Application of FT-IR spectroscopy with various classification and regression models for detection and quantification of Sodium Hydrosulfite in Iranian wheat flour

A. Kazemi[a*], A. Mahmoudi[a]

[a] University of Tabriz, Department of Biosystems Engineering, 29 Bahman Blvd., Tabriz, Iran
[*]Corresponding author: Amir Kazemi
amirkazemi422@gmail.com

Abstract:
Wheat flour is one of the most important and strategic food resources especially in developing countries. The addition of Sodium hydrosulfite to flour for improving some appearance features can have dangerous impacts on the consumer health. Therefore, detection of this harmful substance is significant. In the present study, the potential of Fourier transform-mid infrared (FT-MIR) spectroscopy in 400-4000 cm^{-1} for the fast detection of Sodium hydrosulfite powder in wheat flour was investigated. After getting the spectral data from samples, firstly some preprocessing methods were used to correct harmful and unwanted effects on spectral data, and then Principal Component Analysis (PCA) as unsupervised and Support Vector Machine (SVM) and Artificial Neural Network (ANN) models as supervised classification models and Partial Least Square Regression (PLSR) as regression model were applied to detect and quantify the adulteration in pure flour samples. The best outcomes were the accuracy of 86.66 and 86.70 for SVM and ANN models with S-G + D2 + SNV preprocessing, respectively and $R^2_p = 0.99$ For PLSR model.

Keywords: spectroscopy, adulteration, wheat flour, chemometrics, sodium hydrosulfite

1. Introduction:

Bread as one of the most significant sources of daily requirement components for body (such as proteins, minerals and vitamins) is one of the stable foods for many countries, particularly Iran (Ahamadabadi et al., 2016; GhR, Yunesian, Vaezi, Nabizadeh, & GhA, 2006; Sabeghi, 2004). The usage of bread in Iran is five times more than people in Europe (Malakootian & Dowlatshahi, 2005; Sabeghi, 2004). Among the main ingredients of bread, wheat flour has a special place and has direct relation to the quality of bread and also health of consumer. Therefore, it should get the certificate of Iranian national standard. Wheat contains 78.10% of carbohydrate, 14.70% protein, 2.10% fat, 2.10% minerals and noticeable proportion of vitamins (Adams, Lombi, Zhao, & McGrath, 2002; P. Shewry, 2009; P. R. Shewry et al., 2006; Topping, 2007). According to the statistics of world health organization and food and agriculture organization of united nations, 25 types of food additives are used in each country according to the food safety policy (F. C. Martins, Sentanin, & De Souza, 2019). Then, the maximum acceptable amount of them and also the Esurance of avoidance of any unauthorized additives should be considered. Sodium hydrosulfite also known as Blankit is a white crystalline powder containing inorganic sulfur compounds (Reza et al., 2014). In food industry, this material is applied for nuts, sugar, etc. to avoid browning and bleaching and regeneration of Cellulose fibers (de Carvalho & Schwedt, 2005). Sodium hydrosulfite has been utilized in Iranian bread industry to hide visible defects of bread by affecting the velocity of production process and compensating some visible results of lack of natural fermentation and poor flour quality (Asgari, SeidMohammadi, Faradmal, Moradi, & Yari, 2018). This material has so dangerous effects on human health. Adverse effects of blankit include the elimination and damage of villi in the stomach and intestines in the long term, therefore, it can cause the development of gastrointestinal cancer. It is also known to be an effective factor in causing diabetes (Karami, Alikord, Mokhtari, Sadighara, & Jahed-Khaniki, 2021). Therefore, detection of this harmful material in the human's used diet is essential. In general, different approaches have been applied to quantify Sulfur factors in food, such as titration (Monnier & Wiliams, 1972), liquid and gas chromatography (Rethmeier, Rabenstein, Langer, & Fischer, 1997), high performance ion chromatography (Lavigne-Delcroix, Tusseau, & Proix, 1996), electroanalysis methods include the study of the electrical activity of sulfites, voltammetry (Govaert, Temmerman, & Kiekens, 1999), and amperometry, potentiometric and the method of general evaluation of sulfites in the automated system (Pisoschi et al., 2020). The mentioned techniques encompass some drawbacks such as being high-cost, laborious, and destructive. Then, some other nondestructive, inexpensive and fast methods are required. Fourier Transform infrared (FT-IR) spectroscopy is one of the fingerprint techniques which is widely used to identify components of food and determine possible impurities. FT-IR spectroscopy can be in the middle range (450-4000 cm^{-1}, FT-MIR) or near range (4000-10000 cm^{-1}, FT-NIR) (Pallone, dos Santos Caramês, & Alamar, 2018). FT-MIR comes up with more structural and chemical information than Fourier Transform-Near Infrared (FT-NIR) by the ability of displaying vibrational and rotary stretching process of covalent bonds (Lohumi, Lee, Lee, & Cho, 2015). Some researchers have explored the applicability of spectroscopic techniques to investigate chemical information of materials. Mohamed et al, explored classification of five food powder types (wheat flour, organic wheat flour, rice flour, corn starch, and tapioca starch) and Support Vector Machine (SVM) model had acceptable outcomes for classification of mentioned powders (Mohamed, Solihin, Astuti, Ang,

& Zailah, 2019). In another study, Girolma et al applied FT-IR techniques in different ranges (FT-MIR and FT-NIR) to detect the adulteration of durum wheat pasta with common wheat. Linear Discriminant Analysis (LDA) and Partial Least Square – Discriminant Analysis (PLS-DA) had the results of 80 and 95% for three class dataset and 91 and 97% for two class dataset (De Girolamo et al., 2020). But as far as I know, the applicability of FT-MIR spectroscopy method with combination of ANN for classification and PLSR model for quantification of adulteration of this harmful material in Iranian wheat flour has not been investigated. In the present study, the applicability of FT-MIR spectroscopy combined with chemometric methods and various preprocessing algorithms for detection and quantification of sodium hydrosulfite in wheat flour in Iran was studied.

2. material and methods

In the present research, after preparing samples, spectral data were acquired and preprocessed. The both supervised and unsupervised models were applied. Afterward, the results were analyzed for detection and quantification of pure and adulterated samples. Figure.1. represents the flowchart of flour adulteration detection procedure by FT-MIR spectroscopy.

2.1. sample preparation

Sardari wheat seeds (harvested in 2021) were purchased from a seed modifying center in Bonab, East Azerbaijan, Iran. The gotten seeds were harvested in four distinct places in Iran. Then, geographical variation of samples was considered. Sardari wheat was selected because it is the highest under-harvest wheat variety in Iran. Sodium hydrosulfite (with the purity of 90%) was acquired from a supermarket in Bonab, Iran. First, wheat seeds were milled by a laboratory benchtop mill to get the wheat flours. Then for having homogeneous flour samples, they were passed through a sieve (mesh 420μm). The considered adulterant concentration (w/w) were 10, 15, 20, and 25%. Totally, 150 samples were prepared (25 for pure flour, 25 for sodium hydrosulfite, and 25 samples for 4 adulterant groups). After mixing the adulterant to pure flour with the mentioned levels, they were blended intensely to get the homogenous samples as much as possible. Finally, the prepared samples were poured to microtubes to transfer to the laboratory.

2.2. spectra acquisition

Spectral data were acquired at the central laboratory of Tabriz university with a FT-MIR spectrometer (TENSOR 27, Bruker, Germany) in transmittance mode and with resolution of 1 cm^{-1}. The scanning speed was 20khz and with 64 scans. Each powdered samples were placed on the ATR (single bounce) crystal and pressed until the desired signal density acquired. The crystal was washed with 100% ethanol after testing each sample. For each individual sample, 3 transmittance spectra were acquired and mean spectrum of three replicates was used for further analysis. Finally, the mean spectra were transferred to Excell 2019 version to be prepared for statistical analysis. Multivariate statistical analysis was conducted with Unscrambler v 10.4 (Camo software As, Oslo, Norway, 2011) for PCA and SVM and classification Toolbox in Matlab (Mathworks, Inc., Natick, Massachusetts, USA) for Artificial Neural Networks (ANN).

2.3. preprocessing

Before classification or regression modeling, pretreatment of spectral data is an essential step to remove the unwanted and uninformative data. This can be due to large amount of water in samples, different conditions of samples, and noise in spectra that comes from electronic components in the system (Boysworth & Booksh, 2008; Christy & Kvalheim, 2007; Varmuza & Filzmoser, 2016). The most common applied preprocessing techniques in spectroscopy is divided in to two categories: spectral normalization and spectral derivatives (Rinnan et al., 2009). Spectral normalization techniques which contain Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), and de-trending (DT) can be used for correction of scattering effects. While spectral derivatives including (first and second derivatives and smoothing techniques) are applied for correction of peak overlap and baseline drifts (López-Maestresalas et al., 2019). Both SNV and MSC are the most commonly used algorithms to correct the scatter effects. The difference of SNV and MSC methods is based on the fact that the scatter correction in SNV method is based on the average value of every individual spectra, but in MSC technique a reference spectra (average spectra) is required to contrast the whole spectra with that (Dhanoa, Lister, Sanderson, & Barnes, 1994; Zeaiter, Roger, & Bellon-Maurel, 2005). Among the spectral derivative methods, Savitzky-Golay (S-G) is the most common algorithm for derivation (Savitzky & Golay, 1964). By this method, the data with a window size chosen are fitted by a polynomial for which the degree must also be chosen (Barak, 1995). In present study, S-G (with the window size of ten), SNV, MSV, first and second derivatives and combination of them were applied.

2.4. classification

Spectra contain high volumes of information, which are very difficult to interpret by visual inspection only. Chemometrics is a tool for extracting this information from the multivariate chemical data, using mathematics. Chemometrics is generally applied to explore patterns of association in data; track properties of materials on a continuous basis or to prepare and use multivariate classification models. By utilizing diverse preprocessing techniques, the generation of principal models is triggered and subsequently produces output data. Both unsupervised and supervised techniques for classification were utilized in the study.

2.4.1. unsupervised classification

In the first steps of data exploration, Principal Component Analysis (PCA) is usually applied to recognize any possible separated groups. The main objective of PCA model as an unsupervised modelling method is decreasing the dimensionality of data and preservation of the present variation (Jolliffe, 1986). The reduction of dimensionality is done by defining new variables, principal Components (PCs) that consists linear combinations of the original data (Kamruzzaman, Barbin, ElMasry, Sun, & Allen, 2012). First PC represents the most variance of dataset and the next PCs which are orthogonal to the preceding ones contain the most of the remaining variance (Fodor, 2002). The used data matrix for PCA model in this study consists of 1886 columns (corresponding to the recorded wavenumbers) and 150 rows (corresponding to the number of samples).

2.4.2 supervised classification

2.4.2.1. SVM model

The Support Vector Machine classification (SVMC) is a supervised classification technique that utilizes kernel functions to represents the original space in the format of feature space. It determines the best separation between classes by applying a unique hyperplane to the dataset (Ballabio & Todeschini, 2009; Fletcher, 2009; Vapnik, 1999). The final classification outcomes of SVM are determined by a small number of Support Vectors that are the samples lying on the margins of the model. For building classification model and evaluation of their performance, calibration and test datasets were used, respectively. 70% of data was appointed as training and 30% of data was considered as test dataset. SVs lie to the closest boundaries between classes. In SVM model various kernel functions encompassing linear, Radial Basis Function (RBF), Sigmoid, and polynomial could be employed (Chandrasekaran, Panigrahi, Ravikanth, & Singh, 2019). It is necessary for the correct selection of functions since the type of kernel function directly impacts the model's performance and outcomes (Kazemi, Mahmoudi, & Khojastehnazhand, 2023).

2.4.2.2. ANN model

Recently, Artificial Neural Network (ANN) elucidated from human brain function has been one of the most commonly used modeling technique for classification. The functioning of Neural Networks relies on input, hidden, and output layers, each containing varying numbers of neurons. Neurons have a weight assigned to them based on the model's training and serve as storage for the model's inputs and calculation layers. Randomly assigning weights to neurons sets the foundation for training an ANN model. The present study employed a feed-forward network, a type of neural network methodology, where 70% of data was initially used for training purposes and the remaining 30% for testing purposes.

2.5. regression modeling

After classification of samples, the prediction of adulterated level was done by using partial least squares regression (PLSR). The utilization of PLSR helps to enhance the interconnection between spectral data and the features that need to be quantified. By distinguishing between X and Y variables, PLSR defines a set of new features named latent variables, which are characterized as orthogonal and linear combinations of the X variables (Peng, Cheng, Wang, & Zhu, 2020). In present study, the PLSR model was applied to the FT-MIR spectra to investigate the possibility to predict the percentage of sodium hydrosulfite adulteration in wheat flour. The reliability of the acquired predicted model is explored by using external validation data. 70% of the dataset were used to build calibration model and 30% of the dataset was used for testing the created model.

The assessment of acquired models are done by sensitivity and specificity according to equations 1 and 2 (Kazemi, Mahmoudi, Veladi, Javanmard, & Khojastehnazhand, 2022):

$$\text{Sensitivity (\%)} = \frac{TP}{TP+FN} * 100 \quad (1)$$

$$\text{Specificity (\%)} = \frac{TN}{TN+FP} * 100 \quad (2)$$

Where TP (True Positive) are the number of samples belonging to either pure flour correctly classified as pure samples; FP (False Positive) are the number of mixed samples wrongly classified

as pure samples; TN (True Negative) are the number of mixed samples correctly classified as mixed; FN (False Negative) are those pure samples classified as mixed. These two statistical parameters take values between 0 and 1. The higher their values, the better the classification performance of models.in the regression modeling, Root Mean Square Error (RMSE) of calibration (RMSEC), prediction (RMSEP), and coefficient of determination (R^2) values are important parameters which evaluate the predictive power of a PLS calibration model. Higher predictive power is represented with higher R^2 and lower RMSE (Pebriana, Rohman, Lukitaningsih, & Sudjadi, 2017; Rohman & Salamah, 2018). For PLS calibration models developed to predict the amount of adulteration in adulterated flour, RMSEC and RMSEP can be calculated using equations 3 and 4, where Y_i and $\hat{Y}_i$ are the actual and predicted values of an adulterated sample, respectively. M and N are the number of data in calibration and prediction set, respectively(Sikorska, Khmelinskii, & Sikorski, 2014).

$$\text{RMSEC} = \sqrt{\frac{\sum(\hat{Y}_i - Y_i)^2}{M-1}} \quad (3)$$

$$\text{RMSEP} = \sqrt{\frac{\sum(\hat{Y}_i - Y_i)^2}{N}} \quad (4)$$

$$\text{CCR} = \frac{TP+TN}{TP+FN+TN+FN} \quad (5)$$

3. Results and discussion

3.1. spectra interpretation

FT-MIR spectra of pure wheat flour, sodium hydrosulfite, and adulterated samples with different adulteration levels is displayed in fig.2. due to some peaks overlaps, chemometric tools is necessary to extract information. In most of peaks almost all of the classes showed similar peaks except Blankit. We had major peaks at 1050 cm^{-1}, 1730 cm^{-1}, 2950 cm^{-1}, and 3400 cm^{-1}. But the wavelength of pure Blankit was different and except some peaks like 1050 cm^{-1}, in the majority of peaks of other adulterated classes, it did not have peaks and also in some cases like 1950 cm^{-1} and 2050 cm^{-1} it showed peaks but other classes did not have. The basic bands at 2800-3040 cm^{-1} are related to C-H and C-H_2 symmetric and asymmetric stretching and mainly attributed to band vibrations of the lipids in the flours (Roa, Santagapita, Buera, & Tolaba, 2014). The bands with the maximum at 1640 cm^{-1} are associated to protein band vibrations (Guzmán-Ortiz et al., 2015). Furthermore, spectra show a strong absorption band, from 900-1200 cm^{-1} and C-H bending (1000 cm^{-1}), mainly related to carbohydrates (Rodríguez, Rolandelli, & Buera, 2019).

3.2. PCA model

PCA model as unsupervised modeling was applied to dataset to decrease the dimensions of data as preserving the original variables. The acquired FT-MIR data was processed by PCA model to explore the probable similarities and differences among pure and adulterated flour samples. With the comparison of different applied preprocessing techniques, the result of PCA model with (S-G

+ D1 + SNV) was the best. The obtained score plot of first two PCs (PC1=88% and PC2 = 4%) is shown in fig.3. this figure displays that, all the pure samples were projected on PC1 negative values. Thus, PC1 provided a fairly discrimination between pure and adulterated samples. As it is observable from fig.3, the pure flour samples were gathered and separated well from the adulterated samples. Due to high chemical composition difference of pure Sodium hydrosulfite, the hydrosulfite samples were well-separated and were located on the other side of PC1. Because of similarities of the chemical composition of adulterated samples, there were some misclassifications between different adulterated level groups. Similarities of compositional structure of samples with different adulteration levels can be a reason for misclassification of adulterated samples. Mishra et al, applied PCA model combined with hyperspectral imaging method to detect peanut traces in wheat flour with the presentation of 99.43% of variance, pure and adulterated samples were well-distinguished along PC1 like the present study (Mishra et al., 2015). In addition, the results of PCA model in the present study was in agreement with the result of PCA model for discrimination of wheat flour with other cereal flours (barley, rye, and triticale flour). The score plot represented good discrimination of barley flour samples from wheat flour. However, one type of wheat flour was located very close to other flours (Nur Arslan, 2020).

3.3. SVM and ANN models

Table.1. represents the accuracy of SVM and ANN models as supervised classification methods after applying various kinds of preprocessing methods and combination of them for training datasets. The SVM model was implemented in four different kernel functions (linear, polynomial, Radial Basis Function (RBF), and Sigmoids). In both models, 70% of data was randomly assigned to model training and other 30% were used for model testing. In addition, 5% of 70% neural network model training data was used to validate them. As table.1. shows, the accuracy of SVM model with linear kernel function with S-G + D2 + SNV preprocessing was 86.66% and also 86.70% for ANN models. based on the used preprocessing methods, the ANN model also yielded acceptable results based on the optimal neural network structure shown in Figure.4. Fig.5. represents SVM graphical model after employing S-G + D2 + SNV preprocessing. As table.1. represents, linear kernel had better results for all preprocessing techniques. Linear kernels work well when the underlying relationship between the input features and the target variable is approximately linear. The better performance of linear kernel maybe is because of the nature of the dataset, which is separated or modeled effectively by linear boundaries. Furthermore, in high-dimensional spaces, like spectroscopic data, linear kernels can perform better than more complex kernels. This is because complex kernels can exacerbate the curse of dimensionality making it harder to find a suitable decision boundary. In addition, according to the better outcomes of polynomial kernel in comparison with other kernels, it can be concluded that the structure and nature of dataset tends to simple and linear form. In another study, Yuan et al, employed NIR spectroscopy to detect Sodium hydroxymethanesulfonate in wheat flour. Three algorithms including PLS-DA, advanced K-means dynamic clustering, and LS-SVM were used to establish the calibration models. The outcomes of LS-SVM outperformed other two methods, with the classification accuracy of 94.70% for the prediction (Yuan, Xiang, Yu, & Xu, 2011). However, the outcomes of SVM model in the mentioned research was better than our present study but this point should be mentioned that, the applied SVM model in that study was for classification of two classes

but 86.66% result of present study was for classification of five classes. In spite of the fact that FTIR spectroscopy combined with chemometric methods confirmed its application to detect the adulteration of Sodium hydrosulfite in wheat flour, but there were some limitations in the present study which we hope to be solved or tried in the future studies. The environmental effects like moisture are different from the bakeries or industrial places. Although, due to the fact that the moisture was similar to all of samples, this issue was solved. But for application of this study method in other situations, definitely the conditions of system should be calibrated again. Besides that, the applied technique in the present study can be studied for detection of other adulterants in wheat flour simultaneously. In order to assess the classification ability of each class in SVM model, the confusion matrix was investigated for test dataset (table 2). The results were assessed by calculation of sensitivity, specificity, and accuracy statistical parameters. As it was expected, the highest classification accuracy was for class 1 (pure wheat flour). Also, the accuracy of class C (adulterated with 15% level) was 100% too. The weakest classification result was for class D (20% adulterated). In class E (25% adulterated) 6 samples were classified correctly, and 3 samples were classified for class D. The difference of these classes is 5% adulteration level. Then, the classification result of this class was acceptable too.

3.4. PLSR model

In order to quantify the adulterant in wheat flour, FT-MIR spectroscopy-based regression model (PLSR) was built. This regression model is based on developing algebraic correlation between the quantity of adulteration in wheat flour samples and absorption of sample along different wavelengths. The ideal calibration model was determined based on lowest RMSEC, RMSECV, and highest R_c^2 and R_{cv}^2. The value of mentioned statistical parameters as well as the number of "Latent Variables" (LVs) are presented in table.3. As shown in table3, the best PLS model was obtained with the preprocessing method of S-G using seven LVs, which showed the prediction performance (R_{cv}^2 = 0.994 and RMSECV = 0.123). The similarity of train and test results was representee of model's good ability for prediction of precise levels of adulteration. The performance of PLSR was also externally validated by using the test set of samples, as shown in fig.6. The PLS prediction plot illustrates that PLSR model displayed a very good prediction ability (R_p^2 = 0.992). Fig.6. shows the relationship between reference data and predicted values obtained in the laboratory. Recently, Martins et al, predicted the whey protein in wheat flour by FT-IR spectroscopy and multivariate analysis. The PLSR model was applied to the acquired spectra and the best model of obtained spectra had R_{cal}^2 = 0.99, R_{pre}^2 = 0.98, RMSEC = 3.5, and RMSEP = 3.00 (M. S. Martins et al., 2022). However, the R^2 results were in agreement with this research, but RMSE results were weaker. In other research, Nur Arslan, applied PLSR model to explore the amount of barely flour in wheat flour. The statistical parameters of this study was close to the results of present study (R^2 values were at least 0.994 and RMSECV result was in the range 0.36-1.50%) (Arslan et al., 2020). In another study, the prediction of Azodicarbonamide in wheat flour by visible/near-infrared spectroscopy was investigated by Che et al. By comparing 3 applied models in this research (PLSR, Back Propagation Neural Network, and Radial Basis Function), Radial Basis Function model had the best prediction results with Correlation Coefficient ®, RMSEP 0.99 and 0.54, respectively (Che et al., 2017), and were in agreement with the outcomes of present study.

Conclusion

The presence of Sodium hydrosulfite (Blankit) in wheat flour was investigated by FT-MIR spectroscopy. PCA as unsupervised and SVM and ANN as supervised models were applied to detect the adulteration and PLSR model as regression model was applied to quantify the amount of adulteration. The mentioned chemometric models were built after some preprocessing techniques. The acquired results for detection and quantification of Sodium Hydrosulfite proved that FT-IR spectroscopy can be a reliable method to detect and quantify Sodium hydrosulfite in wheat flour.

Declarations

Conflict of interest

There is no potential conflict of interest between the authors.

Adams, M. L., Lombi, E., Zhao, F. J., & McGrath, S. P. (2002). Evidence of low selenium concentrations in UK bread□making wheat grain. *Journal of the Science of Food and Agriculture, 82*(10), 1160-1165.

Ahamadabadi, M., Saeidi, M., Rahdar, S., Narooie, M. R., Salimi, A., Alipour, V., . . . Biglari, H. (2016). Amount of baking soda and salt in the bread baked in city of zabol. *Iioab Journal, 7*, 518-522.

Arslan, F. N., Akin, G., Karuk Elmas, Ş. N., Üner, B., Yilmaz, I., Janssen, H.-G., & Kenar, A. (2020). FT-IR spectroscopy with chemometrics for rapid detection of wheat flour adulteration with barley flour. *Journal of Consumer Protection and Food Safety, 15*, 245-261.

Asgari, G., SeidMohammadi, A., Faradmal, J., Moradi, M. J., & Yari, K. (2018). The Study of Blankit concentration as an unauthorized additive in hamadan bakeries and risk assessment of food consumption. *Pajouhan Scientific Journal, 16*(4), 21-27.

Ballabio, D., & Todeschini, R. (2009). Multivariate classification for qualitative analysis *Infrared spectroscopy for food quality analysis and control* (pp. 83-104): Elsevier.

Barak, P. (1995). Smoothing and differentiation by an adaptive-degree polynomial filter. *Analytical Chemistry, 67*(17), 2758-2762.

Boysworth, M., & Booksh, K. (2008). Aspects of multivariate calibration applied to near-infrared spectroscopy. *PRACTICAL SPECTROSCOPY SERIES, 35*, 207.

Chandrasekaran, I., Panigrahi, S. S., Ravikanth, L., & Singh, C. B. (2019). Potential of near-infrared (NIR) spectroscopy and hyperspectral imaging for quality and safety assessment of fruits: An overview. *Food Analytical Methods, 12*, 2438-2458.

Che, W., Sun, L., Zhang, Q., Zhang, D., Ye, D., Tan, W., . . . Dai, C. (2017). Application of visible/near□infrared spectroscopy in the prediction of azodicarbonamide in wheat flour. *Journal of food science, 82*(10), 2516-2525.

Christy, A. A., & Kvalheim, O. M. (2007). Latent-variable analysis of multivariate data in infrared spectrometry. *Near-infrared spectroscopy in food science and technology*, 145-162.

de Carvalho, L. M., & Schwedt, G. (2005). Sulfur speciation by capillary zone electrophoresis: Determination of dithionite and its decomposition products sulfite, sulfate and thiosulfate in commercial bleaching agents. *Journal of Chromatography A, 1099*(1-2), 185-190.

De Girolamo, A., Cervellieri, S., Mancini, E., Pascale, M., Logrieco, A. F., & Lippolis, V. (2020). Rapid authentication of 100% italian durum wheat pasta by FT-NIR spectroscopy combined with chemometric tools. *Foods, 9*(11), 1551.

Dhanoa, M., Lister, S., Sanderson, R., & Barnes, R. (1994). The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *Journal of Near Infrared Spectroscopy, 2*(1), 43-47.

Fletcher, T. (2009). Support vector machines explained. *Tutorial paper*, 1-19.

Fodor, I. K. (2002). *A survey of dimension reduction techniques*. Retrieved from

GhR, J. K., Yunesian, M., Vaezi, F., Nabizadeh, R., & GhA, P. (2006). A Survey on Baking Soda Elimination from Iranian Flat Breads in Bakeries of Islamshahr City in 2005. *Res J School Health, Yazd, 1*(2), 21-31.

Govaert, F., Temmerman, E., & Kiekens, P. (1999). Development of voltammetric sensors for the determination of sodium dithionite and indanthrene/indigo dyes in alkaline solutions. *Analytica chimica acta, 385*(1-3), 307-314.

Guzmán-Ortiz, F. A., Hernández-Sánchez, H., Yee-Madeira, H., San Martín-Martínez, E., Robles-Ramírez, M. d. C., Rojas-López, M., . . . Mora-Escobedo, R. (2015). Physico-chemical, nutritional and infrared spectroscopy evaluation of an optimized soybean/corn flour extrudate. *Journal of food science and technology, 52*, 4066-4077.

Jolliffe, I. T. (1986). Principal Component Analysis and Factor Analysis. In I. T. Jolliffe (Ed.), *Principal Component Analysis* (pp. 115-128). New York, NY: Springer New York.

Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W., & Allen, P. (2012). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies, 16*, 316-325.

Karami, M., Alikord, M., Mokhtari, Z., Sadighara, P., & Jahed-Khaniki, G. (2021). Sodium Hydrosulfite (Blankit) in Iranian food as a threat to human health: a review. *Journal of Food Safety and Hygiene*.

Kazemi, A., Mahmoudi, A., & Khojastehnazhand, M. (2023). Detection of sodium hydrosulfite adulteration in wheat flour by FT-MIR spectroscopy. *Journal of Food Measurement and Characterization, 17*(2), 1932-1939.

Kazemi, A., Mahmoudi, A., Veladi, H., Javanmard, A., & Khojastehnazhand, M. (2022). Rapid identification and quantification of intramuscular fat adulteration in lamb meat with VIS–NIR spectroscopy and chemometrics methods. *Journal of Food Measurement and Characterization, 16*(3), 2400-2410.

Lavigne-Delcroix, A., Tusseau, D., & Proix, M. (1996). Validation of a chromatographic chemiluminescence detector. *Sciences des Aliments, 16*, 267-280.

Lohumi, S., Lee, S., Lee, H., & Cho, B.-K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science & Technology, 46*(1), 85-98.

López-Maestresalas, A., Insausti, K., Jarén, C., Pérez-Roncal, C., Urrutia, O., Beriain, M. J., & Arazuri, S. (2019). Detection of minced lamb and beef fraud using NIR spectroscopy. *Food Control, 98*, 465-473.

Malakootian, M., & Dowlatshahi, S. (2005). The quality of the manufactured bread and hygienic conditions of bakeries. *Journal of environmental health science & engineering, 2*(2), 72-78.

Martins, F. C., Sentanin, M. A., & De Souza, D. (2019). Analytical methods in food additives determination: Compounds with functional applications. *Food chemistry, 272*, 732-750.

Martins, M. S., Nascimento, M. H., Barbosa, L. L., Campos, L. C., Singh, M. N., Martin, F. L., . . . Barauna, V. G. (2022). Detection and quantification using ATR-FTIR spectroscopy of whey protein concentrate adulteration with wheat flour. *LWT, 172*, 114161.

Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J. M., Diezma, B., Gorretta, N., & Lleó, L. (2015). Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *Journal of Near Infrared Spectroscopy, 23*(1), 15-22.

Mohamed, M. Y., Solihin, M. I., Astuti, W., Ang, C. K., & Zailah, W. (2019). *Food powders classification using handheld near-infrared spectroscopy and support vector machine.* Paper presented at the Journal of Physics: Conference Series.

Monnier, G., & Wiliams, S. (1972). Determination of sulfur dioxide.(AOAC Official Method 990.28). *analyst, 95*, 119.

Pallone, J. A. L., dos Santos Caramês, E. T., & Alamar, P. D. (2018). Green analytical chemistry applied in food analysis: alternative techniques. *Current Opinion in Food Science, 22*, 115-121.

Pebriana, R. B., Rohman, A., Lukitaningsih, E., & Sudjadi. (2017). Development of FTIR spectroscopy in combination with chemometrics for analysis of rat meat in beef sausage employing three lipid extraction systems. *International journal of food properties, 20*(sup2), 1995-2005.

Peng, L., Cheng, H., Wang, L.-J., & Zhu, D. (2020). Comparisons of the prediction results of soil properties based on fuzzy c-means clustering and expert knowledge from laboratory Visible–Near-Infrared reflectance spectroscopy data. *Canadian Journal of Soil Science, 101*(1), 33-44.

Pisoschi, A. M., Pop, A., Gajaila, I., Iordache, F., Dobre, R., Cazimir, I., & Serban, A. I. (2020). Analytical methods applied to the assay of sulfur-containing preserving agents. *Microchemical Journal, 155*, 104681.

Rethmeier, J., Rabenstein, A., Langer, M., & Fischer, U. (1997). Detection of traces of oxidized and reduced sulfur compounds in small samples by combination of different high-performance liquid chromatography methods. *Journal of chromatography A, 760*(2), 295-302.

Reza, T., Aktar, S., Al Amin, H., Rahman, M., Arefin, A., Mohanto, N. C., . . . Nikkon, F. (2014). In vivo analysis of toxic effect of hydrose used in food preparations in Bangladesh. *Asian Pacific Journal of Tropical Biomedicine, 4*(11), 884-889.

Rinnan, Å., Nørgaard, L., van den Berg, F., Thygesen, J., Bro, R., & Engelsen, S. B. (2009). Data pre-processing. *Infrared spectroscopy for food quality analysis and control*, 29-50.
Roa, D. F., Santagapita, P. R., Buera, M. P., & Tolaba, M. P. (2014). Ball milling of Amaranth starch-enriched fraction. Changes on particle size, starch crystallinity, and functionality as a function of milling energy. *Food and bioprocess technology, 7*, 2723-2731.
Rodríguez, S. D., Rolandelli, G., & Buera, M. P. (2019). Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. *Food chemistry, 274*, 392-401.
Rohman, A., & Salamah, N. (2018). The employment of spectroscopic techniques coupled with chemometrics for authentication analysis of halal pharmaceuticals. *Journal of Applied Pharmaceutical Science, 8*(10), 063-068.
Sabeghi, M. (2004). Interview with Dean of Faculity of flour and Bread. *Journal of Iran dough-baked, 3*, 5-6.
Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry, 36*(8), 1627-1639.
Shewry, P. (2009). The HEALTHGRAIN programme opens new opportunities for improving wheat for nutrition and health: Wiley Online Library.
Shewry, P. R., Powers, S., Field, J. M., Fido, R. J., Jones, H. D., Arnold, G. M., . . . Barro, F. (2006). Comparative field performance over 3 years and two sites of transgenic wheat lines expressing HMW subunit transgenes. *Theoretical and Applied Genetics, 113*, 128-136.
Sikorska, E., Khmelinskii, I., & Sikorski, M. (2014). Vibrational and electronic spectroscopy and chemometrics in analysis of edible oils. *Methods in Food Analysis; Cruz, RMS, Khmelinskii, I., Vieira, M., Eds*, 201-234.
Topping, D. (2007). Cereal complex carbohydrates and their contribution to human health. *Journal of Cereal Science, 46*(3), 220-229.
Vapnik, V. (1999). *The nature of statistical learning theory*: Springer science & business media.
Varmuza, K., & Filzmoser, P. (2016). *Introduction to multivariate statistical analysis in chemometrics*: CRC press.
Yuan, W., Xiang, B., Yu, L., & Xu, J. (2011). A non-invasive method for screening sodium hydroxymethanesulfonate in wheat flour by near-infrared spectroscopy. *Food Analytical Methods, 4*, 550-558.
Zeaiter, M., Roger, J.-M., & Bellon-Maurel, V. (2005). Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *TrAC Trends in Analytical Chemistry, 24*(5), 437-445.

Figures:

Fig.1. the schematic flowchart of the steps of present study

Fig.2.the acquired FT-MIR spectra for flour samples

Fig.3. the score plot of PCA model

Fig.4. The structure of the optimal artificial neural network.

Fig.5. SVM graphical model for classification of samples

Fig.6. The performance of PLSR model for prediction of adulteration levels

Table1. the accuracy of SVM and ANN models

Table.2. the confusion matrix of SVM model for test dataset

Table 3. The results of PLSR model in predicting the adulteration level using different preprocessing methods

Fig.1.

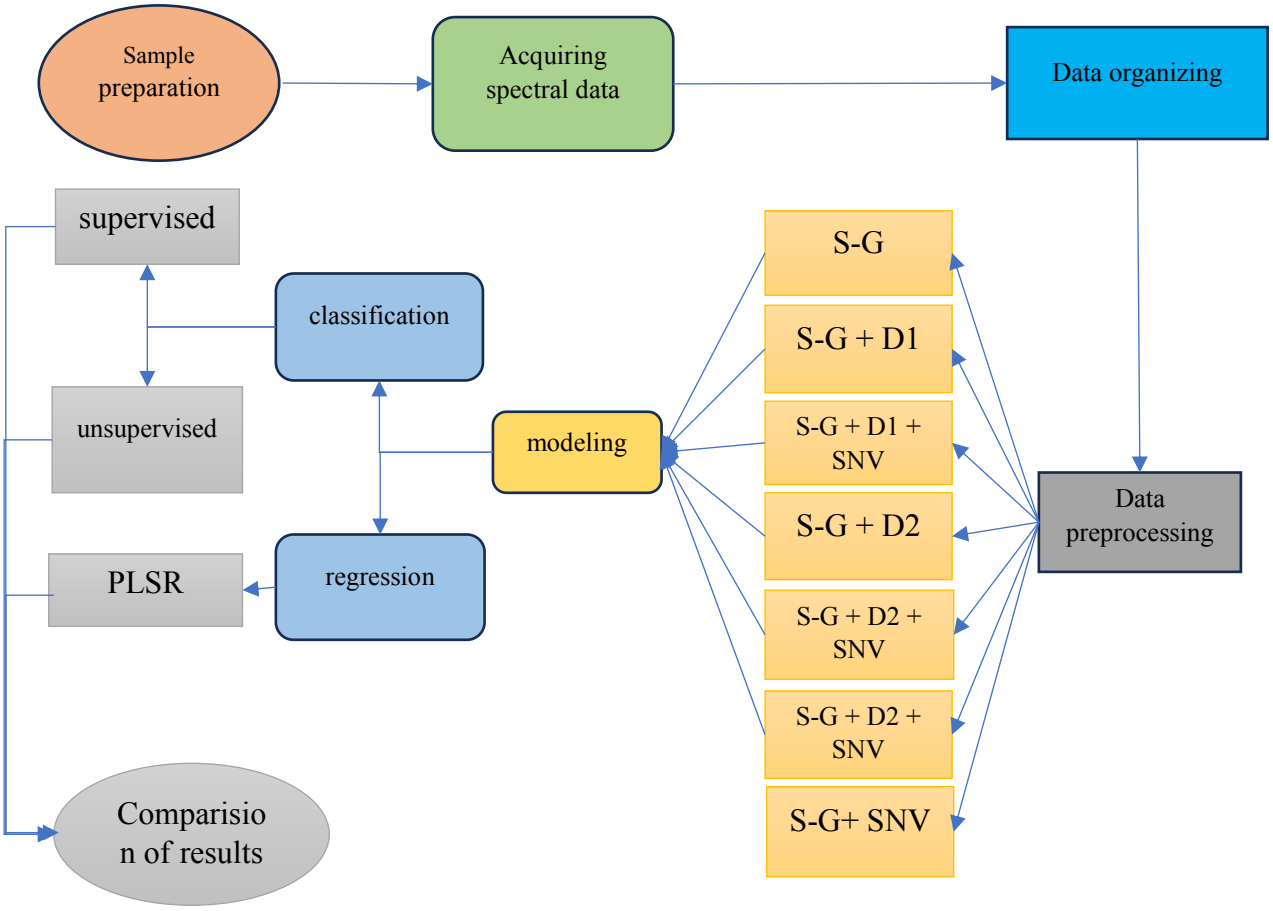
Sample preparation
Acquiring spectral data
Data organizing
supervised
classification
unsupervised
modeling
PLSR
regression
Comparisio n of results
S-G
S-G + D1
S-G + D1 + SNV
S-G + D2
S-G + D2 + SNV
S-G + D2 + SNV
S-G+ SNV
Data preprocessing

Fig.2.

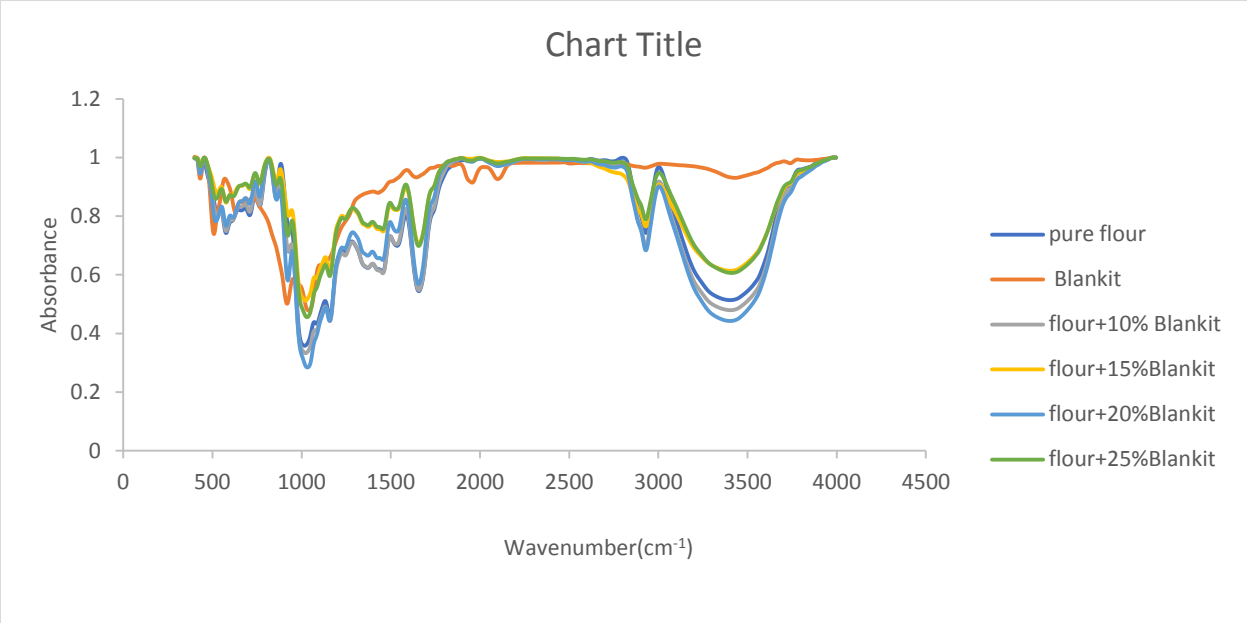

Fig.3.

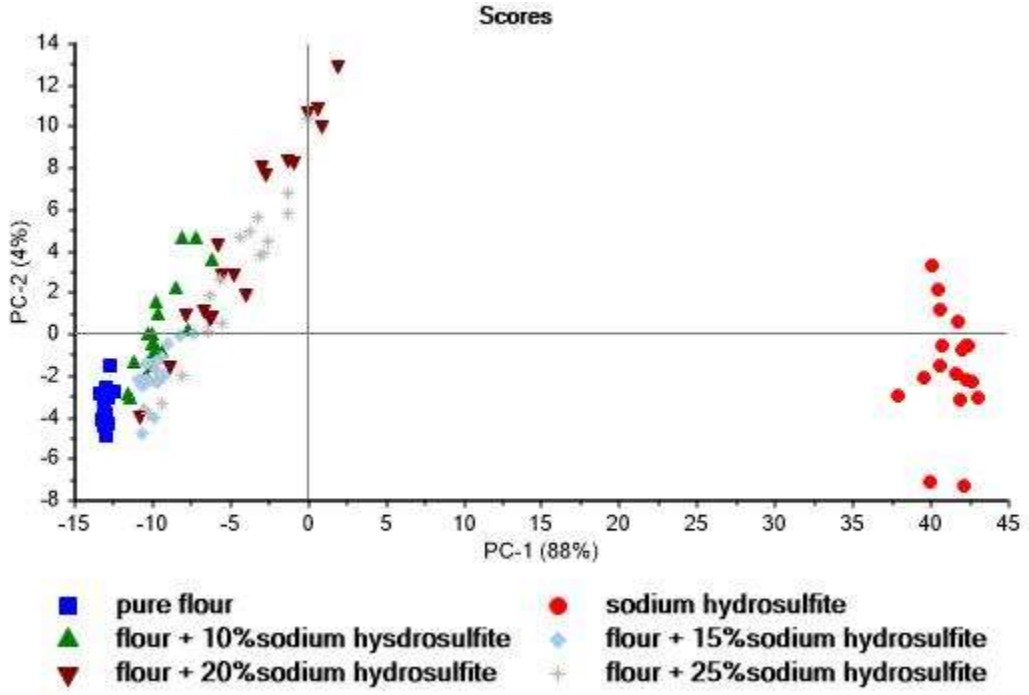

Fig.4.

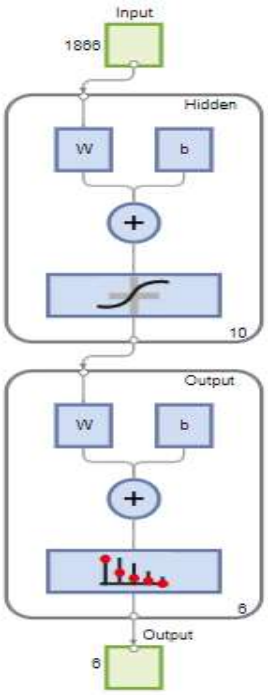

Fig.5.

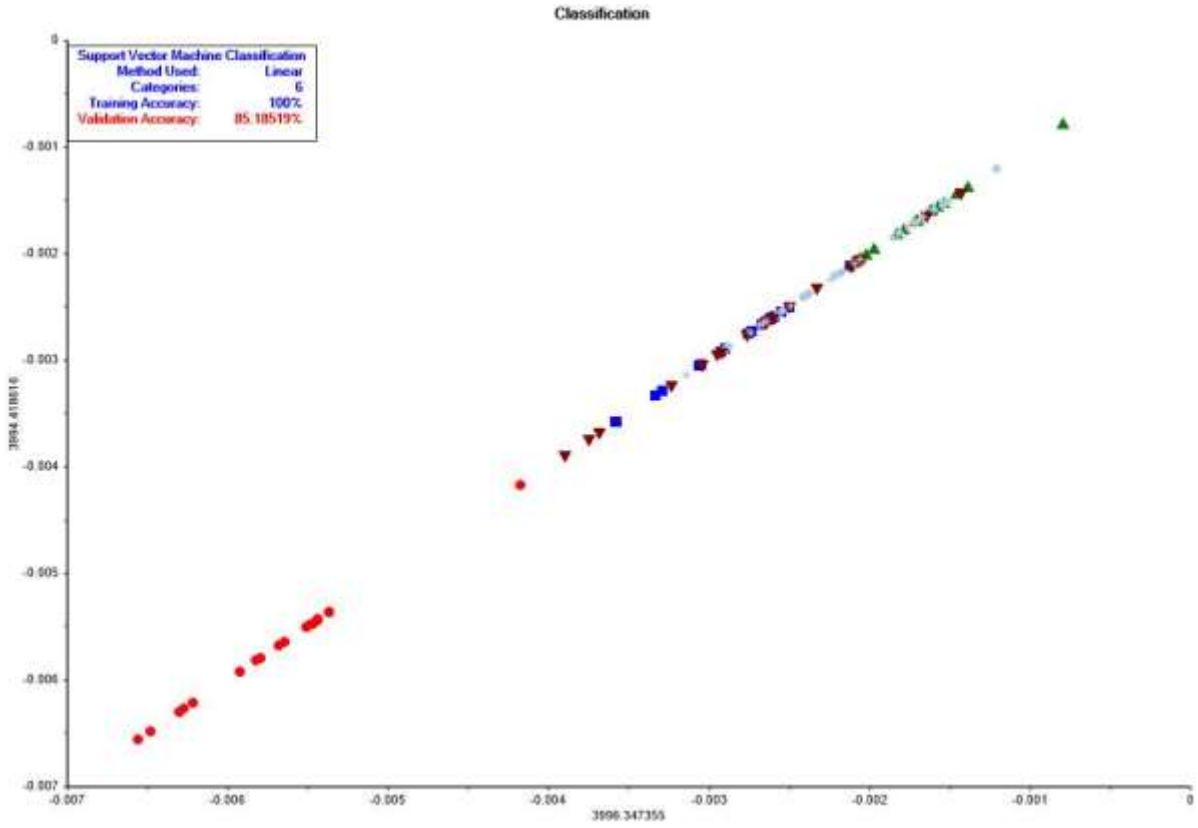
Classification
Support Vector Machine Classification
Method Used: Linear
Categories: 6
Training Accuracy: 100%
Validation Accuracy: 85.18519%
3994.418818
-0.001
-0.002
-0.003
-0.004
-0.005
-0.006
-0.007
-0.007
-0.006
-0.005
-0.004
-0.003
-0.002
-0.001
0
3996.347355

Fig.6.

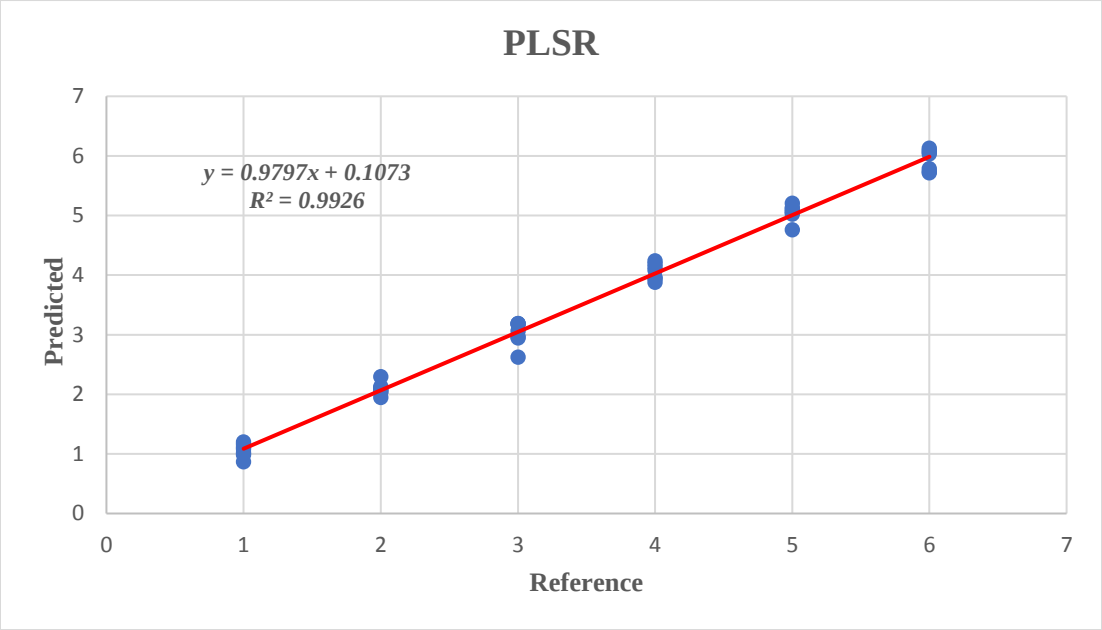

Table.1.

model	SVM												ANN		
Kernel function	linear			polynomial			RBF			Sigmoid			-		
preprocessing	train	val	test	train	val	test	train	val	test	train	val	test	train	val	test
S-G	78.75	68.75	75.55	12.50	11.25	6.66	10	13.75	6.66	17.5	18.75	20	88.9	80	80
S-G +D1	20	17.5	20	20	20	20	20	20	20	20	20	20	84.4	80	75.6
S-G + D1 + SNV	100	81.25	84.44	67.5	50	71.11	52.5	43.75	51.11	1.25	10	2.22	98.9	100	80
S-G + D2	20	20	20	20	20	20	20	20	20	20	20	20	50	60	46.7
S-G + D2 + SNV	**100**	**81.25**	**86.66**	**68.75**	**53.75**	**68.88**	**61.25**	**45**	**55.55**	**1.25**	**23.75**	**4.44**	**100**	**100**	**86.7**
S-G + D2 + MSC	20	20	20	20	20	20	20	20	20	20	20	20	98.9	100	80
S-G + SNV	93.75	70	80	30	30	30	30	28.75	26.66	10	12.5	8.88	85.6	86.7	73.3

Table.2.

	A	B	C	D	E
A	9	0	0	0	0
B	0	9	0	0	0
C	0	1	8	0	0
D	0	1	0	7	1
E	0	0	0	3	6
Sensitivity	1	1	0.88	0.77	0.66
Specificity	1	0.94	1	0.91	0.97
CCR(%)	100	81	100	70	85

Table.3.

preprocessing	LV	Calibration		test	
		R^2	RMSE	R^2	RMSE
SG	7	0.995	0.118	0.994	0.123
S-G + D1	7	1.00	2.46	1.00	2.39
S-G + D1 + SNV	7	0.987	0.36	0.979	0.82
S-G + D2	7	1.00	3.94	1.00	2.87
SG + D2 + SNV	7	0.975	0.265	0.967	0.312
SG + D2 + MSC	7	1.00	2.87	1.00	2.50
S-G+ SNV	7	0.992	0.15	0.989	0.18