Applied Food Research 4 (2024) 100607

Contents lists available at ScienceDirect

Applied Food Research

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

Near infrared reflectance spectroscopy-driven chemometric modeling for predicting key quality traits in lablab bean (*Lablab purpureus* L.) Germplasm

Simardeep Kaur [a], Naseeb Singh [a,*], Ernieca L. Nongbri [a], Mithra T [b], Veerendra Kumar Verma [a], Amit Kumar [a], Tanay Joshi [c], Jai Chand Rana [d], Rakesh Bhardwaj [b,*], Amritbir Riar [c]

[a] *ICAR- Research Complex for North Eastern Hill Region, Umiam, 793103, Meghalaya, India*
[b] *ICAR- National Bureau of Plant Genetic Resources, New Delhi, 110012, India*
[c] *Department of International Cooperation, Research Institute of Organic Agriculture FiBL, Frick, Switzerland*
[d] *The Alliance of Bioversity International & CIAT- India, Office, New Delhi 110012, India*

ARTICLE INFO

Keywords:
Lablab bean
NIRS
Chemometrics
MPLS
Scatter correction
Protein
Phenols
Pre-breeding

ABSTRACT

Lablab bean (*Lablab purpureus* L.) is a multipurpose crop, commonly used for food, feed, and fodder, and its potential as a plant-based meat alternative. Its nutritional diversity, including high protein, starch, and phenolic content, makes it a suitable candidate for nutritional profiling, which is essential for developing nutritionally enhanced varieties. Traditional methods for analyzing its nutritional parameters are labor-intensive, time-consuming, and expensive. This study employs Near-Infrared Reflectance Spectroscopy (NIRS) as a rapid, non-destructive alternative to evaluate 112 Lablab bean genotypes. We developed prediction models for starch, amylose, protein, fat, and phenols using a Modified Partial Least Squares (MPLS) approach, with spectral pre-processing using Standard Normal Variate (SNV) to remove scatter effects and Detrending (DT) to reduce baseline shifts and noise. The models were optimized for derivatives, gap selection, and smoothing, and evaluated using independent test data and key performance metrics including coefficient of determination (R^2), bias, and Residual Prediction Deviation (RPD). The best-performing models were: starch ($R^2 = 0.959$, RPD = 4.57), amylose ($R^2 = 0.737$, RPD = 1.76), protein ($R^2 = 0.911$, RPD = 3.09), fat ($R^2 = 0.894$, RPD = 2.92), and phenols ($R^2 = 0.816$, RPD = 2.36). Statistical tests, including paired *t*-tests, correlation, and reliability analysis, confirmed the robustness of these models. This study presents a first report offering rapid, multi-trait assessment method for evaluating Lablab bean germplasm, demonstrating high predictive accuracy for pre-breeding practices. It has broad applications in developing nutritionally enhanced varieties, supporting plant-based protein alternatives, and optimizing food production processes to meet the growing demand for healthier, sustainable foods.

1. Introduction

The mainstreaming of plant-based meat alternatives is essential in addressing the growing environmental and ethical concerns associated with animal meat production. With the livestock industry contributing significantly to greenhouse gas emissions, deforestation, and water consumption, the shift towards plant-based proteins offers a sustainable solution to reduce the ecological footprint of meat production (Andreani et al., 2023; Jang and Lee, 2024). In countries like India, where nearly 35% of the population adheres to a vegetarian diet, the importance of plant-based meat alternatives is even more pronounced. These alternatives can cater to the dietary preferences of a large segment of the population while supporting global efforts to mitigate climate change and promote sustainable food systems.

Legumes, particularly those that are underutilized but possess significant potential, can play a key role in this transition. Among these, *Lablab purpureus* L., commonly known as the lablab bean, hyacinth bean or Indian bean, emerges as a valuable crop. Lablab bean is traditionally grown and utilized in many regions for its multipurpose uses in food, feed, and fodder. While it has significant potential, it has not yet been widely commercialized or fully integrated into mainstream agriculture. It is an excellent source of protein and carbohydrates including starch, and is rich in antioxidants such as phenolic compounds, thus positioning itself as a future smart food (Kumari et al., 2022; Pandey et al., 2023). The nutritional profile of the lablab bean makes it an ideal candidate for developing plant-based meat products that are not only nutritious but

* Corresponding authors.
E-mail addresses: naseeb501@gmail.com (N. Singh), Rb_biochem@yahoo.com (R. Bhardwaj).

https://doi.org/10.1016/j.afres.2024.100607
Received 21 September 2024; Received in revised form 12 November 2024; Accepted 14 November 2024
Available online 20 November 2024

also appealing to health-conscious consumers. Its adaptability to diverse climates and underutilized potential makes it an ideal candidate for addressing global nutritional needs while promoting sustainable agriculture (Letting et al., 2021; Vishnu and Radhamany, 2022). The increasing interest in plant-based protein sources, along with the bean's nutritional potential, makes it a strong candidate for further research to enhance its commercial value by rapidly selecting genotypes which are nutritionally superior and can meet the growing demand for sustainable, protein-rich foods.

North Eastern Hill (NEH) region of India is a biodiversity hotspot, harboring a vast array of germplasm (Kaur et al., 2024a) with significant variability in key quality traits of lablab bean. The key quality traits including starch, amylose, protein, phenols, and fat are highly significant due to their important roles in defining the nutritional quality and functional properties essential for plant-based meat alternatives. Additionally, they can be effectively measured by advanced spectroscopy-based methods due to the specific wavebands associated with these traits which enables the development of accurate, robust predictive models with high practical applicability. However, to utilize these genotypes in breeding programs and for mainstreaming purposes, it is necessary to assess their quality traits comprehensively Traditional methods for assessing these nutritional traits, such as the Kjeldahl method for protein, enzyme-based kit methods for starch and amylose, Soxhlet extraction for fat, and the Folin-Ciocalteu method for phenols, are widely used for their accuracy. However, these techniques are often expensive, time-consuming, and require specialized equipment and technical expertise, making them less suitable for rapid or large-scale germplasm screening (Kaur et al., 2024b; Padhi et al., 2022). This emphasizes the need for better analytical alternatives, such as Near Infrared Reflectance Spectroscopy (NIRS), which offers a rapid and non-destructive approach to evaluating the chemical composition of plant materials. NIRS works by detecting the absorption of light at specific wavelengths corresponding to C—H, N—H, and O—H bonds, making it a robust tool for analyzing the nutritional content of crops (Bagchi et al., 2016; Bartwal et al., 2023; John et al., 2022). There are several approaches for developing NIRS-based prediction models for key quality traits, with Modified Partial Least Squares (MPLS) being one of the most established methods. MPLS is highly preferred for developing NIRS-based predictive models due to its ability to handle collinearity in spectral data, which is common in NIR spectra (Kaur et al., 2024d; Kondal et al., 2024). It also enhances model accuracy by focusing on the most relevant spectral features, making it robust for small sample sizes and noisy datasets. Furthermore, it efficiently captures the complex relationships between spectral data and nutritional traits, resulting in reliable predictions across diverse germplasm. In addition, spectral preprocessing is important for improving the accuracy and robustness of predictive models by minimizing noise, baseline variations, and scattering effects inherent in raw spectral data. These variations can obscure important spectral features and lead to inaccurate predictions, particularly when working with complex biological samples. By applying techniques like Standard Normal Variate (SNV) and Detrending (DT), we can correct for scatter and baseline shifts, ensuring that the predictive models focus on the relevant chemical information in the spectra (John et al., 2023; Tomar et al., 2021b). Numerous studies have effectively developed NIRS-based predictive models utilizing the spectral pre-processing methods such as SNV and DT along with the MPLS approach for key nutritional traits (John et al., 2023; Kaur et al., 2024b; Padhi et al., 2022; Tomar et al., 2021b).

Though NIRS-based prediction models using MPLS have been developed for important crops such as rice (John et al., 2022), pearl millet (Tomar et al., 2021b), cowpea (Padhi et al., 2022), perilla (Kaur et al., 2024b), mungebean (Bartwal et al., 2023), buckwheat and amaranthus (Shruti et al., 2023), and rice bean (Kaur et al., 2024c), no studies have yet focused on developing NIRS-driven technology for the rapid assessment of key nutritional quality traits in lablab bean. Therefore, the present study focuses on protein, starch, amylose, fat, and total phenols because these components exhibit distinct absorption features in the near-infrared region, associated with molecular vibrations, such as C—H, O—H, and N—H stretching and bending. These vibrations generate overtones and combinations that NIRS can effectively capture, enabling non-destructive, rapid prediction. While other components like dietary fiber or minerals are relevant, their indirect interaction with NIR spectra makes them less suitable for precise prediction using this technique. Therefore, the present study aims to develop NIRS-based prediction models for these five essential nutritional traits using the Modified Partial Least Squares (MPLS) approach. These models were compared using global metrics and their performance was validated with independent data. The developed models can significantly accelerate pre-breeding practices and facilitate the rapid assessment of the extensive germplasm available in national and international repositories. Ultimately, this research paves the way for advancing the integration of plant-based meat alternatives into the mainstream, thus contributing to a more sustainable and nutritious food system.

2. Materials and methods

2.1. Experimental conditions and sample collection

Experiments were conducted using 112 lablab bean seed genotypes, which included local collections, landraces, and accessions from various regions of North Eastern India, specifically Assam, Manipur, Mizoram, Nagaland, Tripura, and Meghalaya. The field experiment was carried out during the 2023–24 growing season (July to February) at the Horticulture Experimental Farm, ICAR-Research Complex for NEH Region, Umiam, Meghalaya, India (Latitude: 25.6506° N, Longitude: 91.8853° E). The plants were arranged in a randomized block design to reduce experimental error and ensure statistical validity. They were grown under terrace conditions, with a plot size of 3.0 × 2.0 m and a spacing of 90.0 × 60.0 cm. Standard agronomic practices including plant protection measures were followed to support optimal plant growth. Seeds were harvested at physiological maturity, sun-dried, hand-cleaned, and further dried in an oven before storage. Approximately 15 g of the dried seeds were ground, homogenized, and sieved through a 1 mm sieve using the Foss Cyclotec™ 1093 Sample Mill (FOSS Analytical, Denmark) to obtain a fine flour, which was then subjected to nutritional analysis. All parameters were analyzed in triplicate, and the mean values were used for model calibration and validation. The overall methodology of the present study is provided in Fig. 1.

2.2. Nutritional parameters estimation

2.2.1. Sample preparation for starch and phenols estimation

Homogenized samples (100 ± 5 mg) were vortexed with 5 mL of 80% ethanol and heated at 80 °C for 30 min. After cooling, samples were mounted on a rotator for 60 min, centrifuged (Model: PR-23; Remi elektrotechnik limited, Vasai-401,208, India) at 16,000 g for 15 min, and the supernatant was extracted two more times using the same method. The final volume of the supernatant was adjusted to 10 mL for total phenols estimation while the pellet was used for starch quantification.

2.2.2. Starch

Total starch was determined using a Megazyme assay kit (TOTAL STARCH (100A) K-TSTA-100A, Megazyme) following AACC 76–13.01 and AOAC 996.11 with slight modifications. The obtained pellet was treated with 200 μL of 80% ethanol and heated in a boiling water bath for 5 min, followed by hydrolysis with α-amylase and amyloglucosidase, converting starch to d-glucose. Glucose was detected at 510 nm using GOPOD reagent, and starch content was calculated as g/100 g DWB using the formula:

$$\text{Starch } \% = \Delta A \times F \times EV \times D/W \times 0.90$$

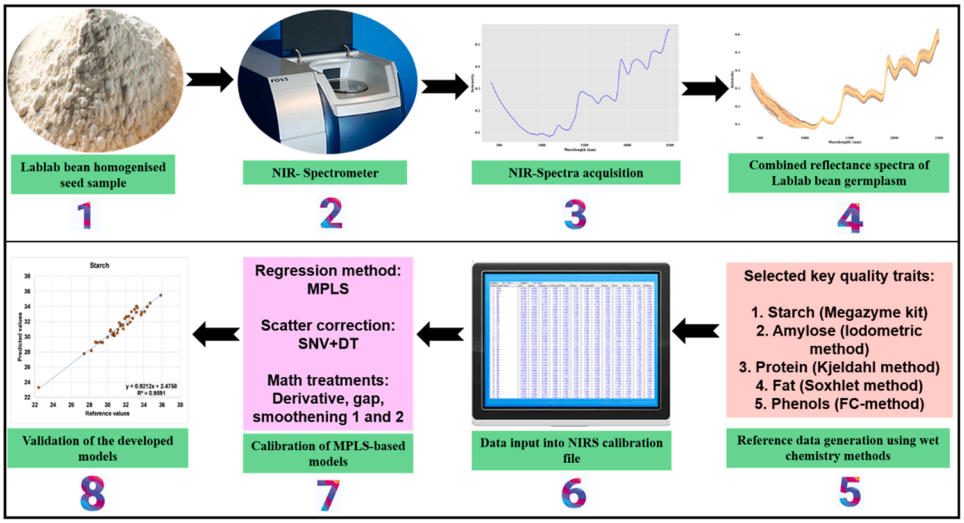

Fig. 1. Overall methodology of the present study.

Where:

- ΔA: Sample absorbance
- F: Factor (GOPOD absorbance for 100 μg glucose)
- EV: Extraction volume (100 mL)
- D: Dilution factor
- W: Sample weight (mg)

2.2.3. *Amylose*

Amylose content was determined using an improved iodometric method (Perez and Juliano, 1978). Briefly, 50 mg of homogenized flour was mixed with 500 μL ethanol in triplicate, vortexed, and heated in a boiling water bath for 20 min. The mixture was then transferred to a 50 mL volumetric flask and diluted with double-distilled water. From this, 500 μL was taken into amber tubes, followed by the addition of 100 μL glacial acetic acid (Qualigens, Q21057, 99.5%) and 200 μL iodine solution. The volume was adjusted to 10 mL with double-distilled water, incubated at room temperature for 20 min, and the absorbance was measured at 620 nm. Amylose content was expressed as g/100 g DWB, using a standard potato amylose (Sigma, 1,002,922,486) curve for quantification.

2.2.4. *Protein*

The Kjeldahl method (AOAC 984.13) was used to estimate the total nitrogen content in the samples. This process involves digesting the sample in sulfuric acid (Avantor, S0530, 98.08%) with a catalyst, distilling the released ammonia into a boric acid solution (Sisco Research Laboratories Pvt. Ltd., 80,266, 99.5%), and titrating it with standard acid to determine the nitrogen content, which is then used to calculate the total protein. The nitrogen (%N) was then converted to percent protein using a conversion factor of 6.25.

2.2.5. *Fat*

The fat content of lablab bean seed flour was estimated using Soxhlet extraction with petroleum ether (Petroleum ether 40–60 °C extra pure AR, Sisco Research Laboratories Pvt. Ltd., 26,440) as the solvent. Three grams of dried seed flour (in triplicates) were placed in thimbles, and the initial weight (W1) was recorded. After oil extraction using the Soxhlet apparatus, the samples were dried until a constant weight (W2) was achieved. The fat content was calculated as ((W2 - W1)/W) *100, where W is the sample weight.

2.2.6. *Total phenols*

Total phenols were determined using a modified method (Tian et al., 2021). A 500 μL extract was evaporated in a water bath at 100 °C, then reconstituted with 3 mL double-distilled water and vortexed. A blank was prepared similarly. Gallic acid standards (0.01–0.05 mg) (Fisher Scientific, 410,862,500,98%) were prepared separately. To each sample, blank, and standard, 500 μL of Folin-Ciocalteu reagent (Folin & Ciocalteus Phenol Reagent AR 2.0N, Sisco Research Laboratories Pvt. Ltd., 39,520) and 2 mL of 20% Na_2CO_3 (Glasil, GSI/GSE258877082020, 99.9%) were added, followed by a 1-hour incubation at room temperature. Absorbance was measured at 650 nm against the blank. Total phenolic content was expressed as GAE g/100 g, using a gallic acid standard curve for quantification.

3. NIR-spectra acquisition

For the acquisition of NIR spectra, homogenized samples of lablab bean seed flour were prepared to ensure standardized conditions. The samples were dried and equilibrated at room temperature (25 °C) for a period of 6 h to standardize temperature and moisture levels, both of which are critical factors that can influence the absorbance and reflectance of NIR waves. To ensure precise and reliable measurements, the NIR spectrometer was calibrated at 20-minute intervals using a reference sample to ensure consistent accuracy throughout the scanning process. Approximately 5.0 gs of the homogenized flour were carefully loaded into a circular ring cup fitted with a quartz window, measuring 3.8 cm in diameter and 1.0 cm in thickness. To achieve uniform packing without any air pockets, a circular cardboard backing was gently pressed onto the samples. The spectral data were collected using a FOSS NIRS DS3 spectrometer (FOSS Nils Foss Alle 1, DK-3400, Hilleroed, Denmark), covering a wavelength range of 400 to 2490 nm. Each spectrum represented the average of 32 scans recorded at 2 nm intervals and was expressed as log (1/R), where R denotes relative reflectance. Following the acquisition, the spectra were extracted using Win ISI Project Manager Software version 1.61 for subsequent analysis.

4. Calibration of NIRS-based predictive models

After obtaining the spectral and reference data for all samples, the reference data was integrated into the NIR file. The dataset was then divided into two subsets based on diversity and homogeneity: a calibration set (75% of the samples), referred to as the "cal" file, and a validation set (25% of the samples), referred to as the "val" file. In the calibration file, Modified Partial Least Squares (MPLS) regression with cross-validation was applied. Preprocessing techniques such as Standard Normal Variate (SNV) and Detrending (DT) were used to correct scatter

effects, aiming to reduce particle size and light path variability. Spectral derivatives were calculated to correct for overlapping absorption bands and baseline shifts to improve the accuracy of the analysis. NIRS calibrations were developed for the spectral range between 400 and 2490 nm through an iterative process that applied various mathematical treatments. These treatments included combinations of derivatives, gaps, and smoothing parameters. For example, in the configuration "2,4,6,2″, "2″ indicates the second derivative, which helps correct overlapping bands and baseline shifts, while "4″ specifies the gap, denoting four data points used in the second derivative calculation. The "6″ and "2″ represent the number of data points for the first and second smoothing steps, respectively. Cross-validation was employed under the scatter correction methods SNV and DT to prevent overfitting and ensure the robustness of the models. Key statistical parameters, including range, standard deviation (SD), standard error for cross-validation (SEC), and the coefficient of determination for internal validation (RSQ internal), were evaluated using Win ISI Project Manager Software version 1.61. Models with lower SEC and higher RSQ values were considered superior. Additionally, the standard error of cross-validation (SEC(V)) and the 1 minus variance ratio (1-VR) were calculated to assess error and cross-validation performance. The mathematical preprocessing treatments were refined iteratively through trial and error, aiming to reduce SEC(V) and increase 1-VR during cross-validation, thereby enhancing the accuracy and reliability of the final predictive models.

5. Validation of the developed models

A range of performance metrics was used to evaluate the predictive accuracy of the models using an independent dataset (val file). The Coefficient of Determination (R^2 or RSQ) quantified how well the model explained the variance in protein content, with values closer to 1.0 indicating better model fit (Eq. (1)). Bias measured systematic errors by calculating the mean difference between reference and predicted values, with lower values reflecting higher accuracy (Eq. (2)). The Corrected Standard Error of Prediction (SEP(C)) assessed prediction precision, with lower values indicating improved accuracy after accounting for bias (Eq. (3)). Lastly, the Residual Prediction Deviation (RPD) evaluated model robustness, where higher values suggested stronger predictive performance (Eq. (4)).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

$$SEP(C) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i - bias)^2}{n}} \quad (3)$$

$$RPD = \frac{StandardDeviation(SD)\ of\ Observed\ Values}{SEP\ (C)} \quad (4)$$

6. Quality control

In this study, a completely randomized design was used to ensure that the spectral acquisition and estimation of nutritional traits were evenly and randomly distributed across the 112 sample units. These samples represented a broad spectrum of nutritional trait levels, from the lowest to the highest, covering the complete diversity. Prior to estimating the nutritional content of the entire germplasm, we standardized our protocols through a pilot study involving 4 randomly selected samples, which allowed us to refine our procedures before proceeding with the full set of samples. Each sample was scanned twice on NIR to identify and correct any potential spectral anomalies, with the average spectrum from these scans being utilized in the subsequent data analyses. Nutritional traits were measured in triplicate to enhance both accuracy and reproducibility, with mean values being used for both the calibration and validation datasets. To ensure the robustness of the developed models, equal representation of samples with low, median, and high nutritional trait levels was maintained in both the calibration and validation sets. This strategy was important to avoid any bias that could lead to underestimation or overestimation of nutritional content across the diverse sample range.

7. Statistical analysis

A paired sample *t*-test was conducted to compare the reference and predicted values at a 95% confidence interval using IBM SPSS version 17.3. Additionally, strict parallel analysis was employed to assess the reliability of the developed models. The reliability score between the predicted values and laboratory-validated samples was also calculated using the same software. Furthermore, correlation analysis was performed to evaluate the relationship between the predicted and reference values, thus providing further insight into the predictive accuracy of the model and alignment with the actual measured data.

8. Results and discussion

8.1. Nutritional composition analysis

Table 1 presents the descriptive statistics for five nutritional traits; starch, amylose, protein, fat, and phenols in 112 lablab bean samples. We observed considerable variability across these traits (presented in a g/100 g dry basis); for instance, starch content ranged from 22.416 to 35.879, with a mean of 31.466, while amylose varied between 13.140 and 16.118, with an average of 14.813. Protein content showed a wide range from 20.987 to 27.145, with a mean value of 24.464. Fat content also exhibited significant variation, ranging from 1.328 to 3.751, with a mean of 2.138. Phenols, though present in smaller quantities, ranged from 0.091 to 0.470, with an average of 0.242. This variability indicates a rich diversity within the germplasm, offering the potential for selecting genotypes with favorable nutritional profiles. This variability can be attributed to several factors, primarily genotype-environment interactions, where different genotypes respond uniquely to environmental conditions. Additionally, inherent genetic diversity within the germplasm, variations in seed maturity at harvest, post-harvest handling, soil nutrient availability, water stress, and other agronomic practices may contribute (Baye et al., 2011; Boye et al., 2024; Kaur et al., 2024b).

8.2. NIR-spectra analysis

Fig. 2 displays the combined NIR spectra of homogenized flour samples from 112 diverse genotypes of lablab bean germplasm, measured across a wavelength range of 400–2490 nm (corresponding to 25,000–4016 cm^{-1}). It is challenging to visually distinguish different regions within the NIR spectra due to the highly overlapping and broad combination bands arising from fundamental vibrational modes (Cozzolino, 2015). This difficulty is amplified in biological materials like lablab bean, which have a complex structural matrix with hydrogen bonding between proteins, starch, amylose, fats, and other biomolecules. These overlapping absorption peaks correspond to combinations and overtones of vibrational modes, particularly N–H, O–H, and C–H, which are associated with proteins, fatty acids, and carbohydrates, respectively. In the spectral region between 2000 and 2222 nm, the peaks are attributed to combinations of C–O and N–H stretching, which are linked to protein content (Kaur et al., 2024b; Plans et al., 2013; Tomar et al., 2021b). Similarly, this region (2000–2222 nm) is also related to O–H groups found in proteins and phenols. A broad peak observed around 1428–1471 nm is associated with the first overtone of O–H stretching in hydroxyl phenol groups and C–H combinations in

Table 1
Descriptive statistics of Lablab bean germplasm collection ($N = 112$) for 5 nutritional traits.

Trait	Starch	Amylose	Protein	Fat	Phenols
Mean	31.466	14.813	24.464	2.138	0.242
Min	22.416	13.140	20.987	1.328	0.091
Max	35.879	16.118	27.145	3.751	0.470
Median	31.546	14.825	24.318	2.105	0.243
SD	2.091	0.558	1.106	0.353	0.063
Range	22.416–35.879	13.140–16.118	20.987–27.145	1.328–3.751	0.091–0.470

*All values are presented in g/100 g of dry basis.

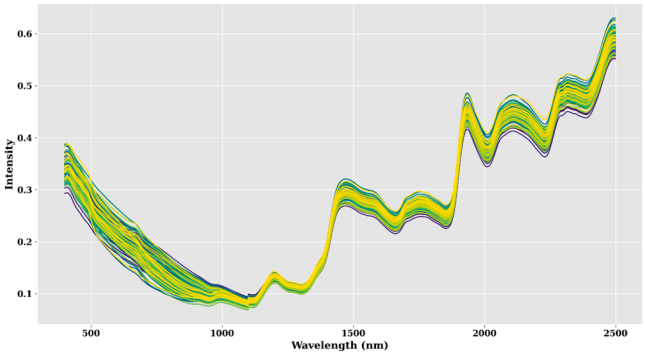

Fig. 2. Combined NIR spectra of lablab bean germplasm ($N = 112$).

aromatic compounds. The absorption peak near 1923 nm is identified with the bending/stretching vibrations of O–H in polysaccharides, which overlaps with water absorption. Furthermore, peaks around 2083 nm (4800 cm^{-1}) are related to the third overtone of polysaccharides' asymmetric C–O–O stretching (Zhang et al., 2017). The sharp absorption band near 1928 nm (5184 cm^{-1}) is linked to the combination of bending and stretching vibrations of O–H in amylose, while the region near 1463 nm (6835 cm^{-1}) corresponds to the symmetric stretching of O–H in amylose. Additionally, peaks around 5800 cm^{-1} are associated with C–O bonds in fats, and the region near 4332 cm^{-1} is related to the second overtone of C–H bending in fats (Kaur et al., 2017; Tomar et al., 2021b). Similar peaks, linked to these functional groups, were observed in several other crops (Bagchi et al., 2016; Bartwal et al., 2023; Chen et al., 2013; John et al., 2022, 2023; Padhi et al., 2022; Tomar et al., 2021b; Zhang et al., 2017). The overlap between NIR bands associated with different traits creates complexities in understanding the relationships between a trait and its corresponding wavelength. We also observed that the presence of absorption bands across multiple spectral regions for a single trait further complicates spectral interpretation, especially for biological samples with identical chemical bonds. For example, the broad absorption region of O–H bonds near 2100 nm can obscure protein amide bond absorption, making trait-specific prediction challenging (Cem Ömer; Kahriman, 2012). These factors may hinder a model's prediction accuracy for traits associated with multiple wavelengths.

8.3. Calibration of models and spectra pre-processing

To develop the robust calibration equations, all samples were first ordered by their 4 nutritional trait values to avoid bias in sub-set divisions. The calibration and validation sets were then selected to cover the full range of concentrations and diversity linked to selected traits. The samples were split (in a ratio of 2:1) into an internal calibration set ($N = 74$) for model training and an external validation set ($N = 38$) to evaluate the performance of the model. Full-range spectral data were used to develop regression models by using techniques like PLS, MPLS, and PCR (using WinISI v1.61 software), which are commonly used for NIR model development. Though we have tested all three methods, MPLS proved to be more stable and accurate than standard PLS. MPLS is a robust statistical approach designed to predict dependent variables based on independent variables especially useful when the predictors are highly collinear or when their number exceeds the observations (Kaur et al., 2024c). MPLS extracts orthogonal components that capture maximum variance in the independent variables while maintaining a strong correlation with the target variable. This dimensionality reduction improves model robustness and accuracy (Khatri et al., 2021). In MPLS, the residuals at each wavelength are standardized after each factor is calculated to allow for better decomposition of the spectroscopic data. This process reduces the impact of irrelevant spectral variations and enhances the calibration by balancing biochemical and spectral information. Thus, MPLS refines the standard PLS approach by normalizing the covariance between the spectral data (x) and the target variable (y) with the covariance of spectral data with itself (4), enhancing the weighting of important wavelengths (Kondal et al., 2024; Westerhaus, 2014). This ensures that the most relevant spectral features are prioritized to enhance model accuracy and reduce multicollinearity (Murphy et al., 2022). Consequently, MPLS was selected as the most suitable method for this study due to its superior handling of spectral and reference data correlations.

NIR spectroscopy faces significant challenges due to interference from factors such as molecular vibrations, light scattering, and path length variations, which can lead to complex spectral distortions like baseline shifts, curvature, and non-linearities. These variations, caused by interactions between light and sample particles result in changes to absorption levels, making linear calibration and spectral interpretation difficult (Beć et al., 2021; Padhi et al., 2022). Path length variations from light scattering generate background signals that fluctuate with wavelength, further complicating spectral analysis. To address these issues, pre-processing of spectral data is essential for developing reliable prediction models. This enhances the signal-to-noise ratio, increases signal variation, and eliminates irrelevant factors unrelated to the property of interest. Common empirical methods for pre-processing include derivatives, multiplicative scatter correction, standard normal variate (SNV), and detrending (DT) (Padhi et al., 2022). SNV normalizes the spectra by removing multiplicative scatter effects, centering each spectrum around its mean and standardizing the values, which is particularly effective in correcting for light scattering and particle size variation (Bi et al., 2014; Wu et al., 2019). Detrending removes baseline shifts and non-linear trends by fitting a polynomial curve to the spectral data and subtracting it, thus eliminating background noise and systematic baseline variations (John et al., 2022; Mills, 2011). Therefore, in our study we used a combination of SNV and DT in our pre-processing steps to develop reliable and robust prediction models (Fig. 3). For the development of calibration equations, the following mathematical treatments (in the sequence of derivative, gap, smoothing 1, and smoothing 2) were identified as the best-performing for each parameter: starch (2,6,6,1), amylose (2,4,4,1), protein (3,6,6,1), fat (2,4,6,1), and phenols (2,4,8,1). For instance, in case of starch (2,6,6,1), '2′ indicates the derivative order, '6′ the gap size, '6′ the first smoothing, and '1′ the second smoothing, respectively. These treatments were selected based on a combination of the highest values for 1-VR (VR= Variance) and RSQ, along with the lowest SEC and SEP(C) values. To eliminate background noise and enhance spectral resolution, second and third derivatives were applied. These derivatives act as high-pass filters and remove low-frequency

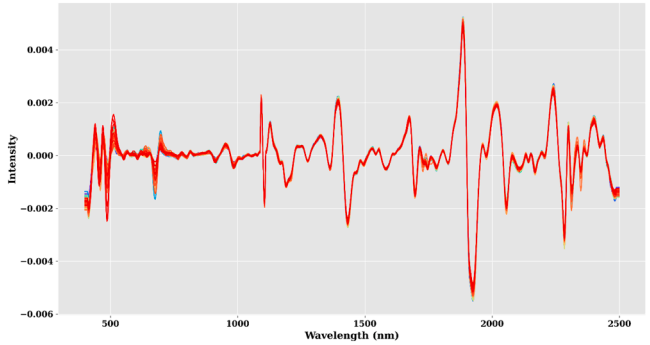

Fig. 3. NIR-spectra (2nd derivative) of Lablab bean germplasm after pre-processing using Standard Normal Variate (SNV) and Detrending (DT) techniques.

background signals and baseline variations, which often obscure weaker spectral features that are critical for analysis. By improving the clarity of minor peaks that are not discernible in the original spectrum, the derivatives significantly improve calibration performance. Additionally, the use of gaps (4, 6) and smoothing techniques (S1–4,6,8, S2–1) was used to reduce noise generated by high-frequency disturbances. This smoothing process ensured that erratic spectral fluctuations did not interfere with the accurate determination of key parameters. Together, these mathematical treatments provided a robust foundation for the development of precise calibration models for various quality traits.

8.4. Validation of the models using independent test data

To validate the developed models, an entirely independent dataset ($N = 38$) was used. The selection of the best-fit models was based on several key criteria, including higher $RSQ_{external}$ and RPD values, along with lower SEP, slope, and bias values. The mean, SD, minimum, maximum values, and validation metrics for both reference and predicted data are presented in Table 2 for all 5 traits. The close agreement between actual and predicted values, with minimal differences observed across these metrics, indicates the reliability of the developed models. For instance, in case of starch, the reference mean value was 31.410, while the predicted mean value was 31.409, showing an almost perfect match between the actual and predicted results. In case of protein, the reference mean value was 24.427, compared to the predicted mean of 24.458, with only a minor difference. This similarity, particularly in mean and SD, suggests that the models can accurately predict the key parameters without significant deviation from the reference laboratory values, further validating their robustness.

$RSQ_{(external)}$, or the coefficient of determination (R^2), is a statistical measure used to evaluate how well the predicted values from a model match the actual values. It ranges from 0 to 1, with values closer to 1 indicating a stronger correlation between the predicted and actual data, meaning the model can explain a larger portion of the variance in the data. In the present study, for starch, the RSQ was very high at 0.959, indicating an excellent fit between the actual and predicted values. For amylose, the RSQ was 0.737, showing a moderate fit but still reasonable for practical use. For protein, the RSQ was 0.911, signifying a strong correlation and reliable prediction. The model for fat also performed well, with an RSQ of 0.894, while in the case of phenols, the RSQ was 0.816, suggesting good predictive ability for this parameter as well (Fig. 4). Bias indicates the systematic difference between the predicted and actual values, with an ideal bias being close to zero, meaning no consistent over- or under-prediction (Wu et al., 2019). In our study, the bias values were nearly negligible: starch (0.001), amylose (−0.125), protein (−0.030), fat (−0.002), and phenols (−0.002). These low bias values suggest that the models have minimal systematic error, with amylose showing a slightly higher under-prediction bias compared to the other parameters, which exhibit almost no bias at all. SEP(C), or Standard Error of Prediction corrected for bias, measures the model's predictive accuracy by accounting for both random error and any systematic bias. A lower SEP(C) value indicates better predictive performance. In our study, the SEP(C) values were: starch (0.494), amylose (0.318), protein (0.355), fat (0.136), and phenols (0.030) (Table 2).

Residual Prediction Deviation (RPD), is the ratio of the standard deviation of reference data to the SEP(C), and it provides a more robust measure of the model's predictive performance. The ideal RPD value should exceed 1. An RPD between 1.5 and 2 suggests the model can distinguish between high and low values of the response variable. Values ranging from 2 to 2.5 indicate the ability of the model to provide approximate quantitative predictions. RPD values of 2.5 or higher represent good predictive accuracy, while values of 3 or above signify excellent model performance (Williams et al., 2017). In our study, the RPD values were: starch (4.57), amylose (1.76), protein (3.09), fat (2.92), and phenols (2.36) (Table 2). The high RPD for starch and protein indicates strong predictive accuracy, while the values for fat and phenols suggest good models for germplasm screening. Amylose, with an RPD of 1.76, indicates a lower performance but may still be useful for approximate estimations.

Several other studies have reported robust NIRS-based prediction models using the MPLS approach, with their results aligning closely with ours in terms of RSQ and RPD values for key traits. For example, in our study, the starch prediction model achieved an R^2 of 0.959 and an RPD of 4.57, indicating high accuracy and reliability (Table 2). These values surpass those reported by (Tomar et al., 2021b) ($R^2 = 0.915$, RPD = 2.71) for pearl millet and as well as (John et al., 2022) ($R^2 = 0.820$, RPD = 2.12) for rice, but slightly lower than those reported by (Padhi et al., 2022) ($R^2 = 0.997$, RPD = 5.32) for cowpea The differences in RPD values could be attributed to varying sample sizes, diversity within the germplasm tested, or differences in the spectral ranges used for calibration in each study. Additionally, variations in the chemical

Table 2
External validation metrics of the developed models.

Parameters		Starch	Amylose	Protein	Fat	Phenols
Mean	Reference value	31.410	14.819	24.427	2.151	0.240
	Predicted value	31.409	14.944	24.458	2.153	0.242
SD	Reference value	2.368	0.607	1.168	0.412	0.061
	Predicted value	2.227	0.554	1.085	0.392	0.070
Minimum value	Reference value	22.416	13.140	20.987	1.328	0.091
	Predicted value	23.315	13.504	21.729	1.397	0.042
Maximum value	Reference value	35.879	16.118	27.145	3.751	0.370
	Predicted value	35.497	16.089	27.051	3.894	0.359
Validation metrics	Math treatment	2,6,6,1	2,4,4,1	3,6,6,1	2,4,6,1	2,4,8,1
	RSQ	0.959	0.737	0.911	0.894	0.816
	Slope	1.041	0.942	1.028	0.993	0.794
	Bias	0.001	−0.125	−0.030	−0.002	−0.002
	SEP(C)	0.494	0.318	0.355	0.136	0.030
	RPD	4.57	1.76	3.09	2.92	2.36

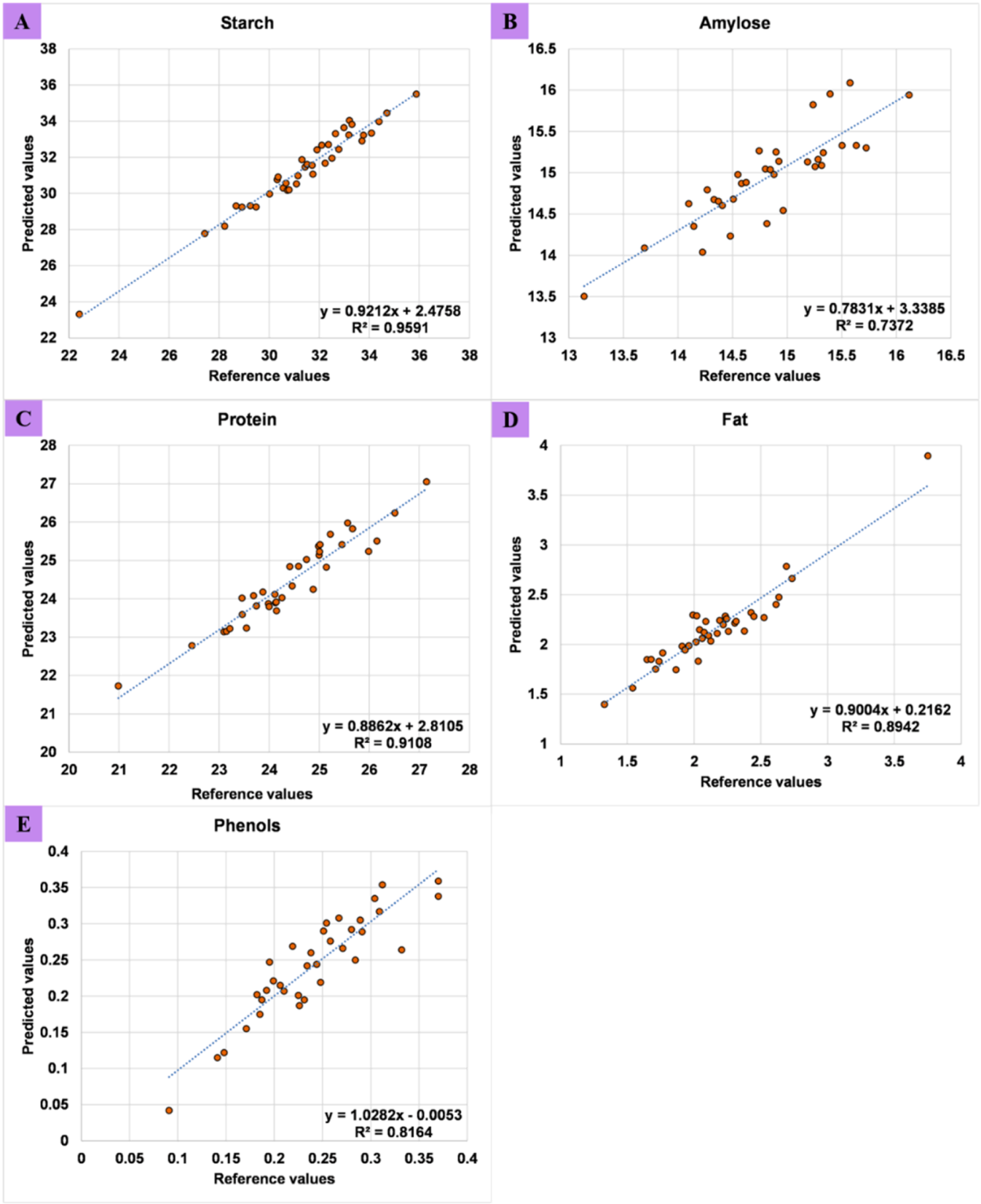

Fig. 4. The figure presents scatter plots showing the relationship between reference and predicted values of key nutritional traits for Lablab bean germplasms. The traits include A: Starch content, B: Amylose content, C: Protein content, D: Fat content, and E: Total phenols, with all values expressed on a g/100 g dry basis.

composition of different crops and geographical locations may contribute to the superior or inferior model performance across studies. For amylose, our model obtained an R^2 of 0.737 and an RPD of 1.76 (Table 2), which, although lower than values in some studies, (Bagchi et al., 2016; John et al., 2022; Tomar et al., 2021b), still demonstrate the model's ability to capture amylose content variation. The relatively lower RPD may be due to the narrower range of amylose content in the Lablab bean germplasm compared to other crops. While our RPD is lower, it offers a foundation for further model refinement, perhaps by increasing sample diversity or improving calibration procedures. The protein prediction model in our study performed well, with an R^2 of 0.911 and an RPD of 3.09, which is consistent with results reported for mung bean (Bartwal et al., 2023) ($R^2 = 0.940$, RPD = 3.84), cowpea (Padhi et al., 2022) ($R^2 = 0.903$, RPD = 2.80), and ricebean ($R^2 = 0.84$, RPD = 2.25) (Kaur et al., 2024c). The slight discrepancies in performance might arise from different methods of protein estimation, differences in the protein content variability within each study's germplasm or due to differences in model calibration techniques used. Despite these differences, our model shows strong predictive reliability for protein content in Lablab bean. For fat, our model achieved an R^2 of 0.894 and an RPD of 2.92, outperforming other legumes such as rice bean and adzuki bean ($R^2 = 0.583$, RPD = 1.1) (John et al., 2023). The better performance of our model could be due to the wider range and diversity of fat content in our calibration and validation datasets along with superior spectral preprocessing techniques (SNV and DT), which helped minimize noise and scatter in the spectra. For phenols, our model recorded an R^2 of 0.816 and an RPD of 2.36, significantly higher than those reported in cowpea ($R^2 = 0.706$, RPD = 1.78) and rice bean ($R^2 = 0.571$, RPD = 1.30) (John et al., 2023; Padhi et al., 2022). This discrepancy could stem from the higher phenolic content variability in our dataset, enabling better model calibration, or differences in extraction methods for phenols, which can impact the NIR spectrum. Lastly, it can be deduced that while the RSQ and RPD values for starch, protein, fat, and phenols in our study are comparable to or better than those reported in similar NIRS-based studies, discrepancies in performance across studies can largely be attributed to variability in sample composition, analysis methods, preprocessing techniques, modelling approach, calibration range, and germplasm diversity as these factors significantly influence the robustness of the predictive models.

Compared to conventional methods, our NIRS-based models offer significant advantages by being more efficient, eco-friendly, and less labor-intensive, allowing for the simultaneous assessment of multiple components in a non-destructive manner. This approach minimizes the need for costly chemical reagents and reduces processing time, making it highly suitable for large-scale applications. The incorporation of spectral pre-processing techniques (SNV and DT) and chemometric methods (MPLS) enhances the reliability and precision of predictions for key nutritional traits. Additionally, these models enable high-throughput screening of Lablab bean germplasm, which is invaluable for plant breeders and researchers focused on developing nutritionally superior varieties. In the food industry, they facilitate rapid assessment of genotypes with desired nutritional profiles. For example, protein-rich genotypes can support the development of plant-based meat alternatives or be used as a high-protein feed for livestock. Starch-rich genotypes are ideal for manufacturing energy-dense foods, such as porridges and flatbreads. On the other hand, genotypes with high amylose and phenol content can be selected for producing low glycemic index foods, catering to the growing demand for healthier diets (John et al., 2022; Tomar et al., 2021b, 2021a).

Furthermore, the application of this model extends to food manufacturing and production optimization, where it can streamline the identification of raw materials with specific nutritional traits, reducing reliance on time-consuming wet chemistry methods. In addition to nutritional profiling, favorable genotypes can be further assessed for agro-morphological traits, accelerating their integration into mainstream agriculture. This approach also enhances breeding efficiency by enabling the early elimination of less promising genotypes and selection of desired chemotypes from diverse backgrounds before further evaluation, thereby reducing breeding cycles and associated costs. Thus, the developed NIR-based models align with the needs of food manufacturing and engineering research by providing a rapid, cost-effective, and sustainable solution for screening, breeding, and optimizing production processes. This approach supports the development of functional foods and nutraceutical products while contributing to the broader goals of sustainable agriculture and global food security.

8.5. Statistical analysis between reference and predicted values

A paired sample *t*-test was performed to evaluate the accuracy of the prediction model by comparing the reference values with the predicted values for key traits. This test helps to determine whether there is a significant difference between the two sets of values, thereby assessing the reliability of the model in predicting the selected parameters. Table 3 presents the paired sample *t*-test results comparing reference (Ref.) and predicted (Pred.) values for starch, amylose, protein, fat, and phenols at a 95% confidence interval. The mean differences between the reference and predicted values are minimal, and all p-values are greater than 0.05, indicating no statistically significant differences. The low SD and standard errors of the mean (SEM) further confirm the accuracy and consistency of the predicted values across the different traits.

In addition to the paired sample *t*-test, we performed correlation and reliability analysis to further assess the model's predictive performance for biochemical parameters in lablab bean germplasm. Table 4 shows high reliability and strong positive correlations between the reference and predicted values across all traits. The correlation coefficients range from 0.859 to 0.979, indicating a strong linear relationship, while the reliability (unbiased) values, ranging from 0.918 to 0.990, confirm the consistency and unbiased nature of the predictions. These results demonstrate the robustness of the model in accurately predicting the biochemical traits.

9. Conclusions

In the present study, we developed Near-Infrared Reflectance Spectroscopy (NIRS)-based prediction models for starch, amylose, protein, fat, and phenols in lablab bean (*Lablab purpureus* L.) using a Modified Partial Least Squares (MPLS) approach. Spectral pre-processing techniques such as Standard Normal Variate (SNV) and Detrending (DT) were applied to improve model accuracy by removing scatter effects and baseline shifts. The models were built on homogenized seed flour and validated using independent test datasets. The best-performing models were: starch ($R^2 = 0.959$, RPD = 4.57), amylose ($R^2 = 0.737$, RPD = 1.76), protein ($R^2 = 0.911$, RPD = 3.09), fat ($R^2 = 0.894$, RPD = 2.92), and phenols ($R^2 = 0.816$, RPD = 2.36). Statistical analyses, including paired sample *t*-test, correlation, and reliability tests, confirmed the robustness of the models. These models provide a rapid and non-destructive approach for screening large germplasm collections, available in national and global repositories and have the potential to accelerate pre-breeding programs aimed at developing nutritionally enriched lablab bean varieties. While this study focused on

Table 3
Paired sample *t*-test at 95% confidence interval.

Pair	Mean	SD	SEM	p-value
Starch (Ref. vs Pred.)	.00024	.49411	.08015	.998
Amylose (Ref. vs Pred.)	−0.12468	.31757	.05446	.069
Protein (Ref. vs Pred.)	−0.03042	.35520	.05920	.611
Fat (Ref. vs Pred.)	−0.00197	.13582	.02203	.929
Phenols (Ref. vs Pred.)	−0.00146	.03029	.00512	.778

*SD= Standard Deviation, SEM= Standard Error of Mean, Ref.- Reference values, Pred.- Predicted values.

Table 4
Reliability (unbiased) test and correlation analysis between reference and predicted values of five nutritional traits in lablab bean germplasm.

Traits	Reliability	Correlation
Starch (Ref. vs Pred.)	0.990	0.979
Amylose (Ref. vs Pred.)	0.918	0.859
Protein (Ref. vs Pred.)	0.977	0.954
Fat (Ref. vs Pred.)	0.974	0.946
Phenols (Ref. vs Pred.)	0.950	0.904

*Ref.- Reference values, Pred.- Predicted values.

homogenized seed flour, future research can explore the development of models using whole grains. Lastly, this study represents the first report of using NIRS and MPLS for rapid, multi-trait screening of lablab bean germplasm, providing a foundation for future research in developing nutritionally enhanced varieties and exploring more advanced modeling techniques.

Ethical statement

The authors confirm that no human or animal subjects were utilized in the experiments conducted for this study.

Funding source

The work was supported by the in-house project of the lead author titled "Nutritional Profiling, Development of NIRS-based Prediction Models, and Genome-wide Association Studies for Key Nutritional Traits in Perilla, Lablab Bean, and Rice Bean." Additionally, it was supported by the International collaborative project" Consumption of Resilient Orphan Crops & Products for Healthier Diets" (CROPS4HD), which is co-funded by the Swiss Agency for Development and Cooperation, Global Programme Food Security (SDC GPFS), and executed in India through FiBL (Research Institute of Organic Agriculture), and Alliance of Bioversity and CIAT with ICAR-NBPGR as a lead partner.

CRediT authorship contribution statement

Simardeep Kaur: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Naseeb Singh:** Writing – original draft, Visualization, Software, Methodology, Data curation. **Ernieca L. Nongbri:** Formal analysis. **Mithra T:** Formal analysis. **Veerendra Kumar Verma:** Writing – review & editing, Resources. **Amit Kumar:** Supervision, Resources. **Tanay Joshi:** Writing – review & editing, Resources. **Jai Chand Rana:** Supervision, Resources, Funding acquisition. **Rakesh Bhardwaj:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Amritbir Riar:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

The authors are thankful to the Directors of ICAR Research Complex for NEH Region, Umiam, and ICAR NBPGR, New Delhi, for providing the necessary facilities to conduct this study.

Data availability

Data will be made available on request.

References

Andreani, G., Sogari, G., Marti, A., Froldi, F., Dagevos, H., & Martini, D. (2023). Plant-based meat alternatives: technological, nutritional, environmental, market, and social challenges and opportunities. *Nutrients, 15*, 452. https://doi.org/10.3390/nu15020452

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

Bartwal, A., John, R., Padhi, S. R., Suneja, P., Bhardwaj, R., Gayacharan, et al. (2023). NIR spectra processing for developing efficient protein prediction Model in mungbean. *Journal of Food Composition and Analysis, 116*, Article 105087. https://doi.org/10.1016/j.jfca.2022.105087

Baye, T. M., Abebe, T., & Wilke, R. A. (2011). Genotype–environment interactions and their translational implications. *Personalized Medicine, 8*, 59–70. https://doi.org/10.2217/pme.10.75

Beć, K. B., Grabska, J., & Huck, C. W. (2021). Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chemistry: A European Journal, 27*, 1514–1532. https://doi.org/10.1002/chem.202002838

Bi, Y., Tang, L., Shan, P., Xie, Q., Hu, Y., Peng, S., et al. (2014). Interference correction by extracting the information of interference dominant regions: Application to near-infrared spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 129*, 542–550. https://doi.org/10.1016/j.saa.2014.03.080

Boye, C., Nirmalan, S., Ranjbaran, A., & Luca, F. (2024). Genotype × environment interactions in gene regulation and complex traits. *Nature Genetics, 56*, 1057–1068. https://doi.org/10.1038/s41588-024-01776-w

Ömer, Cem, & Kahriman, E. (2012). Determination of quality parameters in maize grain by NIR reflectance spectroscopy. *Tarım Bilimleri Dergisi, 18*, 31–42. https://doi.org/10.1501/Tarimbil_0000001190

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*, 241–247. https://doi.org/10.1016/j.jcs.2013.07.002

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

Jang, J., Lee, D.-W., 2024. Advancements in plant based meat analogs enhancing sensory and nutritional attributes. npj Science of Food 8, 50. https://doi.org/10.1038/s41538-024-00292-9.

John, R., Bartwal, A., Jeyaseelan, C., Sharma, P., Ananthan, R., Singh, A. K., et al. (2023). Rice bean-adzuki bean multitrait near infrared reflectance spectroscopy prediction model: a rapid mining tool for trait-specific germplasm. *Frontiers in Nutrition, 10*, Article 1224955. https://doi.org/10.3389/fnut.2023.1224955

John, R., Bhardwaj, R., Jeyaseelan, C., Bollinedi, H., Singh, N., Harish, G. D., et al. (2022). Germplasm variability-assisted near infrared reflectance spectroscopy chemometrics to develop multi-trait robust prediction models in rice. *Frontiers in Nutrition, 946255*, 2022. https://doi.org/10.3389/fnut.2022.946255

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*, 227–233. https://doi.org/10.1007/s12161-016-0572-9

Kaur, S., Godara, S., Singh, N., Kumar, A., Pandey, R., Adhikari, S., et al. (2024a). Multivariate data analysis assisted mining of nutri-rich genotypes from North Eastern Himalayan germplasm collection of perilla (Perilla frutescens L.). *Plant Foods for Human Nutrition*. https://doi.org/10.1007/s11130-024-01220-8

Kaur, S., Singh, N., Dagar, P., Kumar, A., Jaiswal, S., Singh, B. K., et al. (2024b). Comparative analysis of modified partial least squares regression and hybrid deep learning models for predicting protein content in Perilla (Perilla frutescens L.) seed meal using NIR spectroscopy. *Food Bioscience, 61*, Article 104821. https://doi.org/10.1016/j.fbio.2024.104821

Kaur, S., Singh, N., Sharma, P., Ananthan, R., Singh, M., Gayacharan, et al. (2024c). Optimizing protein content prediction in rice bean (Vigna umbellata L.) by integrating near-infrared reflectance spectroscopy, MPLS, deep learning, and key wavelengths selection algorithms. *Journal of Food Composition and Analysis, 135*, Article 106655. https://doi.org/10.1016/j.jfca.2024.106655

Kaur, S., Singh, N., Tomar, M., Kumar, A., Godara, S., Padhi, S. R., et al. (2024d). NIRS-based prediction modeling for nutritional traits in Perilla germplasm from NEH Region of India: comparative chemometric analysis using mPLS and deep learning. *Food Measure*. https://doi.org/10.1007/s11694-024-02856-5

Khatri, P., Gupta, K. K., & Gupta, R. K. (2021). A review of partial least squares modeling (PLSM) for water quality analysis. *Modeling Earth Systems and Environment, 7*, 703–714. https://doi.org/10.1007/s40808-020-00995-4

Kondal, V., Jain, A., Garg, M., Kumar, S., Singh, A. K., Bhardwaj, R., et al. (2024). Gap derivative optimization for modeling wheat grain protein using near-infrared transmission spectroscopy. *Cereal Chemistry*, 10795. https://doi.org/10.1002/cche.10795

Kumari, M., Naresh, P., Acharya, G. C., Laxminarayana, K., Singh, H. S., Raghu, B. R., et al. (2022). Nutritional diversity of Indian lablab bean (Lablab purpureus (L.) Sweet): An approach towards biofortification. *South African Journal of Botany, 149*, 189–195. https://doi.org/10.1016/j.sajb.2022.06.002

Letting, F. K., Venkataramana, P. B., & Ndakidemi, P. A. (2021). Breeding potential of lablab [Lablab purpureus (L.) Sweet]: a review on characterization and bruchid studies towards improved production and utilization in Africa. *Genetic Resources and Crop Evolution, 68*, 3081–3101. https://doi.org/10.1007/s10722-021-01271-9

Mills, T. C. (2011). Dealing with Nonstationarity: Detrending, Smoothing and Differencing. *The foundations of modern time series analysis* (pp. 261–288). London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230305021_10

Murphy, D. J., O' Brien, B., O' Donovan, M., Condon, T., & Murphy, M. D. (2022). A near infrared spectroscopy calibration for the prediction of fresh grass quality on Irish pastures. *Information Processing in Agriculture, 9*, 243–253. https://doi.org/10.1016/j.inpa.2021.04.012

Padhi, S. R., John, R., Bartwal, A., Tripathi, K., Gupta, K., Wankhede, D. P., et al. (2022). Development and optimization of NIRS prediction models for simultaneous multi-trait assessment in diverse cowpea germplasm. *Frontiers in Nutrition, 9*, Article 1001551. https://doi.org/10.3389/fnut.2022.1001551

Pandey, D. K., Singh, S., Dubey, S. K., Mehra, T. S., Dixit, S., & Sawargaonkar, G. (2023). Nutrient profiling of lablab bean (Lablab purpureus) from north-eastern India: A potential legume for plant-based meat alternatives. *Journal of Food Composition and Analysis, 119*, Article 105252. https://doi.org/10.1016/j.jfca.2023.105252

Perez, C. M., & Juliano, B. O. (1978). Modification of the simplified amylose test for milled rice. *Starch Stärke, 30*, 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*, 1643–1651. https://doi.org/10.1016/j.foodres.2013.09.003

Shruti, S. A., Rahman, S. S., Suneja, P., Yadav, R., Hussain, Z., Singh, R., et al. (2023). Developing an nirs prediction model for oil, protein. *Amino Acids and Fatty Acids in Amaranth and Buckwheat. Agriculture, 13*, 469. https://doi.org/10.3390/agriculture13020469

Tian, W., Chen, G., Gui, Y., Zhang, G., & Li, Y. (2021). Rapid quantification of total phenolics and ferulic acid in whole wheat using UV–Vis spectrophotometry. *Food Control, 123*, Article 107691. https://doi.org/10.1016/j.foodcont.2020.107691

Tomar, M., Bhardwaj, R., Kumar, M., Singh, Pal, Krishnan, S., Kansal, V., et al. (2021a). Nutritional composition patterns and application of multivariate analysis to evaluate indigenous Pearl millet ((Pennisetum glaucum (L.) R. Br.) germplasm. *Journal of Food Composition and Analysis, 103*, Article 104086. https://doi.org/10.1016/j.jfca.2021.104086

Tomar, M., Bhardwaj, R., Kumar, M., Singh, S. P., Krishnan, V., Kansal, R., et al. (2021b). Development of NIR spectroscopy based prediction models for nutritional profiling of pearl millet (Pennisetum glaucum (L.)) R.Br: A chemometrics approach. *LWT, 149*, Article 111813. https://doi.org/10.1016/j.lwt.2021.111813

Vishnu, V. S., & Radhamany, P. M. (2022). Assessment of variability in Lablab purpureus (L.) Sweet germplasm based on quantitative morphological and biochemical traits. *Genetic Resources and Crop Evolution, 69*, 1535–1546. https://doi.org/10.1007/s10722-021-01316-z

Westerhaus, M. (2014). Eastern analytical symposium award for outstanding achievements in near infrared spectroscopy: My contributions to near infrared spectroscopy. *NIR News, 25*, 16–20. https://doi.org/10.1255/nirn.1492

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*, 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*, 114–121. https://doi.org/10.1016/j.chemolab.2019.01.005

Zhang, W., Zhu, Y., Liu, Q., Bao, J., & Liu, Qin (2017). Identification and quantification of polyphenols in hull, bran and endosperm of common buckwheat (Fagopyrum esculentum) seeds. *Journal of Functional Foods, 38*, 363–369. https://doi.org/10.1016/j.jff.2017.09.024