Journal of Animal & Plant Sciences, 34(4): 2024, Page: 886-895
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2024.4.0773

ON-SITE IDENTIFICATION OF RAPESEED VARIETIES WITH HANDHELD NEAR-INFRARED SPECTROPHOTOMETER AND MACHINE LEARNING ALGORITHMS

K. Ahmad[1], H. M. Ahmed[2] and A. Shah[3]

[1]Food and Nutrition Division, Nuclear Institute for Food & Agriculture, Peshawar
[2]Plant Breeding & Genetics Division, Nuclear Institute for Food & Agriculture, Peshawar
[3]Chemistry Department, Quaid-i-Azam University, Islamabad.
Corresponding author's email: khahmad314@gmail.com

ABSTRACT

Rapeseed is one of the most important oil crops in the world, but its oil quality and seed yield are affected by the genetic purity of the varieties. Plant variety identification plays a vital role in maintaining genetic purity leading to improve seed business. Current methods for rapeseed variety identification include visual inspection and modern techniques such as DNA profiling. The former which are based on phenotypic character may be prone to error while the later may be expensive and cannot be performed on-site. NIR spectroscopy offers a rapid and non-destructive approach that could overcome these limitations. This study is aimed to evaluate the potential of portable/handheld NIR to make a supervised classification model for the rapeseed varieties. The seed samples (N=225) of three (03) rapeseed varieties were scanned with handheld SCiO NIR sensor and the average of the three scans were used for classification of varieties. The classification model developed by the combination of different pre-processing and classification algorithms were tested on unknown samples (n=75). It was found that all classifiers exhibited good results except Partial Least Square–Discrimination Analysis (Rc^2=0.8). SIMCA classification was tested which correctly identified 96.4% and 93.3% samples from training and test sets respectively followed by Random Forest classifier (F1=0.97) with a success rate of 93.3% on test set. However, Support vector machine (C-SVM type) with a polynomial kernel (3rd degree) gave accurate results after a combination of Standard Normal Variate (SNV) and first order Savitzky-Golay derivative (polynomial degree of 2) with number of smoothing points (window size) of 5. It classified 100% samples of training set and 97.3% samples of test set into their correct classes. Based on initial evaluation of four classification algorithm, it was found that SVM can be better utilized for varietal classification. This study reveals that handheld NIR can be a reliable and useful tool for rapeseed variety identification, which can benefit both the seed industry as well as the farmers.

Keywords: NIR spectroscopy, support vector machine, SIMCA, standard normal variate, Savitzky–Golay, classification

Published first online June 04, 2024 Published final August 25, 2024

INTRODUCTION

Rapeseed is considered to be the second-highest yielding oil crop worldwide and accounted for 12.1% of world major vegetable oil in 2021. Over the last ten years the global production has increased by 20.9% to 70.3 million tons in 2020 (Zheng and Liu, 2022). Rapeseed oil has various applications in food, feed, biofuel, and industrial sectors, and it is valued for its nutritional and environmental benefits. However, rapeseed production and quality are affected by various factors, such as climate change, pests, diseases, and genetic purity (Sendekie, 2020). Therefore, it is essential to develop and maintain high-quality rapeseed varieties that can meet the diverse and growing demands of the market.

Development of a plant variety is a costly task in terms of skill, infrastructure and genetic resources as well as a time-consuming, (approx. 10 years). Therefore, variety identification is an important errand in breeding programs to maintain genetic purity of varieties, prevention of fraud and improvement of seed trade (Xu *et al.*, 2009; Kosmowski and Worku, 2018).

There are several methods employed for variety identification including traditional morphological characters method and modern techniques (Morell *et al.*, 1995). The former, although, fast and less expensive methods but some time misleading due to lack of phenotypic variation causes difficulties for variety discrimination. Modern techniques such as molecular Markers, DNA profiling techniques, are most precise and accurate methods yet require a well-developed laboratory equipment, experienced personnel, and high costs to execute the discrimination study. These are methods are destructive and cannot be performed on-site. Similarly in agriculture sector, we need a large number of samples to be the true representative of a population which also limits the use of modern sophisticated methods for the said purpose. Therefore, the thrust for an accurate, fast, reliable, and cost-effective method of identification of plant varieties is always there to support agriculture research and development.

Recently many researchers used near infrared (NIR) spectroscopy for classification of several agriculture products (Lerma-García *et al*., 2018). NIR operates in small electromagnetic region (e.m.r) from 750-2500 nm responsible for overtones and combination bands of vibration. As the overtone are less probable, therefore, this region of e.m.r. has weak interaction with mater. The weak interactive mode of near infrared region make it difficult to be used for analytical purpose but the same nature has a benefits that NIR can penetrate deep into a sample which make it an excellent region to study interior of a sample (Lammertyn *et al*., 2000). The disadvantage of weak interaction (weak signals) can be overcome by use of some advanced statistical tools e.g. pretreatment of spectral data and use of some algorithm's including Partial Least Square Regression-Discrimination Analysis (PLS-DA), Support Vector Machine (SVM) and Soft Independent Modelling by Class Analogy (SIMCA) collectively called chemometrics (López *et al*., 2017; Kyprianidis and Skvaril, 2017).

The literature review showed that several scientists have used near infrared spectroscopy for successful classification of several fruits and crops instead of laboratory based sophisticated expensive techniques. Abu-Kalaf and his group used NIR in reflectance mode for classification of carrot on sensory characteristics by using multivariate data analysis with an accuracy greater than 84% (Abu-Khalaf *et al*., 2004). Betrand and his team classified wheat (*Triticum aestivum* L.) varieties with 87% success using NIR in reflectance mode in combination with chemometrics (Bertrand *et al*., 1985). Zou used full range of VIS/NIR (300-2500 nm) for the classification of rapeseed varieties using distance discriminant analysis (DDA) and Back Propagation Neural Network (BPNN) (Zou *et al*., 2011). Several researchers applied different statistical tools on spectral data of NIR and developed models for quality assessments of different biochemical traits such as oils, protein, glucosinolates, etc. (Hom *et al*., 2007).

However the main advantage of at-site application can only be obtained by application of portable/handheld NIR devices. As the handy NIR usually operate in a narrow portion of near infrared range due using simple light detector for lowering the cost of device. Which means that much less number of spectral data point is used for model development which usually compromise the high accuracy of analysis. However a reasonable high accuracy can be achieved, if proper statistical tools is applied for smoothing and removal of noise and scattering effect. Recently several manufacturer claimed for the development of pocket size NIR sensor with different spectral ranges. Cen and his group used handheld VIS/NIR for classification of orange varieties using Back Propagation Neural Network (BPNN) combine with PCA and PLS-DA and concluded that PCA+BPNN gave better result (Cen *et al*., 2007). In another study, a portable NIR was used for successful classification of grapevine varieties using SVM, ANN and PLS-DA as classifier algorithms (87.25%). They found that SVM and ANN outperformed PLS-DA (Gutierrez *et al*., 2015). He *et al*. (2007) developed a fast method for successful identification of apple varieties using NIR and chemometrics such as Wavelength Transform (WT), and ANN and claimed a classification model development with 100% accuracy.

The main objective of this study was to develop a model for on-site rapeseed variety differentiation and confirmation, by evaluating the efficiency of a pocket size NIR device to make a classification model for the rapeseed varieties developed at NIFA. The study used a handheld SCiO NIR sensor, which is a portable and affordable device that can scan rapeseed seeds and send the spectral data to a smartphone app for analysis. This study contributes to the field of rapeseed variety identification by demonstrating the potential and feasibility of using a pocket size NIR device and chemo-metrics for rapid and non-destructive classification of rapeseed varieties. This study also benefits the seed industry and the farmers by providing them with a convenient and reliable tool for selecting and verifying rapeseed varieties, which can improve the quality and yield of rapeseed production.

MATERIALS AND METHODS

Rapeseed samples: Three hundred samples of three rapeseed varieties viz Durr-e-NIFA (DN), NIFA Gold (NG) and NIFA Sarson-T20 (NS) developed by Plant Breeding and Genetics Division of Nuclear Institute for Food and Agriculture (NIFA), Peshawar, were collected under the supervision of breeders during harvest season 2020-21 and 2021-22 and stored at oilseed store at room temperature. Three scans of intact grain samples were recorded by using SCiO sample holder and averaged. The sample holder was shaken well before recording each scan sequentially to ensure averaged scan data would be representative of the larger sample. Three hundred samples were randomly split into seventy five samples of each variety as training set (N= 225) and twenty five samples as test set (n=75).

NIR Spectrometer Sensor and Chemometrics: NIR sensor (make: Consumer Physics SCIO) having dimension 18.8 × 40.2 × 67.7 mm and weighing 35 g, was used for recording spectral finger print (McGrath *et al*., 2021). An android app "The Lab" installed on smartphone was used to operate and record the spectra.

The device operate in wavelength range of 740–1070 nm with 1nm resolution (331 data points as dependent variables). The spectra were stored in Consumer Physics cloud database and can be accessed through a window-based app "The Lab web" (https://thelab.consumerphysics.com/), a website of the Consumer Physics which give access to the spectra stored in their cloud (Wiedemair *et al*., 2019). The Lab Web can also be used to view the data in PCA view and to remove the outliers. The Lab Web was used for pre-processing data and developing and testing classification models. The Lab Web allow some pretreatment method such as SNV, Derivatives, de-trending etc. and Random Forest (RF) algorithm to classify the samples with F1 value and confusion matrix to check the performance of the model. The data were downloaded to PC and exported to "The Unscrambler X", v10.4 (CAMO software) for further analysis.

Pre-processing and classification algorithms: In this study several pretreatment methods combined with predictive algorithms were compared by using the Unscrambler X software. NIR spectroscopy data can have some challenges, such as high dimensionality, non-linearity, noise, outliers, and multicollinearity. Therefore, to handle these challenges the performance of four advanced algorithms including Partial Least Squares Discriminant Analysis (PLS-DA), Support Vector Machine (SVM), soft independent modelling by class analogy (SIMCA) and Random Forest (RF) were checked to develop predictive models of grain cultivar identification. Different preprocessing methods including Standard Normal Variate (SNV), smoothing with Savitzky-Golay derivative (SG-d) and Baseline Correction were used to convert the raw spectra to more informative curve. The SNV is used to normalize spectra and reduces scatter and particle size effects while SG-d and baseline correction are employed to enhance spectral features and removes baseline drift by smoothing and differentiating spectra (Tamburini *et al*., 2017). It was found that the combination of SNV and derivative yielded best results in all classification models.

RESULTS AND DISCUSSION

Rapeseed is considered to be the second important oilseed crop having a rich source of oil, proteins and many saturated and unsaturated fatty acid (Raymer, 2002; Reddy, 2017). The varietal identification by naked eye is a real daunting task and most of the time leads to incorrect identification, until and unless some clear cut morphological markers are not available. Instead of costly and destructive methods of seed identification, researchers are trying to make fast, cheap and non-destructive methods for this purpose. The intact seeds of all the three varieties were scanned with pocket size NIR sensor (Fig. 01). Due to poor interaction of near infrared region with matter, the raw spectra usually shows no apparent peaks however, after applying some pretreatments, the spectra can be converted into more informative curve. The raw spectra were pretreated with Standard Normal Variate (SNV), and first order Savitzky-Golay derivative with polynomial order of 2 and different window size followed by model building. As the spectra were recorded on whole grain the scattering effect can be minimized by SNV while derivatives are good mean to elaborate a weak peaks by resolving the overlapped peaks (Barnes *et al*., 1989; Huang *et al*., 2010). After baseline correction, the averaged spectra showed peaks around 914 nm and 960-970 nm which are attributed to 3rd and 2nd overtones of C-H and O-H stretching respectively (Akyar, 2011; Subedi *et al*., 2012; Bantadjan *et al*., 2020) as the rapeseed consist of 45-50% oil (Westad *et al*., 2008). Four classification methods were used along with different pretreatments to get best combination of pretreatment and classification algorithm for classification.

Partial Least Square-Discriminate Analysis (PLS-DA): Partial Least square (PLS) regression is a well-known algorithms commonly used for estimation purposes however it can be applied for classification purposes by selecting some dummy variables for each class(Cen *et al*., 2007). PLS-DA with a random cross validation of twenty segments (15 samples per segment) was used to make a classification model on preprocessed (SNV + SG-d, 1st, 2, W35) spectra of N=225 samples, the score plot is shown in Fig. 02. The values of coefficient of regression for calibration (R^2_c=0.78) and validation (r^2 =0.74) showed that the model would poorly classify the unknown samples. When same model was used to classify the unknown samples (n=75) of test set which were not included in calibration set, 65% of samples were correctly classified to their respected groups. However, by changing the pretreatment conditions i.e. decreasing the window size of SG derivative from 35 to 5 and using full cross validation, a small improvement was observed in model. The R^2c was 0.80 and 84.9% samples of training set and 68% samples of the test set were correctly classified.

As the SCiO NIR sensor operate in a short range of NIR region (331 data points), which is mostly the region of 2nd and 3rd overtones. The absorptions intensities of these overtones are very weak due to less probable nature as compared to 1st overtone vibration. Therefore, ability of PLS-DA for classification may be enhanced by selecting spectra of broader range (i.e including the region of 1st overtone).

Support Vector Machine (SVM): Support Vector Machines (SVM) is a supervised machine learning method which use a kernel function to map the data from the original space to a new feature space. In the new feature space, the kernel function then finds a support vectors for the best classification (Pierna *et al.*, 2004). After applying pretreatment, the spectra were subjected to different SVM classifier by using different kernel and the results are summarized in Table 01 with percent accuracies of different pretreatment and SVM type.

The best training and validation accuracy of 100% and 99.1% was obtained for C-SVM with capacity factor, C=1 after pretreatment combination of SNV and derivative (SG-d, 1st, 2, W5) and the kernel used was polynomial kernel type with degree of 3. The confusion matrix shown in Fig. 03 depicts that all samples of training set were predicted correctly into their rightful classes.

The samples from test set which were not included in calibration/training model, were then subjected to the classifier developed. As shown in Table 01, the combination of preprocessing (SNV+SG-d, 1st, 2, W5) and C-SVM algorithm of polynomial degree 3 classified maximum number of samples into their rightful classes.

Soft Independent Modelling by Class Analogy (SIMCA): Soft Independent Modeling of Class Analogy (SIMCA) is another statistical method for supervised classification of data based on making a principal component analysis (PCA) model for each class separately in the training set. Unknown samples are then compared with all models and assigned to classes according to their closeness to respective PCA model in the training samples (Hibbert, 2016). The score plot of PCA (Fig. 04A) shows that 99% variance in the data could be explained by first two PCs and that NG show maximum variation as compared to DN and NS, the same is evident from the influence plot of F-residual and Hotelling's T^2 (Fig. 04B). The influence plot is used to show the two kinds of outlier. The residual value describes the sample distance to model, whereas the Leverage and Hotelling's T^2 define the ability of model to describe the sample. The samples with high leverage are well described by model but retaining that sample may span an entire component.

Separate PCAs model were built for all three varieties by using Singular Value Decomposition (SVD) algorithm on the preprocessed data. It is evident from Table 02 that almost all samples of training set were described by their respective model except one sample of DN and two samples of NS (false negative). Similarly 22 samples of DN were described by PCA (NG) and 9 by PCA (NS) (false positive). Similarly 5 samples of NG were false positively described by PCA (DN). By changing the pretreatment condition i.e. reducing window size from 35 to 5, all the false negative samples reduced to zero. In new condition all samples were classified by their respective PCAs. However, the doubly classified samples still exist in model (Table 02), therefore some other statistical tools must be applied to decide about their classes.

To decide about the rightful class of a doubly classified samples, is used to calculate the distance of sample from both PCA models (Sim *et al.*, 2004). The PCA of DN, NG and NS plotted on the Coomans' plot, to determine the classification of the unknown sample by measuring its distance from the either of the classes along vertical and horizontal axis (Fig. 05). The shortest distance from any of the classifier from the point of the unknown sample will decide its rightful class. By comparing the distances of all samples from the three models, it was found that 217 samples (96.4%) of training set and 70 (93.3%) samples of test set were correctly classified into their rightful classes.

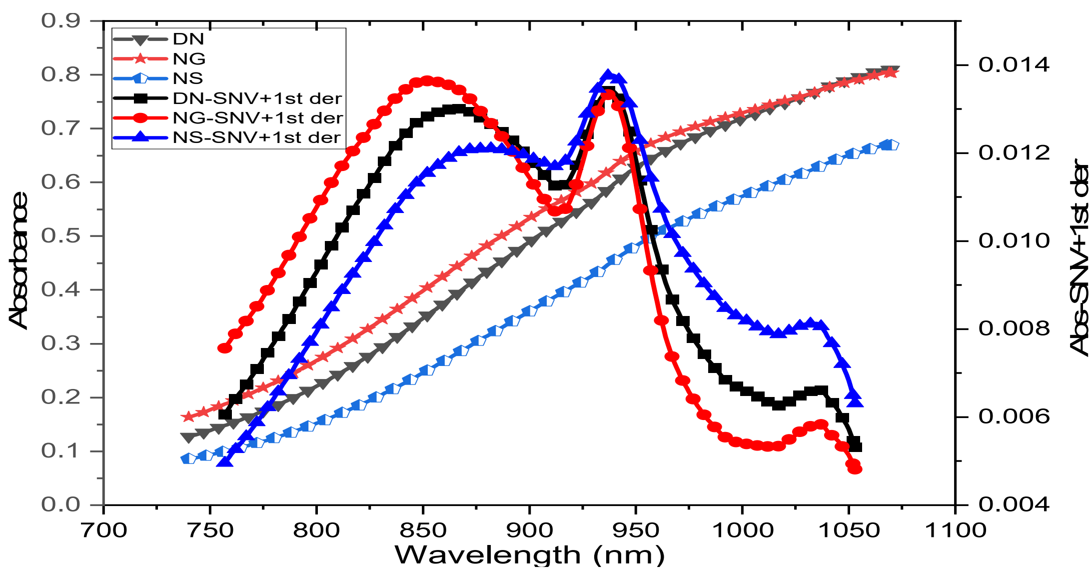

Figure 01. Averaged raw spectra and preprocessed with SNV + 1st order S-Golay derivative (W35)

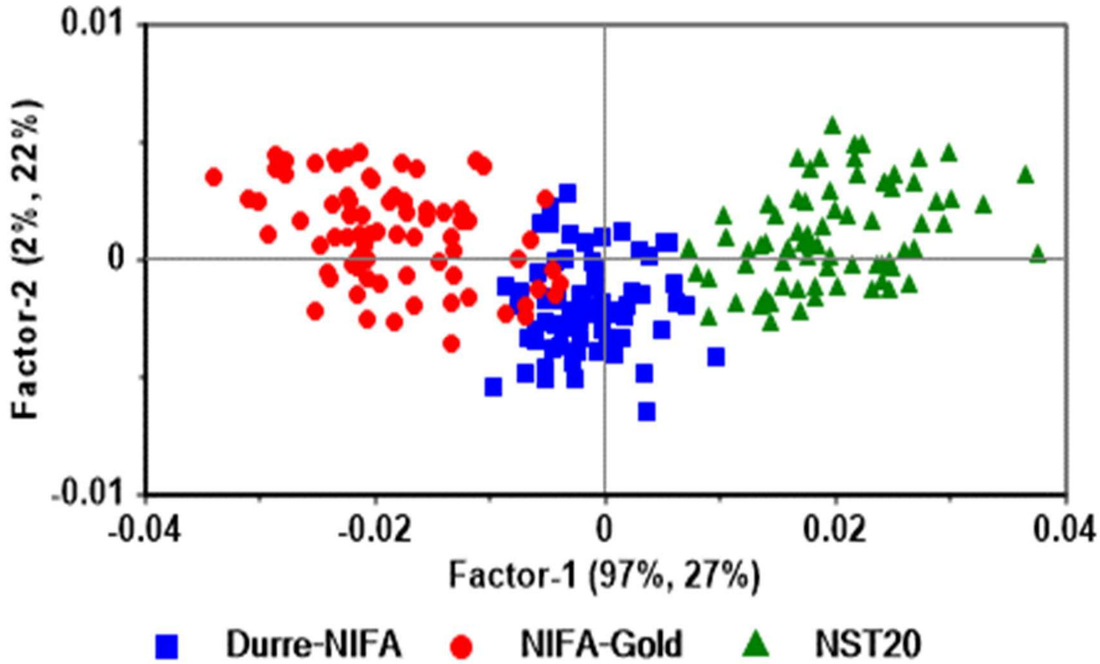

Figure 02. Factor-1 and 2 showing that the first two factors explain 99% variation in data

Table 01. Different combinations of preprocessing before classification (SVM) along with accuracies.

Pretreatment	Kernel Type	Calibration Accuracy (%)	Validation Accuracy (%)	Test set Accuracy (%)
Raw data	Polynomial (nu-SVM)	80.9	79.1	77.3
SG-d, W35	Polynomial (nu-SVM)	43.6	64.0	44.0
SNV	Polynomial (nu-SVM)	94.2	93.8	92.0
SNV+ SG-d, W5	Polynomial (nu-SVM)	93.8	92.9	93.3
SNV+ SG-d, W35	Polynomial (nu-SVM)	93.8	92.9	93.3
SNV+ SG-d, W35	Radial (nu-SVM)	93.8	93.8	93.9
SNV+ SG-d, W35	Polynomial (C-SVM)	98.7	98.7	96.0
SNV+ SG-d, W5	Polynomial (C-SVM)	100	99.1	97.3

Where: SG-d= 1st order S-Golay derivative of polynomial order of 2 and W= window, nu=0.5, C=1 and polynomial degree of SVM = 3

		Actual		
		Durre-NIFA	NIFA-Gold	NST20
Predicted	Durre-NIFA	100%	0%	0
	NIFA-Gold	0	100%	0
	NST20	0	0	100%

Figure 03 Confusion matrix of samples of training set, classified with SVM (polynomial kernel of 3rd degree) after pretreatment of SNV and derivative (SG-d, 1st, 2, W5)

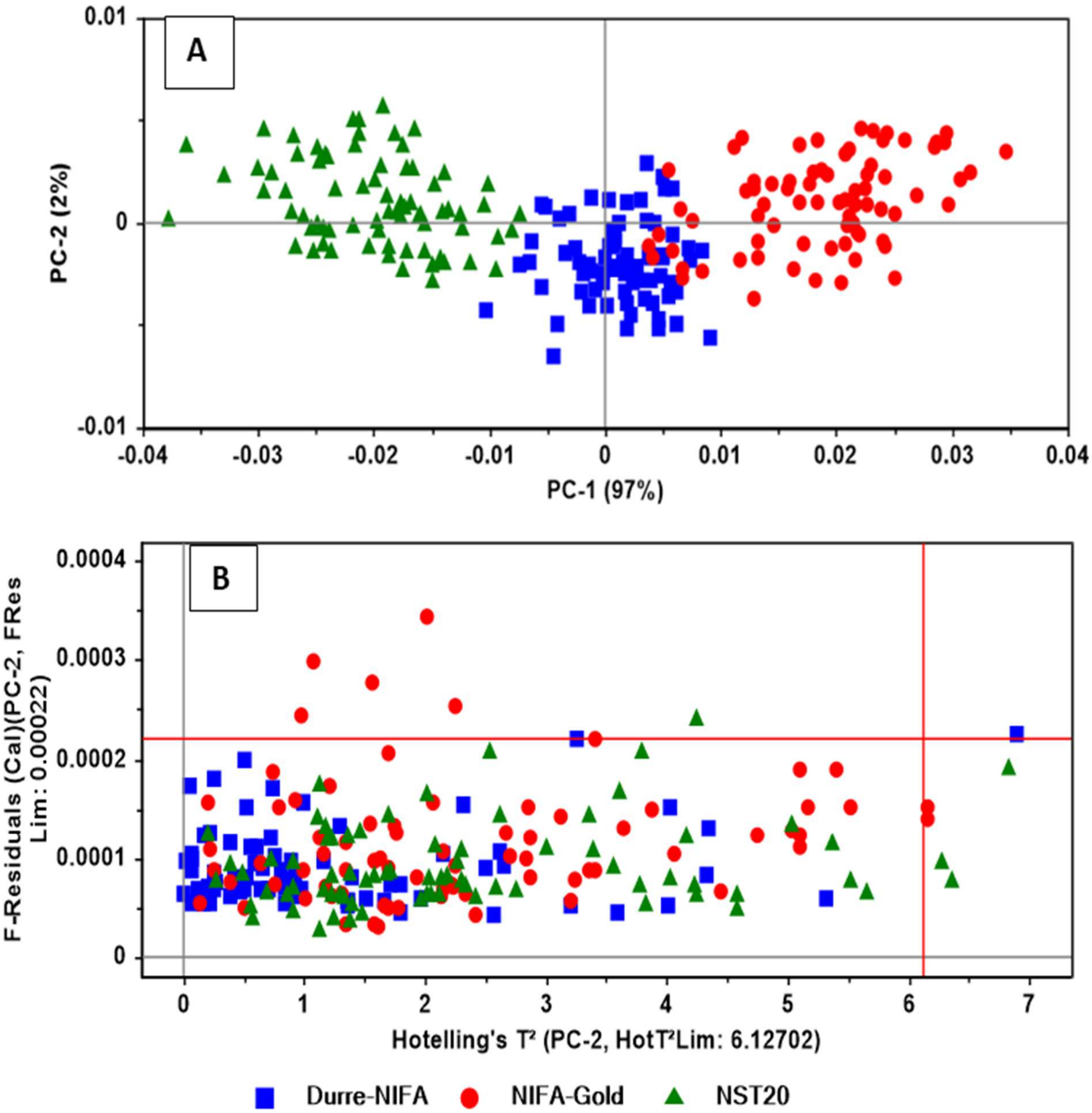

Figure 04. PCA model of all three varieties (A) and influence plot of F-Residual and Hotelling's T^2 (B).

Table 02 Confusion matrix for test set classification (SIMCA-model).

Pretreatment	*Variety*	*PCA(DN)*	*PCA(NG)*	*PCA(NS)*
	DN	24	22	9
SNV + SG-d,1st, 2, W35)	NG	5	25	0
	NS	0	0	23
	DN	25	22	10
SNV + SG-d,1st , 2, W5	NG	5	25	0
	NS	2	0	25

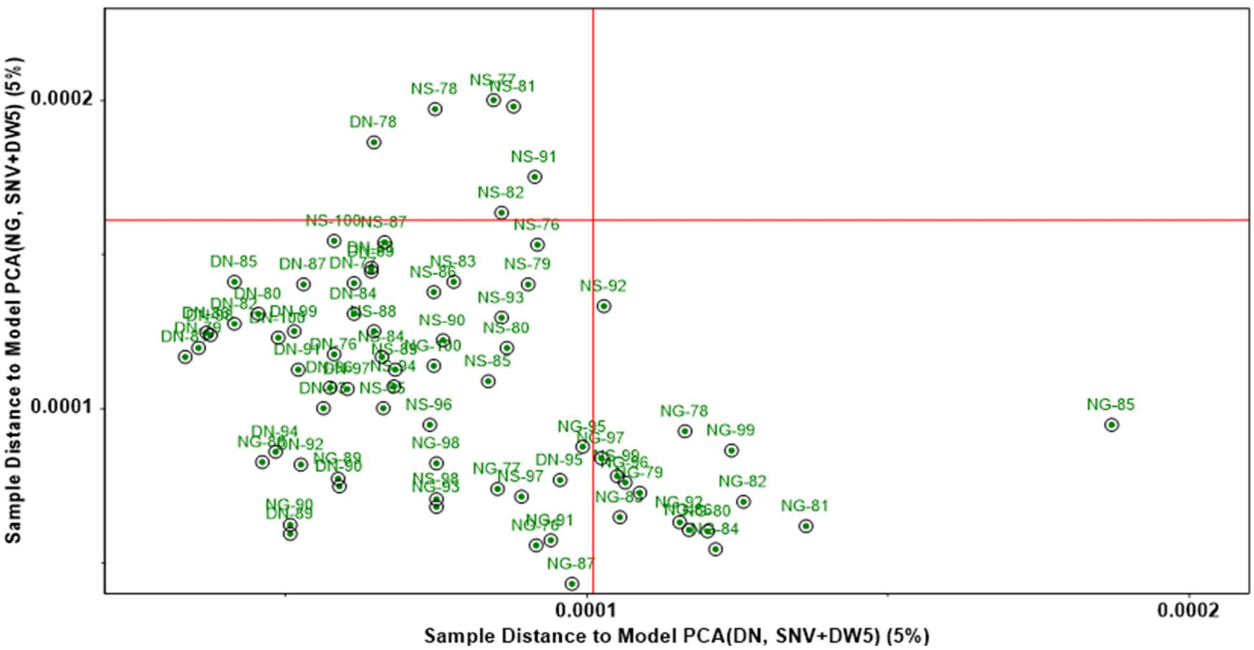

Figure 05 Coomans' Plot of PCA DN and PCA NG showing the distance of samples from each model.

Classification model using the "The Lab web": The spectra collected were also analyzed by SCiO Lab, a web-based app of consumer physics. An android app "the Lab" was used to collect the spectra and data stored at Consumer Physics cloud was accessed by "the Lab Web". The lab web allows the users to select a set of pre-processing functions prior to development of classification models. After preprocessing the averaged spectra are shown in Fig. 06. However, there is a single algorithm available known as Random Forest (RF) algorithm which make a more accurate and stable prediction based on numerous decision trees (Shannon *et al*., 2021). The accuracy (overall proportion of correctly classified samples) of model is checked by F1 score (also called F score) and confusion matrix. F1 score can be applied to multiclass classification and is calculated from the precision and recall of the test (Chicco and Jurman, 2020). The precision (to correctly identifying all samples of a variety) and recall (not to miss any samples of a variety) are calculated by equation 01 and 02.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positve(TP) + False\ Nigative(FN)} \qquad 01$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positve(TP) + False\ Positive(FN)} \qquad 02$$

Where

TP = samples that are correctly classified in a certain class.

FP= samples that don't belong to a certain class and are incorrectly classified in that class.

FN= samples that belong to a certain class and the model fail to identify them in that class..

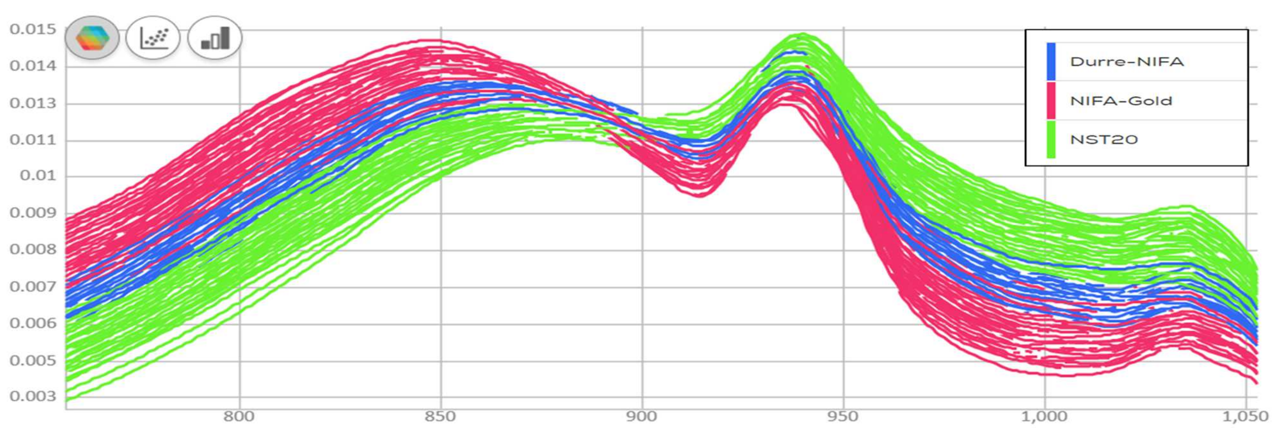

Figure 06 Spectra of three selected varieties after applying preprocessing methods (SNV and 1st order derivative) on "The Lab Web" view.

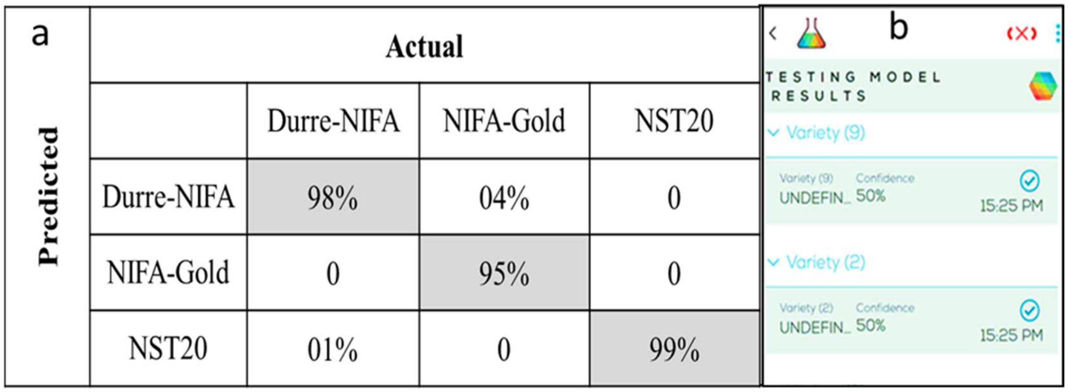

Figure 07 (a) Confusion matrix by Random Forest classifier using preprocessing methods (SNV and 1st order derivative) and (b) testing model on other varieties

Table 03 Summary of all combinations of preprocessing and classification algorithms.

Classification Algorithm	Pretreatment	Correctly classified samples: Training set N=225	Correctly classified samples: Test set n=75
PLS-DA (NIPLAS)	SNV+ SG-d, W35	82.6%	65%
	SNV+ SG-d, W5	84.9%	70%
C-SVM (polynomial kernel)	SNV+ SG-d, W35	98.7%	96.0%
	SNV+ SG-d, W5	100%	97.3%
SIMCA (SVD)	SNV+ SG-d, W35	96.0%	92.6%
	SNV+ SG-d, W5	96.4%	93.3%
Random Forest	SNV+ SG-d, W35	--	93.3%

After checking the several preprocessing methods and their combination, it was revealed that the combination of SNV and derivative (SG-d, 1st, 2, W35) gave best result with F1 score of 0.97 for the classification of three varieties (confusion matrix is given in Fig. 07a) on training set with 10 fold cross validation by samples.

Using the “Test model” option in “The Lab” android app, the model was checked on 75 samples (25 of each) of all the three varieties. 70 of total samples (93.3%) were correctly classified into their rightful classes with a confidence level of 70-99%. However the 4 samples of NG were classified with a confidence level 54-70% which was re-checked thrice to confirm the class. More ever the models developed were checked on samples of other varieties (60 samples) which were not used in calibrations, by using the “test model” option (Fig. 07b). Out of total 56 samples i.e. 93.3% were identified as outliers..

Where: SG-d= 1st order S-Golay derivative of polynomial order of 2 and W= window size (smoothing points)

The study proved that NIR spectroscopy can be used non-destructively for plant variety identification on-site. Ultimately, this research paves the way for future tools that could revolutionize how to identify and manage plant varieties across agriculture, promoting both precision and sustainability. However the accuracy of model can be improved by extending the region of NIR range (750-2500 nm) and including more samples of the varieties of interest

Conclusions: Based on the results, it was concluded that C-SVM (C=1) classification in combination with pretreatment of SNV and 1st order derivative exhibited accurate results on both training and test sets compared to SIMCA and Random Forest classifications. In this study classification models were developed for selected three varieties and sample analyzed qualified over 90% for their rightful class. Therefore, it is suggested that handheld SCiO NIR sensor can be best utilized for classification of all available rapeseed varieties and spot analysis to determine the purity and rightfulness of unknown seed samples class through C-SVM (C=1).

Authors’ contribution: Mr. Khurshid Ahmad framed the basic idea and designed the experiments. He did spectra collection, statistical analysis of the data and writing of manuscript. Hafiz Munir Ahmed designed experimental portion and writing of the manuscript. Dr. Afzal Shah helped in basic idea & discussion.

Conflict of Interest: The authors have no conflicts of interest to declare.

REFERENCES

Abu-Khalaf, N., B. S. Bennedsen, and G. K. Bjørn. 2004. Distinguishing carrot's characteristics by near infrared (NIR) reflectance and multivariate data analysis. Agricul. Enginr. Int.: CIGR J. 3(12):1-17. doi: https://cigrjournal.org/index.php/Ejounral/article/view/511/505

Akyar, I. 2011. Wide Spectra of Quality Control, 1st Ed. Maastricht, (Netherlands). IntechOpen p237.

Bantadjan, Y., R. Rittiron, K. Malithong, and S. Narongwongwattana. 2020. Establishment of an Accurate Starch Content Analysis System for Fresh Cassava Roots Using Short-Wavelength Near Infrared Spectroscopy. ACS Omega 5(25):15468-15475. doi: 10.1021/acsomega.0c01598

Barnes, R., M. S. Dhanoa, and S. J. Lister. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. App. spectr. 43(5):772-777. doi.org/10.1366/0003702894202

Bertrand, D., P. Robert, and W. Loisel. 1985. Identification of some wheat varieties by near infrared reflectance spectroscopy. J. the Science of Food & Agriculture 36(11):1120-1124. doi: https://doi.org/10.1002/jsfa.2740361114

Cen, H., Y. He, and M. Huang. 2007. Combination and comparison of multivariate analysis for the identification of orange varieties using visible and near infrared reflectance spectroscopy. Eur. Food Res. Tech. 225(5):699-705. doi: 10.1007/s00217-006-0470-2

Chicco, D., and G. Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 21(1):1-13. doi:10.1186/s12864-019-6413-7

Gutierrez, S., J. Tardaguila, J. Fernandez-Novales, and M. P. Diago. 2015. Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer. PloS one DOI:10(11):e0143197.

He, Y., X. Li, and Y. Shao. 2007. Fast Discrimination of Apple Varieties Using Vis/NIR Spectroscopy. Int. J. Food Prop. 10(1):9-18. doi: 10.1080/10942910600575666

Hibbert, D. B. 2016. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). Pure & Applied Chemistry 88(4):407-443.

Hom, N. H., H. C. Becker, and C. Möllers. 2007. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. Euphytica 153(1):27-34. DOI:10.1007/s10681-006-9195-3

Huang, J., S. Romero-Torres, and M. Moshgbar. 2010. Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. Am. Pharm. Rev. http://www.americanpharmaceuticalreview. 13(6):116-127.

Kosmowski, F., and T. Worku. 2018. Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. PloS one 13(3):e0193620.

Kyprianidis, K., and J. Skvaril. (2017). Developments in Near-Infrared Spectroscopy 1st Ed., London (UK). IntechOpen p97.

Lammertyn, J., A. Peirs, J. De Baerdemaeker, and B. Nicolaï. 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Biol. Technol. 18(2):121-132. doi: https://doi.org/10.1016/S0925-5214(99)00071-X

Lerma-García, M. J., V. Cortés, P. Talens, and J. M. Barat. 2018. Variety discrimination of fruits, edible plants, and other foodstuffs and beverages by infrared spectroscopy, Compr. Anal. Chem. No. 80. Elsevier. p. 127-163. https://doi.org/10.1016/bs.coac.2018.03.004

López, M. G., A. S. García-González, and E. Franco-Robles. 2017. Carbohydrate analysis by NIRS-Chemometrics: Developments in Near-Infrared Spectroscopy [Internet]. InTech; 2017. Available from: http://dx.doi.org/10.5772/62932 p81.

McGrath, T. F., M. Shannon, O. P. Chevallier, R. Ch, F. Xu, F. Kong, H. Peng, E. Teye, S. Akaba, and D. Wu. 2021. Food Fingerprinting: Using a two-tiered approach to monitor and mitigate food fraud in rice. J. AOAC Int. 104(1):16-28. DOI: https://doi.org/10.1093/jaoacint/qsaa109

Morell, M., R. Peakall, R. Appels, L. Preston, and H. Lloyd. 1995. DNA profiling techniques for plant variety identification. Aust. J. Exp. Agric. 35(6):807-819. https://doi.org/10.1071/EA9950807

Pierna, J. F., V. Baeten, A. M. Renier, R. Cogdill, and P. Dardenne. 2004. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. J. ournal of Chemom. 18(7-8):341-349. DOI: 10.1002/cem.877

Raymer, P. L. 2002. Canola: an emerging oilseed crop. ASHS Press, Alexandria, V.A. pp 122-126 1

Reddy, G. V. 2017. Integrated management of insect pests on canola and other Brassica oilseed crops. CABI

Sendekie, Y. 2020. Review on; Seed genetic purity for quality seed production. Int. J. Sci. Eng. Sci. 4:1-7. ISSN: 2456-7361

Shannon, M., C. Ratnasekhar, T. F. McGrath, A. P. Kapil, and C. T. Elliott. 2021. A two-tiered system of analysis to tackle rice fraud: The Indian Basmati study. Talanta 225:122038. DOI: https://doi.org/10.1016/j.talanta.2020.122038

Sim, C. O., M. R. Hamdan, Z. Ismail, and M. N. Ahmad. 2004. Assessment of herbal medicines by chemometrics–assisted interpretation of FTIR spectra. J. Anal. Chimica Acta 1:14.

Subedi, P. P., K. B. Walsh, and D. W. Hopkins. 2012. Assessment of titratable acidity in fruit using short wave near infrared spectroscopy. Part A: establishing a detection limit based on model solutions. J. Near Infrared Spectrosc. 20(4):449-457.

Tamburini, E., F. Vincenzi, S. Costa, P. Mantovi, P. Pedrini, and G. Castaldelli. 2017. Effects of moisture and particle size on quantitative determination of total organic carbon (TOC) in soils using near-infrared spectroscopy. Sensors 17(10):2366. https://doi.org/10.3390/s17102366

Westad, F., A. Schmidt, and M. Kermit. 2008. Incorporating chemical band-assignment in near infrared spectroscopy regression models. J. Near Infrared Spectrosc. 16(3):265-273. https://doi.org/10.1255/jnirs.786

Wiedemair, V., D. Langore, R. Garsleitner, K. Dillinger, and C. Huck. 2019. Investigations into the performance of a novel pocket-sized near-infrared spectrometer for cheese analysis. Molecules 24(3):428. doi: 10.3390/molecules24030428

Xu, H.-r., P. Yu, X.-p. Fu, and Y.-b. Ying. 2009. On-site variety discrimination of tomato plant using visible-near infrared reflectance spectroscopy. J. Zhejiang Univ. Sci. B 10(2):126-132. doi: 10.1631/jzus.B0820200

Zheng, Q., and K. Liu. 2022. Worldwide rapeseed (Brassica napus L.) research: A bibliometric analysis during 2011–2021. Oil Crop Science 7(4):157-165. https://doi.org/10.1016/j.ocsci.2022.11.004

Zou, Q., H. Fang, F. Liu, W. Kong, and Y. He. 2011. Comparative Study of Distance Discriminant Analysis and Bp Neural Network for Identification of Rapeseed Cultivars Using Visible/Near Infrared Spectra. In: Computer and Computing Technologies in Agriculture IV, Berlin, Heidelberg. p124.