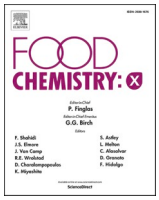
Food Chemistry: X 15 (2022) 100430

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Weight interpretation of artificial neural network model for analysis of rice (*Oryza sativa* L.) with near-infrared spectroscopy

Seungwoo Son [a,1], Donghwi Kim [b,1], Myoung Choul Choi [c], Joonhee Lee [d], Byungjoo Kim [d], Chang Min Choi [c], Sunghwan Kim [a,e,*]

[a] *Department of Chemistry, Kyungpook National University, Daegu 41566, Republic of Korea*
[b] *Oil and POPs Research Group, Korea Institute of Ocean Science and Technology, Geoje 53201, Republic of Korea*
[c] *Center for Scientific Instrumentation, Korea Basic Science Institute, Cheongju 28119, Republic of Korea*
[d] *Organic Metrology Group, Division of Chemical and Biological Metrology, Korea Research Institute of Standards and Science, Daejeon 34113, Republic of Korea*
[e] *Mass Spectrometry Convergence Research Center and Green-Nano Materials Research Center, Daegu 41566, Republic of Korea*

ARTICLE INFO

Keywords:
Artificial neural network
Prediction model
Rice
Nutrients
Near-infrared spectroscopy
Partial least squares

ABSTRACT

Prediction models for major nutrients of rice were built using near-infrared (NIR) spectral data based on the artificial neural network (ANN). Scientific interpretation of the weight values was proposed and performed to understand the wavenumbers contributing to the prediction of nutrients. NIR spectra were acquired from 110 rice samples. Carbohydrate and moisture contents were predicted with values for the determination coefficient, relative root mean square error, range error ratio, and residual prediction deviation of 0.98, 0.11 %, 44, and 7.3, and 0.97, 0.80 %, 27, and 5.8, respectively. The results agreed well with ones reported in the previous studies and acquired by the conventional partial least squares (PLS)-variable importance in projection method. This study demonstrates that the combination of NIR and ANN is a powerful and accurate tool to monitor nutrients of rice and scientific interpretation of weights can be performed to overcome black box nature of the ANN.

Introduction

Along with wheat, rice (*Oryza sativa L.*) is one of the main staple foods worldwide. The general quality evaluation of rice has been conducted for physical properties, such as grain surface, milling date, variety, and production area, as well as for major nutrients, such as carbohydrates, crude proteins, crude fats, and moisture (Rathna Priya, Eliazer Nelson, Ravichandran, & Antony, 2019; Birla et al., 2017). Wet analysis methods are used for the evaluation of the nutrients. The analysis methods differ for each nutritional component, and it takes time for sample preparation and analysis. For example, the acid hydrolysis and Röse – Gottlieb method are mainly used for crude fat analysis, and it takes at least 2 h to analyze a sample (Luo, Xing, Wang, Peng, & Li, 2017; Marto et al., 2018).

Currently, visible/near-infrared (NIR) spectroscopy is widely used in the food industry for the rapid evaluation of nutrients (Sampaio, Castanho, Almeida, Oliveira, Brites, 2019; Burns & Ciurczak, 2007). The NIR analysis method is non-destructive and enables minimal sample preparation. Moreover, it can be done rapidly compared to wet analysis methods and allows simultaneous determination of multiple nutrients. However, when food is analyzed by NIR, the absorption peaks of the major nutrients, such as carbohydrate, crude protein, crude fat, and water, overlap (Burns & Ciurczak, 2007). Therefore, the development and application of an effective data analysis method are very important. Chemometrics is a powerful tool that is gaining momentum in the analysis of NIR data. Principal component analysis (PCA) and partial least squares (PLS) have been widely used to study NIR data. Quantitative and predictive analysis of unknown samples has been done using PLS (Sampaio et al., 2019).

Recently, artificial neural network (ANN) has emerged as an alternative tool for evaluating the complex relationships between variables. ANN has been successfully applied to process NIR data obtained from

Abbreviations: ANN, Artificial neural network; NIR, Near-infrared; NNR, Neural network regression; PCA, Principal component analysis; PLS, Partial least squares; RER, Range error ratio; RPD, Residual prediction deviation; rRMSEC, Relative root mean square error of calibration; rRMSEP, Relative root mean square error of prediction; VIP, Variable importance in projection.

* Corresponding author at: Department of Chemistry, Kyungpook National University, Daegu 41566, Republic of Korea.
E-mail address: sunghwank@knu.ac.kr (S. Kim).

[1] These authors equally contributed to this work.

https://doi.org/10.1016/j.fochx.2022.100430
Received 4 May 2022; Received in revised form 8 August 2022; Accepted 10 August 2022
Available online 12 August 2022

biodiesel (Skvaril, Kyprianidis, & Dahlquist, 2017), asparagus (Richter, Rurik, Gurk, Kohlbacher, Fischer, 2019), green tea (Yu, Low, & Zhou, 2018), wheat flour (Barbon Junior et al., 2020), chicken meat (Kato et al., 2020), and infant formula (Liu et al., 2021). One of the critical advantages of ANN over conventional statistical methods is that both linear and nonlinear relationships can be predicted without prior transformations. In addition, ANN is flexible because it can be used either for regression or classification. Therefore, ANN can be a powerful tool to analyze NIR data.

When ANN is used, avoiding overfitting is very important. Overfitting occurs when the model learns too many details in the training data so that the model becomes effective in predicting data in the training set but not in the test set. Overfitting of ANN is evaluated by having a separate data set for testing. Typically, 20 ~ 30 % of the data are reserved for testing. Another important drawback of the ANN approach is that scientific interpretation is difficult. It is considered to have a 'black box' nature because of the difficulty in extracting useful relational information (Zhang, Beck, Winkler, Huang, Sibanda, & Goyal, 2018). The model can provide good correlation or classification. However, usefulness of ANN for study can be limited unless we understand meaning of the variables used to construct the ANN. Therefore, there has been continuous effort in computer science community to build interpretable and/or explainable neural network (Zhang, Wu, & Zhu, 2018; Zhang, Tiňo, Leonardis, & Tang, 2021; Islam, Ahmed, Barua, Begum, 2022).

Recently, it was proposed that weight interpretation can be used to understand the scientific interpretation ANN model built based on ultrahigh resolution mass spectra obtained from coal contaminated soils. It was shown that the compounds contributing to coal contents of soil could be identified by weight interpretation (Solihat et al., 2022). However, the weight interpretation has not been applied to study data obtained from food. In this study, constructed predictive models for nutrients using an ANN was developed from the NIR spectra of 110 rice samples. Scientific interpretation of the predictive model was based on its weight values, and conventional PLS analysis was implemented for comparison with the results and interpretations of the proposed method. To the best of our knowledge, this is the first study to use the weight analysis for scientific interpretation of the ANN model of NIR data obtained from food.

Materials and methods

Sample preparation

A total of 110 rice samples of different origins and varieties were purchased from the Korean local markets. Rice samples from different origins were used to obtain as wide range of nutrient values as possible. The rice was comminuted to a particle size of 500 μm using a variable speed rotor mill (Pulverisette 14 classic line, Fritsch, GmbH, Idar-Oberstein, Germany), then transferred to a 10 mL transparent vial with a height of 5 cm, and the inlet of the vial was sealed with Teflon tape to minimize external contamination. The rice powder samples were stored at −20 ℃ for 24 h to maintain the same external conditions, and further stored at room temperature for 24 h before NIR and nutrient analysis.

Nutrient analysis

Crude protein

The crude protein was analyzed using an automatic protein analyzer. In the preparation process for decomposition of the test solution, approximately 1 g of rice powder sample was precisely taken and placed in a decomposition tube, and 2 tablets of a decomposition accelerator (1 tablet: 3.5 g K_2SO_4/3.5 mg Se, 1000 Kjeltabs Se/3.5, FOSS, Hillerod, Denmark) and 12 mL of concentrated sulfuric acid were added. Then, it is decomposed at 420 ℃ for 60 min in a decomposition device and cooled to room temperature. 80 mL of distilled water was added to the test solution, 25 mL of a mixing indicator (0.1 % methyl red/0.1 % bromocresol green/4 % boric acid) was mixed, and then put into an erlenmeyer flask and analyzed in an automatic analyzer (AUTOMATIC PROTEIN/FAT ANALYZER, FOSS, Hillerod, Denmark). The crude protein was calculated through Eq. (1). Atomic weight of nitrogen is recorded as (14.01). The sulfuric acid molarity is denoted by M. The nitrogen Kjeldahl coefficient is recorded as F (5.95).

$$Crude\ protein\ (\%) = \frac{(HCl(mL) - Blank\ test\ (mL)) \times M \times 14.01}{Sample\ volume\ (mg)} \times F \times 100 \quad (1)$$

Crude fat

Crude fat analysis was performed using the acid decomposition method and the Roese-Gottlieb method. About 2 g of rice powder sample was put into a beaker, 2 mL of ethanol and 10 mL of hydrochloric acid were added, it was heated while mixing for 20 to 40 min in a water bath at 70 to 80 ℃. After that, 10 mL of ethanol is added, and 10 g of the test solution cooled to room temperature is put into a majonnier tube. Add water to the majonnier tube to make 11 mL test solution, and mix while heating to 40 ℃ ~ 50 ℃. To the test solution, 1.5 mL of ammonium hydroxide and 10 mL of ethanol were added and mixed well. Next, add 25 mL of ether to mix the test solution, open the cap and blow off the ether vapor. Close the cap again, mix for 1 min, add 25 mL of benzine and mix for 1 min. When the supernatant liquid was completely transparent through centrifugation at 600 rpm, the supernatant liquid was transferred to an erlenmeyer flask, and 15 mL of ether and benzine were added to the remaining test solution, and the above operation was repeated 3 times. Finally, wash the cap outlet and funnel of the majonnier tube with a 1:1 vol ratio mixture of ether and benzine, which mix with the test solution in an erlenmeyer flask. After the solvent of the test solution was blown out in a water bath, it was dried for constant-weight in a dryer at 100 ± 2 ℃ and the crude fat was calculated through Eq. (2). The weight of blank sample plate is recorded as (W_0, g). The weight of the sample plate containing the crude fat is marked with (W_1, g). And the sampling weight is denoted by (S, g).

$$Crude\ fat\ (\%) = \frac{W_1 - W_0}{S} \times 100 \quad (2)$$

Moisture

Moisture was analyzed using the atmospheric pressure heating drying method. In this method, the sample is dried under atmospheric pressure at a temperature of 105 ℃ slightly higher than the boiling point of water, and the reduced moisture content is measured. Add 5 g of rice powder sample to the pre-heated and constant-weighted sample plate, and dry for 5 h in a dryer at 110 ℃. Then, after cooling for 30 min in a desiccator, the total weight of the sample plate and the sample is measured. Dry the sample plate for 2 h and repeat the above analysis until the constant- weight is reached. The moisture content of the constant-weighted sample was calculated through Eq. (3). The weight of blank sample plate is recorded as (W_0, g). The weight of the sample plate containing the moisture is marked with (W_1, g). And the sampling weight after drying is indicated by (W_2, g).

$$Moisture\ (\%) = \frac{W_1 - W_2}{W_1 - W_0} \times 100 \quad (3)$$

Ash

Ash was analyzed by direct ashing method. The sample was put in a constant-weight incineration crucible, preliminarily carbonized in an electric muffle furnace (JSMF-270T, JSR, Gongju, Korea), and then incinerated for 12 h so that the entire sample became grayish-white color in an incinerator (J-FM3, JISICO, Seoul, Korea) at 550 to 600 ℃. Then, after cooling to 200 ℃, it was allowed to cool in a desiccator to obtain the constant-weight and calculated as a percentage (%) of the

weight of the sample through Eq. (4). The weight of blank sample crucible is recorded as (W_0, g). The weight of crucible and sample before ashing is marked with (W_1, g). The crucible and ash weight after drying is indicated by (W_2, g).

$$Ash(\%) = \frac{W_2 - W_0}{W_1 - W_0} \times 100 \quad (4)$$

Carbohydrate

Carbohydrates were calculated as 100 minus the sum of the other nutrients content (crude protein, crude fat, moisture and ash).

The list of rice and their nutrient values are listed in supplementary material table S1.

NIR analysis

An FT-IR/NIR spectrophotometer and the NIR reflectance accessory (NIRA; Frontier, PerkinElmer, Inc., Waltham, MA, USA) were used in this study. FT-IR spectra (50 scans per spectrum) were recorded at a resolution of 16 cm^{-1} using the CaF_2 beam splitter and InGaAs detector (optimum range: 10000 ~ 4000 cm^{-1}). The measurement was done from powdered rice sample placed in a 10 mL transparent vial at a height of about 5 cm. 50 scans of data were averaged for each spectrum. After each measurement, the sample was vortexed for 1 min and seven repeated measurements were performed. The NIR was calibrated through a matrix scan. To achieve random sample particle distribution, the sample was blended using a vortex mixer (KMC − 1300 V, Vision Scientific, Co., Daejeon, Korea) for 5 min before each measurement. Each rice sample was analyzed seven times to confirm reproducibility. As a result, a total of 770 spectra were obtained from 110 rice samples. The obtained NIR spectra are provided in Fig. S1.

Neural network regression and PLS regression

Neural network regression

The NIR data were combined using the 'outerjoin' function in MATLAB (version R2021b; MA, USA). The neural network regression (NNR) was performed using the Deep Learning Toolbox in MATLAB. The regression model was built based on randomly selected 616 spectra obtained from 88 rice samples, and the remaining 154 spectra from 22 rice samples were used to test the ANN model. The prediction model for carbohydrate content was generated using regression learner with one hidden layer having 25 nodes. *K*-fold cross-validation with $n = 5$ and sigmoid activation function were used. Data standardization was not used. Models for moisture, fat, and protein contents were built under the same conditions as those used for carbohydrate content.

A feed-forward neural network with three layers (an input layer, a hidden layer, and an output layer) was used in this study. The layers are pictorially described in Fig. 1. The raw NIR spectra was used for the input layer. The hidden layer receives the information from the input layer and processes them according to the Eq. (5). The obtained value is sent to the output layer which will also process the information from the hidden layer and give the output based on the Eq. (6). The interconnection of the nodes between the layers can be divided into two basic classes, namely the feedforward neural network and recurrent neural network.

The input signals are the NIR absorbance values of rice measured at i wavenumber (A_i). There is one input node per wavenumber in a spectrum. The weighted input signals in the input nodes are transferred to the hidden layer. Node i in the input layer is connected to node j in the hidden layer by the weighting factor W_{ij}. These weights are adjusted during the learning process. The value of node m (N_m) in the hidden layer is calculated by Eq. (5).

$$N_m = (w_{1m} \times A_1) + (w_{2m} \times A_2) + \cdots + (w_{nm} \times A_n) + b_m \quad (5)$$

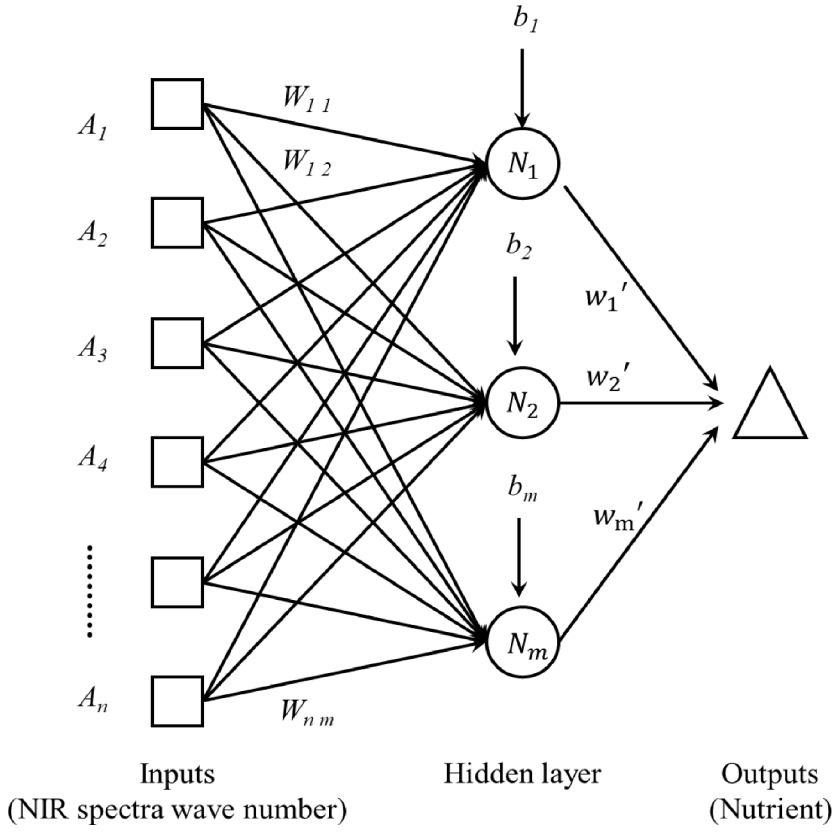


Fig. 1. Schematic representation of the feed-forward neural network used in this study.

The predicted value (V) is calculated by combining the values for the nodes, as shown in Eq. (6).

$$V = (w_1 \times N_1) + (w_2 \times N_2) + \cdots + (w_m \times N_m) + b \quad (6)$$

The final weight value (w_n in Eq. (6)) can be examined to determine the top contribution nodes. The larger the weight value, the greater the contribution. The nodes with a major contribution can be further examined to determine the wavenumbers that contribute to the nodes. The larger the weight of the node (w_{nm} in Eq. (5)), the larger the contribution of the corresponding wavenumber.

PLS regression

The PLS regression analysis was done using the Statistics and Machine Learning Toolbox in MATLAB. The 'plsregress' function was used for the analysis. Variable importance in projection (VIP) scores were calculated as the weighted sum of the squared correlations between the PLS components and the original variable (Cocchi, Biancolillo, & Marini, 2018).

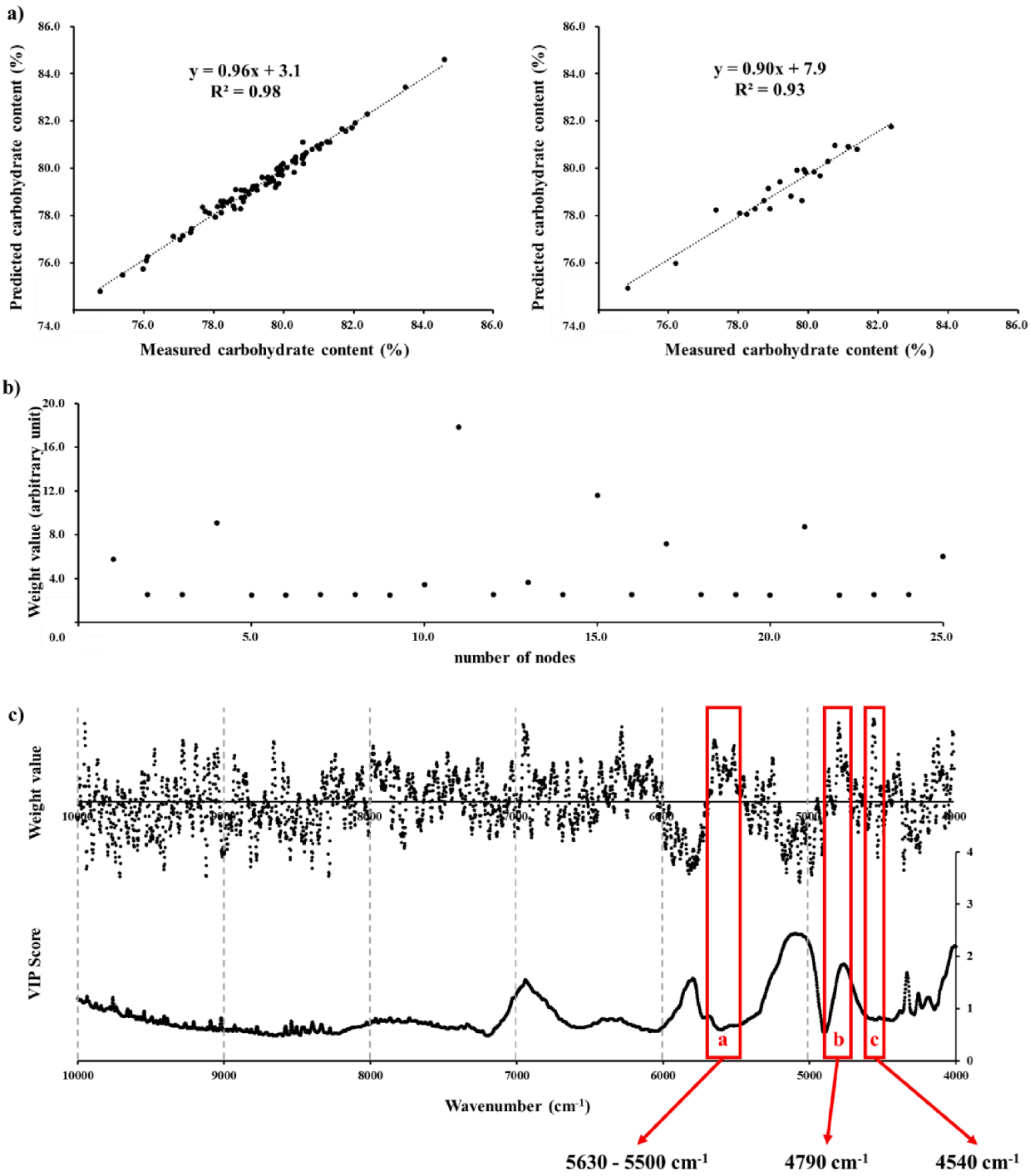


Fig. 2. Plots presenting a) correlation between measured and predicted carbohydrate contents of 88 (left) and 22 (right) rice samples based on the ANN model, b) weights of nodes used to predict the final value (refer to equation (2), and c) weight values of 11th node and VIP score from PLS analysis.

Results and discussion

Carbohydrate of rice

The developed prediction model for carbohydrate content was applied to 616 NIR spectra used for the prediction and 154 NIR spectra reserved for evaluation. The obtained values from the seven replicates were averaged and the averaged values are provided in Fig. 2a. The data obtained from 88 rice samples were used to build the ANN model and ones from 22 rice samples were to test the model. The raw data used to plot Fig. 2b are tabulated in Table S2. There was a good correlation (R^2 = 0.98 and 0.93) between the measured and predicted carbohydrate contents for both data sets. Moreover, the relative root mean square error of calibration (rRMSEC) and of prediction (rRMSEP) were 0.11 % and 0.22 %, respectively, for 616 and 154 NIR data sets, respectively. The range error ratio (RER) and the residual prediction deviation (RPD) were 44 and 7.3 for the prediction model. Based on the results obtained from the evaluation data set, we are confident that overfitting is not a problem for the developed model of carbohydrate content.

To further examine the validity of the proposed model, the weights of the ANN were analyzed. As discussed in section 2.4, one interesting characteristic of ANN is that the weights of the predicted values can be

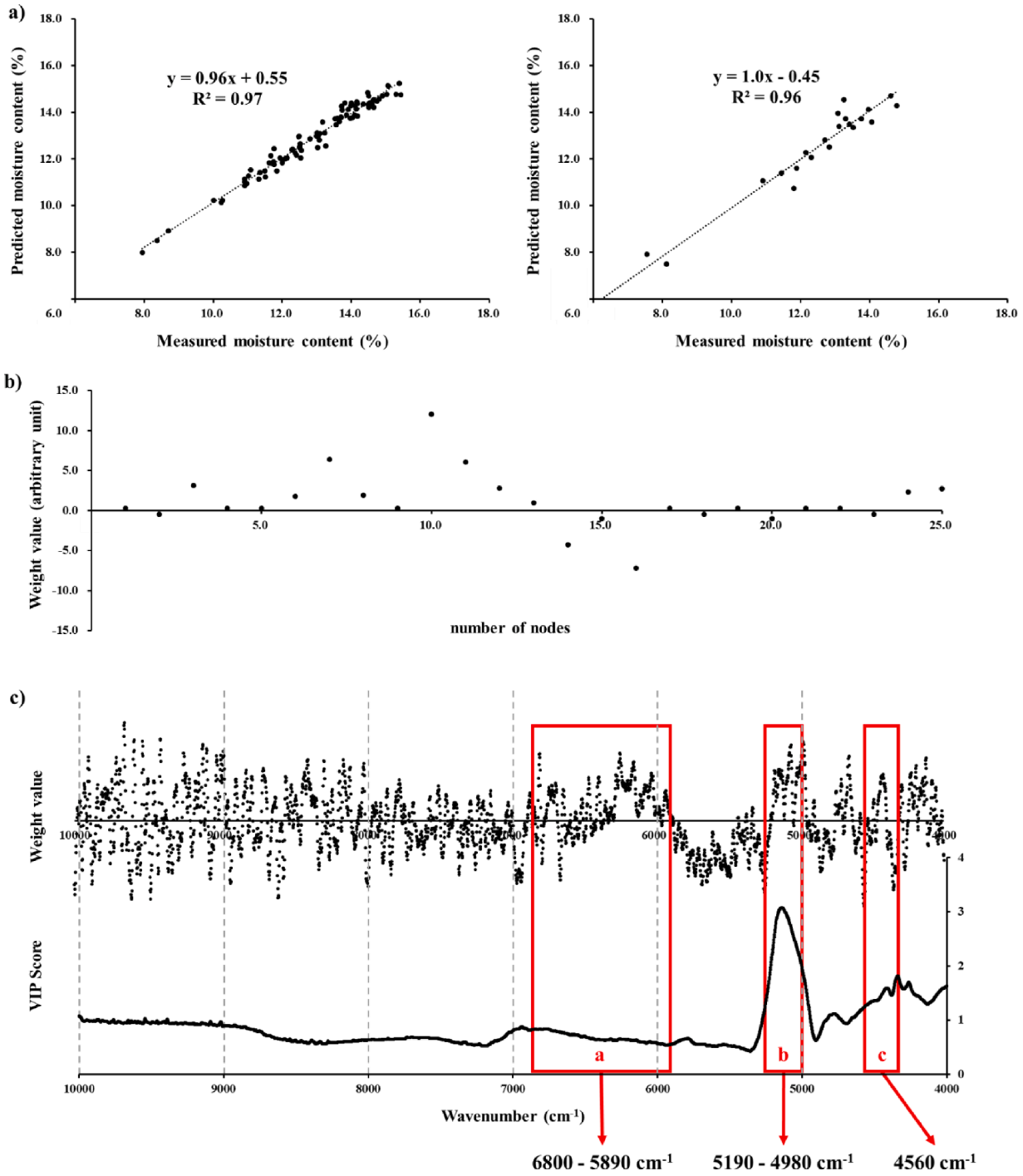


Fig. 3. Plots presenting a) correlation between measured and predicted moisture contents of 88 (left) and 22 (right) rice samples based on the ANN model, b) weights of nodes used to predict the final value (refer to equation (2), and c) weight values of 10th node and VIP score from PLS analysis.

examined to identify the nodes with a major contribution (refer to Eq. (6)). The weights of the predicted values are plotted in Fig. 2b. These data suggest that node 4, 11, 15, and 21 have major contributions to the prediction, and the weights associated with node 11 have the largest values. The weight values of node 11 are provided in Fig. 2c.

High positive weight values were observed at approximately 5630 ~ 5500 cm^{-1}, 4790 cm^{-1}, 4540 cm^{-1} (box a, b, and c in Fig. 2c). In previous studies, it was reported that the OH stretching/CO stretching combination and the CH combinations of polysaccharides presented a NIR peak around 4440 cm^{-1} (Burns & Ciurczak, 2007; Cozzolino, Phan, Netzel, Smyth, & Sultanbawa, 2021; Li, Wang, Du, Diallo, & Xie, 2017) and the peak at 4790 cm^{-1} can be assigned to the CH combinations in sugars (Baeva et al., 2020). The broad peak at 5630 ~ 5500 cm^{-1} can be associated with the CH_2 stretching overtones (Baeva et al., 2020). The previously reported peak locations match well with those observed in this study.

The data presented in Fig. 2b show that the interpretation based on neural network calculation is consistent with the current knowledge on carbohydrate NIR spectra. These findings can serve as evidence of validity for the current approach.

To compare the ANN approach with the conventional method, PLS analysis was performed on the carbohydrate content of rice and NIR data. For the PLS calculation of carbohydrate content, ten PLS components were used. The plot of the percentage of variance explained by the PLS components is presented in Fig. S2a. Six PLS components could explain 84 % of the variance. The prediction results based on the PLS components are presented in Fig. S2b. R^2 was 0.84 when six factors were used. It is apparent that better prediction is observed by employing neural network calculation compared to the conventional PLS method. The VIP scores were calculated, and the results were plotted versus wavenumber (Fig. 2c, bottom). The VIP score can be used to find the variables (in this case, the wavenumber) that contribute to the predicted *y*-values. The peak at 4790 cm^{-1} had high VIP values, and it agreed well with the results from ANN. However, differences between ANN and PLS were also observed. For example, the broad peak between 5630 ~ 5500 cm^{-1} had high weight values but low VIP values.

Moisture of rice

For the evaluation of the developed moisture content prediction model, the training set of 616 NIR data and prediction set of 154 NIR data were respectively applied. The calculation results are shown in Fig. 3a. There was a good correlation (R^2 = 0.97 and 0.96) between the measured and predicted moisture content for both data sets. The rRMSEC and rRMSEP between measured and predicted values were 0.85 % and 1.5 % for the 616 and 154 NIR data sets, respectively. The RER and RPD were 27 and 5.8 for the predictive model. The good correlation obtained from the evaluation data set suggests that overfitting is not a problem with the developed moisture content model.

The wavenumbers contributing to the prediction of moisture content were investigated by examining the weights of the nodes. The weights of the 25 nodes to the output value are plotted in Fig. 3b. Node 7, 10, and 11 had higher weight values than the other nodes, and node 10 had the largest weight value.

All three nodes (7, 10, and 11) had strong and broad positive weight values at 6800 ~ 5890 cm^{-1} and 5190 ~ 4980 cm^{-1} (box a and b in Fig. 3c). It is well documented that moisture produces a broad peak at 6940 cm^{-1} (first overtone of O—H stretches) and another at 5210 cm^{-1} (O—H stretch/deformation combination, second overtone of O—H bends) (Li et al., 2019; Jin, Shi, Yu, Yamada, & Sacks, 2017; Guan, Liu, Huang, Kuang, & Liu, 2019). The positive weight values observed at around 4500 cm^{-1} (box c in Fig. 3c) have been assigned to water bound to minerals or protein in previous studies (Roberts & Cozzolino, 2017; Yüceer & Caner, 2020).

The PLS analysis was performed on the moisture content, and the results are shown in Fig. S3. The PLS of moisture content was calculated through ten PLS components. The percentage of variance explained by the PLS component is shown in Fig. S3a. In Fig. S3b, five PLS components could demonstrate for 92 % of the variance and R^2 was 0.92. It was found that the neural network calculation method was better than the conventional PLS method for predicting moisture content. The VIP scores were calculated, and the VIP score versus wavenumber plot was shown in Fig. 3c (bottom). The moisture prediction results were shown the peaks with high VIP scores at 5190 ~ 4980 cm^{-1} and 4560 cm^{-1}, which matched the ANN results well. However, differences were also observed in ANN and PLS results. For example, the broad peak between 6800 ~ 5890 cm^{-1} had high weight values but low VIP values.

Protein of rice

The calculated protein content prediction model was applied to the NIR spectrum, respectively, and the results are shown in Fig. 4a. The correlation between protein contents measured values and predicted values in both data sets was observed as R^2 = 0.98 (left) and 0.92 (right), indicating that this was a good predicted result. The rRMSEC and rRMSEP between the measured and predicted values were 0.7 % (left) and 1.2 % (right) for 616 and 154 NIR data sets. When the RER and RPD values were checked for accuracy evaluation to verify their suitability for analysis, the RER and RPD were confirmed to be suitable for predictive analysis of protein content at 31 and 7.0 for predictive models.

By examining the weights of the nodes, the wavenumbers contributing to the prediction of protein content were investigated. The weights of 25 nodes to the output value are plotted in Fig. 4b. Node 10, 12, and 23 had appreciably higher weight values than the other nodes, and node 23 had the largest weight value.

The nodes have positive weights at 5950 ~ 5750 cm^{-1} (box a in Fig. 4c), 4880 cm^{-1} (box b in Fig. 4c) and 4560 cm^{-1} (box c in Fig. 4c). An amide bond yields a broad peak at 4650 ~ 4500 cm^{-1}. Especially, the peak at 4900 ~ 4800 cm^{-1} is often used to quantify protein (Ishigaki & Ozaki, 2020; Ishigaki et al., 2021). In our work, the broad peak at 5950 ~ 5750 cm^{-1} (box a in Fig. 4c) also had positive weights. Bands in the region of 6250 ~ 5880 cm^{-1} have been attributed to the overtones of C—H, C—N, and N—H groups of protein (Qiu, Lü, Lu, Xu, Zeng, & Shui, 2018). Therefore, the interpretation based on ANN agrees with the current knowledge on protein analysis from the NIR spectra of rice.

For comparative evaluation of protein content analysis of ANN computations and conventional PLS method, the PLS method was performed with the same data sets. Fig. S4 shown that ten PLS components were used to predict protein content and that 75 % ratio of variance could be explained by six PLS components (Fig. S4a). The prediction results based on six PLS components were presented in Fig. S4b. The correlation R^2 between the measured value of the protein contents and the predicted values was 0.75 for six PLS components. Based on the analysis results, it was confirmed that the protein prediction operation using artificial neural network computation (R^2 = 0.98) was better than the conventional PLS calculation (R^2 = 0.75). The VIP scores have been calculated, and the obtained value versus wavenumber plot is shown in Fig. 4c (bottom). For protein prediction, the weight distribution matched the VIP score well. The peaks at 5950 ~ 5750 cm^{-1}, 4880 cm^{-1} and 4560 cm^{-1} had high VIP scores and weight values.

Fat of rice

A prediction model was generated and evaluated for fat content and applied to 616 and 154 NIR spectra, respectively. The calculated results are provided in Fig. 5a. Like that observed for the carbohydrate, moisture, and protein contents, there was a good correlation (R^2 = 0.92 and 0.87) between the measured and predicted fat contents for both sets of data. However, the rRMSEC and rRMSEP between the measured and predicted values was 8.5 % and 10.7 % for 616 and 154 NIR data sets, respectively. The RER and RPD were 13 and 3.6 for the prediction model. Thus, the prediction was not as accurate as the other nutrients

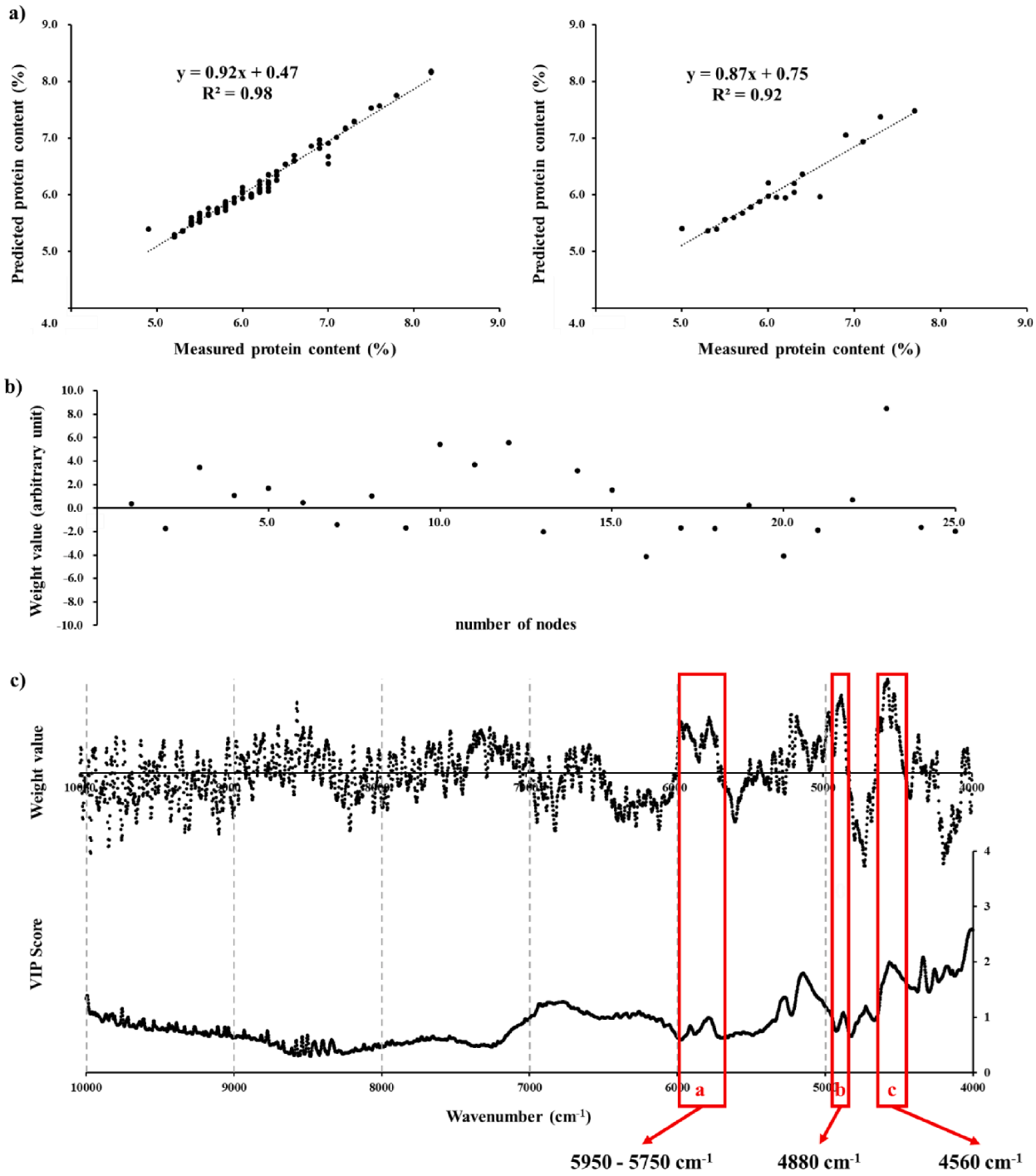


Fig. 4. Plots presenting a) correlation between measured and predicted protein contents of 88 (left) and 22 (right) rice samples based on the ANN model, b) weights of nodes used to predict the final value (refer to equation (2), and c) weight values of 23th node and VIP score from PLS analysis.

described in the above sections. This lower accuracy of prediction can be attributed to the low content of fat in rice, which averaged about 1 % of the total weight of rice.

The weights of 25 nodes to the output value are plotted in Fig. 5b. Examining the weight of nodes to the final output, node 8 and 19 had the highest positive values. The peaks at 5800 cm^{-1} and 4300 cm^{-1} had the largest positive weights (box a and b in Fig. 5c). In previous studies, the peaks with maximum intensity at 5800 cm^{-1} and around 4000 cm^{-1} were mainly associated with rice germ (Malegori et al., 2020). Therefore, the interpretation of the weights is in agreement with the results from previous works.

PLS analysis was performed for the fat content, and the results are shown in Fig. S5. For the PLS calculation of fat content, ten PLS components were used. The percentage of variance explained by the PLS component is shown in Fig. S5a. Six PLS components could explain 86 % of the variance, and the prediction results based on six PLS components are presented in Fig. S5b. R^2 was 0.86 with six PLS components. Fig. 5c (bottom) shown the VIP score and wavenumber plot. For the predicted values of the fat content, the distribution were consistent with the VIP scores. The representative peaks of the fat contents, 5800 cm^{-1} and 4300 cm^{-1}, both had high VIP scores and weight values. Therefore, it was concluded that both techniques successfully identified the

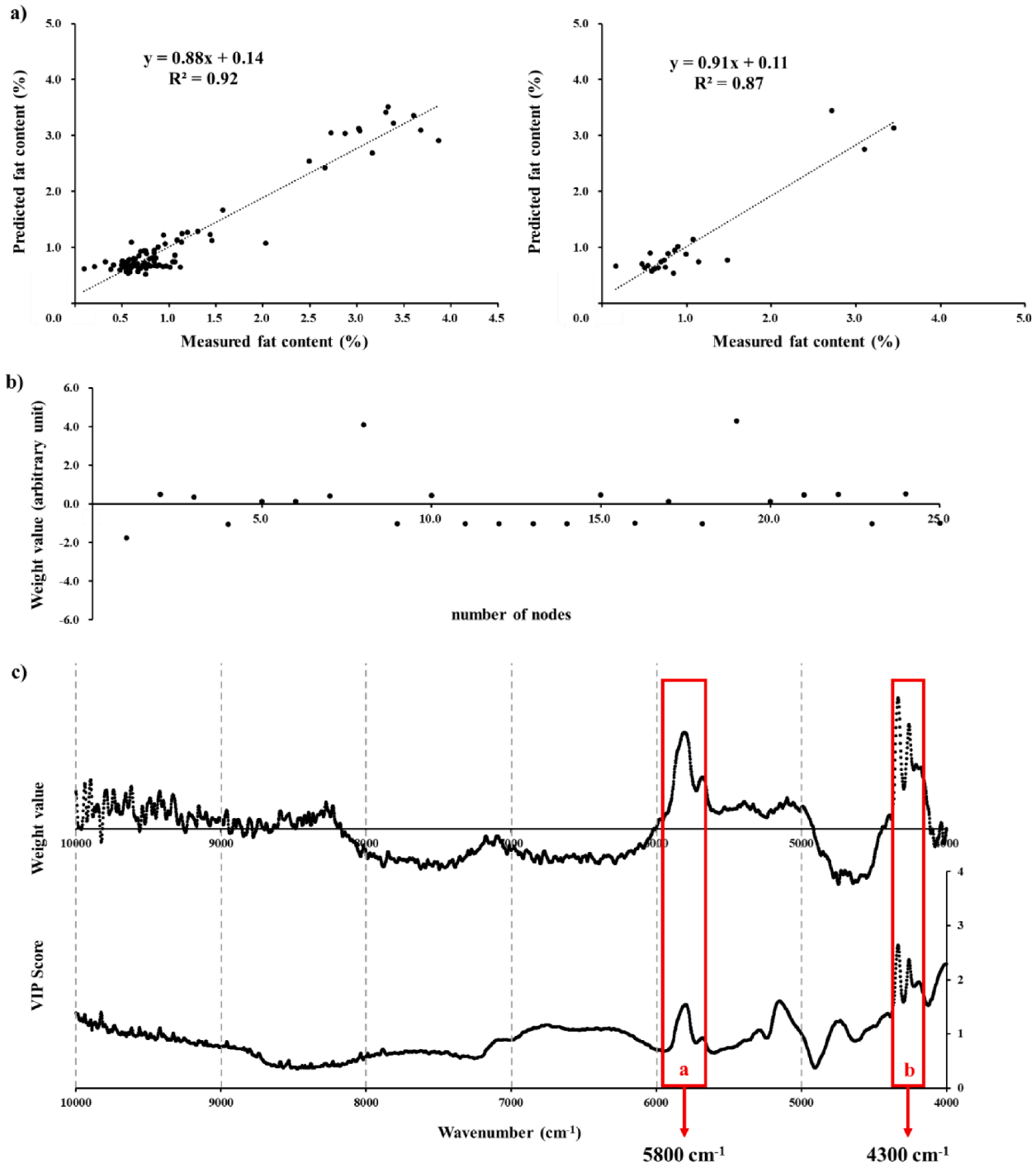


Fig. 5. Plots presenting a) correlation between measured and predicted fat contents of 88 (left) and 22 (right) rice samples based on the ANN model, b) weights of nodes used to predict the final value (refer to equation (2), and c) weight values of 19th node and VIP score from PLS analysis.

important variables (wavenumbers) contributing to fat content.

Conclusions

In this study, ANN was applied to construct prediction models for the carbohydrate, protein, moisture, and fat contents of rice. A very accurate prediction can be achieved based on the proposed method. The RPD values of the ANN based models for carbohydrate, moisture, protein, and fat were 3.27, 5.8, 7.0 and 3.6. The RER values for the nutrients were 44, 27, 31, and 13. Pinto, Ribeiro and Farinas (2018) suggested that RPD and RER values of the model should be larger than 3 and 10, respectively (Pinto, Ribeiro, & Farinas, 2018). Therefore, excellent models for prediction of the nutrients can be built based on ANN. In addition, a scientific interpretation of the weight values was made to understand the wavenumbers contributing to the prediction of nutrients. The interpretation, which was based on the weight values, was in good agreement with the conventional PLS method. In addition, ANN provided improved prediction compared to PLS. This study shows that the NIR – ANN combination is a powerful tool to monitor the nutrient status of rice. The approach used in this study can be applied to other types of food, and further study is currently being conducted in this area.

CRediT authorship contribution statement

Seungwoo Son: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Donghwi Kim:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Visualization. **Myoung Choul Choi:** Conceptualization, Methodology. **Joonhee Lee:** Conceptualization, Methodology, Resources. **Byungjoo Kim:** Conceptualization, Methodology, Resources. **Chang Min Choi:** Conceptualization, Methodology, Resources. **Sunghwan Kim:** Conceptualization, Methodology, Software, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgments

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (NRF-2020R1A4A1018393), Korea institute of Marine Scinece & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (KIMST-20220534), Korea Environment Industry & Technology Institute (KEITI) through Technology Development Project for Safety Management of Household Chemical Products, funded by Korea Ministry of Environment (MOE) (202002960004148501711 7) and the Korea Basic Science Institute [grant number D110100].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2022.100430.

References

Baeva, E., Bleha, R., Sedliaková, M., Sushytskyi, L., Švec, I., Čopíková, J., … Synytsya, A. (2020). Evaluation of the Cultivated Mushroom Pleurotus ostreatus Basidiocarps Using Vibration Spectroscopy and Chemometrics. *Applied Sciences, 10*(22). https://doi.org/10.3390/app10228156

Barbon Junior, S., Mastelini, S. M., Barbon, A. P. A. C., Barbin, D. F., Calvini, R., Lopes, J. F., & Ulrici, A. (2020). Multi-target prediction of wheat flour quality parameters with near infrared spectroscopy. *Information Processing in Agriculture, 7*(2), 342–354. https://doi.org/10.1016/j.inpa.2019.07.001

Birla, D. S., Malik, K., Sainger, M., Chaudhary, D., Jaiwal, R., & Jaiwal, P. K. (2017). Progress and challenges in improving the nutritional quality of rice (*Oryza sativa L.*). *Critical Reviews in Food Science and Nutrition, 57*(11), 2455–2481. https://doi.org/10.1080/10408398.2015.1084992

Burns, D. A., & Ciurczak, E. W. (2007). *Handbook of Near-Infrared Analysis* ((3rd ed.).). CRC Press.

Cocchi, M., Biancolillo, A., & Marini, F. (2018). Chapter Ten - Chemometric Methods for Classification and Feature Selection. In Jaumot, J., Bedia, C., & Tauler, R. (Eds.), *Comprehensive analytical chemistry*, *82*, 265-299. Elsevier. doi: 10.1016/bs.coac.2018.08.006.

Cozzolino, D., Phan, A. D. T., Netzel, M., Smyth, H., & Sultanbawa, Y. (2021). Assessing the interaction between drying and addition of maltodextrin to Kakadu plum powder samples by two dimensional and near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 247*, Article 119121. https://doi.org/10.1016/j.saa.2020.119121

Guan, X., Liu, J., Huang, K., Kuang, J., & Liu, D. (2019). Evaluation of moisture content in processed apple chips using NIRS and wavelength selection techniques. *Infrared Physics & Technology, 98*, 305–310. https://doi.org/10.1016/j.infrared.2019.01.010

Ishigaki, M., Ito, A., Hara, R., Miyazaki, S. I., Murayama, K., Yoshikiyo, K., … Ozaki, Y. (2021). Method of Monitoring the Number of Amide Bonds in Peptides Using Near-Infrared Spectroscopy. *Analytical chemistry, 93*(5), 2758–2766. https://doi.org/10.1021/acs.analchem.0c03424

Ishigaki, M., & Ozaki, Y. (2020). Chapter 6 - Near-infrared spectroscopy and imaging in protein research. In Y. Ozaki, M. Baranska, I. K. Lednev, & B. R. Wood (Eds.), *Vibrational Spectroscopy in Protein Research* (pp. 143-176). Academic Press. doi: 10.1016/B978-0-12-818610-7.00006-2.

Islam, M. R., Ahmed, M. U., Barua, S., & Begum, S. (2022). A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Applied Sciences, 12*(3). https://doi.org/10.3390/app12031353

Jin, X., Shi, C., Yu, C. Y., Yamada, T., & Sacks, E. J. (2017). Determination of Leaf Water Content by Visible and Near-Infrared Spectrometry and Multivariate Calibration in Miscanthus. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/articles/10.3389/fpls.2017.00721.

Kato, N., Padang, R., Scott, C. G., Guerrero, M., Pislaru, S. V., & Pellikka, P. A. (2020). The natural history of severe calcific mitral stenosis. *Journal of the American College of Cardiology, 75*(24), 3048–3057. https://doi.org/10.1016/j.jacc.2020.04.049

Li, D., Li, L., Quan, S., Dong, Q., Liu, R., Sun, Z., & Zang, H. (2019). A feasibility study on quantitative analysis of low concentration methanol by FT-NIR spectroscopy and aquaphotomics. *Journal of Molecular Structure, 1182*, 197–203. https://doi.org/10.1016/j.molstruc.2019.01.056

Li, M., Wang, J., Du, F., Diallo, B., & Xie, G. H. (2017). High-throughput analysis of chemical components and theoretical ethanol yield of dedicated bioenergy sorghum using dual-optimized partial least squares calibration models. *Biotechnology for Biofuels, 10*(1), 206. https://doi.org/10.1186/s13068-017-0892-z

Liu, Y., Zhou, S., Han, W., Li, C., Liu, W., Qiu, Z., & Chen, H. (2021). Detection of Adulteration in Infant Formula Based on Ensemble Convolutional Neural Network and Near-Infrared Spectroscopy. *Foods, 10*(4). https://doi.org/10.3390/foods10040785

Luo, F., Xing, R., Wang, X., Peng, Q., & Li, P. (2017). Proximate composition, amino acid and fatty acid profiles of marine snail Rapana venosa meat, visceral mass and operculum. *Journal of the Science of Food and Agriculture, 97*(15), 5361–5368. https://doi.org/10.1002/jsfa.8425

Malegori, C., Buratti, S., Benedetti, S., Oliveri, P., Ratti, S., Cappa, C., & Lucisano, M. (2020). A modified mid-level data fusion approach on electronic nose and FT-NIR data for evaluating the effect of different storage conditions on rice germ shelf life. *Talanta, 206*, Article 120208. https://doi.org/10.1016/j.talanta.2019.120208

Marto, J., Neves, Â., Gonçalves, L., Pinto, P., Almeida, C., & Simões, S. (2018). Rice water: A traditional ingredient with anti-aging efficacy. *Cosmetics, 5*(2). https://doi.org/10.3390/cosmetics5020026

Pinto, A. S. S., Ribeiro, M. P. A., & Farinas, C. S. (2018). Fast spectroscopic monitoring of inhibitors in the 2G ethanol process. *Bioresource Technology, 250*, 148–154. https://doi.org/10.1016/j.biortech.2017.11.033

Qiu, G., Lü, E., Lu, H., Xu, S., Zeng, F., & Shui, Q. (2018). Single-Kernel FT-NIR spectroscopy for detecting supersweet corn (*Zea mays L. Saccharata Sturt) Seed Viability with Multivariate Data Analysis. Sensors, 18*(4). https://doi.org/10.3390/s18041010

Rathna Priya, T. S., Eliazer Nelson, A. R. L., Ravichandran, K., & Antony, U. (2019). Nutritional and functional properties of coloured rice varieties of South India: A review. *Journal of Ethnic Foods, 6*(1), 11. https://doi.org/10.1186/s42779-019-0017-3

Richter, B., Rurik, M., Gurk, S., Kohlbacher, O., & Fischer, M. (2019). Food monitoring: Screening of the geographical origin of white asparagus using FT-NIR and machine learning. *Food Control, 104*, 318–325. https://doi.org/10.1016/j.foodcont.2019.04.032

Roberts, J. J., & Cozzolino, D. (2017). Wet or dry? The challenges of NIR to analyse soil samples. *NIR News, 28*(4), 3–5. https://doi.org/10.1177/0960336017707884

Sampaio, P. S., Castanho, A., Almeida, A. S., Oliveira, J., & Brites, C. (2019). Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. *European Food Research and Technology, 246*(3), 527–537. https://doi.org/10.1007/s00217-019-03419-5

Skvaril, J., Kyprianidis, K. G., & Dahlquist, E. (2017). Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes: A review. *Applied Spectroscopy Reviews, 52*(8), 675–728. https://doi.org/10.1080/05704928.2017.1289471

Solihat, N. N., Son, S., Williams, E. K., Ricker, M. C., Plante, A. F., & Kim, S. (2022). Assessment of artificial neural network to identify compositional differences in ultrahigh-resolution mass spectra acquired from coal mine affected soils. *Talanta, 248*, Article 123623. https://doi.org/10.1016/j.talanta.2022.123623

Yüceer, M., & Caner, C. (2020). The effects of ozone, ultrasound and coating with shellac and lysozyme–chitosan on fresh egg during storage at ambient temperature. Part II: Microbial quality, eggshell breaking strength and FT-NIR spectral analysis. *International Journal of Food Science & Technology, 55*(4), 1629–1636. https://doi.org/10.1111/ijfs.14422

Yu, P., Low, M. Y., & Zhou, W. (2018). Development of a partial least squares-artificial neural network (PLS-ANN) hybrid model for the prediction of consumer liking scores of ready-to-drink green tea beverages. *International Food Research Journal, 103*, 68–75. https://doi.org/10.1016/j.foodres.2017.10.015

Zhang, Q., Wu, Y. N., & Zhu, S. C. (2018). Interpretable convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8827–8836).

Zhang, Y., Tiňo, P., Leonardis, A., & Tang, K. (2021). A Survey on Neural Network Interpretability. *IEEE Transactions on Emerging Topics in Computational Intelligence, 5*(5), 726–742. https://doi.org/10.1109/TETCI.2021.3100641

Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals Of Translational Medicine*, *6*(11), 216. doi: 10.21037/atm.2018.05.32.