Annals of Agri-Bio Research **28** (1): 129-135, 2023

Impact Score : 0.28 (Scopus)

Geographical Classification with Principal Component Analysis Related to Chemical Attributes and Non-destructive Assessment of Sweetness of Indian Mangoes Using Chemmometrics

AGNIBHA DAS MAJUMDAR, UMA KAMBOJ AND NEHA MUNJAL*

Department of Physics, Lovely Faculty of Technology and Sciences, School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara -144 411 (Punjab), India
**(e-mail: nehamunjal.research@gmail.com; Mobile: 90410 61879)*

(Received: June 25, 2022; Accepted: July 22, 2022)

ABSTRACT

Aim of the present study was to measure the different nutritional parameters with standard methods of Association of Official Agricultural Chemists (AOAC). Different parameters of Indian mango pulp like carbohydrate, crude protein, crude fibre, crude fat, moisture, ash content, pH, total soluble solid (TSS) and energy value were determined along with correlations. The best correlation between the energy, moisture and carbohydrate content was established and expressed with the regression equation. Singular value decomposition under the principal component analysis (PCA) was applied on all chemical data to know the classification of the data. The present study also established a green and non-destructive method to predict the sweetness of mango using near infrared spectroscopy together with chemmometric approach. The spectra was recorded for the 51 different homogenized Indian mango pulp within the range of 700-2500 nm and then wavelength selection method was applied to reduce the wavelengths for prediction. Multiplicative Scatter Correction (MSC) and second order derivative were used to pre-process the original spectra and Partial Least Square regression method to develop the prediction model for sweetness analysis of mango. Coefficient of determination for calibration (R_c^2) as well as for validation (R_v^2) gave the results more than 0.9 for both pH and total soluble solid. The value of coefficient of determination for calibration and the value of root mean square error for calibration (RMSEC) for pH was presented as 0.9945; validation and validation (RMSEV) was presented as 0.9945 and 0. The results indicated that NIR spectroscopy along with chemmometrics gave the promising results and was an efficient method to calculate the sweetness of mango fruit with green and non-destructive way.

Key words: Mango, principal component analysis, non-destructive, near-infrared spectroscopy, sweetness

INTRODUCTION

Fruits, as a vast source of essential nutrients, vitamins and minerals, have an active part in human daily diet. Moreover, dietary fiber present in fruits plays a meaningful role in the case of cardiovascular disease and obesity (Kusumiyati *et al.*, 2021). Mango is a good source of phytochemicals which contain antioxidants like tripenoids, flavonoids, etc. (Bhole and Kumar, 2020). Mango fruit offers a substantial amount of vitamin A, E, C and B_6. The major antioxidant present in mango is commonly known as mangiferin which can have anti-diabetic and anti-inflammatory effects (Akter *et al.*, 2022). Mango fruit can balance the blood sugar level and can handle the glycemic level at pre-diabetic and diabetic stage (Swaroop *et al.*, 2018). In India, mango production covers an area of nearly 1.23 MHa against an annual quantum of 10.99 MT. It happens to grow up within the tropical and subtropical region up to an altitude of almost 1500 m from sea level (FAO, 2019). Totapuri, Baganapalli, Kesar, Alphanso, Chausa and Langra are most common among the Indian varieties (Thakor, 2019).

Mango has a worldwide economical acceptance due to its flavour, texture and chemical constituents and with the growing food industries the demand of tropical fruits has been increasing day by day (Girma *et al.*, 2016). Many food products like ice cream, jam, jelly, fruit cocktail, nectar, squashes, mango shakes, mango toffee have been made from pulp and for this purpose, the companies concerns have to preserve mango pulp obtained from fresh mangoes (Owino and Ambuko, 2021).

Principal component analysis is also known as a dimension reduction method and it can reduce the large number of variables with small number of variables maintaining the same information (Jolliffe, 2021). The principal component analysis is a mathematical tool which can apply to transform a number of correlated variables to uncorrelated variables. Those variables are known as principal components. This method is similar to the multivariate factor analysis and generally it applies on the square symmetric matrix (Beattie and Esmonde-White, 2021). A linear combination of observed variables is expressed as principal components. The main purpose of this analysis was to collect the correct information and ignore the outliers (Granato *et al.*, 2018). Infrared spectroscopy is a non-destructive approach to qualitative and quantitative analysis of any material. In the field of food science and food technology, this method has achieved an important role to solve the several research problems (Yakubu *et al.*, 2022). Near Infrared spectroscopy is classified as a non-invasive, non-destructive, robust and fast method. Grading and sorting is also an important key for exportation and can determine easily rapidly determine with few scans. This method has proven as efficient tool for the qualitative and quantitative analysis of many fruits (Pasquini, 2018). Infrared spectroscopy is a non-destructive approach for qualitative and quantitative analysis of any material. In the field of food science and food technology, this method has achieved an important role to solve the several research problems (Yakubu *et al.*, 2022). Near Infrared spectroscopy is classified as a non-invasive, non-destructive, robust and fast method. Grading and sorting is also an important key for exportation and can determine easily and rapidly with few scans. This method has proven as efficient tool for the qualitative and quantitative analysis of many fruits (Pasquini, 2018). Present research focused on the fruit pulp quality of different indigenous mango varieties. This work can help the agro-food industries to estimate the nutritional parameters and safety status of mango pulp.

MATERIALS AND METHODS

Fully ripe mango of different indigenous varieties was collected from the different mango farms of five states of India, namely, Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal. Collected fruits were brought to the Advanced Analytical Testing Laboratory, Kolkata, West Bengal for chemical analysis. The fruits were then washed properly by distilled water to remove the heat, external dirt, dust, pollutants and the pesticide residues.

Fifty-two different mango samples were sorted and marked. Each mango sample was peeled and the pulp was extracted. The kernel was separated and Lifelong Power pro LLMG02 mixer grinder was used to prepare homogeneous pulp. Pulp samples were then packed into different labelled zip lock pouch bags and finally stored at 4 ^{0}C in refrigerator until future analysis.

The biochemical analysis of the mango homogenized pulp was taken with the AOAC standard method. Crude fiber, crude protein and crude fat were determined with Acid-alkali digestion method, Kjeldahl method and Soxhlet method, respectively. pH content and TSS values were determined with the pH meter and hand refractometer. Results were recorded in the triplicate and the mean value was taken. Energy values were calculated as (Osborne and Voogt).

$$\text{Energy} = [9 \text{ x fat } (\%) + 4 \text{ x carbohydrate } (\%) + 4 \text{ x protein } (\%)]$$

Present study followed Karl Pearson's coefficient of correlation method which can be defined as the mathematical method to measure the intensity or the magnitude of linear relationship between two variables. The correlation among all the biochemical parameters was examined and few of them were found significantly correlated. Correlation analysis and linear regression equation were determined with the IBM SPSS Statistics version 22.0 software. Principal component analysis was carried out with the Clustvis 2.0 web tool (Clustvis 2.0.: https://biit.cs.ut.ee/clustvis/, accessed by 20 June, 2022) for visualizing clustering of multivariate data.

The near infrared spectral data were collected with the NIR DS2500 (Foss) spectrometer between the spectral range of 700-2500 nm. The regression model was developed by importing the NIR spectral data to the data analysis software Unscrambler X (CAMO AS, Trondheim, Norway, version 10.5.1). Multivariate analysis was performed on the chemical data of pH and TSS parameter to check the sweetness of the mango samples. Raw data were plotted and visually non-similar data were identified as the outliers and removed from the data set. pH and TSS parameters of total 52 mango samples were taken for the calibration and validation analysis.

Partial Least Square (PLS) regression method was used to calibrate and validate the data set. Some pre-processing methods, Multiplicative Scatter Correction (MSC) and second order derivative (Using Savitzky Golay method with second order polynomial) were applied on the original raw spectral data to improve the model predictability. In order to search the minimum range of wavelength for good prediction model for sweetness of mango the total range of wavelength was divided into 18 groups. Each group contained a range of 100 nm wavelengths and PLS was applied on each group. The best prediction model was introduced with the coefficient of determination for calibration, coefficient of determination for validation, root mean square error for calibration and root mean square error for validation values. The plot between reference and predicted values of pH and TSS was presented to know the actual predictability using NIR spectroscopy in a non-destructive way.

RESULTS AND DISCUSSION

Descriptive statistics: Mean, standard deviation, maximum, minimum value and the coefficient of variance of all the physical and chemical parameters like moisture, ash, crude fat, crude fiber, crude protein, pH, carbohydrate, total soluble solids (TSS) and chemical energy have been depicted in Table 1. Water content is responsible for the shelf life for the mango fruit and depends upon the different harvesting techniques and growing atmosphere of mangoes. This variation in the moisture content in mangoes may be caused due to different storage conditions. It was reported that fruits were mostly having 70 to 90% moisture content. In this study, every sample was having more than 75% water content, which was significant. Ash content is defined as the total mineral content present in any food sample. These maximum and minimum values of ash content of the particular varieties might be related to the higher and lower concentration of minerals, respectively. The sample from Uttar Pradesh had the highest and sample from West Bengal had the lowest presence of crude protein quantity. It was recorded that the presence of carbohydrate in a mango fruit varied from 10 to 15% in most of the cases. Although the present study reported the lowest percentage of carbohydrate content as 2.81%, which was very less as compared to standard value. This value was observed for the sample collected from West Bengal. The highest value for the same was 16.25% in sample of Maharashtra. The highest energy of any human body can get from the banana fruit, after that mango is in queue. Highest energy value observed in mango fruit was 91.925 Kcal/100 g and lowest value was recorded as 37.915 Kcal/100 g. Total soluble solid or TSS is defined as the total soluble and insoluble sugar present in the mango and is counted as a major quality parameter of mango fruit. This parameter might be changed with storage period of mango with different temperature. Sweetness of mango can be examined with this property. The highest and lowest value recorded for the mango fruit was 17 and 10, respectively. The highest value of pH for mango was 4.79 and the lowest value for the same was 2.74. It was

Table 1. Descriptive statistics of chemical parameters of 52 different samples of Indian mango using standard analytical methods

Chemical parameters	Mean	Standard deviation	Maximum	Minimum	Coefficient of variance
Moisture (%)	82.597	3.777	89.115	76.815	0.045
Ash (%)	0.627	1.082	7.850	0.085	1.726
Crude fat (%)	0.371	0.145	0.605	0.000	0.390
Crude protein (%)	5.562	0.817	7.250	3.750	0.146
Crude fiber (%)	0.592	0.700	0.715	0.470	0.118
pH	3.975	0.443	4.790	2.740	0.111
Carbohydrate (%)	10.248	3.679	16.250	2.810	0.359
TSS (0Brix)	13.278	2.236	17.000	10.000	0.168
Energy (Kcal/100 g)	66.590	15.7865	91.925	37.915	0.237

observed that different varieties of mango had a range from 2.74 to 4.79 of pH value, it can be concluded that mango is an acidic fruit. The trend from this study showed that most of the variety contained the pH more than 4. The variation of the pH content for different varieties has been reasoned for different ripening period of concerned geographical area of current study. The storage temperature and ripening process, soil contamination and the quality of soil differed with the places were responsible for the enormous metabolic activities within the mango fruits, which could lead the chemical changes from variety to variety of different region.

Present research work expressed a range of different values of concerned nutritional parameters. This variation of the nutritional properties generally occurred due to some reasons. The reasons might be soil contamination; the quality of soil differed with the places. The storage temperature and ripening process were responsible for the enormous metabolic activities within the mango fruits, which could lead the chemical changes from variety to variety of different region.

For correlation analysis among all the variables total 52 samples were taken into consideration. The values of all nutritional properties as moisture, ash, crude fat, crude protein, crude fiber, carbohydrate, TSS and pH were used for correlation analysis. All the data were also used for principal component analysis.

For the correlation analysis, Karl Pearson's coefficient or commonly known as covariance method was used. Significant correlations were found between moisture and protein ($P<0.01$), moisture and carbohydrate ($P<0.01$), moisture and energy ($P<0.01$), ash and carbohydrate ($P<0.05$), fat and protein ($P<0.01$), fat and fiber ($P<0.01$), fat and TSS ($P<0.01$), protein and energy ($P<0.01$), fiber and pH ($P<0.01$), fiber and TSS ($P<0.01$) and carbohydrate and energy ($P<0.01$). There was strong correlation between moisture and carbohydrate, moisture and energy and carbohydrate and energy with 'r' value greater than 0.9 (Table 2). The determination of correlation coefficient (R) concluded that there was a strong correlation with moisture between carbohydrate and energy, respectively.

The regression analysis was applied to know the equation between parameters, which predicted their relationship mathematically. The regression equation has been shown below:

$$C = - 0.9\, M + 84.9 \qquad ...(1)$$

$$E= -4.01\, M + 3.98 \times 10\text{\^{}}2 \qquad ...(2)$$

$$E = 4.15\, C + 24.05 \qquad ...(3)$$

Energy was negatively and positively correlated with moisture and carbohydrate with R^2 value of 0.919 and 0.936, respectively, and carbohydrate was negatively correlated with moisture with R^2 value of 0.861. The coefficient of determination R^2 measured the goodness-of-fit of the estimated Sample Regression Plane (SRP) in terms of the proportion of the variation in the dependent variables explained by the fitted sample regression equation. Thus, the values of R^2 as 0.919, 0.936 and 0.861 simply meant that about 91.9, 93.6 and 86.1 of the variation in moisture was explained by the estimated SRP that used energy, the variation in carbohydrate was explained by the estimated SRP that used energy and the variation in moisture was explained by the estimated SRP that used carbohydrate, respectively.

Table 2. Correlation analysis between all the chemical attributes of mango

	Moisture	Ash	Fat	Protein	Fiber	pH	Carbohydrate	TSS	Energy
Moisture	1.000								
Ash	-0.017	1.000							
Fat	-0.125	-0.149	1.000						
Protein	-0.403**	0.029	0.454**	1.000					
Fiber	0.027	0.002	0.382**	0.270	1.000				
pH	-0.153	-0.053	0.092	0.123	0.436**	1.000			
Carbohydrate	-0.928**	-0.278*	0.024	0.160	-0.122	0.133	1.000		
TSS	-0.047	0.163	0.402**	-0.048	0.360**	-0.070	-0.012	1.000	
Energy	-0.959**	-0.265	0.199	0.394**	-0.026	0.157	0.967**	0.013	1.000

*,**Significant at P=0.05 and P=0.01, respectively.

To classify the samples on the basis of nutritional properties unit variance scaling was applied to rows : SVD with imputation was used to calculate principal components. X and Y axis showed principal component 1 and principal component 2 explaining 35.2 and 22% of the total variance for all 51 data points (all five states: Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal; Fig. 1), 36.5 and 21.5% for 37 data points (three states, Bihar, Uttar Pradesh and West Bengal; Fig. 2) and 36.5 and 21.1% for 32 data points (only two states, Uttar Pradesh and West Bengal; Fig. 3), respectively.

The original spectra of the raw NIR data of total 52 different mango samples was taken from the wavelength range of 700 to 2500 nm (Fig. 4). The spectra showed picks at different wavelengths indicating the presence of different organic molecule in the mango pulp sample.

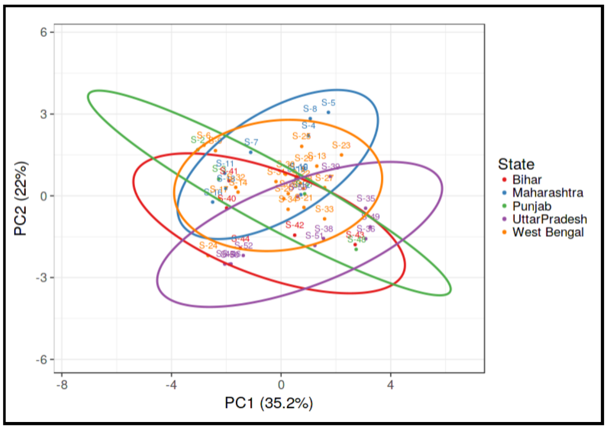

Fig. 1. Classification of mangoes with principal component analysis on basis of five states.

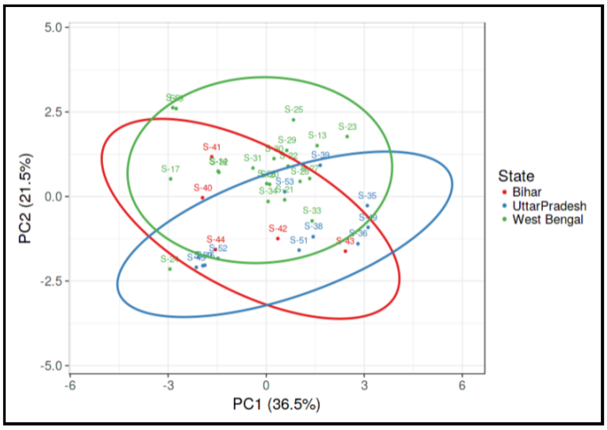

Fig. 2. Classification of mangoes with principal component analysis on basis of three states.

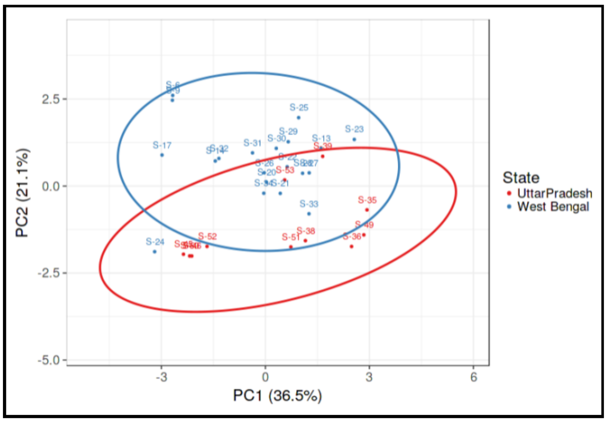

Fig. 3. Classification of mangoes with principal component analysis on basis of two states.

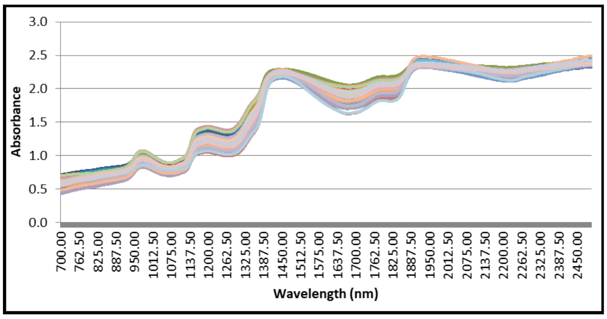

Fig. 4. NIR spectra of different varieties of Indian mangoes with wavelength ranging from 700 to 2500 nm.

To predict the sweetness of the mango, prediction of the two parameters (pH and TSS) was done with non-destructive method. The values of the coefficient of determination for calibration (R_c^2), coefficient of determination for validation (R_v^2), root mean square error for validation (RMSEV) and root mean square error for validation (RMSEV) judged the quality of the prediction model. A good quality regression model had the lower values of RMSEC and RMSEV and higher values of R_c^2 and R_v^2. PLS regression was applied to the raw data for the prediction. Particularly two different pre-processing methods were suitable and applied to the raw spectral data to get the more accurate result. MSC modified the original spectra by bringing all different spectra of different samples more nearly. Derivative computed the overlapping crust and trough of each spectrum and in the present case second order derivative proved as efficient to modify the spectra. After applied MSC and second derivative the results in terms of coefficients and errors were modified for prediction of pH (Table 3) and TSS (Table 4).

Table 3. Effect of pre-processing method on the spectral data of total range of 700-2500 nm for pH

Pre-processing techniques	R_c^2	RMSEC	R_P^2	RMSEV
Original spectra	0.4779	0.0663	0.2955	0.0771
MSC	0.6039	0.0578	0.3935	0.0715
2nd order derivative	0.9905	0.0089	0.9856	0.0010
MSC+2nd order derivative	0.9913	0.0085	0.9868	0.0105

Table 4. Effect of pre-processing method on the spectral data of total range of 700-2500 nm for TSS

Pre-processing techniques	R_c^2	RMSEC	R_P^2	RMSEV
Original spectra	0.3204	0.1085	0.0557	0.1279
MSC	0.3943	0.1024	0.1852	0.1188
2nd order derivative	0.9569	0.0273	0.9411	0.0319
MSC+2nd order derivative	0.9576	0.0270	0.9426	0.0315

The validation results of both pH and TSS were almost similar to the calibration values (Table 5). It clearly confirmed the predictability of the

Table 5. PLS regression models for Indian mango varieties for the wavelength ranges for prediction of pH and TSS

Parameter	Wavelength range	Calibration		Validation	
		R_c^2	RMSEC	R_v^2	RMSEV
pH	700-2500	0.9913	0.0085	0.9868	0.0105
	1100-2300	0.9978	0.0042	0.9970	0.0050
TSS	700-2500	0.9576	0.0270	0.9426	0.0315
	1100-2300	0.9892	0.0136	0.9841	0.0165

R_c^2: Coefficient of determination for calibration and R_v^2: Coefficient of determination for validation, RMSEC : Root mean square error for calibration, RMSEV: Root mean square error for validation and PLS: Partial least square.

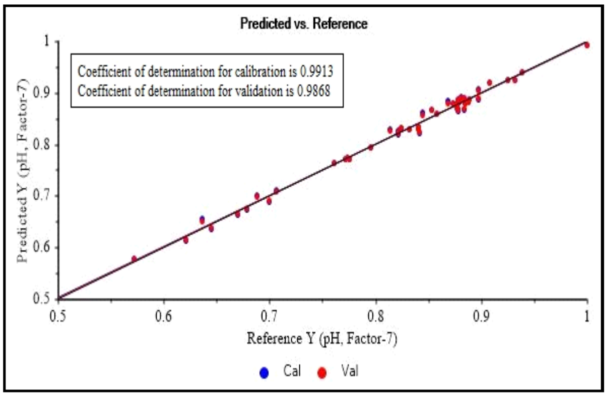

Fig. 5. Scatter plot of mango verities for pH content after multiplicative scatter correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation.

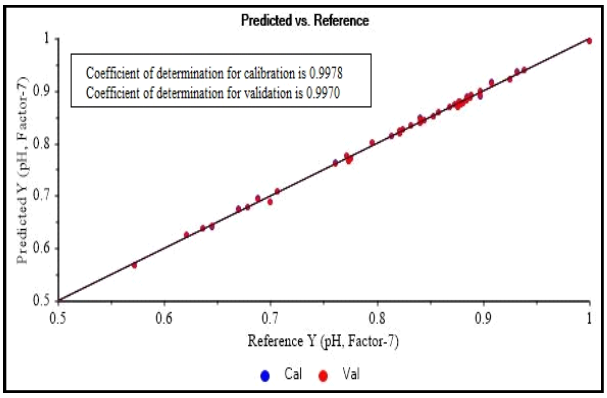

Fig. 6. Scatter plot of mango varieties for pH content after multiplicative scatter correction and second order derivative in wavelength range of 1100-2300 nm for calibration and validation.

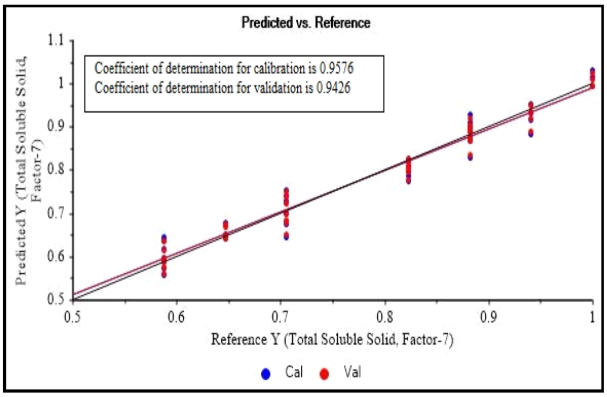

Fig. 7. Scatter plot of mango varieties for total soluble solid content after multiplicative scatter correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation.

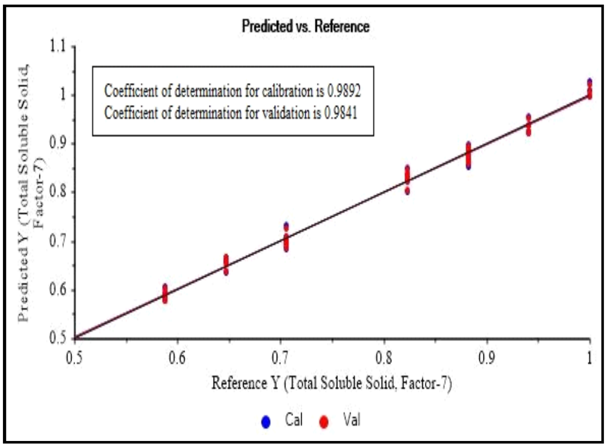

Fig. 8. Scatter plot of mango varieties for total soluble solid content after multiplicative scatter correction and second order derivative in wavelength range of 1100-2300 nm for calibration and validation.

regression model. The similarities in values for calibration and validation removed the probability of over fitting of the model. The comparison of the reference and the predicted values is shown in Figs. 5 to 8. The wavelength range 1100-2300 nm was proved efficient for the good prediction model for the sweetness of mango (Figs. 6 and 8).

CONCLUSION

The research work focused on determination of parameters of Indian mango pulp, namely, carbohydrate, crude protein, crude fiber, crude fat, moisture, ash content, pH, total soluble solid (TSS) and energy value. Secondly, correlation between all these parameters was observed and it was noticed that energy, moisture and carbohydrate content showed the best correlation. Principal component analysis (PCA) was employed using Singular value decomposition algorithm to classify the mango samples on basis of origin. Next work done was to predict the sweetness of mango using near infrared spectroscopy (700-2500 nm) together with chemmometric approach–Partial Least Square regression. It was observed that model showed good accuracy with coefficient of determination for calibration (R_c^2) and validation (R_v^2); results more than 0.9 for both pH and total soluble solid. The results revealed that NIR spectroscopy along with chemmometrics was efficient method to predict the sweetness of mango fruit with green and non-hazardous way.

ACKNOWLEDGEMENTS

The authors are thankful to Dr. Sutapa Bhowmick, Laboratory Head, Advanced Analytical Testing Laboratory, Kolkata, West Bengal and Dr. Sunita Mishra, Principal Scientist, CSIR-CSIO, Chandigarh for providing their facilities to carry out the research work successfully.

REFERENCES

Akter, S., Moni, A., Faisal, G. M., Uddin, M. R., Jahan, N., Hannan, M. A., Asadur, R. and Uddin, M. J. (2022). Renoprotective effects of mangiferin: Pharmacological advances and future perspectives. *Int. J. Environ. Res. Public Health* **19**: 1864.

Beattie, J. R. and Esmonde-White, F. W. (2021). Exploration of principal component analysis: Deriving principal component analysis visually using spectra. *Appl. Spectrosc.* **75**: 361-375.

Bhole, V. and Kumar, A. (2020). Mango quality grading using deep learning technique: Perspectives from agriculture and food industry. In : *Proc. 21st Annual Conference on Information Technology Education, Omaha, Nebraska, USA.* pp. 180-186.

FAO. (2019). Tropical fruits compendium. *http:// www.fao.org/ docrep /meeting/ 022/ am 481t.pdf.*

Girma, G., Garo, G. and Fetena, S. (2016). Influence of post-harvest treatment on physical characteristics and mineral content of mango (*Mangifera indica* L.) fruit in Arba Minch, Southern Ethiopia. *Int. J. Nutr. Food Sci.* **5**: 395.

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L. and Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends Food Sci. Technol.* **72**: 83-90.

Jolliffe, I. (2021). A 50-year personal journey through time with principal component analysis. *J. Multivar. Anal.* **188**: 104820. *https://doi.org/10.1016/j.jmva.2021.104820.*

Kusumiyati, K., Munawar, A. A. and Suhandy, D. (2021). Fast, simultaneous and contactless assessment of intact mango fruit by means of near infrared spectroscopy. *AIMS Agric. Food* **6**: 172-184.

Owino, W. O. and Ambuko, J. L. (2021). Mango fruit processing: Options for small-scale processors in developing countries. *Agriculture* **11** : 1105.

Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives–A review. *Anal. Chim. Acta.* **1026**: 8-36.

Swaroop, A., Bagchi, M., Moriyama, H. and Bagchi, D. (2018). Health benefits of mango (*Mangifera indica* L.) and mangiferin. *Jpn. J. Med.* **1**: 149-154.

Thakor, N. J. (2019). Indian mango–Production and export scenario. *J. Adv. Agric. Technol.* **3**: 80-88.

Yakubu, H. G., Kovacs, Z., Toth, T. and Bazar, G. (2022). The recent advances of near-infrared spectroscopy in dairy production–A review. *Crit. Rev. Food Sci. Nutr.* **62**: 810-831.