overtone ($-CH_2$), 5800 cm^{-1}, related with (C–O) of oils and 4332 cm^{-1} is concerned with (C–H) bending, the second overtone of oils (Kaur et al., 2017). Similar peaks determining these specific functional groups were also found by Chen et al., 2013. The mean reflectance spectrum of pearl millet flour along with all the major peaks is depicted in Fig. 1c, A, B, C, D, E, F.

Understanding of the relationships between the calibrated trait its associated wavelength can become complex due to an overlap between the NIR band vibrations associated with different traits. The presence of a trait in extremely low concentration can also be another limiting factor. These traits usually display low wavelength regression coefficients and NIR absorption band number. Our results indicate that the prediction accuracy of a model is low when a trait is present in less amount with relatively low $RSQ_{external}$ (phenolics, resistant starch, phytic acid and total soluble sugars). The presence of absorption bands in multiple spectral regions related to a trait is another challenge that results in band vibrations overlap. This is identified to be a major limiting factor for the evaluation of biological samples with identical chemical bonds or having components with similar absorption regions despite having dissimilar structures. For example, O–H bonds having wide absorption regions near 2100 nm, screen the alterations in energy caused protein amide bond absorption (Egesel & Kahriman, 2012). These factors can result in poor prediction capacity of the model for a specific trait having multiple wavelengths.

3.2. Derivation, regression, and preprocessing methods for MPLS models (calibration)

To prevent bias in the sub-set divisions, all the parameters were placed in ascending order based on their analyzed biochemical parameter. The calibration and validation set were selected to encompass the full range of concentrations. After the removal of outliers, all the samples were divided into an internal calibration set (N = 53), which was used to build and train the model and an external validation set (N = 34), which is used to test the robustness and accuracy of the model. Regression algorithms based on full-range spectra were used for model development. PLS, MPLS and PCR are the most used regression algorithms for NIR model development. PLS has features of both multiple regression and PCA (principal component analysis). PLS is similar to PCR but uses both spectral information and reference data (physical, chemical, etc.) for forming factors used for a suitable purpose. PLS, MPLS and PCR were used for the model development, but MPLS was found to be more accurate and stable than the standard PLS regression algorithm and was thus used in the present study. In MPLS, the NIR residuals acquired at each wavelength and after each factor was computed and standardized (dividing them by the SD of the residual values at each wavelength) before calculating the next factor (Font et al., 2004). MPLS builds its factors by catching the maximum possible variation in the spectroscopic data by actively utilizing the reference values (physical, chemical, etc.) during spectroscopic data decomposition. This method decreases the effect of irrelevant and large spectroscopic variations in the calibration modelling by counterbalancing the biochemical data and spectroscopic information.

Scattering of light and pathlength variations caused by interactions between the light and sample particles usually causes an alteration in the absorption levels. This makes the linear calibration and spectral interpretation of NIR reflectance spectra very complex and difficult. Pathlength variations resulting from light scattering generates background signals that change with the wavelength causing curvature and baseline shift, which changes with the sample. SNV transformation diminishes the multiplicative effects of particle size and scattering and also decreases the difference between the global signal intensities. SNV works by centring each spectrum around zero by subtracting the mean, and

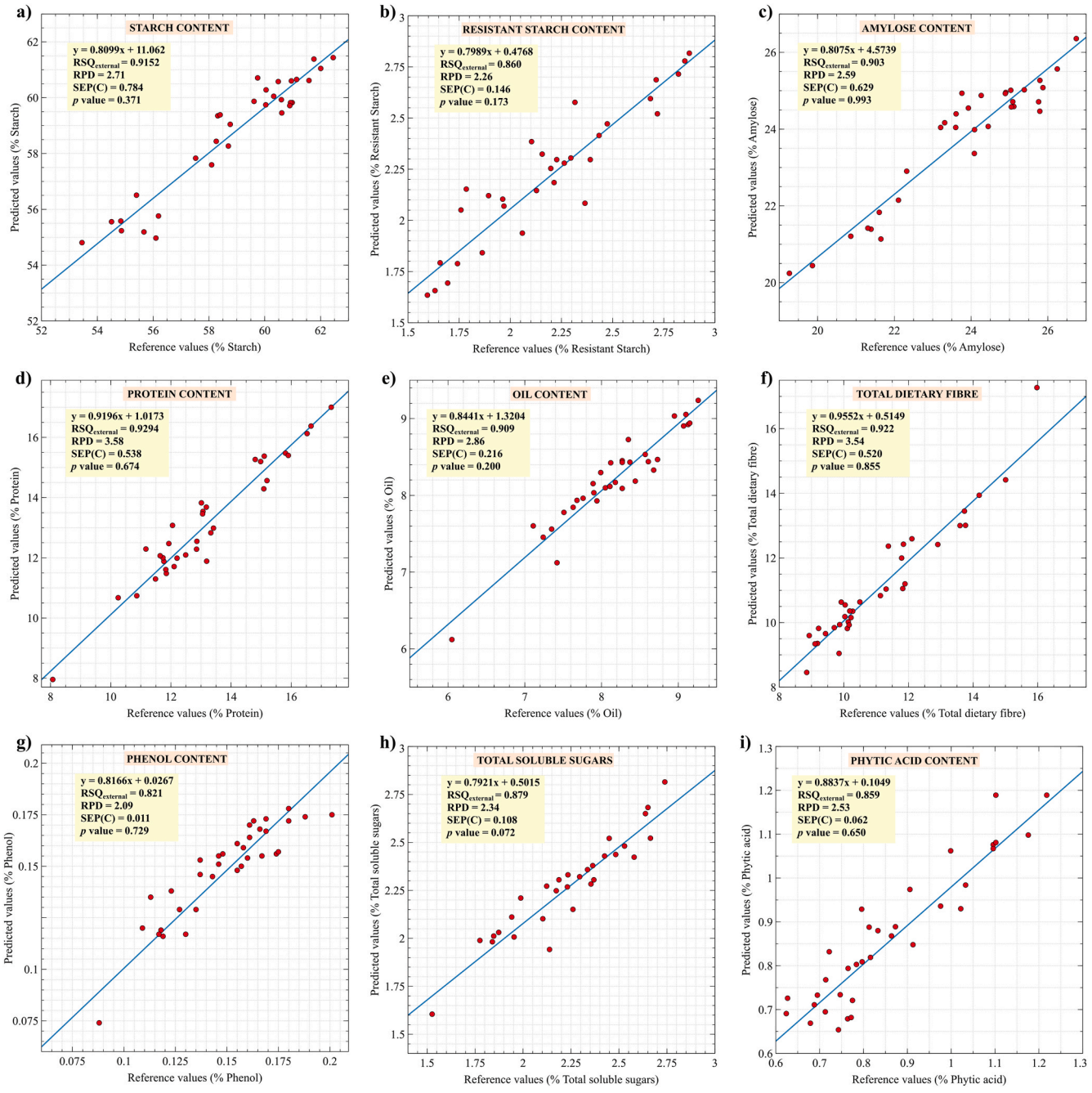

Fig. 2. Scatter plot between the reference versus predicted values for (a) starch, (b) resistant starch, (c) amylose, (d) protein, (e) oil, (f) total dietary fibre, (g) phenolics (h) total soluble sugars, (i) phytic acid. $RSQ_{external}$ – coefficient of determination for validation; RPD – residual prediction deviation; SEP(C) – corrected standard error of performance, *p*-it indicates the probability of achieving the test result under the null hypothesis.

further dividing the value of each signal by the SD of the complete spectrum. The spectral absorption of NIR linearly increases with the wavelength for transparent samples, while it curvilinearly increases in the case of densely packed sample particles. Detrend method is another approach to correct the baseline shift, which is used along with SNV. SNV along with Detrend was used for the development of models in the present study to circumvent any curvilinearity and noise in the NIRS signal baseline.

The development of calibration by MPLS regression for starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars and phytic acid in pearl millet homogenized flour is summarized in Table 1. Several combinations of mathematical treatments were executed for the development of calibration equations. Mathematical treatments "4,5,4,1", "3,4,4,1", "2,8,4,1", "2,4,4,1", "4,7,4,1", and "2,8,4,1" were finalized for the development of calibration equations for various parameters based on highest 1-VR and RSQ values, lowest SEC(V), and SEP(C) values, and the removal of outliers. Derivatives 2, 3, and 4 were used to eliminate the persistent background signals and to improve the visual resolution. Global baseline changes and background signals are a low-frequency phenomenon, thus derivatives can be elucidated as high-pass filters for clearly determining weak peaks that cannot be perceived in the native spectrum. Gaps 5, 4, 8, and 7 and smoothening (S1-4, S2-1) were used to reduce the noise, caused by erratic high-frequency perturbations. As summarized in Table 1, the highest coefficient of correlation ($RSQ_{internal}$) 0.933 was obtained for proteins at "4,5,4,1", 0.803 for phenolics at "3,4,4,1", 0.819

Table 2
Model statistics for the validation set.

TRAIT	N	% RANGE (Calibration)	% RANGE (Predicted)	Math Treatment	$RSQ_{external}$	SLOPE	BIAS	SD	SEP (C)	RPD
Protein	34	8.91–18.15	7.96–17.00	4,5,4,1	0.929	1.011	0.039	1.928	0.538	3.58
Oil	34	5.24–9.99	6.12–9.23	4,5,4,1	0.909	1.076	−0.049	0.618	0.216	2.86
Total dietary fibre	34	7.68–16.18	8.46–17.27	4,5,4,1	0.922	0.966	−0.017	1.843	0.52	3.54
Starch	34	52.49–63.25	54.80–61.43	2,8,4,1	0.915	1.13	0.126	2.132	0.784	2.71
Amylose	34	19.88–26.50	20.24–26.36	2,8,4,1	0.903	1.118	−0.001	1.635	0.629	2.59
Resistant Starch	34	1.49–3.52	1.63–2.82	4,5,4,1	0.860	1.076	−0.037	0.331	0.146	2.26
Total Soluble Sugars	34	1.62–3.22	1.60–2.82	4,7,4,1	0.879	1.109	−0.037	0.253	0.108	2.34
Phenolics	34	0.04–0.21	0.07–0.18	3,4,4,1	0.821	1.001	0.001	0.023	0.011	2.09
Phytic Acid	34	0.54–1.43	0.65–1.19	2,4,4,1	0.859	0.972	−0.005	0.157	0.062	2.53

$RSQ_{external}$ = coefficient of determination for valiation; SD = standard deviation; SEP(C) = standard error of performance; RPD = residual prediction deviation; N=Number of samples; the values of traits are expressed as g/100 g.

for amylose at "2,8,4,1", 0.946 for oils at "4,5,4,1", 0.762 for resistant starch at "4,5,4,1", 0.953 for total dietary fibre at "4,5,4,1", 0.833 for phytic acid at "2,4,4,1", 0.874 for total soluble sugars at "4,7,4,1" and 0.827 for starch at "2,8,4,1".

3.4. Validation of the NIR model

A total of 34 samples were used for validating the developed models. Statistical measures like $RSQ_{external}$, slope, bias, RPD and SEP (C) [Eq. (3)] were used to measure the accuracy and precision of the model. RSQ determines the closeness of the data to the fitted line of regression and the accuracy with which the regression model fits the actual data values. RSQ value of 0.8, indicates that 80% of the data fits the specific regression model. A higher RSQ value suggests a better fit for the model. Highest RSQ for validation was observed for proteins (0.929), followed by total dietary fibre (0.922), starch (0.915), oil (0.909), amylose (0.903), total soluble sugars (0.879), resistant starch (0.860), phytic acid (0.859) and phenolics (0.821). A similar RSQ of 0.894 for protein and 0.860 for oil content was observed in pearl millet by Choudhary et al., 2010. Higher $RSQ_{external}$ between 0.929 and 0.821 indicates a better model fit. In Fig. 2.a, Fig. 2.b, Fig. 2.c, Fig. 2.d, Fig. 2.e, Fig. 2.f, Fig. 2.g, Fig. 2.h and Fig. 2.i the regression plot of predicted values versus reference values for starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars and phytic acid are described respectively. The regression plot of predicted values versus reference values developed through Win ISI III Project Manager software version 3.1 for various traits is represented by Figure S1, S2, S3, S4, S5, S6, S7, S8, S9.

The range for the conventionally calculated values in the calibration and values predicted by the model for various biochemical parameters is summarized in Table 2. Calibration range for proteins was 8.91–18.15 g/100 g, while those predicted by the model were 7.96–17.00 g/100 g, calibration for phenolics was 0.04–0.21 g/100 g while predicted was 0.07–0.18 g/100 g, calibration for amylose was 19.88–26.50 g/100 g while predicted was 20.24–26.36 g/100 g, calibration for oil was 5.24–9.99 g/100 g while predicted was 6.12–9.23 g/100 g, calibration for resistant starch was 1.49–3.52 g/100 g while predicted was 1.63–2.82 g/100 g, calibration for total dietary fibre was 7.68–16.18 g/100 g while predicted was 8.46–17.27 g/100 g, calibration for phytic acid was 0.54–1.43 g/100 g while predicted was 0.65–1.19 g/100 g, calibration for total soluble sugars was 1.62–3.22 g/100 g while predicted was 1.60–2.82 g/100 g, calibration for starch was 52.49–63.25 g/100 g while predicted was 54.80–61.43 g/100 g. The results indicate a good concurrence between the calibration and prediction ranges, depicting the prediction accuracy of the model.

The slope represents the change in predicted values with a one-unit change in reference values. While the ideal value of slope should be 1, any value close to 1 indicates an accurate model. The values of slope ranged from 0.966 to 1.13, where proteins (1.011), phenolics (1.001), total dietary fibre (0.966) and phytic acid (0.972) displayed the most optimum slope values. Similar slope values of 0.909 for proteins, 0.920 for starch, and 0.912 for oil was found by Choudhary et al., 2010 in pearl millet grains. Bias is an important indicator of similarity between reference and predicted values, and in turn determines the model accuracy Eq. (2) (Wu et al., 2019). The ideal value for bias should be equal to zero, which means the reference and predicted values are the same. A positive value of bias indicates that the model overestimates the evaluated parameter. A negative bias value signifies that the model underestimates the evaluated parameter. Oil (−0.049), resistant starch (−0.037), total soluble sugars (−0.037), total dietary fibre (−0.017), phytic acid (−0.005), and amylose (−0.001) displayed negative bias, while phenolics (0.001), proteins (0.039) and starch (0.126) exhibited positive bias. This indicates that the models for oil, resistant starch, total soluble sugars, total dietary fibre, phytic acid, and amylose slightly underestimated these parameters while the models for phenolics, proteins and starch slightly overestimated these parameters.

RPD considers, both the standard error of prediction and variation in the values. This provides a matric to authenticate the model's validity,

Table 3
Paired sample *t*-test for between lab and predicted values.

Pairs	Paired Differences					t value	DF	*p* value
	Mean	SD	SEM	95% Confidence Interval of the Difference				
				Lower	Upper			
Protein analytical - Protein predicted	0.039118	0.537452	0.092172	−0.148408	0.226644	0.424	33	0.674
Oil analytical - Oil Predicted	−0.049182	0.216102	0.037618	−0.125808	0.027444	−1.307	32	0.200
Total dietary fibre analytical - Total dietary fibre predicted	−0.016471	0.520303	0.089231	−0.198013	0.165072	−0.185	33	0.855
Starch analytical - Starch predicted	0.125813	0.784156	0.138621	−0.156906	0.408531	0.908	31	0.371
Amylose analytical - Amylose Predicted	−0.001000	0.629345	0.111253	−0.227903	0.225903	−0.009	31	0.993
Resistant starch analytical - Resistant starch predicted	−0.036645	0.146054	0.026232	−0.090218	0.016928	−1.397	30	0.173
Total soluble sugars analytical - Total soluble sugars predicted	−0.036867	0.108134	0.019742	−0.077245	0.003511	−1.867	29	0.072
Phenolics analytical - Phenolics predicted	0.000647	0.010787	0.001850	−0.003117	0.004411	0.350	33	0.729
Phytic acid analytical - Phytic acid predicted	−0.004853	0.061753	0.010591	−0.026399	0.016694	−0.458	33	0.650

This predicts that there is no statistical difference between between predicted and reference values at 95% confidence interval. DF = Degrees of freedom; SD = standard deviation; SEM = standard error mean; t = test statistic; *p* = probability of achieving the test result under null hypothesis.

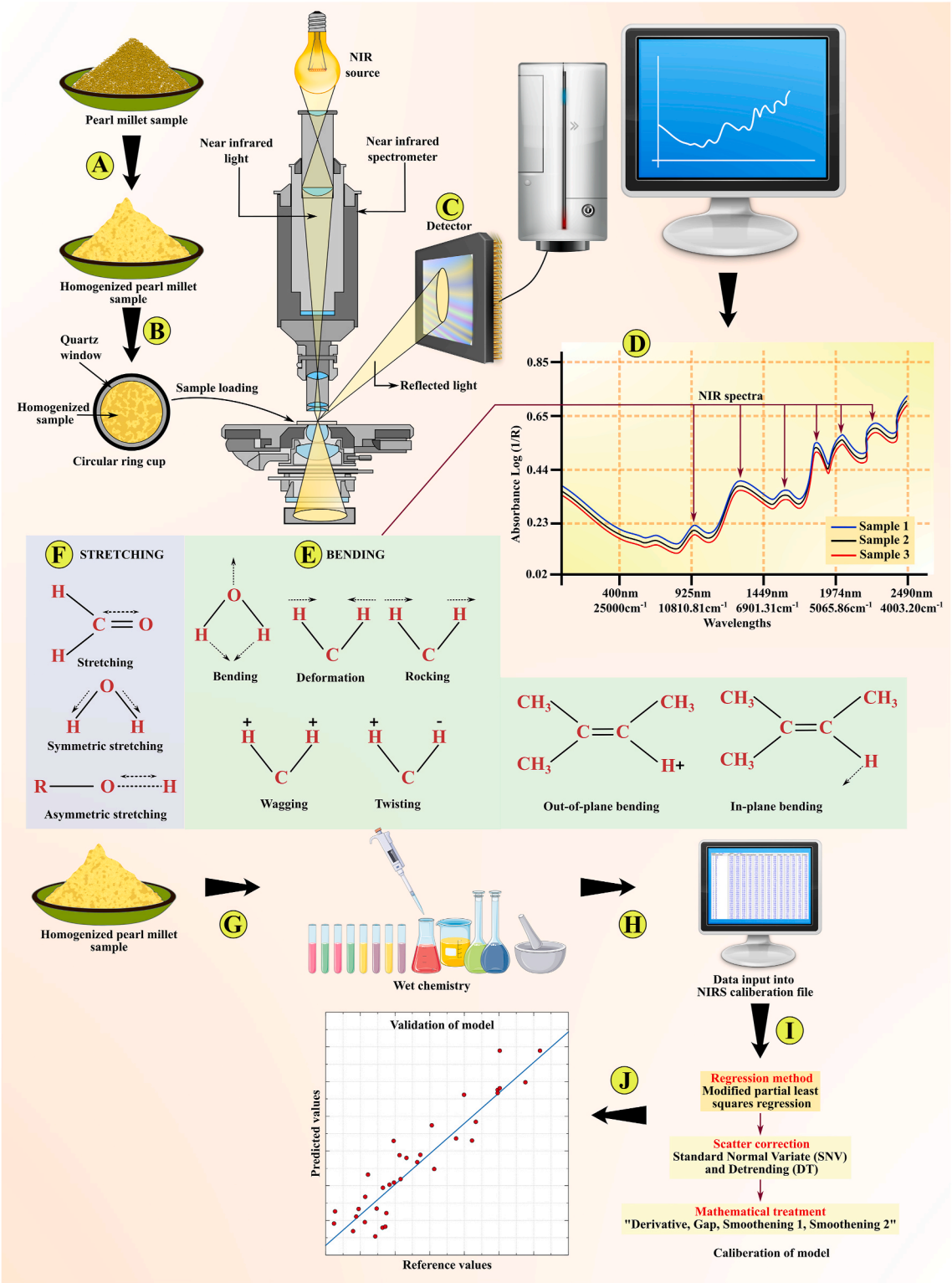

Fig. 3. Overall illustration of the experimental procedure. (A) Pearl millet samples were ground, homogenized and sieved through 1 mm sieve; (B) The spectra were obtained by loading the homogenized sample in a circular ring cup with a quartz window (3.8 cm in diameter and 1 cm in thickness); (C) Detection of the reflected NIR light by the detector; (D) A typical NIRS spectrum indicating the peaks; (E) A molecule absorbs the incoming IR radiation which causes a change in bond length (stretching) and/or (F) bond angle (bending); (G) Wet chemistry analysis of homogenized pearl millet flour for nutritionally relevant biochemical parameters like starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars, and phytic acid; (H) Input of wet chemistry data into the NIR calibration file; (I) Calibration of model using modified partial least squares regression. This was followed by the application of various mathematical algorithms scatter correction and preprocessing the spectral data including the Standard Normal Variate (SNV) and Detrending (DT); (J) Representation of the prediction accuracy by scatter plot between the reference and predicted values.

which is more precise than SEP(C) and can be easily compared across various model validation studies [Eq. (5)] (Prieto et al., 2017). The ideal value of RPD should be higher than 1. An RPD value between 1.5 and 2 can differentiate between high and low values of the response variable. Values between 2 and 2.5 indicate that the model can perform rough quantitative predictions. Values $\geq$ 2.5 corresponds to good and those with $\geq$3 shows excellent prediction accuracy of the model (Williams, 2014). Models for total dietary fibre with RPD (3.54), proteins (3.58) displayed excellent prediction capacity. Models for phytic acid (2.53), amylose (2.59), starch (2.71) and oil (2.86) exhibited excellent prediction capacity. Models for total soluble sugars (2.34), resistant starch (2.26) and phenolics (2.09) can roughly perform good quantitative predictions. A similar RPD value of 4.07, with a bias of 0.07 in foxtail millet was indicated by Chen et al., 2013. As a general thumb rule for the development of robust models, the SEP(C) should not surpass 1.30 times the value of SEC. The slope should have a minimum value of 0.90, the minimum $RSQ_{internal/external}$ should be 0.60 and the bias should not exceed SEC by $\pm$0.6. The models in our study are in agreement with this rule.

As summarized in Table 3, the paired *t*-test at 95% confidence interval, was used to determine whether the mean of a dependent variable is same in the analytical and predicted values for the evaluated biochemical parameters. All the models exhibited *p*-values higher than 0.05 indicating the reliability and accurate prediction capacity of the models. The highest *p*-value was demonstrated by amylose (0.993), followed by total dietary fibre (0.855), phenolics (0.729), proteins (0.674), phytic acid (0.650), starch (0.371), oil (0.200), resistant starch (0.173) and total soluble sugars (0.072). Fig. 3 graphically illustrates the complete experimental procedure along with the major steps.

From a bromatological standpoint, pearl millet grains are used for producing a wide variety of traditional foods including sweets, couscous, flatbreads, porridges, non-alcoholic drinks (*marewa*, *oskikundu*, *mahewu*, *bushera*, *kunun Zaki*, *boza*, *pito*, *pombe*) and alcoholic beverages (*merissa*, *mbeg*, *chibuku shake*, *Dogon millet beer* or *opaque beer*) (Dias-Martins et al., 2018). It is well established that every food product requires a specific biochemical and nutritional composition. Germplasm screened through this technique can be subsequently used for their bromatological analysis and establishing their suitability for a specific food product. For example, the germplasm with a higher content of TSS could be used for synthesizing high glycemic index foods. Germplasm with superior contents of starch, amylose, oil, protein, TDF and lower phytic acid could be more suitable for food processing industries and synthesizing functional foods. These traits determine optimum techno-functional properties and nutritional qualities. Germplasm with higher content of resistant starch, TDF and low TSS could be used for designing low glycemic index foods. Thus this technique can be feasibly used for quickly screening large germplasm and establishing their use for synthesizing a specific food or nutraceutical product.

4. Conclusions

Currently, rapid nondestructive techniques for screening large germplasm collections based on nutritionally relevant biochemical attributes are important in the breeding program for the development of nutritionally rich varieties and food industries. The present work is the first report on the development of effectual MPLS regression models for the rapid quantitative determination of starch, resistant starch, amylose, protein, oil, total dietary fibre, phenolics, total soluble sugars and phytic acid in pearl millet homogenized flour based on NIR spectroscopy and chemometrics. A robust calibration model for routine germplasm screening requires diverse spectral libraries, reliable validation techniques and consistent testing for evaluating real performance. A thorough understanding of spectral data and chemometrics can ensure the reliability of NIRS prediction models for accurately predicting the key physical or biochemical components under consideration. Generally, the prediction errors are consistent and low with the common values of the parameters. The selection of a specific spectral region in many cases can significantly enhance the prediction accuracy.

When compared to conventional analytical methods, these NIR models are more efficient, highly eco-friendly, and less labour-intensive means to simultaneous assess the required components. This study will be of great value for the effective use of NIRS technique for high throughput screening of diverse pearl millet germplasm for nutritional diversity in a non-destructive manner. Chemometrics is now an emerging science for nutritional and biochemical analysis and an important tool for enabling the use of NIRS for accessing the intrinsic nutritional quality of grains like pearl millet. The vibrational spectroscopy associated to multivariate methods in the presently developed models can serve as a potential analytical tool for plant breeding, food industries and inspection agencies for developing, processing, controlling and assessing the nutritional quality of pearl millet grains in a cost-effective and accurate manner. These models could also enable the economical and rapid culling of less desirable genotypes, before their expensive and elaborate evaluations for the necessary traits and identifying more promising genotypes.

CRediT authorship contribution statement

Maharishi Tomar: Investigation, Writing – original draft, Visualization. **Rakesh Bhardwaj:** Investigation, Writing – original draft, Visualization, Conceptualization, Methodology, Supervision, Writing – review & editing. **Manoj Kumar:** Formal analysis, Investigation, review. **Sumer Pal Singh:** Writing – review & editing. **Veda Krishnan:** Writing – review & editing. **Rekha Kansal:** Writing – review & editing. **Reetu Verma:** Formal analysis, Investigation, review. **Vijay Kumar Yadav:** Writing – review & editing. **Anil dahuja:** Writing – review & editing. **Sudhir Pal Ahlawat:** Writing – review & editing. **Jai Chand Rana:** Writing – review & editing. **C. Tara Satyavathi:** Writing – review & editing. **Shelly Praveen:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Archana Sachdev:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the genotype shared by IARI, AICRP on Pearl Millet, Jodhpur.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lwt.2021.111813.

Funding information

The present work is funded by Division of Agricultural Education, ICAR under the Niche Area of Excellence (NAE) Programme (Scheme Strengthening and Development of Higher Agricultural Education in India) (Project Sanction no. Edn. 5(22)/2017-EP&HS, 2019; IARI code: 12/223) and the support of the Global Environment Facility (GEF) of the United Nations Environment Program (UNEP) within the project "Mainstreaming agricultural biodiversity conservation and utilization in the agricultural sector to ensure ecosystem services and reduce vulnerability".

References

Abu-Khalaf, N., & Hmidat, M. (2020). Visible/Near Infrared (VIS/NIR) spectroscopy as an optical sensor for evaluating olive oil quality. *Computers and Electronics in Agriculture, 173*(April), 105445. https://doi.org/10.1016/j.compag.2020.105445

Anne Frank Joe, A., Gopal, A., & Pandian, R. (2020). Performance evaluation of chemometric prediction models-key components of wheat grain. *Journal of Scientific and Industrial Research, 79*(2), 148–152.

Bagchi, T. B., Sharma, S., & Chattopadhyay, K. (2016). Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. *Food Chemistry, 191*, 21–27. https://doi.org/10.1016/j.foodchem.2015.05.038

Bray, H. G., & Thorpe, W. V. (1954). Analysis of phenolic compounds of interest in metabolism. *Methods of Biochemical Analysis, 1*, 27–52. https://doi.org/10.1002/9780470110171.ch2

Carbas, B., Machado, N., Oppolzer, D., Ferreira, L., Brites, C., Rosa, E. A. S., & Barros, A. I. R. N. A. (2020). Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for the determination of nutritional and antinutritional parameters in common beans. *Food Chemistry, 306*, 125509. https://doi.org/10.1016/j.foodchem.2019.125509

Chen, J., Ren, X., Zhang, Q., Diao, X., & Shen, Q. (2013). Determination of protein , total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy. *Journal of Cereal Science, 58*(2), 241–247. https://doi.org/10.1016/j.jcs.2013.07.002

Choudhary, S., Hash, C., Sagar, P., Prasad, K. V. S. V., & Blümmel, M. (2010). Near infrared spectroscopy estimation of pearl millet grain composition and feed quality. In *Proceedings of the 14th international conference on NIR spectroscopy, november 2014* (pp. 87–90). http://www.researchgate.net/profile/Sunita_Choudhary4/publication/268575415_Near_infrared_spectroscopy_estimation_of_pearl_millet_grain_composition_and_feed_quality/links/547070050cf24af340c0a3c0.pdf.

Cozzolino, D. (2015). Foodomics and infrared spectroscopy: From compounds to functionality. *Current Opinion in Food Science, 4*, 39–43. https://doi.org/10.1016/j.cofs.2015.05.003

Cozzolino, D., Fassio, A., Fernández, E., Restaino, E., & La Manna, A. (2006). Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy. *Animal Feed Science and Technology, 129*(3–4), 329–336. https://doi.org/10.1016/j.anifeedsci.2006.01.025

Dias-Martins, A. M., Pessanha, K. L. F., Pacheco, S., Rodrigues, J. A. S., & Carvalho, C. W. P. (2018). Potential use of pearl millet (*Pennisetum glaucum* (L.) R. Br.) in Brazil: Food security, processing, health benefits and nutritional products. *Food Research International, 109*(April), 175–186. https://doi.org/10.1016/j.foodres.2018.04.023

Egesel, C.Ö., & Kahriman, F. (2012). Determination of quality parameters in maize grain by NIR reflectance spectroscopy. *Tarim Bilimleri Dergisi, 18*(1), 31–42. https://doi.org/10.1501/tarimbil_0000001190

Ferreira, D. S., Poppi, R. J., & Lima Pallone, J. A. (2015). Evaluation of dietary fiber of Brazilian soybean (*Glycine max*) using near-infrared spectroscopy and chemometrics. *Journal of Cereal Science, 64*, 43–47. https://doi.org/10.1016/j.jcs.2015.04.004

Font, R., Del Río, M., Fernández-Martínez, J. M., & De Haro-Bailón, A. (2004). Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). *Journal of Agricultural and Food Chemistry, 52*(11), 3563–3569. https://doi.org/10.1021/jf0307649

Gad, H. A., El-Ahmady, S. H., Abou-Shoer, M. I., & Al-Azizi, M. M. (2013). Application of chemometrics in authentication of herbal medicines: A review. *Phytochemical Analysis, 24*(1), 1–24. https://doi.org/10.1002/pca.2378

Gendrin, C., Roggo, Y., & Collet, C. (2008). Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review. *Journal of Pharmaceutical and Biomedical Analysis, 48*(3), 533–553. https://doi.org/10.1016/j.jpba.2008.08.014

Hacisalihoglu, G., Larbi, B., & Mark Settles, A. (2010). Near-infrared reflectance spectroscopy predicts protein, starch, and seed weight in intact seeds of common bean (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry, 58*(2), 702–706. https://doi.org/10.1021/jf9019294

Hansen, J., & Møller, I. (1975). Percolation of starch and soluble carbohydrates from plant tissue for quantitative determination with anthrone. *Analytical Biochemistry, 68*(1), 87–94. https://doi.org/10.1016/0003-2697(75)90682-X

Jiang, H., & Lu, J. (2018). Using an optimal CC-PLSR-RBFNN model and NIR spectroscopy for the starch content determination in corn. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 196*, 131–140. https://doi.org/10.1016/j.saa.2018.02.017

Jukanti, A. K., Gowda, C. L. L., Rai, K. N., Manga, V. K., & Bhatt, R. K. (2016). Crops that feed the world 11. Pearl millet (*Pennisetum glaucum* L.): An important source of food security, nutrition and health in the arid and semi-arid tropics. *Food Security, 8*(2), 307–329. https://doi.org/10.1007/s12571-016-0557-y

Kaur, B., Sangha, M. K., & Kaur, G. (2017). Development of near-infrared reflectance spectroscopy (NIRS) calibration model for estimation of oil content in Brassica juncea and Brassica napus. *Food Analytical Methods, 10*(1), 227–233. https://doi.org/10.1007/s12161-016-0572-9

Kumar, S., Hash, C. T., Singh, G., Basava, R. K., & Srivastava, R. K. (2020). Identification of polymorphic SSR markers in elite genotypes of pearl millet and diversity analysis. *Ecological Genetics and Genomics, 14*, 100051. https://doi.org/10.1016/j.egg.2019.100051. June 2019.

Kumar, S., Parekh, M. J., Patel, C. B., Zala, H. N., Sharma, R., Kulkarni, K. S., Fougat, R. S., Bhatt, R. K., & Sakure, A. A. (2016). Development and validation of EST-derived SSR markers and diversity analysis in cluster bean (*Cyamopsis tetragonoloba*). *Journal of Plant Biochemistry and Biotechnology, 25*(3), 263–269. https://doi.org/10.1007/s13562-015-0337-3

Li, X., Zhang, L., Zhang, Y., Wang, D., Wang, X., Yu, L., Zhang, W., & Li, P. (2020). Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. *Trends in Food Science & Technology, 101*(May), 172–181. https://doi.org/10.1016/j.tifs.2020.05.002

Masithoh, R. E., Lohumi, S., Yoon, W. S., Amanah, H. Z., & Cho, B. K. (2020). Development of multi-product calibration models of various root and tuber powders by fourier transform near infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents. *Heliyon, 6*(10), Article e05099. https://doi.org/10.1016/j.heliyon.2020.e05099

McCleary, B. V., Charmier, L. M. J., & McKie, V. A. (2019). Measurement of starch: Critical evaluation of current methodology. *Starch – Stärke, 71*(1–2). https://doi.org/10.1002/star.201800146

McCleary, B. V., McLoughlin, C., Charmier, L. M. J., & McGeough, P. (2020). Measurement of available carbohydrates, digestible, and resistant starch in food ingredients and products. *Cereal Chemistry, 97*(1), 114–137. https://doi.org/10.1002/cche.10208

McCleary, B. V., Sloane, N., & Draga, A. (2015). Determination of total dietary fibre and available carbohydrates: A rapid integrated procedure that simulates *in vivo* digestion. *Starch - Stärke, 67*(9–10), 860–883. https://doi.org/10.1002/star.201500017

McKie, V. A., & McCleary, B. V. (2016). A novel and rapid colorimetric method for measuring total phosphorus and phytic acid in foods and animal feeds. *Journal of AOAC International, 99*(3), 738–743. https://doi.org/10.5740/jaoacint.16-0029

Pande, R., & Mishra, H. N. (2015). Fourier Transform Near-Infrared Spectroscopy for rapid and simple determination of phytic acid content in green gram seeds (*Vigna radiata*). *Food Chemistry, 172*, 880–884. https://doi.org/10.1016/j.foodchem.2014.09.049

Perez, C. M., & Juliano, B. O. (1978). Modification of the simplified amylose test for milled rice. *Starch - Stärke, 30*(12), 424–426. https://doi.org/10.1002/star.19780301206

Plans, M., Simó, J., Casañas, F., Sabaté, J., & Rodriguez-Saona, L. (2013). Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: Comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. *Food Research International, 54*(2), 1643–1651. https://doi.org/10.1016/j.foodres.2013.09.003

Prieto, N., Pawluczyk, O., Dugan, M. E. R., & Aalhus, J. L. (2017). A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Applied Spectroscopy, 71*(7), 1403–1426. https://doi.org/10.1177/0003702817709299

Ramya, A. R., Ahamed, M. L., Satyavathi, C. T., Rathore, A., Katiyar, P., Raj, A. G. B., Kumar, S., Gupta, R., Mahendrakar, M. D., Yadav, R. S., & Srivastava, R. K. (2018). Towards defining heterotic gene pools in pearl millet [*Pennisetum glaucum* (L.) R. Br.]. *Frontiers of Plant Science, 8*, 1934. https://doi.org/10.3389/fpls.2017.01934

Sáez-Plaza, P., Michałowski, T., Navas, M. J., Asuero, A. G., & Wybraniec, S. (2013). An overview of the kjeldahl method of nitrogen determination. Part I. Early history, chemistry of the procedure, and titrimetric finish. *Critical Reviews in Analytical Chemistry, 43*(4), 178–223. https://doi.org/10.1080/10408347.2012.751786

Shi, H., Lei, Y., Louzada Prates, L., & Yu, P. (2019). Evaluation of near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy techniques combined with chemometrics for the determination of crude protein and intestinal protein digestibility of wheat. *Food Chemistry, 272*, 507–513. https://doi.org/10.1016/j.foodchem.2018.08.075. August 2018.

Shukla, A., Srivastava, N., Suneja, P., Yadav, S. K., Hussain, Z., Rana, J. C., & Yadav, S. (2018). Genetic diversity analysis in buckwheat germplasm for nutritional traits. *Indian Journal of Experimental Biology, 56*(11), 827–837.

Thangavel, K., & Dhivya, K. (2019). Determination of curcumin, starch and moisture content in turmeric by Fourier transform near infrared spectroscopy (FT-NIR). *Engineering in Agriculture, Environment and Food, 12*(2), 264–269. https://doi.org/10.1016/j.eaef.2019.02.003

Tian, W., Chen, G., Zhang, G., Wang, D., Tilley, M., & Li, Y. (2020). Rapid determination of total phenolic content of whole wheat flour using near-infrared spectroscopy and chemometrics. *Food Chemistry*, 128633. https://doi.org/10.1016/j.foodchem.2020.128633. September.

Williams, P. (2014). The RPD statistic: A tutorial note. *NIR News, 25*(1), 22–26. https://doi.org/10.1255/nirn.1419

Williams, P., Dardenne, P., & Flinn, P. (2017). Tutorial: Items to be included in a report on a near infrared spectroscopy project. *Journal of Near Infrared Spectroscopy, 25*(2), 85–90. https://doi.org/10.1177/0967033517702395

Wu, Y., Peng, S., Xie, Q., Han, Q., Zhang, G., & Sun, H. (2019). An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems, 185*(January), 114–121. https://doi.org/10.1016/j.chemolab.2019.01.005

Wu, J. G., & Shi, C. H. (2004). Prediction of grain weight, brown rice weight and amylose content in single rice grains using near-infrared reflectance spectroscopy. *Field Crops Research, 87*(1), 13–21. https://doi.org/10.1016/j.fcr.2003.09.005

Yang, X. S., Wang, L. L., Zhou, X. R., Shuang, S. M., Zhu, Z. H., Li, N., Li, Y., Liu, F., Liu, S. C., Lu, P., Ren, G. X., & Dong, C. (2013). Determination of protein, fat, starch, and amino acids in foxtail millet [*Setaria italica* (L.) Beauv.] by Fourier transform

near-infrared reflectance spectroscopy. *Food Science and Biotechnology, 22*(6), 1495–1500. https://doi.org/10.1007/s10068-013-0243-1

Zeng, L., & Chen, C. (2018). Simultaneous estimation of amylose, resistant, and digestible starch in pea flour by visible and near-infrared reflectance spectroscopy. *International Journal of Food Properties, 21*(1), 1129–1137. https://doi.org/10.1080/10942912.2018.1485027

Zhang, K., Zhou, L., Brady, M., Xu, F., Yu, J., & Wang, D. (2017). Fast analysis of high heating value and elemental compositions of sorghum biomass using near-infrared spectroscopy. *Energy, 118*, 1353–1360. https://doi.org/10.1016/j.energy.2016.11.015