Article

Not peer-reviewed version

Modeling Methods for Evaluation of Texture Properties of Cooked Germinated Brown Rice Using the Near Infrared Spectra of Uncooked Whole Grain

Kannapot Kaewsorn , Thitima Phanomsophon , Pisut Maichoon , Dharma Pokhrel , Pimpen Pornchaloempong , Warawut Krusong , Panmanas Sirisomboon * , Munehiro Tanaka * , Takayuki Kojima

Posted Date: 13 November 2023

doi: 10.20944/preprints202311.0770.v1

Keywords: germinated brown rice; hardness; toughness; texture; near infrared spectroscopy; partial least squares regression; artificial neural network

Article

Modeling Methods for Evaluation of Texture Properties of Cooked Germinated Brown Rice Using the Near Infrared Spectra of Uncooked Whole Grain

Kannapot Kaewsorn [1], Thitima Phanomsophon [2], Pisut Maichoon [2], Dharma Pokhrel [2], Pimpen Pornchaloempong [3], Warawut Krusong [4], Panmanas Sirisomboon [2,*], Munehiro Tanaka [5,*] and Takayuki Kojima [5]

[1] Department of Agricultural Engineering, School of Engineering and Innovation, Rajamangala University of Technology Tawan-Ok, Chon Buri 20110, Thailand; kannapot_ka@rmutto.ac.th
[2] Department of Agricultural Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; thitimap.june@gmail.com, palmpisut1994@gmail.com
[3] Department of Food Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; pimpen.po@kmitl.ac.th
[4] Division of Fermentation Technology, School of Food Industry, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; warawut.kr@kmitl.ac.th
[5] Laboratory of Agricultural Production Engineering, Faculty of Agriculture, Saga University, 1 Honjo-machi, Saga University, Saga 840-8502, Japan; kojimat1733@gmail.com
* Correspondence mune@cc.saga-u.ac.jp (M.T.); panmanas.si@kmitl.ac.th (P.S.)

Abstract: The models of partial least squares (PLS) regression and artificial neural network (ANN) for evaluation of texture properties of cooked germinated brown rice (GBR) using the Fourier transform near infrared (NIR) spectra of uncooked whole grain combined with data separation methods, spectral pretreatment methods were investigated in this study. The ANN was outperformed in evaluation of hardness by back extrusion test of cooked GBR using the smoothing combined with standard normal variate pretreated NIR spectra in the range of 12,500-4,000 cm^{-1} of 188 whole grain samples where the calibration sample set was separated from prediction set by Kennard-Stone method where the best ANN model for hardness from hidden layers of 25 and 8 iteration time provided R^2, r^2, RMSEC, RMSEP, Bias and RPD of 0.9987, 0.9447, 0.1021 N, 0.7699 N, 0.0216 N and 4.3 respectively. The PLS regression of 64 sample KDML GBR group and 64 sample various variety GBR group, provided models for the hardness of the former and the toughness of the latter which developed by using 7506−5446.3, 4605.4−4242.9 cm^{-1} which included the amylose vibration band of 6834 cm^{-1} while toughness model was from 9403.8-6094.3 cm^{-1} where included 6834 and 8316 cm^{-1} vibration bands of amylose which influenced the texture of cooked rice. The PLS regression models for hardness and toughness were with the r^2 of 0.85 and 0.82, and the RPD of 2.9 and 2.4, respectively. The ANN model for hardness of cooked GBR should be implemented to the practical use in GBR production factories for the quality assurance and further updating using more samples and several brands to obtain the robust models.

Keywords: germinated brown rice; hardness; toughness; texture; near infrared spectroscopy; partial least squares regression; artificial neural network

1. Introduction

Rice (*Oryza Sativa* L.), as the world's primary staple food, holds a critical role in providing 20% of calorie intake for nearly half of the global population [1,2]. This essential crop accounts for 19% of dietary energy worldwide [3,4]. Among the leading rice-exporting nations, Thailand, alongside India and Vietnam, stands as a consistent top performer in the global rice export sector [5]. This not only bolsters Thailand's economic prosperity but also solidifies its historical moniker as the "Rice Bowl of the World.

Thailand cultivates a wide variety of rice, with Jasmine rice (Thai Hom Mali) being the most famous [7,8]. Other varieties include glutinous rice, black rice (Riceberry), red cargo rice, and various

 doi:10.20944/preprints202311.0770.v1

fragrant and non-fragrant rice types. Brown rice is a whole grain rice variety that is minimally processed, retaining its outer bran layer and germ.

As Thai consumers become increasingly health−conscious, there is a growing preference for foods that offer health benefits [9]. As many other countries, Thailand faces a rising burden of non−communicable diseases (NCDs) such as obesity, diabetes, and cardiovascular disorders [10]. Germinated brown rice (GBR) has potential in mitigating these health issues, owing to its low glycemic index and antioxidant properties, which provide health benefits such as reducing blood pressure, improving sleepiness, reducing cardiovascular diseases and diabetes regulation, and it may limit weight gain[11,12]. Thus GBR is positioned as a valuable dietary component in the battle against NCDs. . As global demand for healthy and specialty foods continues to grow, Thai GBR has become an export commodity. Its unique nutritional attributes make it an appealing product for international markets, contributing to Thailand's agricultural exports and economy.

If a non-destructive and rapid technique to determine the textural properties of cooked GBR is developed, it holds immense potential for enhancing the quality control process in large−scale commercial rice production.

The texture of cooked rice plays a pivotal role in defining the palatability and overall dining experience [13], refers to the physical feel and structure of the rice grains, including attributes like firmness, chewiness, stickiness, and grain separation [14], which determines the market value and become the driving factors in consumer preferences for rice. However, the consumption of brown rice is limited and some of the barriers are the perceptions of the rough texture and unpalatable taste and increased the length of time for cooking [15].

A number of methods have been employed to improve the textural properties of brown rice, e.g. soaking, gamma radiation, ultrasonic, enzyme treatment, high pressure cooking, freeze thaw cycle treatment and germination [15] and a germination level of 70% is considered the minimum required to produce GBR [16]. Germinated brown rice, often referred to as GBR or GABA rice, is also simply called brown rice that has undergone a natural germination process [17]. The germination process activates enzymes such as α-amylose, proteas, phytase and lipase within the rice grains, leading to various changes in the rice's nutritional content and flavor profile, and resulting in softer and sweeter cooked brown rice [18,19]. Because bran layer contributes to the hard chewy texture usually of favored by consumers, increasing softness is an important attribute in eating quality of cooked GBR [20].

Traditionally, evaluating the texture properties of cooked rice has relied on sensory assessments and instrumental measurements. Onmankhong and Sirisomboon [21] refered to the instrumental texture measurement i.e. low and high compression method on milled cooked rice grain and back extrusion test by mini Ottawa cell on the crumb of cooked parboiled rice. While these kinds of methods are well-established, they may not always be practical for large-scale operations and may be subjected to variations in human judgment.

If a non-destructive and rapid technique to determine the textural properties of cooked GBR is developed, it holds immense potential for enhancing the quality control process in large−scale commercial rice production.

NIR spectroscopy can be used to identify the chemical composition of food products and to provide in-depth identification of changes in physicochemical properties during food processing, as well as to integrate them into control strategies and the physical properties of food e.g. texture of food and their chemical composition e.g. moisture content are strongly correlated and can be reflected in the characteristic NIR spectra [22]. Machine learning (ML) can be used to model such predictive relationships based on NIR spectra, commonly used by principal component regression (PCR), support vector machine regression (SVR), partial least squares regression (PLSR) and back-propagation artificial neural network (BP-ANN) algorithms. For example, PLSR and PCR were combined with the NIR spectra for evaluation of hardness and toughness of cooked parboiled rice, respectively [21], however, the models were only applicable for screening and approximate calibration. PCR, SVR, PLSR and BP-ANN were used in prediction of quality indicators of frozen samples which were drip loss, texture parameters including hardness, chewiness, gumminess and gel strength, respectively and by comparison, the BP-ANN modeling approach performed better than other [22].

 doi:10.20944/preprints202311.0770.v1

In response to these challenges, modern analytical techniques, such as NIR spectroscopy and machine learning, have emerged as powerful tools for predicting and assessing the texture of cooked rice and of other food quality.

This research embarks on the journey of marrying the science of food texture assessment with the NIR, the running giant technology. The overarching objective of this study is to harness the potential of machine learning in conjunction with NIR spectroscopy to predict and evaluate the texture properties of cooked GBR. NIR spectra of uncooked rice grains have the potential to serve as the gateway to understanding the inherent characteristics of the grain, enabling the creation of predictive models capable of discerning textural nuances with precision. In pursuit of this goal, we bring together a diverse collection of rice samples, including both self-generated GBR and a variety of commercial GBR from different brands and cultivars. Through meticulous data collection and rigorous experimentation, we not only explore the correlation between NIR spectra and texture properties but also delve into the repeatability and reproducibility of these measurements.

In this study, two algorithms: Partial least squares regression (PLSR) and Artificial neural network (ANN) were used to develop a model. PLSR is a traditional statistical method, that is a multivariate technique that constructs latent variables, their factors or components, to capture the maximum covariance between predictors and response variables. PLSR particularly effective for high-dimensional datasets with multicollinearity. ANN is a highly advanced deep learning algorithm that is considered to be highly effective for building model overcoming problem of working with non linear relation.

This study, at the intersection of food science and machine learning, not only enhances our understanding of rice texture but also exemplifies the transformative potential of modern technology in shaping the future of food quality assessment. By the conclusion of this exploration, we aspire to unveil a novel approach to evaluating the texture of cooked GBR, ushering in a new era of precision and accuracy in the assessment of rice quality. Up till now, there have been no report on the evaluation of texture properties of cooked germinated brown rice using the near infrared spectra of uncooked whole grain.

2. Materials and Methods

2.1. Rice Samples

Rough rice of *Oryza sativa* L., cultivar Khao Dawk Mali 105 (KDML 105) was collected from a field of P.J. Brand germinated rough rice factory in Chonburi Province, Thailand. The GBR was created using the technique described by Kaewsorn and Sirisomboon [23] and Kaewsorn et al [24]: rough rice water soaking time at room temperature was 24 and 48 hours and seven different incubation intervals (0, 6, 12, 18, 24, 30 and 36 hours) to create germinated rough rice (GRR). The GRR was dried using the fluidized−bed process and air dried process. Prior to the experiment, the GRR sample was dehusked and will be referred to as GBR in this paper. Each treatment condition employed ten kg of GBR. There was one control condition (regular brown rice, 0 h soaking time, and 0 h incubation time) and 14 treatment with 2 replicates resulted in 30 samples used. While 32 commercial types and brands of GBR of various varieties and some of the same varieties indicated were purchased from local department stores in Bangkok, Thailand, and stored at room temperature in the laboratory. The commercial GBR for 32 brands with different varieties were specified in Kaewsorn et al [24]. There were 16 brands for the KDML 105 variety (2 replicates, 32 samples) and 16 brands for the other variations (2 replicates, 32 samples). As a result, 64 samples from local marketplaces were obtained.

2.2. GBR Uncooked Sample Scanning for NIR Spectra

Each sample was emptied from a vacuum bag into the quartz bottom−sampling cup (87 mm diameter and 87.5 mm height) placed in the rotational diffuse reflectance holder of a Bruker Ltd. (Germany). The NIR spectra was measured in diffuse reflection mode with an FT-NIR spectrometer (Bruker Ltd., Germany) at a wavenumber of 12,500−4,000 cm^{-1} (800−2500 nm). At a resolution of 16 cm^{-1}, each rice sample was scanned 64 times. In absorption mode (log 1/R), the scan findings were averaged and recorded. Prior to future usage, the bottom quartz−sampling cup was vacuum−cleaned.

 doi:10.20944/preprints202311.0770.v1

The background compensation was done before scanning of each sample by internal scanning gold plate as reference material. The scanning was done in 25±2 °C air condition room.

2.3. The Approximate Repeatability of NIR Scanning

The scanning was done twice per sub−sample at the same location and there were 2 sub−samples per samples. The standard deviation of the absorption value at each wavenumber of every sub−sample was calculated and the values of every sample was averaged. Then the values of every wavenumber were averaged to obtain the approximate repeatability of NIR scanning. The genuine repeatability can be obtained by scanning the sample at the same location for at least 10 times continuously [25].

2.4. Method of Cooking Rice

The technique of cooking rice utilized by Sirisomboon et al. [26] involved personal rice cookers (RC−10 MM, Toshiba, Thailand) to prepare 200 g of GBR samples using water to rice ratios that were 1.6:1 for GBR. To produce cooked rice with the customary texture that customers want, the required water to rice ratio was employed. The cooked GBR was placed in a plastic cup with a weight of approximately 5 g. In total, 5 cups per sample were prepared.

2.5. Back Extrusion Test for Texture of Cooked GBR

The cooked GBR samples were next submitted to the back extrusion test followed Kaewsorn et al [24], which involved inserting 3 g of cooked rice into a back extrusion test rig (BE) that was pressured from the top entrance of the rice container by a stainless ball for 99 mm of the total height of 100 mm at a ball probe speed of 1 mm/s. Each sample's mean was calculated using 5 duplicate tests. It is determined the hardness, toughness, stickiness, and adhesiveness of cooked GBR. The back extrusion test was performed on 94 samples, with the average of each sample obtained using 5 replications.

2.6. The Repeatability and Reproducibility of the Measurement of Texture Properties

The repeatability and reproducibility of the measurement of texture properties were determined by measuring four duplicates (four pairs) that were randomly subjected during the experiment at different times. These were reported in Kaewsorn et al [24].

2.7. NIR Spectroscopy Modelling by Machine Learning

2.7.1. Calibration Set and Prediction Set Separation

For checking model performance, several methods of sample division for calibration and prediction set were employed as it significantly impacts the model performance [27]. The calibration set should contain enough representative information to model unknown samples in the future [28], be the largest among and coverage of validation data. If the calibration set's value range does not adequately cover the validation set, prediction errors may occur because the model has not seen data with higher or lower values. The validation set is essential for effectively evaluating the model [29]. In this study, we focused on four methods for sampling, with an 80% calibration set and a 20% validation set: Interval sampling (IS), Kennard-Stone (KS), Hold-out cross validation (Hold-out CV), and sorting.

The IS method was obtained by selecting validation samples using the following steps: 1) sorting samples in ascending or descending order of the reference value. 2) dividing the samples into subsets, each containing five samples. 3) selecting the middle sample in each subset as the validation data [30].

The KS method involves selecting samples that are uniformly distributed based on Euclidean distance for distance computations [31]. This method is implemented using the following steps: 1) find the sample that is closest to the mean of the samples to be used as validation data and remove it from the dataset. 2) find the sample that is the most dissimilar to the sample selected in step (1) to be used as validation data and remove it from the dataset. 3) find the sample that is the most dissimilar to the samples that have already been allocated to the validation set based on the minimum distance

 doi:10.20944/preprints202311.0770.v1

from any sample allocated to be validation data and remove this sample from the dataset. 4) repeat step (3) until the desired amount is reached [32].

The CV method using a hold-out strategy involved random sampling without considering the data distribution. The data split proportions could vary, ranging from 90%:10% to 80%:20%, creating two mutually exclusive datasets: the training (calibration) dataset and the test (validation) dataset [33].

The last method is sorting, which is similar to the IS method. It involves dividing the samples into subsets (each subset containing ten samples) and then selecting the 7th to 8th samples in each subset to be validation data.

All four sampling methods provided different information for the calibration and validation sets, including sample distribution in each dataset (S1). To determine which method is most suitable for our data and yields the best predictions, a comparison is needed.

2.7.2. Spectral Pretreatment

The spectral interferences show by a combination of several additive factors, multiplicative factors, polynomial baseline shift and spectral noises, hence, the empirical methods are widely used for spectral pre-processing [34]. Naturally, the raw spectrum may contain noise due to factors such as sample size [35] or moisture [36] that affect to light scattering [37]. This issue can be effectively addressed through spectral pretreatment. Pretreatment techniques play a crucial role in various analytical and data applications as they serve to enhance the quality of data before further analysis is conducted [38].

For modeling by OPUS, v. 7.0, (Bruker, Germany), the following pretreatment algorithms were used in both spectrum pretreatment and model development. NIR absorption spectra were combined with the reference data. After sorting the texture data, the entire spectra data were divided into calibration and prediction sets in a 7:3 ratio. The NIR spectra used for the model development were not pre-processing, constant offset elimination, straight line subtraction, vector normalization (SNV), min−max normalization, multiplicative scatter correction (MSC), first derivatives, second derivatives, first derivatives + straight line subtraction, first derivatives + SNV and first derivatives + MSC.

In case of MATLAB spectral pretreatment was done by no pretreatment where the abbreviation was Raw − raw spectrum, and when there were pretreatment applied the methods including: BL − baseline offset spectrum, MC − mean centering spectrum, MN − mean normalization spectrum, MMN − max−min normalization spectrum, SMT − smoothing spectrum, SMT+SNV − smoothing+standard normal variate spectrum, SMT+MSC − smoothing+multiplicative scatter correction spectrum, SMT+1D − smoothing+1st derivative spectrum, SMT+2D − smoothing+2nd derivative spectrum spectrometer, were used.

The mean centering transformation, the mean of the absorption values of every sample spectrum in each wave band in the spectral data matrix is subtracted from each value in that wave band, hence, the mean centering centers the values corresponding to each band about zero (modified [39]). The mean centering amplifies the different among the sample spectra [40]. The mean normalization and max−min normalization normalize the spectra to have a common feature by dividing each absorption value of each band in the raw spectrum by the average absorption value and the range (subtracting maximum value by minimum value) absorption value, respectively, of the spectrum. Normalization pretreatment correct spectral change caused by small of light path different [40].

Baseline offset correction only removes the baseline shift where every band absorbance of each spectrum is corrected by subtracting either by its absorbance at the first band (or another arbitrarily chosen band) or the median value in a selected range of spectrum [41].

The standard normal variate (SNV) method was employed as a pretreatment step before modeling. SNV operates by centering each spectrum around zero, achieved by subtracting the mean, and then scaling each signal value by the standard deviation of the entire spectrum. SNV is highly effective in removing systematic variations in spectral data, rendering it well-suited for subsequent analyses [42].

The spectra are shifted linearly so that the minimum y−value is equal to zero for the pretreatment of constant offset elimination make the linear baseline shifts eliminated and by subtraction of a straight line pretreatment, in each selected frequency range in the spectrum, the straight line is fitted by partial least squares method and then the straight line is subtracted from the respective spectrum make a linear tilt of the baseline shift eliminated [43]. Min−Max−Normalization (for absorbance

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1

spectra): The spectra are shifted linearly, so that the minimum Y−value equals zero and then the spectra are expanded, so that the maximum Y−value equals two absorbance units where this spectral pretreatment can eliminate the influence of optical path length in changing height of the signal but not its structure in the transmission mode while in the diffuse reflectance mode the effect of different density or different particle sizes can often be minimized [43]. First derivative spectral pretreatment is done by taking derivative on each gap consecutively along the raw spectrum and the if another derivative is done on the first derivative spectrum, it is the second derivative pretreatment spectrum. By these pretreatment methods the baseline of every sample spectrum is same baseline (common baseline), the baseline shift is eliminated. The first derivative makes the peak of raw spectrum to be zero intensity point in the pretreated spectrum and the slope change point in the raw spectra will be peak of the first derivative spectrum. The second derivative pretreatment make the peaks and the overlapping peaks shown but upside down where the NIR radiation absorption by corresponding bond vibration are better shown. The slope change points along the raw spectra will be the peaks of the shoulders of the second derivative spectra. These pretreatments can make if the absorption at the amplified peaks correlated well with the dependent variables, the model performance to be improved. The smoothing pretreatment is needed to suppress noise before derivation if the raw spectrum has noise signal which is in the spike form. However, in the presence of complex interferences or when inappropriate smoothing parameters were used, the result of derivation may be rendered ineffective [34].

Multiplicative scattering correction (MSC) is used to compensate for additive and multiplicative effects in spectral data [41]. The effects are caused by physical different of samples such as different particle sizes, fruit sizes and or density of samples and the uncertainty of spectrometer due to change of humidity and temperature. MSC spectrum is obtained by the linear relationship calculated by ordinary least squares regression between the absorbance of the average spectrum of calibration sample spectra and of the sample raw spectrum where the additive factor and multiplicative factor are calculated for treated on the raw spectrum to transform to be MSC spectrum and the factors are saved for treating on prediction sample set spectra.

2.7.3. Modeling Algorithms

Partial leasts squares regression

Partial Least Squares (PLS) regression is a chemometric algorithm used for modeling to predict dependent variables (Y, in this work is texture) from independent variable (X, in this work is NIR spectral) which helpful to evaluate data in both X and Y with large, noisy, collinear, and even missing variables [44]. PLS was based on principal component regression (PCR) but PLS was created a new variable called latent variables (LVs) combined with regression [45]. The LVs are linear combinations of the original independent and are constructed in a way that maximizes the covariance between the independent and the response variable [46].

In this study, PLS regression was employed to predict the texture (quantification) of cooked GBR by using the spectra of GBR grains by OPUS software (Bruker, Germany) and by MATLAB version: 9.13.0 (R2022b) [47]. In MATLAB calculation, LVs from 1 to 20 were used for modeling to compare their performance.

Artificial neural network

Artificial Neural Networks (ANN) work similarly to the human nervous system in training the independent variables to describe the dependent variables [48]. ANN is a nonlinear model that can be used to handle complicated relationships for classification [49]. It is based on a supervised procedure and consist of input (X), hidden, and output (Y) layers with connected neurons (nodes) to simulate the network and compute weights/bias trade-offs [50,51].

In this work, hidden layers of [5, 10], [10, 10], [15, 10], [20, 10], [25, 10], and [30, 10] were used for modeling. Every set of hidden layers was generated 20 times to find the best-performing model. The modeling was performed by MATLAB version: 9.13.0 (R2022b) [47].

2.7.4. Model Performance Determination

The prediction performance of model was evaluated by the error (e) occurred in each prediction of sample i while i was equal to 1 to n which calculated by subtracting reference measured value (y) by NIR predicted value ($\hat{y}$) of either calibration set or prediction set. The averaged value of y_i ($\bar{y}$)was used together with y_i and $\bar{y}$ to calculate the coefficient of determination.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1

The coefficient of determination of calibration (R^2) and of prediction (r^2) were calculated by equation 1

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

Where, the model should be improved

The root mean square error (RMSE) of calibration (RMSEC) and of prediction (RMSEP) were calculated by equation 2

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n}} \quad (2)$$

The bias was calculated by equation 3

$$\text{Bias} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} = \frac{\sum_{i=1}^{n} e_i}{n} \quad (3)$$

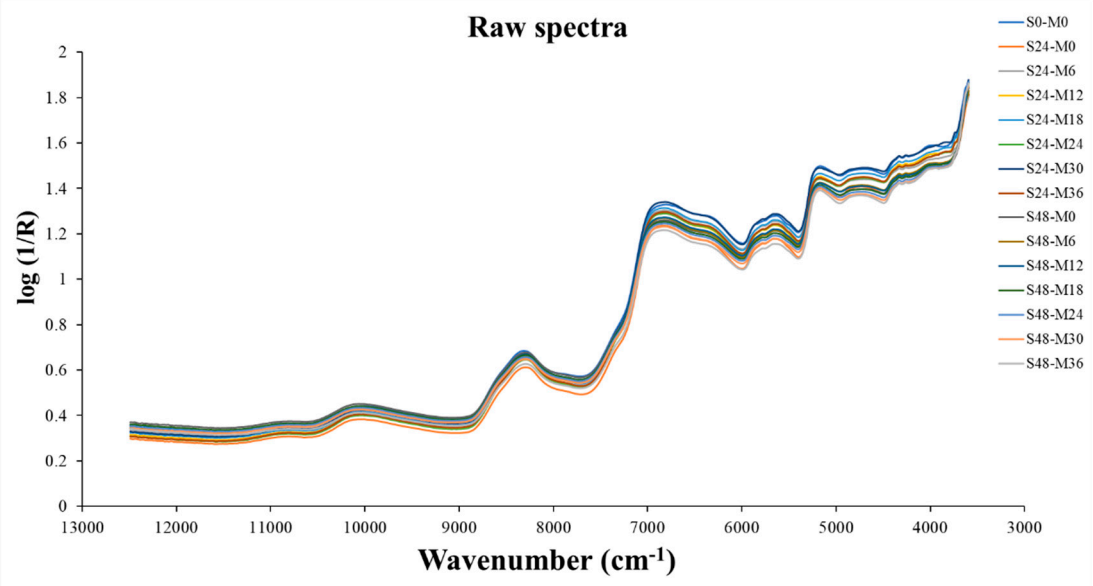
The ratio of prediction to deviation (RPD) was calculated by dividing standard deviation of prediction set (SD) by RMSEP in equation 4

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}} \quad (4)$$

3. Result and Discussion

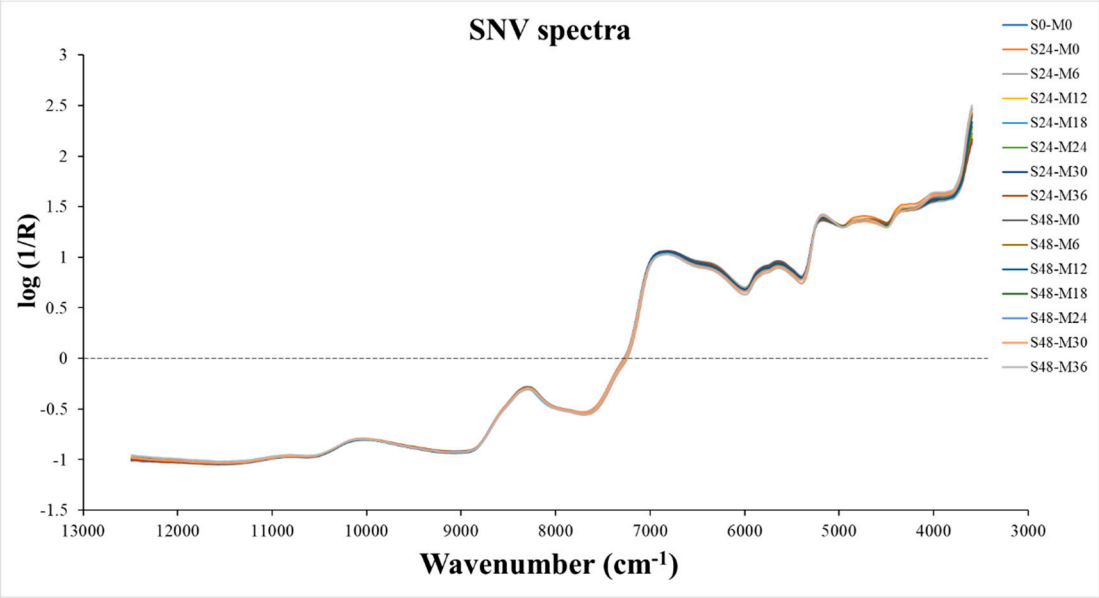
3.1. Spectral Characteristic of Whole Grain GBR

Figure 1a,b show the raw spectra and standard normal variate (SNV) pretreated spectra, respectively, of GBR grain samples on different condition of germinating processes which the structure of the spectra was same as the raw spectra and SNV pretreated spectra, respectively in Figure 1c,d of commercial GRB bought from markets. It was obvious that the GBR spectra from different condition of germinating processes show less baseline shift than the spectra of commercial GBR even with the SNV pretreatment obviously due to different production protocol make the different in constituent and matrix of GBR.

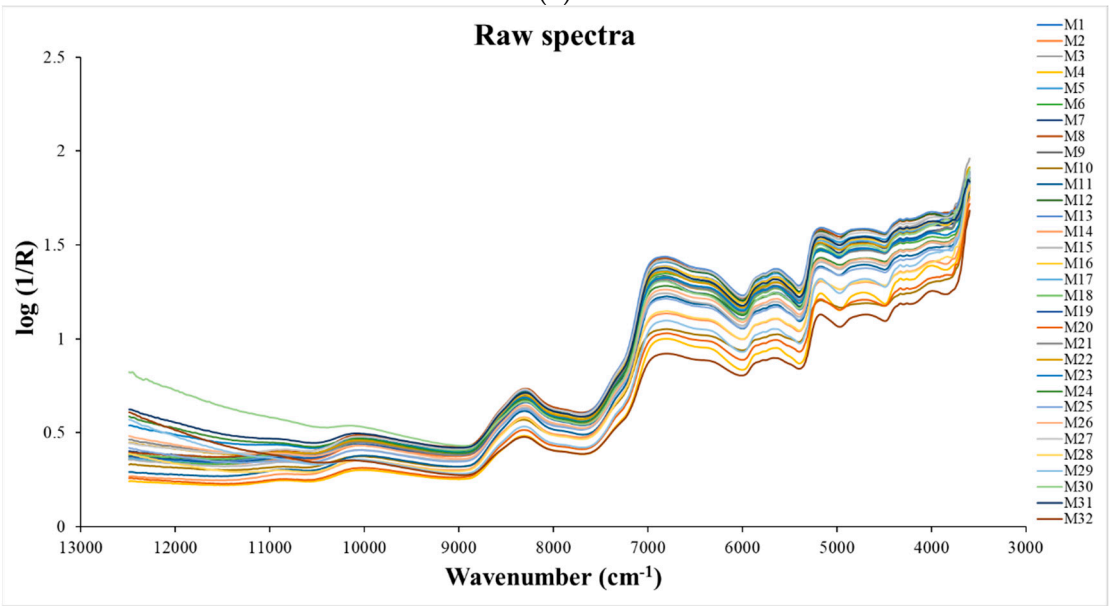


(**a**)

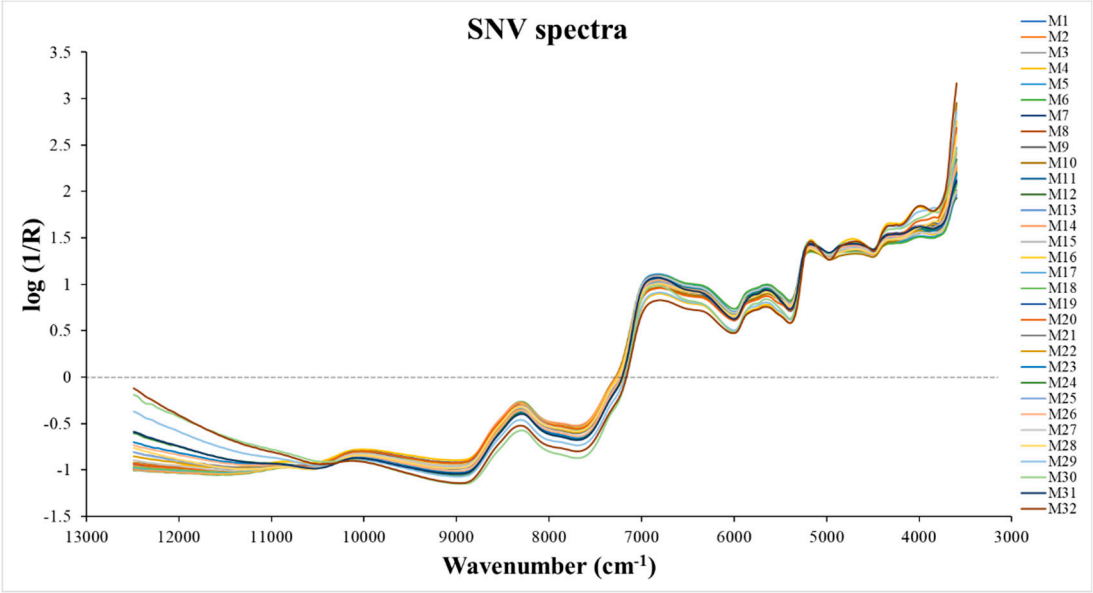
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1



(b)



(c)



(d)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1

Figure 1. a) the raw spectra b) standard normal variate (SNV) pretreated spectra of GBR grain samples on different condition of germinating processes and c) the raw spectra b) SNV pretreated spectra of commercial GRB bought from markets.

Table 1. The statistics of texture properties of cooked GBR using for modeling by OPUS

parameter	treatment	total	cal	pre	Calibration set			Prediction set		
					range	mean	SD	range	mean	SD
Adhesiveness, Nmm	Condition adjusted GBR (24 and 48 soaking hrs)	60	42	18	(-81.15)-(-56.70)	-67.41	6.28	(-77.93)-(-56.99)	-69.09	6.78
Toughness, Nmm		60	42	18	162.32-245.79	201.78	20.50	172.99-245.40	203.87	21.97
Hardness, N		60	42	18	16.55-24.87	20.29	1.85	17.88-23.71	20.86	2.00
Stickiness, N		60	42	18	(-7.67)-(-4.48)	-5.78	0.88	(-7.37)-(-4.65)	-6.15	0.88
Adhesievness, Nmm	KDML (1-16)	64	46	18	(-78.86)-(-39.13)	-64.22	7.71	(-76.41)-(-40.69)	-62.77	10.58
Toughness, Nmm		64	46	18	109.69-240.80	196.85	25.63	112.22-233.45	193.51	35.98
Hardness, N		64	46	18	11.90-24.52	20.24	2.73	12.11-24.27	19.59	3.62
Stickiness, N		64	46	18	(-7.12)-(-2.84)	-5.15	1.04	(-6.85)-(-3.12)	-4.82	1.13
Adhesievness, Nmm	various varieties (17-32)	64	46	18	(-84.93)-(-52.45)	-69.04	8.59	(-83.79)-(-55.22)	-69.44	9.24
Toughness, Nmm		64	46	18	131.94-300.55	197.98	39.34	152.77-276.90	212.36	40.37
Hardness, N		64	46	18	14.23-29.86	20.27	3.33	15.87-27.93	21.93	4.05
Stickiness, N		64	46	18	(-7.20)-(-3.12)	-5.31	0.88	(-6.18)-(-3.18)	-5.08	0.95
Adhesievness, Nmm	Market (1-32)	128	90	38	(-84.93)-(-39.13)	-66.71	8.54	(-83.79)-(-40.69)	-65.94	10.12
Toughness, Nmm		128	90	38	109.69-300.55	197.80	33.01	112.22-276.90	201.73	38.76
Hardness, N		128	90	38	11.90-29.86	20.27	3.12	12.11-27.93	20.69	3.76
Stickiness, N		128	90	38	(-7.20)-(-2.84)	-5.14	0.95	(-7.12)-(-3.12)	-5.17	1.08

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1

parameter	treatment	total	cal	pre	Calibration set			Prediction set		
					range	mean	SD	range	mean	SD
Adhesievness, Nmm	All sample	188	129	59	(-84.93)-(-39.13)	-67.17	7.80	(-83.79)-(-40.69)	-66.43	9.28
Toughness, Nmm		188	132	56	109.69-300.55	198.20	24.74	112.22-276.90	199.41	36.72
Hardness, N		188	133	55	11.90-29.86	20.16	2.37	12.11-27.93	20.45	3.40
Stickiness, N		188	131	57	(-7.67)-(-2.84)	-5.38	0.93	(-7.67)-(-3.12)	-5.49	1.12

Table 2. The statistics of texture properties of cooked GBR samples using for modeling by MATLAB

		Hardness		Toughness		Stickiness		Adhesiveness	
		calibration	prediction	calibration	prediction	calibration	prediction	calibration	prediction
IS	number	150	38	150	38	150	38	150	38
	min	11.90	12.11	109.69	112.22	-7.67	-7.37	-84.93	-83.79
	max	29.86	29.86	300.55	300.55	-2.84	-2.84	-39.13	-39.13
	mean	20.39	20.52	199.75	201.32	-5.39	-5.35	-67.00	-66.69
	SD	3.32	4.51	34.40	46.02	1.10	1.35	9.83	13.68
KS	number	150	38	150	38	150	38	150	38
	min	11.90	15.87	109.69	152.77	-7.67	-7.67	-84.93	-83.79
	max	29.86	29.86	300.55	300.55	-2.84	-3.12	-39.13	-52.45
	mean	20.32	20.79	199.31	203.05	-5.36	-5.48	-66.33	-69.35
	SD	3.27	4.67	34.00	47.23	1.13	1.25	9.97	13.25
sort	number	150	150	150	38	150	38	150	38
	min	11.90	11.90	109.69	112.22	-7.67	-7.37	-84.93	-83.79
	max	29.86	29.86	300.55	275.84	-2.84	-3.12	-39.13	-52.45
	mean	20.46	20.46	200.39	198.77	-5.38	-5.40	-66.83	-67.38
	SD	3.33	3.33	34.39	45.84	1.10	1.34	9.98	13.29
cv	number	151	37	151	37	151	37	151	37
	min	11.90	12.11	109.69	112.22	-7.67	-7.37	-84.93	-83.79
	max	29.86	26.32	300.55	300.55	-2.84	-2.84	-39.13	-39.13
	mean	20.31	20.84	197.92	208.80	-5.41	-5.28	-66.78	-67.59
	SD	3.39	4.33	33.50	48.59	1.09	1.37	9.92	13.57

Table 3. The PLS regression result by OPUS software for prediction of texture quality of cooked GBR by using GBR grains spectra.

treatment	parameter	Pretreatment	Rank	Wavenumber	Calibration			Prediction			
					R^2	RMSEE	RPD	r^2	RMSEP	RPD	Bias
Condition adjusted GBR (24 and 48 soaking hrs)	adhesiev-ness	Constant offset elimination	2	9403.8-6094.3, 5054-4242.9	0.63	3.94	1.64	0.75	3.33	2.13	1.20
	toughness	Constant offset elimination	5	5778-5446.3, 4605.4-4242.9	0.79	10.10	2.17	0.86	8.10	2.73	-2.10
	hardness	no spec	2	6102-5757.3	0.59	1.22	1.55	0.92	0.55	3.95	0.26

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1

treatment	parameter	Pretreatment	Rank	Wavenumber	Calibration			Prediction			
					R^2	RMSEE	RPD	r^2	RMSEP	RPD	Bias
	stickiness	no spec	1	8454.9-7498.3, 4605.4-4242.9	0.12	0.84	1.07	0.03	0.84	1.01	-0.19
KDML (1-16)	adhesiev-ness	first+MSC	9	7506-5446.3	0.84	3.44	2.51	0.74	5.29	2.05	1.64
	toughness	first+MSC	5	9403.8-7498.3, 4605.4-4242.9	0.68	15.50	1.76	0.84	14.00	2.51	1.27
	hardness	**SNV**	**9**	**7506-5446.3, 4605.4-4242.9**	**0.87**	**1.12**	**2.74**	**0.85**	**1.39**	**2.90**	**-0.67**
	stickiness	first+straight	7	7506-4597.7	0.76	0.55	2.06	0.68	0.63	1.94	0.26
various varieties (17-32)	adhesiev-ness	Con off eli	8	7506-6094.3, 5454-4597.7	0.70	5.23	1.81	0.87	3.19	2.91	0.82
	toughness	**SNV**	**9**	**9403.8-6094.3**	**0.84**	**17.40**	**2.53**	**0.82**	**16.90**	**2.35**	**2.45**
	hardness	MSC	10	9403.8-7498.3, 6102-4597.7	0.97	0.63	5.95	0.32	3.25	1.29	1.13
	stickiness	no spec	7	7506-6094.3, 5029.7-4597.7	0.71	0.52	1.84	0.73	0.48	2.00	-0.14
Market(1-32)	adhesievness	SNV	5	9403.8-7498.3, 4605.4-4420.3	0.34	7.13	1.23	0.48	7.20	1.49	2.70
	toughness	no spec	10	9403.8-6094.3, 5454-4597.7	0.64	20.90	1.68	0.71	20.50	1.87	1.61
	hardness	MSC	7	9403.8-6094.3, 5454-4597.7	0.50	2.31	1.41	0.61	2.31	1.62	0.33
	stickiness	first+MSC	5	6102-4597.7	0.42	0.77	1.32	0.44	0.80	1.34	-0.07
total sample	adhesive-ness	Min-Max	7	9403.8-7498.3, 4605.4-4242.9	0.52	5.54	1.45	0.21	8.16	1.14	1.33
	toughness	no spec	10	9403.8-7498.3, 6102-5770.3	0.53	20.10	1.46	0.63	22.20	1.69	-5.11
	hardness	SNV	9	9403.8-7498.3, 6102-4597.7	0.55	1.89	1.50	0.56	2.23	1.51	-0.16
	stickiness	SNV	10	9403.8-6094.3, 5454-4597.7	0.53	0.70	1.45	0.21	0.98	1.14	-0.14

The average raw NIR spectra of germinated brown rice acquired throughout a wave number range of 12500−4000 cm^{-1} appear in the peaks at 10013, 8262, 6781, 6333, 5763, 5608 and 5161 cm^{-1}. The peak at 10013 cm^{-1} (about 990 nm) corresponds to the absorption band of the second overtone associated with starch's O−H stretching, while the peak at 8262 cm^{-1} (1210 nm) relates to the second overtone associated with the CH_2 group's C−H stretching (usually found around 1215 nm), at 6333 cm^{-1} (1579 nm) due to the absorption band associated with the first overtone of C−H stretching of starch (1580 nm), at 5608 cm^{-1} (1783 nm) due to the first overtone of C−H stretching of cellulose (typically found at 1780 nm), and at 5161 cm^{-1} (1938 nm) due to a combination of vibration due to the O−H stretching + O−H deformation of water [52]. Furthermore, the Savizky−Golay second derivative spectra of the GBR samples show that the CH_3, CH_2, CO_2H and cellulose (data not shown) [52].

3.2. Overall Precision Test

The spectral precision of the condition adjusted GBR and the commercial GBR of 32 Brands indicated by the average standard deviation of the absorption value of every wavenumber were 0.00154 and 0.00080, respectively. The spectral precision of whole wheat grains scanned by FT−NIR spectrometer reported by [25] were 0.00310, 0.0034, 0.00494, and 00782 at 10373.4, 8665.5, 8333.3, 5896.3, and 4262.6, respectively. Figure 2 shows the different repeatability with reducing of wavenumber where the beginning and end shows low repeatability and in the middle shows higher repeatability, though the pattern of the two GBR groups were same.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0770.v1

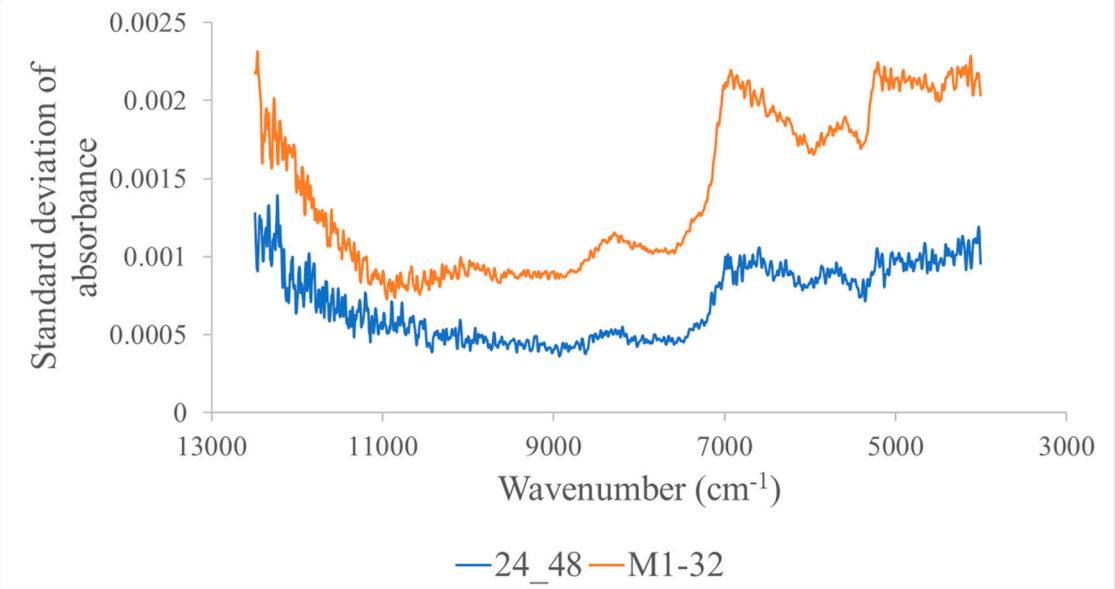


Figure 2. Standard deviation graph of absorbance of condition adjusted GBR samples (Blue) and commercial samples (Red) changed with wavenumber.

Repeatability and reproducibility of texture measurement of cooked GBR of the same sample sets as this experiment which reported by Kaewsorn et al [24] were 1.31 and 1.42 N, 13.97 and 13.34 Nmm, 0.83 and 0.38 N and 2.87 and 12.79 Nmm for hardness, toughness, stickiness and adhesiveness, respectively. These values provided the maximum R^2 for the NIR prediction of 0.8395, 0.8312, 0.4605, and 0.9171 for hardness, toughness, stickiness and adhesiveness, respectively, by calculation using the method used by Sirisomboon and Nawayon [53], Pornchaloempong et al [54] and Lim and Sirisomboon [55] and the statistics of the calibration set obtained by KS method. These maximum R^2 can be obtained when no NIR error but only the reference laboratory error. This indicated the back extrusion test for the stickiness must be researched why to high error obtained or the variation among samples with low standard deviation of stickiness.

3.3. Prediction Performance of PLS Regression Model for Texture of Cooked GBR by Uncooked GBR Grains by OPUS

From Table 3 it is obvious that the prediction using total samples for the texture properties was poor with r^2 only 0.21–0.63 which was the same as the GBR production condition adjusted samples and the commercial 32 brand samples where the r^2 of 0.03–0.92 and 0.44–0.71, respectively but with under–fitting prediction. However, when separated samples of commercial brand sample into KDML GRB group and various variety group, the model performance was better but only for the hardness of the former and the toughness of the latter with the r^2 of 0.85 and 0.82, and the RPD of 2.9 and 2.4, respectively. By Williams guideline [25] the r^2 between 0.83–0.90 indicate the model was usable for most application but with caution including research and the RPD of 2.4 indicate poor model which can be used for rough screening and between 2.5–2.9 the model was fair and could be used for screening. Sampaio et al. [56] showed the pure amylose NIR spectrum where the major peaks of amylose were at 4633, 4996, 5184 and 6834 and 8316 cm^{-1}. The model which was best for hardness prediction was developed by using 7506–5446.3, 4605.4–4242.9 cm^{-1} which included the amylose vibration band which was 6834 cm^{-1} while toughness model was from 9403.8–6094.3 cm^{-1} where included 6834 and 8316 cm^{-1}. Amylose content is correlated with the retrogradation behavior, influencing the textural properties of cooked rice [57,58]

3.4. Prediction Performance of PLS Regression Model for Texture of Cooked GBR by Uncooked GBR Grains by MATLAB Using Total Samples

By calibration and test set separation using IS the hardness, toughness, stickiness and adhesiveness of cooked GBR were poor predicted by PLS regression model using full wavelength range which r^2 were 0.26–0.38, 0.38–0.55, 0.07–0.22 and 0.02–0.41, respectively. In case of KS

 doi:10.20944/preprints202311.0770.v1

separation method, the poor r^2 were 0.46−0.60, 0.45−0.56, −0.00−0.27, 0.05−0.41, respectively. These model predictions were not acceptable obviously.

3.5. Prediction Performance of ANN Model for Texture of Cooked GBR by Uncooked GBR Grains by MATLAB Using Total Samples

By ANN, the model performance was better than those of PLS regression remarkably, where by IS data separation method, the models by using raw and SMT_SNV spectra, showed overfitting prediction for evaluation of every texture parameters while by KS, the overfitting was occurred only on 3 parameters except on the hardness where the R^2 and r^2 when the hidden layer increased from 5−30 were between 0.9172−0.9989 and 0.8133−0.9248 (raw spectra) and 0.9598−0.9992 and 0.8918−0.9523 (SMT_SNV spectra), respectively. The best ANN model for hardness from hidden layers of 25 and 8 iteration time provided R^2, r^2, RMSEC, RMSEP, Bias and RPD of 0.9987, 0.9447, 0.1021 N, 0.7699 N, 0.0216 N and 4.3 respectively. Williams et al [25] indicate the model can be used for most application including quality assurance when the r^2 between 0.92-0.96 and the model can be for any application when RPD more than 4.1 in case of functionality parameters which in this case is texture properties.

From linearized modeling by PLS regression for our case of total samples the r^2 for hardness was less than 0.6 but when the ANN was applied the r^2 was 0.9447 indicating the relationship between NIR spectral data and hardness data of cooked GBR was nonlinear. ANN accurately fit nonlinear variables, which is an advantage compared to multivariate linear analysis [59,60]. ANN have shown its outperformance compared to PLS regression in estimation of texture properties of cooked rice by the models developed by using the spectra of whole grain. Aznan et al. [61] used a portable near-infrared spectrometer coupled with ANN predicted the rice quality traits (color, texture, and pH of cooked rice) of 17 commercial rice types and a high correlation coefficient (R) of 0.98 was obtained. The prediction of hardness and toughness of cooked parboiled rice by using FT-NIR spectra of whole grain parboiled rice combined with PLS regression and principal component regression (PCR), provided the r^2 of prediction of 0.70 and 0.66, respectively, which indicated the lower performance of PLS regression. Sitakalin & Meullenet [62] reported that the sensory texture prediction of cooked rice achieved by ANN model was superior to that of PLS regression model [63].The models developed by using 166 rice flour NIR spectra combined with interval partial least squares (iPLS) and synergy interval PLS (siPLS) were characterized the texture related properties i.e. pasting parameters of rice flour which provided the R between 0.57-0.90, but the ANN provided 0.70-0.99 of R [60]

4. Conclusions

By machine learning the ANN was outperformed in evaluation of hardness by back extrusion test of cooked GBR using the SMT_SNV NIR spectra in the range of 12,500−4,000 cm^{-1} of 188 whole grain samples where the calibration sample set was separated from prediction set by KS method and the model developed can be used for most application including quality assurance use. Though the lower performance, the PLS regression of 64 sample KDML GRB group and 64 sample various variety GBR group, provided models for the hardness of the former and the toughness of the latter which were usable for most application but with caution including research. The model which was best for hardness prediction was developed by using 7506−5446.3, 4605.4−4242.9 cm^{-1} which included the amylose vibration band of 6834 cm^{-1} while toughness model was from 9403.8−6094.3 cm^{-1} where included 6834 and 8316 cm^{-1} vibration bands of amylose which influenced the texture of cooked rice. The hardness and the toughness of cooked rice can be predicted by NIR spectroscopy for the texture reference test was accurate as confirmed by high maximum R^2 and high R^2 while the stickiness and adhesiveness cannot be predicted due to the low maximum R^2 indicating the inaccurate measurement of stickiness by back extrusion test and the uncorrelated between the adhesiveness with NIR absorption of samples even though the maximum R^2 was the highest. The ANN model for hardness of cooked GBR should be implemented to the practical use in GBR production factories for the quality assurance and further updating using more samples and several brands to obtain the robust models.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Supplement data of separation methods.

Author Contributions: Conceptualization, K.K., P.S. and M.T.; validation, K.K., P.P., W.K. and T.K.; formal analysis, K.K., T.P. and P.M.; investigation, K.K.; writing—original draft preparation, K.K., P.S., T.P., D.P. and P.M.; writing—review and editing, K.K., P.S., P.P. W.K. and M.T.; visualization, M.T. and T.K.; supervision, M.T. and T.K.; funding acquisition, P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NIRS research center for agricultural product and food, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.

Institutional Review Board Statement: Not applicable for this study not involving humans or animals.

Informed Consent Statement: Not applicable for this study not involving humans.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors thank the Near infrared spectroscopy research center for agricultural product and food in Department of agricultural engineering, School of engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand for the instrument support and to P.J. Brand germinated rough rice in Chonburi Province, Thailand, for the germinated brown rice production methodology.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Zhou, C.Q.; Lu, C.H.; Mai, L.; Bao, L.J.; Liu, L.Y.; Zeng, E.Y. Response of rice (Oryza sativa L.) roots to nanoplastic treatment at seedling stage. *J. Hazard. Mater.* **2021**, *401*, 123412.
2. Xu, L.; Li, Z.; Zhuang, B.; Zhou, F.; Li, Z.; Pan, X.; Xi, H.; Zhao, W.; Liu, H. Enrofloxacin perturbs nitrogen transformation and assimilation in rice seedlings (Oryza sativa L.). *Sci. Total Environ.* **2022**, *802*, 149900.
3. Ajith, K.; Geethalakshmi, V.; Ragunath, K.P.; Pazhanivelan, S.; Dheebakaran, Ga. Rice Yield Prediction Using MODIS - NDVI (MOD13Q1) and Land Based Observations. *Int.J.Curr.Microbiol.App.Sci* **2017**, *6*, no. 12, 2277-2293.
4. Pame, A.R.P.; Vithoonjit, D.; Meesang, N.; Balingbing, C.; Gummert, M.; Hung, N.V.; Singleton, G.R.; Stuart, A.M. Improving the Sustainability of Rice Cultivation in Central Thailand with Biofertilizers and Laser Land Leveling. *Agronomy* **2023**, *13*, no.2, 587.
5. Madhu, S.; Kamani, D.; Mesara, H. A Study on Production and Export of Rice from India. *Int. Res. J. Mod. Eng. Technol. Sci.* **2023**, *5(3)*, 346.
6. Iemsam-arng, M. (Agricultural Specialist, Approved By: Kelly Stange, PREPARED BY USDA AND GLOBAL AGRICULTURE). Rice Price – Weekly. Report Number TH2023-0047. In *Grain and Feed*, Bangkok, Thailand, Public Distribution Date: 04 August 2023.
7. Srinuttrakul, W.; Mihailova, A.; Islam, M.D.; Liebisch, B.; Maxwell, F.; Kelly, S.D.; Cannavan, A. Geographical Differentiation of Hom Mali Rice Cultivated in Different Regions of Thailand Using FTIR-ATR and NIR Spectroscopy. *Foods* **2021**, *10*, no. 8 1951.
8. Vanavichit, W.; Kamolsukyeunyong, M.; Siangliw, J.L.; Siangliw, S.; Traprab, S.; Ruengphayak, E.; Chaichoompu, C.; Saensuk, E.; Phuvanartnarubal, T.; Toojinda, Tragoonrung, S. Thai Hom Mali Rice: Origin and breeding for subsistence rainfed lowland rice system. *Rice* **2018**, *11*, no. 20.
9. Shafie, F.A.; Rennie, D. Consumer Perceptions Towards Organic Food. *Procedia Soc. Behav. Sci.* **2012**, *49*, 360-367.
10. Wagner, K.-H.; Brath, H. A global view on the development of non-communicable diseases. *Prev. Med.* **2012**, *54*, 38-41.
11. Ismail, M. Germinated Brown Rice and Bioactive Rich Fractions, *Report*, [URL: https://pnc.upm.edu.my/upload/dokumen/20170726114317GErminated_Broen_Rice_and_Bioactive_Rich_Fractions.pdf
12. Kobayashi, K.; Wang, X.; Wang W. Genetically Modified Rice Is Associated with Hunger, Health, and Climate Resilience. *Foods* **2023**, *12*, no. 14, 2776.
13. Nishinari, K.; Peyron, M.-A.; Yang, N.; Gao, Z.; Zhang, K.; Fang, Y.; Zhao, M.; Yao, X.; Hu, B.; Han, L.; Mleko, S.; Tomczyńska-Mleko M.; Nagano, T.; Nitta, Y.; Zhang, Y.; Singh. N.; Meng, A.G.S.; Pongsawat- manit, R.; Gamonpilas, C. The role of texture in the palatability and food oral processing. *Food Hydrocoll.* **2023**, 109095.
14. Tikapunya, T.; Henry, J.R.; Smyth, H. Evaluating the sensory properties of unpolished Australian wild rice. *Food Res. Int.* **2018**, *103*, 406-414.

15. Chao, S.; Mitchell, J.; Prakash, S.; Bhandari, B.; Fukai, S. Effect of germination level on properties of flour paste and cooked brown rice texture of diverse varieties. *J. Cereal Sci.* **2021**, *102*, 103345.
16. Panchan, K.; Naivikul, O. Effect of pre-germination and parboiling on brown rice properties. *Asian J. Food Agro-Ind.* **2009**, *2*, no. 4, 515-524.
17. Munarko, H.; Sitanggang, A.B.; Kusnandar, F.; Budijanto, S. Effect of different soaking and germination methods on bioactive compounds of germinated brown rice. *Int. J. Food Sci* **2021**, *56(9)*, 4540-4548.
18. Nascimento, L.Á.; Abhilasha, A.; Singh, J.; Elias, M.C.; Colussi, R. Rice Germination and Its Impact on Technological and Nutritional Properties: A Review. *Rice Sci.* **2022**, *29(3)*, 201-215.
19. Jiamyangyuen, S.; Ooraikul, B.; The Physico-Chemical, Eating, and Sensorial Properties of Germinated Brown Rice. *J. Food Agric Environ* **2008**, *6*, no. 2, 119.
20. Han, A.; Jinn, J.-R.; Mauromoustakos, A.; Wang, Y.-J. Effect of Parboiling on Milling, Physicochemical, and Textural Properties of Medium- and Long-Grain Germinated Brown Rice. *Cereal Chem.* **2015**, *93(1)*, 47-52.
21. Onmankhong, J.; Sirisomboon, P. Texture Evaluation of Cooked Parboiled Rice Using Nondestructive Milled Whole Grain Near Infrared Spectroscopy. *J. Cereal Sci*, **2021**, *97*, 103151.
22. Jiang, Q.;Zhang, M.; Mujumdar, A. S.; Wang, D. Non-Destructive Quality Determination of Frozen Food Using NIR Spectroscopy-Based Machine Learning and Predictive Modelling. *J. Food Eng*, **2023**, *343*, 111374.
23. Kaewsorn, K.; Sirisomboon, P. Study on Evaluation of Gamma Oryzanol of Germinated Brown Rice by Near Infrared Spectroscopy. *J. Innovative Opt. Health Sci* **2014**, *7(4)*, 1450002.
24. Kaewsorn, K.; Maichoon, P.; Pornchaloempong, P.; Krusong, W.; Sirisomboon, P.; Tanaka, M.; Kojima, T. Evaluation of Precision and Sensitivity of Back Extrusion Test for Measuring Textural Qualities of Cooked Germinated Brown Rice in Production Process. *Foods* **2023**, *12(16)*, 3090.
25. Williams, P.; Manley, M.; Antoniszyn, J. *Near infrared technology: getting the best out of light*. 1st ed.; African Sun Media, 2019.
26. Sirisomboon, P.; Kaewsorn, K.; Thanimkarn, S.; Phetpan, K. Non-linear viscoelastic behavior of cooked white, brown, and germinated brown Thai jasmine rice by large deformation relaxation test. *Int. J. Food Prop* **2017**, *20(7)*, 1547-1557.
27. Sun, Y.; Yuan, M.; Liu, X.; Su, M.; Wang, L.; Zeng, Y.; Zang, H.; Nie, L. A sample selection method specific to unknown test samples for calibration and validation sets based on spectra similarity. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *258*, 119870.
28. Shetty, N.; Rinnan, Å.; Gislum, R. Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration in grasses. *Chemometr. Intell. Lab. Syst.* **2012**, *111(1)*, 59-65.
29. Galvão, R.K.H.; Araujo, M.C.U.; José, G.E.; Pontes, M.J.C.; Silva, E.C.; Saldanha, T.C.B. A method for calibration and validation subset partitioning. *Talanta* **2005**, *67(4)*, 736-740.
30. Li, X.; Wei, Y.; Xu, J.; Feng, X.; Wu, F.; Zhou, R.; Jin, J.; Xu, K.; Yu, X.; He, Y. SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology. *Postharvest Biol. Technol.* **2018**, *143*, 112-118.
31. Biswas, A.; Zhang, Y. Sampling Designs for Validating Digital Soil Maps: A Review. Pedosphere **2018**, *28*, 1-15.
32. Ramirez-Lopez, L.; Schmidt, K.; Behrens, T.; Wesemael, B.V.; Demattê, J.A.M.; Scholten, T. Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma* **2014**, *226-227*, 140-150.
33. Rafało, M. Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis. *ICT Express* **2022**, *8(2)*, 183-188.
34. Bi, Y.; Yuan, K.; Xiao, W.; Wu, J.; Shi, C.; Xia, J.; Zhou, G. A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. *Anal. Chim. Acta* **2016**, *909*, 30-40.
35. Robert, G.; Gosselin, R. Evaluating the impact of NIR pre-processing methods via multiblock partial least-squares. *Anal. Chim. Acta* **2022**, *1189*, 339255.
36. Munnaf, M.A.; Mouazen, A.M. Removal of external influences from on-line vis-NIR spectra for predicting soil or-ganic carbon using machine learning. *CATENA* **2022**, *211*, 106015.
37. Gobrecht, A.; Bendoula, R.; Roger, J.-M.; Bellon-Maurel, V. A new optical method coupling light polarization and Vis–NIR spectroscopy to improve the measurement of soil carbon content. *Soil tillage res.* **2016**, *155*, 461-470.

38. An, C.; Yan, X.; Lu, C.; Zhu, X. Effect of spectral pretreatment on qualitative identification of adulterated bovine colostrum by near-infrared spectroscopy. *Infrared Phys. Technol.* **2021**, *118*, 103869.
39. Yang, L.; Lua, Y.Y.; Jiang, G.; Tyler, B.J.; Linford, M.R. Multivariate analysis of TOF-SIMS spectra of monolayers on scribed silicon. *Anal. Chem.* **2005**, *77(14)*, 4654-4661.
40. Bian, X. Spectral preprocessing methods. In *chemometric methods in analytical spectroscopy technology*. Springer Nature Singapore: Singapore, 2022, pp111-168.
41. Verboven, S.; Hubert, M.; Goos, P. Robust preprocessing and model selection for spectral data. J. Chemom. **2012**, *26(6)*, 282-289.
42. Siesler, H.W.; Kawata, S.; Heise, H.M.; Ozaki, Y. *Near-infrared spectroscopy: principles, instruments, applications*; John Wiley & Sons, 2008.
43. Conzen, J. P. Multivariate calibration. In *A practical guide for developing methods in the quantitative analytical chemistry*. 2nd English ed.; Bruker Optics: Ettlingen, Germany, 2006.
44. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory *Systems* **2001**, *58(2)*, 109-130.
45. Hanrahan, G.; Udeh, F.; Patil, D.G. CHEMOMETRICS AND STATISTICS | Multivariate Calibration Techniques. In *Encyclopedia of Analytical Science*. 2nd ed.; Elsevier Ltd., 2005, pp. 27-32.
46. Lavine, B.K.; Rayens, W.S. Statistical Discriminant Analysis. In *Comprehensive Chemometrics*; Elsevier B.V., 2009.
47. The MathWorks Inc. MATLAB version: 9.13.0 (R2022b), The MathWorks Inc.: Natick, Massachusetts, 2022, https://www.mathworks.com
48. Sadiq, R.; Rodriguez, M.J.; Mian, H.R. Empirical Models to Predict Disinfection By-Products (DBPs) in Drinking Water: An Updated Review. In *Encyclopedia of Environmental Health*. 2nd ed.; Elsevier B.V., 2019, pp. 324-338.
49. Mohseni-Dargah, M.; Falahati, Z.; Dabirmanesh, B.; Nasrollahi, P.; Khajeh, K. Machine learning in surface plasmon resonance for environmental monitoring. In *Artificial Intelligence and Data Science in Environmental Sensing*; Elsevier Inc., 2022.
50. Malekian, A.; Chitsaz, N. Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. In *Advances in Streamflow Forecasting*; Elsevier Inc., 2021, pp. 115-147.
51. Park, Y.-S.; Lek, S. Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. In *Developments in Environmental Modelling*; Elsevier B.V., 2016, Vol. 28, pp. 123-140.
52. Osborne, B.G.; Fearn, T.; Hindle, P.H. Practical NIR spectroscopy with applications in food and beverage analysis. 2nd ed.; Longman scientific and technical: London, England, 1993.
53. Sirisomboon, P.; Nawayon, J. Evaluation of total solids of curry soup containing coconut milk by near infrared spectroscopy. *J. Near Infrared Spectrosc.* **2016**, *24(2)*, 191-198.
54. Pornchaloempong, P.; Sharma, S.; Phanomsophon, T.; Srisawat, K.; Inta, W.; Sirisomboon, P.; Prinyawiwatkul, W.; Nakawajana, N.; Lapcharoensuk, R.; Teerachaichayut, S. Non-destructive quality evaluation of tropical fruit (Mango and Mangosteen) Purée using near-infrared spectroscopy combined with partial least squares regression. *Agriculture* **2022**, *12(12)*, 2060.
55. Lim, C.H.; Sirisomboon, P. Near infrared spectroscopy as an alternative method for rapid evaluation of toluene swell of natural rubber latex and its products. *J. Near Infrared Spectrosc.* **2018**, *26(3)*, 159-168.
56. Sampaio, P.S.; Soares, A.; Castanho, A.; Almeida, A.S.; Oliveira, J.; Brites, C. Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms. *Food Chem.* **2018**, *242*, 196-204.
57. Lu, Z.H.; Sasaki, T.; Li, Y.Y.; Yoshihashi, T.; Li, L.T.; Kohyama, K. Effect of amylose content and rice type on dynamic viscoelasticity of a composite rice starch gel. *Food Hydrocoll.* **2009**, *23(7)*, 1712-1719.
58. Sampaio, P.S.; Brites, C.M. Near-Infrared Spectroscopy and Machine Learning: Analysis and Classification Methods of Rice. In Integrative Advances in Rice Research; Huang, M., IntechOpen, 2022.
59. Vrahatis, M.N.; Magoulas, G.D.; Parsopoulos, K.E.; Plagianakos, V.P. Introduction to Artificial Neural Networks Training and Applications. In 15th Annual Conference of Hellenic Society for Neuroscience, Patras, Greece, 27-29 October 2000.
60. Sampaio, P.S.; Carbas, B.; Brites, C. Development of Prediction Models for the Pasting Parameters of Rice Based on Near-Infrared and Machine Learning Tools. *Appl. Sci.* **2023**, *13(16)*, 9081.
61. Aznan, A.; Gonzalez Viejo, C.; Pang, A.; Fuentes, S. Rapid Assessment of Rice Quality Traits Using Low-Cost Digital Technologies. *Foods* **2022**, *11(9)*, 1181.

 doi:10.20944/preprints202311.0770.v1

62. Sitakalin, C.; Meullenet, J.F.C. Prediction of cooked rice texture using an extrusion test in combination with partial least squares regression and artificial neural networks. *Cereal Chem.* **2001**, *78*, 391–394.
63. Lin, L.H.; Lu, F.M.; Chang, Y.C. Prediction of protein content in rice using a near-infrared imaging system as diagnostic technique. *Int. J. Agric. Biol. Eng.* **2019**, *12(2)*, 195-200.