Multivariate Statistical Analysis-Based Identification of Fake Seeds for Rapid Argo Forensic Application

Yashrajsinh Zala[1], Jayrajsinh Sarvaiya[1*], Poorti Mohindroo[1], Harshad Patel[1], Arnobhab Chattopadhyay[1], Sheetal Buddhadev[2]

[1]Research Scholar, National Forensic Sciences University, Gandhinagar -Gujarat, India.*
[1]Associate Professor and Head of Center, Center FTF, National Forensic Sciences University, Gandhinagar -Gujarat, India.
[1]Research Scholar, National Forensic Sciences University, Gandhinagar -Gujarat, India.
[1]Associate Professor, Center FTF, National Forensic Sciences University, Gandhinagar -Gujarat, India.
[2]Professor, Faculty of Pharmacy, Noble University, Junagadh, Gujarat, India

Corresponding author:
Jayrajsinh Sarvaiya
[1]Associate Professor and Head of Center, Center FTF, National Forensic Sciences University, Gandhinagar -Gujarat, India.jayrajsinh.sarvaiya@gmail.com,

KEYWORDS

NIRS, chemometrics, seed viability, PCA, PLS-DA, LDA, rapid testing.
.

ABSTRACT

Utilising Near-Infrared Reflectance Spectroscopy (NIRS) with chemometric tools like Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA), to differentiate viable and non-viable castor seeds (GCH 7 hybrid). Traditional seed viability tests are accurate but time-consuming and resource-intensive. NIRS offers a rapid, non-destructive alternative for assessing seed viability. Spectral data from 200 viable and 200 non-viable seeds were collected and analysed using PCA and PLS-DA to develop a Linear Discriminant Analysis (LDA) model. The model achieved 99% accuracy in classifying seed viability, demonstrating its potential as a reliable tool for on-spot seed quality assessment. Key spectral markers related to fatty acids, proteins, and functional groups related to castor seed oil were identified. The research highlights the feasibility of integrating NIRS with advanced data analytics for rapid seed viability testing, offering significant benefits to seed testing agencies and farmers by reducing time and resource requirements while maintaining high accuracy.

1. Introduction

The International Seed Testing Association (ISTA) has defined each crop seed's standard Percentage Germination (PG) and shelf life. Once a lot's overall percentage germination falls below its standards, it is considered non-viable and discarded. Percentage germination reduction occurs due to moisture loss, changes in enzyme activities [1], degradation of their cellular structure [2], Mechanical damage from mishandling during transportation [3] and microbial attack [4]. The denaturation of the oil body inside the seed core mainly influences oil seeds. These fatty acid composition changes over time and affect the longevity of oil seeds [5] Incorporating older seeds in the supply chain frequently leads to germination failure and crop loss. Conventional methods for seed testing are time-consuming and limit vigorous surveillance of seed quality in the market.

Figure 01: Importance of Data Analytics-Based Rapid Agro Forensics against onventional Testing

Standardised preliminary seed germination tests such as electrical conductivity, tetrazolium, respiration, respiratory activity, GADA (Glutamic acid decarboxylase activity) test, and stress tests [6] are time-consuming and occasionally inaccurate. Modern methods of germination predictions, such as hyperspectral imaging [7]Potential options include X-ray imaging, oxygen consumption tests, ethanol tests, and volatile organic compound production tests. These techniques are combined with computational data analysis.

Nir Infrared Reflectance Spectroscopy (NIRS) has proven to differentiate the seed quality based on the amount and type of key nutrient content specific to a particular seed, like fatty acid derivatives like oleic acid and ricinoleic acid, which are significant factors affecting castor seed viability [8]. Various investigators report this technique to determine the viability of tomato, spinach, cabbage, and reddish seeds [9]. NIRS can potentially conduct on-site seed quality checks as part of surveillance or seed quality fraud reporting. Data processing tools are a prerequisite for a successful NIRS application in agricultural forensics. Predictability becomes more accurate with the bulk of data and suitable statistical models when market samples are to be screed quickly against pre-loaded standard NIRS data.

Multivariate analysis, such as Principal Component Analysis (PCA) and Partial Least Square Regression Analysis (PLSR), is gaining popularity for statistical applications on NIR spectra, as observed in various research (Table 01). Sample preparation technique, data quality, selection of target spectra window and statistical model validation are key factors determining the prediction accuracy of seed viability when PCA/PLSR is used [10]. Our main objective was to improve the predictability of the statistical model compared to the previously reported NIRS-based oil seed viability test. NIR spectra acquisition by hand-held devices and comparison with a previously created statistical model are suitable for dealing with agricultural fraud cases. The study on GCH 7 hybrid castor oil seeds in our study substantiates this agriculture forensic application.

Table 01: List of literature and key findings

Author(s)	Work Summary	Methodology	Key Findings
Andrade et al. 20 [10]	Study the impact of artificial ageing on the seed Vigor test.	FTIR + PCA and PLS-DA will classify corn seed vigour based on ageing time.	PLS-DA achieved >90% accuracy; identified lipid oxidation and protein degradation as key biomarkers.
Castillejos-Mijangos et al. 2021 [11]	Evaluate the quality control of coffee, cocoa, and spices through FT-MIR spectroscopy.	FT-NIR with PCA & PLS-DA for adulteration detection, compound quantification, and classification.	Detected adulterants, quantified key compounds (e.g., caffeine, curcumin), and classified samples by origin.
Fernández-Cuesta et al. 2014 [8]	Differentiating the oleic acid content into categories of castor seeds	NIRS + PCA and PLSR to differentiate high oleic acid seeds from conventional ones.	NIRS accurately predicted the oleic acid content (R^2 > 0.90) and identified key spectral features markers.
Zhang et al. 2021 [12]	Compared traditional seed testing methods with NIRS for spinach seeds.	Conducted SGT, AAT, and NIRS analysis on spinach seeds.	NIRS provided faster results (2-3 minutes) than SGT (7-14 days).
Gislum et al. 20 [13]	Castor seed quality determination with NIR	NIR + PCA analysis of castor seeds viability	Differentiate viable and non-viable castor seeds with first derivatives results of raw spectra.

2. Material and Methods

2.1 GC-MS analysis

Sample preparation for GC-MS was done by scrapping 10mg of seed core material from 10 viable and non-viable seeds each to make up a 100mg sample for the respective variable. 100ml sample was mixed in 2ml methanol by sonicating for 30min. and then centrifuged at 6000 RPM for 5min. 1ml supernatant was taken from both sample vials and taken to GC-MS analysis

GC-MS analysis was done with a Perkin Elmer (USA) Clarus 680 /SQ8C instrument equipped with an Elite-5ms Capillary Column (30 m x 0.25 mm) and customised parameters (Table 02).

Table 02: GC-MS instrument parameters

Acquisition parameter	Value
Injection volume	1.0 µL
Injection temperature	250°C
Split ratio	10:1
GC oven gradient	Initial temp 60 °C at 0 min, increased to 280°C at the flow rate of 10°C/min for 5 min
Interface temp	250°C
MS source temp	250°C
Solvent delay	3 min
Scan range	50 o 600 Daltons

2.2 Sample collection

Viable and non-viable castor seeds (Gujarat Castor Hybrid 7 or GCH 7) were received from government seed testing laboratories Dantiwada (Gujarat, India) and the Gujarat State Seed Certification Agency (GSSCA) upon request. The report received with each sample mentioned percentage germination, which was later confirmed with the traditional method. Samples were further scrutinised to select the most suitable samples to prepare the PLSR model (Table 03).

Table 03: Selection criteria of seed samples for the experiment

Sr. No.	Seeds type	Percentage Germination
1	Non-viable castor seeds (GCH 7)	<30%
2	Viable castor seeds (GCH 7)	>90%

2.2 Sample preparation

Castor seeds have a natural red coat/shell around them. The shell was removed to expose the inner soft and oily parts. Seeds were diced in 2 parts from its natural split area. A 1 mm cross-section of either side of the seeds was made using a sharp blade. The freshly prepared cross-section was placed in the centre part of the Fourier Transformed Near-infrared (FT-NIR) analyser's solid sample holder. The target area of this NIRS instrument is 5 mm x 7 mm. Seed samples should be made sure to cover the entire target area. Leaving blank spaces in the target gives noise, affecting the overall spectra acquisition.

Figure 02: Step-by-step graphical representation of the sample preparation process.

2.3 Spectral collection of samples

UV-Visible/NIR Spectrophotometer V-770 JASCO (Japan) was used to take spectra with its built-in spectra analyser software (Spectra Manager). 200 viable and 200 non-viable seed samples were tested from 800 nm to 2000 nm range, at 2nm intervals with 32 cm^{-1} resolution. Each spectral data was exported into a CSV/Excel file with built-in software to analyse further in The Unscrambler X (Ver. 10.4.1) (USA).

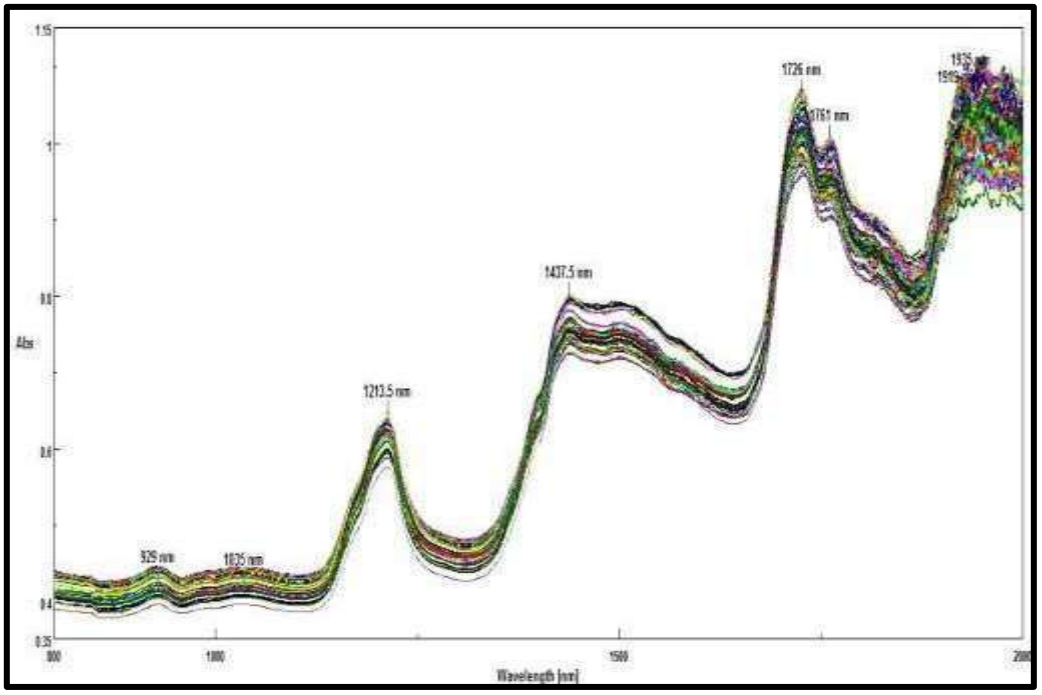

Figure 03: NIR Spectra of viable castor seed

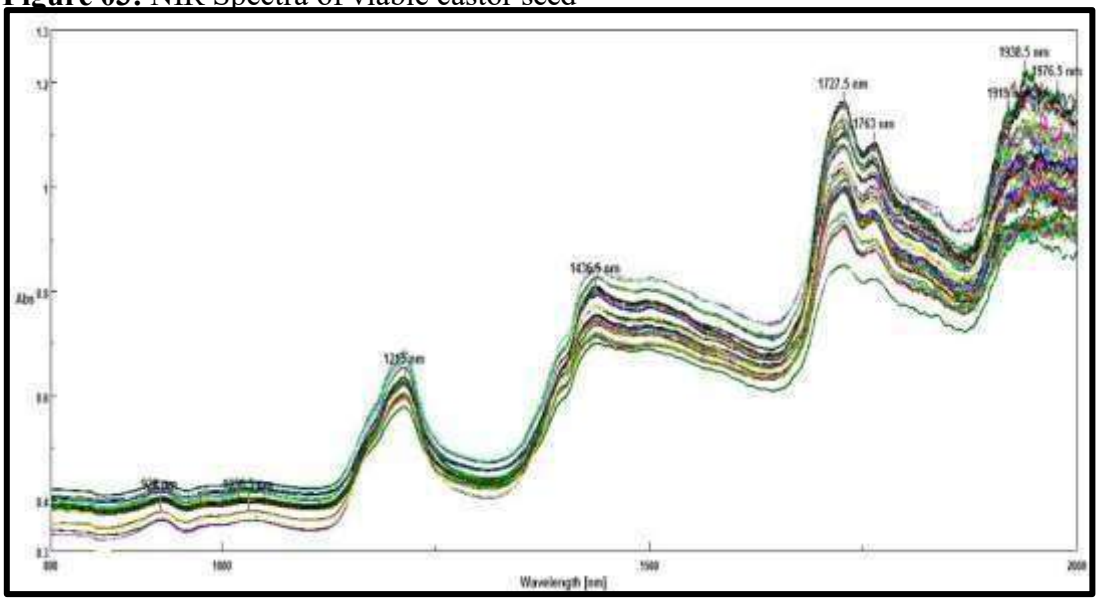

Figure 04: NIR Spectra of non-viable castor seeds

2.4 Data processing for multivariate classification model

Individual data files of all samples were transposed and combined in a master spreadsheet with variables mentioned in columns and spectra mentioned in rows listed, which are data matrix criteria of the Unscrambler X software. The master sheet was imported into analysing software, and variable groups were created for easy analysis.

PCA is the basic workhorse of multivariate data analysis. It assimilates information to variations. When a measured variable exhibits significant systematic variation, it becomes an information attribute [14]. PCA brings that information forward so that humans can generally interpret it in graphical form. Here, a PCA score plot is used to identify potential outliers from the significant groupings. From the graphical correlation representation of PC1 and PC2, potential outliers were observed and removed from each sample group for an improved PCA model. PCA plots are adjacent to every sample score in the matrix. Modified PCA score plots fully classified viable and non-viable groups with required validation.

PLS regression analysis maximises the covariance between individual X and Y values (here, samples and spectral wavelength) of a data matrix that minimises the residual errors[15]. The PLS regression model was prepared from the new data matrix that contained no outliers. The regression model developed in PLS was checked for its predictability by Root Mean Square Error (RMSE) and R2 value for the pulled data.

A line plot of loading values was prepared and analysed to determine which factors influenced the score plot most [16]. Loading weights are specific to PLSR and express how the information in each X variable relates to the variation in Y. Variables with prominent loading weight values are important for the prediction of Y [17]The spectral wavelength that had the most influence on the score plot was identified, and only those wavelength values were carried forward in a new data matrix.

The new data matrix is free from outliers and undesirable wavelengths, and its variables are classified into two categories (viable and non-viable) as a training set for creating the LDA. The LDA is a classification method based on Bayes' formula and is the simplest one [18]. It develops a classification model based on the standard distribution assumption and the assumption that the covariance matrix of all mentioned groups is identical. This means that the variability within each group has the same structure [19]. The only difference between the classification groups is that they have different centres because of their within-group and between-group variance considerations. If a suitable model is made, this factor makes the LDA model an appropriate method to identify adulteration, differences, or changes in chemical properties for the same sample type.

The LDA model was tested by running unknown samples. Ten seed samples from each variable were freshly prepared from a new lot. Their spectra were taken with parameters proportional to those used for the LDA model. Those 20 samples were used to check the model's prediction accuracy as a training set.

3 Result & Discussion

3.1 GC-MS screening for seed constituents

It is reported that multiple biochemical factors in seed affect its longevity [20]. GC-MS is a sophisticated instrument that conducts a comprehensive analysis of the seed's biochemical profile [21]. It can detect specific metabolites that play an essential role in seed deterioration. Our non-targeted GC-MS screening identified the presence of a high amount of tocopherol in viable seeds and its absence in non-viable seeds lesser presence in non-viable seeds during multiple sample runs (Fig 05). This indicates that viable seeds may have a higher oil content without oxidative degradation than non-visible seeds. α-tocopherol is known to prevent oxidative degradation by protecting unsaturated fatty acids during the storage period[22]Andrade et al. have studied and reported an increased content of phosphorus compounds in viable seeds compared to nonviable seeds through FTIR spectroscopy. However, during our seed element profiling by EDXRF (Energy Dispersive X-ray Fluorescence), castor seed samples did not show a significant difference between viable seeds and nonviable seeds as far as phosphorus content is concerned.

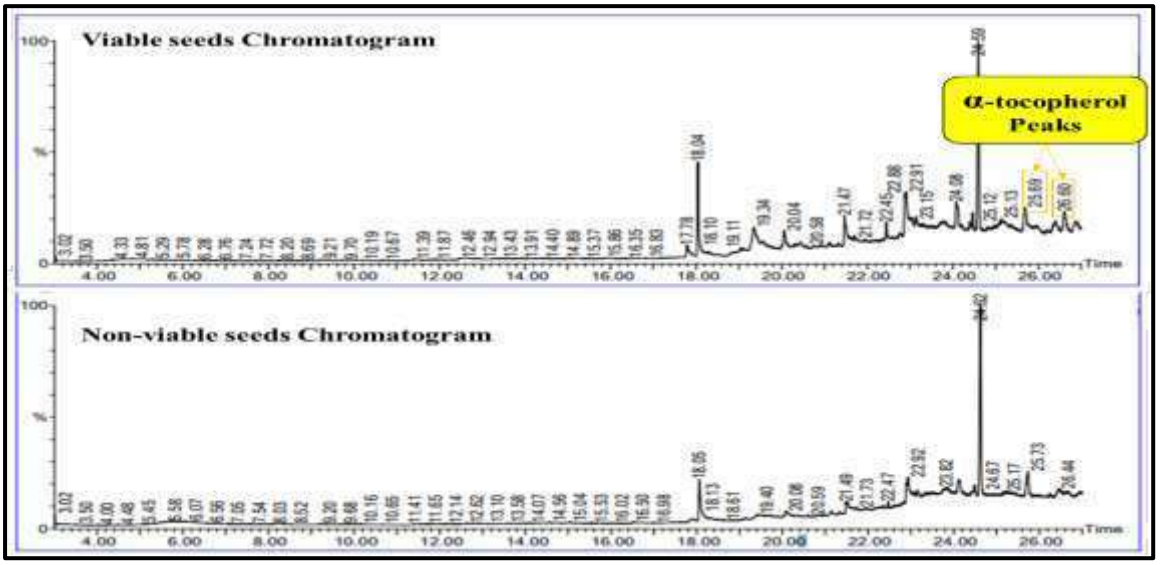

Figure 05: Comparison of viable and non-viable GC-MS chromatogram

3.2 Classification of NIRS result and data processing

PCR plot was formed directly with raw spectra without any smoothing, noise elimination, amplification or conversion of data. Previous studies determined the PCA of viable and non-viable castor seeds after transforming raw NIR spectra into first derivative spectra[13]. We could get similar results with raw NIR spectra using the method used here. Visually distinguishable separate variable groups are observed with the bi-dimensional plot of PC1 and PC2, with PC1 showing 92% variance (Fig. 6A). 10 outliers were identified from each variable and removed for improvement in the plot[23], [24], [25]. The result was improved with PC1 showing 93% variance and a better visual separation of the group observed (Fig. 6B). The software's built-in feature cross-validated samples for PCA, randomly selecting 30 samples and forming their clusters.

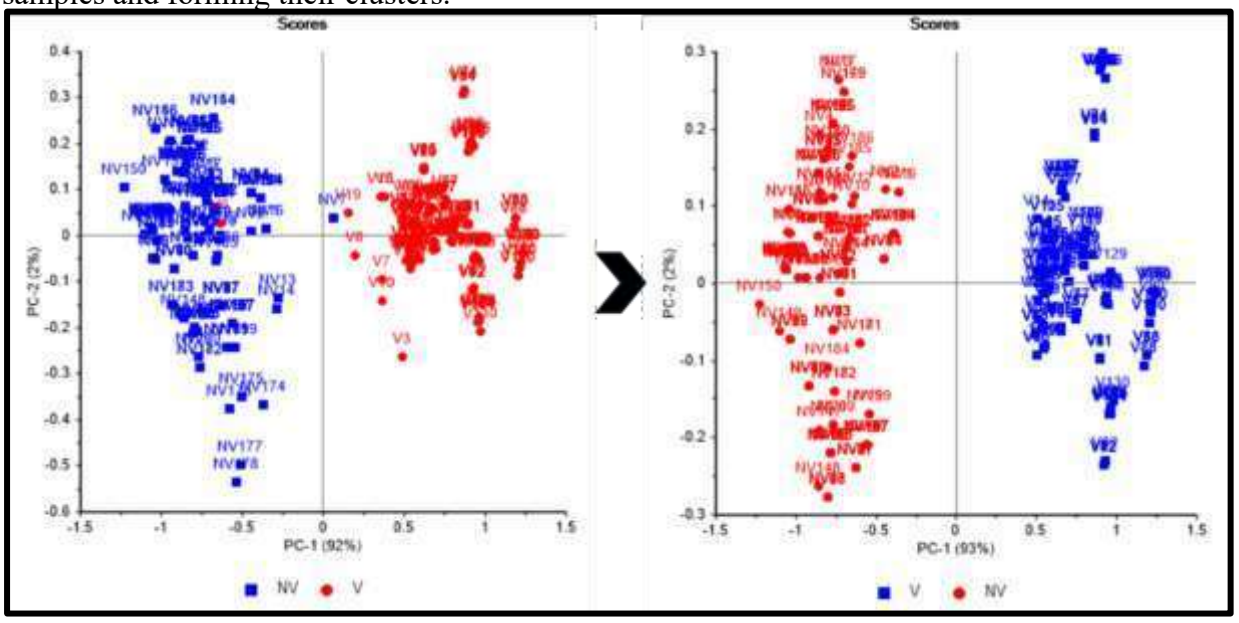

Figure 6A: PCA score plots before removal of outliers

Figure 6B: PCA score plots after the removal of outliers.

PLSR plot showed RMSE value close to zero (Fig.07, Table 04), indicating good data fitness[26], [27]. PLS-DA uses characteristic variables to develop prediction models, while PCA is mainly related to the creation of numeric variable-based models[28].

Table 04: LDA values

Sample	R^2 Value	RMSE Value
Viable seeds samples	0.980	0.008
Non-viable seed samples	0.987	0.008

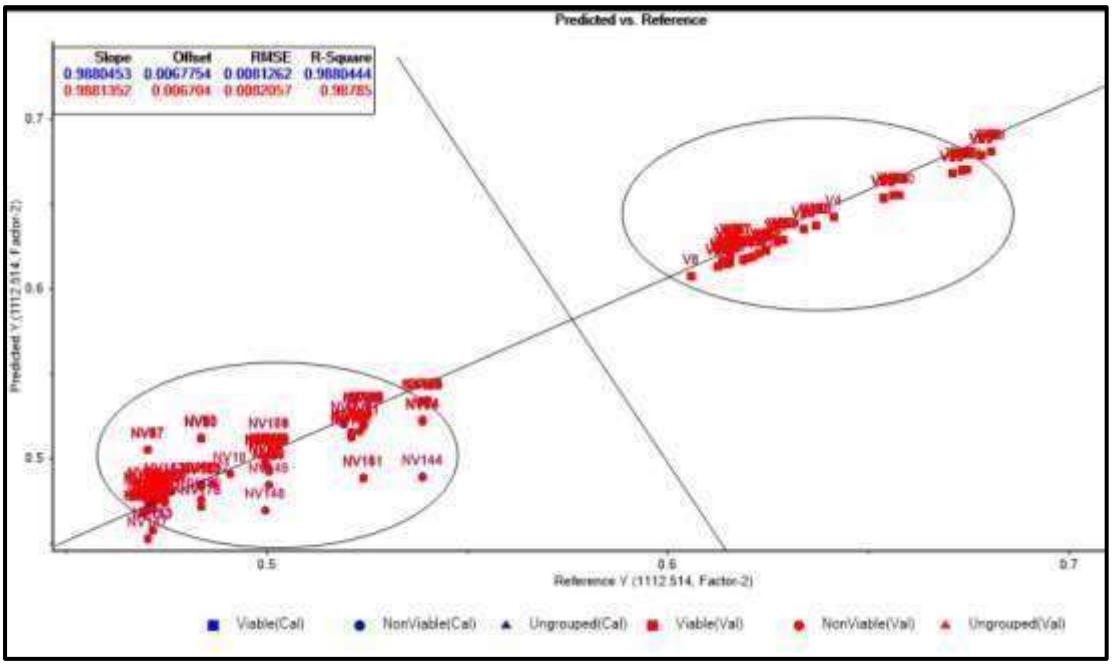

Figure 07: Difference of viable & and non-viable seeds in PLS with RMSE value

Loading weight scores of the PLSR model showed that 12 wavelengths had a prominent impact on score plots. With this technique, it is also easier to find signature peaks which differentiate the values of any variance in a dataset[29]. These peak selection criteria are critical for making an accurate and reliable prediction-yielding PLS-DA model[30]Removing low-impact wavelengths from the matrix before proceeding to LDA model preparation will improve the model's overall accuracy.

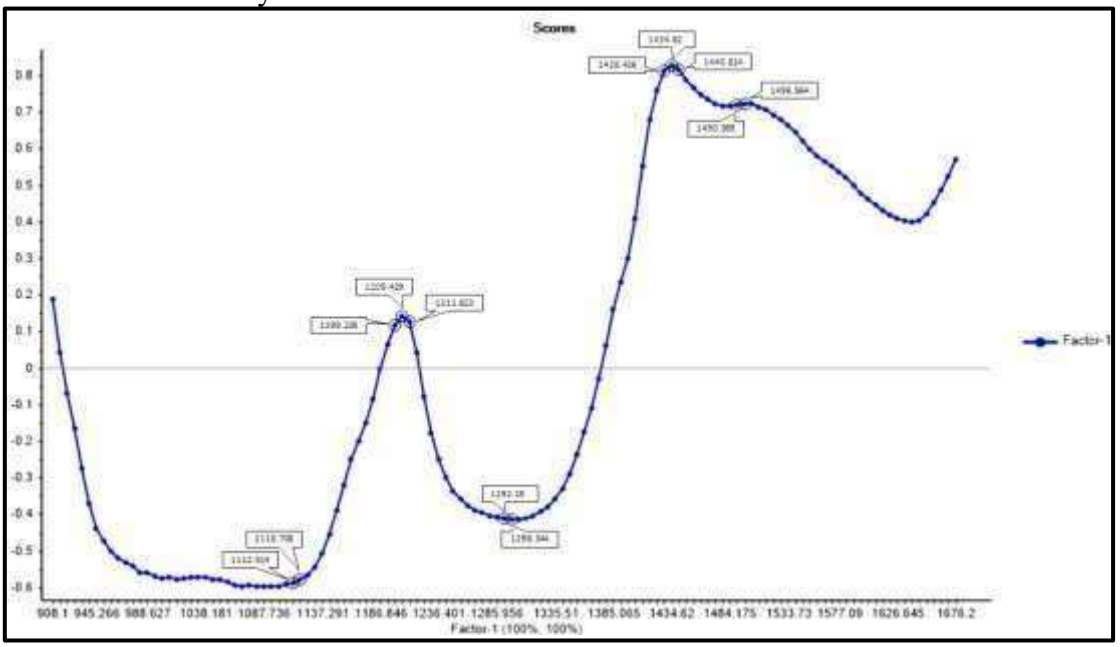

Figure 08: Selection of characteristic peaks for the LDA model.

Table 05: Identified NIRS peaks that were used to create the PLS-DA model

Frequency	Observation	Possible Region of Detection	Reference
1112	Vibration of C-O Ester group	Functional group of Protein	[13]
1118			
1119			
1205	C-H stretching	Functional groups of fatty acids	[13]
1211			
1292	NH stretch and C-H stretching	Functional group of starch and protein	[31]
1295			
1428	O-H overtones	Functional groups of fatty acids (ricinoleic acid)	[32]
1434	C-H stretching	Functional groups of fatty acids	[13]
1440			
1490			
1496			

These 12 selective wavelength values were kept for further analysis, and the rest were not considered for LDA model preparation. Selected wavelengths show essential functional groups related to castor seed contents determined by traditional spectra interpretation (Table 04). Although the NIR spectra range primarily selected target model creation shows fatty acid function groups, they contribute to oil content identification only as per functional group and frequencies relevant to the fatty acid and protein functional group. Moreover, seed germination is also affected by gibberellin and auxin acid growth hormones [33]. The Proteomic of oil seed by various investigators also highlights the presence of fatty acid molecules in viable seed compared to non-viable.

Raw spectral data is now transformed into an accurate data set by outliner elimination at the PCA step and unwanted wavelength elimination at the PLS step. Fine-tuning of raw data gives better accuracy of the LDA model. Such LDA models are considered more reliable because they only have values and information of core properties that differentiate all classes in the model [34]The model prepared here has an accuracy of 99.47%, as observed in its confusion matrix table (Fig. 09). This wavelength-filtered approach to statistical analysis has given better model prediction than a wider wavelength range, as reported in previous studies.

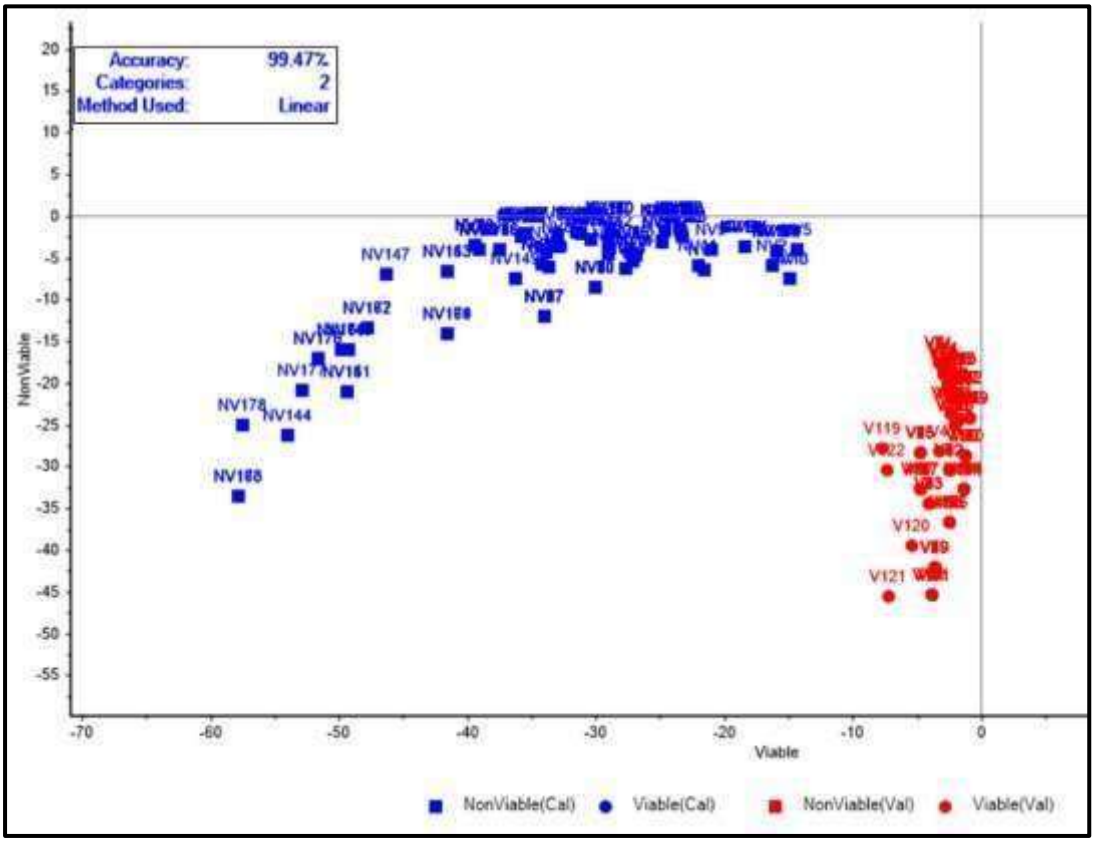

Figure 09: LDA graphical model and its accuracy

Confusion...	Actual	Viable	NonViable
Predicted		1	2
Viable	1	189	1
NonViable	2	1	189

Figure 10: Confusion matrix of LDA plot

Random 10 seed samples (10 viable and 10 non-viable) were taken from another lot of the sample with prior conformation to the Vigour test and used for cross-validation. The 190-190 sample data were used as a training set to classify the LDA model. Appropriate coding was given to those samples (UnV- unknown viable, UnNV- unknown non-viable). This is to validate the prediction of the LDA model in the future because their actual values and class were unknown to the software. The LDA model could identify samples with 100% accuracy (Fig 11), showing the model's fitness for the study's objective. The prediction is more accurate in the model created in this work than in previous reports. Target spectra points differed in both studies, showing the importance of suitable target selection through score/loading score for accurate model creation.

Classified_L...		Viable	NonViable	Class
		1	2	3
UnV1	1	-1.3659	-21.5762	Viable
UnV2	2	-1.7622	-19.2066	Viable
UnV3	3	-1.3752	-32.4164	Viable
UnV4	4	-3.2953	-28.0046	Viable
UnV5	5	-2.4068	-36.2712	Viable
UnV6	6	-2.3774	-23.4450	Viable
UnV7	7	-4.7247	-32.7089	Viable
UnV8	8	-2.4719	-20.6884	Viable
UnV9	9	-1.9548	-24.8747	Viable
UnV10	10	-0.9160	-23.9239	Viable
UnNV1	11	-14.8898	-7.5722	NonViable
UnNV2	12	-16.2094	-5.8815	NonViable
UnNV3	13	-26.8615	-5.4625	NonViable
UnNV4	14	-28.7872	-4.5984	NonViable
UnNV5	15	-14.3299	-4.0843	NonViable
UnNV6	16	-21.6746	-6.5438	NonViable
UnNV7	17	-33.8666	-4.5448	NonViable
UnNV8	18	-32.9330	-3.8404	NonViable
UnNV9	19	-20.8833	-4.0312	NonViable
UnNV10	20	-24.7688	-3.0706	NonViable

Figure 11: Prediction result table showing accurately predicted class

4. Conclusion

Using data analytics tools, spectral-based differentiation among seed types and seed quality can be created, and it can certainly serve as a rapid testing tool once the standard data sets are established. The present study focused on percentage germination identification by oil content in castor seeds, using NIRS to achieve data and develop a model through PLS/PCA. The viability of any GCH 7 castor seed can be quickly determined by taking its NIR spectra at a selective range and predicting it with this LDA model.

A well-prepared LDA model integrated into a remote NIR device as an Artificial Neural Network can work as an on-spot seed viability detection tool. It can benefit all seed testing agencies and farmers by providing fast and accurate results using fewer resources.

5. Conflicts of Interest

No conflict of interest to declare.

6. Acknowledgements

Authors acknowledge Special Research Funding from National Forensic Sciences University to carry out this research.

7. References

[1] J. Yasur and P. U. Rani, “Environmental effects of nanosilver: Impact on castor seed germination, seedling growth, and plant physiology,” Environmental Science and Pollution Research, vol. 20, no. 12, pp. 8636–8648, Dec. 2013, doi: 10.1007/s11356-013-1798-3.

[2] R. M. O. Pires, M. A. B. Àvila, D. G. Leite, H. O. Santos, G. A. Souza, and E. V. R. Von Pinho, “Physiological and enzymatic alterations in sesame seeds submitted to different osmotic potentials,” Genetics and Molecular Research, vol. 16, no. 3, Aug. 2017, doi: 10.4238/gmr16039425.

[3] B. B. Santoso, I. A. Parwata, and I. K. D. Jaya, “Seed Viability and Oil Content of Castor Bean (Ricinus communis L.) as Affected by Packaging Materials During

Storage," International Journal of Applied Sciences and Technology, vol. 5, pp. 56–61, 2015.

[4] J. W. Dalling, A. S. Davis, B. J. Schutte, and A. Elizabeth Arnold, "Seed survival in soil: Interacting effects of predation, dormancy and the soil microbial community," Journal of Ecology, vol. 99, no. 1, pp. 89–95, Jan. 2011, doi: 10.1111/j.1365-2745.2010.01739.x.

[5] S. Balesevic-Tubic, M. Tatic, V. Djordjevic, Z. Nikolic, and V. Djukic, "Seed viability of oil crops depending on storage conditions," Helia, vol. 33, no. 52, pp. 153–159, 2010, doi: 10.2298/HEL1052153B.

[6] S. Basu and S. P. C. Groot, "Seed Vigour and Invigoration," in Seed Science and Technology: Biology, Production, Quality, M. Dadlani and D. K. Yadava, Eds., Singapore: Springer Nature Singapore, 2023, pp. 67–89. doi: 10.1007/978-981-19-5888-5_4.

[7] L. Feng, S. Zhu, F. Liu, Y. He, Y. Bao, and C. Zhang, "Hyperspectral imaging for seed quality and safety inspection: a review," Plant Methods, vol. 15, no. 1, p. 91, 2019, doi: 10.1186/s13007-019-0476-y.

[8] Á. Fernández-Cuesta, J. M. Fernández-Martínez, and L. Velasco, "Identification of high oleic castor seeds by near infrared reflectance spectroscopy," JAOCS, Journal of the American Oil Chemists' Society, vol. 89, no. 3, pp. 431–435, Mar. 2012, doi: 10.1007/s11746-011-1933-6.

[9] M. H. Olesen, N. Shetty, R. Gislum, and B. Boelt, "Classification of Viable and Non-Viable Spinach (Spinacia oleracea L.) Seeds by Single Seed Near Infrared Spectroscopy and Extended Canonical Variates Analysis," J Near Infrared Spectrosc, vol. 19, pp. 171–180, 2011, doi: 10.1255/jnirs.923.

[10] G. C. Andrade, C. M. M. Coelho, and V. G. Uarrota, "Modelling the vigour of maize seeds submitted to artificial accelerated ageing based on ATR-FTIR data and chemometric tools (PCA, HCA and PLS-DA)," Heliyon, vol. 6, no. 2, 2020, doi: 10.1016/j.heliyon.2020.e03229.

[11] L. A. Castillejos-Mijangos, A. Acosta-Caudillo, T. Gallardo-Velázquez, G. Osorio-Revilla, and C. Jiménez-Martínez, "Uses of FT-MIR Spectroscopy and Multivariate Analysis in Quality Control of Coffee, Cocoa, and Commercially Important Spices," Feb. 01, 2022, MDPI. doi: 10.3390/foods11040579.

[12] Y. Cui, W. Ge, J. Li, J. Zhang, D. An, and Y. Wei, "Screening of maize haploid kernels based on near infrared spectroscopy quantitative analysis," Comput Electron Agric, vol. 158, pp. 358–368, Mar. 2019, doi: 10.1016/j.compag.2019.01.038.

[13] R. Gislum, P. Nikneshan, S. Shrestha, A. Tadayyon, L. C. Deleuran, and B. Boelt, "Characterisation of castor (Ricinus communis L.) seed quality using fourier transform near-infrared spectroscopy in combination with multivariate data analysis," Agriculture (Switzerland), vol. 8, no. 4, Apr. 2018, doi: 10.3390/agriculture8040059.

[14] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2065, p. 20150202, 2016.

[15] B.-H. Mevik, "Introduction to the pls Package," 2021.

[16] I.-G. Chong and C.-H. Jun, "Performance of Some Variable Selection Methods When Multicollinearity Is Present," Chemometrics and Intelligent Laboratory Systems, vol. 78, pp. 103–112, 2005, doi: 10.1016/j.chemolab.2005.04.010.

[17] J. B. Wooten, N. E. Kalengamaliro, and D. E. Axelson, "Characterization of bright tobaccos by multivariate analysis of 13C CPMAS NMR spectra," Phytochemistry, vol. 70, no. 7, pp. 940–951, May 2009, doi: 10.1016/j.phytochem.2009.04.015.

[18] G. Li, D. Liang, Q. Huang, S. Jiang, and W. Gao, "Object tracking using incremental 2D-LDA learning and Bayes inference," in 2008 15th IEEE International Conference on Image Processing, IEEE, 2008, pp. 1568–1571.
[19] J. Laimer et al., "Amalgam tattoo versus melanocytic neoplasm - Differential diagnosis of dark pigmented oral mucosa lesions using infrared spectroscopy," PLoS One, vol. 13, no. 11, p. e0207026, Nov. 2018, doi: 10.1371/journal.pone.0207026.
[20] A. YEBOAH et al., "Castor oil (Ricinus communis): a review on the chemical composition and physicochemical properties," Food Science and Technology, vol. 41, no. suppl 2, pp. 399–413, 2021, doi: 10.1590/fst.19620.
[21] S. R. Selvarani, S. Sundareswaran, V. Manonmani, N. Manivannan, V. Gomathi, and K. Raja, "Seed Odour as Oracle: Advanced Analysis of Soybean Seed Volatile Organic Compounds and Statistical Insights," Agricultural Research, 2024, doi: 10.1007/s40003-024-00835-2.
[22] S. E. Sattler, L. U. Gilliland, M. Magallanes-Lundback, M. Pollard, and D. DellaPenna, "Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination," Plant Cell, vol. 16, no. 6, pp. 1419–1432, 2004.
[23] S. F. Møller, J. von Frese, and R. Bro, "Robust methods for multivariate data analysis," J Chemom, vol. 19, no. 10, pp. 549–563, 2005, doi: 10.1002/cem.948.
[24] G. Ivosev, L. Burton, and R. Bonner, "Dimensionality reduction and visualization in principal component analysis," Anal Chem, vol. 80, no. 13, pp. 4933–4944, 2008, doi: 10.1021/ac800226r.
[25] D. Groth, S. Hartmann, S. Klie, and J. Selbig, "Principal components analysis," 2013, Elsevier. doi: 10.1016/B978-0-12-386455-2.00022-2.
[26] A. Gupta and R. Katarya, "PAN-LDA: A latent Dirichlet allocation based novel feature extraction model for COVID-19 data using machine learning," Comput Biol Med, vol. 138, Nov. 2021, doi: 10.1016/j.compbiomed.2021.104920.
[27] S. Qiu, J. Wang, and L. Gao, "Discrimination and characterization of strawberry juice based on electronic nose and tongue: Comparison of different juice processing approaches by LDA, PLSR, RF, and SVM," J Agric Food Chem, vol. 62, no. 27, pp. 6426–6434, Jul. 2014, doi: 10.1021/jf501468b.
[28] M. Lasalvia, V. Capozzi, and G. Perna, "A Comparison of PCA-LDA and PLS-DA Techniques for Classification of Vibrational Spectra," Applied Sciences (Switzerland), vol. 12, no. 11, Jun. 2022, doi: 10.3390/app12115345.
[29] J. J. Workman and L. Weyer, Practical Guide to Interpretive Near-Infrared Spectroscopy. Boca Raton, FL, USA: CRC Press, Taylor and Francis Group, 2007.
[30] D. Ballabio and V. Consonni, "Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA," Analytical Methods, vol. 5, pp. 3790–3798, 2013, doi: 10.1039/C3AY40582F.
[31] M. Al-Amery et al., "Near-infrared spectroscopy used to predict soybean seed germination and vigour," Seed Sci Res, vol. 28, no. 3, pp. 245–252, Sep. 2018, doi: 10.1017/S0960258518000119.
[32] D. Kusumaningrum, H. Lee, S. Lohumi, C. Mo, M. S. Kim, and B. K. Cho, "Non-destructive technique for determining the viability of soybean (Glycine max) seeds using FT-NIR spectroscopy," J Sci Food Agric, vol. 98, no. 5, pp. 1734–1742, Mar. 2018, doi: 10.1002/jsfa.8646.
[33] A. Rastogi et al., "Effect of auxin and gibberellic acid on growth and yield components of linseed (Linum usitatissimum L.)," 2013.
[34] K. Torkkola, "Linear discriminant analysis in document classification," in IEEE ICDM workshop on text mining, 2001.