International Journal of Agriculture, Environment and Biotechnology
Citation: *IJAEB*: **15**(03): 683-691, September 2022
DOI: 10.30954/0974-1712.03.2022.4

ASSOCIATION FOR AGRICULTURE ENVIRONMENT AND BIOTECHNOLOGY

BIOCHEMISTRY

Development and Comparison of Regression Models for Determination of Starch in Chickpea Using NIR Spectroscopy

Madhu Bala Priyadarshi[1]*, Anu Sharma[2], K.K. Chaturvedi[2], Rakesh Bhardwaj[1] and Mohar Singh[1]

[1]ICAR-National Bureau of Plant Genetic Resources (NBPGR), Pusa Campus, New Delhi, India
[2]ICAR-Indian Agricultural Statistics Research Institute, Pusa Campus, New Delhi, India

*Corresponding author: madhu74_nbpgr@yahoo.com (**ORCID ID:** 0000-0002-4667-9579)

Paper No. 1042 **Received:** 21-05-2022 **Revised:** 23-08-2022 **Accepted:** 01-09-2022

ABSTRACT

Crop quality characteristics are rapidly and efficiently assessed using near-infrared spectroscopy. Over the last several decades, NIR spectroscopy's advent and broad application have been an enormous success story in analytical technology development. NIR spectroscopy is frequently used in agricultural and food goods to identify and quantify an unlimited number of analytes. The near-infrared area has a wavelength range of 800 to 2500 nm. Machine learning approaches have proven to be highly successful at predicting various agricultural crop components. The concentration of the starch component in Chickpea (*Cicer arietinum* L.) whole-grain flour was determined using NIR spectroscopy data and machine language algorithms. Starch prediction models are developed using Linear Regression (LR), Artificial Neural Network (ANN), Random Forest (RF), Support Vector Regression (SVR), and Decision Tree Regression (DTR) algorithms. Performance of the models is evaluated using measures, namely, Root Mean Square Error (RMSE), Residual Standard Error (RSE), Coefficient of Determination (R^2), and Adjusted Coefficient of Determination (adjusted R^2). It was observed that LR outperformed all other models in terms of accuracy for predicting starch components from preprocessed spectra, with RMSE, RSE, R^2 and adjusted R^2 values of 0.03, 0.04, 0.98, and 0.97, respectively. The accuracy of the ANN model is similar to that of the LR, with minor differences in RMSE, RSE, R^2 and adjusted R^2, values of 0.03, 0.04, 0.97, and 0.97, respectively.

HIGHLIGHTS

- Near-infrared spectroscopy (NIRS) with machine learning algorithms is one of the most advanced non-destructive component prediction assessment techniques available.
- The NIRS technique has been successfully used for the rapid analysis of starch, moisture, protein, and fat content in many agricultural and food products since its first application in the 1960s.

Keywords: Support Vector Regression, Artificial Neural Network, Chickpea, Near Infrared Spectroscopy, Random Forest, Linear Regression, Partial Least Squares Regression

Legumes are the most important crops due to their nutritional qualities. Legumes' seeds and powder are high in protein, carbs, vitamins and minerals, and dietary fiber (Baljeet *et al.* 2014). Chickpea (*Cicer arietinum* L.) is the world's second-largest legume in overall production, behind soybean and bean, and is primarily grown in warm regions such as India, Pakistan, Iran, Ethiopia, Mexico, and the Mediterranean region (Bar-El Dadon *et al.* 2014). It is a high-protein legume (19-29 g/100 g) (Boye *et al.* 2010), a complex carbohydrate source (60-65 g/100 g), and a good source of B-complex vitamins and minerals (Martelovidal *et al.* 2014).

NIR is a rapid technique that can identify and analyze numerous components in a sample (Acquah *et al.* 2016). Using NIR Spectroscopy (NIRS), different chemical components in samples may be swiftly measured by vibrational absorption modes of the compounds in the spectrum's NIR region (Fassio Cozzolino, 2004). The NIR region contains the doubling frequency and combining bands of

How to cite this article: Priyadarshi, M.B., Sharma, A., Chaturvedi, K.K., Bhardwaj, R. and Singh, M. (2022). Development and Comparison of Regression Models for Determination of Starch in Chickpea Using NIR Spectroscopy. *Int. J. Ag. Env. Biotech.*, **15**(03): 683-691.

Source of Support: None; **Conflict of Interest:** None

several hydrogen-containing groups in fat, protein, moisture, and carbohydrate. The characteristic vibrational information in these organic molecules can be used to determine the chemical composition of samples. NIR Spectroscopy has long been recognized as a powerful tool for analyzing chemical and physical properties without needing sample preprocessing. It has been utilized to evaluate the quality aspects of food and agricultural commodities (Batten *et al.* 1998). NIR Spectroscopy has been used successfully to determine various chemicals in a variety of culinary and industrial crops, including sesame (Sato *et al.* 2003; Kim *et al.* 2007), sweet potato (Lu *et al.* 2006), soybean (Choung *et al.* 2005), perilla and peanut (Oh *et al.* 2000), sunflower (Fassio *et al.* 2004), rice (Wu *et al.* 2004), and maize (Brenna *et al.* 2004) using Partial Least Squares Regression (PLSR).

NIR Spectroscopy is a non-destructive, quick, and straightforward approach to assess the concentration of a component in a sample. This technology is based on the characteristic absorption, transmittance, and reflectance of bands in the NIR region derived from a mixture of N-H, C-H, O-H, and C-O vibrational frequencies. The position of the bands indicates the chemical composition of the components, and the band's strength is related to the amount of hydrogen-containing group (Ghasemi *et al.* 2013).

Supervised predictive modeling uses known data to develop a model capable of predicting values for future events. It creates predictions based on past data analysis. Predictive modeling approaches include ANN, LR, SVR, DTR, and RF. Selecting the most effective predictive modeling technique at the start saves considerable time.

This study was undertaken to apply (i) NIR reflectance spectroscopy, best preprocessing techniques, variable selection techniques, extracting the wavelengths (750-2500 nm) associated with best-predicting starch in chickpea, and (ii) Establishing and comparing five different machine learning prediction models of starch in chickpeas and the wavelength of choice germplasm flour within acceptable agreement with chemical laboratory methods. The machine learning methods considered in this study are LR, ANN, RF, SVR, and DTR to predict starch concentration in chickpea germplasm flour.

MATERIALS AND METHODS

Sample Collection: To create NIR spectroscopy models for predicting the concentration of starch component in chickpea seeds, 237 chickpea germplasm accessions were selected based on variability in seed morphometry from the National Gene Bank ICAR- National Bureau of Plant Genetic Resources (NBPGR), New Delhi. Fig. 1 depicts a flowchart of the developing model for predicting a starch component in chickpeas.

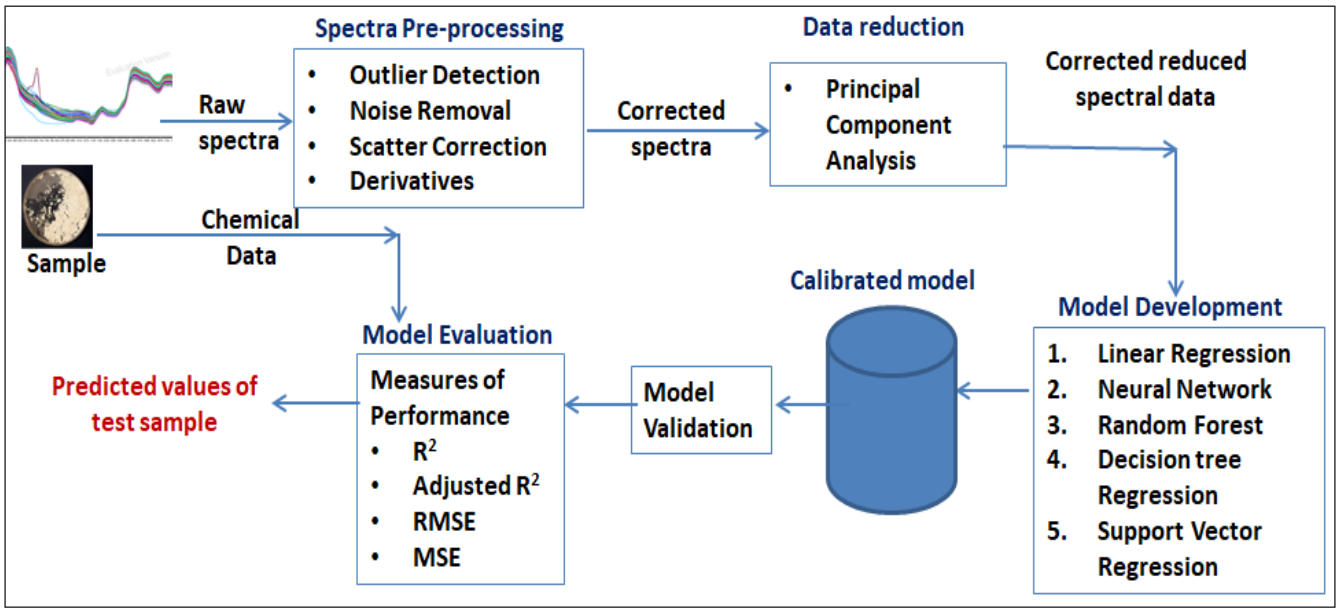

Fig. 1: Flowchart of model development

Collection of Spectral Reflectance data: A near-infrared scanning monochromator in reflectance mode was used for the NIR spectroscopic analysis. Chickpea samples were homogenised in a Foss Cyclotec mill fitted with a 0.5 mm sieve to ensure uniform particle size. To achieve consistent packing, homogenized flour was placed in a circular cuvvete with a glass window and slightly squeezed with the back cover. A Foss NIR spectroscopy 6500 cuvette spinning model was used to capture spectra with a spacing of 2 nm from the 400-2498 nm wavelength range. The reciprocal reflectance's logarithm (log 1/R) was measured at 2 nm intervals in the 400-2500 nm wavelength range. The sample was scanned 32 times, and the resulting spectra were averaged.

Spectra Division: Spectral data suffers from various challenges. Preprocessing needs to be done to get some useful information from spectral data. Before beginning preprocessing, it was also needed to determine which wavelength regions supplied the most valuable qualitative and quantitative data to predict starch component concentration in Chickpea flour. According to (Nawrocka and Lamorska, 2013) starch is present in the NIR wavelength range at 1204, 1464, 1932-2100 and 2290-2324.

These wavelength ranges are characterized by various overtones and C-H, and N-H combinations. According to (Osborne *et al.* 1993), starch bands are at 1200, 1700, 1720, and 1780 nm. According to the literature obtained from different sources, the complete NIR spectra (750 -2500 nm) was divided into five regions to find the best wavelength regions, as shown in table 1. These wavelength regions have got various combinations of first and second overtone regions. Since NIR spectral data contains subtle information in wavelength absorption intensities that is not visible as individual peaks, large wavelength regions were used to develop a model that accurately estimates component concentration in chickpea flour. To select the best prediction wavelength range for improved predictive ability, all five wavelength categories were subjected to PLS regression. PLS regression is a quantitative technique that performs spectral and response variable decomposition. The RMSE is an estimate of the standard error in prediction. R^2, RMSE is used to compare the predictive models of different sets of wavelengths. The RMSE is the average difference between measured and predicted values (Kobayashi and Salam, 2000). Lower RMSE values improve the model's predictive ability: the lower the RMSE value for a model, the better the model's prediction ability. Furthermore, as the number of wavelengths decreased, the model's dimension decreased while its predictive ability remained high (Todeschini *et al.* 1999).

Table 1: NIR wavelength ranges of Starch presence

Sl. No.	Wavelength Range *(with gap of 2 nm)*	No. of Wavelengths	File Name	RMSE	R^2
1	1204-1464	131	Range1	0.254	0.030
2	1700-1780	41	Range2	0.150	0.269
3	1932-2100	85	Range3	0.154	0.187
4	2290-2324	18	Range4	0.154	0.190
5	1204-2324	561	Range5	0.174	0.031

Selection of Preprocessing Strategies: This step covers all data manipulation activities necessary to convert raw data into a format usable with machine learning algorithms, such as data cleaning and feature extraction. NIR spectra are typically vast and include a great deal of information. The majority of the data is unrelated to the subject at hand, and some of it is simply noise. As a result, determining the link between NIR spectra and the predictive variable is a complicated process. Preprocessing is used to remove physical phenomena (noise) from spectra to improve the regression model. Another significant influence on NIR reflectance data is multiplicative interferences by scattering. Scatter correction techniques such as Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), and Standard Normal Variate de-trending were applied to remove scattering effects. SNV computes the average and standard deviation of all spectral data points. The average value is then subtracted from the acquired absorbance, and the result is divided by the calculated standard deviation. SNV-Detrending takes a second-degree least-squares fit polynomial generated from the original data and subtracts it from the data points. MSC is based on the principle that undesirable scatter effects will be removed from the data matrix before data modeling. It accomplishes this by performing two phases: the first is the estimation of correction coefficients, and the second is the correction of the recorded spectrum.

Chemometric analysis

The steps involved in developing models are (1) outlier detection, (2) data preprocessing, (3) dimensionality reduction, (4) model development, and (5) performance estimation of developed models. The flowchart in Figure 1 depicts the model development process.

1. Outlier detection: A data point that stands out from the others is known as an outlier. Outliers are measuring inaccuracies. In the presence of outliers, machine learning model development is in danger of compromising the statistics and distribution of the input data. Outliers can confuse the training process in supervised models, resulting in longer training times or the construction of less precise models. Outliers are visually identified using box plots. The q1 (25^{th} percentile), q2 (50^{th} percentile or median), and q3 (75^{th} percentile) of the data, as well as (q1–1.5*(q3–q1) and (q3 + 1.5*(q3–q1), are plotted in a box plot. Outliers are plotted as points above and below the plot, if any exist. Outliers were identified and removed from preprocessed data in this study. The raw data contained 237 samples at first. The removal of outliers decreased the data set to 192 samples.

2. Preprocessing: Spectral datasets ere preprocessed to maximize the calibration results, and several mathematical treatments using the raw spectrum data were used, with various combinations of smoothing and gap size. For example, in 2,2,2,1, the first number indicates the order of the derivative function (two is the second derivative), the second number is the gap (the length in nm) in data points over which the derivative is calculated, the third number is the number of data points (segment length) used in first smoothing, and the fourth number is the number of data points in second smoothing, which is usually set at 1 if no second smoothing is used (Shenk and Westerhaus, 1993). The optimal combination of data preprocessing was selected as the one providing a pls model with a good compromise of a low RMSE and high R^2 value.

3. Multicollinearity: It is an extreme case in which collinearity exists between three or more variables even though no pair of variables has a robust correlation. This indicates that the predictor variables are redundant. The regression model's solution becomes unstable in the presence of multicollinearity. Because of the multicollinearity, statistical judgments are less reliable. When two or more independent predictor variables are significantly associated, multicollinearity develops in a regression model analysis, resulting in a lack of unique information about the regression model. As a result, when creating a multiple regression model, these variables must be deleted. For continuous variables, linear regression is a supervised learning approach. In machine learning algorithms such as linear regression, there is the possibility that some variables may be multicollinear. Examining the correlation between each pair of explanatory variables is the most straightforward technique to find multicollinearity. If two variables are highly linked, this could be a source of multicollinearity.

Model Development: Machine learning algorithms, LR, ANN, RF, SVR, and DTR were used to develop prediction models utilizing preprocessed spectra. The Comprehensive R Archive Network (CRAN) obtained all model development packages. Creating a linear regression model employs the *lm()* (Chambers, 1992) function. The *neuralnet()* package has been used to develop neural network models. The network employs *Tanh* and *sigmoid* activation functions. In this study, "*randomForest()*" function from the R package "*randomForest*" is used for models of random forest algorithm and "*svm()*" function from the R package "*e1071*". (SVM in R for Data Classification using e1071 Package, 2021) is used for SVR algorithm. For models of decision tree regression, *rpart()* function from the R package "*Recursive Partitioning and Regression Trees (RPART)*" (Breiman, 1984) is used.

Model Evaluation: To evaluate the efficacy of the regression model, the RMSE, RSE, R^2, and adjusted R^2 statistical measures were used. The RMSE is a measure that calculates the difference between expected and actual values. It is a metric for determining how well a regression line matches the data. Residual analysis is used to assess the suitability of a regression model. As the RSE decreases, the fit of a regression model to a dataset improves. However, the greater the RSE, the worse a regression model fits a dataset. A residual plot is created and evaluated to examine the randomness of residuals. The Shapiro-Wilk (SW) test, was used to perform a normality test with a significance level of 0.05 for residues from all models. For predicting

the result of a particular event and to measure changes in one variable due to differences in another variable, R^2 is calculated. R^2 measures the strength of a linear relationship between two variables. The optimum model for each component was chosen in this investigation based on the lowest prediction RMSE and RSE and the highest R^2 and adjusted R^2 value between measured and predicted values.

RESULTS AND DISCUSSION

The descriptive statistics, including mean, and standard deviation (SD) for the starch component of chickpea samples, are shown in Table 2. As shown in Table 1, PLS regression models were created for five categories of wavelength ranges and assessed with RMSE and R^2. Range2 ranging from 1700-1780, has a minimum RMSE value of 0.150 and the highest R^2 value of 0.269. Hence wavelength range 1700-1780 was considered for the prediction of starch component for chickpea flour samples.

Table 2: Descriptive Statistics of Starch Component in Chickpea

	Starch
Mean	31.87
Standard Error	0.26
Median	32.39
Mode	33.81
Standard Deviation	3.60
Sample Variance	12.98
Kurtosis	0.12
Skewness	-0.51
Minimum	20.20
Maximum	40.23
Count	196
CV(%)	11.30
Normality (p-value) [a]	0.004

[a]*Shapiro–Wilks test of normality was used to determine the normality of the data.*

Out of 237 samples, it was discovered that 196 samples had reflectance that could be studied in NIR spectra. In 196 samples, the mean value of the starch component is 31.87, and the Shapiro-Wilk test result for normality is 0.004, indicating that the dataset is significant and suitable for starch component analysis.

Boxplots were used to detect outliers. The raw data at first had 237 samples. After eliminating the outliers, the data set was downsized to 192 samples. Preprocessing on a specific wavelength range of 1700-1780 nm using various mathematical techniques led to the development of the models. The model developed with treatment 1,2,0,2, which means spectra is passed to the first derivative with a gap size of 2 nm. It is smoothed by moving average with a 2 nm gap size. In addition to, spectra is passed to standard normal variate for scatter correction. There were 192 samples for 39 wavelengths after preprocessing.

Multicollinearity was detected, and dimensionality reduction was made using PCA.

- A correlation matrix was created for NIR spectral data with 192 samples for 39 predictor wavelengths. It was found that the 26 predictor variables correlate more significantly than 0.9 percent, indicating multicollinearity in data.
- The preprocessed spectra were decomposed into latent vectors ranked according to the amount of spectral variance explained by PCA. The first ten principal components (PCs) account for approximately 90% of the variation in NIR spectra of chickpea samples.

Prior to model development, data was split into two sets: a calibration set made up of 80% of the data (154 samples), and a validation set made up of 20% of the data (38 samples). The calibration data set, which is split into two parts: training and testing data, contains 75% of the data (116 samples) and 25% of the data (38 samples) respectively. Model development takes place with 10 PCs from PCA. The model generation method for all five algorithms created a unique and resilient model.

An equation connects predictor and response variables in linear regression model, where the exponent (power) of both of these variables is 1. A linear connection mathematically depicts a straight line when plotted as a graph. The general mathematical equation for linear regression is:

$$y = ax + b$$

y = response variable

x = predictor variable

a, b are constants

The mathematical equation for the LR model is given in equation 1.

Table 3: Performance evaluation of five machine learning models developed for prediction of starch component of chickpea

Sl. No.	Model	RMSE	*r*	R^2	Adj. R^2	RSE	RPD	p-value by Shapiro wilk test on residue
1	LR	0.03	0.99	0.98	0.97	0.04	6.50	0.17
2	ANN	0.03	0.99	0.97	0.97	0.04	6.32	0.51
3	RF	0.04	0.98	0.96	0.94	0.04	4.96	0.42
4	SVR	0.03	0.98	0.95	0.94	0.04	4.75	0.03
5	DTR	0.05	0.94	0.88	0.85	0.06	2.96	0.08

$$y = -0.192 + 0.051x_1 + 0.565x_2 + 0.462x_3 + 0.236x_4 - 0.135x_5 - 0.025x_6 + 0.101x_7 - 0.038x_8 + 0.152x_9 - 0.012x_{10} \quad \ldots(1)$$

where y is the starch component and x_i are 10 PCs from PCA.

The RF algorithm combines large-scale regression trees. The random forest model has two parameters: *ntree*, which is equal to 1 tree, and *mtry*, which is the number of input variables per node, which is 1.

Neurons are arranged in a neural network's three layers—input, hidden, and output. Equation 2 represents the neural network, where W stands for the weights vector, X for the inputs vector, and b for the bias. Equation 3 contains the sigmoid activation function that is applied.

$$y = \sum_{i=1}^{n} \left(w_i * X_i + b \right) \quad \ldots(2)$$

$$f(z) = \frac{1}{1 - e^{-y}} \quad \ldots(3)$$

Where $f(z)$ and y are an activation functions. The activation function's output ranges from 0 to 1. Here, the neural network was configured to The parameters of the neural network were set to 2, 8, 0.02, 0.3, 2000 for the hidden layers, nodes for each layer, learning rate, momentum, and iteration.

The accuracy of SVR models depends on how well the loss function, error penalty factor C, and SVR meta-parameters are configured. Additionally, the final models are significantly impacted by the choice of the kernel function. In this study, the commonly used radial basis kernel function (RBF), $K(x, x') = \exp(-|x\text{-}x'^2 / \sigma^2)$, was used. SVR model was created with C as 1, ε as 0.1, and 0.1 for the RBF kernel parameter. Number of support vectors for SVR model is 90.

The decision tree model was developed using *the rpart* algorithm. There are two major processes of *rpart* (1) tree growing and (2) splitting. Tree growing is an expansion of the tree at specific decision points, and tree pruning ignores the subtree with poor decision scores. To develop decision tree model, *rpart (formula = Starch ~. data = train data)* command is used.

RSE values for all models range from 0.075 to 0.145. A residual plot was created to assess the randomness of residues obtained from generated models. It was discovered that all residues are roughly evenly distributed around zero in the plot with no discernible pattern, implying that residues are random (Fig. 3). The Shapiro-Wilk test was used to determine the normality of residues in developed models. The residual p-value obtained for each model is shown in Table 3. The p-values for all models except SVR model are more significant than 0.05, indicating that residual data is normally distributed.

It was discovered that LR and ANN prediction models perform best in the wavelength range 1700-1780 nm, to predict Starch concentration of chickpea. The model created by the LR algorithm was discovered to have the lowest RMSE value of 0.03 and RSE values of 0.04. The R^2 and adjusted R^2 values of 0.98 and 0.97, respectively, are the highest of the five models.

The RMSE, RSE, R^2, and adjusted R^2 values of the ANN model are 0.03, 0.04, 0.97, and 0.97, respectively, with only minimal deviations from those of the LR model in terms of accuracy. Activation function gives NN their non-linearity and expressiveness. An ANN learns throughout the learning phase by changing the weights to forecast the value of the inputs' responses. It is worth noting that the tests were carried out in the lab, and the models' dependability can only be proven once

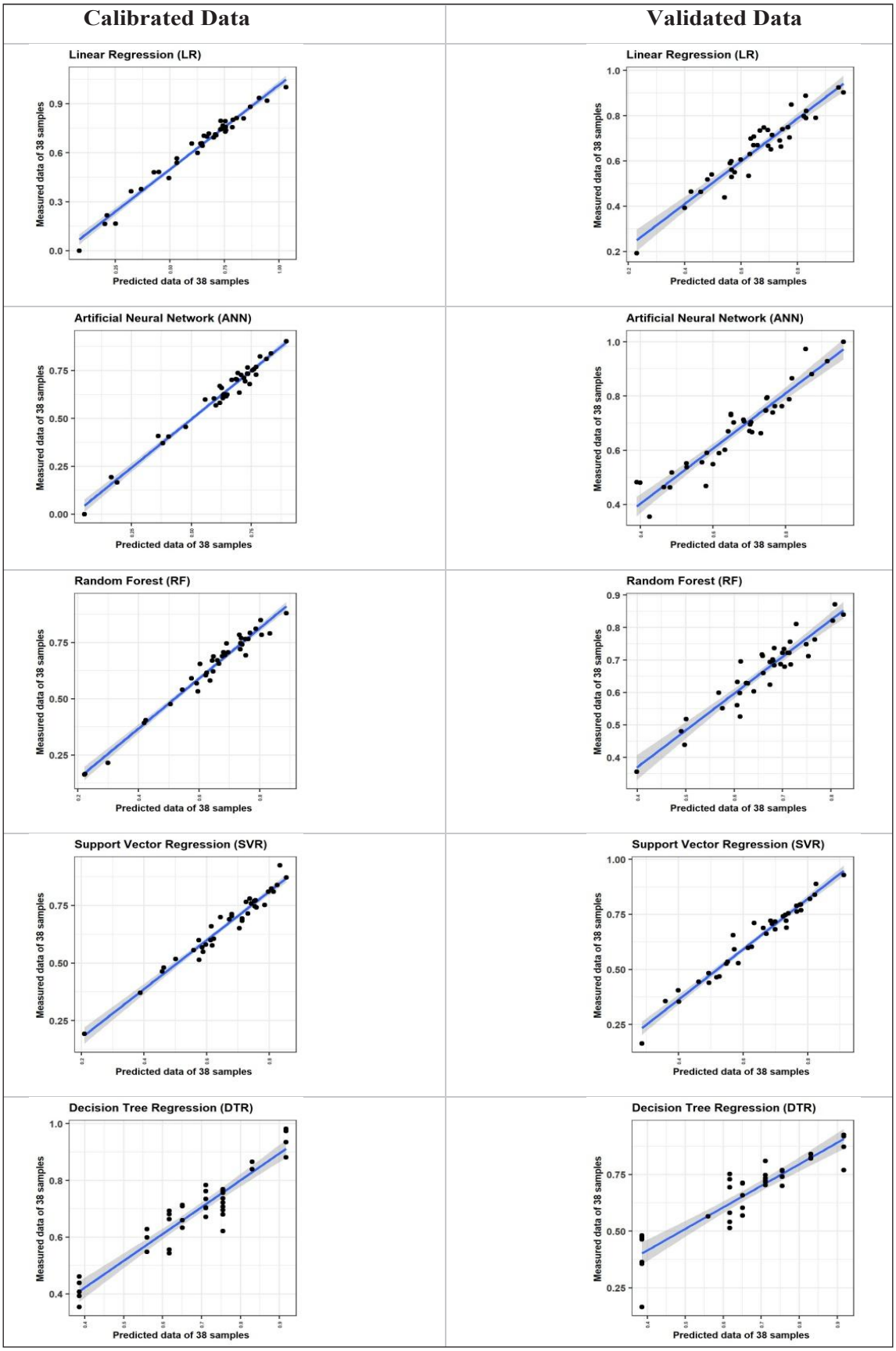

Fig. 2: Scatter plots of the measured versus the predicted values of the five machine learning algorithms for Starch component using calibrated and validated data

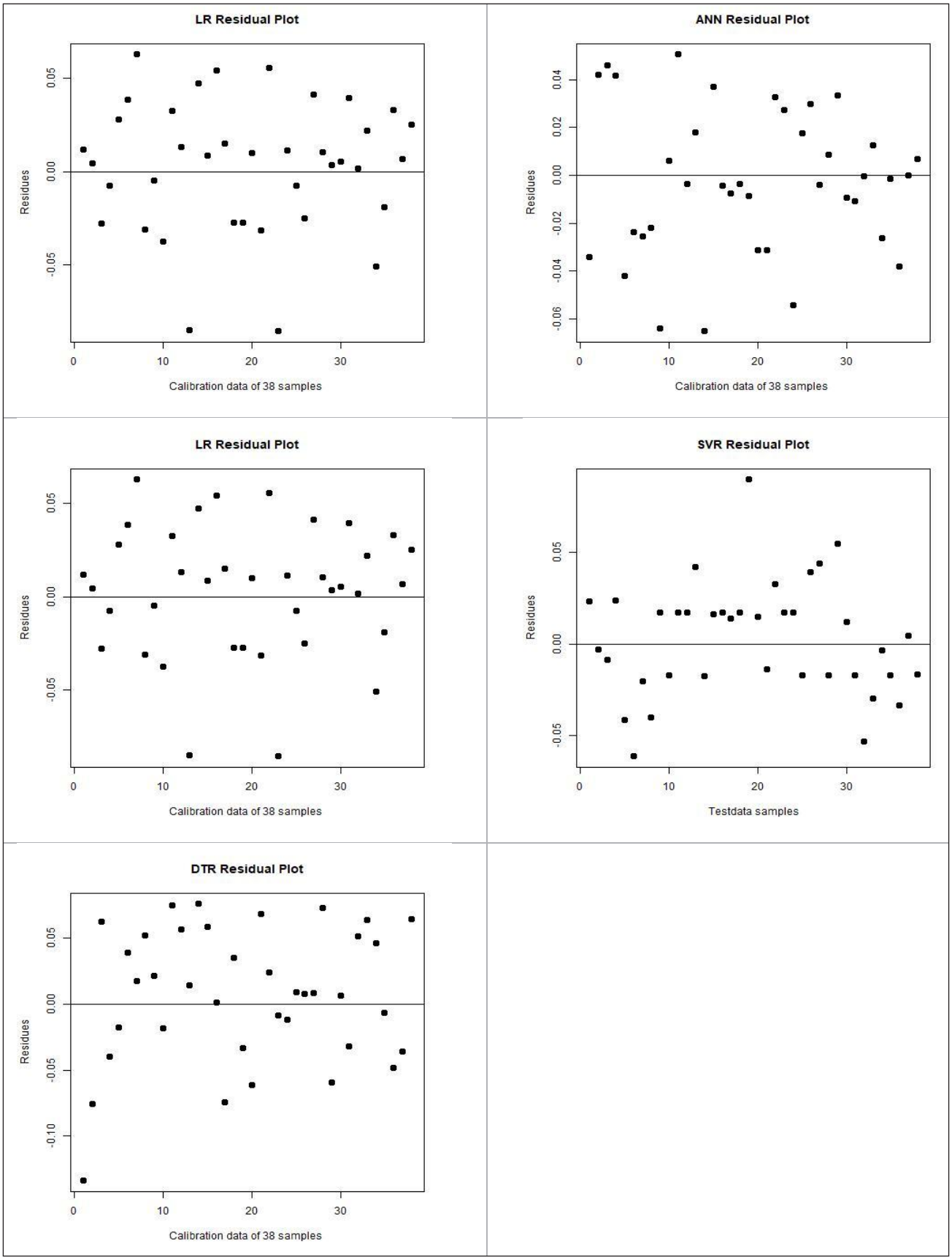

Fig. 3: Residual plots of five machine learning algorithms for Starch component using calibrated data

they've been applied to real-world procedures. Fig. 2 displays scatterplots generated for each model. Fig. 3 displays residual plot.

CONCLUSION

The key findings are as follows: using the variable selection technique of PLS regression, it was discovered that the wavelength range 1700-1780 nm is the best wavelength range for predicting chickpea starch components out of the range of 750-2500 nm. Five machine learning predictive models were created and compared to determine their effectiveness. When these models are compared, the LR and ANN models have better predictive statistics in terms of lower RMSE and RSE, as well as higher R^2 and adjusted R^2. This initiative has the potential to be scaled up to serve as a model for predicting other leguminous crop components. It is a useful tool for predicting the presence of various components in a given sample. It can be used to fingerprint agricultural crops and detect the amounts of several other components in a sample. Due to the non-destructive nature of near-infrared spectroscopy, no or very little sample preparation is required.

ACKNOWLEDGMENTS

This research was conducted as part of the Ph.D. programme at the Indian Council of Agricultural Research-Indian Agricultural Research Institute (ICAR-IARI).

REFERENCES

Acquah, G., Via, B., Billor, N., Fasina, O. and Eckhardt, L. 2016. Identifying plant part composition of forest logging residue using infrared spectral data and linear discriminant analysis, *Sensors*, **16**: 1375.

Baljeet, S.Y., Ritika, B.Y. and Reena, K. 2014. Effect of incorporation of carrot pomace powder and germinated chickpea flour on the quality characteristics of biscuits. *Int. Food Res. J.,* **21**(1): 217-222.

Bar-El Dadon, B.S., Pascual, C.Y. and Reifen, R. 2014. Food allergy and cross-reactivity-chickpea as a test case. *Food Chem.,* **165**: 483-488.

Batten, G.D. 1998. Plant analysis using near-infrared reflectance spectroscopy: the potential and the limitations, *Aust. J. Expl. Agric.,* **38**: 697-706.

Boye, J.I., Zare, F. and Pletch, A. 2010. Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Int. Food Res. J.,* **43**(2): 414-431.

Brenna, O.V. and Berardo, N. 2004. Application of near-infrared reflectance spectroscopy to the evaluation of carotenoids content in Maize. *J. Agric. Food Chem.*, **52**: 5577-5582.

Choung, M.G., Kang, S.T., Han, W.Y., Baek, I.Y., Kim, H.K., Kim, K.S. and Park, S.H. 2005. Determination of fatty acid composition in soybean seed using near-infrared reflectance spectroscopy. *Korean J. Breed. Sci.,* **37***:* 197-202.

Fassio, A. and Cozzolino, D. 2004. Non-destructive prediction of chemical composition in sunflower seeds by near-infrared spectroscopy, *Ind. Crops Prod.,* **20**: 321-329.

Ghasemi, Jahan, Tavakoli, Hossein. 2013. Application of Random Forest Regression to Spectral Multivariate Calibration, *Anal. Method.* https://DOI: 10.1039/C3AY26338J.

Kim, K.S., Si and Choung, Myoung-Gun and Jang, Young. 2007. Use of Near-Infrared Spectroscopy for Estimating Fatty Acid Composition in Intact Seeds of Rapeseed. *J. Crop Sci. Biotechnol.,* **10**: 15-20.

Lu, G., Huang, H. and Zhang, D. 2006. Prediction of sweet potato starch physiochemical quality and pasting properties using near-infrared reflectance spectroscopy. *Food Chem.,* **94**: 632-639.

Martelovidal, M.J. and Vazquez, M. 2014. Evaluation of ultraviolet, visible, and near-infrared spectroscopy for the analysis of wine compounds, *Czech J. Food Sci.,* **32**: 37–47.

Oh, K.W., Choung, M.G., Pae, S.B., Jung, C.S., Kim, B.J., Kwon, Y.C., Kim, J.T. and Kwack, Y.H. 2000. Determination of seed lipid and protein contents in perilla and peanut by near-infrared reflectance spectroscopy. *Korean J. Crop Sci.,* **45**: 339-342.

Osborne, B.G., Fearn, T.A. and Hindle, P.H. 1993. Theory of Near infrared spectrophotometry, In Practical NIR Spectroscopy: With Applications in Food and Beverage Analysis, 2nd ed.; Addison-Wesley Longman Limited: United Kingdom, pp. 13–35.

Sato, T., Maw, A.A. and Katsuta, M. 2003. NIR reflectance spectroscopic analysis of the FA composition in sesame (*Sesamum indicum* L.) seeds. *J. Am. Oil Chem.' Soc.,* **80**: 1157-1162.

Wu, J.G. and Shi, C.H. 2004. Prediction of grain weight, brown rice weight and amylose content in single rice grains using near-infrared reflectance spectroscopy. *Field Crop Res.,* **87**: 3-21.